University of Dundee

Evolution of the endomembrane systems of trypanosomatids

Venkatesh, Divya; Boehm, Cordula; Barlow, Lael D.; Nankissoor, Nerissa N.; O'Reilly, Amanda; Kelly, Steven; Dacks, Joel B.; Field, Mark

Link to publication in Discovery Research Portal

Evolution of the endomembrane systems of trypanosomatids: Conservation and specialisation

Divya Venkatesh[1,2], Cordula Boehm[1], Lael D. Barlow[3], Nerissa N. Nankissoor[3], Amanda O'Reilly[2], Steven Kelly[4], Joel B. Dacks[3] and Mark C. Field[1]*

[1]Wellcome Trust Centre for Anti-Infectives Research, School of Life Sciences, University of Dundee, Dow Street, Dundee, DD1 5EH, UK, [2]Department of Pathology, University of Cambridge, Tennis Court Road, CB2 1PQ, UK [3]Department of Cell Biology, University of Alberta, Edmonton, Alberta, Canada, T6G 2H7, [4]Department of Plant Sciences, University of Oxford, Oxford, OX1 6JP, UK

*Corresponding author: Tel: +44 (0)751-550-7880, email: mfield@mac.com

Abstract

Parasite surfaces support multiple functions required for survival within their hosts, and maintenance and functionality of the surface depends on membrane trafficking. To understand the evolutionary history of trypanosomatid trafficking, where multiple lifestyles and mechanisms of host interactions are known, we examined protein families central to defining intracellular compartments and mediating transport, i.e. Rabs, SNAREs and RabGAPs across all available Euglenid genomes. Bodonids possess a large trafficking repertoire, mainly retained by the cruzi group, with extensive losses in other lineages, particularly African trypanosomes and phytomonads. There are no large-scale expansions or contractions from an inferred ancestor, excluding direct associations between parasitism or host range. However, we observe stepwise secondary losses within Rab and SNARE cohorts (but not RabGAPs). Major changes are associated with endosomal and late exocytic pathways, consistent with the diversity in surface proteomes between trypanosomatids and mechanisms of interaction with the host. Along with the conserved core family proteins, several lineage-specific members of the Rab (but not SNARE) family were found. Significantly, testing predictions of SNARE complex composition by proteomics confirms generalised retention of function across eukaryotes.

Introduction

Membrane trafficking mediates delivery of macromolecules to discrete intracellular compartments from the site of uptake or synthesis. Trafficking is essential to nearly all eukaryotic cells, contributing towards nutrient acquisition, protein processing and turnover, and in multicellular organisms also participates in higher order tissue organization. The importance of membrane transport is reflected in the many diseases associated with trafficking, including diabetes, Alzheimer's disease and cystic fibrosis (Birault et al., 2013; Olkkonen and Ikonen, 2006; Rajendran and Annaert, 2012; Seixas et al., 2013). Development of membrane trafficking was likely a major evolutionary driver enabling the transition from prokaryotic to eukaryotic cells (Field et al., 2011; Dacks et al., 2016). For many pathogens trafficking has special relevance in maintaining the host-parasite interface, the cell surface and both the surface and underlying trafficking apparatus are intimately connected with immune evasion, pathogenesis and life cycle progression (Manna et al., 2013).

Membrane transport requires vesicle formation, translocation, tethering, docking and fusion to release cargo (Bonifacino, 2004), with coordinated action by Rab and ARF GTPases, coat complexes, tethers and SNAREs. Members of these paralogous families encode specificity for individual transport events to define organelles; evolutionary reconstructions suggest stepwise evolution from a simpler ancestral system before the last eukaryotic common ancestor (LECA) (Devos et al., 2004; Devos et al., 2006; Dokudovskaya et al., 2006; Field and Dacks, 2009; Schlacht et al., 2014). A conserved core of Rab, Rab-GAP and SNARE proteins has been derived by reconstructing eukaryotic evolutionary history and broadly supports this model (Arasaki et al., 2015; Diekmann et al., 2011; Elias et al., 2012; Kienle et al., 2009a; Vedovato et al., 2009; Yoshizawa et al., 2006). Beyond this are examples of lineage-specific features in the ESCRT system, sortillins and ARF GTPases in metazoans (Field et al., 2007; Gabernet-Castello et al., 2013; Leung et al., 2008). In parasitic organisms, adaptin complexes, important cargo selectors which are otherwise well conserved in most lineages, are rather variable, possibly connected to specific adaptation (Woo et al., 2015).

Kinetoplastids are unicellular flagellated protists within the supergroup Excavata, and include free-living species and parasites that cause many important human diseases, together with species afflicting animals and plants. Kinetoplastids exhibit varied life-styles, host range, distribution and specialisations over long evolutionary periods. Further, kinetoplastid surfaces are highly divergent between lineages, likely reflecting specialisations within membrane transport (Gadelha et al., 2015; Manna et al., 2013). More remarkable is the highly distinct nature of the proteins and glycoconjugates present at the cell surface, with recent data indicating that many surface proteins are restricted to the kinetoplastids (Adung'a et al., 2013; Field et al., 2007; Jackson 2016; Gadelha et al., 2015). All of these features can be anticipated as leaving an imprint in the evolutionary history of the trafficking system. To address how kinetoplastid trafficking pathways evolved, we examined the representation of Rabs, RabGAPs and SNAREs, across all currently available genome/transcriptome resources.

SNAREs, mediators of membrane fusion, possess a characteristic domain, and are classified as Q- or R-SNAREs (Bock et al., 2001; Fasshauer, 1998). Typically, three Q- (a, b or c) and one R-SNARE form a complex. The complement of Qa and R-SNAREs in the LECA has been established with broad taxonomic sampling (Arasaki et al., 2015; Vedovato et al., 2009), but the Qb and Qc families were only assessed by comparative genomics nearly a decade ago, using the limited sampling of genomes then available (Kloepper et al., 2007; Yoshizawa et al., 2006) and warrant re-examination to solidify definition of the LECA complement. It is key to have a robust estimate of the overall SNARE complement in the LECA as the starting point from which we assume kinetoplastids evolved. Rabs are small GTPases and well established markers of organellar identity (Brighouse et al., 2010; Pereira-Leal and Seabra, 2001). A broad repertoire of ~23 Rabs are predicted in the LECA (Elias et al., 2012). With the exception of metazoa and vascular plants, where great expansion is evident, Rab proteins have evolved mainly through small scale expansions and secondary losses, plus emergence of novel paralogs (Elias et al., 2012; Klöpper et al., 2012). Rabs are regulated by GTPase-activating proteins (GAPs), the majority of which possess a Tre-2/Bub2/Cdc16 (TBC) Rab-binding

domain. LECA possessed about ten TBC subtypes and subsequent expansions has included domain swapping (Gabernet-Castello et al., 2013). The greatest number of TBC innovations are in animals and fungi, although novel subclasses are present in a wide range of lineages. Detailed analysis of both Rab and SNARE repertoires in fungi indicates a simple set with minimal variations between different lineages, despite major diversity in morphology and transitions between single and multicellular forms, *albeit* with evidence for minimization in some species (Kienle et al., 2009b; Pereira-Leal, 2008).

Methods

Sequence data collection:

To identify SNARE sequences, a validated dataset of 26 predicted SNARE sequences obtained for *T. brucei* (Murungi et al., 2014) and 27 sequences from *L. major* (Besteiro et al., 2006) were used to query predicted proteomes for *Trypanosoma brucei brucei 927, T. b. gambiense, T. congolense, T. vivax, T. cruzi, T. grayi, Leishmania major, L. mexicana, L. braziliensis, L. infantum, Bodo saltans, Phytomonas serpens, Phytomonas EM1, Phytomonas HART1* (http://www.genedb.org/ or TriTrypDB) and the heterolobosid *Naegleria gruberi* (Fritz-Laylin et al., 2010) (http://genome.jgi.doe.gov/Naegr1) by BLAST. tBLASTp was used to search transcriptome data from *T. theileri, T. carassii, Trypanoplasma borrelli,* and *Euglena gracilis* (unpublished, SK and MCF). All sequences returned with e-values < 10^{-3} were retained. This dataset was parsed to remove redundancies using >99% sequence identity as criteria for exclusion. ClustalW (Thompson et al., 2002) was used to align each dataset and generate neighbour-joining (NJ) trees (Saitou and Nei, 1987). Sequences which were excluded or weakly clustering were validated for a SNARE domain *via* Interpro (Hunter et al., 2009, http://www.ebi.ac.uk/interpro) and SNARE-DB (Kloepper et al., 2007, http://bioinformatics.mpibpc.mpg.de/snare/index.jsp). Sequences incorporating a SNARE-domain (SNARE, t/v-SNARE, sec20, syntaxin, longin, synaptobrevin) were retained. Sequences in which no domain was detected but were between 70 and 500 residues were also retained.

A pfam Ras domain (PF00071) HMMSCAN (Eddy, 1998) (http://hmmer.org) search was conducted against 50 eukaryotic proteomes, including those listed above to identify Rabs. Sequences with e-values $<10^{-3}$ were retained and used to generate a NJ tree. Kinetoplastid Rab candidates were identified and tentatively assigned on the basis of association with defined Rab sequences. One sequence from each cluster was used as query to define the cluster by reciprocal best hit (rbh) BLAST and another round of rbhBLAST performed using sequences from the first round as queries. Rbh requires that potential positive hits retrieve the original query when used to search. The tree was annotated accordingly to verify initial assignments. Additionally, assignments were made using Rabifier (Diekmann et al., 2011). All collected kinetoplastid sequences were classed as either a tentative Rab subfamily member or as stray. Identical procedures were undertaken with the TBC domain (pfam PF00566) and HMMSCAN to recover sequences.

For analyses addressing the evolution of NPSN and Syp7, a broader eukaryotic sampling and separate phylogenetic methods were used. Homology searches were performed using *Homo sapiens, Saccharomyces cerevisiae* and *Arabidopsis thaliana* sequences; protein sequences from a broad sampling of eukaryotes were retained. Positive BLAST hits for each protein of interest were aligned using MUSCLE v3.8.31 (Edgar, 2004). Where very similar sequences were present from the same or closely related organisms, those that aligned least well were removed to limit overrepresentation of some taxonomic groups in the Hidden Markov Model (HMM). For each potential orthologue identified by HMMER (e-value 0.05), reverse BLAST searches were performed as before. Homology search results were summarized using Coulson Plot Generator (Field et al., 2013).

Phylogenetic reconstruction:

SNARE, Rab and RabGAP datasets were aligned using MAFFT (Katoh et al., 2005) using the E-INS-i strategy and manually edited in Jalview (Waterhouse et al., 2009). The alignment was used to generate maximum-likelihood trees in PhyML v3.0 (Guindon et al., 2010) using the LG model, number of substitution rates – 4/6, starting tree – BioNJ, tree topology search - NNI moves and statistics – aLRT SH-like and bootstrap (100/0 replicates). Bayesian inference was implemented in

MrBayes v3.2 on the CIPRES server (Miller et al., 2010; Ronquist and Huelsenbeck, 2003), generally with 8×10^6 MCMC generations where convergence was achieved, as measured by a splits frequency below 0.01. Substitution models employed for inferring trees were selected using ProtTest v3 (Abascal et al., 2005) for PhyML and the mixed model for MrBayes. One representative from each well supported clade (>0.9, 90, MrBayes, PhyML support respectively) along with a panel of known eukaryotic SNAREs, Rabs and TBC Rab-GAPs were analysed to determine orthology and define subfamilies. A SNARE panel was created from SNARE sequences from *Homo sapiens, Saccharomyces cerevisiae, Arabidopsis thaliana, Phytopthora sojae* and *Entamoeba histolytica* obtained from SNARE-DB*.*

For Qb and Qc, LECA complement datasets were assembled from the searches. Alignments generated for HMMs or phylogenetic analysis were constructed using MUSCLE v3.8.31 (Edgar, 2004) and edited using Mesquite (Maddison and Maddison, 2015). Maximum likelihood phylogenies were constructed using RAxML (Stamatakis, 2014) with the Protein GAMMA model for rate heterogeneity, and the LG4X substitution matrix. Values from 100 bootstraps were mapped onto the best of 20 topologies constructed from the original alignment. Bayesian phylogenetic analysis was performed using MrBayes, and run for 8-10 million MCMC generations with the Mixed substitution model, and the final average standard deviation of split frequencies decreased to less than 0.008. Several trees were constructed iteratively for each dataset, with removal of manually selected sequences to improve resolution.

Trypanosome cell culture:

Procyclic culture form (PCF) *T. b. brucei* Lister 427 were grown as previously described (Brun et al., 1979). Expression of plasmid constructs was maintained using antibiotic selection at the following concentrations: G_{418}/hygromycinB at 25µg/ml, blasticidin at 10µg/ml and puromycin at 2µg/ml.

Expression constructs:

Putative trypanosome SNAREs TbVAMP7C (Tb427.10.790), TbVAMP7A (Tb427.2.5120), TbVAMP7B (Tb427.5.3560), TbYkt6 (Tb927.9.14080) were

amplified from *T. b. brucei* 427 genomic DNA using Hercules DNA polymerase (Aligent Technologies).

For hemagglutinin (HA)-tag fusion, the PCR products containing sequence for a C-terminal HA-epitope were cloned into the PCF expression vector pLew79 (Wirtz et al., 1999) using AvrII and BamHI (TbVAMP7 isoforms) using the following primers (written 5' to 3'): TTGTGTCCTAGGATGCTTATATCTGCCTCCTT pLewVAMP7A_F, ACTCAAGGATCCTTAAGCGTAATCTGGAACATCGTATGGGTACTTTTTGCACTT GAGGTTAG pLewVAMP7A_R, TTGTGTCCTAGGATGCCCATTAAATATAGTTG pLewVAMP7B_F,

ACTCAAGGATCCTTAAGCGTAATCTGGAACATCGTATGGGTATGACTTGC AGTTGGAAAAGT pLewVAMP7B_R,

TTGTGTCCTAGGATGCAGGGAGGAACAAAAT pLewVAMP7C_F,

ACTCAAGGATCCTTAAGCGTAATCTGGAACATCGTATGGGTACTTCTTTT CCTCTTTTTTAC pLewVAMP7C_R

The PCR product of TbYkt6 was cloned into pHD1034 containing an N-terminal HA-epitope using HindIII and AflII using the following primers: GGCCAAGCTTTATACTCCCTGGCAAT pHD1034Ykt6F, CCGTCTTAAGTCACATGACGGTGCAACA pHD1034Ykt6R

Putative SNARE interactors TbSyx16B (Tb427tmp.211.3920), TbSynE (Tb427.03.5570), TbVti1-likeA (Tb427.8.3470), TbVti1-likeB (Tb427.08.1120), TbSyx6-like1 (Tb427.10.1830) and TbSyx8-like (Tb427.10.2340) were also similarly amplified. For 6x cMyc tagging, the product of each gene was cloned into pRPΔOP (with thanks to Dr. Lucy Glover, Institut Pasteur) containing 6x cMyc-epitope using HindIII and XbaI using the following primers:

GCGCGCAAGCTTATGAGCGGGGACGGCGTTGG pRPCTb427.8.1120F,
GCGCGCTCTAGAAACTTTCCCCAGAAACTTCC pRPCTb427.8.1120R,
GCGCGCAAGCTTATGGACGATCCAAGTTGGCA pRPCTb427.3.5570F,
GCGCGCTCTAGATACTTTATGGTACGCAACGA pRPCTb427.3.5570R,
GCGCGCAAGCTTATGTCGTCTCTGCAAGATCC pRPCTb427.10.1830F,
GCGCGCTCTAGAACTAAAGACACAATAGAAGA pRPCTb427.10.1830R,
GCGCGCAAGCTTATGTCTAAACAAGAA 2F_PRP_Tb427.10.2340_C,
GCGCGCTCTAGAAAGTATTAAAAGCAC 2R_PRP_Tb427.10.2340_C,

GCGCGCAAGCTTATGTCATCTGATCTT 3F_PRP_Tb427.08.3470_C,

GCGCGCTCTAGACTTCCAAAATACAAT 3R_PRP_Tb427.08.3470_C,

GCGCGCAAGCTTATGGCGACCCGTGAC 4F_PRP_Tb427.211.3920_C,

GCGCGCTCTAGAAGACAGCATCTTTTG 4R_PRP_Tb427.211.3920_C,

Putative interactors TbVps45 (Tb427.10.6780) and TbSly1 (Tb427tmp.160.0680) were cloned into pMOT vector (Oberholzer et al., 2006) with 3xV5 tag using the following primers: AGGTCCTGTGCACGCCTGCATCGGTGGGACTGGAGTCCTTAACAGTGAAACCT TCCTGAGCCTGCTAGCAGCGCACGCAGGTACCGGGCCCCCCCTCGAG VPS45pMOT_F,

GTATTTTGGTTTCGTTTATTCATACCACCATGCGGAGGCGCAATGTCCCC GCCAAAACAGGCGAGGGCGGCACATGGCGGCCGCTCTAGAACTAGTGGAT VPS45_pMOT_R,

GGTTAGTTATGGCTGTACCGCAATGCTGACGGGGAATGAAGCACTGCGC CAGCTTACTGTTCTTGGTGAAGGAATATCAGGTACCGGGCCCCCCCTCGAG Sly1_pMOT_F,

AAAGCACGTTAGGATAGTATCTGAAAGTGGGAAAACGCCAAATGGCACA AAGACCAAAACGGCCGGGCCGGTGCTGGCGGCCGCTCTAGAACTAGTGGAT Sly1_pMOT_R

All constructs were verified by sequencing and linearized with NotI prior to transfection into cells. Clonal transformants were selected by resistance to antibiotics as relevant to each vector and cell line.

Transfection of PCF T. brucei:

1.6×10^7 cells per transfection were harvested at 4°C, washed in cytomix and resuspended in 500µl cytomix. Electroporation was performed with 5-15 µg of linearized DNA using a Bio-Rad Gene Pulser II (1.5 kV and 25 µF). Cells were transferred to 9.5ml SDM-79 medium and incubated for 6 hours after which selection antibiotics were added. The cells were then diluted into 96-well microtiter plates. Positive transformants were picked into fresh selective medium 10-15 days post-transfection.

Identification of protein-protein interactions:

Interactions between VAMP7C (Tb427.10.790) and other trypanosome proteins were analyzed by immunoisolation. 5×10^{10} procyclic cells habouring VAMP7C tagged at the C-terminus with HA were lysed by mechanical milling in a Retsch Planetary Ball Mill PM200 using liquid nitrogen cooling (Retsch, United Kingdom). Aliquots of powder were thawed in solubilization buffer (50mM Tris-HCL pH8.0, NaCl 100mM, 1% Triton-X100 or CHAPS + 1mM NEM, and 50mM Tris-HCL pH8.0, 100mM NaCl, 5mM EDTA, 1% Triton-X100 and/or 0.5% Triton-X114 + 1mM NEM). VAMP7C::HA was isolated using Pierce anti-HA magnetic beads. All washes were in the same buffer without NEM. Following analysis of an aliquot by SDS-PAGE, affinity-isolated proteins were precipitated in 90% ethanol (v/v). The precipitated samples were trypsinised and analysed by LS-MS^2 mass spectrometry at the University of Dundee Fingerprints Proteomics Service. Peak lists were submitted to ProFound and searched against an in-house *T. brucei* database using data from GeneDB (www.genedb.org). An untagged wildtype cell line treated identically served as background control.

Immunofluorescence microscopy:

Cells were prepared as previously described (Leung et al., 2008). Antibodies were used at the following concentrations: rat anti-HA (Roche) 1:1000; mouse anti-cMyc (SantaCruz Biotech 9E10) 1:500; mouse anti-p67 (James Bangs, University of Wisconsin-Madison) 1:1000; rabbit anti-GRASP (Graham Warren, Vienna, Austria) 1:500 in 20% FCS in PBS (v/v). Wide-field epifluorescence images were acquired using a Nikon Eclipse E600 epifluorescence microscope equipped with a Hamamatsu ORCA CCD camera, and data captured using MetaMorph (Universal Imaging, Marlow, UK).

Western immunoblotting:

Whole cell lysates in SDS-PAGE sample buffer containing 10^7 cells/lane were resolved by SDS-PAGE. Proteins were transferred to polyvinylidene fluoride membranes (Millipore) and blocked using 5% semi-skimmed milk. Antigens were

visualised using standard methods. Primary antibody concentrations were: rat anti-HA (Roche) at 1:10000, mouse anti-cMyc 1:5000. Primary antibody binding was detected using anti-IgG horseradish peroxidase (HRP) conjugates (Sigma-Aldrich) at 1:10000. Detection of HRP-conjugated secondary antibody was by chemiluminescence and luminol.

Results

Confirmation of the LECA complement of Qb and Qc SNAREs:

Whilst recent analyses have addressed most protein families and subfamilies considered here, Qb and Qc SNAREs have not been characterised in a comparable way. In particular, there are conflicting views of the distributions of Qb-SNARE Novel Plant Syntaxin (NPSN) and the Qc-SNARE Syntaxin of Plants 7 (Syp7) (Sanderfoot et al., 2000). As their names imply, both were proposed as plant specific, but other studies suggest orthologues in several protist lineages, including *Trypanosoma* and *Dictyostelium* (Kloepper et al., 2007; Sanderfoot, 2007; Yoshizawa et al., 2006), which prompted some authors to suggest that NPSN, at least, was present in the LECA (Kienle et al., 2009a). To determine if NPSN and Syp7 orthologs are present in non-plant eukaryotes, we performed comparative genomics and phylogenetic analysis. Searches returned NSPN and Syp7 candidate orthologs from the genomes of most eukaryotes except animals and non-basal fungi (Figure 1). A Qb-SNARE phylogeny strongly supports the identity of these putative NPSN orthologues and similarly the Qc-SNARE tree strongly supports identity of Syp7 orthologues (Figure S2F and G). This analysis suggests that both NPSN and Syp7 were present in the LECA and later lost from animals and fungi.

The kinetoplastid complement of trafficking proteins:

We next considered the genomes of eighteen Euglenids including basal bodonids and trypanosomatids. We identified candidate genes for 382 Rabs, 552 SNAREs and 307 TBC Rab-GAPs. Where possible these were assigned to subfamilies by phylogenetic analyses with a reference set of previously defined

sequences from across eukaryotes. Such data are readily available for Rabs and TBC Rab-GAPs (Elias et al., 2012; Gabernet-Castello et al., 2013) and were prepared here for SNAREs using SNARE repertoires from at least one representative of each eukaryotic subgroup, other than the Excavata; *Homo sapiens* and *Saccharomyces cerevisiae* as Opisthokonta*, Dictyostelium discoideum* as Amoebozoa*, Phytopthora sojae* as SAR+CCTH and *Arabidopsis thaliana* for the Archaeplastida. We were able to assign a considerable proportion of the kinetoplastid representatives of all three families (Figures 2, 3, 4, S2). Moreover, 86% of LECA SNAREs, 72% of LECA Rabs and 100% of LECA TBC RabGAPs were identified in one or more members of the lineage, suggesting that the core machinery of trafficking is well conserved (Table 1). Most kinetoplastid proteins were assigned by phylogenetic evidence, but three SNARE proteins, Vti1-like, Syx6-like and Syx8-like, were assigned only by BLAST and reverse BLAST as phylogenetic support was low.

Comparing coding content and trafficking repertoire:

There is loose correspondence between genome and Rab repertoire sizes (Gabernet-Castello et al., 2013). While this may reflect a connection between genome and compartmental complexity, the huge diversity within eukaryotes makes complexity a difficult parameter to define, and analysing a group of organisms where the basic cellular bauplan is well conserved, allows us to re-evaluate this concept (Figure S1). The coding content of *B. saltans,* is considerably larger than the parasites, while the excavates *N. gruberi* and *E. gracilis* have larger genomes (Ebenezer et al., 2017; Fritz-Laylin et al., 2010). In keeping with genome size, *B. saltans, N. gruberi* and *E. gracilis* possess greater numbers of SNAREs, Rabs and TBCs compared to kinetoplastids. Despite comparable coding content for the brucei and cruzi group trypanosomes, the latter have larger Rab, TBC and SNARE repertoires, suggesting further secondary loss in the former. Even *Leishmania* spp. generally have slightly larger numbers of these proteins, despite smaller predicted coding content. Overall, this suggests adaptive shaping of trafficking system gene repertoires rather than simply reflecting genome size.

Anterograde trafficking genes are conserved across kinetoplastids:

The gene complements for SNAREs and Rabs at the kinetoplastid early secretory pathway are highly conserved. In opisthokonts, the SNARE complex: Qa-Syx5, Qb-Bos1, Qc-Bet1 and R-Sec22 mediates ER to Golgi and intra-Golgi transport (Hardwick and Pelham, 1992; Newman et al., 1990). A second Syx5 complex, constituting Qa-Syx5, Qb-Gos1, Qc-Sft1 and R-Ykt6 is thought to exclusively mediate COPI transport within the Golgi in animals and fungi (Ballensiefen et al., 1998). Whilst Sft1 and Bet1 are two alternate Qc SNAREs acting at early anterograde steps (Kloepper et al., 2007), they were unresolved in our phylogenies (Figure S2B). It is likely that Bet1 fulfils the requirements of the kinetoplastid early secretory pathway as two Bet1 paralogs are present, similarly to *H. sapiens*. Significantly *A. thaliana* has an expanded set of four (designated as Bet11, Bet12, Sft11 and Sft12) and *D. discoidium* just one. The other SNAREs in these complexes are all singletons in kinetoplastids bar some species-specific losses (Figure 5A) and duplication of Gos1 in *T. congolense*. Syntaxin 17 cycles between the ER-ERGIC in mammalian cells (Itakura and Mizushima, 2013) and acts in autophagosome formation and mitochondrial dynamics (Arasaki et al., 2015; Itakura et al., 2012). Syntaxin 17 is generally patchily distributed across eukaryotes and absent from all kinetoplastids, signifying a loss since divergence from the Heterolobosea. Overall, these data suggest possible differentiation in early secretory events, consistent with multiple budding pathways at the trypanosome ER, but also indicating lineage-specific evolution of these steps (Sevova and Bangs, 2009).

The retrograde pathway from the Golgi to the ER is mediated by a complex of Qa-Syx18, Qb-Sec20, Qc-Use1, and R-Sec22 in yeast (Dilcher et al., 2003). Except Use1, all are conserved across kinetoplastids, bar losses of Syx18 or Sec20 in *Phytomonas EM1* and *HART1* respectively, and duplication of Sec22 in *T. theileri* (Figure 5A). This suggests that canonical retrograde transport systems are present, consistent with the presence of a KDEL receptor and ER-retrieval signals on major ER proteins (Bangs et al., 1996; Schwartz et al., 2013).

In opisthokonts the late exocytic SNARE complex is a specific ternary complex composed of SynPM-SNAP25-Syb1. SynPM, representing plasma membrane syntaxins, is identified across eukaryotes and likely ancient (Dacks and

Doolittle, 2002). SynPM has a complex history in kinetoplastids, with two clear paralogs, SynPM1 and SynPM2. *L. braziliensis* and *L. major* possess two copies of SynPM1 which are closely related lineage-specific duplications. The bodonid *T. borreli* and cruzi group possess a second SynPM2 protein. No *trans*-membrane domain (TM) was detected in SynPM2 of the duplicated SynPM1 genes in *L. braziliensis* and *L. major,* which are otherwise identical, but for the lack of a TM in one paralog. The TM is also lost from the single SynPM1 in *L. infantum* and *L. mexicana* but retained by *L. donovani.* The localization of *T. brucei* SynPM1 (TbSynPM1) is similar to the TM-containing SynPM1 from *L. major,* which localises close to the flagellar pocket, while as expected the TM-lacking SynPM1 in *L. major* is cytosolic (Besteiro et al., 2006). Overall, this pattern indicates a considerable level of species-specific SynPM paralogs, suggesting a common requirement to differentiate this pathway.

Putative Qbc-type SNAREs are present in *B. saltans*, *T. cruzi* and *T. grayi*, but do not cluster with canonical yeast or human Qbc SNAREs, but instead with SAR-CCTH and Excavate sequences (Figure S2D). The eukaryotic phylogeny of synaptobrevins (Figure S2E) is also complex, and the presence of brevins in different clusters of diverse taxa indicates that they likely emerged independently in several lineages by loss of the longin domain from a 'VAMP7'. Moreover, unlike metazoan brevins, which are truncated at the SNARE domain, longin-less VAMP7 orthologs have an extended N-terminal domain in Apicomplexa and Euglenozoa, indicating a separate origin. A single independently evolved VAMP7 with a divergent/undetected longin domain is present in kinetoplastids, and in *T. brucei* this protein TbVAMP7C has an endosomal localization in juxtaposition to Golgi and lysosomal markers (Figure 6C) rather than the cell membrane.

Rab 1, 8 and 18 are present in the predicted LECA cohort and associated with anterograde pathways (Elias et al., 2012). Interestingly, Rab8 is lost from all kinetoplastids, suggesting differences in post-Golgi trafficking and targeting of material to the flagellum/recycling endosome (Huber et al., 1993; Klöpper et al., 2012). Rab1/18 have a complex history, and paralogous expansion of Rab1 or 18 may have given rise to two lineage-specific Rabs, KSRX1 and UzRX3, which appear across the Euglenids as tandem genes and are also present in the heterolobosids,

but it may be significant that in *T. brucei*, these proteins have no obvious role in trafficking (Natesan et al., 2009). While UzRX3 localisation is similar to Rab1 (early secretory pathway), KSRX1 has a more diffuse localisation. Interestingly, *B. saltans* has two additional relatives of the Rab1/18 clade, suggesting further diversification of these pathways.

Post-Golgi pathway Rabs include Rab4, 14 and 11 and largely well-conserved, with the exception that Rab14 is lost by the brucei group (Figure 5B). This suggests possible further adaptations of endocytic pathways, an important mechanism for immune evasion in mammalian infective African trypanosomes. Like AP-2 however (Manna et al., 2013), Rab14 is present in the other extracellular trypanosomes of the cruzi group (Figure 5B, S20). We also identified a Rab11-like protein (Rab11B) which likely arose from duplication of Rab11 at the base of the kinetoplastids, but this is only retained by *B. saltans*, *Leishmania* spp. and cruzi group (Figure 5B, S20). Rab11B may provide additional flexibility in surface remodeling, and significantly *T. cruzi* Rab11 localizes to the contractive vacuole, regulating trafficking of *trans*-sialidase to the plasma membrane (Niyogi et al., 2014). Hence kinetoplastid Rab11 and 11B may have distinct functions, as noted for Rab11 paralogs in mammals and Archaeplastida (Lapierre et al., 2003; Petrželková and Eliáš, 2014).

Endocytic, retrograde and lysosomal pathway trafficking:

We identified four Qa SNARE proteins (SynE1 and 2, Syx16A and B), four Qb SNARE (Vti1-like A, B and Npsn A, B), four Qc SNAREs (Syp7A, B, Syx6-like and Syx8-like), and four R-SNAREs (VAMP7 A, B, C and D) as orthologs of SNARE proteins involved in endosomal, lysosomal and retrograde trafficking pathways. For several, multiple paralogs exist, particularly in the basal bodonids and the cruzi group (Figure 5A, S3A-D).

Except *B. saltans* and *T. vivax*, SynE1 is present in all kinetoplastids and possibly duplicated in the parasitic bodonid *T. borreli*. SynE2 is also present in all kinetoplastids and duplicated in the cruzi group. Two paralogs of Syx16, Syx16A and

B are conserved across kinetoplastids. Thus, both SynE and Syn16 underwent duplication at least once early in the lineage evolution. As in the SynPM proteins, all Syx16A proteins lack a TMD, while all Syx16B except *T. borreli* retain the TMD. *L. braziliensis* also shows lineage-specific duplication of Syx16A, a surprisingly conserved isoform, which suggests that it must retain a function, and is perhaps a regulatory protein. Paralogous Syx16 sequences from other organisms all retain their TMD, suggesting loss of the TMD is kinetoplastid-restricted. Notably, the Sec1/Munc-like protein (SM) Vps45, which regulates Syn16, is also duplicated at the base of the kinetoplastids (Koumandou et al. 2007), and possibly in the LKCA (Last Kinetoplastid Common Ancestor), suggesting independent regulation of multiple Syn16 complexes.

The Qb (2x Vti1, 2x Npsn), Qc (2x Syp7, 2x Syx6-like) and R-SNAREs (4x VAMP7) are expanded. Specifically, Qb-NpsnA-2, Qc-Syp7B, R-VAMP7D together with Qa-SynPM2 appear to have arisen at the base of the kinetoplastids but retained only in bodo and cruzi groups (Figure 5A). These SNAREs are implicated in plasma membrane trafficking in plants, suggesting they may mediate a similar pathway on the kinetoplastida. In addition, the Syp7B clade also shows evidence for bodo and cruzi group-specific expansions.

Most SNARE losses are scattered with no obvious pattern, but there are clear instances of co-evolutionary loss of components of predicted SNARE complexes. SNAREs Qa-SynPM2, Qbc-SNAP25 and R-VAMP7D, which could form a putative exocytic complex are all lost. Qb-Vti1-like A2, Qb-NpsnA2, and Qc-Syx6-like2, predicted to form complexes with VAMP7, are cioincidentally lost (see Figure 5A). Except SNAP-25 (which shows a broader pattern of loss), these are all SNAREs derived from kinetoplastid-specific expansions, and likely indicates loss of kinetoplastid-specific post-Golgi pathways.

Of the endocytic Rabs (Rab 5, 20/24, 21, 22 and 50), kinetoplastids retain Rab5, 21 and 24, *albeit* with Rab24 bodonid restricted (Figure 5). Early endocytic Rab5 experienced a single basal duplication in kinetoplastids and both paralogs stably retained (Figure S3E), and acquired distinct functions (Pal et al., 2002). This contrasts with the fungal Ypt5, which is less well retained, with several instances of species or clade-specific expansions and losses (Pereira-Leal, 2008). Rab21 is

generally stable with no major expansion in eukaryotes, but there are at least two paralogs at the base of the kinetoplastids (Figure S3E); Rab21 mediates intermediate endosomal trafficking in *T. brucei,* a conserved function with higher eukaryotes (Ali et al., 2014). An expanded Rab21 subfamily is only retained by bodonids and the cruzi group, suggesting additional endosomal pathways. In *Phytomonas* spp., which have lost Rab21, there is clear minimisation. Rab2 and 6, which mainly mediate retrograde trafficking at the Golgi complex, are also stably retained. Two copies of Rab2 are present in the bodonids and *T. grayi*, but cluster separately from the canonical Rab2.

Rab7, which mediates late endocytic trafficking, is represented by a single paralog in trypanosomatids; however, a Rab7-like protein is found in *B. saltans* and *T. borreli* (Figure 5). Rab28 also has an endosomal function in *T. brucei* (Lumb et al., 2011) and except Phytomonads is present across all kinetoplastids. Rab23 and Rab-like IFT27 (intraflagellar transport protein 27), which are involved in the biogenesis of cilia/flagella are both fully retained as expected. Rab32, is involved in the biogenesis of lysosome-related organelles (LROs) autophagosome formation in mammalian cells (Hirota and Tanaka, 2009), and mitochondrial dynamics (Bui et al., 2010; Ortiz-Sandoval et al., 2014). In *T. cruzi* Rab32 is involved in the biogenesis and maintenance of acidocalcisomes (Niyogi et al., 2015), and presents complex evolution in kinetoplastids, present only in the bodonids and cruzi group, where one copy of the canonical Rab32 has undergone (Figure 5B). Many trypanosomes may thus have reduced or alternative pathways for autophagosome biogenesis and/or mitochondrial dynamics, but by contrast, the emergence of Rab32-like proteins likely suggests the development of novel or more complex LROs, particularly in *B. saltans* with up to five Rab32-like proteins.

TBCs associated with endosomal Rabs include TBC-B, which acts on Rab2, 7, 11, and 21 and TBC-D, a GAP for Rab 1, 7, 11 and RabL5 (Fukuda, 2011; Gabernet-Castello et al., 2013). This latter TBC is implicated in recycling of VAMP7 vesicles and endosome to Golgi transport. TBC-B is retained in all kinetoplastids as a single copy but lost in the phytomonads. TBC-D is stably retained across all eukaryotes with occasional duplications but has multiple paralogs in kinetoplastids (Figure 5B). Other TBCs exhibit lineage-specific duplications or losses but there is

no obvious coevolutionary relationship with the Rabs. For example, TBC-F is duplicated in bodonids, four of the five *Leishmania spp.* have duplicated TBC-I (*L. donovani* is missing TBC-I), and *T. grayi* and *T. theileri* have two copies of TBC-RootA. Major TBC losses are present in *Phytomonas spp.,* and all lack TBC-B and E, *P. serpens* and *P.* HART1 also lack TBC-G, H, I and L, while *P. serpens* is specifically missing TBC-M. These Phytomonad TBC families are small when considered relative to the modest Rab complement.

Orphan and unassigned SNAREs, Rabs and TBCs:

Several sequences could not be unambiguously assigned. Examples include lineage-specific KSRabX4, likely a result of a *T. brucei*-specific duplication: only *T. brucei* and *T. gambiense* have two neighbouring copies with ~52% identity, while remaining species possess one paralog (Figure S3E). KSRabX6 is only found in the bodonids and cruzi group and BLAST searches into opistokhonts suggest similarity to Rab5 (Figure 5B, S3E). Several other unassigned Rabs are also present in *B. saltans* and *T. borreli.*

Among SNAREs, a tomosyn-like regulatory R-SNARE is present (Figure S2E) and exhibits patchy distribution, while an unconventional Longin-domain protein with no apparent SNARE domain is patchily distributed. Its structure is reminiscent of plant phytolongins, which are derived from VAMP7 (Vedovato et al., 2009) but the kinetoplastid protein appears closely related to another R-SNARE, Sec22 (Figure S2E). Much of the expansion of lineage-specific paralogs is due to duplication at the base of the lineage and asymmetrical retention, highly suggestive of a genome duplication event at the kinetoplastid root.

Conserved localisation of R-SNAREs:

We are, in essence, using phylogeny and orthology as a predictor of function, which over 10^9 years' divergence between kinetoplastids and opisthokonts, is not necessarily valid. For example, associations between orthologous SNAREs within specific complexes could have changed over such a great period. To address this directly, we tagged VAMP7A, B and C at the C-terminus and Ykt6 at the N-terminus to facilitate overexpression in procyclic *T. brucei* (Figure 6A, B, C and D). All three

VAMP7s are likely largely endosomal. VAMP7A and C are juxtaposed to the lysosome with VAMP7A showing occasional overlap and VAMP7B slightly more distal from the lysosome. Both VAMP7B and C are juxtaposed to the Golgi and appear associated with the Golgi during its duplication (seen as two GRASP-stained puncta during cell division). As material building the new Golgi is derived from the old Golgi stack (Wang and Seemann, 2011), VAMP7B and C may play a role in this transfer given their presence with both the Golgi stacks. The Golgi-SNARE Ykt6 does co-localise with the Golgi. These localisations are consistent with the behaviour of the yeast and human orthologs.

Conservation of VAMP7C interactions:

To examine VAMP7C interactions we used immunoisolation and mass spectrometry. In opisthokonts, VAMP7 forms two complexes, with SynE, Syx8 and Vti1B to mediate lysosomal transport and with Syx6, Syx16 and Vti1A for endosomal trafficking (Jahn and Scheller, 2006). We consistently identified Vti1-likeA and B, Syx8-like, Syx6-like1, Syx16B and SynE (Table 2), a complex composition that is also consistent with localisation (Figure 6E). Opisthokont VAMP7 is also implicated in plasma membrane trafficking, with Qa-SynPM and Qbc SNARE SNAP-25/23 (Jahn and Scheller, 2006) and in plants in complex with Qa-SynPM, Qb-Npsn and Qc-Syp7 (Suwastika et al., 2008; Zheng et al., 2002). We reproducibly recovered TbNpsnA as a candidate VAMP7C interactor, *albeit* with poor emPAI support, but not TbSyp7, suggesting that a different *T. brucei* VAMP7 (TbVAMP7A or B) is required or that the SNARE complex composition for plasma membrane targeting is divergent. Therefore, endosomal interactions between VAMP7 and partners are apparently conserved with the opisthokont SNARE complexes.

Discussion

The kinetoplastids encompass many parasitic species with a particularly broad range of niches and life cycles. Surface components are critical to their success which is partly reflected within the trafficking system, and our analysis here provides a comprehensive view of these processes. The repertoire of Rabs, Rab-GAPs and SNAREs at the LCKA resembles that predicted for LECA, and the free-

living *B. saltans* possesses the largest repertoire, with several kinetoplastid-specific members likely originated in the LCKA. Significantly, we also find that the frequency of paralogous pairs is high, which may indicate a whole genome duplication event. Rather than a precipitous decrease as kinetoplastids transitioned from free living to parasitic forms, there is gradual loss of trafficking genes, and likely corresponding pathway simplification (Jackson et al., 2015). Whilst some LECA proteins, e.g. SNAP-25, Rab24, 32, TBC-N are lost, a large proportion of losses are of lineage-specific duplications, originating in the common excavate or LCKA. Hence radical intracellular remodelling did not accompany the transition to parasitism. The extensive Bodonid repertoire is mainly due to retention of the LECA/LKCA gene complement, and the most prominent expansions appear as Syp7, and Rab32.

The putative common ancestor of African trypanosomes is the single instance of coordinated loss of several SNARE complex subunits, specifically post-Golgi SNAREs and is accompanied by the loss of Rabs predicted to act in similar pathways; phagocytic Rab14 and putative recycling Rab11B. *Leishmania* and trypanosomes have both lost the endosomal Rab32, 21B and 21C, SNAREs Npsn1 and SNAP25, and TBC-ExA and N. The cruzi group exhibit the greatest degree of inter-lineage variation, perhaps reflecting the varied lifestyles and disparate hosts/vectors that this group enjoys. Moreover, bodonids and the cruzi group have several clade and species-specific gains/losses indicating highly dynamic shaping of trafficking genes in these organisms.

Author contributions

DV, LDB, JDB and MCF conceived and designed the study. DV, LDB and NNN performed the experiments and analyses, and generated figures and tables. SK provided data for *T. theileri, T. borreli* and *T. carassi.* AOR performed initial sequence collection. DV, LDB, NNN, JDB and MCF wrote the manuscript, JDB, DV and MCF edited the manuscript.

Acknowledgements

Journal of Cell Science • Advance article

References

Abascal, F., Zardoya, R. and Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105.

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.

Adung'a, V. O., Gadelha, C. and Field, M. C. (2013). Proteomic Analysis of Clathrin Interactions in Trypanosomes Reveals Dynamic Evolution of Endocytosis. *Traffic* **14**, 440–57.

Ali, M., Leung, K. F. and Field, M. C. (2014). The ancient small GTPase Rab21 functions in intermediate endocytic steps in trypanosomes. *Eukaryot. Cell* **13**, 304–19.

Arasaki, K., Shimizu, H., Mogari, H., Nishida, N., Hirota, N., Furuno, A., Kudo, Y., Baba, M., Baba, N., Cheng, J., et al. (2015). A Role for the Ancient SNARE Syntaxin 17 in Regulating Mitochondrial Division. *Dev. Cell* **32**, 304–317.

Ballensiefen, W., Ossipov, D. and Schmitt, H. (1998). Recycling of the yeast v-SNARE Sec22p involves COPI-proteins and the ER transmembrane proteins Ufe1p and Sec20p. *J. Cell Sci.* **111**, 1507–1520.

Bangs, J. D., Brouch, E. M., Ransom, D. M. and Roggy, J. L. (1996). A soluble secretory reporter system in Trypanosoma brucei. Studies on endoplasmic reticulum targeting. *J. Biol. Chem.* **271**, 18387–93.

Besteiro, S., Coombs, G. H. and Mottram, J. C. (2006). The SNARE protein family of Leishmania major. *BMC Genomics* **7**, 250.

Birault, V., Solari, R., Hanrahan, J. and Thomas, D. Y. (2013). Correctors of the basic trafficking defect of the mutant F508del-CFTR that causes cystic fibrosis. *Curr. Opin. Chem. Biol.* **17**, 353–60.

Bock, J. B., Matern, H. T., Peden, A., and Scheller, R. H. (2001). A genomic perspective on membrane compartment organization. *Nature* **409**, 839–41.

Bonifacino, J. (2004). The Mechanisms of Vesicle Budding and Fusion. *Cell* **116**, 153–166.

Brighouse, A., Dacks, J. B. and Field, M. C. (2010). Rab protein evolution and the

history of the eukaryotic endomembrane system. *Cell. Mol. Life Sci.* **67**, 3449–65.

Brun, R., Jenni, L., Tanner, M., Schönenberger, M. and Schell, K. F. (1979). Cultivation of vertebrate infective forms derived from metacyclic forms of pleomorphic Trypanosoma brucei stocks. Short communication. *Acta Trop.* **36**, 387–90.

Bui, M., Gilady, S. Y., Fitzsimmons, R. E. B., Benson, M. D., Lynes, E. M., Gesson, K., Alto, N. M., Strack, S., Scott, J. D. and Simmen, T. (2010). Rab32 modulates apoptosis onset and mitochondria-associated membrane (MAM) properties. *J. Biol. Chem.* **285**, 31590–602.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Dacks, J. B. and Doolittle, W. F. (2002). Novel syntaxin gene sequences from Giardia, Trypanosoma and algae: implications for the ancient evolution of the eukaryotic endomembrane system. *J. Cell Sci.* **115**, 1635–1642.

Dacks, J.B., Field, M.C., Buick, R., Eme, L., Gribaldo, S., Roger, A.J., Brochier-Armanet, C., and Devos, D.P. (2016) The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *J Cell Sci.* **129,** 3695-3703.

de Duve, C. (2007). The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* **8**, 395–403.

Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A. and Rout, M. P. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e80.

Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B. T., Rout, M. P. and Sali, A. (2006). Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2172–2177.

Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M. C. and Pereira-Leal, J. B. (2011). Thousands of Rab GTPases for the cell biologist. *PLoS Comput. Biol.* **7**, e1002217.

Dilcher, M., Veith, B., Chidambaram, S., Hartmann, E., Schmitt, H. D. and

Fischer von Mollard, G. (2003). Use1p is a yeast SNARE protein required for retrograde traffic to the ER. *EMBO J.* **22**, 3664–74.

Dokudovskaya, S., Williams, R., Devos, D., Sali, A., Chait, B. T. and Rout, M. P. (2006). Protease Accessibility Laddering: A Proteomic Tool for Probing Protein Structure. *Structure* **14**, 653–660.

Ebenezer, T.E., Carrington, M., Lebert, M., Kelly, S., and Field, M. C. (2017). The genome of Euglena gracilis. In *Biochemistry, Cell and Molecular Biology of Euglena* (ed. Shigeoka, S., Schwartzbach, S. D.), p. Springer International.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* **14**, 755–763.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Elias, M., Brighouse, A., Gabernet-Castello, C., Field, M. C. and Dacks, J. B. (2012). Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. *J. Cell Sci.* **125**, 2500–2508.

Fasshauer, D. (1998). Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. *Proc. Natl. Acad. Sci.* **95**, 15781–15786.

Field, M. C. and Dacks, J. B. (2009). First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr. Opin. Cell Biol.* **21**, 4–13.

Field, M. C., Gabernet-Castello, C. and Dacks, J. B. (2007). Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv. Exp. Med. Biol.* **607**, 84–96.

Field, M. C., Sali, A. and Rout, M. P. (2011). Evolution: On a bender--BARs, ESCRTs, COPs, and finally getting your coat. *J. Cell Biol.* **193**, 963–72.

Field, H. I., Coulson, R. M. R. and Field, M. C. (2013). An automated graphics tool for comparative genomics: the Coulson plot generator. *BMC Bioinformatics* **14**, 141.

Fritz-Laylin, L. K., Prochnik, S. E., Ginger, M. L., Dacks, J. B., Carpenter, M. L., Field, M. C., Kuo, A., Paredez, A., Chapman, J., Pham, J., et al. (2010). The Genome of Naegleria gruberi Illuminates Early Eukaryotic Versatility. *Cell* **140**, 631–642.

Fukuda, M. (2011). TBC proteins: GAPs for mammalian small GTPase Rab? *Biosci. Rep.* **31**, 159–168.

Gabernet-Castello, C., O'Reilly, A. J., Dacks, J. B. and Field, M. C. (2013). Evolution of Tre-2/Bub2/Cdc16 (TBC) Rab GTPase-activating proteins. *Mol. Biol. Cell* **24**, 1574–83.

Gadelha, C., Zhang, W., Chamberlain, J. W., Chait, B. T., Wickstead, B. and Field, M. C. (2015). Architecture of a host-parasite interface: complex targeting mechanisms revealed through proteomics. *Mol. Cell. Proteomics* 1911–1926.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.

Gurkan, C., Koulov, A. V and Balch, W. E. (2007). An evolutionary perspective on eukaryotic membrane trafficking. *Adv. Exp. Med. Biol.* **607**, 73–83.

Hardwick, K. G. and Pelham, H. R. (1992). SED5 encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex. *J. Cell Biol.* **119**, 513–21.

He, C. Y., Ho, H. H., Malsam, J., Chalouni, C., West, C. M., Ullu, E., Toomre, D. and Warren, G. (2004). Golgi duplication in Trypanosoma brucei. *J. Cell Biol.* **165**, 313–21.

Hirota, Y. and Tanaka, Y. (2009). A small GTPase, human Rab32, is required for the formation of autophagic vacuoles under basal conditions. *Cell. Mol. Life Sci.* **66**, 2913–32.

Hirumi, H. and Hirumi, K. (1994). Axenic culture of African trypanosome bloodstream forms. *Parasitol. Today* **10**, 80–4.

Huber, L. a., Pimplikar, S., Parton, R. G., Virta, H., Zerial, M. and Simons, K. (1993). Rab8, a small GTPase involved in vesicular traffic between the TGN and the basolateral plasma membrane. *J. Cell Biol.* **123**, 35–45.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37**,.

Itakura, E. and Mizushima, N. (2013). Syntaxin 17: The autophagosomal SNARE. *Autophagy* **9**, 917–919.

Itakura, E., Kishi-Itakura, C. and Mizushima, N. (2012). The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. *Cell* **151**, 1256–69.

Jackson, A.P. (2016). Gene family phylogeny and the evolution of parasite cell surfaces. *Mol Biochem Parasitol.* **209**, 64-75.

Jahn, R. and Scheller, R. H. (2006). SNAREs--engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**, 631–643.

Katoh, K., Kuma, K., Toh, H. and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858.

Kienle, N., Kloepper, T. H. and Fasshauer, D. (2009a). Phylogeny of the SNARE vesicle fusion machinery yields insights into the conservation of the secretory pathway in fungi. *BMC Evol. Biol.* **9**, 19.

Kienle, N., Kloepper, T. H. and Fasshauer, D. (2009b). Differences in the SNARE evolution of fungi and metazoa. *Biochem. Soc. Trans.* **37**, 787–91.

Kloepper, T. H., Kienle, C. N. and Fasshauer, D. (2007). An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Mol. Biol. Cell* **18**, 3463–71.

Klöpper, T. H., Kienle, N., Fasshauer, D. and Munro, S. (2012). Untangling the evolution of Rab G proteins: implications of a comprehensive genomic analysis. *BMC Biol.* **10**, 71.

Lapierre, L. A., Dorn, M. C., Zimmerman, C. F. F., Navarre, J., Burnette, J. O. and Goldenring, J. R. (2003). Rab11b resides in a vesicular compartment distinct from Rab11a in parietal cells and other epithelial cells. *Exp. Cell Res.* **290**, 322–331.

Leung, K. F., Dacks, J. B. and Field, M. C. (2008). Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* **9**, 1698–716.

Lumb, J. H., Leung, K. F., DuBois, K. N. and Field, M. C. (2011). Rab28 function

in trypanosomes: interactions with retromer and ESCRT pathways. *J. Cell Sci.* **124**, 3771–3783.

Maddison, W. P. and Maddison, D. R. (2015). Mesquite: a modular system for evolutionary analysis. Version 2.75. 2011. *URL http//mesquiteproject. org.*

Manna, P. T., Kelly, S. and Field, M. C. (2013). Adaptin evolution in kinetoplastids and emergence of the variant surface glycoprotein coat in African trypanosomatids. *Mol. Phylogenet. Evol.* **67**, 123–8.

Miller, M. A., Pfeiffer, W. and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop, GCE 2010*, .

Murungi, E., Barlow, L. D., Venkatesh, D., Adung'a, V. O., Dacks, J. B., Field, M. C. and Christoffels, A. (2014). A comparative analysis of trypanosomatid SNARE proteins. *Parasitol. Int.* **63**, 341–348.

Natesan, S. K. A., Peacock, L., Leung, K. F., Matthews, K. R., Gibson, W. and Field, M. C. (2009). The trypanosome rab-related proteins RabX1 and RabX2 play no role in intracellular trafficking but may be involved in fly infectivity. *PLoS One* **4**,.

Newman, A. P., Shim, J. and Ferro-Novick, S. (1990). BET1, BOS1, and SEC22 are members of a group of interacting yeast genes required for transport from the endoplasmic reticulum to the Golgi complex. *Mol. Cell. Biol.* **10**, 3405–14.

Niyogi, S., Mucci, J., Campetella, O. and Docampo, R. (2014). Rab11 Regulates Trafficking of Trans-sialidase to the Plasma Membrane through the Contractile Vacuole Complex of Trypanosoma cruzi. *PLoS Pathog.* **10**, e1004224.

Niyogi, S., Jimenez, V., Girard-Dias, W., de Souza, W., Miranda, K. and Docampo, R. (2015). Rab32 is essential for maintaining functional acidocalcisomes, and for growth and infectivity of Trypanosoma cruzi. *J. Cell Sci.* **128**, 2363–73.

Oberholzer, M., Morand, S., Kunz, S. and Seebeck, T. (2006). A vector series for rapid PCR-mediated C-terminal in situ tagging of Trypanosoma brucei genes. *Mol. Biochem. Parasitol.* **145**, 117–20.

Olkkonen, V. M. and Ikonen, E. (2006). When intracellular logistics fails--genetic defects in membrane trafficking. *J. Cell Sci.* **119**, 5031–45.

Ortiz-Sandoval, C. G., Hughes, S. C., Dacks, J. B. and Simmen, T. (2014). Interaction with the effector dynamin-related protein 1 (Drp1) is an ancient function of Rab32 subfamily proteins. *Cell. Logist.* **4**, e986399.

Pal, A., Hall, B. S., Nesbeth, D. N., Field, H. I. and Field, M. C. (2002). Differential endocytic functions of Trypanosoma brucei Rab5 isoforms reveal a glycosylphosphatidylinositol-specific endosomal pathway. *J. Biol. Chem.* **277**, 9529–9539.

Pereira-Leal, J. B. (2008). The Ypt/Rab family and the evolution of trafficking in fungi. *Traffic* **9**, 27–38.

Pereira-Leal, J. B. and Seabra, M. C. (2001). Evolution of the Rab family of small GTP-binding proteins. *J. Mol. Biol.* **313**, 889–901.

Petrželková, R. and Eliáš, M. (2014). Contrasting patterns in the evolution of the Rab GTPase family in Archaeplastida. *Acta Soc. Bot. Pol.* **83**, 303–315.

Rajendran, L. and Annaert, W. (2012). Membrane trafficking pathways in Alzheimer's disease. *Traffic* **13**, 759–70.

Rambaut, A. (2009). FigTree v1.3.1. 2006-2009. *Accessed Novemb. 29, 2012* Program package available at http://tree.bio.ed.ac.

Rasband, W. (2012). ImageJ. *U. S. Natl. Institutes Heal. Bethesda, Maryland, USA* //imagej.nih.gov/ij/.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Sanderfoot, A. (2007). Increases in the number of SNARE genes parallels the rise of multicellularity among the green plants. *Plant Physiol.* **144**, 6–17.

Sanderfoot, A. A., Assaad, F. F. and Raikhel, N. V (2000). The Arabidopsis genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiol.* **124**, 1558–1569.

Schlacht, A., Herman, E. K., Klute, M. J., Field, M. C. and Dacks, J. B. (2014). Missing Pieces of an Ancient Puzzle: Evolution of the Eukaryotic Membrane-Trafficking System. *Cold Spring Harb. Perspect. Biol.* **6**, a016048.

Schwartz, K. J., Peck, R. F. and Bangs, J. D. (2013). Intracellular trafficking and glycobiology of TbPDI2, a stage-specific protein disulfide isomerase in Trypanosoma brucei. *Eukaryot. Cell* **12**, 132–141.

Seixas, E., Barros, M., Seabra, M. C. and Barral, D. C. (2013). Rab and Arf proteins in genetic diseases. *Traffic* **14**, 871–85.

Sevova, E. S. and Bangs, J. D. (2009). Streamlined architecture and glycosylphosphatidylinositol-dependent trafficking in the early secretory pathway of African trypanosomes. *Mol. Biol. Cell* **20**, 4739–50.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

Suwastika, I. N., Uemura, T., Shiina, T., Sato, M. H. and Takeyasu, K. (2008). SYP71, a plant-specific Qc-SNARE protein, reveals dual localization to the plasma membrane and the endoplasmic reticulum in Arabidopsis. *Cell Struct. Funct.* **33**, 185–92.

Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinforma.* **Chapter 2**, Unit 2 3.

Ulrich, P. N., Jimenez, V., Park, M., Martins, V. P., Atwood, J., Moles, K., Collins, D., Rohloff, P., Tarleton, R., Moreno, S. N. J., et al. (2011). Identification of contractile vacuole proteins in Trypanosoma cruzi. *PLoS One* **6**, e18013.

Vedovato, M., Rossi, V., Dacks, J. B. and Filippini, F. (2009). Comparative analysis of plant genomes allows the definition of the "Phytolongins": a novel non-SNARE longin domain protein family. *BMC Genomics* **10**, 510.

Wang, Y. and Seemann, J. (2011). Golgi biogenesis. *Cold Spring Harb. Perspect. Biol.* **3**, 1–12.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.

Wirtz, E., Leal, S., Ochatt, C. and Cross, G. M. (1999). A tightly regulated inducible expression system for conditional gene knock-outs and dominant-negative genetics in Trypanosoma brucei. *Mol. Biochem. Parasitol.* **99**, 89–101.

Woo, Y. H., Ansari, H., Otto, T. D., Klinger, C. M., Kolisko, M., Michálek, J., Saxena, A., Shanmugam, D., Tayyrov, A., Veluchamy, A., et al. (2015). Chromerid genomes reveal the evolutionary path from photosynthetic algae to obligate intracellular parasites. *Elife* **4**, e06974.

Yoshizawa, A. C., Kawashima, S., Okuda, S., Fujita, M., Itoh, M., Moriya, Y., Hattori, M. and Kanehisa, M. (2006). Extracting sequence motifs and the phylogenetic features of SNARE-dependent membrane traffic. *Traffic* **7**, 1104–1118.

Zheng, H., Bednarek, S. Y., Sanderfoot, A. A., Alonso, J., Ecker, J. R. and Raikhel, N. V (2002). NPSN11 is a cell plate-associated SNARE protein that interacts with the syntaxin KNOLLE. *Plant Physiol.* **129**, 530–9.
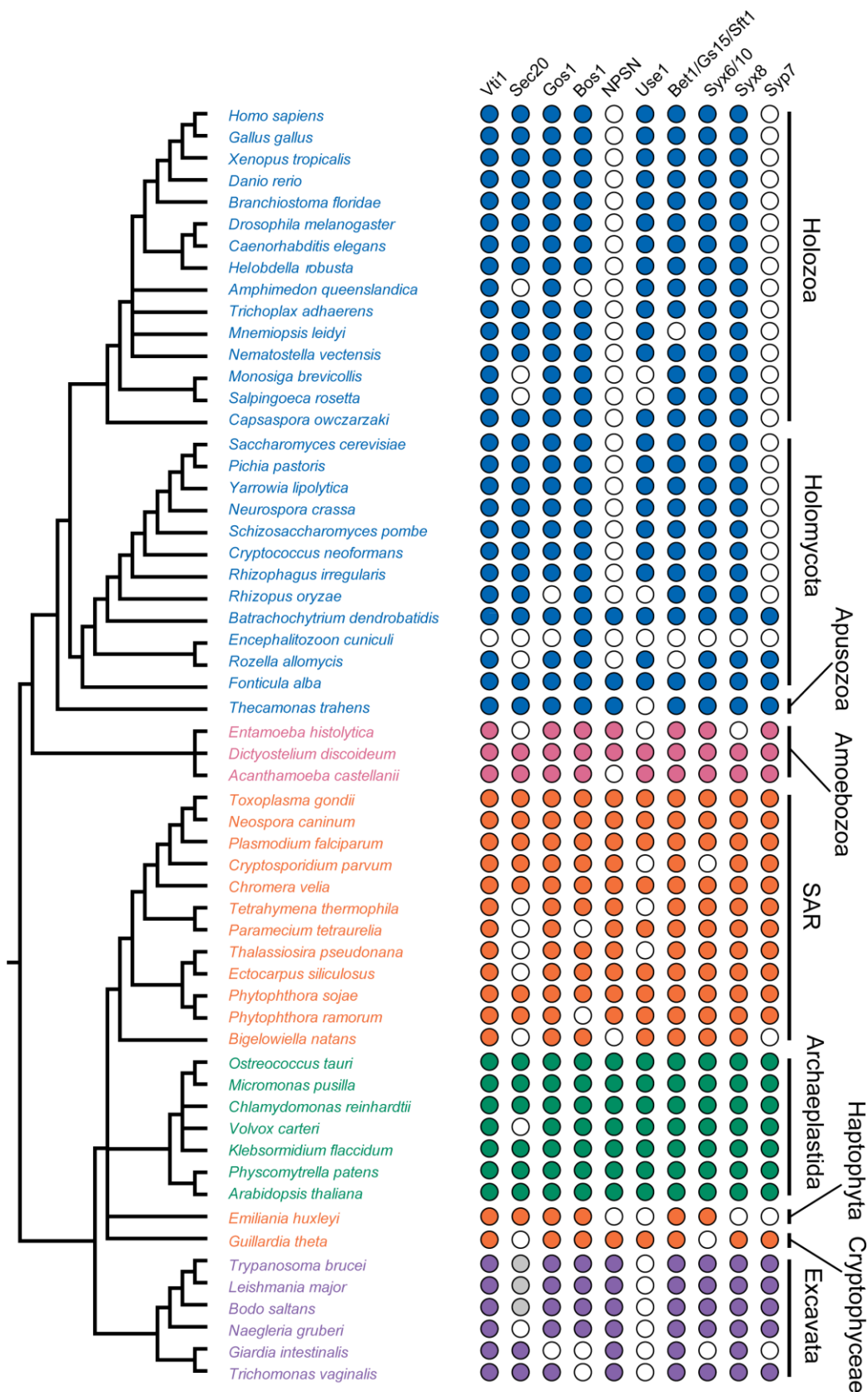
Figures

Figure 1: Distribution of Qb and Qc SNARE family members across a diverse sampling of eukaryotes. Colour-filled circles indicate presence of at least one identified orthologue, while empty circles indicate failure to identify any orthologues. Gray-filled circles indicate identification in downstream analyses focused specifically on kinetoplastid genomes (see Figure 5A). Identified sequences and sources are listed in Tables S4-S6.
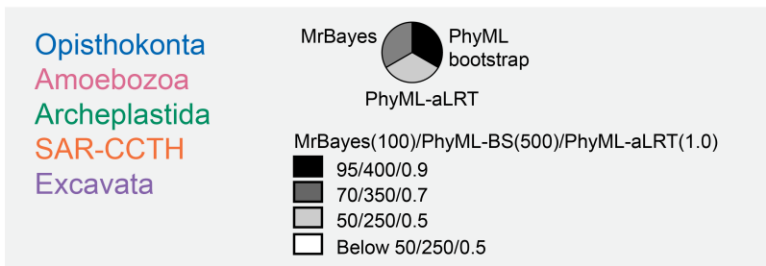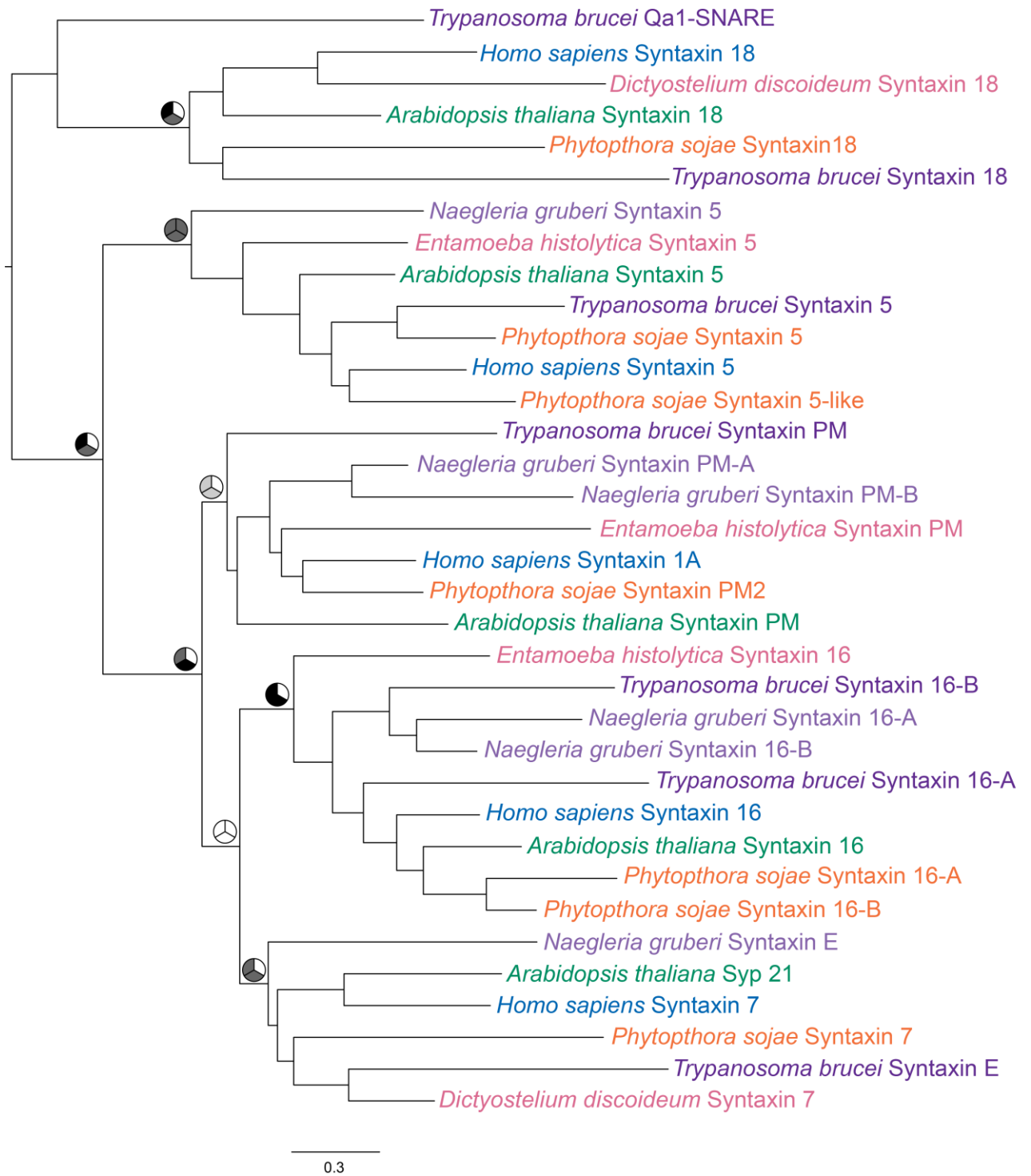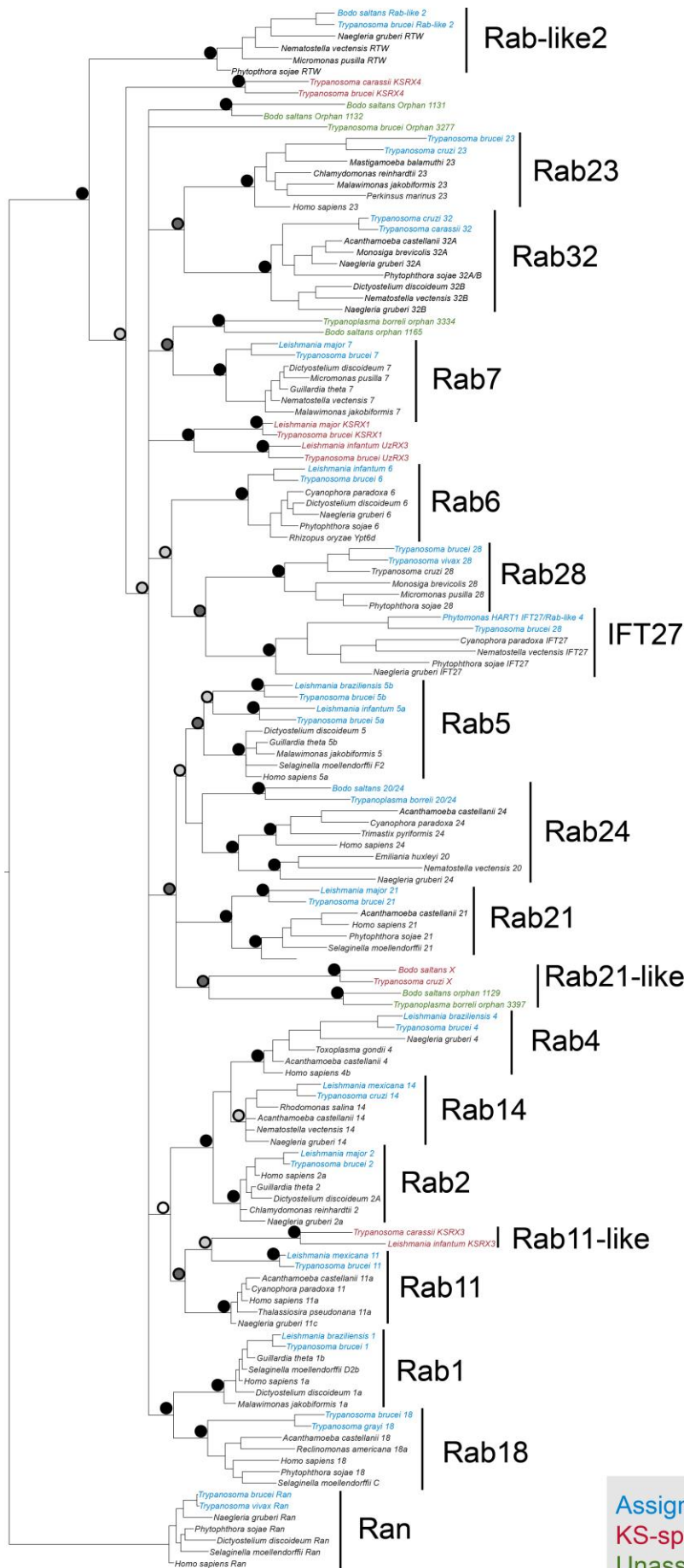
Figure 2: Phylogenetic assignment of kinetoplastid Qa SNAREs. The optimum PhyML topology is presented. Node values are iconised as pie charts for three support values, representing PhyML approximate likelihood ratio test, PhyML Bootstrap and MrBayes posterior probabilities and colour-coded as indicated. Each phylogeny shows one representative kinetoplastid SNARE from each sub-type (Purple) along with eukaryotic representative SNAREs from Opisthokonta (Blue), Amoebozoa (Pink), Archaeplastids (Green), SAR-CCTH (Orange) and Excavata (light Purple). Qa SNAREs showing orthology with eukaryotic orthologs for Syntaxin 18, 5, 16, PM and E.

Rab-like2

Bodo saltans Rab-like 2
Trypanosoma brucei Rab-like 2
Naegleria gruberi RTW
Nematostella vectensis RTW
Micromonas pusilla RTW
Phytopthora sojae RTW

Trypanosoma carassii KSRX4
Trypanosoma brucei KSRX4
Bodo saltans Orphan 1131
Bodo saltans Orphan 1132
Trypanosoma brucei Orphan 3277

Rab23

Trypanosoma brucei 23
Trypanosoma cruzi 23
Mastigamoeba balamuthi 23
Chlamydomonas reinhardtii 23
Malawimonas jakobiformis 23
Perkinsus marinus 23
Homo sapiens 23

Rab32

Trypanosoma cruzi 32
Trypanosoma carassii 32
Acanthamoeba castellanii 32A
Monosiga brevicolis 32A
Naegleria gruberi 32A
Phytophthora sojae 32A/B
Dictyostelium discoideum 32B
Nematostella vectensis 32B
Naegleria gruberi 32B

Trypanoplasma borreli orphan 3334
Bodo saltans orphan 1165

Rab7

Leishmania major 7
Trypanosoma brucei 7
Dictyostelium discoideum 7
Micromonas pusilla 7
Guillardia theta 7
Nematostella vectensis 7
Malawimonas jakobiformis 7

Leishmania major KSRX1
Trypanosoma brucei KSRX1
Leishmania infantum UzRX3
Trypanosoma brucei UzRX3

Rab6

Leishmania infantum 6
Trypanosoma brucei 6
Cyanophora paradoxa 6
Dictyostelium discoideum 6
Naegleria gruberi 6
Phytophthora sojae 6
Rhizopus oryzae Ypt6d

Rab28

Trypanosoma brucei 28
Trypanosoma vivax 28
Trypanosoma cruzi 28
Monosiga brevicolis 28
Micromonas pusilla 28
Phytophthora sojae 28

IFT27

Phytomonas HART1 IFT27/Rab-like 4
Trypanosoma brucei 28
Cyanophora paradoxa IFT27
Nematostella vectensis IFT27
Phytophthora sojae IFT27
Naegleria gruberi IFT27

Rab5

Leishmania braziliensis 5b
Trypanosoma brucei 5b
Leishmania infantum 5a
Trypanosoma brucei 5a
Dictyostelium discoideum 5
Guillardia theta 5b
Malawimonas jakobiformis 5
Selaginella moellendorffii F2
Homo sapiens 5a

Rab24

Bodo saltans 20/24
Trypanoplasma borreli 20/24
Acanthamoeba castellanii 24
Cyanophora paradoxa 24
Trimastix pyriformis 24
Homo sapiens 24
Emiliania huxleyi 20
Nematostella vectensis 20
Naegleria gruberi 24

Rab21

Leishmania major 21
Trypanosoma brucei 21
Acanthamoeba castellanii 21
Homo sapiens 21
Phytophthora sojae 21
Selaginella moellendorffii 21

Rab21-like

Bodo saltans X
Trypanosoma cruzi X
Bodo saltans orphan 1129
Trypanoplasma borreli orphan 3397

Rab4

Leishmania braziliensis 4
Trypanosoma brucei 4
Naegleria gruberi 4
Toxoplasma gondii 4
Acanthamoeba castellanii 4
Homo sapiens 4b

Rab14

Leishmania mexicana 14
Trypanosoma cruzi 14
Rhodomonas salina 14
Acanthamoeba castellanii 14
Nematostella vectensis 14
Naegleria gruberi 14

Rab2

Leishmania major 2
Trypanosoma brucei 2
Homo sapiens 2a
Guillardia theta 2
Dictyostelium discoideum 2A
Chlamydomonas reinhardtii 2
Naegleria gruberi 2a

Rab11-like

Trypanosoma carassii KSRX3
Leishmania infantum KSRX3

Rab11

Leishmania mexicana 11
Trypanosoma brucei 11
Acanthamoeba castellanii 11a
Cyanophora paradoxa 11
Homo sapiens 11a
Thalassiosira pseudonana 11a
Naegleria gruberi 11c

Rab1

Leishmania braziliensis 1
Trypanosoma brucei 1
Guillardia theta 1b
Selaginella moellendorffii D2b
Homo sapiens 1a
Dictyostelium discoideum 1a
Malawimonas jakobiformis 1a

Rab18

Trypanosoma brucei 18
Trypanosoma grayi 18
Acanthamoeba castellanii 18
Reclinomonas americana 18a
Homo sapiens 18
Phytophthora sojae 18
Selaginella moellendorffii C

Ran

Trypanosoma brucei Ran
Trypanosoma vivax Ran
Naegleria gruberi Ran
Phytophthora sojae Ran
Dictyostelium discoideum Ran
Selaginella moellendorffii Ran
Homo sapiens Ran

0.2

Assigned KS
KS-specific
Unassigned 'Orphan'
Reference

MrBayes/PhyML
95%/0.9
80%/0.8
70%/0.7
50%/0.5

Figure 3: Phylogenetic relationships of kinetoplastid Rabs. The consensus Bayesian topology is shown. Nodes are iconised as colour-coded circles according to MrBayes posterior probabilities and PhyML approximate likelihood ratio test as shown in the key. The figure includes two representative kinetoplastid Rabs from each subfamily cluster along with representative Rabs from across eukaryotes. Assigned kinetoplastid-specific Rabs are in blue, lineage-specific Rabs in red, unassigned orphan Rabs in green, while representative eukaryotic Rabs are black. LECS Rabs for which no kinetoplastid orthologs were found (Rabs 8, 22, 34, 50 and Titan) were not included in the analysis. Ran is used as outgroup.
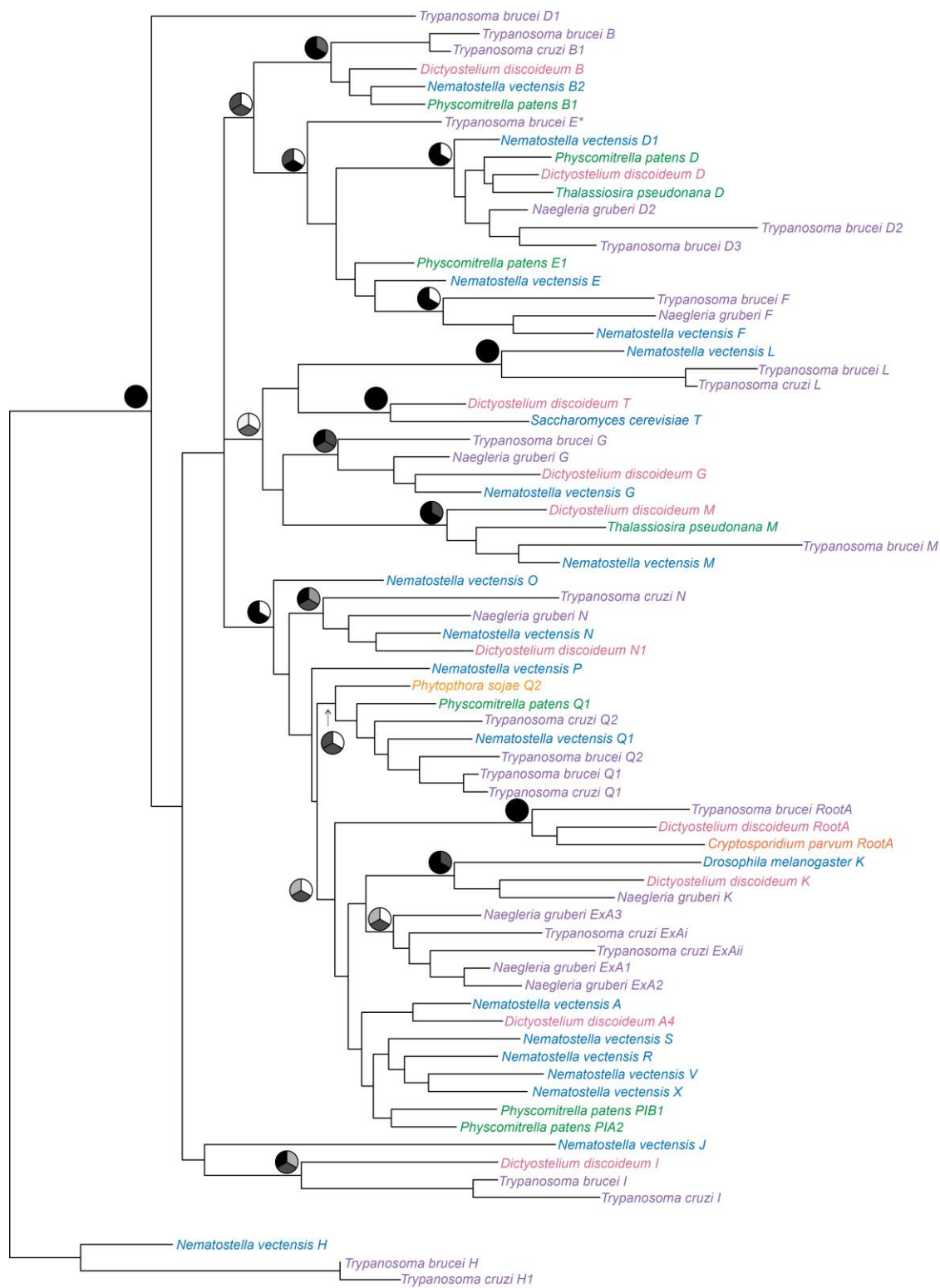
Figure 4: Phylogenetic relationships of kinetoplastid TBC-domain-containing predicted proteins. The optimum PhyML topology is shown. Node values are iconised as pie charts as in Figure 4. Each phylogeny shows one representative kinetoplastid TBC from each sub-type cluster (Purple) along with eukaryotic representative SNAREs from Opisthokonta (Blue), Amoebozoa (Pink), Archaeplastida (Green), SAR-CCTH (Orange) and Excavata (light Purple). TBCs – B, D, F, L, M and Root A are supported by at least two support values above 0.9/90/90 and TBCs G, N, K and I have at least one value above 0.9/90/90 confidence. TBC-Q has two support values above 0.7/70/70 while TBC-ExA has even lower support with one value over 0.5/50/50 and one over 0.7/70/70. In the latter case this suggest that the clade may not be monophyletic. The protein previously described as TBC-E, failed to resolve and is highlighted with an asterisk.

Figure 5: Representation of SNAREs, Rabs and RabGAPs in kinetoplastids.
Individual subtype clades as found in kinetoplastids (assigned to known eukaryotic sub-types or lineage-specific) are shown by columns, with taxa shown as rows, with the hypothetical last common kinetoplastid ancestor as the lowest row. A schematic taxon phylogeny is at left. Black indicates at least one member of the clade was found with phylogenetic support of 80/0.5/50 (MrBayes/PhyMlaLRT/PhyMLbootstrap) or more; gray indicates lower support but above 50 (MrBayes) or in the case of the *LKCA*, indeterminable presence from given data. SNARE subtypes are shown in Figure 5A and are colour-coded as Qa – Purple; Qb – Pink, Qc – Green, Qbc – Gray; and R – Light purple. Subtypes with single asterisks indicate that assignment was via BLAST only due to unresolved phylogeny. Rab and RabGAPs are shown in Figure 5B.

Figure 6: Sub-cellular localisation of *T. brucei* R-SNAREs. Panels A - D: Myc-tagged TbVAMP7A (A), HA-tagged TbVAMP7B (B), TbVAMP7C (C) and TbYkt6 (D) in procyclic cells are shown. Localisation of each SNARE (red) is shown relative to markers (green) for the endoplasmic reticulum (TbBiP) (top), the lysosome (p67) (middle) and the Golgi complex (TbGRASP) (bottom). The nucleus and kinetoplast are stained with DAPI and pseudocoloured in blue. Tags were visualised with rat anti-c-myc antibody or rat anti-HA antibody as appropriate. Panel E: Localisation of candidate TbVAMP7C interactors. HA-tagged VAMP7C (red), is shown relative to putative interactors tagged with myc (top five rows) and V5 (lower two rows), in green. The nucleus and kinetoplast were stained with DAPI. VAMP7C was stained with rat anti-HA, SNARE interactors were stained with mouse anti-c-myc antibody and non-SNARE interactors were stained with mouse anti-V5 antibody. Scale bar is 2μm.

Tables

Table 1: Extent of conservation of LECA trafficking family proteins in kinetoplastids. Putative LECA subtypes of each protein family is presented along with those that were not identified in kinetoplastids. In the final column, proteins predicted to be the last kinetoplastid common ancestor (LKCA) but not in LECA are shown. These include proteins predicted to be derived from expansion of LECA proteins. Numbers in parentheses indicate the number of copies present.

Family	LECA repertoire	Total in LECA	Absent from kinetoplastids	Retention	non-LECA proteins
SNAREs		20		18/21 (86%)	
Qa	Syx18, Syx5, Syx16, SynE, SynPM, Syx 17		Syx17		Qa1, Syx16 (2)
Qb	Sec20, Bos1, Gos1, Vti1, Npsn		none		Vti1-like (3), Npsn (3)
Qc	Use1, Bet1, Syx6, Syx8, Syp7		Use1		Bet1 (2), Syp7 (2), Syx6-like (2)
Qbc	SNAP-25		Very restricted		
R	Sec22, Ykt6, VAMP7, R.reg		none		VAMP7 (4)
Rab	1, 2, 4, 5, 6, 7, 8, 11, 14, 18, 20/24, 21, 22, 23, 28, 32A/B, 34, 50, RTW, IFT27, Titan	22	8, 22, 32B, 34, 50, Titan	16/22 (72%)	Rab11-like, Rab32-like, Rab21 (2), KSRX1, UzRX3, KSRX4
TBC	B, D, E, F, I, L, M, N, Q, RootA	10	none	10/10 (100%)	ExA (2)

Table 2: Proteins identified as potential VAMP7C interactors. Data are arranged by rank order emPAI. Bold indicates an interaction validated by co-immunofluorescence (Figure 6). All experiments and identifications were confirmed in a total of four biological replicates.

| Accession	ID	# of peptides	emPAI	Family
Tb427.10.790	**Handle-VAMP7C**	**10**	**5.74**	**SNARE**
Tb09.211.3920	**Qa-Syx16B**	**10**	**1.98**	**SNARE**
Tb427.03.5570	**Qa-SynE**	**7**	**1.27**	**SNARE**
Tb427.10.1830	**Qc2a-Syp7B**	**6**	**1.25**	**SNARE**
Tb427.08.3470	**Qb2-Vti1-like A**	**6**	**1.22**	**SNARE**
Tb427.08.1120	**Qb3-Vti1like B**	**2**	**0.87**	**SNARE**
Tb427.10.2340	Qc3-Syx8-like	2	0.67	SNARE
Tb427.10.6780	**VPS45**	**10**	**0.67**	**SM protein**
Tb427.10.9950	Qa-Syx18	4	0.32	SNARE
Tb427.10.14200	Qa-Syx5	3	0.28	SNARE
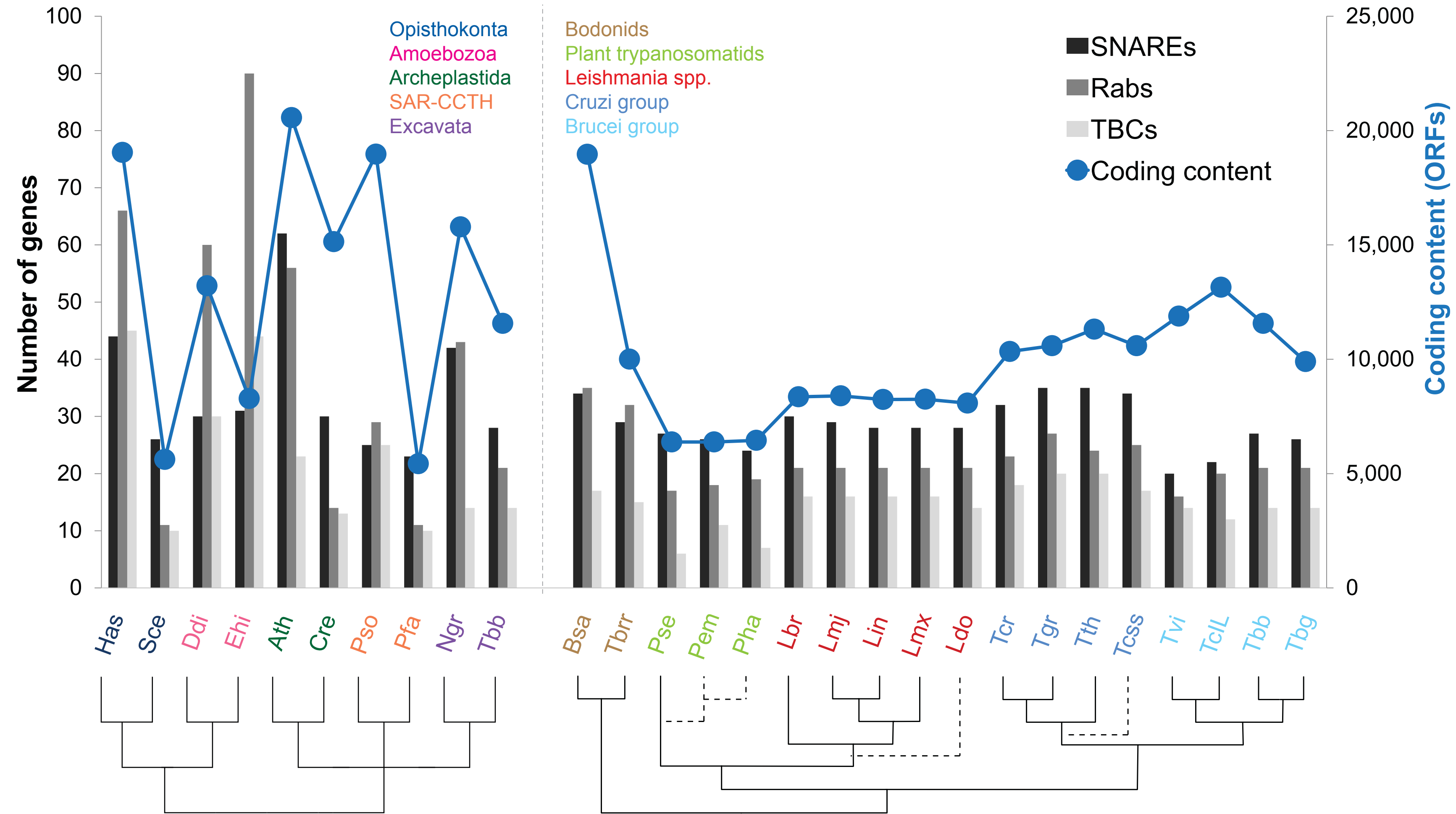
Figure S1. Representation of SNARE, Rab and TBC coding sequences in selected eukaryotic genomes and kinetoplastids. Genomes are arranged by phylogenetic relationships. The five classically recognised, sensu Adl 2005, eukaryotic super groups and each sub-group of kinetoplastida are colour-coded according 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.
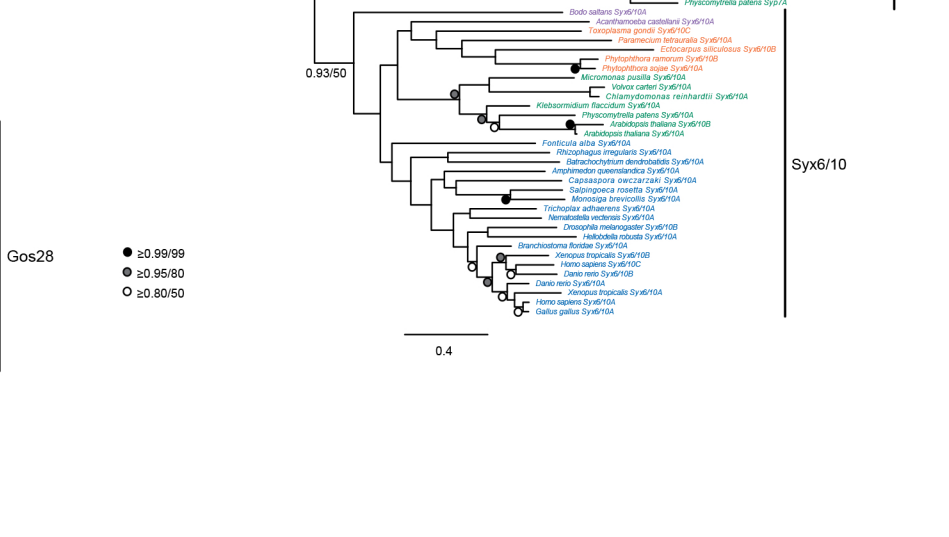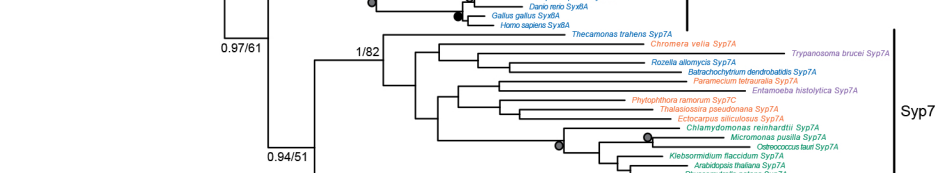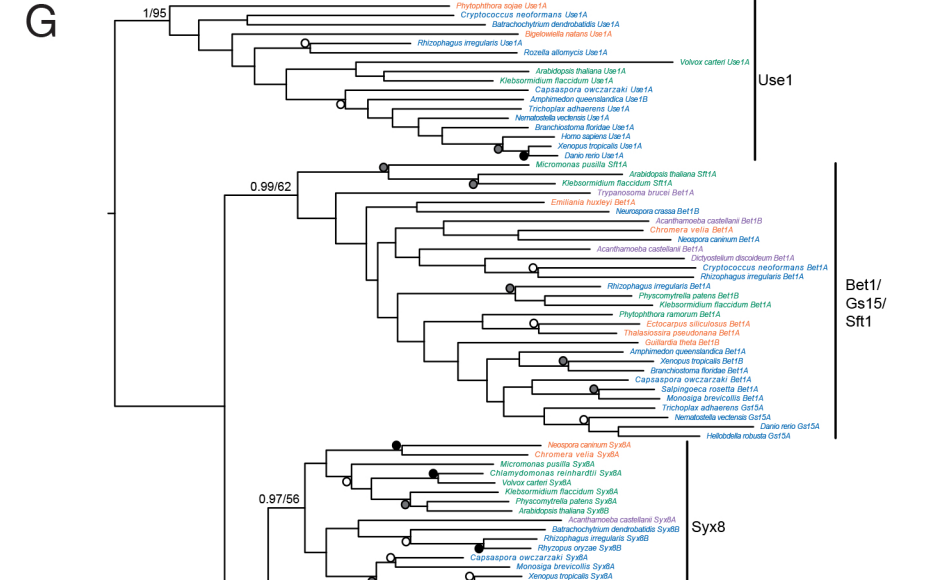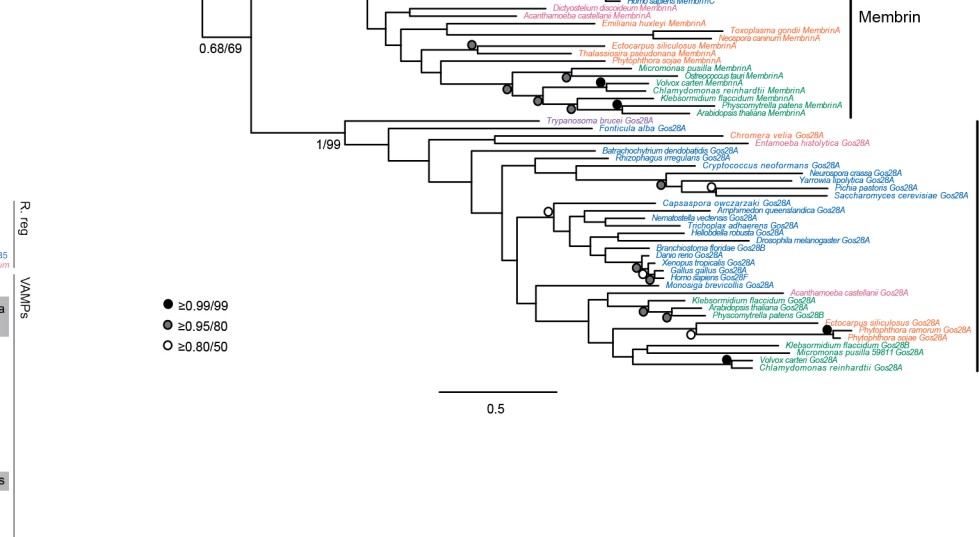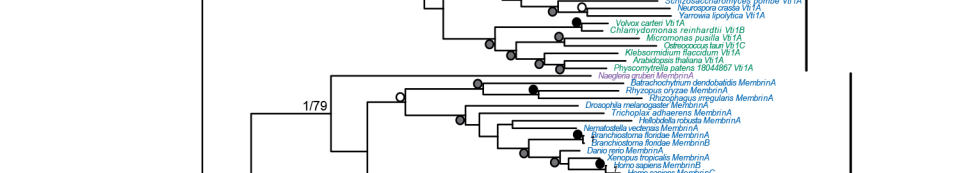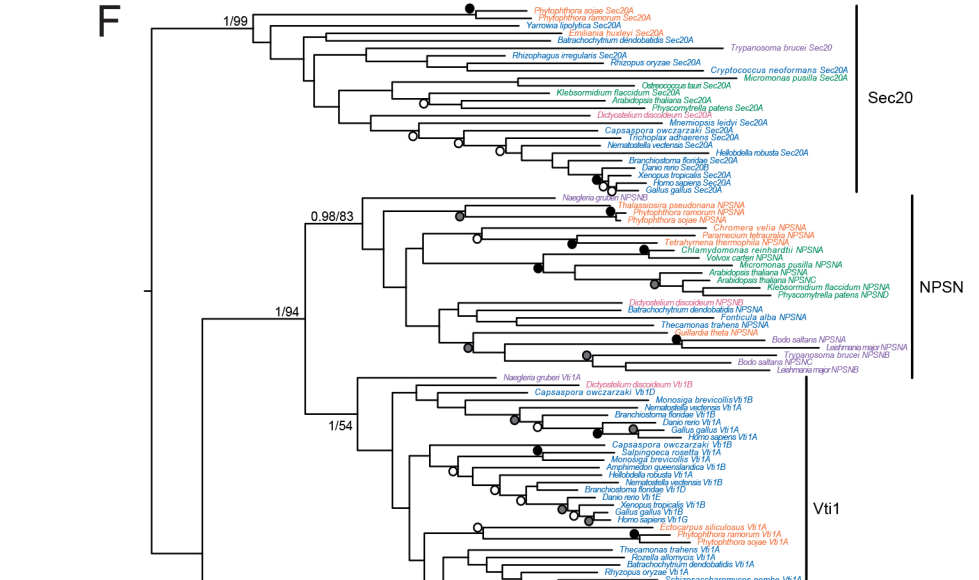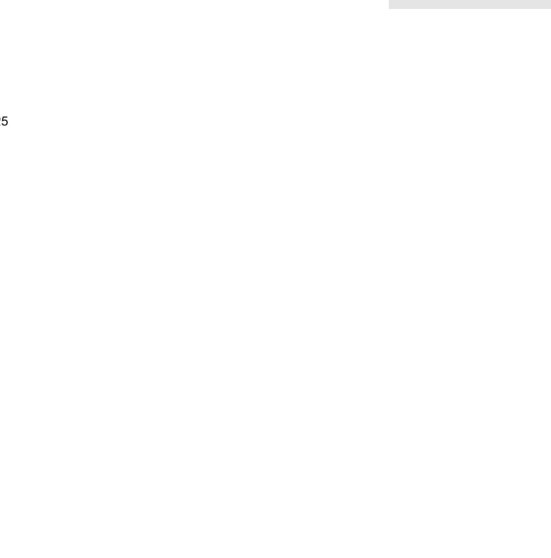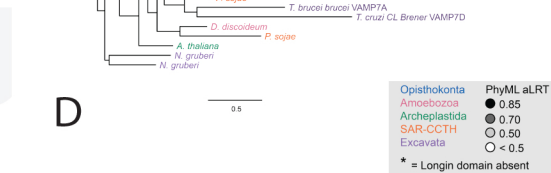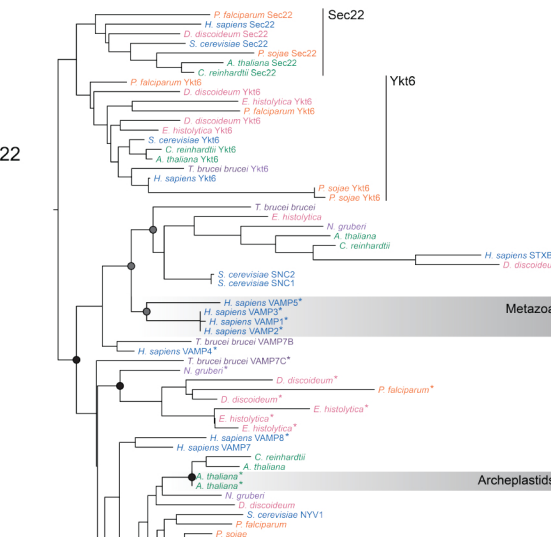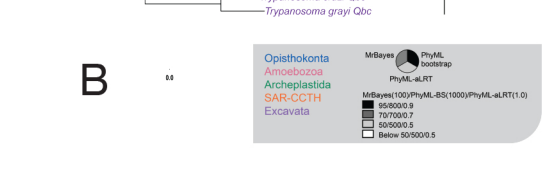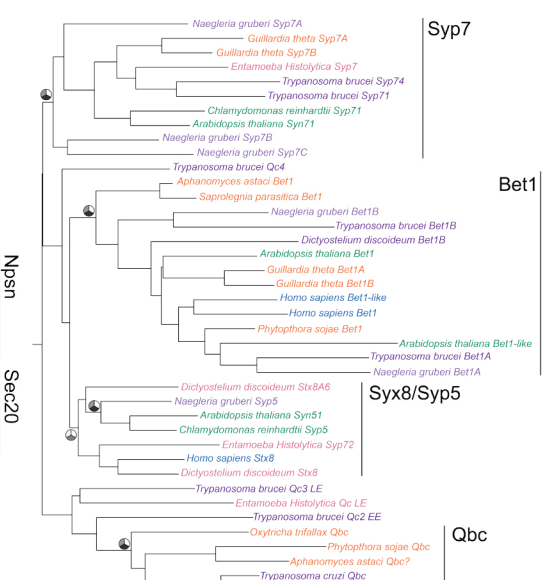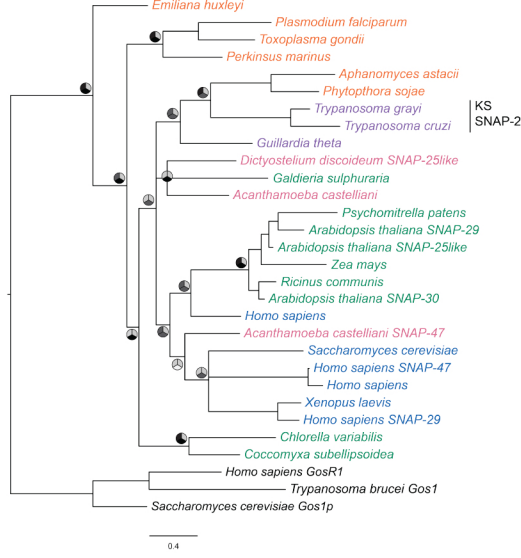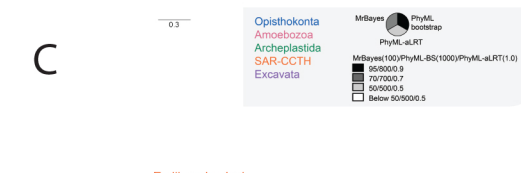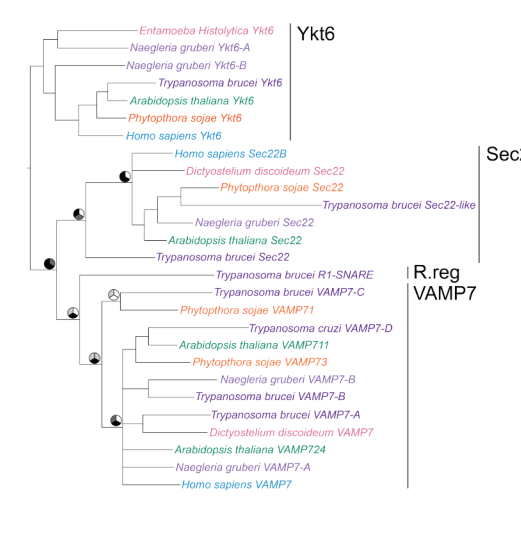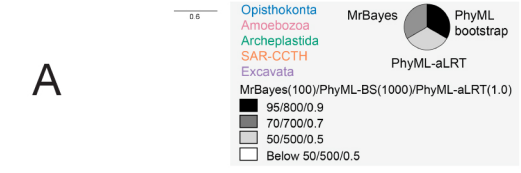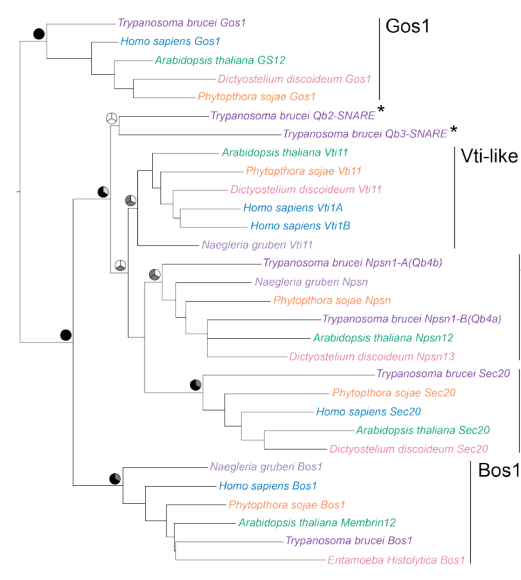
Figures S2. Phylogenetic assignment of kinetoplastid SNAREs (A-C), SNAP-25 (D) and synaptobrevins (E) and Qb (F) and Qc SNAREs (G) confirming NPSN and Syp7 were part of the LECA complement. Best PhyML topology of Qb (S2A) and Qc (S2B), and best Bayesian topology of R (S2C) SNARE phylogeny is presented. 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
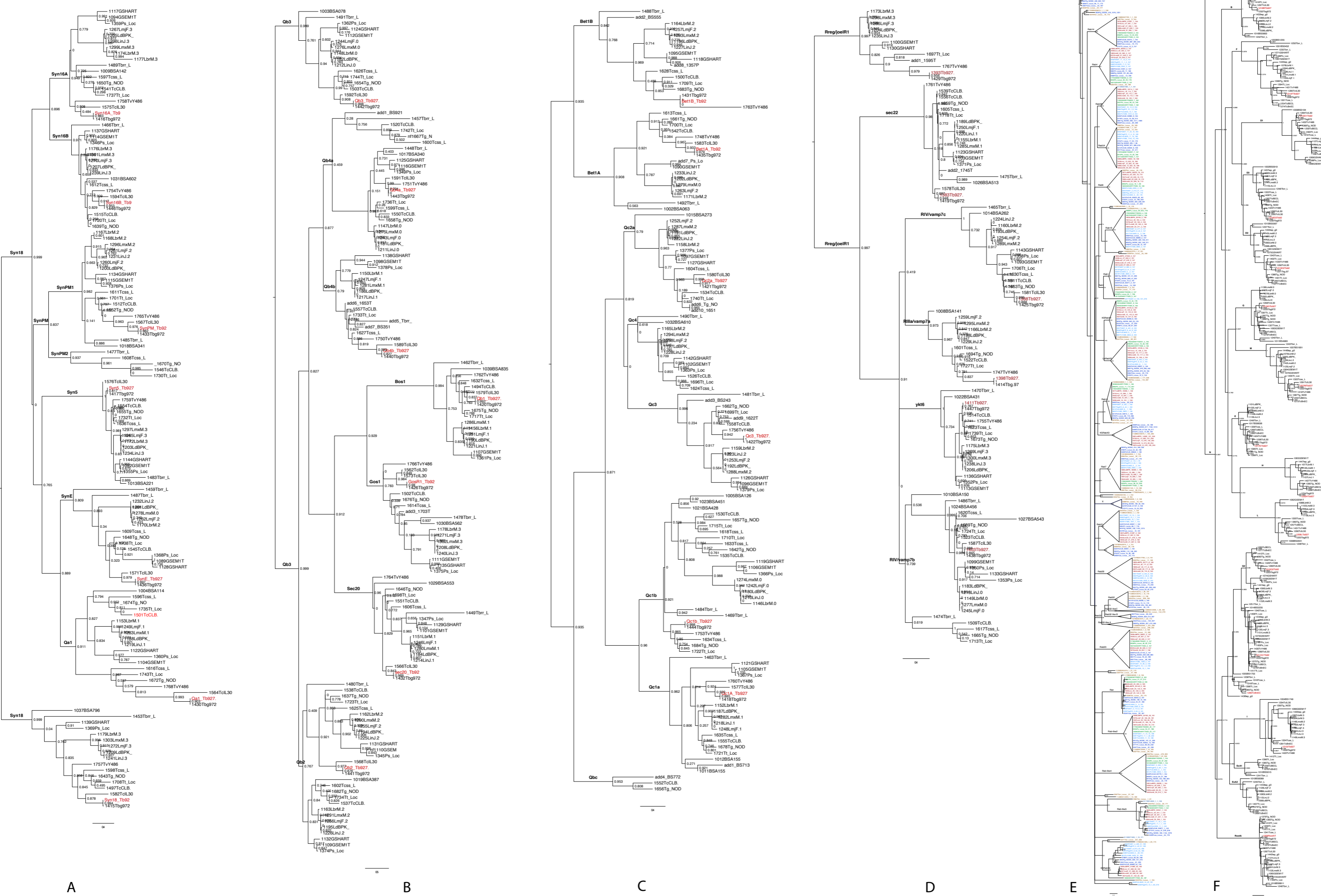
Figure S3A-F. Phylogenetic tree of kinetoplastid SNAREs (A-D), Rabs (E) and TBCs

(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