If you use this data, please cite:

Jackson, AP et al. 2012. A cell-surface phylome for African Trypanosomes. *manuscript submitted*.



Fam22:

VSG-associated hypothetical protein

Type B, e.g. TcIL3000.0.20240 (n=64) Fam22 falls into two very obvious subtypes, with a few sequences combining elements of both. These sequences may be the result of recombination between types A and B.

TcIL3000.0.15980

TcIL3000.0.51630
TcIL3000.0.12650
TcIL3000.0.59710
TcIL3000.0.10060
TcIL3000.0.51870

CIL3000.0.51870 TCIL3000.0.42780

0.1 sub/site

Type A, e.g. TclL3000.0.51280 (n=85)

Fam22 is distributed throughout the *T. congolense* subtelomeric regions (*inset, showing a subtelomeric contig in Artemis; genes are shaded green* (VSG), *orange* (TFR-like) and blue (Fam22)). Fam22 genes are always positioned immediately downstream of VSG copies, (of various lineages). This might suggest that Fam22 is simply an open reading frame within the 3'UTR's of VSG. But it is known that these 3'UTR's are extremely short, perhaps only 15-30bp*.

Coupled with the observation that Fam22 transcripts are observed in all life stages except for bloodstream forms, (J. Donelson, unpublished data), this suggests that Fam22 is cotranscribed with *VSG*. However, it is not known if this mRNA is translated.



NOTES:

Fam22 comprises 157 putative coding regions from *T. congolense* subtelomeric regions. These open reading frames encode a predicted protein of 175-186 amino acids and have no matches to known protein sequences or domains. ~9% of sequences are predicted to have signal peptides (e.g. TcIL3000.0.13450), although this depends on where the start codon is placed, which is ambiguous.

* Rausch *et al.* (1994) *Eur. J. Biochem.* 223: 813-821; Helm *et al.* (2008) *Mol. Biochem. Parasitol.* 168: 34–42.

The maximum likelihood phylogram was estimated from a multiple nucleotide sequence alignment of 564 characters, using PHYML and a GTR+F model. The tree is unrooted.