Chapter 14

Eukaryotic Cell Evolution from a Comparative Genomic Perspective: The Endomembrane System

Joel B. Dacks and Mark C. Field

CONTENTS

Abstract 309
14.1 Introduction 310
14.2 The Endomembrane System 311
14.2.2 Organelles of the Endomembrane System 311
14.2.3 Steps in the Transport Reaction 313
14.2.3.1 Vesicle Budding 313
14.2.3.2 Vesicle Translocation 314
14.2.3.3 Vesicle Fusion 316
14.2.3.4 Regulation and Specificity 316
14.2.4 Vesicular Transport in Non-Opisthohont Lineages 316
14.3 Mining the Databases 319
14.3.1 Candidate Proteins To Be Used as Representative Queries 319
14.3.2 Search Methods 319
14.4 Endomembrane System Component Homologues in Diverse Genomes 320
14.4.1 Bottom Up: Prokaryotic Homologues of Endomembrane System Components 320
14.4.2 Top Down: Reconstructing the Vesicular Transport Machinery of the Last Common Eukaryotic Ancestor 323
14.5 Beyond BLAST: Examples from Functional Studies 326
14.5.1 Genomics and Phylogeny 326
14.5.2 Genomics and Cell Biology 327
14.6 Conclusions 327
14.7 Materials and Methods 328
14.7.2 Search Methods 328
14.7.3 Databases 328
Acknowledgments 329
References 329

Abstract

Comparative genomics provides a powerful tool for both evolutionary and cellular biology. As an example of how comparative genomics can be used in these fields, we examine how

0-415-29904-7/04/$0.00+$1.50

cell biological studies in model systems, together with the rapidly accumulating genomic data from diverse taxa, can be applied to reconstruct complex aspects of the biosynthetic-secretory and endocytic pathways in eukaryotes. The near-universal presence within eukaryotes of the core features of an intracellular transport system serves to highlight the vital role that this elaborate system must play in cell function. The evolution of this system is non-obvious, as prokaryotes have been generally considered to lack primitive or precursor structures that could have given rise to an endomembrane system. We consider, in detail, the proteins involved in vesicular transport, emphasizing a number of insights from selected divergent systems and comparing these with crown eukaryotes. We highlight possible prokaryotic precursors, survey the eukaryotic diversity of vesicular transport machinery and discuss how genomics initiatives have helped push forward cell biological studies of the endomembrane system in diverse organisms. Importantly, the mechanistic details of the transport systems are essentially conserved, indicating an ancient origin for these processes. All the while, increasing complexity in the sense of pathway multiplicity is observed in the vesicular transport system when comparing unicellular eukaryotes to more complex multicellular organisms.

14.1 Introduction

One of the most profound divisions in the biological world is between eukaryotic and non-eukaryotic cells. Prokaryotic organisms exhibit a huge diversity in biochemical and metabolic processes, but are underpinned with a comparatively simple cellular structure. Eukaryotes, in contrast, have massively expanded structural diversity and complexity in cell biological systems. Features such as the membrane-bound nucleus, the cytoskeleton, mitochondria, plastids and a system of functionally connected membrane-bound compartments (collectively referred to as the endomembrane system) are among the major characteristics that set eukaryotes apart from prokaryotes. This division is bridgeable though. Sophisticated homology searches and structural examination have identified prokaryotic homologues for proteins once thought to be strictly eukaryotic (Addinall and Holland, 2002; Kasinsky et al., 2001; van den Ent et al., 2001). Eukaryotes are also not as uniform in their cellular organization as once imagined. Organelles such as mitochondria, peroxisomes and stacked Golgi complexes have been lost or transformed many times in the course of eukaryotic history (Roger, 1999; see also Chapter 2), whereas plastid evolution is an even more sordid tale of theft, kidnapping and metamorphosis (Delwiche, 1999; see also Chapter 3 and Chapter 4). A broad comparative approach across a wide range of taxa is therefore key to making any generalizations about eukaryotic cell biology or evolution.

Two major advances have made the study of eukaryotic cell evolution more tractable. The first is the increasing wealth of molecular information about eukaryote-specific features. For example, it is no longer the cytoskeleton or even microtubules that are the defining characteristic, but the microtubule proteins (tubulins), and their genes, that can be compared (Addinall and Holland, 2002). This is important, because it facilitates a more objective analysis rather than a dependence on morphology or pharmacology, which cannot be quantified accurately in terms of genetic distance or functional divergence. Additionally, significant advances in molecular cell biology have identified many of the gene products involved in meiosis, chromatin organization and the endomembrane system, among other functions. The second advance is genomics, particularly comparative genomics. Prokaryotic genomes are being released at a tremendous rate, and eukaryotic genome initiatives are becoming more common whether as draft or full genomes, expressed sequence tag (EST) or genome sequence survey (GSS) projects. These sequences, organized and annotated into databases, will offer up gene sequences useful for addressing all aspects of eukaryotic evolution. Many of the advances made in molecular cell biology are because of progress in the genome projects of models systems such as mice, *Caenorhabditis*, *Drosophila* and yeasts.

It is likely that eukaryotes emerged from a single common prokaryote-like ancestor. A comparative genomics examination of the transition then from that prokaryote-like state and the emergence of the cellular systems that define eukaryotes can be approached in a number of ways. Prokaryotic homologues of proteins thought to be uniquely eukaryotic can provide hints as to a system's origin. A survey of the components of a system present in a wide diversity of eukaryotes allows an estimate of the complexity already established in the last common eukaryotic ancestor, as well as opens the door to more detailed questions of evolution and function in that system. Such an examination, however, requires methods for searching the genomic databases, candidate proteins to search for and a cell biological system to investigate.

14.2 The Endomembrane System

14.2.1 Evolution of an Endomembrane System: An Important Transition

One of the features that most distinguishes eukaryotes from prokaryotes is the assemblage of internal membrane-bound compartments for protein trafficking that constitutes the biosynthetic-secretory and endocytic pathways. This organellar system sorts, modifies, transports and even captures material (Alberts et al., 1994). Evolving this endomembrane system would have been a crucial step in the transition from the prokaryotic to the eukaryotic condition. The typical eukaryotic cell is 10- to 30-fold larger in linear dimensions than that of a prokaryote, with a consequent volume increase of 10^3- to 10^4-fold, and the result that simple diffusion of macromolecules through the cytoplasm is too slow for biological processes. An efficient transport system allows for increased cellular size, making available novel ecological niches. Endocytosis allows for more efficient heterotrophy and sets the stage for the acquisition of mitochondria and plastids. Targeted protein transport confers the ability to modify and control the composition of the cell surface, which is likely to have been an essential aspect of constructing a complex multicellular state. The evolution of the endomembrane system has been proposed to be the key step in the evolution of eukaryotes (Stanier, 1970).

In the past 5 to 10 years, there has been a huge increase in molecular data accumulated on the endomembrane system and membrane trafficking from cell biological studies. The picture that emerges is of a highly complicated and dynamic network of assembling and disassembling protein complexes required for the transport of material from one compartment to another. Although most of the molecular data have been obtained from animal and fungal model systems (Jahn and Sudhof, 1999), more limited studies in selected organisms will also be outlined.

14.2.2 Organelles of the Endomembrane System

The endoplasmic reticulum (ER), which is contiguous with the nuclear envelope (Figure 14.1), can be considered to be the beginning of the endomembrane system, on account of this being the point of insertion of polypeptides into the secretory pathway. Rough ER (rER) has a studded appearance, because of bound ribosomes, and is the site of synthesis for proteins destined to travel via vesicular transport. At the very heart of this transport process lies the mechanism by which polypeptides are translocated across the ER membrane (cotranslational translocation). The ribosome is targeted to the ER membrane via the signal sequence on the nascent chain of the polypeptide being translated. The information in the signal sequence is read by the signal recognition particle (SRP), a cytoplasmic riboprotein complex consisting of six polypeptides together with a core RNA (Rapoport et al., 1996). Transport vesicles bud from ribosome-free regions of the rER, called transitional elements (Klumperman, 2000), and quickly fuse with other vesicles derived from the same source. They might also fuse with a network of tubules termed the vesicular-tubular compartment (VTC) or ER

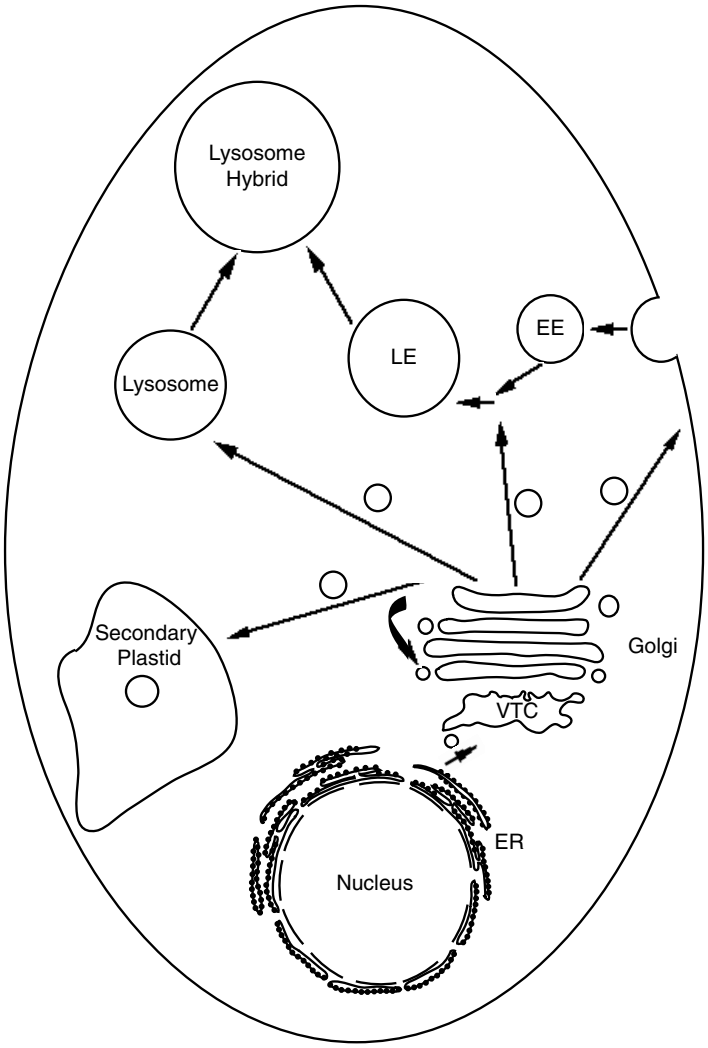

FIGURE 14.1 Organelles and direction of vesicular transport in a hypothetical eukaryotic cell. Straight arrows in this cartoon show the anterograde movement of vesicles between membrane compartments. The curved arrow illustrates retrograde transport in this case alone, as retrograde transport is the only non-controversial function of COPI vesicles. Small circles represent transport vesicles and large ones represent digestive organelles. EE and LE denote early and late endosomes, respectively. ER = endoplasmic reticulum, VTC = vesicular-tubular compartment. Much of the information shown here was derived from studies on mammalian and yeast cells, but the movement of Golgi-derived vesicles to a secondary plastid (as in *Euglena*) is also depicted.

Golgi intermediate compartment (ERGIC). Transport intermediates migrate from the VTC to a complex network of structures at the *cis*-face of the Golgi complex, referred to as the *cis*-Golgi network.

Most familiar as parallel stacks of flattened membrane-bound compartments (cisternae), the Golgi apparatus is the next distinct organelle in the endomembrane system. The Golgi apparatus receives material from the ER and is responsible for modification of proteins and sorting for later transport to various organelles. The compartments receiving material from

the ER are called *cis*-Golgi, the main portions of the stack are the medial-Golgi and stacks that subsequently receive material are termed the *trans*-Golgi cisternae. From these last cisternae, vesicles bud for further transport from an elaborate network of membranes termed the *trans*-Golgi network (TGN). The morphology of the Golgi complex is quite varied among eukaryotes, with distinct flattened stacks in animals, plants and many protozoa; punctate vesicles in most fungi (but not chytrids); and smaller but numerous stacks in algae (Becker and Melkonian, 1996). This structural variation does not necessarily correlate with phylogeny. Although, in general, higher plants tend to have large numbers of small stacks, the organization, number and location is different in comparatively closely related yeasts [*Saccharomyces cerevisiae* and *Picia pastoris* (Glick, 2000)]. Metazoans have typically a single contiguous Golgi ribbon (Shorter and Warren, 2002); however, *Drosophila* can exhibit distinct Golgi morphologies in different life stages (Stanley et al., 1997). Given this diversity, a definition that depends on function and not classical stacked morphology is appropriate.

The progression of material becomes less linear on exiting the Golgi stack (Figure 14.1). In mammals and yeast, vesicles emerge from the TGN and might travel in at least four possible directions: retrograde, i.e., backward to previous compartments within the Golgi or to the ER; anterograde to the plasma membrane; intersect with the endocytic pathway; or be targeted to the lysosome. Each pathway is accompanied by distinct protein factors responsible for sorting, targeting and transporting the vesicle (Pryer et al., 1992). The secondary plastids of some organisms, such as *Euglena*, also receive Golgi-derived vesicles (Sulli and Schwartzbach, 1995; Sulli et al., 1999).

The plasma membrane represents the end point of the secretion–biosynthetic pathway and the beginning of the endocytic system. Vesicles leaving the TGN for the plasma membrane travel to the surface, where they fuse, either releasing their soluble contents or presenting their membrane-bound cargo. At the plasma membrane, endocytic vesicles are created to entrap food, internalize ligand-bound cell surface receptors or take up fluid phase material (Figure 14.1). Vesicles derived from the TGN destined for intracellular compartments fuse either with endocytic vesicles derived from the plasma membrane, called early or sorting endosomes, or with a preexisting late endosome (Figure 14.1). These pathways seem to differ in the components required for vesicle budding from the TGN, but share much of the same machinery for fusion at the lysosome (Bryant and Stevens, 1998; Luzio et al., 2000). The late endosome then fuses with lysosomes to create a hydrolytic organelle involved in degradation. This late endosome–lysosomal compartment is considered by some authors as a separate organelle, the term *lysosome* being reserved for the organelle that contains concentrated hydrolytic enzymes (Figure 14.1; Luzio et al., 2000).

14.2.3 Steps in the Transport Reaction

Regardless of the donating and receiving organelles, the mechanistic process of vesicular transport has many shared features and can therefore be described in a generalized model with three basic steps: vesicle formation and budding from the donor organelle, vesicle movement and finally fusion of the vesicle with the target organelle. The machinery used for vesicular transport between the different organelles is a mixture of components common to a reaction, regardless of location; members of protein families with paralogues specific for transport between two given organelles; and organellar specific complexes.

14.2.3.1 Vesicle Budding

The process begins by recruitment of a small GTPase to the cytosolic side of the membrane at the site of vesicle formation. Initially, the GTPase is GDP-bound, but a guanine exchange factor (GEF) protein catalyzes exchange of GDP for GTP. The GTPase regulates vesicle

formation by recruiting the cytosolic coat proteins required for vesicle budding. Cargo proteins to be transported by the vesicle can be packaged by bulk flow, direct interaction with coat proteins via amino acid motifs in the cargo or via adaptor proteins. After cargo selection, the protein coat polymerizes, deforming the membrane and the ensuing vesicle buds. Figure 14.2 illustrates this generalized model (Springer et al., 1999).

Although the well-characterized types of vesicles built within the cell all conform to the generalized model of vesicle formation, their protein components differ significantly. In anterograde ER to Golgi transport, vesicles are coated with a complex COPII (Kaiser and Ferro-Novick, 1998; Springer et al., 1999). In the creation of COPII vesicles, the GTPase is called Sar1, which binds to the cytosolic face of the ER, with Sec12 acting as its GEF. The Sec23/24 protein complex interacts with the target membrane, in part through Sar1 (the complex is a regulator of the GTPase), and probably also through the cargo and putative cargo receptors. Cargo is concentrated into these exit regions and incorporated into transport vesicles, presumably via retention though proteins of the Emp24 family (Muniz et al., 2000) and also through bulk flow (Klumperman, 2000). The Sec23/24 complex recruits the Sec13/31 coat complex, which, by self-assembly, acts to coat the cytoplasmic surface of the nascent vesicle.

COPI vesicles recycle material from the Golgi apparatus back to the ER. In formation of the COP I complex, a distant paralogue to Sar1, called Arf, binds to the cytosolic portion of the membrane in GDP bound form. The nucleotide diphosphate is then exchanged for a GTP moiety by an Arf GEF in a manner similar to Sar1 activation. Membrane-bound cargo might interact with a preassembled vesicle coat-forming coatomer complex. Coatomer, Arf and ArfGAP (GAP for GTPase-activating protein) complex to form the polymeric coat and vesicle budding occurs (Springer et al., 1999). The formation of COPI vesicles is clearly involved in retrograde transport from the Golgi back to the ER, but it is probable that it is also important in anterograde transport within the Golgi stack (Orci et al., 1997; Schekman and Mellman, 1997).

Many of the remaining vesicles formed in the cell are coated with clathrin or clathrin-related proteins and include vesicles destined for both the endosome and the plasma membrane. In the formation of clathrin-coated vesicles, Arf also acts as the GTPase, with an Arf GEF again providing the GTP exchange (Kirchhausen, 2000; Springer et al., 1999). Heterotetrameric adaptin complexes (AP) bind cargo and provide specificity for particular organellar destinations. AP1 and AP3 at the TGN are involved in targeting material to the late endosome and lysosome, respectively. AP2 is involved in cargo selection for plasma-membrane-derived vesicles entering the endocytic pathway (Kirchhausen, 2000; Robinson and Bonifacino, 2001). Adaptors bind via *cis*-acting amino acid motifs in the cargo or via additional adaptor proteins, such as the mannose-6-phosphate receptors in mammals for the transport of material to the late endosome (Kirchhausen, 2000). Clathrin itself acts as the protein coat, polymerizing and forming the vesicle. In the case of AP3, clathrin is not involved but VPS41 acts as the protein coat polymer and appears to have a domain homologous to clathrin (Kirchhausen, 2000; Robinson and Bonifacino, 2001). Additional coat systems have also been reported, e.g., the GGA family (Golgi-associated, gamma adaptin ear-containing Arf-binding protein). They are also important in mannose-6-phosphate receptor trafficking (Doray et al., 2002) but, at present, it is unclear how these are recruited or function in detail.

14.2.3.2 Vesicle Translocation

After vesicle formation and budding, the vesicle is transported to its eventual target. At some point after vesicle formation, the GTP on the GTPase is hydrolyzed back to GDP via the action of an ArfGAP homologue. The role of this hydrolysis is unclear, although it has

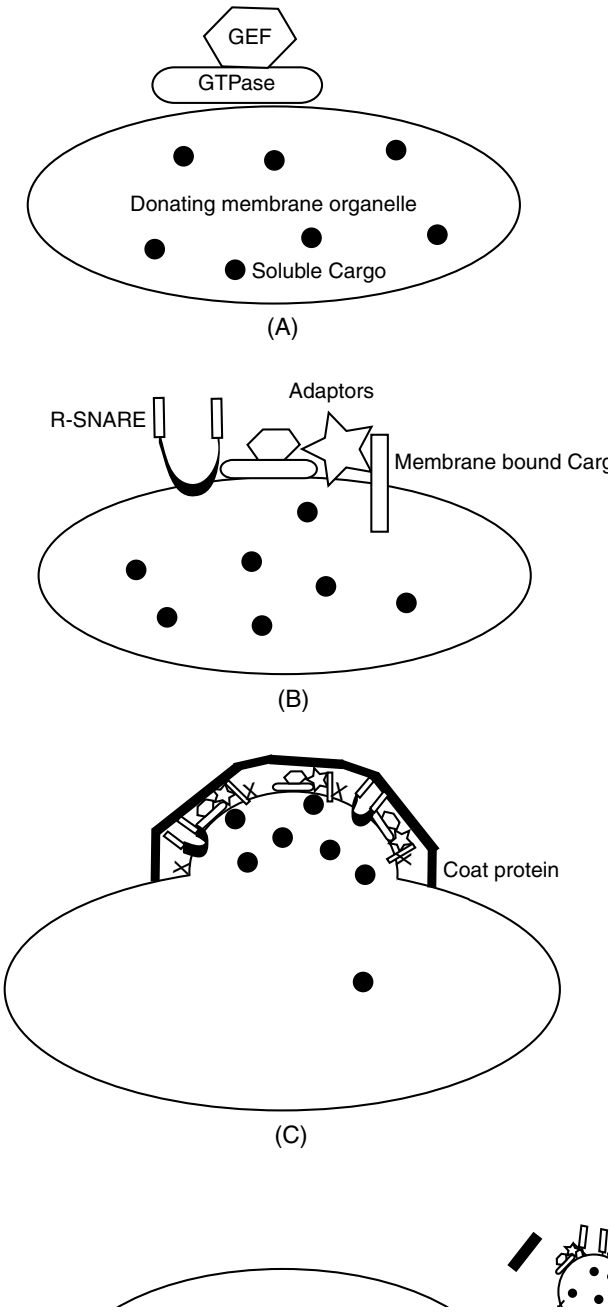

FIGURE 14.2 Generalized cartoon of vesicle formation, budding and movement. (A) GTPase attaches to the membrane and a GEF swaps GDP for GTP. (B) Adaptor proteins and cargo attach to the nucleating site of vesicle formation. (C) Coat proteins arrive and form a scaffolding complex for vesicle formation. Soluble cargo might be incorporated into the vesicle via adaptors or by bulk flow. (D) Vesicle has budded away from the donating membrane, a GAP hydrolyzed GTP and the vesicle uncoats. *Note:* All shapes once named in a panel retain their assignment in subsequent panels.

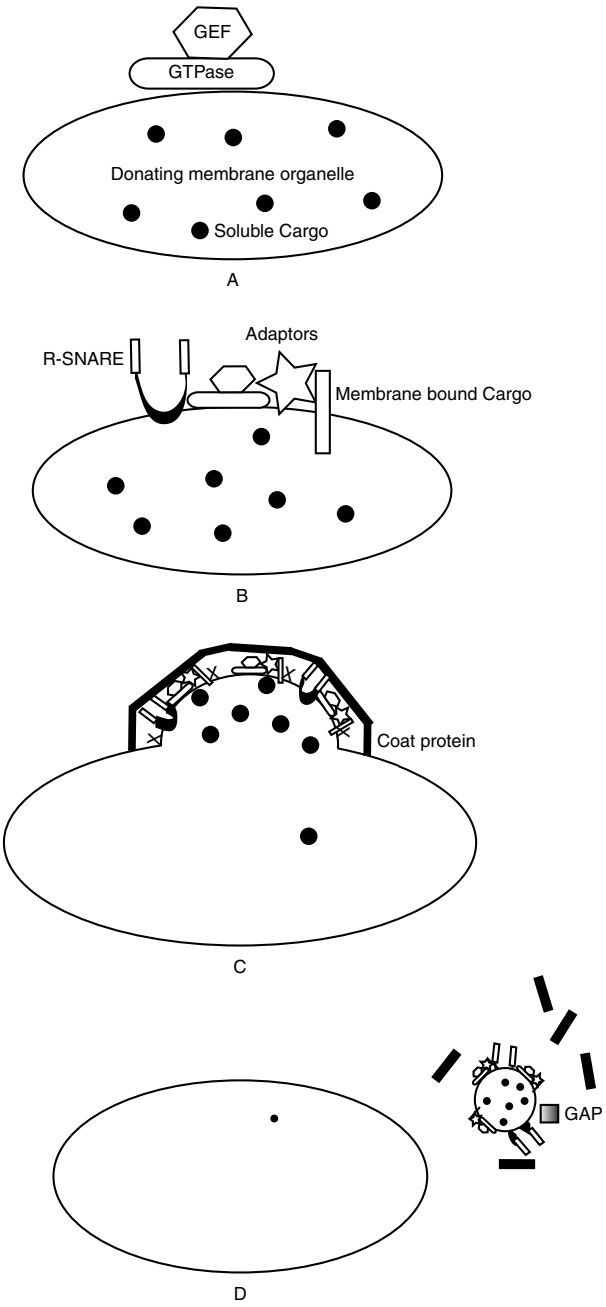

FIGURE 14.2 Generalized cartoon of vesicle formation, budding and movement. (A) GTPase attaches to the membrane and a GEF swaps GDP for GTP. (B) Adaptor proteins and cargo attach to the nucleating site of vesicle formation. (C) Coat proteins arrive and form a scaffolding complex for vesicle formation. Soluble cargo might be incorporated into the vesicle via adaptors or by bulk flow. (D) Vesicle has budded away from the donating membrane, a GAP hydrolyzed GTP and the vesicle uncoats. *Note:* All shapes once named in a panel retain their assignment in subsequent panels.

been suggested that ArfGAPs are involved in signaling and interactions with the cytoskeleton (Donaldson and Lippincott-Schwartz, 2000). It has also been demonstrated that when GTP hydrolysis is blocked, intracellular transport vesicles are unable to uncoat (Tanigawa et al., 1993). Regardless of whether this is a causal relationship, uncoating of the transport vesicles occurs after leaving the donor membrane and before vesicle fusion. Critically, vesicles are transported by interaction with microtubules and appear to be driven through interaction with kinesins (Girod et al., 1999).

14.2.3.3 Vesicle Fusion

The final stage of vesicular transport is the fusion of the vesicle with its target (Figure 14.3). An R-SNARE (soluble NSF attachment protein receptor) is incorporated as membrane cargo during vesicle formation. SNARE proteins are characterized by the presence of extensive coiled-coil regions and can form both *cis*- and *trans*-complexes. SNAREs are classed as Q or R, based on the position of an arginine (R) or glutamine (Q) at a highly conserved position within the coiled-coil domain (Antonin et al., 2002; Fasshauer et al., 1998). On the target membrane, at least one member of the Q-SNARE protein family, syntaxin (Edwardson, 1998), is present. This protein is complexed by a Sec1/Munc18 homologue, which appears to inhibit syntaxin from promiscuous interaction before the appropriate fusion event (Schulze et al., 1994). As the incoming vesicle reaches the target membrane (Figure 14.3B), syntaxin, a SNAP-25 homologue (which also possesses a Q within the critical region) and the R-SNARE form a coiled-coil, four-helix bundle (Hay, 2001). The SNARE complex has been implicated in docking, tethering (Ungermann et al., 1998, 2000) and physical fusion (Nickel et al., 1999) of the membranes. SNARE–SNARE interaction might also provide some specificity for vesicular transport (McNew et al., 2000). After fusion (Figure 14.3D), the *trans*-SNARE bundle is disassembled and recycled via the action of an ATPase, either NSF or p97. Whereas NSF acts in vesicular transport (Edwardson, 1998), its paralogue p97 plays a similar role in the postmitotic reassembly of organelles (Rabouille et al., 1998).

14.2.3.4 Regulation and Specificity

GTPases of the Rab protein family are essentially involved in the vesicular transport process in a variety of steps. They interact with a large number of proteins, both physically and genetically. These include SNARES, docking factors and the cytoskeleton. Most significantly, Rabs act in regulation of the overall process (Armstrong, 2000). Similarly to Sar1 and Arf, Rab proteins are recruited to the membrane as the GDP form, but are rapidly converted to the GTP state, which is the active form. The best-characterized function for Rab proteins is control of vesicle fusion, which is dependent on the GTP form being present on the vesicle membrane. Significantly, target membranes appear to contain a GAP activity, which results in rapid inactivation of the G protein and control of fusion efficiency (Rybin et al., 1996). Rab function is extensive, as the protein interacts with a large number of effector molecules. In the case of mammalian Rab5, at least 20 of these effectors have been identified, including EEA1 (a tethering factor) and several kinases, which most likely influence lipid structure and dynamics (Christoforidis et al., 1999). A highly significant aspect of the Rab gene family is specificity: members of the family are targeted precisely to endomembrane subcompartments. This feature makes these proteins attractive as molecular flags for pathways within the membrane transport system and has been exploited in several systems discussed next.

14.2.4 Vesicular Transport in Non-Opisthohont Lineages

Whereas much of the generalized information given has been developed in yeast and mammals, several divergent systems have been studied at the molecular level in some detail,

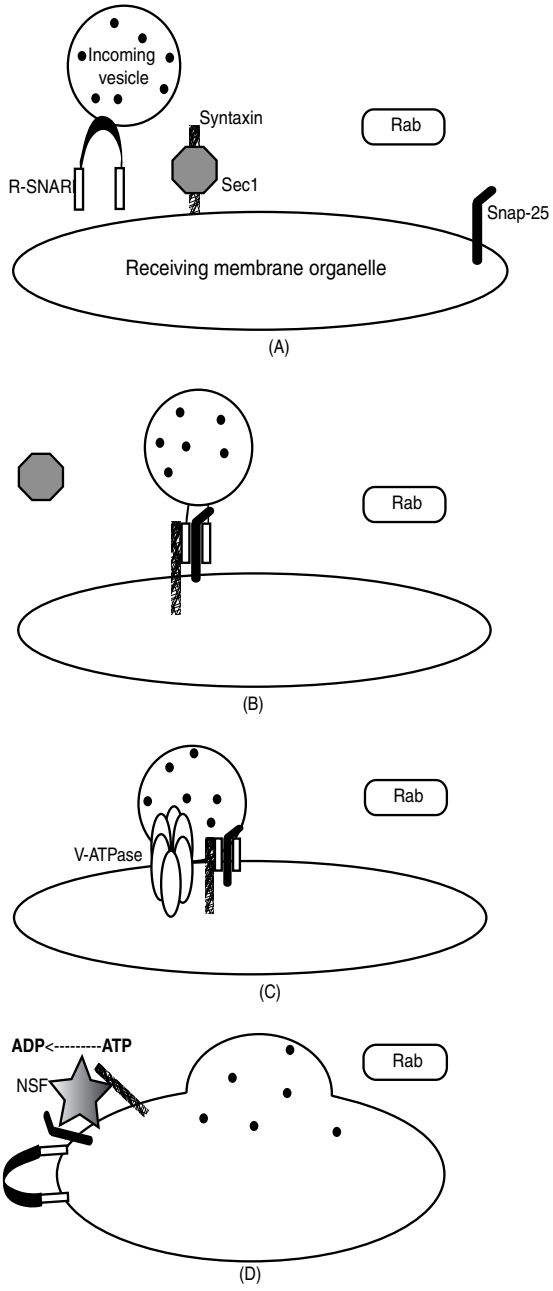

FIGURE 14.3 Generalized cartoon of vesicle fusion. (A) Incoming vesicle, containing cargo and R-SNARE homologue, approaches receiving organelle possessing Snap-25 and syntaxin homologue complexed with a Sec1 homologue. (B) Sec1 releases syntaxin, which forms a coiled coil with the R-SNARE and Snap25 homologues, prompting vesicle docking and tethering. (C) Vesicle fusion begins with the SNARE complex and other proteins (V-ATPase subunits) being implicated in creating the fusion pore. (D) NSF hydrolyzes ATP to dissociate the SNARE complex and recycle components for future rounds of vesicle fusion. Rab is implicated at various steps in the process. *Note:* All shapes once named in a panel retain their assignment in subsequent panels.

primarily because of the importance of these organisms as major disease agents. The best examples are among the apicomplexans *Plasmodium falciparum* and *Toxoplasma gondii* and the kinetoplastid *Trypanosoma brucei*. The last two are the best understood, in part for technical reasons, but also because of the advanced stage of their respective genome projects. The functional analyses in these taxa provide some information that is not forthcoming from sequence analysis alone, and might be informative of the manner in which the endomembrane system has evolved. It must be recognized that these systems are highly developed and are as far from the eukaryote common ancestor as metazoans. Some aspects of these systems have likely arisen from adaptation rather than being a reflection of a true basal or pleisiomorphic state. However, the presence of such systems provides a triangulation point for reconstructing evolutionary processes and importantly can demonstrate whether differences between higher eukaryotes reflect fundamental changes or simply specialization.

In *T. gondii*, members of the ARF family are present, and at least one member (ARF1) shares a clear role in transport through the Golgi complex and secretion with higher eukaryotes (Liendo et al., 2001). A number of Rabs have also been identified and their functions described. *T. gondii* homologues of Rab4, 5, 7 and 11 are present, and data suggest that these proteins also have conserved functions (Stedman et al., 2003). Interestingly, two Rab5 isoforms have been identified, but only one has been studied in detail (Robibaro et al., 2002). In addition, a clear Rab6 homologue is also present; this protein serves to define a retrograde transport pathway that delivers material to the *Toxoplasma* Golgi complex (Stedman et al., 2003). Therefore, once more the basic pathway of the endomembrane system is easily detectable and characterized, suggesting that these pathways emerged rapidly after eukaryotes evolved. Furthermore, indirect biochemical analysis demonstrates that the NSF/SNAP/Rab system is present in this organism (Chaturvedi et al., 1999).

A particularly interesting aspect of the Apicomplexa is the presence of specialized organelles, the rhoptries, dense granules and micronemes, all of which appear to play a role in invasion and establishment of the intracellular environment that the parasite requires for replication. Significantly, targeting of proteins to the rhoptries is dependent on recognition of sequences homologous to lysosomal targeting signals, and it is likely that rhoptry biogenesis is derived from targeting of rhoptry proteins via the early endosome (Robibaro et al., 2001).

T. brucei can be considered representative of the kinetoplastida, and all evidence suggests that other organisms of this order conform to the overall morphology and level of intracellular complexity seen in this paradigm organism. Clathrin has been identified in *Trypanosoma* (Morgan et al., 2001), as has the entire COPI coat (Maier et al., 2001). Other protein families have also been identified, including the subunits for three adaptin complexes; interestingly, the AP2 complex involved in recognition of cargo molecules at the cell surface was not identified (Morgan et al., 2002).

The best-studied family of endomembrane proteins in trypanosomes are the Rab proteins. As in mammals, several Rab proteins are likely involved in ER to Golgi transport: Rab1, 2A and 2B. These three proteins are well conserved, as evidenced by BLAST, and functional data are in good agreement with this assignment (Field et al., 1999; MCF and V. Dhir, unpublished). This is in contrast to *S. cerevisiae*, in which only a single Rab protein appears involved in ER to Golgi transport (the Rab1 homologue, Ypt1p). This suggests that for ER to Golgi transport, yeast is less complex than the trypanosome, and might indicate a loss of function during fungal evolution. For intra-Golgi transport, a similar level of complexity is apparent between yeast (Ypt31p and 32p) and trypanosomes (Rab18 and Rab31). In contrast, mammals have several Rabs that are likely responsible for regulating intra-Golgi transport, consistent with increased pathway complexity among the metazoans.

14.3 Mining the Databases

A comparative genomic study of the evolution of the endomembrane system requires three major pieces: genomic databases (details in Section 14.7.3), specific proteins to search for and methods to do those searches.

14.3.1 Candidate Proteins To Be Used as Representative Queries

It is clear that key proteins or protein families are involved in the generalized steps of the vesicular transport. The common machinery components involved in vesicle formation are Arf/Sar1 GTPases, GAPs and GEFs. As there are at least three major types of vesicles that share some of these common components, it is possible to search for these types of vesicles by looking for a representative component of their respective coat polymers. Clathrin (heavy chain) is the obvious representative for clathrin-coated vesicles, whereas α-COP and Sec31 are used as representatives of COPI and COPII vesicles, respectively. The fusion machinery also provides several attractive search query candidates. Sec1 and SNAREs are excellent examples of protein families with multiple paralogues involved in the same role at various steps of transport. NSF and p97 play keys roles in membrane fusion events and, as such, are good query candidates. Finally, Rab essentiality is undeniable and will also be included. Representatives of each of these protein families or protein complexes were assembled (Table 14.1) and used as query sequences for the subsequent comparative genomic surveys.

TABLE 14.1 Genes Used as Queries for Comparative Genomic Survey

Query	*Gene Family*	*Paralogue*	*Taxon*	*Accession No.*
Arf	ADP ribosylating factor	ARF1	Homo	P32889
Sar1	Secretion-associated, Ras-related	Sar1p	Saccharomyces	NP_015106
ArfGEF	Arf-GTP exchange factor	GEA1	Saccharomyces	P47102
AP	Adaptin	AP2 alpha subunit	Homo	NP_055018
COPII	COP II vesicle coat	Sec31p	Homo	NP_055748
COPI	Coatomer alpha	Alpha-COP	Saccharomyces	P53622
Clathrin	Clathrin	Chc1p	Saccharomyces	NP_011309
ArfGAP	Arf-GTP activating factor	Gcs1	Saccharomyces	NP_010055
R-SNARE	Synaptobrevin	Ykt6p	Saccharomyces	NP_012725
Syntaxin	Syntaxin	Sso1p	Saccharomyces	NP_015092
Sec1	Sec1	Syntaxin-binding protein 2	Homo	XP_008937
Rab	Rab	Ypt52p	Saccharomyces	P36018
NSF	*N*-ethylmaleimide-sensitive factor	NSF	Homo	XP_032173
p97	Transitional ER ATPase	TERA	Homo	P55072

Note: Homo: *Homo sapiens*; Saccharomyces: *Saccharomyces cerevisiae*.

14.3.2 Search Methods

The BLAST search algorithm (Altschul et al., 1997) can be used to find homologues of either DNA or protein sequences (queries) by searching genomic databases containing either sequence type. This algorithm aligns the query sequence with others in the database and assigns it a score based on how similar two sequences are. The reliability of a match by BLAST is measured in expectation (*E*) values and is usually expressed as a negative exponent.

This corresponds to the probability of observing an alignment that scores the same as the alignment between the query and a retrieved database entry, based on chance alone. This value is also corrected for the size of the database. The lower the *E* value, the more significant the match. At some point, the *E* value drops so low that the server may merely state the value as 0. PSI-BLAST is an iterative BLAST program that uses a scoring matrix based on a consensus of retrieved homologues to increase the sensitivity of the subsequent search. This method can also counteract lineage-specific peculiarities for a given search query, such as amino acid compositional bias, rapid evolutionary rate or divergence of a key motif (Altschul et al., 1997).

There are a variety of reasons why a particular protein might not appear in a genomic initiative database, other than its true absence from a genome. EST projects are a snapshot of genes expressed at a given time. If a gene is not expressed at that life cycle stage, or is expressed in low abundance, then it might not be represented. GSS surveys are random samplings of a genome, and so, by chance, a gene of interest simply might not have been encountered when the search was performed. Finally, when looking among diverse eukaryotes, the gene of interest might have diverged so much in that taxon that it is unrecognizable by a BLAST search; this can only be reliably confirmed by functional analysis. If no homologue can be identified in response to a particular query, then stating simply that a homologue was not identified is the most prudent response in the majority of cases.

The conservative nature of the "not identified" label released us from having to use a method that rigorously excludes claims about the lack of a homologue in a genome. Instead, we were able to use a search strategy that was biased against the other major pitfall, false positive identification of homologues. For each protein component, the relevant query sequence was used in a BLAST search against a given database. All sequences retrieved as possible homologues, given a generous cut-off value for significance, were then reciprocally used as queries for a BLAST search. Only those that retrieved the query sequence, and other defined orthologues of it, were noted as true homologues. This struck a balance between allowing for divergent sequence in distantly related taxa (i.e., weak but real BLAST hits) and caution in assigning homology. In cases where the retrieved sequence was a named homologue of the query (implying that functional characterization or at least BLAST identification had already been done), reciprocal BLAST analysis was not performed. For details of the search methodology, refer to Section 14.7.2.

Bioinformatic surveys of diverse genomes were performed by using homologues of the above components as queries in order to examine the origin and evolution of the vesicular transport machinery.

14.4 Endomembrane System Component Homologues in Diverse Genomes

The origins of a eukaryotic cellular system can be approached from the bottom up, looking to diverse prokaryotes for homologues of components of complex eukaryotic cellular machinery, or from the top down by reconstructing from extant taxa a consensus of commonly held and therefore likely ancestral machinery. Both approaches will be taken, using the components listed in Table 14.1, together with BLAST analysis (Altschul et al., 1997).

14.4.1 Bottom Up: Prokaryotic Homologues of Endomembrane System Components

Several of the vesicular transport components have clear homologues in prokaryotic genomes, but these are highly significant and have narrow taxonomic distributions. When the proposed prokaryotic homologue is reciprocally used as a query for a BLAST search

(reciprocal BLAST analysis), the sequences retrieved are not prokaryotic homologues but rather eukaryotic proteins. This pattern seems indicative of a lateral gene transfer (LGT) event, transferring the eukaryotic gene to the prokaryote rather than the gene being the progenitor of a novel eukaryotic gene family. For example, when using ArfGEF as a query in a BLAST search, the RalF protein from *Legionella* is returned at 5e-19. This protein has been shown to have Arf-modulating activity *in vivo* (Nagai et al., 2002). Another protein from *Rickettsia*, identified as a further ArfGEF homologue, is also returned at 2e-15. On reciprocal BLAST analysis, both return each other with *E* values in the range of e-78 and eukaryotic ArfGEFs (e-30), but do not seem to have a wide distribution among prokaryotes. This likely represents a lateral transfer to either *Legionella* or *Rickettsia* and subsequent transfer to the other (Nagai et al., 2002). Similarly, the RecO protein from *Deinococcus* has an identifiable GAP domain at its C-terminal end ($E = 0.051$), but other RecO homologues do not. This might be a case of LGT and domain fusion specifically in this taxon.

The cotranslational system is highly conserved and clearly has its origin with the prokaryotes; in these organisms, the system is used for direct export across the plasma membrane into the periplasmic space (reviewed in Rapoport et al., 1996). Significantly, the minimal functional core of the RNA is conserved in both prokaryotes and eukaryotes. As well, the protein responsible for recognizing the signal sequence (SRP54) in eukaryotes has a bacterial homologue. The remainder of the eukaryotic SRP protein components are absent in prokaryotes. Furthermore, the SRP receptor, responsible for recognizing the ribosome–SRP complex, is also highly conserved between the kingdoms. In addition, in prokaryotes, the nascent protein is translocated through a proteinaceous channel composed of SecY, which exhibits clear homology to a protein performing an analogous function in eukaryotes, Sec61. Interestingly, Sec61 functions in collaboration with several additional polypeptides, including Sec62/63, which are responsible for mediating interactions with machinery in the ER lumen. These are absent from the prokaryotic system (although additional non-conserved polypeptides are also present in the bacterial pathway), highlighting that the basic machinery appears conserved but a number of important (and frequently essential) functions are eukaryote specific. The high degree of conservation seen in the polypeptide translocation system provides a rare and focused insight into evolution of the endomembrane system.

A number of important components of the vesicular transport machinery belong to larger gene families, each having intriguing putative prokaryotic homologues (Table 14.2).

TABLE 14.2 Comparison of Eukaryotic vs. Prokaryotic Endomembrane Component Homologues

Component[a]	*Eukaryotic E value*[b]	*Prokaryotic Homologue*[c]	*Prokaryotic E value*[d]
Rab/Sar/Arf	E-05 to E-98	Putative GTPases	Ψ[e] I2 = E-06 to E-11
α-COP	E-26 to E-101	WD-40 proteins	E-20
Sec31	E-100	WD-40 proteins	E-40 to E-79
p97	E-130 to E-0.0	cdc48 homologues	E-180
NSF	E-50 to E-0.0	cdc48 homologues	E-50

[a] This column lists the protein family, or families, used as queries with specific queries matching their family designation in Table 14.1.

[b] This column lists the range of expectation value scores seen for retrieved eukaryotic homologues.

[c] This column lists the general assignment of prokaryotic sequences assigned as putative homologues.

[d] This column lists the range of expectation value scores seen for putative prokaryotic homologues.

[e] In the case of the Arf/Sar1/Rab searches, two iterations of PSI-BLAST were done before a significant prokaryotic homologue was retrieved.

A BLAST search using Arf, Sar1 or Rab queries produces similarity to each other and a number of other GTPases, mostly eukaryotic. No single, clear, prokaryotic homologue can be said to have given rise to endomembrane system GTPases. Nonetheless, position-specific-interated (PSI)-BLAST searches with Arf, Sar1 or Rab retrieved several GTPases with moderate prokaryote taxonomic distribution. These hit eukaryotic GTPases (E = 5e-13) and elongation factors (E = 4e-04) in reciprocal BLAST analysis. Most likely, an ancestral GTPase gave rise to the eukaryotic GTPases, but which is unclear. Eukaryotic small GTPases being more closely related to each other (E = e-05 to e-98) than to any given prokaryotic homologue suggests that the common ancestor had a very simple GTPase composition, which was expanded following the initial diversification of eukaryotes.

Proteins of both COPI and COPII vesicles (α-COP and Sec31, respectively) possess WD-40 domains. This domain, present in a wide variety of functionally unrelated proteins, is implicated as a scaffolding domain that facilitates protein–protein interactions (Smith et al., 1999). A number of bacterial and archael proteins also have very clear WD-40 domains. BLAST analysis of Sec31 retrieved eukaryotic homologues scoring in the range of E = –100. In the same BLAST search, multiple cyanobacterial sequences were retrieved that, when used as queries in reciprocal BLAST analysis, retrieved diverse prokaryotic sequences from Bacteria and Archaea (E = e-40 to e-77). Use of α-COP as a query in BLAST analysis retrieved eukaryotic sequences from various taxa (E = e-26 to e-101), whereas prokaryotic sequences were obtained with expectation values in the range of e-20. Many of the putative prokaryotic homologues are simply assigned as WD-40 proteins without further functional prediction. This suggests that only the WD-40 domain is conserved and not necessarily homologous functionality. Clearly, Sec31 and α-COP have arisen from one or more ancestral proteins containing such a domain, but this analysis suggests that a true functional homologue is not likely present within the sampled prokaryotic taxa.

The AAA-type ATPase family is a well-defined group of proteins associated with a wide variety of cellular functions (Ye et al., 2001). One member of this family, p97, has been shown to be involved in homotypic membrane fusion events, such as postmitotic reassembly of ER (Latterich et al., 1995) and Golgi (Rabouille et al., 1998). It has also been implicated in a number of additional functional processes, including ubiquitin-dependant protein degradation (Ghislain et al., 1996) and the cell cycle (Moir et al., 1982). A second AAA-type ATPase paralogue, NSF, on the other hand, is known to be involved only in SNARE complex disassembly and recycling (Edwardson, 1998). Clear homologues of AAA-type ATPases can be found in both Bacteria and Archaea (Pamnani et al., 1997). A BLAST search with p97 as the query sequence yields eukaryotic homologues with expectation values ranging from e-130 to 0.0 and prokaryotic homologues with scores of ca. E = e-150. BLAST analysis of NSF retrieves eukaryotic NSF homologues (E = e-50 to 0.0) and prokaryotic sequences in the E = e-50 range as well. A number of indications suggest that p97 might be the ancestral and pleisiomorphic form of the protein (Zhang et al., 2000). BLAST values for prokaryotic homologues are higher when using the p97 version than with the NSF query. Also, the broad spectrum of cellular processes with which p97 is involved suggests that NSF might have been a specialized offshoot. However, because BLAST values might be affected by evolutionary rate, this should be examined by phylogenetic analysis.

Although not a prokaryotic connection, it is still possible to find some evolutionary affinities that go beyond simple paralogue expansion. For example, the adaptin complex has common origins with a subcomplex of COPI (Duden et al., 1991; Schledzewski et al., 1999). Both complexes interact with, and help to form, vesicle coats at the Golgi. The large subunits of each complex are clearly paralogues as are the medium and small subunits. It is proposed that what began as a heterodimer of a large and small subunit duplicated and differentiated to form an ancestral heterotetramer. Further duplications later produced the F-COP complex and then subsequently the AP3, 2 and 1 adaptin complexes (Schledzewski

et al., 1999). Recent data also suggest that the regulatory V1H subunit of the V-ATPase is homologous to the N-terminus of the β-COP and μ-adaptin subunits (Geyer et al., 2002). These studies, however, are restricted in their taxonomic sampling, primarily including sequences from animals, fungi and land plants. A broader sampling would be useful to verify and expand this proposed evolutionary scheme. No clear prokaryotic homologues were identifiable for SNAREs, adaptins, Sec1 and clathrin by BLAST or PSI-BLAST analyses.

14.4.2 Top Down: Reconstructing the Vesicular Transport Machinery of the Last Common Eukaryotic Ancestor

Reconstructing the evolution of a cell biological system is done, in the ideal case, by deducing which components are present in the common ancestor of the group of organisms under consideration. Features found in all descendants of a common ancestor must, barring lateral gene transfer, have been present in that ancestor. Features found in most taxa, and in the deepest branch of a resolved phylogeny, are also most likely to have been present in the common ancestor. To make such deductions, knowledge of the cell biology of the taxa under examination, plus their phylogeny, is required. Systems central to eukaryotic cellular evolution will require a deeply branched phylogeny.

Initially, small subunit ribosomal DNA sequences were used alone to resolve eukaryotic relationships (Sogin, 1991), but for multiple reasons these data are insufficient to construct robust phylogenies and must be supplemented by protein and morphological data (Dacks and Doolittle, 2001; Embley and Hirt, 1998; Philippe and Adoutte, 1998; Philippe et al., 2000; Chapter 2). Issues relating to construction of phylogenies are discussed elsewhere in the book (Chapter 6). Despite analyses showing that the major eukaryotic lineages might have diverged from one another rapidly (Philippe and Adoutte, 1998; Philippe et al., 2000), evidence from improved species sampling, increased availability of molecular data and optimized computational analysis to suggest phylogenetic structure in eukaryotic relationships (Figure 14.4 and references therein). Genomics can further provide us with some of the data needed to better resolve that structure (Dacks and Doolittle, 2001).

Despite these advances, a resolved, broadly sampled and rooted eukaryotic phylogeny does not seem imminent. Nonetheless, reconstructing the evolution of eukaryotic cell biology is still possible in its absence. Rather than looking at a designated deepest taxon, diversity can be used to approximate the last common eukaryotic ancestor. Using comparative genomics to search, among diverse taxa, for proteins known to be functionally important in a particular system allows estimation of the minimal protein machinery present in the last common ancestor of the taxa sampled. The wider the diversity of sampling, the better their last common ancestor approximates the last common ancestor of all eukaryotes. The scheme in Figure 14.4 shows the state of knowledge ca. 2002 regarding eukaryotic relationships. Coded in this picture are major lineages with publicly accessible genomics initiatives. Although there are areas of the eukaryotic tree vastly undersampled, the current diversity of genome initiatives allows us to examine a last common ancestor that is a crude but reasonable approximation of the last common eukaryotic ancestor.

Eukaryotic genomic initiative databases publicly accessible as of September 2002 were searched for homologues of the protein family queries listed in Table 14.1. Most genomes examined have, at least, one member of the protein families identified as important components of the vesicle transport machinery (Table 14.3 and Table 14.4). Several queries were not identified in the *Paramecium* (ciliate) genome, but this project was in its earliest stages at the time of this survey. Given that the apicomplexans (sisters to the ciliates in the alveolate supergroup) possess all the components, it is likely that the scarcity of *Paramecium* components is due to sampling. The shared presence of the basic vesicular transport machinery in most of the surveyed genomes implies that it was already established in an early eukaryotic

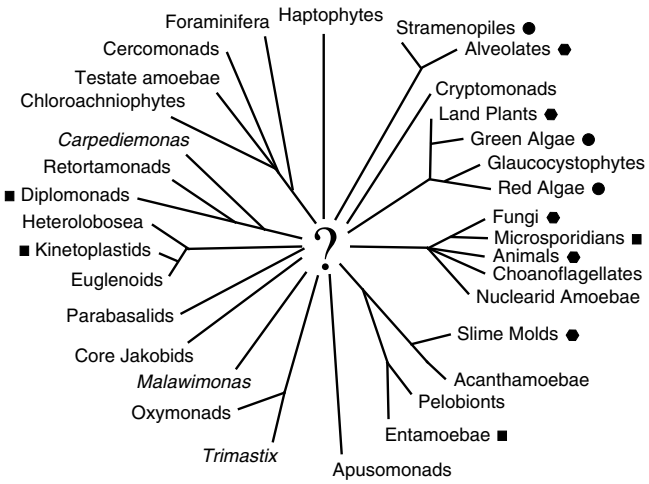

FIGURE 14.4 Schematic of proposed eukaryotic relationships, ca. 2002. This unrooted star phylogeny incorporates references fully cited in Dacks and Doolittle (2001) 419–425, as well as new morphological SSUrDNA and protein data [Archibald et al. (2003) 62–66; Arisue et al. (2002) 1–10; Bapteste et al. (2002) 1414–1419; Silberman et al. (2002) 777–786; Simpson et al. (2002a) 239–248; Simpson et al. 1782–1791]. Taxa with publicly availaable EST projects are noted with a circle. Those with GSS orgenome projects are noted with a square. Taxa with both are marked with a hexagon.

TABLE 14.3 Comparative Genomic Survey of Vesicle Formation and Movement Proteins in Diverse Eukaryotic Genomes

Higher Taxon	*Organism*	*Arf*	*Sar1*	*ArfGEF*	*AP*	*COPII*	*COPI*	*Clathrin*	*ArfGAP*
Fungi	*Saccharomyces*	A	A	A	A	A	A	A	A
Land plants	*Arabidopsis*	A	C	A	A	C	A	C	A
Animal	*Homo*	A	A	A	A	A	A	A	A
Diplomonad	*Giardia*	A	D	D	C	E	B	D	D
Kinetoplastid	*Trypanosoma*	C	E	E	C	E	B	C	E
Apicomplexa	*Plasmodium*	A	E	E	E	C	B	E	C
Slime molds	*Dictyostelium*	C	E	E	D	NI	D	A	C
Entamoebae	*Entamoeba*	C	E	E	E	E	B	E	E
Red Algae	*Porphyra*	D	D	D	D	E	B	D	NI
Stramenopiles	*Phytophthora*	D	E	D	D	D	B	D	D
Green algae	*Chlamydomonas*	A	E	NI	D	NI	B	D	A
Ciliates	*Paramecium*	NI	NI	A	NI	NI	C	C	NI

Note: A: homologues published in separate analyses; B: homologues previously identified in Dacks and Doolittle (2001) 419–425; C: genes not yet published but found in Genbank; D: a gene listed on the respective genome initiative Web site; E: homologue found by reciprocal BLAST analysis in this study; NI: a clear homologue not reliably identified by any of the given criteria. Table last updated in September 2002.

TABLE 14.4 Comparative Genomic Survey of Vesicle Fusion Proteins in Diverse Eukaryotic Genomes

Higher Taxon	*Organism*	*R-SNARE*	*Syntaxin*	*Sec1*	*Rab*	*NSF*	*p97*
Fungi	*Saccharomyces*	A	A	A	A	A	A
Land plants	*Arabidopsis*	A	A	A	A	C	C
Animal	*Homo*	A	A	A	A	A	A
Diplomonad	*Giardia*	B	A	B	A	D	D
Kinetoplastid	*Trypanosoma*	B	A	B	A	C	C
Apicomplexa	*Plasmodium*	B	C	B	B	C	C
Slime molds	*Dictyostelium*	B	A	B	A	A	C
Entamoebae	*Entamoeba*	B	B	B	A	E	C
Red algae	*Porphyra*	B	A	NI	B	NI	D
Stramenopiles	*Phytophthora*	B	A	B	B	E	D
Green algae	*Chlamydomonas*	B	A	B	A	E	D
Ciliates	*Paramecium*	NI-a	NI	NI	C	C	NI

Note: A: homologues published in separate analyses; B: homologues previously identified in Dacks and Doolittle (2001) 419–425; C: genes not yet published but found in Genbank; D: a gene listed on the respective genome initiative Web site; E: homologue found by reciprocal BLAST analysis in this study; NI: a clear homologue not reliably identified by any of the given criteria. In the case of NI-a, an Euplotes (ciliate) homologue has been identified. Table last updated in September 2002.

ancestor and that the basic mechanism of vesicular transport has also been conserved. Similarly, the common presence of clathrin, Sec31, and α-COP homologues in the various genomes suggests that the last common ancestor also had the ability to form the three classes of vesicles seen at present.

Many of the questions surrounding the evolution and complexification of eukaryotic systems are ones of duplications. Some queries involve multiple duplications of closely related proteins, which might be difficult to assess by BLAST alone (see Section 14.5). Others, however, involve deep duplications, yielding paralogues with divergent function such as Sar1/Arf. The Arf protein family is composed of several paralogous subfamilies, each playing a similar role in the formation of clathrin and COPI vesicles as Sar1 does for COPII. The majority of taxa examined have at least one homologue of both Arf and Sar1 present in their genomes (Table 14.3 and Table 14.4), and thus the duplication that gave rise to Arf and Sar1 is likely to have occurred before the divergence of the taxa examined.

The situation with NSF vs. p97 is slightly more complicated. Both proteins are members of a larger AAA-type ATPase family. Most taxa examined seem to have at least one copy of both genes. However, both proteins also retrieve eukaryotic cdc48 homologues as well as a number of uncharacterized cdc48-like ORFs with significant BLAST scores. This makes it quite difficult to distinguish the presence of p97 vs. that of NSF. As well, although the biological function of NSF is well established, p97 seems to have multiple roles in the cell, membrane fusion being only one of them (Ye et al., 2001). As such, the biological significance of the duplication is difficult to assess. Although the story is likely to be infinitely more complex, it is possible to deduce, at a minimum, that the duplication which gave rise to p97 and NSF occurred before the last common ancestor of the taxa tested.

Comparative genomic surveys have also been used to demonstrate a trend of expansion in some of the families involved in vesicular transport in unicellular organisms as compared with multicellular taxa. This is seen quite strikingly in the Rab proteins when comparing *Saccharomyces* (11 Rabs), *Plasmodium* (11) and *Trypanosoma* (16) to *Caenorhabditis* (29), *Homo* (60) and *Arabidopsis* (57); (Bock et al., 2001; Rutherford and Moore, 2002).

14.5 Beyond BLAST: Examples from Functional Studies

BLAST can survey for the presence of relevant protein families in diverse eukaryotic genomes. However, many more detailed questions are beyond the scope of sequence analysis alone and require additional data.

Questions of detailed evolutionary history might require the identification of a gene sequence at the level of its paralogue subfamily within a larger gene family, ideally within a phylogenetic framework. The reliability of such an assignment by BLAST might be compromised, because the algorithm does not take into account evolutionary rate and so a sequence from an organism with a rapid rate might be misidentified. Additionally, many of the databases provide only partial (end reads of cDNAs) or poor-quality gene sequence (single-pass reads of genomic fragments). These should provide enough conserved sequence to yield a broad gene family assignment (such as in the case of Sar1 vs. Arf), but a subfamily identification might be beyond the boundaries of reliability. For identification of close paralogue affiliation, molecular biology, phylogeny and functional assignment might be required. Other detailed questions of evolution within a gene family might involve establishing the relationship of paralogues, and the timing of their expansion relative to various lineage divergences or the relationship of various paralogues relative to an outgroup.

14.5.1 Genomics and Phylogeny

Several phylogenetic studies have used genomic databases for the initial identification of partial sequences, which were then confirmed and expanded through standard molecular biological means. Phylogenetic analysis of the various subunits of the adaptin and COP complexes, as obtained in part by sequencing cDNAs, revealed not only homologies between the seemingly unrelated endomembrane components but also some internal paralogue relationships (Chow et al., 2001; Schledzewski et al., 1999). Similar studies have also been performed on the syntaxin gene family. Syntaxin genes were identified from a variety of protist EST and GSS surveys and further characterized by molecular biological means. Phylogenetic analysis determined that the duplication giving rise to the syntaxin gene families must have occurred early on in eukaryotic evolution (Dacks and Doolittle, 2002), as well as identifying several lineage specific paralogue expansions within the gene family. The Rab protein family has been studied in perhaps the most diverse array of taxa (Bush et al., 1993; Janoo et al., 1999; Langford et al., 2002; Morgan et al., 2002; Rutherford and Moore, 2002; Saito-Nakano et al., 2001; Stedman et al., 2003). Various Rab homologues have been identified from genomics initiatives in *Entamoeba* (Saito-Nakano et al., 2001) and more recently from *Giardia* (Langford et al., 2002). The phylogenetic analysis in this study shows a family with multiple deep duplications giving rise to several clades early on in the history of eukaryotes. On the other hand, phylogenetic analysis of the Rab5 families from mammals, yeast and trypanosomes indicates that the evolution of multiple Rab5 genes postdates the common eukaryotic ancestor as the Rab5 genes for each organism segregate into separate clades (Field et al., 1998). This implies that the common ancestor had a single Rab5 and hence most likely a simplified endocytic system. Interestingly, in yeast there is a high degree of redundancy between the three Rab5 isoforms. In the simpler trypanosomal system, wherein there are only two Rab5 family members, these functions appear fully distinct. The status of the two *Toxoplasma* gene products awaits functional analysis (Robibaro et al., 2002).

Phylogenetic analysis has also shed light on the story of paralogue expansion in both SNAREs and Rab proteins. There is a clear story of convergent, lineage-specific expansion in the plasma-membrane-localized syntaxin families of both metazoa and plants (Dacks and Doolittle, 2002). The *Arabidopsis* genome, in particular, shows a heavily expanded SNARE

complement (Sanderfoot et al., 2000). A similar story is seen with the Rab proteins. A phylogenetic study of Rabs in yeast, mammals and *Arabidopsis* showed not only an expanded Rab content in *Arabidopsis*, as compared with yeast, but also that the mammalian and plant Rabs have expanded separately (Rutherford and Moore, 2002).

14.5.2 Genomics and Cell Biology

As for phylogenetic studies, genomics has allowed for identification of novel vesicular transport component homologues that can then be studied functionally. This has yielded both comforting underlying generalities to the model of endomembrane system organization as well as some surprising differences.

A number of significant finds regarding functional vs. *in silico* assignment of paralogues have emerged so far. A BLAST result alone might not be sufficient to assign functional homology. For example, one trypanosome Rab homologue clearly belongs to the Rab 18 family, based on BLAST. Paradoxically, the trypanosomal protein localizes to the Golgi complex whereas in metazoans Rab18 is associated with an endosomal compartment (Jeffries et al., 2002). Further, for a protein finally assigned as a Rab31 homologue, sequence comparisons were unable to discriminate its assignment between several Rab subfamilies. Functional analysis indicated trypanosomal Rab31 as a Golgi protein, in common with human Rab31 (Field et al., 2000). These observations indicate that detailed reconstruction of pathways within the endomembrane system, based purely on the presence of similar protein factors, is likely to be inaccurate; they probably also reflect the strong emphasis on functional data from higher eukaryotes. Most likely, as studies progress in divergent systems, the information will be of major utility for improving the accuracy of *in silico* assignments. For a second family of proteins, the SNAREs, this will be even more critical as these proteins contain extensive coiled-coil regions and retain limited sequence homology. Assignment of the full SNARE complement took several years, together with functional analysis, in the accessible *S. cerevisiae* system (Lewis and Pelham, 2002). It is clear that BLAST alone will be unable to even identify all SNAREs in divergent systems, let alone assign them a specific function. Phylogenetic analysis and functional cell biology will be even more important to fully understand the role of SNAREs in eukaryotes.

Nonetheless, the generalities of function for both these protein families have been confirmed in diverse taxa. Studies of Rabs have been done in a wide array of organisms, helped in part by sequences derived from genomic initiatives. Syntaxins too, albeit to a lesser extent, have benefited from the genomic windfall (Bogdanovic et al., 2000, 2002; Zhu et al., 2002). This will help establish a generalized model of how the endomembrane system functions have evolved.

14.6 Conclusions

The analysis of prokaryotic homologues raises several points. It appears likely that direct prokaryotic homologues of the proteins involved in vesicle transport are absent. Proteins containing the domains from which the eukaryotic components are built are present, but the occurrence of multiple GTPase or WD-40 domains in prokaryotes is perhaps unsurprising. For the prokaryotic taxa with genomes currently in hand, it is unlikely that any function homologous to vesicle transport is present. However, a recent characterization of the archeon *Ignicoccus* revealed an intracellular vesicle (Rachel et al., 2002). On the other hand, there are well-characterized prokaryotic homologues for various pieces of the protein translocation machinery, which serve similar if not identical roles in the cell (Rapoport et al., 1996). A recent functional study has even shown that when this system is blocked in *E. coli*, stacks of internal membranes with attached ribosomes accumulate in the cell (Herskovits et al., 2002),

eerily reminiscient of ER. Although these and the archael vesicle might only be superficially similar and not truly homologous to the eukaryotic systems, they provide examples of structures similar to those in the endomembrane system arising in a prokaryotic context. Having more than one example of this makes any suggested models of the process more plausible.

From the current survey of eukaryotes, it is clear that the majority of the vesicular transport protein machinery that is well characterized in model systems is present in diverse taxa. This indicates that the entire system is relatively conserved and that the models for the mechanisms of vesicular transport are broadly applicable to eukaryotes beyond yeast and humans. The mere presence of a homologue does not necessarily imply the same function, but the presence of multiple interacting components makes the conservation of mechanism the most parsimonious working hypothesis. This mechanism needs to be tested *in vivo,* however, in diverse eukaryotes. The difference in function will tell exactly how the overall model must be modified to be universally applicable, as well as provide insight into specific evolutionary modifications. The last common eukaryotic ancestor appears to have had a complex endomembrane system. If generalities can be drawn from the evolution of the syntaxin and Rab families, then the elaboration of the vesicular transport components is likely to have begun early in eukaryotic evolution and ballooned on the various incidents of multicellularity. Nonetheless, other protein families will also have to be examined and this detailed picture of the evolution of the protein machinery will ultimately flesh out our understanding of the evolution of the endomembrane system in eukaryotes and provide deeper insight into organisms that are a threat or a benefit to us.

14.7 Materials and Methods

14.7.1 Search Queries

Either animal or fungal representatives of each protein family identified in the introduction were retrieved from Genbank and used as queries for the BLAST analyses. Table 14.1 gives the full listing of queries with their accession numbers. The protein representatives from these taxa were used as queries because the functional characterization of the protein families occurred in these model systems.

14.7.2 Search Methods

Keyword searching was performed at all databases that supported this option, in order to retrieve identified homologues. BLAST analysis was performed at the NCBI BLAST server (http://www.ncbi.nlm.nih.gov/BLAST/). Both the BLASTp algorithm and PSI-BLAST algorithm when necessary were used to search the protein databases. The tBLASTn algorithm was used when searching nucleotide databases. A cut-off value of 0.05 was used when selecting potential homologues, and each retrieved sequence was reciprocally used as a query back to the nr database. Only sequences that retrieved the initial query sequence were deemed legitimate homologues.

Two sets of searches were performed. In September 2001, searches were performed for a subset of the vesicular transport proteins, the results of which were published in November 2001 (Dacks and Doolittle, 2001). A second search was performed in September 2002 for an expanded set of vesicular transport proteins and to search for prokaryotic homologues of the queries listed in Table 14.1. Therefore, the homologues identified as B in Table 14.3 and Table 14.4 were identified in the September 2001 search, and all results are current as of September 2002.

14.7.3 Databases

The nonredundant (nr) database at Genbank was the only database searched when attempting to find prokaryotic homologues. The search for eukaryotic orthologues was

also primarily performed in the nr database. However, the others ESTs database, HTGS and the GSS databases were also searched. Searches were also performed at a number of genome project Web sites, including the *Dictyostelium* cDNA project (http://www.csm.biol.tsukuba.ac.jp/cDNAproject.html), the *Giardia* genome project (http://jbpc.mbl.edu/Giardia-HTML/index2.html), the *Phytophthora* Genome Consortium (https://xgi.ncgr.org/pgc/) as well as the *Chlamydomonas* and *Porphyra* genome projects (http://www.kazusa.or.jp/en/plant/database.html).

Acknowledgments

JBD thanks W. F. Doolittle for helpful discussion, critical reading of the manuscript and supervision. Various discussions with T. Cavalier-Smith and members of the Doolittle and Roger labs have helped to shape many of the ideas in this chapter. P. P. Poon, M. Sogin, L. A. M. Davis and A. McArthur are also thanked for critical reading of this manuscript. MCF thanks members of the Field Laboratory for comments on the manuscript. We acknowledge the various genome projects that have made data publicly available; without these data such discussion would be impossible. Work in this chapter was supported by a grant to W. F. Doolittle (MT4467), a Canadian Institutes of Health Research doctoral research award to JBD and the Wellcome Trust (program grant to MCF).

References

Addinall, S. G. and Holland, B. (2002) The tubulin ancestor, FtsZ, draughtsman, designer and driving force for bacterial cytokinesis. *J. Mol. Biol.* 318: 219–236.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1994) *Molecular Biology of the Cell,* Garland Publishers, New York.

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.

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

Archibald, J. M., Longet, D., Pawlowski, J. and Keeling, P. J. (2003) A novel polyubiquitin structure in cercozoa and foraminifera: Evidence for a new eukaryotic supergroup. *Mol. Biol. Evol.* 20: 62–66.

Arisue, N., Hashimoto, T., Lee, J. A., Moore, D. V., Gordon, P., Sensen, C. W., Gaasterland, T., Hasegawa, M. and Muller, M. (2002) The phylogenetic position of the pelobiont *Mastigamoeba balamuthi* based on sequences of rDNA and translation elongation factors EF-1alpha and EF-2. *J. Euk. Microbiol.* 49: 1–10.

Armstrong, J. (2000) Membrane traffic between genomes. *Genome Biol.* 1: REVIEWS104.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H. (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* 99: 1414–1419.

Becker, B. and Melkonian, M. (1996) The secretory pathway of protists: Spatial and functional organization and evolution. *Microbiol. Rev.* 60: 697–721.

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.

Bogdanovic, A., Bennett, N., Kieffer, S., Louwagie, M., Morio, T., Garin, J., Satre, M. and Bruckert, F. (2002) Syntaxin 7, syntaxin 8, Vti1 and VAMP7 (vesicle-associated membrane protein 7) form an active SNARE complex for early macropinocytic compartment fusion in *Dictyostelium discoideum*. *Biochem. J.* 368: 29–39.

Bryant, N. J. and Stevens, T. H. (1998) Vacuole biogenesis in *Saccharomyces cerevisiae*: Protein transport pathways to the yeast vacuole. *Microbiol. Mol. Biol. Rev.* 62: 230–247.

Bush, J., Franek, K., Daniel, J., Spiegelman, G. B., Weeks, G. and Cardelli, J. (1993) Cloning and characterization of five novel *Dictyostelium discoideum* rab-related genes. *Gene* 136: 55–60.

Chaturvedi, S., Qi, H., Coleman, D., Rodriguez, A., Hanson, P. I., Striepen, B., Roos, D. S. and Joiner, K. A. (1999) Constitutive calcium-independent release of *Toxoplasma gondii* dense granules occurs through the NSF/SNAP/SNARE/Rab machinery. *J. Biol. Chem.* 274: 2424–2431.

Chow, V. T., Sakharkar, M. K., Lim, D. P. and Yeo, W. M. (2001) Phylogenetic relationships of the seven coat protein subunits of the coatomer complex, and comparative sequence analysis of murine xenin and proxenin. *Biochem. Genet.* 39: 201–211.

Christoforidis, S., Miaczynska, M., Ashman, K., Wilm, M., Zhao, L., Yip, S. C., Waterfield, M. D., Backer, J. M. and Zerial, M. (1999) Phosphatidylinositol-3-OH kinases are Rab5 effectors. *Nat. Cell Biol.* 1: 249–252.

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

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.

Delwiche, C. F. (1999) Tracing the thread of plastid diversity through the tapestry of life. *Am. Nat.* 154: S164–S177.

Donaldson, J. G. and Lippincott-Schwartz, J. (2000) Sorting and signaling at the Golgi complex. *Cell* 101: 693–696.

Doray, B., Ghosh, P., Griffith, J., Geuze, H. J. and Kornfeld, S. (2002) Cooperation of GGAs and AP-1 in packaging MPRs at the trans-Golgi network. *Science* 297: 1700–1703.

Duden, R., Griffiths, G., Frank, R., Argos, P. and Kreis, T. E. (1991) Beta-COP, a 110 kDa protein associated with non-clathrin-coated vesicles and the Golgi complex, shows homology to beta-adaptin. *Cell* 64: 649–665.

Edwardson, J. M. (1998) Membrane fusion: All done with SNAREpins? *Curr. Biol.* 8: R390–393.

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.

Field, H., Farjah, M., Pal, A., Gull, K. and Field, M. C. (1998) Complexity of trypanosomatid endocytosis pathways revealed by Rab4 and Rab5 isoforms in *Trypanosoma brucei*. *J. Biol. Chem.* 273: 32102–32110.

Field, H., Sherwin, T., Smith, A. C., Gull, K. and Field, M. C. (2000) Cell-cycle and developmental regulation of TbRAB31 localisation, a GTP-locked Rab protein from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 106: 21–35.

Field, H., Ali, B. R., Sherwin, T., Gull, K., Croft, S. L. and Field, M. C. (1999) TbRab2p, a marker for the endoplasmic reticulum of *Trypanosoma brucei*, localises to the ERGIC in mammalian cells. *J. Cell Sci.* 112: 147–156.

Geyer, M., Yu, H., Mandic, R., Linnemann, T., Zheng, Y. H., Fackler, O. T. and Peterlin, B. M. (2002) Subunit H of the V-ATPase binds to the medium chain of adaptor protein complex 2 and connects Nef to the endocytic machinery. *J. Biol. Chem.* 277: 28521–28529.

Ghislain, M., Dohmen, R. J., Levy, F. and Varshavsky, A. (1996) Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *EMBO J.* 15: 4884–4899.

Girod, A., Storrie, B., Simpson, J. C., Johannes, L., Goud, B., Roberts, L. M., Lord, J. M., Nilsson, T. and Pepperkok, R. (1999) Evidence for a COP-I-independent transport route from the Golgi complex to the endoplasmic reticulum. *Nat. Cell Biol.* 1: 423–430.

Glick, B. S. (2000) Organization of the Golgi apparatus. *Curr. Opin. Cell Biol.* 12: 450–456.

Hay, J. C. (2001) SNARE complex structure and function. *Exp. Cell Res.* 271: 10–21.

Herskovits, A. A., Shimoni, E., Minsky, A. and Bibi, E. (2002) Accumulation of endoplasmic membranes and novel membrane-bound ribosome-signal recognition particle receptor complexes in *Escherichia coli*. *J. Cell Biol.* 159: 403–410.

Jahn, R. and Sudhof, T. C. (1999) Membrane fusion and exocytosis. *Annu. Rev. Biochem.* 68: 863–911.

Janoo, R., Musoke, A., Wells, C. and Bishop, R. (1999) A Rab1 homologue with a novel isoprenylation signal provides insight into the secretory pathway of *Theileria parva*. *Mol. Biochem. Parasitol.* 102: 131–143.

Jeffries, T. R., Morgan, G. W. and Field, M. C. (2002) TbRAB18, a developmentally regulated Golgi GTPase from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 121: 63–74.

Kaiser, C. and Ferro-Novick, S. (1998) Transport from the endoplasmic reticulum to the Golgi. *Curr. Opin. Cell. Biol.* 10: 477–482.

Kasinsky, H. E., Lewis, J. D., Dacks, J. B. and Ausio, J. (2001) Origin of H1 linker histones. *FASEB J.* 15: 34–42.

Kirchhausen, T. (2000) Clathrin. *Annu. Rev. Biochem.* 69: 699–727.

Klumperman, J. (2000) Transport between ER and Golgi. *Curr. Opin. Cell Biol.* 12: 445–449.

Langford, T. D., Silberman, J. D., Weiland, M. E., Svard, S. G., McCaffery, J. M., Sogin, M. L. and Gillin, F. D. (2002) *Giardia lamblia*: Identification and characterization of Rab and GDI proteins in a genome survey of the ER to Golgi endomembrane system. *Exp. Parasitol.* 101: 13–24.

Latterich, M., Frohlich, K. U. and Schekman, R. (1995) Membrane fusion and the cell cycle: Cdc48p participates in the fusion of ER membranes. *Cell* 82: 885–893.

Lewis, M. J. and Pelham, H. R. (2002) A New Yeast Endosomal SNARE related to mammalian syntaxin 8. *Traffic* 3: 922–929.

Liendo, A., Stedman, T. T., Ngo, H. M., Chaturvedi, S., Hoppe, H. C. and Joiner, K. A. (2001) *Toxoplasma gondii* ADP-ribosylation factor 1 mediates enhanced release of constitutively secreted dense granule proteins. *J. Biol. Chem.* 276: 18272–18281.

Luzio, J. P., Rous, B. A., Bright, N. A., Pryor, P. R., Mullock, B. M. and Piper, R. C. (2000) Lysosome-endosome fusion and lysosome biogenesis. *J. Cell Sci.* 113: 1515–1524.

Maier, A. G., Webb, H., Ding, M., Bremser, M., Carrington, M. and Clayton, C. (2001) The coatomer of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 115: 55–61.

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.

Moir, D., Stewart, S. E., Osmond, B. C. and Botstein, D. (1982) Cold-sensitive cell-division-cycle mutants of yeast: Isolation, properties, and pseudoreversion studies. *Genetics* 100: 547–563.

Morgan, G. W., Allen, C. L., Jeffries, T. R., Hollinshead, M. and Field, M. C. (2001) Developmental and morphological regulation of clathrin-mediated endocytosis in *Trypanosoma brucei*. *J. Cell Sci.* 114: 2605–2615.

Morgan, G. W., Hall, B. S., Denny, P. W., Field, M. C. and Carrington, M. (2002) The endocytic apparatus of the kinetoplastida. Part II: Machinery and components of the system. *Trends Parasitol.* 18: 540–546.

Muniz, M., Nuoffer, C., Hauri, H. P. and Riezman, H. (2000) The Emp24 complex recruits a specific cargo molecule into endoplasmic reticulum-derived vesicles. *J. Cell Biol.* 148: 925–930.

Nagai, H., Kagan, J. C., Zhu, X., Kahn, R. A. and Roy, C. R. (2002) A bacterial guanine nucleotide exchange factor activates ARF on *Legionella* phagosomes. *Science* 295: 679–682.

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.

Orci, L., Stamnes, M., Ravazzola, M., Amherdt, M., Perrelet, A., Sollner, T. H. and Rothman, J. E. (1997) Bidirectional transport by distinct populations of COPI-coated vesicles. *Cell* 90: 335–349.

Pamnani, V., Tamura, T., Lupas, A., Peters, J., Cejka, Z., Ashraf, W. and Baumeister, W. (1997) Cloning, sequencing and expression of VAT, a CDC48/p97 ATPase homologue from the archaeon *Thermoplasma acidophilum*. *FEBS Lett.* 404: 263–268.

Philippe, H. and Adoutte, A. (1998) The molecular phylogeny of Eukaryota: Solid facts and uncertainties. In *Evolutionary Relationships among Protozoa* (Coombs, G., Vickerman, K., Sleigh, M. and Warren, A., Eds.), Kluwer, Dordrecht, pp. 25–56.

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. R. Soc. Lond. B Biol. Sci.* 267: 1213–1221.

Pryer, N. K., Wuestehube, L. J. and Schekman, R. (1992) Vesicle-mediated protein sorting. *Annu. Rev. Biochem.* 61: 471–516.

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.

Rachel, R., Wyschkony, I., Riehl, S. and Huber, H. (2002) The ultrastructure of *Ignicoccus*: Evidence for a novel outer membrane and for intracellular vesicle budding in an archeon. *Archaea* 1: 9–18.

Rapoport, T. A., Jungnickel, B. and Kutay, U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* 65: 271–303.

Robibaro, B., Hoppe, H. C., Yang, M., Coppens, I., Ngo, H. M., Stedman, T. T., Paprotka, K. and Joiner, K. A. (2001) Endocytosis in different lifestyles of protozoan parasitism: Role in nutrient uptake with special reference to *Toxoplasma gondii*. *Int. J. Parasitol.* 31: 1343–1353.

Robibaro, B., Stedman, T. T., Coppens, I., Ngo, H. M., Pypaert, M., Bivona, T., Nam, H. W. and Joiner, K. A. (2002) *Toxoplasma gondii* Rab5 enhances cholesterol acquisition from host cells. *Cell Microbiol.* 4: 139–152.

Robinson, M. S. and Bonifacino, J. S. (2001) Adaptor-related proteins. *Curr. Opin. Cell Biol.* 13: 444–453.

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

Rutherford, S. and Moore, I. (2002) The *Arabidopsis* Rab GTPase family: Another enigma variation. *Curr. Opin. Plant Biol.* 5: 518–528.

Rybin, V., Ullrich, O., Rubino, M., Alexandrov, K., Simon, I., Seabra, M. C., Goody, R. and Zerial, M. (1996) GTPase activity of Rab5 acts as a timer for endocytic membrane fusion. *Nature* 383: 266–269.

Saito-Nakano, Y., Nakazawa, M., Shigeta, Y., Takeuchi, T. and Nozaki, T. (2001) Identification and characterization of genes encoding novel Rab proteins from *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 116: 219–222.

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.

Schekman, R. and Mellman, I. (1997) Does COPI go both ways? *Cell* 90: 197–200.

Schledzewski, K., Brinkmann, H. and Mendel, R. R. (1999) Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. *J. Mol. Evol.* 48: 770–778.

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.

Shorter, J. and Warren, G. (2002) Golgi architecture and inheritance. *Annu. Rev. Cell Dev. Biol.* 18: 379–420.

Silberman, J. D., Simpson, A. G., Kulda, J., Cepicka, I., Hampl, V., Johnson, P. J. and Roger, A. J. (2002) Retortamonad flagellates are closely related to diplomonads: Implications for the history of mitochondrial function in eukaryote evolution. *Mol. Biol. Evol.* 19: 777–786.

Simpson, A. G., Radek, R., Dacks, J. B. and O'Kelly, C. J. (2002a) How oxymonads lost their groove: An ultrastructural comparison of *Monocercomonoides* and excavate taxa. *J. Euk. Microbiol.* 49: 239–248.

Simpson, A. G., Roger, A. J., Silberman, J. D., Leipe, D. D., Edgcomb, V. P., Jermiin, L. S., Patterson, D. J. and Sogin, M. L. (2002b) Evolutionary history of "early-diverging" eukaryotes: The excavate taxon *Carpediemonas* is a close relative of *Giardia*. *Mol. Biol. Evol.* 19: 1782–1791.

Smith, T. F., Gaitatzes, C., Saxena, K. and Neer, E. J. (1999) The WD repeat: A common architecture for diverse functions. *Trends Biochem. Sci.* 24: 181–185.

Sogin, M. L. (1991) Early evolution and the origin of eukaryotes. *Curr. Opin. Genet. Dev.* 1: 457–463.

Springer, S., Spang, A. and Schekman, R. (1999) A primer on vesicle budding. *Cell* 97: 145–148.

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

Stanley, H., Botas, J. and Malhotra, V. (1997) The mechanism of Golgi segregation during mitosis is cell type-specific. *Proc. Natl. Acad. Sci. USA* 94: 14467–14470.

Stedman, T. T., Sussmann, A. R. and Joiner, K. A. (2003) *Toxoplasma gondii* Rab6 mediates a retrograde pathway for sorting of constitutively secreted proteins to the Golgi complex. *J. Biol. Chem.* 278: 5433–5443.

Sulli, C. and Schwartzbach, S. D. (1995) The polyprotein precursor to the *Euglena* light-harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing. *J. Biol. Chem.* 270: 13084–13090.

Sulli, C., Fang, Z., Muchhal, U. and Schwartzbach, S. D. (1999) Topology of *Euglena* chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J. Biol. Chem.* 274: 457–463.

Tanigawa, G., Orci, L., Amherdt, M., Ravazzola, M., Helms, J. B. and Rothman, J. E. (1993) Hydrolysis of bound GTP by ARF protein triggers uncoating of Golgi-derived COP-coated vesicles. *J. Cell Biol.* 123: 1365–1371.

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.

van den Ent, F., Amos, L. A. and Lowe, J. (2001) Prokaryotic origin of the actin cytoskeleton. *Nature* 413: 39–44.

Ye, Y., Meyer, H. H. and Rapoport, T. A. (2001) The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. *Nature* 414: 652–656.

Zhang, X., Shaw, A., Bates, P. A., Newman, R. H., Gowen, B., Orlova, E., Gorman, M. A., Kondo, H., Dokurno, P., Lally, J., Leonard, G., Meyer, H., van Heel, M. and Freemont, P. S. (2000) Structure of the AAA ATPase p97. *Mol. Cell* 6: 1473–1484.

Zhu, J., Gong, Z., Zhang, C., Song, C. P., Damsz, B., Inan, G., Koiwa, H., Zhu, J. K., Hasegawa, P. M. and Bressan, R. A. (2002) OSM1/SYP61: A syntaxin protein in *Arabidopsis* controls abscisic acid-mediated and non-abscisic acid-mediated responses to abiotic stress. *Plant Cell* 14: 3009–3028.