Novel syntaxin gene sequences from *Giardia*, *Trypanosoma* and algae: implications for the ancient evolution of the eukaryotic endomembrane system

Joel B. Dacks* and W. Ford Doolittle

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, N.S., B3H 4H7, Canada.

*Author for correspondence (e-mail: jdacks@is2.dal.ca)

Accepted 25 January 2002
Journal of Cell Science 115, 1635-1642 (2002) © The Company of Biologists Ltd

Summary

SNAP receptors or SNARES are crucial components of the intracellular membrane system of eukaryotes. The syntaxin family of SNAREs have been shown to have roles in neurotransmission, vesicular transport, membrane fusion and even internal membrane compartment reconstruction. While syntaxins and SNAREs in general have been well characterized in mammalian and yeast models, little is known about their overall distribution across eukaryotic diversity or about the evolution of the syntaxin gene family. By combining bioinformatic, molecular biological and phylogenetic approaches, we demonstrate that various syntaxin homologs are not only present in 'eukaryotic crown taxa' but across a wide range of eukaryotic lineages. The alignment of evolutionarily diverse syntaxin paralogs shows that an isoleucine residue critical to nSec1–syntaxin complex formation and the characteristic syntaxin glutamine residue are nearly universally conserved, implying a general functional importance for these residues. Other identified functional residues involved in botulism toxicity and calcium-binding-protein interactions are also compared. The presence of Golgi-related syntaxins in the intestinal parasite *Giardia intestinalis* provides further evidence for a cryptic Golgi in this 'adictyosomal' taxon, and another likely case of secondary reduction in this parasite. The phylogeny of syntaxins shows a number of nested duplications, including a case of parallel evolution in the plasma membrane-associated syntaxins, and ancestral duplications in the other syntaxin paralogs. These speak to ancient events in the evolution of the syntaxin system and emphasize the universal role of the syntaxins in the eukaryotic intracellular compartment system.

Key words: SNARE, Protist, Golgi

Introduction

The eukaryotic endomembrane system and trafficking between its various components underlie basic cellular processes such as secretion and phagocytosis, as well as more complicated organismal features such as neuronal function. The evolution of this system represents a crucial but poorly studied step in the transition from prokaryotes to eukaryotes: some authors consider it the key event in this transition (Stanier, 1970). The improved internal transport, secretory and heterotrophic capabilities that come with an endomembrane system would have allowed the first eukaryote to exploit a number of new ecological and evolutionary opportunities.

SNAREs have been implicated in vesicle tethering (Ungermann et al., 2000), docking (Ungermann et al., 1998), and fusion (Nickel et al., 1999). Some authors have suggested that SNAREs form the minimal fusion machinery (Sollner et al., 1993) and are responsible for the specificity of vesicle transport in the eukaryotic cell (McNew et al., 2000). Other workers downplay the importance of SNARES, particularly in fusion (Peters et al., 2001) and as the minimal machinery (Wickner and Haas, 2000).

One class of T-snares, the syntaxins, form a clearly delineated protein family based on primary and secondary structure (Bennett et al., 1993). Syntaxins have three N-terminal regulatory helices (Parlati et al., 1999), denoted A, B and C, interspaced with linker regions of variable size. These are linked to the SNARE motif (Fasshauer et al., 1998), a coiled-coil forming helix that generally ends in a transmembrane spanning domain, although some syntaxins lack this anchor (Low et al., 2000).

Syntaxin proteins can themselves be classified into various paralog families (Bennett et al., 1993), each associated with either a step in the transport pathway, or an intracellular location. There are a number of plasma membrane (PM) localized syntaxins (syntaxins 1-4, 11, knolle, syr 1, sso 1 and -2) collectively referred to herein as syntaxin PM homologs. Other syntaxin paralogs may be associated with endoplasmic reticulum (ER), Golgi or endosomes. Syntaxins have also been implicated in reassembly of Golgi (Rabouille et al., 1998) and transitional ER (Roy et al., 2000), perhaps suggesting a role for syntaxins in maintaining organellar stability and the identity of an intracellular compartment.

The majority of functional work on SNAREs has been performed using animal neuronal (Bennett et al., 1992) and fungal secretion models (Banfield et al., 1995; Sollner et al., 1993). However, with the exception of a few studies from plants (Lauber et al., 1997; Leyman et al., 1999) and a single syntaxin from the slime mold *Dictyostelium discoideum*

(Bogdanovic et al., 2000), there is little understanding of the diversity of syntaxins among eukaryotes. In a recent bio-informatic study we suggested that an early eukaryotic ancestor had a relatively complex endomembrane machinery including a primitive syntaxin (Dacks and Doolittle, 2001). This BLAST-based analysis used only partial and publicly available genome sequences and so was not suited to examine syntaxin evolution in-depth. A more phylogenetically diverse sampling of full syntaxin sequences would allow us to: (1) determine details of events in the evolution of the syntaxin superfamily; and (2) more effectively deduce functional constraints on syntaxin primary structure from patterns of evolutionary conservation. In addition, while there have been several phylogenetic classifications of syntaxins, some failed to test the robustness of their result by resampling (Bogdanovic et al., 2000; Sanderfoot et al., 2000) and none accounted for important variables in phylogenetic reconstruction such as among site rate variation or invariant sites (Bogdanovic et al., 2000; Sanderfoot et al., 2000; Wang et al., 1997). These are well known to have major effects on phylogenetic analyses (Efron et al., 1996; Felsenstein, 1978; Lockhart et al., 1996; Yang, 1994) and so a thorough phylogenetic analysis of the syntaxin gene family is warranted.

We have undertaken to expand the available taxonomic diversity of syntaxin sequences using a combined bioinformatic and molecular biology approach. We have identified and sequenced seven different syntaxins from various protists and present the first rigorous molecular phylogenetic analysis of the syntaxin gene family.

Materials and Methods

EST identification, gene amplification and sequencing

Expressed sequence tags (EST) were identified using the BLAST algorithm (Altschul et al., 1997) at NBCI (www.ncbi.nlm.nih.gov/BLAST), as well as by searching the *Dictyostelium discoideum* (www.csn.biol.tsukuba.ac.jp), *Chlamydomonas reinhardii* and *Porphyra yezoensis* (www.kazusa.or.jp/en/plant/database.html) and *Phytophthora* EST (www.ncgr.org/pgcl) databases (Asamizu et al., 1999; Nikaido et al., 2000; Qutob et al., 2000). Additional syntaxin searching was performed in *Giardia intestinalis* by searching the *Giardia* genome site (http://hermes.mbl.edu/baypaul/Giardia.html/index.html) or by downloading the *G. intestinalis* incomplete genome and BLASTing locally. Upon identifying a BLAST hit, that sequence was then BLASTed back to the NR database and only queries that also returned BLAST hits to syntaxins were used for further analysis. The various ESTs and their clone identification numbers are listed in Table 1.

The syntaxins identified from the *Giardia* genome project were based on single pass genomic reads. It was therefore necessary to amplify the genes and confirm their double strand sequences. This was done using exact match primers GsynAXF1- TCATCGCTCCTA-GCTACG and GsynXXR2-GTACAGTGCAGCATTTGGCG for *Giardia* syntaxin PM and primers Gsyn7X1F-GCTCAAACTTGTC-GAAGG and Gsyn7X2R-TAAGCACAGCTCATTGCC for *Giardia* syntaxin 16.

5′ and 3′ fragments of a syntaxin 5 homolog were detected in the *Trypanosoma brucei* GSS database. The complete Open Reading Frame (ORF) was obtained (including the missing internal portion) by amplifying the gene from *T. brucei* genomic DNA using exact match primers TBS5X1F-CTCCAACTATGGTTGTAGAGC and TBS5X2R-ATTTCATTGCCTTGAGACGGC designed to the respective GSS fragments. All amplified fragments were cloned into Topo 2.1 vector (Invitrogen Carlsbad VA). Plasmids resulting from these amplifications and all ESTs with the exception of the MY-F08 clones were transformed into TOP10 cells.

The syntaxins identified from ESTs were either based on single pass or partial sequencing reads (not covering the entire ORF). The complete sequencing of each ORF was performed on an ABI 377 sequencer with two clones of each syntaxin ORF sequenced fully in both directions.

Alignments

To aid in alignment of the syntaxins, each syntaxin sequence was analyzed using the secondary structure prediction software at the EMBL site (www.embl-heidelberg.de/Services/index.html). This was followed by alignment of the predicted helix regions of all syntaxins using Clustal X (Thompson et al., 1997). The raw Clustal output was then pared down by eye and only regions of unambiguously alignable sequence were used. This resulted in a final, global alignment of 68 taxa and 87 sites corresponding to the conserved SNARE motif of representatives from all syntaxin families. Several sub-datasets consisting of pairwise combinations of the five syntaxin families were also constructed. Paralog specific alignments were made, yielding a plasma membrane syntaxin alignment of 24 taxa and 112 sites and a non-plasma membrane alignment of 18 and 123 sites. All alignments are available upon request.

Phylogeny

Protein maximum likelihood (ML) analysis was done either using Puzzle (Strimmer and von Haeseler, 1997) with a gamma correction for among-site-rate variation plus a correction for invariant sites (8 plus 1 rate categories) estimated from the dataset or by using ProtML 2.2 (Adachi and Hasegawa, 1996) with a q1000 search for each dataset. Relative estimated log likelihood values (RELLs) were calculated using Mol2con (A. Stoltzfus, personal communication). The topology shown for each dataset is the best ProtML tree, but with branch lengths estimated in Puzzle to incorporate gamma and invariant sites. Maximum likelihood distance analyses were performed using Puzzle (Strimmer and von Haeseler, 1997) to calculate ML distance matrices in coordination with Puzzleboot

Table 1. Syntaxin sequences from this study

Taxon	Assignment	Accession no.	Orf size	Clone name	Clone acc. no.	BLAST	Top hit
T. brucei	Syntaxin 5	AF404745	327 AA	N/A	N/A	2.00E-11	STX3-*C.elegans*
P. sojae	Syntaxin 5	AF404748	321 AA	5-9d-MY.seq	Pgi:S:2018	1.00E-31	Sed5-*O.sativa*
P. infestans	Syntaxin 6	AF404749	248 AA	MY-38-F-08	Pgi:S:5087	1.00E-16	Putative protein
C. reinhardii	Syntaxin 6	AF404746	225 AA	CM011a08_r	AV386929	4.00E-21	Syntaxin of Plants 61
G. intestinalis	Syntaxin 16	AF404743	271 AA	N/A	N/A	1.00E-11	U00064_6-*C.elegans*
G. intestinalis	Syntaxin PM	AF404744	307 AA	N/A	N/A	7.00E-13	SyntaxinA-*C.elegans*
P. yezoensis	Syntaxin PM	AF404747	346 AA	PM059d08_r	AV434474	2.00E-09	Syntaxin-*S.puporea*

The table lists the syntaxin ORFs obtained by PCR or EST sequencing. Syntaxins are listed by organism, syntaxin family assignment, registered clone name and accession number, as well as the new GenBank number for the completely sequenced ORF. In addition, the size of the ORF in amino acids is given along with the top BLAST hit and the BLASTp score.

(A. Roger and M. Holder; http://members.tripod.de/korbi/puzzle/). These matrices were then analyzed using Neighbor from the Phylip package (Felsenstein, 1995) with jumbling. All bootstrap support values are based on 100 replicates.

Results

BLAST identification of clones

To identify putative syntaxin sequences from as wide a range of eukaryotes as possible, we searched various EST, GSS or eukaryotic genome databases. Partial syntaxin clones or sequences identified in this way were completed either by PCR amplification of the gene or cloning of the cDNA. The full sequence was determined by double stranded sequencing of the whole ORF for each syntaxin sequence in Table 1. Once we obtained the entire ORF for each putative syntaxin, BLAST analysis of the full length sequences was used to confirm that they were indeed syntaxin homologs and to give us preliminary sub-family assignments. BLAST scores for each amplified sequence or EST along with its top BLAST hit are listed in Table 1. Although the top hit for the *Phytophthora infestans* syntaxin 6 homolog is a hypothetical ORF, the next seven hits were to either syntaxin 6 or the closely related paralog syntaxin 10, with BLAST scores of e-16 to e-12. Similarly the *Chlamydomonas reinhardii* syntaxin 6 BLASTed to syntaxin of Plant 61, but was followed by syntaxin 6 and 10 with scores ranging down to 4e-10. As seen in Table 1, the other five putative syntaxin sequences BLASTed to specific syntaxin paralog families, both giving us confidence of their identity as syntaxins and providing a preliminary guide as to their phylogenetic affinity.

Physical attributes of the new syntaxin ORFs

The predicted proteins of all syntaxin sequences obtained were well within the normal size range for syntaxin proteins (Table 1) and probably share the characteristic conserved secondary structure. The *Trypanosoma brucei* syntaxin 5 sequence had a slightly longer linker region between helices 1 and 2 (as estimated by secondary structure prediction software) than other syntaxin 5 orthologues. The *Porphyra yezoensis* syntaxin PM ORF encodes a C-terminal extension of 54 amino acids after the end of its transmembrane domain. Given the previous data on syntaxin membrane insertion (Bennett et al., 1993), this region presumably forms a luminal/extracellular extension.

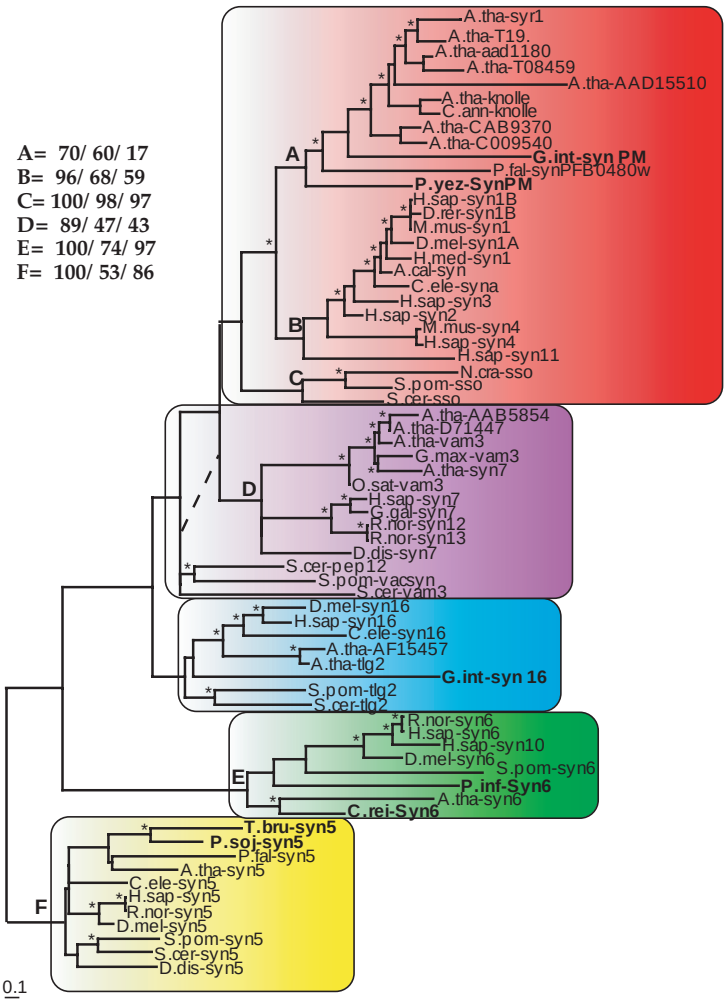

Fig. 1. Global unrooted phylogeny of syntaxin paralogs. This figure illustrates the basic resolution of syntaxin paralogs into five families of varying support and placement of new syntaxin sequences provided by this study (shown in bold). Support values for critical nodes are shown in the inset table with ProtML RELL values/Quartet Puzzling values/Puzzleboot values. For clarity and because internal relationships of the sub-families are more rigorously tested in later analyses, internal nodes for the clades are simply denoted with an asterisk if the ProtML RELL support is greater than 50%. The dashed line linking the fungal endosomal syntaxins to the metazoan and plants ones denotes that all endosomal syntaxins were monophyletic in consensus trees, but did not appear so in the single best tree. Coloured boxes denote the different syntaxin paralog families, with syntaxin PM homologs in red, endosomal homologs in purple, TLG2/syntaxin 16 homologs in blue, syntaxin 6 in green and syntaxin 5 in yellow.

Global phylogeny

In order to determine the evolutionary affinities of the syntaxin genes obtained, a global dataset was assembled of syntaxins from as broad a taxonomic range as possible. Because of the limited sequence conservation of the syntaxin genes, predicted secondary structure of the deduced protein sequences was used as a guide for sequence alignment. So as to be sure of the homology of the regions analyzed, only the unambiguously alignable coiled-coil forming region and the transmembrane domain were used in the global alignment. This yielded a dataset of 68 taxa and 87 aligned sites, which was analyzed using maximum-likelihood (ML) and ML distance methods. As seen in Fig. 1, the syntaxin 5 and syntaxin 6 families were robustly delineated and separated from the other syntaxin genes (bootstrap support values E and F). The syntaxin 16/TLG2, endosomal and plasma-membrane-associated syntaxin families were reconstructed, but not strongly supported.

Because the syntaxin 16/TLG2, endosomal localized and syntaxin PM clades were poorly supported in the global phylogeny, we sought to test their monophyly by explicitly

Table 2. Outgroup analysis testing syntaxin family robustness

	Syntaxin 7	Syntaxin 16	Syntaxin 5	Syntaxin 6
Syntaxin PM	100/25/94	71/54/69	100/75/100	100/29/100
Syntaxin 7		62/64/66	100/98/100	100/98/100
Syntaxin 16			100/94/100	100/99/99
Syntaxin 5				100/90/100

Values shown are ProtML RELLs, Quartet Puzzling values and ML distance bootstrap values, respectively. The 'syntaxin 7' category encompasses syntaxin 7, Vam3 and Vam3-like sequences, all other categories correspond to the sequences enclosed by the coloured boxes in Fig. 1.

asking whether one syntaxin family was robustly separated from another. This was done by performing phylogenetic analysis on all pairwise combinations of syntaxin families, the results of which are summarized in Table 2. In accordance with the global phylogeny, the support for the monophyly of syntaxin 5 and 6 families is very strong. Moreover, the syntaxin 7/Vam 3 and syntaxin PM clades were now reconstructed with significantly improved bootstrap support. The low Puzzle support values for the syntaxin PM family may be due to a recently documented problem of quartet puzzling methods in dealing with sequences with high rates of evolutionary change (Ranwez and Gascuel, 2001). However, Bootstrap support for the separation between the syntaxin 16/TLG2 clade and the syntaxin 7 and syntaxin PM clades was only moderate. In no case was there significant conflict between the placement of the outgroup roots within a syntaxin family, thus discounting the possibility of strong paraphyly. Overall, this suggests that the syntaxin paralogs are monophyletic and allows us to assign, with moderate confidence, the *Giardia* syntaxin 16 sequence, as well as the *Giardia* and *Porphyra* PM syntaxins. Thus all of the syntaxins that we obtained were assignable to specific families, despite their long branch nature in several cases.

Plasma membrane syntaxin phylogeny.

To better assess internal relationships within the plasma membrane associated syntaxins, a sub-dataset was assembled. This enabled the unambiguous alignment of helix A in addition to the regions previously aligned, bringing the total number of sites in this dataset to 112. Unrooted phylogenies clearly delineated the syntaxin PM families of animals, plants and fungi (Fig. 2). There appears to be at least two separate cases of expansion in the syntaxin PM family, one in the metazoan line (Fig. 2, pink box) and one in the streptophytes (Fig. 2, orange box).

Within the animal sequences, the syntaxin 1 genes from both vertebrates and invertebrates (Fig. 2, bracket A) were robustly separated from the human and mouse syntaxin PM subfamilies. In addition, the syntaxin 4 proteins were separated from the syntaxins 1, 2 and 3 with moderate support (Fig. 2, bracket B). Sanderfoot et al. noted that the SNARE complement of *Arabidopsis thaliana* had been expanded (Sanderfoot et al., 2000); this is seen particularly in the syntaxin PM family. The plant syntaxins are robustly separated from the red algal syntaxin and form a number of internally resolved clades. From our analysis it appears likely that the expanded complement will be common to higher plants since the *Capsicum annum* knolle protein is robustly placed with the *Arabidopsis thaliana* knolle. This indicates that the duplication of some plant SNAREs occurred before the separation of these two plant lineages.

Non-plasma membrane syntaxin phylogeny

A dataset of representative syntaxin 5, 16/TLG2 and endosomal homologs was also constructed. This allowed for the unambiguous alignment of helix A from these paralogs, although reliable alignment was not possible between these and the syntaxin PM sequences. The new dataset included 18 taxa and 123 sites. The syntaxin 5 clade was robustly reconstructed under all methods of phylogenetic analysis. The syntaxin 16/TLG2 clade was consistently recovered, despite the long branch nature of the *G. intestinalis* sequence (Fig. 3).

Discussion

This study represents the first rigorous phylogenetic analysis and detailed examination of syntaxin evolution using sequences from beyond the 'eukaryotic crown'. It has

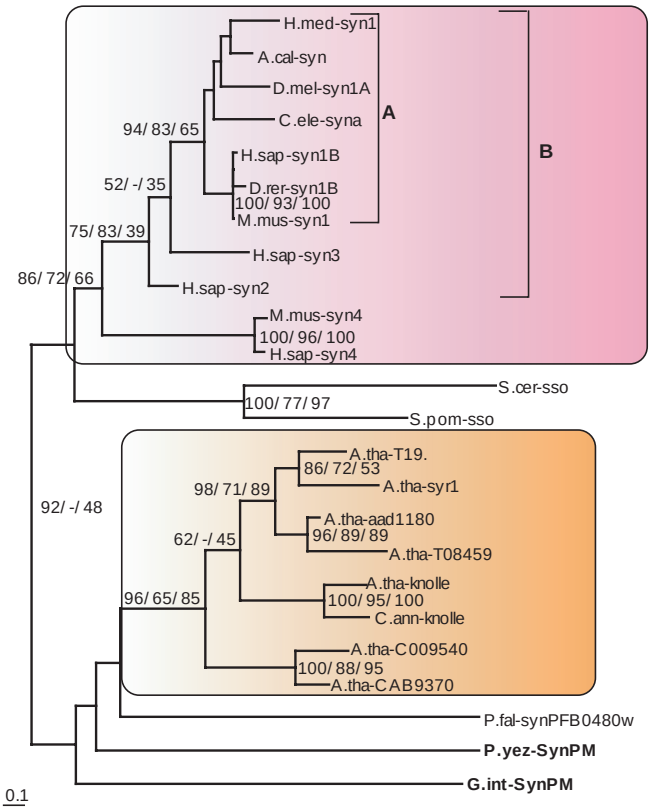

Fig. 2. Unrooted syntaxin PM phylogeny. This data set of 24 taxa and 112 positions shows the two independent evolutionary diversifications of syntaxin paralogs involved in Golgi to plasma membrane transport. The syntaxins derived from the animal diversification are enclosed in the pink box, while the ones in the higher land plant diversification are contained in the orange box. Values for ProtML RELL, Quartet Puzzling and Puzzleboot are shown at all nodes supported over 50% by at least one method. Bracket A shows the separation of syntaxin 1 homologs, while bracket B shows the separation of syntaxins 1, 2 and 3 from syntaxin 4 sequences. New syntaxin sequences provided by this study are shown in bold.

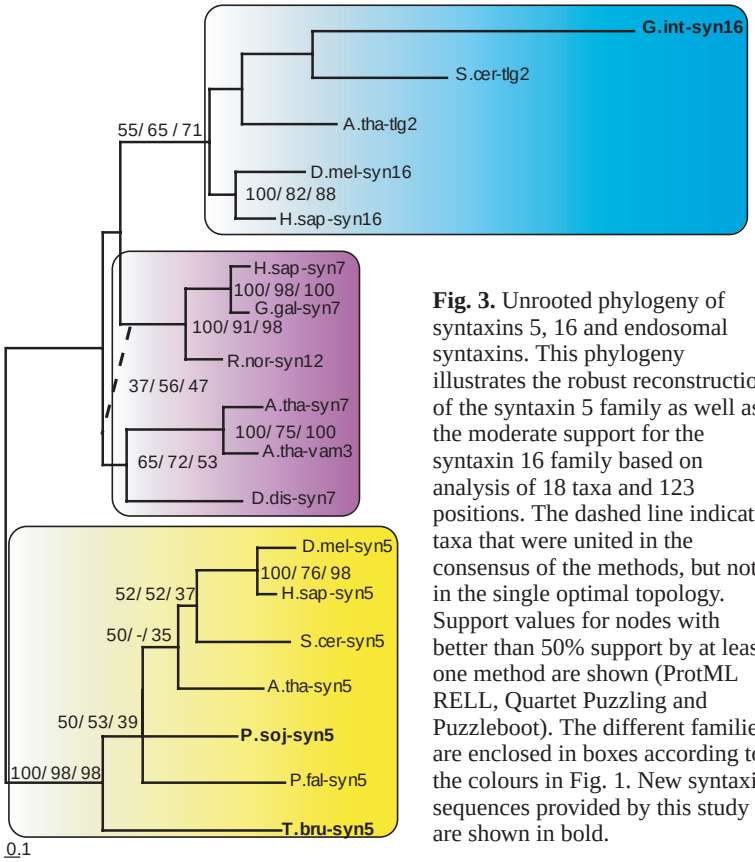

Fig. 3. Unrooted phylogeny of syntaxins 5, 16 and endosomal syntaxins. This phylogeny illustrates the robust reconstruction of the syntaxin 5 family as well as the moderate support for the syntaxin 16 family based on analysis of 18 taxa and 123 positions. The dashed line indicates taxa that were united in the consensus of the methods, but not in the single optimal topology. Support values for nodes with better than 50% support by at least one method are shown (ProtML RELL, Quartet Puzzling and Puzzleboot). The different families are enclosed in boxes according to the colours in Fig. 1. New syntaxin sequences provided by this study are shown in bold.

implications for the evolution of syntaxins and the eukaryotic endomembrane system, as well as allowing us to examine the conservation of identified functional residues and assess their general relevance to syntaxin function.

Evolution of syntaxin functional residues

Syntaxins are classified as 'Q snares' (Fasshauer et al., 1998) based on the presence of a functionally critical glutamine (Q) residue; this glutamine was reliably alignable and conserved in all syntaxins that we obtained (Fig. 4Aa).

Calcium ions are known to play an important role in membrane fusion events (reviewed by Wickner and Haas, 2000). Bezprozvanny et al. identified Ala240 and Val244 in syntaxin 1A as critical to the protein's interaction with α_{1B}, the pore-forming unit of the N-type Ca^{2+} channels (Bezprozvanny et al., 2000). When compared across syntaxins, there appears to be little pattern to the conservation at these positions (Fig. 4Ac,d). There is variability even within the animal syntaxin 1A versus -B, but the *Giardia* syntaxin PM, and some *Arabidopsis* syntaxin PM homologs do have the conserved residues. As well the animal syntaxin 7 and syntaxin 5 proteins have the appropriate alanine and valines. On the other extreme the *Dictyostelium* syntaxin 5 has a large portion of this region deleted entirely. The functional implication of this variability is a matter for experimental investigation.

Neuronal sec1 binds to syntaxin 1A in animal nerve cells and appears to act as a regulatory inhibitor (Schulze et al., 1994). Studies with neuronal sec1 showed that mutating several residues including isoleucine 236, leucine 165 and glutamate 166, in syntaxin 1A significantly decreased sec1 binding (Misura et al., 2000). We found that the isoleucine residue is conserved in all paralogs that we identified with the exception of syntaxin 6, where it was replaced with leucine or phenylalanine (Fig. 4Ab). This underlines the general importance of this position for sec1 binding, and its potential as a target for site-directed mutagenesis studies aimed at examining differential syntaxin 6 function or mechanism with respect to the other syntaxins. Unfortunately, it was not possible to align the region containing leucine 165 and glutamate 166 between the various syntaxin paralogs. However, it was possible to align it within the syntaxin PM proteins (Fig. 4B). The homologous position to Leu165 is always hydrophobic, including a leucine in the *G. intestinalis* and a valine in the *P. yezoensis* syntaxin PM sequences, respectively. The next position is slightly less well conserved with an aspartate in the *Giardia* sequence and a glutamine in the *Porphyra* sequence. These two positions are also well conserved in the *S. cerevisiae* sso1 gene and in the *A. thaliana* syr1 and knolle sequences, indicating the general functional importance of this region. As this region is one of the binding sites for botulism toxin, it should be of considerable interest for studies aimed at the understanding of syntaxin function.

The minimal syntaxin complement of the 'adictyosomal' taxon *Giardia intestinalis* might include Golgi-associated syntaxins

The protozoan *Giardia intestinalis* causes the disease Giardiasis and is the most commonly found intestinal parasite throughout the world (Adams and Perkin, 1985). Based on molecular evidence and its lack of mitochondria and several other membrane bounded organelles, it has also been proposed as one of the earliest evolving eukaryotes (Cavalier-Smith, 1998; Leipe et al., 1993; see also Embley and Hirt, 1998). *Giardia*, in its feeding form, possesses no clearly identifiable Golgi apparatus. However, it is possible to induce the appearance of Golgi structure and function in *Giardia* during encystation (Lujan et al., 1995). Therefore, the adictyosomal status of *Giardia* is a matter of contention.

We amplified, sequenced and analyzed two clear syntaxin paralogs from *Giardia intestinalis*. Interestingly, both of the paralog families to which the *Giardia* syntaxins were assigned are associated with transport steps involving the Golgi apparatus (Abeliovich et al., 1998; Sollner et al., 1993). This seems to provide further evidence for the cryptic presence of a Golgi apparatus in *Giardia*.

Syntaxin complements of animals, fungi and *Arabidopsis* range from 7 to 24 known genes (Bock et al., 2001; Sanderfoot et al., 2000). In our search of the *Giardia* genome only two clearly identifiable syntaxin sequences were found. We have amplified, but not analyzed two other potential syntaxin homologs from the *Giardia* genome database (J.B.D. and W.F.D., unpublished). Although other syntaxins may be present, it seems unlikely that sufficient syntaxin paralogs will

A

```
R. nor-syn1A      LENSIRELHDMFMDMAMLVESQGEMIDRIEYNVEHAVDYVE
H. sap-syn1B      LETSIRELHDMFVDMAMLVESQGEMIDRIEYNVEHSVDYVE
H. sap-syn2       LETSIRELHEMFMDMAMFVETQGEMINNIERNVMNATDYVE
H. sap-syn3       LESSIKELHDMFMDIAMLVENQGEMLDNIELNVMHTVDHVE
H. sap-syn4       LERSIRELHDIFTFLATEVEMQGEMINRIEKNILSSADYVE
S. cer-sso1a      LEKSMAELTQLFNDMEELVIEQQENVDVIDKNVEDAQLDVE
A. tha-syr1       IERSLLELHQVFLDMAALVEAQGNMLNDIESNVSKASSFVM
A. tha-knolle     IEKSLLELHQVFLDMAVMVESQGEQMDEIEHHVINASHYVA
P. yez-synPM      LAGSLTELHAMFVDMGLLVNQQTELLNNIEANVEKTKVETV
G. int-synPM      IQKTAQEIHQLTMDAAMMCEQQSRLIEQIETNVLHAREAVQ
S. cer-vam3       IHTAVQEVNAIFHQLGSLVKEQGEQVTTIDENISHLHDNMQ
H. sap-syn7       LEADIMDINEIFKDLGMMIHEQGDVIDSIEANVENAEVHVQ
A. tha-vam3       IHQQIGEVNEIFKDLAVLVNDQGVMIDDIGTHIDNSRAATS
D. dis-syn7       IEQSIVEINEIFVDLSGLVAEQGVMINTIEASLESTTINTK
A. tha-syn7       IEDQIRDVNGMFKDLALMVNHQGNIVDDISSNLDNSHAATT
D. dis-syn5       ISSTINQLEGIFTQLANLVSMQGEVIERIDLNVS-------
H. sap-syn5       IESTIVELGSIFQQLAHMVKEQEETIQRIDENVLGAQLDVE
S. cer-syn5       IESTIQEVGNLFQQLASMVQEQGEVIQRIDANVDDIDLNIS
T. bru-syn5       IEAAVVEVGEMFNDFTRLVHEQNEIVLRIDTNVETSLRHVN
P. soj-syn5       IESHIVDIGQLFGRLSTLIHEQGDLVRRIDDNVEDSLVNVS
A. tha-syn5       VESRITELSGIFPQLATMVTQQGELAIRIDDNMDESLVNVE
A. tha-tlg2       VVESVNDLAQIMKDLSALVIDQGTIVDRIDYNIENVATTVE
G. int-syn16      ITTGIAEIANIITQMSELIYEQGTVLDRIDANVYTAVGYAE
H. sap-syn16      IVQSISDLNEIFRDLGAMIVEQGTVLDRIDYNVEQSCIKTE
S. cer-tlg2       LARGVLEVSTIFREMQDLVVDQGTIVDRIDYNLENTVVELK
H. sap-syn6       VSGSIGVLKNMSQRIGGELEEQAVMLEDFSHELESTQSRLD
A. tha-syn6       LSKSVQRIGGVGLTIHDELVAQERIIDELDTEMDSTKNRLE
C. rei-syn6       IEQAVIRIGRQGREIGNELAEQERMLDELDQDVDTTHSRLK
P. inf-syn6       LHSDITRLHGVTVEISSEVKHQNKMLDDLTDDVDEAQERMN
S. pom-syn6       VYDTIGNIRGQAALMGEELGQQADLLDTLDNSIETTNSKLR
                                        a      b      c   d
```

B

```
R. nor-syn1A        SEELEDML
M. mus-syn2         DDELEEML
H. sap-syn3         DEELEEML
H. sap-syn4         DEELEQML
G. int-synPM        DAELDFVI
P. yez-synPM        EADVQAAL
P. fal-synPM        DEDISTFL
S. cer-sso1a        EDEVEAAI
A. tha-knolle       DEMIEKII
A. tha-syr1         EETVEKLI
                       ab
```

Fig. 4. Aligned functional regions for syntaxins. This figure illustrates representative syntaxin homologs from evolutionarily diverse eukaryotes aligned in regions containing identified functional residues. (A) Region of the SNARE motif including the epimorphin region, which corresponds to positions 205-246 of *R. norvegicus* Syn1A. Aa shows the universally conserved Gln residue characteristic of syntaxins. Ab shows the identified Ile 236 residue from syntaxin 1A shown to be important for nSec1–syntaxin 1A complex formation. Ac and Ad underline the identified residues involved in Ca^{2+} channel interactions with syn1. (B) Botulism toxin binding region from syntaxin PM homologs, which corresponds to positions 162-170 of *R. norvegicus* Syn1A. Ba and Bb show the two residues also identified to be important for nSec1–syntaxin complex formation.

be identified in *Giardia* to equal the standard eukaryotic diversity. Syntaxin diversity in *Giardia* may be reduced to a minimal complement as *Giardia* is known to have undergone reduction in response to its parasitic lifestyle (Roger, 1999). Since our two analyzed sequences have each been assigned to specific paralogs, this seems to indicate loss rather than divergence of *Giardia* prior to the duplication of the syntaxin families. It will be interesting to see, when the entire *Giardia* genome is complete, what is the final syntaxin complement and to compare it with other, non-reduced, single-celled eukaryotes.

Duplications in animal and plant plasma membrane syntaxins

The syntaxin PM family contains two sets of nested duplications, one in the animal and one in the plant lineage (Fig. 2). This expansion of the syntaxins involved in Golgi to plasma membrane transport seems to have occurred twice independently and represents an interesting case of parallel evolution. Since the red algal (*Porphyra*) syntaxin PM homolog emerges prior to the monophyletic green plant clade, it appears that the expansion of the syntaxin PM sub-family occurred after the red and green algal divergence (Moreira et al., 2000). A syntaxin PM from a green alga might allow us to further narrow the time frame for the beginning of the expansion process in plants. Having a comparable sampling of syntaxins from a plant in addition to *Arabidopsis* as well as further information regarding the biological roles of the predicted syntaxins from the *A. thaliana* genome should shed light on the evolution of this syntaxin sub-family within green algae and land plants.

In the animal lineages, syntaxins 1-4 are well characterized. All are involved in exocytosis but, while syntaxin 4 is constitutively sent to the basolateral region of epithelial cells, syntaxins 2 and 3 are apically associated (Low et al., 2000). Since syntaxin 4 emerges prior to the clade formed by the syntaxin 1-3 paralogs (Fig. 2, bracket B), it is possible that the first duplication of animal PM syntaxins was associated with the evolution of cell polarity in metazoa. Syntaxin 1 is found in the nerve synapse and is involved in neurotransmitter release (Bennett et al., 1992). Since syntaxin 1 is clearly separated from syntaxins 2 and 3 (Fig. 2, bracket A) and since syntaxin 1 is present in both vertebrate and invertebrates, we propose that the syntaxin 1 duplication is associated with the evolution of the nervous system and occurred prior to the vertebrate/invertebrate split.

Paralog duplications and the ancient nature of the syntaxin system

Recently we noted that syntaxin protein genes were present in diverse eukaryotic genomes, prompting us to speculate the presence of a primitive syntaxin in an early eukaryotic ancestor (Dacks and Doolittle, 2001). The phylogeny of the syntaxin families based on full length protozoan gene sequences allows us to deduce the timing of divergence versus duplication events in the evolution of the syntaxin superfamily. If a taxon contains at least two syntaxins from different known paralog families, then the duplication events giving rise to those families must have occurred prior to the divergence of that taxon. Since *Phytophthora* species have both syntaxins 5 and 6, the duplication giving rise to these two syntaxins occurred prior to the origin of *Phytophthora* and by extension the rest of the heterokonts. The same argument can be made, although less strongly in accordance with the lower bootstrap support, for *Dictyostelium discoideum* (syntaxins 5 and 7), *Plasmodium falciparum* (syntaxins 5 and PM) and *Giardia intestinalis* (syntaxins 16/TLG2 and PM). Finally the robust assignment of the *Trypanosoma brucei* syntaxin 5 indicates that the duplication that gave rise to the syntaxin 5 family occurred prior to the divergence of *Trypanosoma*. Similar logic can be applied for *Chlamydomonas*, but with respect to the syntaxin 6 family duplication.

More generally, we have expanded evidence for the presence of syntaxin paralogs from a broad range of eukaryotes, speaking to the universal role of the different syntaxin families in the vesicular transport machinery. We can now more strongly infer the ancient nature of the syntaxin system and conclude that the diversification into paralog families began early in eukaryotic evolution. This raises the possibility that syntaxins could have been involved in the origin or early evolution of the endomembrane system. One major obstacle in first evolving a permanent internal membrane system would be establishing a system that is both stable in the identity of its compartments and its maintenance within the cell, and yet dynamic enough to accommodate incoming and outgoing vesicles. In vitro reconstitution assays have shown that syntaxins not only play a role in vesicular transport but also in organellar reconstruction (Rabouille et al., 1998; Roy et al., 2000). This implies that they could be partially responsible for the 'identity' of an organellar compartment and might have been able to fulfill the early role necessary for internal membrane stability and flexibility.

Syntaxins are critical components of the vesicular transport machinery from human to yeast. Here we have assembled and extended the evidence that they are likely to be present in all eukaryotes, even those deemed primitive. By analyzing syntaxins from across eukaryotic diversity, we begin to get a glimpse of some events in the evolution of the syntaxin system, from recent animal or plant-specific paralog expansions all the way back to the earliest evolutionary stages of the eukaryotic internal membrane system.

We thank Lesley Davis for technical support, and Jeffrey Gerst, Tom Cavalier-Smith, Pak Poon and Robert Hirt for helpful discussion and critical reading of this manuscript. We also want to thank Stephen Hadjuk and Andrew Roger for their generous gifts of *T. brucei* and *G. intestinalis* genomic DNA, respectively. We acknowledge Kazusa DNA research for supplying *Chlamydomonas*, and *Porphyra* ESTs, and thank the *Dictyostelium* cDNA project in Japan with support by Japan Society for the Promotion of Science (RFTF96L00105) and Ministry of Education, Science, Sports and Culture of Japan (08283107) for the unpublished sequence of the *Dictyostelium* syntaxin 5. In addition we are grateful to Mark Gijzen and Wageningen University for the *Phytophthora sojae* and *P. infestans* ESTs, respectively. This work was supported by CIHR grant MT4467 to W.F.D. and a CIHR doctoral research award and Walter C. Sumner Memorial Fellowship to J.B.D.

References

Abeliovich, H., Grote, E., Novick, P. and Ferro-Novick, S. (1998). Tlg2p, a yeast syntaxin homolog that resides on the Golgi and endocytic structures. *J. Biol. Chem.* **273**, 11719-11727.

Adachi, J. and Hasegawa, M. (1996). MOLPHY Version 2.3, Programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**, 1-150.

Adams, R. T. and Perkin, J. E. (1985). The prevalence of intestinal parasites in children living in an unincorporated area in rural northern Florida. *J. Sch. Health* **55**, 76-78.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.

Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S. (1999). A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. *DNA Res.* **6**, 369-373.

Banfield, D. K., Lewis, M. J. and Pelham, H. R. (1995). A SNARE-like protein required for traffic through the Golgi complex. *Nature* **375**, 806-809.

Bennett, M. K., Calakos, N. and Scheller, R. H. (1992). Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. *Science* **257**, 255-259.

Bennett, M. K., Garcia-Arraras, J. E., Elferink, L. A., Peterson, K., Fleming, A. M., Hazuka, C. D. and Scheller, R. H. (1993). The syntaxin family of vesicular transport receptors. *Cell* **74**, 863-873.

Bezprozvanny, I., Zhong, P., Scheller, R. H. and Tsien, R. W. (2000). Molecular determinants of the functional interaction between syntaxin and N-type Ca^{2+} channel gating. *Proc. Natl. Acad. Sci. USA* **97**, 13943-13948.

Bock, J. B., Matern, H. T., Peden, A. A. and Scheller, R. H. (2001). A genomic perspective on membrane compartment organization. *Nature* **409**, 839-841.

Bogdanovic, A., Bruckert, F., Morio, T. and Satre, M. (2000). A syntaxin 7 homologue is present in *Dictyostelium discoideum* endosomes and controls their homotypic fusion. *J. Biol. Chem.* **275**, 36691-36697.

Cavalier-Smith, T. (1998). A revised six-kingdom system of life. *Biol. Rev. Camb. Philos. Soc.* **73**, 203-266.

Dacks, J. B. and Doolittle, W. F. (2001). Reconstructing/deconstructing the earliest eukaryotes. How comparative genomics can help. *Cell* **107**, 419-425.

Efron, B., Halloran, E. and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93**, 7085-7090.

Embley, T. M. and Hirt, R. P. (1998). Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 624-629.

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

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **25**, 401-410.

Felsenstein, J. (1993). PHYLIP: Phylogeny Inference Package (version 3.2). *Cladistics* **5**, 164-166.

Lauber, M. H., Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz, W. and Jurgens, G. (1997). The *Arabidopsis* KNOLLE protein is a cytokinesis-specific syntaxin. *J. Cell Biol.* **139**, 1485-1493.

Leipe, D. D., Gunderson, J. H., Nerad, T. A. and Sogin, M. L. (1993). Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* **59**, 41-48.

Leyman, B., Geelen, D., Quintero, F. J. and Blatt, M. R. (1999). A tobacco syntaxin with a role in hormonal control of guard cell ion channels. *Science* **283**, 537-540.

Lockhart, P. J., Larkum, A. W., Steel, M., Waddell, P. J. and Penny, D. (1996). Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. USA* **93**, 1930-1934.

Low, S. H., Miura, M., Roche, P. A., Valdez, A. C., Mostov, K. E. and Weimbs, T. (2000). Intracellular redirection of plasma membrane trafficking after loss of epithelial cell polarity. *Mol. Biol. Cell* **11**, 3045-3060.

Lujan, H. D., Marotta, A., Mowatt, M. R., Sciaky, N., Lippincott-Schwartz, J. and Nash, T. E. (1995). Developmental induction of Golgi structure and function in the primitive eukaryote *Giardia lamblia*. *J. Biol. Chem.* **270**, 4612-4618.

McNew, J. A., Parlati, F., Fukuda, R., Johnston, R. J., Paz, K., Paumet, F., Sollner, T. H. and Rothman, J. E. (2000). Compartmental specificity of cellular membrane fusion encoded in SNARE proteins. *Nature* **407**, 153-159.

Misura, K. M., Scheller, R. H. and Weis, W. I. (2000). Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. *Nature* **404**, 355-362.

Moreira, D., Le Guyader, H. and Phillippe, H. (2000). The origin of red algae and the evolution of chloroplasts. *Nature* **405**, 69-72.

Nickel, W., Weber, T., McNew, J. A., Parlati, F., Sollner, T. H. and Rothman, J. E. (1999). Content mixing and membrane integrity during membrane fusion driven by pairing of isolated v-SNAREs and t-SNAREs. *Proc. Natl. Acad. Sci. USA* **96**, 12571-12576.

Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Tabata, S. (2000). Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res.* **7**, 223-227.

Parlati, F., Weber, T., McNew, J. A., Westermann, B., Sollner, T. H. and Rothman, J. E. (1999). Rapid and efficient fusion of phospholipid vesicles

by the alpha-helical core of a SNARE complex in the absence of an N-terminal regulatory domain. *Proc. Natl. Acad. Sci. USA* **96**, 12565-12570.

Peters, C., Bayer, M. J., Buhler, S., Andersen, J. S., Mann, M. and Mayer, A. (2001). Trans-complex formation by proteolipid channels in the terminal phase of membrane fusion. *Nature* **409**, 581-588.

Qutob, D., Hraber, P. T., Sobral, B. W. and Gijzen, M. (2000). Comparative analysis of expressed sequences in *Phytophthora sojae*. *Plant Physiol.* **123**, 243-254.

Rabouille, C., Kondo, H., Newman, R., Hui, N., Freemont, P. and Warren, G. (1998). Syntaxin 5 is a common component of the NSF- and p97-mediated reassembly pathways of Golgi cisternae from mitotic Golgi fragments in vitro. *Cell* **92**, 603-610.

Ranwez, V. and Gascuel, O. (2001). Quartet-based phylogenetic inference: improvements and limits. *Mol. Biol. Evol.* **18**, 1103-1116.

Roger, A. J. (1999). Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146-S163.

Roy, L., Bergeron, J. J., Lavoie, C., Hendriks, R., Gushue, J., Fazel, A., Pelletier, A., Morre, D. J., Subramaniam, V. N., Hong, W. et al. (2000). Role of p97 and syntaxin 5 in the assembly of transitional endoplasmic reticulum. *Mol. Biol. Cell* **11**, 2529-2542.

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.

Schulze, K. L., Littleton, J. T., Salzberg, A., Halachmi, N., Stern, M., Lev, Z. and Bellen, H. J. (1994). rop, a *Drosophila* homolog of yeast Sec1 and vertebrate n-Sec1/Munc-18 proteins, is a negative regulator of neurotransmitter release in vivo. *Neuron* **13**, 1099-1108.

Sollner, T. H., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P. and Rothman, J. E. (1993). SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**, 318-324.

Stanier, R. (1970). Some aspects of the biology of cells and their possible evolutionary significance. In *Organization and Control in Prokaryotic and Eukaryotic Cells* (ed. H. Charles and B. Knight), pp. 1-38. Cambridge, UK: Syndics of the Cambridge University Press.

Strimmer, K. and von Haeseler, A. (1996). Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**, 964-969.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876-4882.

Ungermann, C., Price, A. and Wickner, W. (2000). A new role for a SNARE protein as a regulator of the Ypt7/Rab-dependent stage of docking. *Proc. Natl. Acad. Sci. USA* **97**, 8889-8891.

Ungermann, C., Sato, K. and Wickner, W. (1998). Defining the functions of trans-SNARE pairs. *Nature* **396**, 543-548.

Wang, H., Frelin, L. and Pevsner, J. (1997). Human syntaxin 7: a Pep12p/Vps6p homologue implicated in vesicle trafficking to lysosomes. *Gene* **199**, 39-48.

Wickner, W. and Haas, A. (2000). Yeast homotypic vacuole fusion: a window on organelle trafficking mechanisms. *Annu. Rev. Biochem.* **69**, 247-275.

Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**, 306-314.