

Multivesicular bodies in the enigmatic amoeboflagellate *Breviata anathema* and the evolution of ESCRT 0

Emily K. Herman^{1,*}, Giselle Walker^{2,*}, Mark van der Giezen³ and Joel B. Dacks^{1,†}

¹Department of Cell Biology, School of Molecular and Systems Medicine, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB T6G 2H7, Canada

²Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

³Centre for Eukaryotic Evolutionary Microbiology, Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

*These authors contributed equally to this work

†Author for correspondence (dacks@ualberta.ca)

Accepted 6 October 2010

Journal of Cell Science 124, 613–621

© 2011. Published by The Company of Biologists Ltd

doi:10.1242/jcs.078436

Summary

Endosomal sorting complexes required for transport (ESCRTs) are heteromeric protein complexes required for multivesicular body (MVB) morphogenesis. ESCRTs I, II, III and III-associated are ubiquitous in eukaryotes and presumably ancient in origin. ESCRT 0 recruits cargo to the MVB and appears to be opisthokont-specific, bringing into question aspects of the current model of ESCRT mechanism. One caveat to the restricted distribution of ESCRT 0 was the previous limited availability of amoebozoan genomes, the supergroup closest to opisthokonts. Here, we significantly expand the sampling of ESCRTs in Amoebozoa. Our electron micrographic and bioinformatics evidence confirm the presence of MVBs in the amoeboflagellate *Breviata anathema*. Searches of genomic databases of amoebozoans confirm the ubiquitous nature of ESCRTs I–III-associated and the restriction of ESCRT 0 to opisthokonts. Recently, an alternate ESCRT 0 complex, centering on Tom1 proteins, has been proposed. We determine the distribution of Tom1 family proteins across eukaryotes and show that the Tom1, Tom1L1 and Tom1L2 proteins are a vertebrate-specific expansion of the single Tom1 family ancestor, which has indeed been identified in at least one member of each of the major eukaryotic supergroups. This implies a more widely conserved and ancient role for the Tom1 family in endocytosis than previously suspected.

Key words: Endocytosis, Eukaryote, Membrane trafficking

Introduction

Multivesicular bodies (MVBs) are crucial eukaryotic organelles involved in ubiquitin-mediated endocytic processes, underlying cellular acquisition of nutrients, and downregulation of receptors (Williams and Urbe, 2007). They are generally sized at 400–500 nm (Gruenberg and Stenmark, 2004) and contain intraluminal vesicles (ILVs) that are uniformly round and 50 nm in diameter or smaller (Williams and Urbe, 2007). ILVs are created by the invagination and inward budding of the membrane, a process modulated by a set of protein components collectively known as the endosomal sorting complexes required for transport, or ESCRTs.

There are five soluble ESCRT complexes that work at the cytosolic face of the MVB and are responsible for recruiting the proper cargo to ILVs, as well as the budding and scission events (Hurley, 2008). An emerging model of ESCRT function (Fig. 1) in mammalian and yeast systems identifies the ESCRT 0 complex as recognizing and binding ubiquitylated cargo from sites such as the plasma membrane and recruiting it to the MVB. The mechanism of this process is unclear because a recent study has shown that depletion of the human Vps27 (an ESCRT 0 component) by siRNA does not have a significant effect on epidermal growth factor receptor (EGFR) endocytosis (Raiborg et al., 2008). Interaction between the ESCRT I component Vps23 with the P[S/T]xP domain in Vps27 recruits ESCRT I to MVBs where it is responsible for cargo sorting and recruitment of ESCRTs II and III. These

subcomplexes interact with each other, as well as with the membrane, to mediate inward budding. The ESCRT III-associated machinery has been shown to induce scission of the vesicle (Wollert and Hurley, 2010) and one particular component, the AAA-type ATPase Vps4, is responsible for disassembly of the other ESCRTs (Saksena et al., 2009).

The ESCRT machinery is not only functionally essential but ancient as well. Comparative genomic studies have shown that the vast majority of protein components composing ESCRT complexes I–III-associated are present in the diversity of eukaryotic taxa (Field et al., 2007; Leung et al., 2008; Slater and Bishop, 2006), well beyond the model systems of yeast and Metazoa. This implies that the last eukaryotic common ancestor (LECA) possessed an ESCRT machinery of near-modern complexity. Comparative experimental characterization in organisms from various eukaryotic supergroups (Adl et al., 2005) suggests similarity and conservation of ESCRT function as well, with organelles resembling MVBs identified in diverse eukaryotes (Haas et al., 2007; Hurley, 2008; Leung et al., 2008; Yang et al., 2004). Delving further back in evolutionary time, it is apparent that gene duplications gave rise to two sets of components in the ESCRT III and III-associated machinery, the Vps20/Vps32/Vps60 and the Vps2/Vps24/Vps46 families (Leung et al., 2008). This implies a model of an ancestral dimeric ESCRT III complex composed of a progenitor protein from each family (Leung et al., 2008). Such a model is bolstered

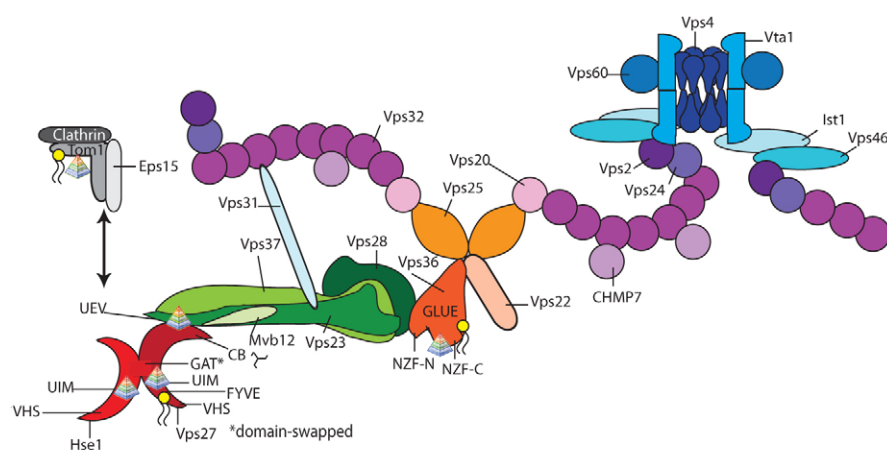


Fig. 1. Model of ESCRT complex assembly. The diagram is based loosely upon one previously published (Raiborg and Stenmark, 2009) incorporating information from other sources (Hurley, 2008; Im et al., 2009; Kostelansky et al., 2007; Prag et al., 2007; Shestakova et al., 2010; Xiao et al., 2008). The complexes are marked as follows: ESCRT 0 components in red, ESCRT I in green, ESCRT II in orange, ESCRT III in purple, and ESCRT III-associated in blue. Relevant domains are identified. The GAT domain of ESCRT 0 is a heterodimer due to domain-swapping. The multicoloured pyramids indicate binding sites of ubiquitylated cargo, and the 'lipids' (yellow circles with tails) indicate phosphatidylinositol 3-phosphate binding sites. The Tom1 complex marked in grey is based on the ancestral complex proposed previously (Blanc et al., 2009).

by the recent discovery of ESCRT III homologs in Archaea that are involved in cell division (Samson et al., 2008), providing a path of origin for the ESCRT machinery in eukaryotes (Field and Dacks, 2009).

By contrast, ESCRT 0 components appear to be opisthokont-specific (Field et al., 2007; Leung et al., 2008), raising questions of the origin of this machinery and the generality of the current model of ESCRT mechanism. However, at the time of the most recent and exhaustive comparative genomic analysis of the ESCRT machinery to date (Leung et al., 2008), only a limited number of genomes were available from the nearest supergroup to the Opisthokonta, i.e. the Amoebozoa. The possibility therefore exists that undersampling might explain the ESCRT 0 distribution. We have therefore undertaken an investigation of ESCRT machinery in the Amoebozoa, with emphasis on the enigmatic amoeba *Breviata anathema*.

Originally mis-identified as the pelobiont *Mastigamoeba invertens*, *B. anathema* is an amoeboid flagellate (Fig. 2A), 5–10 µm in size (Walker et al., 2006). It lacks canonical mitochondria, possessing instead a multi-lobed double-membrane-bounded organelle, postulated to be a hydrogenosome (Walker et al., 2006). *B. anathema* was recently thrust into the spotlight as a clear counterexample to the prominent hypothesis of the bikont–unikont rooting of the eukaryotic tree (Roger and Simpson, 2009). The organism possesses two basal bodies supported by flagellar root-like structures, as found in bikont organisms (Walker et al., 2006), but is evolutionarily placed within the unikont clade as a basal amoebozoan (Minge et al., 2009). *Breviata* is therefore a crucial sampling point for any investigation into the evolution of the ESCRT machinery in the Amoebozoa. We previously provided a single transmission electron microscopy (TEM) image of a putative MVB organelle in *B. anathema* (Walker et al., 2006), which

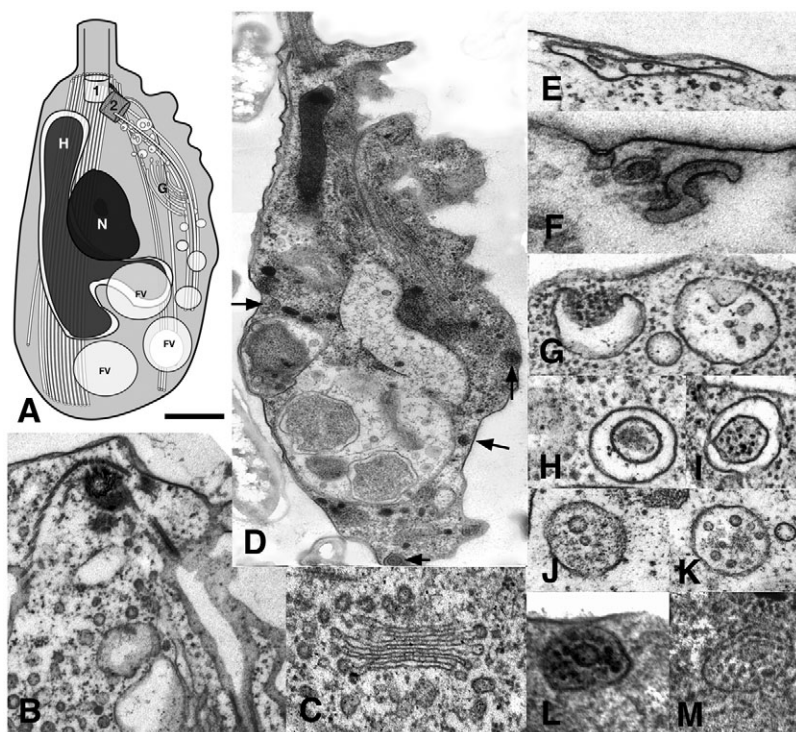


Fig. 2. Ultrastructure of the endocytic system in *Breviata anathema*. (A) Diagram showing the general ultrastructure: a flagellated cell with pseudopodia, with one large nucleus (N) surrounded by a branching hydrogenosome-like organelle (H). In the flagellar apparatus there are two basal bodies (1, 2), with associated microtubular roots, those of basal body 2 subtending the area of the feeding groove, which also contains the Golgi dictyosome (G). There are food vacuoles (FV) in the posterior of the cell. (B) TEM image showing longitudinal section through the cell, with the flagellar apparatus at the top. The vesicular area and part of the feeding groove (subtended by microtubular roots) can be seen. (C) Golgi dictyosome. (D) Whole cell, showing the flagellar apparatus at the top, the pseudopodial area to the right, many large food vacuoles and some MVBs (arrows). (E–M) MVBs or other stages in the endocytic system. Scale bar: 500 nm (B, D); 300 nm (C); 125 nm (E); 200 nm (F–I, L, M); 400 nm (J, K). Panels D, F and M are reproduced from Walker et al. (Walker et al., 2006) with permission from the authors.

prompted us to further investigate this potential organelle by TEM and using comparative genomics to look for ESCRT homologs. The latter sequence-based approach was expanded to explore the representation of ESCRT machinery in available amoebozoan genomic databases. Recent evidence raised the possibility of an alternative ESCRT-0-like machinery centered on the Tom1 protein family (Blanc et al., 2009; Yanagida-Ishizaki et al., 2008). Tom1 family homologs have been identified in opisthokonts, as well as in *Dictyostelium discoideum* and the multicellular plants *Arabidopsis thaliana* and *Oryza sativa* (Blanc et al., 2009; Winter and Hauser, 2006). We have therefore extended our investigation beyond conventional ESCRT machinery and performed comparative genomic and phylogenetic investigations to explore the evolution and diversity of the Tom1 family.

We here provide electron micrographic evidence and novel ESCRT component sequence data from *B. anathema*, thus confirming and extending the evidence for the presence of MVBs in this organism. Our comparative genomic study expands the identification of the machinery of ESCRTs I–III-associated into a wider diversity of amoebozoan organisms, while bolstering ESCRT 0 as an opisthokont-specific innovation. Finally, Tom1 family homologs were identified in at least one representative of each eukaryotic supergroup, suggesting it as an ancient and widely present eukaryotic cellular component.

Results

Electron microscopy

We used TEM to ascertain whether *B. anathema* does possess MVBs, as suggested by figure 24 of Walker et al. (Walker et al., 2006) (here shown in Fig. 2M). *B. anathema* cells have a vesicular area immediately proximal to the flagellar apparatus, bounded by the microtubular roots associated with the second basal body and extending to the posterior of the cell where food vacuoles are found (Fig. 2A,B). In this area, a Golgi dictyosome is usually seen (Fig. 2C) and the cell membrane is extended in pseudopodia, creating a feeding groove bound by microtubular roots (Fig. 2B). Multivesicular bodies are found only in this area in most cells, but occasionally MVBs are visible in the whole posterior of the cell (Fig. 2D, arrowheads). MVBs are disc-shaped, up to 500 nm in diameter and ca. 50 nm deep (Fig. 2E). They contain smaller vesicles (mostly ca. 20 nm in diameter, some up to about 200 nm) and granules (Fig. 2E–M).

Breviata ESCRT machinery

Given the MVB-like organelles that we observed, we predicted the presence of ESCRT components in *B. anathema* as well. To test this hypothesis, we searched for sequences encoding ESCRT machinery in our on-going expressed sequence tag (EST) survey of *B. anathema* (M.v.d.G., G.W. and J.B.D., unpublished). For sequences recovered,

in most cases, we were able to assemble a large-enough coding region from overlapping EST reads to unambiguously propose a homology assignment by BLASTp analysis (Table 1). If multiple reads were not available, or if the sequence was not sufficient to yield a clear result by homology searching, the full sequence was obtained by double-strand sequencing of the insert.

Additionally, in the cases of the paralogous Vps24 (Fig. 3) and SNF7 (Fig. 4) families, phylogenetic analysis was performed to verify the orthology of the sequences. In the case of the Vps24 analysis (Fig. 3), the Vps46, Vps24 and Vps2b clades were well resolved, with Vps24 and Vps2b emerging from a paraphyletic assemblage of Vps2a homologs. All sequences were clearly assigned orthology, with the exception of *Entamoeba histolytica* Vps2 that grouped with Vps24 homologs but was also robustly excluded from that clade (Fig. 3). In the case of the SNF7 homologs (Fig. 4), the Vps60 and Vps20 clades were robustly reconstructed, whereas the Vps32 clade was recovered but without statistical support. Consequently, any candidate Vps32 homologs were assigned as such on the basis of their BLASTp results and on their exclusion from the Vps20 and Vps60 clades (Fig. 4).

We were able to identify clones encoding an extensive set of ESCRT machinery from *B. anathema* (Table 1, Fig. 5). Although no ESCRT II subunits were identified, the ESCRT I component Vps28 was found. A near-complete set of ESCRT III and III-associated machinery was found, including Vps2, Vps20 and Vps32 as well as Vps31, Vps4 and Vps46 (Figs 3–5).

Amoebozoan ESCRT comparative genomics

The presence of ESCRT components in *Breviata*, as well as those previously identified in *D. discoideum* and *E. histolytica* (Leung et al., 2008), prompted us to perform a comparative genomic analysis of publicly available amoebozoan databases to investigate the conservation and diversity of ESCRT machinery in this supergroup. The draft genome sequence of *Acanthamoeba castellanii* and public EST datasets of *Physarum polycephalum*, *Hartmannella vermiformis* and *Mastigamoeba balamuthi* were searched.

No ESCRT 0 components were found in any of the amoebae sampled. When searching with ESCRT 0 queries, some candidate sequences were retrieved that shared a domain with either Hse1 or Vps27, usually a VHS, FYVE, UIM or SH3 domain. However, reverse BLAST searching revealed that these amoebozoan proteins were not homologs due to failure to retrieve Vps27 homologs as their top BLAST hit or based on domain structure (see Materials and Methods for criteria). By contrast, despite the inherently incomplete nature of EST projects, at least four ESCRT subunits from multiple ESCRT complexes were found in each amoeba, with organismal specifics detailed below.

The plasmodial slime mold, *P. polycephalum*, was found to possess subunits from ESCRT I (Vps37), ESCRT II (Vps 36) and

Table 1. Clones encoding ESCRT machinery from *M. balamuthi* and *B. anathema*

Organism	Annotation	GenBank nucleotide accession	<i>H. sapiens</i> BLAST	<i>S. cerevisiae</i> BLAST
<i>M. balamuthi</i>	Vps37	GU292811	4.00E-06	6.00E-09
<i>M. balamuthi</i>	Vps24	GU256250	2.00E-35	2.00E-25
<i>B. anathema</i>	Vps28	HM773427	1.00E-37	2.00E-28
<i>B. anathema</i>	Vps2	HM773426	6.00E-42	9.00E-37
<i>B. anathema</i>	Vps20	HM773425	8.00E-26	9.00E-15
<i>B. anathema</i>	Vps32	HM773428	4.00E-13	6.00E-09
<i>B. anathema</i>	Vps4	HM773429	7.00E-81	1.00E-74
<i>B. anathema</i>	Vps31	HM773430	3.00E-17	3.00E-03
<i>B. anathema</i>	Vps46	HM773431	4.00E-37	4.00E-31

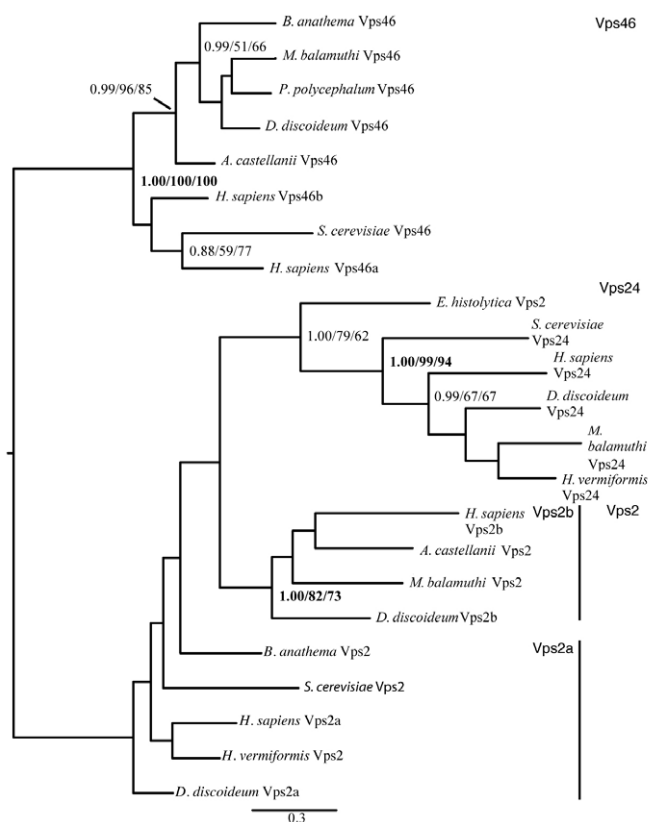


Fig. 3. Phylogeny of Vps24 family homologs. Orthology of the various amoebozoan Vps24 family homologs identified in this study. The phylogeny is rooted on the ESCRT III-associated subunit Vps46 because this is proposed as the most ancient duplication in the Vps24 family (Leung et al., 2008). The tree diagram shown for this and subsequent phylogenies is the best Bayesian topology, with support values listed in the order of Bayesian posterior probability values/PhyML bootstrap values/RAxML bootstrap values. Values are only shown for nodes with support better than 0.80/50%/50%. The vertical bars denote the Vps2, Vps24 and Vps46 clades, and relevant support nodes are shown in bold.

ESCRT III-associated (Vps4, and 46) (Fig. 5; supplementary material Tables S1, S2). Because components were identified from three of the four subcomplexes and yet only the latter two proteins are known in model organisms to have a direct interaction, this probably represents an incomplete picture of the *P. polycephalum* ESCRT complement. Surprisingly though, we did identify a charged multivesicular body protein 7 (CHMP7) homolog. Until recently, this protein was thought to be opisthokont-specific, but it has now been found with a patchy distribution across the eukaryotes (Leung et al., 2008). Of the amoebae sampled thus far, only the slime molds *P. polycephalum* and *D. discoideum* have been shown to possess CHMP7 homologs.

Acanthamoeba was found to possess multiple copies of some ESCRT components, namely Vps28 and Vps4 and an extensive ESCRT complement, including Vps22, Vps36, Vps2, Vta1 and Vps46 (Fig. 5; supplementary material Tables S1, S2). Because *Acanthamoeba* has Vps36, it could theoretically bind ubiquitinated cargo as well as the membrane via a GLUE domain (supplementary material Table S2). However, the other proteins present are mostly ESCRT disassembly proteins because all ESCRT III components except for Vps2 appear to be absent from the genome database. The ESCRT complement of *H. vermiformis* includes Vps25, Vps2,

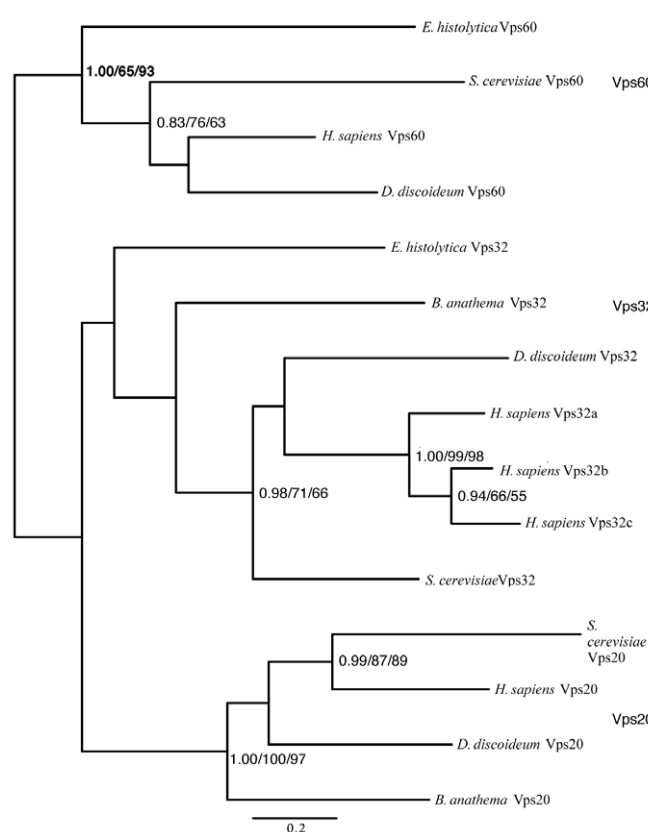


Fig. 4. Phylogeny of SNF7 family homologs. Orthology of the various amoebozoan SNF7 family homologs identified in this study. The phylogeny is rooted on the ESCRT III-associated subunit Vps60 because this is proposed as the most ancient duplication in the SNF7 family (Leung et al., 2008). The vertical bars denote the Vps32, Vps20 and Vps60 clades, and relevant support nodes (Bayesian posterior probability/PhyML bootstrap/RAxML bootstrap) are shown in bold.

Vps24 and Vps31 (Fig. 5; supplementary material Tables S1, S2). Though few components were found, it appears that they could interact and perhaps function with only the addition of Vps20 and Vps32.

Mastigamoeba had several components of each ESCRT complex, excluding ESCRT 0. Of ESCRT I, it has Vps28, Vps37 and two copies of Vps23. Homologs of Vps22 and Vps25 (ESCRT II), Vps2 and Vps24 (ESCRT III), and Vps4 and Vps46 (ESCRT III-associated) were also identified (Fig. 5; supplementary material Tables S1, S2). In the case of Vps37 and Vps24 there was insufficient information in the public database to robustly determine homology. Consequently, clones encoding these ESTs were obtained and fully sequenced (Table 1). The ESCRT components found thus far for *M. balamuthi* are capable of binding cargo (Vps23), participating in budding and scission events (Vps22, Vps25) and in disassembly (Vps4 and Vps46). Interestingly, no members of the SNF7 family were found, but because the sequences were retrieved from an EST project, this probably represents sampling bias.

Tom1 evolution

In the absence of obvious ESCRT 0 complexes in the majority of eukaryotic groups, the question arises of what, if any, machinery

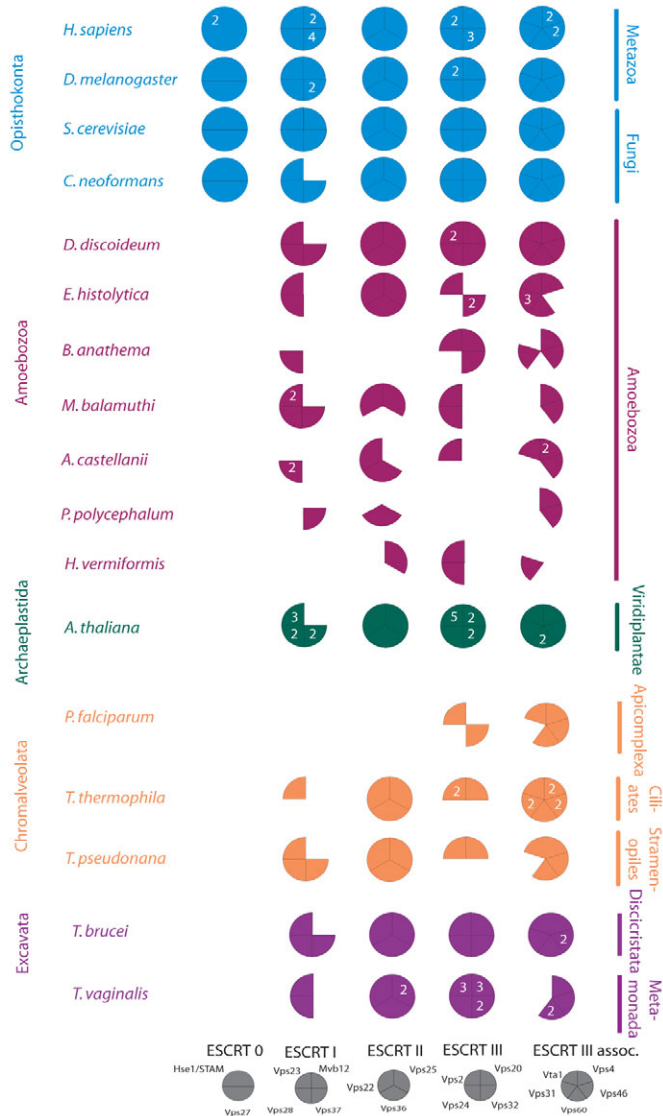


Fig. 5. Coulsen plot showing the distribution of ESCRT complexes. Filled segments indicate the presence of homologs. Missing segments indicate that a homolog was not found. The number of paralogs is indicated by the white numbers on the segments. The components represented by each sector are shown at the bottom in grey. Note that CHMP7 was found in *Physarum polycephalum*, but this factor is not included as a sector in the ESCRT III pie. The diagram shows the amoebozoan taxa sampled along with information from representatives of the other major eukaryotic supergroups. The figure is redrawn from, and incorporates, information from Leung et al. (Leung et al., 2008).

performs the analogous function of recruiting ubiquitylated cargo to the MVB. The ESCRT 0 components Vps27 and Hse1 are both VHS domain-containing proteins, as are the Tom1 protein family (Puertollano, 2005; Raiborg and Stenmark, 2009). Recent evidence has been presented, in multiple and diverse eukaryotes, that Tom1 and related proteins bind ubiquitylated cargo (Blanc et al., 2009; Katoh et al., 2004) and the ESCRT I component Tsg101/Vps23 (Yanagida-Ishizaki et al., 2008). Consequently, it has been proposed that Tom1 proteins might have a role as central components of an ancient alternative ESCRT 0 machinery (Blanc et al., 2009). In order to assess the distribution and evolution of the Tom1 family,

we performed BLAST and HMMer homology searches in 36 genomes from organisms across the diversity of eukaryotes. We were able to identify Tom1 homologs in at least one representative organism from all supergroups searched (supplementary material Table S1). Importantly, with the exception of the multicellular plants and the Metazoa, most taxa possessed a single Tom1 family homolog. Initial phylogenetic analysis provided little resolution but did allow for identification of closely related, lineage-specific duplicates that were removed from subsequent rounds of analysis (supplementary material Fig. S1). Further analysis allowed us to resolve the evolution of the metazoan Tom1 family (Fig. 6), demonstrating that the duplications giving rise to the Tom1, Tom1-like1 and Tom1-like2 paralogs occurred prior to the divergence of humans and fish. We therefore refer to Tom1 family proteins found in organisms outside of vertebrates as Tom1esc proteins.

Several motifs have been found in the human and *Dictyostelium* Tom1 family homologs, including a clathrin box (Blanc et al., 2009; Yamakami et al., 2003) and a P[S/T]xP motif (Blanc et al., 2009; Puertollano, 2005), as well as NPF repeats in the C-terminal portion of the *Dictyostelium* Tom1 (Blanc et al., 2009). This prompted us to examine the Tom1esc proteins in other eukaryotes for similar motifs. The clathrin box motif has been described as Lφφφ(–), signifying leucine followed by a bulky hydrophobic residue, a polar residue, another bulky hydrophobic residue and a negatively charged residue (Dell’Angelica, 2001). We were unable to find clear clathrin box motifs in the non-metazoan candidates. NPF repeats bind Eps15, part of the machinery involved in clathrin-mediated endocytosis (Polo et al., 2003). NPF sequences were found near the N-terminus in several Metazoa, and near the C-terminus in *Cryptococcus neoformans*, several archaeplastids and *Phytophthora ramorum*. P[S/T]xP motifs required for binding Vps23 and/or Tsg101 were again found in several opisthokonts, archaeplastids and the excavate *Naegleria gruberi*.

Discussion

The ESCRT machinery is well known to be functionally crucial in model organisms (generally yeast and metazoans) and ancient in eukaryotic cells. There are, however, key open questions regarding the nature of the machinery that binds ubiquitylated cargo for recruitment to the MVB and the variability and conservation of the ESCRTs and MVBs in diverse eukaryotic taxa. The independent identification of MVB machinery in poorly characterized eukaryotes is crucial for addressing the latter point.

Together, the electron micrographs of MVB-like compartments (Fig. 2) and the expressed genes encoding ESCRT machinery (Table 1, Fig. 5) strongly imply the presence of a functional MVB in *Breviata*. Although some eukaryotes (e.g. Apicomplexa) appear to have dispensed with ESCRT complexes I and II (Leung et al., 2008), our identification of Vps28 suggests that this is not the case for *B. anathema*. These complexes are probably present and should be identifiable by further sequencing efforts. It is also possible that the lack of ESCRT 0 components is also due to low gene expression or incomplete sampling. However, given the sampling of complete genomes from diverse additional eukaryotes that have failed to identify ESCRT 0 genes, we suggest that it is much more likely that our result represents a legitimate absence. The systematic placement of *B. anathema* as a basal amoebozoan strengthens the conclusion that the ESCRT machinery is a common feature of this supergroup. It has also been proposed that *Breviata* is a separate eukaryotic lineage of unclear affiliation (Parfrey et al., 2010) and not an amoebozoan. Should this be the case, it would only increase

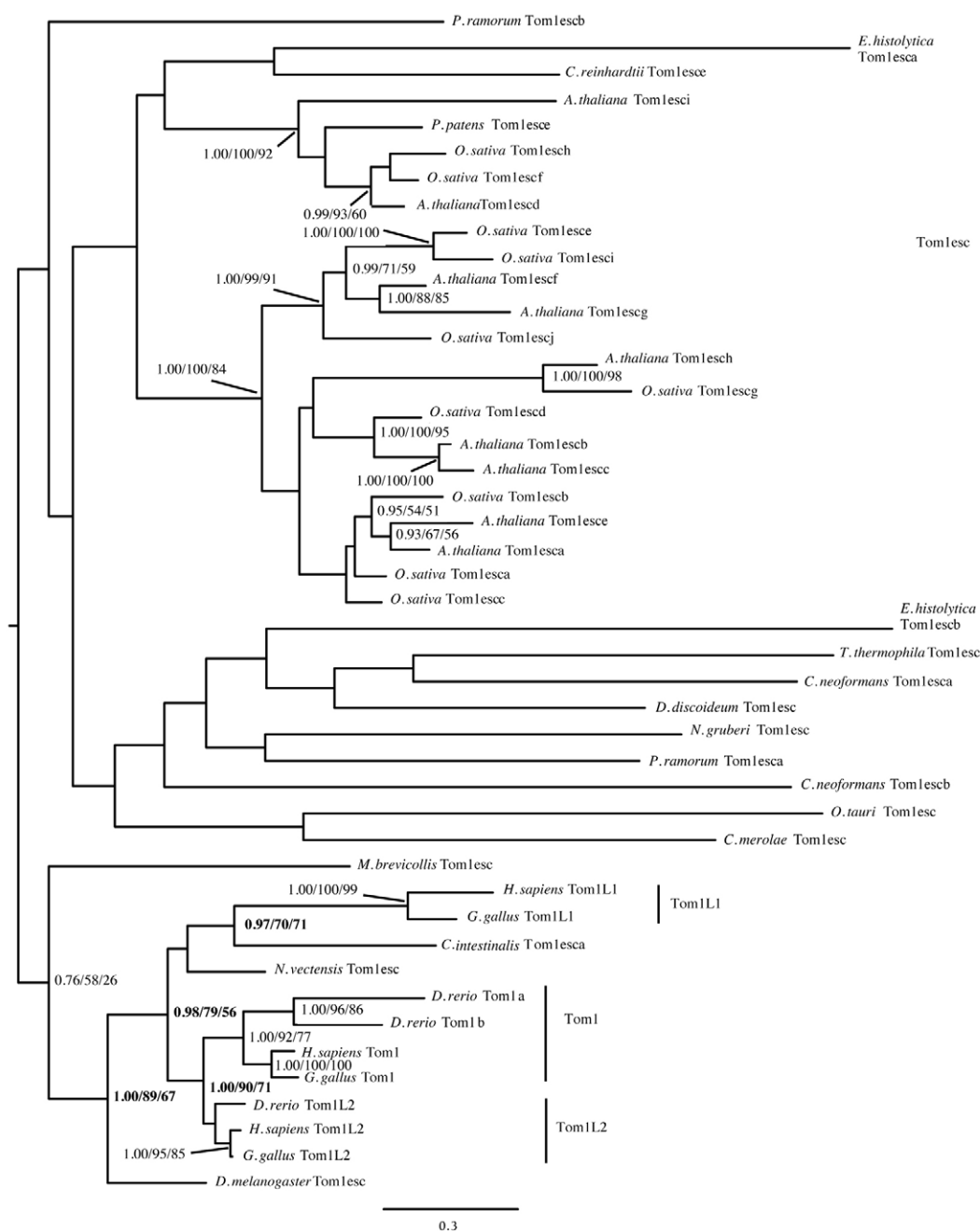


Fig. 6. Phylogeny of Tom1 family homologs. Orthology of the various Tom1 family homologs identified in this study. This phylogeny is arbitrarily rooted on the metazoan Tom1 homologs to highlight the family expansion in that lineage. The vertical bars denote the Tom1esc, Tom1L1, Tom1 and Tom1L2 clades, and relevant support nodes (Bayesian posterior probability/PhyML bootstrap/RAXML bootstrap) are shown in bold.

the importance of our independent identification of MVBs in this lineage, with reference to the conclusion that MVBs and the ESCRT machinery are indeed conserved features of eukaryotes and were present in the LECA.

The limited sampling of Amoebozoa in the study by Leung and co-workers left open the possibility that ESCRT 0 components were present in that supergroup, but lost from the two amoebozoans sampled: *D. discoideum* (a highly derived cellular slime mold) and *E. histolytica* (a highly derived gut parasite). Our more extensive sampling of amoebozoan taxa confirms and extends the conclusion that ESCRT 0 is not present beyond the Opisthokonta. On the other hand, our identification, in diverse amoebozoans, of ESCRT components from all other subcomplexes emphasizes the ubiquitous nature of this remaining ESCRT machinery. Functional MVBs are suggested by the fact that we were able to identify interacting

components in each taxon, as well as coexpression of these genes in taxa that have EST projects. Furthermore, the identification of a second CHMP7 homolog reinforces the idea that this component plays a role in ESCRT function in diverse eukaryotes.

Because ESCRT 0 (Vps27 and Hse1) is an opisthokont-specific innovation, the possibility has been raised of an alternate route for sorting ubiquitylated cargo to the MVB. In mammalian cells, Tom1-related proteins bind clathrin, ubiquitin and Tsg101 and have been implicated in EGFR internalization (Liu et al., 2009). In *Dictyostelium*, the single Tom1 protein has been shown to bind clathrin, ubiquitin and Tsg101, as well as an Eps15 homolog (Blanc et al., 2009). It was also shown to localize to punctae and colocalize with ubiquitylated proteins.

We found patchy distribution of Tom1esc proteins, but homologs were found in at least one member from each eukaryotic

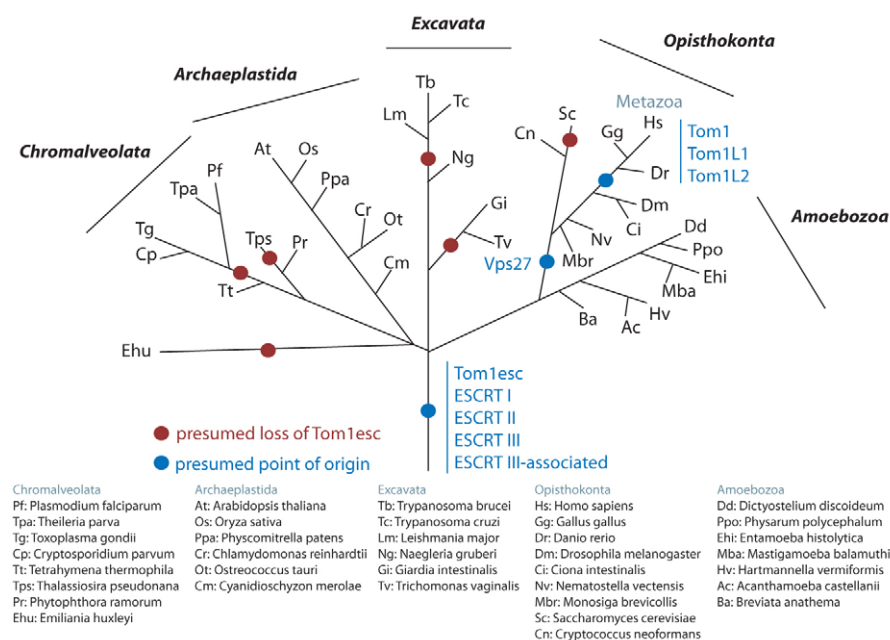


Fig. 7. Evolution of the ESCRTs and Tom1esc in the eukaryotes. A key to the abbreviations used is provided in the bottom of the figure. The blue circles indicate the deduced origins of each component and the red circles indicate where Tom1esc appears to have been lost. Tom1esc and the ESCRT I–III-associated complexes were acquired prior to the LECA. Although Vps27 is opisthokont-specific, the Tom1, Tom1L1 and Tom1L2 proteins are a vertebrate-specific expansion. There have also been losses of Tom1esc in several taxa. In the Amoebozoa, Tom1esc was only found in *D. discoideum* and *E. histolytica*. However, our sampling of several other amoebozoans is from EST projects only and so we have not marked losses in that group.

super group, implying that the LECA did possess an ancient Tom1 protein (Fig. 7). Although loss of Tom1esc homologs has probably occurred frequently in eukaryotes, we also note that failure to identify a homolog could be the result of high sequence divergence, despite our use of the most sensitive homology searching algorithms available. Incompleteness of some genomic database might also have played a role. As additional genome sequences become available, a more detailed and accurate pattern of Tom1 retention might become apparent. Although several organisms express Tom1 proteins that have clathrin box motifs, the NPF sequence and P[S/T]xP motifs, many other Tom1esc proteins might either have highly variant motifs or have lost the motif. The former is a probable explanation for the lack of clathrin box conservation, because the canonical L ϕ p ϕ (–) ‘rule’ has been shown to have many exceptions (Dell’Angelica, 2001). On the other hand, the latter explanation of motif loss suggests that Tom1esc proteins lacking these motifs might not bind the same components as the human and *Dictyostelium* Tom1 family proteins, and therefore might function in an alternate manner.

On the basis of experimental data concerning human and *Dictyostelium* Tom1 family proteins, Blanc and colleagues proposed an ancestral alternative ESCRT 0 complex composed of Eps15, clathrin and Tom1 ‘contributing to the sorting of ubiquitinated proteins to the MVB formation machinery’ (Blanc et al., 2009). Although some data (Blanc et al., 2009) might not fit that explanation entirely, the proposal warrants further experimental investigation. From our data, Tom1 at least has the potential to be a widely conserved eukaryotic cellular component. Although clathrin is a very well-conserved component of the endocytic machinery in diverse eukaryotes, Eps15 is an opisthokont-specific innovation (Field et al., 2007). Nonetheless, Eps15R is an ancient component and thus potentially the actual piece of this putative complex (Field et al., 2007). If the Tom1esc complex does play the role of cargo recruitment and chaperoning to the MVB then, from its phylogenetic distribution, it is likely to have been the original set of components, either replaced or perhaps supplemented by the ESCRT 0 complex in opisthokonts.

Regardless of whether these proteins do form a full complex and whether they are involved in MVB formation or another endocytic process, it seems likely that the component parts of this putative assembly are widely present in eukaryotes and were present in our ancestor approximately 1.5 billion years ago (Yoon et al., 2004).

Materials and Methods

Cultures and microscopy

Two isolates of *B. anathema* were studied: culture 50338 was originally obtained from the American Type Culture Collection (Edgcomb et al., 2002; Minge et al., 2009; Stiller et al., 1998; Walker et al., 2006), and the other isolated by Jeff Silberman (University of Arkansas, Fayetteville, AR), verified as identical to 50388 by ultrastructure and sequencing of the 18S small subunit of the ribosomal RNA gene. Cultures were maintained in 10-ml Falcon tubes of ATCC medium 1773, with mixed bacteria. For electron microscopy, 1 ml of culture was taken from the dense bacterial growth at the bottom of culture tubes, placed in a new tube and rapidly swamped, using a Gilson pipette, with 10 ml of an ice-cold fixation cocktail (5% v/v glutaraldehyde, 0.5% w/v osmium tetroxide, 80 mg K₃[Fe(CN)₆], 50 mM cacodylate buffer pH 7.4). The mixture was left on ice for 30 minutes, then washed in cacodylate buffers in a descending series of concentrations. Cells were then trapped in agar, dehydrated through a series of increasing concentrations of ethanol, and embedded in Spurr’s low viscosity resin (Agar Scientific), which was allowed to infiltrate for up to a week before polymerizing overnight at 60°C. Blocks were serially sectioned at either 70 nm (ATCC culture 50388) (Fig. 2B–D,F,L,M) or 50 nm (Nebraska culture) (Fig. 2E,G–K) with a diamond knife, using either a Reichert Ultracut E or a Leica EM UC6 ultramicrotome, respectively. Serial sections were placed on pioloform-coated grids after the method of Rowley and Moran (Rowley and Moran, 1975). Thin sections were examined using, respectively, either a Hitachi H-7100 or a FEI Tecnai-12 TEM fitted with a goniometer stage.

Isolation of clones and sequencing

Genes encoding ESCRT components in *B. anathema* were identified from a database of 6937 ESTs, as part of an on-going gene survey project (M.v.d.G., G.W. and J.B.D., unpublished). Sequences were assembled from individual EST reads, with manual assessment of base quality using the chromatograms. In all cases, the coding region for each gene was determined from at least 2 \times coverage. In order to obtain this coverage or to clarify homology-searching results, additional sequence information was required for the putative Vps28, Vps4, Vps2 and Vps31 homologs of *B. anathema* as well as the putative Vps37 and Vps24 sequences from *M. balamuthi*. Consequently, clones encoding these ESTs were sequenced using standard methods to at least 2 \times coverage. cDNA clones MBE00019398 coding for Vps37 (accession GU292811) and MBE00002967 coding for Vps24 (accession GU256250) (Table 1) from *M. balamuthi* were generously provided by Andrew Roger (Dalhousie University, Halifax, Canada).

Homology searching

Functionally verified *Homo sapiens*, and *Saccharomyces cerevisiae* ESCRT sequences (Hurley, 2008), as well as their *A. thaliana* homologs, were used as queries for BLAST searches (Altschul et al., 1997) to identify homologs of ESCRT components in EST databanks of *M. balamuthi*, *H. vermiformis* and *P. polycephalum* at the National Center for Biotechnology Information (NCBI, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and in the genome of *A. castellanii* at the Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/microbial-detail.xsp?project_id=163).

All identified homologs available in the query organism were used as queries for BLASTp searches against protein databases or tBLASTn searches against nucleotide databases. The BLOSUM62 substitution matrix was used as the default scoring matrix. Only sequences that returned with an E-value of 0.05 or less were considered acceptable candidates. Reciprocal BLAST searches were then done in NCBI by using the amoebozoan sequences to search the genome of the original query (*H. sapiens*, *S. cerevisiae* or *A. thaliana*). The following criteria were used to infer homology: the original query or the same protein with a different GenBank ID must be recovered in the reciprocal BLAST as the top hit and have an acceptable E-value (<0.05), and the original query or its clear ortholog must be recovered as the top hit in the non-redundant database.

In addition to BLAST, HMM searches for Vps27 and Tom1 family member homologs were performed using the program HMMer v 2.3.2. In order to obtain VHS-GAT domain-containing proteins, HMM profiles were constructed from Vps27 and Tom1 family homologs from organisms across the diversity of eukaryotes. Conceptual proteomes were downloaded and searched manually for the following organisms: *Danio rerio* was found at the Vertebrate Genome Annotation database (VEGA, <http://vega.sanger.ac.uk/>); *Nematostella vectensis*, *Monosiga brevicollis*, *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Emiliania huxleyi*, *P. ramorum*, *Thalassiosira pseudonana* and *N. gruberi* were found at the Joint Genome Institute (JGI, <http://www.jgi.doe.gov/>); *Drosophila melanogaster* data were found at Flybase (<http://flybase.org/>); *Cryptococcus neoformans* and *S. cerevisiae* were found at the BROAD Institute (<http://www.broadinstitute.org/>); *D. discoideum* was found at dictyBase (<http://dictybase.org/>) and *E. histolytica* was found at the Sanger Institute (<http://www.sanger.ac.uk/>); *A. thaliana* data were found at The Arabidopsis Information Resource (TAIR, <http://www.arabidopsis.org/>); *Cyanidioschyzon merolae* data were found at the *C. merolae* genome project site (<http://merolae.biol.s.u-tokyo.ac.jp/>); *Physcomitrella patens* was found at Phytozome (<http://www.phytozome.net/>); and *O. sativa* was found at PlantGDB (<http://www.plantgdb.org/>). The following organismal genomic databases were found at the Eukaryotic Pathogen Database Resources (EuPathDB, <http://eupathdb.org/eupathdb/>): *Plasmodium falciparum*, *Toxoplasma gondii*, *Cryptosporidium parvum*, *Giardia intestinalis*, *Trypanosoma brucei*, *Leishmania major* and *Trypanosoma cruzi*. *Theileria parva* data were found at the NCBI (<http://www.ncbi.nlm.nih.gov/>). *Tetrahymena thermophila* was found at the J. Craig Venter Institute (JCVI, <http://www.jcvi.org/>). The cut-off for the reciprocal BLAST of candidates was an E-value of 0.05.

Criteria for homology to Vps27 were based on retrieving Vps27 as the top reciprocal match in searches of both the human and non-redundant database, as well as the presence of VHS, FYVE and UIM domains, and being whole or mostly complete proteins. Attempts were made to retrieve additional sequence data from the relevant genome project database if the protein sequence was incomplete. In some cases, this clarified homology using BLAST. In other cases of incomplete sequence, the absence of a VHS domain was used as a criterion to exclude the protein as irresolvable for the time being.

Criteria for homology to the Tom1 family were based on retrieval of a Tom1-related protein as the top reciprocal match, as well as the presence of VHS and GAT domains. Incomplete sequences were treated as above. For the Tom1 homologs the databases for *Ciona intestinalis* and *Gallus gallus* were additionally searched by BLASTp, using criteria for homology as described above.

Alignments and phylogenetics

Phylogenetic analysis was performed for the Vps24, Snf7 and Tom1 families. The Vps24 dataset included 23 sequences: the eight sequences from the query organisms and 15 amoebozoan sequences. The Snf7 dataset included 15 sequences, eight of which were query sequences and seven of which were amoebozoan. An initial dataset of all Tom1 and Vps27 homologs, which contained 84 taxa and 219 positions, was constructed. Finally, a dataset composed of verified Tom1 homologs (see criteria above) and additional sequences from *M. brevicollis*, *N. vectensis*, *G. gallus* and *C. intestinalis*, but removing closely related lineage-specific duplicates as well as sequences that failed the above homology criteria, was assembled to contain 45 taxa and 206 positions. All alignments are available from the authors upon request.

Gene sequences acquired from BLAST searching nucleotide databases were translated into proteins using the online ExPASy Translate tool (<http://www.expasy.ch/tools/dna.html>). Because the *A. castellanii* sequences were predicted from genomic contigs, introns had to be predicted and removed in silico using Sequencher 4.9 (Gene Codes) before translation into proteins.

All protein sequences were then aligned using MUSCLE v3.6 (Edgar, 2004), and the alignment was manually adjusted. Only regions of unambiguous homology were retained for analysis. ProtTest v. 2.4 (Abascal et al., 2005) was used to find the best model of protein evolution for the sequences, incorporating correction for invariable sites as well as a four-category gamma correction for rate variation among sites.

MR BAYES v. 3.2.1 (Ronquist and Huelsenbeck, 2003) was used to search treespace using 1,000,000 MCMC generations. Consensus trees were generated using a burn-in value of 25%: in each case. This was validated by plotting likelihood versus generations to ensure that no trees were included prior to the likelihood plateau. Two independent runs, each of four chains, were performed, with convergence of the results confirmed by ensuring a splits frequency of <0.1. Posterior probabilities of nodes were then applied to the most likely tree in each Bayesian MCMC analysis. Additionally, PhyML v. 2.4.4 (Guindon and Gascuel, 2003) and RAXML-VI-HP v. 2.2.3 (Stamatakis, 2006) were used for maximum-likelihood analyses, and to generate ML-bootstrap values based on 100 pseudo-replicates of each dataset. These values were then applied to the most likely tree from each of the Bayesian analyses. The tree diagram used in phylogenetic figures (Figs 3, 4 and 6) was the best Bayesian topology, with support values listed in the order of Bayesian posterior probability values/PhyML bootstrap values/RAXML bootstrap values.

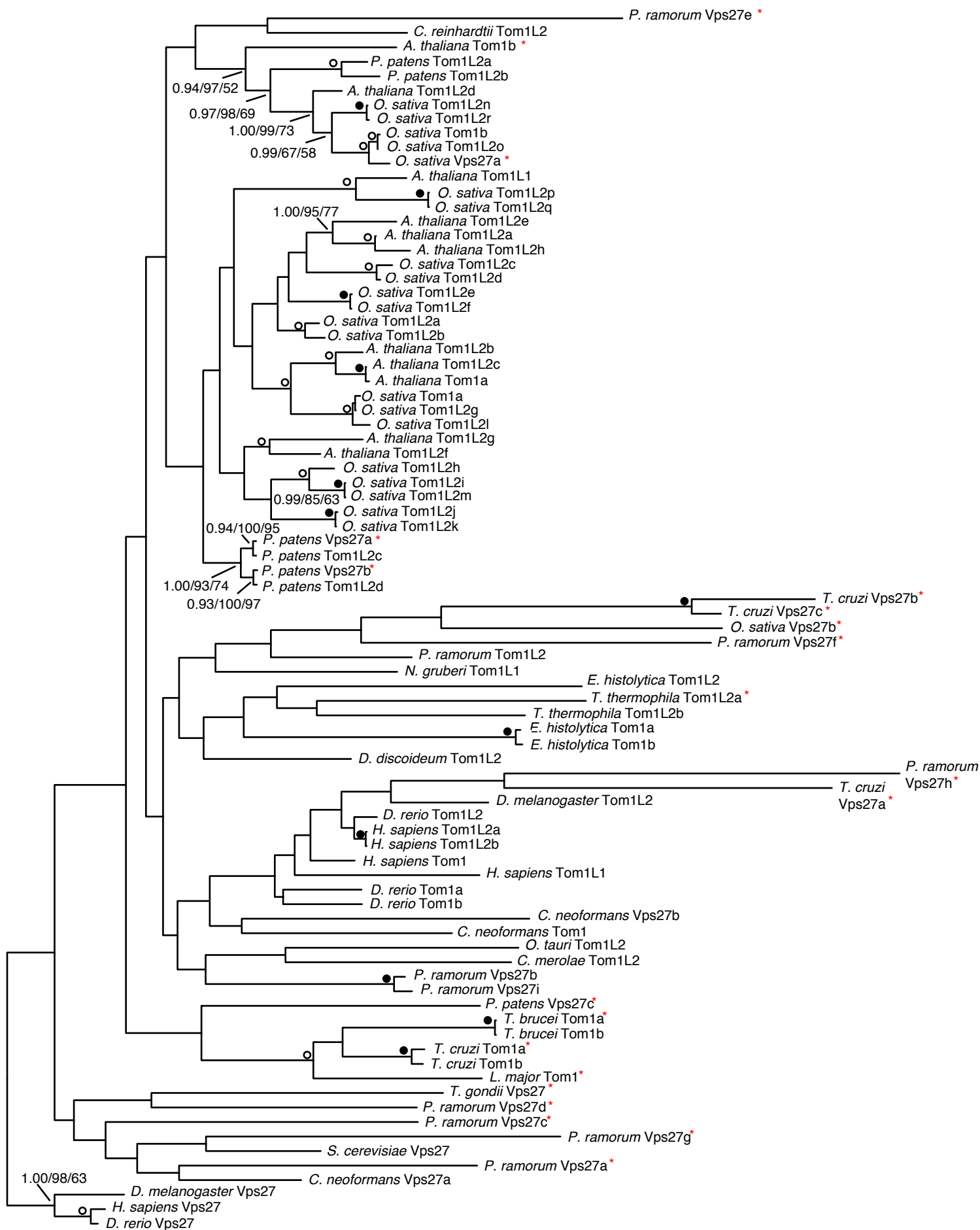
We wish to recognize the contribution of the various genome projects sampled in this study for making their data publicly available. We also wish to thank Robert Mullen (University of Guelph, Guelph, Canada) and Mark Field (University of Cambridge, Cambridge, UK) for critical comments, and Kamran Shalchian-Tabrizi (University of Oslo, Norway) and Jeffrey Silberman (University of Arkansas, Fayetteville, AR) for collaborative work on *Breviata*. This work was supported by a CoSyst-BBSRC grant to M.v.d.G., G.W. and J.B.D., as well as an NSERC Discovery Grant to J.B.D. E.K.H. was supported by a Heritage Summer Studentship. M.v.d.G. is grateful for support from the Wellcome Trust. Deposited in PMC for release after 6 months.

Supplementary material available online at
<http://jcs.biologists.org/cgi/content/full/124/4/613/DC1>

References

- Abascal, F., Zardoya, R. and Posada, D. (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.
- Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S. et al. (2005). The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**, 399-451.
- 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.
- Blanc, C., Charette, S. J., Mattei, S., Aubry, L., Smith, E. W., Cosson, P. and Letourneur, F. (2009). Dictyostelium Tom1 participates to an ancestral ESCRT-0 complex. *Traffic* **10**, 161-171.
- Dell'Angelica, E. C. (2001). Clathrin-binding proteins: got a motif? Join the network! *Trends Cell Biol.* **11**, 315-318.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.
- Edgcomb, V. P., Simpson, A. G., Zettler, L. A., Nerad, T. A., Patterson, D. J., Holder, M. E. and Sogin, M. L. (2002). Pelobionts are degenerate protists: insights from molecules and morphology. *Mol. Biol. Evol.* **19**, 978-982.
- Field, M. C. and Dacks, J. B. (2009). First and last ancestors: reconstructing evolution of the endomembrane system with ESCRTs, vesicle coat proteins, and nuclear pore complexes. *Curr. Opin. Cell Biol.* **21**, 4-13.
- Field, M. C., Gabernet-Castello, C. and Dacks, J. B. (2007). Reconstructing the evolution of the endocytic system: insights from genomics and molecular cell biology. *Adv. Exp. Med. Biol.* **607**, 84-96.
- Gruenberg, J. and Stenmark, H. (2004). The biogenesis of multivesicular endosomes. *Nat. Rev. Mol. Cell Biol.* **5**, 317-323.
- Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704.
- Haas, T. J., Sliwinski, M. K., Martinez, D. E., Preuss, M., Ebine, K., Ueda, T., Nielsen, E., Odorizzi, G. and Otegui, M. S. (2007). The Arabidopsis AAA ATPase SKD1 is involved in multivesicular endosome function and interacts with its positive regulator LYST-INTERACTING PROTEIN5. *Plant Cell* **19**, 1295-1312.
- Hurley, J. H. (2008). ESCRT complexes and the biogenesis of multivesicular bodies. *Curr. Opin. Cell Biol.* **20**, 4-11.
- Im, Y. J., Wollert, T., Boura, E. and Hurley, J. H. (2009). Structure and function of the ESCRT-II-III interface in multivesicular body biogenesis. *Dev. Cell* **17**, 234-243.
- Katoh, Y., Shiba, Y., Mitsuhashi, H., Yanagida, Y., Takatsu, H. and Nakayama, K. (2004). Tollip and Tom1 form a complex and recruit ubiquitin-conjugated proteins onto early endosomes. *J. Biol. Chem.* **279**, 24435-24443.
- Kostelansky, M. S., Schluter, C., Tam, Y. Y., Lee, S., Ghirlando, R., Beach, B., Conibear, E. and Hurley, J. H. (2007). Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. *Cell* **129**, 485-498.
- Leung, K. F., Dacks, J. B. and Field, M. C. (2008). Evolution of the multivesicular body ESCRT machinery: retention across the eukaryotic lineage. *Traffic* **9**, 1698-1716.
- Liu, N. S., Loo, L. S., Loh, E., Seet, L. F. and Hong, W. (2009). Participation of Tom1L1 in EGF-stimulated endocytosis of EGF receptor. *EMBO J.* **28**, 3485-3499.

- Minge, M. A., Silberman, J. D., Orr, R. J., Cavalier-Smith, T., Shalchian-Tabrizi, K., Burki, F., Skjaeveland, A. and Jakobsen, K. S. (2009). Evolutionary position of breviate amoebae and the primary eukaryote divergence. *Proc. Biol. Sci.* **276**, 597-604.
- Parfrey, L. W., Grant, J., Tekle, Y. I., Lasek-Nesselquist, E., Morrison, H. G., Sogin, M. L., Patterson, D. J. and Katz, L. A. (2010). Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst. Biol.* **59**, 518-533.
- Polo, S., Confalonieri, S., Salcini, A. E. and Di Fiore, P. P. (2003). EH and UIM: endocytosis and more. *Sci. STKE* **2003**, re17.
- Prag, G., Watson, H., Kim, Y. C., Beach, B. M., Ghirlando, R., Hummer, G., Bonifacino, J. S. and Hurley, J. H. (2007). The Vps27/Hse1 complex is a GAT domain-based scaffold for ubiquitin-dependent sorting. *Dev. Cell* **12**, 973-986.
- Puertollano, R. (2005). Interactions of TOM1L1 with the multivesicular body sorting machinery. *J. Biol. Chem.* **280**, 9258-9264.
- Raiborg, C. and Stenmark, H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* **458**, 445-452.
- Raiborg, C., Malerød, L., Pedersen, N. M. and Stenmark, H. (2008). Differential functions of Hrs and ESCRT proteins in endocytic membrane trafficking. *Exp. Cell Res.* **314**, 801-813.
- Roger, A. J. and Simpson, A. G. (2009). Evolution: revisiting the root of the eukaryote tree. *Curr. Biol.* **19**, R165-R167.
- Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3, Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.
- Rowley, J. C., 3rd and Moran, D. T. (1975). A simple procedure for mounting wrinkle-free sections on formvar-coated slot grids. *Ultramicroscopy* **1**, 151-155.
- Saksena, S., Wahlman, J., Teis, D., Johnson, A. E. and Emr, S. D. (2009). Functional reconstitution of ESCRT-III assembly and disassembly. *Cell* **136**, 97-109.
- Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L. and Bell, S. D. (2008). A role for the ESCRT system in cell division in archaea. *Science* **322**, 1710-1713.
- Shestakova, A., Hanono, A., Drosner, S., Curtiss, M., Davies, B. A., Katzmann, D. J. and Babst, M. (2010). Assembly of the AAA ATPase Vps4 on ESCRT-III. *Mol. Biol. Cell* **21**, 1059-1071.
- Slater, R. and Bishop, N. E. (2006). Genetic structure and evolution of the Vps25 family, a yeast ESCRT-II component. *BMC Evol. Biol.* **6**, 59.
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.
- Stillier, J. W., Duffield, E. C. and Hall, B. D. (1998). Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **95**, 11769-11774.
- Walker, G., Dacks, J. B. and Embley, T. M. (2006). Ultrastructural description of *Breviata anathema*, n. gen., n. sp., the organism previously studied as "*Mastigamoeba invertens*". *J. Eukaryot. Microbiol.* **53**, 65-78.
- Williams, R. L. and Urbe, S. (2007). The emerging shape of the ESCRT machinery. *Nat. Rev. Mol. Cell Biol.* **8**, 355-368.
- Winter, V. and Hauser, M. T. (2006). Exploring the ESCRTing machinery in eukaryotes. *Trends Plant Sci.* **11**, 115-123.
- Wollert, T. and Hurley, J. H. (2010). Molecular mechanism of multivesicular body biogenesis by ESCRT complexes. *Nature* **464**, 864-869.
- Xiao, J., Xia, H., Zhou, J., Azmi, I. F., Davies, B. A., Katzmann, D. J. and Xu, Z. (2008). Structural basis of Vta1 function in the multivesicular body sorting pathway. *Dev. Cell* **14**, 37-49.
- Yamakami, M., Yoshimori, T. and Yokosawa, H. (2003). Tom1, a VHS domain-containing protein, interacts with tollip, ubiquitin, and clathrin. *J. Biol. Chem.* **278**, 52865-52872.
- Yanagida-Ishizaki, Y., Takei, T., Ishizaki, R., Imakagura, H., Takahashi, S., Shin, H. W., Katoh, Y. and Nakayama, K. (2008). Recruitment of Tom1L1/Srcasm to endosomes and the midbody by Tsg101. *Cell Struct. Funct.* **33**, 91-100.
- Yang, M., Coppens, I., Wormsley, S., Baeovova, P., Hoppe, H. C. and Joiner, K. A. (2004). The *Plasmodium falciparum* Vps4 homolog mediates multivesicular body formation. *J. Cell Sci.* **117**, 3831-3838.
- Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809-818.



Organism	Annotation	Accession	Alignment identifier
Final Sequences			
	Vps27/Tom1		
H. sapiens	Vps27	NP_004703	HSV27
D. rerio	Vps27	XP_002661278	DRV27
G. gallus	Vps27	XP_426233	GGV27
C. intestinalis	Vps27	NP_001071892	CIV27
N. vectensis	Vps27	182116	NVV27
D. melanogaster	Vps27	NP_722830	DMV27
M. brevicollis	Vps27	2665	MBV27
C. neoformans	Vps27	XP_570941	CNV27
S. cerevisiae	Vps27	P40343	SCV27
H. sapiens	Tom1	NP_005479	HST1
H. sapiens	Tom1L1	NP_005477	HST1L1
H. sapiens	Tom1L2	NP_001076437	HST1L2
D. rerio	Tom1a	AAI16549	DRT1a
D. rerio	Tom1b	AAH56566	DRT1b
D. rerio	Tom1L2	XP_688819	DRT1L2
G. gallus	Tom1	NP_990475	GGT1
G. gallus	Tom1L1	XP_415646	GGT1L1
G. gallus	Tom1L2	XP_414813	GGT1L2
C. intestinalis	Tom1esc	XP_002128241	CIT1E
N. vectensis	Tom1esc	100273	NVT1E
D. melanogaster	Tom1esc	NP_648315	DMT1E
M. brevicollis	Tom1esc	13859	MBT1E
C. neoformans	Tom1esc a	XP_570354	CNT1Ea
C. neoformans	Tom1esc b	XP_572930	CNT1Eb
D. discoideum	Tom1esc	XP_635855	DDT1E
E. histolytica	Tom1esc a	XP_651355	EHT1Ea
E. histolytica	Tom1esc b	XP_653676	EHT1Eb
A. thaliana	Tom1esc a	NP_195002	ATT1Ea
A. thaliana	Tom1esc b	NP_564138	ATT1Eb
A. thaliana	Tom1esc c	NP_177823	ATT1Ec
A. thaliana	Tom1esc d	NP_197190	ATT1Ed
A. thaliana	Tom1esc e	NP_187491	ATT1Ee
A. thaliana	Tom1esc f	NP_181375	ATT1Ef
A. thaliana	Tom1esc g	NP_195796	ATT1Eg
A. thaliana	Tom1esc h	NP_201169	ATT1Eh
A. thaliana	Tom1esc i	NP_563762	ATT1Ei
C. reinhardtii	Tom1esc	401111	CRT1E
O. tauri	Tom1esc	11249	OTT1E
C. merolae	Tom1esc	CME025C	CMT1E
O. sativa	Tom1esc a	NP_001046532	OST1Ea
O. sativa	Tom1esc b	NP_001042484	OST1Eb
O. sativa	Tom1esc c	NP_001057544	OST1Ec
O. sativa	Tom1esc d	NP_001067438	OST1Ed
O. sativa	Tom1esc e	NP_001055830	OST1Ee
O. sativa	Tom1esc f	NP_001047818	OST1Ef
O. sativa	Tom1esc g	NP_001060807	OST1Eg
O. sativa	Tom1esc h	AAL58181	OST1Eh
O. sativa	Tom1esc i	NP_001044671	OST1Ei
O. sativa	Tom1esc j	NP_001055241	OST1Ej
P. patens	Tom1esc	196153	PPT1E
P. patens	Tom1esc	XP_001010380	TTT1E
P. ramorum	Tom1esc a	75625	PRT1Ea
P. ramorum	Tom1esc b	76391	PRT1Eb
N. gruberi	Tom1esc	68824	NGT1E
	ESCRTs		
B. anathema	Vps28	HM773427	
B. anathema	Vps2	HM773426	BAVPS2
B. anathema	Vps20	HM773425	BAVPS20
B. anathema	Vps32	HM773428	BAVPS32
B. anathema	Vps4	HM773429	
B. anathema	Vps31	HM773430	
B. anathema	Vps46	HM773431	BAVPS46
M. balamuthi	Vps23	EC703007	
M. balamuthi	Vps23	EC698844	
M. balamuthi	Vps28	EC711962	
M. balamuthi	Vps37	GU292811	
M. balamuthi	Vps22	EC710818	
M. balamuthi	Vps25	EC701301	
M. balamuthi	Vps2	EC707267	MBVPS2
M. balamuthi	Vps24	GU256250	MBVPS24
M. balamuthi	Vps4	EC708596	
M. balamuthi	Vps46	EC711152	MBVPS46
A. castellanii	Vps28	Contig5296 Contigs_from_Unchanged_contigs Contig3550	
A. castellanii	Vps22	Contig7727 Contigs_from_Unchanged_contigs Contig6633	
A. castellanii	Vps36	Contig16524 Contigs_from_Unchanged_contigs Contig17568	
A. castellanii	Vps2	Contig1459 CBINs/cbin297.fa.contigs cbin297.fa.Contig1	ACVPS2
A. castellanii	Vps4	Contig724 CBINs/cbin1496.fa.contigs cbin1496.fa.Contig1	
A. castellanii	Vps4	Contig2440 Contigs_from_Unchanged_contigs Contig1	
A. castellanii	Vps46	Contig2560 Contigs_from_Unchanged_contigs Contig155	ACVPS46
A. castellanii	Vta1	Contig2144 CBINs/cbin793.fa.contigs cbin793.fa.Contig2	
P. polycephalum	Vps37	EL564641	
P. polycephalum	Vps36	EL576160	
P. polycephalum	CHMP7	EL572472	
P. polycephalum	Vps4	EL5475991	
P. polycephalum	Vps46	EL577848	PPVPS46
H. vermiformis	Vps25	EC132121	
H. vermiformis	Vps2	EC132715	HVVPS2
H. vermiformis	Vps24	EC132318	HVVPS24
H. vermiformis	Vps31	combined EC128183 and EC133221	
H. sapiens	Vps20	NP_078867	HSVPS20
H. sapiens	Vps32a	NP_054888	HSVPS32A
H. sapiens	Vps32b	NP_789782	HSVPS32B
H. sapiens	Vps32c	NP_689497	HSVPS32C
H. sapiens	Vps60	NP_057494	HSVPS60
S. cerevisiae	Vps20	NP_013794	SCVPS20
S. cerevisiae	Vps32	NP_013125	SCVPS32
S. cerevisiae	Vps60	NP_010774	SCVPS60
D. discoideum	Vps20	XP_637438	DDVPS20
D. discoideum	Vps32	XP_643404	DDVPS32
D. discoideum	Vps60	XP_636927	DDVPS60
E. histolytica	Vps32	XP_656010	EHVPS32
E. histolytica	Vps60	XP_650324	EHVPS60
H. sapiens	Vps2a	NP_055268	HSVPS2A
H. sapiens	Vps2b	NP_054762	HSVPS2B
H. sapiens	Vps24	NP_057163	HSVPS24
H. sapiens	Vps46a	NP_002759	HSVPS46A
H. sapiens	Vps46b	NP_065145	HSVPS46B
S. cerevisiae	Vps2	NP_012924	SCVPS2
S. cerevisiae	Vps24	NP_012883	SCVPS24
S. cerevisiae	Vps46	NP_012961	SCVPS46
D. discoideum	Vps2a	XP_629585	DDVPS2A
D. discoideum	Vps2b	XP_635908	DDVPS2B
D. discoideum	Vps24	XP_638395	DDVPS24
D. discoideum	Vps46	XP_647634	DDVPS46
E. histolytica	Vps2	XP_654699	EHVPS2
Sequences from the first round of Vps27/Tom1 phylogenetics			
H. sapiens	Vps27	NP_004703	HSV27
H. sapiens	Tom1	NP_005479	HST1
H. sapiens	Tom1L1	NP_005477	HST1L1
H. sapiens	Tom1L2a iso1	NP_001028723	HST1L2a
H. sapiens	Tom1L2b iso2	NP_001076437	HST1L2b
S. cerevisiae	Vps27	P40343	SCV27
D. melanogaster	Vps27	NP_722830	DMV27
D. melanogaster	Tom1L2	NP_648315	DMT1L2
C. neoformans	Vps27a	XP_570941	CNV27a
C. neoformans	Tom1	XP_570354	CNT1
C. neoformans	Vps27b	XP_572930	CNV27b
D. rerio	Vps27	XP_002661278	DRV27
D. rerio	Tom1a	AAI16549	DRT1a
D. rerio	Tom1b	AAH56566	DRT1b
D. rerio	Tom1L2	XP_688819	DRT1L2
N. vectensis	Vps27	182116	NVV27
N. vectensis	Tom1L2	100273	NVT1L2
M. brevicollis	Vps27	2665	MBV27
M. brevicollis	Tom1L2	13859	MBT1L2
D. discoideum	Tom1L2	XP_635855	DDT1L2
E. histolytica	Tom1a	XP_651355	EHIT1a
E. histolytica	Tom1b	XP_001913512	EHIT1b
E. histolytica	Tom1L2	XP_653676	EHIT1L2
A. thaliana	Tom1L2a	NP_195002	ATT1L2a
A. thaliana	Tom1L2b	NP_564138	ATT1L2b
A. thaliana	Tom1L2c	NP_177823	ATT1L2c
A. thaliana	Tom1L2d	NP_197190	ATT1L2d
A. thaliana	Tom1L2e	NP_187491	ATT1L2e
A. thaliana	Tom1L2f	NP_181375	ATT1L2f
A. thaliana	Tom1L2g	NP_195796	ATT1L2g
A. thaliana	Tom1L2h	CAA18585	ATT1L2h
A. thaliana	Tom1L1	NP_201169	ATT1L1
A. thaliana	Tom1a	AAC00635	ATT1a
A. thaliana	Tom1b	NP_973770	ATT1b
A. thaliana	Tom1c	NP_563762	ATT1c
C. reinhardtii	Tom1L2	401111	CRT1L2
O. tauri	Tom1L2	11249	OTT1L2
C. merolae	Tom1L2	CME025C	CMT1L2
O. sativa	Vps27a	NP_001065496	OSV27a
O. sativa	Vps27b	BAF12780	OSV27b
O. sativa	Tom1a	EEE51810	OST1a
O. sativa	Tom1b	EAY79641	OST1b
O. sativa	Tom1L2a	NP_001046532	OST1L2a
O. sativa	Tom1L2b	EEE56715	OST1L2b
O. sativa	Tom1L2c	NP_001042484	OST1L2c
O. sativa	Tom1L2d	EAY73130	OST1L2d
O. sativa	Tom1L2e	EEC80523	OST1L2e
O. sativa	Tom1L2f	NP_001057544	OST1L2f
O. sativa	Tom1L2g	NP_001067438	OST1L2g
O. sativa	Tom1L2h	NP_001055830	OST1L2h
O. sativa	Tom1L2i	NP_001044671	OST1L2i
O. sativa	Tom1L2j	EAY97588	OST1L2j
O. sativa	Tom1L2k	NP_001055241	OST1L2k
O. sativa	Tom1L2l	EEC67842	OST1L2l
O. sativa	Tom1L2m	EEC71726	OST1L2m
O. sativa	Tom1L2n	NP_001047818	OST1L2n
O. sativa	Tom1L2o	AAL58181	OST1L2o
O. sativa	Tom1L2p	EAZ05339	OST1L2p
O. sativa	Tom1L2q	NP_001060807	OST1L2q
O. sativa	Tom1L2r	ABR25582	OST1L2r
P. patens	Vps27a	114284	PPV27a
P. patens	Vps27b	124457	PPV27b
P. patens	Vps27c	204293	PPV27c
P. patens	Tom1L2a	196153	PPT1L2a
P. patens	Tom1L2b	205924	PPT1L2b
P. patens	Tom1L2c	114284	PPT1L2c
P. patens	Tom1L2d	124457	PPT1L2d
T. gondii	Vps27	TGGT1_058390	TGV27
T. thermophila	Tom1L2a	XP_001023917	TTT1L2a
T. thermophila	Tom1L2b	XP_001010380	TTT1L2b
P. ramorum	Vps27a	96230	PRV27a
P. ramorum	Vps27b	75625	PRV27b
P. ramorum	Vps27c	80414	PRV27c
P. ramorum	Vps27d	95441	PRV27d
P. ramorum	Vps27e	78138	PRV27e
P. ramorum	Vps27f	72659	PRV27f
P. ramorum	Vps27g	84876	PRV27g
P. ramorum	Vps27h	72804	PRV27h
P. ramorum	Vps27i	86789	PRV27i
P. ramorum	Tom1L2	76391	PRT1L2
T. brucei	Tom1a	XP_829141	TBT1a
T. brucei	Tom1b	CBH18087	TBT1b
L. major	Tom1	XP_001687100	LMT1
T. cruzi	Tom1a	XP_803318	TCT1a
T. cruzi	Tom1b	XP_813111	TCT1b
N. gruberi	Tom1L1	68824	NGT1L1

Organism	Annotation	Accession	Domain	Pfam	Missing domains
B. anathema	Vps28	HM773427	Vps28	PF03997	
B. anathema	Vps2	HM773426	Snf7	PF03357	
B. anathema	Vps20	HM773425	Snf7	PF03357	
B. anathema	Vps32	HM773428	Snf7	PF03357	
B. anathema	Vps4	HM773429	AAA ATPase, Vps4_C	PF00004, PF09336	MIT domain (PF04212)
B. anathema	Vps31	HM773430	Not found		Bro1 domain (PF03097)
B. anathema	Vps46	HM773431	Snf7	PF03357	
M. balamuthi	Vps23	EC703007	Not found		UEV (PF05743) and Vps23 core (PF09454) domains
M. balamuthi	Vps23	EC698844	Vps23_core	PF09454	UEV domain (PF05743)
M. balamuthi	Vps28	EC711962	Vps28	PF03997	
M. balamuthi	Vps37	GU292811	Mod_r	PF07200	
M. balamuthi	Vps22	EC710818	EAP/Vps36	PF04157	
M. balamuthi	Vps25	EC701301	ESCRT-II	PF05871	
M. balamuthi	Vps2	EC707267	Snf7	PF03357	
M. balamuthi	Vps24	GU256250	Snf7	PF03357	
M. balamuthi	Vps4	EC708596	MIT, AAA ATPase	PF04212, PF00004	Vps4_C domain (PF09336)
M. balamuthi	Vps46	EC711152	Snf7	PF03357	
A. castellanii	Vps28	Contig5296 Contigs_from_Unchanged_contigs Contig3550	Vps28	PF03997	
A. castellanii	Vps22	Contig7727 Contigs_from_Unchanged_contigs Contig6633	EAP30/Vps36	PF04157	
A. castellanii	Vps36	Contig16524 Contigs_from_Unchanged_contigs Contig17568	EAP30/Vps36	PF04157	
A. castellanii	Vps2	Contig1459 CBINs/cbin297.fa.contigs cbin297.fa.Contig1	Snf7	PF03357	
A. castellanii	Vps4	Contig724 CBINs/cbin1496.fa.contigs cbin1496.fa.Contig1	MIT, AAA ATPase	PF04212, PF00004	Vps4_C domain (PF09336)
A. castellanii	Vps4	Contig2440 Contigs_from_Unchanged_contigs Contig1	MIT, AAA ATPase, Vps4_C	PF04212, PF00004, PF09336	
A. castellanii	Vps46	Contig2560 Contigs_from_Unchanged_contigs Contig155	Snf7	PF03357	
A. castellanii	Vta1	Contig2144 CBINs/cbin793.fa.contigs cbin793.fa.Contig2	DUF605/Vta1-like	PF04652	
P. polycephalum	Vps37	EL564641	Mod_r	PF07200	
P. polycephalum	Vps36	EL576160	Vps36_ESCRT-II	PF11605	
P. polycephalum	CHMP7	EL572472	Not found		Snf7 domain (PF03357)
P. polycephalum	Vps4	EL5475991	MIT, AAA ATPase	PF04212, PF00004	Vps4_C domain (PF09336)
P. polycephalum	Vps46	EL577848	Snf7	PF03357	
H. vermiformis	Vps25	EC132121	ESCRT-II	PF05871	
H. vermiformis	Vps2	EC132715	Snf7	PF03357	
H. vermiformis	Vps24	EC132318	Snf7	PF03357	
H. vermiformis	Vps31	combined EC128183 and EC133221	BRO1	PF03097	