

SCIENTIFIC REPORTS



OPEN

Membrane Trafficking Modulation during *Entamoeba* Encystation

Emily Herman¹, Maria A. Siegesmund², Michael J. Bottery^{1,3}, Ronny van Aerle^{1,2,4}, Maulood Mohammed Shather², Elisabet Caler^{5,6}, Joel B. Dacks¹ & Mark van der Giezen^{1,2}

Entamoeba histolytica is an intestinal parasite that infects 50–100 million people and causes up to 55,000 deaths annually. The transmissible form of *E. histolytica* is the cyst, with a single infected individual passing up to 45 million cysts per day, making cyst production an attractive target for infection control. Lectins and chitin are secreted to form the cyst wall, although little is known about the underlying membrane trafficking processes supporting encystation. As *E. histolytica* does not readily form cysts *in vitro*, we assessed membrane trafficking gene expression during encystation in the closely related model *Entamoeba invadens*. Genes involved in secretion are up-regulated during cyst formation, as are some *trans*-Golgi network-to-endosome trafficking genes. Furthermore, endocytic and general trafficking genes are up-regulated in the mature cyst, potentially preserved as mRNA in preparation for excystation. Two divergent dynamin-related proteins found in *Entamoeba* are predominantly expressed during cyst formation. Phylogenetic analyses indicate that they are paralogous to, but quite distinct from, classical dynamins found in human, suggesting that they may be potential drug targets to block encystation. The membrane-trafficking machinery is clearly regulated during encystation, providing an additional facet to understanding this crucial parasitic process.

Entamoeba histolytica is one of the most common causes of death due to parasites in the world¹. It is thought to infect 50–100 million people each year of whom 55,500 succumb due to dysentery, amoebic colitis and invasive disease such as amoebic liver abscess^{2,3}. Although a global pathogen, it is a major issue in tropical areas within the Indian subcontinent, Africa and South and Central America⁴. The majority of *E. histolytica* infections are asymptomatic cyst spreaders, but 10% of infections will develop disease³, with symptoms including dysentery, colitis, and amoebic granulomas⁵. If arteries are perforated during infection, amoebae can spread to the liver and brain causing abscesses, to the heart causing pericarditis, and to the lungs causing pleuropulmonary disease⁵. Invasive infection and dysentery is commonly treated with nitroimidazole-derived drugs⁶ and the aminoglycoside paromomycin, but surgical intervention may be necessary in cases of fulminant colitis^{7,8}.

E. histolytica parasites are transmitted by ingestion of cyst-contaminated water. While the average infectious dose is >1,000 organisms, a single infectious individual can pass up to 45 million cysts in their stool per day⁹. After exiting the host, cysts must survive the drying, nutrient poor, and thermo-variable conditions of the environment for several weeks to months. Once taken up, cysts face the harsh acidity of the stomach and the majority of the small intestine, rich with digestive enzymes, as excystation occurs later in the terminal ileum¹⁰. After excystation, the *Entamoeba* trophozoite divides and colonises the colon, in some cases causing tremendous tissue destruction. It is here that trophozoites form the hardy and infectious cysts that are excreted to the environment⁵.

The robustness and prodigious quantity of cysts shed by infected individuals means that communities in the developing nations that lack proper sanitation are particularly affected⁴. As a diarrheal disease, *Entamoeba* infection influences child mortality, development, and worker productivity, promoting poverty¹¹. In regions with unreliable water treatment and waste disposal infrastructure, and where access to medicines for treatment and symptom management may be insufficient, controlling the spread of the environmental cyst form of *Entamoeba* is of utmost importance. Several groups have sought to elucidate the systems and molecules that are involved

¹Department of Cell Biology, Faculty of Medicine and Dentistry, University of Alberta, T6G 2H7, Edmonton, Alberta, Canada. ²Biosciences, University of Exeter, Stocker Road, Exeter, EX4 4QD, UK. ³Department of Biology, University of York, Heslington, York, YO10 5DD, UK. ⁴Centre for Environment, Fisheries, and Aquaculture Science (Cefas), Barrack Road, The Nothe, Weymouth, Dorset, DT4 8UB, UK. ⁵J. Craig Venter Institute, 9714 Medical Center Drive, Rockville, MD, 20850, USA. ⁶Present address: National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), 6701, Rockledge Drive, Room 9144, Bethesda, MD, 20892-7950, USA. Maria A. Siegesmund and Michael J. Bottery contributed equally to this work. Correspondence and requests for materials should be addressed to J.B.D. (email: dacks@ualberta.ca) or M.v.d.G. (email: m.vandergiezen@exeter.ac.uk)

in encystation, recently reviewed in Mi-ichi *et al.* (2016) including Gal-lectin production, catecholamine pathway signalling, cholesterol sulphate synthesis, heat shock protein 90, chitin metabolism, proteolytic systems, and potentially enolase¹². However, little is known about the underlying cellular trafficking events that facilitate encystation, and furthermore, how the quiescent cyst is prepared for eventual excystation.

The process of forming a stable, resistant cyst relies on the specific, ordered secretion of cyst wall proteins. It is thought that the cyst wall is assembled in three phases, according to the “wattle-and-daub” model¹³. In the foundation stage, the lectin Jacob is trafficked to the cell surface and there binds constitutively expressed Gal/GalNAc lectins^{13,14}. During the “wattle” stage, chitin is synthesized and secreted, likely crosslinked by Jacob’s tandemly arranged chitin-binding domains. Jacob lectins and chitin are seen in separate vesicles in early encystation (12 hours post induction, hpi), and have begun to accumulate at the cyst wall at 24–36 hpi¹³. Although the timing is not clear, the chitin-cleaving enzyme chitinase and deacetylases trim and deacetylate extracellular chitin^{15,16}. Finally, the Jessie3 lectin, which binds chitin and may also self-aggregate, solidifies the cyst wall in the “daub” phase, making it impermeable to small molecules¹³. Jessie3 is observed in vesicles beginning at 36 hpi, and is found in the cyst wall at 72 hpi¹³. The heavy secretory load clearly implicates the membrane trafficking system (MTS) in encystation.

Membrane trafficking is the movement of proteins and other macromolecules throughout the organelles of the endomembrane system, including endo- and exocytosis. It is indeed a defining feature of eukaryotic cells, and has been shown to be critical for parasite infection¹⁷ and evasion of the immune system¹⁸. Several studies have implicated the MTS in *Entamoeba* parasitism^{19–21}, and specifically encystation. Membrane trafficking events are regulated by small GTPases such as Rabs, involved with vesicle fusion dynamics, and Arfs, which regulate the assembly of protein coats on vesicles. Fourteen Rab genes were found to be upregulated during encystation by a microarray screen²². A targeted bioinformatics analysis has shown that *Entamoeba* has expanded its Rab protein family through gene duplication; it has over 100 Rab proteins²³, many of which lack homologues in human²⁴, raising the possibility that some divergent MTS proteins may be potential targets for drugs abrogating cyst formation. Furthermore, the *E. histolytica* genome paper reported that the amoebae contain basic vesicle transport machinery common to eukaryotic cells, and in addition to Rabs, it has also expanded its Arf GTPase complement²⁵. However, an in-depth assessment of the *Entamoeba* MTS, and how it relates to encystation, has not yet been done. Due to its importance in cyst formation, we have taken a comparative genomics and transcriptomics approach to define the MTS gene complement of *Entamoeba*, and then assessed how these genes are modulated during encystation. A recent transcriptomic analysis of *E. histolytica* pathogenesis relative to its environment highlights the relevance of large-scale genomic approaches to understanding multiple aspects of this parasite’s biology and lifestyle²⁶.

One membrane trafficking factor potentially involved in cyst formation in *E. histolytica* is the dynamin family of enzymes. Dynamins are large GTPase proteins involved in membrane remodelling^{27,28}, although the exact mechanism by which dynamin drives membrane fission is still debated²⁹. In general, dynamins first polymerise around the neck of a vesicle forming a helix, and then induce scission through tube constriction and/or membrane destabilization, by hydrolysing GTP. They are essential for myriad cellular processes, such as endocytosis, organelle division, cytokinesis, vesicle scission, and cytoskeletal organisation^{29,30}. In a previous study of *Giardia intestinalis*, an unrelated enteric cyst-forming parasite, its single dynamin-related protein (Drp) was shown to be necessary in cyst formation³¹. It is associated with encystation-specific vesicles, and expression of a dominant negative Drp mutant blocks *Giardia* encystation³¹. Because of the involvement of dynamin in membrane trafficking, and the requirement of Drp function for *Giardia* stage conversion, we focused particularly on the expression of dynamin-related proteins during *Entamoeba* encystation.

As *E. histolytica* cysts do not form readily *in vitro*, the closely related model *Entamoeba invadens* is often used to study cyst formation. *E. invadens* is a reptile pathogen, and encystation is easily induced under axenic conditions by glucose starvation³². We induced encystation in *E. invadens* and performed RNA-Seq on mRNA from the trophozoite (0 hpi), early encystation stage (24 hpi), late encystation stage (48 hpi), and the mature cyst (72 hpi), and identified MTS gene expression patterns during this process. In addition to showing the up-regulation of previously identified encystation genes such as Jacob, Jessie and chitinase, we found that the secretory pathway is generally up-regulated during cyst formation, as well as TGN-endosome trafficking. Furthermore, we found that the mature cyst contains specific mRNAs associated with clathrin-mediated endocytosis, potentially in preparation for excystation. Two highly divergent Drps are also implicated by their predominant expression and by their changed localization during encystation.

Results

Transcriptome assembly. We induced encystation in *E. invadens* and performed transcriptomics in trophozoites, during early and late encystation, and in the mature cyst. Transcriptome sequencing resulted in 40 to 160 million reads per time point, with the fewest being at 48 hpi (late stage encystation) and the highest being at 0 hpi (trophozoite). After pre-processing to trim adapters and filter poor-quality reads, the number of reads were 34.3 million reads at 0 hpi, 27.8 million reads at 24 hpi, 9.1 million reads at 48 hpi, and 26.4 million reads at 72 hpi. These reads were mapped to the *E. invadens* IP-1 genome with PASA identifying 116 potential new transcripts. The *E. invadens* genome contains 11,553 predicted proteins, and only 473 have expression values of 0 FPKM at all time points. Therefore, approximately 96% of the *E. invadens* genome is transcribed at least at one time point during encystation. Of the 116 novel transcripts, only 26 retrieved any other sequence when used to search the NCBI non-redundant nucleotide and protein databases, and 11 of these are related to transposable element sequences (Supplementary Table S1). All genes are listed with expression values in Supplementary Table S1.

***E. histolytica* and *E. invadens* have a similar repertoire of MTS genes.** In order to support the validity *E. invadens* as a model to study the MTS during encystation, we used comparative genomics to assess

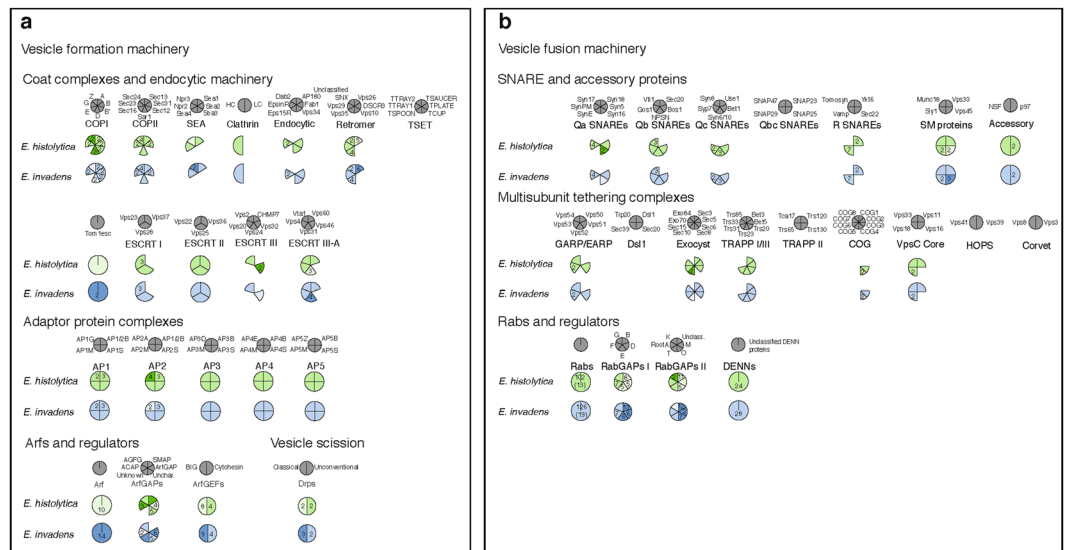


Figure 1. Coulson plot showing vesicle formation machinery (a) and vesicle fusion machinery (b) in *E. histolytica* and *E. invadens*. Filled sectors indicate presence of a homologue while empty sectors indicate that it could not be identified. Numbers indicate paralogs. Sectors are shaded lighter or darker to indicate where the two *Entamoeba* species differ in copy number.

the complement of MTS genes in *E. histolytica* and *E. invadens*. The predicted proteomes of *E. invadens* and *E. histolytica*, and the *E. invadens* transcriptome generated in this study, were examined for the presence of MTS genes using molecular data from human and other model organisms using the homology searching tool BLASTP. Figure 1a and b show the presence and paralogue number of vesicle formation and vesicle fusion machinery, respectively, in the two *Entamoeba* genomes. Orthologue accession numbers for all MTS components searched are available in Supplementary Table S2. In general, the presence/absence of MTS genes is conserved between the two species, with 246 of 367 identified genes showing one to one orthology. Excluding Arf and Rab GTPases and their regulators, there are 11 genes for which *E. invadens* and *E. histolytica* have a different number of paralogs, and in most of these cases, only one or two duplications have occurred.

The Arf and Rab GTPases, the TBC (Tre-2/Bub2/Cdc16) proteins (Rab GTPase Activating Proteins; Rab GAPs), and the DENN domain-containing proteins (Guanine nucleotide Exchange Factors for Rabs; RabGEFs) are highly paralogous gene families. Phylogenetic analyses of Arf and Rab GTPases and their regulators were undertaken in order to determine 1:1 orthologues and independent expansions within *Entamoeba*. Supplementary Table S3 is a summary of this work, with the corresponding phylogenies found in Supplementary Figures S1M–N, S1R–T, S2A–B, and Supplementary Files 3 and 4. With the exception of the Arfs, more than half of these proteins have direct 1:1 orthology between *E. invadens* and *E. histolytica*, while less than 1/3 of the proteins are ‘singletons’ (i.e. one or more gene family members in one *Entamoeba* species with no clear orthologues in the other). The rest of the proteins make up orthologous groups with some expansion or loss events (i.e. 1 + :1 or 1:1 + orthology). Overall, the vast majority of GTPases and their regulators share orthology between the two species (>75%), although there have been multiple independent expansions in both species. While there is divergence in this one sub-system of the membrane-trafficking machinery between the two organisms, the overall similarity of the remaining MTS complement in *E. histolytica* and *E. invadens* is high. 130 of 155 genes have 1:1 orthology, and there are 61 cases where a factor deduced as present in the LECA was not identified in either genome. Therefore, we suggest that *E. invadens* is a good proxy for studying encystation in the human pathogen *E. histolytica*.

The vesicle formation machinery (Fig. 1a) is largely complete relative to that of human and yeast model systems, despite the high sequence divergence of *Entamoeba* spp. genes. The only complex that appears to be absent given its complete lack of identified subunits is TSET, a recently discovered vesicle coat complex that has been partially or fully lost numerous times in eukaryotes³³. Although other complexes show partial losses, it is not clear whether these are real gene absences or false negatives due to high sequence divergence, and generally enough members of the complex are present that it may still function even if some subunits are truly missing. There has been a clear *Entamoeba* lineage-specific duplication of the coat complex COPI, which is responsible for retrograde transport in the Golgi, and between the Golgi and ER³⁴. There has also been a notable expansion of the Vps26 subunit of the retromer complex in both species, which plays a role in cargo preference for the complex in yeast and human^{35,36}. Retromer is responsible for the recycling of lysosomal cargo receptors from the endosomes to the TGN in yeast and human³⁷, and in *Entamoeba*, it has been shown to be targeted to pre-phagosomal vacuoles (PPVs), an *Entamoeba*-specific endolysosomal organelle important for pathogenesis²¹. Finally, there are duplications in the large subunits of the adaptor protein (AP) complexes 1 and 2, which interact with the vesicle coat proteins in TGN-endosome and endocytic trafficking, respectively^{38,39}.

Common losses are observed in both *Entamoeba* species for the vesicle fusion machinery (Fig. 1b). Anterograde trafficking from the ER to the Golgi is supported by the presence of SNARE proteins Syntaxin 5, Bos1, Ykt6⁴⁰, *inter, alia* and potentially Syntaxin of plants 7 (Syp7)⁴¹, as well as the SM protein Sly1 and the majority of the multi-subunit tethering complex (MTC) TRAPP1⁴². Conversely, neither the SNARE proteins associated with retrograde Golgi-ER trafficking, nor the MTC Dsl1, appear to be present. Only two subunits were identified of the Conserved Oligomeric Golgi (COG) complex, an MTC involved in retrograde Golgi trafficking^{43,44}. However, as the COPI coat complex is present (and certain subunits duplicated), and Golgi-ER retrograde trafficking is an integral feature of ER resident protein recycling and cisternal maturation^{34,45–47}, it is likely that this trafficking pathway is functional.

Genes involved in endosome-to-TGN trafficking are conserved and in some cases expanded; for example, Syntaxin 16, Vti1, Syntaxins 6/10, and Vesicle Associated Membrane Proteins (VAMPs)⁴⁰, *inter, alia* the Sec. 1/ Munc18-like protein Vps45⁴⁸, and a partial GARP MTC complex⁴⁹. This mirrors the expansions of the retromer subunit Vps26, as retromer is the coat complex that functions in this pathway. The endosome-to-lysosome trafficking pathway appears to be somewhat incomplete, as the SNAREs syntaxin 7 and syntaxin 8 could not be confidently identified, and while three out of four VpsC core components were found, we could not identify any of the HOPS (Vps41, Vps39) or Corvet (Vps3, Vps8) accessory proteins.

Two secretory SNARE subfamilies have also undergone expansion. While the SNARE proteins involved in constitutive versus regulated secretion are not well established outside of mammals⁵⁰, there is evidence for an expansion of plasma membrane syntaxins, as well as VAMPs, which can function at the plasma membrane. The exocytic SM protein Munc18 was identified⁵¹, as well as a partial, but potentially functional, exocyst complex; the MTC involved in exocytosis⁵². This expansion certainly raises the possibility that extra machinery may support secretion of cyst components. In order to determine which MTS proteins are specifically involved in this process, we assessed their expression patterns during encystation.

Genes differentially expressed during encystation fall into ten clusters of expression patterns. Statistical pair-wise comparisons of each time point were conducted to identify significantly differentially expressed genes using edgeR⁵³. This resulted in a set of 9,073 genes identified as differentially expressed between at least two time points, of which 4,987 were considered to be significant (defined as FDR < 0.01, fold change > 2). As the dispersion between genes was set *a priori* rather than estimated from the data, we chose not to focus on differential expression of individual genes. We opted instead to cluster genes based on their expression patterns, and then identify the membrane trafficking pathways for which a clear pattern is observed.

Ten clusters with elevated expression at one or more time point were identified using *k*-means clustering (Fig. 2). In general, the subclusters represent three over-arching patterns: genes whose expression increases during early and/or late encystation, but decreases in the mature cyst; genes whose expression decreases during encystation; and genes whose expression increases in the mature cyst. The first pattern – genes up-regulated during encystation – includes subclusters 1, 2, 5, and 9. Within this group, it is possible to further distinguish genes whose expression remains constant in both early (24 hpi) and late encystation (48 hpi, subcluster 2) from those whose expression peaks in late encystation (subclusters 1, 5, and 9). Subclusters 3 and 8 are characterized by a general decrease in expression during the encystation process, including in the mature cyst (72 hpi). Finally, genes whose expression peaks in the mature cyst make up subclusters 4, 7, and 10. Subcluster 7 is particularly intriguing, as its members are generally down-regulated during early and late encystation, but return to expression levels equivalent to or slightly higher than in the trophozoite (0 hpi). Subcluster 10, on the other hand, shows a pattern of steadily increasing expression throughout encystation and in the mature cyst. Subcluster 6 shows a pattern of downregulation between the trophozoite stage and early encystation (24 hpi); however, the magnitude of this change is low relative to the other expression patterns, and therefore we have chosen to focus on the other subclusters. Jacob lectins are found in subcluster 9, except Eijacob6 and Eijacob7 which are found in subclusters 4 and 5, respectively. Jessie lectins are found in subcluster 5. This is congruent with the notion that the Jacob proteins act at an earlier stage, as subcluster 9 describes genes that have high expression during early and late encystation, and Jessie proteins act later, as subcluster 5 describes genes specifically up-regulated in late encystation. Furthermore, all four chitinases have expression patterns in agreement with previous RT-PCR work⁵⁴ and are also members of subclusters 5 and 9.

The variance between these clusters can be accounted for within the first two principle components of the data set (Supplementary Figure S3). The majority of the variance is explained by principle component 1, accounting for 58.4% of the total variability within the data set. This component is dominated by the contrast in expression of genes within trophozoite (0 hpi) and within late encystation (48 hpi). Principle component 2 accounts for 33% of the variance within the data set, representing the contrast in expression of genes between the encystation (24 hpi) and cyst (72 hpi) stages. Differences within expression between trophozoite (0 hpi) and within late encystation (48 hpi) play a prominent role in distinguishing the expression of these clusters.

Membrane trafficking gene expression during encystation. Of approximately 400 MTS genes, 223 were identified as differentially expressed and could be clustered into the groups described above during encystation (Supplementary Table S1). Strikingly, we found that 1/4 of the ~400 trafficking components searched for in *E. invadens* encode multiple paralogs where two or more are found in subclusters with different general expression patterns. These findings point to a complex MTS with potentially functionally distinct paralogs that can be swapped in and out of protein complexes to finely tune trafficking processes during encystation. Rather than a blanket increase or decrease in expression of some trafficking factors over others, this suggests that the MTS is highly modulated and a critical part of encystation.

We identified 18 genes involved in secretion in subclusters 1, 2, 5, or 9, which show a pattern of up-regulation during encystation, and which do not have paralogs with opposing expression patterns (Supplementary

Gene expression profiles

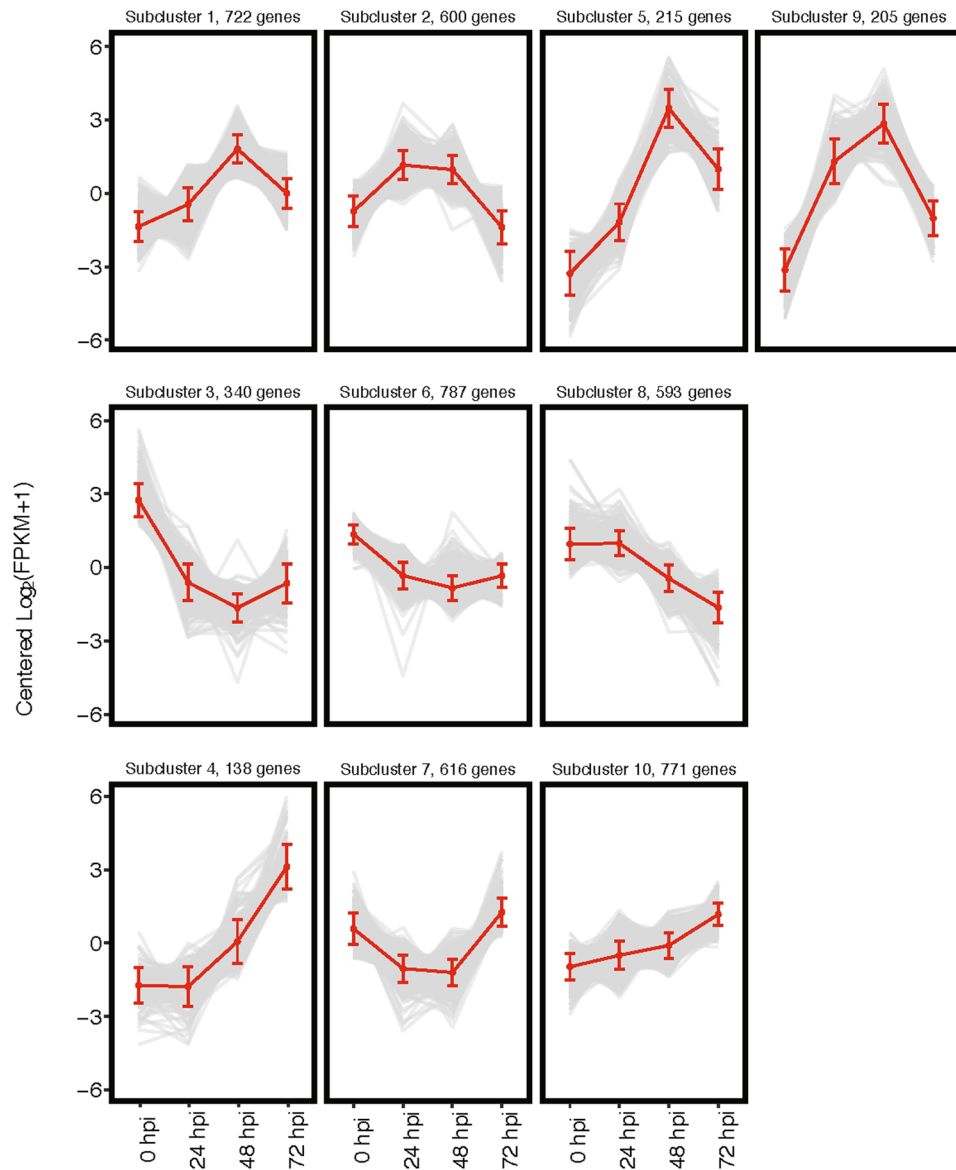


Figure 2. Differentially expressed transcripts cluster into ten expression profiles. Plots of expression profiles (centred \log_2 FPKM) of transcripts that are significantly differently expressed at least one time point ($FDR < 0.01$ \log_2 fold change > 2) split into ten subclusters defined using both k -means and MCL clustering algorithms. The number of genes in each cluster is provided above respective plots. Red points represent the cluster's mean expression profile.

Table S2). Several members of the COPII vesicle coat complex, responsible for the initial ER-to-Golgi step of trafficking secreted material, are up-regulated during encystation (Sar1⁵⁵ and two paralogues of Sec. 24). Intra-Golgi and Golgi-ER retrograde trafficking is also represented in these subclusters, which include two of the six subunits of the COPI vesicle coat complex required for intra-Golgi and Golgi-ER trafficking, four BIG-like ArfGEF paralogues, as well as several SNARE proteins that function in this pathway (Syntaxin 5, Gos1, Bos1, Bet1) and the SM protein Sly1, which helps disassemble this SNARE complex. Post-Golgi secretory SNAREs are members of these subclusters: Syntaxin PM (two paralogues), and Syp7 as well as the cognate SM protein Munc18. Syp7 is a plant syntaxin thought to be involved in secretion in plants⁴¹, and due to its absence in human, it may represent a therapeutic target to disrupt encystation.

Rab1 and EhRabA, thought to function in early secretion^{56,57}, appear to be down-regulated during encystation. However, they may play multiple roles in trafficking: while the *Entamoeba* Rab1 homologue has not been functionally characterized, Rab1 proteins in *Dictyostelium* cells are recruited to phagosomes⁵⁸, and EhRabA has been shown to be involved in cell motility⁵⁹.

The TGN-endosome recycling pathway is more modestly represented in subclusters 1, 2, 5, and 9, with five members. These include the SNARE protein Syntaxin 16, its cognate SM protein Vps45, EpsinR, and Fab1 (PIKfyve in mammals), which as well as a subunit of the GARP multi-subunit tethering complex, Vps51, is involved in the tethering of vesicles during fusion.

Overall, there is a weak endocytic/phagocytic signal in subclusters 3 and 8 (five members), whose expression decreases during encystation. These include both paralogs of Rab8, two cytohesin-like ArfGEFs, and Rab5. Contrary to this, the phagocytic Rab EhRabB appears to be up-regulated during encystation. However, this gene is known to be activated by heat shock⁶⁰, and may be activated as part of the stress response cascade that has been proposed to regulate encystation⁶¹, rather than as part of an increase in phagocytic machinery. The *E. invadens* genome has many genes with multiple paralogs, and it is often the case that while one paralog of a gene appears to be down-regulated during encystation (in subcluster 3 or 8), the other paralog does not change in expression. As such, there are few genes and pathways that clearly exemplify a pattern of downregulation. Despite the high modulation and differential expression of MTS components, it is relatively rare for two paralogs of any given gene to have ‘opposing’ expression patterns, i.e. one paralog is a member of a subcluster whose expression increases during encystation, and the other is part of a subcluster whose expression decreases under those same conditions. The majority of these cases occur in highly paralogous gene families, such as the TBC-B family of RabGAPs, or the exocytic SNARE protein Vamp7. Given the extensive duplications and diversification in these gene families, we are hesitant to use their subcluster membership as evidence of an emphasized MTS pathway.

Curiously, there are 57 MTS genes whose expression increases from late encystation (48 hpi) to the mature cyst (72 hpi; subclusters 4, 7, 10). The most obvious example is the ESCRT complex proteins, where 7 of 17 are found in these subclusters. The ESCRT complexes are responsible for multivesicular body biogenesis, a mechanism by which plasma membrane and cytosolic proteins can be targeted to the lysosome for degradation. Other members of these subclusters that are involved in endosome recycling and endo-lysosomal function include the adaptor protein complex AP1 γ subunit (two paralogs), EpsinR, Syntaxin 6 or 10, two putative DENN2 paralogs, Rab7, TBC-B, and TBC-F. Clathrin heavy chain is also highly expressed in the mature cyst. There are 7 secretory pathway genes that make up these subclusters, including two members of the COPII coat complex (Sec. 23 and Sec. 31), two non-human SNARE proteins thought to function in ER-to-plasma membrane trafficking in plants (NPSN11 and Syp7), a plasma membrane syntaxin, the COPI subunit COPB, and two BIG-like ArfGEFs. Finally, there are three ArfGAP1, one SMAP, and three uncharacterized ArfGAP paralogs in these subclusters, although their interacting Arf proteins and therefore functions are unknown. Other small G proteins and their regulators that could not be classified, or whose function is unknown, include five RabGEFs, three RabGAPs, two Arf proteins, and nine Rabs. These mRNAs are highly expressed in the mature cyst, although it is not clear whether they are translated into protein products at this stage. As the mature cyst is quiescent, it is possible that this is done to prepare for excystation, so that the cell can quickly perform early endocytic, exocytic and recycling functions prior to beginning the gene expression program seen in trophozoites. This may be analogous to increases in mRNA stability seen in late stage trophozoites of the malaria parasite *Plasmodium falciparum*, which has been suggested to allow nascent merozoites to rapidly activate their development cycle upon invasion of a new erythrocyte⁶². However, two proteomic analyses of *E. histolytica* have identified several membrane trafficking proteins (among others) present in the cyst^{63,64}, suggesting that at least some MTS proteins are retained.

Presence of divergent *Entamoeba* dynamin proteins during encystation. One final membrane-trafficking gene of interest was identified as a member of subcluster 5, a dynamin-related protein (Drp). DRPs have been shown to be involved in membrane trafficking steps such as vesicle scission and vesicle fusion, as well as other cellular processes requiring membrane deformation including cytokinesis. Given that the single dynamin encoded by the enteric parasite *G. intestinalis* is required in its encystation³¹, we asked whether Drps might also play a role in encystation in *Entamoeba*.

The *E. histolytica* genome encodes four dynamin homologues, EhDrp1 (EHI_013180), EhDrp2 (EHI_052740), EhDrp3 (EHI_174650), and EhDRP4 (EHI_139440). We further identified their homologues in the *E. invadens* genome: EiDrp1 (EIN_081190), EiDrp2 (EIN_114780), EiDrp2a (EIN_428080), EiDrp3 (EIN_376410) and EiDrp4 (EIN_080030), based on sequence similarity.

While Drp 1–4 are comparable in length and molecular weight to the other dynamin family members, Drp3 and Drp4 appear to differ in domain structure from other dynamins. While *Entamoeba* Drp1 and 2 contain the classic N-terminal large GTPase domain and C-terminal GTPase effector domain, only the N-terminal large domain was identified in Drp3 and Drp4 using Pfam⁶⁵, SMART⁶⁶, and NCBI’s Conserved Domain Database⁶⁷. However, Drp3 and Drp4 are alignable in the region downstream of the GTPase domain, suggesting sequence divergence and perhaps non-identified functional domains.

Phylogenetic analysis of DRPs in *Entamoeba* spp. To assess how the *Entamoeba* Drps are phylogenetically related to other dynamins, we performed extensive global and amoebozoan phylogenetic analyses. The initial dataset contained 133 dynamins and dynamin related sequences (Supplementary Figure S4, which was then separated into Drp1 and Drp2 sequences (Supplementary Figure S5), and Drp3 and Drp4 sequences, with an emphasis on amoebozoan taxa (Fig. 3). The phylogenetic reconstructions clearly show conservation of the orthologues among the *Entamoeba* species. *Entamoeba* Drp1 and Drp2 clustered separately from other functionally characterized members of the dynamin protein family (Supplementary Figure S4) but their function is difficult to discern. However, *Entamoeba* Drp3 and Drp4 formed a well-supported relationship with free-living non-parasitic Amoebozoa, and the archaeplastid and Stramenopile lineages involved in cell and chloroplast division (Supplementary Figure S4 and Fig. 3). *Entamoeba* Drp3 and Drp4, together with the amoebozoan *D. discoideum* DlpC and *Polysphondylium pallidum* Drp homologue (EFA76188) form a well-supported clade, with the *Entamoeba* Drp4 clade being the outgroup to the other sequences. Importantly, there is no such human dynamin

Phylogeny of Drp3 and Drp4 proteins

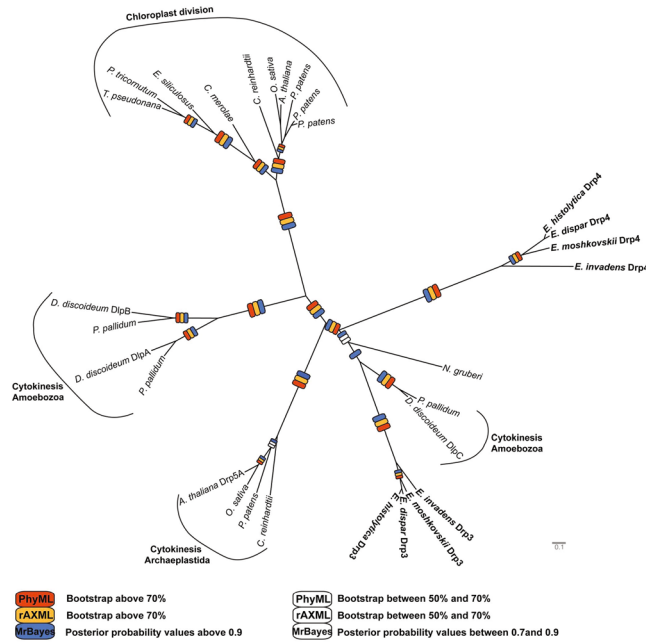


Figure 3. Phylogenetic analysis of the unusual dynamin-like proteins Drp3 and Drp4. This protein family contains dynamins involved in chloroplast division and cytokinesis and no homologs have been identified in humans. The PhyML Maximum Likelihood topology is shown. Bootstrap values and posterior probability values were calculated using PhyML, RAXML and MrBayes. The dataset contained 29 sequences with 572 informative residues (see Supplementary Table S7 for accession numbers).

orthologue, therefore these divergent *Entamoeba* DRPs may be prospective therapeutic targets if they are relevant to cyst formation.

Surprisingly, the basal red alga *Cyanidioschyzon merolae* groups with the phaeophycete taxa *Thalassiosira pseudonana*, *Phaeodactylum tricornutum* and *Ectocarpus siliculosus*, whose secondary plastid originates from a red algal endosymbiont (Fig. 3). This might suggest that these chloroplast-containing stramenopiles acquired this dynamin through the secondary plastid endosymbiosis event. This reconstruction is in agreement with an earlier analysis that suggested a link between cytokinesis and chloroplast division⁶⁸.

Divergent Drps are significantly expressed in late encystation stage in *E. invadens*. Identification of these unusual Drps prompted us to look more closely at their expression during encystation. To this end, exponentially growing cells were glucose starved to induce cyst formation. In order to see more specifically how these Drps are expressed, we performed reverse transcription PCR and quantitative PCR on mRNA sampled at the following time points: 0 hpi, 24 hpi, 28 hpi, 32 hpi, 36 hpi, 40 hpi, 44 hpi, and 72 hpi. Cyst formation was monitored by fluorescent microscopy using Calcofluor White, which stains the chitin cell wall (Supplementary Figure S6). In addition to monitoring cyst formation, total RNA was isolated at each time point and the expression of chitinases (known encystation markers) was analysed (Supplementary Figure S7). In agreement with previous studies⁵⁴, chitinases are expressed during cyst formation, with the highest expression between 32 and 44 hpi.

During cyst formation, reverse transcription PCR revealed that all *E. invadens* Drps are up-regulated, however the degree of which is stage-specific for individual genes (Fig. 4a). EiDrp1 and EiDrp2 are expressed at low levels in the trophozoite stage, but EiDrp1 expression increases after 24 hpi and remains high until the late cyst stage, while EiDrp2 expression peaks at 28–32 hpi, and drops but remains constant after 36 hpi. Drp2a is constitutively highly expressed, suggesting a general cellular role for this gene.

RT-PCR, qPCR and Western blotting results show that the two divergent Drps are predominantly expressed during cyst formation (Fig. 4 and Supplementary Figure S8). To directly compare gene expression, we determined the PCR amplification efficiency for each primer pair, with values of 106% for EiDrp3, 104% for EiDrp4, and 109% for the reference gene SKI-interacting protein SKIP. EiDrp3 expression increases during early encystation (28 hpi), peaks at mid-encystation (36 hpi), and drops during late encystation (40 + hpi). EiDrp4 expression peaks in late encystation, at 44 hpi, and drops once again in the mature cyst (72 hpi). The RNA-Seq and subcluster analysis also indicates that the expression of both EiDrp3 and EiDrp4 increases during encystation, as they are members of subclusters 1 and 10, respectively. These genes are therefore likely to be involved in cyst formation, and represent potential therapeutic targets as they are highly divergent from human dynamins.

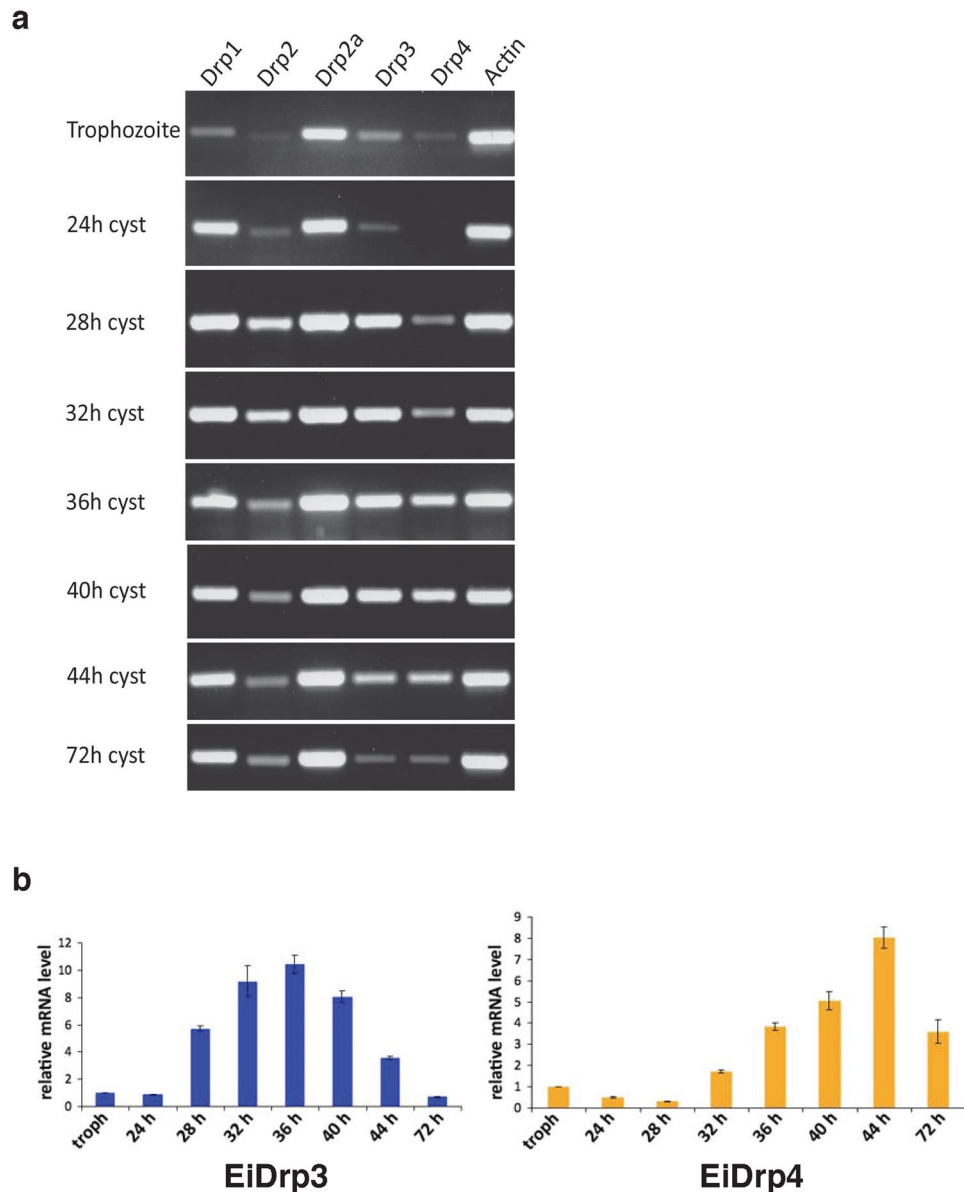


Figure 4. Semi-quantitative expression of all *E. invadens* Drps and quantitative expression of divergent EiDrp3 and EiDrp4. *E. invadens* dynamin expression was assessed during cyst formation using semi-quantitative RT-PCR (a). Comparative analysis indicates that the unusual Drp3 and Drp4 dynamins that are absent in humans are only expressed during cyst formation. Stages shown are trophozoites, early cysts (24–28 hpi), mid cysts (32–40 hpi) and late and mature cysts (44–72 hpi). Actin expression was used as a loading control. Quantitative real-time PCR analysis of unusual *Entamoeba* dynamins during cyst formation (b). Relative quantification of *E. invadens* Drp3 (left) and Drp4 (right) expression in trophozoites versus cysts is shown. Time point of RNA extraction is indicated on the x-axis (early cysts: 24–28 hpi, mid cysts: 32–40 hpi and late and mature cysts: 44–72 hpi) and relative mRNA levels is shown on the y-axis. Standard deviations are based on triplicate experiments.

Divergent Drps change cellular locations during encystation. We have clearly demonstrated that the highly divergent Drp3 and Drp4 are upregulated during encystation. If these proteins play some role in encystation, it is likely that their cellular localization would change during the course of the encystation process. To investigate the possibility, we produced recombinant *E. invadens* cell lines producing hemagglutinin-tagged versions of Drp3 and Drp4 expressed using the pEiNEO-LUC vector developed in the Bhattacharya laboratory⁶⁹. This vector has been demonstrated to continuously express inserted genes under control of the promoter of ribosomal protein gene L3 during stage conversion from trophozoite to cyst and back. Due to the asynchronous nature of *E. invadens* encystation⁷⁰, we observed several different localization patterns variably across the time points (Supplementary Figure S9); however, the pattern representing the majority of cells at each time point is shown in Fig. 5. The localization of both Drp3 and Drp4 in the trophozoite seems punctate and diffuse, and Drp3 is localized to fewer but larger structures than Drp4 (Fig. 5). For Drp3, the localization stays similar for the first

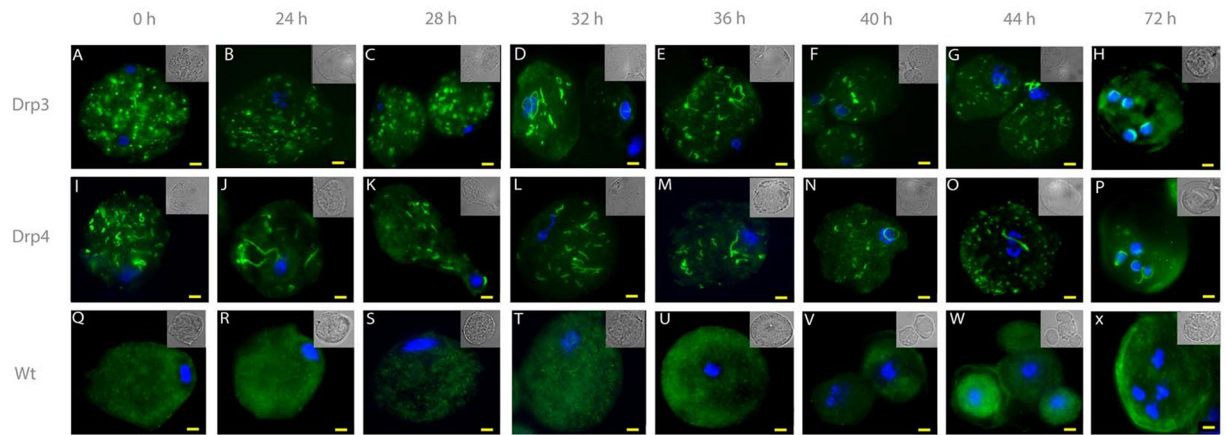


Figure 5. Localisation of divergent Drps during encystation of *Entamoeba invadens*. Hemagglutinin-tagged *E. invadens* Drp3 and Drp4 were stably overexpressed using p*Ei*NEO-LUC⁶⁹ under the control of the ribosomal protein gene L3 promoter. The HA-tag was visualised using anti-HA antibodies (ThermoFisher) as primary antibodies and secondary antibodies conjugated with Alexa Fluor 488 (Abcam). Shown are representative images of the most frequently observed phenotype at each time point, based on data in Supplementary Figure S9. Trophozoites and encysting cells were collected at indicated time points. Wildtype *E. invadens* IP-1 were treated the same to indicate absence of cross-reactivity. Nuclei are stained using DAPI and a bright field image is shown in the top right of each image. Bar equals 3 μ m.

28 hours after which the number of punctate structures reduces and more elongated structures appear (Fig. 5 and Supplementary Figure S9). The final cyst localization seems close to the nucleus suggesting a possible association with the ER. Drp4 on the other hand seems more or less similar throughout encystation but also localizes around the nucleus in the final cyst stage (Fig. 5 and Supplementary Figure S9).

Discussion

E. invadens is a well-established model system in which to study cyst formation of the related human pathogen *E. histolytica*. As cyst wall formation by secretion of various lectins and chitin is critical to pathogenicity in *Entamoeba*, we have focused on the role of the membrane trafficking system (MTS) in this process.

The MTS gene complement of *E. invadens* and *E. histolytica* is highly similar. As these species are not immediate sister taxa⁷¹, these data might be used as a proxy for studying other *Entamoeba* species. Vesicle formation machinery presence is well-conserved relative to that of the model systems of human and yeast. Only the ESCRT family of proteins is relatively incomplete, although we cannot exclude the possibility that the potentially missing genes are present but highly divergent, and refractory to identification by BLAST searches alone. However, even in their absence, it is likely that the ESCRT complex is capable of generating multivesicular bodies, especially given that many ESCRT components show patterns of increased expression in the mature cyst. MVBs are sorting organelles; usually their contents are destined for lysosomal degradation, but can also be secreted via an unconventional pathway involving autophagy machinery⁷².

Overall, vesicle fusion machinery is less well-conserved. None of the SNARE proteins involved in retrograde trafficking from the Golgi to the ER could be found (Syntaxin 18, Sec. 20, Use1, Sec. 22)⁴⁰. Additionally, the COG complex, which is involved in intra-Golgi trafficking⁴³, appears to be heavily reduced (only two subunits out of eight could be positively identified). Dsl1 appears to be missing, but it is typically partially or completely lost in organisms that lack peroxisomes⁷³. In these cases where very few or no members of a complex could be identified, false negatives due to divergence are much less likely. *Entamoeba* does not have a ‘stacked’ Golgi body, instead, large Golgi ‘vesicles’ surround the nucleus, as visualized by confocal microscopy of amoebae stained with anti-Arf antibodies⁷⁴. One may speculate that these potential losses of intra-Golgi and Golgi-ER trafficking machinery are correlated with *Entamoeba*’s alternative Golgi organelle structure. However, the COPI coat complex, which also supports retrograde Golgi-ER trafficking is nearly complete, and almost all subunits have undergone gene duplication events in both *Entamoeba* species. This suggests that the typical Golgi-to-ER trafficking pathway is functional and may even have two functionally distinct COPI coat complexes. Additionally, two paralogues of the SM protein that functions in this step are present in both *Entamoeba* species, lending further support for this trafficking pathway. These results suggest that Golgi-ER trafficking in *Entamoeba* differs from what is required in mammalian and yeast cells^{75,76}, in that it is apparently accomplished without the SNAREs and MTCs that tether incoming vesicles to target membranes in opisthokont systems.

The MTC complexes of the endocytic system are similarly reduced. GARP, which supports endosome-TGN trafficking⁴⁹, appears to be missing one subunit (Vps52), and while the core TRAPP and VpsC core subunits are largely conserved, the subunits that are specific to TRAPP II (another early endosome-TGN trafficking tether)⁷⁷, and both Corvet and HOPS (early and late endosome/lysosome trafficking, respectively)⁷⁸ could not be found. Our results corroborate those shown in Klinger *et al.* 2013, in which neither Dsl1 nor Corvet were identified⁷³. However, other MTS proteins that support these functions are present, such as the retromer complex and adaptor

protein complexes AP1 and AP3. Again, these pathways are likely functional, but perhaps do not require the canonical MTCs for vesicle fusion. In both this case and that of the Golgi-to-ER trafficking pathway above, it is possible that analogous machinery has evolved to take the place of the more canonical components. This would make for attractive therapeutic targets, but can only be addressed by molecular or biochemical means. Our identification of machinery (e.g. the VpsC or COPI complexes) that could serve as bait for interrogation to identify protein interactors should stimulate these types of analyses in the future.

Between *E. histolytica* and *E. invadens*, the presence and number of paralogues of MTS proteins is relatively similar, with the exception of Arf and Rab GTPases, and their regulators. While there has been significant innovation in both species, we do not suspect that this is related to encystation, as that is a feature of both organisms. Ras family GTPases commonly vary in number expanded in closely related taxa⁷⁹, and in this case, may instead be an adaptation to their respective hosts, or other lifestyle factors. As shown in Supplementary Table S3, one-to-one orthology was observed in at least 50% of the Rab GTPases, and Arf and Rab regulators, suggesting a balance of functional retention and innovation. Given that the small GTPases are variably expanded, it is consistent that the number of cognate GAPs and GEFs similarly fluctuates. However, there are six *E. invadens*-specific Rab genes that are members of subclusters 1, 2, 5, and 9 – which are up-regulated during encystation – suggesting they may be involved in this process. Therefore, we cannot overlook the possibility that there may be some differences in the trafficking events underlying encystation in the two *Entamoeba* species. There are also a number of *E. histolytica*-specific Rabs, and again, these may be relevant to encystation as it occurs in their respective hosts.

The two pathways that chiefly characterize encystation are the secretory pathway, and TGN-endosome recycling. The purpose of the former is clear: to support Jacob and Jessie lectin secretion and therefore cyst formation. This correlates with the increase in expression of chitin-cleaving chitinases during encystation (Supplementary Figure S7)⁵⁴. The role of the TGN-endosome recycling pathway in encystation is less clear. During encystation, vesicles containing cyst-forming material are transported to the cell surface, and the extra membrane from this exocytic process must be retrieved. It may be that endosome-to-TGN trafficking functions to retrieve surface proteins that have been internalized with this membrane, thus preventing them from being trafficked to the lysosome for degradation.

It is interesting to consider that many of the genes that are differentially expressed during encystation are small monomeric G proteins and their regulators, i.e. Arf and Rabs. This suggests that these proteins that control vesicle formation and fusion kinetics may be the ‘gatekeepers’ of cyst formation. Our work and others’ has shown that Rabs are particularly differentially expressed during cyst formation²². Unfortunately, little work has been done to investigate the function of Rabs in *Entamoeba*, other than that of Rabs 5²⁰, 7²¹, 8⁸⁰, 11^{81–83}, RabA⁵⁷, and RabB⁶⁰. Therefore, these unstudied Rabs represent potential therapeutic targets, in particular, those proteins that are highly expressed during encystation.

It has been previously shown that *E. histolytica* Rab11A is involved in encystation⁸¹, and Rab11 paralogues are also involved in protein recycling, as are their human orthologues^{81–83}. In *E. invadens*, Rab11A is highly but consistently expressed at time points during encystation, while Rab11B is a member of subcluster 6, and appears to decrease slightly in expression during encystation. By contrast, Rab11C and Rab11D expression increases in early encystation, peaks at late encystation, and drops again in the mature cyst (subcluster 2), implicating Rab11C and Rab11D in cyst formation. Much of the work on *Entamoeba* Rabs has examined their involvement in phagocytosis and pathogenesis. Rab5 and Rab7, particularly, have been shown to be involved in phagosome and pre-phagosomal vacuole maintenance^{20,21,84,85}. Three Rab7 paralogues are members of subclusters 1 and 2, suggesting that they may be involved in encystation, in contrast to other paralogues that have different expression patterns. This again lends support to the idea of a highly specialized MTS in *Entamoeba* where gene families encode functionally divergent paralogues.

The only clear pattern of downregulation during cyst formation is seen in some endocytic and phagocytic trafficking components, which is congruent with the idea that the amoeba is preparing to become a quiescent cyst. We propose that part of this preparation involves generating mRNAs for various endocytic and degradative trafficking genes that may be kept untranslated for rapid translation during excystation, similar to the mRNA localization that occurs in *Drosophila* embryogenesis⁸⁶. Proteomics analyses of *E. histolytica* cysts have shown a handful of MTS proteins, among many others, present in mature cysts^{63,64}, raising the possibility that these mRNAs may indeed be translated. Two proteins identified were clathrin heavy chain and a TBC-B homologue (EHI_094140). Our analysis shows that the *E. invadens* orthologues of both genes are members of subcluster 10, whose expression peaks in the mature cyst. However, at this time, we cannot differentiate between long-lived proteins generated during encystation and untranslated mRNA pools kept in the cyst for excystation. Regardless, this raises the question of a mechanism to prevent mRNA translation, degradation, or protein functioning in the quiescent cyst. Cysts can remain infective for weeks under optimal conditions⁸⁷, suggesting that there is a mechanism by which mRNA (or protein) is stored in an untranslated or inactive manner during this time. Interestingly, there is evidence for the accumulation of mRNA in cysts of the amoeba *Acanthamoeba*⁸⁸, and the ciliate *Colpoda inflata*⁸⁹, suggesting that mRNA storage in cysts occurs in other organisms, although the mechanisms have not yet been deduced.

A similar study to ours was performed by Ehrenkaufner *et al.*⁹⁰, which assessed genome-wide transcriptional changes in *E. invadens* using RNA-Seq. They generated gene expression profiles, and identified Gene Ontology (GO) terms enriched in each profile or profile group. Genes annotated with the GO term ‘vesicle-mediated transport’ were found to be significantly enriched in profiles with a pattern of increased gene expression during encystation. We found that overall, the membrane trafficking gene expression patterns we assigned are largely consistent with the profile assignments by Ehrenkaufner *et al.*⁹⁰. Furthermore, consistent with our findings, they identified genes involved in RNA metabolism expressed in the mature cyst, potentially preconfiguring the cell for rapid excystation.

We identified four dynamin related proteins in *E. histolytica* and five in *E. invadens*. *Entamoeba* Drp1 and Drp2(a) both resemble canonical dynamin related proteins, while Drp3 and Drp4 do not. Immunofluorescence studies have previously shown EhDrp1 to bind to the nuclear membrane⁹¹, while the function of Drp2/Drp2a has not been studied. Although the phylogeny is not internally well-resolved, these proteins do group with other canonical dynamins, including metazoan dynamins. Drp3 and Drp4, however, form a well-supported clade with other amoebozoan, *Naegleria*, plant, and stramenopile and alveolate sequences, to the exclusion of any metazoan sequence. These homologues have been reported to work in cytokinesis in *Dictyostelium* and *Arabidopsis* and suspected to work in chloroplast division in both plant and Stramenopiles, based on phylogenetic affinity⁶⁸. However, the *Entamoeba* Drp3 and Drp4 proteins are solely expressed during encystation, during which time two rounds of nuclear replication occur, but cytokinesis does not^{92,93}. It is likely that the original function of this Drp subfamily is cytokinesis, based on this function being retained in two evolutionarily distant organisms, suggesting *Entamoeba* Drp3 and Drp4 have been repurposed in encystation, although their function is not known. It is worth noting that the RNASeq expression patterns were confirmed by RT-PCR, qPCR and Western blotting, serving not only to give confidence about the dynamic expression, but also as a validation of the RNASeq data generally.

The cellular location of these divergent Drps changed dramatically during encystation from a punctate diffuse localization to one surrounding the nucleus in the mature cyst suggesting an association with the ER. Further co-localization and knock-down experiments will allow us to understand the role of these unusual Drps more fully. The fact that these Drps do not have human orthologues raises the possibility of their use as therapeutic targets. In addition to these, we identified a SNARE protein, Syp7, which is specifically up-regulated during encystation, and lacks a human orthologue⁹⁴. While disrupting them would not treat an infection, it could reduce parasite spread in a population. Currently, luminal amoebicides, which target organisms in the intestinal lumen rather than invasive tissue infection, include diloxanide furoate, iodoquinol, and paromomycin. Diloxanide furoate has an efficacy >50%, while the other two have an efficacy rate >80%, and are relatively safe^{95,96}. However, some asymptomatic cases still fail to respond to treatment with these drugs, and it has been previously suggested that high rates of genetic polymorphisms in a limited geographical region, and extensive geographic diversity in general, may be responsible^{6,97}. Additionally, drug resistance is an ever-growing problem, and *E. histolytica* resistance to treatment with metronidazole, which is effective against invasive *Entamoeba*, has been reported⁹⁸. The development of new amoebicidal drugs is therefore still a valuable pursuit, especially drugs that limit cyst production, which is key in areas with poor sanitation.

In conclusion, we have shown, using the model *E. invadens*, that the membrane trafficking system is modulated during the encystation process. The normal secretory pathway supports exocytosis of cyst wall factors, and TGN-endosome recycling may be a mechanism to control membrane homeostasis. Furthermore, early endocytic factors such as clathrin, as well as phagocytic genes, appear to be upregulated in the mature cyst, suggesting that mRNAs or perhaps protein is stored in the cyst in preparation for excystation. Finally, *Entamoeba* expresses two highly divergent dynamin-related proteins specifically during encystation. Orthologues in other taxa are involved in cytokinesis, suggesting that the *Entamoeba* proteins have undergone neofunctionalization. These represent potential drug targets to be explored in future, as an additional way to lower diarrheal disease burden around the world.

Methods

Strains, Culture Conditions and Cyst Formation. Trophozoites of *E. invadens* IP-1 were cultured in 50 mL culture flasks with unvented lids in axenic LYI-S-2 medium at 20 °C⁹⁹. To induce encystation, trophozoites in logarithmic growth phase were transferred into axenic glucose-free 50% LY-G medium¹⁰⁰. For RNA-Seq experiments, samples were taken at 0, 24, 48, and 72 hours post-induction (hpi) of cyst formation; to examine the expression of dynamin-related proteins in cyst formation, samples were taken 0, 24, 28, 32, 36, 40, 44 and 72 hpi in accordance with previous studies^{13,54,70,101}. The 72 hpi mature cyst sample was treated with 0.05% Sarkosyl (Sigma Aldrich) to destroy any remaining trophozoites. Cells were pelleted and washed in phosphate-buffered saline for subsequent RNA and protein extraction. *Escherichia coli* strain DH5 α was maintained in LB containing 100 μ g/ml ampicillin when used to support vector maintenance.

Amplification and Cloning of EiDrp3 and EiDrp4. Genes encoding EiDrp3 (GenBank accession number EIN_376410) and EiDrp4 (EIN_080030) were PCR amplified from genomic DNA of *E. invadens* IP-1 using primers adding hemagglutinin (HA) tags at the 3' end, for EiDrp3: 5'-ATG TCT ATC TTG AAT GAA CCA G-3' and 5'-CTA TGC ATA GTC TGG AAC GTC ATA TGG ATA CAT TTC TCG CTC TTT GG-3'; and for EiDrp4: 5'-ATG ACA TCA ACA CAG ACA ATG CG-3' and 5'-CTA TGC ATA GTC TGG AAC GTC ATA TGG ATA AAA GTT CAA TTT CTT TTT AAG TTG AGG-3'. The cycling parameters of PCR were: denaturation at 95 °C for 45 s; annealing at 50–56 °C for 45 s; elongation at 72 °C for 1 min/kb; 35 cycles, using GoTaq G2 Hot Start Green Master Mix (Promega). PCR products were purified with QIAquick PCR purification kit (Qiagen) and cloned into pGEM-T-Easy and confirmed by sequencing. The genes were subsequently cloned into the *E. invadens* expression vector pEiNEO-LUC¹⁰², a very kind gift of Professor Sudha Bhattacharya's laboratory (Jawaharlal Nehru University, New Delhi, India). To facilitate cloning, *Kpn*I and *Bam*HI sites were added via PCR: for EiDrp3: 5'-AGA AGA GGT ACC ATG TCT ATC TTG AAT GAA CCA G-3' and 5'-TCT TCT GGA TCC CTA CAT TTC TCG CTC TTT GG-3'; for EiDrp4: 5'-AGA AGA GGT ACC ATG ACA TCA ACA CAG ACA ATG CG-3' and 5'-TCT TCT GGA TCC CTA AAA GTT CAA TTT CTT TTT AAG TTG AGG-3'. Constructs were confirmed by sequencing.

Transfection of *E. invadens*. Transfection was performed by electroporation according to Singh *et al.* 2012¹⁰². Briefly, *E. invadens* cells were grown up to 70% confluency, placed on ice for 15 min and harvested by

centrifugation at 400 g for 5 min at 4 °C. Cells were washed once with PBS pH 7.0 and once washed with incomplete cytomix buffer (120 mM KCl, 0.15 mM CaCl₂, 10 mM K₂HPO₄/KH₂PO₄ pH 7.5, 25 mM HEPES, 2 mM EGTA, 5 mM MgCl₂, final pH 7.8). Subsequently, 2 × 10⁶ trophozoites were re-suspended in 600 µl of complete cytomix buffer (5 mM reduced glutathione and 2 mM ATP). Cells were transferred to a 0.4 mm electroporation cuvette. 100 µg of plasmid DNA was added. Two consecutive pulses of 3000 V/cm were applied at capacitance of 25 µF using a Bio-Rad Gene Pulser Xcell Electroporation System. Electroporated cells were incubated on ice for 10 min, transferred to borosilicate tubes containing 12 ml of LYI-S medium supplemented with 15% heat inactivated adult bovine serum and 2% vitamin mix. 10 µg of G418 was added after 24 hrs. Cells were subcultured with G418 containing media every 72 hrs.

Fluorescence microscopy. Immunofluorescence was performed according to Siegesmund *et al.*¹⁰³. Trophozoites and cysts were harvested at 400 g/5 min, attached to poly-L-lysine coated glass slides for 20 min, fixed in 4% paraformaldehyde (PFA) for 45 min, and rehydrated with 1x PBS for 30 min. Trophozoites were permeabilized with 0.2% Triton X-100 in 1x PBS at room temperature for 1 hour. Cysts were permeabilized overnight at 4 °C. Cells were blocked with 2% BSA at room temperature for 2 hours and incubated with anti-HA primary antibodies (ThermoFisher) diluted to 1:500 in 0.2% Triton X-100, 0.02% sodium azide and 2% BSA for 1 hour at room temperature. Slides were washed twice with 0.2% Triton X-100 in 1x PBS for 3–5 min. and incubated with secondary antibodies conjugated with Alexa Fluor 488 (Abcam) diluted to 1:500 in 0.2% Triton X-100, 0.02% sodium azide, 1% BSA in 1x PBS, for 1 hour at room temperature. Slides were washed with 0.2% Triton X-100, 0.5% BSA in 1x PBS, 3 times for 5 min. each. Slides were mounted with Fluoroshield Mounting Medium with DAPI (Abcam) and covered immediately. Cells were examined using an Olympus IX-81 inverted microscope equipped with the appropriate filter combinations and a 100x objective. Cyst formation was monitored by Calcofluor White staining; two drops of 0.05% Calcofluor White solution (Fluka) were added to a microscope slide with 40 µL encysting cell suspension and incubated for 5 minutes at room temperature before applying the coverslip. Cysts were observed under ultraviolet light by fluorescent microscopy. Spores from the ascomycete fungus *Magnaporthe grisea* were used as a positive control. For a negative control, *E. invadens* mature cysts were not treated with Calcofluor White (Supplementary Figure S10).

RNA extraction for RT-PCR and qPCR. Trophozoites and the various cyst stages were harvested and washed in PBS. RNA was isolated using Tripure (Roche). Total RNA isolated was treated with DNaseI (Ambion) for 1 hour at 37 °C to remove any genomic DNA. To test for successful DNA removal, RNA samples were subjected to PCR using gene specific primers. Purified RNA was quantified spectrophotometrically using a Nanodrop ND-1000 (Thermo-Fisher). First strand cDNA synthesis was started from 2 µg RNA using Promega ML-MLRT reverse transcriptase in combination with random primers.

Transcription analysis during cyst formation. PCR was performed for 30 cycles at 95 °C for 45 s, 53 °C for 45 s, and 72 °C for 45 s using the GoTaq Master Mix (Promega). Prior to PCR, the annealing temperature for the primer pairs was determined experimentally by gradient PCR using a mix of cDNA samples from each time point. The chitinase primers and corresponding annealing temperatures were taken from a previous study⁵⁴. All other primers used in this study are listed in Supplementary Table S4. Quantitative real-time PCR experiments were carried out using a Stratagene Mx3005P real-time PCR system, Brilliant II Sybr Green qPCR Master Mix (Agilent) and cDNA prepared as described above. PCR was performed for 40 cycles at 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s. Primer specificities and PCR amplification efficiencies were analysed by melt curve and standard curve experiments. The melt curve was established for each primer pair by increasing temperatures at the end of the PCR reaction from 55 to 94 °C, and monitoring fluorescence. The standard curve for each primer pair was constructed by using serial diluted cDNA templates for PCR. qPCR experiments were performed in technical triplicates. The reference gene of choice was the SKI-interacting protein SKIP (Accession number: AANW01003268), which has been shown to express stably during cyst formation in *E. invadens*¹⁰⁴.

RNA-seq analysis. cDNA libraries were constructed for each time point (0, 24, 48, 72 hpi) using the Illumina TruSeq RNA Library Preparation kit and run on an Illumina HiSeq. 2000 sequencer, generating 100 bp paired-end sequences. Raw reads were trimmed, clipped and filtered using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The first 12 bp were trimmed of the reads using the fastx_trimmer, and adapter sequences were removed using the fastx_clipper tool (using 'AGATCGGAAGAGC' as adapter string). The 3' end of the reads were quality-trimmed using a Perl script written by Joe Fass (Bioinformatics Core Facility at the UC Davis Genome Center, USA) (Trim.slidingWindow.pl; see Supplementary File 1), that uses a sliding window to trim the first base with a quality Phred score <20, in the first sliding window with a quality Phred score <20. Reads less than 30 bp were removed and processed sequences were paired using a Python script (https://github.com/lexnederbragt/denovo-assembly-tutorial/blob/master/scripts/pair_up_reads.py). Transcripts were assembled *de novo* using CLC Genomics Workbench software package (CLC, Denmark) clustered using CD-HIT-EST¹⁰⁵ and subsequently used to correct and refine the *E. invadens* IP-1 genomic sequence using a hybrid genomic guided and *de novo* approach to identify potential new transcripts (PASA)¹⁰⁶. All raw sequence data are accessible at the NCBI Sequence Read Archive through accession numbers MF361146 - MF361261.

The assembled transcripts were preliminary annotated against the *E. invadens* IP-1 (JCVI) and *E. histolytica* HM-1:IMSS (JCVI) annotated transcripts and proteomes (publically available at amoebaDB.org). All-against-all nucleotide BLASTN and BLASTX (translating a nucleotide query in six reading frames) searches were conducted¹⁰⁷. Annotations were assigned to the RNA-Seq transcripts when hits had an E- value of at least 1e-3 and coverage of at least 50%. When no adequate hit was available the transcript was annotated as hypothetical. A second genome sequence of *E. invadens* IP-1 independently constructed by The Sanger Institute (NCBI GenBank,

accession number AANW00000000) was also used to annotate the RNA-Seq transcripts. Independent BLASTN was conducted for each RNA-Seq transcript against the NCBI Nucleotide BLAST database using BLAST +¹⁰⁸. Annotations were accepted for BLASTN hits with an E-value < 1e-2 with a minimum coverage of 50%.

Gene expression analysis. The Trinity package^{109,110} and edgeR⁵³ were used to determine the levels of gene expression using the mapped transcripts. Within the Trinity package, RSEM was used to calculate a Maximum Likelihood estimation of abundance based on the number of reads mapped at both an isoform level and gene level, given as fragments per kilobase of exon per million fragments mapped (FPKM)¹¹⁰. Differentially expressed genes were identified using edgeR, with a dispersion value of 0.4 (recommended for datasets with no biological replicates)⁵³. Transcripts with a false discovery rate < 0.01 and a minimum log₂ fold change of 2-fold were classified as significantly differently expressed between two time points. TMM-normalized FPKM values for each gene across the four time points (0 hpi, 24 hpi, 48 hpi and 72 hpi) were calculated with default parameters¹¹¹ producing expression profiles for each gene. As the Expectation Maximization algorithm of RSEM assigns reads to isoforms based on the most likely relative abundances of the transcripts, in some cases alternative transcripts were identified as very lowly expressed at only one time point, and these appeared to not be biologically significant. Therefore, only transcripts with the highest average expression across conditions were considered in our analyses.

Significantly differentially expressed genes (FDR < 0.01, fold change > 2) were clustered in order to identify groups of similarly regulated genes using *k*-means clustering (Supplementary Table S5). The optimum number of clusters was determined using the median number of clusters proposed by 18 independent indexes (R package NbClust)¹¹². Hierarchical clustering of differentially expressed transcripts with similar expression patterns using Euclidean distance was conducted using scripts from the Trinity package (clustering using 'analyze_diff_expr.pl' with the parameters -P 1e-2 and -C 1 and clusters split using 'define_clusters_by_cutting_tree.pl', parameters -K 10). To determine which variables (time points) within the data set were contributing the largest proportion of variation between the clusters, defined using the above techniques, Principal Component Analysis (PCA) was carried out. Scaled FPKM values for each gene at each of the four time points were used for the PCA analysis and a biplot was created using custom R script (Supplementary File 2) (R Core Team, 2013).

Comparative genomics. For all membrane trafficking components, *Homo sapiens* sequences that have been functionally characterized were used as queries to search the individual transcript files for each of the four time points, as well as the combined set of *E. invadens* transcripts. Additionally, searches were repeated in the predicted proteome of the *E. invadens* genome at AmoebaDB (<http://amoebadb.org/amoeba/>). In cases where an MTS component has not been described in human, the sequence from the organism in which it is best described or evolutionarily is closest to *E. invadens* was chosen; these include *Arabidopsis thaliana*, *Naegleria gruberi*, *Saccharomyces cerevisiae*, and *Trypanosoma brucei*. tBLASTn searches were performed with a BLOSUM62 matrix and an E-value cut-off of 0.05 to identify potential homologues, which were then confirmed by reverse BLASTx searches into the *H. sapiens* genome. To be considered homologous, *E. invadens* sequences must retrieve the original human query sequence or a clear orthologue with an E-value < 0.05. As *E. invadens* sequences tend to be divergent, those that retrieved a clear orthologue of the human query by searching the NCBI non-redundant database with an acceptable E-value were also considered orthologous. Rab proteins were not searched for using BLAST, as a comparative genomic analysis of Rabs in *E. invadens* and *E. histolytica* was previously done by Nakada-Tsukui *et al.* 2010²³.

Dynamins were identified by searching the *E. histolytica* genome using previously classified dynamin sequences from across the eukaryotic tree of life. The four *E. histolytica* dynamins, EhDrp1 (EHI_013180), EhDrp2 (EHI_052740), EhDrp3 (EHI_174650), and EhDRP4 (EHI_139440), were then used to search all available *Entamoeba* genomes on Sanger (<http://www.sanger.ac.uk/>) and AmoebaDB (<http://amoebadb.org/amoeba/>).

Phylogenetics. Several MTS gene families are highly paralogous, and phylogenetic trees were generated to confirm the identity of homologues found via BLAST. Sequences were aligned using MUSCLE v3.8.31¹¹³, visualized using Mesquite v3.03¹¹⁴, and then manually masked and trimmed to include only homologous positions. Masked alignments are available upon request. A list of sequence accessions, corresponding gene identifiers, and tree meta-data are listed in Supplementary Table S6. ProtTest v3.4¹¹⁵ was used to determine the best-fit model of sequence evolution, which was LG¹¹⁶ with gamma-distributed rate variation across sites (+G) and observed amino acid frequencies (+F) in the following trees: adaptin GADEZ subfamily, ESCRT Snf7 family, the TBC/RabGAP family, and the DENN RabGEF family. Both the Rab and Arf trees to assess *Entamoeba* spp. sequence orthology also used this model. The LG + G model was used in ArfGEF and SNARE Qc subfamily trees, and the WAG + G + F model¹¹⁷ was used in ArfGAP trees. Phylobayes v3.3^{118–120} and MrBAYES v3.2.2^{121,122} were used for Bayesian inference of phylogeny, and RAXML v8.1.3¹²³ was used for Maximum-Likelihood analyses. Phylobayes was run until the largest discrepancy observed across all bipartitions was less than 0.1 and at least 100 sampling points were achieved, MrBAYES was used to search treespace for a minimum of one million MCMC generations, sampling every 1000 generations, until the average standard deviation of the split frequencies of two independent runs (with two chains each) was less than 0.01. Consensus trees were generated using a burn-in value of 25%, well above the likelihood plateau in each case. RAXML was run with 100 pseudoreplicates. MTS gene phylogenies can be found in Supplementary Figures S1 and S2.

The dynamin sequences from *E. invadens*, *Entamoeba dispar* and *Entamoeba moshkovskii* were extracted from the respective genome projects and aligned with homologous protein sequences found in Archaeplastida, stramenopiles and *N. gruberi* using ClustalW in Seaview 4.2.12¹²⁴. Manual masking and trimming was performed in Seaview. Three datasets were constructed including a global dynamin-family phylogeny, a separate Drp1/2 phylogeny and a Drp3/4 phylogeny. The large dataset contained 133 sequences with 229 informative sites. The Drp1 and Drp2 dataset contained 66 sequences with 230 informative sites and the Drp3 and Drp4 dataset contained 29 sequences with 572 informative residues. A complete list of taxa and accession numbers can be found in

Supplementary Table S7. RAxML¹²³ and the Maximum Likelihood program PhyML v3.0¹²⁵ were used to generate trees for all datasets. MrBAYES^{121,122} was run for the Drp3/Drp4 dataset. Modelgenerator v.0.85¹²⁶ was used to determine the best-fit model of sequence evolution, which was the LG model¹¹⁶ for all dynamin trees. For individual phylogenies, the following additional parameters were used: global dynamin phylogeny, LG + G; the Drp1/Drp2 phylogeny, LG + I + G (estimated proportion of invariant sites, + I); and Drp3/Drp4, LG + G. Bootstrap resampling was performed on ML trees with 100 replications for all three analyses. In addition, the Bayesian analysis for the Drp3/Drp4 dataset was performed using a mixed model accommodating 4 rate + inv categories containing 4 chains each. One million generations were calculated and trees sampled every 1,000 generations. The model stabilized rapidly and 200 trees were discarded as burn-in.

References

1. WHO, PAHO & UNESCO. Report: a consultation with experts on amoebiasis (1997).
2. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095–2128 (2012).
3. Verkerke, H. P., Petri, W. A. & Marie, C. S. The dynamic interdependence of amebiasis, innate immunity, and undernutrition. *Semin. Immunopathol.* **34**, 771–785 (2012).
4. Ximénez, C., Morán, P., Rojas, L., Valadez, A. & Gómez, A. Reassessment of the epidemiology of amebiasis: State of the art. *Infect. Genet. Evol.* **9**, 1023–1032 (2009).
5. Stanley, S. L. Amoebiasis. *Lancet* **361**, 1025–1034 (2003).
6. Ali, V. & Nozaki, T. Current therapeutics, their problems, and sulfur-containing-amino-acid metabolism as a novel target against infections by ‘mitochondriate’ protozoan parasites. *Clin. Microbiol. Rev.* **20**, 164–187 (2007).
7. Ishida, H. *et al.* Fulminant amoebic colitis with perforation successfully treated by staged surgery: A case report. *J. Gastroenterol.* **38**, 92–96 (2003).
8. Gupta, S. S., Singh, O., Shukla, S. & Raj, M. K. Acute fulminant necrotizing amoebic colitis: a rare and fatal complication of amoebiasis: a case report. *Cases J.* **2**, 6557 (2009).
9. Kenneth, J. R. & Ray, C. G. *Sherrie Medical Microbiology*. Vasa <https://doi.org/10.1036/0838585299> (2004).
10. Nozaki, T. & Bhattacharya, A. *Amebiasis*. <https://doi.org/10.1007/978-4-431-55200-0> (2015).
11. Mondal, D., Petri, W. A., Sack, R. B., Kirkpatrick, B. D. & Haque, R. Entamoeba histolytica-associated diarrheal illness is negatively associated with the growth of preschool children: evidence from a prospective study. *Trans. R. Soc. Trop. Med. Hyg.* **100**, 1032–1038 (2006).
12. Mi-ichi, F., Yoshida, H. & Hamano, S. Entamoeba Encystation: New Targets to Prevent the Transmission of Amebiasis. *PLOS Pathog.* **12**, e1005845 (2016).
13. Chatterjee, A. *et al.* Evidence for a ‘wattle and daub’ model of the cyst wall of Entamoeba. *PLoS Pathog.* **5** (2009).
14. Frisardi, M. *et al.* The Most Abundant Glycoprotein of Amebic Cyst Walls (Jacob) Is a Lectin with Five Cys-Rich, Chitin-Binding Domains. *Infect. Immun.* **68**, 4217–4224 (2000).
15. Das, S. *et al.* The cyst wall of Entamoeba invadens contains chitosan (deacetylated chitin). *Mol. Biochem. Parasitol.* **148**, 86–92 (2006).
16. Vega, H. D. *et al.* Cloning and expression of chitinases of Entamoeba. *Mol. Biochem. Parasitol.* **85**, 139–147 (1997).
17. Jimenez-Ruiz, E., Morlon-Guyot, J., Daher, W. & Meissner, M. Vacuolar protein sorting mechanisms in apicomplexan parasites. *Mol. Biochem. Parasitol.* 1–8, <https://doi.org/10.1016/j.molbiopara.2016.01.007> (2015).
18. Manna, P. T., Boehm, C., Leung, K. F., Natesan, S. K. & Field, M. C. Life and times: Synthesis, trafficking, and evolution of VSG. *Trends Parasitol.* **30**, 251–258 (2014).
19. Nozaki, T. & Nakada-Tsukui, K. Membrane trafficking as a virulence mechanism of the enteric protozoan parasite Entamoeba histolytica. *Parasitol. Res.* **98**, 179–183 (2006).
20. Saito-Nakano, Y., Yasuda, T., Nakada-Tsukui, K., Leippe, M. & Nozaki, T. Rab5-associated vacuoles play a unique role in phagocytosis of the enteric protozoan parasite Entamoeba histolytica. *J. Biol. Chem.* <https://doi.org/10.1074/jbc.M403987200> (2004).
21. Nakada-Tsukui, K., Saito-Nakano, Y., Ali, V. & Nozaki, T. A Retromerlike Complex Is a Novel Rab7 Effector That Is Involved in the Transport of the Virulence Factor Cysteine Protease in the Enteric Protozoan Parasite Entamoeba histolytica. *Mol. Biol. Cell* **16**, 5294–5303 (2005).
22. De Cádiz, A. E., Jeelani, G., Nakada-Tsukui, K., Caler, E. & Nozaki, T. Transcriptome analysis of encystation in Entamoeba invadens. *PLoS One* **8**, e74840 (2013).
23. Nakada-Tsukui, K., Saito-Nakano, Y., Husain, A. & Nozaki, T. Conservation and function of Rab small GTPases in Entamoeba: Annotation of E. invadens Rab and its use for the understanding of Entamoeba biology. *Exp. Parasitol.* **126**, 337–347 (2010).
24. Saito-Nakano, Y., Loftus, B. J., Hall, N. & Nozaki, T. The diversity of Rab GTPases in Entamoeba histolytica. *Exp. Parasitol.* **110**, 244–252 (2005).
25. Loftus, B. *et al.* The genome of the protist parasite Entamoeba histolytica. *Nature* **433**, 865–868 (2005).
26. Weber, C. *et al.* Extensive transcriptome analysis correlates the plasticity of Entamoeba histolytica pathogenesis to rapid phenotype changes depending on the environment. *Sci. Rep.* **6**, 35852 (2016).
27. van der Bliek, A. M. & Meyerowitz, E. M. Dynammin-like protein encoded by the Drosophila shibire gene associated with vesicular traffic. *Nature* **351**, 411–414 (1991).
28. Chen, M. S. *et al.* Multiple forms of dynamin are encoded by shibire, a Drosophila gene involved in endocytosis. *Nature* **351**, 583–586 (1991).
29. Ferguson, S. M. & Camilli, P. De. Dynammin, a membrane-remodelling GTPase. *Nat. Publ. Gr.* **13**, 75–88 (2012).
30. Praefcke, G. J. K. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nat. Rev. Mol. Cell Biol.* **5**, 133–47 (2004).
31. Gaechter, V., Schraner, E., Wild, P. & Hehl, A. B. The single dynamin family protein in the primitive protozoan Giardia lamblia is essential for stage conversion and endocytic transport. *Traffic* **9**, 57–71 (2008).
32. Vázquezdelara-Cisneros, L. G. & Arroyo-Begovich, A. Induction of Encystation of Entamoeba invadens by Removal of Glucose from the Culture Medium. *J. Parasitol.* **70**, 629–633 (1984).
33. Hirst, J. *et al.* Characterization of TSET, an ancient and widespread membrane trafficking complex. *Elife* **2014**, 1–18 (2014).
34. Cosson, P. & Letourneur, F. Coatamer interaction with di-lysine endoplasmic reticulum retention motifs. *Science* **263**, 1629–1631 (1994).
35. Seaman, M. N. J. Recycle your receptors with retromer. *Trends Cell Biol.* **15**, 68–75 (2005).
36. Fjorback, A. W. *et al.* Retromer binds the FANSHY sorting motif in SorLA to regulate amyloid precursor protein sorting and processing. *J. Neurosci.* **32**, 1467–80 (2012).
37. Wassmer, T. *et al.* A loss-of-function screen reveals SNX5 and SNX6 as potential components of the mammalian retromer. *J. Cell Sci.* **120**, 45–54 (2006).
38. Hirst, J. & Robinson, M. S. Clathrin and adaptors. *Biochim. Biophys. Acta - Mol. Cell Res.* **1404**, 173–193 (1998).

39. Robinson, M. S., Sahlender, D. A. & Foster, S. D. Rapid Inactivation of Proteins by Rapamycin-Induced Rerouting to Mitochondria. *Dev. Cell* **18**, 324–331 (2010).
40. Jahn, R. & Scheller, R. H. SNAREs — engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**, 631–643 (2006).
41. Suwastika, I. N., Uemura, T., Shiina, T., H Sato, M. & Takeyasu, K. SYP71, a plant-specific Qc-SNARE protein, reveals dual localization to the plasma membrane and the endoplasmic reticulum in Arabidopsis. *Cell Struct. Funct.* **33**, 185–192 (2008).
42. Hong, W. & Lev, S. Tethering the assembly of SNARE complexes. *Trends Cell Biol.* **24**, 35–43 (2014).
43. Ungar, D. *et al.* Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. *J. Cell Biol.* **157**, 405–415 (2002).
44. Walter, D. M., Paul, K. S. & Waters, M. G. Purification and characterization of a novel 13 S hetero-oligomeric protein complex that stimulates *in vitro* Golgi transport. *J. Biol. Chem.* **273**, 29565–29576 (1998).
45. Ishii, M., Suda, Y., Kurokawa, K. & Nakano, A. COPI is essential for Golgi cisternal maturation and dynamics Authors: *J. Cell Sci.* <https://doi.org/10.1242/jeb.108043> (2016).
46. Martínez-Menárguez, J. A. *et al.* Peri-Golgi vesicles contain retrograde but not anterograde proteins consistent with the cisternal progression model of intra-Golgi transport. *J. Cell Biol.* **155**, 1213–1224 (2001).
47. Lanoix, J. *et al.* GTP hydrolysis by arf-1 mediates sorting and concentration of Golgi resident enzymes into functional COP I vesicles. *EMBO J.* **18**, 4935–4948 (1999).
48. Burd, C. G., Peterson, M., Cowles, C. R. & Emr, S. D. A novel Sec. 18p/NSF-dependent complex required for Golgi-to-endosome transport in yeast. *Mol. Biol. Cell* **8**, 1089–1104 (1997).
49. Conibear, E. & Stevens, T. H. Vps52p, Vps53p, and Vps54p form a novel multisubunit complex required for protein sorting at the yeast late Golgi. *Mol. Biol. Cell* **11**, 305–323 (2000).
50. Kloeppe, T. H., Kienle, C. N. & Fasshauer, D. SNAREing the basis of multicellularity: Consequences of protein family expansion during evolution. *Mol. Biol. Evol.* **25**, 2055–2068 (2008).
51. Hata, Y., Slaughter, C. A. & Südhof, T. C. Synaptic vesicle fusion complex contains unc-18 homologue bound to syntaxin. *Nature* **366**, 347–351 (1993).
52. Hsu, S. C. *et al.* The mammalian brain rsc. 6/8 complex. *Neuron* **17**, 1209–1219 (1996).
53. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
54. Makioka, A., Kumagai, M., Hiranuka, K., Kobayashi, S. & Takeuchi, T. Different structure and mRNA expression of Entamoeba invadens chitinases in the encystation and excystation. *Parasitol. Res.* **109**, 417–423 (2011).
55. Nakano, A. & Muramatsu, M. A Novel GTP-binding Protein, Sarlp, Is Involved in Transport from the Endoplasmic Reticulum to the Golgi Apparatus. *J. Cell Biol.* **109**, 2677–2691 (1989).
56. Zahraoui, A., Touchot, N., Chardin, P. & Tavitian, A. The human rab genes encode a family of GTP-binding proteins related to yeast YPT1 and SEC. 4 products involved in secretion. *J. Biol. Chem.* **264**, 12394–12401 (1989).
57. Welter, B. H. & Temesvari, L. A. Overexpression of a mutant form of EhRabA, a unique rab GTPase of entamoeba histolytica, alters endoplasmic reticulum morphology and localization of the Gal/GalNAc adherence lectin. *Eukaryot. Cell* **8**, 1014–1026 (2009).
58. Gotthardt, D. *et al.* Proteomics Fingerprinting of Phagosome Maturation and Evidence for the Role of a G during Uptake. *Mol. Cell. Proteomics* **5**, 2228–2243 (2006).
59. Bosch, D. E. & Siderovski, D. P. G protein signaling in the parasite Entamoeba histolytica. *Exp. Mol. Med.* **45**, e15 (2013).
60. Romero-Diaz, M. *et al.* Structural and functional analysis of the Entamoeba histolytica EhrabB gene promoter. *BMC Mol Biol* **8**, 82 (2007).
61. Hendrick, H. M. *et al.* Phosphorylation of Eukaryotic Initiation Factor-2 α during Stress and Encystation in Entamoeba Species. *PLoS Pathog.* **12**, 1–21 (2016).
62. Shock, J. L., Fischer, K. F. & DeRisi, J. L. Whole-genome analysis of mRNA decay in Plasmodium falciparum reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol.* **8**, R134 (2007).
63. Luna-Nácar, M. *et al.* Proteomic study of Entamoeba histolytica trophozoites, cysts, and cyst-like structures. *PLoS One* **11**, 1–24 (2016).
64. Ali, I. K. M. *et al.* Proteomic analysis of the cyst stage of Entamoeba histolytica. *PLoS Negl. Trop. Dis.* **6** (2012).
65. Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–51 (2006).
66. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–4 (2000).
67. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2010).
68. Miyagishima, S., Kuwayama, H., Urushihara, H. & Nakanishi, H. Evolutionary linkage between eukaryotic cytokinesis and chloroplast division by dynamin proteins. *Proc. Natl. Acad. Sci. USA* **105**, 15202–7 (2008).
69. Singh, N., Ojha, S., Bhattacharya, A. & Bhattacharya, S. Stable transfection and continuous expression of heterologous genes in Entamoeba invadens. *Mol. Biochem. Parasitol.* **184**, 9–12 (2012).
70. Segovia-Gamboa, N. C. *et al.* Entamoeba invadens, encystation process and enolase. *Exp. Parasitol.* **125**, 63–69 (2010).
71. Clark, C. G. *et al.* New insights into the phylogeny of Entamoeba species provided by analysis of four new small-subunit rRNA genes. *Int. J. Syst. Evol. Microbiol.* **56**, 2235–2239 (2006).
72. Duran, J. M., Anjard, C., Stefan, C., Loomis, W. F. & Malhotra, V. Unconventional secretion of Acb1 is mediated by autophagosomes. *J. Cell Biol.* **188**, 527–536 (2010).
73. Klinger, C. M., Klute, M. J. & Dacks, J. B. Comparative genomic analysis of multi-subunit tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in apicomplexa. *PLoS One* **8**, e76278 (2013).
74. Ghosh, S. K. *et al.* Chitinase secretion by encysting Entamoeba invadens and transfected Entamoeba histolytica trophozoites: Localization of secretory vesicles, endoplasmic reticulum, and Golgi apparatus. *Infect. Immun.* **67**, 3073–3081 (1999).
75. VanRheenen, S. M., Reilly, B. A., Chamberlain, S. J. & Waters, M. G. Dsl1p, an essential protein required for membrane traffic at the endoplasmic reticulum/Golgi interface in yeast. *Traffic* **2**, 212–31 (2001).
76. Zolov, S. N. & Lupashin, V. V. Cog3p depletion blocks vesicle-mediated Golgi retrograde trafficking in HeLa cells. *J. Cell Biol.* **168**, 747–759 (2005).
77. Cai, H., Zhang, Y., Pypaert, M., Walker, L. & Ferro-Novick, S. Mutants in trs 120 disrupt traffic from the early endosome to the late Golgi. *J. Cell Biol.* **171**, 823–833 (2005).
78. Balderhaar, H., kleine, J. & Ungermann, C. CORVET and HOPS tethering complexes - coordinators of endosome and lysosome fusion. *J. Cell Sci.* **126**, 1307–16 (2013).
79. Rojas, A. M., Fuentes, G., Rausell, A. & Valencia, A. The Ras protein superfamily: Evolutionary tree and role of conserved amino acids. *J. Cell Biol.* **196**, 189–201 (2012).
80. Hanadate, Y., Saito-Nakano, Y., Nakada-Tsukui, K. & Nozaki, T. Endoplasmic reticulum-resident Rab8A GTPase is involved in phagocytosis in the protozoan parasite Entamoeba histolytica. *Cell. Microbiol.* **18**, 1358–1373 (2016).
81. McCugan, G. C. & Temesvari, L. a. Characterization of a Rab11-like GTPase, EhRab11, of Entamoeba histolytica. *Mol. Biochem. Parasitol.* **129**, 137–146 (2003).
82. Mitra, B. N., Saito-Nakano, Y., Nakada-Tsukui, K., Sato, D. & Nozaki, T. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite Entamoeba histolytica. *Cell. Microbiol.* **9**, 2112–2125 (2007).

83. Temesvari, L. A., Harris, E. N., Stanley, S. L. & Cardelli, J. A. Early and late endosomal compartments of *Entamoeba histolytica* are enriched in cysteine proteases, acid phosphatase and several Ras-related Rab GTPases. *Mol. Biochem. Parasitol.* **103**, 225–241 (1999).
84. Okada, M. & Nozaki, T. New insights into molecular mechanisms of phagocytosis in *Entamoeba histolytica* by proteomic analysis. *Archives of Medical Research*, <https://doi.org/10.1016/j.arcmed.2005.10.003> (2006).
85. Klinger, C. M. *et al.* Resolving the homology–function relationship through comparative genomics of membrane-trafficking machinery and parasite cell biology. *Mol. Biochem. Parasitol.* <https://doi.org/10.1016/j.molbiopara.2016.07.003> (2016).
86. Lasko, P. mRNA localization and translational control in *Drosophila* oogenesis. *Cold Spring Harb. Perspect. Biol.* **4**, 1–15 (2012).
87. Petri, W. A. Jr & Singh, U. Diagnosis and Management of Amebiasis. *Clin. Infect. Dis.* **29**, 1117–1125 (1990).
88. Byers, T. J., Kim, B. G., King, L. E. & Hugo, E. R. Molecular Aspects of the Cell Cycle and Encystment of *Acanthamoeba*. *Rev. Infect. Dis.* **13**, S378–84 (1991).
89. Benitez, L. & Gutiérrez, J. C. Encystment-specific mRNA is accumulated in the resting cysts of the ciliate *Colpoda inflata*. *Biochem. Mol. Biol. Int.* **41**, 1137–1141 (1997).
90. Ehrenkauf, G. M. *et al.* The genome and transcriptome of the enteric parasite *Entamoeba invadens*, a model for encystation. *Genome Biol.* **14**, R77 (2013).
91. Jain, R., Shrimal, S., Bhattacharya, S. & Bhattacharya, A. Identification and partial characterization of a dynamin-like protein, EhDLP1, from the protist parasite *Entamoeba histolytica*. *Eukaryot. Cell* **9**, 215–223 (2010).
92. Levine, N. D. Protozoan parasites of domestic animals and of man (1973).
93. Singh, N., Bhattacharya, S. & Paul, J. *Entamoeba invadens*: Dynamics of DNA synthesis during differentiation from trophozoite to cyst. *Exp. Parasitol.* **127**, 329–333 (2011).
94. Bassham, D. C., Brandizzi, F., Otegui, M. S. & Sanderfoot, A. A. in *The Arabidopsis Book* **65**, 1 (2008).
95. Blessmann, J. & Tannich, E. Treatment of Asymptomatic Intestinal *Entamoeba histolytica* Infection. *N. Engl. J. Med.* **347**, 1376–1384 (2002).
96. Pollard, J. & McCarthy, J. in *Kucers' The Use of Antibiotics Sixth Edition: A Clinical Review of Antibacterial, Antifungal and Antiviral Drugs* (eds. Grayson, M. L. *et al.*) 2107–2109 (CRC Press, 2010).
97. Haghghi, A., Kobayashi, S., Takeuchi, T., Thammapalerd, N. & Nozaki, T. Geographic diversity among genotypes of *Entamoeba histolytica* field isolates. *J. Clin. Microbiol.* **41**, 3748–56 (2003).
98. Singh, J. P. & Kashyap, A. A comparative evaluation of percutaneous catheter drainage for resistant amebic liver abscesses. *Am. J. Surg.* **158**, 58–62 (1989).
99. Clark, C. G. & Diamond, L. S. Methods for cultivation of luminal parasitic protists of clinical importance. *Clin. Microbiol. Rev.* **15**, 329–341 (2002).
100. Sanchez, L., Enea, V. & Eichinger, D. Identification of a developmentally regulated transcript expressed during encystation of *Entamoeba invadens*. *Mol. Biochem. Parasitol.* **67**, 125–135 (1994).
101. Eichinger, D. Encystation of *Entamoeba* parasites. *BioEssays* **19**, 633–639 (1997).
102. Singh, N., Ojha, S., Bhattacharya, A. & Bhattacharya, S. Establishment of a transient transfection system and expression of firefly luciferase in *Entamoeba invadens*. *Mol. Biochem. Parasitol.* **183**, 90–3 (2012).
103. Siegesmund, M. A., Hehl, A. B. & van der Giezen, M. Mitosomes in trophozoites and cysts of the reptilian parasite *Entamoeba invadens*. *Eukaryot. Cell* **10**, 1582–1585 (2011).
104. Ehrenkauf, G. M., Haque, R., Hackney, J., Eichinger, D. J. & Singh, U. Identification of developmentally regulated genes in *Entamoeba histolytica*: Insights into mechanisms of stage conversion in a protozoan parasite. *Cell. Microbiol.* **9**, 1426–1444 (2007).
105. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
106. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
107. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
108. Camacho, C. *et al.* BLAST plus: architecture and applications. *BMC Bioinformatics* **10**, 1 (2009).
109. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech* **29**, 644–652 (2011).
110. Li, B. & Dewey, C. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
111. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. 1–9, <https://doi.org/10.1186/gb-2010-11-3-r25> (2010).
112. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **61** (2014).
113. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
114. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis (2015).
115. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest-HPC: Fast selection of best-fit models of protein evolution. *Bioinformatics* **6586 LNCS**, 177–184 (2011).
116. Le, S. Q. & Gascuel, O. An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307–1320 (2008).
117. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
118. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
119. Lartillot, N. & Philippe, H. Computing Bayes Factors Using Thermodynamic Integration. *Syst. Biol.* **55**, 195–207 (2006).
120. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7**(Suppl 1), S4 (2007).
121. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
122. Huelsenbeck, J. P., Ronquist, F., Nielsen, R. & Bollback, J. P. Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science (80-)*. **294**, 2310–2314 (2001).
123. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
124. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–4 (2010).
125. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
126. Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McInerney, J. O. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* **6**, 29 (2006).

Acknowledgements

The authors thank Audrey Farbos, Karen Moore and Konrad Paszkiewicz for facilitating the sequencing experiments. The Exeter Sequencing Facility was funded by a Wellcome Trust Institutional Strategic Support Award (WT097835MF). MMS acknowledges support from the Iraqi Ministry of Higher Education and Scientific Research (MOHESR) and the University of Baghdad. MvdG is grateful for support from the University of Exeter and Higher Education Funding Council for England (HEFCE). We thank Marwan Al-Maqtoofi and Christian Hacker (both University of Exeter) for technical support. Emily Herman was supported by a Graduate studentship from Alberta Innovates Health Solutions and a Vanier Graduate Scholarship. Work in the Dacks lab is supported by NSERC Discovery Grant (RES0021028). JBD is the Canada Research Chair Tier II in Evolutionary Cell Biology. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Author Contributions

M.v.d.G. and J.B.D. designed and supervised experiments; M.A.S. and M.M.S. performed the laboratory experiments; E.H., M.A.S., M.J.B., M.M.S., R.v.A. and E.C. contributed to the data analysis; E.H., M.A.S. and M.J.B. organized, designed, and wrote the paper; E.H., J.B.D. and M.v.d.G. critically revised the manuscript. All authors reviewed the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-12875-6>.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017