If you use these data, please cite: wellcome trust Jackson, AP et al. 2012. A cell-surface phylome for African Trypanosomes. manuscript submitted. Cene Key: Tb927. Fam49: T. brucei **Invariant surface** 0.25 rep/site TcIL3000. T. congolense glycoprotein TvY486 TclL3000.0.38640 T. vivax 0.84/ Tb927.5.309b 0.97/ TclL3000.0.51750 TclL3000.0.51760 Tb927.5.610 0.93/61 Tb927.5.620 0.94/-Tb927.5.630 Tb11.47.0001 65kDa ISG-like TclL3000.0.08580 1.00/98 Tb927.5.410 0.70/ Tb927.5.420 Tb927.5.430 1.00/9 — TclL3000.0.25130 - TclL3000.0.22060 — TclL3000.0.42190 1.00/100 TclL3000.0.29690 TclL3000.0.29700 TclL3000.0.06490 1.00/100 0.89/ 1.00/-TclL3000.0.12460 0.98/9 TclL3000.0.15990 1.00/8 1.00/100 TclL3000.0.42200 0.64 TclL3000.0.42210 - TclL3000.0.22080 0.70/96 TclL3000.0.18780 TclL3000.0.31810 1.00/-.2.3270 Tb927.2.3300 Tb927.2.3290 Tb927.2.3280 Tb927.2.3310 65kDa ISG 1.00/100 1.00/92 rb927.5.1410 64kDa ISG 0927.5.1390 vY486 0014380 486 05039 1.00/90 1.00/99 0 87/59 1 00/10 030320 VĂÑR 0013080 102450 486 0045640 1.00/84 1.00/100 75kDa ISG 1.00/9 Tb927.5 Tb927.5.370 Tb927.5.390 0 73/ Tb927.5.380 TclL3000.0.15690 TclL3000.0.35530 0.81/-亡TclL3000.0.27250 TclL3000.0.35560 Tcll.3000.0.15280 1.00/ Y486\_0003770 Y486\_0009410 ISG are unique to the African trypanosomes, absent in T. cruzi or Leishmania. They split into two clades in this Bayesian phylogeny, roughly corresponding to 65kDa-like and 75kDa-like genes. T. brucei lineages have orthologous lineages in T. congolense, indicating that the ISG repertoire is conserved between these species, although genomic locations may not be. Within the 65kDa-like clade, there is no orthology between individual loci in T. brucei/T. congolense and T. vivax (the sequences of which are monophyletic). This indicates that either ISG diversified independently in these two lineages, or, more likely, gene duplication proceeds quickly enough for ancestral sequence types to be replaced by derived, species-specific sequences. ISG loci are typically arranged in tandem arrays and found at strand-switch regions or core/sub-telomeric boundaries in T. brucei. The draft data indicate that this is also true in T. congolense and T. vivax. NOTES: Fam49 comprises the 64, 65, and 75kDa invariant surface glycoproteins as well as related sequences not previously annotated as ISG. While conserved across all African trypanosomes, ISG tend to be found at several species-specific loci throughout the core chromosomes and subtelomeric regions. All members encode proteins with a predicted signal peptide and a transmembrane helix towards the C-terminus. The Bayesian phylogram was estimated from a multiple protein sequence alignment of 508 characters, using MrBayes under default settings. The tree is mid-point rooted as no homolog exists in T. cruzi. Selected nodes are supported by posterior probabilities and non-parametric bootstraps generated from a maximum likelihood analysis under a GTR+Γ model.