An Elaborate Classification of SNARE Proteins Sheds Light on the Conservation of the Eukaryotic Endomembrane System

Tobias H. Kloepper,*† C. Nickias Kienle,*† and Dirk Fasshauer†

*Research Group Algorithms in Bioinformatics, Center for Bioinformatics, University of Tübingen, 72076 Tübingen, Germany; and †Research Group Structural Biochemistry, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, 37077 Göttingen, Germany

Proteins of the SNARE (soluble *N*-ethylmalemide–sensitive factor attachment protein receptor) family are essential for the fusion of transport vesicles with an acceptor membrane. Despite considerable sequence divergence, their mechanism of action is conserved: heterologous sets assemble into membrane-bridging SNARE complexes, in effect driving membrane fusion. Within the cell, distinct functional SNARE units are involved in different trafficking steps. These functional units are conserved across species and probably reflect the conservation of the particular transport step. Here, we have systematically analyzed SNARE sequences from 145 different species and have established a highly accurate classification for all SNARE proteins. Principally, all SNAREs split into four basic types, reflecting their position in the four-helix bundle complex. Among these four basic types, we established 20 SNARE subclasses that probably represent the original repertoire of a eukaryotic cenancestor. This repertoire has been modulated independently in different lines of organisms. Our data are in line with the notion that the ur-eukaryotic cell was already equipped with the various compartments found in contemporary cells. Possibly, the development of these compartments is closely intertwined with episodes of duplication and divergence of a prototypic SNARE unit.

INTRODUCTION

The elaborate endomembrane system of eukaryotes is thought to have evolved by invagination of the plasma membrane during perfection of a phagotrophic lifestyle. Subsequently, primitive endomembranes differentiated into the various spatially and functionally separated compartments found in contemporary eukaryotes (Roger, 1999; Cavalier-Smith, 2002). Material exchange is mediated by cargo-loaded vesicles that bud from the donor and eventually fuse with the acceptor compartment. This allows the cells to take up nutrients through the endocytic pathway. Conversely, newly synthesized proteins and lipids are transported within the cell through the exocytic pathway. As each organelle must maintain its identity, vesicular trafficking is tightly regulated (Bonifacino and Glick, 2004). Although some protists possess highly divergent compartments, it is becoming clear that the underlying molecular machineries involved in vesicular trafficking are highly conserved among all eukaryotes. Key players in the concluding step, the fusion of a vesicle with its acceptor membrane, are the so-called SNARE (soluble *N*-ethylmalemide–sensitive factor attachment protein receptor) proteins (reviewed in Hong, 2005; Jahn and Scheller, 2006). They form a family of small cytoplasmically oriented membrane-associated proteins that comprise a relatively simple domain architecture. Their characteristic is the so-called SNARE motif, an extended segment arranged in heptad repeats. In most SNAREs the motif is C-terminally connected to a single transmembrane domain by a short linker.

As general mechanism it is thought that the SNAREs on the transport vesicle form a tight complex with the SNAREs in the acceptor membrane. Directional complex assembly between the membranes, starting from the N-terminal tips of the SNARE motifs toward the C-terminal membrane anchors, is thought to pull the lipid bilayers together and to initiate membrane merger. The very similar crystal structures of three, only distantly related SNARE units (Sutton *et al.,* 1998; Antonin *et al.,* 2002b; Zwilling *et al.,* 2007), have strengthened the view that they all form similar "nanomachines," consisting of an elongated, parallel four-helix bundle. In its interior, 16 layers of mostly hydrophobic residues are formed, which are highly conserved; in particular, a hydrophilic layer in the center, consisting of three glutamine (Q) residues and one arginine (R) residue, is almost unchanged throughout the SNARE family, leading to a classification into Q- and R-SNAREs (Weimbs *et al.,* 1997; Fasshauer *et al.,* 1998). The existence of other asymmetric layers in the bundle implies that the three Q-SNARE helices are also prototypes that define distinct subfamilies. Accordingly, the SNARE motifs of a SNARE unit are classified into Qa-, Qb-, Qc-, and R-SNAREs ("QabcR-complex"; Bock *et al.,* 2001). However, it is still not entirely clear whether all SNARE sequences can be classified into four basic types. The analysis of SNARE sequences is also made difficult by the fact that the current Hidden Markov model (HMM) profiles used in the Pfam and SMART databases often do not recognize SNARE motifs nor do they allow for an unambiguous classification of SNARE types.

An initial phylogenetic survey had indicated that all SNAREs build functional units that are in accord with the structural "QabcR-rule" (Bock *et al.,* 2001). These SNARE units are thought to participate in different membrane traffic steps within the cell. Yet, several SNARE proteins seem to function in more than one trafficking step, rendering it difficult to specify unique SNARE sets (Banfield, 2001; Pelham, 2001; Jahn and Grubmuller, 2002; Jahn and Scheller, 2006). In addition, the original classification of SNARE proteins (Bock *et al.,* 2001) was based on only ~100 sequences from only few organisms and did not include some more diverged SNARE types (Lewis *et al.,* 1997; Lewis and Pelham, 2002; Burri *et al.,* 2003; Dilcher *et al.,* 2003). In the subsequent years, additional complete genomes shed more light onto the conservation of the SNARE machinery, yet some SNAREs, in particular from protists, appeared to be rather atypical (Sanderfoot *et al.,* 2000; Dacks and Doolittle, 2002, 2004; Gupta and Brent Heath, 2002; Uemura *et al.,* 2004; Besteiro *et al.,* 2006; Schilde *et al.,* 2006; Sutter *et al.,* 2006; Ayong *et al.,* 2007; Kissmehl *et al.,* 2007; Sanderfoot, 2007). These studies, however, did not provide a universal classification scheme, as usually only a subset of sequences was examined. Thus, it remains unclear how the remarkable morphological diversity of eukaryotes is reflected in their SNARE repertoires. For example, so far the highest number of SNARE proteins was discovered in green plants (Sanderfoot *et al.,* 2000; Sutter *et al.,* 2006; Sanderfoot, 2007), but it is unclear so far whether the additional SNAREs mediate novel trafficking steps or whether they are simply variations in the given repertoire. This calls for an exhaustive comparison of SNARE repertoires in different organisms. Evidently, a better understanding of the number, distributions, and interactions of SNAREs in different eukaryotic kingdoms might shed more light onto the organization, conservation, and possibly the origins of the eukaryotic endomembrane system.

Here, we present a detailed classification of the different members of the SNARE family, reflecting their participation in different trafficking steps. Nineteen of 20 generated HMM profiles achieve at least a 95% accuracy and can therefore be used as a highly significant method to classify SNARE proteins. Finally we provide an interactive web interface to de novo classify SNAREs and to access our collected information.

MATERIALS AND METHODS

Sequences

We started with a set of ≈150 SNARE proteins from the five species *Homo sapiens, Drosophila melanogaster, Caenorhabditis elegans, Saccharomyces cerevisiae,* and *Arabidopsis thaliana.* These were grouped into the four previously established subgroups (Bock *et al.,* 2001). We aligned each subgroup using *muscle* (Edgar, 2004) and used an eye-by-eye verification to assure the quality of the alignment. We extracted the 53 amino acid long SNARE motif of each subgroup and used HMMER, with standard settings and calibration, to build HMM profiles (hmmer.janelia.org; Durbin *et al.,* 1998) and to search the nonredundant (nr) database at National Center for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov). We gathered ~800 sequences and used the SNARE motif to perform a classification analysis. With the results of the classification analysis we searched the nr-database at NCBI again and gathered more than 2000 protein sequences. We finally added more than 1000 protein sequences from different genome projects (DOE Joint Genome Institute [JGI], genome.jgi-psf.org/; and Broad Institute; www.broad. mit.edu/seq/). After removing sequences with more than one occurrence, sequences that seemed to be misassembled or sequences who failed an eye-by-eye verification, we obtained a set of 2165 SNARE sequences. For these sequences we removed insertions and added gaps according to the aligned HMMER sequence.

Classification Analysis

We reconstructed the phylogenetic tree at two different levels of evolution, the complete tree of all SNAREs, and trees of each of the four main subgroups (data not shown). Because of the large number of sequences, we defined a core set of 11 representative species (short name in brackets), for which we calculated the individual species trees and a general tree based on all 11

species: *Arabidopsis thaliana (ArTh), Cryptococcus neoformans (CrNe), Danio rerio (DaRe), Homo sapiens (HoSa), Neurospora crassa (NeCr), Schistosoma japonicum (ScJa), Ostreococcus lucimarinus (OsLu), Phytophthora sojae (PhSo), Schizosaccharomyces pombe (ScPo), Trichomonas vaginalis (TrVa),* and *Trypanosoma cruzi (TrCr).* For each set of sequences gathered, we started the analysis by building an IQPNNI (Important Quartet Puzzling and Nearest Neighbor Interchange) tree (Vinh le and Von Haeseler, 2004) from the SNARE motifs using the Jones, Taylor, and Thornton (JTT-) distance matrix (Jones *et al.,* 1992) and a gamma distribution with four categories to take rate-heterogeneity across sites into account. Afterward we used Likelihood-Mapping (Strimmer and von Haeseler, 1997) for each edge in the IQPNNI tree to estimate the accuracy of the topology. For more confidence, we also used the *phylip* package (Felsenstein, 1989) to run a distance-based bootstrap analysis with 1000 replicates. We used standard settings for *seqboot*, again the JTT-distance matrix (Jones *et al.,* 1992) and also a gamma distribution (with parameter approximation from *treepuzzle*; Schmidt *et al.,* 2002), for *protdist* and standard options for *neighbor*. Whenever necessary we used a random seed of 9. Because bootstrap values have been shown to be systematically biased, we used the almost unbiased (AU) test (Shimodaira, 2002) to correct for this. The site wise log-likelihoods needed for the AU-test were obtained using a modified version of *phyml* (Guindon and Gascuel, 2003) and the test was performed using *consel* (Shimodaira and Hasegawa, 2001). Finally we estimated the root of the four main subgroups using the trees of the four main subgroups with the Qa subgroups as outgroups for the Qb and R groups and the R subgroups as outgroups for the Qa and Qc subgroups.

Cluster Analysis

Together with the results of the phylogenetic analysis, biological knowledge and the domain structure of the complete protein sequences, we assigned each sequence motif to one functional subgroup. The resulting subgroups were analyzed for their PPR and sensitivity. Whenever we found motifs with <95% PPR, we joined the subgroup with the subgroup causing most false positive hits. Motifs with <95% sensitivity were broken apart into smaller and more sensitive subgroups.

### A Web Interface for the De Novo Classification of SNAREs and Access to Our Results

We implemented a Web-based interface for access to our results (http://bioinformatics.mpibpc.mpg.de/snare/). It is divided into three sections. The first section is dedicated to the access to our collected information, which can be searched for groups, species, and protein names. We programmed two different views, one for an overview and one for the detailed sequence information, such as the position of the found motif and its expectation value. The second section presents an interface for the submission of new sequences to our HMM models. We implemented a 0.1 expectation value cutoff to minimize false-positive results. The results display the best three hits and the position of the motif in the alignment. The final section contains the SNARE tree generated in Nexus format (Maddison *et al.,* 1997) that can be analyzed in detail with SplitsTree (Huson and Bryant, 2006).

RESULTS

For our study, we used only the highly conserved SNARE motif, which allows for an ungapped alignment of 53 amino acids (Figure 1). We started with a known set of approximate 150 SNARE proteins (Bock *et al.,* 2001). We extracted HMM profiles for the four main groups of the SNARE family and searched the nonredundant (nr) database from the NCBI site (www.ncbi.nlm.nih.gov) with these. We gathered ~800 protein sequences for a phylogenetic analysis. This allowed to refine the four main groups into 20 distinct subgroups. We then extracted HMM profiles for these subgroups and gathered ≈3600 protein sequences from the nr-database, the est-database, and some genome projects. Rigorously removing splice variants and sequences with low certainty we kept a total of 2165 sequences. One hundred forty-five species occurred within the analysis, and we are certain that we have found an almost complete set of SNAREs in about half of these species. In detail, our analysis included SNAREs from 59 animals, 41 fungi, 18 plants, 25 protists, and 2 viruses.

The discovery of SNAREs encoded in the genome of viruses was somewhat unexpected, but these proteins might be used during host cell invasion. The R-SNARE (group IV) encoded by the Coccolithovirus, which infects the microalga

Figure 1. Schematic representation of the different domain arrangements found in SNARE proteins. The four-helix bundle structure of the neuronal SNARE complex is shown as ribbon diagram on the top right (blue, red, and green for synaptobrevin 2, syntaxin 1a, and SNAP-25a, respectively). The layers (-7 until $+8$) in the core of the bundle are indicated by virtual bonds between the corresponding Cα positions. The structure of the central 0-layer is shown in detail on the top left (Sutton *et al.,* 1998). Below the domain architecture of two most important types of SNARE proteins are depicted. The highly conserved SNARE motif is indicated by a yellow box and the adjacent transmembrane domain by a black box. In addition, the structures of two types of autonomous N-terminal domains: a three-helix bundle (Fernandez *et al.,* 1998; Lerman *et al.,* 2000) and a profilin-like domain (Gonzalez *et al.,* 2001) are shown.
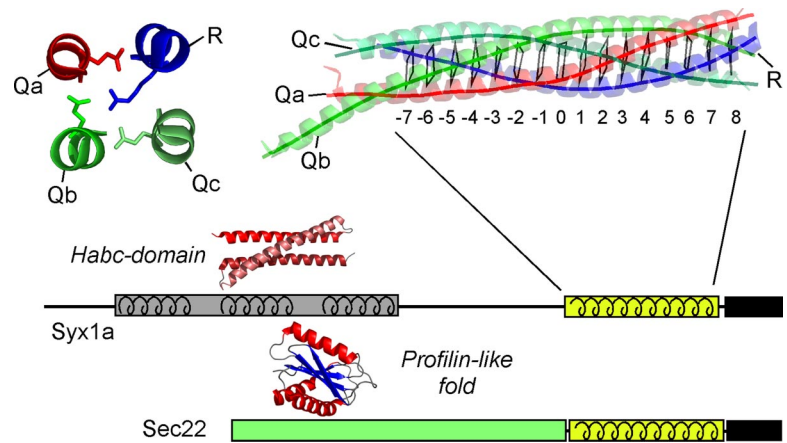
Emiliana huxleyi, had previously been reported (Wilson *et al.,* 2005). The other viral SNARE sequence was found in the genome of the mimivirus, which grows in the amoeba *Acanthamoeba polyphaga* (Raoult *et al.,* 2004). It encodes for a SNAP-25-like (25-kDa synaptosome-associated protein) protein containing two SNARE motifs (Qbc-SNARE).

Even though we were not aiming at finding new SNAREs in well-studied genomes, we identified four new SNARE proteins in vertebrate genomes. One of these new SNAREs (SNAP-47) is a Qbc-SNARE and has been described recently by our group (Holt *et al.,* 2006). The three others were Qa-SNAREs (often referred to as syntaxins [Syx]). We named them Syx20, a homolog of Syx7 (Qa.III.b), Syx19, and Syx21, the latter two

homologues of Syx11 (Qa.IV). Detailed information to all collected SNARE proteins can be found on our publicly accessible Web-based interface (http://bioinformatics.mpibpc.mpg.de/snare/).

Evidently, a high quality of the obtained data set is indispensable for discriminating the different types of SNARE proteins within a cell and for pinpointing evolutionary changes within the repertoires. An assessment of our generated HMM profiles showed that our classification is well suited to clearly separate nearly all true positives from false positives (Figure 2A). Furthermore, for each HMM profile we estimated the positive predictive rate (PPR), a value that gives the rate at which a positive is a true positive, and the sensitivity, a value

Figure 2. Quality assessment of the generated SNARE protein classification. (A) The absolute number of true positives is shown in dependence to the expectation value of the correct HMM profile of a search through the Feb. 2006 nr-database together with the rate of false positives found by an eye-by-eye verification of the results. The classification separates the largest part of the true positively found sequence motifs very well from the false positives. False positives are only found with rather large expectation values. The first false positive is found at an expectation value of 0.012. The false positive rate reaches the 1% margin at an expectation value of ~0.1 and the 5% margin at ~5. (B) To estimate the PPR (■), the rate at which a positive is a true positive, and the sensitivity (▨), the percentage of correctly found true positives, of the generated HMM profiles one need to know the true and false negatives of a search. Because this is almost impossible to know for a random database, we used a resampling method. We randomly gathered 95% of the SNARE motifs in each class to generate a new HMM profile. The 95% sampling assured each class to consist of at least 10 sequences and was also efficient to minimize variation within the generated motif. The other SNAREs were used as the search database. For a comparable expectation value we set the database size to a fixed size of one million sequences. We repeated the resampling 1000 times to minimize variation. We assumed the profile with the best expectation value to be the correct class. All 20 profiles achieved at least a 95% PPR and, besides one, a sensitivity of 97%. If we count a true positive as a true positive whenever it has at least achieved the second best expectation value, the PPR rises above the 99% margin for all profiles.
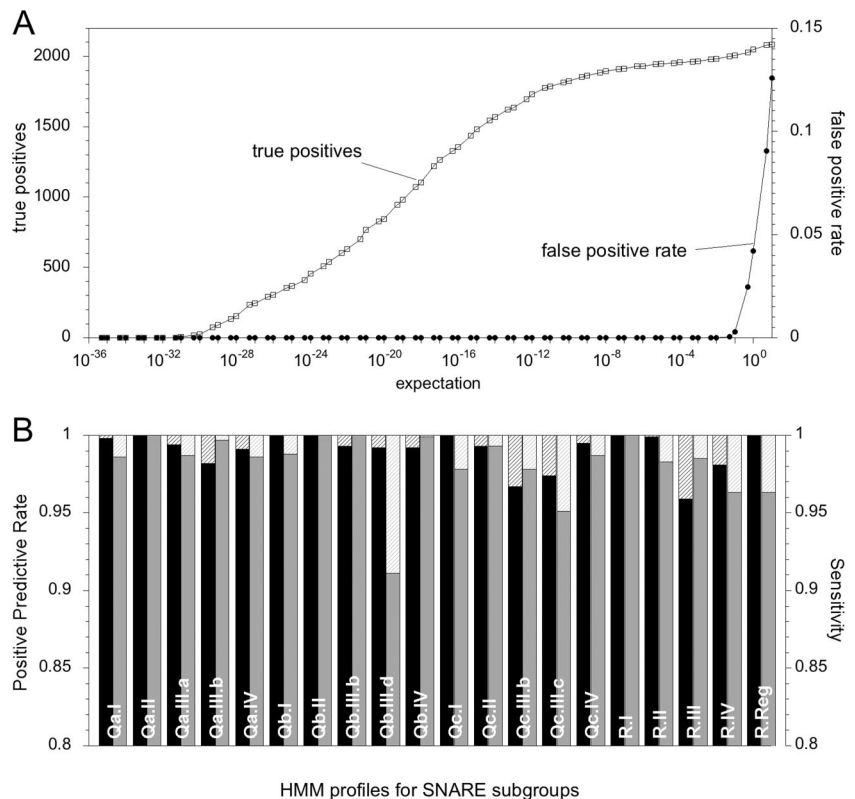
that describes the percentage of correctly found true positives, by a resampling approach. All profiles achieved at least 95% PPR and reached, with one exception, at least 95% sensitivity (Figure 2B). Profiles such as the Qa.II top out at a 100% PPR and 100% sensitivity. The only profile not matching our goal is the one of the Qb.III.d-class with a sensitivity of only 91%. As this group contains the smallest number of sequences, it might be that as new sequences become available, the sensitivity of this group improves above the 95% margin.

It is very likely these subgroups established by bioinformatic analysis correspond to functional orthologues, but the sequence information is not sufficient to pinpoint the exact transport step of the different SNARE types. With the available biological information, we have therefore tentatively assigned the 20 subgroups to the main routes of vesicle trafficking of a "typical" eukaryotic cell (Table 1). In addition, we have grouped them—following the structural QabcR rule—into basic SNARE units. We started by allocating the five Qa-SNARE subgroups, which are thought to denote the target organelle, to the major compartments. These compartments were numbered according to the arrangement of the secretory pathway: endoplasmic reticulum (ER; I), Golgi apparatus (II), *trans*-Golgi network (TGN; III.a), digestive endosomal compartments (III.b), and plasma membrane (IV). According to biological informa-

Table 1. Distribution of the basic SNARE repertoire within the cell and their allocation into functional units

	Qa group	Qb group	Qc group	R group
I. Endoplasmic reticulum (ER)	Qa.I m: *Syx18* f: *Ufe1* p: *Syp8*	Qb.I *Sec20*	Qc.I *Use1*	R.I *Sec22*
II. Golgi apparatus ER-Golgi	Qa.II m: *Syx5* f: *Sed5* p: *Syp3*	Qb.II m/p: *Membrin* f: *Bos1*	Qc.II *Bet1*	R.II *Ykt6*
Intra-Golgi		Qb.II m: *Gos28* f: *Gos1*	Qc.II m: *Gs15* f: *Sft1*	
III.a. *trans*-Golgi network (TGN)	Qa.III.a m: *Syx16* f: *Tlg2* p: *Syp4*			
III.b. Endosomal compartments Early	Qa.III.b m: *Syx7/13* f: *Pep12* p: *Syp2*	Qb.III.b *Vti1*	Qc.III.b m: *Syx6* f: *Tlg1* p: *Syp5/6*	R.III m/p: *Vamp7* f: *Nyv1*
Late (lysosome/vacuoles)			Qc.III.c m: *Syx8* f: *Syx8/Vam7* p: *Syp7*	
Plant/protist specific		Qb.III.d p: *Npsn*		
IV. Secretion (+ Cytokinesis/Sporulation)	Qa.IV m: *Syx1* f: *Sso* p: *Syp1*	Qb.IV (SNAP.b) 1. Helix of Qbc m: *SNAP-25* f: *Sec9*	Qc.IV (SNAP.c) 2. Helix of Qbc m: *SNAP-25* f: *Sec9*	R.IV m: *Syb1* f: *Snc1*
Regulatory				R.Reg *Tomosyn/Sro*

We classified SNARE proteins into four main groups and into 20 subgroups using phylogenetic analysis. Genuine complexes are composed of four different SNARE motifs each belonging to one of the four main groups ("QabcR" composition). Putative SNARE units have been assigned to the basic transport steps. In addition to the fusogenic SNARE proteins, a regulatory R-SNARE without membrane anchor, tomosyn, exists. Note that the two groups Qb.IV and Qc.IV are part of one protein, "Qbc-SNARE" and are connected by an extended linker region. Our analysis suggests that Qbc-proteins have arisen from a merger of two formerly independent Qb- and Qc-SNARE genes. The function of Npsn (Qb.III.d), which is restricted to plants and several protists, is not established so far. In the trees it is close to Vti1 (Qb.III.b) and the first helix of Qbc-SNAREs (Qb.IV). Hence, it is possible that it is involved in endosomal trafficking or secretion. The most commonly used names for the different SNARE types are given. For historical reasons, the names used for homologous SNAREs are often different in the different eukaryotic kingdoms. The different names used for metazoa (m), fungi (f), and plants (p) are listed. The names syntaxin and synaptobrevin (the secretory R-SNARE of metazoa that is also referred to as VAMP) are abbreviated as Syx and Syb, respectively. Several plant Q-SNAREs have been named Syntaxin of plants (Syp), yet in our databank we often did not make use of this somewhat confusing nomenclature. Moreover, several, more special names of the markedly increased SNARE repertoire of vertebrates are not listed.

tion, we assigned the other SNAREs to the trafficking steps toward the respective target compartments. For example, the SNARE set thought to be involved in retrograde transport from the Golgi toward the ER has been combined into group I (Qa.I: Syx18/Ufe1; Qb.I: Sec20; Qc.I: Use1; R.I: Sec22). The consecutive transport steps, from the ER to the Golgi and within the Golgi, are mediated by only one Qa.II-SNARE (Syx5/Sed5). The interacting Qb.II- and Qc.II-groups usually contain two factors each, suggesting that each factor participates only in one transport step. Furthermore, the interaction partners of the TGN-localized Qa.III.a-group (Syx16/Tlg2) are not established without doubt. As several SNAREs have been reported to engage in promiscuous interactions (Banfield, 2001; Pelham, 2001; Jahn and Scheller, 2006), it seems possible that Syx16/Tlg2 can interact with the SNAREs from the adjacent Golgi apparatus (II) and the endosomal compartment (III.b and III.c). It should also be kept in mind that it is still debated for some SNAREs in which trafficking step and which complex they participate. Hence, the groups listed in Table 1 might reflect the predominantly formed complexes, but interactions with SNAREs of neighboring compartments are possible as well.

Notably, the SNAREs of the secretory routes involving the ER (group I), the Golgi apparatus (group II), but also in the TGN (Syx16/Tlg2) rarely exhibited an increase in the number of the involved genes, even in multicellular organisms. The high conservation of these SNARE groups suggests that these trafficking routes in particular are highly preserved between all eukaryotes. In contrast, several SNARE proteins involved in endosomal trafficking seem to have adapted much more to the different demands and feeding types in

different eukaryotes. The original exocytotic SNARE set comprises only three SNARE proteins, a Qa-SNARE (Qa.IV), an R-SNARE (R.IV), and a SNAP-25-like protein that combines one Qb- and one Qc-motif in one protein (Qb.IV and Qc.IV, "Qbc-SNARE"). In many organisms, however, the number of secretory SNAREs has increased.

Almost all SNARE groups found are highly conserved throughout all inspected species. Several eukaryotes, in particular fungi and green algae, encompassed a relatively simple repertoire of SNARE, often consisting of only one member for each subgroup. This strongly suggests that the 20 subgroups in Table 1 represent the basic set of SNAREs of the eukaryotic ancestor. Changes in the SNARE repertoire in different eukaryotic organisms appear to have arisen by multiplications and diversifications within this original set. The high conservation of the basic SNARE groups, suggests that the earliest eukaryotes were probably already equipped with an elaborated endomembrane system with the main routes of vesicle trafficking established. Interestingly, due to genome duplications (Aury *et al.*, 2006) the unicellular ciliate *Paramecium tetraurelia* with its stunningly complex subcellular organization holds even more SNARE proteins (~70) than multicellular organisms like higher green plants (e.g., 62 in *Arabidopsis*) or vertebrates (e.g., 41 in *Homo*). This indicates that an increase in the number of SNARE proteins is not necessarily associated with the emergence of multicellularity. Interestingly, although the *Paramecium* sequences are often clearly more deviated, they still fitted into our classification scheme. Thus, it seems likely that the more complex membrane traf-

Figure 3. General outline of the phylogenetic tree of SNARE proteins. To gain insights into the evolutionary relationship of SNARE proteins, we constructed trees of the SNARE sets of 11 representative species and one larger tree containing the SNAREs of all 11 species. The IQPNNI-trees were generated as described in *Materials and Methods*. All original IQPNNI-trees can be downloaded from our webpage (http://bioinformatics.mpibpc.mpg.de/snare/). As an example, the tree of the 25 SNAREs of *Schistosoma japonicum* (ScJa) is shown. It shows the typical layout that we found for the species examined in our analysis. Basically, all SNAREs split into four well-supported groups that represent the four different positions of the four-helix bundle SNARE core complexes (see Figure 1). Each of the four main groups segregates into the distinct subgroups (see Table 1 for nomenclature) found by our classification. The labels on the tree edges represent the likelihood mapping (first) and AU support values (right). Interestingly the inner most split (Qa and Qc vs. R and Qc) is well supported. This clustering can be observed in all species trees calculated, except those of the two fungi species. Furthermore, the SNARE set involved in ER-transport, group (I), is quite diverged from the other SNARE groups and seems to have undergone a rather specific evolution. Note that the two SNARE motifs of Qbc-SNAREs are indicated by a B or C for the Qb- or the Qc-helix, respectively.
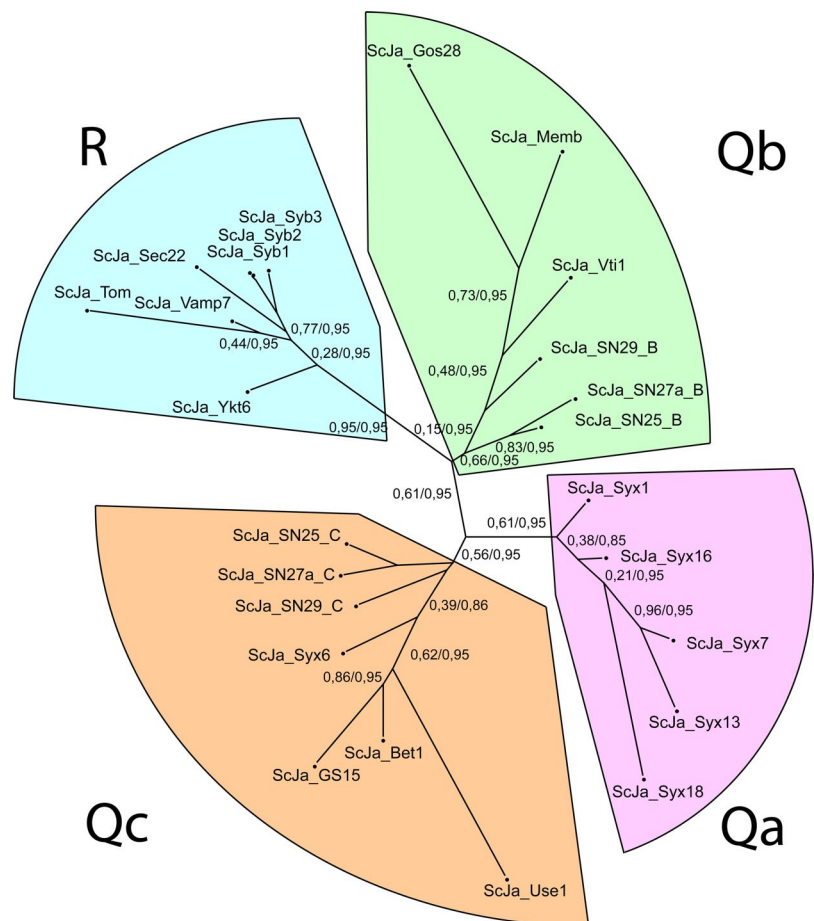
ficking pathways of *Paramecium* is build on the basic subcellular organization known from "typical" eukaryotic cells.

Finally, we constructed 11 phylogenetic trees, each containing the complete or nearly complete set of SNARES of a different representative eukaryotic species and one tree containing the SNARES of all 11 species. The tree obtained for the trematode *Schistosoma japonicum* is shown in Figure 3. The tree of this "lower" animal represents well the general outline of SNARE trees obtained for most eukaryotic species. All other trees can be found in the supplemental section on our Webpage (http://bioinformatics.mpibpc.mpg.de/snare/). These trees confirm the fundamental splitting of SNAREs into the four main branches Qa-, Qb-, Qc-, and R-SNAREs (Bock *et al.,* 2001). These main groups very likely reflect the principal structural arrangement of SNARE complexes, strengthening the view that all SNAREs assemble into canonical four-helix bundle structures. Each of the four fundamental branches segregated into distinct subgroups (Figure 3), corresponding to our HMM profiles (Table 1). The root of each subgroup branch usually consists of the SNAREs involved in secretion (group IV). This suggests, assuming that the different SNARE sets in present-day eukaryotes evolved by duplication and diversification of a prototypic set, that the secretory SNAREs appear to be relatively unchanged. In contrast the SNAREs involved in ER transport (group I) seem to have deviated more.

DISCUSSION

The conserved mechanism and structure of SNARE proteins has long been recognized (reviewed in Hong, 2005; Jahn and Scheller, 2006), but a comprehensive classification was lacking so far. Our first attempt to classify SNAREs using psi-BLAST together with hierarchical clustering resulted in a clearly inferior accuracy. A recently published survey of the SNARE family using the cluster approach aids this impression (Yoshizawa *et al.,* 2006), because several well-known SNAREs were not found and sequences were clustered together that may not be orthologues.

We systematically identified and classified SNAREs using HMM-profiles and a phylogenetic analysis. Because sequence orthology is a reliable predictor of function, our classification scheme provides valuable insights into the catalyzed transport steps. In addition, our classification can be used as a highly significant method to assign a distinct class to a de novo found SNARE sequence. As a highly beneficial tool for cell biology, we have implemented a Web-based interface to access our results and to query new sequences according to our classification.

SNAREs from Protists with More Derived Endomembrane Systems Fit Well into Our Classification

As most SNARE sequences in our study stem from genomes of animals, plants, and fungi, the current classification scheme is especially suited to classify SNAREs from these kingdoms. For several orthologous SNAREs, the tree also quite accurately reflects the phylogenetic relationship of species within these kingdoms. However, morphological and functional adaptations of trafficking routes and organelles, mostly of the endosomal and secretory pathways, also influenced the pattern of the SNARE tree. It has been proposed that the deepest eukaryotic divide places plants together with most protists on one side and fungi, animals, and some amoebae on the other side (Philippe *et al.,* 2000; Stechmann and Cavalier-Smith, 2002; Richards and Cavalier-Smith, 2005). If correct, comparisons of the SNARE sets of animals, fungi, and plants would be largely sufficient to outline the

bauplan of the secretory apparatus of all eukaryotes. Nevertheless, in the phylogenetic tree containing all 11 species (Supplementary Material), SNAREs from some protists are often isolated. We believe that this will change, as more sequences from other protists will be included in our classification. This phenomenon is visible already for the apicomplexa, for which several species were included (*Plasmodium, Theilleria,* and *Cryptosporidium*). Similarly, SNAREs of the kinetoplastid parasites *Trypanosoma brucei, T. cruzi,* and *Leishmania major* usually also congregated, reflecting their close relationship. In most protists with entire genomes available, we discovered a reasonable collection of SNARE proteins, supplementing the so far established repertoires (Dacks and Doolittle, 2002, 2004; Besteiro *et al.,* 2006; Schilde *et al.,* 2006; Yoshizawa *et al.,* 2006; Ayong *et al.,* 2007; Kissmehl *et al.,* 2007). For example, for *Plasmodium falciparum* we found 22 SNAREs with almost all different subgroups represented. For *L. major* we discovered 28 SNARE sequences, but so far we found only one type (Sec20, Qb.I) of the more deviated Q-SNAREs involved in ER-transport (group I). Before drawing conclusions, however, one has to keep in mind that, although several protist genomes have been completely sequenced, the available data still appear slightly fragmentary.

A number of contemporary anaerobic eukaryotes were until recently thought to primitively lack mitochondria and to possess a more primitive endomembrane system. For example, the parasitic protist *Giardia* was thought to lack a morphologically identifiable Golgi apparatus. Formerly, *Giardia* was thought to have diverged before the acquisition of such organelles. However, *Giardia* was found not only to contain rudimentary mitochondria (Embley and Martin, 2006), it also possess several key factors of membrane trafficking, among them SNARE proteins (Dacks and Doolittle, 2002, 2004; Marti *et al.,* 2003a,b). In fact, with only 15 detected SNAREs, *Giardia* has a clearly smaller set as compared with most other unicellular eukaryotic organisms. Principally, the SNAREs of *Giardia* are compatible with our classification; however, SNAREs belonging to group II, which is involved in Golgi transport, appear to lack. Because *Giardia* does not possess a conventional Golgi apparatus (Marti *et al.,* 2003a,b), a possible explanation is that an entire SNARE unit, and hence the vesicular trafficking route mediated by these SNAREs, has been lost in this organism. Another possibility, however, is that the group II SNARE proteins in *Giardia* are too derived to be detected by our HMM profiles.

In contrast to the compact set in *Giardia,* we found 31 SNAREs in the genome of *Entamoeba histolytica.* Remarkably, in this organism the Q-SNAREs involved in ER transport (group I) appear to lack. As the sequences of this SNARE group are often more derived, we again cannot rule out that our current HMM profiles do not detect these sequences in this organism. Because we found at least one member of the group I SNAREs in all major groups of eukaryotes that are represented by the current data set, it is likely that this vesicular trafficking step was also present in the common ancestor.

Thus, although the SNAREs of several different protozoan parasites often form long branches in our trees, our analysis does not support the notion that their endomembrane system is more primitive but rather might be a degenerate feature of their parasitic lifestyle. Nevertheless, because of sparse biological data for these protist groups, our classification of protist SNAREs ought to be verified in the future.

Evolution of the SNARE Apparatus Is Linked to the Development of the Endomembrane System

Our analysis suggests that the 20 basic SNARE subgroups discriminated by our HMM profiles represent the repertoire of SNARE proteins of the eukaryotic ancestor. Almost certainly, in the eukaryotic ancestor these 20 SNARE types operated already in vesicular trafficking steps between the major intracellular compartments that are conserved in almost all contemporary eukaryotes. It has been proposed that the different intracellular compartments of a eukaryotic cell, along with the molecular machineries involved in vesicular trafficking, emerged by events of duplication and diversification of a simpler endomembrane system of a more primitive ancestor (Roger, 1999; Cavalier-Smith, 2002). Possibly, the development of the 20 basic SNARE types is closely intertwined with the development of different intracellular compartments. In fact, it had been discussed earlier that the machineries involved in vesicular trafficking appear not only to be conserved through phylogeny of species but also throughout the different compartments of the cell (Bock *et al.,* 2001), although the species included in this study by far did not represent the entire eukaryotic diversity. In subsequent studies, the presence of different conserved syntaxin subfamilies (i.e., mostly Qa-SNAREs) in several diverse eukaryotes supported the notion that these SNAREs diverged early in eukaryotic evolution (Dacks and Doolittle, 2002, 2004). The early diversification of SNARE proteins is also corroborated by the fact that other, distantly related eukaryotes encompass relatively conserved SNARE sets (Besteiro *et al.,* 2006; Schilde *et al.,* 2006; Yoshizawa *et al.,* 2006; Ayong *et al.,* 2007; Kissmehl *et al.,* 2007; and our analysis). Notably, our study substantiates the view that all SNARE proteins principally split into the four major phylogenetic classes: Qa, Qb, Qc, and R (Bock *et al.,* 2001), suggesting that the prototypic unit was composed of four different SNARE proteins, able to assemble into a tight four-helix bundle between two fusing membranes. Thus, it is likely that during the early stages of eukaryotic evolution entire SNARE units rather than the single subunits were duplicated. These prototypic SNARE units diverged afterward. In fact, we detected some patterns of coevolution in the different SNARE units, in particular in the Q-SNAREs of group I, but promiscuous interactions between different SNARE subunits may have precluded further diversifications. The phylogenetic trees obtained from SNARE proteins of different eukaryotic species (Figure 3 and Supplementary Material), together with the relative simple domain architecture of the SNAREs (Figure 1), allow for some tentative speculations about the nature of a prototypic SNARE machinery: Because all ancestral R-SNARE types appear to contain an N-terminal domain with a profilin-like fold (Figure 1; Gonzalez *et al.,* 2001; Tochio *et al.,* 2001; Wen *et al.,* 2006), they may have originated from a common ancestor that contained this N-terminal extension. Note that this domain has been lost in the secretory R.IV-SNAREs in fungi and animals. Similarly, all Qa-SNAREs exhibit a very similar domain architecture, carrying an N-terminal three-helix bundle structure (Fernandez *et al.,* 1998; Lerman *et al.,* 2000; Misura *et al.,* 2000; Munson *et al.,* 2000; Dulubova *et al.,* 2001). Interestingly, several of the Qb- and the Qc-SNAREs possess an N-terminal three-helix bundle as well (Antonin *et al.,* 2002a; Misura *et al.,* 2002; Fridmann-Sirkis *et al.,* 2006). This hints at a common origin of all three main Q-SNARE groups from a prototypic Q-SNARE. Hence, a scenario can be envisioned in which originally a trimer of a prototypic Q-SNAREs on the target membrane interacted with a prototypic R-SNARE on the vesicular membrane. However, this partition between the two membranes is so far only established for the secretory SNAREs, whereas in other trafficking steps the distribution of the four SNARE subunits is still heavily debated. Therefore, it is challenging to bring our data into line with biological knowledge. For example, a different notion about the prototypic SNARE unit might be supported by the fact that in the species trees (e.g., Figure 3) the four basic SNARE groups partition into two elementary groups, one containing the R- and Qb-SNAREs and one containing the Qa- and Qc-SNAREs. These splitting can be observed in species trees of animals, protists, and plants. Interestingly, the two main branches each unite the two diagonally opposite helices of the four-helix bundle SNARE complex. Notably, these two main branches in general exhibit a comparable topology of the subgroup splits, possibly reflecting coevolution.

The two trees obtained from fungi species show a different splitting, but can be brought into line as can be seen from the tree containing all eleven species (data set in Supplementary Materials). In this tree the support of the inner splitting is somewhat decreased compared with a tree without the fungi species (data not shown). A more detailed phylogenetic analysis of the evolution of SNAREs within the different eukaryotic kingdoms seems to be of interest but is beyond the scope of this article.

Analogous themes of duplication and divergence for other important factors involved in intracellular membrane trafficking have been exposed and implicated in the evolution of the eukaryotic endomembrane system, for example, for the organelle-specific Rab GTPases (Pereira-Leal and Seabra, 2001; Jekely, 2003) and tethering factors involved in vesicular trafficking (Koumandou *et al.,* 2007). Similarly, coat protein–based budding of transport vesicles is mediated by different but homologous protein machineries at different donor organelles: COPI, COPII, and clathrin (McMahon and Mills, 2004). Related protein machineries have been even implicated in the establishment of the nuclear envelope (Devos *et al.,* 2004; Mans *et al.,* 2004) and the eukaryotic cilium (Jekely and Arendt, 2006). Very likely these factors evolved from a prototypic unit that was able to generate areas of highly curved membranes. Together, the development of such prototypic protein machineries provided the raw material for the intricate evolution of the endomembrane system of the eukaryotic cell.

REFERENCES

Antonin, W., Dulubova, I., Arac, D., Pabst, S., Plitzner, J., Rizo, J., and Jahn, R. (2002a). The N-terminal domains of syntaxin 7 and vti1b form three-helix bundles that differ in their ability to regulate SNARE complex assembly. J. Biol. Chem. *277,* 36449–36456.

Antonin, W., Fasshauer, D., Becker, S., Jahn, R., and Schneider, T. R. (2002b). Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. Nat. Struct. Biol. *9,* 107–111.

Aury, J. M. *et al.* (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia.* Nature *444,* 171–178.

Ayong, L., Pagnotti, G., Tobon, A. B., and Chakrabarti, D. (2007). Identification of *Plasmodium falciparum* family of SNAREs. Mol. Biochem. Parasitol. *152,* 113–122.

Banfield, D. K. (2001). SNARE complexes—is there sufficient complexity for vesicle targeting specificity? Trends Biochem. Sci. *26,* 67–68.

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

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.

Bonifacino, J. S., and Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. Cell *116*, 153–166.

Burri, L., Varlamov, O., Doege, C. A., Hofmann, K., Beilharz, T., Rothman, J. E., Sollner, T. H., and Lithgow, T. (2003). A SNARE required for retrograde transport to the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA *100*, 9873–9877.

Cavalier-Smith, T. (2002). The phagotrophic origin of eukaryotes and phylogenetic classification of *Protozoa.* Int. J. Syst. Evol. Microbiol. *52*, 297–354.

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., and Doolittle, W. F. (2004). Molecular and phylogenetic characterization of syntaxin genes from parasitic protozoa. Mol. Biochem. Parasitol. *136*, 123–136.

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*, e380.

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*, 3664–3674.

Dulubova, I., Yamaguchi, T., Wang, Y., Sudhof, T. C., and Rizo, J. (2001). Vam3p structure reveals conserved and divergent properties of syntaxins. Nat. Struct. Biol. *8*, 258–264.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). The theory behind profile HMMs. New York: Cambridge University Press.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Embley, T. M., and Martin, W. (2006). Eukaryotic evolution, changes and challenges. Nature *440*, 623–630.

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. (1989). PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics *5*, 164–166.

Fernandez, I., Ubach, J., Dulubova, I., Zhang, X., Sudhof, T. C., and Rizo, J. (1998). Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. Cell *94*, 841–849.

Fridmann-Sirkis, Y., Kent, H. M., Lewis, M. J., Evans, P. R., and Pelham, H. R. (2006). Structural analysis of the interaction between the SNARE Tlg1 and Vps51. Traffic *7*, 182–190.

Gonzalez, L. C., Jr., Weis, W. I., and Scheller, R. H. (2001). A novel snare N-terminal domain revealed by the crystal structure of Sec22b. J. Biol. Chem. *276*, 24203–24211.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. *52*, 696–704.

Gupta, G. D., and Brent Heath, I. (2002). Predicting the distribution, conservation, and functions of SNAREs and related proteins in fungi. Fungal Genet. Biol. *36*, 1–21.

Holt, M., Varoqueaux, F., Wiederhold, K., Takamori, S., Urlaub, H., Fasshauer, D., and Jahn, R. (2006). Identification of SNAP-47, a novel Qbc-SNARE with ubiquitous expression. J. Biol. Chem. *281*, 17076–17083.

Hong, W. (2005). SNAREs and traffic. Biochim. Biophys. Acta *1744*, 493–517.

Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.

Jahn, R., and Grubmuller, H. (2002). Membrane fusion. Curr. Opin. Cell Biol. *14*, 488–495.

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

Jekely, G. (2003). Small GTPases and the evolution of the eukaryotic cell. Bioessays *25*, 1129–1138.

Jekely, G., and Arendt, D. (2006). Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. Bioessays *28*, 191–198.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. *8*, 275–282.

Kissmehl, R., Schilde, C., Wassmer, T., Danzer, C., Nuehse, K., Lutter, K., and Plattner, H. (2007). Molecular identification of 26 syntaxin genes and their assignment to the different trafficking pathways in *Paramecium.* Traffic *8*, 523–542.

Koumandou, V. L., Dacks, J. B., Coulson, R. M., and Field, M. C. (2007). Control systems for membrane fusion in the ancestral eukaryote; evolution of tethering complexes and SM proteins. BMC Evol. Biol. *7*, 29.

Lerman, J. C., Robblee, J., Fairman, R., and Hughson, F. M. (2000). Structural analysis of the neuronal SNARE protein syntaxin-1A. Biochemistry *39*, 8470–8479.

Lewis, M. J., and Pelham, H. R. (2002). A new yeast endosomal SNARE related to mammalian syntaxin 8. Traffic *3*, 922–929.

Lewis, M. J., Rayner, J. C., and Pelham, H. R. (1997). A novel SNARE complex implicated in vesicle fusion with the endoplasmic reticulum. EMBO J. *16*, 3017–3024.

Maddison, D. R., Swofford, D. L., and Maddison, W. P. (1997). NEXUS: an extensible file format for systematic information. Syst. Biol. *46*, 590–621.

Mans, B. J., Anantharaman, V., Aravind, L., and Koonin, E. V. (2004). Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. Cell Cycle *3*, 1612–1637.

Marti, M., Li, Y., Schraner, E. M., Wild, P., Kohler, P., and Hehl, A. B. (2003a). The secretory apparatus of an ancient eukaryote: protein sorting to separate export pathways occurs before formation of transient Golgi-like compartments. Mol. Biol. Cell *14*, 1433–1447.

Marti, M., Regos, A., Li, Y., Schraner, E. M., Wild, P., Muller, N., Knopf, L. G., and Hehl, A. B. (2003b). An ancestral secretory apparatus in the protozoan parasite *Giardia intestinalis.* J. Biol. Chem. *278*, 24837–24848.

McMahon, H. T., and Mills, I. G. (2004). COP and clathrin-coated vesicle budding: different pathways, common approaches. Curr. Opin. Cell Biol. *16*, 379–391.

Misura, K. M., Bock, J. B., Gonzalez, L. C., Jr., Scheller, R. H., and Weis, W. I. (2002). Three-dimensional structure of the amino-terminal domain of syntaxin 6, a SNAP-25 C homolog. Proc. Natl. Acad. Sci. USA *99*, 9184–9189.

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.

Munson, M., Chen, X., Cocina, A. E., Schultz, S. M., and Hughson, F. M. (2000). Interactions within the yeast t-SNARE Sso1p that control SNARE complex assembly. Nat. Struct. Biol. *7*, 894–902.

Pelham, H. R. (2001). SNAREs and the specificity of membrane fusion. Trends Cell Biol. *11*, 99–101.

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.

Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., and Le Guyader, H. (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc. Biol. Sci. *267*, 1213–1221.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004). The 1.2-megabase genome sequence of Mimivirus. Science *306*, 1344–1350.

Richards, T. A., and Cavalier-Smith, T. (2005). Myosin domain evolution and the primary divergence of eukaryotes. Nature *436*, 1113–1118.

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

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.

Schilde, C., Wassmer, T., Mansfeld, J., Plattner, H., and Kissmehl, R. (2006). A multigene family encoding R-SNAREs in the ciliate Paramecium tetraurelia. Traffic *7*, 440–455.

Schmidt, H. A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502–504.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.

Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.

Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. Science *297*, 89–91.

Strimmer, K., and von Haeseler, A. (1997). Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. USA *94*, 6815–6819.

Sutter, J. U., Campanoni, P., Blatt, M. R., and Paneque, M. (2006). Setting SNAREs in a different wood. Traffic *7*, 627–638.

Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. Nature *395*, 347–353.

Tochio, H., Tsui, M. M., Banfield, D. K., and Zhang, M. (2001). An autoinhibitory mechanism for nonsyntaxin SNARE proteins revealed by the structure of Ykt6p. Science *293*, 698–702.

Uemura, T., Ueda, T., Ohniwa, R. L., Nakano, A., Takeyasu, K., and Sato, M. H. (2004). Systematic analysis of SNARE molecules in *Arabidopsis*: dissection of the post-Golgi network in plant cells. Cell Struct. Funct. *29*, 49–65.

Vinh le, S., and Von Haeseler, A. (2004). IQPNNI: moving fast through tree space and stopping in time. Mol. Biol. Evol. *21*, 1565–1571.

Weimbs, T., Low, S. H., Chapin, S. J., Mostov, K. E., Bucher, P., and Hofmann, K. (1997). A conserved domain is present in different families of vesicular fusion proteins: a new superfamily. Proc. Natl. Acad. Sci. USA *94*, 3046–3051.

Wen, W., Chen, L., Wu, H., Sun, X., Zhang, M., and Banfield, D. K. (2006). Identification of the yeast R-SNARE Nyv1p as a novel longin domain-containing protein. Mol. Biol. Cell *17*, 4282–4299.

Wilson, W. H. *et al.* (2005). Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science *309*, 1090–1092.

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.

Zwilling, D., Cypionka, A., Pohl, W. H., Fasshauer, D., Walla, P. J., Wahl, M. C., and Jahn, R. (2007). Early endosomal SNAREs form a structurally conserved SNARE complex and fuse liposomes with multiple topologies. EMBO J. *26*, 9–18.