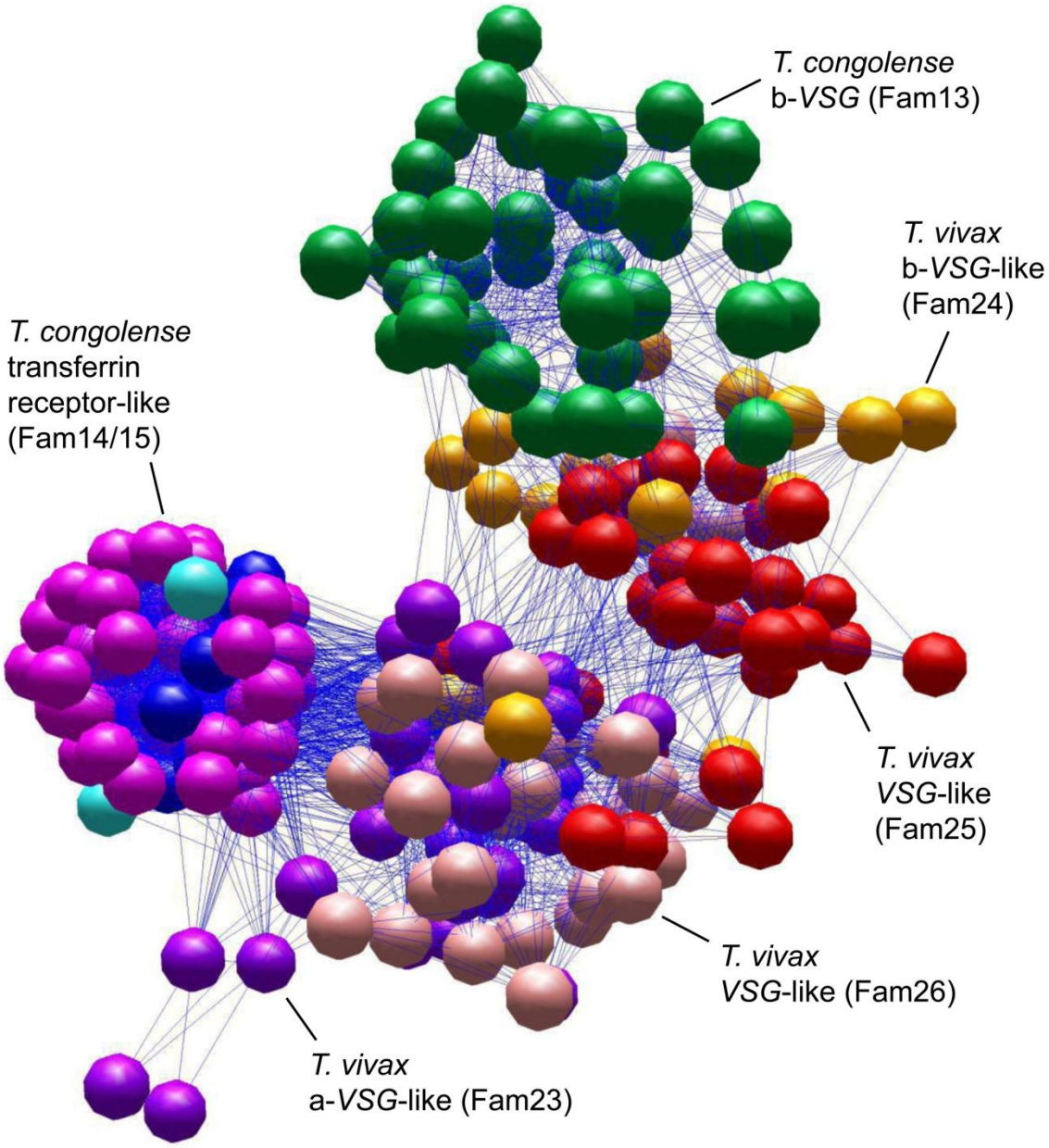
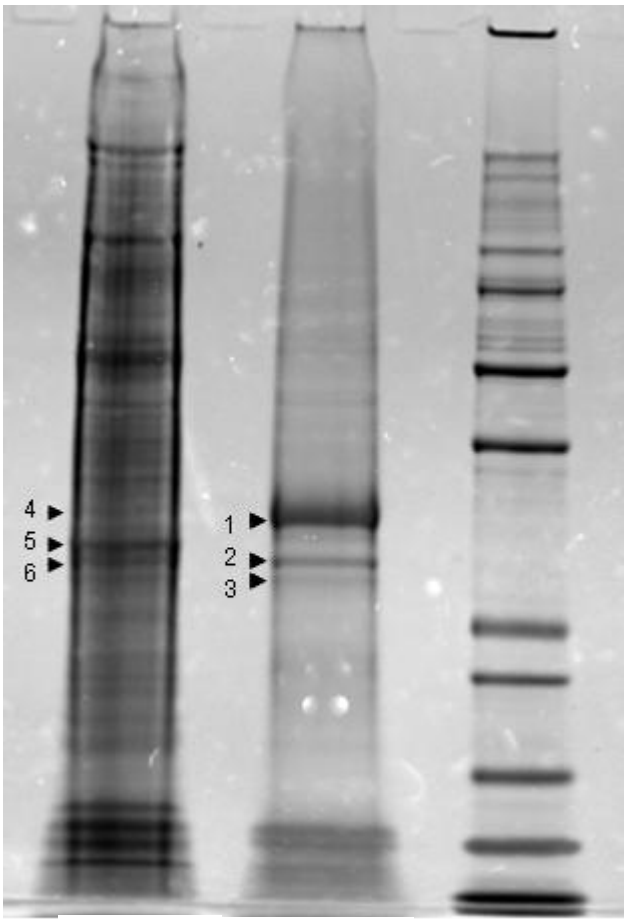


Supplementary Figure 1. A clustering network of *T. vivax* VSG-like protein sequences. The network was created with BioLayout Express v3.0 using pairwise FASTA scores for *T. vivax* VSG-like protein sequences: Fam23 (a-VSG-like; purple), Fam24 (b-VSG-like; orange), Fam25 (red) and Fam26 (pink). The figure shows that Fam23 clusters closest to a-VSG-like sequences (i.e. transferrin receptor-like proteins) from *T. congolense* (pink) and *T. brucei* (blue), and Fam24 clusters closest to b-VSG sequences from *T. congolense* (green). Forming an arc of FASTA connections between a- and b-VSG are Fam25 and Fam26. On this basis, these two *T. vivax*-specific sequence families appear intermediate.

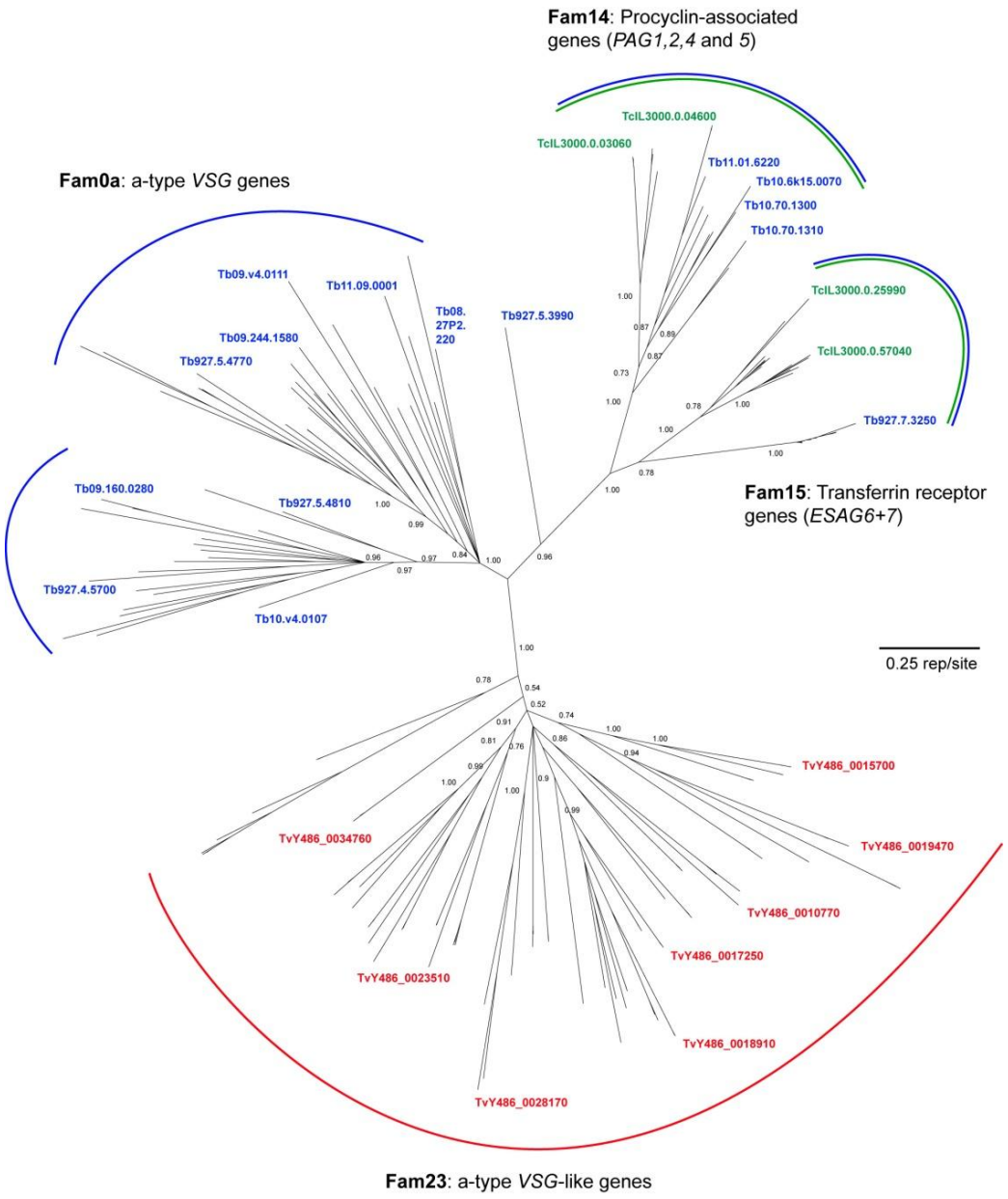
A.



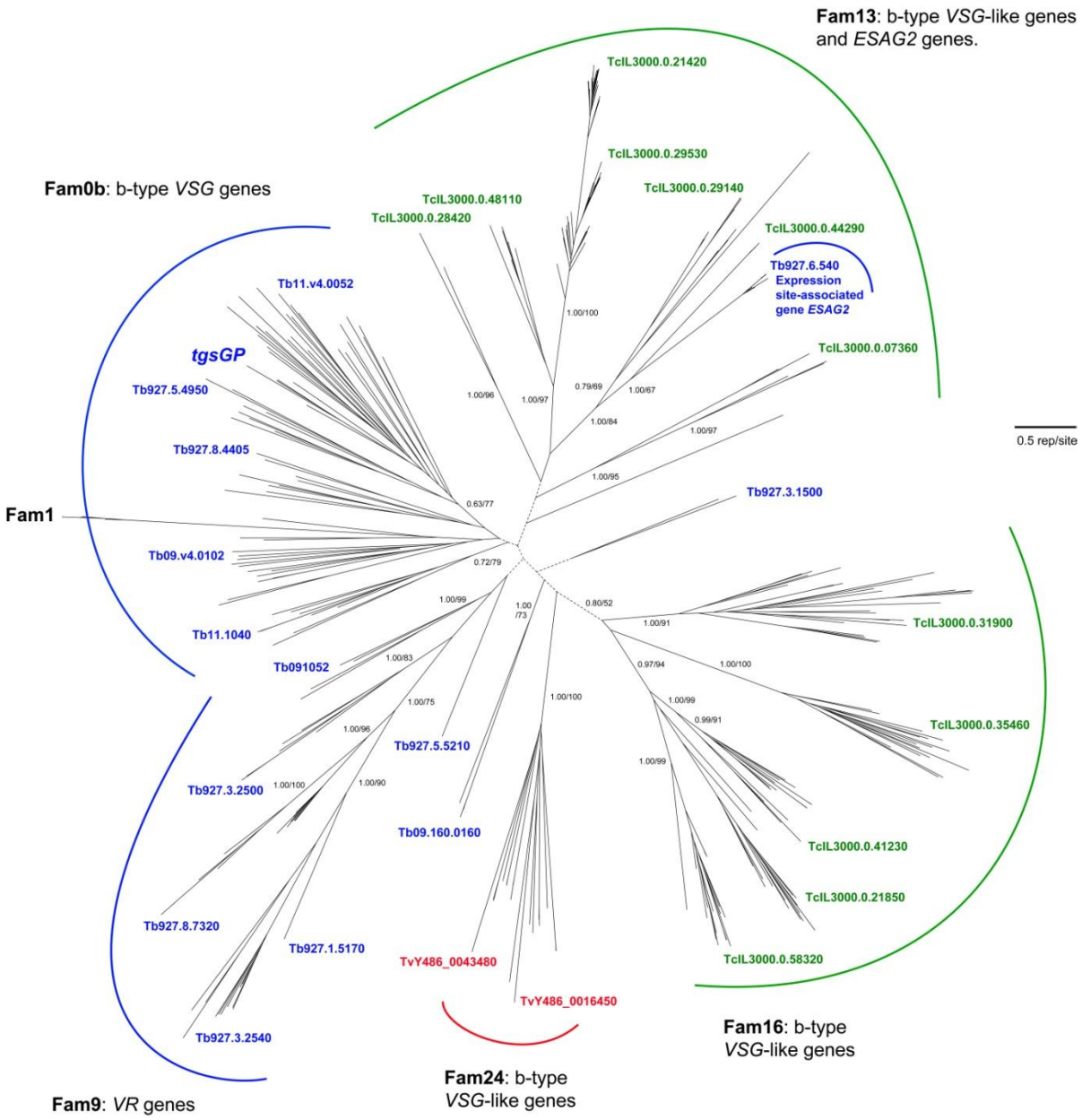
B.

MTAGAARILLALFAGCFCVLARNAAGGSDKA
IAGDDMGK**VCAASSALK**ATAAR**AGEEMLGAE**
ERTRHAVGNASAWQHTVQEQAANMSGTGKQA
AAWAKEETDKRIAKAADALRRVQRAALSVAR
RATRAAARI**DEMVILFTTYTSK**TSSGLACVK
AGSTR**TKPTSYGDGALTWAAGKGSTLK**GCAD
EKWTKGTTDLATDAAKLAETTKALGKLGAGT
AGKLFDGATTSSGVQHACPLLSGSGSSAITT
DYAQIYDTQYDSSTLTQMGGLWQVSAKSNAV
VLDLVEEDTDTVETSKKQPLKQLR**ADAAALW**
KSINATEPDTEEAETLEQALQQLAALPATAF
SVGTQSGERT**AEEWVQLEATLAQQK**KHAAKR
GPTRNDAGTAEQEGAQAITGGEATTVTDGRG
EETHTQATTGDKSAASSGTQRLYRAWALLAA
NTLGANARGAQHSGARGRLA

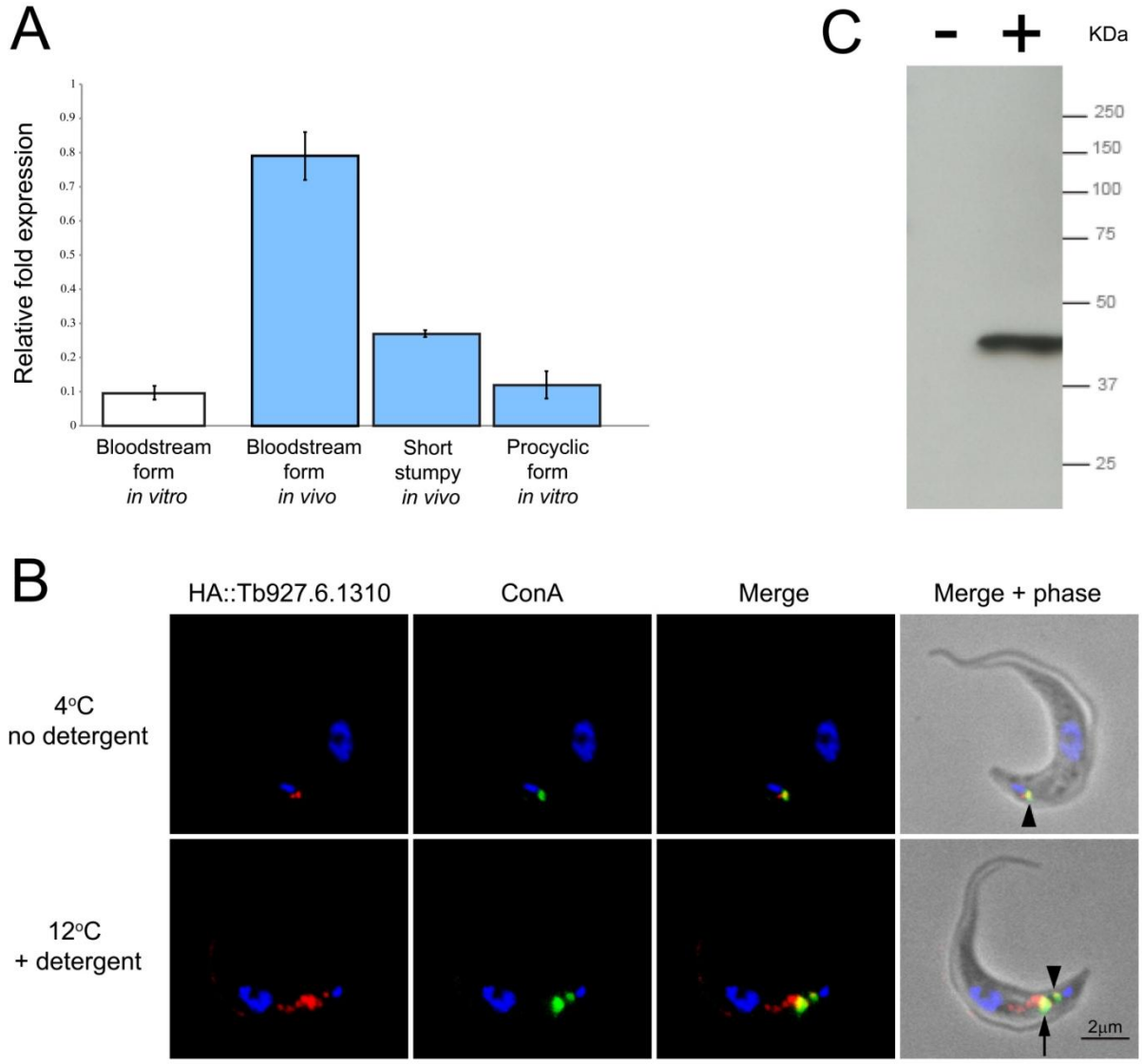
Supplementary Figure 2. Sequence of an expressed VSG in *Trypanosoma vivax* Y486. **A**. Relapse specific protein band yields a putative VSG. Trypanosome lysates derived from days 7 and 14 of a single infection initiated with a population of unknown VSG composition were separated by SDS-PAGE and stained with sypro-ruby. Bands were excised and analysed by mass-spec. The identities of proteins were derived from Mascot analysis, and the numbered bands were identified as: 1, putative VSG; 2,3,5,6. GAPDH (predicted isoforms in *T. vivax* are 35.6, 38.8 and 39.1 kDa); 4: actin (predicted 40.9 kD). **B.** Tryptic peptide sequences (underlined) from band 1 in **A**., determined by mass-spec analysis, mapped onto the protein predicted for TvY486_0027060.



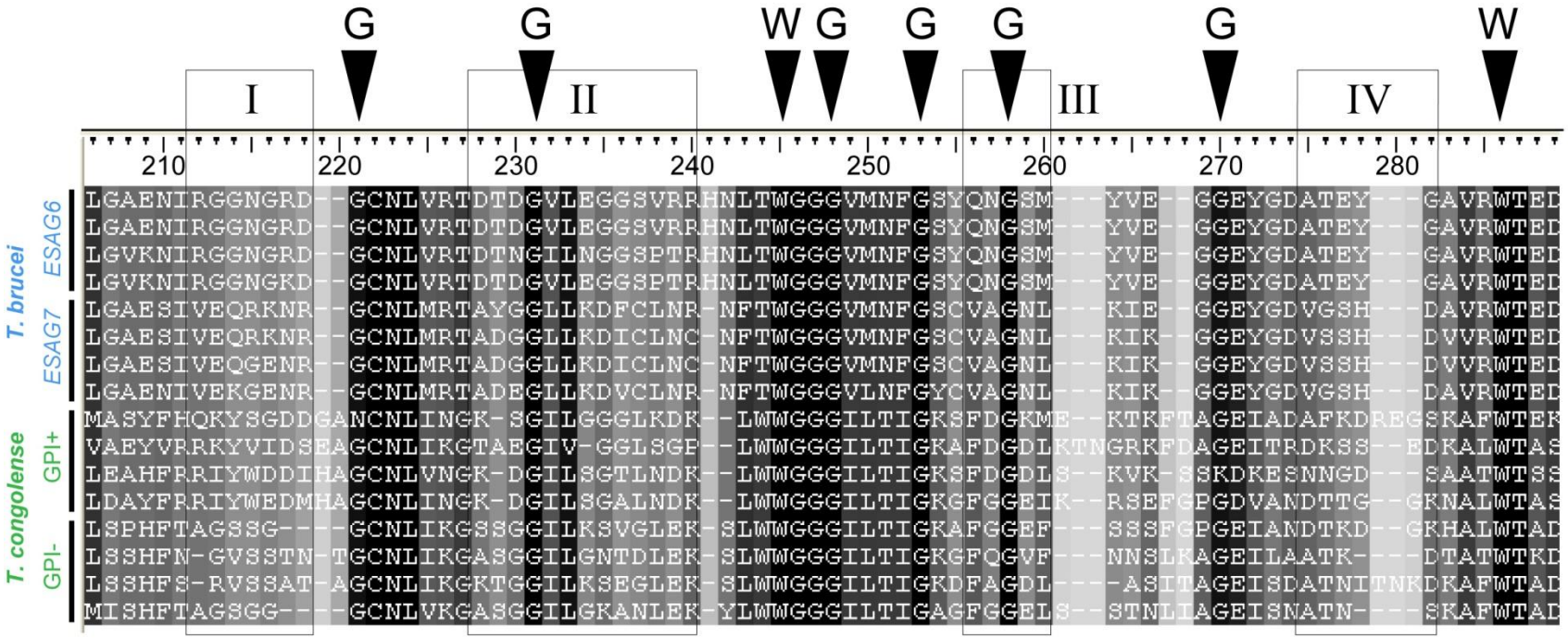
Supplementary Figure 3. a-VSG phylogeny. A Bayesian consensus phylogram was estimated from a multiple protein sequence alignment of 421 characters, using MrBayes with a WAG model and corrections for rate heterogeneity. Four MCMC chains were run for 5000000 generations to encourage maximal convergence of model parameters. The tree is unrooted. Selected nodes are supported by posterior probability values and non-parametric bootstraps generated from a maximum likelihood analysis using an LG model with rate heterogeneity, executed using RAxML. a-VSG from *T. brucei*, Fam23 from *T. vivax* and transferrin receptor-like genes from both *T. brucei* and *T. congolense* form three robust clades. Example gene names using their GeneDB/TriTrypDB locus identifers. Fam0a splits into two distinct clades corresponding to sequences with (upper) and without (lower) the 'GRIDE' motif generally characteristic of a-type VSG families.



Supplementary Figure 4. b-*VSG* phylogeny. A maximum likelihood phylogram was estimated from a multiple protein sequence alignment of 660 characters, using RAxML with an LG model and corrections for rate heterogeneity. The tree is unrooted. Basal nodes are particularly unstable and have very little or no support, these regions are indicated by dashed branches. Selected nodes are supported by non-parametric bootstraps and posterior probability values generated from a Bayesian analysis of the same data set in MrBayes. The tree contains major clades that correspond with BLAST-defined surface phylome families. In this tree canonical *T. brucei* VSG include the atypical *VSG*-related Fam1 gene family. The b-*VSG*-derived *tgsGP* sequence is also shown. All *VR* genes are positioned outside of canonical *T. brucei VSG* but are not monophyletic. *ESAG2* from *T. brucei* is nested within Fam13 *T. congolense* genes.



Supplementary Figure 5. Expression of a *T. brucei*-specific, VSG-related protein (Fam1). **A**. Quantitative reverse transcriptase (qRT) PCR analysis of Fam1 RNA expression in three different life cycle stages. RNA levels are relative to a constitutive gene (*Rab11*) RNA expression level set to 1.0. Error bars show the standard error of the mean for each triplicate experiment. **B**. Immunofluorescence analysis. Tb927.6.1310 was N-terminally HA epitope-tagged and expressed in bloodstream-form cells. Cells expressing HA::Tb927.6.1310 were loaded with FITC- concanavalin-A in serum-free media and incubated at either 4˚C (upper panel) or 12˚C (lower panel). ConA is restricted to the flagellar pocket at the lower temperature, whereas it accumulates within Rab5A positive early endosomes at 12˚C. Columns in each panel (from left to right); HA epitope-tag (red); FITC ConA (green); fluorescence merge and merged images from phase and fluorescence. DAPI-stain of the nucleus and kinetoplast is shown in blue. HA::Tb927.6.1310 colocalises with ConA at the flagellar pocket (indicated with an arrow head) and early endosomes (indicated with a whole arrow). Note that internal structures will not be visualized in the upper panel. The abundance of Tb927.6.1310 mRNA is rather low, and the data here suggest that while the protein is able to access the cell surface (the flagellar pocket is a specific sub-domain of the plasma membrane), the Tb927.6.1310 protein is restricted from the bulk plasma membrane. **C**. Western blot analysis of whole cell lysates from non-transfected (-) and transfected (+) trypanosomes.



Supplementary Figure 6. An amino acid sequence alignment for the ligand-binding domain of the transferrin receptor (ESAG6/7) in *T. brucei* and homologous sequences from *T. congolense*. The most conserved residues are shaded darkest and these describe a series of conserved glycine and tryptophan residues (labeled at top); the conserved motif WGGG denotes the site of a conserved peptidic turn. Four variable regions (I-IV) that are required for transferrin binding are shown (Salmon et al. 1997). Glycine residues in II and III are conserved throughout. The scale refers to positions in the complete Fam15 sequence alignment, available from the cell-surface phylome page (www.genedb.org/Page/trypanosoma_surface_phylome).