**Figure S1. Representation of SNARE, Rab and TBC coding sequences in selected eukaryotic genomes and kinetoplastids. Genomes are arranged by phylogenetic relationships.** The five classically recognised, sensu Adl 2005, eukaryotic super groups and each sub-group of kinetoplastida are colour-coded according to the colour key on either side of the dividing dashed line respectively. Blue symbols and solid line represent the total coding content of the respective organism by total number of predicted ORFs (reads are shown on the y-axis, right). Numbers of SNARE, Rab and TBC ORFs are represented by dark, medium and light gray bars respectively (x-axis, left).

Adl, S. M., Simpson, A. G. B., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredricq, S. et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**, 399-451.
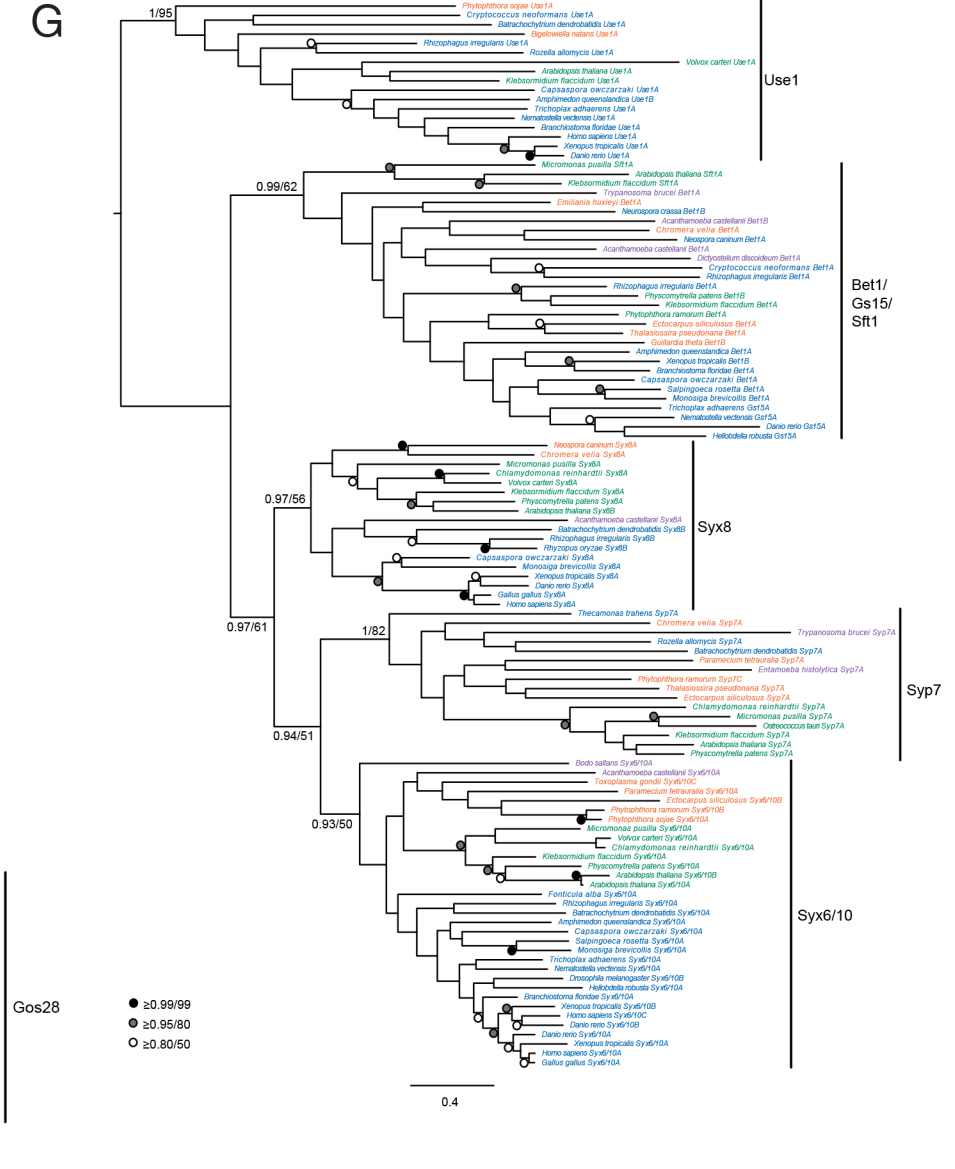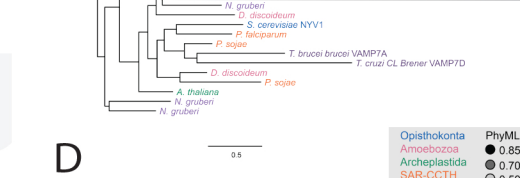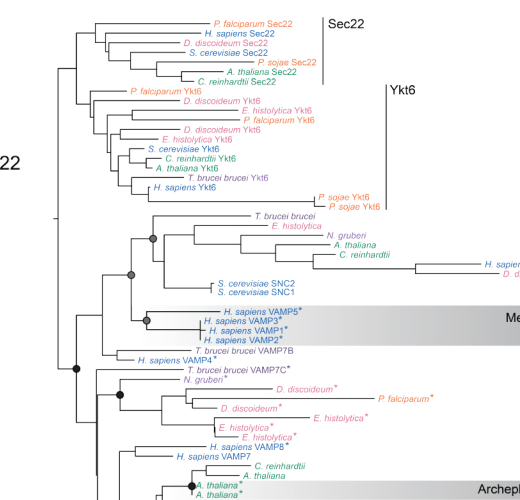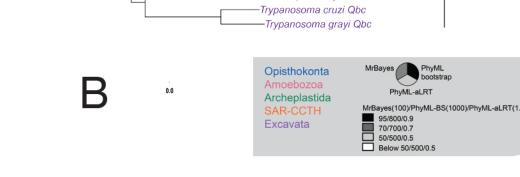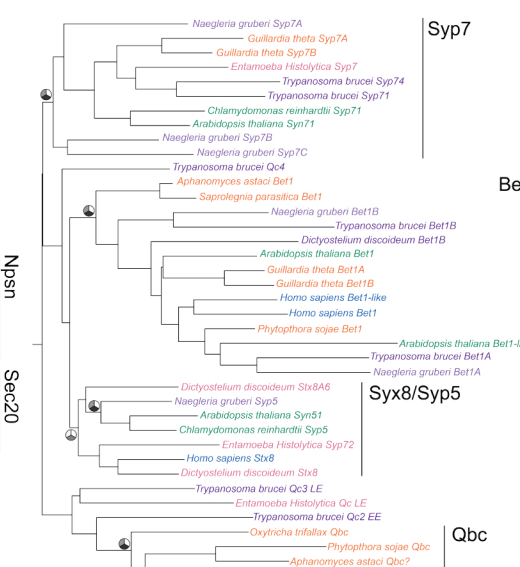
Figures S2. Phylogenetic assignment of kinetoplastid SNAREs (A-C), SNAP-25 (D) and synaptobrevins (E) and Qb (F) and Qc SNAREs (G) confirming NPSN and Syp7 were part of the LECA complement. Best PhyML topology of Qb (S2A) and Qc (S2B), and best Bayesian topology of R (S2C) SNARE phylogeny is presented. Node values are iconised as pie charts for three support values each representing PhyML approximate likelihood ratio test, PhyML Bootstrap and MrBayes posterior probabilities and colour-coded as shown in the key. Each phylogeny shows one representative kinetoplastid SNARE from each sub-type cluster (Purple) along with eukaryotic representative SNAREs from Opisthokonta (Blue), Amoebozoa (Pink), Archaeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple). In S2A, Qb SNAREs showing orthology with eukaryotic orthologs for Gos1, Npsn, Sec20 and Bos1. Qb2 and Qb3 are putative Vti SNAREs on the basis of BLAST and reverse BLAST into *H. sapiens* and *S. cerevisiae* genomes as well as our proteomic data. In S2B, Qc SNAREs showing orthology with Bet1 and Syp7 and a putative QbcSNARE is shown. Qc2, 3 and 4 have not been sufficiently confidently placed. In S2C, R-SNAREs showing orthology with eukaryotic Ykt6, Sec22 and VAMP7. R1-SNARE can be deduced to be an R.reg Tomosyn-like SNARE from the sequence length and domain structure, but its orthologs are not presented due to formation of long branches, possibly because of the derived nature of the proteins. Note Sec22-like protein has no SNARE domain but only a single longin domain. Figure S2D shows the best PhyML topology rooted on Qb-Gos1 sequences is present. Eukaryotic representative SNAP-25like SNAREs identified in Opisthokonta (Blue), Amoebozoa (Pink), Archaeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple) are shown. Note expansions in archaeplastids and opisthokonts are lineage specific. Higher-level relationships between clusters is not resolved, but presence of several separate clusters indicates divergence of sequences in different lineages. Kinetoplastid Qbc-like sequences are found to cluster with representative stramenopile sequences (*A. astacii* and *P. sojae*), marked with a vertical line. Figure S2E shows all R-SNARE synaptobrevin domain (IPR01388) containing sequences from selected eukaryotic representatives were analysed. PhyML bootstrap and Bayesian analyses were inconclusive due to very low supports and unresolved relationships respectively so only the PhyML aLRT analysis is shown. Statistical support at key nodes are presented as circles filled in in gray-scale according to the key shown. Eukaryotic representative sequence are colour-coded as Opisthokonta (Blue), Amoebozoa (Pink), Archeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple). Sec22 and Ykt6 are conserved compared to other VAMPs. R.reg forms long branches likely due to derived nature of the sequences. 'Brevin'-like VAMPs, lacking the N-terminal longin domain are marked with an asterisk (*). Their presence in different clusters indicates a likely convergent lineage specific evolution. Robust reconstruction of an NPSN (F) and Syp7 (G) clade, including sequences from all supergroups, confirms the ancient origin of NPSN and Syp7. Accession numbers and sources are listed in Tables S4-S6.
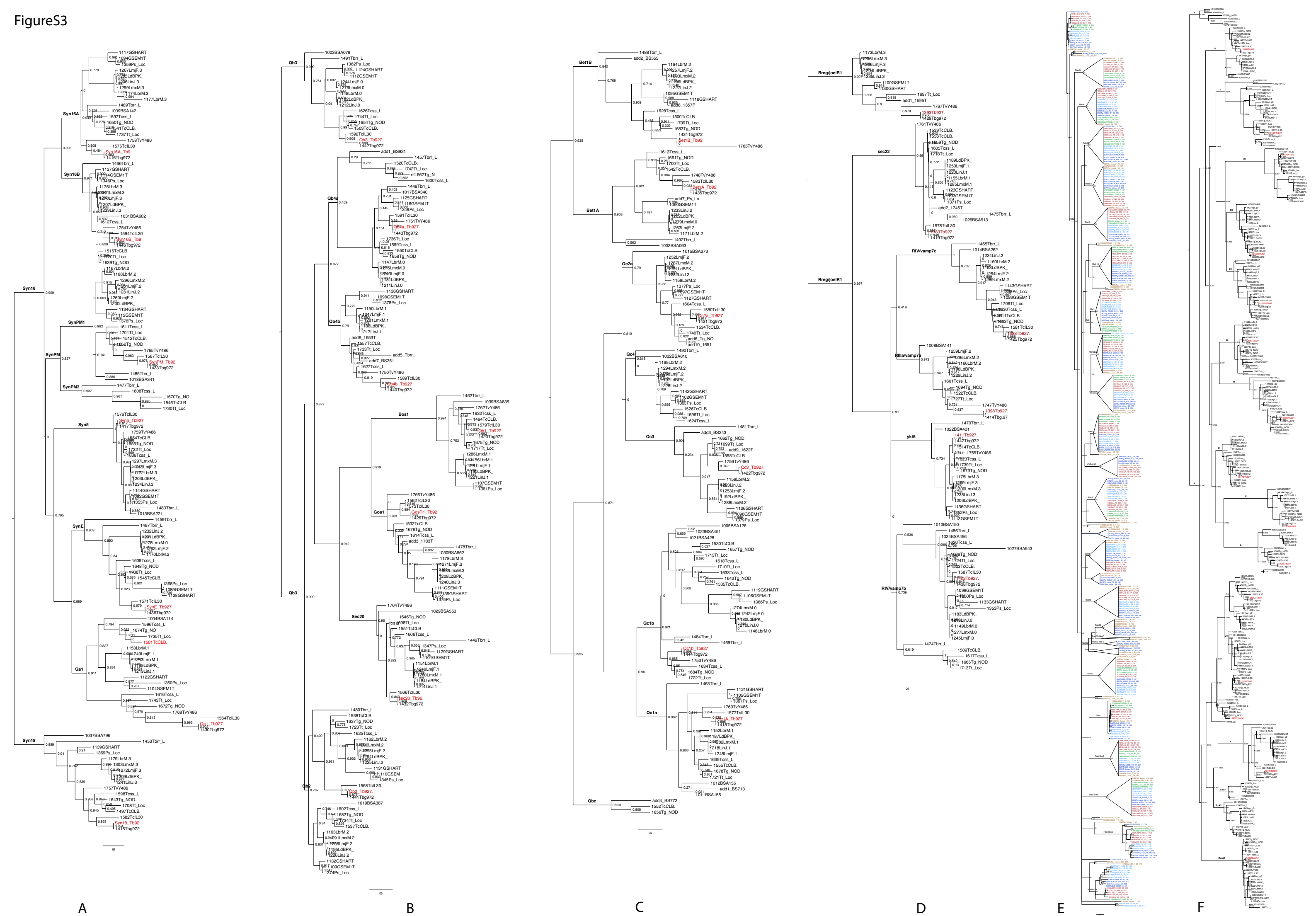
FigureS3

A   B   C   D   E   F

Figure S3A-F. Phylogenetic tree of kinetoplastid SNAREs (A-D), Rabs (E) and TBCs

(F). PhyML topology of all kinetoplastid Qa (S3A), Qb (S3B), Qc (S3C) and R (S3D) SNAREs

is shown. PhyML approximate likelihood ratio test values are shown at nodes, and clades

are labelled with assigned identity on branches. T. brucei IDs are shown in red. In S3E,

MrBayes topology of all kinetoplastid Rab sequences is shown. Mr Bayes posterior

probabilities (100) are indicated at nodes and clades are labelled with assigned identity on

branches. In S3F, PhyML topology of all kinetoplastid TBC RabGAPs is shown. PhyML

approximate likelihood ratio test values are shown at nodes, and clades are labelled with

assigned identity on branches.

Table S1. Accession numbers. Sequences in each dataset (Rab, TBC, and SNARE) were given unique 4-digit ID codes that can be cross-referenced with the trees to check sub-family assignment and the fasta files to pull out the sequence.

Click here to Download Table S1

Table S2. Raw data from proteomic analysis. Raw data of mass spectrometry results from four buffer conditions as described in methods. WT = results from untagged cell line, 790 = results from TbVAMP7C::HA.

Click here to Download Table S2

Table S3. Protein sequence data accessed for analysis of Qb and Qc SNARE LECA complement. Resources are listed by species name. References to relevant publications are also included as appropriate.

Click here to Download Table S3

Table S4. Qb and Qc SNAREs identified in a diverse sampling of eukaryotes. Accession numbers and annotations are listed for sequences identified through a combination of homology searching and phylogenetics analysis, as described in the methods. Sheet 1: Qb, sheet 2: Qc. Sources of sequence data are listed in Table S4

Click here to Download Table S4