matches are for the same genes, so that the substrate specificity of these enzymes is unclear and needs to be examined biochemically.

The *E. histolytica* genome contains one sequence each for the α and β chains of protein farnesyltransferase (EC 2.5.1.58), which were previously cloned and characterised as recombinant proteins (Kumagai *et al.*, 2004).

In addition to the protein farnesyltransferase, a protein geranylgeranyltransferase I (EC 2.5.1.59) β chain has recently been cloned and expressed together with the protein farnesyltransferase α chain (Makioka *et al.*, 2006). The heterodimeric molecule had protein geranylgeranyltransferase activity of unusually broad substrate specificity. The α and β chains of the protein (Rab-)geranylgeranyltransferase II (EC 2.5.1.60) have also been cloned, as cDNAs (M. Kumagai, A. Makioka, T. Takeuchi and T. Nozaki, unpublished data).

The *E. histolytica* genome encodes candidate enzymes for the modification of prenylated proteins. There are two highly divergent proteins both identified as CAAX prenyl proteases (EC 3.4.24.84). CAAX is the carboxy terminus of the substrate protein in which C is the prenylated cysteine residue, A is an aliphatic amino acid and X is the terminal residue. The proteases cleave after the modified cysteine. After the processing step, a prenylcysteine carboxyl methyltransferase (EC 2.1.1.100) methylates the carboxy-terminal residue; there are two divergent candidate genes for this enzyme.

Taken together, the *E. histolytica* genome contains all the necessary genes to encode the pathway from isopentenyl diphosphate to a processed farnesylated or geranylgeranylated protein. The source of the starting material, isopentenyl diphosphate, remains unknown at this time, but there may be a previously unknown pathway for its synthesis or *E. histolytica* may be able to aquire it from its environment.

4.5.1.2. Fatty acid biosynthesis *E. histolytica* encodes an unusual 138 kDa acetyl-CoA carboxylase with 2 bacterial-type carboxylase domains, an acetyl-CoA carboxylase and a pyruvate carboxylase. Since no biotin carboxylase domain is found in the *E. histolytica* genome, it was proposed that the enzyme removes a carboxyl group from oxaloacetate and transfers it to acetyl-CoA, forming malonyl-CoA and pyruvate (Jordan *et al.*, 2003; Loftus *et al.*, 2005). This fusion protein has not been identified in any organisms other than *Giardia* and *Entamoeba*.

In the classical pathway of fatty acid biosynthesis, starting from acetyl-CoA sequential two-carbon units are added from malonyl-CoA. In each round of extension, the β-keto group is reduced in three steps before a new two-carbon unit is added. The whole pathway is carried out in a large fatty acid synthase complex, where the growing chain is linked to an acyl carrier protein. *E. histolytica* lacks this classical pathway. There are,

however, plant homologues of fatty acid chain elongases such as *Arabidopsis thaliana* KCS1 (Todd *et al.*, 1999). There are eight putative fatty acid elongases in the *E. histolytica* genome, and all are very similar to each other. These enzymes could be involved in elongation of fatty acids taken up from the host or food sources, but their function and substrate specificity are unknown at this time.

4.5.2. Phospholipid metabolism

Phospholipids amount to 60–70% of the total lipids in *E. histolytica* (Sawyer *et al.*, 1967). So far little information is available at the biochemical level on how phospholipids are synthesised, acquired or remodelled. The genome project has revealed a number of genes, indicating that the phospholipid metabolism could be more complex than expected.

4.5.2.1. Phospholipid biosynthesis In order to produce phospholipids one has to generate the important intermediate phosphadidate (1,2-diacylglycerol 3-phosphate) by phosphorylation and acylation of glycerol. *E. histolytica* contains one gene for a glycerol kinase (EC 2.7.1.30). The second step would be the transfer of the acyl group to glycerol-3-phosphate by glycerol-3-phosphate *O*-acyltransferase (EC 2.3.1.15), but no candidate gene for this enzyme has been found in the genome. There are, however, two potential 1-acylglycerol-3-phosphate *O*-acyltransferases (EC 2.3.1.51) that could attach the second acyl group. After the attachment of the acyl groups, and in preparation for the attachment of the activated aminoalcohols, the phosphate is removed by phosphadidate phosphatase (EC 3.1.3.4), for which there is one gene, resulting in a diacylglycerol.

The activation of ethanolamine (EC 2.7.1.82) or choline (EC 2.7.1.32) for attachment to the phosphadidate starts with phosphorylation. There are two genes identified as choline/ethanolamine kinases that share 37% amino acid identity. Next, ethanolamine phosphate and choline phosphate are converted into CDP-ethanolamine (EC 2.7.7.14) and CDP-choline (EC 2.7.7.15), respectively. The genome encodes 2 enzymes sharing 57% sequence identity that are identified as ethanolamine-phosphate cytidylyltransferases. The substrate specificity of these enzymes needs to be examined on the biochemical level. Finally, the activated ethanolamine or choline is attached to diacylglycerol by the enzymes ethanolaminephosphotransferase (EC 2.7.8.1) or diacylglycerol cholinephosphotransferase (EC 2.7.8.2) producing phosphatidylethanolamine or phosphatidylcholine, respectively. For these activities a total of eight possible genes are found that share varying degrees of sequence similarity.

In *E. histolytica*, an alternative pathway of phospholipid biosynthesis could involve the biosynthesis of phosphatidylserine. In this pathway, the phosphatidate itself is activated by CTP in a reaction catalysed by phosphatidate cytidylyltransferase (EC 2.7.7.41) resulting in CDP-diacylglycerol. Three genes have been identified. Phosphatidylserine synthase then catalyses the reaction of CDP-diacylglycerol with serine to give phosphatidylserine (EC 2.7.8.8); one gene has been found.

Some organisms can form phosphatidylethanolamine from phosphatidylserine using a decarboxylase, but such an enzyme appears to be absent from the *E. histolytica* genome. There are, however, several candidate methyltransferases of yet unknown substrate specificity, which might be able to generate phosphatidylcholine from phosphatidylethanolamine.

Taken together, large portions of the pathways needed to generate the most important phospholipids can be assembled from genes tentatively identified to date in the *E. histolytica* genome. The first acylation of glycerol 3-phosphate to lysophosphatidate remains an important gap. As *E. histolytica* could potentially aquire all the necessary phospholipids from the host, the functional relevance of the described biosynthetic pathways may not be high.

Finally, two additional interesting enzymes present in *E. histolytica* should be mentioned. The first was previously characterised using cDNA sequences and recombinant proteins as L-*myo*-inositol 1-phosphate synthase (EC 5.5.1.4; Lohia *et al.*, 1999). This enzyme catalyses the complicated isomerisation of glucose 6-phosphate to L-*myo*-inositol 1-phosphate. Inositol is found in phosphatidylinositol (PI) and in GPI-anchors of some membrane proteins, as well as playing a major role in signal transduction via the secondary messenger 1,4,5-inositol trisphosphate. There are three *myo*-inositol 1-phosphate synthase genes, all highly similar to each other and to the previously sequenced cDNA.

The second is phospholipid-cholesterol acyltransferase (EC 2.3.1.43), which transfers an acyl group from phospholipids such as phosphatidylcholine to cholesterol giving a cholesterol ester. The genome contains seven genes for this enzyme. So far nothing is known about the importance of cholesterol esters for *E. histolytica*.

4.5.2.2. Phospholipid degradation Phospholipids are degraded by phospholipases. Whereas phospholipases A1 (EC 3.1.1.32) and A2 (EC 3.1.1.4) cleave the acyl residues in the 1 or 2 position of the glycerol core, phospholipases C (EC 3.1.4.3) and D (EC 3.1.4.4) cleave at the phosphate, phospholipase C on the glycerol side, and phospholipase D on the aminoalcohol side. In *E. histolytica* phospholipase A activity has been implicated in virulence (Ravdin *et al.*, 1985), as it liberates toxic fatty acids and

lysophospholipids (Said-Fernandez and Lopez-Revilla, 1988). Phospholipases A have been found in two forms, a membrane-bound Ca-dependent form active at alkaline pH and a soluble Ca-independent form active at acid pH (Long-Krug *et al.*, 1985; Vargas-Villarreal *et al.*, 1998). The genome encodes 11 potential phospholipases A with predicted pI values between 4.8 and 8.8 and various degrees of sequence similarity. In addition, the *E. histolytica* genome encodes three potential phospholipases D.

Finally, there are two highly similar genes for phospholipases C, but these are homologous to PI-specific phospholipases C (EC 3.1.4.11) and most likely do not cleave PI or phosphatidylcholine but GPI-anchors instead. So far there are no studies using individual recombinant phospholipases, and it is not yet known how much these enzymes may contribute to the virulence of *E. histolytica*.

4.6. Coenzyme A biosynthesis and pantothenate metabolism

Analysis of the genome revealed a complete lack of known folate-dependent enzymes and folate transporters, suggesting this cofactor is not utilised by *E. histolytica*. This is at odds with a study on the nutritional requirements of *E. histolytica* in which folate was found to be essential for growth (Diamond and Cunnick, 1991). More experimental research will be needed to resolve this discrepancy. Most organisms require folate as a cofactor for several reactions of amino acid metabolism and for synthesis of thymidylate, a component of DNA. The microsporidian *E. cuniculi*, which possesses the smallest-known eukaryotic genome, still contains a folate transporter and several folate-dependent enzymes (Katinka *et al.*, 2001). In eukaryotes possessing mitochondria or chloroplasts, folate is required for the formylation of methionine on the initiator tRNA used for organelle protein synthesis. Although *E. histolytica* possesses a mitochondrion-derived organelle, the mitosome, there is no organellar genome (Leon-Avila and Tovar, 2004) and so no need for organellar protein synthesis. The most important metabolic consequences of the loss of folate metabolism for *E. histolytica* are therefore the absence of thymidylate synthesis and methionine recycling, although it remains possible that *E. histolytica* possesses folate-independent enzymes carrying out these steps.

Phosphopantothenoyl-cysteine decarboxylase (EC 4.1.1.36) and phosphopantothenoyl-cysteine synthetase (EC 6.3.2.5, synonymous with phosphopantothenate-cysteine ligase) exist as a fusion protein in *E. histolytica*, as in Bacteria and Archaea. The amino- and carboxyl-terminal domains possess decarboxylase and synthetase activity, respectively (Kupke, 2002, 2004; Kupke *et al.*, 2000; Strauss *et al.*, 2001). The role of this enzyme in coenzyme A biosynthesis is not well understood in *E. histolytica* as the other necessary enzymes are absent.

4.7. Nucleic acid metabolism

Like many protistan parasites, *E. histolytica* lacks *de novo* purine synthesis (Reeves, 1984). The genome reveals that nucleic acid metabolism of *E. histolytica* is similar to that of the other lumenal parasites *G. intestinalis* and *T. vaginalis* in lacking pyrimidine synthesis and thymidylate synthase (Aldritt *et al.*, 1985; Wang and Cheng, 1984). In addition, *E. histolytica* appears to lack ribonucleotide reductase, a characteristic shared with *G. intestinalis* (Baum *et al.*, 1989). Ribonucleotide reductase was found, however, in genomic sequences of the species *E. invadens* and *E. moshkovskii*, indicating that the enzyme was lost or replaced relatively recently. Among eukaryotes, the loss of these areas of nucleic acid metabolism is otherwise rare. The enzymes were likely lost during adaptation to living in an organic nutrient rich environment.

4.8. Missing pieces

Several important enzymes and pathways could not be found within the genome and their presumed sequence divergence from known enzymes and pathways labels them as possible drug targets once they are identified. Phosphopyruvate carboxylase, which reversibly converts phosphoenolpyruvate to oxaloacetate, is a central enzyme of carbon metabolism in *E. histolytica* (Reeves, 1970), but could not be identified. Isoprenyl-PP synthesis and aminoethylphosphonate synthesis are also likely to be present, but no candidate genes could be identified.

4.9. Transporters

A total of 174 transporters were identified within the genome, a number intermediate between the 62 transporters of *P. falciparum* and the 286 transporters of *S. cerevisiae* (http://membranetransport.org). *E. histolytica* has a number of ion transporters similar to those of yeast, but fewer identifiable nutrient and organellar transporters. Both *Plasmodium* and *Entamoeba* have reduced metabolisms and take up many complex nutrients. The higher number of transporters in *Entamoeba* suggests that they may be more substrate specific than the *Plasmodium* transporters or that they may have a higher level of redundancy.

Since glucose transport activity has experimentally been characterised in *E. histolytica* and glucose is thought to be the major energy source, it was surprising to find no homologues of known hexose transporters in the genome. Most hexose transporters belong to the sugar porter subfamily of the major facilitator superfamily (TC 2.A.1.1), members of which are found in prokaryotes, animals, fungi, plants and other protists, including *D. discoideum*, but no proteins of this family were found in the

E. histolytica genome. A group of candidate monosaccharide transporters found within the genome are related to the glucose/ribose porter family from prokaryotes (TC 2.A.7.5). These transporters consist of two related domains, and the *Entamoeba* proteins appear to have the N-terminal and C-terminal domains switched relative to the bacterial proteins. Functional characterisation of transporter-encoding genes will be necessary for a more complete picture.

5. THE CYTOSKELETON

The eukaryotic cytoskeleton is composed of three main elements: actin microfilaments, tubulin-based microtubules and intermediate filaments. Despite the fact that *E. histolytica* is very motile and performs phagocytosis very efficiently, its cytoskeletal components are simple. No genes encoding homologues of intermediate filament network proteins, including keratins, desmin and vimentin, have been identified in *E. histolytica*, providing further evidence that these particular cytoskeletal components are rather poorly conserved in evolution. In contrast, microfilament and microtubule components have been readily identified. Additional detail is given in the supplementary tables for this section.

5.1. Actin and microfilaments

Genome information suggests that *E. histolytica* has a greater dependence than other protists on an actin-rich cytoskeletal network. Microfilament proteins are represented by actin and several actin-binding proteins, although there are notable differences with respect to analogous proteins in other eukaryotes. There are eight actin genes in the *E. histolytica* genome, in addition to six others that encode divergent actins. Three divergent actins surprisingly contain an extra N-terminal domain with as yet unknown functional characteristics. Examples of hybrid actins are rather scarce and have been found as ubiquitin fusions (Archibald *et al.*, 2003). The functional significance of these *E. histolytica* hybrid actins is as yet unknown.

Under physiological salt concentrations, monomeric actin assembles into polymers of F-actin, thus building microfilaments. Actin assembles and disassembles in an extremely dynamic and highly controlled process that is dependent on many different actin-binding proteins (Winder and Ayscough, 2005). The *E. histolytica* genome encodes homologues of actin-binding proteins involved in the severing, bundling, cross-linking and capping of filamentous actin. The number and variety of actin-binding proteins support the view that the actin-rich cytoskeleton is very dynamic in *E. histolytica*.

Since the spontaneous polymerisation of actin monomers is inhibited by the action of sequestering proteins such as thymosin $\beta 4$ and profilin, efficient actin polymerisation requires the intervention of an actin polymerisation-promoting factor. The best-described promoting factors are the Arp2/3 complex and the formin protein family.

The Arp2/3 complex is composed of two actin-related proteins (Arp2 and Arp3, which act as a template for new actin filaments) and works in conjunction with five additional subunits: ARPC1 to 5 (Vartiainen and Machesky, 2004). All subunits have clearly been identified in the *E. histolytica* genome, and among these the Arp2 and Arp 3 subunits are the best conserved. The Arp2/3 complex's ability to nucleate new actin filaments is stimulated by its interaction with nucleation promoting factors such as the Wiskott–Aldrich syndrome protein (WASP) or the suppressor of cAMP-receptor (SCAR) factor. Surprisingly, no proteins with homology to WASP/SCAR components were found in the genome, suggesting that actin nucleation depends on the activity of other, as yet unidentified proteins.

In contrast, *E. histolytica* possesses six genes coding for formins, which have emerged as potent regulators of actin dynamics in eukaryotic cells through their ability to increase actin filament assembly (Higgs and Peterson, 2005). Formins control rearrangements of the actin cytoskeleton, especially in the context of cytokinesis and cell polarisation. Members of this family have been found to interact with Rho-GTPases, profilin and other actin-associated proteins. The precise nature of this polymerisation-accelerating activity differs from one formin to another: some nucleate filaments *de novo*, some require profilin for effective nucleation, while yet others seem to use filament severing as their basic mechanism. However, the formin homology 2 domain (FH2, comprising roughly 400 amino acids) is central to formin activity (Otomo *et al.*, 2005; Xu *et al.*, 2004). Actin nucleation by formins is thought to occur by stabilisation of an unfavourable nucleation intermediate, possibly through FH2 domains binding to monomers in the same manner that they bind to barbed ends (an activity influenced by profilin). The formin homologues from *E. histolytica* all contain an FH2 domain, suggesting that they are potential actin nucleation factors.

Once nucleated, actin filaments are able to grow rapidly by addition of monomers at their barbed ends. Filaments are regulated by several mechanisms (Winder and Ayscough, 2005). Filament length is controlled by capping proteins: barbed end cappers (such as capping protein and gelsolin) block addition of new monomers and thus act to decrease the overall length of the filament. In addition, gelsolin severs actin filaments, thereby rapidly increasing actin dynamics. Actin filaments appear to be significantly shorter in *E. histolytica* when compared with those from fibroblasts and stress fibres are not formed in this amoeba. Although

E. histolytica actin has been shown not to bind DNase I (Meza *et al.*, 1983), the inferred amino acid sequence indicates conservation of all the residues likely to participate in this binding event—suggesting that post-translational modifications of actin monomers may prevent DNAse I -actin binding. It remains to be determined whether such modifications of actin participate in the regulation of actin polymerisation. The genome encodes multiple genes associated with filament capping and severing, as well as candidates for proteins that cross-link actin filaments and thus organise them into a supramolecular network. The organisation of actin into networks and higher-order structures is crucial for both cell shape and function. These structures can be responsible for overall cell shape and related processes, such as bundle formation through α-actinin activity, for example. The arrangement of actin filaments into cross-linked arrays is also mediated by proteins with multiple actin-binding domains, which allows a more perpendicular arrangement of actin filaments. Examples of this type of protein are the large, flexible filamin dimer (Vargas *et al.*, 1996) and the spectrin tetramer. Genome analysis has now identified many candidate genes for actin-binding proteins in *E. histolytica*, and additional protein partners of this versatile family responsible for cytoskeleton regulation are likely to emerge from curation of the sequence and cellular studies of cell motility and phagocytosis in this parasite.

5.2. Tubulins and microtubules

E. histolytica has a lower dependence on a tubulin-based cytoskeleton than most other eukaryotic cells. Protein homologues of the basic (α, β and γ) tubulins are present, although other tubulins more characteristic of organisms with basal bodies and flagella (e.g., ε- and δ-tubulins) are absent from *E. histolytica* (Dutcher, 2001). Nine different tubulins (grouped into multigene families) exist in most eukaryotic cells. Microtubules (MTs) composed of α- and β-tubulins are intranuclear in *E. histolytica* (Vayssie *et al.*, 2004), and this raises the question of how such structures are modulated within the nucleus, given that MT dynamics require MT nucleation-based renewal at the minus end and MT capping at the plus end. Proteins involved in MT nucleation act in concert with γ-tubulin (which is also intra-nuclear in *E. histolytica*), and this parasite possesses at least one homologue to the Spc98 factor, a component of the MT-nucleating Tub4p-γ tubulin complex. In contrast, no homologues of EB1, CLIP-170, APC (all involved in MT capping) or centrins (which operate at the MT organising centre) have yet been identified, suggesting that other factors (or mechanical constraints within the MT) may be required in blocking MT growth. *E. histolytica* does encode candidate proteins involved in MT severing or chromosome segregation. All these proteins are good candidates for experimental analysis of the mechanisms of intranuclear

MT localisation and turnover as well as of the trafficking of tubulins between the cytoplasm and nucleus.

There is little information available on the precise organisation of microtubules and F-actin cytoskeleton during *E. histolytica* motility. In many eukaryotic cells, F-actin–microtubule interactions can be observed in lamellipodia at all stages. Interestingly, microtubules preferentially grow along actin bundles in filopodia, suggesting that a physical link between the structures exists (Leung *et al.*, 2002). Multifunctional MT-associated proteins (MAPs, like MAP1B, MAP2 and plakins) are promising candidates for acting as such links, either via dimerisation of MAPs with single microtubule and actin binding sites or via direct bridging of the two cytoskeletons (e.g., via plakins, which contain binding sites for both microtubules and actin within a single molecule). Plakin homologues have not been identified in the *E. histolytica* genome but a MAP is present. Furthermore, proteins with domains that can bind to actin (and potentially to MT) have been described in *E. histolytica*—the actin binding protein ABP-120 gelation factor, for example (Vargas *et al.*, 1996).

5.3. Molecular motors

The distribution of intracellular factors and vesicles is performed using three sets of molecular transporters: myosin along microfilaments and kinesin and dynein along MTs. Although *E. histolytica* is a highly motile cell, stress fibres and cytoplasmic MTs have never been observed. The fluidity of the parasite's cytoplasm may be related to features of its molecular motors some of which are very surprising. The myosin family of actin filament-based molecular motors consists of at least 20 structurally and functionally distinct classes. The human genome contains nearly 40 myosin genes, representing 12 of these classes. Remarkably, *E. histolytica* is the first reported instance of a eukaryote with only one unconventional myosin. This myosin heavy chain (myosin IB) belongs to the type I myosin family, of which 12 are present in the *Dictyostelium* genome (Eichinger *et al.*, 2005).

All members of the myosin family share a common structure composed of three modules: the head, neck and tail domains. The N-terminal region harbours the motor unit, which uses ATP to power movement along the actin filaments. By interacting with specific proteins and 'cargoes' the tail is responsible for the myosin's specific function and location. In particular, the presence of an SH3 domain in the tail region is important for linking these myosin I molecules with the endocytic machinery and the Arp2/3 complex. Protistan class I myosins are able to recruit the Arp2/3 complex towards the CARMIL adapter protein and Acan125. These homologous adapters consist of multiple, leucine-rich repeat sequences and bear two carboxyl-terminal polyproline motifs

that are ligands for the myosin I SH3 domains. CARMIL has been shown to bind the Arp2/3 complex via an acidic motif similar to those found in WASP. In view of the fact that *E. histolytica* does not have WASP homologues, the discovery of a CARMIL homologue through proteomic analysis of *E. histolytica* phagosomes (Marion *et al.*, 2005) provides an important clue for understanding actin nucleation in *E. histolytica*. Interestingly, myosin IB in *E. histolytica* plays a structural role in the actin network due to its ability to cross-link filaments (Marion *et al.*, 2004). The cytoskeletal structuring activity of myosin IB regulates the gelation state of cell cytoplasm and the dynamics of cortical F-actin during phagocytosis.

The most-studied myosin has been the conventional or class II myosin. This double-headed molecule is composed of two heavy chains and two pairs of essential and regulatory light chains. The heavy chain tail consists of an α-helical, coiled coil protein able to form a parallel dimer that in turn can self-associate into bipolar, thick filaments. This enables myosin II to operate in huge filament arrays, which drive high speed motility. In addition to myosin IB, *E. histolytica* also has a conventional myosin II heavy chain (very closely related to its homologue in *Dictyostelium*), which has been reported to be involved in crucial phases of parasite motility, surface receptor capping and phagocytosis (Arhets *et al.*, 1998). *E. histolytica*'s sole isoform shapes the actin network and maintains cytoskeletal integrity. Candidate genes for the regulatory and essential light chain activities were also found, and these possess the EF hand domains necessary for Ca^{2+} binding.

Directional transport along the MTs depends on dynein and kinesin, both MT-associated motor proteins that convert the chemical energy from ATP hydrolysis into movement. These motors are unidirectional and move towards either the MT plus- or minus-ends (Mallik and Gross, 2004). Kinesins and dyneins have been implicated in a wide range of functions—principally intracellular organelle transport during interphase and spindle function during mitosis and meiosis. Members of the dynein family are minus-end directed, although this remains to be confirmed for a few uncharacterised, vertebrate, cytoplasmic dynein heavy chains. It has not yet been reliably established that the *E. histolytica* genome contains a dynein heavy chain gene, although a dynein light chain gene is present: improvements in gene assembly should provide us with more information on this high molecular mass protein.

Kinesins are microtubule-dependent molecular motors that play important roles in intracellular transport and cell division. Even though the motor domain is found within the N-terminus in most kinesins (N-type), it is located within the middle or C-terminal domains in some members of the family (M-type and C-type kinesins, respectively) (Asbury, 2005). The position of the motor domain dictates the polarity of the movement of kinesin along the MT: whereas N- and M-type kinesins

are plus-end directed, the C-type kinesins are minus-end directed. Humans possess 31 different kinesins and trypanosomes have more than 40. The *E. histolytica* genome sequence predicts only six kinesin-encoding genes (four N-type, two C-type and no M-type homologues have been found). One of the N-kinesins also contains a domain homologous to the HOOK protein required for the correct positioning of microtubular structures within the cell (Walenta *et al.*, 2001). Considering that *E. histolytica* MTs are intranuclear, the study of kinesin function and trafficking should help elucidate what is likely to be a very interesting MT functional mechanism.

6. VESICULAR TRAFFIC

The requirement for nutritional uptake from the extracellular milieu in the host intestine imposes a heavy reliance on endocytic and phagocytic activities in *Entamoeba* (Espinosa-Cantellano and Martínez-Palomo, 2000). Proliferating trophozoites secrete a number of peptides and proteins including cysteine proteases (Que and Reed, 2000) and amoebapores (Leippe, 1999) required for bacterial cell killing and degradation as well as being implicated in virulence (Petri, 2002). During encystation, the cells also secrete substrates used for the formation of the cyst wall (Eichinger, 1997). Electron micrographic studies have revealed a complex membrane organisation. The trophozoites contain numerous vesicles and vacuoles varying in size and shape (Clark *et al.*, 2000; Mazzuco *et al.*, 1997). Intracellular transport of both endocytosed and synthesised molecules between compartments is regulated by the elaborate orchestration of vesicle formation, transport, docking and fusion to the target compartment (Bonifacino and Glick, 2004; Kirchhausen, 2000). More detail is given in the supplementary tables for this section.

6.1. Complexity of vesicle trafficking

Among a number of molecules and structures involved in vesicular trafficking, three types of coated vesicles, named coatomer protein (COP) I, COPII, and clathrin-coated vesicles are the best characterised (Bonifacino and Glick, 2004; Kirchhausen, 2000). COPI vesicles primarily mediate transport from the Golgi to the ER and between the Golgi cisternae, while COPII vesicles are involved in the transport from the ER to the *cis*-Golgi. The clathrin-dependent pathway has a few independent routes: from the plasma membrane to endosomes, from the Golgi to endosomes and from endosomes to the Golgi. It has been well established that certain subfamilies of Ras-like small GTPases, widely conserved among eukaryotes, regulate both the formation of transport vesicles and their docking and fusion to the target organelles. The ARF and secretion-associated

Ras-related protein (Sar) families of GTPases regulate the formation of COPI and COPII vesicles (Memon, 2004), respectively. In contrast, the Rab family of GTPases (Novick and Zerial, 1997) is involved in the targeting and fusion of vesicles to the acceptor organelles together with the tethering machinery SNARE (a soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor) (Chen and Scheller, 2001). Since individual coat proteins, small GTPases, SNAREs and their associated proteins show distinct intracellular distributions in both unicellular and multicellular organisms, they are believed to play a critical role in the determination of membrane trafficking specificity (Chen and Scheller, 2001; Munro, 2004; Novick and Zerial, 1997). It is generally believed that the total number of proteins involved in the membrane traffic reflects the complexity and multiplicity of its organism. The total number of the putative amoebic genes encoding Arf/Sar, Rab, SNARE, and coat proteins together with those from *S. cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Homo sapiens* and *A. thaliana*, is shown in Table 2.7. *E. histolytica* reveals complexity similar to yeast, fly and worm in the case of Sar/Arf and SNAREs, while the number of genes encoding three coat proteins [COPI, COPII and adapter proteins (APs)] was higher in *E. histolytica* than these organisms and comparable to that in mammals and plants. In contrast, the number of Rab proteins in *E. histolytica* is exceptionally high, exceeding that in mammals and plants.

6.2. Proteins involved in vesicle formation

6.2.1. COPII-coated vesicles and Sar1 GTPase

COPII components were originally discovered in yeast using genetic and biochemical approaches (reviewed in Bonifacino and Glick, 2004). COPII vesicles mediate the transport from the ER to the Golgi and consists of three major cytosolic components and a total of five essential proteins: the Sec23p–Sec24p complex, the Sec13p–Sec31p complex and the small GTPase Sar1p (Barlowe *et al.*, 1994). Sar1p and Sec23p–Sec24p complex are involved in the formation of the membrane–proximal layer of the coat, while Sec13p–Sec31p complex mediates the formation of the second membrane–distal layer (Shaywitz *et al.*, 1997). These proteins are well conserved among various organisms (Table 2.7). *E. histolytica* encodes one each of Sar1, Sec13 and Sec31, 2 of Sec23 and 5 proteins corresponding to Sec24 (Table 2.7). The yeast and human genomes also encode multiple Sec24 isotypes (3 and 4, respectively). Although Sec24 isotypes have been shown to be responsible for the selection of transmembrane cargo proteins in yeast (Peng *et al.*, 2000; Roberg *et al.*, 1999), the significance of the Sec24 redundancy in *E. histolytica* is not clear. Additional regulatory proteins participate in COPII assembly in yeast, including Sec16p, a putative scaffold protein (Espenshade *et al.*, 1995), and Sec12p, a GEF

TABLE 2.7 The number of genes encoding representative proteins involved in vesicular trafficking in *E. histolytica*

Protein	*E. histolytica*	*S. cerevisiae*	*C. elegans*	*D. melanogaster*	*H. sapiens*	*A. thaliana*	References
Sar1	1	1	1	1	2	4	1,2
COPII	9	6	5	4	9	12	1
Arf	10	6	11	11	27	17	3
COPI	11	7	7	7	9	9	1
AP-1		5	7	5	8	9	
AP-2		4	5	5	5	6	
AP-3		4	4	4	7	4	
AP-4		0	0	0	4	4	
AP total	18	13	16	14	24	23	1
Rab	91	11	29	29	60	57	1,4
Qa	8	7	9	7	12	18	
Qb	10(b + c)	6	7	5	9	11	
Qc		8	4	5	8	8	
R	10	5	6	5	9	14	
SNARE total	28	24	23	20	35	54	1,5,6
NSF	1	1	1	2	1	1	1,7
SNAP	1	1	1	3	1	3	
Sec1	5	4	6	5	7	6	8

References: (1): Bock *et al.* (2001); (2): Wennerberg *et al.* (2005); (3): Kahn *et al.* (2006); (4): Pereira-Leal and Seabra (2001); (5): Burri and Lithgow (2004); (6): Uemura *et al.* (2004); (7): Sanderfoot *et al.* (2000); (8): Boehm *et al.* (2001).

for Sar1p (Barlowe and Schekman, 1993). Homologues of Sec12p and Sec16p appear to be absent in *E. histolytica*. The p24 protein is a non-essential component of vesicle formation (Springer *et al.*, 2000), and in yeast it functions as a cargo adaptor through binding to Sec23p (Kaiser, 2000; Schimmoller *et al.*, 1995). *E. histolytica* encodes four p24 proteins, fewer than in yeast and humans, which have eight. GAP Sec23p is also present in *E. histolytica*; this activates the intrinsic GTPase activity of Sar1p after the formation of COPII vesicle and inactivates the function of Sar1p (Yoshihisa *et al.*, 1993), resulting in the uncoating of COPII vesicles.

6.2.2. COPI-coated vesicles and Arf GTPases

COPI-coated vesicles, which mediate transport from the Golgi to the ER and between the Golgi cisternae (Kirchhausen, 2000), consist of seven proteins (α, β, β', γ, δ, ε and ζ-COP) (Hara-Kuge *et al.*, 1994). The number of proteins making up the COPI coat, and thus the complexity of COPI components, varies among organisms (Table 2.7). While human possesses two isotypes of γ-COP and ζ-COP, yeast has a single gene for each. In humans, the two isotypes of γ-COP and ζ-COP form three different COPI complexes ($\gamma1/\zeta1$, $\gamma1/\zeta2$ and $\gamma2/\zeta1$), which have different intracellular distributions (Wegmann *et al.*, 2004). This implies that COPI-coated vesicles are also involved in functions other than Golgi-to-ER transport (Whitney *et al.*, 1995). In *E. histolytica*, the COPI complex appears more heterogeneous: *E. histolytica* encodes two isotypes each of γ-COP, δ-COP and α-COP and three isotypes of β-COP. In contrast, *E. histolytica* lacks ϵ-COP, which is known to stabilise α-COP (Duden *et al.*, 1998). It has been shown in yeast that all genes encoding components of COPI coat except for Sec28p, the yeast ε-COP homologue, are essential for growth (Duden *et al.*, 1998).

Recruitment of COPI to the Golgi membrane requires the association of a GTP-bound GTPase called Arf (Donaldson *et al.*, 1992; Kahn *et al.*, 2006). Arf was initially identified because of its ability to stimulate the ADP-ribosyltransferase activity of cholera toxin A (Kahn and Gilman, 1984). To recruit the COPI coat, Arfs are activated by a Sec7 domain-containing protein, Arf-GEF, which is a target of a fungal metabolite brefeldin A (Helms and Rothman, 1992; Sata *et al.*, 1998). Among Arf family proteins, Arf1 is involved in the formation of COPI-coated vesicles in the retrograde transport from the Golgi to ER, and is also involved in the assembly of clathrin-AP1 (see next section) on the *trans*-Golgi network (TGN) (Stamnes and Rothman, 1993), clathrin-AP3 on endosomes (Ooi *et al.*, 1998) and the recruitment of AP-4 to the TGN (Boehm *et al.*, 2001). The specific roles of Arfs3–5 are less clear, although Arf4 and Arf5 show *in vitro* activities similar to Arf1. Functional cooperativity of Arfs in the vesicular formation has also been demonstrated recently. At least two of four human Arf isotypes (Arf1, Arf3–5) are essential for a retrograde

pathway from the Golgi to the ER, in the secretory pathway from the Golgi to the TGN and in the recycling from endosomes to the plasma membrane (Volpicelli-Daley *et al.*, 2005). In contrast to these Arfs, Arf6 regulates the assembly of actin filaments and is involved in endocytosis on the plasma membrane (Radhakrishna and Donaldson, 1997).

GTPases that share significant similarity to Arf but do not either activate cholera toxin A or rescue *S. cerevisiae Arf* mutants are known as Arls (Arf-like GTPases) (Lee *et al.*, 1997a). Arl1 is involved in endosome-to-Golgi trafficking (Lu *et al.*, 2001, 2004). Other Arls (Arls 2–11) and Arf-related proteins (Arp or ArfRP 1–2) have been localised to the cytosol, nucleus, cytoskeleton and mitochondria (Burd *et al.*, 2004; Pasqualato *et al.*, 2002). The number of Arf, Arl and Arf-related proteins varies among organisms (Table 2.7). Among 27 members identified in humans, only about a half dozen Arf/Arl/Arp proteins, including Arf1–6 and Arl1 (Wennerberg *et al.*, 2005), have been shown to function in membrane traffic (Lu *et al.*, 2001). The localisation and function of the remaining Arf/Arl/Arp remains unclear.

E. histolytica encodes 10 Arf/Arl proteins (Table 2.7). Only 2 *E. histolytica* Arfs (A1 and A2) have a high percentage identity to human Arfs 1, 3, 5 and 6 and yeast Arfs 1–3 (57–76% identity), while the remaining 8 Arf/Arl fall into 3 groups (A4–6, B1–3 and C) and are equally divergent from one another and from other organisms. Both the intracellular distributions and the specific steps in vesicular trafficking mediated by these *Entamoeba* Arf/Arl proteins are unknown. It is worth noting that five of these Arfs lack a conserved glycine residue at the second amino acid position of the amino terminus; this glycine is known to be myristylated and essential for membrane association in other organisms (Randazzo *et al.*, 1995). *Eh*ArfA4 also lacks one of the conserved GTP-binding consensus regions (Box2). Similar deletion of GTP-binding domains has also been observed in proteins belonging to the Rab family (see Section 6.3.1).

6.2.3. Clathrin-coated vesicle and its adaptor proteins

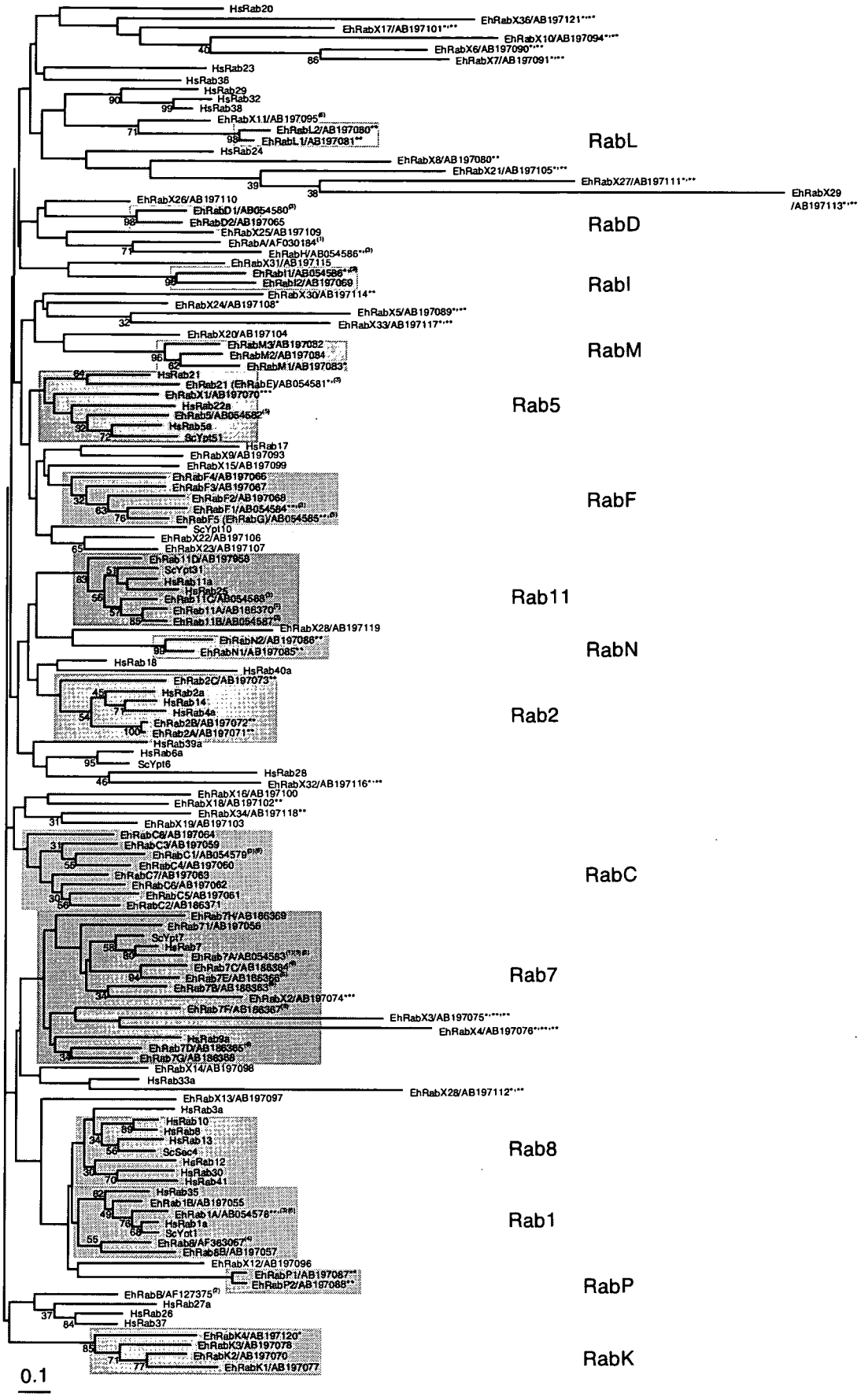
Clathrin-coated vesicles and pits, as demonstrated by electron microscopy, are often indicative of clathrin-mediated endocytosis. However, there is no clear ultrastructural evidence for their occurrence in *Entamoeba* (Chavez-Munguia *et al.*, 2000). Interestingly, heavy- but not light-chain clathrin is encoded in the genome. Since a majority of proteins, including adaptor proteins (APs, adaptins), known to be involved in the assembly of clathrin-coated vesicles are encoded in *E. histolytica*, the fundamental mechanisms and components of clathrin-mediated endocytosis are probably present in this organism, but are likely to be divergent from other eukaryotes. AP is a cytosolic heterotetramer that mainly mediates the integration of membrane proteins into clathrin-coated vesicles in the secretory and endocytic pathways (Boehm and Bonifacino, 2001;

Kirchhausen, 2000). AP is composed of two large, one medium and one small subunit (Keen, 1987). Four major types of AP complexes (AP1–4) have been identified (Boehm and Bonifacino, 2001; Nakatsu and Ohno, 2003). AP-2 (consisting of α, $\beta 2$, $\sigma 2$ and $\mu 2$) mediates endocytosis from the plasma membrane (Conner and Schmid, 2003; Motley *et al.*, 2003), while AP-1 (γ, $\beta 1$, $\sigma 1$ and $\mu 1$A) (Meyer *et al.*, 2000), AP-3 (δ, $\beta 3$A $\sigma 3$ and $\mu 3$A) (Le Borgne *et al.*, 2001; Vowels and Payne, 1998) and AP-4 (ϵ, $\beta 4$, $\sigma 4$ and $\mu 4$) (Aguilar *et al.*, 2001) play a role in the Golgi-endosome, endosomal-lysosomal or the Golgi/lysosome sorting pathway, respectively. AP-4, which is present only in mammals and plants (Boehm and Bonifacino, 2001), was also identified in non-clathrin-coated vesicles mediating the transport from TGN to the plasma membrane or endosomes (Hirst *et al.*, 1999). A few isotypes of AP-1 and AP-3, for example, AP-1B (γ, $\beta 1$, $\sigma 1$ and $\mu 1$B) and AP-3B (δ, $\beta 3$B, $\sigma 3$ and $\mu 3$B), showed tissue-specific expression (Faundez *et al.*, 1998; Folsch *et al.*, 1999). *E. histolytica* encodes 10 large subunits (α, β, γ, δ and ϵ), 4 medium subunits (one each of $\mu 1$ and $\mu 2$ and two $\mu 3$) and 4 small subunits ($\epsilon 1-\epsilon 4$). This suggests that *E. histolytica* produces four types of AP complex, as in humans and plants.

6.3. Proteins involved in vesicle fusion

6.3.1. Rab GTPases

The docking and fusion of transport vesicles to a specific target compartment requires the appropriate Rab protein. Specific interaction of a Rab with its effector molecules in conjunction with the interaction between SNAREs plays a central role in vesicle fusion (Zerial and McBride, 2001). In general, the complexity of the Rab gene family correlates with the degree of multicellularity. For example, *S. pombe*, *S. cerevisiae*, *C. elegans*, *D. melanogaster* and *H. sapiens* consist of 1, 1, $\sim 10^3$, 10^9 and 10^{13} cells, and have 7, 11, 29, 29 and 60 Rab genes, respectively (Pereira-Leal and Seabra, 2001). It has also been shown that in multicellular organisms, Rab proteins are expressed in a highly coordinated (i.e., tissue-, organ-, or developmental stage-specific) fashion (Seabra *et al.*, 2002; Zerial and McBride, 2001). *E. histolytica* possesses an extremely high number of Rab genes—91 (Fig. 2.6). Among its 91 Rabs only 22, including *Eh*Rab1, *Eh*Rab2, *Eh*Rab5, *Eh*Rab7, *Eh*Rab8, *Eh*Rab11, *Eh*Rab21, and their isotypes showed >40% identity to Rabs from other organisms. The 69 remaining *E. histolytica* Rab proteins showed only moderate similarity (<40% identity) and represent unique, presumably *Entamoeba*-specific, Rab proteins. Approximately one-third of Rab proteins form 15 subfamilies, including Rab1, Rab2, Rab7, Rab8, Rab11 and RabC-P, each of which contains up to 9 isoforms. Interestingly, ~70% of *E. histolytica* Rab genes contain one or more introns (Saito-Nakano *et al.*, 2005). SNARE genes are also intron-rich whereas the Sar/Arf GTPase and the three coat protein genes have a low


HsRab20
EhRabX36/AB197121
EhRabX17/AB197101
EhRabX10/AB197094
EhRabX6/AB197090
EhRabX7/AB197091
HsRab23
HsRab36
HsRab29
HsRab32
HsRab38
EhRabX11/AB197095
EhRabL2/AB197080
EhRabL1/AB197081
RabL
HsRab24
EhRabX8/AB197080
EhRabX21/AB197105
EhRabX27/AB197111
EhRabX29 /AB197113
EhRabX26/AB197110
EhRabD1/AB054580
EhRabD2/AB197065
EhRabX25/AB197109
EhRabA/AF030184
EhRabH/AB054586
RabD
EhRabX31/AB197115
EhRabI1/AB054586
EhRabI2/AB197069
RabI
EhRabX30/AB197114
EhRabX24/AB197108
EhRabX5/AB197089
EhRabX33/AB197117
EhRabX20/AB197104
EhRabM3/AB197082
EhRabM2/AB197084
EhRabM1/AB197083
RabM
HsRab21
EhRab21 (EhRabE)/AB054581
EhRabX1/AB197070
HsRab22a
EhRab5/AB054582
HsRab5a
ScYpt51
Rab5
HsRab17
EhRabX9/AB197093
EhRabX15/AB197099
EhRabF4/AB197066
EhRabF3/AB197067
EhRabF2/AB197068
EhRabF1/AB054584
EhRabF5 (EhRabG)/AB054585
RabF
ScYpt10
EhRabX22/AB197106
EhRabX23/AB197107
EhRab11D/AB197959
ScYpt31
HsRab11a
HsRab25
EhRab11C/AB054588
EhRab11A/AB186370
EhRab11B/AB054587
Rab11
EhRabX28/AB197119
EhRabN2/AB197088
EhRabN1/AB197085
RabN
HsRab18
HsRab40a
EhRab2C/AB197073
HsRab2a
HsRab14
HsRab4a
EhRab2B/AB197072
EhRab2A/AB197071
Rab2
HsRab39a
HsRab6a
ScYpt6
HsRab28
EhRabX32/AB197116
EhRabX16/AB197100
EhRabX18/AB197102
EhRabX34/AB197118
EhRabX19/AB197103
EhRabC8/AB197064
EhRabC3/AB197059
EhRabC1/AB054579
EhRabC4/AB197060
EhRabC7/AB197063
EhRabC6/AB197062
EhRabC5/AB197061
EhRabC2/AB186371
RabC
EhRab7H/AB186369
EhRab7I/AB197056
ScYpt7
HsRab7
EhRab7A/AB054583
EhRab7C/AB186364
EhRab7E/AB186365
EhRab7B/AB186363
EhRabX2/AB197074
EhRab7F/AB186367
EhRabX3/AB197075
EhRabX4/AB197076
Rab7
HsRab9a
EhRab7D/AB186385
EhRab7G/AB186388
EhRabX14/AB197098
HsRab33a
EhRabX28/AB197112
EhRabX13/AB197097
HsRab3a
HsRab10
HsRab8
HsRab13
ScSec4
HsRab12
HsRab30
HsRab41
Rab8
HsRab35
EhRab1B/AB197055
EhRab1A/AB054576
HsRab1a
ScYpt1
EhRab8/AF363067
EhRab8B/AB197057
Rab1
EhRabX12/AB197096
EhRabP1/AB197087
EhRabP2/AB197088
RabP
EhRabB/AF127375
HsRab27a
HsRab26
HsRab37
EhRabK4/AB197120
EhRabK3/AB197078
EhRabK2/AB197070
EhRabK1/AB197077
RabK
0.1

frequency of introns. The high frequency of introns in the Rab and SNARE gene families may indicate the presence of post-transcriptional regulation of these genes.

Although Rab proteins generally possess a CXC or CC at the carboxyl terminus, 25 *E. histolytica* Rabs have an atypical carboxyl terminus, such as CXXX, XCXX, XXCX, XXXC, or no cysteine at all. The enzyme(s) involved in the lipid modification of these unusual Rab proteins remain poorly understood (see Section 4.5.1.1). It is also worth noting that >20 *E. histolytica* Rabs lack or contain only a degenerate form of the consensus sequence for structural elements such as the GTP-binding regions and the Switch I and II regions, implicated in the binding to GEF, GAP, effectors or guanine nucleotides (Saito-Nakano *et al.*, 2005). These non-conventional *Eh*Rabs are not pseudo-genes since at least some of the genes are known to be expressed as mRNA (Saito-Nakano *et al.*, 2001). It has been shown that neither *Eh*Rab5 nor *Eh*Rab7A rescued the corresponding yeast mutant (Saito-Nakano *et al.*, 2004). Therefore, many, if not all, *E. histolytica* Rabs may have lost functional interchangeability with their homologues in other organisms despite the relatively high percentage of sequence identities. Classification and annotation of the *E. histolytica* Rab proteins have been previously described (Saito-Nakano *et al.*, 2005).

One of the peculiarities of *E. histolytica* Rab proteins was demonstrated by the unprecedented function of *Eh*Rab7A, which plays an important role in the transport of cysteine proteases via interaction with the retromer complex. The *E. histolytica* retromer complex consists of three components, Vps26, Vps29 and Vps35, rather than the 4–5 found in yeast and mammals (Nakada-Tsukui *et al.*, 2005). Homologues of Vps5, Vps17 and sorting nexins are not encoded in the genome. It has been suggested that the *Eh*Rab7A–retromer interaction, mediated by direct binding of *Eh*Rab7A to a unique carboxyl-terminal region of Vps26, regulates intracellular trafficking of cysteine proteases, and possibly other hydrolases as well, by modulating the recycling of a putative cysteine protease receptor

FIGURE 2.6 A phylogenetic tree of Rab proteins from *E. histolytica*, human and yeast. The number on the nodes represents the bootstrap proportions (%) of 1,000 pseudo samples; only bootstrap proportions >30% are shown. *E. histolytica* Rab proteins are indicated in bold. Tentative subfamilies that revealed significant similarity (>40% identity) to their human or yeast counterpart are shaded dark, while *Entamoeba*-specific subfamilies have light shading. The scale bar indicates 0.1 substitutions at each amino acid position. *: *Eh*Rab proteins that lack the conserved effector region, switch regions or GTP-binding boxes. **: *Eh*Rab proteins that possess a non-conventional carboxyl-terminus or lack carboxyl-terminal cysteines. ***: Rab proteins that were not classified as isotypes based on <40% identity to other members of the subfamily. References on tree: (1): Temesvari *et al.* (1999); (2): Rodríguez *et al.* (2000); (3): Saito-Nakano *et al.* (2001); (4): Juarez *et al.* (2001); (5): Saito-Nakano *et al.* (2004); (6): Okada *et al.* (2005).

from lysosomes and phagosomes to the Golgi or post-Golgi compartment (Nakada-Tsukui *et al.*, 2005).

6.3.2. SNARE and their accessory proteins

The final step in membrane trafficking is the fusion of a transport vesicle with its target membrane, which is mediated by the SNARE family of proteins. SNAREs are integral membrane proteins that are present on both donor and acceptor membranes and form a stable complex to tether the two membranes. It is believed that the formation of a SNARE complex pulls the vesicle and target membrane together and provides the energy to drive fusion of the lipid bilayers (Chen and Scheller, 2001; Chen *et al.*, 1999). In a prototypical model, a SNARE complex, which consists of four helices, is formed at each fusion site (Hanson *et al.*, 1997; Poirier *et al.*, 1998). For instance, the fusion of synaptic vesicles with the presynaptic nerve terminus is mediated by the formation of a complex comprising one helix each from syntaxin 1A [Qa-SNARE, also termed target-SNARE (t-SNARE)] and VAMP2 [R-SNARE, vesicular SNARE (v-SNARE)] and two helices from SNAP-25 (Qb- and Qc-SNARE).

The complexity of SNAREs has remained largely unchanged in yeast, fly and worm, but has increased remarkably in mammals and plants (Table 2.7) indicating that although expansion of SNARE repertoires occurs, a set of core SNAREs is sufficient to mediate vesicular fusion of most pathways in multicellular organisms. *E. histolytica* encodes 28 putative SNAREs, 18 Q-SNAREs and 10 R-SNAREs, which is comparable to the complexity in humans and plants. A notable peculiarity of SNAREs in *E. histolytica* is the lack of a group of proteins possessing two helices (Qb and Qc SNAREs) such as SNAP-25. Thus, the prototype model of membrane tethering by a combination of four helices (from Qa, R to Qb/Qc) does not appear to be possible in this organism.

A group of proteins that interact directly with the syntaxin subfamily, including the prototypical member yeast Sec1p and mammalian Munc-18, are essential cytosolic proteins peripherally associated with membranes (Toonen and Verhage, 2003). They are presumed to be chaperones, putting syntaxins into the conformations required for interaction with other SNAREs (Dulubova *et al.*, 1999; Yang *et al.*, 2000). Sec1/Munc-18 proteins are also conserved in *E. histolytica* (there are five *Sec1* genes). Two additional important components involved in the recycling of fusion machinery, *N*-ethylmaleimide sensitive factor (NSF) (Beckers *et al.*, 1989) and soluble NSF attachment protein (SNAP) (Clary *et al.*, 1990; Mayer *et al.*, 1996) are also found in *E. histolytica*.

Other proteins involved in vesicle fusion are the saposin-like proteins mentioned earlier (Section 3.3). The membrane-fusogenic activity of the *E. histolytica* SAPLIPs may play a role in vesicle fusion (Winkelmann

et al., 2006), but how they interface with the Rab/SNARE processes remains to be determined.

6.4. Comparisons and implications

While the fundamental machinery of vesicular trafficking is conserved in *E. histolytica*, the high activity of the endocytic and biosynthetic transport pathways in this organism appears to have resulted in the dramatic expansion of the Rab gene repertoire. The diversity and complexity of Rab proteins present in *E. histolytica* likely reflect the vigorous dynamism of membrane transport and the reliance on Rab proteins for the specificity of vesicular trafficking. The high degree of Rab complexity observed in *E. histolytica* (91) has no precedent in other organisms, although the incomplete genome of *T. vaginalis* appears to encode 65 Rabs (Lal *et al.*, 2005) while *Dictyostelium* encodes 50 (Eichinger *et al.*, 2005). Rab proteins have been extensively studied in *T. brucei* and the recent completion of *T. brucei*, *Trypanosoma cruzi* and *Leishmania major* genomes led to identification of all Rab genes in these haemoflagellates (Ackers *et al.*, 2005; Berriman *et al.*, 2005; Quevillon *et al.*, 2003). Among the 16 Rabs present in *T. brucei*, there are only 3 Rab proteins (RabX1-X3) that appear to be unique to kinetoplastids. *T. brucei* encodes 13 Rab proteins homologous to those in humans, suggesting significant conservation of the Rab-dependent core endomembrane systems in kinetoplastids. *P. falciparum* possesses only 11 Rab genes all of which are considered orthologues of yeast and mammalian Rabs, although Rab5a, 5b and 6 revealed unique features (Quevillon *et al.*, 2003). Interestingly, some of these Rabs are expressed in a stage-dependent manner (Quevillon *et al.*, 2003). The comparatively small number of Rabs in these protists reinforces the tremendous diversity and complexity of Rabs seen in *E. histolytica* (Table 2.7).

In marked contrast to the complexity of Rab proteins in *E. histolytica*, the number of SNARE proteins, the other major components of vesicular fusion, is comparable to that in yeast. The apparent disparity in the number of Rab and SNARE proteins suggests one of three possibilities: (1) *Eh*Rab proteins share a single SNARE complex as an interacting partner (Huber *et al.*, 1993; Rowe *et al.*, 2001; Torii *et al.*, 2004), (2) a majority of *Eh*Rabs do not require SNARE proteins for membrane fusion (Demarque *et al.*, 2002), (3) some *Eh*Rabs are primarily involved in cellular functions other than membrane fusion, like Arl GTPases (Burd *et al.*, 2004; Pasqualato *et al.*, 2002). Genome-wide surveys of SNAREs in other protists are not available. The three major types of coatomer protein, which are conserved in *E. histolytica*, are also conserved in kinetoplastids (Berriman *et al.*, 2005). However, in contrast to *E. histolytica*, *T. brucei* does not possess multiple isotypes of COPI and II components except for Sec24, which has two isotypes. *T. cruzi* encodes all four AP complexes