Student Theses and Dissertations

2016

Tracking the In Vivo Dynamics of Antigenic Variation in the African Trypanosome

Monica R. Mugnier

Follow this and additional works at: http://digitalcommons.rockefeller.edu/
student_theses_and_dissertations

Part of the Life Sciences Commons

TRACKING THE IN VIVO DYNAMICS OF ANTIGENIC

VARIATION IN THE AFRICAN TRYPANOSOME

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Monica R. Mugnier

June 2016

TRACKING THE IN VIVO DYNAMICS OF ANTIGENIC

VARIATION IN THE AFRICAN TRYPANOSOME

Monica Mugnier, Ph.D.

The Rockefeller University 2016

Trypanosoma brucei, a causative agent of African sleeping sickness in humans

and nagana in animals, constantly changes its dense variant surface glycoprotein (VSG)

coat to avoid elimination by the immune system of its mammalian host, using an

extensive repertoire of dedicated genes. Although this process, referred to as antigenic

variation, is the major mechanism of pathogenesis for *T. brucei*, the dynamics of *VSG*

expression in *T. brucei* during an infection are poorly understood.

In this thesis, I describe the development of VSG-seq, a method for quantitatively

examining the diversity of expressed *VSGs* in any population of trypanosomes. Using

VSG-seq, I monitored *VSG* expression dynamics *in vivo* during both acute and chronic

mouse infections. My experiments revealed unexpected diversity within parasite

populations, and the expression of as much as one-third of the functional genomic *VSG*

repertoire after only one month of infection.

In addition to suggesting that the host-pathogen interaction in *T. brucei* infection

is substantially more dynamic and nuanced than previously expected, this observed

diversity highlighted the importance of the mechanisms by which *T. brucei* diversifies its

genome-encoded *VSG* repertoire. During infection, the parasite can form mosaic *VSGs*,

novel variants that arise through recombination events within the parasite genome during

infection. Though these novel variants had been identified previously, little was known

about the mechanisms by which they form. VSG-seq facilitated the identification of mosaic *VSGs* during the infection, which allowed me to track their formation over time. My results provide the first temporal data on the formation of these variants and suggest that mosaic *VSGs* likely form at sites of *VSG* transcription.

VSG-seq, which is based on the *de novo* assembly of *VSGs*, obviates the requirement for a reference genome for the analysis of expressed *VSG* populations. This allows the method to be used for the high-resolution study of *VSG* expression in any strain of *T. brucei,* whether in the lab or in the field. To this end, I have applied VSG-seq to samples grown *in vitro*, parasites isolated from natural infections, and extravascular parasites occupying various tissues *in vivo*. These extensions of the method reveal new aspects of *T. brucei* biology and demonstrate the potential of high-throughput approaches for studying antigenic variation, both in trypanosomes and in any pathogen that uses antigenic variation as a means of immune evasion.

For my Dad.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. Introduction

1.1 *Trypanosoma brucei*

African sleeping sickness, or human African trypanosomiasis (HAT), is a vector-borne disease affecting sub-Saharan Africa. It is fatal if left untreated, most existing drugs are extremely toxic, and the economic burden of sleeping sickness, along with its zoonotic counterpart, *nagana*, is estimated to be at least 1.5 billion USD per year[1]. The economic burden of this disease is due largely to the animal infection, which commonly affects livestock and can be devastating to endemic areas. In fact, there are probably only approximately 20,000 cases of HAT per year[2]. However, underreporting is a problem in the poor rural areas typically affected by HAT, so estimating its prevalence can be difficult. Nevertheless, the diseases caused by African trypanosomes represent a major barrier to economic development in affected regions.

The causative agents of sleeping sickness and nagana are a group of protozoan parasites known as the African trypanosomes, of which there are two human-infective subspecies: *Trypanosoma brucei gambiense*, which is found in west and central Africa and causes a chronic infection lasting months to years, and *T. brucei rhodesiense*, which is found in east Africa and causes an acute form of the disease lasting weeks to months. Humans are resistant to *T. brucei brucei* as well as a number of other species of *Trypanosoma,* meaning that only animals are infected by these parasites. *T. brucei* is limited geographically to the range of its vector, the tsetse (*Glossina* sp.), but two African trypanosome species, *T. evansi* and *T. equiperdum*, which are actually probably subspecies of *T. brucei*[3], have evolved new modes of transmission, expanding their range

beyond sub-Saharan Africa into Asia and South America. Thus, the range of the African trypanosomes and the diseases they cause is large and has the potential to expand. Of all of the African trypanosomes, however, *T. brucei* is the most commonly studied in the laboratory and will be the focus of this thesis.

T. brucei's entirely extracellular life cycle (Figure 1.1) begins when its vector, the tsetse, takes a blood meal from a mammal, transferring metacyclic stage parasites to a new mammalian host. These metacyclic forms move from the skin to the bloodstream where they transform into long slender bloodstream form parasites. The long slender forms divide quickly. As their density increases, they undergo quorum sensing and some of the long slender population transforms into non-dividing short stumpy bloodstream form parasites, which are primed for transmission back to the fly[4,5].

When a fly bites an infected mammal, these short stumpy trypanosomes are taken up and move to the midgut of the fly, where they transform into procyclic form parasites. The procyclic forms then move to the fly's salivary glands, where they further develop into epimastigote form parasites which can undergo meiosis, allowing for genetic exchange between parasites[6,7]. The epimastigotes divide for some time in the salivary gland before finally becoming metacyclics once again. These parasites can then be transmitted to a new host at the fly's next blood meal.

During a mammalian infection, long slender and short stumpy trypanosomes reside both in the blood and in tissues[8] but appear to always live extracellularly. In the early stages of infection, patients experience non-specific symptoms such as fever and headaches. Eventually, parasites cross the blood-brain barrier and enter the central nervous system, though the timing of this transit can vary[9,10]. This infection of the brain

and cerebrospinal fluid defines late-stage infection, and results in severe neuropathology, including the disturbance of the circadian rhythm that gives sleeping sickness its name[11]. The drugs required to treat this stage of infection are difficult to administer and frequently toxic[12]. Although prevention of African trypanosomiasis would be ideal, there is little hope for a vaccine against *T. brucei*[13,14]. One major reason why a vaccine proves unlikely is *T. brucei's* capacity for antigenic variation, the parasite's primary mechanism of immune evasion.
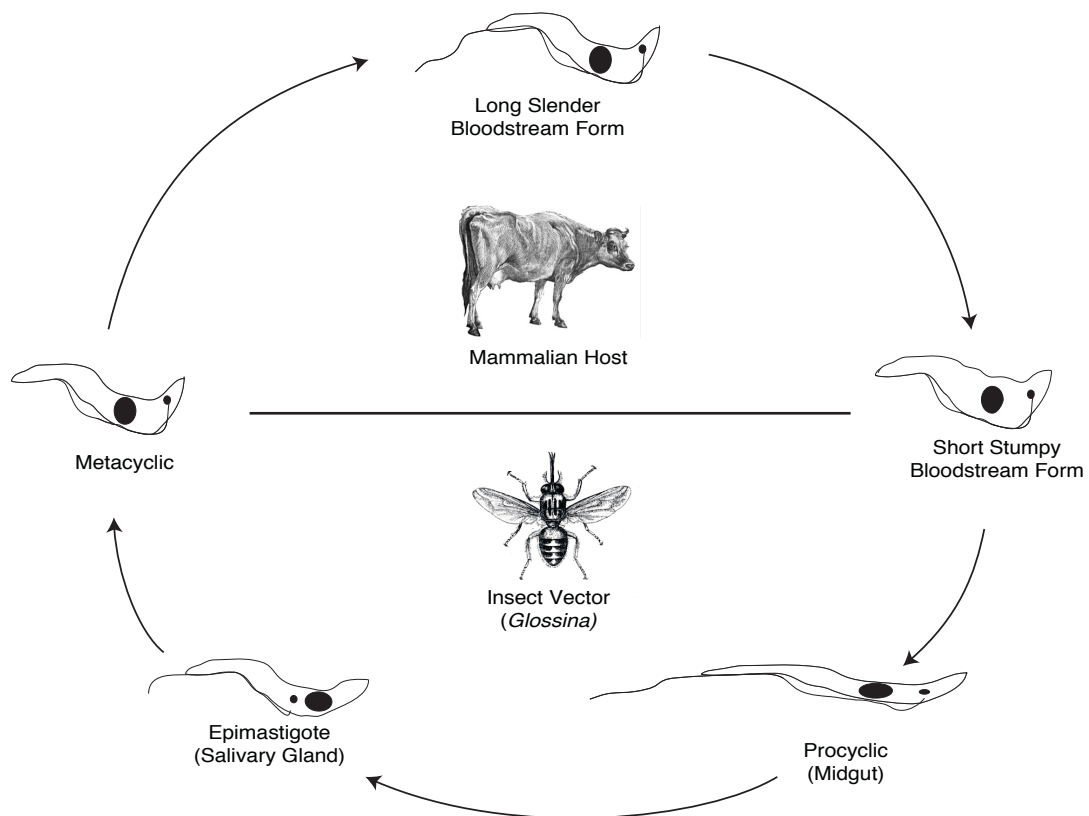
Figure 1.1. The life cycle of *T. brucei*.

1.2 The variant surface glycoprotein and antigenic variation

In 1909, Ross and Thomson applied a new method for counting parasites to the blood samples of a patient infected with *T. b. gambiense.* Their methodical counting revealed the periodic peaks and valleys in parasitemia characteristic of *T. brucei* infections[15](Figure 1.2), and they and others hypothesized that these waves resulted from some kind of habituation to host antibody[16]. Their guess was not too far off. Decades later, the antigenic component of *T. brucei* responsible for the waves of parasitemia during infection was identified as variant surface glycoprotein (VSG)[17], and these waves, we now understand, are the result of a sort of habituation: *T. brucei* periodically "switches" an antigenic VSG coat, allowing the parasite to escape recognition by host antibody.
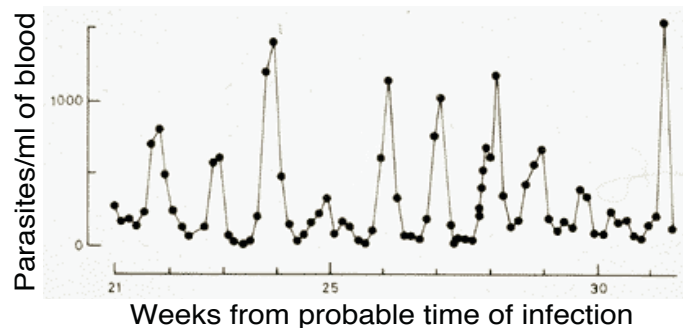
Figure 1.2. Waves of parasitemia during infection.
Image adapted from Ross and Thomson, 1910.

VSG, a GPI-anchored glycoprotein, is extremely densely packed on the cell surface of metacyclic and bloodstream form parasites[18], comprising ~95% of the surface proteome[19] with approximately 10^7 identical copies of the protein covering the parasite[17,20]. Due to its abundance, VSG is effectively the only protein "seen" by the host immune system. Though the host may mount a response to other trypanosome proteins,

4

no other protein can elicit a significant or protective immune response[21,22]. Therefore,

changing the VSG coat renders a parasite temporarily invisible to the host, prolonging

infection.

Each VSG is composed of a relatively conserved membrane-proximal C-terminal

domain and a more variable N-terminal domain (Figure 1.3). Only two VSG crystal

structures exist, and these structures only contain VSG N-termini[23,24]. Interestingly, the

structures are superimposable despite only 16% sequence identity[24]. Analysis of the

sequence of VSG-encoding genes has identified conserved cysteine residues that may

play a role in maintaining VSG structure while allowing for such significant sequence

divergence[25].
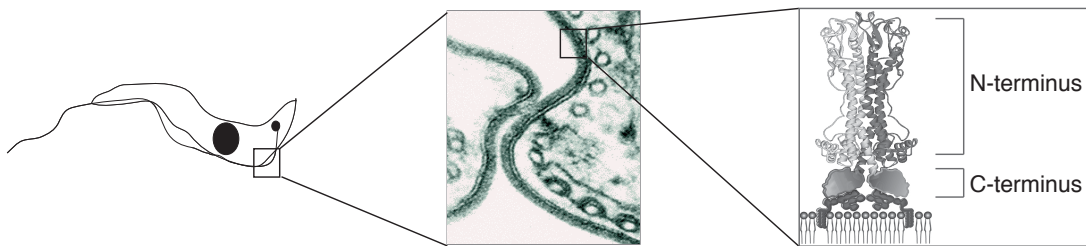
Figure 1.3. The VSG coat.
The Variant Surface Glycoprotein (VSG) densely coats the parasite membrane, as shown
by electron micrograph (courtesy of George A.M. Cross) displayed in. 10 million VSG
molecules pack on the surface of the parasite cell (image courtesy of Markus Engstler).

The parasite takes advantage of its set of structurally similar but antigenically

diverse VSGs during antigenic variation, the mechanism by which *T. brucei* changes its

VSG coat to avoid recognition by the host immune system. Throughout infection, the

parasite turns on and off expression of individual VSG-encoding genes, from a genomic

repertoire of ~2000 different genes[26,27]. As the host immune system recognizes and clears

a *VSG* expressed within a population of parasites, a minority of parasites will have

already changed their VSG coat, turning on a new, and likely antigenically distinct, *VSG*. These "switchers" will eventually be recognized by the immune system, but not before another set of parasites has turned on another *VSG* and escaped immune clearance.

The process of antigenic variation is common to many pathogens, including viruses[28], bacteria[29], and other protozoa[30] such as the malaria parasite *Plasmodium falciparum*[31]. *T. brucei*, then, can be thought of as a tractable model for an important process common to many organisms relevant to public health.

1.3 *VSG* expression

In bloodstream form parasites, *VSGs* are expressed from one of ~15 telomeric bloodstream expression sites (BESs)[32](Figure 1.4A). Only one BES is ever transcribed at a time, though the mechanisms controlling this monoallelic expression are not well understood. Each BES is made up of a promoter, a collection of expression site associated genes (ESAGs), a 70-bp repeat sequence, and a *VSG*. The number and identity of ESAGs can vary among sites, and it is not clear whether these genes play a role in antigenic variation; many are also expressed from other loci within the genome and during other life cycle stages[33,34]. The 70-bp repeat adjacent to the *VSG* can vary in length between BESs, and the sequence is speculated to play a role in antigenic variation by allowing *VSG* switching to diverse genomic donor sequences[35-38].

The entire BES, from just downstream of the promoter[39] through the *VSG,* is polycistronically transcribed by RNA Polymerase I (Pol I)[40]. This polymerase is exclusively responsible for transcription of ribosomal RNA in other organisms, and its role in *VSG* expression may be due to the high level of expression required to produce

sufficient VSG to cover *T. brucei's* surface. This long Pol I-transcribed unit is then processed into individual transcripts and, as is the case for all *T. brucei* mRNAs, a 5′ spliced-leader sequence is trans-spliced onto each gene, including the *VSG*[41](Figure 1.4B). Though *VSG* sequences are quite divergent from one another at the DNA level, they all contain a conserved 14 base pair (bp) sequence in their 3′ untranslated region (UTR)[42]. This sequence may play a role in maintaining monoallelic expression of *VSG,* although its precise role in the process is not well understood.
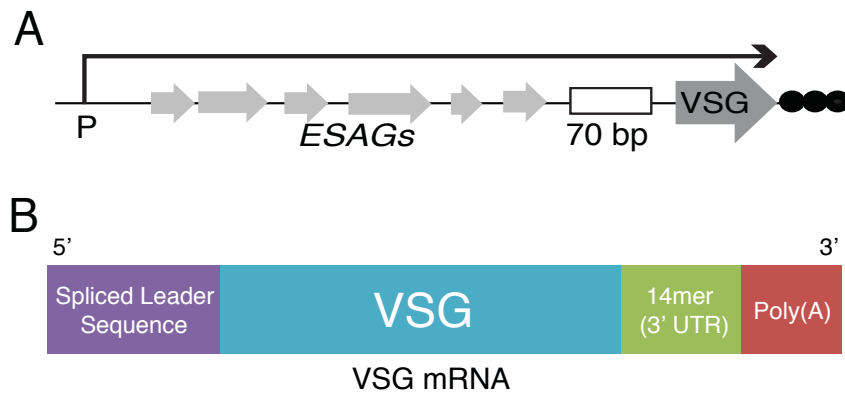
Figure 1.4. *VSG* expression.

(A) Schematic of a BES. (B) Schematic of a mature *VSG* mRNA.

Though not the focus of this thesis, metacyclic *VSGs* are expressed from a specialized expression site in metacyclic form parasites[43,44]. This site is also telomeric, but it lacks *ESAGS* and the 70-bp repeat sequence. Metacyclic expression sites have a promoter sequence distinct from that of bloodstream forms[26,45,46].

1.4 *VSG* switching

There are two broad mechanisms by which *T. brucei* switches *VSG* expression. The first mechanism is transcriptional, or *in situ*, switching, during which transcription at one BES is turned off and transcription at a second BES is turned on[47] (Figure 1.5A). It is not clear how an *in situ* switch is initiated. The second mechanism is recombination-based switching. This can occurs in two ways: duplicative gene conversion (GC) and telomere exchange (TE). Telomere exchange occurs through a homologous recombination event in which telomere ends are swapped between two chromosomes, resulting in a new *VSG* in the active BES, but retaining the previously active *VSG*[48] (Figure 1.5B). GC, on the other hand, occurs though a homologous recombination event in which a *VSG* located in a distant site is copied into the active BES. This results in the loss of the previously expressed *VSG* and duplication of the donor *VSG* sequence (Figure 1.5C). This type of switch can be initiated by a DNA double-strand break (DSB)[49,50] within the active BES, which is then repaired by gene conversion. A number of genes known to be involved in homologous recombination in eukaryotes have been implicated in this process[51,52]. While it is thought that recombinatorial switching results from naturally occurring DSBs, and that subtelomeric fragility may play a role in the induction of breaks in a BES[50,53,54], it is not clear whether this process is wholly stochastic.
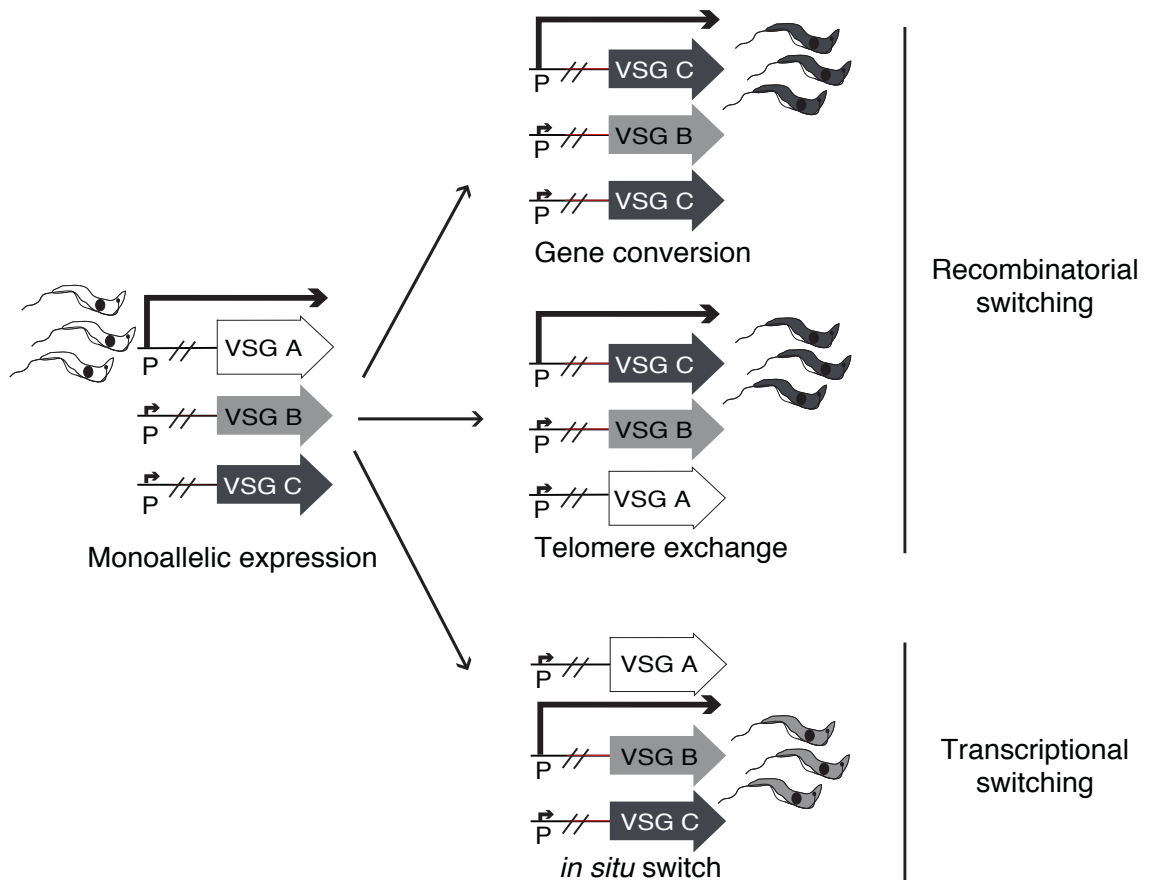
Figure 1.5. Mechanisms of *VSG* switching.

T. brucei can switch expression to an antigenically distinct *VSG* by recombinatorial (top, GC and TE) or transcriptional (bottom) switching.

1.5 The genomic *VSG* repertoire

GC based switching, in particular, is extremely important because it allows access to the extensive genomic repertoire of *VSGs*. The key to antigenic variation is the extent of this repertoire: recent work sequencing the "VSGnome" of the *T. brucei* Lister 427 strain revealed more than 2000 different *VSG*-encoding genes[26]. The telomeric BESs contain one set of *VSGs,* but these only represent a small fraction of the repertoire. The majority of *VSG* genes and pseudogenes exist elsewhere in the genome (Figure 1.6). The

rest of the *VSG* archive can be found within intra-chromosomal arrays[27,55] or in the minichromosomes (MCs). MCs are small 30-150kb chromosomes consisting only of *VSG* sequences, a 177-bp repeat sequence, and telomeres[56]. The enrichment for *VSG* within MCs suggest these tiny chromosomes play a role in antigenic variation, but this role has not yet been elucidated. Interesting, the majority of MC *VSGs* are complete *VSGs* that should produce a full-length and functional VSG[26]. This is in contrast to the rest of the *VSG* archive, which is primarily composed of incomplete *VSGs* or *VSG* pseudogenes. In fact, studies of two strains of *T. brucei*, the Lister 427 strain and the TREU927 genome strain, show >80% of genomic *VSG* sequences are incomplete or pseudogenous[26,57].
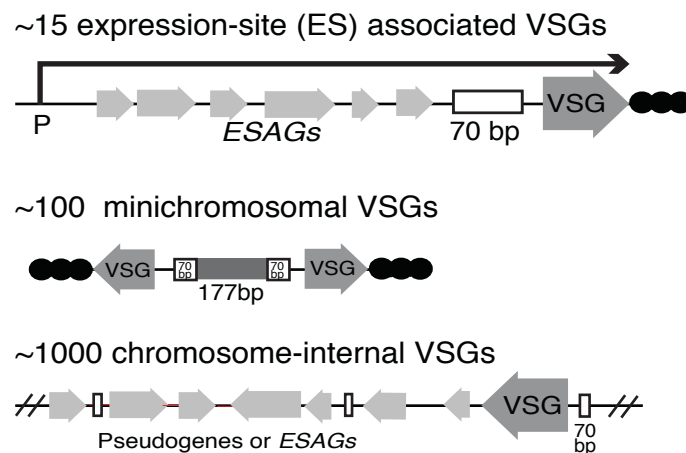
Figure 1.6. The genomic *VSG* archive.
The extensive repertoire of VSG genes is located in various sites throughout the genome. Around 15 *VSGs* are associated with expression sites, where they are transcribed with a cohort of Expression Site Associated Genes (*ESAGs*). Expression sites also contain repetitive 70bp repeat elements. Around 100 *VSGs* are located on ~100 minichromosomes. Finally, a large number of *VSGs* are located in chromosome-internal regions, where they are often found in combination with *ESAGs* and repetitive elements.

1.7 Mosaic *VSGs*

The presence of so many pseudogenes might at first suggest a very limited repertoire of *VSG* genes from which the parasite can draw during an infection. Although there are ~2000 *VSG* sequences in the genome, only ~400 are obviously capable of being used. What purpose could a repertoire of partial *VSG* genes serve? It appears that this repertoire of partial *VSGs* and *VSG* pseudogenes may actually be extremely important to *T. brucei* infection. Various studies have identified expressed *VSGs* that appear to have undergone incomplete gene conversion events, in which part of a *VSG* recombines with another *VSG*, resulting in new, chimeric, variants[58-63]. This chimeric *VSG* sequence is referred to as a "mosaic *VSG*" (Figure 1.7). Such mosaic variants are usually detected *in vivo* after the first few waves of parasitemia (~3 weeks of infection in a mouse)[64].

Reports of mosaic *VSGs* appearing only after the initial peaks of infection suggest that not only is the genomic repertoire of *VSGs* extended by the formation of mosaics, but that this extension of the repertoire may be required to sustain a chronic infection. That is, the repertoire of functional *VSG* genes is insufficient for escape from the host immune system over the course the months- or years-long infections observed in the wild[65]. Notably, the use of recombination for generating antigenic diversity within the genome is common to a number of pathogens that undergo antigenic variation, including *Borellia burgdorferi*[66], *Anaplasma marginale*[67], *and P. falciparum*[68], so this may be a common strategy for immune evasion. Some studies have suggested an even greater extension of the *VSG* repertoire through the introduction of point mutations during gene conversion events[69,70], although others argue that this is not a significant source of *VSG* diversity[71].
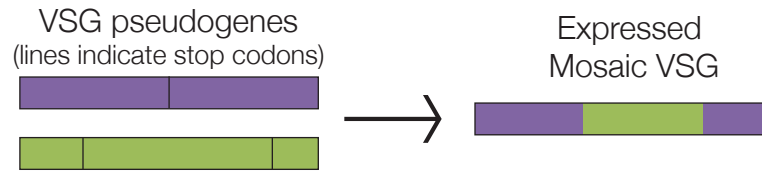
Figure 1.7. Mosaic *VSG* can form by recombination between pseudogenes.

1.7 Mechanisms of mosaic formation

Although many instances of mosaic *VSGs* have been described, very little is known about when or where these variants form. It has been hypothesized that gene conversion in general and *VSG* switching in particular are facilitated by microhomology-mediated end-joining in *T. brucei*[72,73]. Thus, this mechanism is expected to be involved in the formation of mosaic *VSGs* as well. A recent study attempted to tackle the formation of mosaic *VSGs in vivo* in depth by cloning and sequencing hundreds of expressed *VSGs* over the course of mouse infections[64]. In this extensive analysis, mosaic *VSG* donor sequences did indeed typically show homology with one another. It did not appear, however, that homology at the site of gene conversion was required for the formation of mosaic *VSG* genes; there were mosaic *VSGs* in which the boundaries of segmental gene conversion between donors showed absolutely no homology[64]. However, it is important to note that determining the precise source of donor sequence for each mosaic can be complicated, as the genomic *VSG* archive is complex and only partly known in this study[64]. Frequently *VSG* sequences share significant stretches of homology[26,57] that can complicate analysis.

Whether DSBs initiate the formation of mosaic *VSGs*, as is the case for GC switching in general, is not clear. Moreover, because most studies are limited to analysis of expressed *VSGs* and not the evolution of the genomic archive, is not clear where in the

12

genome mosaic *VSGs* form. The gene conversion events resulting in mosaic *VSGs* could

occur within a telomeric BES or throughout the silent *VSG* archive. Because previous

studies have been somewhat limited in resolution[61,62,74-76], it has also been impossible to

determine how frequently mosaic variants form. Although long recognized as an

important aspect of antigenic variation in *T. brucei*, the mechanisms of mosaic *VSG*

formation remain largely mysterious.

1.8 Order in *VSG* switching

Just as mosaic *VSGs* generally appear late in infection, it has been observed in

many *in vivo* studies that other *VSGs* are more likely to arise early in infection[47,77-80].

While some work suggests that this can be explained in part by differences in growth rate

(i.e. a variant that grows the fastest will appear first)[78,81], this view has been challenged[82].

Telomeric *VSGs* are thought to arise early in infection[47], followed by intact but non-

telomeric variants[27,55]. Finally, mosaic *VSGs* comprised of pseudogenes tend to appear

after the other sets of functional *VSGs* have been expressed[64].

The factors driving this preference are not well understood, but a number of

groups have attempted to explain the predicted order of *VSG* switching with modeling.

One model suggests that the order observed during infection arises solely from parasite-

intrinsic factors, rather than the host milieu[81]. Intuitively, it makes sense for genomic

location to determine the probability of a switch to a given *VSG*: a *VSG* within a BES

only requires that BES be activated, and a complete *VSG* near a telomere provides

straightforward homology for transposition the active BES. Mosaic variants, on the other

hand, are likely to be subject to frequent failure due to the requirement that a recombinant

VSG produce an in-frame open reading frame and a properly folding protein. Thus, only when the repertoire of complete and easily expressed *VSGs* has been exhausted do these harder-to-produce variants emerge.

Other models, however, suggest that the activation probabilities of *VSGs* in different locations are not the only players in determining *VSG* order. Some suggest that genomic location in combination with the effects of immune cross-reactivity to certain VSGs[83,84] determines *VSG* order during infection, while other models propose that the host immune system is the chief player in determining *VSG* order, primarily through cross-reactive epitopes[85,86]. A discussion of the specific ways in which the host immune response could affect *VSG* order and switching dynamics will follow in sections 1.12 through 1.15.

1.9 Dynamics of antigenic variation

The extensive modeling of *VSG* switching *in vivo* highlights the fact that the dynamics of this process are a lingering question in *T. brucei* biology. Despite so many attempts at modeling[5,78,81,82,84,85], little is known about the kinetics of *VSG* expression during infection. To date, the resolution of most studies of *VSG* switching in *T. brucei in vivo* has been quite low, examining only a handful of variants at any given time, such that these kinetics could not be measured directly[47,79,80,87,88]. Moreover, many factors on both the parasite and host side of this system work together to produce the observed peaks and valleys of parasitemia during infection, the *VSGs* arising within those peaks, and the *VSGs* cleared from those peaks. Parasite-intrinsic factors include the genomic archive of *VSGs* and the molecular mechanisms driving switching discussed previously, as well as

the rate at which switching occurs in a population. On the host side, the recognition of parasites by the host immune system and the dynamics and mechanisms of clearance of each variant will also likely affect the kinetics of observed *VSG* expression *in vivo*.

1.10 Antigenic variation *in vivo*: Rates of switching

A major parasite-intrinsic factor likely to affect *VSG* expression dynamics during infection is the rate at which switching occurs in *T. brucei*. Despite *T. brucei's* large repertoire of *VSGs*, it is thought that the rate of switching must be tightly controlled. If switching is too slow, the immune system will recognize the expressed *VSG* and clear the infection, preventing transmission to the tsetse. If switching is too fast, it is possible that the *VSG* repertoire will be exhausted prematurely. Many measurements of *VSG* switching rates have been made, both *in vivo* and *in vitro.* In laboratory adapted strains the rates of switching have been measured to be between 10^{-5} to 10^{-7} switches/parasite/population doubling both *in vivo*[87,89,90] and *in vitro*[89-91]. Many interpret the *in vitro* data to suggest that the process of switching is purely stochastic, but some studies have shown the rate of switching in naturally occurring (fly-transmitted) infections to be much higher, perhaps as high as 1 in 100[87,88]. These findings suggest that there may be some environmental cue initiating or influencing *VSG* switching.

Determining the true rate of *VSG* switching during infection presents experimental challenges, however. Trypanosomes expressing different *VSGs* may grow at different rates, or trypanosomes may switch to a *VSG* for which the immune system has already mounted a full or partial response. As a result, even if a switch occurs, that switched variant may never establish in the measured population. Additionally, the non-

dividing short stumpy form of *T. brucei* appears to be incapable of switching[4]. The number of short stumpy parasites present at any time fluctuates throughout infection[5], and thus the number of parasites capable of switching at any given time also fluctuates throughout infection.

1.11 Antigenic variation *in vivo*: Diversity

One measure that may indirectly indicate the rate at which switching occurs *in vivo,* however, is *VSG* diversity within parasite populations. If switching occurs infrequently, populations should be relatively less diverse. Typically, antigenic variation is depicted by a schematic showing only one or a few variants per parasitemic peak (Figure 1.8). This model suggests a rate of switching just sufficient for immune evasion that also most efficiently preserves the *VSG* repertoire. Experimental data, however, suggest this is unlikely to be the case. The most comprehensive study of expressed *VSG* populations to date estimated, but did not measure directly, as many as ~100 *VSGs* could be expressed at any given time during an infection[64]. Other studies have also detected multiple *VSGs* expressed at the same time, though the experiments looked at so few *VSGs* that these observations are hard to interpret[47,64,79,80,88]. A recent study of parasites isolated from human infections also showed VSG diversity higher than the typical depiction of antigenic variation would imply[92]. This suggests that *VSG* diversity is not an artifact of the mouse infection model. It is also important to note that *T. brucei* is extracellular but not exclusively vascular. Trypanosomes can sequester in extravascular spaces[93], and most studies so far have examined only populations in the bloodstream. This could lead to an underestimation of diversity and switching frequency *in vivo*.
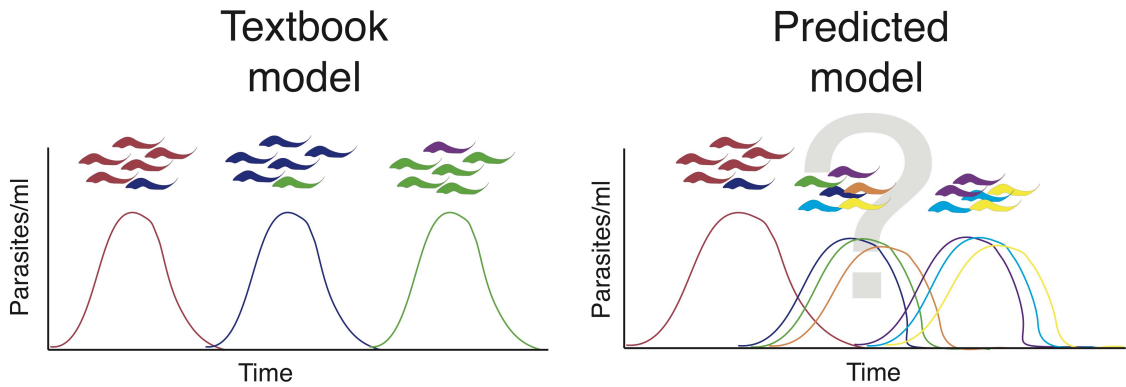
Figure 1.8. *VSG* diversity *in vivo.*

The textbook depiction of *VSG* diversity *in vivo* (left) shows only a handful of variants at each peak, but low-throughput experiments predict much higher diversity (right). The predicted model had not been tested at the time I started my studies.

1.12 The B cell response to *T. brucei*

The dynamics of antigenic variation during an infection are likely to be affected by *T. brucei's* interaction with its host, which determines the rate and mechanisms by which parasites are cleared, likely in response to each novel VSG. Given the parasite's extracellular lifestyle, it is no surprise that the immune response to *T. brucei* is largely B cell mediated. A number of studies, however, highlight the complexity of the B cell response in the arms race between *T. brucei* and its host. One study reported non-specific polyclonal B cell activation in response to *T. brucei* infection[94], although VSG-specific antibodies can also be detected during infection[95,96]. In addition, certain B cell populations have been reported to be lost during *T. brucei* infection, specifically IgM marginal zone B cells in the spleen. This loss of marginal zone B cells coincides with a loss of vaccine-induced protection against *Bordetella pertussis* in vaccinated mice challenged during *T. brucei* infection[97]. Overall, these studies and others[98,99] suggest that B cell exhaustion may take place during *T. brucei* infection. This could occur either

directly through an interaction between the trypanosome and host, or perhaps more likely, through the constant presentation of new antigens via antigenic variation over the long course of infection. This, of course, could have an effect on the outcomes of antigenic variation *in vivo*. If B cell memory is lost, for example, it is likely that antigenic variants that occurred earlier in infection will reappear. Additionally, the diversity of VSG populations might increase over time, in synchrony with a weakening B cell response. Thus, an understanding of the dynamics of antigenic variation could shed light on the immune response to *VSG*, and an understanding of the immune response to *VSG* could shed light on the dynamics of antigenic variation.

In terms of antibody isotype, it is generally thought that IgM B cells mediate the response to *T. brucei* and VSG. This conclusion is based on a study tracking VSG-specific antibody titers during *T. brucei* infections[96] that showed IgM appearing earliest in infection and persisting longer than other isotypes. VSG-specific IgG did not appear until after the first peak of parasitemia was cleared, leading the authors to conclude that the IgM response was responsible for clearance of each variant. On the other hand, IgM$^{-/-}$ mice are still capable of maintaining a chronic infection[100]. Further complicating matters, IgM$^{-/-}$ mice express compensatory IgD[101], an isotype whose function in the immune response is an enduring mystery[102], as well as VSG-specific IgG2a and IgG3. Yet another study suggests that the anti-VSG antibody response is different from the response to other surface proteins[103] in terms of isotype. This study did not determine whether this variability in antibody response could vary among different *VSGs*. Thus, the importance of IgM, in comparison to other isotypes, in the VSG-specific immune response remains an open question.

1.13 Antibody-parasite interactions

The specific role of IgM in parasite recognition and clearance, though not yet clear, is critical for understanding how antibody interacts with *VSG*. Generally, it is thought that only the most N-terminal regions of the VSG can be recognized by antibody, due to the density of the VSG coat. The coat, extending ~12-15nM from the cell membrane[104,105] is thought to "shield" invariant surface proteins as well as the relatively buried C-terminal domain of VSG from binding by host antibody[106]. This shielding of epitopes is important for immune evasion, as it prevents recognition of invariant epitopes, both on invariant surface proteins and in the relatively conserved VSG C-terminal domain.

The specific role of each antibody isotype in recognizing VSG is important because it may determine the degree to which these epitopes are shielded. There is evidence that IgG can reach at least partially into the VSG monolayer[107], though the integrity of the VSG coat is suspect in some experimental situations. Additionally, epitopes have been described that appear more membrane-proximal than expected, quite close to the buried end of the VSG, rather than at the exposed and variable N-terminus[105]. On the other hand, modeling of the protein structure of a handful of mosaic VSGs showed that many of the distinct residues on related mosaics occurred in the N-terminal portion of the VSG, likely to be exposed on the trypanosome surface and thus most immunogenic[64]. If the primary isotype recognizing VSG in a natural infection is IgM, VSG-internal epitopes may not be relevant to parasite clearance. Therefore, the epitopes recognized by the host are likely to be determined by the antibody isotype dominating the B cell response to VSG as well as by each specific antibody-VSG interaction.

Besides potentially shielding internal and invariant epitopes, the VSG coat functions to selectively degrade VSG-specific antibodies, thus dampening the host antibody response[108]. All endocytosis and exocytosis occurs through the flagellar pocket of the parasite[109], and VSG is continuously endocytosed at this location[110]. The parasite quickly degrades antibody bound to endocytosed VSG while shuttling the VSG, now cleared of bound antibody, back to the surface. Though the precise molecular mechanism that underlies this process is not completely clear, it appears to occur through hydrodynamic-flow-mediated forces, with the bound antibody acting as a "molecular sail". Again, the specific isotypes recognizing VSG affect immune evasion: due to hydrodynamic flow, antibody-bound VSG reaches the flagellar pocket more quickly than bare VSG, and the bulky pentameric IgM molecule is cleared more quickly than other isotypes (e.g. dimeric IgG). Thus the rate of antibody degradation by *T. brucei*, which can result in evasion of complement-mediate lysis and opsonization, depends on the specific host antibodies targeting each VSG.

1.14 Mosaic VSGs and antibody cross-reactivity

The possibility of immune cross-reactivity between VSGs could also have consequences on the dynamics of antigenic variation *in vivo*. Mosaic *VSGs* are particularly interesting in this respect because they, by definition, share sequence homology with other *VSGs* in the genome. This is not exclusive to mosaic *VSGs*, however, as many genomic *VSGs* exist as members of closely related families[26] that share stretches of sequence homology. Thus, the question arises: if similar *VSGs* are expressed during the same infection, can pre-existing immunity against a prior variant affect the

survival of a second variant? Studies of mosaic VSG epitopes suggest this is likely to be the case. A number of old studies in *T. brucei* and *T. equiperdum,* a related species of trypanosome that infects horses, actually identified mosaic *VSGs* by their immune cross-reactivity[61,62,74-76]. In these experiments, monoclonal antibodies against a parent *VSG* were found to bind epitopes on other *VSGs*, which were later identified as mosaics. Some studies have shown, however, that point mutations alone can be sufficient to alter monoclonal antibody binding[69,111]. The recent study by Hall et al. identified mosaics with similar sequence that were nonetheless antigenically distinct, as antibodies raised against one variant would not cross-react with a related variant from the same infection. Thus, it is likely that VSG mosaicism can result in a mixture of unique and cross-reactive epitopes. The relative prevalence of each type of epitope and their relative influence of VSG dynamics and switching outcomes, however, is unknown.

1.15 The T cell response to *T. brucei* and T cell recognition of VSG

The T cell response to *T. brucei* has also been characterized[112] and it appears to be central to eliciting an IFN-γ response during infection, which is a major determinant in resistance and susceptibility to infection[113]. Interestingly, although T cells have access to any VSG peptide, including the much more conserved VSG C-terminal domain, on study showed that T cells only present peptides from the more variable N-terminal domain [114]. Somehow, the parasite manages to avoid immune recognition at invariant regions of the VSG. Thus, both arms of the adaptive immune system primarily recognize the variable, and exposed, N-terminal domain of VSG.

1.16 Outstanding questions

The interaction between host and VSG, though it has been well studied throughout the years, is not yet well understood. It is still unclear precisely how VSG is recognized by antibody or how that results in parasite clearance. It is still unclear how often switching occurs *in vivo*. The determinants of order in *VSG* switching are still unknown. The mechanisms of mosaic *VSG* formation remain a mystery.

Overall, antigenic variation in *T. brucei* in the mammalian host occurs in the context of an extremely complex system. The first step towards understanding this complex system is to characterize it. The dynamics of this process are affected by the innate and the adaptive host immune system, density-dependent differentiation of the parasite, the probability of switching to different members of the genomic archive, and the relative fitness of individual parasites, at the very least. Nearly all of the studies on *VSG* switching rate and *VSG* switching outcomes have measured less than ten *VSGs*, for which either the sequence of the *VSG* gene was known or antibodies against the *VSG* existed[47,79,80,87,88]. Even more recent work cloning hundreds of *VSGs* over the course of infection[64] could only detect the handful of variants present at any given time, though they suggest this was an underestimate.

In this thesis I describe the development and application of a high-resolution method for measuring *VSG* expression *in vivo*. I have used this approach to describe the dynamics of antigenic variation during infection, which should provide a foundation upon which to better understand the mechanisms driving this process.

CHAPTER 2. Development of a broadly applicable method for quantifying *VSG* expression

2.1 Introduction

Given the limitations of existing methods for quantifying and identifying *VSGs* expressed in a population of parasites, I sought to develop a new method that could both identify and measure expression of *VSGs* in parasite populations. Ideally, the method would identify any *VSG*, as compared to methods relying on anti-VSG antibody which limit analysis to one or a few antigenic types[49,88,89] and would be quite sensitive, so that it could be applied to the small parasite samples available from natural infections, for which parasitemia is often extremely low[115]. Furthermore, such a method would ideally be able to identify any *VSG*, including mosaic *VSGs*, the novel recombinants that arise during infection through segmental gene conversion and are likely to be important for long-term infection[64]. Along with its utility in identifying novel mosaic *VSGs,* a method capable of identifying any *VSG* would be useful for the analysis of wild parasite populations, for which no genomic *VSG* reference exists.

Given these goals, high-throughput sequencing seemed an ideal approach for examining *VSG* expression in parasite populations. Producing hundreds of millions of reads, high-throughput sequencing is extremely sensitive, quantitative and, in combination with existing tools for transcriptome assembly[116], capable of identifying expressed *VSGs de novo*. In this chapter, I describe the optimization of a high-throughput approach for sequencing *VSGs*, termed "VSG-seq", which is quantitative and sensitive, works on small parasite samples, and does not require a reference genome.

2.2 Optimization of VSG-seq library preparation strategy

There are many possible NGS approaches one can imagine that could solve the problems delineated above, and to identify the best approach for preparing VSG-seq sequencing libraries, I compared four possible protocols (Figure 2.1). Each protocol was designed to identify which aspects of the preparation were important for maintaining accuracy in quantification while also maximizing sensitivity and efficiency. I prepared starting material for all library preparations by mixing four *T. brucei* cell lines, each expressing a different *VSG,* in equal proportions. These control samples would allow me to assess the accuracy of quantification using each method.
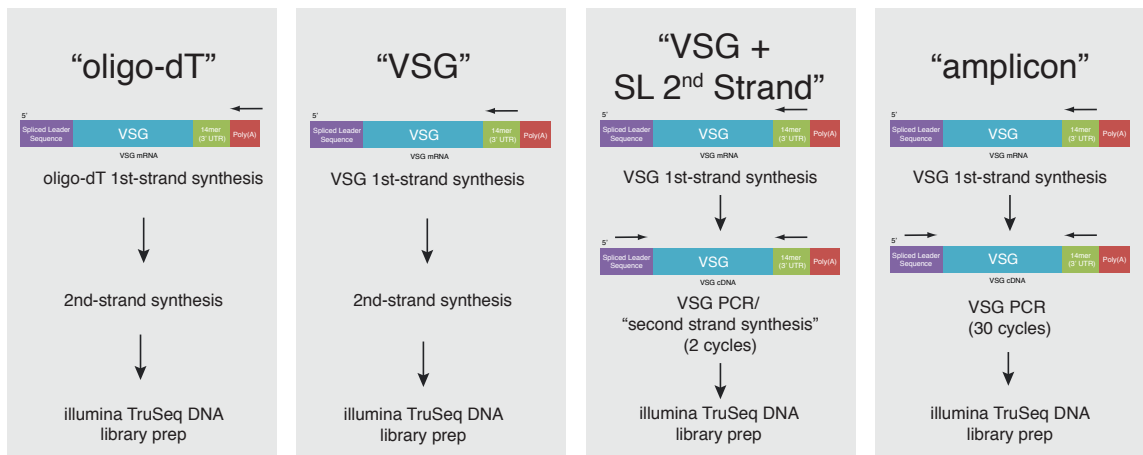
Figure 2.1. Schematic of four library preparation protocols.

The first approach, "oligo-dT", was used as a baseline for comparison. This protocol was a typical RNA-seq library preparation in which poly(A) mRNA is isolated and then fragmented, followed by generation of double-stranded cDNA (dscDNA) using random hexamers. The dscDNA was then used as the input material for an Illumina DNA sequencing library preparation kit.

The second approach, "VSG", tested the degree of enrichment for *VSG* sequence accomplished by using a *VSG*-specific primer for 1st-strand synthesis. As an alternative to random hexamers, a primer against the conserved 14-mer in all *VSG* 3'UTRs was used to prime cDNA synthesis. 2nd-strand synthesis was performed as in the oligo-dT preparation, followed by enzymatic fragmentation of dscDNA. The fragmented material was used as input material for the same kit as in the oligo-dT protocol. Ideally this method would remain quantitative but, because the majority of sequenced material would be *VSG* mRNA, increase sensitivity.

The third approach, "VSG + SL 2nd Strand", tested whether an alternative 2nd strand synthesis protocol, in which 2 rounds of *VSG*-specific PCR were performed to create dscDNA, could increase sensitivity even further. This method might also increase the proportion of reads coming specifically from full-length *VSG* mRNAs.

The final protocol, "Amplicon", tested whether simply sequencing *VSG* cDNA amplicons could increase sensitivity and enrichment for *VSG* sequence without compromising quantification. This protocol used the same *VSG*-specific primers as in the previous method, but increased the number of cycles of *VSG* PCR to 35.

100bp single-end sequencing was performed on all prepared libraries, and *VSG* expression was measured by aligning reads to a reference genome made up of only the sequences of the *VSGs* mixed in the control populations. The number of reads aligning to each *VSG* was then counted. To calculate the percent of the population represented by each *VSG*, which should be 25% in all cases, the proportion of reads aligning to each *VSG* was divided by the total number of reads aligning to any *VSG*. It is important to note

that, because reads were aligned to a reference of only four genes, this analysis did not assess any possible noise or mapping issues that could result from VSG-seq analysis.

This analysis revealed that all approaches were fairly quantitative (Figure 2.2A), showing values within 2-3 fold of the actual representation of each *VSG*. This variability could be due to variations in expression levels between individual *VSGs* or BESs, although the variation observed for each *VSG* was not consistent between protocols. This experiment was performed without replicates, however, so it is hard to interpret such variation. Overall, no method appeared significantly more biased than the others in terms of quantification.
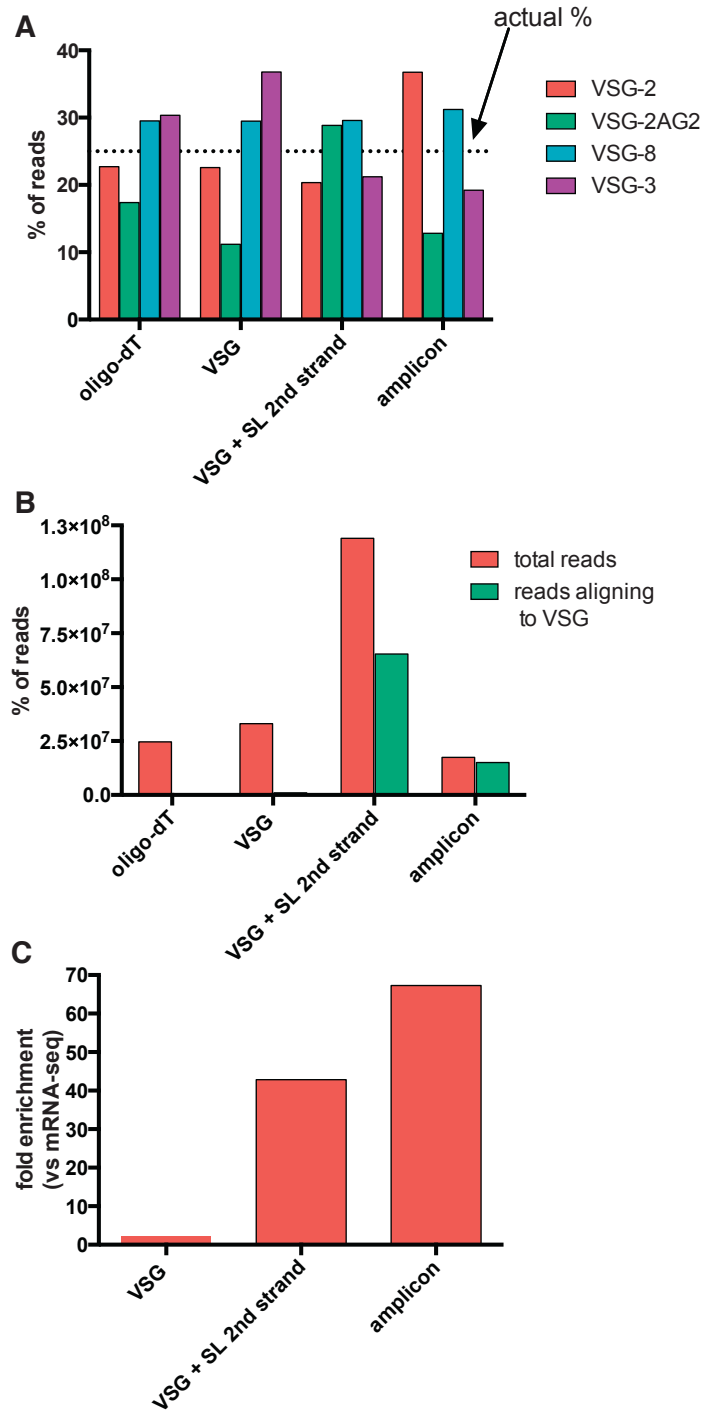
Figure 2.2. *VSG* expression levels and enrichment measured by each protocol.
(A) Graph showing *VSG* expression measured by each method. (B) Graph showing the
total number of reads and number of reads aligning to *VSG* in each preparation. (C)
Graph showing enrichment for *VSG* sequence, relative to standard mRNA-seq, in each
library.

2.2.1 Enrichment for *VSG* sequences

Because I observed little bias in quantification of *VSG* expression between methods, I chose to investigate whether enrichment for *VSG* sequence varied between library preparations. This was important for two reasons: (1) increased sequencing depth could improve sensitivity; and (2) enrichment for *VSG* would imply fewer "wasted" non-*VSG* reads, providing more information per sequencing run and potentially allowing a higher degree of multiplexing per sequencing lane.

Both "VSG" and "oligo-dT" showed a very small proportion of reads mapping to *VSG* (Figure 2.2B&C). Approximately 1% of mRNA sequence is *VSG* mRNA, which is expected given the high level *VSG* expression required in bloodstream form cells. Using only a *VSG*-specific primer for reverse transcription did not result in significant enrichment for *VSG* sequence. However, using a combination of this primer and a primer against the spliced-leader sequence resulted in the majority of reads mapping to *VSG* (Figure 2.2B). When compared to the proportion of reads mapping to *VSG* after the "oligo-dT" protocol (Figure 2.2C), "VSG + SL 2nd strand" increased enrichment to over 40-fold, while amplicon sequencing increased enrichment to nearly 70-fold.

2.2.2 Read coverage across *VSG* genes

Another concern in developing VSG-seq, besides enrichment for *VSG* sequence, is *VSG* sequence coverage. While enrichment can be high, this does not guarantee that coverage is uniform. Because one of the goals of VSG-seq was *de novo* assembly of *VSGs*, good coverage of each *VSG* is critical; otherwise, it would be impossible to reconstruct full-length *VSG* sequences. The reconstruction of full-length transcripts is

particularly important in the case of *VSGs*, because the 3′ end of the *VSG*, from which

VSG is reverse transcribed, is the more conserved end of the *VSG*, and a reference

database composed of 3′ *VSG* fragments would severely underestimate the extent of *VSG*

diversity in any sample. While reverse transcriptase may be processive enough to

transcribe full-length *VSG*, the efficiency of this reverse transcription in the case of *VSGs*

is unknown. To examine this, read coverage was visualized in Integrative Genomics

Viewer (IGV)[117,118], which visually depicts the mapping of each read to the reference

genome (Figure 2.3). Only the PCR-based approaches showed unbiased coverage, further

supporting the use of these methods for VSG-seq. Because amplicon sequencing would

allow for the sequencing of the smallest samples, it was chosen for further analysis and

optimization.

In summary, I tested four different protocols, and chose an amplicon sequencing

approach based on the following attributes: (1) Enrichment. Amplicon sequencing

provided the most *VSG* sequence per library; (2) Coverage. Amplicon sequencing

provided uniform read coverage across the entire *VSG* transcript; (3) Quantification.

Amplicon sequencing was similarly accurate in quantification to the other methods

tested; and (4) Input requirements. Amplicon sequencing required the smallest amount of

starting material, because *VSG* cDNA is amplified by PCR during the library preparation
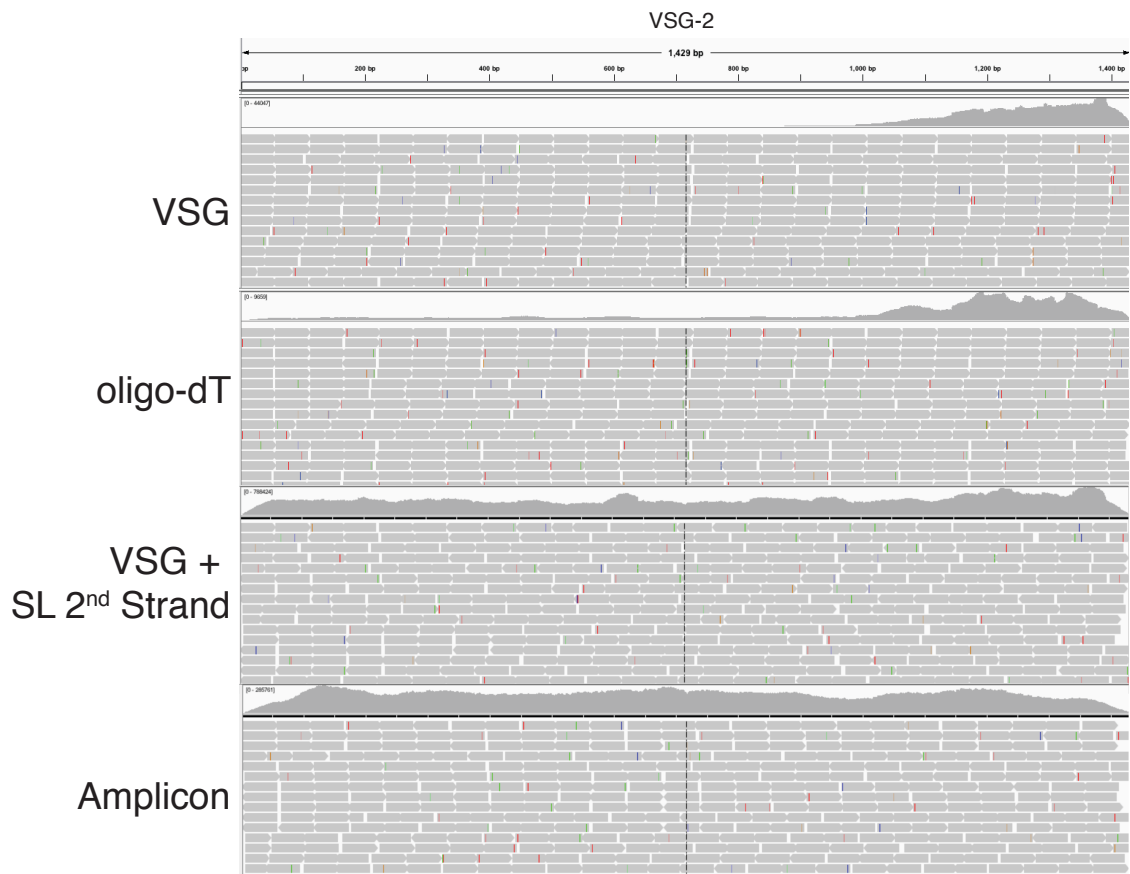
Figure 2.3. Read coverage across *VSG* transcripts.

Screenshots from IGV visualization of alignments of reads to *VSG-2* from each library preparation. Each row shows an alignment of a subset of reads (bottom) and overall read coverage(top) for the protocol noted.

2.3 Optimizing VSG-seq analysis: mappability

After identifying *VSG* amplicon sequencing as the best approach for VSG-seq, I

optimized the analysis parameters for this approach. In the previous analysis, sequences

were only mapped to a reference genome containing *VSGs* known to be in the sample.

This simplified quantification, but ignored a major issue with sequencing *VSGs*:

mapping. Portions of *VSG* sequences are frequently very similar or identical[26], such that

it is likely that a given 100-bp read cannot map uniquely to certain regions within the

VSG it arose from. Uniquely mapping reads are required for accurate quantification of

expression, however, and quantification of multiply mapping reads (without the aid of

some complicated analysis algorithms that have not yet been developed) will always

result in the over- and/or underestimation of *VSG* expression.

An analysis of the annotated Lister 427 VSGnome reference confirmed that

mapping is a concern when analyzing *VSG* sequences. To assess the "mappability" (the

proportion of reads which can uniquely map) of the VSGnome, 100bp reads were created

from each reference database, representing all of the possible reads that could result from

each *VSG* sequence. These simulated reads were then mapped back to the reference,

allowing for only uniquely mapped reads with no more than 2 mismatches per read. A

number of databases were analyzed, using different sets of *VSGs*, and the mappability of

each *VSG* was calculated by dividing the number of reads mapping to that *VSG* by the

number of reads that would map to the *VSG* if its sequence were completely unique. For

example, if only 50% of the sequence of a *VSG* is unique to that *VSG*, that variant will be

50% "mappable". This analysis shows that as the number of *VSGs* in a reference database

increases, the mappability of the *VSGs* within that database decreases (Figure 2.4). A

simple count of reads mapping to each *VSG*, then, will result in an underestimate of expression for *VSGs* that are less unique and an inability to quantify the expression of *VSGs* that have no unique sequence at all.

To assess the degree to which mappability affects quantification in a real sequencing experiment, I aligned the sequences generated from the amplicon sequencing library described earlier in this chapter to the Lister 427 reference VSGnome and quantified abundance either using cufflinks[119], an algorithm with little correction for mappability, or MULTo[120], a program specifically designed to correct for the mappability of each gene. This analysis shows that correction for mappability does indeed affect expression measurements generally (Figure 2.5A) and in the case of the specific *VSGs* in the control library (Figure 2.5B). Cufflinks, which does not correct for the uniqueness of each sequence, hugely overestimates the abundance of VSG-2, while it underestimates VSG-832. These errors are proportional to the mappability of each VSG. Thus, VSG-seq can only accurately measure *VSG* expression when a tool to correct for mappability, such as MULTo, is used. All further analyses, unless otherwise noted, used MULTo for quantification of *VSG* expression.
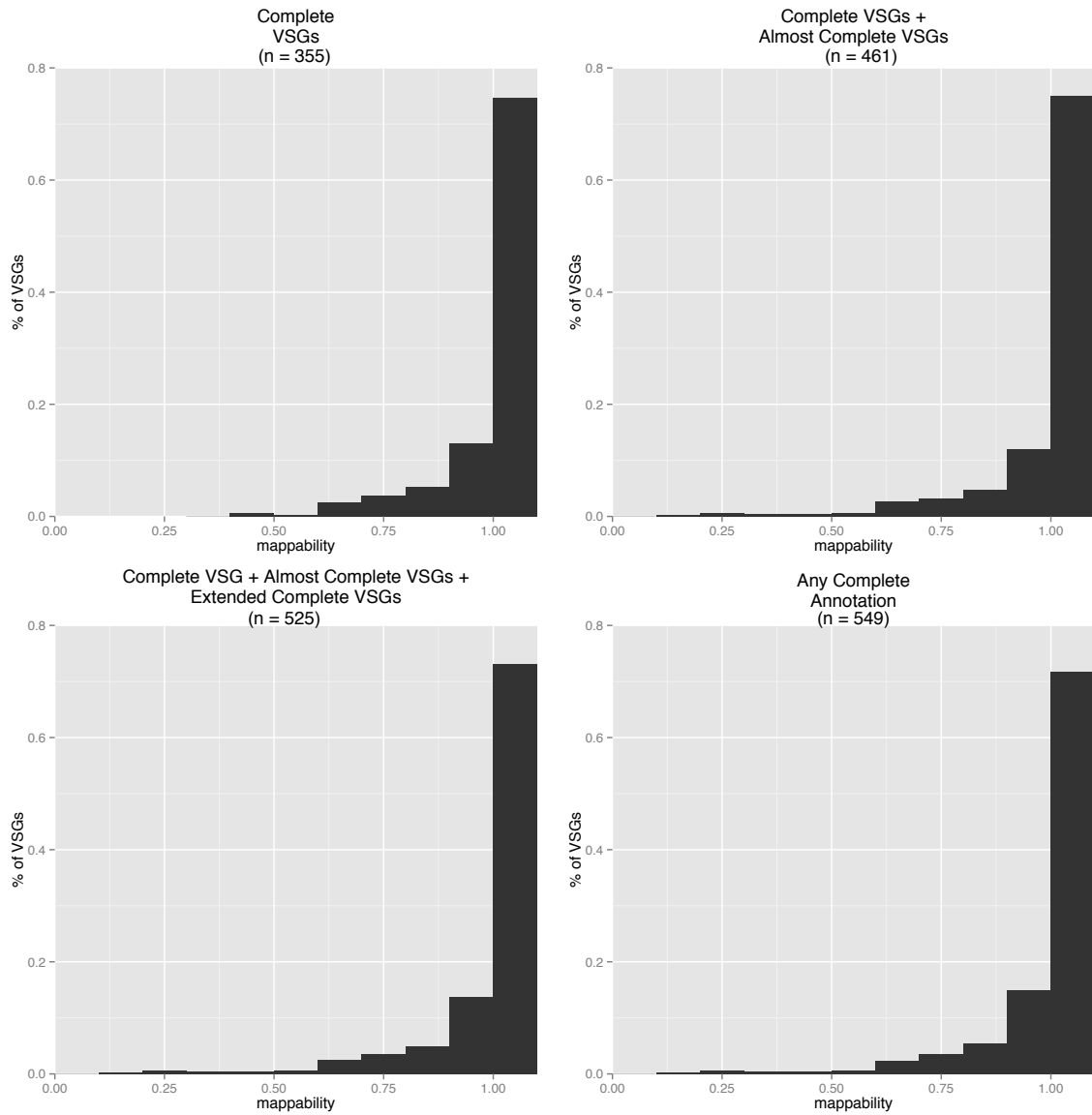
Figure 2.4. Mappability of different VSG databases.

Histograms showing the mappability (or uniqueness) of VSGs in databases of various
sizes.

2.4 Characterizing VSG-seq

After identifying the ideal parameters for preparing and analyzing VSG-seq

libraries, I wanted to assess the final protocol. Specifically, I sought to determine the

sensitivity of the method along with whether the raw number of input cells could affect

the limit of detection or quantification of *VSG* expression. To do this, I again prepared

control mixtures of parasites. An outline of this experiment is provided in Figure 2.6.

Briefly, I prepared two sets of mixtures in triplicate, from seven lines, each expressing a

different *VSG*. In one set (Library A), cells were mixed in ten-fold dilutions, and in the

other (Library B) cells were mixed in equal proportions. After preparing these mixtures,

either 1 million or 10 million cells were used to make VSG-seq libraries. For these

experiments, and all further VSG-seq experiments, a newer, low-input DNA sequencing

kit from Illumina (NexteraXT) was used to make sequencing libraries from *VSG*

amplicons.
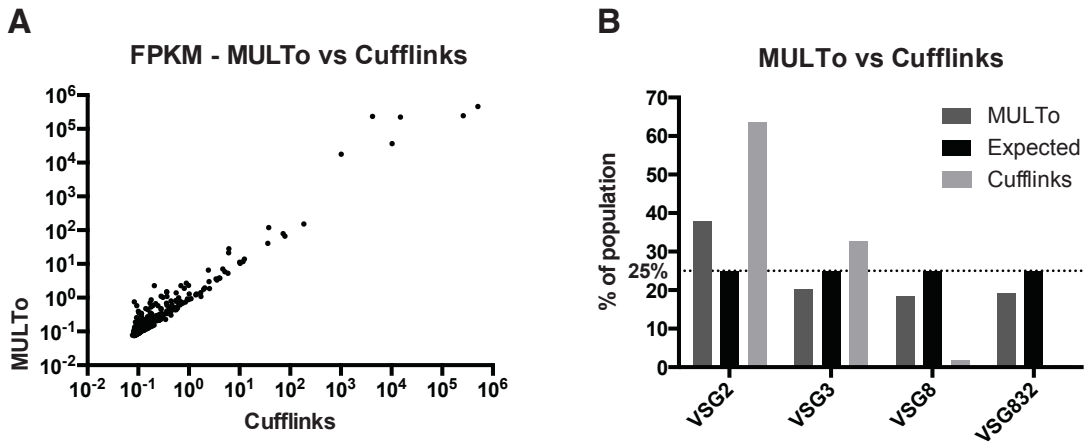
Figure 2.5. Comparison of quantification using MULTo or cufflinks.
(A) Scatter plot showing measured FPKM for all *VSGs* using each algorithm. (B) Bar
chart showing relative expression of each *VSG* in control mixture. The black bar
("Expected") represents the proportion of cells expressing that *VSG* in each control
mixture, and the gray bars represent quantification for each library using MULTo or
cufflinks.
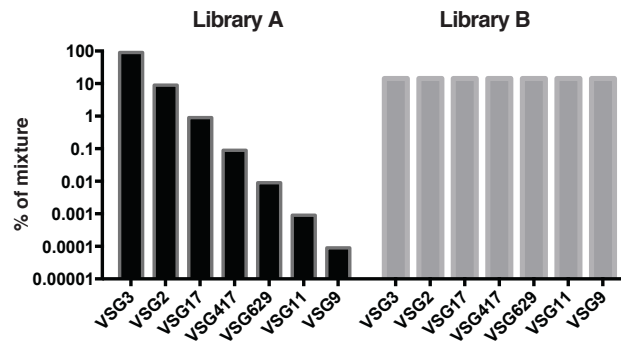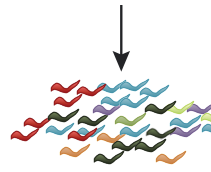
Figure 2.6. Schematic of creation of control libraries for validation.

2.4.1 Assembly of *VSGs* and relationship to number of input cells

Although previous analyses show that *VSG* expression could be quantified using the Lister 427 reference, this reference is useless for analysis of any strain without a reference VSGnome or for the analysis of expression of mosaic *VSGs*, because mosaic variants generated by segmental gene conversion during the course of an infection are, by definition, absent from any reference genome. Thus, it was critical to test whether a useful *VSG* reference could be assembled *de novo*. I used Trinity, a transcriptome reconstruction method designed specifically for the reconstruction of alternative isoforms, for *de novo* assembly of *VSGs*[116]. I reasoned that this tool might be suitable for the reconstruction of *VSG* transcripts, which often share many stretches of homology, especially in the case of mosaic *VSGs*. I tested Trinity's assembly using Library A in order to determine whether assembly required some level of coverage for accurate reconstruction of a *VSG* transcript. This analysis revealed that Trinity could successfully assemble full-length or nearly full-length *VSG* sequences from as few as nine cells in the control mixture (Figure 2.7). It is important to note that Trinity also assembled *VSG* transcripts that were not expressed in the control mixtures, which probably represent transcriptional noise in each cell.
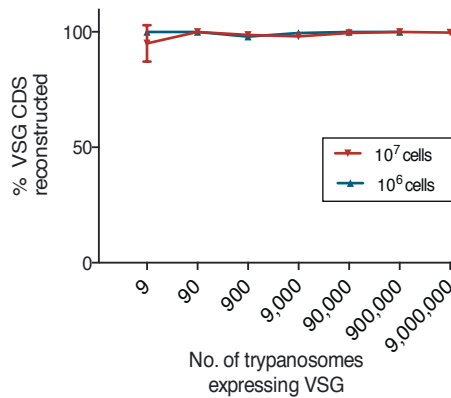
Figure 2.7. Efficiency of assembly using Trinity.

Graph showing the fraction of each *VSG* assembled (mean +/- SD) compared to the number of cells in the population.

2.4.2 Limits of quantification of *VSG* expression

After aligning reads to the Trinity-assembled reference genome (in the case of Library A; Library B was quantified using the Lister 427 reference VSGnome), I determined that VSG-seq is capable of quantifying a variant's presence within the population, for variants present above 0.1% of the population (Figure 2.8A), within a ~2-5-fold range (Figure 2.8B). VSG-seq could quantitatively detect variants present on 0.01% of parasites; all assembled *VSG* sequences that were not present in the control mixtures were measured at less than 0.01% of the population (Figure 2.8C), while variants present at <0.1% of the control mixture all showed values between 0.01-0.1%. The apparent overestimation of less abundant *VSGs* in this control experiment is likely a result of low-level switching in the more abundant components of the mixture, or low-level transcription of silent *VSGs*. The limits of detection and quantification for VSG-seq appear to be independent of starting cell number, as control mixtures made from 10^6 or 10^7 cells showed similar results.
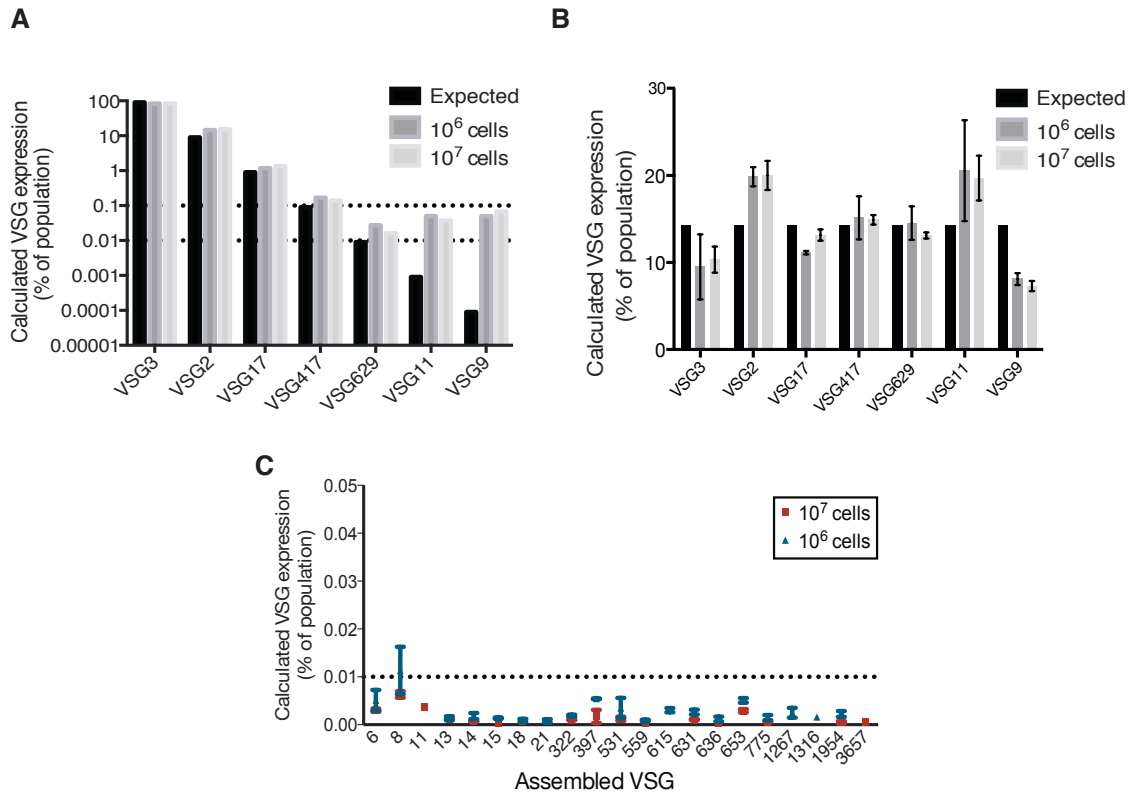
Figure 2.8. Quantification of *VSG* expression in control libraries.

(A) Quantification of *VSG* expression in control library (mean +/- SD), in which cells were mixed in ten-fold dilutions. The black bar ("Expected") represents the proportion of cells expressing that *VSG* in each control mixture, and the gray bars represent quantification for each library using VSG-seq. (B) Quantification of *VSG* expression in control library (mean +/- SD) in which cells were mixed in equal proportions. Note that the analysis of this library was performed using the Lister427 VSGnome database, rather than an assembled reference genome. (C) *VSG* expression calculated for *VSGs* absent from the control mixture in Figure 1 (mean +/- SD). Although some *VSGs* are assembled and quantified below 0.01%, these do not appear to correspond with VSG protein in the population. The single point above 0.01% is likely the result of a parasite that has undergone switching in one replicate, as this *VSG* is expression-site associated and is frequently switched to in this strain.

38

2.5 Discussion

These experiments demonstrate that VSG-seq is both quantitative and sensitive. The optimized method, based on sequencing of *VSG* amplicons, maintains accuracy in quantification even after 35 cycles of PCR, so even very small samples can be analyzed. Most importantly, the approach relies on *de novo* assembly of *VSG* transcripts from sequencing reads to identify *VSGs* present in a parasite population. This approach, an alternative to methods relying on anti-VSG antibodies, can be applied to any sample, regardless of one's knowledge of the parasite's genomic *VSG* repertoire.

Of course, there are aspects of this method that can complicate analysis and which are important to keep in mind when analyzing VSG-seq data. First, the limit of detection of VSG-seq is neither a number of cells nor a number of reads, but rather a proportion of the sample of parasites analyzed. This is likely due to transcriptional noise in every cell. In these control samples, at least, it is likely that this noise arises from similar genomic sources (silent BESs or chromosome-internal *VSGs* which fall near a promoter are possibilities) and these transcripts are present at a level of about 1 per 10,000 *VSG* transcripts. Thus, when analyzing a sample of 10^6 parasites, any VSG present on less than 100 parasites will not rise above the level of this noise. This is unfortunate, because these extremely rare variants could be extremely interesting biologically. An approach that could determine *VSG* expression in individual cells, such as single genome sequencing[121], could be used to determine the precise level of transcriptional noise in individual parasites. Another complicating aspect of this detection limit is that smaller samples (<10,000) allow, in theory, for detection of all variants. These control experiments only tested samples isolated from 10^6 parasites; further characterization of libraries made from

smaller numbers would be beneficial to understanding the limitations of VSG-seq. In patient samples, for example, which may only contain hundreds of parasites, an understanding of the limits of detection for library preparations made from small samples may be critical to interpretation of VSG-seq data.

Another aspect of this method that is important to consider is the issue of mappability. Quantification relies on corrections for mappability but in cases where sets of *VSGs* are very similar, there may be no unique region of certain *VSGs* and no straightforward way to correct for this. Although this has not been a major issue in my experiments, in some situations, perhaps later in infection when mosaic *VSGs* dominate, this could be an important consideration for accurate VSG-seq analysis.

Along these same lines, read mapping can limit the diversity detectable within a population. Due to the error rate of high-throughput sequencing[122], it is typically to allow read mapping with mismatches. To effectively map reads to assembled *VSGs*, this analysis requires that assembled *VSGs* sharing >98% sequence identity be "collapsed" into one representative *VSG* sequence. Thus, some degree of diversity at the nucleotide level can be lost during this analysis. An alternative approach would be to allow no mismatches during read mapping, but this could compromise the sensitivity of the approach.

Overall, however, these experiments have shown that VSG-seq works well on samples isolated from *in vitro* cultures. In the next chapter, I will examine the feasibility of the technique when applied to samples collected from the blood of infected animals.

CHAPTER 3. *VSG* expression during acute *T. brucei* infection

3.1 Introduction

After optimizing VSG-seq *in vitro*, I next piloted the approach in samples collected from mice infected with the well-characterized but highly virulent Lister 427 strain of *T. brucei*. This strain was chosen because we have some knowledge of its switching behavior *in vitro*, to which *in vivo* results can be compared[49,54]. Although the Lister 427 strain has been well characterized *in vitro*, little is known about antigenic variation in the strain *in vivo*. The strain was ideal for a pilot experiment because the Lister 427 reference VSGnome could be used to validate assembled *VSGs*. This pilot experiment would demonstrate the applicability of VSG-seq to samples collected from animals, in addition to possibly providing new information about this strain's switching behavior *in vivo.*

3.2 Infection dynamics vary between mice

For this pilot, five Balb/cByJ mice were infected intraperitoneally (i.p.) with ~5 wild-type Lister 427 parasites expressing VSG-2. Five parasites were used to initiate infection in order to keep infections clonal, so that any variants detected at later times would represent true switch events and not the outgrowth of a rare variant present upon infection. Indeed, in experiments described in Chapter 6, I found that an increased parasite inoculum does appear to increase observed diversity. Parasitemia in tail blood was monitored daily by hemacytometer. Once parasitemia was detectable, parasites were collected by submandibular bleed. The vast majority of red blood cells (RBCs) were

41

removed from each sample using magnetic-activated cell sorting (MACS) with anti-Ter119 beads, which I determined to be a better method for smaller numbers of parasites than the DEAE cellulose protocol used extensively in the past[123].
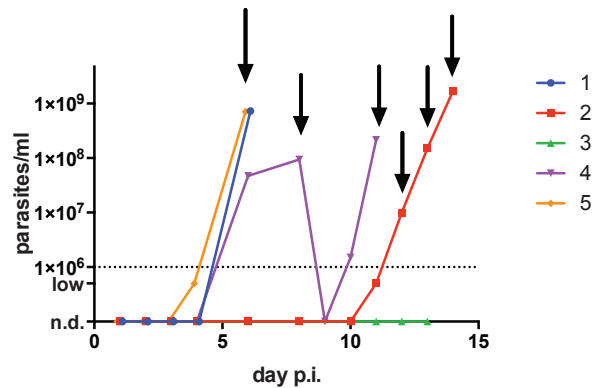
Figure 3.1. Parasitemia during acute infection.
Graph showing parasitemia in each of five mice during infection, as measured by hemacytometer. "low" indicates parasites could be detected, but parasitemia was too low to quantify accurately. Arrows indicate when samples were collected for VSG-seq; samples were only collected in those mice in which parasitemia was measureable.

The dynamics of infection in each of the five mice varied significantly (Figure 3.1). One mouse never established an infection, which was most likely due to the dilution of injected parasites, suggesting that these were in fact clonal infections. If, as intended, the injection of ~5 parasites effectively results in 1-2 parasites initiating each infection, infrequently a mouse will receive an inoculum containing no parasites and will not become infected, as seen in this experiment. Two mice (Mice 1 and 5) succumbed to infection at day 6 p.i., without clearing the first wave of parasitemia, a common occurrence in infections with this particular *T. brucei* strain, which reaches high levels of

parasitemia due to its inability to respond to quorum sensing factors[124]. Two mice, however, managed to control infection beyond the first week. Interestingly, one of these mice (Mouse 2) did not show detectable parasitemia ($>10^5$/ml) until day 11. The last mouse (Mouse 4) showed more typical dynamics, clearing a first peak of parasitemia around day 6, followed by a second parasitemic peak before succumbing to infection. These variable dynamics presented an interesting scenario in which to investigate antigenic variation: do they relate to the repertoire of *VSGs* expressed in each infection, or are they related to the individual immune systems of each otherwise congenic mouse? Although these mice are clonal, their immune systems, and their B cell repertoires in particular, are highly and divergently polyclonal[125].

3.3 Analysis of variants in mice unable to control the first peak of parasitemia

Analysis of the two mice that could not control the first wave of parasitemia revealed switching behavior similar to what is observed *in vitro* in this strain. At the end of both infections, >99% of parasites express the starting variant, VSG-2, and a small number of parasites, almost all at <0.1%, express 4-5 other VSG types. 7 unique variants were identified between the two infections, of which two are present in both infections. Five of these variants are known to be present in a BES[32]. While the genomic position of VSG-12 is uncertain, this variant is commonly seen in populations of switchers in Lister 427[54,126]. Interestingly, a new *VSG* was assembled from and expressed in both populations. It is likely that this variant is absent from the Lister 427 reference database because this database is based on the Lister 427 "single marker" line[127], and the cells used in this experiment were the wild-type line from which the single marker line was derived.
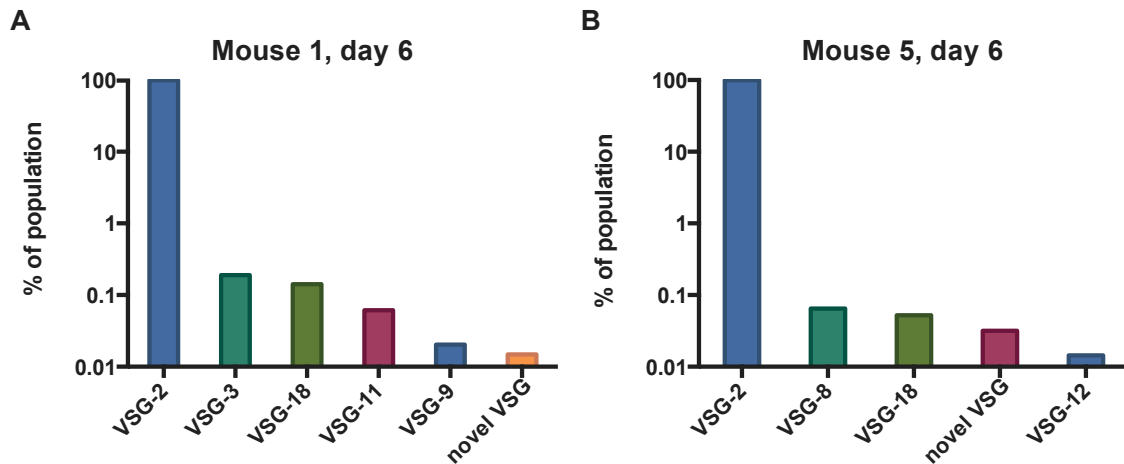
Figure 3.2. Early *VSG* expression.

Quantification of *VSG* expression in Mouse 1 (A) and Mouse 5 (B). Note that the y-axis is on a log scale and colors do not correspond between graphs.

Overall, the population of switchers in each mouse is quite similar and consistent with switched populations observed *in vitro*, suggesting that antigenic variation may not be influenced by host factors, at least in this strain at this early stage of infection.

3.4 Analysis of first relapse peaks in mice that control infection temporarily

Analysis of the later time points collected from Mice 2 and 4 showed surprising *VSG* diversity, at least in contrast to the simplified though prevalent model of switching in which one or a few variants make up each parasitemic peak. In both infections, 9 variants besides the starting variant, VSG-2, could be detected (Figure 3.2 and Figure 3.3). In total 14 variants are detected between both infections, 5 of which appear in both. All variants detected appeared in at least one of the infections described in the previous

section, again demonstrating that switching preferences in this strain *in vivo* resemble those observed *in vitro,* and that switching preference is consistent between infections even with different dynamics.
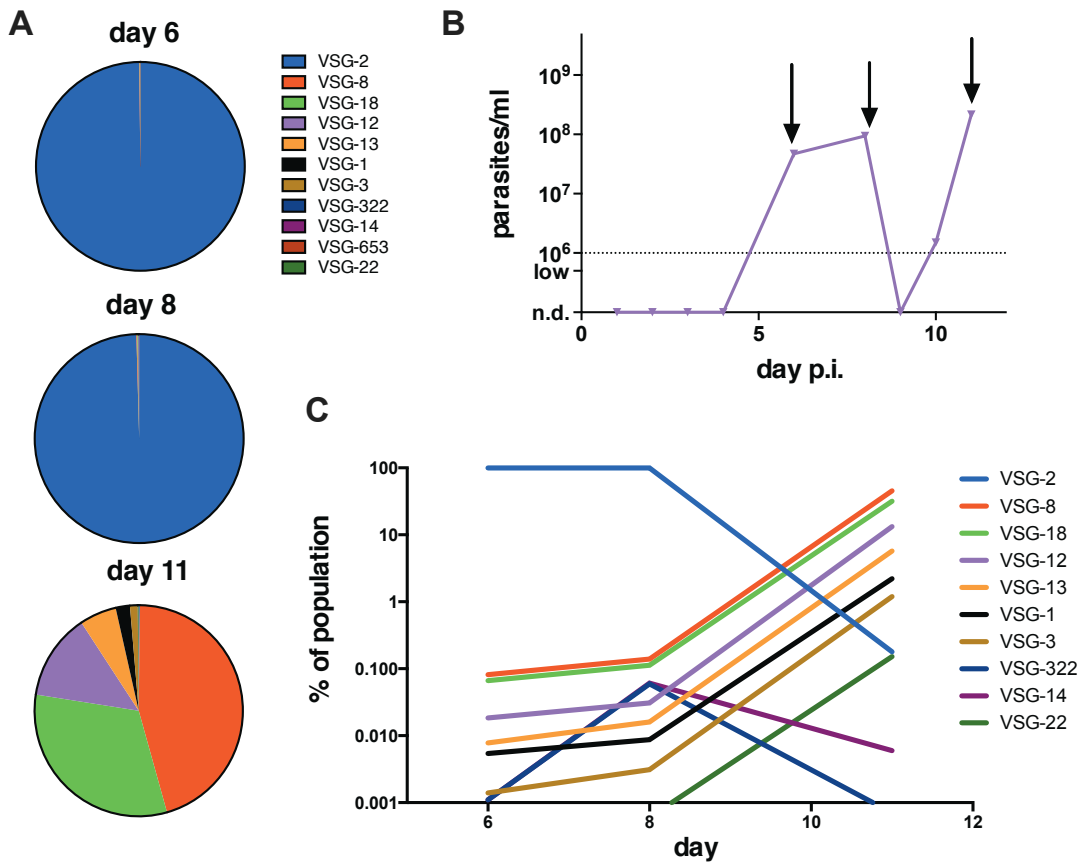
Figure 3.3. *VSG* expression across initial and first relapse peaks.

(A) Pie charts showing relative expression of each *VSG* at each time point analyzed in Mouse 4. (B) Graph showing parasitemia in this infection, with collection times indicated by an arrow. (C) Line graph showing change in expression in each *VSG* over time.

Examining each infection individually reveals a number of other interesting aspects of *VSG* expression *in vivo*. Mouse 4 shows a peak of parasitemia with approximately the same timing as the peaks in Mice 1 and 5 and, indeed, the makeup of this peak resembles that of Mice 1 and 5. Analysis of the later time points in this infection show that the rare variants from day 6 make up the more abundant variants on day 11, in roughly the same proportions. Interestingly, however, two *VSGs* appear on day 8 but are absent from the population on day 11, suggesting that there may be some competition between *VSGs in vivo*, causing certain variants to fail within a population.
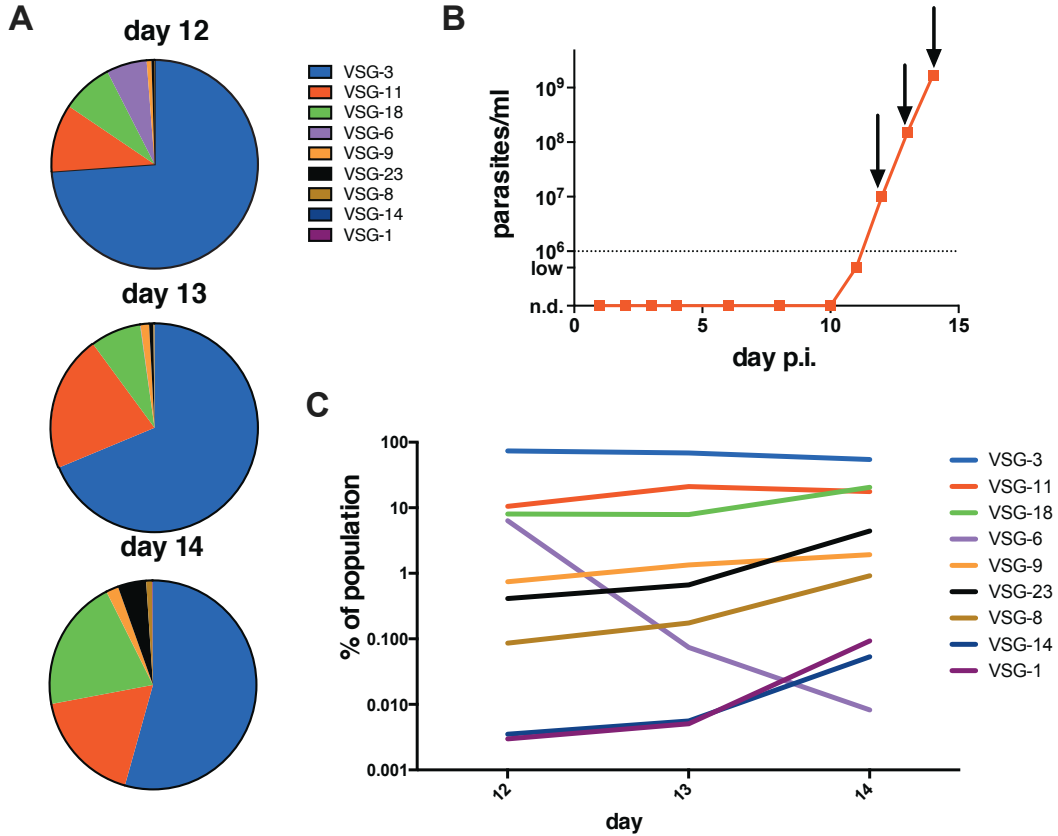
Figure 3.4. *VSG* expression in the absence of an initial parasitemic peak.
(A) Pie charts showing relative expression of each *VSG* at each time point analyzed in Mouse 2. (B) Graph showing parasitemia in this infection, with collection times indicated by an arrow. (C) Line graph showing change in expression in each *VSG* over time.

Mouse 2 never shows a first parasitemic peak, but nevertheless shows clearance of VSG-2 by day 12, the first time populations were analyzed. This shows that *VSGs* can be expressed and cleared by the host immune system even when parasitemia is very low or undetectable, and that variants need not be abundant to be cleared.

3.5 Discussion

Overall, these experiments demonstrate that VSG-seq works well when applied to samples collected from an infected animal. All *VSGs* assembled matched annotated *VSGs* from the Lister 427 strain, and most early variants were BES *VSGs,* consistent both with observations *in vitro* and with the prevailing model of switching hierarchy in *T. brucei*, in which telomeric *VSGs* appear earliest in infection. Moreover, the experiments demonstrated the power of *de novo* assembly for identifying *VSGs*, as a new *VSG* was discovered in two infections.

The variation in infection dynamics among the four infections raises interesting questions about *VSG* dynamics *in vivo*. Though all infections were initiated with genetically identical parasites, control of the infections was quite variable, suggesting that the host immune system could play a large role in the measureable outcomes of *VSG* switching. Specifically, the dynamics observed in Mouse 2 show that some variants may be cleared even when present in small numbers. This is likely due to the germline B cell repertoire unique to this mouse[125]. That is, this Mouse 2 may have had pre-existing antibodies recognizing VSG-2, resulting in early clearance of this variant and suppression of the first parasitemia peak. Thus, although switching may occur in a rough order, with some probability for certain variants to be chosen before others, the observable outcomes

of switching may be largely shaped by the B cell repertoire of a given host. The dynamics of this particular infection also emphasize the importance of analyzing parasite populations even at very low parasitemia. During the time when parasitemia was undetectable, VSG-2 was cleared and another variant took over as the major *VSG* within the population. The dynamics of this process were missed because samples were not collected during this time.

Despite the variation in infection dynamics between these four infections, all of the infections have one thing in common: substantial and unexpected *VSG* diversity, particularly as infection progressed. However, because these later time points also represented terminal time points, it remained unclear whether this was a common feature of antigenic variation *in vivo* or a consequence of terminal parasitemia at the time of sample collection. To determine whether this diversity is characteristic of long term *T. brucei* infection, which is a more relevant biological context, I next investigated *VSG* expression dynamics during infections with a less virulent trypanosome strain that produces chronic infections in mice.

CHAPTER 4. VSG expression *in vivo* during chronic *T. brucei* infection

4.1 Introduction

After analysis of acute infections with a virulent strain of *T. brucei,* it became

clear that a more chronic model of infection would be required to answer questions about

long-term *VSG* expression dynamics *in vivo.* Such an infection model is more relevant to

natural infections, which can last months or even years[65], and certain aspects of antigenic

variation that may be critical to natural infections, such as the appearance of mosaic

VSGs or the effects of immune system exhaustion on *VSG* dynamics, are likely only to be

measurable after weeks of infection. To analyze *VSG* dynamics during a chronic

infection, I used the EATRO1125 strain of *T. brucei*[128,129], which produces infections of

at least 30 days, or approximately 4 parasitemic peaks, in a mouse.

4.2 *VSG* diversity and dynamics during early infection

To measure *VSG* expression within populations of *T. brucei*, I infected four mice

with ~5 EATRO1125 parasites. This strain had originally expressed *VSG* AnTat1.1 [128,129]

but was now heterogeneous, such that each infection was initiated with a distinct *VSG*. I

tracked *VSG* expression dynamics for 30 days (Figure 4.1 and Figure 4.2; the raw data

produced by these experiments and used for these analyses can be found in the published

version of this work[130]). A few variants made up the majority of the population at each

time point analyzed, but, surprisingly, each sample also contained many rare variants that

would have been undetectable using previous methods for characterizing *VSG*

populations. Variants typically persisted for approximately 7 days followed by quick

clearance from the population, although some variants deviated from this pattern. Notably, all infections showed great diversity even within parasitemic valleys. VSG-seq identified an average of 28 variants at each time point analyzed during the first 30 days of infection (Table 4.1).
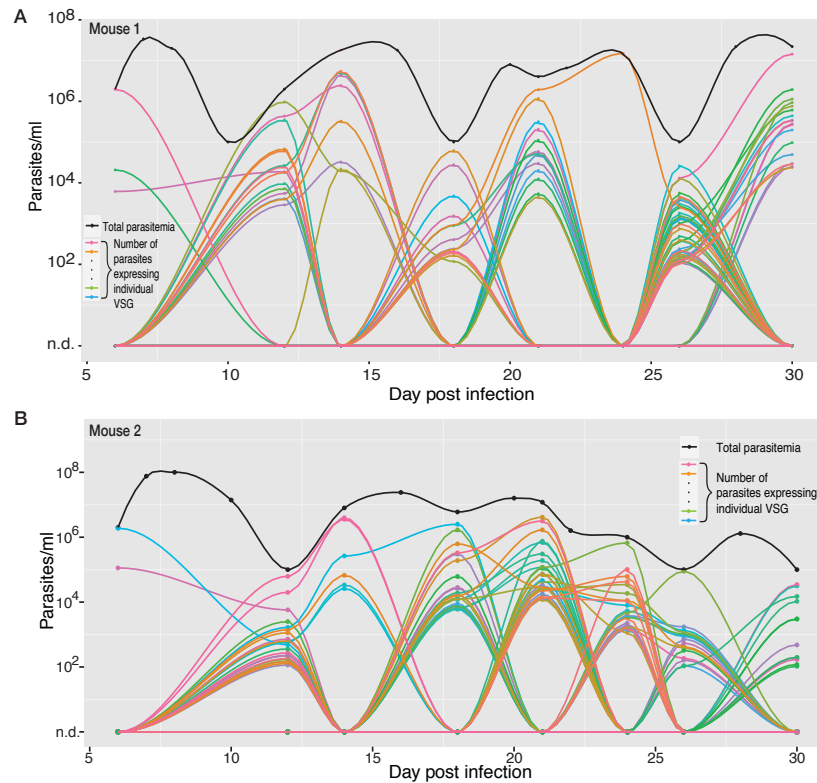
Figure 4.1. Dynamics of *VSG* expression during early infection in (A) Mouse 1 and (B) Mouse 2.

Each colored line represents an individual *VSG's* presence in the population, while the black line represents total parasitemia. Only variants present at greater than 0.1% of the population at that time point are shown. When parasitemia could not be measured by hemacytometer (<106/ml), parasitemia is artificially set at 105/ml to allow for visualization of the population. Note that, because there are so many *VSGs* expressed during infection, colors are difficult to distinguish; overall, variants do not reappear later in the same infection. A smooth curve connects points where expression or parasitemia was measured; these curves are for visualization and do not imply the actual kinetics of variant expression between points.

Figure 4.2. Dynamics of *VSG* expression during early infection in (A) Mouse 3 and (B) Mouse 4.

Each colored line represents an individual *VSG's* presence in the population, while the black line represents total parasitemia. Only variants present at greater than 0.1% of the population at that time point are shown. When parasitemia could not be measured by hemacytometer ($<10^6$/ml), parasitemia is artificially set at 10^5/ml to allow for visualization of the population. Note that, because there are so many *VSGs* expressed during infection, colors are difficult to distinguish; overall, variants do not reappear later in the same infection. A smooth curve connects points where expression or parasitemia was measured; these curves are for visualization and do not imply the actual kinetics of variant expression between points.

Number of VSGs expressed												
Day	d6	d7	d12	d14	d18	d21	d24	d26	d30	d96	d99	d102	d105
Mouse 1	15		18	21	34	36	6	83	33				
Mouse 2	10		34	15	53	57	36	29	39				
Mouse 3		9	13	9	22	21	14	30	21	58	66	30	32
Mouse 4		7	23	17	79	39	12	39	25				

Table 4.1. Variants present during infection.

Table of the number of *VSGs* present at each time point. Any variants quantified as greater than 0.01% of the population are included.
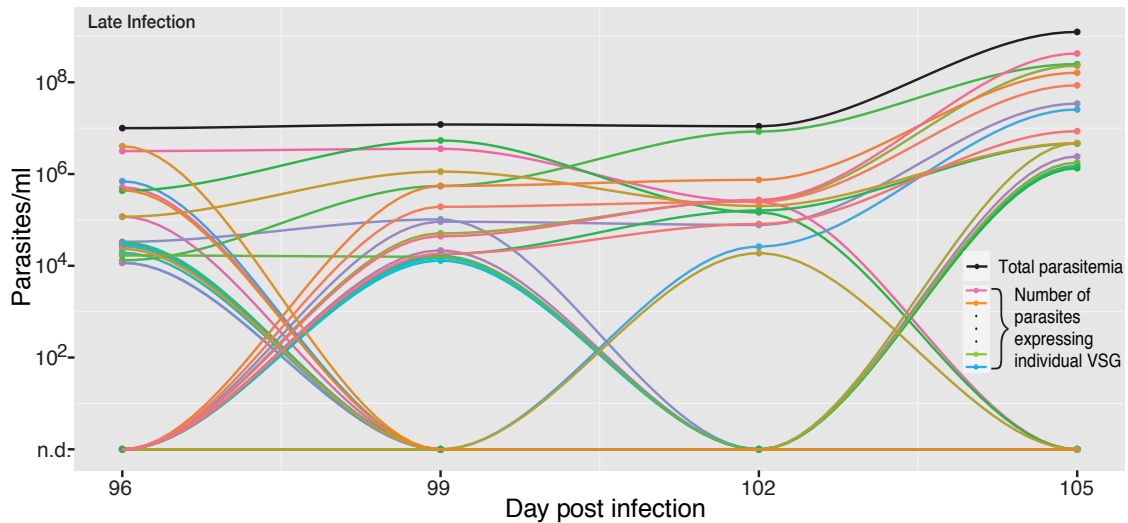
Figure 4.3. Dynamics of *VSG* expression during late infection (d96-105) for Mouse 3.

Each colored line represents an individual *VSG's* presence in the population, while the black line represents total parasitemia. Only variants present at greater than 0.1% of the population at that time point are shown. When parasitemia could not be measured by hemacytometer ($<10^6$/ml), parasitemia is artificially set at 10^5/ml to allow for visualization of the population. Note that, because there are so many *VSGs* expressed during infection, colors are difficult to distinguish; overall, variants do not reappear later in the same infection. A smooth curve connects points where expression or parasitemia was measured; these curves are for visualization and do not imply the actual kinetics of variant expression between points.

4.3 *VSG* diversity and dynamics during late infection

One mouse (Mouse 3) survived much longer than the other three (106 days, compared to 41-72 days), allowing for close analysis of longer-term infection dynamics. Although in the later stages of this infection *VSG* dynamics did appear qualitatively different with variants persisting longer before clearance (Figure 4.3), possibly due to immune system exhaustion, parasite populations remained diverse, with 30-66 variants detectable at each sampling (Table 4.1). The increased survival and lower diversity in this mouse are also likely due to the polyclonal germline B cell repertoire of this mouse[125], rather than the initiating *VSG,* as *VSG* identity has not been shown to affect trypanosome growth rate or induction of the immune response[77,131].

4.4 There is a preference for certain *VSGs* during infection

To see if these infections showed any bias or hierarchy in *VSG* expression[47,78-80], I compared the *VSG* repertoires of all four mice during the first 30 days of infection, in which 192 *VSGs* were expressed in total. Although each infection initiated with a different major *VSG*, the majority of variants (86%) appeared in more than one infection, and nearly half (46%) appeared in all four infections (Figure 4.4A). 97 *VSGs* were expressed in Mouse 3 from days 96-105. I compared later-occurring *VSGs* with those expressed early in Mice 1, 2, and 4, and found none in common, even though early variants from Mouse 3 also appeared frequently in Mice 1, 2, and 4 (Figure 4.4B). Overall, these experiments revealed striking diversity within each infection, but surprisingly frequent occurrence of the same *VSGs* in different infections. Unfortunately, these experiments cannot be used to address the question of order or hierarchy in

switching during infection because they each initiate with different *VSGs* and a

comprehensive assembly of the genomic *VSG* archive for this strain was not finished at

the time of analysis.
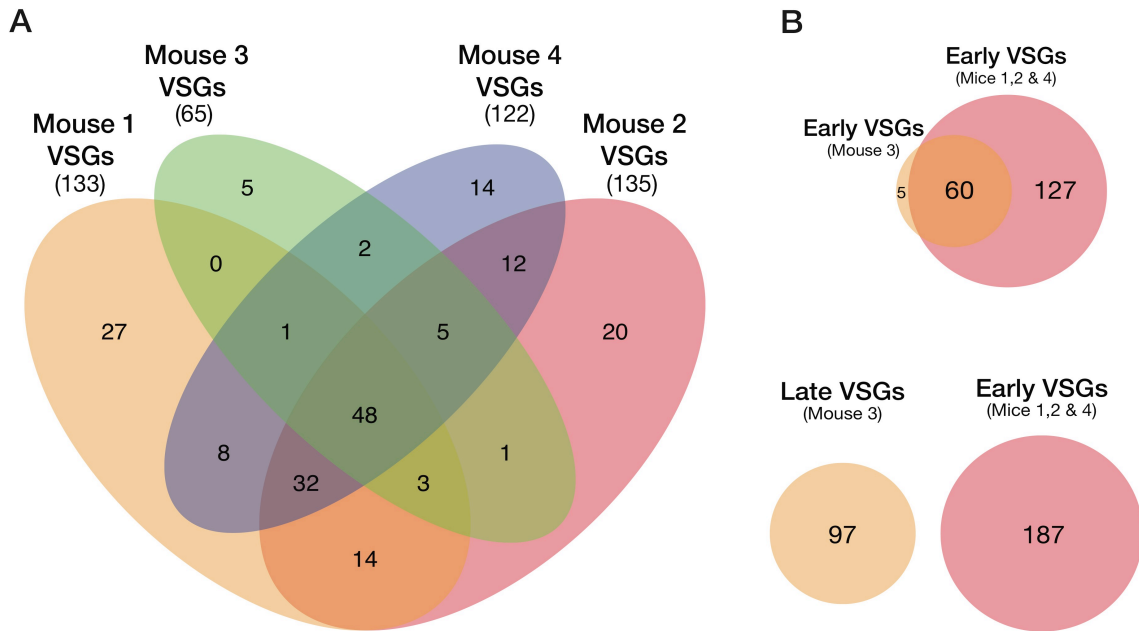
Figure 4.4. Overlap in expressed *VSGs* across infections.

(A) Intersection of sets of *VSGs* expressed during early infection (d6-d30). The total

number of *VSGs* is listed in parentheses below the mouse number. (B) Venn diagrams

showing intersection of *VSGs* expressed early in infection (*VSGs* from Mouse 1, 2, or 4

vs. *VSGs* from Mouse 3, d7-30) and intersection of *VSGs* expressed early in infection

with *VSGs* expressed late in infection (*VSGs* from Mouse 1, 2, or 4 vs. *VSGs* from Mouse

3, d96-105).

4.5 Many expressed *VSGs* never establish *in vivo*

Within these diverse parasite populations, many variants appeared transiently at

low levels. I have termed these 'minor' variants. By examining the fate of every *VSG*, I

found that at any time during the first 30 days of infection about half (53%) of the

54

variants present never reach 1% of the population (Figure 4.5A). This mirrored the observation of failed variants during acute infection described in Chapter 3. The frequency with which variants failed suggested this could be an important phenomenon affecting the measureable outcomes of *VSG* switching, and led to the question of what determines *VSG* fitness.
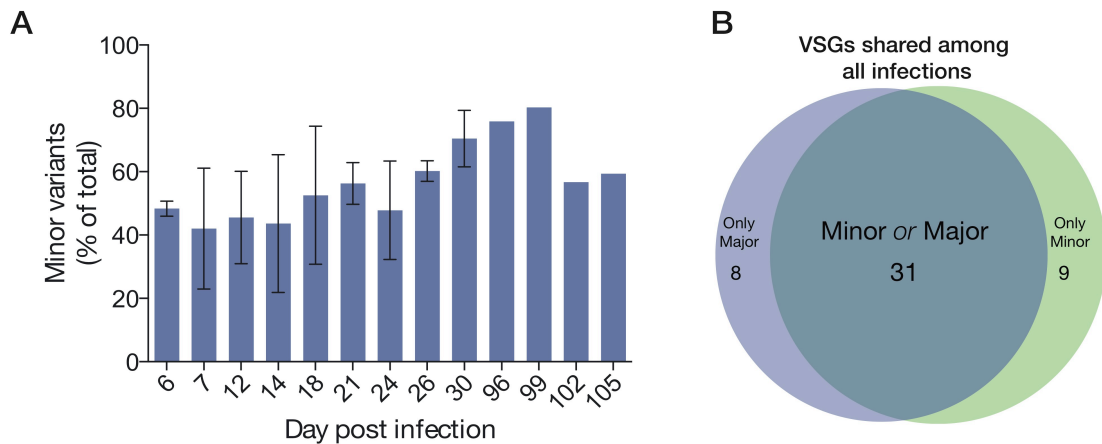
Figure 4.5. Emergence of minor variants during infection.
(A)Minor variants present at each time point (mean +/- SD). A minor variant is arbitrarily defined as any *VSG* that never exceeds 1% of the population during the course of infection in a single mouse. Major variants are any variant that exceeds 1% of the population at some point during infection. (B) Venn diagram comparing the fates of *VSGs* appearing in all four infections.

To investigate the possibility that certain *VSGs,* when expressed, conferred some intrinsic fitness to parasites within the population, I compared the fates of *VSGs* that appeared in all four infections. Of the 48 *VSGs* that appeared in all infections, few were consistently dominant and few were only ever expressed as a minor variant (Figure 4.5B). This implies that variant success is not determined by the expressed *VSG* alone. Instead, variant success is likely to be determined by interactions between the parasite

and the humoral immune response in each animal. However, there were a handful of

VSGs that were consistently either major or minor. It is possible that these variants do

confer some fitness advantage or disadvantage that overcomes the effects of immune

cross-reactivity or the germline B cell repertoire.

4.6 Detection of mosaic *VSGs*

T. brucei's genomic *VSG* repertoire consists of a high proportion of incomplete

VSG genes or pseudogenes[26,27]. Indeed, the 289 *VSGs* observed in these infections may

represent over half of the complete *VSG* repertoire (~400 complete and predicted to be

functional *VSGs* for the Lister427 strain[26], and ~500 for EATRO1125, according to a

recent assembly of the VSGnome for this strain completed in collaboration with George

Cross). The 65-135 *VSGs* observed before day 30 could represent up to one-third of the

pre-existing repertoire. Thus, there is pressure on the parasite to diversify its genomic

repertoire to sustain infection, namely by the formation of mosaic *VSGs*.

Though previous studies had noted that mosaics tend to arise later in infection, or

at least after about 3 weeks of infection in a mouse, they did not determine when these

variants form within the genome, or how[61,64,76]. It is unknown whether mosaic *VSGs* form

at the active expression site or within the silent repertoire prior to expression. To identify

possible mosaics, I compared expressed *VSG* sequences to two independently assembled

genomes for this parasite graciously made available by Keith Matthews and Isabel Roditi.

Because of limitations in the amount of material available at each time point, I could

choose only a few candidates for validation. Candidate *VSGs* were identified based on

three criteria: (1) <80% of the *VSG* length aligned to any sequence in either genome,

which suggested a novel variant absent from the parent genome; (2) the *VSG* was expressed in only one infection, which provided evidence the *VSG* resulted from a novel recombination event; and (3) the *VSG* was expressed at >0.1% of the population, meaning it would be more easily detected by PCR. To test that these were true mosaics and to determine when they formed within the genome, I used VSG-specific primers to confirm their absence from the genome of the parental strain and presence within gDNA collected during infection. I identified three mosaic *VSGs* using this approach. In each case, the mosaic *VSG* was only detectable by PCR when it was also being expressed within the parasite population (Figure 4.6). With the caveat that parasites were not sampled daily, this suggests that mosaic formation occurs, at least in these cases, shortly before expression with subsequent transposition into the active expression site, or directly within the active expression site.

Figure 4.6. Identification of mosaic *VSGs*.

(A) Transient expression of a mosaic *VSG* in the population, and PCR confirmation of the mosaic is shown below. The black line represents total parasitemia at each day post infection, and the green line represents the number of parasites expressing the mosaic *VSG*. "n.q." (not quantifiable) indicates that the *VSG* is detectable within the population, but not quantifiable. "n.d." indicates that the *VSG* is not detectable within the population. Below the graph are products from PCR of gDNA at each time point, using either primers specific for the mosaic *VSG* or the control gene, *ura3*. This *VSG* could not be amplified when first detected by VSG-seq, likely because of low cell numbers in the DNA sample (probably less than 10 cells). (B) and (C) depict mosaics from late infection with PCR confirmation of the mosaic shown below.

A Mosaic VSG c5477_g4_i1_d24_1624 (Mouse 2)

B Mosaic VSG c754_g1_i2_d105_1452 (Mouse 3)

C Mosaic VSG c370_g1_i1_d102_1713 (Mouse 3)
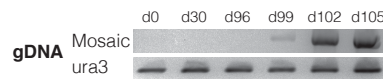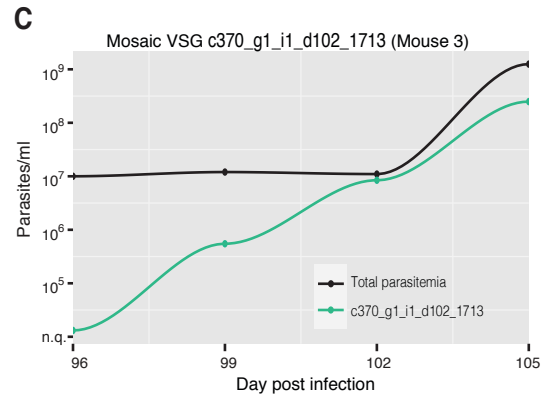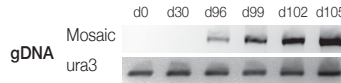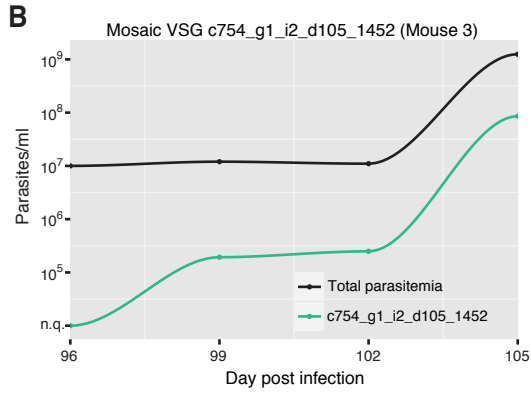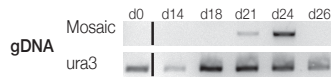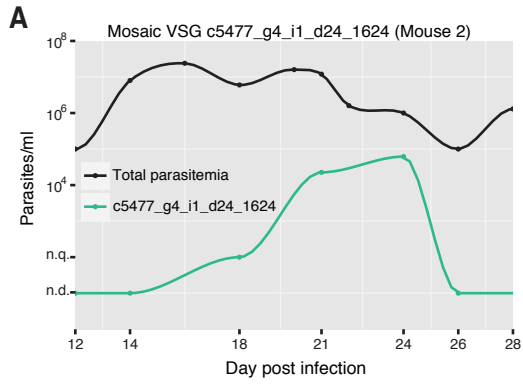
4.7 Discussion

My results indicate that *VSG* switching does not occur at a rate just sufficient for immune evasion, with only a few variants present at any time. Instead, as suggested in the previous experiments in acute *T. brucei* infections, high *VSG* diversity seems to be a universal aspect of infection that persists throughout infection. This was suggested previously but never measured directly[64]. In contrast to experiments in the Lister 427 strain, however, diversity in EATRO1125 infections appears to be much higher. Whereas 9 *VSGs* were observed at the first relapse peak in acute infections, an average of over 28 variants were observed at any point during a chronic infection. It is hard to say whether this difference in diversity is meaningful, as only a handful of mice have been analyzed so far for either strain, but the difference could be due to differences in switching rates, as it has been suggested, though never concretely quantified, that pleomorphic parasite strains producing chronic infections switch at a higher rate than lab-adapted strains[88].

Beyond simple *VSG* diversity, antigenic similarity between VSGs may result in the quick elimination of certain variants by cross-reacting antibodies, rendering such variants useless, and making the effective *VSG* repertoire even smaller than the intact genomic repertoire would suggest. Thus, recombinatorial mechanisms that expand the pre-existing *VSG* repertoire may be critical for sustaining a chronic infection, and mosaic *VSG* formation is likely necessary to increase repertoire diversity as infection progresses. My results show that much of the intact *VSG* repertoire is likely to have been expended early in an infection, as a result of expression and subsequent recognition by the immune system. Therefore the pre-existing repertoire of complete *VSGs* within the genome is probably insufficient to support the sometimes years-long infections observed in the

wild[65]. Although parasitemia is much lower in natural hosts, pre-existing immunity is common in native mammals[132], providing an alternative pressure to diversify the *VSG* repertoire.

In addition to highlighting the importance of mosaic *VSGs* in a chronic infection, these experiments demonstrate that VSG-seq can be used to successfully identify these novel variants. More importantly, these experiments give the first mechanistic insight into mosaic formation. By tracking the presence of a mosaic *VSG* in the genome over time, I found that mosaic *VSGs* probably form primarily within the active expression site, rather than elsewhere in the genome followed by subsequent transposition as a complete and functional *VSG* into the BES, because the mosaic *VSGs* identified were never detectable at the genomic level before they could be detected at the RNA level. Due to the lack of a well-annotated reference VSGnome, it was impossible to determine the patterns of mosaic formation, but this will be an interesting question in the future. Where did recombination occur in these examples of mosaicism? Based on the location of the primers used for confirmation, it appears that all of the mosaic *VSGs* identified formed through recombination in the 3′ end of the *VSG*. This mechanism of mosaicism has been observed before[59,60,63,133] and is referred to by some as 3′ donation[64]. The fact that all mosaics identified might have formed by this mechanism is interesting, as it suggests that the majority of recombination events could occur in the presumably shielded C-terminus of the VSG, which may or may not play a role in immune recognition. What, then is the purpose of such *VSG* diversification? It will be interesting to follow up on these experiments and determine whether the introduction of mutation into the C-terminal domain of the *VSG* is ever capable of conferring resistance to host antibody.

Finally, these experiments revealed a significant preference for parasites to switch to certain *VSGs,* another aspect of switching long speculated to be true[47,78-80]. The infections described in this chapter, however, were each initiated with a different *VSG*. This suggests that preference is at least somewhat independent of the initiating *VSG*, but also makes a detailed analysis of order in *VSG* switching, if any such order exists, very difficult. A specific investigation of order in *VSG* switching that takes advantage of an annotated VSGnome for this parasite strain will be needed to tease out the determinants of hierarchy in *VSG* switching.

CHAPTER 5. Further applications of VSG-seq

5.1 Introduction

VSG-seq revealed interesting aspects of *VSG* dynamics *in vivo* in experimental mouse infections, but one of the most exciting aspects of the approach is its broad applicability. It can be used on any *T. brucei* population, including parasites grown *in vitro*, parasites from natural infection, or even parasites within other niches outside of the bloodstream. A fast, high-resolution approach to profiling any of these populations was previously unavailable. In this chapter, I will discuss further applications of VSG-seq and the mechanistic insight these applications can provide.

5.2 VSG switching *in vitro*

VSG-seq can be used to assess switching outcomes for experiments performed in culture. Recently, a MACS protocol[49] in which parasites are coated in anti-VSG magnetic beads has been used for isolating populations of switched parasites *in vitro*. The antibody used is typically against the dominant or starting *VSG* in a population, retaining parasites expressing this *VSG* on the magnetic column while switched parasites are collected in the flow-through. The best way to measure switching outcomes in the isolated switcher population, however, has been through the laborious process of cloning individual switched cells, followed by sequencing of VSG cDNA isolated from each clone. In collaboration with Galadriel Hovel-Miner, I tested VSG-seq for measuring outcomes in an experimental switch population as part of a project investigating the role of 70-bp repeats in antigenic variation in *T. brucei*.

Earlier work from several labs had led to the hypothesis that 70-bp repeats play a role in antigenic variation. This prevailing, but unproven, hypothesis speculated that these repeats provided recombinatorial homology that facilitated the "choosing" of a *VSG* donor sequence in switching[35-38]. My collaborator had already determined, using a number of mutant parasite lines, that this repeat sequence, adjacent to most *VSGs,* indeed plays a role in selection of non-BES *VSGs* during a *VSG* switch. She found that mutants lacking the 70-bp repeats at the active BES (Δ70) are restricted in their ability to access the entire repertoire and switch only to *VSGs* located at other telomeric BESs (Hovel-Miner, et al, in submission). Additionally, this Δ70 mutant switched very frequently, at levels orders of magnitude higher than a wild type strain *in vitro.* She had determined that introduction of a minimal repeat, consisting of a 70-bp repeat dimer, reduced levels of switching to wild type levels, whereas a reversed dimer sequence at the active BES switched as frequently as a Δ70 mutant. The key experiment to demonstrate the role of this repeat with respect to *VSG* donor selection, however, needed to show that a dimer allowed access to the non-BES archive, while a reversed dimer sequence did not. Previously, this would have required additional rounds of laborious cloning and sequencing of individual *VSGs* after isolation of the switched population by MACS.

VSG-seq enabled us to assess repertoire access by the two mutants in a faster, less biased, and higher-resolution way. I applied VSG-seq to switched populations isolated by MACS from biological triplicates of each line to determine the switching outcomes for each. I observed that the line lacking repeats in the active BES (Δ70) could occasionally select *VSGs* from sites other than BESs (Figure 5.1). This was in contrast to previous data using cloning to identify switching outcomes, which showed absolutely no selection from

sites other than BESs, and highlights the increased resolution of VSG-seq. Nevertheless, introduction of the repeat Dimer resulted in an increase in the number of *VSGs* selected when compared with the no-repeat line (average of 18 *VSGs* in Δ70-ISceI and 35 *VSGs* in Dimer). This increase in diversity of *VSG* selection appears to be due to an increase in selection of MC *VSGs*. The Δ70 strain selected 1 MC *VSG*, whereas the Dimer strain selected an average of 11 MC *VSGs*. There was an additional, albeit smaller, increase in the selection of *VSGs* arising from undermined loci. In contrast, the strain carrying a reversed dimer sequence did not result in a significant increase in the selected *VSG* repertoire. These data confirmed that two 70-bp repeats at the active BES are sufficient to allow selection of *VSGs* from non-BES sites, and that the 70-bp repeat sequence is required for this selection, as a reversed Dimer sequence prevents the selection of *VSGs* outside of BESs. Thus, VSG-seq can facilitate the quick profiling of switching outcomes towards an understanding of switching mechanisms *in vitro*.
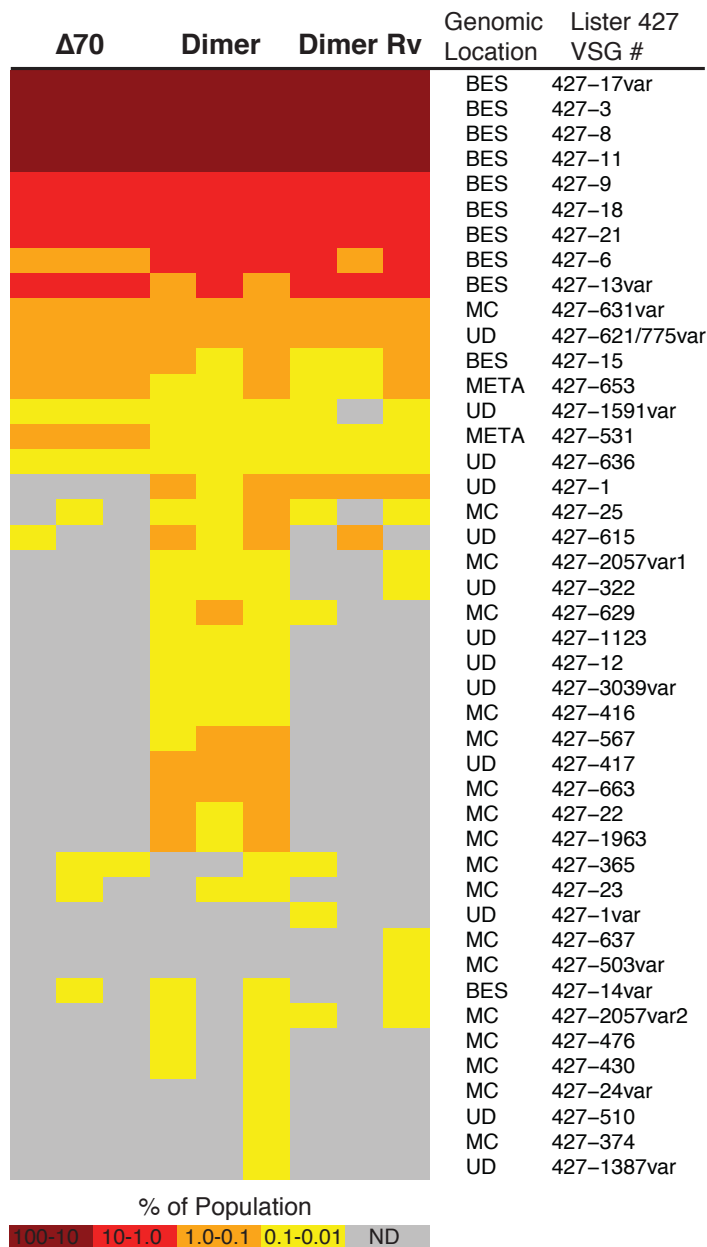
Figure 5.1. VSG-seq analysis of 70-bp repeat mutant strains.

Switching outcomes in populations of MACS-isolated switchers from three biological replicates of Δ70, Dimer, & Dimer Rv are shown in a heat map, where color intensity reflects the proportion of each *VSG* in the population. The Lister 427 *VSG* number and its predicted genomic location are shown on the right. "var" indicates that the assembled *VSG* sequence had minor sequence variations from the most similar 427 reference *VSG*.

5.3 Antigenic variation field isolates

One of my primary goals in developing VSG-seq was to develop a method that could be applied to *T. brucei* field isolates. In collaboration with Veerle Lejon and Stijn Deborggraeve, I tested the feasibility of applying VSG-seq to samples collected from natural human infections. The samples analyzed were collected from parasitologically confirmed stage 1 or stage 2 HAT patients in the Democratic Republic of the Congo.

RNA was extracted from 2.5 mL of blood from each patient. An estimate of parasitemia for each patient was made microscopically by examining the buffy coat of 10ml of blood, and by QPCR[134] on the isolated RNA. QPCR was used to infer the relative parasitemia of samples and did not provide an absolute measure of parasitemia. All samples analyzed, except for one (RB0017S), showed >50 parasites microscopically. RB0017S showed 21-50 parasites in the isolated buffy coat, and its relatively lower parasitemia was confirmed by QPCR. Though sensitive for diagnosis, visual counting of parasites concentrated in the buffy coat is not particularly accurate for estimating parasitemia[135], and these counts do not indicate a parasitemia of 5 parasites/mL. Rather, these results show that all patients analyzed presented with relatively high parasitemia for infection with *T. b. gambiense.*

The *VSG* PCR step of the VSG-seq library preparation was somewhat inefficient in these samples, producing many nonspecific bands. This was probably due to the high level of host RNA present in each sample, relative to parasite RNA. To improve enrichment for *VSG* sequence, an alteration to the library preparation was made in which PCR products ranging in size from 1kb to 2kb were gel extracted and used as input material for creating sequencing libraries, rather than using all PCR products.

This approach appears to have worked, as *VSG* sequences were assembled from all five samples analyzed. Analysis parameters were difficult to set, however, due to the unknown parasitemia in each sample. To be conservative in my analysis, I only counted a *VSG* as present within the population if it was present at >1% of the population. In *T. b. gambiense* infection, parasitemia can be as low as <100 parasites/mL[135], though this is unlikely given the estimates of parasitemia in these samples. If parasitemia were indeed <100 parasites/mL, a detection cutoff of 0.1% could result in significant overestimations of diversity.

Using this relatively conservative cutoff, *VSG* diversity in each sample ranged from 2 to 7 *VSGs* (Figure 5.2A). Using a detection cutoff of 0.01%, which assumes at least 10,000 parasites per sample or 4,000 parasites/ mL of blood, diversity ranges from 7 to 11 *VSGs* per sample. This is somewhat lower than observed in mouse infections but perhaps unsurprising, given the low parasitemia typical of *T. b. gambiense* infection, and the likelihood of pre-existing immunity against certain variants in these patients[132].

A comparison of expressed *VSGs* showed some overlap between samples (Figure 5.2B). RB0011S and RB0030S contain the same two *VSGs*, and one *VSG* is in common between RB0011S, RB0017S, and RB0030S. Interestingly, there was no overlap between RB001S or RB002S and the other samples.

It is important to consider the relationships between each sample when interpreting the overlap in expressed *VSG* repertoires. RB001S and RB002S came from the same village around the same time, while RB0011S and RB0017S were isolated around the same time in the same clinic. RB0030S was the outlier in this set, collected far from the other samples. With respect to time, all samples were collected within a one-

month period. The overlap between *VSG* repertoires is interesting considering the relationships between these samples; those with the most in common are not necessarily those with any temporal or geographic relationship.
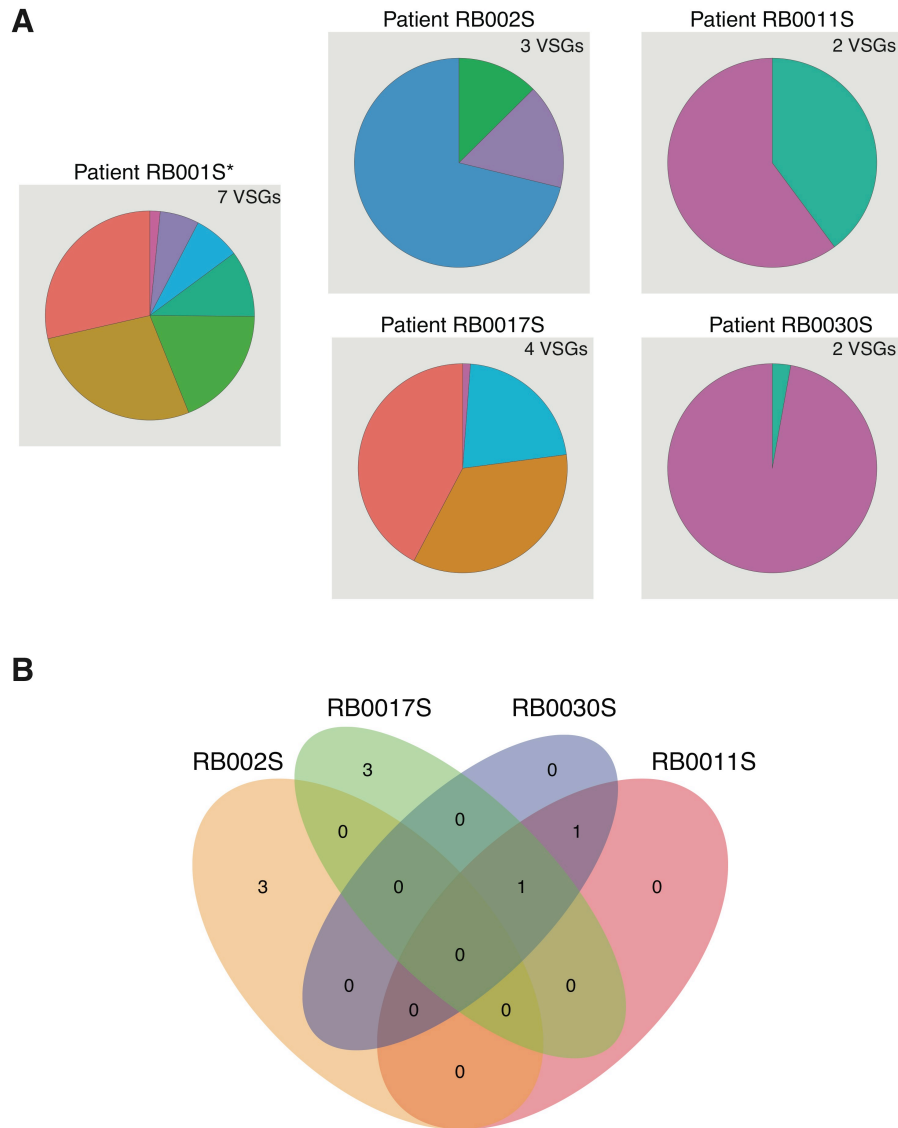
Figure 5.2. VSG-seq analysis of human *T. gambiense* infections.
(A) Pie charts showing *VSG* expression in each sample, using a 1% detection cutoff. Colors in each chart represent the same *VSGs*, with the exception of the chart for Patient RB001S (*), which was analyzed separately and has no *VSGs* in common with the other samples. (B) Venn diagram showing overlap in *VSG* repertoire between patients RB002S-RB0030S.

Analysis of the identity of *VSGs* expressed in each sample also revealed something unusual: the most abundant *VSG* identified in patient RB0017S is identical to a common BES-associated Lister 427 *VSG*, VSG-11. Using less strict analysis parameters, this *VSG* also appears in RB002S (0.02%) and RB0030S (0.7%). No other common *VSGs* were identified in this analysis. The significant overlap in expressed *VSG* repertoires between these infections may provide hints at both the diversity of *VSG* repertoires between parasite subspecies as well as the extent of the expressed *VSG* repertoire in the wild.

There is of course, a major caveat to these observations, which is that it is not yet clear whether using such small starting samples affects quantification. In these samples, the amount of parasite RNA is low, and host RNA is in significant excess. Control experiments, along with possible optimization of the method, will need to be performed to show that high levels of contaminating RNA do not affect quantification. Although more experiments are needed to shore up these findings, they suggest not only that VSG-seq will work on samples isolated from natural infections, but also that there are interesting things to be found in wild *T. brucei* populations. The genomic repertoire has been compared between some species and subspecies of African trypanosome[136,137] but not the expressed repertoire, which these data show may be distinct and important.

5.4 Antigenic variation in extravascular parasite populations

Finally, VSG-seq provides the opportunity to investigate parasites beyond those found in the bloodstream of the host. Even in the earliest days of *T. brucei* research, it had been recognized that parasites could be sequestered in specific tissue niches, the

central nervous system (CNS) in particular, where they could be protected from chemotherapy, to later re-enter the bloodstream and re-establish infection[138,139]. Although it had been thought that *T. brucei* existed primarily in the bloodstream until late in infection, when it invades the CNS and neuropathology appears in patients, recent work in mice shows that parasites invade extravascular sites throughout the body after only 6 days of infection. In addition to the CNS, parasites can be found in the heart, lung, liver, spleen, and adipose tissue and, in all cases, parasites are found extracellularly in the interstitium of these tissues. Surprisingly, by day 28 p.i. this extravascular population represents the majority of parasites in the mouse (Luisa Figueiredo, in submission). Parasites residing in adipose tissue have been found to take advantage of novel metabolic pathways to utilize fat as a nutrient source, while CNS-resident parasites are morphologically different from the well-characterized bloodstream form *T. brucei*[9]. The unique properties of parasites residing in other tissues have not yet been characterized, but together these results suggest that the extravascular niche is not simply a side effect of high parasitemia at the late stages of infection, as once thought. Instead, it is likely that this niche is an important phase of the parasite's lifecycle, one that has not yet been characterized with respect to antigenic variation.

VSG-seq provides an opportunity to examine the expressed *VSG* repertoire within extravascular populations. With this in mind, in collaboration with Luisa Figueiredo, I performed a pilot experiment on samples collected from two mice infected with EATRO1125 *T. brucei* expressing the *VSG* AnTat1.1. One set of samples was collected from day 6 p.i. and the other from day 28. In both cases, blood was first collected,

followed by perfusion of the animal with PBS. RNA was isolated from blood and from brain, fat, heart, and lung. This material was used for VSG-seq.

Analysis of these samples showed the number of *VSGs* expressed in each tissue did not correlate with the number of parasites present in that tissue (Figure 5.3A), though generally diversity was higher on day 28 than day 6, consistent with previous observations. The number of *VSGs* expressed in parasites in blood was somewhat higher than previously observed, but this was likely due to the fact that these infections were initiated with 5000 parasites, rather than only 5. It is possible, however, that this difference in diversity is the result of differences in the host mouse strain, as these experiments were performed in C57Bl/6 mice, rather than the Balb/c mice used in Chapter 3 and 4.

Comparison of *VSG* populations in each tissue revealed a complex situation within the host. On both day 6 (Figure 5.3B) and day 28 (Figure 5.3C) *VSGs* could be found that were unique to an individual tissue. Some tissues seemed to contain more unique *VSGs* than others but, with only two infections, it is hard to interpret these data. Nonetheless, the high diversity across tissues can be taken as evidence that this is not well-mixed system; a simple survey of parasites in the blood is not representative of the tissue niche.
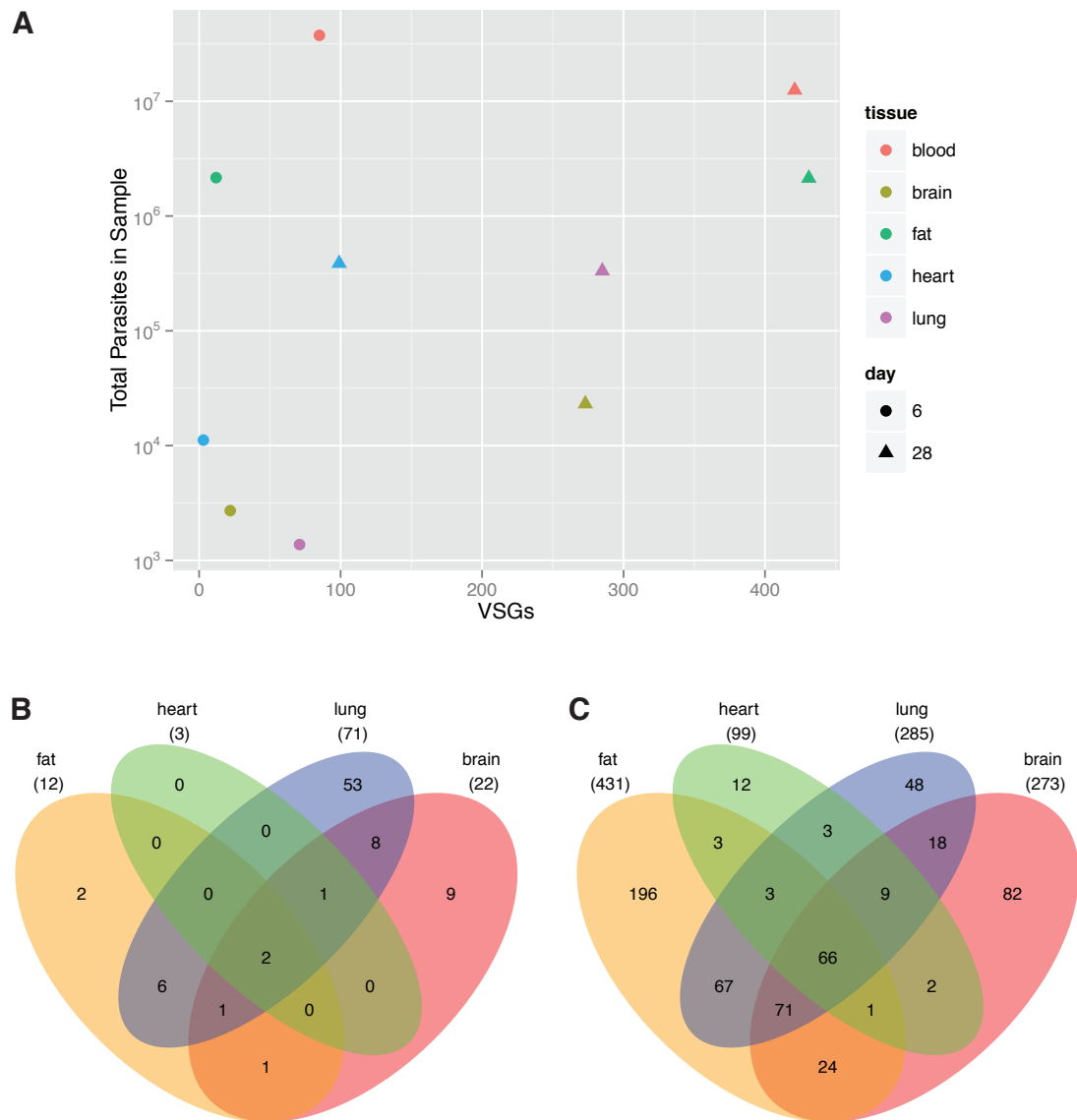
Figure 5.3. *VSG* expression in extravascular spaces.

(A) Scatter plot showing the number of *VSGs* expressed in each tissue compared to the number of parasites in the sample analyzed. (B) Venn diagram showing overlap in expressed *VSG* repertoire between each tissue on day 6 post-infection. (C) Venn diagram showing overlap in expresses *VSG* repertoire between each tissue on day 28 post-infection.

73

Comparison of *VSG* populations in each tissue to those in the blood may hint at possible roles for this extravascular niche. VSG-seq analysis of perfused tissues on day 6 p.i. showed that while nearly every *VSG* expressed in other tissues was also present in the blood, over half of the *VSGs* identified in the lung were unique to that space (Figure 5.4). Could the lung be a site for initiating *VSG* switching or diversification? While the situation on day 28 of infection was more complicated (Figure 5.4), there was one striking observation resulting from this comparison: while the initiating *VSG* (AnTat1.1) had been cleared from the blood and the heart, it persisted in the brain, fat, and lung. In fact, the initiating *VSG* was expressed in 1.4% of all parasites in the brain. Thus, certain extravascular spaces appear to be immune-privileged, which could be important for parasite dynamics *in vivo*.
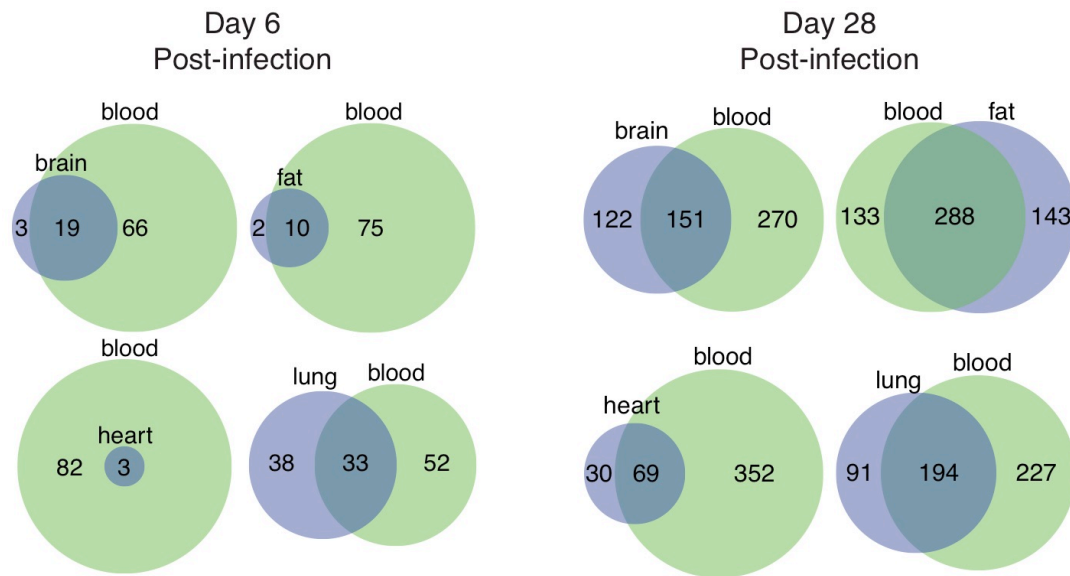
Figure 5.4. Overlap in *VSG* repertoire between extravascular spaces and blood. (A) Venn diagrams showing the overlap in expressed *VSGs* between each tissue and blood on day 6 post-infection. (B) Venn diagrams showing the overlap in expressed *VSGs* between each tissue and blood on day 28 post-infection.

5.5 Discussion

These experiments show that VSG-seq can be applied successfully to many types of samples, and in all cases has the potential to provide insight into the mechanisms of antigenic variation in *T. brucei*. My experiments examining the role of 70-bp repeats in *VSG* donor selection in antigenic variation confirmed previous observations, but provided much higher resolution than possible using previous approaches. Moreover, the method is much less laborious than other strategies for profiling switch repertoires.

Studies of *VSG* repertoires in wild isolates provide evidence that this approach could be easily applied to any natural infection. Although some optimization and validation will be needed, the findings from this small experiment hint at interesting possibilities. *VSGs* known to be present in *T. b. brucei* were found in infections presumed to be *T. b. gambiense*. This could suggest co-infection, the transfer of genetic material between parasites[6,7], or that the expressed *VSG* repertoire is, for some other reason, somewhat constrained. An understanding of the extent of and limitations on the wild *VSG* repertoire could affect our understanding of the epidemiology and molecular biology of this infection in the wild, as well as our approaches to treatment and prevention of African trypanosomiasis.

Analysis of parasites in extravascular spaces in experimental mouse infections reveals a new aspect of *T. brucei* antigenic variation. Based on these data, is it clear that the population in the bloodstream of an animal is not representative of the entire parasite population *in vivo*. This could have significant implications for parasite biology. For example, the presence of many unique variants in the lung early in infection may imply a role for this site in antigenic variation. The persistence of certain variants after four

weeks of infection indicates an immune privileged niche in certain extravascular spaces that could serve as a refuge for switching *in vivo*. Taken together, the set of experiments described in this chapter demonstrates the promise of high-throughput methods for tackling many questions in the field of *T. brucei* antigenic variation.

CHAPTER 6. Discussion

6.1 Development of VSG-seq

An understanding of the dynamics of *VSG* expression and switching is an important step towards understanding the mechanisms driving immune evasion in *T. brucei* infection as well as the host response to *VSG*. For technical reasons, however, it has been challenging to measure *VSG* expression *in vivo* in a high-resolution way. Some have tried to address the question of *VSG* dynamics in other ways, either by mathematical modeling[5,78,81,82,84,85] or by using low-throughput approaches[47,64,79,80,87,88]. While our understanding of the genomic *VSG* repertoire has improved greatly in recent years due to next-generation sequencing approaches, our understanding of how that repertoire is accessed *in vivo* has not yet benefitted from more modern techniques.

To address this gap in our understanding of *VSG* switching, I developed a next-generation sequencing approach, termed VSG-seq, for measuring *VSG* expression in high resolution. Based on *de novo* assembly, the method obviates the requirement for a reference genome, and allows for the identification and quantification of *VSG* expression from any strain of *T. brucei*. The method, orders of magnitude higher in resolution than previous approaches, can detect *VSGs* expressed on as few as 0.01% of the parasites in a population.

6.2 Confirmation of existing hypotheses

Using VSG-seq, I profiled *VSG* expression in both acute and chronic mouse infections. My results confirmed many predictions made based on low-throughput studies. Indeed, the predictions made by Hall, et al.[64], of an average of 32 and as many as 95 *VSGs* at any time were extremely accurate: I measured an average of 28 and as many as 83 *VSGs* expressed as one time. It is telling that these estimates are somewhat higher than my measurements, as the limit of detection of VSG-seq prevents the method from sampling every *VSG* in a population. Moreover, recent work on samples collected from sleeping sickness patients[92] along with my own results from sequencing patient samples suggest that this *VSG* diversity is not simply an artifact of the mouse model and may be clinically relevant.

In line with predictions of a hierarchy in *VSG* switching, I observed that *VSG* populations in chronic infections were distinct at different stages of infection and that there was a clear preference for expression of certain *VSGs* during the first 30 days of infection. Additionally, my work again confirmed the observation that mosaic *VSGs* dominate only after the first peaks of infection, as I was only able to detect these variants on day 18 p.i. or later.

These support previous predictions and lay the foundation for a more in-depth study of switching hierarchy and order. Because the chronic infections in this study were each initiated with a different *VSG*, explicit questions of order could not be addressed. I hope to address these questions explicitly in the near future, however, by initiating sets of infections with specific *VSGs* and comparing the switching outcomes. This of course will require the generation of a good catalogue of BES-associated and genomic *VSGs* in the

EATRO1125 strain. Once generated, such a catalogue can be used to assess the relationship between *VSG* order and genomic position. Are the same set of *VSGs* switched to no matter what *VSG* begins an infection, consistent with a model in which only genomic location matters, or is preference relative to the starting *VSG*, supporting a model in which homology in or around the expressed *VSG* plays a larger role? This experiment should determine whether hierarchy is independent of initiating *VSG*, as the experiment described in Chapter 4, comparing infections beginning with different variants, suggests.

6.3 New questions about antigenic variation: Minor variants

In addition to lending to support to existing hypotheses, my work brought to light aspects of infection that could not have been accessed using other approaches. During infection, many *VSGs* are expressed at a single time, but most of these variants never establish to levels at which they could have been detected using low-resolution techniques. The presence of these minor variants both supports the hypothesis that immune cross-reactivity plays a role in determining variant order during infection[85,86], as well as the concept that the complete *VSG* repertoire (~400 *VSGs* in Lister 427, ~500 in EATRO1125) is insufficient to support the long *T. brucei* infections observed in the field. Many of those 400-500 variants fail too quickly to contribute significantly to immune evasion. Thus, *VSG* diversification through the formation of mosaic *VSGs* or other yet-to-be identified mechanisms may play a more significant role in sustaining infection than previously appreciated.

Perhaps most interestingly, the identification of minor variants raises a new question relevant to understanding *VSG* expression dynamics: what determines variant fitness? My data suggest that fitness is not solely intrinsic to the *VSG*. Although I have proposed antibody cross-reactivity or the individual host's unique germ line B cell repertoire to be factors that determine variant success or failure, this has certainly not been proven. To address this hypothesis, one could examine, for example, the likelihood of an individual variant to be minor after the expression of an antigenically similar *VSG* during the same infection. If the host immune system is primed to recognize a variant because an earlier VSG was antigenically similar, that variant should not establish in the population as successfully as the previously expressed VSG. Addressing this experimentally would require an understanding of antigenic "classes" of VSG based on the ability of VSGs that are different in primary sequence to elicit similar antibody responses. There are only two VSG crystal structures, however, so it is difficult to predict which variants are likely to be similar at the structural level based on sequence alone. Currently, little is known about how antibody binds VSG, but patterns of VSG success and failure *in vivo* might inform our understanding of the antibody-VSG interaction. Perhaps pairs of variants in which one frequently "fails" when expressed during the same infection as the other are in fact structurally and antigenically similar.

6.4 New questions about antigenic variation: The wild *VSG* repertoire

The identification of minor variants in the bloodstream during infection raised new questions about the biology of antigenic variation in *T. brucei*, but the application of VSG-seq to patient samples and extravascular parasites raised even more. The study of

field isolates presented in Chapter 5, although quite preliminary, has potentially

interesting implications for the wild *VSG* repertoire. The expressed repertoire of *VSGs* in

the field could be important for understanding both parasite ecology as well as host

immunity. My data show significant overlap in expressed *VSG* repertoires between

geographically distinct infections. Does this mean the expressed *VSG* repertoire in the

field is actually quite limited? If so, what restricts the repertoire to such a degree? Could

host immunity play a role in the detectable expressed *VSG* repertoire?

Additionally, my data show one *VSG* that is expressed in at least one patient, and

possibly three, which is BES-associated in the Lister 427 strain of *T. b. brucei*. This again

suggests a limited *VSG* repertoire, but in this case across *T. brucei* subspecies. Does this

reflect a shared ancestral *VSG* repertoire, sexual recombination between parasite

subspecies in the tsetse, or coinfection?

Perhaps this limited repertoire is unsurprising. One of the most common HAT

field diagnostics is the card agglutination test for trypanosomiasis (CATT) [135,140,141],

which diagnoses patients based on the presence of antisera against a three *VSGs*

commonly expressed by parasites in West Africa. Thus, the repertoire regionally appears

to be constrained enough to use the presence of an anti-VSG antibody response for

diagnosis. Interestingly, however, these variants do not predominate throughout Africa,

and one study showed the absence of the gene encoding one of these variants in some

trypanosome strains isolated in Cameroon[142]. Thus, an understanding of the wild

expressed *VSG* repertoire and how it changes in different regions could lead to the

development of diagnostics specific for different regions in Africa. Although these data

are quite preliminary, they suggest there is interesting biology happening in natural *T. brucei* infections that will be worth investigating in the future.

Another source of interesting biology in natural *VSGs* repertoires is in the tsetse salivary glands. Because only hundreds of metacyclic parasites need to be injected by the tsetse for transmission[143], and these parasites initiate nearly all infection, could there be a limited inoculating repertoire? Studies of the genomic set of metacyclic *VSGs* show only between 5 and 27 metacyclic *VSGs* in the genome[26,144], and another study shows some evidence that metacyclic *VSGs* are less likely to rearrange to form mosaic variants, further limiting this repertoire[145]. How extensive is the wild metacyclic expressed *VSG* repertoire? If the repertoire is in fact limited, this could have dramatic consequences for therapeutic approaches for HAT, including vaccination.

6.5 New questions about antigenic variation: Extravascular parasites

The study of parasites residing in extravascular spaces in tissues described later in Chapter 6, though once again very preliminary, has significant implications for our understanding of antigenic variation *in vivo*. A surprising number of *VSGs* are expressed exclusively in the lung. This suggests that the lung may be a site of early switching and diversification. If this is the case, it may indicate the existence of a trigger for switching or diversification unique to that tissue. One such trigger could be oxygen tension, which is higher in the lung than in other tissues[146], and which is known to affect the expression of a major surface glycoprotein in procyclic forms of the parasite[147]. Such a hypothesis could be tested easily by varying oxygen concentrations in parasite cultures and measuring the effect on antigenic variation. Of course other factors, such as pH, could

also serve as a trigger for switching. So far, there is no known environmental factor that can initiate *VSG* switching, but these data suggest that one might exist.

An alternative explanation for the increased diversity in the lung is the lack of an antibody response within that tissue, which would preclude specific clearance of any variants and allow parasites expressing antigenically similar *VSGs* to survive and grow. The persistence of early expressed *VSGs* in certain spaces (brain, fat, and lung) long after they have been cleared from the blood supports a model in which increased *VSG* diversity results from immune-privileged sites in the host. This observation, once confirmed, could be followed up by a study of the B cell populations and antibody repertoires, both general and VSG-specific, within each potentially immune privileged tissue. Such experiments might reveal certain antibody isotypes that are present or absent from each site. For example, although IgM is thought to be the primary isotype involved in parasite clearance[96], the role of IgM and other isotypes in the context of extravascular spaces is not known. Thus, it is possible that IgM is required for parasite clearance in tissues and that early variants persist in certain spaces because the bulky IgM molecule cannot diffuse efficiently into those spaces. On the other hand, if VSG-specific antibody or B cells can be detected but parasites expressing that *VSG* have not been cleared, it may be that *T. brucei* modulates the host B cell response in these spaces. Very little is known about immune privilege or the extravascular B cell response, but sequencing of the expressed extravascular *VSG* repertoire may have indirectly shed light on both of these phenomena.

Moreover, it might be possible to take advantage of immune-privileged spaces, if they exist, experimentally. Due to antibody cross-reactivity, it is likely that many variants

never establish in the bloodstream. This may be especially true for mosaic *VSGs,* which are more likely to be antigenically similar to a previously expressed *VSG*, particularly if they do form in the active BES. The early clearance of cross-reactive or less fit variants in the blood leads to a skewed view of the *VSG* population *in vivo*. Immune-privileged spaces may provide a better context for studying the mechanisms of switching and *VSG* diversification, as all variants would be detected in such sites.

The application of VSG-seq to tissue-resident parasites has hinted at two potential roles for these spaces. In addition to this, a more extensive study of parasite populations over time in each space could provide information on the presence of variants in each niche over time. This kind of data could then be used to create a map of *T. brucei's* route through the tissues of its host, something about which very little is currently known. Another major unanswered question that would be informed by an extension of these experiments is whether *VSGs* have a role in sequestration of trypanosomes in different physiological niches, as is the case for the *var* gene family in *P. falciparum* infection[148]. Could certain *VSGs* be lung- or brain-specific?

6.6 New answers to old questions: Mosaic *VSGs*

Although many aspects of antigenic variation are informed by this work, nearly every finding in these investigations points towards one theme: *VSG* diversification. The set of expressed *VSGs* at any given time during is large. Of those many variants, half are destined to fail. An examination of the parasites in tissues reveals another layer of diversity. Over and over, it becomes clear that the genomic *VSG* repertoire alone is

insufficient to sustain an infection, and the parasite must diversify its *VSG* repertoire through the formation of mosaic *VSGs.*

The idea that *VSG* diversification is critical to sustaining a chronic *T. brucei* infection is certainly not a new one[88]. Despite the obvious importance of the process, the mechanisms of *VSG* diversification in general, and mosaic *VSG* formation in particular, have remained elusive and experimentally intractable. VSG-seq, then, presents a new avenue for tackling this historically difficult question. My investigation of chronic infections revealed the first temporal data on the formation mosaic *VSGs*, suggesting that these novel variants form within the active expression site. The patterns of mosaic formation suggested by my data also raise interesting questions about the immune response to VSG. What use could mosaic formation in the buried C-terminal end of the VSG serve? It will be interesting to investigate whether mutations in the 3′ of the *VSG* can ever confer resistance to host antibody. The mosaics identified by VSG-seq may be ideal candidates for such analysis, because they are present at relatively high levels and should, to some degree, be capable of evading host antibody. Thus, the mosaics detectable by this method may be not only the variants most relevant to a natural infection, but also the variants most likely to confer an antigenic advantage *in vivo*. Thus, beyond informing antigenic variation in *T. brucei*, they may provide insight into the antibody-VSG interaction.

Taken together, my results show that VSG-seq is not only descriptive. Instead, this approach can provide mechanistic insight into antigenic variation and the creation of mosaic *VSGs*. This high-resolution and quantitative technology can now be used to ask not only whether mosaic *VSGs* form, but when, where, and how. There is also no reason

why this approach cannot be applied to other systems, so questions of repertoire diversification can extend beyond the African trypanosome. Using VSG-seq, I have shown that high-throughput bioinformatics approaches are key to achieving the high-resolution picture of *VSG* switching needed to understand the mechanisms of antigenic variation *in vivo*.

CHAPTER 7. Methods

7.1 Cell Culture

Control libraries (Chapter 2) were made using mixtures of Lister 427 clones each expressing a different *VSG*. Each line has an antibiotic resistance marker at the promoter of the active expression site and was grown *in vitro* in HMI-9 with antibiotic selection to minimize *in situ* switching. To make the libraries, parasites were counted with a hemacytometer and mixed in known proportions. RNA was isolated from these mixtures immediately and used for the preparation of VSG-seq libraries.

7.2 Infections

For infections investigating parasite diversity in blood, female Balb/cByJ (Jackson Labs) mice were infected intraperitoneally with ~5 parasites (pleomorphic or monomorphic), and parasitemia was counted every 2 days by hemacytometer. 50–100 µl blood was collected by submandibular bleed for the isolation of parasites and preparation of sequencing libraries. Mice were sacrificed when they began displaying pathological symptoms.

For infections investigating parasite diversity in tissues, 6-10 week old male C57BL/6J mice (Charles River, France) were infected intraperitoneally with 2000 *T. brucei* EATRO1125 parasites. Parasitemia was tracked daily by hemacytometer and animals were sacrificed on either day 6 or day 28 by CO_2 narcosis.

7.3 Isolation of parasites from blood

Serum was removed from blood samples and saved, and the remaining material was incubated with Ter-119 microbeads (Miltenyi). Samples were then washed twice in HMI-9 and applied to a MACS column. The flow-through was collected and these red-blood-cell-depleted samples were divided for DNA (10%) and RNA (90%) isolation. DNA was isolated from a portion of the parasite sample using DNAzol (Invitrogen), according to the manufacturer's protocol.

7.4 Isolation of tissue-resident parasites

For experiments examining tissue-resident parasites, mice were immediately perfused transcardially following euthanasia with pre-warmed heparinized saline using a peristaltic pump. Approximately 30 mg of each organ was collected into Trizol (Life Technologies).

7.5 RNA isolation

For all experiments except for the investigation of parasites in human infections and in mouse tissues, RNA was isolated using RNA STAT-60 (Tel-Test, Inc.), according to the manufacturer's protocol.

For experiments examining tissue-resident parasites, RNA was isolated from perfused tissue samples with Trizol (Life Technologies) according to manufacturer's instructions. Blood was isolated before perfusion and RNA was isolated from blood samples using Trizol LS (Life Technologies).

For human infection samples, samples were collected from parasitologically

confirmed stage 1 or stage 2 HAT patients in D.R. Congo. 2.5 ml of blood was collected in PAXgene tubes (Qiagen) in the field, which were shipped and stored at -80°C. RNA was extracted using a PAXgene Blood RNA Kit (Qiagen).

7.6 VSG-seq Library Preparations for optimization

Four approaches to VSG-seq library preparation were tested, each with slight variations in the protocol up until the production of fragmented dscDNA. Sequencing libraries were prepared from this fragmented material using the Illumina TruSeq DNA Sample Prep Kit v1 (FC-121-1001) according to manufacturer's instructions. All libraries were sequenced on an Illumina HiSeq 2000 producing 100bp single-end reads.

Prep 1: "VSG"

For this prep, RNA samples were DNase treated using RQ1 RNase-Free DNase (Promega). After DNase treatment, samples were cleaned by phenol/chloroform extraction, and polyA mRNA was isolated using Sera-mag oligo(dT) beads, followed by first strand synthesis with SuperScriptIII Reverse Transcriptase (Invitrogen) using the VSG 3'UTR primer (see primers). Double-stranded cDNA was produced by combining 10X Second Strand synthesis Buffer (NEB), 10mM dNTPs (NEB), *E. coli* DNA Ligase (NEB), *E. coli* DNA polymerase (NEB), and RNase H (Invitrogen) for 2.5 hours at 16ºC. After dscDNA was produced, DNA was fragmented to about 200bp using NEBNext dsDNA fragmentase (NEB) for 43 minutes at 37ºC.

Prep 2: "oligo-dT"

This prep, which sequenced all polyA mRNA, rather than just VSG mRNA, was performed exactly as the "VSG" prep, except that mRNA was chemically fragmented (4 minutes at 94ºC in 30 mM Mg^{2+}) to ~200nt and precipitated, and first-strand was performed using random hexamers. cDNA was not fragmented with dsDNA fragmentase after second-strand synthesis.

Prep 3: "VSG + SL 2nd Strand"

This protocol was performed as the "VSG" preparation above except, in place of the second-strand synthesis step, 2 cycles of PCR were performed, using Phusion polymerase (New England Biolabs) and the SL and SP6-14mer primers (see primers), to specifically produce double-stranded *VSG* cDNA.

Prep 4: "Amplicon"

This sequencing preparation was identical to the "VSG + SL 2nd Strand" preparation, except 30 cycles of PCR, rather than 2 cycles, were performed.

7.7 Optimized VSG-seq library preparation

RNA samples were DNase treated using RQ1 RNase-Free DNase (Promega). After DNase treatment, samples were cleaned by phenol/chloroform extraction or using AMPure RNAClean XP beads (Agencourt), and first strand synthesis was performed using SuperScript III Reverse Transcriptase (Life Technologies), using a primer specific for the conserved 14-mer in the *VSG* 3´-UTR (see primers, all-VSG-3´-UTR). 1st strand

products were cleaned up using AmPureXP beads (Agencourt). 22 cycles of PCR (or 35 cycles in the case of human infection samples), using Phusion polymerase (New England Biolabs), was performed to amplify *VSG* cDNA, using SL and SP6-14mer primers (see primers). PCR products were cleaned using AmPureXP beads and DNA concentration was measured using a QuBit HS DNA kit (Life Technologies), according to manufacturer's protocol. Libraries were prepared from *VSG* PCR products using the Nextera XT DNA Sample Prep Kit (Illumina), according to the manufacturer's protocol.

For control libraries in Chapter 2 and acute and chronic infections described in Chapters 2 and 3, 100bp single-end sequencing was performed on an Illumina HiSeq 2000. For libraries described in Chapter 5, made from tissue samples, parasites isolated from *in vitro* cultures, and human infection samples, 150bp single-end sequencing was performed on an Illumina NextSeq 500. For human infection samples described in Chapter 5, libraries were prepared with one alteration to the protocol described above. Because of the amount of contaminating host RNA and very low parasitemia, PCR products between 1-1.5 kb were isolated by gel extraction using a QiaQuick Gel Extraction kit (Qiagen), after the amplification of VSG cDNA. 150bp paired-end sequencing was performed on these libraries by an Illumina MiSeq (for sample RB001S) or an Illumina NextSeq 500 (for all other samples).

7.8 VSG-seq Analysis

Adapters were trimmed using trim_galore(v0.3.3) and SP6 sequences (from *VSG* PCR step) were trimmed using cutadapt(v1.3)[149].

Because many *VSGs* show stretches of homology, quantification of expression can be complicated, whether allowing unique or multiply mapping reads during alignment. To address this issue, I used MULTo (v1.0) to create a mappability file for each reference genome[120]. Sequencing reads were aligned to the reference genome (either assembled or the Lister 427 VSGnome reference at http://129.85.245.250/index.html), using bowtie (v1.0.0)[150] allowing only uniquely mapping reads and no more than two mismatches per read (*-v 2 –m 1*). Then, MULTo's rpkmforgenes.py was used to calculate the FPKM value for each *VSG*. The percent of the population expressing each *VSG* was calculated by dividing the FPKM for a *VSG* by the total FPKM. The number of parasites expressing a *VSG* was calculated by multiplying the total number of parasites by the percent of the population expressing that *VSG*.

When a reference VSGnome was assembled, sequencing reads were assembled using Trinity [116] using *--normalize_reads*. Once libraries were assembled, in-house scripts were used to identify contigs with an open reading frame (ORF) of >900bp (where an ORF is identified as start codon to stop codon, or start codon to end of contig). These ORFs were then compared against the Lister 427 reference VSGnome (http://129.85.245.250/index.html) using BLASTn (v2.2.28+), and ORFs with an alignment with an e-value < 1e^{-10} were considered *VSGs*. These ORF sequences were then merged using cd-hit-est (cd-hit v4.6.1)[151,152] with the following parameters: *-c 0.9 -n 8 -r 1 -G 1 -g 1 -b 20 -s 0.0 -aL 0.0 -aS 0.5*.

These potential *VSG* sequences were then evaluated by hand using NCBI's

BLASTn nr/nt database. Sequences that matched chromosomal sequences or non-*VSG*

genes were removed.

Because some *VSG* sequences were very similar, and the alignment program

bowtie[150] cannot distinguish VSGs >98% similar in sequence using the described

alignment parameters, sequences were then merged again using cd-hit-est (*-c 0.98 -n 8 -r*

1 -G 1 -g 1 -b 20 -s 0.0 -aL 0.0 -aS 0.5). This final merged contig file represents the final

reference genome for each experiment.

For analysis of 70-bp repeat modified strains (Chapter 5), assembly and

quantification was performed as described above, followed by a comparison to the Lister

427 reference database: assembled *VSG* sequences, when compared to the most similar

427 *VSG*, were identified as the 427 *VSG* when they had either 100% identity over >99%

of the length of the assembled ORF or >99% identity over 100% of the assembled ORF.

Otherwise, assembled *VSGs* were referred to as variants of the most similar Lister 427

VSG. After identifying the corresponding 427 *VSG,* noise (*VSGs* measured below the

limit of detection, 0.01%) and contamination (the starting *VSG, 427-2*) were removed.

The relative abundance of each remaining expressed *VSG* was then calculated using its

measured FPKM. These data were then used to create a heatmap using heatmap.2 from

the gplots package in R (https://cran.r-project.org/web/packages/gplots/gplots.pdf).

7.8 Mosaic Identification

Candidate mosaics were identified by comparing *VSG* sequences to two independently assembled genomes for this strain, which were kindly provided by Keith Matthews and Isabel Roditi. Candidate mosaics were *VSGs* with <80% of their length aligning to any sequence in either genome, and expressed in only one infection at >0.1% of the population. Primers were designed around the likely region of recombination (see primers) and PCR was performed using HotStarTaq (Qiagen). Mosaic amplicons were cloned using a Topo-TA cloning kit (Invitrogen) and sequenced to confirm validation.

7.9 Primer sequences

Primer	Sequence (5´→3´)
SL	ACA GTT TCT GTA CTA TAT TG
SP6-VSG14mer	GAT TTA GGT GAC ACT ATA GTG TTA AAA TAT ATC
All-VSG-3´UTR	GTG TTA AAA TAT ATC
TbURA3F	CGG CAG CAG TTC TCG AGT
TbURA3R	TGG CGT GTA CCT TGA GGC
c5477_g4_i1_d24-F	GGA ATC ACA TCG CGG TTC CCA AAG
c5477_g4_i1_d24-R	CAA TTT TGC CAT CTG CCC CTC CTG
c754_g1_i2_d105-F	CTT CGA GAC GAC ACC AAA AGC CAC
c754_g1_i2_d105-R	GGC CGC AAA TGC AGA AGA AAC CAT TAG
370_g1_i1_d102-F	GTA AAC TTG AAC CGG CGG TGG C
370_g1_i1_d102-R	CAG AGC ACT TTT TGG CAT CGG AGT T

Table 8.1. Primer Sequences.
Table of all primer sequences used for library preparations and mosaic *VSG* validation.

94

REFERENCES

1. World Health Organization. Working to overcome the global impact of neglected tropical diseases: first WHO report on neglected tropical diseases. Geneva : World Health Organization; 2010.

2. Simarro PP, Cecchi G, Franco JR, Paone M, Diarra A, Priotto G, Mattioli RC, Jannin JG. Monitoring the Progress towards the Elimination of Gambiense Human African Trypanosomiasis. Matovu E, editor. PLoS neglected tropical diseases. 2015;9(6):e0003785.

3. Lai D-H, Hashimi H, Lun Z-R, Ayala FJ, Lukes J. Adaptations of Trypanosoma brucei to gradual loss of kinetoplast DNA: Trypanosoma equiperdum and Trypanosoma evansi are petite mutants of T. brucei. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(6):1999–2004.

4. Reuner B, Vassella E, Yutzy B, Boshart M. Cell density triggers slender to stumpy differentiation of Trypanosoma brucei bloodstream forms in culture. Molecular and biochemical parasitology. 1997;90(1):269–280.

5. Macgregor P, Savill NJ, Hall D, Matthews KR. Transmission Stages Dominate Trypanosome Within-Host Dynamics during Chronic Infections. Cell Host and Microbe. 2011;9(4):310–318.

6. Gibson W, Peacock L, Ferris V, Williams K, Bailey M. The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. Parasites & vectors. 2008;1(1):1.

7. Peacock L, Ferris V, Sharma R, Sunter J, Bailey M, Carrington M, Gibson W. Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(9):3671–3676.

8. Seed JR, Effron HG. Simultaneous presence of different antigenic populations of Trypanosoma brucei gambiense in Microtus montanus. Parasitology. 1973;66(2):269–278.

9. Mogk S, Meiwes A, Boßelmann CM, Wolburg H, Duszenko M. The lane to the brain: how African trypanosomes invade the CNS. Trends in parasitology. 2014;30(10):470–477.

10. Bentivoglio M, Kristensson K. Tryps and trips: cell trafficking across the 100-year-old blood-brain barrier. Trends in neurosciences. 2014;37(6):325–333.

11. Buguet A, Tapie P, Bert J. Reversal of the sleep/wake cycle disorder of sleeping sickness after trypanosomicide treatment. Journal of sleep research. 1999;8(3):225–235.

12. Steverding D. The development of drugs for treatment of sleeping sickness: a historical review. Parasites & vectors. 2010;3(1):15.

13. Magez S, Caljon G, Tran T, Stijlemans B, Radwanska M. Current status of vaccination against African trypanosomiasis. Parasitology. 2010;137(14):2017–2027.

14. La Greca F, Magez S. Vaccination against trypanosomiasis. Human vaccines. 2011;7(11):1225–1233.

15. Ross R, Thomson D. A Case of Sleeping Sickness Studied by Precise Enumerative Methods: Regular Periodical Increase of the Parasites Disclosed. Proceedings of the Royal Society B: Biological Sciences. 1910;82(557):411–415.

16. Ross R, Thomson D. A Case of Sleeping Sickness Studied by Precise Enumerative Methods: Further Observations. Proceedings. Biological sciences / The Royal Society. 1911;83(563):187–205.

17. Cross GA. Identification, purification and properties of clone-specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology. 1975;71(3):393–417.

18. Vickerman K, Luckins AG. Localization of variable antigens in the surface coat of Trypanosoma brucei using ferritin conjugated antibody. Nature. 1969;224(5224):1125–1126.

19. Grünfelder CG, Engstler M, Weise F, Schwarz H, Stierhof Y-D, Boshart M, Overath P. Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. Traffic. 2002;3(8):547–559.

20. Jackson DG, Owen MJ, Voorheis HP. A new method for the rapid purification of both the membrane-bound and released forms of the variant surface glycoprotein from Trypanosoma brucei. The Biochemical journal. 1985;230(1):195–202.

21. Ziegelbauer K, Overath P. Organization of two invariant surface glycoproteins in the surface coat of Trypanosoma brucei. Infection and immunity. 1993;61(11):4540–4545.

22. Sullivan L, Wall SJ, Carrington M, Ferguson MAJ. Proteomic selection of immunodiagnostic antigens for human African trypanosomiasis and generation of a prototype lateral flow immunodiagnostic device. Tschudi C, editor. PLoS neglected tropical diseases. 2013;7(2):e2087.

23. Freymann D, Down J, Carrington M, Roditi I, Turner M, Wiley D. 2.9 A resolution structure of the N-terminal domain of a variant surface glycoprotein from Trypanosoma brucei. Journal of molecular biology. 1990;216(1):141–160.

24. Blum ML, Down JA, Gurnett AM, Carrington M, Turner MJ, Wiley DC. A structural motif in the variant surface glycoproteins of Trypanosoma brucei. Nature. 1993;362(6421):603–609.

25. Carrington M, Miller N, Blum M, Roditi I, Wiley D, Turner M. Variant specific glycoprotein of Trypanosoma brucei consists of two domains each having an independently conserved pattern of cysteine residues. Journal of molecular biology. 1991;221(3):823–835.

26. Cross GAM, Kim H-S, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of Trypanosoma brucei Lister 427. Molecular and biochemical parasitology. 2014;195(1):59–73.

27. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, et al. The genome of the African trypanosome Trypanosoma brucei. Science (New York, NY). 2005;309(5733):416–422.

28. Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GCM, Vervaet G, Skepner E, Lewis NS, Spronken MIJ, Russell CA, et al. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science (New York, NY). 2013;342(6161):976–979.

29. Palmer GH, Bankhead T, Lukehart SA. "Nothing is permanent but change-" antigenic variation in persistent bacterial pathogens. Cellular microbiology. 2009;11(12):1697–1705.

30. Nash T. Surface antigen variability and variation in Giardia lamblia. Parasitology today (Personal ed.). 1992;8(7):229–234.

31. Kirkman LA, Deitsch KW. Recombination and Diversification of the Variant Antigen Encoding Genes in the Malaria Parasite Plasmodium falciparum. Microbiology spectrum. 2014;2(6):437–449.

32. Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, Brooks K, Churcher C, Fahkro S, Goodhead I, et al. Telomeric expression sites are highly conserved in Trypanosoma brucei. PloS one. 2008;3(10):e3527.

33. Alexandre S, Paindavoine P, Tebabi P, Pays A, Halleux S, Steinert M, Pays E. Differential expression of a family of putative adenylate/guanylate cyclase genes in Trypanosoma brucei. Molecular and biochemical parasitology. 1990;43(2):279–288.

34. Morgan RW, El-Sayed NM, Kepa JK, Pedram M, Donelson JE. Differential expression of the expression site-associated gene I family in African trypanosomes. The Journal of biological chemistry. 1996;271(16):9771–9777.

35. Liu AY, Van der Ploeg LH, Rijsewijk FA, Borst P. 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. Journal of molecular biology. 1983;167(1):57–75.

36. Aline R, MacDonald G, Brown E, Allison J, Myler P, Rothwell V, Stuart K. (TAA)n within sequences flanking several intrachromosomal variant surface glycoprotein genes in Trypanosoma brucei. Nucleic acids research. 1985;13(9):3161–3177.

37. Campbell DA, van Bree MP, Boothroyd JC. The 5'-limit of transposition and upstream barren region of a trypanosome VSG gene: tandem 76 base-pair repeats flanking (TAA)90. Nucleic acids research. 1984;12(6):2759–2774.

38. McCulloch R, Rudenko G, Borst P. Gene conversions mediating antigenic variation in Trypanosoma brucei can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. Molecular and cellular biology. 1997;17(2):833–843.

39. Pham VP, Qi CC, Gottesdiener KM. A detailed mutational analysis of the VSG gene expression site promoter. Molecular and biochemical parasitology. 1996;75(2):241–254.

40. Günzl A, Bruderer T, Laufer G, Schimanski B, Tu L-C, Chung H-M, Lee P-T, Lee MG-S. RNA polymerase I transcribes procyclin genes and variant surface glycoprotein gene expression sites in Trypanosoma brucei. Eukaryotic Cell. 2003;2(3):542–551.

41. Parsons M, Nelson RG, Watkins KP, Agabian N. Trypanosome mRNAs share a common 5' spliced leader sequence. Cell. 1984;38(1):309–316.

42. Berberof M, Vanhamme L, Tebabi P, Pays A, Jefferies D, Welburn S, Pays E. The 3'-terminal region of the mRNAs for VSG and procyclin can confer stage specificity to gene expression in Trypanosoma brucei. The EMBO Journal. 1995;14(12):2925–2934.

43. Lenardo MJ, Rice-Ficht AC, Kelly G, Esser KM, Donelson JE. Characterization of the genes specifying two metacyclic variable antigen types in Trypanosoma brucei rhodesiense. Proceedings of the National Academy of Sciences. 1984;81(21):6642–6646.

44. Barnes DA, Mottram JC, Agabian N. Bloodstream and metacyclic variant surface glycoprotein gene expression sites of Trypanosoma brucei gambiense. Molecular and biochemical parasitology. 1990;41(1):101–114.

45. Nagoshi YL, Alarcon CM, Donelson JE. The putative promoter for a metacyclic VSG gene in African trypanosomes. Molecular and biochemical parasitology. 1995;72(1-2):33–45.

46. Ginger ML, Blundell PA, Lewis AM, Browitt A, Günzl A, Barry JD. Ex vivo and in vitro identification of a consensus promoter for VSG genes expressed by metacyclic-stage trypanosomes in the tsetse fly. Eukaryotic Cell. 2002;1(6):1000–1009.

47. Liu AY, Michels PA, Bernards A, Borst P. Trypanosome variant surface glycoprotein genes expressed early in infection. Journal of molecular biology. 1985;182(3):383–396.

48. Rudenko G, McCulloch R, Dirks-Mulder A, Borst P. Telomere exchange can be an important mechanism of variant surface glycoprotein gene switching in Trypanosoma brucei. Molecular and biochemical parasitology. 1996;80(1):65–75.

49. Boothroyd CE, Dreesen O, Leonova T, Ly KI, Figueiredo LM, Cross GAM, Papavasiliou FN. A yeast-endonuclease-generated DNA break induces antigenic switching in Trypanosoma brucei. Nature. 2009;459(7244):278–281.

50. Glover L, Alsford S, Horn D. DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. PLoS pathogens. 2013;9(3):e1003260.

51. McCulloch R, Barry JD. A role for RAD51 and homologous recombination in Trypanosoma brucei antigenic variation. Genes & development. 1999;13(21):2875–2888.

52. Hartley CL, McCulloch R. Trypanosoma brucei BRCA2 acts in antigenic variation and has undergone a recent expansion in BRC repeat number that is important during homologous recombination. 2008;68(5):1237–1251.

53. Dreesen O, Li B, Cross GAM. Telomere structure and function in trypanosomes: a proposal. Nature reviews Microbiology. 2007;5(1):70–75.

54. Hovel-Miner GA, Boothroyd CE, Mugnier MR, Dreesen O, Cross GAM, Papavasiliou FN. Telomere length affects the frequency and mechanism of antigenic variation in Trypanosoma brucei. PLoS pathogens. 2012;8(8):e1002900.

55. Van der Ploeg LH, Valerio D, De Lange T, Bernards A, Borst P, Grosveld FG. 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 research. 1982;10(19):5905–5923.

56. Wickstead B, Ersfeld K, Gull K. The small chromosomes of Trypanosoma brucei involved in antigenic variation are constructed around repetitive palindromes. Genome research. 2004;14(6):1014–1024.

57. Marcello L, Barry JD. 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 research. 2007;17(9):1344–1352.

58. Bernards A, Van der Ploeg LH, Frasch AC, Borst P, Boothroyd JC, Coleman S, Cross GA. Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3' end. Cell. 1981;27(3 Pt 2):497–505.

59. Michels PA, Liu AY, Bernards A, Sloof P, Van der Bijl MM, Schinkel AH, Menke HH, Borst P, Veeneman GH, Tromp MC, et al. Activation of the genes for variant surface glycoproteins 117 and 118 in Trypanosoma brucei. Journal of molecular biology. 1983;166(4):537–556.

60. Pays E, Houard S, Pays A, Van Assel S, Dupont F, Aerts D, Huet-Duvillier G, Gomés V, Richet C, Degand P. Trypanosoma brucei: the extent of conversion in antigen genes may be related to the DNA coding specificity. Cell. 1985;42(3):821–829.

61. Kamper SM, Barbet AF. Surface epitope variation via mosaic gene formation is potential key to long-term survival of Trypanosoma brucei. Molecular and biochemical parasitology. 1992;53(1-2):33–44.

62. Barbet AF, Myler PJ, Williams RO, McGuire TC. Shared surface epitopes among trypanosomes of the same serodeme expressing different variable surface glycoprotein genes. Molecular and biochemical parasitology. 1989;32(2-3):191–199.

63. Aline RF, Myler PJ, Gobright E, Stuart KD. Early expression of a Trypanosoma brucei VSG gene duplicated from an incomplete basic copy. The Journal of eukaryotic microbiology. 1994;41(1):71–78.

64. Hall JPJ, Wang H, Barry JD. Mosaic VSGs and the scale of Trypanosoma brucei antigenic variation. PLoS pathogens. 2013;9(7):e1003502.

65. Checchi F, Filipe JAN, Haydon DT, Chandramohan D, Chappuis F. Estimates of the duration of the early and late stage of gambiense sleeping sickness. BMC infectious diseases. 2008;8(1):16.

66. Coutte L, Botkin DJ, Gao L, Norris SJ. Detailed analysis of sequence changes occurring during vlsE antigenic variation in the mouse model of Borrelia burgdorferi infection. Coburn J, editor. PLoS pathogens. 2009;5(2):e1000293.

67. Zhuang Y, Futse JE, Brown WC, Brayton KA, Palmer GH. Maintenance of antibody to pathogen epitopes generated by segmental gene conversion is highly dynamic during long-term persistent infection. Infection and immunity. 2007;75(11):5185–5190.

68. Claessens A, Hamilton WL, Kekre M, Otto TD, Faizullabhoy A, Rayner JC, Kwiatkowski D. Generation of antigenic diversity in Plasmodium falciparum by structured rearrangement of Var genes during mitosis. Deitsch K, editor. PLoS genetics. 2014;10(12):e1004812.

69. Lu Y, Alarcon CM, Hall T, Reddy LV, Donelson JE. A strand bias occurs in point mutations associated with variant surface glycoprotein gene conversion in Trypanosoma rhodesiense. Molecular and cellular biology. 1994;14(6):3971–3980.

70. Lu Y, Hall T, Gay LS, Donelson JE. Point mutations are associated with a gene duplication leading to the bloodstream reexpression of a trypanosome metacyclic VSG. Cell. 1993;72(3):397–406.

71. Graham VS, Barry JD. Is point mutagenesis a mechanism for antigenic variation in Trypanosoma brucei? Molecular and biochemical parasitology. 1996;79(1):35–45.

72. Glover L, McCulloch R, Horn D. Sequence homology and microhomology dominate chromosomal double-strand break repair in African trypanosomes. Nucleic acids research. 2008;36(8):2608–2618.

73. Glover L, Jun J, Horn D. Microhomology-mediated deletion and gene conversion in African trypanosomes. Nucleic acids research. 2011;39(4):1372–1380.

74. Barbet AF, Davis WC, McGuire TC. Cross-neutralization of two different trypanosome populations derived from a single organism. Nature. 1982;300(5891):453–456.

75. Roth CW, Longacre S, Raibaud A, Baltz T, Eisen H. The use of incomplete genes for the construction of a Trypanosoma equiperdum variant surface glycoprotein gene. The EMBO Journal. 1986;5(5):1065–1070.

76. Roth C, Bringaud F, Layden RE, Baltz T, Eisen H. Active late-appearing variable surface antigen genes in Trypanosoma equiperdum are constructed entirely from pseudogenes. Proceedings of the National Academy of Sciences of the United States of America. 1989;86(23):9375–9379.

77. Gray AR. Antigenic variation in a strain of Trypanosoma brucei transmitted by Glossina morsitans and G. palpalis. Journal of general microbiology. 1965;41(2):195–214.

78. Myler PJ, Allen AL, Agabian N, Stuart K. Antigenic variation in clones of Trypanosoma brucei grown in immune-deficient mice. Infection and immunity. 1985;47(3):684–690.

79. Miller EN, Turner MJ. Analysis of antigenic types appearing in first relapse populations of clones of Trypanosoma brucei. Parasitology. 1981;82(1):63–80.

80. Morrison LJ, Majiwa P, Read AF, Barry JD. Probabilistic order in antigenic variation of Trypanosoma brucei. International journal for parasitology. 2005;35(9):961–972.

81. Lythgoe KA, Morrison LJ, Read AF, Barry JD. Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(19):8095–8100.

82. Kosinski RJ. Antigenic variation in trypanosomes: a computer analysis of variant order. Parasitology. 1980;80(2):343–357.

83. Gjini E, Haydon DT, Barry JD, Cobbold CA. Critical Interplay between Parasite Differentiation, Host Immunity, and Antigenic Variation in Trypanosome Infections. The American naturalist. 2010 Aug 17.

84. Gjini E, Haydon DT, Barry JD, Cobbold CA. Linking the antigen archive structure to pathogen fitness in African trypanosomes. Proceedings. Biological sciences / The Royal Society. 2013;280(1753):20122129–20122129.

85. Antia R, Nowak MA, Anderson RM. Antigenic variation and the within-host dynamics of parasites. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(3):985–989.

86. Recker M, Nee S, Bull PC, Kinyanjui S, Marsh K, Newbold C, Gupta S. Transient cross-reactive immune responses can orchestrate antigenic variation in malaria. Nature. 2004;429(6991):555–558.

87. Turner CM. The rate of antigenic variation in fly-transmitted and syringe-passaged infections of Trypanosoma brucei. FEMS microbiology letters. 1997;153(1):227–231.

88. Turner CM, Barry JD. High frequency of antigenic variation in Trypanosoma brucei rhodesiense infections. Parasitology. 1989;99 Pt 1:67–75.

89. Lamont GS, Tucker RS, Cross GA. Analysis of antigen switching rates in Trypanosoma brucei. Parasitology. 1986;92 (Pt 2):355–367.

90. Doyle JJ, Hirumi H, Hirumi K, Lupton EN, Cross GA. Antigenic variation in clones of animal-infective Trypanosoma brucei derived and maintained in vitro. Parasitology. 1980;80(2):359–369.

91. Aitcheson N, Talbot S, Shapiro J, Hughes K, Adkin C, Butt T, Sheader K, Rudenko G. VSG switching in Trypanosoma brucei: antigenic variation analysed using RNAi in the absence of immune selection. Molecular microbiology. 2005;57(6):1608–1622.

92. Eyford BA, Ahmad R, Enyaru JC, Carr SA, Pearson TW. Identification of Trypanosome proteins in plasma from African sleeping sickness patients infected with T. b. rhodesiense. Sturtevant J, editor. PloS one. 2013;8(8):e71463.

93. Seed JR, Edwards R, Sechelski J. The ecology of antigenic variation. The Journal of protozoology. 1984;31(1):48–53.

94. Diffley P. Trypanosomal surface coat variant antigen causes polyclonal lymphocyte activation. Journal of immunology (Baltimore, Md : 1950). 1983;131(4):1983–1986.

95. Musoke AJ, Nantulya VM, Barbet AF, Kironde F, McGuire TC. Bovine immune response to African ;trypanosomes: specific antibodies to variable surface glycoproteins of Trypanosoma brucei. Parasite immunology. 1981;3(2):97–106.

96. Dempsey WL, Mansfield JM. Lymphocyte function in experimental African trypanosomiasis. V. Role of antibody and the mononuclear phagocyte system in variant-specific immunity. Journal of immunology (Baltimore, Md : 1950). 1983;130(1):405–411.

97. Radwanska M, Guirnalda P, De Trez C, Ryffel B, Black S, Magez S. Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. PLoS pathogens. 2008;4(5):e1000078.

98. Corsini AC, Clayton C, Askonas BA, Ogilvie BM. Suppressor cells and loss of B-cell potential in mice infected with Trypanosoma brucei. Clinical and experimental immunology. 1977;29(1):122–131.

99. Bockstal V, Guirnalda P, Caljon G, Goenka R, Telfer JC, Frenkel D, Radwanska M, Magez S, Black SJ. T. brucei infection reduces B lymphopoiesis in bone marrow and truncates compensatory splenic lymphopoiesis through transitional B-cell apoptosis. Wynn TA; Wynn TA, editors. PLoS pathogens. 2011;7(6):e1002089.

100. Magez S, Schwegmann A, Atkinson R, Claes F, Drennan M, DE Baetselier P, Brombacher F. The role of B-cells and IgM antibodies in parasitemia, anemia, and VSG switching in Trypanosoma brucei-infected mice. PLoS pathogens. 2008;4(8):e1000122.

101. Lutz C, Ledermann B, Kosco-Vilbois MH, Ochsenbein AF, Zinkernagel RM, Köhler G, Brombacher F. IgD can largely substitute for loss of IgM function in B cells. Nature. 1998;393(6687):797–801.

102. Chen K, Cerutti A. New insights into the enigma of immunoglobulin D. Immunological reviews. 2010;237(1):160–179.

103. Radwanska M, Magez S, Michel A, Stijlemans B, Geuskens M, Pays E. 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. Infection and immunity. 2000;68(2):848–860.

104. Vickerman K. On the surface coat and flagellar adhesion in trypanosomes. Journal of cell science. 1969;5(1):163–193.

105. Schwede A, Macleod OJS, Macgregor P, Carrington M. How Does the VSG Coat of Bloodstream Form African Trypanosomes Interact with External Proteins? Gubbels M-J, editor. PLoS pathogens. 2015;11(12):e1005259.

106. Schwede A, Jones N, Engstler M, Carrington M. The VSG C-terminal domain is inaccessible to antibodies on live trypanosomes. Molecular and biochemical parasitology. 2011;175(2):201–204.

107. Hsia R, Beals T, Boothroyd JC. Use of chimeric recombinant polypeptides to analyse conformational, surface epitopes on trypanosome variant surface glycoproteins. Molecular microbiology. 1996;19(1):53–63.

108. Engstler M, Pfohl T, Herminghaus S, Boshart M, Wiegertjes G, Heddergott N, Overath P. Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. Cell. 2007;131(3):505–515.

109. Overath P, Stierhof YD, Wiese M. Endocytosis and secretion in trypanosomatid parasites - Tumultuous traffic in a pocket. Trends in cell biology. 1997;7(1):27–33.

110. Engstler M, Thilo L, Weise F, Grünfelder CG, Schwarz H, Boshart M, Overath P. Kinetics of endocytosis and recycling of the GPI-anchored variant surface glycoprotein in Trypanosoma brucei. Journal of cell science. 2004;117(Pt 7):1105–1115.

111. Baltz T, Giroud C, Bringaud F, Eisen H, Jacquemot C, Roth CW. Exposed epitopes on a Trypanosoma equiperdum variant surface glycoprotein altered by point mutations. The EMBO Journal. 1991;10(7):1653–1659.

112. Schleifer KW, Filutowicz H, Schopf LR, Mansfield JM. Characterization of T helper cell responses to the trypanosome variant surface glycoprotein. Journal of immunology (Baltimore, Md : 1950). 1993;150(7):2910–2919.

113. Hertz CJ, Filutowicz H, Mansfield JM. Resistance to the African trypanosomes is IFN-gamma dependent. Journal of immunology (Baltimore, Md : 1950). 1998;161(12):6775–6783.

114. Dagenais TR, Demick KP, Bangs JD, Forest KT, Paulnock DM, Mansfield JM. T-cell responses to the trypanosome variant surface glycoprotein are not limited to hypervariable subregions. Infection and immunity. 2009;77(1):141–151.

115. Camara M, Camara O, Ilboudo H, Sakande H, Kaboré J, N'Dri L, Jamonneau V, Bucheton B. Sleeping sickness diagnosis: use of buffy coats improves the sensitivity of the mini anion exchange centrifugation test. Tropical medicine & international health : TM & IH. 2010;15(7):796–799.

116. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology. 2011;29(7):644–652.

117. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24–26.

118. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 2013;14(2):178–192.

119. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology. 2010;28(5):511–515.

120. Storvall H, Ramsköld D, Sandberg R. Efficient and comprehensive representation of uniqueness for next-generation sequencing by minimum unique length analyses. PloS one. 2013;8(1):e53822.

121. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, Rock D, Falloon J, Davey RT, Dewar RL, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. Journal of clinical microbiology. 2005;43(1):406–413.

122. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Ishikawa S, Linak MC, Hirai A, Takahashi H, et al. Sequence-specific error profile of Illumina sequencers. Nucleic acids research. 2011;39(13):e90–e90.

123. Lanham SM, Godfrey DG. Isolation of salivarian trypanosomes from man and other mammals using DEAE-cellulose. Experimental parasitology. 1970;28(3):521–534.

124. Vassella E, Reuner B, Yutzy B, Boshart M. Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway. Journal of cell science. 1997;110 (Pt 21):2661–2671.

125. Lu J, Panavas T, Thys K, Aerssens J, Naso M, Fisher J, Rycyzyn M, Sweet RW. IgG variable region and VH CDR3 diversity in unimmunized mice analyzed by massively parallel sequencing. Molecular Immunology. 2014;57(2):274–283.

126. Navarro M, Cross GA. DNA rearrangements associated with multiple consecutive directed antigenic switches in Trypanosoma brucei. Molecular and cellular biology. 1996;16(7):3615–3625.

127. Wirtz E, Leal S, Ochatt C, Cross GA. A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. Molecular and biochemical parasitology. 1999;99(1):89–101.

128. Claes F, Vodnala SK, van Reet N, Boucher N, Lunden-Miguel H, Baltz T, Goddeeris BM, Büscher P, Rottenberg ME. Bioluminescent imaging of Trypanosoma brucei shows preferential testis dissemination which may hamper drug efficacy in sleeping sickness. PLoS neglected tropical diseases. 2009;3(7):e486.

129. Van Meirvenne N, Janssens PG, Magnus E. Antigenic variation in syringe passaged populations of Trypanosoma (Trypanozoon) brucei. 1. Rationalization of the experimental approach. Annales de la Société belge de médecine tropicale. 1975;55(1):1–23.

130. Mugnier MR, Cross GAM, Papavasiliou FN. The in vivo dynamics of antigenic variation in Trypanosoma brucei. Science (New York, NY). 2015;347(6229):1470–1473.

131. Seed JR. Competition among serologically different clones of Trypanosoma brucei gambiense in vivo. The Journal of protozoology. 1978;25(4):526–529.

132. Marcello L, Barry JD. From silent genes to noisy populations-dialogue between the genotype and phenotypes of antigenic variation. The Journal of eukaryotic microbiology. 2007;54(1):14–17.

133. Thon G, Baltz T, Giroud C, Eisen H. Trypanosome variable surface glycoproteins: composite genes and order of expression. Genes & development. 1990;4(8):1374–1383.

134. González-Andrade P, Camara M, Ilboudo H, Bucheton B, Jamonneau V, Deborggraeve S. Diagnosis of trypanosomatid infections: targeting the spliced leader RNA. The Journal of molecular diagnostics : JMD. 2014;16(4):400–404.

135. Chappuis F, Loutan L, Simarro P, Lejon V, Büscher P. Options for field diagnosis of human african trypanosomiasis. Clinical microbiology reviews. 2005;18(1):133–146.

136. Hutchinson OC, Picozzi K, Jones NG, Mott H, Sharma R, Welburn SC, Carrington M. Variant Surface Glycoprotein gene repertoires in Trypanosoma brucei have diverged to become strain-specific. BMC genomics. 2007;8(1):234.

137. Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, Brown R, Browne H, Corton N, Hauser H, et al. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(9):3416–3421.

138. Jennings FW, Whitelaw DD, Holmes PH, Chizyuka HG, Urquhart GM. The brain as a source of relapsing Trypanosoma brucei infection in mice after chemotherapy. International journal for parasitology. 1979;9(4):381–384.

139. Ehrlich P. Chemotherapeutische trypanosomen-studien. 1907.

140. Truc P, Lejon V, Magnus E, Jamonneau V, Nangouma A, Verloo D, Penchenier L, Büscher P. Evaluation of the micro-CATT, CATT/Trypanosoma brucei gambiense, and LATEX/T b gambiense methods for serodiagnosis and surveillance of human African trypanosomiasis in West and Central Africa. Bulletin of the World Health Organization. 2002;80(11):882–886.

141. Sullivan L, Fleming J, Sastry L, Mehlert A, Wall SJ, Ferguson MAJ. Identification of sVSG117 as an immunodiagnostic antigen and evaluation of a dual-antigen lateral flow test for the diagnosis of human African trypanosomiasis. Raper J, editor. PLoS neglected tropical diseases. 2014;8(7):e2976.

142. Dukes P, Gibson WC, Gashumba JK, Hudson KM, Bromidge TJ, Kaukus A, Asonganyi T, Magnus E. Absence of the LiTat 1.3 (CATT antigen) gene in Trypanosoma brucei gambiense stocks from Cameroon. Acta tropica. 1992;51(2):123–134.

143. Fairbairn H, Burtt E. The Infectivity to Man of a Strain of Trypanosoma rhodesiense transmitted cyclically by Glossina morsitans through Sheep and Antelope : Evidence that Man requires a Minimum Infective Dose of Meta-cyelic Trypanosomes. Ann trop Med Parasit. 1946;40(3/4):270–313.

144. Turner CMR, Barry JD, Maudlin I, Vickerman K. An estimate of the size of the metacyclic variable antigen repertoire of Trypanosoma brucei rhodesiense. Parasitology. 1988;97(02):269–276.

145. Lenardo MJ, Esser KM, Moon AM, Van der Ploeg LH, Donelson JE. Metacyclic variant surface glycoprotein genes of Trypanosoma brucei subsp. rhodesiense are activated in situ, and their expression is transcriptionally regulated. Molecular and cellular biology. 1986;6(6):1991–1997.

146. McKeown SR. Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response. The British journal of radiology. 2014;87(1035):20130676.

147. Vassella E, Abbeele Den JV, Bütikofer P, Renggli CK, Furger A, Brun R, Roditi I. A major surface glycoprotein of trypanosoma brucei is expressed transiently during development and can be regulated post-transcriptionally by glycerol or hypoxia. Genes & development. 2000;14(5):615–626.

148. Kraemer SM, Smith JD. A family affair: var genes, PfEMP1 binding, and malaria disease. Current opinion in microbiology. 2006;9(4):374–380.

149. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17:10–12.

150. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009;10(3):R25.

151. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics (Oxford, England). 2006;22(13):1658–1659.

152. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England). 2012;28(23):3150–3152.