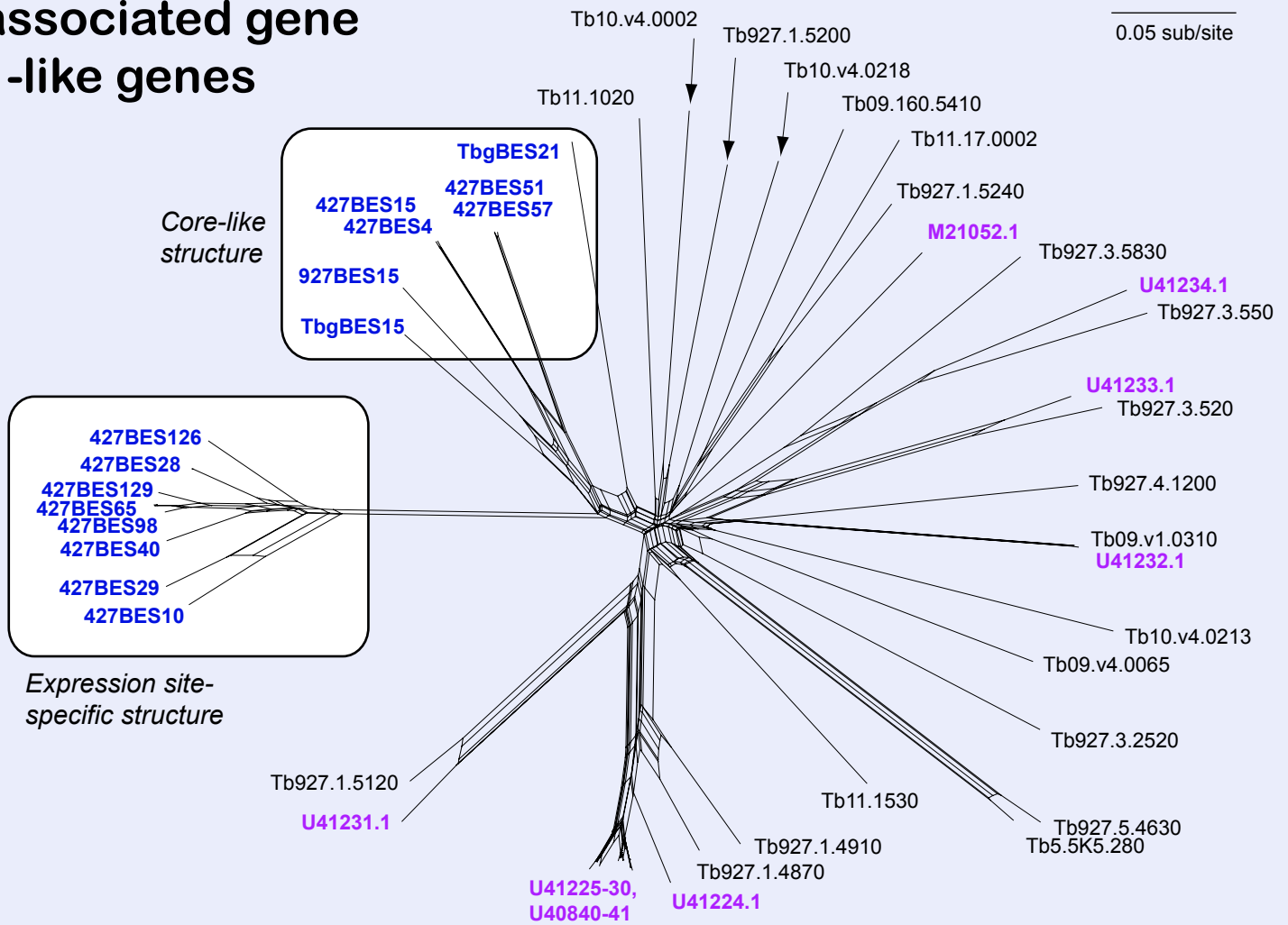


If you use this data, please cite:

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

Fam8: Expression-site associated gene 1-like genes



NOTES: Fam8 comprises all *ESAG1* genes from *T. brucei* 427 telomeres (*Hertz-Fowler *et al.* (2008) *PLoS one*, **3**: e3527), and all homologous genes found on core chromosomes in *T. brucei* 927.

Of all components of the expression site, *ESAG1* (Cully *et al.* 1985, 1986) is perhaps the most enigmatic. This gene is present in most expression sites immediately upstream of the 70bp repeat, and related forms are found throughout the *T. brucei* 927 genome at locations in sub-telomeric and strand-switch regions. The predicted *ESAG1* protein has no discernable similarity with known structures, and no homologs to *ESAG1* are present in either *T. congolense* or *T. vivax*.

In an effort to characterize the gene, Morgan *et al.* (1996 *J Biol Chem.* 271(16):9771-7) obtained 20 distinct *ESAG1* sequences from a *T. brucei* cDNA library, showing that all were expressed in the bloodstream stages, and demonstrating clear evidence for recombination between loci. When these sequences are combined with genome data in a phylogenetic network, it is clear that Morgan *et al.* did not sample *ESAG1 sensu stricto*, but rather various transcripts belonging to conserved subtelomeric loci.

The *ESAG1* phylogenetic network shows that ES-linked and non-ES copies are reciprocally monophyletic. Like other ESAG families, this suggests that there is an ES-specific sequence type and that other types are excluded from the expression site; although the network also implies that some ES-linked copies are more closely related to subtelomeric genes than others (e.g. "TbgBES21"). Examining the sequence alignment it is clear that this could be due to recombination between ES-linked and subtelomeric loci, as seen previously (Morgan *et al.* 1996).

An unrooted phylogenetic network of *ESAG1*-like genes in *T. brucei*. is shown. The network was estimated using Splittree v4 from a nucleotide sequence alignment of 1015 characters including all *ESAG1*-like genes from *T. brucei* 927 as well as *ESAG1 sensu stricto* (shaded blue) from *T. brucei* 927, 427 and *T. b. gambiense*. For comparison, *ESAG1*-like cDNA obtained from *T. b. rhodesiense* (Morgan *et al.* 1996) are also included and identified by their GenBank accession numbers (shaded purple).