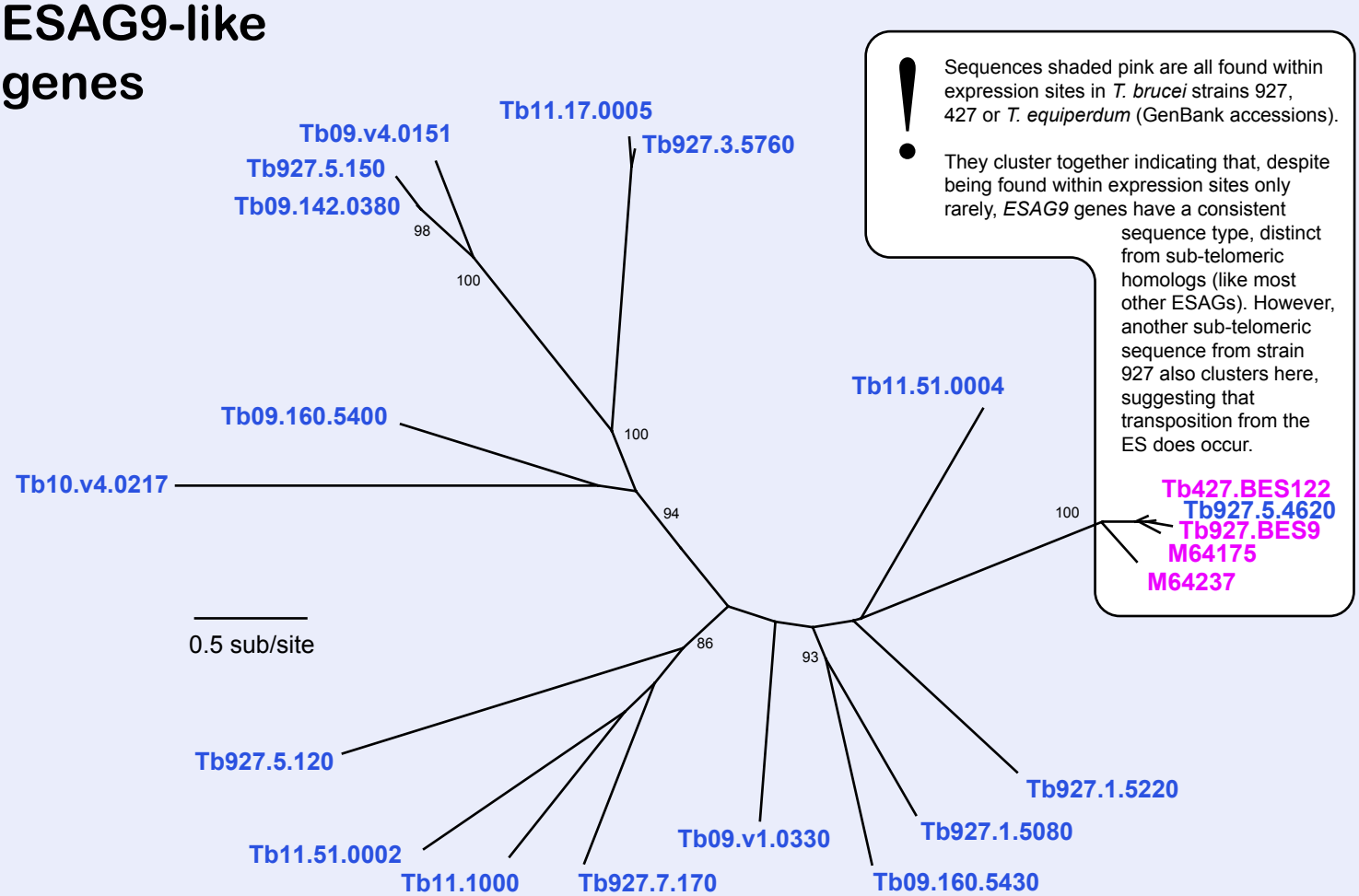
If you use this data, please cite:

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

Fam2: **ESAG9-like genes**



Sequences shaded pink are all found within expression sites in *T. brucei* strains 927, 427 or *T. equiperdum* (GenBank accessions).

They cluster together indicating that, despite being found within expression sites only rarely, *ESAG9* genes have a consistent sequence type, distinct from sub-telomeric homologs (like most other ESAGs). However, another sub-telomeric sequence from strain 927 also clusters here, suggesting that transposition from the ES does occur.

NOTES:

Fam2 comprises the *ESAG9* gene family, which was first described in *T. brucei* by Florent *et al*. (1991) and shown to have bloodstream stage-specific expression emanating from both telomeric and sub-telomeric loci. *ESAG9* from *T. brucei* 927 encode short proteins (252-297 amino acids) with predicted signal peptides and GPI anchors, indicating a cell surface role.

Numerous subtelomeric loci are present in *T. brucei* 927, although these are uniformly conserved in other strains (Barnwell *et al.* 2010), while *ESAG9* is only infrequently found in expression sites. Only one of 21 expression sites (ES) described in *T. brucei* 427 contained *ESAG9* (Hertz-Fowler et al. 2008). The same ES-specific sequence type occurs in multiple *T. brucei* strains, which may indicate that ES-linked *ESAG9* originated through duplicative transposition from a single sub-telomeric locus. However, Tb927.5.4620 is a subtelomeric gene in *T. brucei* 927 with a sequence that is almost identical to ES-linked *ESAG9* in another strain (Barnwell et al. 2010); so secondary transpositions from the expression site back to the sub-telomere could also have occurred.

It may be that there is no functional distinction between gene copies inside and outside of the ES, and perhaps this explains why *ESAG9* is not ubiquitous among expression sites.

We have found no homologs to *ESAG9* in *T. congolense* or *T. vivax,* suggesting that it is *T. brucei*-specific. However, a recent study (Barnwell *et al*. 2010 J Cell Sci. 2010 123(19):3401-11) noted the similarity of *ESAG9* to Mucin-Associated Surface Proteins (MASP) in *T. cruzi*. MASP are a principal component of the *T. cruzi* cell glycocalyx and a highly diverse gene family (Bartholomeu *et al*. 2009). It may be that this lineage has persisted in African trypanosomes in much lower abundance within the sub-telomeres since the split with *T. cruzi*.

Several loci are not annotated as *ESAG9*-like sequences at the time of writing, and appear as hypothetical proteins: Tb09.142.0370, Tb927.5.150, Tb09.v4.0151, Tb927.3.5760,

The Bayesian phylogram was estimated from a multiple nucleotide sequence alignment of 1149bp, using MrBayes with a GTR+Γ model and default settings. The tree is unrooted. Nodes are supported by posterior probability values and non-parametric bootstraps.