



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/228377/>

Version: Published Version

Article:

Escrivani, Douglas O, Scheidt, Viktor, Tinti, Michele et al. (2023) Competition among variants is predictable and contributes to the antigenic variation dynamics of African trypanosomes. PLOS PATHOGENS. e1011530. ISSN: 1553-7366

<https://doi.org/10.1371/journal.ppat.1011530>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

RESEARCH ARTICLE

Competition among variants is predictable and contributes to the antigenic variation dynamics of African trypanosomes

Douglas O. Escrivani[☉], Viktor Scheidt^{☉a☉}, Michele Tinti, Joana Faria^{☉b}, David Horn^{☉*}

Wellcome Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dow Street, Dundee, United Kingdom

☉ These authors contributed equally to this work.

^a Current address: Dr. Brill & KEBOS GmbH, Institute for Hygiene and Microbiology, Hamburg, Germany.^b Current address: York Biomedical Research Institute, Department of Biology, University of York, Heslington, York, United Kingdom.* d.horn@dundee.ac.uk**OPEN ACCESS**

Citation: Escrivani DO, Scheidt V, Tinti M, Faria J, Horn D (2023) Competition among variants is predictable and contributes to the antigenic variation dynamics of African trypanosomes. *PLoS Pathog* 19(7): e1011530. <https://doi.org/10.1371/journal.ppat.1011530>

Editor: F. Nina Papavasiliou, German Cancer Research Centre: Deutsches Krebsforschungszentrum, GERMANY

Received: October 28, 2022

Accepted: July 3, 2023

Published: July 17, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1011530>

Copyright: © 2023 Escrivani et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The high-throughput RNA-seq data generated in this study have been deposited in the Short Read Archive (SRA) under

Abstract

Several persistent pathogens employ antigenic variation to continually evade mammalian host adaptive immune responses. African trypanosomes use variant surface glycoproteins (VSGs) for this purpose, transcribing one telomeric VSG expression-site at a time, and exploiting a reservoir of (sub)telomeric VSG templates to switch the active VSG. It has been known for over fifty years that new VSGs emerge in a predictable order in *Trypanosoma brucei*, and differential activation frequencies are now known to contribute to the hierarchy. Switching of approximately 0.01% of dividing cells to many new VSGs, in the absence of post-switching competition, suggests that VSGs are deployed in a highly profligate manner, however. Here, we report that switched trypanosomes do indeed compete, in a highly predictable manner that is dependent upon the activated VSG. We induced VSG gene recombination and switching in *in vitro* culture using CRISPR-Cas9 nuclease to target the active VSG. VSG dynamics, that were independent of host immune selection, were subsequently assessed using RNA-seq. Although trypanosomes activated VSGs from repressed expression-sites at relatively higher frequencies, the population of cells that activated minichromosomal VSGs subsequently displayed a competitive advantage and came to dominate. Furthermore, the advantage appeared to be more pronounced for longer VSGs. Differential growth of switched clones was also associated with wider differences, affecting transcripts involved in nucleolar function, translation, and energy metabolism. We conclude that antigenic variants compete, and that the population of cells that activates minichromosome derived VSGs displays a competitive advantage. Thus, competition among variants impacts antigenic variation dynamics in African trypanosomes and likely prolongs immune evasion with a limited set of antigens.

accession codes PRJEB40415 [<https://www.ncbi.nlm.nih.gov/bioproject/PRJEB40415>] for the main time-course, and PRJNA852534 [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA852534>] for the independent 5- and 25-day samples and switched clones. Scripts to align the RNA-seq data and to compute read-counts have been deposited in GitHub [https://github.com/mtinti/VSG_Competition/] and Zenodo [<https://doi.org/10.5281/zenodo.7942978>].

Funding: The work was funded by Wellcome Investigator Awards to D.H. [100320/Z/12/Z and 217105/Z/19/Z] and a Wellcome Ph.D. studentship award to V.S. [203816/Z/16/Z]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Human and animal parasites, including those that cause sleeping sickness, malaria and giardiasis, have evolved effective mechanisms to counter host adaptive immunity. These parasite genomes typically incorporate large gene-families encoding variant surface antigens that are deployed during infection, allowing parasites to remain ‘one step ahead’ of host immune responses. It remains unclear, however, how these reservoirs of surface antigens are effectively deployed. We generated a complex mixture of African trypanosome variants and quantitatively monitored their dynamic behavior over time. The results confirm differential activation rates, and also reveal differential growth rates, which are predictable. Differential growth can facilitate persistence through sparing, rather than profligate, presentation of variant antigens.

Introduction

The African trypanosome, *Trypanosoma brucei*, is a protozoan parasite, transmitted by tsetse flies, that causes persistent diseases in both humans and livestock; sleeping sickness and nagana, respectively. Exclusively extracellular bloodstream and tissue-resident trypanosomes can divide every 6–7 h, although substantially slower growth has been reported for adipose tissue forms [1]. These cells express a super-abundant variant surface glycoprotein (VSG), that forms a dense coat on each cell; comprising approximately 10% of total cell protein [2]. Indeed, translation in each *T. brucei* cell produces up to 80,000 VSG molecules every minute [3]. In the advanced second stage of infection, trypanosomes breach the blood-brain barrier and invade the central nervous system, causing severe and, without treatment, often lethal disease. These trypanosomes undergo antigenic variation by spontaneously switching the VSG, exploiting a reservoir of hundreds of intact VSG genes, even assembling mosaic VSGs from incomplete VSGs [4,5], and thereby avoiding clearance by the host’s adaptive immune system. The consequent and ongoing ‘arms race’ at the parasite-host interface, involving antigenically distinct VSGs, and anti-VSG antibodies, is the result of two of the most prolific protein diversifying mechanisms in biology [6].

Only one VSG is expressed at a time by each *T. brucei* cell [7]. The active VSG is transcribed in one of fourteen alternative sub-telomeric ‘bloodstream’ expression sites (ESs) in the ‘427’ strain [8] and switching is typically triggered by spontaneous DNA double-strand breaks that result in homologous recombination and gene conversion, replacing the active VSG [9–12]. In the first weeks of infection, donor VSG templates are typically from repressed ESs or from the ends of approximately 100 minichromosomes [13–15]. Repressed ESs, since they have promoters, can also be activated *in situ* [7]. Mosaic VSGs typically arise at low frequency and appear later during infection [4,5]. Thus, *T. brucei* is a remarkably persistent parasite that, as a result of antigenic variation, sustains host immune evasion for years or even decades [16].

Based on published switching rates that typically range from 10^{-3} to 10^{-4} per cell division [17], African trypanosomes could simultaneously present large numbers of VSGs to the host immune system in a highly profligate manner, which could rapidly exhaust the repertoire. Even a low parasitemia of one million cells, switching at a rate of 10^{-4} per cell division, has the potential to simultaneously present up to 400 different VSGs every day. Indeed, approximately 80 [5] to 190 VSGs [18] have been detected simultaneously *in vivo* in mouse models. The long-term persistence of African trypanosome infections, however, indicates that more complex mechanisms are likely at play. Variants may grow at different rates and may fail to elicit an

effective immune response until they constitute a substantial proportion of the overall parasitemia, for example [19].

It first became clear in the 1960s that *T. brucei* VSGs appear in a predictable order *in vivo* [20–22]. Further illustrating this point, several populations, all isolated on day eighteen of a chronic infection, were found to have independently activated the same VSG [23,24]. It has subsequently become clear that parasite-intrinsic factors contribute to the frequency of VSG activation. In particular, VSG location is important; a telomeric variant sharing more extensive flanking homology with the active variant, including upstream tracts of ‘70-bp repeats’ [13], is more likely to be activated by recombination, for example [25–27]. Beyond differential VSG activation frequency, roles for switch intermediates [28] and density dependent parasite differentiation [29] have also been considered in impacting antigenic variation dynamics. A more recent mathematical model [19], and analysis of data from mouse infections [5], suggested that *T. brucei* parasites expressing different-length VSGs grow at different rates [19]. Indeed, evidence supporting the view that variants compete was initially reported more than forty years ago [22,30].

Here, we use CRISPR-Cas9 to trigger VSG switching and to generate a heterogeneous mixture of antigenic variants in cell culture. We then quantitatively monitor VSG abundance and dynamics over time using RNA-seq. Our findings reveal antigenic variation dynamics that are driven by the preferential activation of expression site derived VSGs and a subsequent competitive advantage of populations of cells expressing minichromosome derived VSGs. Further analysis of switched clones revealed VSG length differences and a transcriptome pattern that coincided with differential growth. Our findings explain features of VSG population dynamics that have been observed in mouse models and that likely allow these parasites to prolong immune evasion using a limited set of variant antigen genes.

Results

Cas9-induced DNA breaks trigger antigenic variation

Each *T. brucei* VSG displays a number of characteristic features (Fig 1A). An N-terminal signal sequence targets the protein to the secretory pathway and a C-terminal glycosylphosphatidylinositol (GPI) signal directs post-translational membrane anchoring. Despite a remarkable level of primary sequence divergence, a pattern of cysteine residues [31] directs protein folding and maintains a characteristic structure [32]. A number of different N- (A-C) and C-terminal domain type (1–4) VSGs are derived based on these features and two examples are shown in Fig 1A; C-terminal domain types 1 and 3 typically contain eight Cys residues and are longer than types 2 and 4, which typically contain four Cys residues [31].

To assess VSG population dynamics, we sought to establish a complex antigenically variant *T. brucei* population in *in vitro* cell culture. Prior studies employed the I-SceI meganuclease to introduce a DNA break adjacent to the active VSG and to initiate efficient VSG switching [9, 10, 12, 13]. These studies also indicated that resection from the break promoted recombination within upstream tandem 70-bp repeats, followed by VSG gene conversion (see Fig 1B). We more recently developed an inducible CRISPR-Cas9 system in *T. brucei* [33], and reasoned that this would likely provide a more versatile approach for programming DNA breaks that trigger VSG switching, without prior modification of the target sequence. We first selected three distinct Cas9-single guide RNA (sgRNA) target sequences within the active bloodstream VSG expression site (BES1) in our experimental strain; two target sites located between the 70-bp repeats and the active VSG and one within the active VSG coding sequence (Fig 1B). VSG replacement is expected to remove these Cas9-sgRNA target sites, thereby preventing further VSG switching (Fig 1B).

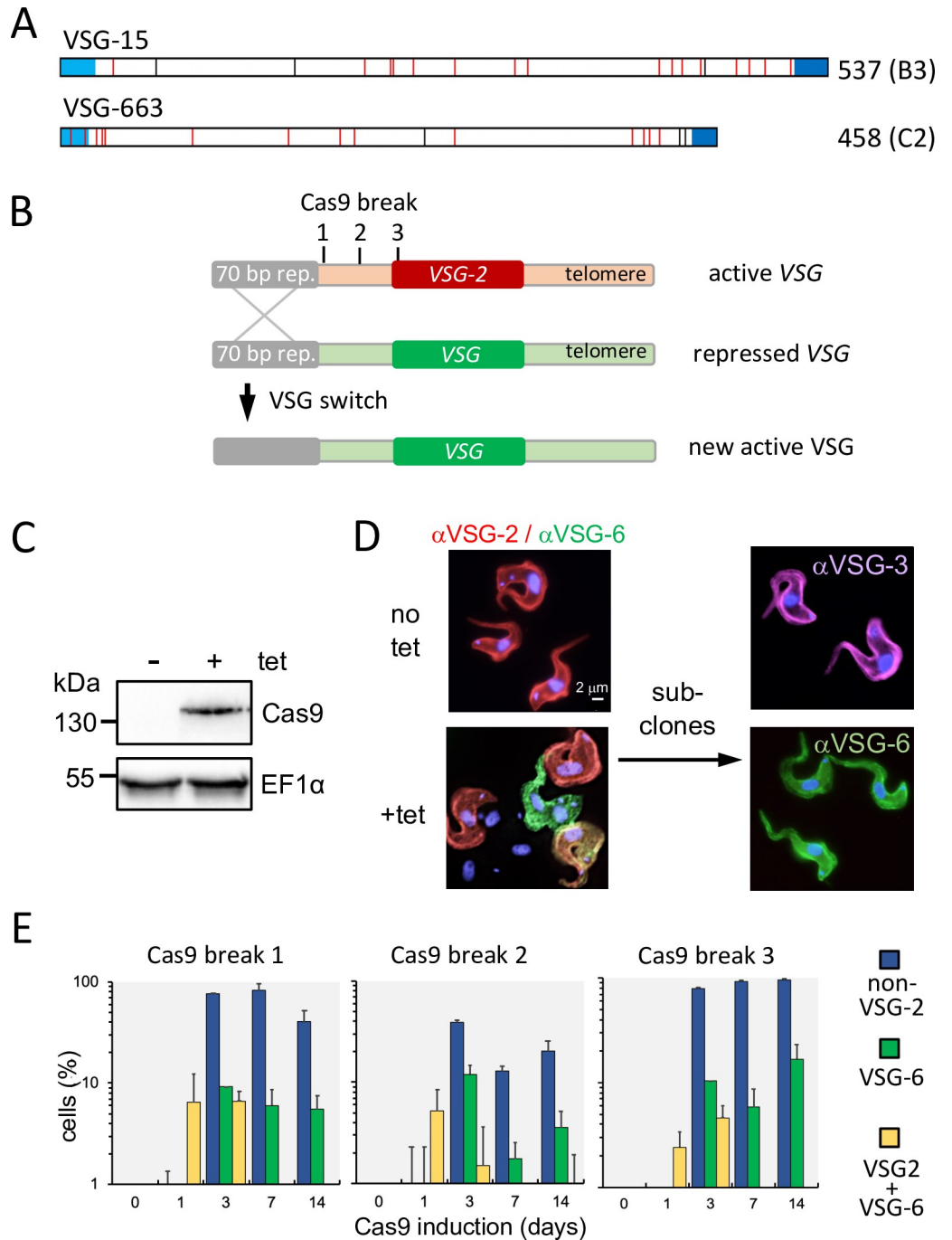


Fig 1. Cas9-induced DNA breaks trigger antigenic variation. (A) The schematic shows two examples of VSGs with salient features highlighted; the N-terminal signal sequence, Cys residues (red), predicted N-glycosylation sites (black) and C-terminal GPI anchor signal. Length, in amino acids, and class (B3 or C2) are indicated. See S2 Fig for more examples. (B) The schematic shows the active telomeric VSG (VSG-2) and the sites targeted by Cas9-sgRNAs to introduce DNA double-strand breaks. Recombination within 70-bp repeats triggers gene conversion and duplicative replacement of VSG-2 with a new VSG. (C) The protein blot shows robust inducible expression of Cas9. EF1α was used as loading control. (D) Immunofluorescence microscopy reveals switched and intermediate switching cells following Cas9-induction. Two switched sub-clones assessed by immunofluorescence microscopy are also shown. (E) An immunofluorescence microscopy time-course assay, with continuous Cas9 induction. Two biological replicates for each sgRNA. n > 90 cells for each sample at each time-point.

<https://doi.org/10.1371/journal.ppat.1011530.g001>

Protein blotting confirmed that Cas9 was tetracycline-inducible (Fig 1C). Using one sgRNA strain initially, we carried out immunofluorescence analysis to demonstrate a homogeneous population expressing VSG-2 prior to Cas9 induction, and a mixed population following Cas9 induction; cells that had switched to VSG-6 or another VSG, as well as switch intermediates, were detected (Fig 1D). By sub-cloning these populations and assessing a panel of clones using VSG-generic RT-PCR and Sanger sequencing, we demonstrated switching to different VSGs. Immunofluorescence analysis revealed stable expression of new switched VSGs; two example clones are shown in Fig 1D, and no subsequent switching was detected in pairs of clones expressing VSG-3 or VSG-6 ($n > 700$ cells). To compare the efficiency of each sgRNA in triggering antigenic variation, we induced Cas9 and monitored VSG-2 and VSG-6 expression over time (Fig 1E); duplicate clones were $>99.5\%$ homogeneous for VSG-2 expression prior to Cas9-induction. The analysis revealed that Cas9 induction triggered high-efficiency VSG switching when combined with each of the three sgRNAs tested; break-site 3, within the active VSG, generated the highest proportion of switched cells. We were able to detect switch intermediates with both VSG-2 and VSG-6 at the surface on day-1 and day-3 after Cas9-induction, but not before and not thereafter, while cells that no longer expressed detectable VSG-2 at the surface were first detected on day-3 (Fig 1E). Thus, our approach provided tightly regulated and highly efficient Cas9-based triggers for VSG switching.

RNA-seq revealed relative switching frequency to distinct groups of VSGs

We selected sgRNA break-site 3, within the active VSG, for subsequent analyses (see Fig 1B and 1E). We induced Cas9 for three days and allowed cells to complete the switching process for a further two days. We then split these cells into three technical replicate cultures, took the first mRNA samples for RNA-seq (day-5) and then took subsequent mRNA samples for RNA-seq on days 9, 13, 17 and 25 (S1 Data). We also monitored VSG-2 and VSG-6 expression in these populations and, consistent with the analysis above (Fig 1E), found a high proportion of cells throughout the time-course that no longer expressed VSG-2 (Fig 2A). As expected, switch intermediates, with both VSG-2 and VSG-6 at the surface, were no longer detectable on day 5, nor thereafter; prior studies also indicated full replacement of VSG coats ~ 4.5 days after switching [34]. Thus, our RNA-seq analysis from day 5 to day 25 represented, as intended, a post-switching time-course.

RNA-seq reads were initially aligned at high-stringency [35] to the VSG-ESs [8] to determine whether switched cells typically replaced only the DNA segment containing the VSG at the active ES (see Fig 2B). As expected, and consistent with recombination within 70-bp repeats, multiple ES-VSGs are expressed in the day-5 populations, while the ESAGs in their respective ESs are not expressed (Fig 2C; BES3, 7 and 12 are shown). In contrast, the ESAGs in the initially active VSG-ES, BES1, are expressed (Fig 2C; top panel). Thus, transient Cas9-induction yielded cells that typically replaced the VSG at the active ES, while retaining the same set of transcribed ESAGs (see Fig 2B).

We next considered the cohort of expression site derived VSGs (ES-VSGs) and minichromosome derived VSGs (MC-VSGs) that could be monitored in our switched populations. The Lister 427 *T. brucei* strain used here is the most commonly used trypanosome model for studies on antigenic variation. This strain benefits from excellent annotation of a full set of sequenced VSG-ESs; in the form of intact cloned fragments of up to 60 kbp [8], and several annotated monocistronic VSG-ESs, typically expressed in the metacyclic stage [13, 36–40]; these closely related subtelomeric sequences typically remain unassembled in short-read genome sequencing projects [41]. Thus, we collected the set of ES-VSG sequences, together with a set of MC-VSG sequences, and RNA-seq reads were aligned at high-stringency to these

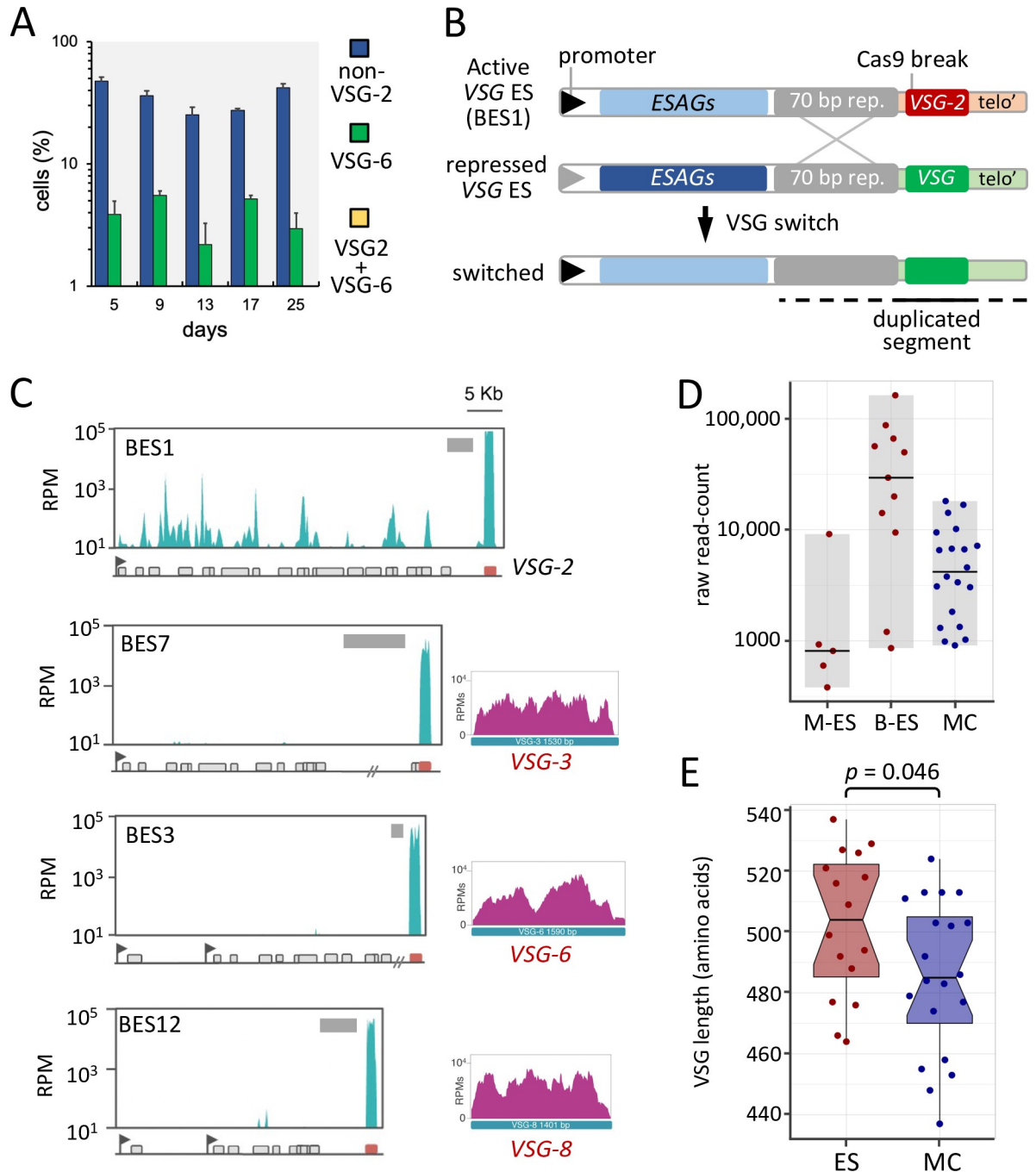


Fig 2. RNA-seq confirms that switched cells retain common ESAGs. (A) An immunofluorescence microscopy time-course assay, beginning on day-5; after three days of growth in tetracycline, to induce Cas9 with sgRNA-3, and a further two days after removal of tetracycline. Three technical replicates. $n > 90$ cells for each sample at each time-point. RNA was extracted from these samples for RNA-seq analysis. (B) The schematic shows the active telomeric VSG-ES (BES1) and the site targeted by sgRNA-3. Recombination with a silent bloodstream ES triggers replacement of VSG-2 with a new ES-VSG, while retaining the BES1 ESAGs. (C) RNA-seq analysis of VSG-ESs. Only the VSGs (3, 6 and 8) are activated, but not their respective ESAGs (grey boxes below each read-mapping profile, in BES7, 3 and 12), as illustrated in B. Values expressed in RPMs (reads per million). Flags indicate VSG-ES promoters; grey blocks indicate the extent of the 70-bp repeats. (D) The graph shows VSG expression levels as determined by RNA-seq on day-5 as jittered dots; averages of three replicates. The number of VSGs is: metacyclic ES-VSGs (M-ES) = 5, bloodstream ES-VSGs (B-ES) = 11, minichromosomal VSGs (MC) = 20. The horizontal bars indicate the mean for each group. Values from [S1 Data](#). (E) The graph shows the length of the activated VSG set as jittered dots. ES-VSGs, $n = 16$; MC-VSGs $n = 20$. The summary of the data is shown as a boxplot, with the box indicating the IQR, the whiskers showing the range of values that are within $1.5 \times \text{IQR}$ and a horizontal line indicating the median. The notches represent for each median the 95% confidence interval (approximated by $1.58 \times \text{IQR} / \sqrt{n}$). The p value was derived using a two-tailed t -test.

<https://doi.org/10.1371/journal.ppat.1011530.g002>

VSGs. Recognizing the challenge of working with a complex gene family, we inspected the read-mapping profiles for evidence of expression of the full coding sequence (CDS). VSGs for which mapped reads were restricted to the conserved C-terminal sequence (see [S1 Fig](#)) were removed. This yielded thirty-six activated VSGs ([S2 Fig](#)) that were readily detected by RNA-seq in our switched populations (see [Materials and Methods](#)); five metacyclic ES-VSGs, eleven polycistronic bloodstream ES-VSGs and twenty MC-VSGs. The read-mapping profiles for three example ES-VSGs are shown in [Fig 2C](#) (VSGs 3, 6 and 8 in BES7, 3 and 12, respectively).

Taking the N-terminal 400 amino acids predicted for each VSG, which comprises the divergent domain, and the C-terminal 100 amino acids, we constructed phylogenetic trees, which revealed a highly divergent set of N-terminal sequences ([S2 Fig](#)). Based on these analyses, and to improve mapping fidelity, we took only the first 1.2 kbp of each VSG CDS and remapped the RNA-seq data ([S1 Data](#)). An assessment of relative read-count in our earliest RNA-seq samples (day 5) then revealed differential activation frequencies for each VSG and for each cohort of VSGs ([Fig 2D](#)). Bloodstream ES-VSGs were on average eight times more likely than MC-VSGs to be activated and nineteen times more likely than metacyclic ES-VSGs to be activated. Prior to further analysis, we also assessed VSG length distribution, since a prior study [[19](#)] suggested a role for VSG length in driving antigenic variation dynamics in a mouse model [[5](#)]. The VSG length range seen in the set we detected by RNA-seq extends from 437 to 537 amino acids, a 23% differential ([Fig 2E](#)). Notably, ES-VSGs were significantly longer than MC-VSGs; seven of the eight longest VSGs (>515 aa) were from ESs, while the five shortest VSGs were from minichromosomes. We also note prior analysis here, indicating that all ESs appear to encode unique VSGs [[8](#)], and suggesting that cells with a duplicated VSG present in two ESs are often subsequently lost from the population or modified.

Switched populations expressing MC-derived VSGs display a competitive advantage

To investigate VSG dynamics in switched populations, we next turned to analysis of the full 20-day time-course, using RNA-seq to quantify those switched variants present in the population. A principal component analysis, monitoring read-counts for the activated set of VSGs, showed that the population was dynamic, and also that the three technical replicates followed a similar and remarkably consistent path ([Fig 3A](#)). A striking pattern emerged when we compared the two major groups of VSGs; cells expressing VSGs derived from ESs were relatively diminished in abundance over time, while VSGs derived from minichromosomes increased in abundance over time; again, the behavior of the three populations was remarkably consistent ([Fig 3B](#)). Thus, the transition from ES-VSGs to MC-VSGs, previously thought to be driven by the adaptive immune response, in fact occurs in the absence of an adaptive immune response. Notably, these findings are consistent with our observation above regarding the diversity of VSGs in ESs, which suggested that cells with a duplicated VSG present in two ESs are subsequently lost from the population or modified. An analysis of individual VSGs from each group during the time-course revealed the underlying source of these trends in more detail. Indeed, ten of thirteen VSGs that initially diminished in abundance were derived from ESs; χ^2 test $p = 0.005$ ([Fig 3C](#)).

To determine whether competition among variants, and the competitive advantage displayed by populations of cells expressing MC-VSGs, were reproducible, we generated another independent 'sgRNA break-site 3' strain and again induced VSG switching as described above. RNA-seq analysis was carried out using triplicate switched populations, from which mRNA samples were prepared on day-5 and on day-25 after switching was induced. A comparison of read-counts for the activated set of VSGs on day 5 revealed remarkably similar activation

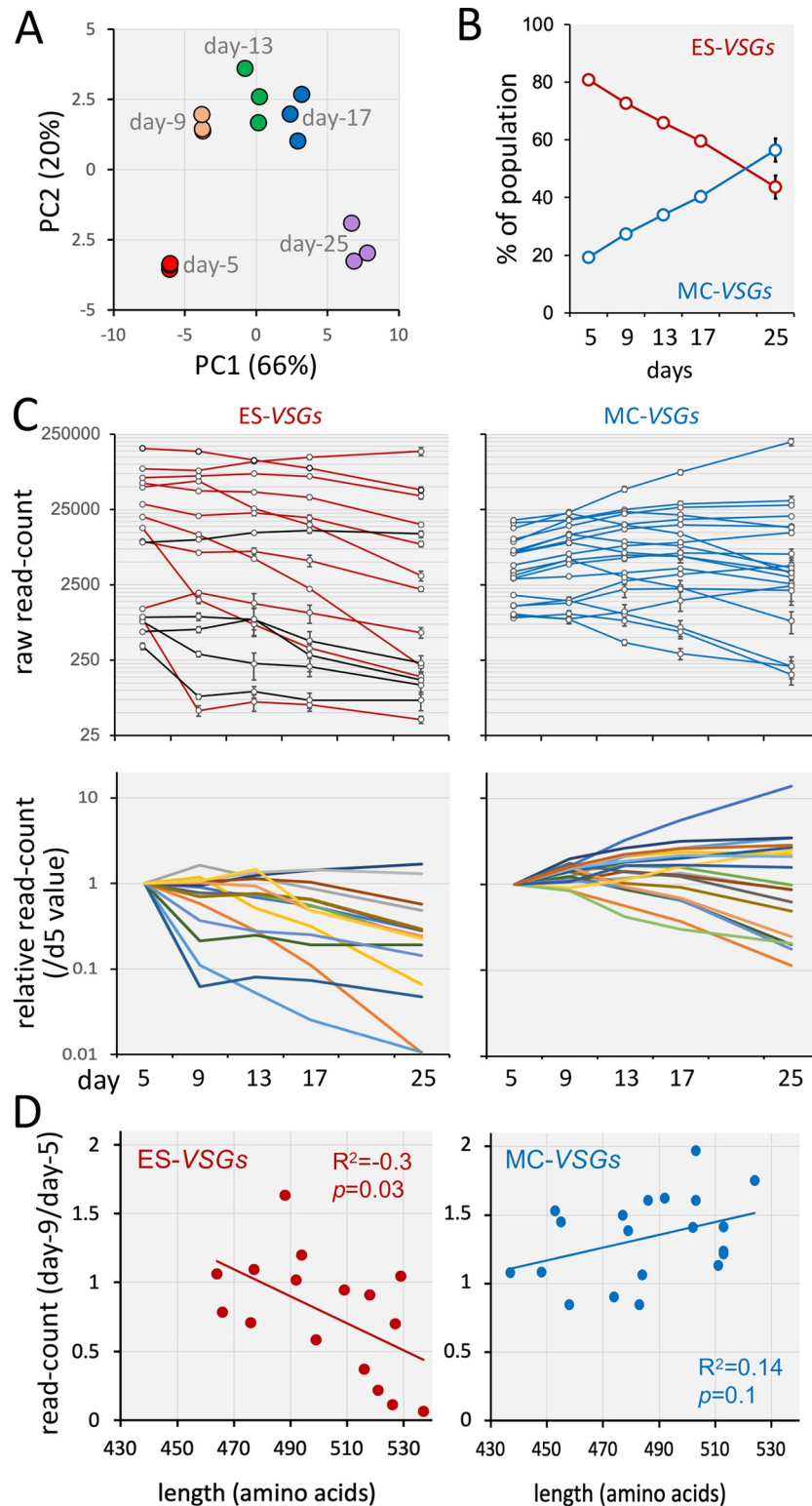


Fig 3. VSG dynamics are related to VSG template location and VSG length. (A) A principal component analysis of read-counts (S1 Data) for the activated VSGs (n = 36) in each replicate during the RNA-seq time-course. (B) Relative read-counts for ES-VSGs (n = 16) and MC-VSGs (n = 20) over the RNA-seq time-course. Error bars, SD; three replicates. (C) Read-counts for individual ES-VSGs and MC-VSGs over the RNA-seq time-course. The upper panels show bloodstream ES-VSGs, red; metacyclic ES-VSGs, black; MC-VSGs, blue. Error bars, SD; three replicates. The

lower panels show the same dataset but displayed relative to the day-5 values to emphasize the different behavior for the two sets of VSGs. (D) The graphs show fold-change in read-counts for either ES-VSGs or MC-VSGs between the day-5 and day-9 time-points relative to VSG length. R_2 and p values were derived using regression analysis in Excel.

<https://doi.org/10.1371/journal.ppat.1011530.g003>

frequencies for individual VSGs ($R^2 = 0.47$; $p = 4^{-6}$), those derived from both expression-sites and minichromosomes (S3 Fig). A comparison of read-counts over the 20-day timeframe also revealed a remarkably similar pattern of relative growth for populations expressing individual VSGs ($R^2 = 0.89$; $p = 8^{-18}$), again including those derived from both expression-sites and minichromosomes (S3 Fig).

Given the similar activation rates and subsequent growth rates observed above, we suspected that analysis of the second independent 'sgRNA break-site 3' strain would also reveal diminished abundance over time for cells expressing VSGs derived from ESs and increased abundance over time for VSGs derived from minichromosomes, as observed in Fig 3B. This was indeed the case, and once again, the behavior of the three replicate populations was remarkably consistent (S3 Fig). An analysis of individual VSGs from each group revealed the underlying source of these trends in more detail (S3 Fig). We concluded that competition among variants, and the competitive advantage displayed by populations of cells expressing minichromosome derived VSGs, were highly reproducible in independent biological replicates.

To determine whether growth differences could also be observed in isolated clones, we compared a switched clone expressing an ES-VSG (VSG-3) and three switched clones expressing MC-VSGs; two expressing VSG-1 and one expressing VSG-23. Cells were seeded at 10^5 / ml in duplicate, split each day, and counted after three days. Doubling times were 6.3 h, except for one of the VSG-1 expressing clones, which displayed a doubling time of 6.5 h (S4 Fig). We next mixed VSG-3 expressing cells 50:50 with each of the other clones and counted the proportion of VSG-3 negative cells by microscopy over two weeks of competitive growth. In all three cases, the VSG-3 negative, MC-VSG expressing cells increased to 70–75% of the population, within only 3-days in three cases (S4 Fig). These results suggest “that the growth rate of clones of trypanosomes grown individually may not correlate with their relative growth rates when grown in mixtures”, as stated more than forty years ago [30], and as also observed in an even earlier independent study [22].

VSG dynamics in switched populations transiently correlate with VSG length

We were interested in the possibility that trypanosomes expressing different-length VSGs grow at different rates [19]. To determine whether there was a relationship between VSG expression dynamics and VSG-length, we assessed changes in VSG abundance at the earliest time-points. When considering the full set of thirty-six VSGs, we saw no significant correlation between length and abundance changes ($p = 0.1$); perhaps unsurprising given the distinct behavior of ES-VSGs and MC-VSGs (Fig 3B and 3C). In contrast, we saw an inverse and significant correlation between length and abundance changes for ES-VSGs ($R^2 = -0.3$; $p = 0.03$) and a weak correlation between length and abundance changes for MC-VSGs ($R^2 = 0.14$; $p = 0.1$) (Fig 3D). Other parameters failed to achieve a significant correlation with VSG abundance changes for either cohort of VSGs; including VSG mass (Fig 4A), number of predicted N-glycosylation sites per VSG (Fig 4B) or number of Cys residues per VSG (Fig 4C). VSG activation frequency (Fig 4D) and VSG abundance changes between day-9 and day-13 (Fig 4E) also failed to achieve a significant correlation with VSG length for either cohort of VSGs. These results

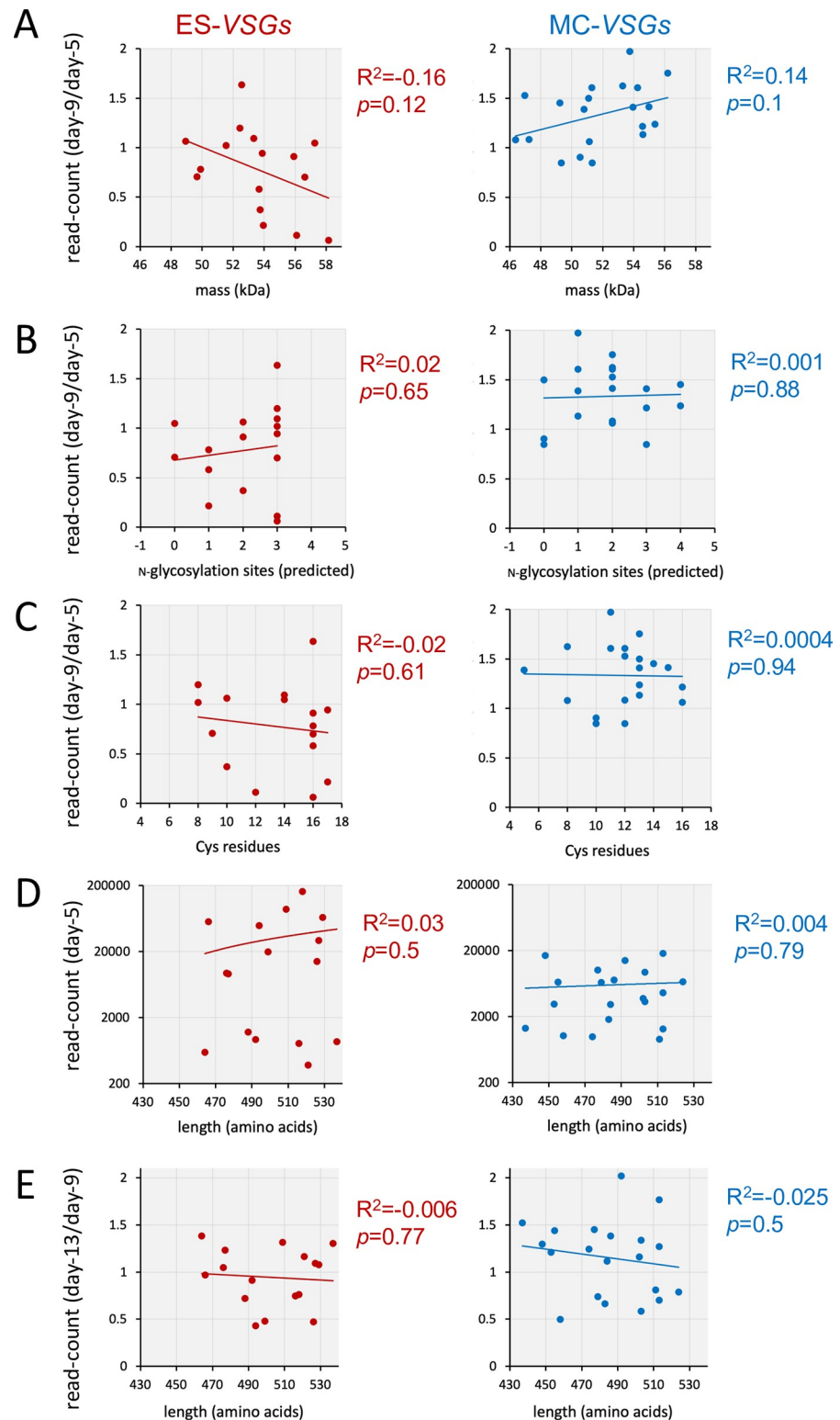


Fig 4. VSG dynamics and relationship to other features of VSGs. (A) The graphs show fold-change in read-counts for VSGs between the day-5 and day-9 time-points relative to VSG mass. (B) As in A but relative to predicted N-glycosylation sites. (C) As in A but relative to number of Cys residues. (D) The graphs show read-counts for VSGs at the day-5 time-point (a measure of VSG activation rate) relative to VSG length. (E) The graphs show fold-change in read-counts for VSGs between the day-9 and day-13 time-points relative to VSG length. Further comparisons between

the day-13/17 and day-17/25 time-points and VSG length yielded R_2 values <0.1 and p values >0.3 . All R_2 and p values were derived using regression analysis in Excel.

<https://doi.org/10.1371/journal.ppat.1011530.g004>

suggest that the length of the expressed VSG does indeed impact trypanosome growth rate, as suggested previously [19], based on an analysis of VSG population dynamics *in vivo* [5].

Differential growth is associated with wider transcriptome differences

We next asked whether contemporary switched cells growing at different rates displayed transcriptome differences that extended beyond the active VSG. A principal component analysis, monitoring the transcriptome over the 20-day time-course, again revealed a dynamic population, and also a remarkably consistent path for the three technical replicates (Fig 5A). Notably, the wider transcriptome appeared to diverge only transiently. A comparison of RNA-seq data from the day-5 samples relative to either the day-9 or day-25 samples confirmed this view (Fig 5B) and revealed significant differences in gene expression at earlier time-points, as highlighted by transiently reduced expression of genes encoding nucleolar proteins, tRNA synthetases and glycolytic enzymes (Fig 5B). These results supported the hypothesis that differential growth is indeed associated with wider changes in the transcriptome.

To further explore this hypothesis, we generated a panel of switched clones by once again inducing DNA breaks at sgRNA site-3 (see Fig 1B). Among thirty-six independent clones analyzed, we identified clones expressing several different ES-VSGs (VSG-3, 6, 8, 9, 11, 13, 17, 631) and MC-VSGs (VSG-24, 25, 476, 832, 3039), as determined by VSG-generic RT-PCR and Sanger sequencing (see Fig 5C). A comparison with results shown in Fig 2D suggests $>30\%$ under-sampling of clones that activate ES-VSGs here, potentially reflecting declining growth-rates that are incompatible with clone expansion in some cases. Since the changes we sought to characterize may be transient and/or may result in the loss of clones that activate ES-VSGs, we assessed growth rates at the earliest time-point possible; 5 days after subcloning. The analysis revealed different growth rates associated with different VSGs (Fig 5C), and based on the results, six clones were selected for RNA-seq analysis; four clones that activated ES-VSGs, two slower-growing (VSG-6 or VSG-8) and two faster-growing (VSG-9 or VSG-11), and two faster-growing clones that activated MC-VSGs (VSG-25 or VSG-832). The relative growth of these clones is shown in Fig 5C (right-hand panel).

RNA-seq analysis confirmed loss of VSG-2 (no reads mapped) and exclusive expression of the expected new VSG by each clone (av. 99.6% of VSG reads mapped). A principal component analysis, comparing the wider transcriptomes of these clones, revealed clustering of the faster-growing clones, while the pair of slower-growing clones were outliers (Fig 5D). Indeed, the slower-growing clones displayed divergent transcriptome profiles, that correlated with growth-rate, and also the signatures described above, characterized by negative correlation between growth-rate and relative expression of nucleolar proteins, glycolytic enzymes and tRNA synthetases (Fig 5E). These results support the view that activation of an ES-VSG, when compared to activation of a minichromosomal VSG, often results in relatively slow growth that is correlated with transcriptome remodeling.

Discussion

African trypanosomes are remarkably persistent parasites, which can maintain infections in the bloodstream of immune-competent hosts for years. In experimental *T. brucei* infections, a VSG hierarchy is observed, which can prolong infection, but how this arises is not fully understood. Although competition among VSG variants was reported more than forty years ago [22,

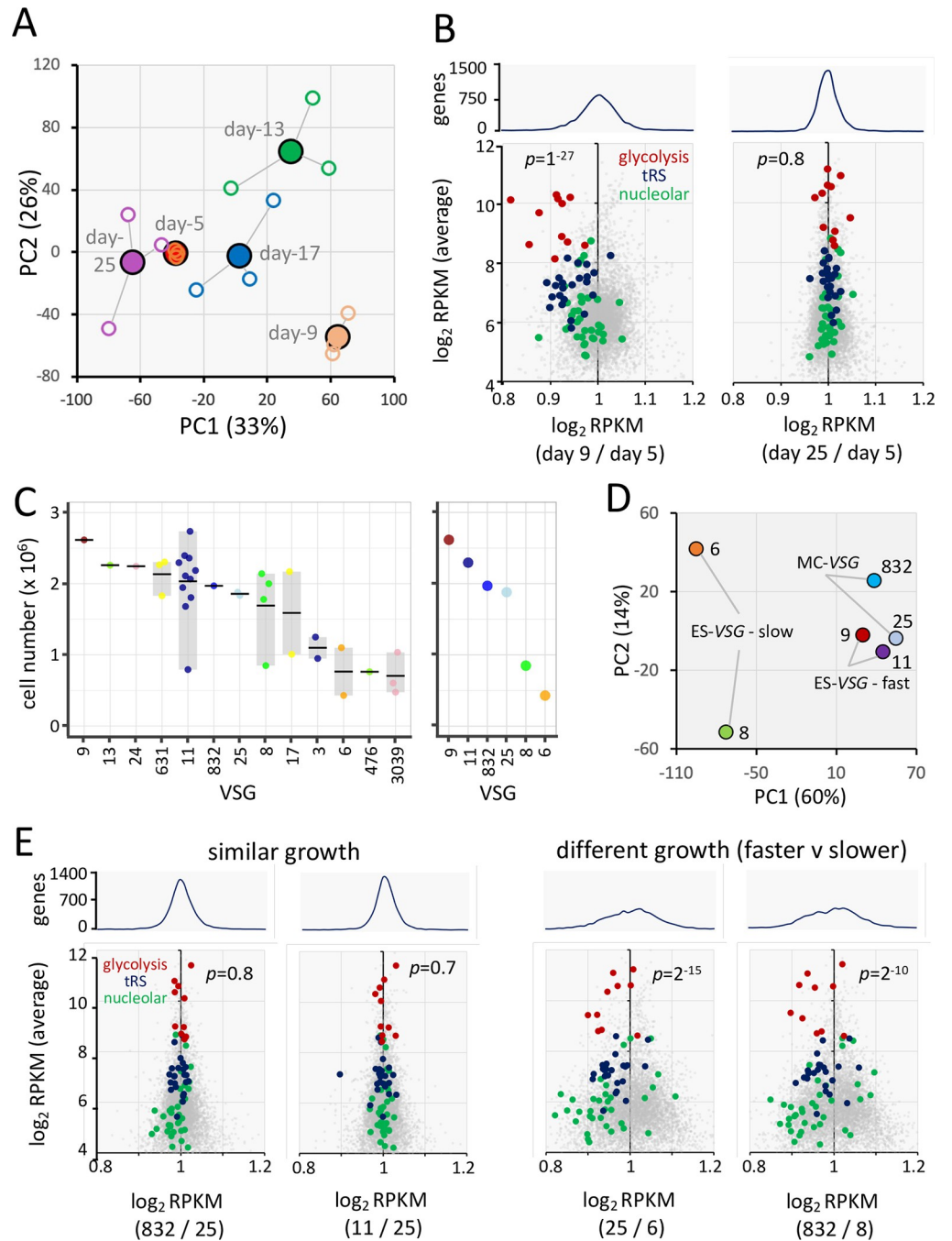


Fig 5. Differential growth is associated with wider transcriptome differences. (A) A principal component analysis of \log_2 RPKM (reads per kilobase per million) values (S1 Data) for the transcriptome ($n = 7255$) during the RNA-seq time-course. The closed datapoints represent averages while the open datapoints represent each replicate. (B) The scatterplots show transcriptome differences between the day-5 and either day-9 or day-25 samples, highlighting transcripts encoding nucleolar proteins ($n = 35$), tRNA synthetases ($n = 23$) or glycolytic enzymes ($n = 11$). p values were derived using a χ^2 test for increased v decreased expression of these 69 genes. The upper panels indicate data distribution. (C) The graphs show cell numbers for 36 clones and six of those clones selected for RNA-seq (right-hand panel), also indicating the VSG expressed by each clone. The data are shown as jittered dots and, when $n > 1$, the horizontal line indicates the mean and the box indicates the range of values. (D) A principal component analysis of \log_2 RPKM values (S1 Data) for the transcriptomes of the six selected clones. (E) The scatterplots show transcriptome differences between the pairs of clones indicated. Other details as in B above.

<https://doi.org/10.1371/journal.ppat.1011530.g005>

30], more recent studies have focused on switch-rates rather than subsequent growth rates. One recent study did report changes in the average length of VSGs expressed in a mouse model, however, pointing to a role in differential growth [19]. Our goal was to experimentally assess factors that contribute to the VSG hierarchy, specifically to assess the hypothesis that cells expressing different VSGs compete. Using inducible Cas9 to introduce programmed DNA breaks at the active VSG-ES, we generated a highly heterogeneous population of cells expressing different VSGs. VSG dynamics were then monitored using RNA-seq. We show that variants do indeed compete. Thus, parasite-intrinsic factors that drive population structure are more sophisticated than previously appreciated.

Our Cas9-based approach allowed us to generate a complex population of cells expressing different VSGs at the telomeric end of the same single-copy expression site on chromosome 6a. RNA-seq analysis then provided a measure of relative VSG activation frequency, at the earliest post-switching time-point, and relative growth rate, at subsequent time-points. We detected activation of sixteen distinct ES-VSGs and twenty distinct MC-VSGs which could be effectively monitored during the time-course. In terms of relative VSG activation frequency, we were able to monitor three distinct groups of VSGs; those within polycistronic ESs, which typically have long tracts of 70-bp repeats [8]; those within monocistronic ESs [40], which typically have short tracts of 70-bp repeats; and minichromosomal or MC-VSGs [38], which also typically have short tracts of 70-bp repeats [42]. Although the length of 70-bp repeat tracts remains unknown at many sites, we found that the frequency of VSG activation was broadly consistent with the previously reported lengths of these tracts [8,38,40,42]. Indeed, VSG-14 and VSG-15, the polycistronic ES-associated VSGs reported to have the shortest adjacent tracts of 70-bp repeats [8], were activated at a lower frequency than any other polycistronic ES-associated VSG. Thus, recombination efficiency appears to be dependent on the length of 70-bp repeat substrate available; as is also the case for homologous recombination amongst other *T. brucei* sequences [43]. Mathematical modelling indicated that VSG length-dependent activation rates have the potential to allow *T. brucei* to persist for moderately longer *in vivo* [19]. Our findings, however, do not provide support for VSG length-dependent activation rates.

Our time-course analysis revealed competition among *T. brucei* cells expressing VSGs of different lengths. The set of VSGs analyzed here ranged from 437 to 537 amino acids. Relatively few of the annotated VSGs have been shown to be functional, but a VSGnome collection comprises 325 apparently intact genes ranging in length from 436–582 amino acids [38]. This 33%, 146 amino acids differential suggests that the capacity for VSG-length dependent differential growth may be greater than sampled here. We suggest that differential VSG activation-frequency and differential growth both contribute to prolonging persistence *in vivo*; VSG flanking homology (70-bp repeats in particular) primarily determines the activation frequency while VSG template location, and also VSG length, determine subsequent growth rate. Further transcriptome analysis revealed a signature associated with differential growth. Specifically, recently switched, and slower-growing trypanosomes expressing ES-VSGs displayed increased expression of genes involved in nucleolar function, protein translation and energy metabolism.

Why do cells that activate ES-VSGs grow slower than cells that activate MC-VSGs? We suggest a mechanism based on homology-dependent interference, proposed previously as a mechanism that promotes allelic exclusion and transcriptional repression [44]. An active VSG that is identical to the repressed ES-located VSG from which it was copied may itself be subject to competition and interference. Indeed, all ES-located VSGs appear to be unique [8], supporting the view that cells with pairs of identical ES-associated VSGs, although they may arise often, are typically lost from the population, or modified. This feature may also present a potential route for assembling mosaic VSGs, recognized as important for sustaining more persistent

infections. It has remained unclear how mosaic VSGs are assembled through multiple recombination steps without disrupting VSG function; especially given the high proportion of VSG pseudogenes and gene fragments present in the genome [45]. Our findings now suggest a solution to this conundrum. Cells that activate ES-VSGs may escape immune surveillance due to homology-dependent interference and slower growth, as detailed above. These populations may then circulate at low density for sufficient periods of time to allow for the assembly of mosaic VSGs. Cells expressing defective mosaic VSGs would be rapidly eliminated, since continuous VSG expression is required for proliferation and viability [46], while functional mosaics would support continued growth. Notably, mosaic formation would also relax the homology-dependent interference proposed above. Our findings can also explain why the average ES-VSG is significantly longer than the average MC-VSG; because cells expressing longer MC-derived VSGs (that are not present in a second expression site) have a competitive advantage.

It has been known since the 1960s that VSGs emerge in a hierarchical order during a *T. brucei* infection [21,23], but our understanding of the mechanisms underpinning this behavior has remained incomplete. We now show that the activated VSGs themselves, their locus of origin in the genome and their length, can make important contributions to the hierarchy, thereby supporting a more persistent infection than would otherwise be possible [19]. Notably, competition among variants was reported more than forty years ago [22,30], and we report similar competition here. In particular, we find that populations of trypanosomes that activate MC-VSGs exhibit a competitive advantage, highlighting the importance of the minichromosomal repertoire for establishing an infection. Thus, we suggest that a complex set of VSG variants emerges, but that relatively few achieve a sufficient density to trigger an effective immune response. This represents a feint attack strategy [19], whereby *T. brucei* diverts the host acquired immune response with faster-growing variants, allowing slower-growing variants to persist before eventually emerging from below-threshold sub-populations. Such an economical use of VSGs allows *T. brucei* to stay one step ahead of the host immune response. These behaviors likely delay the exhaustion of available VSGs and help to explain why *T. brucei* is such a persistent parasite.

Materials and methods

T. brucei growth and manipulation

T. brucei 2T1-T7-Cas9 cells [33] were grown in HMI-11 medium at 37°C and 5% CO₂, with hygromycin (1 µg/mL) and blasticidin (2 µg/mL). Transfections were performed using electroporation. sgRNA expression cassettes 1–3 were introduced into *T. brucei* as described [33]; in the absence of tetracycline. In this case, transformed clones were selected in hygromycin, blasticidin and phleomycin (2 µg/mL).

Plasmid construction

The sgRNA constructs 1–3 were assembled as described [33] and using the following oligonucleotides: gRNA1fw (AGGGATGGATTTGAAGTGATGATG), gRNA1rv (AAACCATCATC ACTTCAAATCCAT); gRNA2fw (AGGGAGAATTCGAGACGAAGAGC), gRNA2rv (AAACGCTCTTCGTCTCCGAATTCT); gRNA3fw (AGGGTGCCCTTCCAATCAGGAGGCC), gRNA3rv (AAACGGCCTCCTGATTGGAAGGCA). Correct assembly was confirmed by sequencing.

Western blotting

Protein blotting was carried out according to standard protocols. PBS-T-M (PBS, 0.1% Tween-20, 5% milk powder) was used to block the membrane for 1 h. The >70 kDa part of the membrane was used for Cas9 detection with a mouse α -Cas9 primary antibody (1:1000, Abcam) and α -mouse HRP secondary antibody (1:2000, BioRad), and the <70 kDa part of the membrane was used for EF1 α detection with a mouse α -EF1 α primary antibody (1:20000, Merck Millipore) and an α -mouse HRP secondary antibody (1:2000, BioRad). Primary antibodies were incubated at 4°C overnight and secondary antibodies were incubated for 1 h at RT. Membranes were developed using ECL solution (GE healthcare) and images were acquired using a ChemiDoc XRS+ (BioRad).

Immunofluorescence microscopy

Cells were washed in PBS and fixed in 3% formaldehyde in PBS at 37°C for 10 min, plus 5 min at RT. Following washing in PBS, 1% BSA and H₂O, the cells were resuspended in 30 μ L 1% BSA. 5 μ L of a suspension was placed on each well of a 12-well, 5 mm slide (Thermo scientific) and dried overnight at RT. Cells were then re-hydrated for 5 min in PBS and blocked for 10 min with 50% FBS. Following two washes in PBS, cells were incubated in primary antibody for 1 h at RT; rat α -VSG-2 (1:10,000), rabbit α -VSG-6 (1:10,000), mouse α -VSG-3 coupled with Alexa 488 (1:500). Following three washes in PBS, cells were incubated in secondary antibody for 1 h at RT; α -rat rhodamine (1:2000 Pierce), α -rabbit Alexa 488 (1:2000, Life Technologies). Cells were washed three more times in PBS and mounted in medium containing DAPI (Vecta-Shield). Microscopy was performed using a Zeiss Axiovert 200M inverted light microscope with a magnification of 10 x 60 with oil immersion. Images were acquired using ZenPro (Carl Zeiss, Germany) and the same parameters and then equally processed using Fiji v1.5.2e.

Cas9-induced antigenic variation

T. brucei 2T1-T7-Cas9 cells with the appropriate sgRNA, as described above, were induced to express Cas9 by the addition of tetracycline (1 μ g/mL). For the RNA-seq time-course (see below), cells expressing sgRNA-3 were grown in tetracycline for 3-days (1 μ g/mL) and then washed in HMI-11 to remove tetracycline. This culture was split into three parallel cultures on day 5 following tetracycline addition and the cells were subsequently split to 1 x 10⁴ cells/mL in 50 mL every two days. Approximately 10 million cells were collected from each of the three flasks for RNA extraction using an RNeasy Mini Kit (Qiagen) on days 5, 9, 13, 17 and 25 following tetracycline addition, or on days 5 and 25 only for the second 'sgRNA-3' biological replicate; samples were taken for immunofluorescence microscopy in parallel. Following cell lysis, samples were stored at -80°C and further processed in parallel. Individual switched clones were generated by inducing *T. brucei* 2T1-T7-Cas9 cells with sgRNA-3 as above. Following tetracycline removal, the cells were allowed to recover for 2-days, and cells were then subcloned by limiting dilution in 500 μ L of HMI-11 in 48 well-plates. Five days after subcloning 100 μ L of medium was collected for each clone and VSG-2 expression was assessed by immunofluorescence microscopy. The number of parasites in each well for those clones that no longer expressed VSG-2 was quantified using a CASY Cell Counter (Cambridge Bioscience). These clones were grown for a further 2-days for RNA extraction as described above. Additional clones were selected for growth analysis in isolation, and in competition, using immunofluorescence microscopy.

RNA sequencing and analysis

Briefly, for RNA-seq, polyadenylated transcripts were enriched using polydT beads and reverse-transcribed. For the time course, sequencing was on a HiSeq 4000 platform (Illumina): approximately 40 million 100 b paired-end reads were generated per sample (BGI, Hong-Kong). For analysis of switched clones, and for analysis of the second 'sgRNA-3' biological replicate, sequencing was on a DNBSEQ-G400 platform (MGI): approximately 70 million 100 b paired-end reads were generated per sample (BGI, Hong-Kong). The forward and reverse paired-end reads were aligned to the *T. brucei* TREU927 reference genome, v55 downloaded from TriTrypDB [47], and to the set of VSG-ESs from the *T. brucei* Lister 427 strain [8], using bowtie2 [48], with the 'very-sensitive-local' pre-set alignment option. The alignments were converted to BAM format, reference sorted and indexed with SAMtools [49]. Fragment counts were determined from the BAM files using featureCounts [50] with parameters: -p (pair end) -B (both ends successfully aligned) -C (skip fragments that have their two ends aligned to different chromosome) -M (count multi-mapping) -O (match overlapping features) -t CDS (count level) -g gene_id (summarization level). Genes with low counts were filtered out using edgeR [51] and RPKM values were extracted with cqn.

The alignments to the set of '1,200 bp' truncated VSGs were filtered for MapQ > 1 with SAMtools. We considered a set of ES-VSG sequences, and a set of MC-VSG sequences [8,37,38,40], and only considered these VSGs to be activated if they registered >10-fold the proportion of VSG reads typically observed for silent VSGs; these 'background' values were 0.004% for B-ES VSGs, 0.002% for M-ES VSGs, and 0.00003% for MC-VSGs. Average proportions for activated VSGs in the day-5 samples were 2.8% for B-ES VSGs, 0.15% for M-ES VSGs, and 0.37% for MC-VSGs; 671-, 65- and 14,195-fold above 'background', respectively. Additional VSGs in preliminary analyses included subtelomeric array VSGs (VSG-4 and VSG-5), but these VSGs failed to register sufficient reads to qualify as activated. The scripts to align the RNA-seq data and to compute read-counts have been deposited in GitHub and Zenodo. To generate base-pair resolution plots, reads per kilobase per million were extracted using deepTools [52], plotted on Prism9 and then annotated and edited in Illustrator. Sanger sequencing of VSG mRNA was carried out in-house. Total RNA samples (1 µg) were used for first strand synthesis using a primer that anneals to all VSG 3'-UTRs (DH3, GACTAGTGTAAAATA-TATCA) and M-MLV reverse transcriptase (Promega). VSG cDNA was then generated by PCR using hot start Taq DNA polymerase (NEB), DH3 primer, and a primer for the spliced leader sequence (DH5, GACTAGTTTCTGTAATAT). PCR steps were five cycles of 30 s at 94°C, 30 s at 50°C, and 1 min at 72°C, followed by 13 cycles of 30 s at 94°C, 30 s at 55°C, and 1 min at 72°C. PCR products were resolved on agarose gels and cDNA was extracted using a QIAquick Gel Extraction Kit (Qiagen). VSG sequencing was performed using Applied Biosystems Big-Dye Ver 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer.

VSG annotation

Signal peptides were predicted using the SignalP-6.0 tool at <https://services.healthtech.dtu.dk/service.php?SignalP> or earlier iterations (SignalP-5.0, or SignalP-3.0). GPI modification sites were predicted using the big-PI predictor at https://mendel.imp.ac.at/gpi/gpi_server.html. N-glycosylation sites were predicted using NetNGlyc at <https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0>. Phylogenetic analysis was carried out using ETE 3 at <https://www.genome.jp/tools-bin/ete> [53].

Supporting information

S1 Fig. Filtering of the minichromosomal VSG set. The VSG set was filtered based on RNA-seq read-mapping profiles. VSGs displaying full coding sequence activation were included in the analysis. VSGs with mapped reads restricted to the C-terminal coding sequence were excluded from further analysis. These results are explained by the presence of common sequences in multiple VSGs. The 3'-terminal 268 nucleotides of VSG-444 are shared with VSG-18, while the 3'-terminal 584 nucleotides of VSG-514 are shared with VSG-17, for example. VSGs included in the analysis were truncated to 1,200 bp to remove shared sequence. (TIF)

S2 Fig. Annotation and phylogenetic analysis of the activated VSG set. (A) The schematic shows the activated set of VSG proteins with salient features highlighted. Signal peptides were predicted using the SignalP tool. GPI modification sites were predicted using the big-PI predictor. N-glycosylation sites were predicted using NetNGlyc. O-glycosylation sites are from [32]. VSG domain types are indicated. (B) The phylogenetic tree shows the activated set of VSGs analyzed using ETE 3 [53]; the N-terminal 400 amino acids only. (C) As in B but for the C-terminal 100 amino acids only. An asterisk indicates undefined VSG C-terminal type. (TIF)

S3 Fig. Reproducibility of VSG activation and MC-VSG associated competitive advantage. (A) The graph shows VSG expression levels as determined by RNA-seq read-count on day 5 after inducing switching in two independent biological replicate strains. $n = 36$ VSGs. (B) The graph shows relative VSG expression levels as determined by relative RNA-seq read-count over 20 days of growth following switching in two independent biological replicate strains. $n = 36$ VSGs. (C) Relative read-counts for ES-VSGs ($n = 16$) and MC-VSGs ($n = 20$) at day-5 and day-25 in the second replicate strain, and as determined by RNA-seq; three replicates, error bars, SD (not visible). (D) Read-counts for individual ES-VSGs ($n = 16$) and MC-VSGs ($n = 20$) at day-25 relative to the day-5 values in the second replicate strain. (TIF)

S4 Fig. Growth and competitive growth of individual clones. (A) The graph shows doubling times calculated for individual clones expressing the VSGs indicated. Duplicate flasks were counted after three days of growth; quadruplicate flasks in the case of the VSG-3 expressing clone. The horizontal bars indicate average values. (B) The VSG-3 expressing cells were mixed 50:50 with each of the other clones shown in A. Cells were then stained with α VSG-3 on the days indicated, and the proportion of VSG-3 negative cells were counted by microscopy. (TIF)

S1 Data. RNA-seq data. Read-counts and RPKM values for the time course, for the set of six switched clones, and for the second replicate clone. Excel format. (XLSX)

Acknowledgments

We thank E. Rico for advice on use of the inducible Cas9 system; S. Duncan and M. Ferguson for advice on N-glycosylation and S. Hutchinson (Institut Pasteur, Paris), N. Siegel and R. Cosentino (LMU, Munich) for advice on VSG sequences. We also thank the Sequencing & Services team (MRC PPU, School of Life Sciences, University of Dundee, Scotland).

Author Contributions

Conceptualization: David Horn.

Data curation: Michele Tinti, Joana Faria, David Horn.

Formal analysis: Douglas O. Escrivani, Viktor Scheidt, Michele Tinti, Joana Faria.

Funding acquisition: David Horn.

Investigation: Douglas O. Escrivani, Viktor Scheidt.

Methodology: Douglas O. Escrivani, Viktor Scheidt.

Supervision: David Horn.

Visualization: Michele Tinti, Joana Faria.

Writing – original draft: David Horn.

Writing – review & editing: Douglas O. Escrivani, Viktor Scheidt, Michele Tinti, David Horn.

References

1. Trindade S, De Niz M, Costa-Sequeira M, Bizarra-Rebello T, Bento F, Dejung M, et al. Slow growing behavior in African trypanosomes during adipose tissue colonization. *Nat Commun.* 2022; 13(1):7548. Epub 2022/12/09. <https://doi.org/10.1038/s41467-022-34622-w> PMID: 36481558; PubMed Central PMCID: PMC9732351.
2. Horn D. Antigenic variation in African trypanosomes. *Mol Biochem Parasitol.* 2014; 195(2):123–9. Epub 2014/05/27. <https://doi.org/10.1016/j.molbiopara.2014.05.001> PMID: 24859277; PubMed Central PMCID: PMC4155160.
3. Manna PT, Boehm C, Leung KF, Natesan SK, Field MC. Life and times: synthesis, trafficking, and evolution of VSG. *Trends Parasitol.* 2014; 30(5):251–8. Epub 2014/04/16. <https://doi.org/10.1016/j.pt.2014.03.004> PMID: 24731931; PubMed Central PMCID: PMC4007029.
4. Hall JP, Wang H, Barry JD. Mosaic VSGs and the scale of *Trypanosoma brucei* antigenic variation. *PLoS Pathog.* 2013; 9(7):e1003502. Epub 2013/07/16. <https://doi.org/10.1371/journal.ppat.1003502> PMID: 23853603; PubMed Central PMCID: PMC3708902.
5. Mugnier MR, Cross GA, Papavasiliou FN. The in vivo dynamics of antigenic variation in *Trypanosoma brucei*. *Science.* 2015; 347(6229):1470–3. Epub 2015/03/31. <https://doi.org/10.1126/science.aaa4502> PMID: 25814582; PubMed Central PMCID: PMC4514441.
6. Aresta-Branco F, Erben E, Papavasiliou FN, Stebbins CE. Mechanistic similarities between antigenic variation and antibody diversification during *Trypanosoma brucei* infection. *Trends Parasitol.* 2019; 35(4):302–15. Epub 2019/03/04. <https://doi.org/10.1016/j.pt.2019.01.011> PMID: 30826207.
7. Faria J, Luzak V, Muller LSM, Brink BG, Hutchinson S, Glover L, et al. Spatial integration of transcription and splicing in a dedicated compartment sustains monogenic antigen expression in African trypanosomes. *Nat Microbiol.* 2021; 6(3):289–300. Epub 2021/01/13. <https://doi.org/10.1038/s41564-020-00833-4> PMID: 33432154; PubMed Central PMCID: PMC7610597.
8. Hertz-Fowler C, Figueiredo LM, Quail MA, Becker M, Jackson A, Bason N, et al. Telomeric expression sites are highly conserved in *Trypanosoma brucei*. *PLoS One.* 2008; 3(10):e3527. Epub 2008/10/28. <https://doi.org/10.1371/journal.pone.0003527> PMID: 18953401; PubMed Central PMCID: PMC2567434.
9. Boothroyd CE, Dreesen O, Leonova T, Ly KI, Figueiredo LM, Cross GA, et al. A yeast-endonuclease-generated DNA break induces antigenic switching in *Trypanosoma brucei*. *Nature.* 2009; 459(7244):278–81. Epub 2009/04/17. <https://doi.org/10.1038/nature07982> PMID: 19369939; PubMed Central PMCID: PMC2688456.
10. Glover L, Alford S, Horn D. DNA break site at fragile subtelomeres determines probability and mechanism of antigenic variation in African trypanosomes. *PLoS Pathog.* 2013; 9(3):e1003260. Epub 2013/04/05. <https://doi.org/10.1371/journal.ppat.1003260> PMID: 23555264; PubMed Central PMCID: PMC3610638.
11. Sima N, McLaughlin EJ, Hutchinson S, Glover L. Escaping the immune system by DNA repair and recombination in African trypanosomes. *Open Biol.* 2019; 9(11):190182. Epub 2019/11/14. <https://doi.org/10.1098/rsob.190182> PMID: 31718509; PubMed Central PMCID: PMC6893398.

12. Thivolle A, Mehnert AK, Tihon E, McLaughlin E, Dujeancourt-Henry A, Glover L. DNA double strand break position leads to distinct gene expression changes and regulates VSG switching pathway choice. *PLoS Pathog.* 2021; 17(11):e1010038. Epub 2021/11/13. <https://doi.org/10.1371/journal.ppat.1010038> PMID: [34767618](https://pubmed.ncbi.nlm.nih.gov/34767618/); PubMed Central PMCID: PMC8612549 statement: "The authors declare that they have no conflict of interest."
13. Hovel-Miner G, Mugnier MR, Goldwater B, Cross GA, Papavasiliou FN. A conserved DNA repeat promotes selection of a diverse repertoire of *Trypanosoma brucei* surface antigens from the genomic archive. *PLoS Genet.* 2016; 12(5):e1005994. Epub 2016/05/07. <https://doi.org/10.1371/journal.pgen.1005994> PMID: [27149665](https://pubmed.ncbi.nlm.nih.gov/27149665/); PubMed Central PMCID: PMC4858185.
14. Liu AY, Michels PA, Bernardis A, Borst P. Trypanosome variant surface glycoprotein genes expressed early in infection. *J Mol Biol.* 1985; 182(3):383–96. Epub 1985/04/05. [https://doi.org/10.1016/0022-2836\(85\)90198-6](https://doi.org/10.1016/0022-2836(85)90198-6) PMID: [4009712](https://pubmed.ncbi.nlm.nih.gov/4009712/).
15. Robinson NP, Burman N, Melville SE, Barry JD. Predominance of duplicative VSG gene conversion in antigenic variation in African trypanosomes. *Mol Cell Biol.* 1999; 19(9):5839–46. Epub 1999/08/24. <https://doi.org/10.1128/MCB.19.9.5839> PMID: [10454531](https://pubmed.ncbi.nlm.nih.gov/10454531/); PubMed Central PMCID: PMC84433.
16. Jamonneau V, Ilboudo H, Kabore J, Kaba D, Koffi M, Solano P, et al. Untreated human infections by *Trypanosoma brucei gambiense* are not 100% fatal. *PLoS Negl Trop Dis.* 2012; 6(6):e1691. Epub 2012/06/22. <https://doi.org/10.1371/journal.pntd.0001691> PMID: [22720107](https://pubmed.ncbi.nlm.nih.gov/22720107/); PubMed Central PMCID: PMC3373650.
17. Turner CM. The rate of antigenic variation in fly-transmitted and syringe-passaged infections of *Trypanosoma brucei*. *FEMS Microbiol Lett.* 1997; 153(1):227–31. Epub 1997/08/01. <https://doi.org/10.1111/j.1574-6968.1997.tb10486.x> PMID: [9252591](https://pubmed.ncbi.nlm.nih.gov/9252591/).
18. Jayaraman S, Harris C, Paxton E, Donachie AM, Vaikkinen H, McCulloch R, et al. Application of long read sequencing to determine expressed antigen diversity in *Trypanosoma brucei* infections. *PLoS Negl Trop Dis.* 2019; 13(4):e0007262. Epub 2019/04/04. <https://doi.org/10.1371/journal.pntd.0007262> PMID: [30943202](https://pubmed.ncbi.nlm.nih.gov/30943202/); PubMed Central PMCID: PMC6464242.
19. Liu D, Albergante L, Newman TJ, Horn D. Faster growth with shorter antigens can explain a VSG hierarchy during African trypanosome infections: a feint attack by parasites. *Sci Rep.* 2018; 8(1):10922. Epub 2018/07/22. <https://doi.org/10.1038/s41598-018-29296-8> PMID: [30026531](https://pubmed.ncbi.nlm.nih.gov/30026531/); PubMed Central PMCID: PMC6053454.
20. Gray AR. Antigenic variation in clones of *Trypanosoma brucei*. I. Immunological relationships of the clones. *Ann Trop Med Parasitol.* 1965; 59:27–36. Epub 1965/03/01. <https://doi.org/10.1080/00034983.1965.11686278> PMID: [14297353](https://pubmed.ncbi.nlm.nih.gov/14297353/).
21. Myler P, Nelson RG, Agabian N, Stuart K. Two mechanisms of expression of a predominant variant antigen gene of *Trypanosoma brucei*. *Nature.* 1984; 309(5965):282–4. Epub 1984/05/17. <https://doi.org/10.1038/309282a0> PMID: [6325951](https://pubmed.ncbi.nlm.nih.gov/6325951/).
22. Seed JR. Competition among serologically different clones of *Trypanosoma brucei gambiense* in vivo. *J Protozool.* 1978; 25(4):526–9. Epub 1978/11/01. <https://doi.org/10.1111/j.1550-7408.1978.tb04179.x> PMID: [368326](https://pubmed.ncbi.nlm.nih.gov/368326/)
23. Lee MG, Van der Ploeg LH. Frequent independent duplicative transpositions activate a single VSG gene. *Mol Cell Biol.* 1987; 7(1):357–64. Epub 1987/01/01. <https://doi.org/10.1128/mcb.7.1.357-364.1987> PMID: [3031467](https://pubmed.ncbi.nlm.nih.gov/3031467/); PubMed Central PMCID: PMC365076.
24. Timmers HT, de Lange T, Kooter JM, Borst P. Coincident multiple activations of the same surface antigen gene in *Trypanosoma brucei*. *J Mol Biol.* 1987; 194(1):81–90. Epub 1987/03/05. [https://doi.org/10.1016/0022-2836\(87\)90717-0](https://doi.org/10.1016/0022-2836(87)90717-0) PMID: [3612801](https://pubmed.ncbi.nlm.nih.gov/3612801/).
25. Lythgoe KA, Morrison LJ, Read AF, Barry JD. Parasite-intrinsic factors can explain ordered progression of trypanosome antigenic variation. *Proc Natl Acad Sci U S A.* 2007; 104(19):8095–100. Epub 2007/04/28. <https://doi.org/10.1073/pnas.0606206104> PMID: [17463092](https://pubmed.ncbi.nlm.nih.gov/17463092/); PubMed Central PMCID: PMC1876577.
26. McCulloch R, Morrison LJ, Hall JPJ. DNA recombination strategies during antigenic variation in the African trypanosome. *Microbiol Spectr.* 2015; 3(2):MDNA3-0016-2014. Epub 2015/06/25. <https://doi.org/10.1128/microbiolspec.MDNA3-0016-2014> PMID: [26104717](https://pubmed.ncbi.nlm.nih.gov/26104717/).
27. Morrison LJ, Majiwa P, Read AF, Barry JD. Probabilistic order in antigenic variation of *Trypanosoma brucei*. *Int J Parasitol.* 2005; 35(9):961–72. Epub 2005/07/08. <https://doi.org/10.1016/j.ijpara.2005.05.004> PMID: [16000200](https://pubmed.ncbi.nlm.nih.gov/16000200/).
28. Agur Z, Abiri D, Van der Ploeg LH. Ordered appearance of antigenic variants of African trypanosomes explained in a mathematical model based on a stochastic switch process and immune-selection against putative switch intermediates. *Proc Natl Acad Sci U S A.* 1989; 86(23):9626–30. Epub 1989/12/01. <https://doi.org/10.1073/pnas.86.23.9626> PMID: [2594790](https://pubmed.ncbi.nlm.nih.gov/2594790/); PubMed Central PMCID: PMC298551.

29. Gjini E, Haydon DT, Barry JD, Cobbold CA. Critical interplay between parasite differentiation, host immunity, and antigenic variation in trypanosome infections. *Am Nat.* 2010; 176(4):424–39. Epub 2010/08/19. <https://doi.org/10.1086/656276> PMID: 20715972.
30. Miller EN, Turner MJ. Analysis of antigenic types appearing in first relapse populations of clones of *Trypanosoma brucei*. *Parasitology.* 1981; 82(1):63–80. Epub 1981/02/01. <https://doi.org/10.1017/s0031182000041871> PMID: 6163126.
31. 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. *J Mol Biol.* 1991; 221(3):823–35. Epub 1991/10/05. [https://doi.org/10.1016/0022-2836\(91\)80178-w](https://doi.org/10.1016/0022-2836(91)80178-w) PMID: 1942032.
32. Pinger J, Nestic D, Ali L, Aresta-Branco F, Lilić M, Chowdhury S, et al. African trypanosomes evade immune clearance by O-glycosylation of the VSG surface coat. *Nat Microbiol.* 2018; 3(8):932–8. Epub 2018/07/11. <https://doi.org/10.1038/s41564-018-0187-6> PMID: 29988048; PubMed Central PMCID: PMC6108419.
33. Rico E, Jeacock L, Kovarova J, Horn D. Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. *Sci Rep.* 2018; 8(1):7960. Epub 2018/05/23. <https://doi.org/10.1038/s41598-018-26303-w> PMID: 29785042; PubMed Central PMCID: PMC5962531.
34. Pinger J, Chowdhury S, Papavasiliou FN. Variant surface glycoprotein density defines an immune evasion threshold for African trypanosomes undergoing antigenic variation. *Nat Commun.* 2017; 8(1):828. Epub 2017/10/12. <https://doi.org/10.1038/s41467-017-00959-w> PMID: 29018220; PubMed Central PMCID: PMC5635023.
35. Hutchinson S, Glover L, Horn D. High-resolution analysis of multi-copy variant surface glycoprotein gene expression sites in African trypanosomes. *BMC Genomics.* 2016; 17(1):806. Epub 2016/10/21. <https://doi.org/10.1186/s12864-016-3154-8> PMID: 27756224; PubMed Central PMCID: PMC5070307.
36. Chandra M, Dakovic S, Foti K, Zeelen JP, van Straaten M, Aresta-Branco F, et al. Structural similarities between the metacyclic and bloodstream form variant surface glycoproteins of the African trypanosome. *PLoS Negl Trop Dis.* 2023; 17(2):e0011093. Epub 2023/02/14. <https://doi.org/10.1371/journal.pntd.0011093> PMID: 36780870; PubMed Central PMCID: PMC9956791.
37. Cosentino RO, Brink BG, Siegel TN. Allele-specific assembly of a eukaryotic genome corrects apparent frameshifts and reveals a lack of nonsense-mediated mRNA decay. *NAR Genom Bioinform.* 2021; 3(3):lqab082. Epub 2021/09/21. <https://doi.org/10.1093/nargab/lqab082> PMID: 34541528; PubMed Central PMCID: PMC8445201.
38. Cross GA, Kim HS, Wickstead B. Capturing the variant surface glycoprotein repertoire (the VSGnome) of *Trypanosoma brucei* Lister 427. *Mol Biochem Parasitol.* 2014; 195(1):59–73. Epub 2014/07/06. <https://doi.org/10.1016/j.molbiopara.2014.06.004> PMID: 24992042.
39. Hutchinson S, Foulon S, Crouzols A, Menafrá R, Rotureau B, Griffiths AD, et al. The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of *Trypanosoma brucei* in the tsetse fly salivary glands. *PLoS Pathog.* 2021; 17(9):e1009904. Epub 2021/09/21. <https://doi.org/10.1371/journal.ppat.1009904> PMID: 34543350; PubMed Central PMCID: PMC8509897 following competing interests: A.D.G is a shareholder in 1CellBio.
40. Kolev NG, Ramey-Butler K, Cross GA, Ullu E, Tschudi C. Developmental progression to infectivity in *Trypanosoma brucei* triggered by an RNA-binding protein. *Science.* 2012; 338(6112):1352–3. Epub 2012/12/12. <https://doi.org/10.1126/science.1229641> PMID: 23224556; PubMed Central PMCID: PMC3664091.
41. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. *Science.* 2005; 309(5733):416–22. Epub 2005/07/16. <https://doi.org/10.1126/science.1112642> PMID: 16020726.
42. Wickstead B, Ersfeld K, Gull K. The small chromosomes of *Trypanosoma brucei* involved in antigenic variation are constructed around repetitive palindromes. *Genome Res.* 2004; 14(6):1014–24. Epub 2004/06/03. <https://doi.org/10.1101/gr.2227704> PMID: 15173109; PubMed Central PMCID: PMC419779.
43. Barnes RL, McCulloch R. *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. *Nucleic Acids Res.* 2007; 35(10):3478–93. Epub 2007/05/05. <https://doi.org/10.1093/nar/gkm249> PMID: 17478508; PubMed Central PMCID: PMC1904282.
44. Glover L, Hutchinson S, Alsford S, Horn D. VEX1 controls the allelic exclusion required for antigenic variation in trypanosomes. *Proc Natl Acad Sci U S A.* 2016; 113(26):7225–30. Epub 2016/05/27. <https://doi.org/10.1073/pnas.1600344113> PMID: 27226299; PubMed Central PMCID: PMC4932947.

45. 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 Res.* 2007; 17(9):1344–52. Epub 2007/07/27. <https://doi.org/10.1101/gr.6421207> PMID: [17652423](https://pubmed.ncbi.nlm.nih.gov/17652423/); PubMed Central PMCID: PMC1950903.
46. Smith TK, Vasileva N, Gluenz E, Terry S, Portman N, Kramer S, et al. Blocking variant surface glycoprotein synthesis in *Trypanosoma brucei* triggers a general arrest in translation initiation. *PLoS One.* 2009; 4(10):e7532. Epub 2009/10/27. <https://doi.org/10.1371/journal.pone.0007532> PMID: [19855834](https://pubmed.ncbi.nlm.nih.gov/19855834/); PubMed Central PMCID: PMC2762041.
47. Aslett M, Aurrecochea C, Berriman M, Brestelli J, Brunk BP, Carrington M, et al. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 2010; 38(Database issue): D457–62. Epub 2009/10/22. <https://doi.org/10.1093/nar/gkp851> PMID: [19843604](https://pubmed.ncbi.nlm.nih.gov/19843604/); PubMed Central PMCID: PMC2808979.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9(4):357–9. Epub 2012/03/06. <https://doi.org/10.1038/nmeth.1923> PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/); PubMed Central PMCID: PMC3322381.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009; 25(16):2078–9. Epub 2009/06/10. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/); PubMed Central PMCID: PMC2723002.
50. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014; 30(7):923–30. Epub 2013/11/15. <https://doi.org/10.1093/bioinformatics/btt656> PMID: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/).
51. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010; 26(1):139–40. Epub 2009/11/17. <https://doi.org/10.1093/bioinformatics/btp616> PMID: [19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/); PubMed Central PMCID: PMC2796818.
52. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 2014; 42(Web Server issue):W187–91. Epub 2014/05/07. <https://doi.org/10.1093/nar/gku365> PMID: [24799436](https://pubmed.ncbi.nlm.nih.gov/24799436/); PubMed Central PMCID: PMC4086134.
53. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 2016; 33(6):1635–8. Epub 2016/02/28. <https://doi.org/10.1093/molbev/msw046> PMID: [26921390](https://pubmed.ncbi.nlm.nih.gov/26921390/); PubMed Central PMCID: PMC4868116.