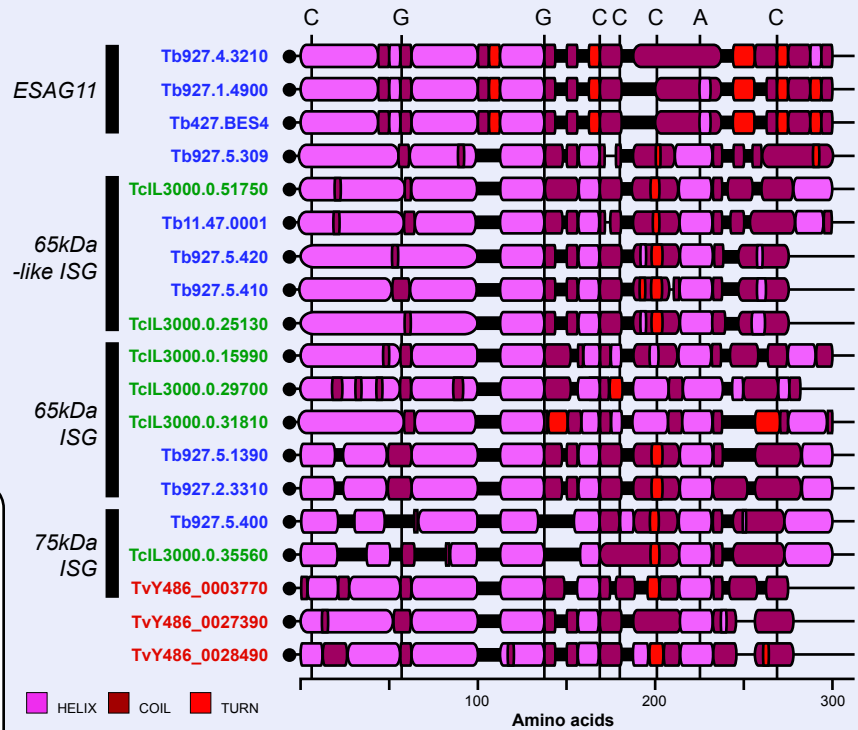
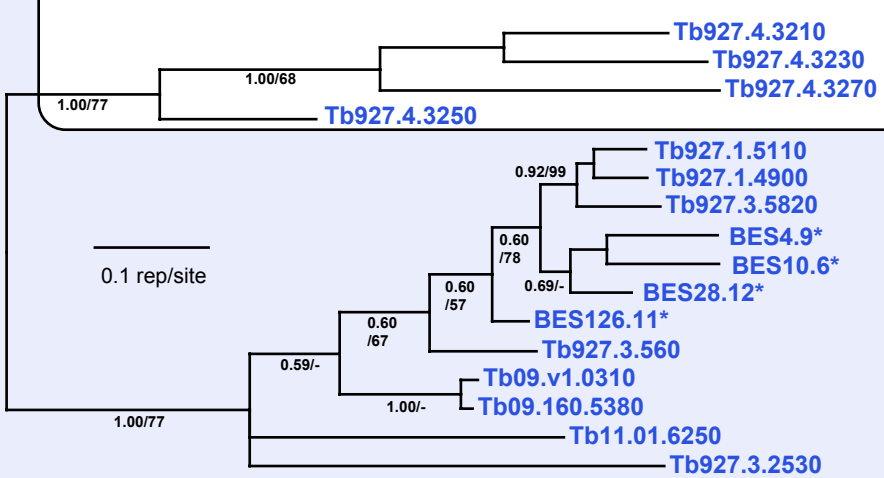
If you use this data, please cite:

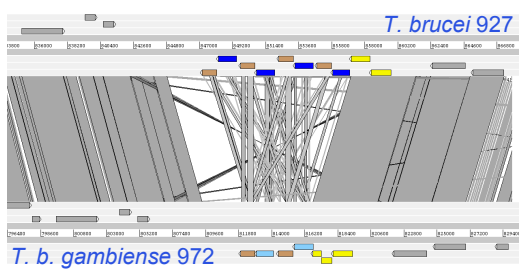
Jackson, AP et al. 2012. A cell-surface phylome for African Trypanosomes. *manuscript submitted.*

Fam3:
Expression-site associated gene 11-like genes



This clade includes four tandem gene copies with a conserved genomic position with *T. brucei* and closely associated with both *ESAG2*-related genes (Fam13) and another *T. brucei*-specific gene family (Fam4) encoding a putative GPI-anchored protein *(left; showing an ACT comparison of a strand-switch region conserved in* T. brucei *and* T. b. gambiense *containing Fam3 (blue) alternating with Fam4 (brown) and followed downstream by two Fam13 gene copies (i.e. VSG/ESAG2-like).*





NOTES:

Fam3 comprises a selection of *ESAG11* genes from *T. brucei* 427 (marked with asterisks), as well as all homologous genes found on the core chromosomes of *T. brucei* 927. The Bayesian phylogram was estimated from a multiple protein sequence alignment of 357 characters, using MrBayes under default settings. The tree is midpoint-rooted. Nodes are supported by posterior probability values and non-parametric bootstraps generated from a maximum likelihood analysis using an LG model with rate heterogeneity.

ESAG11 was initially described occupying a locus between *ESAG1* and *2* in the bloodstream expression site (Redpath *et al.* 2000. Mol Biochem Parasitol. 111: 223-8). Subsequently, various homologs in sub-telomeric and strand-switch regions were identified; generally these encode intact proteins, in contrast with *ESAG11* which are almost always pseudogenes. *ESAG11* is atypical in that there are essentially no structural distinctions between ES-linked and most sub-telomeric copies, and so ES-linked genes are not monophyletic. This may be due to secondary transpositions from the expression site to the sub-telomeres. Therefore, it cannot be determined whether genes inside or outside the expression site evolved first.

The conserved cysteine residues common to all members of the invariant surface glycoprotein (ISG) family were also found in ESAG11, and we propose that *ISG* and *ESAG11* are homologous. Searching all African trypanosome predicted proteins with a HMM designed against all ISG (with variable C-termini removed) found only matches to self and then ESAG11 (Tb927.1.4900; p = 7.3x10-5); ESAG11 matches straddled the inclusion threshold. A comparison of predicted secondary structures (shown, top right) demonstrates that ESAG11 and ISG share a series of α-helices towards the N-terminus, followed by several conserved coils. It also demonstrates that *ESAG11* is more divergent than any recognized ISG. Homology between ESAG11 and ISG is consistent with its original description, since *ESAG11* was detected in a screen of a *T. brucei* bloodstream-form cDNA expression library using an antiserum raised against 64kDa ISG (Redpath et al. 2000).