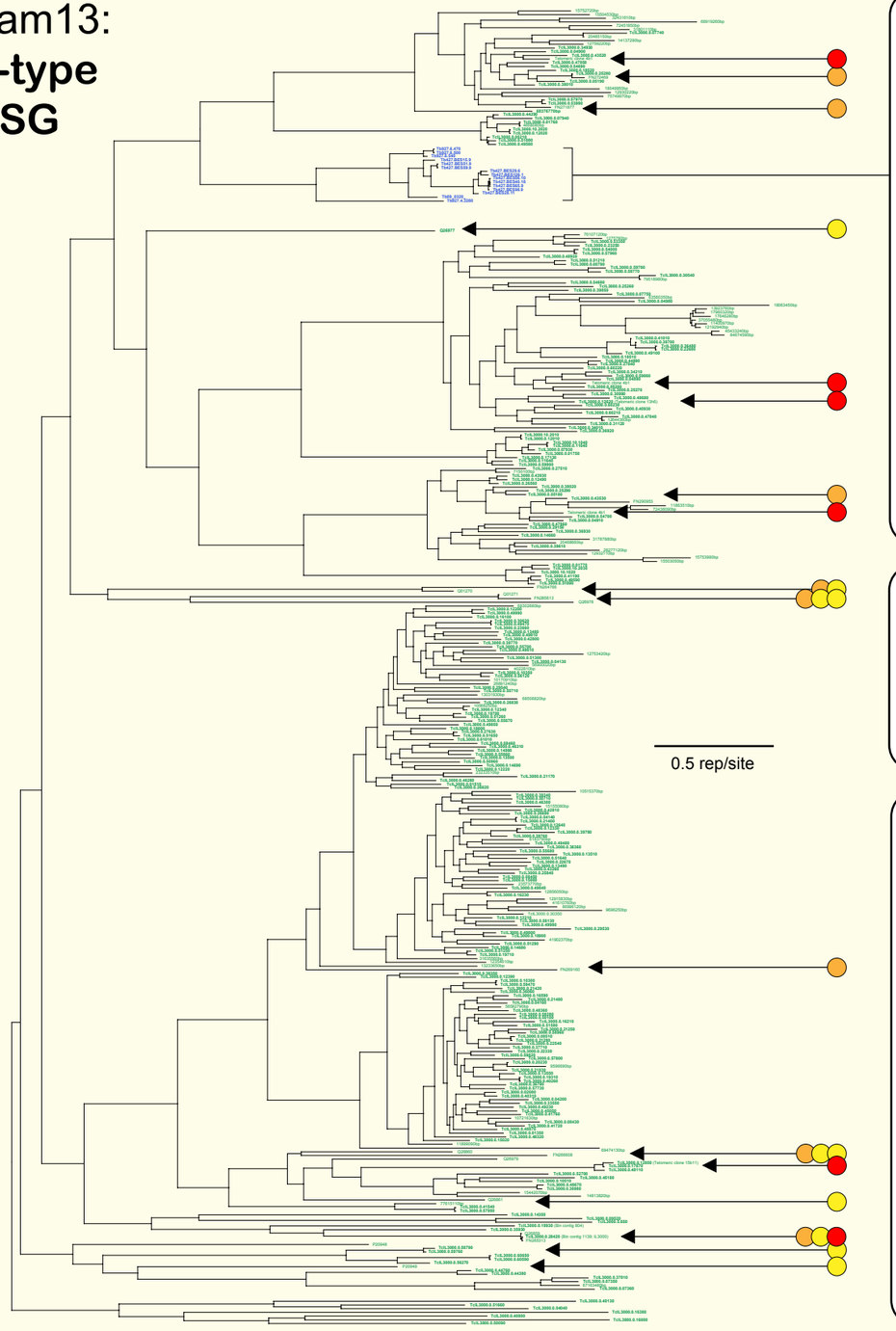
If you use this data, please cite:
Jackson, AP et al. 2012. A cell surface phylome for African Trypanosomes. *Manuscript submitted.*

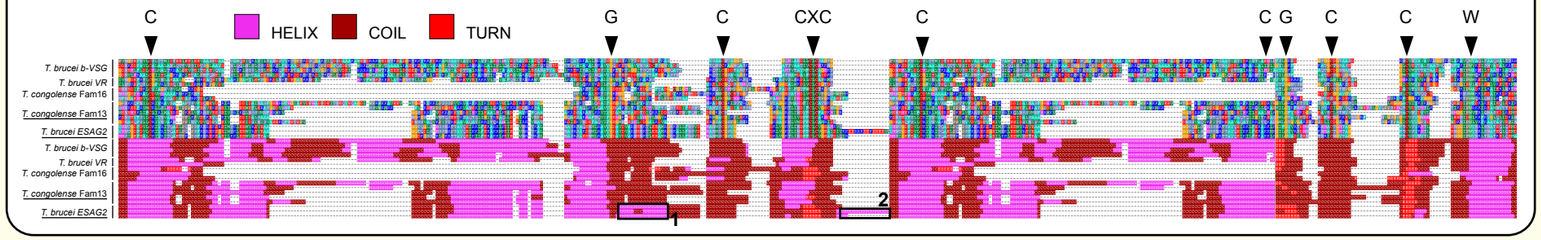
Fam13: **b-type VSG**



ESAG2 is a member of Fam13

A comparison of primary and secondary structures for various b-type VSG-like proteins is shown below. The amino acid sequences for canonical b-VSG, VR and ESAG2 in *T. brucei* are aligned with Fam13 and Fam16 b-type VSG from *T. congolense*; the signal peptide and variable C-terminal domains, (which do not align), have been removed. This demonstrates that the essential stucture and length of the b-VSG is conserved across all families, and well conserved residues are indicated with arrows. Below are the predicted secondary structures for these amino acid sequences, presented in the same order. Again, the arrangement of helix and coil regions is well conserved. The phylogeny clearly shows that ESAG2 from the *T. brucei* telomeric expression site, as well as GRESAG2 from the core chromosomes, form a clade nested within Fam13. *ESAG2* is more closely related to b-*VSG*-like genes in *T. congolense* than any other gene in *T. brucei*. This is reflected in their secondary structures; although ESAG2 is notable for two unique helices that are indicated by black boxes 1 and 2 below, its secondary structure is otherwise well matched to Fam13.

To explain the affinity between *ESAG2*, which does not function as a variant antigen, and some members of Fam13, which contains many genes encoding proven and putative variant antigens (see elsewhere on this page), we must suggest that these genes are descended from the same ancestral b-type *VSG* family in the common ancestor of *T. brucei* and *T. congolense.* It follows that *ESAG2* has acquired a novel function uniquely in *T. brucei*, and rather than being derived directly from canonical b-VSG in *T. brucei*, has evolved from a much older lineage that still contributes to antigenic variation in *T. congolense*. Although the function of *ESAG2* is not understood, it is an example of how *VSG* can be co-opted to novel roles, in this case associated with expression of the active variant antigen.



Published *T. congolense VSG*

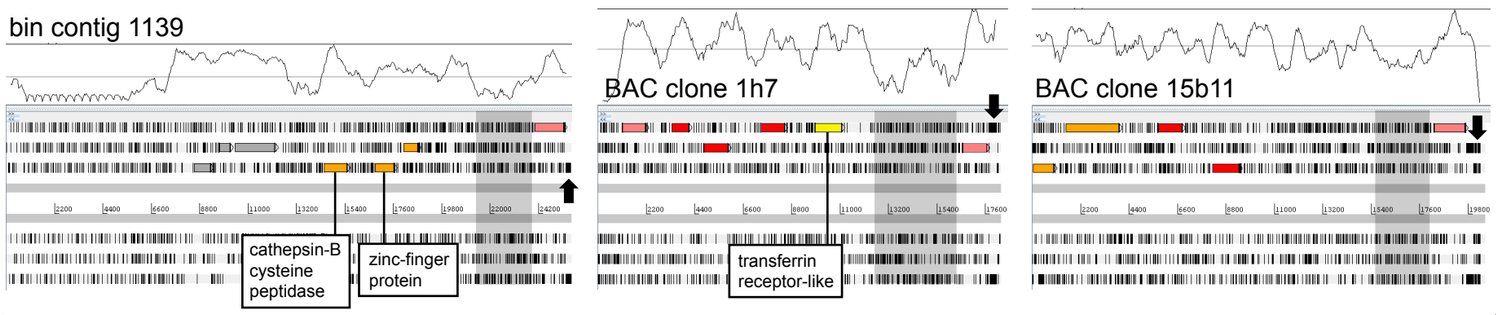
Of the 11 known VSG sequences deposited in UNIPROT, 10 cluster within Fam13, as indicated by a yellow circle. In most cases these VSG were isolated from different *T. congolense* strains*, and so their sequences are divergent and cluster outside of the major clades in this phylogeny. This demonstrates two points: 1) Fam13 contains proven variant antigens of diverse repertoire; and 2) we can expect substantial variation in Fam13 repertoire other strains.

VSG expressed sequence tags

A survey of EST in different *T. congolense* life stages identified the principal *VSG* transcripts expressed by IL3000 and associated with antigenic variation**. Where these are long enough to be aligned, these EST are included in the phylogeny and marked with an orange circle. Expressed *VSG* are placed throughout Fam13, providing further evidence that this family encodes functional variant antigens.

Telomeric *VSG*

Various *VSG* were observed adjacent to telomeric repeats on unplaced bin contigs and BAC clones specifically selected for the telomeric repeat. Seven such telomeric *VSG* that belong to Fam13 are included in the phylogeny and marked with a red circle. These genes are phylogenetically widespread and there appears to be nothing structurally distinctive about them, as in *T. brucei* where *VSG* within the telomeric expression sites are sampled from across the *VSG* repertoire. The panels below show Artemis screenshots of three telomeric regions; six reading frames are shown with a scale in bp, vertical black marks represent stop codons, coloured boxes represent genes, the telomeric repeat is marked with a black arrow, and base composition (%GC) is shown along the top. In all three cases the telomeric-proximal *VSG* (at right) is preceded upstream by a 'barren region' (shaded grey) that shows depressed GC content. Other *VSG,* of both *T. congolense* families (Fam13 in pink; Fam16 in red), are also found upstream. The left panel (bin contig 1139) contains the putative expression site of the expressed VSG in the genome strain (IL3000); this region also contains a *T. congolense*-specific cathepsin-B cysteine peptidase and a *T. congolense*-specific zinc-finger protein.



NOTES: **Fam13** comprises 302 *T. congolense* genes homologous to *ESAG2* in *T. brucei* and to known *T. congolense VSG,* as well as ESAG2 and related genes from *T. brucei*. The predicted proteins range in length from 391 to 457 amino acids and all have predicted signal peptides and GPI anchors. With the exception of two genes on chromosome 5, all gene copies are currently on unplaced contigs, which correspond to subtelomeric regions of megabase chromosomes, intermediate or minichromosomes.

The maximum likelihood phylogram was estimated from a multiple nucleotide sequence alignment of 649 characters (i.e. full-length sequences). The phylogram was estimated with PHYML using a LG model withrate heterogeneity. The tree is rooted using an outgroup comprised of less related b-type *VSG* from *T. congolense* belonging to Fam16. Secondary structures were estimated using PSIpred (Bryson K, et al. (2005) *Nucl. Acids Res*. **33**: W36-38). *Strickler, et al. (1987) *Biochemistry* **26**: 796-805; Eshita et al. (1992) *Gene* **113**: 139-148; Rausch et al. (1994) *Eur J Biochem*. **223**: 813-821. ** Helm *et al*. (2008) *Mol. Biochem. Parasitol.* **168**: 34–42.