Transcriptomic Profiling of *Trypanosoma Congolense* Mouthpart Parasites from Naturally Infected Flies

Sara Silva Pereira (✉ ssilvapereira@medicina.ulisboa.pt)

Instituto de Medicina Molecular <https://orcid.org/0000-0002-6590-6626>

Kawira Mathenge

ICIPE: International Centre for Insect Physiology and Ecology

Daniel Masiga

ICIPE: International Centre for Insect Physiology and Ecology

Andrew Jackson

University of Liverpool, Institute of Infection, Veterinary and Ecological Sciences

---

Research Article

Keywords:

Posted Date: January 31st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1292471/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Animal African Trypanosomiasis, or nagana, is a veterinary disease caused by African trypanosomes transmitted by tsetse flies. In Sub-Saharan Africa, *Trypanosoma congolense* is the major cause of nagana in livestock, resulting in high animal morbidity and mortality and extensive production losses. In the tsetse fly, parasites colonise the midgut and eventually reach the mouthparts, from where they can be transmitted as the fly feeds on vertebrate hosts such as cattle. Despite the extreme importance of mouthpart form parasites for disease transmission, very few global expression profile studies have been done in these parasite forms. Here, we collected tsetse flies from the Shimba Hills National Reserve, a wildlife area in southeast Kenya, diagnosed *T. congolense* infections, sequenced the transcriptomes of the *T. congolense* parasites colonising the mouthparts of the flies, and compared their expression profiles with those of parasites collected from experimental fly infections. We found little correlation between mouthpart parasites from natural and experimental fly infections. Furthermore, we performed differential gene expression analysis between mouthpart and bloodstream form parasites and identified several surface-expressed genes and 152 novel hypothetical proteins differentially expressed in mouthpart parasites. Finally, we profiled variant antigen expression and discovered a VSG gene constitutively expressed in metacyclic and bloodstream form parasites, that may be subject to novel developmental regulation. Our study is the first to provide trypanosome transcriptomes of parasites from naturally infected tsetse flies and highlights the need to confirm findings from experimental models of infection in epidemiologically-relevant settings.

Introduction

Animal African Trypanosomiasis (AAT), or nagana, is a livestock disease caused by the blood parasites African trypanosomes, particularly *Trypanosoma congolense*, *T. brucei*, and *T. vivax*, and transmitted by tsetse flies. In Sub-Saharan Africa, *T. congolense*, particularly of the savannah subtype, is the most pathogenic trypanosome species [1]. The life cycle of *T. congolense* alternates between the bloodstream of a mammalian host (usually livestock) and the tsetse fly vector. In the tsetse fly, the parasites first colonise the midgut as procyclic forms, but quickly cross the proventriculus towards the mouthparts of the fly. Here, they differentiate into epimastigotes and then into metacyclic forms, which are infective to mammals. Infection results in substantial animal mortality and economic loss. Animal productivity can be reduced by 38%, and milk and animal off-take by 12 and 10%, respectively [2]. Symptoms including anaemia and abortion lead to further production losses. With drug resistance widespread and vector control impractical, there is an urgent need for an AAT vaccine. However, a vaccine is considered impossible because of antigenic variation, a mechanism through which trypanosomes change their Variant Surface Glycoproteins (VSG) to prevent a long-lasting, effective immune response. Previous vaccine attempts were directed at bloodstream stage of the parasite. Yet, metacyclic parasites elicit the first immune response in the host. These parasites also express VSG on their surface, but they are not antigenically variant. A small set of VSGs has been found to be expressed by *T. congolense* metacyclic

forms in three independent studies [3–5], which offers some possibility that this form of *T. congolense* may be amenable to vaccination.

Recently, we have developed variant antigen profiling, a method to describe and quantify the VSG repertoire for any isolate from sequencing reads [6]. Thus, we can examine metacyclic VSG expression to identify whether our VSGs of interest are constitutively expressed in natural tsetse fly populations of distinct strains and genetic backgrounds. We collected wild tsetse flies from the Shimba Hills National Reserve (Kenya), diagnosed *T. congolense* mouthpart infections and sequenced their transcriptomes. Here, we show that gene expression profiles of trypanosomes from natural and experimental fly infections are considerably different. Additionally, we show that the pattern of metacyclic VSG expression in natural infections is non-random. Further dissecting the dynamics of metacyclic VSG expression in natural trypanosome infections can help to identify constitutively expressed or metacyclic-specific VSG genes that may offer an effective vaccine target.

Methods

Sample Collection, Preparation and Sequencing

A total of 3237 flies were collected alive from the Shimba Hills National Reserve and its periphery using Vavoua traps [7] over the course of 10 days. Flies were killed by decapitation and midguts dissected to identify positive trypanosome infections. Mouthparts of flies with positive midguts were further dissected and diagnosed according to the description of Peel (1962) [8]. When metacyclic parasites were visible, the hypopharynx and the parasite suspension that was released during dissection were collected and immediately frozen in liquid nitrogen. Sixty positive hypopharynxes were pooled into six groups of ten flies. Total RNA was extracted from the hypopharynxes using AllPrep RNA/protein kit (Qiagen, UK) according to the manufacturer's protocol, yielding RNA outputs of 108–374 ng per sample. Total RNA was used to prepare cDNA libraries for Multiplexed Spliced-Leader Sequencing using the *T. congolense* splice leader sequence [9] as complementary strand amplification primer. The protocol described by Cuypers et al. (2017) [10] was followed up to library indexing. After amplification of adaptor-ligated cDNA, we used the NEBNext® Ultra™ II FS DNA Library Prep Kit (NEB) for fragmentation, indexing, and amplification, as previously described [11]. Samples were sequenced on the Illumina HiSeq 4000 platform as 2x150bp reads, yielding 1.5 to 2.6x10⁷ mappable reads per sample.

Read Processing and Species Identification

Paired reads were mapped to the tsetse fly [12], *T. brucei* Lister 427 [13], *T. congolense* IL3000 [14], and *T. vivax* Y486 [15] genomes using Hisat2 [16] under default settings. One sample yielded very few trypanosome-mappable reads and thus was not used in downstream analyses.

Differential Expression Analysis

Reads mapping to *T. congolense* IL3000 were used in subsequent differential expression analyses. Alignment files were parsed through SAMtools (v1.9) [17] and transcript assembly and read count

estimation were done using Stringtie [18], following the developers' protocol [19]. Transcripts with less than 10 read counts were removed using EdgeR [20] and differential expression was estimated with Limma-voom [21] and DESeq2 [22]. Transcripts were considered differentially expressed between parasites probosces and in the bloodstream if they had log2 fold change (logFC) > 1.5 and adjusted p -value < 0.01 in both limma-voom and DESeq2.

Functional Annotation of Differentially Expressed Genes

Gene ontology (GO) categories related to biological processes, molecular function and cell component were identified using the annotated GO term list for IL3000 genome available in TritrypDB and TopGo [23]. Enriched GO terms among DE gene products were established by Fisher's exact test at a false discovery rate (FDR) p -value ≤ 0.05 .

Variant Antigen Profiling

For variant antigen profiling estimation, reads mapping to the tsetse genome were discarded. Remaining reads were assembled *de-novo* using Trinity [24]. Transcript abundances were estimated in a reference-free manner using kallisto [25] as per the pre-compiled scrip available in the Trinity package. Assembled transcriptomes were screened for VSG transcripts using VAPPER [6], an automated tool for the analysis of VSG repertoires from genomic or transcriptomic sequencing data. It relies on a set of conserved protein motifs that allow the identification and quantification of each of the 15 phylotypes (i.e. phylogenetically-related lineages) that comprise the *T. congolense* VSG repertoire [5].

Results

Tsetse flies circulating in the Shimba Hills National Reserve are mostly infected with *T. congolense*

Over the course of 10 days, we collected 3237 live tsetse flies, of which 108 (3.33%) had positive *T. congolense* midgut infections. Of those, 62 (57.41%) had detectable metacyclic parasites in their mouthparts. We detected neither *T. brucei* parasites in the midgut nor *T. vivax* parasites in the proboscis. However, given the challenge in distinguishing African trypanosome species by microscopy, the possibility of mixed infections cannot be completely excluded.

We made six pools of 10 probosces, extracted RNA and sequenced them on the Illumina MiSeq platform. One of the samples was discarded post-sequencing due to low quality. For each sample, we obtained 7 to 13 million read pairs (Mean $\pm$ SEM, 1.02E7 $\pm$ 9.73E05) (Table 1), of which 22.46% $\pm$ 3.99 mapped to the IL3000 genome and only 23.48% $\pm$ 2.70 mapped to the tsetse genome. As the trypanosome-enrichment method performed at the cDNA library preparation stage (SL-sequencing) cannot accurately select between African trypanosome species, it was necessary to check for contamination from *T. vivax* and *T. brucei*. We found a negligible number of reads mapping to these species (1.67% $\pm$ 0.23 and 0.46% $\pm$ 0.13, respectively), which suggests that the proboscis of flies collected in the Shimba Hills National Reserve are mostly, if not only, colonised by *T. congolense*.

Table 1 Sequencing and mapping statistics for transcriptomes produced from mouthpart parasites collected at the Shimba Hills National Reserve (Kenya). Read mapping was performed with Hisat2 [16].

Sample	Number of Reads (#)	# Tsetse	% Tsetse	# IL3000	% IL3000	# TvY486	% TvY486	# Tb427	% Tb427
A	1.29E+07	2.22E+06	17.23%	3.22E+06	24.95%	2.71E+05	2.10%	5.23E+04	0.41%
B	9.72E+06	1.98E+06	20.34%	2.34E+06	24.07%	1.23E+05	1.27%	2.40E+04	0.25%
C	1.29E+07	2.36E+06	18.32%	4.78E+06	37.07%	1.89E+05	1.46%	1.32E+05	1.02%
D	1.07E+07	3.27E+06	30.43%	1.33E+06	12.38%	2.59E+05	2.41%	4.93E+04	0.46%
E	7.08E+06	2.20E+06	31.09%	9.79E+05	13.82%	7.76E+04	1.10%	1.22E+04	0.17%
Min	7.08E+06	1.98E+06	17.23%	9.79E+05	12.38%	7.76E+04	1.10%	1.22E+04	0.17%
Max	1.29E+07	3.27E+06	31.09%	4.78E+06	37.07%	2.71E+05	2.41%	1.32E+05	1.02%
Mean	1.07E+07	2.41E+06	23.48%	2.53E+06	22.46%	1.84E+05	1.67%	5.39E+04	0.46%
SEM	9.73E+05	2.01E+05	2.70%	6.14E+05	3.99%	3.36E+04	0.23%	1.87E+04	0.13%

Expression profiles of mouthpart parasites from naturally-infected tsetse flies are poorly reproduced in experimental tsetse infections

We compared the transcriptomes of parasites from each pool of naturally infected flies and we observed that they are reproducible and correlate well (average $R^2=0.94\pm0.04$, Pearson Correlation). This shows that mouthpart parasites from naturally infected tsetse flies collected in the Shimba Hills National Reserve have similar transcriptomes. We then asked how the transcriptomes from parasites collected from wild flies compared to mouthpart parasites from experimental infections and found that such reproducibility is lost. Specifically, our data correlates weakly to the transcriptomes previously published by us [5] and by Awuoche et. al (2018) [26] ($R^2 = 0.11$ and $R^2 = 0.31$) (Figure 1A-C). Nonetheless, variation is large even between experimental infections, as the correlation between these two previously published transcriptomes is even lower ($R^2 = 0.10$), despite similar experimental approaches. Correlation analysis of the most-abundant 1000 transcripts from the average read counts of natural infections transcriptomes does not improve correlation with those obtained from experimental infections ($R^2 = 0.09$). This suggests that transcriptomes from experimental infections may not be a good proxy of wild trypanosomes.

To further understand the differences between transcriptomes, we searched for differentially expressed genes between parasites collected from experimental and natural infections. We did not detect any differentially expressed genes between natural infections transcriptomes and those from Awuoche *et al.* (2018) [27]. This was surprising given the low correlation between the transcriptomes, and is likely a reflection of the low number of replicates in the latter dataset, which reduces the statistical power of the analysis. In contrast, we identified 757 differentially-expressed genes (9.5%, corresponding to 321 up-regulated and 436 down-regulated genes) between the present natural infections and our previous experimental infections [5] (Supplementary Table 1). These differences may be explained by strain

diversity, infection mode, tsetse rearing conditions, and number of pooled probosces, among other factors.

To better understand the variation in gene expression profiles between natural and experimental datasets, we compared the experimental and natural infection transcriptomes to previously published bloodstream-form transcriptomes, using limma-voom [21]. We detected 316 differentially-expressed genes in the mouthpart transcriptome produced by Awuoché *et al.* (2018) [27], 2672 in the sample previously published by us [5], and 1208 genes in the transcriptomes from the natural infections presented in the current study. However, only 110 of the DE genes are shared between all three datasets; 431 are shared between our experimental and natural infection datasets; 39 between Awuoché's and our natural infection transcriptomes; and 104 are common to both experimental datasets.

Together, the strong correlation between the five samples collected from wild flies, the weak correlation between mouthpart transcriptomes of different sources and the little overlap between the genes DE in the mouthparts compared to the bloodstream suggest that there is wide variation in the expressed gene profile of mouthpart transcriptomes. However, the weak correlation between both experimental datasets, indeed in the same magnitude of the correlation between natural and experimental datasets, suggest that such variation is not due to the wildness of the flies.

Differential gene expression shows abundance of hypothetical proteins upregulated in mouthpart parasites

Having established that trypanosomes from natural fly infections have a characteristic, but reproducible, expression profile, we proceeded to investigate how they differ from cultured *T. congolense* IL3000 bloodstream forms [28]. As previously described, these bloodstream-form transcriptomes were produced from parasites collected from the blood of infected laboratory mice at the first peak of parasitaemia and the ascending parasitaemia phase preceding it [28]. The library sizes (a proxy for sequencing depth) of mouthpart transcriptomes ranged between 2.48×10^6 and 1.16×10^7 per sample (average of $6.61 \times 10^6 \pm 2.48 \times 10^6$), whereas for bloodstream form transcriptomes libraries ranged from 2.99×10^7 to 3.57×10^7 (average of $3.35 \times 10^7 \pm 1.90 \times 10^6$). Of the 10315 transcripts annotated in the IL3000 genome, we detected 8835 transcripts with more than 10 read counts. Only these were considered in downstream analyses, as a measure to reduce the noise associated with low read counts (Figure 2A). We have also normalised read counts for the differences in sequencing depth (Figure 2B).

We observed a clear separation between the expression profiles of mouthpart and bloodstream form parasites. The transcriptomes of mouthpart parasites are more variable than the bloodstream-forms, reflected by longer distances in the multi-dimensional scaling plot (Figure 2C). This was expected, as they potentially represent not only a mixture of parasite strains circulating in the Shimba Hills National Reserve, but also a mixture of parasites in different life stages (i.e. epimastigotes, pre-metacyclics, metacyclics).

We have subjected our samples to differential expression analysis using Limma-voom [21] (Figure 2D) and DESeq2 [22] ($\log_2FC > |1.5|$ and $p\text{-value} < 0.01$). Limma-voom detected a total of 1208 differentially-expressed genes (648 down-regulated and 560 up-regulated), whereas DESeq2 detected 1276 (Supplementary Table 2). Of these, 1062 were identified by both methods, corresponding to 88% and 83% of the total, respectively (Figure 3A).

Our results suggest a strong contribution of epimastigote parasites to the mouthpart parasite population. The list of genes up-regulated in mouthpart parasites is rich in hypothetical proteins, GTPase-related genes, zinc-finger proteins, and histones (Figure 3B, red), which is consistent with previous observations in experimental tsetse fly infections [26]. Down-regulated genes relate to VSG, several other surface genes, including ESAG2-like (most likely transferrin receptors), adenylate cyclases, and ISG, RHS genes, and trans-sialidases (Figure 3B, blue). This is also reflected within the top 25 most down-regulated genes (Figure 3C). Indeed, within the pool of DE genes, we found 111 genes to belong to 32 of the 74 previously-described non-VSG cell-surface gene families [29]. Fam50, encoding brucei alanine-rich proteins (BARP), is the most represented cell-surface gene family amongst up-regulated genes (N=13), whereas Fam51, encoding adenylate cyclases, is the most common amongst down-regulated genes (N=20) (Figure 3D). Six gene families had members within both the up- and down-regulated gene lists, which may indicate developmental regulation within these gene families. These were Fam46 (major surface protease – gp63), Fam49 (invariant surface glycoproteins), Fam51 (adenylate cyclases), Fam58 (MFS transporters), Fam63 (lipases), and Fam77 (hypothetical proteins conserved in *T. brucei*, *T. congolense*, and *T. vivax*).

We subjected the list of up and down-regulated genes to GO term enrichment analysis using TopGo and the GO term annotation of the IL3000 genome available at TritrypDB. We observed that the biological processes significantly enriched amongst the up-regulated gene list are quite varied, including terms relating to mitosis, cell-wall organisation, glutathione catabolic process, or protein retention (Figure 3E). In contrast, the down-regulated processes relate to biosynthetic, metabolic and signalling processes. In terms of molecular functions, we observed down-regulation of ribosome processes and mostly upregulation of DNA binding, protein heterodimerization activity and enzymatic functions (Figure 3E). Enrichment analysis of cellular component also reflects lower protein synthesis in mouthpart forms, which is consistent with the quiescent state of metacyclic parasites (Figure 3E). In contrast, there seems to be an increase in nucleus and nucleosome activity in mouthpart parasites, which was also observed by Awuoche *et al.* [26] in their comparison to cardia parasites. The GO term analysis confirms that mouthpart forms are less physiologically active than bloodstream form parasites.

Interestingly, we found that 143 out of the 511 up-regulated genes and 144 of the 491 down-regulated genes are annotated as hypothetical proteins. Whilst 135 of these hypothetical proteins have already been detected by Awuoche *et al.* [26] as differentially expressed in mouthpart parasites compared to

cardia parasites, we found 83 additional hypothetical proteins that are down-regulated and 69 that are upregulated in mouthpart parasites compared to bloodstream forms (Supplementary Table 3).

Variant Antigen Profiles of *T. congolense* metacyclic parasites show a reproducible expression pattern

An infected proboscis contains both epimastigotes and metacyclic parasites. Whilst distinguishing between epimastigote and metacyclic transcriptomes is not possible without parasite purification or enrichment, we can analyse metacyclic VSG expression because VSGs are not expressed by epimastigotes. Therefore, we estimated the variant antigen profiles of each sample with VAPPER [6] (Figure 4A). VAPPER allows the comparison of VSG diversity by searching for diagnostic protein motifs that allow grouping of individual VSG genes into 15 conserved phylotypes. By comparing the representation of individual phylotypes, we can detect changes in VSG expression patterns [5]. Although we have successfully produced trypanosome transcriptomes from the mouthparts of heavily-infected individual flies before [5], in the present study we made pools of 10 probosces to ensure that we obtained enough material for reliable transcriptomes. As such, we committed to profiling the VSG repertoire across a cohort, rather than a clonal population. We observed that all five samples have an overall similar pattern of VSG expression, which is consistent with previous observations from experimental infections [5]. However, we detected subtle fluctuations in VSG phylotype (P) expression patterns (Figure 4A). Specifically, we observed the highest variation in the abundance of P1, P8, P4, P15, and P10, whereas P14, P6, P5, P2, and P12 were the least variable between samples. We also estimated VSG diversity in the bloodstream-form parasites. These displayed a very high degree of reproducibility between replicates, reflecting their origin from a single infection with the same blood stabilate [28].

When we compared the expressed VSG profiles of metacyclic parasites to the average genomic VSG profile of *T. congolense* (estimated from 97 genomes, available from the VAPPER database [6]), quite considerable variation was revealed. We observed that all phylotypes, with the exception of P1, P3, P7, and P13, were statistically different (independent t-test, p -value < 0.01) (Figure 4B). P2, P5, P6, P12, P14, and P15 were less abundant in the transcriptomes than in their relative frequency in the genomic repertoire (independent t-test, p -value < 0.001), whilst P4, P8, P9, P10, P11 were more abundant (independent t-test, p -value < 0.01). The differences in abundance of phylotypes P4, P7, P8, P9, P11, P12, and P15 are consistent with previous observations from experimental infections with two different strains (1/148 and TC13) [5, 6].

The contribution of phylotype 8 to the expressed VSG profiles of metacyclic parasites is noteworthy, especially taking into consideration its small influence in the genome collection, and in the repertoires of bloodstream form parasites. Upon more detailed analysis, we discovered that its abundance is mostly due to the high expression of gene TcIL3000.A.H_000381200 (formerly TcIL3000_0_09520). This VSG-like gene is present in all samples from both mouthpart and bloodstream form parasites and, despite being slightly more abundant in metacyclic than in bloodstream form parasites, it is not differentially expressed ($\log_2\text{FC} = 2.32$, p -value = 0.12).

Discussion

T. congolense is the most prevalent and pathogenic African trypanosome species in East Africa [30]. Its development in the tsetse fly is essential for transmission and successful establishment in the mammalian host. In this work, we produced the first transcriptomes of mouthpart *T. congolense* parasites collected from naturally-infected, wild tsetse flies, compared them to available transcriptomes of mouthpart and bloodstream form parasites, and described their variant antigen repertoires. We showed that transcriptomes from experimental fly infections poorly reproduce the expression profiles observed in the field. We identified a set of novel hypothetical proteins upregulated in mouthpart parasites, as well as variant antigens constitutively expressed in both metacyclic and bloodstream form parasites.

Mouthpart trypanosomes from natural and experimental fly infections have distinct gene expression profiles

□

Our results show that the expression profiles of mouthpart trypanosomes from natural and experimental fly infections correlate weakly. This is consistent with intrinsic variation between parasite populations, both in experimental and natural conditions. The disparities in gene expression can be attributed to the biology of either the parasite or the vector. From a parasite point of view, intrinsic variability within trypanosome strains could explain the weak correlation.

Experimental transcriptomes were produced from parasites strains TC13 [26] and 1/148 [5], whilst in the field we would expect a multitude of strains. However, it is unlikely that strain differences explain all the disparities observed because, at least in the bloodstream form stage, strains with high genetic distance still share more than 80% of transcriptome identity (e.g. strains IL3000 and 1/148 are highly divergent, but share 84% of their bloodstream form transcriptome (Sara Silva Pereira & Luísa Figueiredo, unpublished results)).

It appears that vector biology may play a large role in parasite expression profiles, be it in terms of the tsetse genetic background, or its microbiome. In the Shimba Hills National Reserve, tsetse flies of the species *Glossina pallidipes* predominate [31, 32], while the flies examined in previous studies were *Glossina morsitans morsitans* [5, 26]. Moreover, laboratory-infected flies are always infected as teneral (i.e. newly emerged, previously unfed), whereas in the Shimba Hills National Reserve, a previous study estimated that only approximately 37% of the infected tsetse flies are young adults [31]. It is well described that teneral flies are more susceptible to trypanosome infections [33]. It has also been reported that the fly microbiome and its dynamics during the tsetse fly's development affects susceptibility to trypanosome infections because it modulates the tsetse fly's immune response [34]. Furthermore, it has been shown experimentally that the absence of gut microbiome compromises the integrity of the

peritrophic matrix, a chitinous, sheath-like structure that separates the midgut epithelia from the lumen, thus increasing trypanosome susceptibility [34]. Finally, we cannot ignore the role of the additional barriers faced by trypanosomes in the wild may have on parasite gene expression. For instance, one pertinent aspect that remains unresolved is the impact of present and past co-infections in the gut microbiome, immune response, and the overall consequence to subsequent trypanosome infection establishment. Together, our results highlight the importance of translating gene expression studies to epidemiologically relevant settings.

Our results also highlighted the importance of surface-expressed genes in the adaptation of *T. congolense* to the different host or vector environments. In particular, we observed considerable developmental regulation amongst cell-surface phylome families (including Fam 46 Fam46 (major surface protease – gp63), Fam49 (invariant surface glycoproteins), Fam51 (adenylate cyclases), Fam58 (MFS transporters), Fam63 (lipases), and Fam77 (hypothetical proteins conserved in *T. brucei*, *T. congolense*, and *T. vivax*). Surface-expressed genes are responsible for many of the species-specific adaptations of trypanosomes. This was perceived in the initial comparative genomics analyses [15, 35], as gene families encoding cell surface proteins are often non-homologous and lineage-specific [29]. Furthermore, the fact that we have identified 152 novel hypothetical proteins differentially expressed in mouthpart parasites, shows that there is still a lot of hidden gene functionality waiting to be described in trypanosomes.

VSG expression is dynamic, but reproducible

We observed that the variant antigen profiles of both metacyclic and bloodstream-form parasites were reproducible across biological replicates. This reproducibility is remarkable, given the potential diverse origin of the mouthpart parasites, but agrees with preceding findings from experimental infections [5]. Together, our data provides further evidence that metacyclic *VSG* expression is dynamic, but reproducible and non-random. The most parsimonious explanation for this observation is that, at least at the infrapopulation level, *T. congolense* VSGs are not arbitrarily selected from the genome repertoire for expression in the metacyclic life stage. This concept of *VSG* expression obeying a pre-determined, probabilistic order agrees with current knowledge of *T. brucei* [36, 37] and *T. vivax* [11] antigenic variation in the mammalian host. However, in the tsetse fly, *T. brucei* metacyclic VSGs seem to be arbitrarily selected from the genomic repertoire as their expression patterns are not reproducible over time in either natural or experimental conditions [38].

We have also identified one *VSG* gene, belonging to P8, that is expressed in both metacyclic and bloodstream-form parasites (TcIL3000.A.H_000381200, formerly TcIL3000_0_09520). The same *VSG* transcript was detected in experimental metacyclic form transcriptomes in two independent experiments [5, 26]. In 2018, when we first discovered P8 to be highly represented in the variant antigen profiles of metacyclic parasites from experimental fly infections [5], we postulated that this gene(s) could be potentially exploitable for vaccine design, if reproducible in nature. Now, we confirm that TcIL3000.A.H_000381200 is enriched in natural fly infections and that is also expressed by bloodstream form parasites. We propose that this gene might be a constitutive *VSG*, and therefore likely to have a

distinct function required by cells. Examples of VSGs acquiring novel functions beyond antigenic variation are plenty within trypanosome genomes [39]. They include ESAG2 and the transferrin receptors, both surface proteins expressed in the cell body and the flagellar pocket, respectively, during the bloodstream form stage [40]. To consider TcIL3000.A.H_000381200 a promising vaccine candidate, it must encode an immunogenic surface protein. Despite the common belief that effective vaccines against African trypanosome infections are virtually impossible to design due to the very successful immune evasion strategies employed by the parasites [41], researchers have recently discovered an invariant surface antigen that induces long-lasting, protective immunity to *T. vivax* infections in mice [42]. In another study, single-cell RNA sequencing in *T. brucei* mouthpart forms revealed a metacyclic-specific surface-expressed gene that showed potential as a transmission blocking antigen (i.e. SGM1.7, belonging to Fam10) [43]. Therefore, as we continue to dissect the variability of surface-expressed genes, the possibility of a vaccine to trypanosomiasis becomes more plausible.

In conclusion, this study provides insight into the gene expression profile of *T. congolense* mouthpart forms in an epidemiological and biologically-relevant context. Our results expose the need to dissect the functionality of novel hypothetical proteins, numerous genes encoding putative cell-surface proteins, as well as constitutively-expressed *VSG* genes, as they may be promising candidates for drug treatment and vaccinology strategies against trypanosomiasis.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in NCBI/SRA, under project accession number PRJNA795045.

Competing interests

The authors declare that they have no competing interests.

Author's Contributions

Conceived and designed the experiments: SSP, DM, APJ. Performed the experiments: SSP, KM. Analysed the data: SSP, KM, APJ. Contributed reagents/materials/analysis tools: DM, APJ. Wrote the paper: SSP, APJ. Obtained funding: SSP, APJ. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BB/M022811/1 and BB/R021139/1) to APJ. SSP is funded by Marie Skłodowska-Curie Individual Standard European Fellowship, under grant agreement no. 839960.

Acknowledgements

We would like to thank Ms. Dianah Katiti for her assistance in tsetse fly screening, as well as Mr James Kabii, Mr Richard Tumba and Mr David Mbuvi for their assistance in tsetse fly trapping.

References

1. Bengaly Z, Sidibe I, Ganaba R, Desquesnes M, Boly H, Sawadogo L. Comparative pathogenicity of three genetically distinct types of *Trypanosoma congolense* in cattle: clinical observations and haematological changes. *Vet Parasitol.* 2002;108:1–19. doi:10.1016/S0304-4017(02)00164-4.
2. Swallow B. Impacts of trypanosomiasis on African agriculture. Int Livest Res Institute, Nairobi, Kenya. 1999; 1–46. Available.
3. Eshita Y, Urakawa T, Hirumi H, Fish WR, Majiwa PAO. Metacyclic form-specific variable surface glycoprotein-encoding genes of *Trypanosoma (Nannomonas) congolense*. *Gene.* 1992;113:139–48. doi:10.1016/0378-1119(92)90389-7.
4. Helm JR, Hertz-Fowler C, Aslett M, Berriman M, Sanders M, Quail M, et al. Analysis of expressed sequence tags from the four main developmental stages of *Trypanosoma congolense*. *Mol Biochem Parasitol.* 2009;168:34–42. doi:10.1016/j.molbiopara.2009.06.004.
5. Silva Pereira S, Casas-Sanchez A, Haines LR, Absolomon K, Ogugo M, Sanders M, et al. Variant antigen repertoires in *Trypanosoma congolense* populations and experimental infections can be profiled from deep sequence data with a set of universal protein motifs. *Genome Res.* 2018;28:1383–94. doi:10.1101/gr.234146.118.
6. Silva Pereira S, Heap J, Jones AR, Jackson AP. VAPPER. High-throughput variant antigen profiling in African trypanosomes of livestock. *Gigascience.* 2019;8:1–8. doi:10.1093/gigascience/giz091.
7. Laveissière C, Grébaut P. An economical trap for tsetse flies: the “Vavoua” trap. *Trop Med Parasitol.* 1990;41:185–92.
8. Peel E. Identification of metacyclic trypanosomes in the hypopharynx of tsetse flies, infected in nature or in the laboratory. *Trans R Soc Trop Med Hyg.* 1962;56:339–41.
9. González-Andrade P, Camara M, Ilboudo H, Bucheton B, Jamonneau V, Deborggraeve S. Diagnosis of trypanosomatid infections: Targeting the spliced leader RNA. *J Mol Diagnostics.* 2014;16:400–4. doi:10.1016/j.jmoldx.2014.02.006.
10. Cuypers B, Domagalska MA, Meysman P, Muylder G, De, Vanaerschot M, Imamura H, et al. Multiplexed Spliced-Leader Sequencing: A high-throughput, selective method for RNA-seq in Trypanosomatids. *Sci Rep.* 2017;7:3725. doi:10.1038/s41598-017-03987-0.

11. Silva Pereira S, Neto KJG, de AC, Duffy CW, Richards P, Noyes H, Ogugo M, et al. Variant antigen diversity in *Trypanosoma vivax* is not driven by recombination. *Nat Commun.* 2020;11:844. doi:10.1038/s41467-020-14575-8.
12. International Glossina Genome Initiative. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. *Science.* 2014;344:380–6. doi:10.1126/science.1249656.
13. Müller LSM, Cosentino RO, Förstner KU, Guizetti J, Wedel C, Kaplan N, et al. Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature.* 2018. doi:10.1038/s41586-018-0619-8.
14. Abbas AH, Silva Pereira S, D'Archivio S, Wickstead B, Morrison LJ, Hall N, et al. The structure of a conserved telomeric region associated with variant antigen loci in the blood parasite *Trypanosoma congolense*. *Genome Biol Evol.* 2018;10:2458–73. doi:10.1093/gbe/evy186.
15. Jackson AP, Berry A, Aslett M, Allison HC, Burton P, Vavrova-Anderson J, et al. Antigenic diversity is generated by distinct evolutionary mechanisms in African trypanosome species. *Proc Natl Acad Sci U S A.* 2012;109:3416–21. doi:10.1073/pnas.1117313109.
16. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. *Nat Methods.* 2015. doi:10.1038/nmeth.3317.
17. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.
18. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015. doi:10.1038/nbt.3122.
19. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016. doi:10.1038/nprot.2016.095.
20. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010. doi:10.1093/bioinformatics/btp616.
21. Law CW, Chen Y, Shi W, Smyth GK. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 2014. doi:10.1186/gb-2014-15-2-r29.
22. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010. doi:10.1186/gb-2010-11-10-r106.
23. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for Gene Ontology. *R Packag. version 2.26.0. R Packag version 2260.* 2016. doi:10.1038/ncomms8832.
24. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, et al. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc.* 2013;8:1–43. doi:10.1038/nprot.2013.084.De.
25. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34:525–7. doi:10.1038/nbt.3519.
26. Awuoche EO, Weiss BL, Mireji PO, Vigneron A, Nyambega B, Murilla G, et al. Expression profiling of *Trypanosoma congolense* genes during development in the tsetse fly vector *Glossina morsitans*

- morsitans. *Parasit Vectors*. 2018;11:1–18. doi:doi.org/10.1186/s13071-018-2964-8.
27. Awuoche EO, Weiss BL, Vigneron A, Mireji PO, Aksoy E, Nyambega B, et al. Molecular characterization of tsetse's proboscis and its response to *Trypanosoma congolense* infection. *PLoS Negl Trop Dis*. 2017;11:e0006057.
28. Silvester E, Ivens A, Matthews KR. A gene expression comparison of *Trypanosoma brucei* and *Trypanosoma congolense* in the bloodstream of the mammalian host reveals species-specific adaptations to density-dependent development. *PLoS Negl Trop Dis*. 2018;12:e0006863. doi:10.1371/journal.pntd.0006863.
29. Jackson AP, Allison HC, Barry JD, Field MC, Hertz-Fowler C, Berriman M. A Cell-surface Phylome for African Trypanosomes. *PLoS Negl Trop Dis*. 2013;7:e2121. doi:10.1371/journal.pntd.0002121.
30. Giordani F, Morrison LJ, Rowan TG, De Koning HP, Barrett MP. The animal trypanosomiasis and their chemotherapy: A review. *Parasitology*. 2016;143:1862–89. doi:10.1017/S0031182016001268.
31. Gachoki S, Groen T, Vrieling A, Okal M, Skidmore A, Masiga D. Satellite-based modelling of potential tsetse (*Glossina pallidipes*) breeding and foraging sites using teneral and non-teneral fly occurrence data. *Parasites Vectors*. 2021;14:1–18. doi:10.1186/s13071-021-05017-5.
32. Ebhodaghe FI, Bastos ADS, Kidambasi KO, Kalayou S, Masiga DK, Okal MN. Molecular characterization of *Trypanosoma vivax* in tsetse flies confirms the presence of the virulent Tvv4 genotype in Kenya: Potential implications for the control of trypanosomiasis in Shimba Hills. *Infect Genet Evol*. 2021;93:104953. doi:10.1016/j.meegid.2021.104953.
33. Kubi C, Van Den Abbeele J, De Deken R, Marcotty T, Dorny P, Van Den Bossche P. The effect of starvation on the susceptibility of teneral and non-teneral tsetse flies to trypanosome infection. *Med Vet Entomol*. 2006;20:388–92. doi:10.1111/j.1365-2915.2006.00644.x.
34. Weiss BL, Wang J, Maltz MA, Wu Y, Aksoy S. Trypanosome Infection Establishment in the Tsetse Fly Gut Is Influenced by Microbiome-Regulated Host Immune Barriers. *PLoS Pathog*. 2013;9. doi:10.1371/journal.ppat.1003318.
35. 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;309:416–22. doi:10.1126/science.1112642.
36. Morrison LJ, Majiwa P, Read AF, Barry JD. Probabilistic order in antigenic variation of *Trypanosoma brucei*. *Int J Parasitol*. 2005;35:961–72. doi:10.1016/j.ijpara.2005.05.004.
37. Jayaraman S, Harris C, Paxton E, Donachie A-M, Vaikkinen H, McCulloch R, et al Application of long read sequencing to determine expressed antigen diversity in *Trypanosoma brucei* infections. Acosta-Serrano A, editor. *PLoS Negl Trop Dis*. 2019;13: e0007262. doi:10.1371/journal.pntd.0007262.
38. Barry JD, Crowe JS, Vickerman K. Instability of the *Trypanosoma brucei rhodesiense* metacyclic variable antigen repertoire. *Nature*. 1983;306:699–701.
39. Silva Pereira S, Jackson AP, Figueiredo LM. Evolution of the variant surface glycoprotein family in African trypanosomes. *Trends Parasitol*. 2021. doi:10.1016/j.pt.2021.07.012.

40. Gadelha C, Zhang W, Chamberlain JW, Chait BT, Wickstead B, Field MC. Architecture of a host-parasite interface: complex targeting mechanisms revealed through proteomics. *Mol Cell proteomics*. 2015;14:1911–26. doi:10.1074/mcp.M114.047647.
41. La Greca F, Magez S. Vaccination against trypanosomiasis: can it be done or is the trypanosome truly the ultimate immune destroyer and escape artist? *Hum Vaccin*. 2011;7:1225–33. doi:10.4161/hv.7.11.18203.
42. Autheman D, Crosnier C, Clare S, Goulding DA, Brandt C, Harcourt K, et al. An invariant *Trypanosoma vivax* vaccine antigen induces protective immunity. *Nature*. 2021;595:96–100. doi:10.1038/s41586-021-03597-x.
43. Vigneron A, O’neill MB, Weiss BL, Savage AF, Campbell OC, Kamhawi S, et al. Single-cell RNA sequencing of *Trypanosoma brucei* from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens. *Proc Natl Acad Sci*. 2020;117:2613–21. doi:10.1073/pnas.1914423117/-/DCSupplemental.

Figures

Figure 1

Correlation between transcriptomes of mouthpart trypanosomes from natural and experimental fly infections. A. Correlation of transcriptomes from samples B and C (natural infections), shown as representatives of remaining samples. $R^2=0.98$, Pearson’s correlation. B. Correlation of average transcript abundance from transcriptomes obtained from natural infections and those presented by Awuoche et al. (2018). $R^2=0.31$, Pearson’s correlation. C. Correlation of average transcript abundance from transcriptomes obtained from natural infections and those presented by Silva Pereira et al. (2018). $R^2=0.11$, Pearson’s correlation.

Figure 2

Differential gene expression analysis between mouthpart (MP) and bloodstream form (BSF) *T. congolense* parasites. A. Read count density expressed as log-counts per million ($\log_2$ CPM) per sample before and after filtering transcripts with less than 10 read counts. B. Distribution of $\log_2$ CPM values per sample before and after normalisation for library size. C. Multi-dimensional scaling plot showing the distances between gene expression profiles of each experimental group. D. Volcano plot showing change in expression between conditions and transcript abundance, expressed as $\log_2$ fold change and $\log_2$ cpm.

Differentially expressed genes ($\log_2FC > |1.5|$ and $p\text{-value} < 0.01$) are highlighted in blue (downregulated) or red (upregulated in mouthpart parasites).

Figure 3

Differential gene expression analysis between mouthpart and bloodstream form parasites. A. Overlap between differentially-expressed (DE) genes detected by Limma and DESeq2. B. Diagram showing most represented words amongst the functional annotation of the differentially expressed genes. Size of the words is proportional to their frequency. C. Annotation and fold change of the 25 genes most up- and down-regulated in mouthpart parasites. D. Frequency of DE genes belonging to non-VSG cell-surface gene families. Numbering in pie slices represent number of DE genes per family. Gene families are colour-coded according to key. Red represents genes up-regulated in mouthpart parasites, whilst blue represents down-regulated genes. E. Gene Ontology (GO) Enrichment Analysis of differentially-expressed genes between mouthpart and bloodstream form parasites. Red represents up-regulated GO terms in mouthpart parasites, blue represents down-regulated. Numbers adjacent to the bars represent the number of significant genes that contributed to each GO term.

Figure 4

Variant antigen profiles of *T. congolense* are non-random and partially reproducible. A. Variant Antigen Profiles of *T. congolense* metacyclic parasites collected from tsetse flies circulating in the Shimba Hills National Forest and of bloodstream form parasites from culture [28]. B. Comparison of average phylotype proportion (adjusted for transcript abundance) in transcriptomic samples from naturally-infected, fly-derived *T. congolense* metacyclics and the average VSG genomic profile of *T. congolense* [6] ($\text{mean} \pm \text{SEM}$). Statistical analysis reveals that phylotype (P)2, P5, P6, P12, P14, and P15 are underrepresented in the transcriptomes (independent t-test, $p\text{-value} < 0.001$), while P4, P8, P9, P10, and P11 are significantly overrepresented (independent t-test, $p\text{-value} < 0.001$) in fly-derived metacyclic transcriptomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [GraphAbstract.png](#)

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)