Running title: *Entamoeba histolytica* genome

Structure and Content of the *Entamoeba histolytica* Genome

C. Graham Clark[1,*], U. Cecilia M. Alsmark[2], Margit Hofer[3], Yumiko Saito-Nakano[4], Vahab Ali[5], Sabrina Marion[6,#], Christian Weber[6], Chandrama Mukherjee[7], Iris Bruchhaus[8], Egbert Tannich[8], Matthias Leippe[9], Thomas Sicheritz-Ponten[10], Peter G. Foster[11], John Samuelson[12], Christophe J. Noël[2], Robert P. Hirt[2], T. Martin Embley[2], Carol A. Gilchrist[13], Barbara J. Mann[13], Upinder Singh[14], John P. Ackers[1], Sudha Bhattacharya[15], Alok Bhattacharya[16], Anuradha Lohia[7], Nancy Guillén[6], Michael Duchêne[3], Tomoyoshi Nozaki[5] and Neil Hall[17]

1: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
2: Division of Biology, Newcastle University, Newcastle NE1 7RU, UK.
3: Department of Specific Prophylaxis and Tropical Medicine, Center for Physiology and Pathophysiology, Medical University of Vienna, A-1090 Vienna, Austria
4: Department of Parasitology, National Institute of Infectious Diseases, Tokyo, Japan
5: Department of Parasitology, Gunma University Graduate School of Medicine, Maebashi, Japan
6: Institut Pasteur, Unité Biologie Cellulaire du Parasitisme and INSERM U786, F-75015 Paris, France
7: Department of Biochemistry, Bose Institute, Kolkata 700054, India
8: Bernhard Nocht Institute for Tropical Medicine, D-20359 Hamburg
9: Zoologisches Institut der Universität Kiel, D-24098 Kiel, Germany
10: Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, DK-2800 Lyngby, Denmark.
11: Department of Zoology, Natural History Museum, London, SW7 5BD, UK.
12: Department of Molecular and Cell Biology, Boston University Goldman School of Dental Medicine, Boston, MA 02118, USA
13: Department of Medicine, Division of Infectious Diseases, University of Virginia Health Sciences Center, Charlottesville, VA 22908, USA
14: Departments of Internal Medicine, Microbiology, and Immunology, Stanford University School of Medicine, Stanford, CA 94305, USA
15: School of Environmental Sciences, Jawaharlal Nehru University, New Delhi 110067, India
16: School of Life Sciences and Information Technology, Jawaharlal Nehru University, New Delhi 110067, India
17: The Institute for Genomic Research, Rockville, MD 20850, USA

*Corresponding Author
Current address: Cell Biology and Biophysics Program, European Molecular Biology Laboratory, 69117 Heidelberg, Germany

ABSTRACT

The intestinal parasite *Entamoeba histolytica* is one of the first protists for which a draft genome sequence has been published. Although the genome is still incomplete, it is unlikely that many genes are missing from the list of those already identified. In this review we summarise the features of the genome as they are currently understood and provide previously unpublished analyses of many of the genes.

1. INTRODUCTION

Entamoeba histolytica is one of the most widespread and clinically important parasites, causing both serious intestinal (amoebic colitis) and extraintestinal (amoebic liver abscess) diseases throughout the world. A recent WHO estimate (WHO, 1998) places *E. histolytica* second after *Plasmodium falciparum* as causing the most deaths annually (70,000) among protistan parasites.

Recently a draft of the complete genome of *E. histolytica* was published (Loftus *et al.*, 2005) making it one of the first protist genomes to be sequenced. The *E. histolytica* genome project was initiated in 2000 with funding from the Wellcome Trust and the National Institute of Allergy and Infectious Diseases to the Wellcome Trust Sanger Institute and The Institute for Genomic Research (TIGR) in the UK and the USA, respectively. The publication describing the draft sequence concentrated on the expanded gene families, metabolism and the role of horizontal gene transfer in the evolution of *E. histolytica.* In this review we summarise the structure and content of the *E. histolytica* genome in comparison to other sequenced parasitic eukaryotes, provide a description of the current assembly and annotation, place the inferred gene content in the context of what is known about the biology of the organism, and discuss plans for completing the *E. histolytica* genome project and extending genome sequencing to other species of *Entamoeba.*

The fact that the genome sequence is still a draft has several important consequences. The first is that a few genes may be missing from the sequence data we have at present, although the number is likely to be small. For example, at least one gene (amoebapore B) is not present in the genome data despite it having been cloned, sequenced and the protein extensively characterised well before the start of the genome project. The second consequence is that the assembly contains a number of large duplicated regions that may be assembly artifacts, meaning that the number of gene copies is over-estimated in several cases. These problems cannot as yet be resolved but should be eventually as more data becomes available. Nevertheless, it is important to remember these issues when reading the rest of this article.

As the number of genes in *E. histolytica* runs into several thousands it is not possible to discuss all of them. However, we have generated a number of tables that identify many genes and link them to their entries in GenBank using the relevant protein identifier. Only a few tables are included in the text of this review, but the others are available on line as supplementary material - http://pathema.tigr.org/pathema/entamoeba_resources.shtml. The *E. histolytica* genome project data are being 'curated' at TIGR and it is on that site that the most current version of the assembled genome will be found. The 'Pathema' database will hold the data and the annotation (http://pathema.tigr.org/). The gene tables are also linked to the appropriate entry in the Pathema database and the links will be maintained as the genome structure is refined over time.

Reference is made throughout the text to other species of *Entamoeba* where data are available. *Entamoeba dispar* is the sister species to *E. histolytica* and infects humans without causing symptoms. *Entamoeba invadens* is a reptilian parasite that causes invasive disease, primarily in snakes and lizards, and is widely used as a model for *E. histolytica* in the study of encystation although the two species are not very closely related (Clark *et al.*, 2006b). Genome projects for both these species are underway at TIGR and it is anticipated that

high quality draft sequences will be produced for both in the near future. It is hoped that the *E. dispar* sequence will prove useful in identifying genomic differences linked to disease causation while that of *E. invadens* will be used to study patterns of gene expression during encystation. Small-scale genome surveys have been performed for two other species: *E. moshkovskii*, which is primarily a free-living species although it occasionally infects humans, and *E. terrapinae*, a reptilian commensal species (http://www.sanger.ac.uk/Projects/Comp_Entamoeba/).

2. GENOME STRUCTURE

2.1 The *E. histolytica* Genome Sequencing, Assembly and Annotation Process

The first choice to be made in the genome project was perhaps the easiest - the identity of the strain to be used for sequencing. A significant majority of the existing sequence data prior to the genome project was derived from one strain: HM-1:IMSS. This culture was established in 1967 from a rectal biopsy of a Mexican man with amoebic dysentery and axenised shortly thereafter. It has been used widely for virulence, immunology, cell biology and biochemistry in addition to genetic studies. In an attempt to minimise the effects of long-term culture cryopreserved cells that had been frozen in the early 1970s were revived and this uncloned culture used to generate the DNA for sequencing.

Before undertaking a genome scale analysis it is important to understand the quality and provenance of the underlying data. The *E. histolytica* genome was sequenced by whole genome shotgun approach with each center generating roughly half of the reads. Several different DNA libraries containing inserts of different sizes were produced using DNA that had been randomly sheared and sequences were obtained from both ends of each cloned fragment. The Phusion assembler (Mullikin and Ning, 2003) was used to assemble the 450,000 short reads into larger contigs (contiguous sequences), resulting in 1819 genome fragments that were approximately 12X deep, which means that each base has

been sequenced 12 times, on average. While the genome shotgun sequence provides high coverage of each base it is inevitable that there will be misassemblies and sequencing errors in the final consensus particularly towards each end of the contigs. Another problem with draft sequence is that it contains gaps, and while most of these will be small and will mostly contain repetitive non-coding "junk" sequence, some of the gaps will probably contain genes. This makes it impossible to be absolutely certain of the absence of particular genes in *E. histolytica* and, in some cases, the presence or absence of particular biological pathways. Due to the high repeat content and low GC content (24.1%) of the *E. histolytica* genome, closure of the remaining gaps is likely to be a lengthy process. Therefore it was decided to undertake and publish an analysis of the genome draft following assembly of the shotgun reads.

Annotation of the protein coding regions of the genome was initially carried out using two genefinders (GlimmerHMM (Majoros *et al.*, 2004) and Phat (Cawley *et al.*, 2001)) previously used successfully on another low G+C genome, that of *P. falciparum*. The software was re-trained specifically for analysis of the *E. histolytica* genome. The training process involved preparing a set of 600 manually edited genes to be used as models with the subsequent genefinding then being carried out on all of the assembled contigs to generate a 'complete' gene set. Predicted gene functions were generated automatically by homology searches using public protein and protein-domain databases, with subsequent refinement of identifications being carried out by manual inspection. For particular genes and gene families of special interest, members of the *Entamoeba* scientific community were involved throughout this process as expert curators with each individual assisting in the analysis and annotation of their genes of interest. Therefore although the manual curation of the genome has not been systematic, those areas of biology that are of primary interest to the *Entamoeba* community have been annotated most thoroughly. The publication of the genome by Loftus *et al* therefore represents a "first draft" of the complete genome sequence and the level of annotation is similar to the

initial publications of other genomes such as *Drosophila* (Adams *et al.*, 2000; Myers *et al.*, 2000) and human (Lander *et al.*, 2001).

2.2 Karyotype and Chromosome Structure.

The current *E. histolytica* genome assembly is approximately 23.7 million basepairs (Mbp) in size (Table 1). This figure is not likely to be a very accurate measure. In part this is due to misassembly of repetitive regions, which will cause the genome to appear smaller, and in part because of the possibility of aneuploidy in some regions of the genome, which would cause them to appear more than once in the assembly. Overall, however, this size is not inconsistent with data from pulse-field gels (Willhoeft and Tannich, 1999) and kinetic experiments (Gelderman *et al.*, 1971a,b) making the *E. histolytica* genome comparable in size (24 Mbp) to that of *Plasmodium falciparum* (23 Mbp) (Gardner *et al.*, 2002), *Trypanosoma brucei* (26 Mbp) (Berriman *et al.*, 2005), and the free living amoeba *Dictyostelium discoideum* (34 Mbp) (Eichinger *et al.*, 2005).

The current assembly does not represent complete chromosomes. Analysis of pulse-field gels predicts 14 chromosomes ranging in sizes from 0.3 to 2.2 Mb and possibly a ploidy of four (Willhoeft and Tannich, 1999). There is no current information regarding the size and nature of the centromeres and there are no contigs that appear to contain likely centromeric regions based on comparisons with other organisms. A search for signature telomeric repeats within the data indicates that these are either not present in the genome, not present in our contigs, or are diverged enough to be unidentifiable. However, there is circumstantial evidence that the chromosome ends may contain arrays of tRNA genes (see 2.4 below).

2.3 Ribosomal RNA Genes

The organisation of the structural RNA genes in *E. histolytica* is unusual with the rRNA genes carried exclusively on 24 kb circular episomes (Bhattacharya

et al., 1998) that have two transcription units in an inverted repeat. These episomes are believed to make up about 20% of the total cellular DNA; indeed, roughly 15% of all of the sequencing reads generated in the genome project were derived from this molecule with the exception of certain libraries where attempts were made to exclude it. There are thought to be numerous other circular DNA molecules of varying sizes present with unknown functions (Dhar *et al.*, 1995; Lioutas *et al.*, 1995)but unfortunately they have not yet been identified in the genome shotgun sequence data. The exact reasons for this are unknown but the small size of the DNA may have prevented proper shearing during the library construction process. These molecules represent an intriguing unsolved aspect of the *E. histolytica* genome.

2.4 tRNA Genes

Perhaps the most unusual structural feature identified in the *E. histolytica* genome is the unprecedented number and organisation of its tRNA genes (Clark *et al.*, 2006a). Over 10% of the sequence reads contained tRNA genes and these are (with a few exceptions) organised in linear arrays. The array organisation of the tRNAs was immediately obvious in some cases from the presence of more than one repeat unit in individual sequence reads and in other cases from their presence in both reads from the two ends of the same clone. However because of the near complete identity of the array units they were impossible to assemble by the software used and therefore the size of the arrays cannot be estimated accurately.

By manual assembly of tRNA gene-containing reads, 25 distinct arrays with unit sizes ranging from under 500 bp to over 1750 bp were identified (Clark *et al.*, 2006a). The arrayed genes are predicted to be functional because of the 42 acceptor types found in arrays none has been found elsewhere in the genome. These array units encoded between one and five tRNAs and a few tRNA genes are found in more than one unit. Three arrays also encode the 5S RNA and one encodes what is thought to be a small nuclear RNA. Experimental quantitative

hybridisations suggest a copy number of between about 70 and 250 for various array units. In total it is estimated that there are about 4500 tRNA genes in the genome. The frequency of a particular tRNA isoacceptor appears to be independent of the codon usage in *E histolytica* protein-coding genes.

Between the genes in the array units are complex, non-coding, short tandem repeats ranging in size from 5 to over 36 bp. Some variation in short tandem repeat number is observed between copies of the same array unit but this variation is usually minor and not visible when inter-tRNA PCR amplification is performed. However, these regions often exhibit substantial variation when different isolates of *E. histolytica* are compared and this is the basis of a recently described genotyping method for this organism (Ali *et al.*, 2005).

There is indirect evidence to suggest that the tRNA arrays are present at the ends of chromosomes. Although allelic *E. histolytica* chromosomes often differ substantially in size in pulse-field gels, a central protein-encoding region appears to be conserved as DNA digested with rare cutting enzymes gives only a single band in Southern blots when most protein-coding genes are used as probes. In contrast, when some tRNA arrays are used as probes on such blots, the same number of bands is seen in digested and undigested DNA. It is therefore tempting to conclude that the tRNA genes are at the ends of the chromosomes and to speculate that these repeat units may perform a structural role. In *D. discoideum* it is thought that rDNA may function as a telomere in some cases (Eichinger *et al.*, 2005) and the tRNA arrays in *E. histolytica* may perform a similar role.

The chromosomal regions flanking the tRNA arrays are generally devoid of protein coding genes but often contain incomplete transposable elements (see next section) and other repetitive sequences (Clark *et al.*, 2006a). This is also consistent with a telomeric location.

2.5 LINEs

The *E. histolytica* genome is littered with transposable elements. There are two major types autonomous LINEs (Long Interspersed Elements) of which there are three subtypes (EhLINE 1, 2 and 3) and there are two types of SINEs (Short Interspersed Elements) (Eh SINE1 and 2) (Table 2a). The classification of these elements and their organisation has been reviewed recently (Bakre *et al.*, 2005). Phylogenetic analysis of the EhLINEs places them in the R4 clade of non- Long Terminal Repeat (LTR) elements, a mixed clade of elements that includes members from nematodes, insects, and vertebrates (Van Dellen *et al.*, 2002a). Analysis of the *E. histolytica* genome shows no evidence for the presence of LTR retrotransposons and very few DNA transposons (of the *Mutator* family) (Pritham *et al.*, 2005).

All copies of EhLINEs examined encode non-conservative amino acid changes, frame shifts, and/or stop codons and no copy with a continuous open reading frame (ORF) has yet been found. This suggests that the majority of these elements are inactive. However, a large number of EhLINE1 copies do contain long ORFs without mutations in the conserved protein motifs of the RT and EN domains, suggesting that inactivity is quite recent. ESTs corresponding to EhLINEs have been found suggesting that transcription of these elements still occurs. Although most R4 elements insert in a site-specific manner, EhLINEs do not show strict site-specificity and are widely dispersed in the genome. They are quite frequently found close to protein-coding genes and inserted near T-rich stretches (Bakre *et al.*, 2005).

All three EhLINE subtypes are of approximately equal size ranging from 4715 to 4811 bp in length. Individual members within an EhLINE family typically share >85% identity, while between families they are <60% identical. By aligning the available sequences, each EhLINE can be interpreted to encode a single predicted ORF that spans almost the entire element (EhLINE1, 1589 aa; EhLINE2, 1567 aa; EhLINE3, 1587 aa). However, a precise 5bp duplication at

nucleotide position 1442 in about 80% of the copies of EhLINE1 creates a stop codon, dividing the single ORF in two. Similarly in 92% of EhLINE2 copies, the single ORF contains a precise deletion of two nucleotides at position 1272, resulting in two ORFs. Very few intact copies of EhLINE3 are found. The location of the stop codon leading to two ORFs appears to be conserved since in both EhLINE1 and EhLINE2 the size of ORF1 is about half that of ORF2 (Bakre *et al.*, 2005). Among the identifiable domains in the predicted proteins are reverse transcriptase (RT) and a restriction enzyme-like endonuclease (EN). The putative 5' and 3' untranslated regions are very short (3-44 bp).

EhLINEs 1 and 2 appear to be capable of mobilising partner SINEs (see next section) for which abundant transcripts have been detected in *E. histolytica*. Putative LINE/SINE partners can be assigned on the basis of conserved sequences at the 3' -ends of certain pairs, which otherwise showed no sequence similarity. The relevance of this assignment for the EhLINE1/SINE1 pair has recently been demonstrated (Mandal *et al.*, 2004).

2.6 SINEs

The two EhSINEs are clearly related to the EhLINEs as they have a conserved 3' sequence. They are nonautonomous, non-LTR retrotransposons (nonautonomous SINEs). The genetic elements encoding the abundant polyadenylated but untranslatable transcripts found in *E. histolytica* cDNA libraries (initially designated IE elements (Cruz-Reyes and Ackers, 1992; Cruz-Reyes *et al.*, 1995) or *ehapt2* (Willhoeft *et al.*, 2002)) have now been designated EhSINE1 (Van Dellen *et al.*, 2002a; Willhoeft *et al.*, 2002). BLAST searching with representative examples of the first 44 EhSINE1s detected has identified 90 full-length (= 99% complete) copies and at least a further 120 partial (= 50% of full length) copies in the genome. Length variation is observed among EhSINE1s and is largely due to variable numbers of internal 26-27 bp repeats (Ackers, unpublished). The majority contain two internal repeats and cluster closely around 546 bp in length.

A second *E. histolytica* SINE (EhSINE2) has recently been described (Van Dellen *et al.*, 2002a; Willhoeft *et al.*, 2002). Examination of the four published sequences again suggests the presence of variable numbers of short (20 bp) imperfect repeats. BLAST searching identified a total of 47 full-length (= 99%) and at least 60 partial copies in the genome. The 3'-end of EhSINE2 shows high similarity (76%) to the 3' end of EhLINE2.

A polyadenylated transcript designated UEE1 found commonly in cDNA libraries from *E. dispar* (Sharma *et al.*, 1999) is also a non-LTR retrotransposon. A single copy of a UEE1-like element has been identified in the *E. histolytica* genome and is here designated EhSINE3. There is no significant sequence identity between EhSINE3 and EhLINE3 but the 3' end of EhSINE3 is very similar to that of EhLINE1.

Analysis of an *E. histolytica* EST library identified over 500 significant hits to both EhSINE1 and EhSINE2. No convincing transcript from EhSINE3 could be identified although the nearly identical *E. dispar* UEE elements (EdSINE1; Shire and Ackers, submitted) are abundantly transcribed.

A very abundant polyadenylated transcript, *ehapt1*, was described by Willhoeft *et al.* (1999) in a cDNA library. However, only a small number of partial matches could be found in the current *E. histolytica* assembly and only 10-20 strong hits in the much larger *E. histolytica* EST library now available. *ehapt1* does not appear to be a SINE element and its nature is currently unclear. The lack of matches in the genome suggests either that it is encoded in regions missing from the current assembly or that it contains numerous introns.

2.7 Other Repeats

The *E. histolytica* genome contains a number of other repetitive elements whose functions are not always clear. There are over 75 genes encoding

leucine-rich tandem repeats (LRR) of the type found in BspA-like proteins of the *Treponema pallidum* LRR (TpLRR) subfamily, which has a consensus sequence of LxxIxIxxVxxIgxxAFxxCxx (Davis *et al*., 2006). These proteins generally have a surface location and may be involved in cell-cell interaction. Genes encoding such proteins are found mainly in Bacteria and some Archaea; so far they have been identified in only one other eukaryote, *Trichomonas vaginalis* (Hirt *et al.*, 2002). An extensive description of the BspA-like proteins *of E. histolytica* has recently been published (Davis *et al.*, 2006) and one of them has been shown to be surface exposed (Davis *et al.*, 2006).

E. histolytica stress sensitive protein (Ehssp) 1 is a dispersed, polymorphic and multicopy gene family (Satish *et al.*, 2003) and is present in ca. 300 copies per haploid genome as determined by hybridisation (Table 2a). The average Ehssp1 ORF is 1 kb in length with a centrally-located acidic-basic region (ABR) that is highly polymorphic. Unlike other such domains no clear repetitive motifs are present. The protein has, on average, 21% acidic (aspartate and glutamate) and 17% basic (arginine and lysine) amino acids, most of which are located in the ABR. The ABR varies in size from 5 to 104 amino acids among the various copies. No size polymorphism is seen outside the central ABR domain. The genes have an unusually long 5' untranslated region (UTR; 280 nucleotides). Only one or a few copies of the gene are transcribed during normal growth, but many are turned on under stress conditions. Homologues of this gene are present in *E. dispar*, but there is very little size polymorphism in the *E. dispar* gene family.

Eukaryotic genomes usually contain numerous microsatellite loci with repeat sizes of 2-3 basepairs. With the exception of di- and tri-nucleotides made up entirely of A+T such sequences are rare in the *E. histolytica* genome. In contrast, two dispersed repeated sequences of unknown function occur far more frequently than would be expected at random. Family 16 has a 42 base consensus sequence and occurs approximately 38 times in the genome while

family 17 has a 27 base consensus sequence and occurs 35 time in the genome (Table 2b). The significance of these sequences remains to be determined.

2.8 Gene Number

The current assembly predicts that the genome contains around 10,000 genes, almost twice as many as seen in *P. falciparum* (Gardner *et al.*, 2002) or *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996) but closer to that of the free living protist *Dictyostelium discoideum* (ca. 12,500; Eichinger *et al*., 2005). It should be remembered that this number will change as the assembly improves, and is likely to decrease somewhat. Nevertheless, the comparatively large gene number when compared to some other parasitic organisms reflects both the relative complexity of *E. histolytica* and the presence of large gene families, despite the loss of certain genes as a consequence of parasitism. Gene loss and gain can both represent an adaptive response to life in the human host. Gene loss is most evident in the reconstruction of metabolic pathways of *E. histolytica*, which show a consistent pattern of loss of synthetic capacity as a consequence of life in an environment rich in complex nutrient sources. Similarly, analyses of expanded gene families with identifiable functions indicate that many are directly associated with the ability to sense and adapt to the environment within the human host and the ability to ingest and assimilate the nutrients present. One consequence of these gene family expansions being linked to phagocytosis of bacteria and other cells may be an association between many of these gene families and pathogenicity.

2.9 Gene Structure

Most *E. histolytica* genes comprise only a single exon; however as many as 25% may be spliced and 6% contain two or more introns. Therefore mRNA splicing is far less common than in the related protist *D. discoideum* or the malaria parasite *P. falciparum.* The genome contains all of the essential machinery for splicing (section 2.14) and a comparison of intron positions suggests that *D. discoideum* and *E. histolytica* have both lost introns since their

shared common ancestor with *P. falciparum*, although many more have been lost in the *E. histolytica* lineage. A good example of this intron loss is the vacuolar ATP synthase subunit D gene (Figure 1). This protein is highly conserved but the number of introns in each gene varies. *P. falciparum* has 5 introns, *D. discoideum* has two and *E. histolytica* has one. The positions of three of the five *P. falciparum* introns are conserved in one of the other species which suggests that these three (at least) were present in the common ancestor and that intron loss has led to the lower number seen in *E. histolytica* today. This loss is consistent with reverse transcriptase mediated 3' intron loss (Roy and Gilbert, 2005) as the 5'-most introns are retained. It would appear that this process has been more active in the *E. histolytica* and *D. discoideum* lineages than in *P. falciparum*, possibly because *Plasmodium* lacks a reverse transcriptase.

2.10 Gene Size

Genes in *E. histolytica* are surprisingly short, not only due to the loss of introns but also in the predicted lengths of the proteins they code for. On average the predicted length of a protein in *E. histolytica* is 389 amino acids (aa) which is 129 aa and 372 aa shorter than in *D. discoideum* and *P. falciparum* respectively. In fact the protein length distribution is most similar to that of the microsporidian *Encephalitozoon cuniculi* (Figure 2) which has a very compact genome of 3Mb and less than 2000 genes. Direct comparison of orthologous genes between *E. histolytica* and its closest sequenced relative *D. discoideum* demonstrates this phenomenon quite well, with the majority of *E. histolytica* proteins being shorter than the *D. discoideum* counterpart (Hall, unpublished). Protein length is normally very well conserved among eukaryotes so the reason for protein shortening is unclear. It has been postulated that in bacteria reduced protein lengths reflects a reduced capacity for signaling (Zhang, 2000). This would not seem to be the case here as the number of genes identified as having a role in signaling suggests quite the opposite. An alternative theory is that as *E.*

histolytica has reduced organelles it is possible that its proteins contain fewer or simpler targeting signals.

2.11 Protein Domain Content

The most common protein family (Pfam) domains of *E. histolytica* are shown in Table 3. The domains that are unusually common in *E. histolytica* reflect some of the more unusual aspects of the biology of this protist. For example, the Rab and Rho families that are involved in signaling and vesicle trafficking are among the most common domains in *E. histolytica* while in other species they are not often among the top 50 families. This could well be due to the fact that *E. histolytica* has a 'predatory' life style and these domains are intimately involved in environmental sensing, endocytosis and delivery of lysosomes to the phagosome. There are also a number of domains involved in actin dynamics and cytoskeletal rearrangement that are not common in non-phagocytic species, such as the gelsolin and SH3 domains. Myb domains are the most common transcription regulatory domains in *E. histolytica*; this domain is also common in plants where the proteins regulate many plant-specific pathways (Ito, 2005). An important finding from an initial analysis was the presence of unusual multidomain proteins, including five proteins containing both RhoGEF and Arf-GAP domains, suggesting a mechanism for direct communication between the regulators of vesicle budding and cytoskeletal rearrangement. Over 80 receptor kinases were identified (section 7.2.2), each containing a kinase domain and a C rich extracellular domain. These kinases fall into distinct classes depending on the presence of CXC or CXXC repeats. There are also domains that are common in most other sequenced genomes but rare or missing from *E. histolytica.* For example, most mitochondrial carrier domain proteins are not needed in *E. histolytica* as it lacks a normal mitochondrion (section 8).

2.12 Translation-Related Proteins

Two of the predicted tRNAs (Ile^{TAT} and Tyr) need to be spliced due to the presence of an intron. tRNA introns are distinct in structure from those in protein-coding genes and require a distinct splicing machinery. The expected enzymes required for this splicing are present as are a number of tRNA modification enzymes (including those for sythesising queuine and pseudouridine) and rRNA methylases that act on specific bases in their respective RNA molecules. The expected panel of tRNA synthetases necessary for aminoacylating the tRNAs is also present, with one or two gene copies for each type.

The majority of ribosomal protein genes are well-conserved in *E. histolytica* and only the gene for large subunit protein L41 could not be identified. The missing protein is only 25 amino acids in length, 17 of which are arginines or lysines, which would make it difficult to identify in this A+T-rich genome, but it is highly conserved, having been reported from Archaea to mammals. However, it also appears to be dispensable, as *S. cerevisiae* can grow relatively normally after deletion of both its copies (Yu and Warner, 2001). Nevertheless, deletion of L41 in *S. cerevisiae* reduces the level of 80S ribosomes, suggesting that it is involved in ribosomal subunit association, reduces peptidyl transferase activity, and increases translocation (Dresios *et al.*, 2003). In addition, L41 has been shown to interact with the beta subunit of protein kinase CKII and to stimulate phosphorylation of DNA topoisomerase II alpha by CKII (Lee *et al.*, 1997b). If this gene is truly absent from *E. histolytica* it may have important consequences for the cell.

No genes for mitochondrial ribosomal proteins were found. Their absence is not surprising since *E. histolytica* lacks typical mitochondria (see section 8 below).

In eukaryotic translation, elongation factor EF-1 is activated upon GTP binding and forms a ternary complex with aminoacyl tRNAs and ribosomes. EF-1 beta and delta subunits work as GDP-GTP exchange factors to cycle EF-1 alpha

between two forms while EF-1 gamma provides structural support for the formation of this multimeric complex. EF2 assists in the translocation of tRNAs on the mRNA by exactly one codon. *E. histolytica* has most of the expected factors except for EF-1 delta, a protein involved in exchanging GDP with GTP. This is also absent from *S. cerevisiae* and *P. falciparum*. It is likely that EF-1 beta carries out this activity. It is thought that the EF-1 complex can exist in two forms, EF-1-alpha/beta/gamma and EF-1-alpha/delta/gamma. In *E. histolytica*, probably only the former complex exists.

Eukaryotes typically have two polypeptide release factors, eRF1 and eRF3. Both of these factors have been found in *E. histolytica*.

2.13 Analysis of Cell Cycle Genes

Alternation of DNA duplication and chromosome segregation is a hallmark in the cell cycle of most eukaryotes. Carefully orchestrated processes coordinate an ensemble of cell cycle regulating 'checkpoint' proteins ensure that progeny cells receive an exact copy of the parental genetic material (Hartwell and Weinert, 1989). Unlike most eukaryotes, *Entamoeba histolytica* cells can reduplicate their genome several times before cell division occurs (Gangopadhyay *et al.*, 1997). Approximately 5-20% of the trophozoites (depending on the growth phase) of axenic culture are multi-nucleated. Additionally, DNA reduplication may occur without nuclear division so that single nuclei contain 1X -6X or more genome contents (Das and Lohia, 2002). Thus axenically cultured *E. histolytica* trophozoites display heterogeneity in their genome content suggesting that eukaryotic cell cycle checkpoints are either absent or altered in this organism. Around 200 genes have been identified in yeast that play a direct role in cell cycle progression.

2.13.1 DNA replication initiation and DNA duplication

The DNA replication licensing system is one of the crucial mechanisms that ensures the alternation of S-phase with mitosis in most cells (Tye, 1999).

Initiation of DNA replication involves binding of the replicative helicases to DNA replication origins in late mitosis. Loading of the replicative helicase Mcm2-7 proteins is preceded by formation of the pre-replicative complex (pre-RC) and its subsequent activation. Formation of pre-RC requires the ordered assembly of the origin recognition complex (ORC), Cdc6, Cdt1 and the Mcm2–7 proteins. The pre-RC is activated by the protein kinase Cdc7p and its regulatory subunit Dbf4 (Masai and Arai, 2002). Other factors that regulate the transition from pre-RC to replication initiation are Mcm10p, Cdc45p, TopBP1, RecQL4 and the GINS complex (Gregan *et al.*, 2003; Machida *et al.*, 2005; Merchant *et al.*, 1997; Wohlschlegel *et al.*, 2002). Two other Mcm proteins – Mcm8 and Mcm9 - have been identified in metazoan systems and are believed to be part of the replicative helicase (Maiorano *et al.*, 2006). Replication origin licensing is inactivated during S-phase but Mcm2–9p may function as a helicase that unwinds DNA ahead of the replication fork during S-phase (Maiorano *et al.*, 2006). Once S-phase has begun, the formation of new pre-RC is kept in check by high CDK activity and by the activity of the protein geminin (Bell and Dutta, 2002).

A detailed analysis of the *E. histolytica* genome shows that homologues of several proteins required for DNA replication initiation are absent. These include ORC (Origin Recognition Complex) 2-6, Cdt1, geminin, Cdc7/Dbf4 and Mcm10. A single gene encoding a homologue of the archaeal and human Cdc6/Orc1p (Capaldi and Berger, 2004) was identified. This suggests that DNA replication initiation in *E.histolytica* is likely similar to archaeal replication initiation where a single Cdc6p/ORC1p replaces the hetero-hexameric ORC complex (Kelman and Kelman, 2004). Several proteins described from metazoa, such as Cdt1, geminin, Mcm8 and Mcm9, have not been found in yeast. Surprisingly, Mcm8 and Mcm9 were identified in the *E.histolytica* genome.

Of the four known checkpoint genes that regulate DNA replication in *S. cerevisiae* only Mec1 and Mrc1 have homologues in *E. histolytica*. *E. histolytica* homologues of several proteins involved in G1-S transitions are absent, such as Sic1, Chk1. The S-phase checkpoint genes p21, p27, p53 and retinoblastoma (RB) required for transition from G1 to S-phase in humans were absent in *E. histolytica*. Chk1 and Chk2 genes encode kinases that act downstream from the ATM and ATR kinases (intra-S phase checkpoint genes). The Chk1 homologue is absent but a Chk2 homologue has been identified in *E. histolytica* and partially characterised (Iwashita *et al.*, 2005).

2.13.2 Chromosome segregation and cell division

A large number of genes are known to regulate different events during the transition from G2-Mitosis - spindle formation checkpoint, chromosome segregation, mitosis, exit from mitosis, and cytokinesis - in *S.cerevisiae*. Many of the proteins required by yeast for kinetochore formation have no obvious homologues in *E. histolytica* suggesting that amoeba kinetochores may have an altered composition and structure. Proteins of the Anaphase Promoting Complex (APC) regulate transition from metaphase to anaphase. With the exception of APC11, none of the APC proteins could be identified in *E.histolytica*. In contrast two genes encoding CDC20 homologues, which are known to activate the APC complex, were identified in *E.histolytica* along with ubiquitin and related proteins (Wöstmann *et al.*, 1992), indicating that although most APC subunit homologues were absent the pathway of proteasomal degradation for regulation of cell cycle proteins may still be functional in *E.histolytica*. Effectors of the apoptotic pathway and meiosis were also largely absent.

2.13.3 CDKs and cyclins

The CDC28 gene encodes the single cyclin dependant kinase (CDK) in *S. cerevisiae* and regulates cell cycle progression by binding to different cyclins at the G1/S or G2/M boundaries (Reed, 1992; Surana *et al.*, 1991; Wittenberg *et*

al., 1990). Similarly, *Schizosaccharomyces pombe* also encodes a single CDK (*cdc2*) (Simanis and Nurse, 1986). Mammals and plants can encode multiple CDKs and an equally large number of cyclins (Morgan, 1995; Vandepoele *et al.*, 2002). Association of different CDKs with specific cyclins regulates the cell cycle in different developmental stages as well as in specific tissues. CDKs belong to the serine/threonine family of kinases with a conserved PSTAIRE domain where cyclins are believed to bind (Jeffrey *et al.*, 1995; Morgan, 1996) although some mammalian and plant CDKs have been shown to have divergent PSTAIRE motifs. This heterogeneity may or may not affect cyclin binding (Poon *et al.*, 1997). The *E. histolytica* genome encodes at least 9 different CDKs among which not even one has the conserved PSTAIRE motif. The closest homologue of the CDC28/cdc2 gene, which shows only conservative substitutions in the PSTAIRE motif (PVSTVRE), was cloned previously (Lohia and Samuelson, 1993). The remaining 8 CDK homologues exhibit even greater divergence in this motif. Eleven putative cyclin homologues with a high degree of divergence have been found. Identifying their CDK/cyclin partner along with their roles in the cell cycle is a major task that lies ahead. Some of the CDKs may not function by associating with their functional cyclin partners but may play a role in regulating global gene expression, either by activation from non-cyclin proteins or by other mechanisms (Nebreda, 2006).

E. histolytica presents a novel situation where the eukaryotic paradigm of a strictly alternating S-phase and mitosis is absent. Discrete G1, S and G2 populations of cells are not routinely found in axenic cultures. Instead cells in S-phase show greater than 2x genome contents, suggesting that the G2 phase is extremely short and irregular. This observation together with the absence of a large number of checkpoint genes suggests that regulation of genome partitioning and cell division in *E. histolytica* may be additionally dependant on extracellular signals. *E. histolytica* must however contain regulatory mechanisms to ensure that its genome is maintained and transmitted with precision even in the absence of the expected checkpoint controls. The

discovery of these mechanisms will be crucial to our understanding of how the *E. histolytica* cell divides.

2.14 Transcription

RNA polymerase II transcription in *E. histolytica* is known to be α-amanitin-resistant (Lioutas and Tannich, 1995). The F homology block of the RNA polymerase II large subunit has been identified as the putative α-amanitin binding site. This block is highly divergent in the α-amanitin resistant *Trichomonas vaginalis* RNA polymerase II (Quon *et al.*, 1996). The *E. histolytica* RPB1 homologue also diverges from the consensus in this region but, interestingly, it is also quite dissimilar to the *T. vaginalis* sequence.

The heptapeptide repeat (TSPTSPS) common to other eukaryotic RNA polymerase II large subunit C terminal domains (CTD) is not present in the *E. histolytica* protein. Indeed, the *E. histolytica* CTD is not similar to any other RNA polymerase II domain in the current database. However, the CTD of the *E. histolytica* enzyme does remain proline/serine-rich (these amino acids constitute 40% of the CTD sequence). The *E. histolytica* CTD also retains the potential to be highly phosphorylated: of the 24 serines, 6 threonines and 3 tyrosines within the CTD, 9 serines, 3 threonines and 1 tyrosine are predicted to be within potential phosphorylation sites. It is therefore possible that, despite its divergence, modification of the CTD by kinases and phosphatases could modulate protein-protein interactions as is postulated to occur in other RNA polymerases (Yeo *et al.*, 2003). In yeast, phosphorylation of the CTD regulates association with the mediator protein (Davis *et al.*, 2002; Kang *et al.*, 2001; Kornberg, 2001). The yeast mediator protein complex consists of 20 subunits. However, perhaps due to the divergence of the CTD, only two of these proteins have been identified in *E. histolytica* (Med7 and Med10). Homologues of the Spt4 and Spt5 elongation factors, also thought to interact with the CTD, have been identified.

The RNA polymerase core is composed of 12 putative subunits in *S. cerevisiae* (Young, 1991), while *S. pombe* contains a subset of 10 of these proteins, lacking the equivalents of subunits 4 and 9 (Yasui *et al.*, 1998). In *E. histolytica* only 10 of the RNA polymerase subunits have been identified, identifiable homologues of subunits 4 and 12 being absent. While the homologue to subunit 9 was present it lacks the first of the two characteristic zinc binding motifs of this protein and the DPTLPR motif in the C terminal region. A similar sequence, DPTYPK, is however present and a homologue of the TFIIE large subunit Tfa1, which is proposed to interact with this region of the protein, has been identified (Hemming and Edwards, 2000; Van Mullem *et al.*, 2002). The conserved N terminal portion (residues 1-52) of Rpb9 is thought to interact with both Rpb1 and Rpb2 in *S. cerevisiae* (Hemming and Edwards, 2000) and homologues of these have been identified.

The core promoter of *E. histolytica* has an unusual tripartite structure consisting of the three conserved elements TATA, GAAC and INR (Purdy *et al.*, 1996; Singh and Rogers, 1998; Singh *et al.*, 2002; Singh *et al.*, 1997). Singh and Rogers (1998) have speculated that the GAAC motif may be the binding site of a second or alternative *E. histolytica* DNA binding protein in the preinitiation complex. It is therefore of interest that, in addition to the *E. histolytica* TATA-binding protein (TBP), two other proteins contain the TATA-binding motif (Hernandez *et al.*, 1997). TBP is a subunit of the TFIID general transcription factor (GTF) which in other organisms is required for the recognition of the core promoter. In light of the variation in the core promoter previously mentioned, and the divergence in proteins that bind to the core promoter in other parasitic protists, it is not surprising that only six of the 14 evolutionary conserved subunits of TFIID, TBP Associated Factors (TAFs) 1, 5, 6, 10, 12 and 13 were identified. Homologues of some of the global regulatory subunits of the Ccr4/Not complex, which interacts with TBP and TAFs 1 and 13, have also been identified.

TAFs 5, 6, 10 and 12 are also components of the histone acetyltransferase (HAT) complexes in other organisms as is SPT6 and 16 (Carrozza *et al.*, 2003). While all known components of the HAT complexes have by no means been identified or the role of the previously unknown bromodomain containing proteins encoded in the *E. histolytica* genome, histone acetylation complexes are known to be active in *E. histolytica* (Ramakrishnan *et al.*, 2004). Other potential members of chromatin remodeling complexes of *E. histolytica* include the TBP interacting helicase (RVB1 & 2) and the SNF2 subunit of the SWI/SNF complex.

Homologues of some of the other GTFs (TFII E, F and H) but not the large or small subunits were identified. In contrast to the difficulty identifying some of the GTFs, the *E. histolytica* spliceosomes components U1, U2, U4/6, U5 and the Prp19 complex have all been identified. In fact homologues of ten of the fourteen "core" snRNP proteins, two of the U1 specific snRNPs, seven of the ten U2 specific snRNPs, five of the six U5 specific snRNPs, three of the U4/6 specific snRNPs, and four of the nine subunits of the Prp19 complex have been found. In fact *E. histolytica* has homologues of approximately 80% of the *S. cerevisiae* splicing machinery (Jurica and Moore, 2003).

Like *G. intestinalis, E. histolytica* has short 5' untranslated regions on its mRNAs. However, unlike those of *G. intestinalis, E. histolytica* mRNA has been shown to be capped (Ramos *et al.*, 1997; Vanacová *et al.*, 2003). Identification of homologues of the Ceg1 RNA guanylyltransferase - an enzyme which adds an unmethylated GpppRNA cap to new transcripts - and of Abd1 - which methylates the cap to form m7GpppRNA - gives new insight into the probable cap structure in *E. histolytica* (Hausmann *et al.*, 2001; Pillutla *et al.*, 1998). It has been proposed that the capping enzymes interact with the phosphorylated CTD of RNA polymerase (Schroeder *et al.*, 2000). The CTD of *E. histolytica* large subunit is, as discussed earlier, not well conserved but contains several probable phosphorylation sites.

mRNAs in *E. histolytica* are polyadenylated and the polyadenylation signal is found within the short 3' untranslated region (Bruchhaus *et al.*, 1993; Li *et al.*, 2001). However only eight of the eighteen yeast Cleavage and Polyadenylation Specificity Factor (CPSF) subunits are identifiable in *E. histolytica*.

3. VIRULENCE FACTORS

3.1 Gal/GalNAc Lectin

One of the hallmarks of *E. histolytica* pathogenicity is contact-dependent killing of host cells. *E. histolytica* is capable of killing a variety of cells types including human intestinal epithelium, erythrocytes, neutrophils, and lymphocytes (Burchard and Bilke, 1992; Burchard *et al.*, 1992a; Burchard *et al.*, 1992b; Guerrant *et al.*, 1981; Ravdin and Guerrant, 1981). Cytolysis occurs as a step-wise process that begins with adherence to target cells via galactose/N-acetyl D-galactosamine-inhibitable (Gal/GalNAc) lectin (Petri *et al.*, 1987; Ravdin and Guerrant, 1982). Adherence via the Gal/GalNAc lectin is a requirement for cell killing because in the presence of galactose or GalNAc targets cells are not killed by the amoebae. Target cell death occurs within 5 to 15 minutes and is often followed by phagocytosis. Inhibition of the Gal/GalNAc lectin with galactose or specific antibody also blocks phagocytosis (Bailey *et al.*, 1990). Resistance to lysis by the complement system is also mediated in part by the Gal/GalNAc lectin. The lectin contains a CD59-like domain that likely helps protect the trophozoites from complement; CD59 is a surface antigen of many blood cells known to have this property (Braga *et al.*, 1992).

The Gal/GalNAc lectin is a membrane complex that includes heavy (Hgl) 170 kDa, and light (Lgl) 30-35 kDa subunits linked by disulphide bonds, and a non-covalently associated intermediate (Igl) 150 kDa subunit (Cheng *et al.*, 2001; Petri *et al.*, 1989). The structure and function of the Gal/GalNAc lectin has recently been reviewed (Petri *et al.*, 2002). The heavy subunit is a type1 transmembrane protein while the light and intermediate subunits have

glycosylphosphatidylinositol (GPI) anchors (Cheng *et al.*, 2001; McCoy *et al.*, 1993). Gal/GalNAc lectin subunits do not share any significant protein identity or similarity to any other known proteins, though Hgl and Igl have some very limited regions of similarity with known classes of proteins that will be discussed below.

3.1.1 The heavy (Hgl) subunit

Based on pulse-field gel electrophoresis there are five loci in the genome with similarity to the Hgl subunit. However, the current genome assembly only identifies two complete genes, one of which corresponds to Hgl2 (Tannich *et al.*, 1991b). The predicted proteins encoded by these loci are 92% identical. In initial assemblies there were three other sequences with high similarity to the Hgl subunit that were pseudogenes. These pseudogenes may account for the additional loci detected by pulse-field gel electrophoresis. The large size of these genes means that assembly problems may also be affecting our interpretation.

Hgl subunit sequences can be divided in to domains based on amino acid content and distribution (Figure 3). The amino-terminal domain of ca. 200 amino acids consist of 3.2% cysteine and 2.1% tryptophan residues. The next domain, also ca. 200 amino acids, is completely devoid of these two amino acids. The C-terminal domain of ca. 930 amino acids is cysteine-rich, comprising 10.8 % cysteine. The number and spacing of all predicted tryptophan and cysteine residues are 100% conserved in the two complete genes. Although a portion of the C-terminal domain can be said to contain cysteine-rich pseudo- repeats, there is no clear repetitive nature to the protein (Tannich *et al.*, 1991b). The Hgl subunit has a single transmembrane domain and a highly conserved 41 amino acid cytoplasmic domain. In addition to these two *hgl* genes, the genome contains a newly identified divergent member of the Hgl gene family (XP_650534). This ORF shares 43% similarity with the two

other Hgl isoforms, and is predicted to encode a protein with an almost identical domain structure to that of Hgl described above.

3.1.2 The light (Lgl) subunit

The Lgl subunit is encoded by five genes (*lgl1-5*) that share 74-85% amino acid identity. A sequence corresponding to Lgl2 is missing from the current genome assembly. The light subunits range from 270 to 294 amino acids in length. Each isoform has a 12 amino acid signal peptide, 5 conserved cysteine residues, and a GPI-anchor addition site. Lgl1 has two potential glycosylation sites. Lgl2 has one of these sites, Lgl3 has one different site, and Lgl4 and Lgl5 have none.

3.1.3 The intermediate (Igl) subunit

The Igl subunit was first identified by a monoclonal antibody that blocked amoebic adherence to and cytotoxicity for mammalian cells (Cheng *et al.*, 1998). Co-purification of the Hgl, Lgl, and Igl suggests that these three subunits form a complex (Cheng *et al.*, 1998; Cheng *et al.*, 2001). The Igl subunit also has galactose-binding activity (Cheng *et al.*, 1998) and can serve as protective antigen in vaccine trials (Cheng and Tachibana, 2001). There are two loci that encode Igl subunits (Cheng *et al.*, 2001) and the predicted amino acid sequences are 81% identical. The Igl subunit, like the Hgl subunit, does not have any recognisable carbohydrate-binding domain.

3.1.4 Conservation of Gal/GalNAc lectin subunits in other species of Entamoeba

There are clearly identifiable orthologues of the Hgl and Lgl subunits among the limited sequences of *E. dispar, E. invadens, E. moshkovskii,* and *E. terrapinae* available at present (Dodson *et al.*, 1997; Pillai *et al.*, 1997; Wang *et al.*, 2003). Because these genomes are incomplete it is possible that as yet unidentified family members will show greater similarity to the *E. histolytica* sequences. Nevertheless, the Lgl subunit is quite conserved among the five

Entamoeba species. For instance, the *E. terrapinae* gene is 56% identical and 62 % similar to *E. histolytica* Lgl1 over a span of 201 amino acids. The Hgl subunits are more diverse. The *E. dispar* Hgl orthologue is highly similar to the *E. histolytica* subunit (86%) but the other species show more diversity, including the region that corresponds to the CRD. However, the number and positions of the cysteine residues are highly conserved, as is the sequence of the cytoplasmic domain, showing only a few changes. It is difficult to put precise numbers to these similarities because the complete sequences of Hgl subunits from the other species are not present in the database. The character of the conservation of the Hgl subunits suggests that the ligand specificity is different for the Hgl subunits of each species but the signaling functions of the cytoplasmic domains are similar, if not perhaps identical. Only *E. dispar* has an identifiable Igl subunit. The other three species clearly have paralogues of the CXXC repeat family to which Igl belongs, but their similarity to Igl is mostly restricted to the CXXC and CXC repeat motifs.

3.2 Cysteine endopeptidases

Entamoeba histolytica is characterised by its extraordinary capacity to invade and destroy human tissues. The main lytic activity has been attributed to cysteine endopeptidases. This class of enzymes, which is found in all organisms, plays a major role in the pathogenicity of *E. histolytica* as demonstrated in a large number of in vitro and in vivo studies (Ankri *et al.*, 1999; Gadasi and Kessler, 1983; Keene *et al.*, 1990; Li *et al.*, 1995; Luaces and Barrett, 1988; Lushbaugh *et al.*, 1985; Reed *et al.*, 1989; Schulte and Scholze, 1989; Stanley *et al.*, 1995). Most striking are results from laboratory animal infections showing that *E. histolytica* trophozoites with reduced cysteine proteinase activity are greatly impaired in their ability to induce amoebic disease (Ankri *et al.*, 1999; Stanley *et al.*, 1995). In addition, the discovery that *E. histolytica* cysteine proteinases possess interleukin-1β convertase activity suggests that these enzymes use a mechanism that is novel in microbial pathogenicity (Zhang *et al.*, 2000).

Thiol-dependent proteolytic activity in *E. histolytica* was first attributed to a neutral sulphydryl proteinase (McLaughlin and Faubert, 1977) and later to a cytotoxic proteinase (Lushbaugh *et al.*, 1984). Other terms that have been used to describe closely related or identical enzymes are cathepsin B (Lushbaugh *et al.*, 1985), neutral proteinase (Keene *et al.*, 1990), histolysin (Luaces and Barrett, 1988) (later changed to histolysain; Luaces *et al.*, 1992)), and amoebapain (Scholze *et al.*, 1992). *E. histolytica* cysteine endopeptidases were found to be secreted (Leippe *et al.*, 1995) and localised in lysosome-like vesicles or at the surface of the cell (Garcia-Rivera *et al.*, 1999; Jacobs *et al.*, 1998). Molecular cloning has revealed a large number of cysteine endopeptidase genes in the *E. histolytica* genome (Bruchhaus *et al.*, 2003; Garcia-Rivera *et al.*, 1999; Reed *et al.*, 1993; Tannich *et al.*, 1991c; Tannich *et al.*, 1992). Interestingly, most of these genes are not expressed during in vitro cultivation (Bruchhaus *et al.*, 2003). As our current knowledge of *E. histolytica* biology and pathogenicity is mostly based on analysis of cultured cells, the function of most of the cysteine endopeptidases and their precise role in *E. histolytica* virulence is largely unknown.

Homology searches using conserved active site regions revealed that the *E. histolytica* genome contains at least 44 genes coding for cysteine endopeptidases. Of these, the largest group is structurally related to the C1 papain superfamily (Table 4), whereas a few others are more similar to family C2 (calpain-like cysteine proteinases), C19 (ubiquitinyl hydrolase), C54 (autophagin), and C65 (otubain), respectively (Table 5).

Phylogenetic analyses of the 36 C1-family members revealed that they represent 3 distinct clades (A, B, C), consisting of 12, 11 and 13 members, respectively. Clade A and B members correspond to the two previously described subfamilies of *E. histolytica* cysteine proteinases, designated EhCP-A and EhCP-B (Bruchhaus *et al.*, 2003). In contrast, clade C represents a new

group of *E. histolytica* cysteine endopeptidases that has not been described before. EhCP-A and EhCP-B subfamily members are classical pre-pro enzymes with an overall cathepsin L-like structure (Barrett 1998) as indicated by the presence of an ERFNIN motif in the pro region of at least 21 of the 23 EhCP-A and EhCP-B enzymes (Figure 4). Interestingly, biochemical studies with purified EhCP-A indicated a cathepsin B-like substrate specificity (Scholze and Schulte, 1988). This is likely due to the substitution of an alanine residue by acidic or charged amino acids in the postulated S2 pocket, corresponding to residue 205 of the papain sequence (Barrett 1998). As reported previously (Bruchhaus *et al.*, 2003), the EhCP-A and EhCP-B subfamilies differ in the length of the pro regions as well as of the catalytic domains, and have distinct sequence motifs in the N-terminal regions of the mature enzymes (DWR vs. PCNC). Moreover, none of the EhCP-A subfamily but 10 of the 11 Eh CP-B sequences contain hydrophobic stretches near or at the C-terminus, some of which are predicted to constitute transmembrane helices (TMH) or GPI-attachment moieties. This finding is consistent with previous reports on surface localisation of *E. histolytica* cysteine proteinases but, so far, studies on the cellular localisation of the various EhCP-B molecules have not been reported.

In contrast to the EhCP-A and EhCP-B subfamilies, primary structure prediction indicates that EhCP-C members are not pre-pro enzymes, as they lack hydrophobic signal sequences as well as identifiable pro regions. Instead, they contain a hydrophobic region located 11 to 28 amino acids from the N-terminus, which is predicted to form a TMH (Figure 4). Therefore, this new group of molecules appears to be membrane associated via a signal anchor. All EhCP-C enzymes have a conserved motif of the sequence $H/I(X)_6L/ICP$ in the C-terminal half but they differ substantially in their pI, with values ranging from 4.6 to 8.8. As there is no example of a structurally related cysteine endopeptidase corresponding to the EhCP-C subfamily in other organisms, the specific functions of this group of molecules remain completely unknown.

In addition to the large number of C1 superfamily members, the *E. histolytica* genome contains 2 genes encoding cysteine endopeptidases homologous to family C2 or calpain-like cysteine proteases (EhCALP1 and EhCALP2). Enzymes of this class contain several calcium-binding domains and have been shown to participate in a variety of cellular processes including remodeling of the cytoskeleton and membranes, signal transduction pathways and apoptosis.

Another 4 genes were identified coding for enzymes with homology to the peptidase family C54 also termed autophagins (EhAUTO1-4). The process of autophagy has been studied in human and yeast cells (Kirisako *et al.*, 2000; Marino *et al.*, 2003). Autophagy is a mechanism for the degradation of intracellular proteins and the removal of damaged organelles. During this process the cellular components become enclosed in double membranes and are subsequently degraded by lysosomal peptidases. Autophagins seem to be important for cytoplasm-to-vacuole targeting.

Two other genes encoding putative cysteine endopeptidases of *E. histolytica* show homology to the C19 and C65 families. These two groups of enzymes are known to be involved in ubiquitin degradation. Family C19 are ubiquitinyl hydrolases described as having ubiquitin-specific peptidase activity in humans. C65 or otubains are a group of enzymes with isopeptidase activity, which releases ubiquitin from polyubiquitin.

In summary, the *Entamoeba* genome contains a considerable number of endopeptidase genes. Elucidation of the precise role of each of the various enzymes will be a major challenge but may help us to understand the mechanism(s) of virulence and other unique properties of this protistan parasite.

3.3 Amoebapores and related proteins

In the lysosome-like granular vesicles of *E. histolytica* is found a family of small proteins, amoebapores, that are cytolytic towards human host cells,

display potent antibacterial activity, and cause ion channel formation in artificial membranes (for a review see Leippe (1997)). Three amoebapore isoforms have been isolated and biochemically characterised, and their primary structure has been elucidated by molecular cloning of the genes encoding their precursors (Leippe *et al.*, 1991; Leippe *et al.*, 1992; Leippe *et al.*, 1994b). These membrane-permeabilising polypeptides are discharged by *E. histolytica* into bacteria-containing phagosomes to combat growth of engulfed microorganisms (Andrä *et al.*, 2003). Because of their potent cytolytic activity against human cells in vitro (Berninghausen and Leippe, 1997; Leippe *et al.*, 1994a), amoebapores have been viewed as a crucial element of the machinery use by the parasite to kill host cells. Trophozoites of *E. histolytica* lacking the major isoform amoebapore A, whether through antisense inhibition of translation (Bracha *et al.*, 1999) or epigenetic silencing of the gene (Bracha *et al.*, 2003), became avirulent demonstrating that this protein plays a key role in pathogenesis. Relatives of these protistan polypeptides are found in granules of porcine and human cytotoxic lymphocytes where they are termed NK-lysin and granulysin, respectively. All of these polypeptides are 70-80 amino acids in length and are characterised by a compact alpha-helical, disulphide-bonded structure known as the saposin-like fold. The structures of the amoebic and mammalian polypeptides have been solved and compared (Anderson *et al.*, 2003; Hecht *et al.*, 2004; Leippe *et al.*, 2005; Liepinsh *et al.*, 1997). The biological activities have also been measured in parallel (Bruhn *et al.*, 2003; Gutsmann *et al.*, 2003) to evaluate the similarities and differences of these effector molecules from organisms whose evolutionary paths diverged very early. As they are active against both prokaryotic and eukaryotic target cells, they may be viewed as broad-spectrum effector molecules.

In the genome of *E. histolytica*, 16 genes coding for putative saposin-like proteins (SAPLIPs) were identified. All of these genes are transcribed by cells growing in axenic culture (Winkelmann *et al.*, 2006). Like amoebapores, the predicted proteins all contain one C-terminal SAPLIP domain and (with one

exception) a putative signal peptide (Table 6). As a transmembrane domain is not apparent in these proteins, it may well be that they are secretory products stored in the cytoplasmic vesicles and act synergistically with the amoebapores. However, only four of them have a similar size to amoebapores, the others being considerably larger (up to 1009 residues). At present, it is not clear whether these larger gene products represent precursor molecules that are processed further. None of the novel SAPLIPs contain the conserved unique histidine residue at the C-terminus that is a key residue for the pore-forming activity of amoebapores (Andrä and Leippe, 1994; Hecht *et al.*, 2004; Leippe *et al.*, 2005). Indeed, it has recently been shown that recombinant SAPLIP3 has no pore-forming or bactericidal activity, although it does cause membrane fusion in vitro (Winkelmann *et al.*, 2006). This is in agreement with the experimental evidence for only three pore-forming entities being present in trophozoite extracts. Therefore, it is most likely that the three amoebapores are the sole pore-forming molecules of the parasite. However, the lipid-interacting activity present in all SAPLIP proteins (Munford *et al.*, 1995) and a function that helps to kill bacterial prey may well characterise all members of the amoebapore/SAPLIP superfamily of this voraciously phagocytic cell.

3.4 Antioxidants

Entamoeba histolytica trophozoites usually reside and multiply within the human gut, which constitutes an anaerobic or microaerophilic environment. However, during tissue invasion, the amoebae are exposed to an increased oxygen pressure and have to eliminate toxic metabolites such as reactive oxygen or nitrogen species (ROS/RNS) produced by activated phagocytes during the respiratory burst. *E. histolytica* lacks a conventional respiratory electron transport chain that terminates in the reduction of O_2 to H_2O. However, *E. histolytica* does respire and tolerates up to 5% oxygen in the gas phase (Band and Cirrito, 1979; Mehlotra, 1996; Weinbach and Diamond, 1974). Thus, *E. histolytica* trophozoites must use different antioxidant enzymes for the removal of ROS, RNS and oxygen (Figure 5).

Among the enzymes in the first line of oxidative defence are superoxide dismutases (SODs), which are metalloproteins that use copper/zinc (Cu/Zn), manganese (Mn) or iron (Fe) as metal cofactors. SODs catalyse the dismutation of superoxide radical anions to form H_2O_2 and O_2 (Fridovich, 1995). Analysis of the *E. histolytica* genome revealed only a single gene coding for a FeSOD and no sequences encoding MnSOD or Cu/ZnSOD. This reflects the situation found in most protistan parasites and is consistent with biochemical studies previously performed on *E. histolytica* lysates (Tannich *et al.*, 1991a).

Entamoeba histolytica lacks the tripeptide glutathione (Fahey *et al.*, 1984), which constitutes the major low molecular weight thiol found in almost all aerobic cells (Sies, 1999). Instead, *E. histolytica* uses cysteine as its principal low molecular weight thiol (Ariyanayagam and Fairlamb, 1999; Fahey *et al.*, 1984; Nozaki *et al.*, 1999). As expected, coding sequences for enzymes that use glutathione as a cofactor, such as glutathione-S-transferase, glutathione-dependent peroxidase, glutathione reductase or glutaredoxin, are all absent from the *E. histolytica* genome. In addition, genes encoding catalases and peroxidases are also missing, as previously suggested (Sykes and Band, 1977; Weinbach and Diamond, 1974).

Other genes were identified that code for proteins involved in detoxification of H_2O_2, including one with homology to rubrerythrin. Rubrerythrin is a non-haeme iron protein thought to be able to reduce H_2O_2 as part of an oxidative stress protection system (Weinberg *et al.*, 2004). So far, the nature of its redox partner is unknown in *E. histolytica* and it remains to be determined whether protection against oxidative stress is indeed its main function. Another group of H_2O_2-detoxifying proteins identified in *E. histolytica* are peroxiredoxins. Peroxiredoxins are known from a wide variety of organisms. They are able to reduce H_2O_2 as well as peroxynitrite with the use of electrons provided by thiols. In addition to involvement in the detoxification of reactive oxygen

species peroxiredoxins seem to play a role in other processes such as signalling and differentiation (Hofmann *et al.*, 2002; Rhee *et al.*, 2005; Wood *et al.*, 2003a,b). All peroxiredoxins contain a conserved cysteine residue that undergoes a cycle of peroxide-dependent oxidation and thiol-dependent reduction during the reaction. The whole protein family can be divided into three classes based on the number and position of active site Cys residues (2-Cys, atypical 2-Cys, and 1-Cys peroxiredoxins; Wood *et al.*, 2003a,b). In *E. histolytica* five different genes coding for peroxiredoxins were identified (Prx1-5). They all belong to the 2-Cys peroxiredoxin family. Four of them (Prx1-4) share 98% sequence identity and have an unusual N-terminal Cys-rich repeat (KECCKKECQEKECQEKECCC) of unknown function. In contrast, the fifth peroxiredoxin (Prx5) lacks the cysteine-rich N-terminal extension and shares only 30% identity with Prx1-4. Biochemical studies have shown that *E. histolytica* peroxiredoxins are able to detoxify H_2O_2 and cumene hydroperoxide (Bruchhaus *et al.*, 1997; Poole *et al.*, 1997). Moreover, up-regulation of peroxiredoxin and FeSOD was associated with metronidazole resistance in cultured *E. histolytica* trophozoites (Samarawickrema *et al.*, 1997; Wassmann *et al.*, 1999).

Reactions catalysed by peroxiredoxins are dependent on the presence of physiological thiols like thioredoxin (Rhee *et al.*, 2005; Wood *et al.*, 2003b). Thioredoxins are small proteins involved in thiol-redox processes (Holmgren, 2000). They contain two redox-active site cysteine residues of the motif CXXC (Watson *et al.*, 2004). Five genes coding for classical cytoplasmic thioredoxins were identified in the *E. histolytica* genome (Trx1-5). These thioredoxins have a length of 103-114 amino acids and share 25 – 47 % sequence identity. Trx1-3 have identical active site motifs of the sequence WCGPC, whereas the active sites of Trx4 and Trx5 have the sequences SCPSC and WCKDC, respectively. In addition, another five thioredoxin-related proteins were identified (Trx6-10). All have a signal sequence of 15 to 19 amino acid residues and the active site motif WCGHC, which is also known from the active site of protein disulphide

isomerases. However, in contrast to the latter group of enzymes, the *E. histolytica* thioredoxin-related molecules contain only one rather than two active-site motifs and only two of the proteins have an ER membrane retention signal (Freedman *et al.*, 2002). Thus it remains to be determined whether the thioredoxin-related molecules of *E. histolytica* do constitute protein disulphide isomerases or whether they undertake other functions within the cell.

Thioredoxins are kept in the reduced state by the enzyme thioredoxin reductase which catalyses the reduction of oxidised thioredoxin by NADPH using FAD and its redox-active disulphide (Nakamura, 2005). Two different genes with homology to thioredoxin reductases have been previously described from *E. histolytica* (thioredoxin reductase (TrxR) and NADPH:flavin oxidoreductase (p34)). They share about 87% sequence identity and both contain the two conserved sequence motifs forming the FAD and NAD(P)H binding domains. p34 was shown to catalyse the NADPH-dependent reduction of oxygen to H_2O_2 as well as of disulphides like DTNB and cystine (Bruchhaus *et al.*, 1998; Lo and Reeves, 1980). Therefore, in addition to disulphide reductase activity the enzyme has H_2O_2-forming NADPH oxidase activity. It was also shown that p34 can transfer reducing equivalents to peroxiredoxin, converting the protein from its non-active, oxidised form back into its active, reduced form (Bruchhaus *et al.*, 1997). However, it is unlikely that peroxiredoxin is directly reduced by p34 *in vivo*. It is more likely that *E. histolytica* contains a classical thioredoxin redox system consisting of thioredoxin reductase, thioredoxin and peroxiredoxin (Poole *et al.*, 1997).

In addition to genes coding for proteins with homology to thioredoxin reductase, four other gene families were identified that encode various flavoproteins. One of these families includes 4 members that have between 53% and 61% sequence identity to A-type flavoproteins (flavorubredoxin/flavodiiron). A-type flavoproteins belong to a large family of enzymes that are widespread among anaerobic and facultatively anaerobic

prokaryotes. In addition to bacteria, homologous genes are also found in the genomes of the pathogenic amitochondriate protistan parasites *Trichomonas vaginalis* and *Giardia intestinalis* (Andersson *et al.*, 2003; Sarti *et al.*, 2004). The A-type flavoproteins are made up of two independent structural modules. The N-terminal region forms a metallo-beta-lactamase-like domain, containing a non-haeme di-iron site, whereas the C-terminal region is a flavodoxin-like domain, containing one FMN moiety. These enzymes have significant nitric oxide reductase activity (Gomes *et al.*, 2002; Sarti *et al.*, 2004). For *Escherichia coli* it is known that the nitric oxide reductase (FIRd) receives electrons from a NADH:oxidoreductase (FIRd-red). Consistent with that situation, the *E. histolytica* genome contains a gene encoding an NADH oxidase with 25% sequence identity to several bacterial FIRd-reds.

The three other *E. histolytica* gene families with homology to iron-sulphur flavoproteins (families B-D) are characterised by the presence of a flavodoxin-like domain forming a typical FMN binding site. Family B and family C consist of three members each, which share sequence identity of 42% and 46%, respectively. Family D consists of two members, which share only 33% sequence identity. At present, the function of the various flavodoxin-like molecules remains to be determined and deserves to be investigated fully, particularly as to whether they do indeed have antioxidant capacity.

4. METABOLISM

Biochemical analysis of *E. histolytica* metabolism has a long history (Reeves, 1984), dating back to shortly after the development of culture media that allowed the generation of substantial numbers of axenic cells. The genome sequence has confirmed most of the predicted metabolic pathways shown biochemically to be present or absent in *E. histolytica* in the past. As with most parasites, secondary loss of biosynthetic pathways is a recurring theme. However, a few surprises have also been uncovered. Every single enzyme involved in metabolism cannot realistically be discussed in this review. In this

section the only the major energy generating and biosynthetic aspects of metabolism will be covered. Enzyme names, EC numbers and accession numbers are given in the the supplementary table for this section.

4.1 Energy Metabolism

4.1.1 Glycolysis

E. histolytica lacks a functional tricarboxylic acid (TCA) cycle and oxidative phosphorylation. It is not able to convert organic substrates such as glucose into H_2O and CO_2, but has to rely on the energy generated by various types of substrate level phosphorylation (Reeves, 1984). Glycolysis is the major pathway of ATP generation, but in addition the genome project has identified a number of genes that could result in more ATP generation through the catabolism of amino acids. These enzymes will be described further below. As *E. histolytica* lacks compartmentalised energy generation, it has been classified as a type I amitochondriate protist (Martin and Müller , 1998) in contrast to the type II amitochondriate protists containing hydrogenosomes, such as *Trichomonas vaginalis*. Nevertheless, it does contain a mitochondrial remnant, the mitosome (see section 8).

In *E. histolytica*, glycolysis appears to be localised in the cytosol. This is in contrast to trypanosomes where a major part is carried out in the glycosomes (Parsons, 2004) and the pathway is regarded as a potential target for chemotherapy (Opperdoes and Michels, 2001). The kinetic properties of recombinant *E. histolytica* glycolysis enzymes have recently been studied by Saavedra *et al.* (2005). Their analysis suggested that fructose-1,6-bisphosphate aldolase, phosphoglycerate mutase, glyceraldehyde-3-phosphate dehydrogenase, and pyruvate phosphate dikinase might be regulating the glycolytic flux.

4.1.1 (a) Hexokinases

Glucose taken up by *E. histolytica* is phosphorylated by two hexokinase (EC 2.7.1.1) isoenzymes (Hxk1 and Hxk2). The two *E. dispar* isoenzymes are shifted towards a slightly more basic pI, which is the basis of the classical biochemical method for distinguishing *E. histolytica* from *E. dispar* by starch gel electrophoresis (Farri *et al.*, 1980). The pI differences among the two *E. histolytica* isoforms (Ortner *et al.*, 1995) and between the two species (Ortner *et al.*, 1997b) are the result of genetic differences that lead to different amino acid sequences and charge differences. Hxk1 phosphorylates glucose and mannose, while Hxk2 phosphorylates mainly glucose and is much less active with mannose as a substrate (Kroschewski *et al.*, 2000).

4.1.1 (b) Glucose-6-phosphate isomerase

Glucose 6-phosphate is converted to fructose 6-phosphate by glucose-6-phosphate isomerase (EC 5.3.1.9). The genome has 2 genes for this enzyme, which code for proteins that differ only by a single insertion/deletion of 7 amino acid residues. Glucose-6-phosphate isomerase is another of the enzymes for the classical differentiation of *Entamoeba* zymodemes by starch gel electrophoresis (Sargeaunt, 1987).

4.1.1 (c) Phosphofructokinases

The main phosphofructokinase activity in *E. histolytica* is PPi-dependent (EC 2.7.1.90; Reeves *et al.*, 1976). There is a single gene (Deng *et al.*, 1998) encoding this 60 kDa enzyme. The gene is a candidate for lateral transfer from bacteria (Loftus *et al.*, 2005) (see section 10). The enzyme is expressed at a tenfold higher level and displays about tenfold higher activity than a second phosphofructokinase of 48 kDa (XP_653373) (Chi *et al.*, 2001). The substrate specificity of the smaller enzyme is disputed. Whereas Bruchhaus *et al.* (1996) reported that this minor enzyme also used PPi as phosphate donor, Chi *et al.* (2001) found only an ATP-dependent activity. The 48 kDa and 60 kDa enzymes are highly divergent with less than 20% sequence identity. Interestingly, the specificity of the 60 kDa phosphofructokinase can be changed

from PPi to ATP by mutation of a single amino acid residue (Chi and Kemp, 2000). The authors concluded that ATP rather than PPi was the primordial high energy compound. In the genome, there are two additional genes encoding isoforms of the 48 kDa enzyme, which have not been studied at the protein level.

4.1.1 (d) Fructose-1,6-bisphosphate aldolase

Fructose 1,6-bisphosphate is cleaved to glyceraldehyde 3-phosphate and dihydroxyacetone 3-phosphate by fructose-1,6-bisphosphate aldolase (EC 4.1.2.13). The enzyme, a Class II aldolase (Marsh and Lebherz, 1992) has been cloned (XP_650373) and exhibits strong sequence similarity to eubacterial aldolases (Sanchez *et al.*, 2002). A second gene (XP_655966) encodes a protein differing from the first by a single deletion of 28 amino acids flanked by short divergent stretches. These bacterial-type aldolases are also found in *Trichomonas vaginalis*, *Giardia intestinalis* and other protists (Sanchez *et al.*, 2002). *E. histolytica* has no gene coding for a Class I aldolase like those found in animals, which might make aldolase an interesting target for chemotherapy.

4.1.1 (e) Triose-phosphate isomerase

Triose-phosphate isomerase (EC 5.3.1.1) converts dihydroxyacetone 3-phosphate into glyceraldehyde 3-phosphate. The gene was previously cloned (Landa *et al.*, 1997), and is highly similar to the annotated gene product. This dimer-forming enzyme represents the first *E. histolytica* protein for which the structure has been solved by X-ray crystallography (Rodriguez-Romero *et al.*, 2002).

4.1.1 (f) Glyceraldehyde-3-phosphate dehydrogenase

Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) oxidises and phosphorylates glyceraldehyde 3-phosphate to 1,3-bisphosphoglycerate in two coupled reactions using NAD^+ as cofactor (Reeves, 1984). The genome project revealed five putative genes, three of which encode the identical protein

sequence of 36.0 kDa and a predicted pI of 7.04. The fourth gene product, XP_648981, differs from these three only by a 13 amino acid deletion, while XP_650370 is a clearly distinct 34.8 kDa isoform with a lower predicted pI of 5.80. Interestingly, the isoforms XP_650356 and XP_650370 of different pI are encoded within the same contig.

4.1.1 (g) Phosphoglycerate kinase

Phosphoglycerate kinase has an unusual substrate (Reeves and South, 1974), transferring the high energy phosphate group from 3-phosphoglyceroyl phosphate to GDP leading to the formation of GTP (EC 2.7.2.10). There is one candidate gene encoding a 45 kDa protein.

4.1.1 (h) Phosphoglycerate mutase

Phosphoglycerate mutase (Reeves, 1984) isomerises 3-phosphoglycerate to 2-phosphoglycerate (EC 5.4.2.1). Five divergent putative genes for this enzyme are found in the genome. Two gene products of 62 kDa were classified as 2,3-bisphosphoglycerate-independent phosphoglycerate mutases (XP_649031 and XP_654182); they differ only at their C-termini and display significant similarity to bacterial phosphoglycerate mutases. The three other genes are very divergent. XP_ 651808 was identified as a candidate for lateral gene transfer (Loftus *et al.*, 2005) (see section 10). The remaining two gene products XP_649053 and XP_657284 are related to genes found in both prokaryotes and eukaryotes.

4.1.1 (i) Enolase (2-phosphoglycerate dehydratase)

Enolase (EC 4.2.1.11) converts 2-phosphoglycerate to phosphoenolpyruvate. The gene has been cloned (Beanan and Bailey, 1995) and the protein characterised (Hidalgo *et al.*, 1997) previously. The 47 kDa gene product is a typical eukaryotic enolase (XP_649161). A carboxy-terminally truncated incomplete ORF is also found.

4.1.1 (j) Pyruvate, orthophosphate dikinase and pyruvate kinase

In *E. histolytica*, both activities forming ATP and pyruvate from phosphoenolpyruvate have been found. The exergonic pyruvate kinase reaction uses ADP (Saavedra *et al.*, 2004), and the pyruvate, orthophosphate dikinase uses AMP and PPi in a slightly endergonic reaction (Varela-Gomez *et al.*, 2004). The dikinase activity is found in C4 plants where it is involved in phosphoenolpyruvate generation for gluconeogenesis. In *E. histolytica* it was discovered long before the pyruvate kinase (Reeves, 1968).

The cloning of pyruvate, orthophosphate dikinase (EC 2.7.9.1) was reported by two groups. The published sequences (Bruchhaus and Tannich, 1993; Saavedra Lira *et al.*, 1992) are highly similar or identical to XP_657332 and XP_654666. In addition there are two shorter related open reading frames.

In the genome 3 putative pyruvate kinase genes (EC 2.7.1.40) have been identified. The 3 are identical except for an amino-terminal deletion in XP_648240 and an internal deletion in XP_653635.

4.1.1 (k) Pyruvate:ferredoxin oxidoreductase (PFOR) and ferredoxin

PFOR (EC 1.2.7.1) is an enzyme of major importance to *E. histolytica*, as the parasite lacks NAD^+-dependent pyruvate dehydrogenase and pyruvate decarboxylase (Reeves, 1984). No evidence for the latter two genes was found in the genome, confirming the biochemical results. PFOR oxidatively decarboxylates pyruvate to acetyl-CoA. The electrons are transferred to ferredoxin which, in its reduced form, can activate and reduce metronidazole, the major anti-amoebic drug (Müller , 1986). The activated form of metronidazole can potentially react with a number of biomolecules and is able to cleave the parasite DNA. In human cells, metronidazole is not activated and is much less toxic. In *T. vaginalis*, down-regulation of PFOR is one mechanism of producing metronidazole resistance (Kulda, 1999); however PFOR expression appears unaltered in partially resistant *E. histolytica*

(Samarawickrema *et al.*, 1997; Wassmann *et al.*, 1999). All eukaryotic PFOR genes, including that of *E. histolytica*, appear to have been acquired during an ancient lateral gene transfer event from bacteria (Horner *et al.*, 1999; Rotte *et al.*, 2001). There are two putative PFORs in the *E. histolytica* genome displaying minor sequence differences.

The genome contains 7 ferredoxin genes in total with 5 quite divergent sequences. All are related to eubacterial and archaeal ferredoxins (Nixon *et al.*, 2002). The gene pairs XP_655183 / XP_655182 and XP_654311 / XP_652694 are identical. The other three gene products represent more divergent open reading frames. The deduced proteins have similar molecular masses, between 6.1 kDa and 8.8 kDa, and different predicted isoelectric points between 4.2 and 8.6.

4.1.1 (l) Acetyl-CoA synthetase (acetate thiokinase)

The normal fate of acetyl-CoA in mitochondriate organisms is entry into the tricarboxylic acid cycle. However this pathway is absent from *E. histolytica*. Instead, the cleavage energy of the thioester bond of acetyl-CoA can be used to generate one ATP molecule. One of the known acetyl-CoA synthetases generates ATP from ADP and Pi (EC 6.2.1.13). Such an enzyme has been characterised by Reeves *et al.* (1977) and cloned (Field *et al.*, 2000), and reported to be a 77 kDa protein. The common acetyl-CoA synthetase activity that produces ATP from AMP and PPi (EC 6.2.1.1) appears to be absent in *E. histolytica*.

4.1.1 (m) Aldehyde and alcohol dehydrogenases

The *E. histolytica* genome encodes a complex system of alcohol and/or aldehyde dehydrogenases. In total, there are 25 predicted genes, 3 of which are on the list of lateral gene transfer (LGT) candidates.

Alcohol dehydrogenase ADH1 was the first alcohol dehydrogenase to be characterised in *E. histolytica* (Reeves *et al.*, 1971) and is a NADPH-dependent enzyme (EC 1.1.1.2). The gene was previously cloned (Kumar *et al.*, 1992); in the genome three genes are almost identical to that sequence, while one (XP_652772) has 67% identity.

Fermentation in *E. histolytica* uses the bifunctional alcohol dehydrogenase / aldehyde dehydrogenase NADH-dependent enzyme ADH2, which belongs to the ADHE family and has both alcohol dehydrogenase and aldehyde dehydrogenase activities (Lo and Reeves, 1978). Under anaeobic conditions, reduction of the acetyl-CoA generated by PFOR to ethanol is one way to regenerate the NAD+ used by glyceraldehyde-3-phosphate dehydrogenase. ADH2 first reduces acetyl-CoA to an enzyme-bound hemiacetal which is then hydrolysed to acetaldehyde (EC 1.2.1.10) and further reduced to ethanol (EC 1.1.1.1). If the enzyme is also able to work in the reverse direction, *E. histolytica* would be able to generate acetyl-CoA and energy from ethanol in the presence of oxygen. This would explain older reports of ethanol stimulated oxygen uptake in *E. histolytica* (Weinbach and Diamond, 1974). The enzyme is closely related to AdhE from *E. coli* and other bacteria (Reid and Fewson, 1994) and there is strong support for its aquisition by LGT (Andersson *et al.*, 2006; Field *et al.*, 2000; Loftus *et al.*, 2005) (see section 10). Like its bacterial homologue, ADH2 appears to form helical rods that sediment with membrane fractions (Avila *et al.*, 2002). Two groups have previously cloned ADH2 (Bruchhaus and Tannich, 1994; Yang *et al.*, 1994), and in total the genome contains 5 full-length ADH2 genes and one that is truncated. All share between 98% and 100% sequence identity.

In total, there are 11 alcohol dehydrogenase ADH3 genes in the genome, two of which been reported previously (Kimura *et al.*, 1996; Rodriguez *et al.*, 1996). The recombinant enzyme characterised by Rodriguez *et al.* (1996) was NADPH-specific, like ADH1. There are five genes similar to these previously

reported sequences. The rest of the ADH3 sequences fall into two groups of 3 similar sequences. All 11 ADH3 sequences are between 44% and 100% identical on the amino acid level. XP_649823 was originally on the list of LGT candidates (Loftus *et al.*, 2005), and a similarity to ADH3 sequences of gram-negative bacteria had been noted before (Nixon *et al.*, 2002). However a related sequence is now known to exist in *T. vaginalis* also (see section 10).

The genome encodes three additional distinct alcohol dehydrogenases. XP_656535 is a putative Zn-containing enzyme, and is on the list of LGT candidates. XP_652753 has been annotated as a Fe-containing alcohol dehydrogenase and XP_652262 simply as putative alcohol dehydrogenase.

One NADPH-dependent aldehyde dehydrogenase encoding gene (ALDH1) is present and was reported previously (Zhang *et al.*, 1994).

4.1.2 Energy storage: the glycogen metabolism

E. histolytica uses glycogen as its major energy store. Glycogen is a polymer of alpha-1,4-linked glucose chains with alpha-1,6 branch points, which in *E. histolytica* has a compact structure as suggested by branch points every 5-6 glucose residues (Bakker-Grunwald *et al.*, 1995). The cytoplasm of trophozoites contains numerous glycogen granules which were first observed by electron microscopy (Rosenbaum and Wittner, 1970) and later characterised biochemically (Takeuchi *et al.*, 1977). A glycogen phosphorylase activity (EC 2.4.1.1), associated with the glycogen granules, generates glucose 1-phosphate from orthophosphate and the linear portion of various glucopolysaccharides (Werries and Thurn, 1989). The genome contains at least 6 putative full-length and truncated genes encoding glycogen phosphorylases, two of which were cloned by Wu and Müller (2003). These authors noted a marked sequence divergence in those regions of the enzymes involved in regulation by phosphorylation and concluded that classical regulation by phosphorylation may not occur.

Glycogen phosphorylase degrades the linear chains only down to the alpha-1,6 branch points. The remaining core molecule is called limit dextrin. Degradation can proceed further with the help of a debranching enzyme that has been purified (Werries *et al.,* 1990). It exhibits activities of both amylo-1,6-glucosidase (EC 3.2.1.33) and 4-alpha-glucanotransferase (EC 2.4.1.25). The genome contains two genes putatively encoding a full-length (XP_653608) and a truncated glycogen debranching enzyme. The deduced molecular mass of the large protein is 166 kDa which corresponds to the biochemical data (Werries *et al.*, 1990).

Glucose 1-phosphate is isomerised to glucose 6-phosphate by phosphoglucomutase (EC 5.4.2.2) before entering the glycolytic pathway. The isoelectric points of the phosphoglucomutases from *E. histolytica* and *E. dispar* differ which was exploited for differentiation of the two species by starch gel electrophoresis (Sargeaunt *et al.,* 1978). The migration properties are reproduced by recombinant enzymes and are the result of primary sequence differences (Ortner *et al.,* 1997a). *E. histolytica* has one gene coding for this important enzyme, and in addition there are two distantly related members of the phosphoglucomutase / phosphomannomutase family.

Genes encoding the enzymes involved in glycogen biosynthesis in *E. histolytica* have been identified: a glycogen synthase (EC 2.4.1.11) of 155 kDa and two putative branching enzymes (EC 2.4.1.18). The glycogen precursor UDP-glucose is generated from UTP and glucose 1-phosphate by UTP:glucose-1-phosphate uridylyltransferase (EC 2.7.7.9). Two UTP-hexose-1-phosphate uridyltransferases have been characterised biochemically, a larger glucose 1-phosphate-specific enzyme of 45 kDa and a less specific enzyme of 40 kDa reported to use both galactose 1-phosphate and glucose 1-phosphate (Lobelle-Rich and Reeves, 1983). The genome contains one larger open reading frame encoding a putative UTP:glucose-1-phosphate uridylyltransferase of 54.7 kDa

and two smaller ones encoding enzymes of 46.3 kDa with high similarity identified as UTP:N-acetyl-glucosamine-1-phosphate uridyltransferases. These enzymes are interesting in that they could possibly be involved in the activation of N-acetyl-glucosamine 1-phosphate as a precursor of the chitin cyst wall.

4.1.3 Catabolism of sugars other than glucose

4.1.3 (a) Activation of fructose and galactose for glycolysis

Neither Hxk1 nor Hxk2 can use fructose or galactose as a substrate, but there are 2 genes encoding bacterial-type enzymes that may do so, a 33 kDa fructokinase, which is one of the candidates for lateral gene transfer to the *E. histolytica* lineage (see section 10), and a 43 kDa galactokinase. The fructokinase groups with bacterial fructose 6-kinases (EC 2.7.1.4), and the galactokinase groups with galactose 1-kinases (EC 2.7.1.6). This substrate specificity has been noted before (Reeves, 1984). Fructose 6-phosphate enters as an intermediate of the glycolytic pathway (see 4.1.1 (c)). As described above (4.1.2), galactose 1-phosphate can be activated to UDP-galactose (Lobelle-Rich and Reeves, 1983), and then epimerised to UDP-glucose by UDP-glucose 4-epimerase (EC 5.1.3.2) (Reeves, 1984). In the genome, a single candidate 38 kDa ORF for the latter enzyme has been identified. The UDP-bound glucose can then be used either for the synthesis of glycogen or fed into the glycolysis pathway via glucose 1-phosphate and glucose 6-phosphate. This efficient pathway allows *E. histolytica* to grow on galactose instead of glucose (Reeves, 1984).

4.1.3 (b) Anomerisation of aldoses

The 1-position in the pyranose form of aldoses has a hydroxyl group that can be in either the α- or β-configuration. These forms can be interconverted by means of an aldose 1-epimerase (EC 5.1.3.3) an enzyme that has recently been characterised (Villalobo *et al.*, 2005). There is a single gene encoding this product.

4.1.3 (c) Activation of pentoses

Two gene candidates encoding pentose-activating enzymes have been identified in the *E. histolytica* genome: a 35 kDa ribokinase (EC 2.7.1.15) and a 56 kDa xylulokinase (EC2.7.1.17). The latter is another bacterial-type sequence putatively acquired by lateral gene transfer.

4.1.3 (d) Interconversion of hexoses and pentoses

The pathway of interconversion between hexoses and pentoses in *E. histolytica* was described many years ago (Reeves, 1984; Susskind *et al.*, 1982). A transketolase (EC 2.2.1.1) converts fructose 6-phosphate and glyceraldehyde 3-phosphate into xylulose 5-phosphate and erythrose 4-phosphate. Erythrose 4-phosphate and dihydroxyacetone phosphate are condensed by the glycolytic enzyme fructose-1,6-bisphosphate aldolase to sedoheptulose 1,7-bisphosphate, an extended substrate specificity of the aldolase. Phosphofructokinase then is able to remove a phosphate group forming diphosphate and sedoheptulose 7-phosphate. This molecule and glyceraldehyde 3-phosphate are then converted by transketolase to the pentoses ribose 5-phosphate and xylulose 5-phosphate. A transaldolase activity is absent (Reeves, 1984) consistent with there being no such gene in the genome. In contrast, 7 gene products were identified as likely transketolases: three highly similar proteins of 73 kDa and four truncated versions.

4.2 Amino acid catabolism

4.2.1 General features

As discussed above, glycolysis under anaerobic conditions can use only part of the energy contained in glucose for ATP generation. *E. histolytica* is capable not only of taking up amino acids (Reeves, 1984), but also using them for the generation of energy, as suggested by Zuo and Coombs (1995). The genome has revealed a number of unusual genes, often with bacterial affinities, coding for enzymes of amino acid catabolism (Anderson and Loftus, 2005).

In many cases, the degradation of amino acids starts with a transamination reaction (EC 2.6.1. -) generating a 2-ketoacid. The *E. histolytica* genome has five ORFs identified as aminotransferases. These ORFs are distinct from each other with the exception of XP_655090 and XP_655099, which differ only by one insertion and are LGT candidates. So far there is no enzymological data on this group of enzymes, so their substrate specificities in *E. histolytica* are unknown.

Both amino acid degradation and glycolysis have 2-ketoacids as intermediates. Pyruvate is one common intermediate, as amino acid degradation can produce either pyruvate or other 2-ketoacids. PFOR (see 4.1.1 (k)) is known to have a relaxed specificity, and in addition to pyruvate it can oxidatively decarboxylate 2-ketobutanoate, oxaloacetate, and 2-ketoglutarate (Samarawickrema *et al.*, 1997). The reaction generates CoA-thioesters with the potential of producing one ATP per molecule.

The amino acids asparagine, aspartate, serine, alanine, tryptophan, cysteine, threonine, methionine, glutamine, and glutamate can all be transformed into one of these 2-ketoacids in one or very few steps. This underlines the major importance of the PFOR in the energy metabolism of *E. histolytica*. The enzyme is indispensable, and as it always generates reduced ferredoxin it will always activate metronidazole. Consequently, it would be very difficult for *E. histolytica* to become resistant to metronidazole.

4.2.1 Aspartate and asparagine

E. histolytica takes up asparagine and aspartate in the presence or absence of glucose (Zuo and Coombs, 1995). Four putative asparaginases (EC 3.5.1.1) are found in the genome. Three are identical and share only 48% amino acid identity with the fourth (XP_656586). Asparaginase mediates the formation of aspartate from aspargine by releasing ammonia. The predicted sequences appear to possess a signal sequence, as suggested by TargetP program

(www.cbs.dtu.dk/services/TargetP/), which is reminiscent of a periplasmic isotype (EcA, type II) (Swain *et al.*, 1993) that is up-regulated under anaerobic and carbon-restricted conditions (Cedar and Schwartz, 1967).

Aspartate can be converted to fumarate and ammonia by aspartate ammonia-lyase (aspartase, EC 4.3.1.1). Addition of a water molecule by fumarase (EC 4.2.1.2) produces malate. The genome encodes a putative fumarase that is related to bacterial Class I fumarases. The aspartase is a member of the bacterial Class II fumarase / aspartase protein family (Woods *et al.*, 1988), and also on the list of LGT candidates.

Aspartate is also decomposed into oxaloacetate and ammonia by aspartate aminotransferase, with the concomitant production of glutamate from 2-oxoglutarate. Oxaloacetate is then converted into malate via malate dehydrogenase (EC 1.1.1.37) and, since *E. histolytica* lacks both a functional TCA cycle and phosphoenolpyruvate carboxykinase, the malate generated can be oxidatively decarboxylated to pyruvate by malic enzyme (EC 1.1.1.39). Both of these enzymes are present in *E. histolytica*. Two very similar genes have been identified as encoding malic enzyme and are LGT candidates.

4.2.2 Serine, threonine

Serine and threonine are also taken up by *E. histolytica* in the presence and absence of glucose (Zuo and Coombs, 1995). Serine can be deaminated by the pyridoxal phosphate-dependent serine dehydratase (L-serine ammonia-lyase, EC 4.3.1.17) to pyruvate and ammonia. The enzyme was characterised by Takeuchi *et al.* (1979) who showed that addition of serine to the culture medium stimulated oxygen consumption. In an analogous reaction, threonine dehydratase (threonine ammonia-lyase, EC 4.3.1.19) breaks down threonine to 2-oxobutanoate. Both ketoacids can then be oxidised by PFOR to acetyl-CoA or proprionyl-CoA. Both catabolic reactions can be carried out by the same enzyme, as has been shown in yeast for example (Ramos and Wiame, 1982). In

the *E. histolytica* genome annotation, four gene products have been annotated as threonine dehydratases, but none as serine dehydratase. XP_650405 and XP_652480 are identical while XP_655614 and XP_657171 share 95% and 37% identity with the others, respectively. The exact substrate specificities of these 4 putative serine / threonine dehydratases have not been reported.

Degradation of serine via the non-phosphorylated serine pathway, by the sequential reactions of L-serine: pyruvate aminotransferase (EC 2.6.1.51), D-glycerate dehydrogenase (EC 1.1.1.29), and D-glycerate kinase (EC 2.7.1.31) (Snell, 1986) results in the glycolytic intermediate 3-phosphoglycerate. The genome encodes several putative aminotransferases (see section 4.2.1), but it is not yet known if serine is among their substrates. An unusual bacterial-type NADPH-dependent D-glycerate dehydrogenase was characterised by Ali *et al.* (2003) and there are 2 genes encoding D-glycerate dehydrogenases, one of which (XP_648124) is among the weaker LGT candidates (see section 10). The genome also contains 2 genes encoding identical glycerate kinases. The enzyme has recently been characterised by Ali and Nozaki (unpublished).

4.2.3 Methionine, homocysteine and cysteine

Methionine γ-lyase (EC 4.4.1.11) decomposes methionine to methanethiol (mercaptomethane), ammonia, and 2-oxobutanoate. In *E. histolytica*, two methionine γ–lyases, EhMGL1 and EhMGL2, of similar molecular weights have been characterised (Tokoro *et al.*, 2003). These two isoenzymes show marked differences in substrate specificity, isoelectric point, enzymological and biochemical parameters (Tokoro *et al.*, 2003). Both enzymes can also act on other amino acids. In addition to degrading methionine, both EhMGL1 (pI 6.01) and EhMGL2 (pI 6.63) can convert homocysteine to hydrogen sulphide, ammonia and 2-oxobutanoate. EhMGL2 also decomposes cysteine to hydrogen sulphide, ammonia, and pyruvate, whereas EhMGL1 is only weakly active against cysteine. Decomposition of homocysteine by methionine γ-lyase is essential since this parasite lacks the other known enzymes capable of

destroying this toxic amino acid. In the genome, three open reading frames correspond to EhMGL1 and one to EhMGL2. So far, the only eukaryotes known to possess methionine γ–lyases are *E. histolytica* and *T. vaginalis* (Lockwood and Coombs, 1991). As the enzymes are absent from the human host and important for the generation of metabolic energy, they could be targets for chemotherapy (Coombs and Mottram, 2001; Tokoro *et al.*, 2003).

In addition to serving as a source of metabolic energy, another important role of methionine is as a donor of methyl groups via S-adenosylmethionine synthetase (synonymous with methionine adenosyltransferase, EC 2.5.1.6). Seven gene candidates were identified, four full-length and three truncated. The S-adenosylhomoserine left after the transfer of the activated methyl group can be hydrolysed by S-adenosylhomocysteine hydrolase (EC 3.3.1.1) giving adenosine and homocysteine. Two candidate genes with identical sequences and one truncated form are present.

However, *E. histolytica* lacks the remaining enzymes for the reverse transsulphuration pathway (forming cysteine from methionine) (Nozaki *et al.*, 2005), i.e. cystathionine β-synthase and cystathionine γ-lyase. In addition, *E. histolytica* lacks all enzymes involved in the forward transsulphuration (forming methionine from cysteine) including cobalamin-dependent methionine synthase (EC 2.1.1.13) or cobalamin-independent methionine synthase (EC 2.1.1.14), which suggests that *E. histolytica* is capable of neither converting homocysteine to cystathionine nor recycling homocysteine to methionine.

E. histolytica lacks the methylthioadenosine cycle enzymes except for two, 5'-methylthioadenosine/S-adenosyl homocysteine nucleosidase (EC 3.2.2.9) and aspartate aminotransferase (AT, EC 2.6.1.1). The significance of these two enzymes in *E. histolytica* is unknown.

4.2.4 Arginine

In *G. intestinalis* and *T. vaginalis* the arginine deiminase (EC 3.5.3.6) pathway is important for energy generation (Knodler *et al.*, 1994; Linstead and Cranshaw, 1983; Schofield and Edwards, 1994), generating one ATP molecule from the breakdown of arginine to ornithine. In contrast, no arginine deiminase gene or dihydrolase pathway was detected in the *E. histolytica* genome.

In *E. histolytica*, arginine can either be degraded by arginase (EC 3.5.3.1) via ornithine or arginine decarboxylase (EC 4.1.1.19) via agmatine. The arginine decarboxylase reaction uses up protons and may be involved in the acid resistance needed for the passage of cysts through the human stomach Anderson and Loftus, 2005). Another function suggested for arginine degradation was that it depletes arginine as a substrate for human macrophages, preventing NO synthesis and amoebicidal activity (Elnekave *et al.*, 2003). Both enzymes could also be important for the generation of the polyamine putrescine (see 4.3 below). The genome contains a single gene encoding a 96 kDa polypeptide annotated as ornithine/arginine/lysine decarboxylase, the substrate specificity of which has not yet been examined on the recombinant protein level. There is a single gene encoding a putative 33 kDa arginase.

4.2.5 Glutamate, glutamine

In aerobic organisms, the 2-oxoglutarate generated from glutamate in a transaminase reaction enters the citric acid cycle for further catabolism. In *E. histolytica*, which also contains transaminases, 2-oxoglutarate can be oxidised by PFOR to give succinyl-CoA from which one molecule of ATP can be generated.

Several other gene products of *E. histolytica* could act on glutamine and glutamate. The genome lacks a glutaminase (EC 3.5.1.2) to carry out the simple hydrolysis of glutamine. Instead there is a putative glucosamine-fructose-6-phosphate aminotransferase (EC 2.6.1.16), which uses the energy in the amide

group of glutamine to generate glucosamine 6-phosphate from fructose 6-phosphate. This product may be used for cyst wall biosynthesis.

4.2.6 Tryptophan

Tryptophan can be degraded to indole, pyruvate, and ammonia by the PLP-dependent enzyme tryptophanase (EC 4.1.99.1), for which one candidate gene exists. To date, tryptophanase has only been found in bacteria, never in eukaryotes, and it is also on the list of LGT candidates.

4.2.7 Alanine: a possible special case

Alanine could potentially be transformed into pyruvate by alanine aminotransferase (synonymous with alanine:pyruvate transaminase, EC 2.6.1.2). However, *E. histolytica* is reported to excrete alanine (Zuo and Coombs, 1995) suggesting that this enzyme is not used under the culture conditions tested. Conceivably, the purpose of the excretion process may be to carry excess nitrogen out of the cell in the absence of a functional urea cycle.

4.2.8 Catabolism of other amino acids

Most of the enzymes for branched-chain amino acid metabolism are missing in *E. histolytica* but leucine, isoleucine, and valine could be transformed to 2-oxoisocaproate, 2-oxo-3-methylvalerate, and 2-oxovalerate, respectively, by a putative branched-chain amino acid aminotransferase (EC 2.6.1.42), one of the aminotransferases mentioned above (section 4.2). This could produce ammonia or transfer the amino group to 2-oxoglutarate to form glutamate. Subsequent oxidative decarboxylation to give the respective CoA-derivatives could be envisaged but so far no gene candidates for the necessary dehydrogenases have been identified.

One gene encodes a putative histidine ammonia-lyase (EC 4.3.1.3), which is responsible for the decomposition of histidine into urocanate and ammonia. Other than the formation of ammonia, the significance of this enzyme is not

clear since the downstream enzymes involved in histidine catabolism from urocanate to glutamate were not found.

Currently, there is little information regarding the fate of the amino acids glycine, proline, phenylalanine, tyrosine, and lysine in *E. histolytica.* No genes for the catabolic enzymes necessary were detected except for a LGT candidate bacterial-type 96 kDa broad-specificity ornithine/arginine/lysine decarboxylase that may be acting on lysine.

4.3 Polyamine Metabolism

The absence of *S*-adenosyl-L-methionine decarboxylase (EC 4.1.1.50), which converts *S*-adenosyl methionine into decarboxylated *S*-adenosyl methionine, spermidine synthase (EC 2.5.1.16), and spermine synthase (EC 2.5.1.22) suggests a complete lack of polyamine metabolism in this parasite (Anderson and Loftus, 2005). However, as mentioned above, *E. histolytica* possesses genes encoding arginase and arginine decarboxylase. Both could be involved in the production of putrescine via agmatine and agmatinase (EC 3.5.3.11) or via ornithine and ornithine decarboxylase (EC 4.1.1.17). The high putrescine concentration in trophozoites demonstrated by NMR spectroscopy (9.5 mM) (Bakker-Grunwald *et al.*, 1995) reinforces the physiological significance of putrescine. However, the fate of putrescine is unknown as neither spermine nor spermidine has been demonstrated in *E. histolytica*.

There is controversy regarding the presence or absence of trypanothione, a spermidine-containing thiol, in *E. histolytica*. Trypanothione is a major thiol in trypanosomes and leishmania (Fairlamb and Cerami, 1992) and contains two molecules of glutathione joined by a spermidine linker. The first reports detected the presence of trypanothione in *E. histolytica* (Ondarza *et al.*, 1997) but were contradicted soon after (Ariyanayagam and Fairlamb, 1999). More recently another study reaffirmed its presence (Ondarza *et al.*, 2005). However, the gene encoding trypanothione reductase reported from *E. histolytica* strain

HK-9 (AF503571) has no homologue in the genome of HM-1:IMSS. Although this matter has not been resolved, there is general agreement that the major thiol in *E. histolytica* is cysteine (Fahey *et al.*, 1984).

The *E. histolytica* genome encodes a 46 kDa ornithine decarboxylase with similarity to both plant and vertebrate enzymes and there is also the 96 kDa ornithine/arginine/lysine decarboxylase (see section 4.2.4). Only the former enzyme has been characterised at the biochemical level (Arteaga-Nieto *et al.*, 2002) and has been shown to be insensitive to difluoromethylornithine (DFMO), as is *E. histolytica* (Gillin *et al.*, 1984).

The conversion of arginine into putrescine via agmatine, in a reaction initiated by arginine decarboxylase, is generally present in bacteria and plants. Although arginine decarboxylase is present in *E. histolytica*, agmatinase (EC 3.5.3.11), which further catalyses conversion of agmatine to putrescine and urea, appears absent. However, one gene identified as a 33 kDa arginase also shares 21% sequence identity with human mitochondrial agmatinase and therefore its substrates need to be examined on the biochemical level to see whether the enzyme can act on arginine, agmatine, or both. At present, the role of arginine decarboxylase in *E. histolytica* is not clear, although as mentioned above this enzyme may also be involved in acid resistance in *E. histolytica.*

4.4 Biosynthesis of Amino Acids

4.4.1 Cysteine and serine

One of the areas in which reduction of metabolism is most evident is in amino acid biosynthesis. Biosynthetic pathways for most amino acids other than serine and cysteine (Ali *et al.*, 2003; Ali *et al.*, 2004a; Nozaki *et al.*, 1998a; Nozaki *et al.*, 1999) have been lost in *E. histolytica*. Similarly, *P. falciparum*, which predominantly acquires amino acids from host haemoglobins, lacks biosynthesis of most amino acids (Gardner *et al.*, 2002). Intracellular concentrations of some amino acids (glutamate, leucine, valine, and proline, in

descending order of abundance) are very high in *E. histolytica* ranging from 6-21 mM (Bakker-Grunwald *et al.*, 1995). In particular, the glutamate and proline concentrations are much higher in the cells than in the growth medium (21 and 7.3 mM vs. 5.9 and 1.8 mM, respectively). Glutamate accounts for over one third of the total amino acid pool (Bakker-Grunwald *et al.*, 1995), and is likely to play a central role in homeostasis not only of amino acids but of energy metabolism in general. Thus, it is likely that these amino acids are actively taken up by as-yet unidentified amino acid transporters.

Retention of the serine and cysteine biosynthetic pathways when the others have been lost is likely related to the physiological importance of cysteine, which is the major intracellular thiol of this parasite. The cysteine biosynthetic pathway consists of two major steps, catalysed by serine acetyltransferase (EC 2.3.1.30), which produces *O*-acetylserine from serine and acetyl-coenzyme A, and cysteine synthase (EC 2.5.1.47), which subsequently transfers an alanyl moiety from *O*-acetylserine to sulphide to produce cysteine. *E. histolytica* possesses three genes each for cysteine synthase and serine acetyltransferase. Cysteine synthase 1 and 2 were considered to be allelic isotypes (Nozaki *et al.*, 1998b), while cysteine synthase 3 appears to be distinct, with only 83% identity to cysteine synthase 1 and 2. In contrast, all three serine acetyltransferase genes seem to be distinct, showing only 48-73% identity (Ali and Nozaki, unpublished). It was previously shown that cysteine synthase 1/2 and serine acetyltransferase 1 are unique in that (a) they do not form a heterocomplex, in contrast to other organisms (Bogdanova and Hell, 1997; Droux *et al.*, 1998) and (b) serine acetyltransferase 1 is sensitive to allosteric inhibition by both L-cysteine and L-cystine (Nozaki *et al.*, 1999). Since all variants of these two enzymes lack organelle-targeting sequences, the significance of the multiple isotypes is unknown. It is important to determine subcellular distribution and specific functions of these isotypes to understand the significance of the redundancy. As this pathway is absent in humans, it is a rational target for development of new chemotherapeutic drugs against amoebiasis.

Serine is synthesised de novo utilising the glycolytic intermediate 3-phosphoglycerate, in a pathway that includes three sequential reactions catalysed by D-phosphoglycerate dehydrogenase (EC 1.1.1.95), phospho-L-serine aminotransferase (EC 2.6.1.52), and *O*-phospho L-serine phosphatase (EC 3.1.3.3). Although the final enzyme has not yet been enzymologically and functionally analysed, the first two enzymes have been characterised (Ali and Nozaki, 2006; Ali *et al*., 2004a).

4.4.2 Interconversion of glutamate-glutamine and aspartate-asparagine

The single step interconversions of glutamate and glutamine, catalysed by glutamate synthase (EC 1.4.1.13) and glutamine synthetase (EC 6.3.1.2), and of aspartate and asparagine by asparagine synthase (EC 6.3.5.4) are found in *E. histolytica*. There are two isotypes of glutamine synthetase with 47% amino acid identity and 5 candidate genes. NADPH-dependent glutamate synthase (EC 1.4.1.13) catalyses the formation of two glutamates from glutamine and 2-oxo-glutarate in bacteria, yeast, and plants, and together with glutamine synthetase is involved in ammonia fixation under ammonia-restricted conditions. NADPH-dependent glutamate synthase is normally composed of two large and two small subunits (Petoukhov *et al.*, 2003). Although three genes encoding the small subunit are present, the large subunit appears to be absent in *E. histolytica*. These putative NADPH-dependent glutamate synthase small subunits share 80% amino acid identity and show 44% amino acid identity to homologues from the Archaea. The similarity to archaeal-type glutamate synthase (Nesbo *et al.*, 2001) suggests that the *E. histolytica* small subunits may function as a glutamate synthase without the large subunit, as shown for gltA from the archaean *Pyrococcus* (Jongsareejit *et al.*, 1997).

The two enzymes that catalyse interconversion between aspartate and asparagine, aspartate ammonia ligase (EC 6.3.1.1) and asparaginase (EC 3.5.1.1; see 4.2.1), are present in *E. histolytica*. Two types of aspartate

ammonia ligases, AsnA and AsnB, are known from other organisms: the former utilises only ammonia, while the latter uses both ammonia and glutamine as amide donors in a reverse reaction. Mammals possess only AsnA, whereas prokaryotes have both AsnA and AsnB (Boehlein *et al.*, 1996; Nakamura *et al.*, 1981). Interestingly, *E. histolytica* possesses only the AsnB homologue. Thus, the amoebic enzyme is likely involved in the formation of glutamate from glutamine, in addition to asparagine formation from aspartate.

4.4.3 Synthesis of glutamate and aspartate

Glutamate can be formed from 2-oxo-glutarate and ammonia in a reversible reaction catalysed by glutamate dehydrogenase (EC 1.4.1.2), which is present in *E. histolytica*. It is known that this enzyme plays a dominant role in ammonia fixation under ammonia-non-restricted conditions as this reaction consumes no ATP. In addition, glutamate dehydrogenase is also involved in gluconeogenesis from glutamate.

Aspartate ammonia lyase (synonymous with aspartase, EC 4.3.1.1), which decomposes aspartate into fumarate and ammonia in a reversible reaction, is also present in *E. histolytica* (see 4.2.1 above).

4.5 Lipid Metabolism

For *E. histolytica*, the lack of oxidative phosphorylation means that the high energy content of lipids such as fatty acids cannot be exploited. Therefore lipids such as phospholipids and cholesterol are primarily membrane components in *E. histolytica* (Das *et al.*, 2002; Sawyer *et al.*, 1967). Although these components are acquired mainly from their food or from the human host, *E. histolytica* does have some capability for biosynthesis, as well as extending and remodeling lipids, and for attaching lipids to proteins.

4.5.1 Lipid biosynthetic capabilities

4.5.1 (a) Polyisoprene biosynthesis and protein prenylation

Cholesterol is an important membrane constituent generated from C_5 isoprene precursors. *E. histolytica* trophozoites in axenic culture need cholesterol in their growth medium (Reeves, 1984), and it is likely that they acquire it from their human host. Reeves (1984) even cites several studies which show that hypercholesteremia in the host increases the damage inflicted by amoebic infection. *E. histolytica* lacks several enzymes for the classical sterol biosynthesis pathway (Schroepfer, 1981). The first stage of sterol biosynthesis is the formation of isopentenyl- or dimethylallyl diphosphate. In the *E. histolytica* genome no candidate genes for the generation of these intermediates were found, neither for the mevalonate pathway nor for the mevalonate-independent methylerythritol 4-phosphate (MEP) pathway that operates in bacteria and plants (Hunter *et al.*, 2003; Rohmer *et al.*, 1993). In a later step towards cholesterol sythesis, two molecules of C_{15} farnesyl diphosphate are dimerised to give C_{30} presqualene diphosphate (EC 2.5.1.21). This enzyme activity and those catalysing the subsequent steps also appear to be absent. The genome data thus support the long standing conclusion that cholesterol biosynthesis is absent from *E. histolytica*.

Unexpectedly, the *E. histolytica* genome appears to encode enzymes involved in the intermediate stages of cholesterol biosynthesis from C_5 isopentenyl diphosphate to C_{15} farnesyl diphosphate. The latter compound, and the larger C_{20} compound geranylgeranyl diphosphate, may serve as precursors for the hydrophobic modification of GTP-binding proteins allowing them to bind to membranes (Grunler *et al.*, 1994). Protein prenylation is a ubiquitous process. It is important in human cell biology, health and disease (McTaggart, 2006), but it is also essential for parasites such that protein farnesylation has been proposed as a potential novel target for anti-parasitic chemotherapy (Maurer-Stroh *et al.*, 2003) including anti-*E. histolytica* chemotherapy (Ghosh *et al.*, 2004).

The first enzyme in this pathway is the isopentenyl-diphosphate delta-isomerase which catalyses the conversion of isopentenyl diphosphate to dimethylallyl

diphosphate (EC 5.3.3.2). There is a single gene encoding this enzyme that is of presumed bacterial origin and is on the list of LGT candidates. The two isomeric C_5 isoprenyl diphosphates undergo condensation to C_{10} geranyl diphosphate, catalysed by geranyl-diphosphate synthase (EC 2.5.1.1). Farnesyl-diphosphate synthase (EC 2.5.1.10) then adds another C_5 unit to give C_{15} farnesyl diphosphate. Finally geranylgeranyl-diphosphate synthase (EC 2.5.1.29) adds another C_5 prenyl unit to give C_{20} geranylgeranyl diphosphate. The genome contains five putative prenyl transferase genes, which all have been annotated as geranylgeranyl-diphosphate synthases. Their sequences are highly similar, with the exception that the open reading frames are disrupted in two of them (XP_650479 and XP_655958). These prenyl transferases appear to be of bacterial origin as well, and XP_650913 is on the list of LGT candidates. When searching for geranyl-diphosphate synthase or farnesyl-diphosphate synthase in the *E. histolytica* genome, the closest matches are for the same genes, so that the substrate specificity of these enzymes is unclear and needs to be examined biochemically.

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

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

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

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

4.5.1 (b) Fatty acid biosynthesis

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

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

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

4.5.2 Phospholipid metabolism

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

4.5.2 (a) Phospholipid biosynthesis

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

The activation of ethanolamine (EC 2.7.1.82) or choline (EC 2.7.1.32) for attachment to the phosphadidate starts with phosphorylation. There are two genes identified as choline/ethanolamine kinases that share 37% amino acid identity. Next, ethanolamine phosphate and choline phosphate are converted to

CDP-ethanolamine (EC 2.7.7.14) and CDP-choline (EC 2.7.7.15), respectively. The genome encodes two enzymes sharing 57% sequence identity that are identified as ethanolamine-phosphate cytidylyltransferases. The substrate specificity of these enzymes needs to be examined on the biochemical level. Finally the activated ethanolamine or choline is attached to diacylglycerol by the enzymes ethanolaminephosphotransferase (EC 2.7.8.1) or diacylglycerol cholinephosphotransferase (EC 2.7.8.2) producing phosphatidylethanolamine or phosphatidylcholine, respectively. For these activities a total of 8 possible genes are found that share varying degrees of sequence similarity.

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

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

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

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

The second is phospholipid-cholesterol acyltransferase (EC 2.3.1.43), which transfers an acyl group from phospholipids such as phosphatidylcholine to cholesterol giving a cholesterol ester. The genome contains 7 genes for this enzyme. So far nothing is known about the importance of cholesterol esters for *E. histolytica.*

4.5.2 (b) Phospholipid degradation

Phospholipids are degraded by phospholipases. Whereas phospholipases A1 (EC 3.1.1.32) and A2 (EC 3.1.1.4) cleave the acyl residues in the 1 or 2 position of the glycerol core, phospholipases C (EC 3.1.4.3) and D (EC 3.1.4.4) cleave at the phosphate, phospholipase C on the glycerol side, and phospholipase D on the aminoalcohol side. In *E. histolytica* phospholipase A activity has been implicated in virulence (Ravdin *et al.*, 1985) as it liberates toxic fatty acids and lysophospholipids (Said-Fernandez and Lopez-Revilla, 1988). Phospholipases A have been found in two forms, a membrane-bound Ca-dependent form active at alkaline pH and a soluble Ca-independent form active at acid pH (Long-Krug *et al.*, 1985; Vargas-Villarreal *et al.*, 1998). The genome encodes 11 potential phospholipases A with predicted pI values between 4.8 and 8.8 and various degrees of sequence similarity. In addition the *E. histolytica* genome encodes 3 potential phospholipases D.

Finally, there are 2 highly similar genes for phospholipases C, but these are homologous to phosphatidylinositol-specific phospholipases C (EC 3.1.4.11) and most likely do not cleave phosphatidylinositol or phosphatidylcholine but GPI-anchors instead. So far there are no studies using individual recombinant phospholipases, and it is not yet known how much these enzymes may contribute to the virulence of *E. histolytica*.

4.6 Coenzyme A Biosynthesis and Pantothenate Metabolism

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

Phosphopantothenoyl-cysteine decarboxylase (EC 4.1.1.36) and phosphopantothenoyl-cysteine synthetase (EC 6.3.2.5, synonymous with

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

4.7 Nucleic Acid Metabolism

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

4.8 Missing Pieces

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

4.9 Transporters

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

Since glucose transport activity has been experimentally characterised in *E. histolytica* and glucose is thought to be the major energy source, it was surprising to find no homologues of known hexose transporters in the genome. Most hexose transporters belong to the sugar porter subfamily of the major facilitator superfamily (TC 2.A.1.1), members of which are found in prokaryotes, animals, fungi, plants, and other protists, including *D. discoideum*, but no proteins of this family were found in the *E. histolytica* genome. A group of candidate monosaccharide transporters found within the genome is related to the glucose/ribose porter family from prokaryotes (TC 2.A.7.5). These transporters consist of two related domains, and the *Entamoeba* proteins appear to have the N-terminal and C-terminal domains switched relative to the bacterial proteins. Functional characterisation of transporter-encoding genes will be necessary for a more complete picture.

5. THE CYTOSKELETON

The eukaryotic cytoskeleton is composed of three main elements: actin microfilaments, tubulin-based microtubules and intermediate filaments. Despite the fact that *E. histolytica* is very motile and performs phagocytosis very efficiently, its cytoskeletal components are rather simple. No genes encoding homologues of intermediate filament network proteins, including keratins, desmin and vimentin, have been identified in *E. histolytica*, providing further

evidence that these particular cytoskeletal components are rather poorly conserved in evolution. In contrast, microfilament and microtubule components have been readily identified.

5.1 Actin and Microfilaments

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

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

Since the spontaneous polymerisation of actin monomers is inhibited by the action of sequestering proteins such as thymosin β4 and profilin, efficient actin polymerisation requires the intervention of an actin polymerisation-promoting factor. The best described promoting factors are the Arp2/3 complex and the formin protein family.

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

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

Once nucleated, actin filaments are able to grow rapidly by addition of monomers at their barbed ends. Filaments are regulated by several mechanisms (Winder and Ayscough, 2005). Filament length is controlled by capping proteins: barbed end cappers (such as capping protein and gelsolin) block addition of new monomers and thus act to decrease the overall length of the filament. In addition, gelsolin severs actin filaments, thereby rapidly increasing actin dynamics. Actin filaments appear to be significantly shorter in *E. histolytica* when compared with those from fibroblasts and stress fibres are not formed in this amoeba. Although *E. histolytica* actin has been shown not to bind DNase I (Meza *et al.*, 1983), the inferred amino acid sequence indicates conservation of all the residues likely to participate in this binding event - suggesting that post-translational modifications of actin monomers may prevent DNAse I-actin binding. It remains to be determined whether such modifications of actin participate in the regulation of actin polymerisation. The genome encodes multiple genes associated with filament capping and severing, as well as candidates for proteins that cross-link actin filaments and thus organise them into a supramolecular network. The organisation of actin into networks and higher-order structures is crucial for both cell shape and function. These structures can be responsible for overall cell shape and related processes, such as bundle formation through α-actinin activity, for example. The arrangement of actin filaments into cross-linked arrays is also mediated by proteins with multiple actin-binding domains, which allows a more perpendicular arrangement of actin filaments. Examples of this type of protein are the large, flexible filamin dimer (Vargas *et al.*, 1996) and the spectrin tetramer. Genome analysis has now identified many candidate genes for actin-binding proteins in *E. histolytica*, and additional protein partners of this versatile family responsible for cytoskeleton regulation are likely to emerge from curation of the sequence and cellular studies of cell motility and phagocytosis in this parasite.

5.2 Tubulins and microtubules

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

There is little information available on the precise organisation of microtubules and F-actin cytoskeleton during *E. histolytica* motility. In many eukaryotic cells, F-actin-microtubule interactions can be observed in lamellipodia at all stages. Interestingly, microtubules preferentially grow along actin bundles in filopodia, suggesting that a physical link between the structures exists (Leung *et al.*, 2002). Multifunctional MT-associated proteins (MAPs, like MAP1B, MAP2 and plakins) are promising candidates for acting as such links, either via dimerisation of MAPs with single microtubule and actin binding sites or by

direct bridging of the two cytoskeletons (for example via plakins, which contain binding sites for both microtubules and actin within a single molecule). Plakin homologues have not been identified in the *E. histolytica* genome but a MAP is present. Furthermore, proteins with domains that can bind to actin (and potentially to MT) have been described in *E. histolytica* - the ABP-120 gelation factor, for example (Vargas *et al.*, 1996).

5.3 Molecular motors

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

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

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

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

Directional transport along the MTs depends on dynein and kinesin, both MT-associated motor proteins which convert the chemical energy from ATP hydrolysis into movement. These motors are unidirectional and move towards either the MT plus- or minus- ends (Mallik and Gross, 2004). Kinesins and dyneins have been implicated in a wide range of functions - principally intracellular organelle transport during interphase and spindle function during

mitosis and meiosis. Members of the dynein family are minus-end directed, although this remains to be confirmed for a few uncharacterised, vertebrate, cytoplasmic dynein heavy chains. It has not yet been reliably established that the *E. histolytica* genome contains a dynein heavy chain gene, although a dynein light chain gene is present: improvements in gene assembly should provide us with more information on this high molecular mass protein.

Kinesins are microtubule-dependent molecular motors that play important roles in intracellular transport and cell division. Even though the motor domain is found within the N-terminus in most kinesins (N-type), it is located within the middle or C-terminal domains in some members of the family (M-type and C-type kinesins, respectively) (Asbury, 2005). The position of the motor domain dictates the polarity of the movement of kinesin along the MT: whereas N- and M-type kinesins are plus-end directed, the C-type kinesins are minus-end directed. Humans possess 31 different kinesins and trypanosomes have more than 40. The *E. histolytica* genome sequence predicts only six kinesin-encoding genes (four N-type, two C-type and no M-type homologues have been found). One of the N-kinesins also contains a domain homologous to the HOOK protein required for the correct positioning of microtubular structures within the cell (Walenta *et al.*, 2001). Bearing in mind that *E. histolytica* MTs are intranuclear, the study of kinesin function and trafficking should help us elucidate what is likely to be a very interesting MT functional mechanism.

6. VESICULAR TRAFFIC

The requirement for nutritional uptake from the extracellular milieu in the host intestine imposes a heavy reliance on endocytic and phagocytic activities in *Entamoeba* (Espinosa-Cantellano and Martínez-Palomo, 2000). Proliferating trophozoites secrete a number of peptides and proteins including cysteine proteases (Que and Reed, 2000) and amoebapores (Leippe, 1999) required for bacterial cell killing and degradation as well as being implicated in virulence (Petri, 2002). During encystation, the cells also secrete substrates used for the

formation of the cyst wall (Eichinger, 1997). Electron micrographic studies have revealed a complex membrane organisation. The trophozoites contain numerous vesicles and vacuoles varying in size and shape (Clark *et al.*, 2000; Mazzuco *et al.*, 1997). Intracellular transport of both endocytosed and synthesised molecules between compartments is regulated by the elaborate orchestration of vesicle formation, transport, docking and fusion to the target compartment (Bonifacino and Glick, 2004; Kirchhausen, 2000).

6.1 Complexity of Vesicle Trafficking

Among a number of molecules and structures involved in vesicular trafficking, three types of coated vesicles, named coatomer protein (COP) I, COPII, and clathrin-coated vesicles are the best characterised (Bonifacino and Glick, 2004; Kirchhausen, 2000). COPI vesicles primarily mediate transport from the Golgi to the endoplasmic reticulum (ER) and between the Golgi cisternae, while COPII vesicles are involved in the transport from the ER to the *cis*-Golgi. The clathrin-dependent pathway has a few independent routes: from the plasma membrane to endosomes, from the Golgi to endosomes, and from endosomes to the Golgi. It has been well established that certain subfamilies of Ras-like small GTPases, widely conserved among eukaryotes, regulate both the formation of transport vesicles and their docking and fusion to the target organelles. The ADP-ribosylation factor (Arf) and secretion-associated Ras-related protein (Sar) families of GTPases regulate the formation of COPI and COPII vesicles (Memon, 2004), respectively. In contrast, the Rab family of GTPases (Novick and Zerial, 1997) is involved in the targeting and fusion of vesicles to the acceptor organelles together with the tethering machinery SNARE (a soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor) (Chen and Scheller, 2001). Since individual coat proteins, small GTPases, SNAREs, and their associated proteins show distinct intracellular distributions in both unicellular and multicellular organisms, they are believed to play a critical role in the determination of membrane trafficking specificity (Chen and Scheller, 2001; Munro, 2004; Novick and Zerial, 1997). It is generally believed that the

total number of proteins involved in the membrane traffic reflects the complexity and multiplicity of its organism. The total number of the putative amoebic genes encoding Arf/Sar, Rab, SNARE, and coat proteins together with those from *S. cerevisiae, C. elegans, D. melanogaster, H. sapiens,* and *A. thaliana,* is shown in Table 7. *E. histolytica* reveals complexity similar to yeast, fly, and worm in case of Sar/Arf and SNAREs, while the number of genes encoding three coat proteins (COPI, COPII, and Adapter Proteins (APs)) was higher in *E. histolytica* than these organisms and comparable to that in mammals and plants. In contrast, the number of Rab proteins in *E. histolytica* is exceptionally high, exceeding that in mammals and plants.

6.2 Proteins Involved in Vesicle Formation

6.2.1 COPII-coated vesicles and Sar1 GTPase

COPII components were originally discovered in yeast using genetic and biochemical approaches (reviewed in Bonifacino and Glick (2004)). COPII vesicles mediate the transport from the ER to the Golgi and consists of three major cytosolic components and a total of five essential proteins: the Sec23p-Sec24p complex, the Sec13p-Sec31p complex, and the small GTPase Sar1p (Barlowe *et al.*, 1994). Sar1p and Sec23p-Sec24p complexare involved in the formation of the membrane-proximal layer of the coat, while Sec13p-Sec31p complex mediates the formation of the second membrane-distal layer (Shaywitz *et al.*, 1997). These proteins are well conserved among various organisms (Table 7). *E. histolytica* encodes one each of Sar1, Sec13 and Sec31, two of Sec23, and five proteins corresponding to Sec24 (Table 7). The yeast and human genomes also encode multiple Sec24 isotypes (3 and 4, respectively). Although Sec24 isotypes have been shown to be responsible for the selection of transmembrane cargo proteins in yeast (Peng *et al.*, 2000; Roberg *et al.*, 1999), the significance of the Sec24 redundancy in *E. histolytica* is not clear.

Additional regulatory proteins participate in COPII assembly in yeast, including Sec16p, a putative scaffold protein (Espenshade *et al.*, 1995), and Sec12p, a

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

6.2.2 COPI-coated vesicles and Arf GTPases

COPI-coated vesicles, which mediate transport from the Golgi to the ER and between the Golgi cisternae (Kirchhausen, 2000), consist of seven proteins (α, β, β', γ, δ, ε, and ζ-COP) (Hara-Kuge *et al.*, 1994). The number of proteins making up the COPI coat, and thus the complexity of COPI components, varies among organisms (Table 7). While human possesses two isotypes of γ-COP and ζ-COP, yeast has a single gene for each. In humans, the two isotypes of γ-COP and ζ-COP form three different COPI complexes (γ1/ζ1, γ1/ζ2, and γ2/ζ1), which have different intracellular distributions (Wegmann *et al.*, 2004). This implies that COPI-coated vesicles are also involved in functions other than Golgi-to-ER transport (Whitney *et al.*, 1995). In *E. histolytica*, the COPI complex appears more heterogeneous: *E. histolytica* encodes two isotypes each of γ-COP, δ-COP, and α-COP and three isotypes of β-COP. In contrast, *E. histolytica* lacks ε-COP, which is known to stabilise α-COP (Duden *et al.*, 1998). It has been shown in yeast that all genes encoding components of COPI coat except for Sec28p, the yeast ε-COP homologue, are essential for growth (Duden *et al.*, 1998).

Recruitment of COPI to the Golgi membrane requires the association of a GTP-bound GTPase called Arf (Donaldson *et al.*, 1992; Kahn *et al.*, 2006). Arf was initially identified due to its ability to stimulate the ADP-ribosyltransferase activity of cholera toxin A (Kahn and Gilman, 1984). To recruit the COPI coat, Arfs are activated by a Sec7 domain-containing protein, Arf-GEF, which is a target of a fungal metabolite brefeldin A (Helms and Rothman, 1992; Sata *et al.*, 1998). Among Arf family proteins, Arf1 is involved in the formation of COPI-coated vesicles in the retrograde transport from the Golgi to ER, and is also involved in the assembly of clathrin-AP1 (see next section) on the *trans*-Golgi network (TGN) (Stamnes and Rothman, 1993), clathrin-AP3 on endosomes (Ooi *et al.*, 1998), and the recruitment of AP-4 to the TGN (Boehm *et al.*, 2001). The specific roles of Arfs3-5 are less clear, although Arf4 and Arf5 show *in vitro* activities similar to Arf1. Functional cooperativity of Arfs in the vesicular formation has also been demonstrated recently. At least two of four human Arf isotypes (Arf1, Arf3-5) are essential for a retrograde pathway from the Golgi to the ER, in the secretory pathway from the Golgi to the TGN, and in the recycling from endosomes to the plasma membrane (Volpicelli-Daley *et al.*, 2005). In contrast to these Arfs, Arf6 regulates the assembly of actin filaments and is involved in endocytosis on the plasma membrane (Radhakrishna and Donaldson, 1997).

GTPases that share significant similarity to Arf, but do not either activate cholera toxin A or rescue *S. cerevisiae Arf* mutants are known as Arls (Arf-like GTPases) (Lee *et al.*, 1997a). Arl1 is involved in endosome-to-Golgi trafficking (Lu *et al.*, 2001; Lu *et al.*, 2004). Other Arls (Arls 2-11) and Arf-related proteins (Arp or ArfRP 1-2) have been localised to the cytosol, nucleus, cytoskeleton and mitochondria (Burd *et al.*, 2004; Pasqualato *et al.*, 2002). The number of Arf, Arl, and Arf-related proteins varies among organisms (Table 7). Among 27 members identified in humans, only about a half dozen Arf/Arl/Arp proteins, including Arf1-6 and Arl1 (Wennerberg *et al.*, 2005), have been

shown to function in membrane traffic (Lu *et al.*, 2001). The localisation and function of the remaining Arf/Arl/Arp remained unclear.

E. histolytica encodes ten Arf/Arl proteins (Table 7). Only two *E. histolytica* Arfs (A1 and A2) have a high percentage identity to human Arfs 1, 3, 5, and 6 and yeast Arfs 1-3 (57-76% identity), while the remaining eight Arf/Arl fall into three groups (A4-6, B1-3, and C) and are equally divergent from one another and from other organisms. Both the intracellular distributions and the specific steps in vesicular trafficking mediated by these *Entamoeba* Arf/Arl proteins are unknown. It is worth noting that five of these Arfs lack a conserved glycine residue at the second amino acid position of the amino terminus; this glycine is known to be myristylated and essential for membrane association in other organisms (Randazzo *et al.*, 1995). *Eh*ArfA4 also lacks one of the conserved GTP-binding consensus regions (Box2). Similar deletion of GTP-binding domains has also been observed in proteins belonging to the Rab family (see section 6.3.1).

6.2.3 Clathrin-coated vesicle and its adaptor proteins

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

subunit (Keen, 1987). Four major types of AP complexes (AP1-4) have been identified (Boehm and Bonifacino, 2001; Nakatsu and Ohno, 2003). AP-2 (consisting of α, β2, σ2, and μ2) mediates endocytosis from the plasma membrane (Conner and Schmid, 2003; Motley *et al.*, 2003), while AP-1 (γ, β1, σ1, and μ1A) (Meyer *et al.*, 2000), AP-3 (δ, β3A, σ3, and μ3A) (Le Borgne *et al.*, 2001; Vowels and Payne, 1998), and AP-4 (ε, β4, σ4, and μ4) (Aguilar *et al.*, 2001), play a role in the Golgi-endosome, endosomal-lysosomal, or the Golgi/lysosome sorting pathway, respectively. AP-4, which is present only in mammals and plants (Boehm and Bonifacino, 2001), was also identified in non-clathrin-coated vesicles mediating the transport from TGN to the plasma membrane or endosomes (Hirst *et al.*, 1999). A few isotypes of AP-1 and AP-3, e.g., AP-1B (γ, β1, σ1, and μ1B) and AP-3B (δ, β3B, σ3, and μ3B), showed tissue specific expression (Faundez *et al.*, 1998; Folsch *et al.*, 1999). *E. histolytica* encodes ten large subunits (α, β, γ, δ, and ε), four medium subunits (one each of μ1 and μ2, and two μ3), and four small subunits (ε1–ε4). This suggests that *E. histolytica* produces four types of AP complex, as in humans and plants.

6.3 Proteins Involved in Vesicle Fusion

6.3.1 Rab GTPases

The docking and fusion of transport vesicles to a specific target compartment requires the appropriate Rab protein. Specific interaction of a Rab with its effector molecules in conjunction with the interaction between SNAREs plays a central role in vesicle fusion (Zerial and McBride, 2001). In general, the complexity of the Rab gene family correlates with the degree of multicellularity. For example, *S. pombe*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, and *H. sapiens* consist of one, one, ca. 10^3, 10^9, and 10^{13} cells, and have 7, 11, 29, 29, and 60 Rab genes, respectively (Pereira-Leal and Seabra, 2001). It has been also shown that in multicellular organisms, Rab proteins are expressed in a highly coordinated (i.e. tissue-, organ-, or

developmental stage-specific) fashion (Seabra *et al.*, 2002; Zerial and McBride, 2001). *E. histolytica* possesses an extremely high number of Rab genes - 91 (Figure 6). Among its 91 Rabs only 22, including *Eh*Rab1, *Eh*Rab2, *Eh*Rab5, *Eh*Rab7, *Eh*Rab8, *Eh*Rab11, *Eh*Rab21, and their isotypes showed >40% identity to Rabs from other organisms. The 69 remaining *E. histolytica* Rab proteins showed only moderate similarity (<40% identity) and represent unique, presumably *Entamoeba*-specific, Rab proteins. Approximately one third of Rab proteins form 15 subfamilies, including Rab1, Rab2, Rab7, Rab8, Rab11, and RabC-P, each of which contains up to 9 isoforms. Interestingly, approximately 70% of *E. histolytica* Rab genes contain one or more introns (Saito-Nakano *et al.*, 2005). SNARE genes are also intron-rich whereas the Sar/Arf GTPase and the three coat protein genes have a low frequency of introns. The high frequency of introns in the Rab and SNARE gene families may indicate the presence of post-transcriptional regulation of these genes.

Although Rab proteins generally possess a CXC or CC at the carboxyl terminus, twenty-five *E. histolytica* Rabs have an atypical carboxyl terminus, such as CXXX, XCXX, XXCX, XXXC, or no cysteine at all. The enzyme(s) involved in the lipid modification of these unusual Rab proteins remain poorly understood (see 4.5.1 (a)). It is also worth noting that >20 *E. histolytica* Rab lack or contain only a degenerate form of the consensus sequence for structural elements such as the GTP-binding regions and the Switch I and II regions, implicated in the binding to GEF, GAP, effectors, or guanine nucleotides (Saito-Nakano *et al.*, 2005). These non-conventional *Eh*Rabs are not pseudogenes since at least some of the genes are known to be expressed as mRNA (Saito-Nakano *et al.*, 2001). It has been shown that neither *Eh*Rab5 nor *Eh*Rab7A rescued the corresponding yeast mutant (Saito-Nakano *et al.*, 2004). Therefore, many, if not all, *E. histolytica* Rabs may have lost functional interchangeability with their homologues in other organisms despite the relatively high percentage of sequence identities. Classification and annotation

of the *E. histolytica* Rab proteins has been previously described (Saito-Nakano *et al.*, 2005).

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

6.3.2 SNARE and their accessory proteins

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

The complexity of SNAREs has remained largely unchanged in yeast, fly, and worm, but has increased remarkably in mammals and plants (Table 7) indicating that although expansion of SNARE repertoires occurs, a set of core SNAREs is sufficient to mediate vesicular fusion of most pathways in multicellular organisms. *E. histolytica* encodes 28 putative SNAREs, 18 Q-SNAREs and 10 R-SNAREs, which is comparable to the complexity to humans and plants. A notable peculiarity of SNAREs in *E. histolytica* is the lack of a group of proteins possessing two helixes (Qb and Qc SNAREs) such as SNAP-25. Thus, the prototype model of membrane tethering by a combination of four helixes (from Qa, R, and Qb/Qc) does not appear to be possible in this organism.

A group of proteins that interact directly with the syntaxin subfamily, including the prototypical member yeast Sec1p and mammalian Munc-18, are essential cytosolic proteins peripherally associated with membranes (Toonen and Verhage, 2003). They are presumed to be chaperones, putting syntaxins into the conformations required for interaction with other SNAREs (Dulubova *et al.*, 1999; Yang *et al.*, 2000). Sec1/Munc-18 proteins are also conserved in *E. histolytica* (there are 5 Sec1 genes). Two additional important components involved in the recycling of fusion machinery, *N*-ethylmaleimide sensitive factor (NSF) (Beckers *et al.*, 1989) and soluble NSF attachment protein (SNAP) (Clary *et al.*, 1990; Mayer *et al.*, 1996) are also found in *E. histolytica*.

Other proteins involved in vesicle fusion are the saposin-like proteins mentioned earlier (Section 3. 3). The membrane-fusogenic activity of the *E. histolytica* SAPLIPs may play a role in vesicle fusion (Winkelmann *et al.*, 2006) but how they interface with the Rab/SNARE processes remains to be determined.

6.4 Comparisons and Implications

While the fundamental machinery of vesicular trafficking is conserved in *E. histolytica*, the high activity of the endocytic and biosynthetic transport pathway in this organism appears to have resulted in the dramatic expansion of the Rab gene repertoire. The diversity and complexity of Rab proteins present in *E. histolytica* likely reflect the vigorous dynamism of membrane transport and the reliance on Rab proteins for the specificity of vesicular trafficking. The high degree of Rab complexity observed in *E. histolytica* (91) has no precedent in other organisms, although the incomplete genome of *Trichomonas vaginalis* appears to encode 65 Rabs (Lal *et al.*, 2005) while *Dictyostelium* encodes 50 (Eichinger *et al.*, 2005). Rab proteins have been extensively studied in *Trypanosoma brucei* and the recent completion of *T. brucei*, *T. cruzi* and *Leishmania major* genomes led to identification of all Rab genes in these haemoflagellates (Ackers *et al.*, 2005; Berriman *et al.*, 2005; Quevillon *et al.*, 2003). Among the 16 Rab present in *T. brucei*, there are only three Rab proteins (RabX1-X3) that appear to be unique to kinetoplastids. *T. brucei* possesses 11 Rab proteins homologous to those in humans, suggesting significant conservation of the Rab-dependent core endomembrane systems in kinetoplastids. *Plasmodium falciparum* possesses only 11 Rab genes all of which are considered orthologues of yeast and mammalian Rabs, although Rab5a, 5b, and 6 revealed unique features (Quevillon *et al.*, 2003). Interestingly, some of these Rabs are expressed in a stage-dependent manner (Quevillon *et al.*, 2003). The comparatively small number of Rabs in these protists reinforces the tremendous diversity and complexity of Rabs seen in *E. histolytica* (Table 7).

In marked contrast to the complexity of Rab proteins in *E. histolytica*, the number of SNARE proteins, the other major components of vesicular fusion, is comparable to that in yeast. The apparent disparity in the number of Rab and SNARE proteins suggests one of three possibilities: 1) *Eh*Rab proteins share a single SNARE complex as an interacting partner (Huber *et al.*, 1993; Rowe *et*

al., 2001; Torii *et al.*, 2004), 2) a majority of *Eh*Rabs do not require SNARE proteins for membrane fusion (Demarque *et al.*, 2002), 3) some *Eh*Rabs are primarily involved in cellular functions other than membrane fusion, like Arl GTPases (Burd *et al.*, 2004; Pasqualato *et al.*, 2002). Genome-wide surveys of SNAREs in other protists are not available. The three major types of coatomer protein, which are conserved in *E. histolytica*, are also conserved in kinetoplastids (Berriman *et al.*, 2005). However, in contrast to *E. histolytica, T. brucei* does not possess multiple isotypes of COPI and II components except for Sec24, which has two isotypes. *T. cruzi* encodes all four AP complexes while *L. major* and *T. brucei* lack AP-4 or AP-2, respectively, which suggests that the repertoire of AP complexes in kinetoplastids is variable and species-specific. Although low similarity of the *E. histolytica* components to either yeast or mammalian orthologues make unequivocal assignment of *Entamoeba* AP complexes challenging, tentative assignments have been made. It is likely that *E. histolytica* encodes four kinds of AP complex corresponding to APs 1-4.

6.5 Glycosylation and Protein Folding.

6.5.1 Asparagine-linked glycan precursors.

Mammals, plants, *Dictyostelium*, and most fungi synthesise asparagine-linked glycans (N-glycans) by means of a common 14-sugar precursor dolichol-PP-$Glc_3Man_9GlcNAc_2$ (Figures 7 and 8) (Helenius and Aebi, 2004). This lipid-linked precursor is made by at least fourteen glycosyltransferases, which are present in the cytosolic aspect or lumen of the ER. The reducing end of the glycan contains two N-acetylglucosamines, while nine mannoses are present on three distinct arms. Three glucoses are added to the left arm, which is the same arm that is involved in the quality control (QC) of protein folding (see next section) (Trombetta and Parodi, 2003).

Entamoeba is missing luminal glucosylating and mannosylating enzymes and so makes the truncated, 7-sugar N-glycan precursor dolichol-PP-$Man_5GlcNAc_2$ (Figures 7 and 8) (Samuelson *et al.*, 2005). Five mannoses on this N-glycan

include the left arm, which is involved in the quality control of protein folding. In contrast, *Entamoeba* is missing the middle and the right arms, which are involved in N-glycan associated QC of protein degradation (see next section). Because *Dictyostelium*, which is phylogenetically related to *Entamoeba*, makes a complete 14-sugar N-glycan precursor, it is likely that *Entamoeba* has lost sets of glycosyltransferases in the ER lumen (Samuelson *et al.*, 2005). Similarly, secondary loss of glycosyltransferases best explains the diversity of N-glycan precursors in fungi, which contain 0-14 sugars, and apicomplexa, which contain 2-10 sugars (Samuelson *et al.*, 2005).

The 14-sugar N-glycan precursor of mammals, plants, *Dictyostelium*, and most fungi is transferred to the nascent peptide by an oligosaccharyltransferase (OST), which is composed of a catalytic peptide and 6-7 non-catalytic peptides (Kelleher and Gilmore, 2006). In contrast, the *Entamoeba* OST contains a catalytic peptide and just three non-catalytic peptides, while other protists (e.g. *Giardia* and *Trypanosoma*) have an OST with a single catalytic peptide. This reduced complexity does not likely affect the site of N-glycan addition to the nascent peptides, which is NxS or NxT (the so-called sequon) (Kornfeld and Kornfeld, 1985).

6.5.2 N-glycans and quality control of protein folding.

Protein folding in the lumen of the ER is a complex process that involves N-glycan-dependent and N-glycan-independent QC systems (Helenius and Aebi, 2004; Trombetta and Parodi, 2003). *Entamoeba* has four of five systems present in higher eukaryotes for protein folding (Figure 9).

1) *Entamoeba* has the minimum component parts for N-glycan-dependent QC of protein folding (Helenius and Aebi, 2004; Trombetta and Parodi, 2003; Banerjee, Robbins, and Samuelson, unpublished data). These include a UDP-glucose-dependent glucosyltransferase (UGGT), which adds a single glucose to the left arm of the N-glycans of misfolded proteins and so form $GlcMan_5GlcNAc_2$ (Figure 7). The glucosylated N-glycan is then bound and

refolded by the lectin calreticulin (CRT), which is a chaperone that works with a protein disulfide isomerase (PDI) to make and break disulfide bonds. A glucosidase (Gls2) removes glucose from the well-folded protein, which is transferred to the Golgi by a mannose-binding lectin (ERGIC-53). The *Entamoeba* system is similar to that of mammals and fungi, which add glucose to the $Man_9GlcNAc_2$ precursor to make $GlcMan_9GlcNAc_2$ (Figure 7). Mammals have a second glucosidase to remove glucose from the $Glc_3Man_9GlcNAc_2$ precursor (Figure 7).

2) *Entamoeba* has N-glycan-independent QC of protein folding within the lumen of the ER, which includes the chaperones Hsp70 and Hsp90 (also known as BiP and Grp94, respectively) (Figure 9) (Helenius and Aebi, 2004; Trombetta and Parodi, 2003; Banerjee, Cui, Robbins, and Samuelson, unpublished data). Also involved in this QC system are PDIs; DnaJ proteins that increase the ATPase activity of Hsp70 and Hsp90; and peptidyl-prolyl cis-trans isomerases (PPIases). This N-glycan-independent QC system for protein folding is present in all eukaryotes (Banerjee, Cui, Robbins, and Samuelson, unpublished data).

3) *Entamoeba* and all other eukaryotes have an N-glycan-independent system for ER-associated degradation (ERAD) of misfolded proteins (Figure 9) (Hirsch *et al.*, 2004; Banerjee, Cui, Robbins, and Samuelson, unpublished data). This system is composed of proteins (Sec61 and Der1) that dislocate misfolded proteins from the ER lumen to the cytosol. There a complex of proteins (Cdc48, Npl4, and Ufd1) ubiquinate misfolded proteins, which are then degraded in the proteasome. In contrast, *Entamoeba* and the vast majority of eukaryotes are missing an N-glycan-dependent system of ERAD of misfolded proteins (Helenius and Aebi, 2004; Trombetta and Parodi, 2003; Banerjee, Cui, Robbins, and Samuelson, unpublished data). In this system, the middle arm of $Man_9GlcNAc_2$ is trimmed to $Man_8GlcNAc_2$, which is recognised by a unique mannose-binding lectin (EDEM) before dislocation into the cytosol for degradation (Figure 9).

4) *Entamoeba* has a transmembrane kinase (Ire1), which recognises misfolded proteins in the lumen of the ER and triggers the unfolded protein response (Figure 9) (Patil and Walter, 2001; Banerjee, Cui, Robbins, and Samuelson, unpublished data; and see section 7.2.2). The amoebic unfolded protein response is likely to be different from those of mammals and fungi, because *Entamoeba* is missing an important downstream target, which is a transcription factor called Hac1.

6.5.3 Unique N-glycans.

Mammals make complex N-glycans in the Golgi by trimming back the precursor to $Man_3GlcNAc_2$ and then adding N-acrtyl glucosamine, galactose, sialic acid, and fucose (Figure 8) (Hubbard and Ivatt, 1981). In each case, the activated sugars (UDP-GlcNAc, UDP-Gal, CMP-sialic acid, and GDP-fucose) are transferred from the cytosol to the lumen of the Golgi by a specific nucleotide-sugar transporter (NST) (Hirschberg *et al.*, 1998). In turn, each activated sugar is added to the N-glycans by a specific glycosyltransferase. *Entamoeba* N-glycans are remarkable for two properties. First, the most abundant N-glycan is unprocessed $Man_5GlcNAc_2$ (Figure 7) (Magnelli, Ratner, Robbins, and Samuelson, unpublished data). This N-glycan is recognised by the mannose-binding lectin Concanavalin A, which caps glycoproteins on the *Entamoeba* surface (Silva *et al.*, 1975). Unprocessed $Man_5GlcNAc_2$ is also recognised by the anti-retroviral lectin cyanovirin, which binds $Man_9GlcNAc_2$ on the surface of gp120 (Adams *et al.*, 2004; Magnelli, Ratner, Robbins, and Samuelson, unpublished data). This result suggests the possibility that the anti-retroviral lectin may be active against numerous protists.

Second, complex N-glycans of *Entamoeba*, which are built upon the same $Man_3GlcNAc_2$ core as higher eukaryotes, contain just two additional sugars (galactose and glucose) (Figure 9, D and H) (Magnelli, Ratner, Robbins, and Samuelson, unpublished data). Galactose is added first to both arms of $Man_3GlcNAc_2$, and then glucose is added to galactose. To make these complex N-glycans, *Entamoeba* has NSTs for glucose (UDP-Glc) and galactose (UDP-

Gal) (Bredeston *et al.*, 2005). Glucose is also transferred to N-glycans during the QC of protein folding in the ER, while both galactose and glucose are transferred to proteophosphoglycans (PPGs) (see next section) (Moody-Haupt *et al.*, 2000). Because the complex N-glycans of *Entamoeba* are unique, it is possible that they may be targets of anti-amoebic antibodies.

6.5.4 O-glycans and GPI anchors.

The surface of *E. histolytica* trophozoites is rich in glycoconjugates as shown by the ability of many lectins and carbohydrate specific antibodies to recognise the cell surface (Srivastava *et al.*, 1995; Zhang *et al.*, 2002). Proteophosphoglycans (PPG) constitute the major glycoconjugate of the *E. histolytica* cell surface. PPG is anchored to the cell surface through a GPI moiety (Bhattacharya *et al.*, 1992). The structure of the PPG GPI has been tentatively determined (Moody-Haupt *et al.*, 2000). In most eukaryotes, phosphatidylinositol (PI) is glycosidically linked to the reducing end of de-acetylated glucosamine followed by three mannoses which are in turn attached to the ethanolamine that links the protein to the GPI. However, the GPI anchor of *E. histolytica* PPG was found to have a unique backbone that is not observed in other eukaryotes, namely Gal-Man-Man-GlcN-myoinositol. The intermediate and light subunits of the *E. histolytica* Gal/GalNAc lectin, among other cell surface molecules, are anchored to the cell surface through GPI anchors. Though the structure of the GPI anchors is not known, they are thought to be functionally important (Ramakrishnan *et al.*, 2000). In humans, 23 genes are known to participate in the biosynthesis of GPI anchors. However, only 15 of these were identified in *E. histolytica* (Vats *et al.*, 2005). Interestingly, all the catalytic subunits were identified in *E. histolytica*, the missing genes encoding the accessory subunits suggesting that the biosynthetic pathway may not be significantly different from that in other eukaryotes. The presence of the pathway was also confirmed by detecting the biochemical activities of the first two enzymes - N-acetyl glucosamine transferase and deacetylase. In addition, antisense inhibition of the deacetylase blocked GPI

anchor biosynthesis and reduced virulence of the parasite (Vats *et al.*, 2005). A novel GIPL (glycosylated inositol phospholipid) was also identified in *E. histolytica* (Vishwakarma *et al.*, 2006). Structural studies indicate that a galactose residue is attached to glucosamine as the terminal sugar instead of mannose. This suggests that *E. histolytica* is capable of synthesising unusual GPI-containing glycoconjugates not observed in other organisms.

In PPG, glycans are attached to a peptide backbone by an *O*-Phosphodiester-linkage (O-P glycans). The *E. histolytica* O-P-glycans have galactose at the reducing end followed by a chain of glucoses. *Entamoeba invadens* also has O-P-glycans on its cyst wall proteins but the reducing sugar is a deoxysugar rather than galactose (Van Dellen *et al.*, 2006b). While *Dictyostelium* also has O-P-glycans on glycoproteins in its spore wall, glycoproteins with O-P-glycans are absent from the vast majority of animals and plants (West, 2003).

6.5.5 Significance.

The unique glycans of *Entamoeba* lead to three important evolutionary inferences. First, much of the diversity of eukaryotic N-glycans is due to secondary loss of enzymes that make the 14-sugar lipid-linked precursor, which was present in the common ancestor to extant eukaryotes. Despite the truncated N-glycan precursor, *Entamoeba* has conserved the relatively complex N-glycan-dependent QC system for protein folding. Third, the unique N-glycans and O-P-linked glycans are based upon a novel set of glycosyltransferases, which are present in *Entamoeba* and remain to be characterised molecularly.

7. PROTEINS INVOLVED IN SIGNALLING

7.1 Phosphatases

The combined actions of protein kinases and phosphatases regulate many cellular activities through reversible phosphorylation of proteins. These activities include such basic functions as growth, motility, and metabolism. Although it was once assumed that kinases played the major regulatory role, it is now clear that phosphatases can also be critical participants in some cellular

events (Li and Dixon, 2000). There are few publications on the role of phosphatases in *E. histolytica*, however, several investigators have established a role for phosphatases in proliferation, and growth. Chaudhuri *et al.* (1999) observed that there was an increase in phospho-tyrosine levels in serum starved, growth inhibited, *E. histolytica* cultures. Upon the additional serum and subsequent growth simulation, an increase in tyrosine phosphatase activity occurred. These investigators also demonstrated that genistein, a tyrosine kinase inhibitor, had no effect on growth, while the addition of sodium orthovanadate, a phosphatase inhibitor, produced a major decrease in cell proliferation. Membrane-bound and secreted acid phosphatase activities have been detected in *E. histolytica* (Aguirre-Garcia *et al.*, 1997; Anaya-Ruiz *et al.*, 1997). The secreted acid phosphatase activity is absent from *E. dispar* (Talamas-Rohana *et al.*, 1999). This secreted acid phosphatase was found to have phosphotyrosine hydrolase activity, and caused cell rounding and detachment of HeLa cells (Anaya-Ruiz *et al.*, 2003), suggesting that phosphatase activity contributes to the virulence of the organism.

There are four families of phosphatases (Stark, 1996). Members of the PPP (protein phosphatase P) family are serine/threonine phosphatases, and include PP1, PP2A, and PP2B (calcineurin-like) classes. The PPM (protein phosphatase M) family phosphatases also dephosphorylate serine/threonine residues but are unrelated to the PPP family proteins. A third family consists of protein tyrosine phosphatases (PTP) and dual phosphatases. Low molecular weight phosphatases make up the fourth family. In eukaryotic cells, greater than 99% of protein phosphorylation is on serine or threonine residues (Chinkers, 2001). Human cells have about 500 serine/threonine phosphatases and 100 tyrosine phosphatases (Hooft van Huijsduijnen, 1998; Hunter, 1995). *Saccharomyces cerevisiae* has 31 identified or putative protein phosphatases (Stark, 1996). *E. histolytica* has over 100 putative protein phosphatases. Only a few of these phosphatases have potential transmembrane domains. Some *E. histolytica* phosphatases have varying numbers of leucine-rich-repeats (LRR).

The LRR domain is thought to be a site for protein:protein interactions (Hsiung *et al.*, 2001; Kobe and Deisenhofer, 1994). LRR domains have been found in a few kinases, but had not been identified in any phosphatases until recently (Gao *et al.*, 2005).

7.1.1 Serine/Threonine Protein Phosphatases

Members of the PPP family of protein phosphatases are closely related metalloenzymes, and complex with regulatory subunits. In contrast, PPM family members are generally monomeric, ranging 42-61 kDa in size. By Blast analysis, the serine/threonine protein phosphatases of *E. histolytica* are most closely related to PPP phosphatases PP2A, PP2B, and PPM phosphatase PP2C.

7.1.1 (a) PP2A and PP2B (Calcineurin-like) serine/threonine phosphatases

PP2A phosphatases are trimeric enzymes consisting of catalytic, regulatory, and variable subunits (Wera and Hemmings, 1995). Calcineurin is a calcium-dependent protein serine/threonine phosphatase (Rusnak and Mertz, 2000). Orthologues of calcineurin are widespread, from yeast to mammalian cells. Calcineurin is a heterodimeric complex with catalytic (CaNA) and regulatory (CaNB) subunits. CaNA ranges in size from 58-64 kDa. Its conserved domain structure includes a catalytic domain, a CaNB-binding domain, a calmodulin binding domain, and an autoinhibitory (AI) domain. The binding of CaNB and calmodulin activates CaNA. CaNB subunit is 19 kDa, contains 4 EF hand calcium binding motifs and has similarity to calmodulin. The binding of calmodulin releases the autoinhibitory domain and results in activation of the phosphatase. Deletion of the AI domain results in a constitutively active protein. Calcineurin is specifically inhibited by cyclosporin A and FK506. Cyclosporin A and FK506 first bind to specific proteins, cyclophilin A and FK506BP, respectively, then bind to CaNA at the CaNB binding site. Cyclophilin A has been identified in *E. histolytica* and treatment with cyclosporin A decreases growth and viability (Carrero *et al.*, 2000; Carrero *et al.*, 2004; Ostoa-Saloma *et al.*, 2000).

The *E. histolytica* genome has 51 PP2A and calcineurin-like protein phosphatases. The Pfam motif that classifies proteins as PPP phosphatases is Metallophos (PF00149, calcineurin-like phosphoesterase). This motif is also found in a large number of proteins involved in phosphorylation, including DNA polymerase, exonucleases and other phosphatases. The genome annotation identifies three loci as CaNA orthologues. However, due to the similarity among this family of phosphatases, it is difficult to tell by sequence analyses alone those that are calcium-dependent. Identification of CaNA will have to be confirmed experimentally.

Two of the PPM phosphatases contain a TPR domain (PF00515). TPR is thought to be involved in protein:protein interactions (Das *et al.*, 1998). Activities that have been ascribed to TPR include regulatory roles, lipid binding and auto-inhibition.

7.1.1 (b) PP2C phosphatases

PP2C phosphatases are also widespread and are often involved in terminating/attenuating phosphorylation during the cell cycle or in response to environmental stresses such as osmotic and heat shock (Kennelly, 2001). Thirty-five genes were identified as PP2C phosphatases. These proteins can be divided into three broad categories: 1) PP2C domain only- small (235-381 amino acids), 2) PP2C domain only- large (608-959 amino acids), and 3) PP2C with LRR domains.

7.1.2 Tyrosine phosphatases (PTP)

Tyrosine phosphorylation-dephosphorylation is a key regulatory mechanism for many aspects of cell biology, and development (Li and Dixon, 2000). PTPs are a large class of enzymes that have catalytic domains of ~300 amino acids. Forty of these residues are highly conserved (Hooft van Huijsduijnen, 1998). PTPs can be divided into membrane (receptor) and non-membrane (soluble)

PTPs (Li and Dixon, 2000). The soluble PTP group includes those that contain conserved SH2, PEST, Ezrin, PDZ, or CH2 domains. Two other classes of PTPs are the low molecular weight and dual phosphatases. *Saccharomyces cerevisiae* lacks classic PTPs but does contain dual phosphatases, such as the MAP kinase kinases.

E. histolytica has only four potential PTPs none of which are receptor PTPs, (*i.e.* PTPs with recognisable transmembrane spanning regions). Two of the PTPs (XM_650778, XM_645883) are 350 and 342 amino acids in length and share 48% identity. Neither of these phosphatases has any other recognisable conserved domain. Non-receptor type 1 PTPs are the closest match to these proteins (Li and Dixon, 2000). Membrane and secreted forms of a PTP that cross-react with anti human PTP1B have been reported in *E. histolytica* (Aguirre-García *et al.*, 2003; Talamas-Rohana *et al.*, 1999). Both forms have an apparent molecular weight of 55 kDa and disrupt host actin stress fibers. However, since none of the putative PTPs identified by the genome project appear to encode secreted or membrane forms it is unlikely that these loci represent these previously reported PTP1B cross-reacting proteins.

A third PTP contains a protein tyrosine phosphatase like protein (PTPLA) domain (PF04387). The PTPLA domain is related to the catalytic domains of tyrosine kinases, but it has an arginine for proline substitution at the active site (Uwanogho *et al.*, 1999). It is not yet clear whether this family of proteins actually has phosphatase activity or serves some other regulatory role.

An orthologue of a low molecular weight PTP has also been identified. Low molecular weight protein tyrosine phosphatases have been found in bacteria, yeast, and mammalian cells (Ramponi and Stefani, 1997). They are not similar to other PTPs except in the conserved catalytic domain.

7.1.3 Dual-specificity protein phosphatases

Dual specificity PTPs (DSP) can hydrolyse both tyrosine and serine/threonine residues, though they hydrolyse phosphorylated tyrosine substrates 40-500 fold faster (Zhang and Van Etten, 1991). In other organisms, DSPs are found mostly in the nucleus and have roles in cell cycle control, nuclear dephosphorylation and inactivation of MAP kinase.

The *E. histolytica* genome has 23 sequences related to DSPs. They fall into three main subclasses: those with the DSP domain only, those with DSP plus a variable number (1-5) of LRRs, and those with the Rhodanese homology domain (RHOD; IPR001763). Rhodanese is a sulphurtransferase involved in cyanide detoxification. Its active site, RHOD, is also found in the catalytic site of the dual specificity phosphatase CDC25 (Bordo and Bork, 2002).

7.1.4 Leucine Rich Repeats (LRRs)

LRRs are tandem arrays of 20-29 amino acid, leucine-rich motifs. LRRs have been found in a number of proteins with varied functions including enzyme inhibition, regulation of gene expression, morphology and cytoskeleton formation (Kobe and Deisenhofer, 1994). LRRs are thought to provide versatile sites for protein:protein interaction and have been found linked to a variety of secondary domains. Most LRRs form curved horseshoe-shaped structures with "a parallel beta sheet on the concave side and mostly helical elements on the convex side" (IPR001611).

The LRR_1 Pfam is the second most abundant Pfam domain found in the *E. histolytica* genome (Table 3). The LRR motifs in *E. histolytica* most closely resemble the LRR found in BspA (section 2.7; Davis *et al.*, 2006). Several *E. histolytica* proteins that contain LRRs are associated with other recognised domains. These include the protein phosphatases PP2C and DSP, as well as protein kinase (PK), F-box (PF00646), gelsolin/villin headpiece (IPR007122),

DNA J (IPR001623), Band 41 (B41;IPR000299), WD-40 (IPR001680), and Zinc binding (IPR000967) domains. The association of LRRs with phosphatases is unusual. One published example is the phosphatase that dephosphorylates the kinase Akt (Gao *et al.*, 2005). Fungal adenylate cyclases have both LRR and PP2C-like domains but this is not a wide spread feature of adenylate cyclases in other species (Mallet *et al.*, 2000; Yamawaki-Kataoka *et al.*, 1989). The LRR may be a site for interaction with phosphorylated residues in *E. histolytica*. This speculation is supported by the example of the Grr1 protein of yeast, which contains an F-box and a LRR (Hsiung *et al.*, 2001). Grr1 is involved in ubiquitin-dependent proteolysis. The LRR domain of Grr1 binds to phosphorylated targets in the proteosome complex. Another example is the fission yeast phosphatase regulatory subunit, Sds22, which also has LRRs (MacKelvie *et al.*, 1995). The LRR containing phosphatases of *E. histolytica* may represent fusions of regulatory and catalytic subunits.

7.2 Kinases

7.2.1. Cytosolic kinases.

Eukaryotic protein kinases are a superfamily of enzymes, which are important for signal transduction and cell-cycle regulation. Six families of Serine/Threonine kinases (STKs), which include AGC, Ste, CK1, CaMK, CMGC, and TKL (tyrosine kinase-like), have conserved aspartic acid and lysine amino acids in their active sites and phosphorylate serine or threonine on target proteins (Hanks and Hunter, 1995). Tyrosine kinases (TK), which lack active site lysine, phosphorylate tyrosine on target proteins. Phosphorylated tyrosine is in turn recognised by Src-homology 2 (SH2) domains that are present on some kinases and other proteins. All seven families of protein kinases are present in metazoa and in *D. discoideum*, while plants lack TK, and *S. cerevisiae* lacks both TK and TKL.

Over 150 predicted *E. histolytica* cytosolic kinases, those that lack signal peptides and trans-membrane helices, can be identified, including

representatives of each of the seven groups of kinases (AGC, CAMK, CK1, CMGC, STE, TKL, and TK) (Loftus *et al.*, 2005; Cui and Samuelson, unpublished data). Two predicted *E. histolytica* TKs, which group with human TKs in phylogenetic trees, contain an AAR peptide in the active site and a Kelch domain at the C-terminus (Gu and Gu, 2003). Four cytosolic protein kinases contain C-terminal SH2 domains, which bind phosphorylated tyrosine residues. Phosphotyrosine has been identified in *E. histolytica* using specific antibodies (Hernandez-Ramirez *et al.*, 2000). The thirty-five predicted cytosolic *E. histolytica* TKLs include some that contain Leu-rich repeats (LRR) and ankyrin repeats at their N-termini. In contrast, the vast majority of *Entamoeba* cytosolic kinases lack accessory domains.

7.2.2. Receptor-kinases.

Five distinct families of eukaryotic proteins have an N-terminal ectoplasmic domain, a single transmembrane helix, and a C-terminal cytoplasmic kinase domain (Blume-Jensen and Hunter, 2001). Ire-1 transmembrane kinases, which are present in *S. cerevisiae*, plants, and metazoa, detect unfolded proteins in the lumen of the ER and help splice a transcription factor mRNA by means of a unique C-terminal ribonuclease (Patil and Walter, 2001). Receptor tyrosine kinases (RTKs), which include growth hormone and epidermal growth factor (EGF) receptors, are restricted to metazoa and have a diverse set of N-terminal ectoplasmic domains and a conserved C-terminal cytosolic TK (Schlessinger, 2000). Receptor serine/threonine kinases (RSK) of metazoa and receptor-like kinases (RLKs) of plants each contain a C-terminal TKL domain (Massague *et al.*, 2000; McCarty and Chory, 2000; Shiu and Bleecker, 2001). Phylogenetic analyses suggest that plant RLKs, animal RSKs, and animal RTKs each form monophyletic groups, and that plant RLKs closely resemble cytosolic TKLs of animals called Pelle or IRAK (Shiu and Bleecker, 2001).

E. histolytica contains >80 novel receptor RSKs, each of which has a N-terminal signal sequence, a conserved ectoplasmic domain, a single

transmembrane helix (TMH), and a cytosolic kinase domain (Beck *et al.*, 2005). The largest group of *E. histolytica* RSKs has a CXXC-rich ectoplasmic domain with 6 to 31 internal repeats that each contains 4 to 6 cysteine residues (Figure 10). Very similar CXXC-rich domains are present in the ectoplasmic domain intermediate subunit of the Gal/GalNAc lectin (section 3.1.3). CXXC-rich domains are also present in hypothetical secreted proteins of *E. histolytica*, while cysteine-rich domains are also present in the heavy subunit of the Gal/GalNAc lectin and at the cytosolic aspect of some cysteine proteases (Figure 10).

Ectoplasmic domains of other large families of *Entamoeba* RSKs have one or two 6-Cys domains at the N-terminus and four 6-Cys domains proximal to the plasma membrane. There are no plasma membrane proteins or secreted proteins with similar domains. A minority of RSKs do not contain Cys-rich ectoplasmic domains. Numerous *Entamoeba* RSKs are expressed at the same time, but the specific ligands for the *Entamoeba* RSKs have not been identified (Beck *et al.*, 2005).

As discussed in the section on protein folding (6.5.2), *Entamoeba* has an Ire1 transmembrane kinase, which recognises misfolded proteins in the lumen of the ER and triggers the unfolded protein response (Figure 8).

7.2.3 Significance

While most protists lack TK, TKL, receptor-kinases, and Ire1 *E. histolytica* has all four. It is very likely that the *E. histolytica* receptor-kinases, which are extensively duplicated, will have important roles in pathogensis (Beck *et al.*, 2005; Okada *et al.*, 2005). Similarly, trimeric G-proteins and the associated adenyl-cyclases likely have important roles in cyst formation and virulence (Coppi *et al.*, 2002; Frederick and Eichinger, 2004).

7.3 Calcium Binding Proteins

Ca^{2+} signaling plays a crucial role in the pathogenesis of many protozoan parasites, including *E. histolytica* (Ravdin *et al.*, 1985). Many of the calcium-mediated processes are carried out with the help of calcium binding proteins (CaBPs). CaBPs have been identified and characterised in almost all eukaryotic systems. Some of these, such as calmodulin (CaM) and troponin C, have been studied extensively. A number of CaBPs have also been identified in *E. histolytica*. Among these are two related EF-hand containing proteins, grainin 1 and granin 2, which are likely to be localised in intracellular granules (Nickel *et al.*, 2000). Another protein, URE3-BP, was shown to have a transcription regulatory function (Gilchrist *et al.*, 2001). The CaM-dependent secretion of collagenases from electron dense granules has been demonstrated using *E. histolytica* lysate. However, there is as yet no direct molecular evidence for the presence of CaM in *E. histolytica* (Muñoz *et al.*, 1991). The CaM-like protein EhCaBP1 has four canonical EF-hand Ca^{2+} binding domains but no functional similarity to CaM (Yadava *et al.*, 1997). Inducible expression of EhCaBP1 antisense RNA demonstrated this protein's role in actin-mediated processes (Sahoo *et al.*, 2004).

Analysis of the whole genome revealed presence of 27 CaBPs with multiple EF-hand calcium binding domains (Bhattacharya *et al.*, 2006). Many of these proteins are architecturally very similar but functionally distinct from CaM. Moreover, functional diversity was also observed among closely related CaBPs, such as EhCaBP1 and EhCaBP2 (79% identical at the amino acid level; Chakrabarty *et al.*, 2004). Analysis of partial EST and proteomic databases combined with Northern blots and RT-PCR shows that at least one third of these genes are expressed in trophozoites, suggesting that many if not all of the 27 are functional genes (Bhattacharya *et al.*, 2006).

What are the roles of these proteins in the context of *E. histolytica* biology? At present the function of only two EhCaBPs are known, EhCaBP1 and URE3-BP. The rest of the proteins are likely to be Ca^{2+} sensors involved in a number of

different signal transduction pathways. After binding Ca^{2+} these may undergo conformational changes and the bound form then activates downstream target proteins. It is not clear why *E. histolytica* would need so many Ca^{2+} sensors when many other organisms do not. It is likely that with Ca^{2+} being involved in many functions, some of which are localised in different cellular locations, the various CaBPs may participate in different functions that are spatially and temporally separated.

8. THE MITOSOME

One of the expectations for the *E. histolytica* genome project was that it would identify the function of the mitochondrial remnant known as the mitosome (Tovar *et al.*, 1999) or crypton (Mai *et al.*, 1999). Under the microscope mitosomes are ovoid structures smaller than 0.5 μm in diameter (Leon-Avila and Tovar, 2004). While it is now clear that no mitochondrial genome still persists, from both genome sequencing and cellular localisation data (Leon-Avila and Tovar, 2004), the protein complement of the organelle is still somewhat obscure. The number of identifiable mitosomal proteins remains very small and does not provide great insight into the organelle's function. Genes encoding mitochondrial-type chaperonins (cpn60, hsp10 and mt-hsp70) have been identified and appear to be synthesised with amino-terminal signal sequences. The importation machinery has been shown to be conserved with that in true mitochondria (Mai *et al.*, 1999; Tovar *et al.*, 1999) but none of the proteins involved in mitosomal protein import have been identified with certainty.

Other genes encoding putative mitosomal proteins include pyridine nucleotide transhydrogenase (which moves reducing equivalents between NAD and NADP, and acts as a proton pump (Clark and Roger, 1995); only an incomplete gene is present in the assembly), and ADP/ATP transporter (Chan *et al.*, 2005), a P-glycoprotein-like protein (Pgp6), and a mitochondrial type thioredoxin, although the latter two are identified based largely on their amino terminal

extensions. The only enzymatic pathway that is normally mitochondrial in location is iron-sulphur cluster synthesis. Genes encoding homologues of both IscS/NifS and IscU/NifU proteins are present, but uniquely among eukaryotes the *E. histolytica* homologues are not of mitochondrial origin, having been acquired by distinct lateral gene transfer from an ε-proteobacterium (Ali *et al.*, 2004b; van der Giezen *et al.*, 2004). The location of these proteins appears to be cytoplasmic as determined by immunofluorescence, using antibodies against both the native proteins as well as detection of epitope-tagged proteins in transformed *E. histolytica* (Ali and Nozaki, unpublished). The same pathway has been localised to mitosomes in *Giardia* and is also retained in all other organisms with remnant mitochondria. Given the apparently unique non-compartmentalised nature of iron-sulphur cluster synthesis in *E. histolytica* the location of the proteins needs to be confirmed by immuno-electron-microscopy; such experiments are currently underway (Ali and Nozaki, unpublished). The function of the *E. histolytica* mitosome therefore remains an enigma.

9. ENCYSTATION

The infectious stage of *Entamoeba histolytica*, and also that most often used for diagnosis, is the quadrinucleate cyst. Because it is not possible to encyst *E. histolytica* in axenic culture, *Entamoeba invadens*, which is a reptilian parasite, has been used as a model organism for encystation (Eichinger, 2001; Wang *et al.*, 2003). The *E. invadens* cyst wall is composed of three parts: deacetylated chitin (also known as chitosan), lectins that bind chitin (e.g. Jacob and Jessie) or cyst wall glycoproteins (e.g. plasma membrane Gal/GalNAc lectin), and enzymes that modify chitin or cyst wall proteins (e.g. chitin deacetylase, chitinase, and cysteine proteases) (Figure 11).

9.1 Chitin synthases

Chitin fibrils, which are homopolymers of β-1,4-linked N-acetyl glucosamine (GlcNAc), are synthesised by chitin synthases. Chitin synthases share common ancestry with cellulose synthases and hyaluronan synthase. They are

transmembrane proteins with a catalytic domain in the cytosol (Bulawa, 1993), where UDP-GlcNAc is made into a homopolymer and is threaded through the transmembrane domains into the extracellular space. In *Saccharomyces cerevisiae*, four accessory peptides, encoded by the Chs4-7 genes, are necessary for the function of its chitin synthases (Trilla *et al.*, 1999). Remarkably, the *E. histolytica* chitin synthase 2 (EhChs2) complements a *S. cerevisiae* chs1/chs3 mutant and the function of EhChs2 is independent of the four accessory peptides (Van Dellen *et al.*, 2006a). This result suggests the possibility that chimaeras of *E. histolytica* and *S. cerevisiae* chitin synthases may be used to map domains in the *S. cerevisiae* chitin synthase that interact with the accessory peptides.

9.2 Chitin Deacetylases

Chitin fibrils in the cyst wall are modified by deacetylases and chitinases (see section 9.3). There are two *E. invadens* chitin deacetylases, which convert chitin to chitosan (Das *et al.*, 2006). Chitosan is a mixture of N-acetyl glucosamine and glucosamine and so has a positive charge. It is also present in spore walls of *S. cerevisiae* and in lateral walls of *Mucor* (Kafetzopoulos *et al.*, 1993; Mishra *et al.*, 1997). It is likely that the positive charge of chitosan fibrils contributes to the binding of cyst wall proteins, all of which are acidic (de la Vega *et al.*, 1997; Frisardi *et al.*, 2000; Van Dellen *et al.*, 2002b). Monosaccharide analyses of the *E. invadens* cyst walls following treatment with SDS to remove proteins strongly suggest that chitosan is the only sugar homopolymer present (Das *et al.*, 2006).

9.3 Chitinases

Entamoeba species encode numerous chitinases with a conserved type 18 glycohydrolase domain (de la Vega *et al.*, 1997). Recombinant *Entamoeba* chitinases have both endo- and exo-chitinase activities. Two other domains are important in *Entamoeba* chitinases: 1) At the N-terminus is a unique 8-Cys chitin-binding domain (CBD), which is also present as a single domain in *E.*

histolytica Jessie lectins (Figure 11) (Van Dellen *et al.*, 2002b). Chitinase and Jessie-3 lectin bind to the *E. invadens* cyst wall by means of this 8-Cys CBD (Van Dellen *et al.*, submitted). This *E. histolytica* chitinase CBD has the same function as CBDs in chitinases of fungi, nematodes, insects, and bacteria, but has no sequence similarity (i.e. it has arisen by convergent evolution) (Shen and Jacobs-Lorena, 1999). 2) Between the CBD and chitinase domains of *Entamoeba* species are low complexity sequences that contain heptapeptide repeats (Ghosh *et al.*, 2000). These polymorphic repeats may be used to distinguish isolates of *E. histolytica* within the same population and may be able to discriminate among isolates from New and Old World (Haghighi *et al.*, 2003). These polymorphic repeats, which are rich in serine and resemble mucin-like domains in other glycoproteins, may also be the sites for addition of *O*-phosphodiester linked sugars (see section 6.5.4).

9.4 Jacob lectins

Chitin fibrils in the cyst wall of *E. invadens* are cross-linked by Jacob lectins, which contain 3 to 5 unique 6-Cys CBDs (Frisardi *et al.*, 2000). *E. invadens* has at least nine genes encoding Jacob lectins, and the mRNA levels from each gene increase during encystation (Van Dellen *et al.*, submitted). In addition, at least six Jacob lectin proteins are present in *E. invadens* cyst walls (Van Dellen *et al.*, submitted). Between the CBDs, Jacob lectins have low complexity sequences that are rich in serine as in the case of chitinase [5]. Jacob lectins are post-translationally modified in two ways. First, they are cleaved by cysteine proteinases at conserved sites in the serine- and threonine-rich spacers between CBDs. Second, they have *O*-phosphodiester-linked sugars added to serine and threonine residues. *O*-phosphodiester-linked glycans are also present in proteophosphoglycans (PPGs) on the surface of *E. histolytica* trophozoites (Moody-Haupt *et al.*, 2000).

9.5 Gal/GalNAc lectins

The Gal/GalNAc lectins present on the surface of *E. histolytica* trophozoites have been described above (section 3.1) and in the literature (Mann *et al.*, 1991; Petri *et al.*, 2002). Their possible role in encystation is suggested by two independent experiments. First, the signal for encystation likely depends in part on aggregation of *E. invadens*, which is inhibited by exogenous galactose (Coppi and Eichinger, 1999). Aggregated *E. invadens* secrete catecholamines, which in an autocrine manner stimulate amoebae to encyst (Coppi *et al.*, 2002). Second, in the presence of excess galactose, *E. invadens* forms wall-less cysts that contain four nuclei and makes Jacob lectins and chitinase (Frisardi *et al.*, 2000). Because *E. invadens* trophozoites have a Gal/GalNAc lectin on their surface that is capable of binding sugars on Jacob lectin, and because Jacob lectins have no carboxy-terminal transmembrane helix or GPI-anchor, it is likely that the cyst wall is bound to the plasma membrane by the Gal/GalNAc lectin.

9.6 Summary and Comparisons

Similar to the cyst wall of *Giardia*, the cyst wall of *E. invadens* is a single homogeneous layer and contains a single homopolymer, chitosan (Figure 11) (Frisardi *et al.*, 2000; Gerwig *et al.*, 2002; Shen and Jacobs-Lorena, 1999). In contrast, *S. cerevisiae* spore walls have multiple layers and contain β-1,3-glucans in addition to chitin, while *Dictyostelium* walls have multiple layers and contain N-acetyl galactosamine polymers in addition to cellulose (West, 2003; Yin *et al.*, 2005).

Similar to *Dictyostelium* and in contrast to fungi, the vast majority of *Entamoeba* cyst wall glycoproteins are released by SDS (Van Dellen *et al.*, submitted; Frisardi *et al.*, 2000; West, 2003; Yin *et al.*, 2005). While some *Dictyostelium* cyst wall proteins have been shown to be cellulose-binding lectins, all of the proteins bound to the cyst wall of *E. invadens* have 6-Cys CBDs (Jacob lectins) or 8-Cys CBDs (Jessie 3 lectin and chitinase) (Frisardi *et al.*, 2000; Van Dellen *et al.*, 2002b; Van Dellen *et al.*, submitted). In the same

way that *Giardia* cyst wall protein 2 is cleaved by a cysteine proteinase, Jacob lectins are cleaved by an endogenous cysteine proteinase at sites between chitin-binding domains (Touz *et al.*, 2002).

Like *Dictyostelium* spore coat proteins and insect peritrophins, cysteine-rich lectin domains of *E. invadens* cyst wall proteins are separated by serine- and threonine-rich domains that are heavily glycosylated (Frisardi *et al.*, 2000; West, 2003; Yin *et al.*, 2005; Van Dellen *et al.*, submitted). *S. cerevisiae* cyst wall proteins also have extensive serine- and threonine-rich domains that are heavily glycosylated (Yin *et al.*, 2005). These glycans likely protect proteins in cyst walls or fungal walls from exogenous proteases. While glycoproteins of the *E. invadens* cyst wall and *Dictyostelium* spore coat contain *O*-phosphodiester-linked glycans, *S. cerevisiae* wall glycoproteins contain *O*-glycans (Gemmill and Trimble, 1999; West *et al.*, 2005).

Like *S. cerevisiae*, *E. invadens* has enzymes in its wall that modify chitin (Yin *et al.*, 2005). Similar to chitinases of *S. cerevisiae* and bacteria, *E. invadens* chitinase has a CBD in addition to the catalytic domain (Kuranda and Robbins, 1991). It is likely that the CBD is present to localise chitinase to the cyst wall (*E. invadens*) or cell wall (*S. cerevisiae*). Finally, while *E. invadens* uses catecholamines as autocrines for encystation, *Dictyostelium* uses cAMP as an autocrine for sporulation (Coppi *et al.*, 2002; Kriebel and Parent, 2004). An important goal of future research will be to translate what is known about the *E. invadens* cyst wall to that of *E. histolytica*.

10. EVIDENCE OF LATERAL GENE TRANSFER IN THE *E. hISTOLYTICA* GENOME

Lateral (or horizontal) gene transfer (LGT) plays a significant role in prokaryotic genome evolution, contributing up to ~20% of the content of a given genome (Doolittle *et al.*, 2003). LGT has therefore been an important means of acquiring new phenotypes, such as resistance to antibiotics and new

physiological and metabolic capabilities, that may permit or facilitate adaptation to new ecological niches (Koonin *et al.*, 2001; Lawrence, 2005b; Ochman *et al.*, 2000). More recently, data from microbial eukaryote genomes suggest that LGT has also played a role in eukaryotic genome evolution, particularly among protists that eat bacteria (Andersson, 2005; Doolittle, 1998; Doolittle *et al.*, 2003; Lawrence, 2005a; Richards *et al.*, 2003). *Entamoeba histolytica* lives in the human gut, an environment that is rich in microorganisms and where LGT is thought to be common between bacteria (Shoemaker *et al.*, 2001). The *E. histolytica* genome thus provides a nice model for investigating prokaryote to eukaryote LGT. In the original genome description (Loftus *et al.*, 2005) 96 putative cases of LGT were identified using phylogenetic analyses of the *E. histolytica* proteome. These have now been reanalysed in the light of more recently published (August 2005) eukaryotic and prokaryotic genomes. This has allowed evaluation of how previous inferences were influenced by the sparse sampling of eukaryotic and prokaryotic genes and species available at the time of the original analysis. Sparse gene and species sampling is, and is likely to remain, a very serious problem for reconstructing global trees and inferring LGT (Andersson *et al.*, 2001; Richards *et al.*, 2003; Salzberg *et al.*, 2001). Thus, although ecologists differ in their claims for the extent of the unsampled microbial world, they all agree that those species in culture, and the even smaller subset for which genome data exist, represent the smallest tip of a very large iceberg.

10.1 How Do The 96 LGT Cases Stand Up?

As before (Loftus *et al.*, 2005), Bayesian and maximum likelihood distance bootstrap phylogenetic analyses were used to identify putative LGT using the following *ad hoc* conservative criteria: Putative LGT was inferred where either no other eukaryote possessed the gene, or where the *E. histolytica* sequence was grouped with bacteria and separated from other eukaryotes by at least two strongly supported nodes (bootstrap support >70%, posterior probabilities >0.95). In cases where tree topologies were more weakly supported but still

suggested a possible LGT, bootstrap partition tables were examined for partitions where the *E. histolytica* sequence clustered with another eukaryote. If no such partitions were found that gene was considered to be a putative LGT. Table 8 lists the results of the new analyses and also gives BlastP statistics for each sequence.

A total of 41 LGT remain as strongly supported as before based upon the original criteria. For the remaining 55 tree topologies, support for recent LGT into the *Entamoeba* lineage is not as strong as before. For 27 of these 55 trees, two strongly supported nodes separating *E. histolytica* from other eukaryotes has been reduced to only one well-supported node. However, close scrutiny of the bootstrap partition tables for these trees revealed that, as before, there are no trees in which *E. histolytica* is found together with another eukaryote. Thus, LGT still remains the strongest hypothesis to explain 68 (70%) of the original 96 tree topologies. In a further 14 cases, the position of *E. histolytica* among prokaryotes and eukaryotes was not well supported. The taxonomic sampling of eukaryotes in these trees is very patchy and the trees do not depict consensus eukaryotic relationships. Thus, although the trees do not fulfill the conservative criteria for LGT they also do not provide strong support for the alternative hypothesis, that the *E. histolytica* genes were vertically inherited from a common ancestor shared with all other eukaryotes.

In nine trees *E. histolytica* either clustered with a single newly published eukaryotic sequence, or such a relationship could not be ruled out. In six of these nine trees *E. histolytica* and *Trichomonas vaginalis* grouped together, and two trees grouped *E. histolytica* with the diatom *Thalassiosira* (for example see Figure 12). Such trees are also not easy to explain within the current consensus for eukaryotic relationships (Baldauf, 2003). Similar topologies have been previously reported for other eukaryotes (Andersson, 2005). The explanations advanced to explain the absence of the gene in other eukaryotes include massive gene loss from multiple eukaryotic lineages, or LGT between the

eukaryotic lineages concerned. *Entamoeba* species can ingest both eukaryotes and prokaryotes and it has been suggested that LGT between eukaryotes, subsequent to one lineage acquiring the gene from a prokaryote, could explain such peculiar tree topologies and sparse distribution (Andersson, 2005). The fact that six of the nine cases recover a relationship between *Entamoeba* and *Trichomonas*, whose relatives often share the same niche, is consistent with this idea. In prokaryotes, recent large-scale analyses support the hypothesis that species from the same environment may share a set of niche specific genes (Beiko *et al.*, 2005; Mira *et al.*, 2004).

For five trees, the *E. histolytica* gene now appears to be present in eukaryotes from a different taxonomic group and the analysis cannot exclude a common origin for all eukaryotic sequences. Thus, for about 5% of the original 96 cases the simplest explanation is no longer LGT, but vertical inheritance from a common ancestor shared with other eukaryotes.

10.2 Where Do The Genes Come From?

As before, certain prokaryotic groups are favoured as the potential donors of LGT genes in the *E. histolytica* genome (Loftus *et al.*, 2005). In 15 well-resolved trees *E. histolytica* is recovered next to a member of the Bacteroidetes/Chlorobii group. Bacteroidetes/Chlorobii are abundant members of the intestinal microflora (Shoemaker *et al.*, 2001) providing plenty of opportunity for LGT to occur. Members of the Bacteroidetes/Chlorobii and Fusobacterium (one tree) groups are all obligate anaerobes. This bias is consistent with the idea that prokaryotic and eukaryotic cohabitants of the same anaerobic niche are sharing genes (Andersson *et al.*, 2001; Beiko *et al.*, 2005; Lawrence, 2005a). For example, Figure 13 shows an intriguing example where the *T. vaginalis* gene clusters with members of the Bacteroidetes/Chlorobii and *E. histolytica* clusters with *Fusobacterium*.

10.3 What Kinds of Gene Are Being Transferred?

Most of the 68 laterally transferred genes that can be assigned to a functional category encode enzymes involved in metabolism (Figure 14). This is consistent with the complexity hypothesis, which posits that LGT of genes involved in processing a single substrate are more likely to be transferred than those genes encoding proteins that interact with many other cellular components, such as ribosomal proteins for example (Jain *et al.*, 1999). Mapping the LGT enzymes on the *E. histolytica* metabolic pathway (Loftus *et al.*, 2005) indicates that LGT has affected some important pathways, including iron-sulphur cluster biosynthesis, amino acid metabolism, and nucleotide metabolism. Since only eight of the 68 LGT have obvious homologues in the human genome, the proteins are potentially specific to the parasite and may thus be worth exploring as potential drug targets. The rest of the LGT cases involve hypothetical or unclassified proteins.

11. MICROARRAY ANALYSIS

Microarray-based analyses can be utilised in conjunction with genome sequencing to assign functional roles to annotated genes and to clarify genomic architecture. A number of groups have utilised DNA microarrays in *E. histolytica* (made from random genomic DNA fragments or long or short oligonucleotides based on annotated genes) to successfully study transcriptional differences between virulent and avirulent *E. histolytica* as well transcriptional responses to heat shock, collagen and calcium exposure, tissue invasion, and cyst development (Debnath *et al.*, 2004; Gilchrist *et al.*, 2006; MacFarlane and Singh, 2006; Weber *et al.*, 2006; Davis *et al.*, 2007; Ehrenkaufer *et al.*, 2007). Additionally, using a genomic DNA microarray, comparative genomic hybridisations (CGH) between strains and species of *Entamoeba* have been performed (Shah *et al.*, 2005).

Some interesting aspects of amoebic biology have been uncovered using DNA microarray based expression profiling. To investigate the hypothesis that virulence determinants will be more highly expressed in virulent strains, the

transcriptomes of virulent and avirulent *Entamoeba* species and strains have been studied. It has been confirmed that a number of known virulence determinants have decreased expression in avirulent *Entamoeba* (MacFarlane and Singh, 2006; Davis *et al.*, 2007). A genomic DNA microarray composed of 2,110 genes identified 29 genes with decreased expression in both an attenuated *E. histolytica* strain (Rahman) and the avirulent *E. dispar* (strain SAW760) (MacFarlane and Singh, 2006), while an oligonucleotide microarray composed of 6,242 genes identified 152 genes with a higher level of expression in the virulent *E. histolytica* HM-1:IMSS than in the attenuated Rahman strain (Davis *et al.*, 2007). A majority of these genes are annotated as hypothetical and whether these genes encode novel virulence factors will require genetic analysis of their functions. A peroxiredoxin gene identified as having decreased expression in *E. histolytica* Rahman has been shown to be a virulence factor (Davis *et al.*, 2006), indicating that these comparisons between virulent and avirulent strains are likely to be a fruitful avenue of investigation.

In other microarray based studies, the large family of transmembrane receptor kinases identified in *E. histolytica* has been found to be differentially expressed under *in vitro* trophozoite culture conditions (Beck *et al.*, 2005). One can easily envision that these kinases may have roles in signaling, allowing the parasite to adapt to its ever changing environmental milieu. A substantial transcriptional response to heat shock has been demonstrated (Weber *et al.*, 2006), and interestingly lectin gene family members were identified as being differentially regulated under heat shock conditions.

The most comprehensive microarray data to date used a whole genome short oligonucleotide microarray (based on the Affymetrix platform) to profile the transcriptional changes that occur as the parasite colonises and invades the host colon (Gilchrist *et al.*, 2006). Using a mouse model of colitis, in which the microscopic features replicate human disease and substantial pathology can be seen, the transcriptional response of parasites was assayed soon after

colonisation (1 day after injection into the caecum) and in a long-term (29 days) disease state. Overall, 326 genes were modulated at day 1 after infection, 109 at 29 days after infection, and 88 at both time points. A number of the well-characterised "virulence determinants" in *E. histolytica* were highly expressed under all conditions tested and not transcriptionally modulated, although some members of the cysteine proteinase gene family were highly regulated during tissue invasion. A summary of the genes and gene families that have been identified as being transcriptionally active under the conditions mentioned above are listed in Table 9.

The life cycle of *E. histolytica* involves transition between the trophozoite stage, responsible for colonisation as well as invasive disease and the cyst, responsible for infection transmission. Despite its central role, little is known about cyst development in *E. histolytica*, largely due to our inability to generate *E. histolytica* cysts in axenic culture. Using a whole genome microarray and xenic cultures of recently isolated *E. histolytica* strains that contained spontaneously produced cysts, a cyst transcriptome was developed that identified 1,439 developmentally regulated genes (672 cyst-specific and 767 trophozoite-specific genes; Ehrenkaufer *et al.*, 2007). This first large-scale insight into encystation indicates that ca. 15% of *E. histolytica* genes are transcriptionally controlled in this developmental pathway. Among the genes identified were a number of stage-specific cysteine proteinases, transmembrane kinases, transcriptional regulators, and other potential initiators of the developmental cascade. Future characterisation of these genes and pathways will provide important insights into developmental processes in this parasite.

The above microarray studies used expression data to identify interesting genes and pathways potentially involved in amoebic pathogenesis or development. In another application of microarrays, comparative genomic hybridisations (CGH) identified a number of interesting genomic characteristics of *Entamoeba* (Shah *et al.*, 2005). The *E. histolytica* genome project revealed that a large number of

genes are multi-copy or members of highly similar gene families. Due to the repetitive nature of the genome there has been difficulty with genome assembly and thus the large number of gene duplications could have represented an assembly artifact. The data from CGH confirmed the high copy number of a significant portion (ca. 14%) of the genome and validated the genome assembly. Additionally, genome-wide genetic diversity was demonstrated among strains of *E. histolytica* (Shah *et al.*, 2005) including the observation that the attenuated *E. histolytica* strain Rahman had a unique genetic pattern suggesting the possibility that a genomic signature may correlate with invasive potential. Since genome sequencing for different *E. histolytica* strains, including clinical isolates, is unlikely the promise of CGH to study genetic diversity and identify genotype-phenotype associations is substantial.

E. dispar, the closely related but avirulent species, had been identified early on as having some genetic divergence from the virulent *E. histolytica*. CGH analysis of *E. histolytica* and *E. dispar* revealed a significant amount of difference between the two species. Whether the genetic drift in these genes is responsible for the non-invasive phenotype of *E. dispar* is not known, but the work has highlighted a number of genes for further functional analyses.

Taken together the DNA microarray analyses of *Entamoeba* have been useful to begin to dissect the genome of this parasite and provide functional context to the genes identified in the genome sequencing effort. Future directions will include analysis of the parasite transcriptome in invasive hepatic disease as well as further characterisation of the developmental conversion to the cyst form. Those data may be useful in the development of novel diagnostic and therapeutic options. Additionally, genetic approaches can now be applied to definitively assign a role for these genes in amoebic biology and pathogenesis.

12. FUTURE PROSPECTS FOR THE *E. HISTOLYTICA* GENOME

Although the genome of *E. histolytica* is not yet complete it has already revealed much about the biology of the parasite. There appear to be forces acting to compact the genome, leading to a reduction in the coding region and intron length of genes, and resulting in the loss of numerous metabolic pathways. However, there are also opposing evolutionary forces as many gene families have expanded. This applies particularly to genes involved in signaling and trafficking that allow the parasite to sense and respond to its environment, a necessary adaptation for a predatory protist. Unfortunately, it is difficult at present to understand the genome structure on a macro scale due to the fragmented nature of the current assembly. In other parasites, genome structure has been vital to unraveling important biological processes, such as antigenic variation in *T. brucei* and identification of rifin genes in *P. falciparum.* Until the *E. histolytica* genome is complete we will not know what else remains to be uncovered. Efforts are already underway to complete the genome by first generating a HAPPY map (Dear and Cook, 1993). Over 2000 markers are being designed at approximately 25 kb intervals across all contigs. Using PCR, co-segregation analysis allows the identification of contigs that are physically linked in the genome. This will allow the ordering and orientation of the contigs and will facilitate gap closure. Shotgun genome sequencing projects of *E. invadens* and *E. dispar* are underway (Loftus and Hall, 2005). At present the *E. invadens* genome appears to assemble with fewer problems than were encountered with that of *E. histolytica.* It is anticipated that an essentially complete *E. invadens* genome sequence will be obtained, enabling extensive comparative analyses to be made, and facilitating the study of pathogenicity, host interaction and the evolutionary forces acting on the genome.

ACKNOWLEDGEMENTS

This review is dedicated to Louis S. Diamond without whose pioneering studies the sequencing of the genome would have been impossible. The genome sequencing was supported by grants from the Wellcome Trust (064057) and the National Institute of Allergy and Infectious Disease (5R01AI046516-03).

UCMA is supported by a Wellcome Trust project grant (075796) awarded to TME and RPH. Analyses by MH and MD were supported by grant P15960 from the Austrian Science Fund, Vienna, Austria. Analyses by CM and AL were supported by grant 1RO3 TW007314-01 from the Fogarty International Center of the National Institutes of Health, USA. Analyses by NG were supported by an INCO-DEV grant in the fifth framework program of the European Union. Analyses by IB were supported by the Deutsche Forschungsgemeinschaft (DFG), BR 1744/7-1. NH would like to thank Lynn Schriml and Regina Cer for preparing the web pages with the Tables and ML would like to thank Julia Winkelmann for preparing table 6.

REFERENCES

Ackers, J. P., Dhir, V., and Field, M. C. (2005). A bioinformatic analysis of the RAB genes of *Trypanosoma brucei*. *Molecular and Biochemical Parasitology* **141**, 89-97.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-95.

Adams, E. W., Ratner, D. M., Bokesch, H. R., McMahon, J. B., O'Keefe, B. R., and Seeberger, P. H. (2004). Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. *Chemistry & Biology* **11**, 875-81.

Aguilar, R. C., Boehm, M., Gorshkova, I., Crouch, R. J., Tomita, K., Saito, T., Ohno, H., and Bonifacino, J. S. (2001). Signal-binding specificity of the mu4 subunit of the adaptor protein complex AP-4. *Journal of Biological Chemistry* **276**, 13145-52.

Aguirre-Garcia, M. M., Rosales-Encina, J. L., and Talamas-Rohana, P. (1997). Secreted *Entamoeba histolytica* acid phosphatase (SAP). *Archives of Medical Research* **28 (**Spec No), 184-5.

Aguirre-García, M. M., Anaya-Ruiz, M., and Talamás-Rohana, P. (2003). Membrane-bound acid phosphatase (MAP) from *Entamoeba histolytica* has phosphotyrosine phosphatase activity and disrupts the actin cytoskeleton of host cells. *Parasitology* **126**, 195-202.

Aldritt, S. M., Tien, P., and Wang, C. C. (1985). Pyrimidine salvage in *Giardia lamblia*. *Journal of Experimental Medicine* **161**, 437-45.

Ali, V., and Nozaki, T. (2006). Biochemical and functional characterization of phosphoserine aminotransferase from *Entamoeba histolytica*, which possesses both phosphorylated and non-phosphorylated serine metabolic pathways. *Molecular and Biochemical Parasitology* **145**, 71-83.

Ali, V., Shigeta, Y., and Nozaki, T. (2003). Molecular and structural characterization of NADPH-dependent D-glycerate dehydrogenase from the enteric parasitic protist *Entamoeba histolytica*. *Biochemical Journal* **375**, 729-736.

Ali, V., Hashimoto, T., Shigeta, Y., and Nozaki, T. (2004a). Molecular and biochemical characterization of D-phosphoglycerate dehydrogenase from *Entamoeba histolytica*. A unique enteric protozoan parasite that possesses both phosphorylated and nonphosphorylated serine metabolic pathways. *European Journal of Biochemistry* **271**, 2670-81.

Ali, V., Shigeta, Y., Tokumoto, U., Takahashi, Y., and Nozaki, T. (2004b). An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. *Journal of Biological Chemistry* **279**, 2670-81.

Ali, I. K.M., Zaki, M., and Clark, C. G. (2005). Use of PCR amplification of tRNA gene-linked short tandem repeats for genotyping *Entamoeba histolytica*. *Journal of Clinical Microbiology* **43**, 5842-7.

Anaya-Ruiz, M., Rosales-Encina, J. L., and Talamas-Rohana, P. (1997). Membrane acid phosphatase (MAP) from *Entamoeba histolytica*. *Archives of Medical Research* **28 (**Spec No), 182-3.

Anaya-Ruiz, M., Perez-Santos, J. L., and Talamas-Rohana, P. (2003). An ecto-protein tyrosine phosphatase of *Entamoeba histolytica* induces cellular detachment by disruption of actin filaments in HeLa cells. *International Journal for Parasitology* **33**, 663-670.

Anderson, I. J., and Loftus, B. J. (2005). *Entamoeba histolytica*: observations on metabolism based on the genome sequence. *Experimental Parasitology* **110**, 173-177.

Anderson, D. H., Sawaya, M. R., Cascio, D., Ernst, W., Modlin, R., Krensky, A., and Eisenberg, D. (2003). Granulysin crystal structure and a structure-derived lytic mechanism. *Journal of Molecular Biology* **325**, 355-65.

Andersson, J. O. (2005). Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences* **62**, 1182-97.

Andersson, J. O., Doolittle, W. F., and Nesbo, C. L. (2001). Genomics. Are there bugs in our genome? *Science* **292**, 1848-50.

Andersson, J. O., Sjogren, A. M., Davis, L. A., Embley, T. M., and Roger, A. J. (2003). Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Current Biology* **13**, 94-104.

Andersson, J. O., Hirt, R. P., Foster, P. G., and Roger, A. J. (2006). Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *BMC Evolutionary Biology* **6**, 27.

Andrä, J., and Leippe, M. (1994). Pore-forming peptide of *Entamoeba histolytica*. Significance of positively charged amino acid residues for its mode of action. *FEBS Lett* **354**, 97-102.

Andrä, J., Herbst, R., and Leippe, M. (2003). Amoebapores, archaic effector peptides of protozoan origin, are discharged into phagosomes and kill bacteria by permeabilizing their membranes. *Developmental and Comparative Immunology* **27**, 291-304.

Ankri, S., Stolarsky, T., Bracha, R., Padilla-Vaca, F., and Mirelman, D. (1999). Antisense inhibition of expression of cysteine proteinases affects *Entamoeba histolytica*-induced formation of liver abscess in hamsters. *Infection and Immunity* **67**, 421-422.

Archibald, J. M., Teh, E. M., and Keeling, P. J. (2003). Novel ubiquitin fusion proteins: ribosomal protein P1 and actin. *Journal of Molecular Biology* **328**, 771-8.

Arhets, P., Olivo, J. C., Gounon, P., Sansonetti, P., and Guillen, N. (1998). Virulence and functions of myosin II are inhibited by overexpression of light meromyosin in *Entamoeba histolytica*. *Molecular Biology of the Cell* **6**, 1537-1547.

Ariyanayagam, M. R., and Fairlamb, A. H. (1999). *Entamoeba histolytica* lacks trypanothione metabolism. *Molecular and Biochemical Parasitology* **103**, 61-9.

Arteaga-Nieto, P., López-Romero, E., Teran-Figueroa, Y., Cano-Canchola, C., Luna Arias, J. P., Flores-Carreón, A., and Calvo-Méndez, C. (2002). *Entamoeba histolytica*: purification and characterization of ornithine decarboxylase. *Experimental Parasitology* **101**, 215-222.

Asbury, C. L. (2005). Kinesin: world's tiniest biped. *Current Opinion in Cell Biology* **17**, 89-97.

Avila, E. E., Martinez-Alcaraz, E. R., Barbosa-Sabanero, G., Rivera-Baron, E. I., Arias-Negrete, S., and Zazueta-Sandova, l. R. (2002). Subcellular localization of the NAD^+-dependent alcohol dehydrogenase in *Entamoeba histolytica* trophozoites. *Journal of Parasitology* **88**, 217-222.

Bailey, G. B., Gilmour, J. R., and McCoomer, N. E. (1990). Roles of target cell membrane carbohydrate and lipid in *Entamoeba histolytica* interaction with mammalian cells. *Infection and Immunity* **58**, 2389-91.

Bakker-Grunwald, T., Martin, J. B., and Klein, G. (1995). Characterization of glycogen and amino acid pool of *Entamoeba histolytica* by C^{13}-NMR spectroscopy. *Journal of Eukaryotic Microbiology* **42**, 346-349.

Bakre, A. A., Rawal, K., Ramaswamy, R., Bhattacharya, A., and Bhattacharya, S. (2005). The LINEs and SINEs of *Entamoeba histolytica*: comparative analysis and genomic distribution. *Experimental Parasitology* **110**, 207-13.

Baldauf, S. L. (2003). The deep roots of eukaryotes. *Science* **300**, 1703-6.

Band, R. N., and Cirrito, H. (1979). Growth response of axenic *Entamoeba histolytica* to hydrogen, carbon dioxide, and oxygen. *Journal of Protozoology* **26**, 282-6.

Barlowe, C., and Schekman, R. (1993). SEC12 encodes a guanine-nucleotide-exchange factor essential for transport vesicle budding from the ER. *Nature* **365**, 347-9.

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., Ravazzola, M., Amherdt, M., and Schekman, R. (1994). COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. *Cell* **77**, 895-907.

Barrett , A. J. (1998). Cysteine peptidase, pp. 543-798. In A. J. Barrett, N. D. Rawlings, and J. F. Woessner (Eds). *Handbook of proteolytic enzymes*, Academic Press, Inc.,, San Diego, Calif.

Baum, K. F., Berens, R. L., Marr, J. J., Harrington, J. A., and Spector, T. (1989). Purine deoxynucleoside salvage in *Giardia lamblia*. *Journal of Biological Chemistry* **264**, 21087-90.

Beanan, M. J., and Bailey, G. B. (1995). The primary structure of an *Entamoeba histolytica* enolase. *Molecular and Biochemical Parasitology* **69**, 119-21.

Beck, D. L., Boettner, D. R., Dragulev, B., Ready, K., Nozaki, T., and Petri, W. A. J. (2005). Identification and gene expression analysis of a large family of transmembrane kinases related to the Gal/GalNAc lectin in *Entamoeba histolytica*. *Eukaryotic Cell* **4**, 722-732.

Beckers, C. J., Block, M. R., Glick, B. S., Rothman, J. E., and Balch, W. E. (1989). Vesicular transport between the endoplasmic reticulum and the Golgi stack requires the NEM-sensitive fusion protein. *Nature* **339**, 397-8.

Beiko, R. G., Harlow, T. J., and Ragan, M. A. (2005). Highways of gene sharing in prokaryotes. *Proceedings of the National Academy of Sciences USA* **102**, 14332-7.

Bell, S. P., and Dutta, A. (2002). DNA replication in eukaryotic cells. *Annual Review of Biochemistry* **71**, 333-74.

Berninghausen, O., and Leippe, M. (1997). Calcium-independent cytolysis of target cells induced by *Entamoeba histolytica*. *Archives of Medical Research* **28** (Spec No), 158-60.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., Bohme, U., Hannick, L., Aslett, M. A., Shallom, J., Marcello, L., Hou, L., Wickstead, B., Alsmark, U. C., Arrowsmith, C., Atkin, R. J., Barron, A. J., Bringaud, F., Brooks, K., Carrington, M., Cherevach, I., Chillingworth, T. J., Churcher, C., Clark, L. N., Corton, C. H., Cronin, A., Davies, R. M., Doggett, J., Djikeng, A., Feldblyum, T., Field, M. C., Fraser, A., Goodhead, I., Hance, Z., Harper, D., Harris, B. R., Hauser, H., Hostetler, J., Ivens, A., Jagels, K., Johnson, D., Johnson, J., Jones, K., Kerhornou, A. X., Koo, H., Larke, N., Landfear, S., Larkin, C., Leech, V., Line, A., Lord, A., Macleod, A., Mooney, P. J., Moule, S., Martin, D. M., Morgan, G. W., Mungall, K., Norbertczak, H., Ormond, D., Pai, G., Peacock, C. S., Peterson, J., Quail, M. A., Rabbinowitsch, E., Rajandream, M. A., Reitter, C., Salzberg, S. L., Sanders, M., Schobel, S., Sharp, S., Simmonds, M., Simpson, A. J., Tallon, L.,

Turner, C. M., Tait, A., Tivey, A. R., Van Aken, S., Walker, D., Wanless, D., Wang, S., White, B., White, O., Whitehead, S., Woodward, J., Wortman, J., Adams, M. D., Embley, T. M., Gull, K., Ullu, E., Barry, J. D., Fairlamb, A. H., Opperdoes, F., Barrell, B. G., Donelson, J. E., Hall, N., Fraser, C. M., *et al.* (2005). The genome of the African trypanosome *Trypanosoma brucei. Science* **309**, 416-22.

Bhattacharya, A., Prasad, R., and Sacks, D. L. (1992). Identification and partial characterization of a lipophosphoglycan from a pathogenic strain of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **56**, 161-168.

Bhattacharya, S., Som, I., and Bhattacharya, A. (1998). The ribosomal DNA plasmids of *Entamoeba*. *Parasitology Today* **14**, 181-185.

Bhattacharya, A., Padhan, N., Jain, R., and Bhattacharya, S. (2006). Calcium-binding proteins of *Entamoeba histolytica*. *Archives of Medical Research* **37**, 221-5.

Blume-Jensen, P., and Hunter, T. (2001). Oncogenic kinase signalling. *Nature* **411**, 355-65.

Bock, J. B., Matern, H. T., Peden, A. A., and Scheller, R. H. (2001). A genomic perspective on membrane compartment organization. *Nature* **409**, 839-41.

Boehlein, S. K., Schuster, S. M., and Richards, N. G. (1996). Glutamic acid gamma-monohydroxamate and hydroxylamine are alternate substrates for *Escherichia coli* asparagine synthetase B. *Biochemistry* **35**, 3031-7.

Boehm, M., and Bonifacino, J. S. (2001). Adaptins: the final recount. *Molecular Biology of the Cell* **12**, 2907-20.

Boehm, M., Aguilar, R. C., and Bonifacino, J. S. (2001). Functional and physical interactions of the adaptor protein complex AP-4 with ADP-ribosylation factors (ARFs). *EMBO Journal* **20**, 6265-76.

Bogdanova, N., and Hell, R. (1997). Cysteine synthesis in plants: protein-protein interactions of serine acetyltransferase from *Arabidopsis thaliana*. *Plant Journal* **11**, 251-62.

Bonifacino, J. S., and Glick, B. S. (2004). The mechanisms of vesicle budding and fusion. *Cell* **116**, 153-66.

Bordo, D., and Bork, P. (2002). The rhodanese/Cdc25 phosphatase superfamily. Sequence-structure-function relations. *EMBO Reports* **3**, 741-6.

Bracha, R., Nuchamowitz, Y., Leippe, M., and Mirelman, D. (1999). Antisense inhibition of amoebapore expression in *Entamoeba histolytica* causes a decrease in amoebic virulence. *Molecular Microbiology* **34**, 463-72.

Bracha, R., Nuchamowitz, Y., and Mirelman, D. (2003). Transcriptional silencing of an amoebapore gene in *Entamoeba histolytica*: molecular analysis and effect on pathogenicity. *Eukaryotic Cell* **2**, 295-305.

Braga, L. L., Ninomiya, H., McCoy, J. J., Eacker, S., Wiedmer, T., Pham, C., Wood, S., Sims, P. J., and Petri, W. A. (1992). Inhibition of the complement membrane attack complex by the galactose-specific adhesin of *Entamoeba histolytica*. *Journal of Clinical Investigation* **90**, 1131-1137.

Bredeston, L. M., Caffaro, C. E., Samuelson, J., and Hirschberg, C. B. (2005). Golgi and endoplasmic reticulum functions take place in different subcellular compartments of *Entamoeba histolytica. Journal of Biological Chemistry* **280**, 32168-76.

Bruchhaus, I., and Tannich, E. (1993). Primary structure of the pyruvate phosphate dikinase in *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **62**, 153-6.

Bruchhaus, I., and Tannich, E. (1994). Purification and molecular characterization of the NAD^+-dependent acetaldehyde/alcohol dehydrogenase from *Entamoeba histolytica*. *Biochemical Journal* **303**, 743-8.

Bruchhaus, I., Leippe, M., Lioutas, C., and Tannich, E. (1993). Unusual gene organization in the protozoan parasite *Entamoeba histolytica. DNA and Cell Biology* **12**, 925-33.

Bruchhaus, I., Jacobs, T., Denart, M., and Tannich, E. (1996). Pyrophosphate-dependent phosphofructokinase of *Entamoeba histolytica:* molecular

cloning, recombinant expression and inhibition by pyrophosphate analogues. *Biochemical Journal* **316**, 57-63.

Bruchhaus, I., Richter, S., and Tannich, E. (1997). Removal of hydrogen peroxide by the 29 kDa protein of *Entamoeba histolytica*. *Biochemical Journal* **326**, 785-9.

Bruchhaus, I., Richter, S., and Tannich, E. (1998). Recombinant expression and biochemical characterization of an NADPH:flavin oxidoreductase from *Entamoeba histolytica*. *Biochemical Journal* **330**, 1217-21.

Bruchhaus, I., Loftus, B. J., Hall, N., and Tannich, E. (2003). The intestinal protozoan parasite *Entamoeba histolytica* contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. *Eukaryotic Cell* **2**, 501-9.

Bruhn, H., Riekens, B., Berninghausen, O., and Leippe, M. (2003). Amoebapores and NK-lysin, members of a class of structurally distinct antimicrobial and cytolytic peptides from protozoa and mammals: a comparative functional analysis. *Biochemical Journal* **375**, 737-44.

Bulawa, C. E. (1993). Genetics and molecular biology of chitin synthesis in fungi. *Annual Review of Microbiology* **47**, 505-34.

Burch, D. J., Li, E., Reed, S., Jackson, T. F. H. G., and Stanley, S. L., Jr. (1991). Isolation of a strain-specific *Entamoeba histolytica* cDNA clone. *Journal of Clinical Microbiology* **29**, 696-701.

Burchard, G. D., and Bilke, R. (1992). Adherence of pathogenic and non-pathogenic *Entamoeba histolytica* strains to neutrophils. *Parasitology Research* **78**, 146-153.

Burchard, G. D., Moslein, C., and Brattig, N. W. (1992a). Adherence between *Entamoeba histolytica* trophozoites and undifferentiated or DMSO-induced HL-60 cells. *Parasitology Research* **78**, 336-340.

Burchard, G. D., Prange, G., and Mirelman, D. (1992b). Interaction of various *Entamoeba histolytica* strains with human intestinal cell lines. *Archives of Medical Research* **23**, 193-195.

Burd, C. G., Strochlic, T. I., and Gangi Setty, S. R. (2004). Arf-like GTPases: not so Arf-like after all. *Trends in Cell Biology* **14**, 687-94.

Burri, L., and Lithgow, T. (2004). A complete set of SNAREs in yeast. *Traffic* **5**, 45-52.

Capaldi, S. A., and Berger, J. M. (2004). Biochemical characterization of Cdc6/Orc1 binding to the replication origin of the euryarchaeon *Methanothermobacter thermoautotrophicus*. *Nucleic Acids Research* **32**, 4821-32.

Carrero, J. C., Petrossian, P., Olivos, A., Sanchez-Zerpa, M., Ostoa-Soloma, P., and Laclette, J. P. (2000). Effect of cyclosporine A on *Entamoeba histolytica*. *Archives of Medical Research* **31**, S8-9.

Carrero, J. C., Lugo, H., Perez, D. G., Ortiz-Martinez, C., and Laclette, J. P. (2004). Cyclosporin A inhibits calcineurin (phosphatase 2B) and P-glycoprotein activity in *Entamoeba histolytica*. *International Journal for Parasitology* **34**, 1091-7.

Carrozza, M. J., Utley, R. T., Workman, J. L., and Cote, J. (2003). The diverse functions of histone acetyltransferase complexes. *Trends in Genetics* **19**, 321-9.

Cawley, S. E., Wirth, A. I., and Speed, T. P. (2001). Phat: a gene finding program for *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **118**, 167-74.

Cedar, H., and Schwartz, J. H. (1967). Localization of the two-L-asparaginases in anaerobically grown *Escherichia coli*. *Journal of Biological Chemistry* **242**, 3753-5.

Chakrabarty, P., Sethi, D. K., Padhan, N., Kaur, K. J., Salunke, D. M., Bhattacharya, S., and Bhattacharya, A. (2004). Identification and characterization of EhCaBP2. A second member of the calcium-binding protein family of the protozoan parasite *Entamoeba histolytica*. *Journal of Biological Chemistry* **279**, 12898-12908.

Chan, K. W., Slotboom, D. J., Cox, S., Embley, T. M., Fabre, O., van der Giezen, M., Harding, M., Horner, D. S., Kunji, E. R., Leon-Avila, G.,

and Tovar, J. (2005). A novel ADP/ATP transporter in the mitosome of the microaerophilic human parasite *Entamoeba histolytica*. *Current Biology* **15**, 737-742.

Chaudhuri, S., Choudhury, N., and Raha, S. (1999). Growth stimulation by serum in *Entamoeba histolytica* is associated with protein tyrosine dephosphorylation. *FEMS Microbiology Letters* **178**, 241-249.

Chavez-Munguia, B., Espinosa-Cantellano, M., Castanon, G., and Martinez-Palomo, A. (2000). Ultrastructural evidence of smooth endoplasmic reticulum and golgi-like elements in *Entamoeba histolytica* and *Entamoeba dispar*. *Archives of Medical Research* **31**, S165-7.

Chen, Y. A., Scales, S. J., Patel, S. M., Doung, Y. C., and Scheller, R. H. (1999). SNARE complex formation is triggered by Ca2+ and drives membrane fusion. *Cell* **97**, 165-74.

Chen, Y. A., and Scheller, R. H. (2001). SNARE-mediated membrane fusion. *Nature Reviews. Molecular and Cellular Biobgy* **2**, 98-106.

Cheng, X. J., and Tachibana, H. (2001). Protection of hamsters from amebic liver abscess formation by immunization with the 150- and 170-kDa surface antigens of *Entamoeba histolytica*. *Parasitology Research* **87**, 126-30.

Cheng, X. J., Tsukamoto, H., Kaneda, Y., and Tachibana, H. (1998). Identification of the 150-kDa surface antigen of *Entamoeba histolytica* as a galactose- and N-acetyl-D-galactosamine-inhibitable lectin. *Parasitology Research* **84**, 632-9.

Cheng, X. J., Hughes, M. A., Huston, C. D., Loftus, B., Gilchrist, C. A., Lockhart, L. A., Ghosh, S., Miller-Sims, V., Mann, B. J., Petri, W. A., Jr., and Tachibana, H. (2001). Intermediate subunit of the Gal/GalNAc lectin of *Entamoeba histolytica* is a member of a gene family containing multiple CXXC sequence motifs. *Infection and Immunity* **69**, 5892-8.

Chi, A., and Kemp, R. G. (2000). The primordial high energy compound: ATP or inorganic pyrophosphate? *Journal of Biological Chemistry* **275**, 35677-9.

Chi, A. S., Deng, Z., Albach, R. A., and Kemp, R. G. (2001). The two phosphofructokinase gene products of *Entamoeba histolytica. Journal of Biological Chemistry* **276**, 19974-81.

Chinkers, M. (2001). Protein phosphatase 5 in signal transduction. *Trends in Endocrinology and Metabolism* **12**, 28-32.

Clark, C. G., and Roger, A. J. (1995). Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **92**, 6518-6521.

Clark, C. G., Espinosa Cantellano, M., and Bhattacharya, A. (2000). *Entamoeba histolytica*: an overview of the biology of the organism, pp. 1-45. In J. I. Ravdin (Ed.). *Amebiasis.*, Imperial College Press, London.

Clark, C. G., Ali, I. K., Zaki, M., Loftus, B. J., and Hall, N. (2006a). Unique organisation of tRNA genes in *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **146**, 24-9.

Clark, C. G., Kaffashian, F., Tawari, B., Windsor, J. J., Twigg-Flesner, A., Davies-Morel, M. C. G., Blessmann, J., Ebert, F., Peschel, B., Le Van, A., Jackson, C. J., Macfarlane, L., and Tannich, E. (2006b). New insights into the phylogeny of *Entamoeba* species provided by analysis of four new small-subunit rRNA genes. *International Journal of Systematic and Evolutionary Microbiology* **56**, 2235-2239.

Clary, D. O., Griff, I. C., and Rothman, J. E. (1990). SNAPs, a family of NSF attachment proteins involved in intracellular membrane fusion in animals and yeast. *Cell* **61**, 709-21.

Conner, S. D., and Schmid, S. L. (2003). Differential requirements for AP-2 in clathrin-mediated endocytosis. *Journal of Cell Biology* **162**, 773-9.

Coombs, G. H., and Mottram, J. C. (2001). Trifluoromethionine, a prodrug designed against methionine gamma-lyase-containing pathogens, has efficacy in vitro and in vivo against *Trichomonas vaginalis*. *Antimicrobial Agents and Chemotherapy* **45**, 1743-5.

Coppi, A., and Eichinger, D. (1999). Regulation of *Entamoeba invadens* encystation and gene expression with galactose and N-

acetylglucosamine. *Molecular and Biochemical Parasitology* **102**, 67-77.

Coppi, A., Merali, S., and Eichinger, D. (2002). The enteric parasite *Entamoeba* uses an autocrine catecholamine system during differentiation into the infectious cyst stage. *Journal of Biological Chemistry* **277**, 8083-90.

Cruz-Reyes, J. A., and Ackers, J. P. (1992). A DNA probe specific to pathogenic *Entamoeba histolytica*. *Archives of Medical Research* **23**, 271-275.

Cruz-Reyes, J., ur-Rehman, T., Spice, W. M., and Ackers, J. P. (1995). A novel transcribed repeat element from *Entamoeba histolytica*. *Gene* **166**, 183-184.

Das, S., and Lohia, A. (2002). Delinking of S phase and cytokinesis in the protozoan parasite *Entamoeba histolytica. Cellular Microbiology* **4**, 55-60.

Das, A. K., Cohen, P. W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO Journal* **17**, 1192-9.

Das, S., Stevens, T., Castillo, C., Villasenor, A., Arredondo, H., and Reddy, K. (2002). Lipid metabolism in mucous-dwelling amitochondriate protozoa. *International Journal for Parasitology* **32**, 655-75.

Das, S., Van Dellen, K., Bulik, D., Magnelli, P., Cui, J., Head, J., Robbins, P. W., and Samuelson, J. (2006). The cyst wall of *Entamoeba invadens* contains chitosan (deacetylated chitin). *Molecular and Biochemical Parasitology* **148**, 86-92.

Davis, J. A., Takagi, Y., Kornberg, R. D., and Asturias, F. A. (2002). Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular Cell* **10**, 409-15.

Davis, P. H., Zhang, Z., Chen, M., Zhang, X., Chakraborty, S., and Stanley, S. L., Jr. (2006). Identification of a family of BspA like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats. *Molecular and Biochemical Parasitology* **145**, 111-6.

Davis, P. H., Schulze, J., and Stanley, S. L., Jr. (2007). Transcriptomic comparison of two *Entamoeba histolytica* strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. *Molecular and Biochemical Parasitology* **151**, 118-28.

de la Vega, H., Specht, C. A., Semino, C. E., Robbins, P. W., Eichinger, D., Caplivski, D., Ghosh, S., and Samuelson, J. (1997). Cloning and expression of chitinases of Entamoebae. *Molecular and Biochemical Parasitology* **85**, 139-147.

Dear, P. H., and Cook, P. R. (1993). Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Research* **21**, 13-20.

Debnath, A., Das, P., Sajid, M., and McKerrow, J. H. (2004). Identification of genomic responses to collagen binding by trophozoites of *Entamoeba histolytica*. *Journal of Infectious Diseases* **190**, 448-57.

Demarque, M., Represa, A., Becq, H., Khalilov, I., Ben-Ari, Y., and Aniksztejn, L. (2002). Paracrine intercellular communication by a Ca2+- and SNARE-independent release of GABA and glutamate prior to synapse formation. *Neuron* **36**, 1051-61.

Deng, Z., Huang, M., Singh, K., Albach, R. A., Latshaw, S. P., Chang, K. P., and Kemp, R. G. (1998). Cloning and expression of the gene for the active PPi-dependent phosphofructokinase of *Entamoeba histolytica*. *Biochemical Journal* **329**, 659-64.

Dhar, S. K., Choudhury, N. R., Bhattacharya, A., and Bhattacharya, S. (1995). A multitude of circular DNAs exist in the nucleus of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **70**, 203-206.

Diamond, L. S., and Cunnick, C. C. (1991). A serum-free, partly defined medium, PDM-805, for the axenic cultivation of *Entamoeba histolytica* Schaudinn, 1903 and other *Entamoeba*. *Journal of Protozoology* **38**, 211-216.

Dodson, J. M., Clark, C. G., Lockhart, L. A., Leo, B. M., Schroeder, J. W., and Mann, B. J. (1997). Comparison of adherence, cytotoxicity, and Gal/GalNAc lectin gene structure in *Entamoeba histolytica* and *Entamoeba dispar*. *Parasitology International* **46**, 225-235.

Donaldson, J. G., Cassel, D., Kahn, R. A., and Klausner, R. D. (1992). ADP-ribosylation factor, a small GTP-binding protein, is required for binding of the coatomer protein beta-COP to Golgi membranes. *Proceedings of the National Academy of Sciences USA* **89**, 6408-12.

Doolittle, W. F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trend in Genetics* **14**, 307-311.

Doolittle, W. F., Boucher, Y., Nesbo, C. L., Douady, C. J., Andersson, J. O., and Roger, A. J. (2003). How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **358**, 39-57; discussion 57-8.

Dresios, J., Panopoulos, P., Suzuki, K., and Synetos, D. (2003). A dispensable yeast ribosomal protein optimizes peptidyltransferase activity and affects translocation. *Journal of Biological Chemistry* **278**, 3314-22.

Droux, M., Ruffet, M. L., Douce, R., and Job, D. (1998). Interactions between serine acetyltransferase and O-acetylserine (thiol) lyase in higher plants: structural and kinetic properties of the free and bound enzymes. *European Journal of Biochemistry* **255**, 235-45.

Duden, R., Kajikawa, L., Wuestehube, L., and Schekman, R. (1998). epsilon-COP is a structural component of coatomer that functions to stabilize alpha-COP. *EMBO Journal* **17**, 985-95.

Dulubova, I., Sugita, S., Hill, S., Hosaka, M., Fernandez, I., Sudhof, T. C., and Rizo, J. (1999). A conformational switch in syntaxin during exocytosis: role of munc18. *EMBO Journal* **18**, 4372-82.

Dutcher, S. K. (2001). The tubulin fraternity: alpha to eta. *Current Opinion in Cell Biology* **13**, 49-54.

Ehrenkaufer, G.M., Haque, R., Hackney, J.A., Eichinger, D.J., Singh, U., 2007. Identification of developmentally regulated genes in *Entamoeba histolytica*: insights into mechanisms of stage conversion in a protozoan parasite. *Cellular Microbiology* (in press).

Eichinger, D. (1997). Encystation of *Entamoeba* parasites. *Bioessays* **19**, 633-639.

Eichinger, D. (2001). Encystation in parasitic protozoa. *Current Opinion in Microbiology* **4**, 421-426.

Eichinger, L., Pachebat, J. A., Glockner, G., Rajandream, M. A., Sucgang, R., Berriman, M., Song, J., Olsen, R., Szafranski, K., Xu, Q., Tunggal, B., Kummerfeld, S., Madera, M., Konfortov, B. A., Rivero, F., Bankier, A. T., Lehmann, R., Hamlin, N., Davies, R., Gaudet, P., Fey, P., Pilcher, K., Chen, G., Saunders, D., Sodergren, E., Davis, P., Kerhornou, A., Nie, X., Hall, N., Anjard, C., Hemphill, L., Bason, N., Farbrother, P., Desany, B., Just, E., Morio, T., Rost, R., Churcher, C., Cooper, J., Haydock, S., van Driessche, N., Cronin, A., Goodhead, I., Muzny, D., Mourier, T., Pain, A., Lu, M., Harper, D., Lindsay, R., Hauser, H., James, K., Quiles, M., Madan Babu, M., Saito, T., Buchrieser, C., Wardroper, A., Felder, M., Thangavelu, M., Johnson, D., Knights, A., Loulseged, H., Mungall, K., Oliver, K., Price, C., Quail, M. A., Urushihara, H., Hernandez, J., Rabbinowitsch, E., Steffen, D., Sanders, M., Ma, J., Kohara, Y., Sharp, S., Simmonds, M., Spiegler, S., Tivey, A., Sugano, S., White, B., Walker, D., Woodward, J., Winckler, T., Tanaka, Y., Shaulsky, G., Schleicher, M., Weinstock, G., Rosenthal, A., Cox, E. C., Chisholm, R. L., Gibbs, R., Loomis, W. F., Platzer, M., Kay, R. R., Williams, J., Dear, P. H., Noegel, A. A., Barrell, B., and Kuspa, A. (2005). The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**, 43-57.

Elnekave, K., Siman-Tov, R., and Ankri, S. (2003). Consumption of L-arginine mediated by *Entamoeba histolytica* L-arginase (EhArg) inhibits

amoebicidal activity and nitric oxide production by activated macrophages. *Parasite Immunology* **25**, 597-608.

Espenshade, P., Gimeno, R. E., Holzmacher, E., Teung, P., and Kaiser, C. A. (1995). Yeast SEC16 gene encodes a multidomain vesicle coat protein that interacts with Sec23p. *Journal of Cell Biology* **131**, 311-24.

Espinosa-Cantellano, M., and Martínez-Palomo, A. (2000). Pathogenesis of intestinal amebiasis: from molecules to disease. *Clinical Microbiology Reviews* **13**, 318-331.

Fahey, R. C., Newton, G. L., Arrick, B., Overdank-Bogart, T., and Aley, S. B. (1984). *Entamoeba histolytica*: a eukaryote without glutathione metabolism. *Science* **224**, 70-2.

Fairlamb, A. H., and Cerami, A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. *Annual Review of Microbiology* **46**, 695-729.

Farri, T. A., Sargeaunt, P. G., Warhurst, D. C., Williams, J. E., and Bhojnani, R. (1980). Electrophoretic studies of the hexokinase of *Entamoeba histolytica* groups I to IV. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **74**, 672-3.

Faundez, V., Horng, J. T., and Kelly, R. B. (1998). A function for the AP3 coat complex in synaptic vesicle formation from endosomes. *Cell* **93**, 423-32.

Field, J., Rosenthal, B., and Samuelson, J. (2000). Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Molecular Microbiology* **38**, 446-455.

Folsch, H., Ohno, H., Bonifacino, J. S., and Mellman, I. (1999). A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells. *Cell* **99**, 189-98.

Frederick, J., and Eichinger, D. (2004). *Entamoeba invadens* contains the components of a classical adrenergic signaling system. *Molecular and Biochemical Parasitology* **137**, 339-43.

Freedman, R. B., Klappa, P., and Ruddock, L. W. (2002). Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Reports* **3**, 136-40.

Fridovich, I. (1995). Superoxide radical and superoxide dismutases. *Annual Review of Biochemistry* **64**, 97-112.

Frisardi, M., Ghosh, S. K., Field, J., Van Dellen, K., Rogers, R., Robbins, P., and Samuelson, J. (2000). The most abundant glycoprotein of amebic cyst walls (Jacob) is a lectin with five Cys-rich, chitin-binding domains. *Infection and Immunity* **68**, 4217-24.

Gadasi, H., and Kessler, E. (1983). Correlation of virulence and collagenolytic activity in *Entamoeba histolytica*. *Infection and Immunity* **39**, 528-31.

Gangopadhyay, S. S., Ray, S. S., Kennady, K., Pande, G., and Lohia, A. (1997). Heterogeneity of DNA content and expression of cell cycle genes in axenically growing *Entamoeba histolytica* HM1:IMSS clone A. *Molecular and Biochemical Parasitology* **90**, 9-20.

Gao, T., Furnari, F., and Newton, A. C. (2005). PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth. *Molecular Cell* **18**, 13-24.

Garcia-Rivera, G., Rodriguez, M. A., Ocadiz, R., Martinez, L. p. M. C., Arroyo, R., Gonzalez-Robles, A., and Orozco, E. (1999). *Entamoeba histolytica* : a novel cysteine protease and an adhesin form the 112 kDa surface protein. *Molecular Microbiology* **33**, 556-68.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M. S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002). Genome sequence

of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498-511.

Gelderman, A. H., Bartgis, I. L., Keister, D. B., and Diamond, L. S. (1971a). A comparison of genome sizes and thermal denaturation-derived base composition of DNAs from several members of *Entamoeba* (*histolytica* group). *Journal of Parasitology* **57**, 912-916.

Gelderman, A. H., Keister, D. B., Bartgis, I. L., and Diamond, L. S. (1971b). Characterization of the deoxyribonucleic acid of representative strains of *Entamoeba histolytica*, *E. histoytica*-like amebae, and *E. moshkovskii*. *Journal of Parasitology* **57**, 906-11.

Gemmill, T. R., and Trimble, R. B. (1999). Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochimica et Biophysica Acta* **1426**, 227-37.

Gerwig, G. J., van Kuik, J. A., Leeflang, B. R., Kamerling, J. P., Vliegenthart, J. F., Karr, C. D., and Jarroll, E. L. (2002). The *Giardia intestinalis* filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. *Glycobiology* **12**, 499-505.

Ghosh, S., Frisardi, M., Ramirez-Avila, L., Descoteaux, S., Sturm-Ramirez, K., Newton-Sanchez, O. A., Santos-Preciado, J. I., Ganguly, C., Lohia, A., Reed, S., and Samuelson, J. (2000). Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (Demographic sweep) and transcontinental spread of diploid parasites. *Journal of Clinical Microbiology* **38**, 3815-3821.

Ghosh, S., Chan, J. M., Lea, C. R., Meints, G. A., Lewis, J. C., Tovian, Z. S., Flessner, R. M., Loftus, T. C., Bruchhaus, I., Kendrick, H., Croft, S. L., Kemp, R. G., Kobayashi, S., Nozaki, T., and Oldfield, E. (2004). Effects of bisphosphonates on the growth of *Entamoeba histolytica* and *Plasmodium* species in vitro and in vivo. *Journal of Medicinal Chemistry* **47**, 175-87.

Gilchrist, C. A., Holm, C. F., Hughes, M. A., Schaenman, J. M., Mann, B. J., and Petri, W. A., Jr. (2001). Identification and characterization of an *Entamoeba histolytica* upstream regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *Journal of Biological Chemistry* **276**, 11838-43.

Gilchrist, C. A., Houpt, E., Trapaidze, N., Fei, Z., Crasta, O., Asgharpour, A., Evans, C., Martino-Catt, S., Baba, D. J., Stroup, S., Hamano, S., Ehrenkaufer, G., Okada, M., Singh, U., Nozaki, T., Mann, B. J., and Petri, W. A., Jr. (2006). Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Molecular and Biochemical Parasitology* **147**, 163-76.

Gillin, F. D., Reiner, D. S., and McCann, P. P. (1984). Inhibition of growth of *Giardia lamblia* by difluoromethylornithine, a specific inhibitor of polyamine biosynthesis. *Journal of Protozoology* **31**, 161-3.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science* **274**, 546, 563-7.

Gomes, C. M., Giuffre, A., Forte, E., Vicente, J. B., Saraiva, L. M., Brunori, M., and Teixeira, M. (2002). A novel type of nitric-oxide reductase. *Escherichia coli* flavorubredoxin. *Journal of Biological Chemistry* **277**, 25273-6.

Gregan, J., Lindner, K., Brimage, L., Franklin, R., Namdar, M., Hart, E. A., Aves, S. J., and Kearsey, S. E. (2003). Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding. *Molecular Biology of the Cell* **14**, 3876-87.

Grunler, J., Ericsson, J., and Dallner, G. (1994). Branch-point reactions in the biosynthesis of cholesterol, dolichol, ubiquinone and prenylated proteins. *Biochimica et Biophysica Acta* **1212**, 259-77.

Gu, J., and Gu, X. (2003). Natural history and functional divergence of protein tyrosine kinases. *Gene* **317**, 49-57.

Guerrant, R. L., Brush, J., Ravdin, J. I., Sullivan, J. A., and Mandell, G. L. (1981). Interaction between *Entamoeba histolytica* and human polymorphonuclear neutrophils. *Journal of Infectious Diseases* **143**, 83-93.

Gutsmann, T., Riekens, B., Bruhn, H., Wiese, A., Seydel, U., and Leippe, M. (2003). Interaction of amoebapores and NK-lysin with symmetric phospholipid and asymmetric lipopolysaccharide/phospholipid bilayers. *Biochemistry* **42**, 9804-12.

Haghighi, A., Kobayashi, S., Takeuchi, T., Thammapalerd, N., and Nozaki, T. (2003). Geographic diversity among genotypes of *Entamoeba histolytica* field isolates. *Journal of Clinical Microbiology* **41**, 3748-3756.

Hanks, S. K., and Hunter, T. (1995). Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB Journal* **9**, 576-96.

Hanson, P. I., Roth, R., Morisaki, H., Jahn, R., and Heuser, J. E. (1997). Structure and conformational changes in NSF and its membrane receptor complexes visualized by quick-freeze/deep-etch electron microscopy. *Cell* **90**, 523-35.

Hara-Kuge, S., Kuge, O., Orci, L., Amherdt, M., Ravazzola, M., Wieland, F. T., and Rothman, J. E. (1994). En bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. *Journal of Cell Biology* **124**, 883-92.

Hartwell, L. H., and Weinert, T. A. (1989). Checkpoints: controls that ensure the order of cell cycle events. *Science* **246**, 629-34.

Hausmann, S., Ho, C. K., Schwer, B., and Shuman, S. (2001). An essential function of *Saccharomyces cerevisiae* RNA triphosphatase Cet1 is to stabilize RNA guanylyltransferase Ceg1 against thermal inactivation. *Journal of Biological Chemistry* **276**, 36116-24.

Hecht, O., Van Nuland, N. A., Schleinkofer, K., Dingley, A. J., Bruhn, H., Leippe, M., and Grotzinger, J. (2004). Solution structure of the pore-

forming protein of *Entamoeba histolytica. Journal of Biological Chemistry* **279**, 17834-41.

Helenius, A., and Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry* **73**, 1019-49.

Helms, J. B., and Rothman, J. E. (1992). Inhibition by brefeldin A of a Golgi membrane enzyme that catalyses exchange of guanine nucleotide bound to ARF. *Nature* **360**, 352-4.

Hemming, S. A., and Edwards, A. M. (2000). Yeast RNA polymerase II subunit RPB9. Mapping of domains required for transcription elongation. *Journal of Biological Chemistry* **275**, 2288-94.

Hernandez, R., Luna-Arias, J. P., and Orozco, E. (1997). Comparison of the *Entamoeba histolytica* TATA-binding protein (TBP) structure with other TBP. *Archives of Medical Research* **28 (**Spec No), 43-5.

Hernandez-Ramirez, V. I., Anaya-Ruiz, M., Rios, A., and Talamas-Rohana, P. (2000). *Entamoeba histolytica*: tyrosine kinase activity induced by fibronectin through the beta1-integrin-like molecule. *Experimental Parasitology* **95**, 85-95.

Hidalgo, M. E., Sanchez, R., Perez, D. G., Rodriguez, M. A., Garcia, J., and Orozco, E. (1997). Molecular characterization of the *Entamoeba histolytica* enolase gene and modelling of the predicted protein. *FEMS Microbiology Letters* **148**, 123-9.

Higgs, H. N., and Peterson, K. J. (2005). Phylogenetic analysis of the formin homology 2 domain. *Molecular Biology of the Cell* **16**, 1-13.

Hirsch, C., Jarosch, E., Sommer, T., and Wolf, D. H. (2004). Endoplasmic reticulum-associated protein degradation: one model fits all? *Biochimica et Biophysica Acta* **1695**, 215-23.

Hirschberg, C. B., Robbins, P. W., and Abeijon, C. (1998). Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annual Review of Biochemistry* **67**, 49-69.

Hirst, J., Bright, N. A., Rous, B., and Robinson, M. S. (1999). Characterization of a fourth adaptor-related protein complex. *Molecular Biology of the Cell* **10**, 2787-802.

Hirt, R. P., Harriman, N., Kajava, A. V., and Embley, T. M. (2002). A novel potential surface protein in *Trichomonas vaginalis* contains a leucine-rich repeat shared by micro-organisms from all three domains of life. *Molecular and Biochemical Parasitology* **125**, 195-9.

Hofmann, B., Hecht, H. J., and Flohe, L. (2002). Peroxiredoxins. *Biological Chemistry* **383**, 347-64.

Holmgren, A. (2000). Antioxidant function of thioredoxin and glutaredoxin systems. *Antioxididants & Redox Signalling* **2**, 811-20.

Hooft van Huijsduijnen, R. (1998). Protein tyrosine phosphatases: counting the trees in the forest. *Gene* **225**, 1-8.

Horner, D. S., Hirt, R. P., and Embley, T. M. (1999). A single eubacterial origin of eukaryotic pyruvate: ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. *Molecular Biology and Evolution* **16**, 1280-91.

Hsiung, Y. G., Chang, H. C., Pellequer, J. L., La Valle, R., Lanker, S., and Wittenberg, C. (2001). F-box protein Grr1 interacts with phosphorylated targets via the cationic surface of its leucine-rich repeat. *Molecular and Cellular Biology* **21**, 2506-20.

Hubbard, S. C., and Ivatt, R. J. (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annual Review of Biochemistry* **50**, 555-83.

Huber, L. A., Pimplikar, S., Parton, R. G., Virta, H., Zerial, M., and Simons, K. (1993). Rab8, a small GTPase involved in vesicular traffic between the TGN and the basolateral plasma membrane. *Journal of Cell Biology* **123**, 35-45.

Hunter, T. (1995). Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling. *Cell* **80**, 225-36.

Hunter, W. N., Bond, C. S., Gabrielsen, M., and Kemp, L. E. (2003). Structure and reactivity in the non-mevalonate pathway of isoprenoid biosynthesis. *Biochemical Society Transactions* **31**, 537-42.

Ito, M. (2005). Conservation and diversification of three-repeat Myb transcription factors in plants. *Journal of Plant Research* **118**, 61-9.

Iwashita, J., Sato, Y., Kobayashi, S., Takeuchi, T., and Abe, T. (2005). Isolation and functional analysis of a chk2 homologue from *Entamoeba histolytica*. *Parasitology International* **54**, 21-7.

Jacobs, T., Bruchhaus, I., Dandekar, T., Tannich, E., and Leippe, M. (1998). Isolation and molecular characterization of a surface-bound proteinase of *Entamoeba histolytica*. *Molecular Microbiology* **27**, 269-76.

Jain, R., Rivera, M. C., and Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proceedings of the National Academy of Sciences USA* **96**, 3801-3806.

Jeffrey, P. D., Russo, A. A., Polyak, K., Gibbs, E., Hurwitz, J., Massague, J., and Pavletich, N. P. (1995). Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313-20.

Jongsareejit, B., Rahman, R. N., Fujiwara, S., and Imanaka, T. (1997). Gene cloning, sequencing and enzymatic properties of glutamate synthase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1. *Molecular & General Genetics* **254**, 635-42.

Jordan, I. K., Henze, K., Fedorova, N. D., Koonin, E. V., and Galperin, M. Y. (2003). Phylogenomic analysis of the *Giardia intestinalis* transcarboxylase reveals multiple instances of domain fusion and fission in the evolution of biotin-dependent enzymes. *Journal of Molecular Microbiology and Biotechnology* **5**, 172-89.

Juarez, P., Sanchez-Lopez, R., Stock, R. P., Olvera, A., Ramos, M. A., and Alagon, A. (2001). Characterization of the Ehrab8 gene, a marker of the late stages of the secretory pathway of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology.* **116**, 223 - 228.

Jurica, M. S., and Moore, M. J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Molecular Cell* **12**, 5-14.

Kafetzopoulos, D., Thireos, G., Vournakis, J. N., and Bouriotis, V. (1993). The primary structure of a fungal chitin deacetylase reveals the function for two bacterial gene products. *Proceedings of the National Academy of Sciences USA* **90**, 8005-8.

Kahn, R. A., and Gilman, A. G. (1984). Purification of a protein cofactor required for ADP-ribosylation of the stimulatory regulatory component of adenylate cyclase by cholera toxin. *Journal of Biological Chemistry* **259**, 6228-34.

Kahn, R. A., Cherfils, J., Elias, M., Lovering, R. C., Munro, S., and Schurmann, A. (2006). Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins. *Journal of Cell Biology* **172**, 645-50.

Kaiser, C. (2000). Thinking about p24 proteins and how transport vesicles select their cargo. *Proceedings of the National Academy of Sciences USA* **97**, 3783-5.

Kang, J. S., Kim, S. H., Hwang, M. S., Han, S. J., Lee, Y. C., and Kim, Y. J. (2001). The structural and functional organization of the yeast mediator complex. *Journal of Biological Chemistry* **276**, 42003-10.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., and Vivares, C. P. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450-3.

Keen, J. H. (1987). Clathrin assembly proteins: affinity purification and a model for coat assembly. *Journal of Cell Biology* **105**, 1989-98.

Keene, W. E., Hidalgo, M. E., Orozco, E., and McKerrow, J. H. (1990). *Entamoeba histolytica*: correlation of the cytopathic effect of virulent trophozoites with secretion of a cysteine proteinase. *Experimental Parasitology* **71**, 199-206.

Kelleher, D. J., and Gilmore, R. (2006). An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R-62R.

Kelman, L. M., and Kelman, Z. (2004). Multiple origins of replication in archaea. *Trends in Microbiology* **12**, 399-401.

Kennelly, P. J. (2001). Protein phosphatases : a phylogenetic perspective. *Chemical Reviews* **101**, 2291-312.

Kimura, A., Hara, Y., Kimoto, T., Okuno, Y., Minekawa, Y., and Nakabayashi, T. (1996). Cloning and expression of a putative alcohol dehydrogenase gene of *Entamoeba histolytica* and its application to immunological examination. *Clinical and Diagnostic Laboratory Immunology* **3**, 270-274.

Kirchhausen, T. (2000). Three ways to make a vesicle. *Nature Reviews. Molecular and Cellular Biology* **1**, 187-98.

Kirisako, T., Ichimura, Y., Okada, H., Kabeya, Y., Mizushima, N., Yoshimori, T., Ohsumi, M., Takao, T., Noda, T., and Ohsumi, Y. (2000). The reversible modification regulates the membrane-binding state of Apg8/Aut7 essential for autophagy and the cytoplasm to vacuole targeting pathway. *Journal of Cell Biology* **151**, 263-76.

Knodler, L. A., Edwards, M. R., and Schofield, P. J. (1994). The intracellular amino acid pools of *Giardia intestinalis, Trichomonas vaginalis*, and *Crithidia luciliae*. *Experimental Parasitology* **79**, 117-25.

Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. *Trends in Biochemical Sciences* **19**, 415-21.

Koonin, E. V., Makarova, K. S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology* **55**, 709-42.

Kornberg, R. D. (2001). The eukaryotic gene transcription machinery. *Biological Chemistry* **382**, 1103-7.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annual Review of Biochemistry* **54**, 631-64.

Kriebel, P. W., and Parent, C. A. (2004). Adenylyl cyclase expression and regulation during the differentiation of *Dictyostelium discoideum*. *IUBMB Life* **56**, 541-6.

Kroschewski, H., Ortner, S., Steipe, B., Scheiner, O., Wiedermann, G., and Duchêne, M. (2000). Differences in substrate specificity and kinetic properties of the recombinant hexokinases HXK1 and HXK2 from *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **105**, 71-80.

Kulda, J. (1999). Trichomonads, hydrogenosomes and drug resistance. *International Journal for Parasitology* **29**, 199-212.

Kumagai, M., Makioka, A., Takeuchi, T., and Nozaki, T. (2004). Molecular cloning and characterization of a protein farnesyltransferase from the enteric protozoan parasite *Entamoeba histolytica*. *Journal of Biological Chemistry* **279**, 2316-23.

Kumar, A., Shen, P. S., Descoteaux, S., Pohl, J., Bailey, G., and Samuelson, J. (1992). Cloning and expression of an $NADP^+$-dependent alcohol dehydrogenase gene of *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **89**, 10188-10192.

Kupke, T. (2002). Molecular characterization of the 4'-phosphopantothenoylcysteine synthetase domain of bacterial dfp flavoproteins. *Journal of Biological Chemistry* **277**, 36137-45.

Kupke, T. (2004). Active-site residues and amino acid specificity of the bacterial 4'-phosphopantothenoylcysteine synthetase CoaB. *European Journal of Biochemistry* **271**, 163-72.

Kupke, T., Uebele, M., Schmid, D., Jung, G., Blaesse, M., and Steinbacher, S. (2000). Molecular characterization of lantibiotic-synthesizing enzyme EpiD reveals a function for bacterial Dfp proteins in coenzyme A biosynthesis. *Journal of Biological Chemistry* **275**, 31838-46.

Kuranda, M. J., and Robbins, P. W. (1991). Chitinase is required for cell separation during growth of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **266**, 19758-67.

Lal, K., Field, M. C., Carlton, J. M., Warwicker, J., and Hirt, R. P. (2005). Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology* **143**, 226-35.

Landa, A., Rojo-Dominguez, A., Jimenez, L., and Fernandez-Velasco, D. A. (1997). Sequencing, expression and properties of triosephosphate isomerase from *Entamoeba histolytica*. *European Journal of Biochemistry* **247**, 348-55.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.

Lawrence, J. G. (2005a). Common themes in the genome strategies of pathogens. *Current Opinion in Genetics and Development.*

Lawrence, J. G. (2005b). Horizontal and vertical gene transfer: the life history of pathogens. *Contributions to Microbiology* **12**, 255-71.

Le Borgne, R., Planque, N., Martin, P., Dewitte, F., Saule, S., and Hoflack, B. (2001). The AP-3-dependent targeting of the melanosomal glycoprotein QNR-71 requires a di-leucine-based sorting signal. *Journal of Cell Science* **114**, 2831-41.

Lee, F. J., Huang, C. F., Yu, W. L., Buu, L. M., Lin, C. Y., Huang, M. C., Moss, J., and Vaughan, M. (1997a). Characterization of an ADP-ribosylation factor-like 1 protein in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **272**, 30998-1005.

Lee, J. H., Kim, J. M., Kim, M. S., Lee, Y. T., Marshak, D. R., and Bae, Y. S. (1997b). The highly basic ribosomal protein L41 interacts with the beta subunit of protein kinase CKII and stimulates phosphorylation of DNA topoisomerase IIalpha by CKII. *Biochemical and Biophysical Research Communications* **238**, 462-7.

Leippe, M. (1997). Amoebapores. *Parasitology Today* **13**, 178-83.

Leippe, M. (1999). Antimicrobial and cytolytic polypeptides of amoeboid protozoa: effector molecules of primitive phagocytes. *Developmental and Comparative Immunology* **23**, 267-79.

Leippe, M., Ebel, S., Schoenberger, O. L., Horstmann, R. D., and Müller -Eberhard, H. J. (1991). Pore-forming peptide of pathogenic *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **88**, 7659-63.

Leippe, M., Tannich, E., Nickel, R., van der Goot, G., Pattus, F., Horstmann, R. D., and Müller -Eberhard, H. J. (1992). Primary and secondary structure of the pore-forming peptide of pathogenic *Entamoeba histolytica*. *EMBO Journal* **11**, 3501-6.

Leippe, M., Andrä, J., and Müller -Eberhard, H. J. (1994a). Cytolytic and antibacterial activity of synthetic peptides derived from amoebapore, the pore-forming peptide of *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **91**, 2602-6.

Leippe, M., Andrä, J., Nickel, R., Tannich, E., and Müller -Eberhard, H. J. (1994b). Amoebapores, a family of membranolytic peptides from cytoplasmic granules of *Entamoeba histolytica*: isolation, primary structure, and pore formation in bacterial cytoplasmic membranes. *Molecular Microbiology* **14**, 895-904.

Leippe, M., Sievertsen, H. J., Tannich, E., and Horstmann, R. D. (1995). Spontaneous release of cysteine proteinases but not of pore-forming peptides by viable *Entamoeba histolytica. Parasitology* **111**, 569-74.

Leippe, M., Bruhn, H., Hecht, O., and Grotzinger, J. (2005). Ancient weapons: the three-dimensional structure of amoebapore A. *Trends in Parasitology* **21**, 5-7.

Leon-Avila, G., and Tovar, J. (2004). Mitosomes of *Entamoeba histolytica* are abundant mitochondrion-related remnant organelles that lack a detectable organellar genome. *Microbiology* **150**, 1245-1250.

Leung, C. L., Green, K. J., and Liem, R. K. (2002). Plakins: a family of versatile cytolinker proteins. *Trends in Cell Biology* **12**, 37-45.

Li, E., Yang, W. G., Zhang, T., and Stanley, S. L., Jr. (1995). Interaction of laminin with *Entamoeba histolytica* cysteine proteinases and its effect on amebic pathogenesis. *Infection and Immunity* **63**, 4150-3.

Li, L., and Dixon, J. E. (2000). Form, function, and regulation of protein tyrosine phosphatases and their involvement in human diseases. *Seminars in Immunology* **12**, 75-84.

Li, Y., Chen, Z. Y., Wang, W., Baker, C. C., and Krug, R. M. (2001). The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. *RNA* **7**, 920-31.

Liepinsh, E., Andersson, M., Ruysschaert, J. M., and Otting, G. (1997). Saposin fold revealed by the NMR structure of NK-lysin. *Nature Structural Biology* **4**, 793-5.

Linstead, D., and Cranshaw, M. A. (1983). The pathway of arginine catabolism in the parasitic flagellate *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology* **8**, 241-52.

Lioutas, C., and Tannich, E. (1995). Transcription of protein-coding genes in *Entamoeba histolytica* is insensitive to high concentrations of α–amanitin. *Molecular and Biochemical Parasitology.* **73**, 259-261.

Lioutas, C., Schmetz, C., and Tannich, E. (1995). Identification of various circular DNA molecules in *Entamoeba histolytica*. *Experimental Parasitology* **80**, 349-352.

Lo, H. S., and Reeves, R. E. (1978). Pyruvate-to-ethanol pathway in *Entamoeba histolytica*. *Biochemical Journal* **171**, 225-30.

Lo, H., and Reeves, R. E. (1980). Purification and properties of NADPH:flavin oxidoreductase from *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **2**, 23-30.

Lobelle-Rich, P. A., and Reeves, R. E. (1983). Separation and characterization of two UTP-utilizing hexose phosphate uridylyltransferases from *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **7**, 173-82.

Lockwood, B. C., and Coombs, G. H. (1991). Purification and characterization of methionine gamma-lyase from *Trichomonas vaginalis*. *Biochemical Journal* **279**, 675-82.

Loftus, B. J., and Hall, N. (2005). *Entamoeba*: still more to be learned from the genome. *Trends in Parasitology* **21**, 453.

Loftus, B., Anderson, I., Davies, R., Alsmark, U. C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R. P., Mann, B. J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M. A., Rabbinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillen, N., Gilchrist, C., Stroup, S. E., Bhattacharya, S., Lohia, A., Foster, P. G., Sicheritz-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N. M., Petri, W. A., Jr., Clark, C. G., Embley,

T. M., Barrell, B., Fraser, C. M., and Hall, N. (2005). The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865-8.

Lohia, A., and Samuelson, J. (1993). Cloning of the *Eh cdc2* gene from *Entamoeba histolytica* encoding a protein kinase p34^{cdc2} homologue. *Gene* **127**, 203-207.

Lohia, A., Hait, N. C., and Lahiri Majumder, A. (1999). L-*myo*-inositol 1-phosphate synthase from *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **98**, 67-79.

Long-Krug, S. A., Fischer, K. J., Hysmith, R. M., and Ravdin, J. I. (1985). Phospholipase A enzymes of *Entamoeba histolytica*: description and subcellular localization. *Journal of Infectious Diseases* **152**, 536-41.

Lu, L., Horstmann, H., Ng, C., and Hong, W. (2001). Regulation of Golgi structure and function by ARF-like protein 1 (Arl1). *Journal of Cell Science* **114**, 4543-55.

Lu, L., Tai, G., and Hong, W. (2004). Autoantigen Golgin-97, an effector of Arl1 GTPase, participates in traffic from the endosome to the trans-golgi network. *Molecular Biology of the Cell* **15**, 4426-43.

Luaces, A. L., and Barrett, A. J. (1988). Affinity purification and biochemical characterization of histolysin, the major cysteine proteinase of *Entamoeba histolytica*. *Biochemical Journal* **250**, 903-9.

Luaces, A. L., Pico, T., and Barrett, A. J. (1992). The ENZYMEBA test: detection of intestinal *Entamoeba histolytica* infection by immuno-enzymatic detection of histolysain. *Parasitology* **105**, 203-5.

Lushbaugh, W. B., Hofbauer, A. F., and Pittman, F. E. (1984). Proteinase activities of *Entamoeba histolytica* cytotoxin. *Gastroenterology* **87**, 17-27.

Lushbaugh, W. B., Hofbauer, A. F., and Pittman, F. E. (1985). *Entamoeba histolytica*: purification of cathepsin B. *Experimental Parasitology* **59**, 328-36.

MacFarlane, R. C., and Singh, U. (2006). Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species:

potential implications for amebic pathogenesis. *Infection and Immunity* **74**, 340-51.

Machida, Y. J., Hamlin, J. L., and Dutta, A. (2005). Right place, right time, and only once: replication initiation in metazoans. *Cell* **123**, 13-24.

MacKelvie, S. H., Andrews, P. D., and Stark, M. J. (1995). The *Saccharomyces cerevisiae* gene SDS22 encodes a potential regulator of the mitotic function of yeast type 1 protein phosphatase. *Molecular and Cellular Biology* **15**, 3777-85.

Mai, Z., Ghosh, S., Frisardi, M., Rosenthal, B., Rogers, R., and Samuelson, J. (1999). Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("Crypton") in the microaerophilic protozoan parasite *Entamoeba histolytica*. *Molecular and Cellular Biology* **19**, 2198-2205.

Maiorano, D., Lutzmann, M., and Mechali, M. (2006). MCM proteins and DNA replication. *Current Opinion in Cell Biology* **18**, 130-6.

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878-9.

Makioka, A., Kumagai, M., Takeuchi, T., and Nozaki, T. (2006). Characterization of protein geranylgeranyltransferase I from the enteric protist *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **145**, 216-25.

Mallet, L., Renault, G., and Jacquet, M. (2000). Functional cloning of the adenylate cyclase gene of *Candida albicans* in *Saccharomyces cerevisiae* within a genomic fragment containing five other genes, including homologues of CHS6 and SAP185. *Yeast* **16**, 959-66.

Mallik, R., and Gross, S. P. (2004). Molecular motors: strategies to get along. *Current Biology* **14**, R971-82.

Mandal, P. K., Bagchi, A., Bhattacharya, A., and Bhattacharya, S. (2004). An *Entamoeba histolytica* LINE/SINE pair inserts at common target sites cleaved by the restriction enzyme-like LINE-encoded endonuclease. *Eukaryotic Cell* **3**, 170-179.

Mann, B. J., Torian, B. E., Vedvick, T. S., and Petri, W. A., Jr. (1991). Sequence of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **88**, 3248-3252.

Marino, G., Uria, J. A., Puente, X. S., Quesada, V., Bordallo, J., and Lopez-Otin, C. (2003). Human autophagins, a family of cysteine proteinases potentially implicated in cell degradation by autophagy. *Journal of Biological Chemistry* **278**, 3671-8.

Marion, S., Wilhelm, C., Voigt, H., Bacri, J. C., and Guillen, N. (2004). Overexpression of myosin IB in living *Entamoeba histolytica* enhances cytoplasm viscosity and reduces phagocytosis. *Journal of Cell Science* **117**, 3271-9.

Marsh, J. J., and Lebherz, H. G. (1992). Fructose-bisphosphate aldolases: an evolutionary history. *Trends in Biochemical Sciences* **17**, 110-3.

Martin, W., and Müller , M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37-41.

Masai, H., and Arai, K. (2002). Cdc7 kinase complex: a key regulator in the initiation of DNA replication. *Journal of Cellular Physiology* **190**, 287-96.

Massague, J., Blain, S. W., and Lo, R. S. (2000). TGFbeta signaling in growth control, cancer, and heritable disorders. *Cell* **103**, 295-309.

Maurer-Stroh, S., Washietl, S., and Eisenhaber, F. (2003). Protein prenyltransferases: anchor size, pseudogenes and parasites. *Biological Chemistry* **384**, 977-89.

Mayer, A., Wickner, W., and Haas, A. (1996). Sec18p (NSF)-driven release of Sec17p (alpha-SNAP) can precede docking and fusion of yeast vacuoles. *Cell* **85**, 83-94.

Mazzuco, A., Benchimol, M., and De Souza, W. (1997). Endoplasmic reticulum and Golgi-like elements in *Entamoeba*. *Micron* **28**, 241-247.

McCarty, D. R., and Chory, J. (2000). Conservation and innovation in plant signaling pathways. *Cell* **103**, 201-9.

McCoy, J. J., Mann, B. J., Vedvick, T. S., Pak, Y., Heimark, D. B., and Petri, W. A., Jr. (1993). Structural analysis of the light subunit of the *Entamoeba histolytica* galactose-specific adherence lectin. *Journal of Biological Chemistry* **268**, 24223-24231.

McLaughlin, J., and Faubert, G. (1977). Partial purification and some properties of a neutral sulfhydryl and an acid proteinase from *Entamoeba histolytica*. *Canadian Journal of Microbiology* **23**, 420-5.

McTaggart, S. J. (2006). Isoprenylated proteins. *Cellular and Molecular Life Sciences* **63**, 255-67.

Mehlotra, R. K. (1996). Antioxidant defense mechanisms in parasitic protozoa. *Critical Reviews in Microbiology* **22**, 295-314.

Memon, A. R. (2004). The role of ADP-ribosylation factor and SAR1 in vesicular trafficking in plants. *Biochimica et Biophysica Acta* **1664**, 9-30.

Merchant, A. M., Kawasaki, Y., Chen, Y., Lei, M., and Tye, B. K. (1997). A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in *Saccharomyces cerevisiae. Molecular and Cellular Biology* **17**, 3261-71.

Meyer, C., Zizioli, D., Lausmann, S., Eskelinen, E. L., Hamann, J., Saftig, P., von Figura, K., and Schu, P. (2000). mu1A -adaptin-deficient mice: lethality, loss of AP-1 binding and rerouting of mannose 6-phosphate receptors. *EMBO Journal* **19**, 2193-203.

Meza, I., Sabanero, M., Cázares, F., and Bryan, J. (1983). Isolation and characterization of actin from *Entamoeba histolyica*. *Journal of Biological Chemistry* **258**, 3936-3941.

Mira, A., Pushker, R., Legault, B. A., Moreira, D., and Rodriguez-Valera, F. (2004). Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics. *BMC Evolutionary Biology* **4**, 50.

Mishra, C., Semino, C. E., McCreath, K. J., de la Vega, H., Jones, B. J., Specht, C. A., and Robbins, P. W. (1997). Cloning and expression of two chitin deacetylase genes of *Saccharomyces cerevisiae*. *Yeast* **13**, 327-36.

Moody-Haupt, S., Patterson, J. H., Mirelman, D., and McConville, M. J. (2000). The major surface antigens of *Entamoeba histolytica* trophozoites are GPI-anchored proteophosphoglycans. *Journal of Molecular Biology* **297**, 409-420.

Morgan, D. O. (1995). Principles of CDK regulation. *Nature* **374**, 131-4.

Morgan, D. O. (1996). The dynamics of cyclin dependent kinase structure. *Current Opinion in Cell Biology* **8**, 767-72.

Motley, A., Bright, N. A., Seaman, M. N., and Robinson, M. S. (2003). Clathrin-mediated endocytosis in AP-2-depleted cells. *Journal of Cell Biology* **162**, 909-18.

Müller , M. (1986). Reductive activation of nitroimidazoles in anaerobic microorganisms. *Biochemical Pharmacology* **35**, 37-41.

Mullikin, J. C., and Ning, Z. (2003). The phusion assembler. *Genome Research* **13**, 81-90.

Munford, R. S., Sheppard, P. O., and O'Hara, P. J. (1995). Saposin-like proteins (SAPLIP) carry out diverse functions on a common backbone structure. *Journal of Lipid Research* **36**, 1653-63.

Muñoz, M. de L., Moreno, M. A., Pérez-Garcia, J. N., Tovar, G. R., and Hernandez, V. I. (1991). Possible role of calmodulin in the secretion of *Entamoeba histolytica* electron-dense granules containing collagenase. *Molecular Microbiology* **5**, 1707-1714.

Munro, S. (2004). Organelle identity and the organization of membrane traffic. *Nature Cell Biology* **6**, 469-72.

Myers, E. W., Sutton, G. G., Delcher, A. L., Dew, I. M., Fasulo, D. P., Flanigan, M. J., Kravitz, S. A., Mobarry, C. M., Reinert, K. H., Remington, K. A., Anson, E. L., Bolanos, R. A., Chou, H. H., Jordan, C. M., Halpern, A. L., Lonardi, S., Beasley, E. M., Brandon, R. C., Chen, L., Dunn, P. J., Lai, Z., Liang, Y., Nusskern, D. R., Zhan, M., Zhang,

Q., Zheng, X., Rubin, G. M., Adams, M. D., and Venter, J. C. (2000). A whole-genome assembly of *Drosophila*. *Science* **287**, 2196-204.

Nakada-Tsukui, K., Saito-Nakano, Y., Ali, V., and Nozaki, T. (2005). 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*. *Molecular Biology of the Cell* **16**, 5294-303.

Nakamura, H. (2005). Thioredoxin and its related molecules: update 2005. *Antioxididants & Redox Signalling* **7**, 823-8.

Nakamura, M., Yamada, M., Hirota, Y., Sugimoto, K., Oka, A., and Takanami, M. (1981). Nucleotide sequence of the asnA gene coding for asparagine synthetase of *E. coli* K-12. *Nucleic Acids Research* **9**, 4669-76.

Nakatsu, F., and Ohno, H. (2003). Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. *Cell Structure and Function* **28**, 419-29.

Nebreda, A. R. (2006). CDK activation by non-cyclin proteins. *Current Opinion in Cell Biology* **18**, 192-8.

Nesbo, C. L., L'Haridon, S., Stetter, K. O., and Doolittle, W. F. (2001). Phylogenetic analyses of two "archaeal" genes in *Thermotoga maritima* reveal multiple transfers between archaea and bacteria. *Molecular Biology and Evolution* **18**, 362-75.

Nickel, R., Jacobs, T., Urban, B., Scholze, H., Bruhn, H., and Leippe, M. (2000). Two novel calcium-binding proteins from cytoplasmic granules of the protozoan parasite *Entamoeba histolytica*. *FEBS Letters* **486**, 112-6.

Nixon, J. E. J., Wang, A., Field, J., Morrison, H. G., McArthur, A. G., Sogin, M. L., Loftus, B. J., and Samuelson, J. (2002). Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. *Eukaryotic Cell* **1**, 181-190.

Novick, P., and Zerial, M. (1997). The diversity of Rab proteins in vesicle transport. *Current Opinion in Cell Biology* **9**, 496-504.

Nozaki, T., Arase, T., Shigeta, Y., Asai, T., Leustek, T., and Takeuchi, T. (1998a). Cloning and bacterial expression of adenosine-5'-triphosphate sulfurylase from the enteric protozoan parasite *Entamoeba histolytica*. *Biochimica et Biophysica Acta* **1429**, 284-91.

Nozaki, T., Asai, T., Kobayashi, S., Ikegami, F., Noji, M., Saito, K., and Takeuchi, T. (1998b). Molecular cloning and characterization of the genes encoding two isoforms of cysteine synthase in the enteric protozoan parasite *Entamoeba histolytica. Molecular and Biochemical Parasitology* **97**, 33-44.

Nozaki, T., Asai, T., Sanchez, L. B., Kobayashi, S., Nakazawa, M., and Takeuchi, T. (1999). Characterization of the gene encoding serine acetyltransferase, a regulated enzyme of cysteine biosynthesis from the protist parasites *Entamoeba histolytica* and *Entamoeba dispar*. Regulation and possible function of the cysteine biosynthetic pathway in Entamoeba. *Journal of Biological Chemistry* **274**, 32445-52.

Nozaki, T., Ali, V., and Tokoro, M. (2005). Sulfur-containing amino acid metabolism in parasitic protozoa. *Advances in Parasitology* **60**, 1-99.

Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304.

Okada, M., Huston, C. D., Mann, B. J., Petri, W. A., Jr., Kita, K., and Nozaki, T. (2005). Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica. Eukaryotic Cell* **4**, 827-31.

Ondarza, R. N., Tamayo, E., Hurtado, G., Hernandez, E., and Iturbe, A. (1997). Isolation and purification of glutathionyl-spermidine and trypanothione from *Entamoeba histolytica. Archives of Medical Research* **28**, 73-5.

Ondarza, R. N., Hurtado, G., Iturbe, A., Hernandez, E., Tamayo, E., and Woolery, M. (2005). Identification of trypanothione from the human pathogen *Entamoeba histolytica*, by mass spectrometry and chemical analysis. *Biotechnology and Applied Biochemistry* **42**, 175-181.

Ooi, C. E., Dell'Angelica, E. C., and Bonifacino, J. S. (1998). ADP-Ribosylation factor 1 (ARF1) regulates recruitment of the AP-3 adaptor complex to membranes. *Journal of Cell Biology* **142**, 391-402.

Opperdoes, F. R., and Michels, P. A. (2001). Enzymes of carbohydrate metabolism as potential drug targets. *International Journal for Parasitology* **31**, 482-90.

Ortner, S., Plaimauer, B., Binder, M., Scheiner, O., Wiedermann, G., and Duchêne, M. (1995). Molecular analysis of two hexokinase isoenzymes from *Entamoeba histolytica*. *Molecular and Biochemical Parasitology.* **73**, 189-198.

Ortner, S., Binder, M., Scheiner, O., Wiedermann, G., and Duchêne, M. (1997a). Molecular and biochemical characterization of phosphoglucomutases from *Entamoeba histolytica* and *Entamoeba dispar. Molecular and Biochemical Parasitology* **90**, 121-129.

Ortner, S., Clark, C. G., Binder, M., Scheiner, O., Wiedermann, G., and Duchêne, M. (1997b). Molecular biology of the hexokinase isoenzyme pattern that distinguishes pathogenic *Entamoeba histolytica* from nonpathogenic *Entamoeba dispar*. *Molecular and Biochemical Parasitology*. **86**, 85-94.

Ostoa-Saloma, P., Carrero, J. C., Petrossian, P., Herion, P., Landa, A., and Laclette, J. P. (2000). Cloning, characterization and functional expression of a cyclophilin of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology.* **107**, 219-225.

Otomo, T., Tomchick, D. R., Otomo, C., Panchal, S. C., Machius, M., and Rosen, M. K. (2005). Structural basis of actin filament nucleation and processive capping by a formin homology 2 domain. *Nature* **433**, 488-94.

Parsons, M. (2004). Glycosomes: parasites and the divergence of peroxisomal purpose. *Molecular Microbiology* **53**, 717-24.

Pasqualato, S., Renault, L., and Cherfils, J. (2002). Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. *EMBO Reports* **3**, 1035-41.

Patil, C., and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Current Opinion in Cell Biology* **13**, 349-55.

Peng, R., De Antoni, A., and Gallwitz, D. (2000). Evidence for overlapping and distinct functions in protein transport of coat protein Sec24p family members. *Journal of Biological Chemistry* **275**, 11521-8.

Pereira-Leal, J. B., and Seabra, M. C. (2001). Evolution of the Rab family of small GTP-binding proteins. *Journal of Molecular Biology* **313**, 889-901.

Petoukhov, M. V., Svergun, D. I., Konarev, P. V., Ravasio, S., van den Heuvel, R. H., Curti, B., and Vanoni, M. A. (2003). Quaternary structure of *Azospirillum brasilense* NADPH-dependent glutamate synthase in solution as revealed by synchrotron radiation x-ray scattering. *Journal of Biological Chemistry* **278**, 29933-9.

Petri, W. A. Jr (2002). Pathogenesis of amebiasis. *Current Opinion in Microbiology* **5**, 443-447.

Petri, W. A., Jr., Smith, R. D., Schlesinger, P. H., Murphy, C. F., and Ravdin, J. I. (1987). Isolation of the galactose-binding lectin that mediates the in vitro adherence of *Entamoeba histolytica. Journal of Clinical Investigation* **80**, 1238-44.

Petri, W. A., Jr., Chapman, M. D., Snodgrass, T., Mann, B. J., Broman, J., and Ravdin, J. I. (1989). Subunit structure of the galactose and N-acetyl-D-galactosamine-inhibitable adherence lectin of *Entamoeba histolytica*. *Journal of Biological Chemistry* **264**, 3007-3012.

Petri, W. A., Haque, R., and Mann, B. J. (2002). The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica. Annual Review of Microbiology* **56**, 39-64.

Pillai, D. R., Britten, D., Ackers, J. P., Ravdin, J. I., and Kain, K. C. (1997). A gene homologous to *hgl2* of *Entamoeba histolytica* is present and expressed in *Entamoeba dispar. Molecular and Biochemical Parasitology* **87**, 101-105.

Pillutla, R. C., Yue, Z., Maldonado, E., and Shatkin, A. J. (1998). Recombinant human mRNA cap methyltransferase binds capping enzy me/RNA polymerase IIo complexes. *Journal of Biological Chemistry* **273**, 21443-6.

Poirier, M. A., Xiao, W., Macosko, J. C., Chan, C., Shin, Y. K., and Bennett, M. K. (1998). The synaptic SNARE complex is a parallel four-stranded helical bundle. *Nature Structural Biology* **5**, 765-9.

Poole, L. B., Chae, H. Z., Flores, B. M., Reed, S. L., Rhee, S. G., and Torian, B. E. (1997). Peroxidase activity of a TSA-like antioxidant protein from a pathogenic amoeba. *Free Radical Biology & Medicine* **23**, 955-9.

Poon, R. Y., Lew, J., and Hunter, T. (1997). Identification of functional domains in the neuronal Cdk5 activator protein. *Journal of Biological Chemistry* **272**, 5703-8.

Pritham, E. J., Feschotte, C., and Wessler, S. R. (2005). Unexpected diversity and differential success of DNA transposons in four species of *Entamoeba* protozoans. *Molecular Biology and Evolution* **22**, 1751-63.

Purdy, J. E., Pho, L. T., Mann, B. J., and Petri, W. A., Jr. (1996). Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Molecular and Biochemical Parasitology* **78**, 91-103.

Que, X., and Reed, S. L. (2000). Cysteine proteinases and the pathogenesis of amebiasis. *Clinical Microbiology Reviews* **13**, 196-206.

Quevillon, E., Spielmann, T., Brahimi, K., Chattopadhyay, D., Yeramian, E., and Langsley, G. (2003). The *Plasmodium falciparum* family of Rab GTPases. *Gene* **306**, 13-25.

Quon, D. V., Delgadillo, M. G., and Johnson, P. J. (1996). Transcription in the early diverging eukaryote *Trichomonas vaginalis*: an unusual RNA

polymerase II and alpha-amanitin-resistant transcription of protein-coding genes. *Journal of Molecular Evolution* **43**, 253-62.

Radhakrishna, H., and Donaldson, J. G. (1997). ADP-ribosylation factor 6 regulates a novel plasma membrane recycling pathway. *Journal of Cell Biology* **139**, 49-61.

Ramakrishnan, G., Lee, S., Mann, B. J., and Petri, W. A., Jr. (2000). *Entamoeba histolytica*: deletion of the GPI anchor signal sequence on the Gal/GalNAc lectin light subunit prevents Its assembly into the lectin heterodimer. *Experimental Parasitology* **96**, 57-60.

Ramakrishnan, G., Gilchrist, C. A., Musa, H., Torok, M. S., Grant, P. A., Mann, B. J., and Petri, W. A., Jr. (2004). Histone acetyltransferases and deacetylase in *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **138**, 205-16.

Ramos, F., and Wiame, J. M. (1982). Occurrence of a catabolic L-serine (L-threonine) deaminase in *Saccharomyces cerevisiae. European Journal of Biochemistry* **123**, 571-6.

Ramos, M. A., Mercado, G., C,, Salgado, L. M., Sanchez-Lopez, R., Stock, R. P., Lizardi, P. M., and Alagón, A. (1997). *Entamoeba histolytica* contains a gene encoding a homologue to the 54kDa subunit of the signal recognition particle. *Molecular and Biochemical Parasitology* **88**, 225-235.

Ramponi, G., and Stefani, M. (1997). Structure and function of the low Mr phosphotyrosine protein phosphatases. *Biochimica et Biophysica Acta* **1341**, 137-56.

Randazzo, P. A., Terui, T., Sturch, S., Fales, H. M., Ferrige, A. G., and Kahn, R. A. (1995). The myristoylated amino terminus of ADP-ribosylation factor 1 is a phospholipid- and GTP-sensitive switch. *Journal of Biological Chemistry* **270**, 14809-15.

Ravdin, J. I., and Guerrant, R. L. (1981). Role of adherence in cytopathogenic mechanisms of *Entamoeba histolytica*. Study with mammalian tissue

culture cells and human erythrocytes. *Journal of Clinical Investigation* **68**, 1305-13.

Ravdin, J. I., and Guerrant, R. L. (1982). Separation of adherence, cytolytic, and phagocytic events in the cytopathogenic mechanisms of *Entamoeba histolytica*. *Archivos de Investigacion Medica (Mexico)* **13** (Suppl 3), 123-8.

Ravdin, J. I., Murphy, C. F., Guerrant, R. L., and Long-Krug, S. A. (1985). Effect of antagonists of calcium and phospholipase A on the cytopathogenicity of *Entamoeba histolytica*. *Journal of Infectious Diseases* **152**, 542-9.

Reed, S. I. (1992). The role of p34 kinases in the G1 to S-phase transition. *Annual Review of Cell Biology* **8**, 529-61.

Reed, S. L., Keene, W. E., and McKerrow, J. H. (1989). Thiol proteinase expression and pathogenicity of *Entamoeba histolytica*. *Journal of Clinical Microbiology* **27**, 2772-7.

Reed, S., Bouvier, J., Sikes Pollack, A., Engel, J. C., Brown, M., Hirata, K., Que, X., Eakin, A., Hagblom, P., Gillin, F., and McKerrow, J. H. (1993). Cloning of a virulence factor of *Entamoeba histolytica*. Pathogenic strains possess a unique cysteine proteinase gene. *Journal of Clinical Investigation* **91**, 1532-1540.

Reeves, R. E. (1968). A new enzyme with the glycolytic function of pyruvate kinase. *Journal of Biological Chemistry* **243**, 3202-4.

Reeves, R. E. (1970). Phosphopyruvate carboxylase from *Entamoeba histolytica*. *Biochimica et Biophysica Acta* **220**, 346-9.

Reeves, R. E. (1984). Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Advances in Parasitology* **23**, 105-142.

Reeves, R. E., and South, D. J. (1974). Phosphoglycerate kinase (GTP). An enzyme from *Entamoeba histolytica* selective for guanine nucleotides. *Biochemical and Biophysical Research Communications* **58**, 1053-7.

Reeves, R. E., Montalvo, F. E., and Lushbaugh, T. S. (1971). Nicotinamide-adenine dinucleotide phosphate-dependent alcohol dehydrogenase.

Enzyme from *Entamoeba histolytica* and some enzyme inhibitors. *International Journal for Biochemistry* **2**, 55-64.

Reeves, R. E., Serrano, R., and South, D. J. (1976). 6-phosphofructokinase (pyrophosphate). Properties of the enzyme from *Entamoeba histolytica* and its reaction mechanism. *Journal of Biological Chemistry* **251**, 2958-62.

Reeves, R. E., Warren, L. G., Susskind, B., and Lo, H. S. (1977). An energy-conserving pyruvate-to-acetate pathway in *Entamoeba histolytica*. Pyruvate synthase and a new acetate thiokinase. *Journal of Biological Chemistry* **252**, 726-31.

Reid, M. F., and Fewson, C. A. (1994). Molecular characterization of microbial alcohol dehydrogenases. *Critical Reviews in Microbiology* **20**, 13-56.

Rhee, S. G., Chae, H. Z., and Kim, K. (2005). Peroxiredoxins: A historical overview and speculative preview of novel mechanisms and emerging concepts in cell signaling. *Free Radical Biology & Medicine* **38**, 1543-52.

Richards, T. A., Hirt, R. P., Williams, B. A., and Embley, T. M. (2003). Horizontal gene transfer and the evolution of parasitic protozoa. *Protist* **154**, 17-32.

Roberg, K. J., Crotwell, M., Espenshade, P., Gimeno, R., and Kaiser, C. A. (1999). LST1 is a SEC24 homologue used for selective export of the plasma membrane ATPase from the endoplasmic reticulum. *Journal of Cell Biology* **145**, 659-72.

Rodriguez-Romero, A., Hernandez-Santoyo, A., del Pozo Yauner, L., Kornhauser, A., and Fernandez-Velasco, D. A. (2002). Structure and inactivation of triosephosphate isomerase from *Entamoeba histolytica*. *Journal of Molecular Biology* **322**, 669-75.

Rodriguez, M. A., Baez-Camargo, M., Delgadillo, D. M., and Orozco, E. (1996). Cloning and expression of an *Entamoeba histolytica* $NADP^+$-dependent alcohol dehydrogenase gene. *Biochimica et Biophysica Acta* **1306**, 23-26.

Rodríguez, M. A., García-Pérez, R. M., García-Rivera, G., López-Reyes, I., Mendoza, L., Ortiz-Navarrete , V., and Orozco, E. (2000). An *Entamoeba histolytica* Rab-like encoding gene and protein: function and cellular location. *Molecular and Biochemical Parasitology* **108**, 199-206.

Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahm, H. (1993). Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochemical Journal* **295**, 517-24.

Rosenbaum, R. M., and Wittner, M. (1970). Ultrastructure of bacterized and axenic trophozoites of *Entamoeba histolytica* with particular reference to helical bodies. *Journal of Cell Biology* **45**, 367-82.

Rotte, C., Stejskal, F., Zhu, G., Keithly, J. S., and Martin, W. (2001). Pyruvate : NADP+ oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a biochemical relic linking pyruvate metabolism in mitochondriate and amitochondriate protists. *Molecular Biology and Evolution* **18**, 710-20.

Rowe, J., Calegari, F., Taverna, E., Longhi, R., and Rosa, P. (2001). Syntaxin 1A is delivered to the apical and basolateral domains of epithelial cells: the role of munc-18 proteins. *Journal of Cell Science* **114**, 3323-32.

Roy, S. W., and Gilbert, W. (2005). The pattern of intron loss. *Proceedings of the National Academy of Sciences USA* **102**, 713-8.

Rusnak, F., and Mertz, P. (2000). Calcineurin: form and function. *Physiological Reviews* **80**, 1483-521.

Saavedra, E., Olivos, A., Encalada, R., and Moreno-Sanchez, R. (2004). *Entamoeba histolytica*: kinetic and molecular evidence of a previously unidentified pyruvate kinase. *Experimental Parasitology* **106**, 11-21.

Saavedra, E., Encalada, R., Pineda, E., Jasso-Chavez, R., and Moreno-Sanchez, R. (2005). Glycolysis in *Entamoeba histolytica*. Biochemical characterization of recombinant glycolytic enzymes and flux control analysis. *FEBS Journal* **272**, 1767-83.

Saavedra Lira, E., Robinson, O., and Pérez Montfort, R. (1992). Partial nucleotide sequence of the enzyme pyruvate, orthophosphate dikinase of *Entamoeba histolytica* HM-1:IMSS. *Archives of Medical Research* **23**, 39-40.

Sahoo, N., Labruyere, E., Bhattacharya, S., Sen, P., Guillen, N., and Bhattacharya, A. (2004). Calcium binding protein 1 of the protozoan parasite *Entamoeba histolytica* interacts with actin and is involved in cytoskeleton dynamics. *Journal of Cell Science* **117**, 3625-34.

Said-Fernandez, S., and Lopez-Revilla, R. (1988). Free fatty acids released from phospholipids are the major heat-stable hemolytic factor of *Entamoeba histolytica* trophozoites. *Infection and Immunity* **56**, 874-9.

Saito-Nakano, Y., Nakazawa, M., Shigeta, Y., Takeuchi, T., and Nozaki, T. (2001). Identification and characterization of genes encoding novel Rab proteins from *Entamoeba histolytica. Molecular and Biochemical Parasitology* **116**, 219-22.

Saito-Nakano, Y., Yasuda, T., Nakada-Tsukui, K., Leippe, M., and Nozaki, T. (2004). Rab5-associated vacuoles play a unique role in phagocytosis of the enteric protozoan parasite *Entamoeba histolytica*. *Journal of Biological Chemistry* **279**, 49497-507.

Saito-Nakano, Y., Loftus, B. J., Hall, N., and Nozaki, T. (2005). The diversity of Rab GTPases in *Entamoeba histolytica. Experimental Parasitology* **110**, 244-52.

Salzberg, S. L., White, O., Peterson, J., and Eisen, J. A. (2001). Microbial genes in the human genome: lateral transfer or gene loss? *Science* **292**, 1903-6.

Samarawickrema, N. A., Brown, D. M., Upcroft, J. A., Thammapalerd, N., and Upcroft, P. (1997). Involvement of superoxide dismutase and pyruvate:ferredoxin oxidoreductase in mechanisms of metronidazole resistance in *Entamoeba histolytica*. *Journal of Antimicrobial Chemotherapy* **40**, 833-40.

Samuelson, J., Banerjee, S., Magnelli, P., Cui, J., Kelleher, D. J., Gilmore, R., and Robbins, P. W. (2005). The diversity of dolichol-linked precursors

to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. *Proceedings of the National Academy of Sciences USA* **102**, 1548-53.

Sanchez, L., Horner, D., Moore, D., Henze, K., Embley, T., and Müller , M. (2002). Fructose-1,6-bisphosphate aldolases in amitochondriate protists constitute a single protein subfamily with eubacterial relationships. *Gene* **295**, 51-9.

Sanderfoot, A. A., Assaad, F. F., and Raikhel, N. V. (2000). The *Arabidopsis* genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiology* **124**, 1558-69.

Sargeaunt, P. G. (1987). The reliability of *Entamoeba histolytica* zymodemes in clinical diagnosis. *Parasitology Today* **3**, 40-43.

Sargeaunt, P. G., Williams, J. E., and Grene, J. D. (1978). The differentiation of invasive and non-invasive *Entamoeba histolytica* by isoenzyme electrophoresis. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **72**, 519-521.

Sarti, P., Fiori, P. L., Forte, E., Rappelli, P., Teixeira, M., Mastronicola, D., Sanciu, G., Giuffre, A., and Brunori, M. (2004). *Trichomonas vaginalis* degrades nitric oxide and expresses a flavorubredoxin-like protein: a new pathogenic mechanism? *Cellular and Molecular Life Sciences* **61**, 618-23.

Sata, M., Donaldson, J. G., Moss, J., and Vaughan, M. (1998). Brefeldin A-inhibited guanine nucleotide-exchange activity of Sec7 domain from yeast Sec7 with yeast and mammalian ADP ribosylation factors. *Proceedings of the National Academy of Sciences USA* **95**, 4204-8.

Satish, S., Bakre, A. A., Bhattacharya, S., and Bhattacharya, A. (2003). Stress-dependent expression of a polymorphic, charged antigen in the protozoan parasite *Entamoeba histolytica*. *Infection and Immunity* **71**, 4472-4486.

Sawyer, M. K., Bischoff, J. M., Guidry, M. A., and Reeves, R. E. (1967). Lipids from *Entamoeba histolytica*. *Experimental Parasitology* **20**, 295-302.

Schimmoller, F., Singer-Kruger, B., Schroder, S., Kruger, U., Barlowe, C., and Riezman, H. (1995). The absence of Emp24p, a component of ER-derived COPII-coated vesicles, causes a defect in transport of selected proteins to the Golgi. *EMBO Journal* **14**, 1329-39.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. *Cell* **103**, 211-25.

Schofield, P. J., and Edwards, M. R. (1994). Biochemistry - is *Giardia* opportunistic in its use of substrates? , pp. 171-183. In R. C. A. Thompson, J. A. Reynoldson, and A. J. Lymbery (Eds). *Giardia: from molecules to disease*, CAB International, Wallingford, UK.

Scholze, H., and Schulte, W. (1988). On the specificity of a cysteine proteinase from *Entamoeba histolytica*. *Biomedica Biochimica Acta* **47**, 115-23.

Scholze, H., Lohden-Bendinger, U., Müller , G., and Bakker-Grunwald, T. (1992). Subcellular distribution of amebapain, the major cysteine proteinase of *Entamoeba histolytica*. *Archives of Medical Research* **23**, 105-8.

Schroeder, S. C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes & Development* **14**, 2435-40.

Schroepfer, G. J., Jr. (1981). Sterol biosynthesis. *Annual Review of Biochemistry* **50**, 585-621.

Schulte, W., and Scholze, H. (1989). Action of the major protease from *Entamoeba histolytica* on proteins of the extracellular matrix. *Journal of Protozoology* **36**, 538-43.

Seabra, M. C., Mules, E. H., and Hume, A. N. (2002). Rab GTPases, intracellular traffic and disease. *Trends in Molecular Medicine* **8**, 23-30.

Shah, P. H., MacFarlane, R. C., Bhattacharya, D., Matese, J. C., Demeter, J., Stroup, S. E., and Singh, U. (2005). Comparative genomic

hybridizations of *Entamoeba* strains reveal unique genetic fingerprints that correlate with virulence. *Eukaryotic Cell* **4**, 504-15.

Sharma, R., Azam, A., Bhattacharya, S., and Bhattacharya, A. (1999). Identification of novel genes of non-pathogenic *Entamoeba dispar* by expressed sequence tag analysis. *Molecular and Biochemical Parasitology* **99**, 279-85.

Shaywitz, D. A., Espenshade, P. J., Gimeno, R. E., and Kaiser, C. A. (1997). COPII subunit interactions in the assembly of the vesicle coat. *Journal of Biological Chemistry* **272**, 25413-6.

Shen, Z., and Jacobs-Lorena, M. (1999). Evolution of chitin-binding proteins in invertebrates. *Journal of Molecular Evolution* **48**, 341-7.

Shiu, S. H., and Bleecker, A. B. (2001). Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences USA* **98**, 10763-8.

Shoemaker, N. B., Vlakamis, H., and Slyers, A. A. (2001). Evidence for extensive resistance gene transfer among *Bacteroides* spp. and among *Bacteroides* and other genera in the human colon. *Applied and Environmental Microbiology* **67**, 561-568.

Sies, H. (1999). Glutathione and its role in cellular functions. *Free Radical Biology & Medicine* **27**, 916-21.

Silva, P. P., Martinez-Palomo, A., and Gonzalez-Robles, A. (1975). Membrane structure and surface coat of *Entamoeba histolytica.* Topochemistry and dynamics of the cell surface: cap formation and microexudate. *Journal of Cell Biology* **64**, 538-50.

Simanis, V., and Nurse, P. (1986). The cell cycle control gene cdc2+ of fission yeast encodes a protein kinase potentially regulated by phosphorylation. *Cell* **45**, 261-8.

Singh, U., and Rogers, J. B. (1998). The novel core promoter element GAAC in the hgl5 gene of *Entamoeba histolytica* is able to direct a transcription start site independent of TATA or initiator regions. *Journal of Biological Chemistry* **273**, 21663-8.

Singh, U., Purdy, J., Mann, B. J., and Petri, W. A., Jr. (1997). Three conserved cis-acting sequences in the core promoter control gene expression in the protozoan parasite *Entamoeba histolytica. Archives of Medical Research* **28 Spec No**, 41-2.

Singh, U., Gilchrist, C. A., Schaenman, J. M., Rogers, J. B., Hockensmith, J. W., Mann, B. J., and Petri, W. A. (2002). Context-dependent roles of the *Entamoeba histolytica* core promoter element GAAC in transcriptional activation and protein complex assembly. *Molecular and Biochemical Parasitology.* **120**, 107-116.

Snell, K. (1986). The duality of pathways for serine biosynthesis is a fallacy. *Trends in Biochemical Sciences* **11**, 241–243.

Springer, S., Chen, E., Duden, R., Marzioch, M., Rowley, A., Hamamoto, S., Merchant, S., and Schekman, R. (2000). The p24 proteins are not essential for vesicular transport in *Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences USA* **97**, 4034-9.

Srivastava, G., Anand, M. T., Bhattacharya, S., and Bhattacharya, A. (1995). Lipophosphoglycan is present in distinctly different form in different *Entamoeba histolytica* strains and absent in *Entamoeba moshkovskii* and *Entamoeba invadens. Journal of Eukaryotic Microbiology* **42**, 617-22.

Stamnes, M. A., and Rothman, J. E. (1993). The binding of AP-1 clathrin adaptor particles to Golgi membranes requires ADP-ribosylation factor, a small GTP-binding protein. *Cell* **73**, 999-1005.

Stanley, S. L., Jr., Zhang, T., Rubin, D., and Li, E. (1995). Role of the *Entamoeba histolytica* cysteine proteinase in amebic liver abscess formation in severe combined immunodeficient mice. *Infection and Immunity* **63**, 1587-90.

Stark, M. J. (1996). Yeast protein serine/threonine phosphatases: multiple roles and diverse regulation. *Yeast* **12**, 1647-75.

Strauss, E., Kinsland, C., Ge, Y., McLafferty, F. W., and Begley, T. P. (2001). Phosphopantothenoylcysteine synthetase from *Escherichia coli*. Identification and characterization of the last unidentified coenzyme A

biosynthetic enzyme in bacteria. *Journal of Biological Chemistry* **276**, 13513-6.

Surana, U., Robitsch, H., Price, C., Schuster, T., Fitch, I., Futcher, A. B., and Nasmyth, K. (1991). The role of CDC28 and cyclins during mitosis in the budding yeast *S. cerevisiae*. *Cell* **65**, 145-61.

Susskind, B. M., Warren, L. G., and Reeves, R. E. (1982). A pathway for the interconversion of hexose and pentose in the parasitic amoeba *Entamoeba histolytica*. *Biochemical Journal* **204**, 191-6.

Swain, A. L., Jaskolski, M., Housset, D., Rao, J. K., and Wlodawer, A. (1993). Crystal structure of *Escherichia coli* L-asparaginase, an enzyme used in cancer therapy. *Proceedings of the National Academy of Sciences USA* **90**, 1474-8.

Sykes, D. E., and Band, R. N. (1977). Superoxide dismutase and peroxide activity of *Acanthamoeba* and two anaerobic *Entamoeba* species. *Journal of Cell Biology* **75**, 85a.

Takeuchi, T., Weinbach, E. C., and Diamond, L. S. (1977). *Entamoeba histolytica*: localization and characterization of phosphorylase and particulate glycogen. *Experimental Parasitology* **43**, 107-14.

Takeuchi, T., Weinbach, E. C., Gottlieb, M., and Diamond, L. S. (1979). Mechanism of L-serine oxidation in *Entamoeba histolytica*. *Comparative Biochemistry and Physiology B* **62**, 281-5.

Talamas-Rohana, P., Aguirre-Garcia, M. M., Anaya-Ruiz, M., and Rosales-Encina, J. L. (1999). *Entamoeba dispar* contains but does not secrete acid phosphatase as does *Entamoeba histolytica*. *Experimental Parasitology* **92**, 219-222.

Tannich, E., Bruchhaus, I., Walter, R. D., and Horstmann, R. D. (1991a). Pathogenic and nonpathogenic *Entamoeba histolytica*: identification and molecular cloning of an iron-containing superoxide dismutase. *Molecular and Biochemical Parasitology* **49**, 61-71.

Tannich, E., Ebert, F., and Horstmann, R. D. (1991b). Primary structure of the 170-kDa surface lectin of pathogenic *Entamoeba histolytica*. *Proceedings of the National Academy of Sciences USA* **88**, 1849-1853.

Tannich, E., Scholze, H., Nickel, R., and Horstmann, R. D. (1991c). Homologous cysteine proteinases of pathogenic and nonpathogenic *Entamoeba histolytica.* Differences in structure and expression. *Journal of Biological Chemistry* **266**, 4798-803.

Tannich, E., Nickel, R., Buss, H., and Horstmann, R. D. (1992). Mapping and partial sequencing of the genes coding for two different cysteine proteinases in pathogenic *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **54**, 109-11.

Temesvari, L. A., Harris, E. N., Stanley, J., S.L., and Cardellia, J. A. (1999). Early and late endosomal compartments of *Entamoeba histolytica* are enriched in cysteine proteases, acid phosphatase and several Ras-related Rab GTPases. *Molecular and Biochemical Parasitology.* **103**, 225-241.

Todd, J., Post-Beittenmiller, D., and Jaworski, J. G. (1999). KCS1 encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. *Plant Journal* **17**, 119-30.

Tokoro, M., Asai, T., Kobayashi, S., Takeuchi, T., and Nozaki, T. (2003). Identification and characterization of two isoenzymes of methionine γ-lyase from *Entamoeba histolytica*: a key enzyme of sulfur-amino acid degradation in an anaerobic parasitic protist that lacks forward and reverse trans-sulfuration pathways. *Journal of Biological Chemistry* **278**, 42717 - 42727.

Toonen, R. F., and Verhage, M. (2003). Vesicle trafficking: pleasure and pain from SM genes. *Trends in Cell Biology* **13**, 177-86.

Torii, S., Takeuchi, T., Nagamatsu, S., and Izumi, T. (2004). Rab27 effector granuphilin promotes the plasma membrane targeting of insulin granules via interaction with syntaxin 1a. *Journal of Biological Chemistry* **279**, 22532-8.

Touz, M. C., Nores, M. J., Slavin, I., Carmona, C., Conrad, J. T., Mowatt, M. R., Nash, T. E., Coronel, C. E., and Lujan, H. D. (2002). The activity of a developmentally regulated cysteine proteinase is required for cyst wall formation in the primitive eukaryote *Giardia lamblia*. *Journal of Biological Chemistry* **277**, 8474-81.

Tovar, J., Fischer, A., and Clark, C. G. (1999). The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. *Molecular Microbiology* **32**, 1013-1021.

Trilla, J. A., Duran, A., and Roncero, C. (1999). Chs7p, a new protein involved in the control of protein export from the endoplasmic reticulum that is specifically engaged in the regulation of chitin synthesis in *Saccharomyces cerevisiae*. *Journal of Cell Biology* **145**, 1153-63.

Trombetta, E. S., and Parodi, A. J. (2003). Quality control and protein folding in the secretory pathway. *Annu Rev Cell Dev Biol* **19**, 649-76.

Tye, B. K. (1999). MCM proteins in DNA replication. *Annual Review of Biochemistry* **68**, 649-86.

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

Uwanogho, D. A., Hardcastle, Z., Balogh, P., Mirza, G., Thornburg, K. L., Ragoussis, J., and Sharpe, P. T. (1999). Molecular cloning, chromosomal mapping, and developmental expression of a novel protein tyrosine phosphatase-like gene. *Genomics* **62**, 406-16.

Van Dellen, K., Field, J., Wang, Z., Loftus, B., and Samuelson, J. (2002a). LINEs and SINE-like elements of the protist *Entamoeba histolytica*. *Gene* **297**, 229-39.

Van Dellen, K., Ghosh, S. K., Robbins, P. W., Loftus, B., and Samuelson, J. (2002b). *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys chitin-binding domains. *Infection and Immunity* **70**, 3259-63.

Van Dellen, K. L., Bulik, D. A., Specht, C. A., Robbins, P. W., and Samuelson, J. C. (2006a). Heterologous expression of an *Entamoeba histolytica* chitin synthase in *Saccharomyces cerevisiae*. *Eukaryotic Cell* **5**, 203-6.

Van Dellen, K. L., Chatterjee, A., Ratner, D. M., Magnelli, P. E., Cipollo, J. F., Steffen, M., Robbins, P. W., and Samuelson, J. (2006b). Unique posttranslational modifications of chitin-binding lectins of *Entamoeba invadens* cyst walls *Eukaryotic Cell* **5**, 836-848.

van der Giezen, M., Cox, S., and Tovar, J. (2004). The iron-sulfur cluster assembly genes iscS and iscU of *Entamoeba histolytica* were acquired by horizontal gene transfer. *BMC Evolutionary Biology* **4**, 7.

Van Mullem, V., Wery, M., Werner, M., Vandenhaute, J., and Thuriaux, P. (2002). The Rpb9 subunit of RNA polymerase II binds transcription factor TFIIE and interferes with the SAGA and elongator histone acetyltransferases. *Journal of Biological Chemistry* **277**, 10220-5.

Vanacová, S., Liston, D. R., Tachezy, J., and Johnson, P. J. (2003). Molecular biology of the amitochondriate parasites, *Giardia intestinalis*, *Entamoeba histolytica* and *Trichomonas vaginalis*. *International Journal for Parasitology* **33**, 235-55.

Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S., and Inze, D. (2002). Genome-wide analysis of core cell cycle genes in *Arabidopsis*. *Plant Cell* **14**, 903-16.

Varela-Gomez, M., Moreno-Sanchez, R., Pardo, J. P., and Perez-Montfort, R. (2004). Kinetic mechanism and metabolic role of pyruvate phosphate dikinase from *Entamoeba histolytica*. *Journal of Biological Chemistry* **279**, 54124-30.

Vargas-Villarreal, J., Olvera-Rodriguez, A., Mata-Cardenas, B. D., Martinez-Rogriguez, H. G., Said-Fernandez, S., and Alagon-Cano, A. (1998). Isolation of an *Entamoeba histolytica* intracellular alkaline phospholipase A2. *Parasitology Research* **84**, 310-4.

Vargas, M., Sansonetti, P., and Guillen, N. (1996). Identification and cellular localization of the actin-binding protein ABP-120 from *Entamoeba histolytica*. *Molecular Microbiology* **22**, 849-57.

Vartiainen, M. K., and Machesky, L. M. (2004). The WASP-Arp2/3 pathway: genetic insights. *Current Opinion in Cell Biology* **16**, 174-81.

Vats, D., Vishwakarma, R. A., Bhattacharya, S., and Bhattacharya, A. (2005). Reduction of cell surface glycosylphosphatidylinositol conjugates in *Entamoeba histolytica* by antisense blocking of *E. histolytica* GlcNAc-phosphatidylinositol deacetylase expression: effect on cell proliferation, endocytosis, and adhesion to target cells. *Infection and Immunity* **73**, 8381-92.

Vayssie, L., Vargas, M., Weber, C., and Guillen, N. (2004). Double-stranded RNA mediates homology-dependent gene silencing of gamma-tubulin in the human parasite *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **138**, 21-8.

Villalobo, E., Wender, N., and Mirelman, D. (2005). *Entamoeba histolytica*: molecular characterization of an aldose 1-epimerase (mutarotase). *Experimental Parasitology* **110**, 298-302.

Vishwakarma, R. A., Anand, M. T., Arya, R., Vats, D., and Bhattacharya, A. (2006). Glycosylated inositol phospholipid from *Entamoeba histolytica*: identification and structural characterization. *Molecular and Biochemical Parasitology* **145**, 121-4.

Volpicelli-Daley, L. A., Li, Y., Zhang, C. J., and Kahn, R. A. (2005). Isoform-selective effects of the depletion of ADP-ribosylation factors 1-5 on membrane traffic. *Molecular Biology of the Cell* **16**, 4495-508.

Vowels, J. J., and Payne, G. S. (1998). A dileucine-like sorting signal directs transport into an AP-3-dependent, clathrin-independent pathway to the yeast vacuole. *EMBO Journal* **17**, 2482-93.

Walenta, J. H., Didier, A. J., Liu, X., and Kramer, H. (2001). The Golgi-associated hook3 protein is a member of a novel family of microtubule-binding proteins. *Journal of Cell Biology* **152**, 923-34.

Wang, C. C., and Cheng, H. W. (1984). The deoxyribonucleoside phosphotransferase of *Trichomonas vaginalis*. A potential target for anti-trichomonial chemotherapy. *Journal of Experimental Medicine* **160**, 987-1000.

Wang, Z., Samuelson, J., Clark, C. G., Eichinger, D., Paul, J., Van Dellen, K., Hall, N., Anderson, I., and Loftus, B. (2003). Gene discovery in the *Entamoeba invadens* genome. *Molecular and Biochemical Parasitology* **129**, 23-31.

Wassmann, C., Hellberg, A., Tannich, E., and Bruchhaus, I. (1999). Metronidazole resistance in the protozoan parasite *Entamoeba histolytica* is associated with increased expression of iron-containing superoxide dismutase and peroxiredoxin and decreased expression of ferredoxin 1 and flavin reductase. *Journal of Biological Chemistry* **274**, 26051-6.

Watson, W. H., Yang, X., Choi, Y. E., Jones, D. P., and Kehrer, J. P. (2004). Thioredoxin and its role in toxicology. *Toxicological Sciences* **78**, 3-14.

Weber, C., Guigon, G., Bouchier, C., Frangeul, L., Moreira, S., Sismeiro, O., Gouyette, C., Mirelman, D., Coppee, J. Y., and Guillen, N. (2006). Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *Entamoeba histolytica*. *Eukaryotic Cell* **5**, 871-5.

Wegmann, D., Hess, P., Baier, C., Wieland, F. T., and Reinhard, C. (2004). Novel isotypic gamma/zeta subunits reveal three coatomer complexes in mammals. *Molecular and Cellular Biology* **24**, 1070-80.

Weinbach, E. C., and Diamond, L. S. (1974). *Entamoeba histolytica*. I. Aerobic metabolism. *Experimental Parasitology* **35**, 232-43.

Weinberg, M. V., Jenney, F. E., Jr., Cui, X., and Adams, M. W. (2004). Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus* is a rubredoxin-dependent, iron-containing peroxidase. *Journal of Bacteriology* **186**, 7888-95.

Wennerberg, K., Rossman, K. L., and Der, C. J. (2005). The Ras superfamily at a glance. *Journal of Cell Science* **118**, 843-6.

Wera, S., and Hemmings, B. A. (1995). Serine/threonine protein phosphatases. *Biochemical Journal* **311**, 17-29.

Werries, E., and Thurn, I. (1989). Breakdown of glucopolysaccharides in *Entamoeba histolytica* by phosphorylase. *Journal of Protozoology* **36**, 607-12.

Werries, E., Franz, A., and Geisemeyer, S. (1990). Detection of glycogen-debranching system in trophozoites of *Entamoeba histolytica. Journal of Protozoology* **37**, 576-80.

West, C. M. (2003). Comparative analysis of spore coat formation, structure, and function in *Dictyostelium*. *International Review of Cytology* **222**, 237-93.

West, C. M., van der Wel, H., Coutinho, P. M., and Henrissat, B. (2005). Glycosyltransferase genomics in *Dictyostelium*, pp. 235-264. In W. F. Loomis, and A. Kuspa (Eds). *Dictyostelium Genomics*, Horizon Bioscience, Norfolk, UK.

Whitney, J. A., Gomez, M., Sheff, D., Kreis, T. E., and Mellman, I. (1995). Cytoplasmic coat proteins involved in endosome function. *Cell* **83**, 703-13.

WHO (1998). *The world health report 1998. Life in the 21st century: a vision for all.* World Health Organization. Geneva, Switzerland.

Willhoeft, U., and Tannich, E. (1999). The electrophoretic karyotype of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **99**, 41-53.

Willhoeft, U., Buss, H., and Tannich, E. (1999). Analysis of cDNA expressed sequence tags from *Entamoeba histolytica*: identification of two highly abundant polyadenylated transcripts with no overt open reading frames. *Protist* **150**, 61-70.

Willhoeft, U., Buss, H., and Tannich, E. (2002). The abundant polyadenylated transcript 2 DNA sequence of the pathogenic protozoan parasite

Entamoeba histolytica represents a nonautonomous non-long-terminal-repeat retrotransposon-like element which is absent in the closely related nonpathogenic species *Entamoeba dispar*. *Infection and Immunity* **70**, 6798-804.

Winder, S. J., and Ayscough, K. R. (2005). Actin-binding proteins. *Journal of Cell Science* **118**, 651-4.

Winkelmann, J., Leippe, M., and Bruhn, H. (2006). A novel saposin-like protein of *Entamoeba histolytica* with membrane-fusogenic activity. *Molecular and Biochemical Parasitology* **147**, 85-94.

Wittenberg, C., Sugimoto, K., and Reed, S. I. (1990). G1-specific cyclins of *S. cerevisiae*: cell cycle periodicity, regulation by mating pheromone, and association with the p34CDC28 protein kinase. *Cell* **62**, 225-37.

Wohlschlegel, J. A., Dhar, S. K., Prokhorova, T. A., Dutta, A., and Walter, J. C. (2002). *Xenopus* Mcm10 binds to origins of DNA replication after Mcm2-7 and stimulates origin binding of Cdc45. *Molecular Cell* **9**, 233-40.

Wood, Z. A., Poole, L. B., and Karplus, P. A. (2003a). Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* **300**, 650-3.

Wood, Z. A., Schroder, E., Robin Harris, J., and Poole, L. B. (2003b). Structure, mechanism and regulation of peroxiredoxins. *Trends in Biochemical Sciences* **28**, 32-40.

Woods, S. A., Schwartzbach, S. D., and Guest, J. R. (1988). Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochimica et Biophysica Acta* **954**, 14-26.

Wöstmann, C., Tannich, E., and Bakker-Grunwald, T. (1992). Ubiquitin of *Entamoeba histolytica* deviates in 6 amino acid residues from the consensus of all other known ubiquitins. *FEBS Letters* **308**, 54-58.

Wu, G., and Müller , M. (2003). Glycogen phosphorylase sequences from the amitochondriate protists, *Trichomonas vaginalis, Mastigamoeba balamuthi, Entamoeba histolytica* and *Giardia intestinalis*. *Journal of Eukaryotic Microbiology* **50**, 366-72.

Xu, Y., Moseley, J. B., Sagot, I., Poy, F., Pellman, D., Goode, B. L., and Eck, M. J. (2004). Crystal structures of a Formin Homology-2 domain reveal a tethered dimer architecture. *Cell* **116**, 711-23.

Yadava, N., Chandok, M. R., Prasad, J., Bhattacharya, S., Sopory, S. K., and Bhattacharya, A. (1997). Characterization of EhCaBP, a calcium binding protein of *Entamoeba histolytica* and its binding proteins. *Molecular and Biochemical Parasitology* **84**, 69-82.

Yamawaki-Kataoka, Y., Tamaoki, T., Choe, H. R., Tanaka, H., and Kataoka, T. (1989). Adenylate cyclases in yeast: a comparison of the genes from *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences USA* **86**, 5693-7.

Yang, W., Li, E., Kairong, T., and Stanley, S. L., Jr. (1994). *Entamoeba histolytica* has an alcohol dehydrogenase homologous to the multifunctional *adhE* gene product of *Escherichia coli*. *Molecular and Biochemical Parasitology* **64**, 253-260.

Yang, B., Steegmaier, M., Gonzalez, L. C., Jr., and Scheller, R. H. (2000). nSec1 binds a closed conformation of syntaxin1A. *Journal of Cell Biology* **148**, 247-52.

Yasui, K., Ishiguro, A., and Ishihama, A. (1998). Location of subunit-subunit contact sites on RNA polymerase II subunit 3 from the fission yeast *Schizosaccharomyces pombe*. *Biochemistry* **37**, 5542-8.

Yeo, M., Lin, P. S., Dahmus, M. E., and Gill, G. N. (2003). A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5. *Journal of Biological Chemistry* **278**, 26078-85.

Yin, Q. Y., de Groot, P. W., Dekker, H. L., de Jong, L., Klis, F. M., and de Koster, C. G. (2005). Comprehensive proteomic analysis of *Saccharomyces cerevisiae* cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. *Journal of Biological Chemistry* **280**, 20894-901.

Yoshihisa, T., Barlowe, C., and Schekman, R. (1993). Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. *Science* **259**, 1466-8.

Young, R. A. (1991). RNA polymerase II. *Annual Review of Biochemistry* **60**, 689-715.

Yu, X., and Warner, J. R. (2001). Expression of a micro-protein. *Journal of Biological Chemistry* **276**, 33821-5.

Zerial, M., and McBride, H. (2001). Rab proteins as membrane organizers. *Nature Reviews. Molecular and Cellular Biology* **2**, 107-17.

Zhang, J. (2000). Protein-length distributions for the three domains of life. *Trends in Genetics* **16**, 107-9.

Zhang, Z. Y., and VanEtten, R. L. (1991). Pre-steady-state and steady-state kinetic analysis of the low molecular weight phosphotyrosyl protein phosphatase from bovine heart. *Journal of Biological Chemistry* **266**, 1516-25.

Zhang, W. W., Shen, P. S., Descoteaux, S., and Samuelson, J. (1994). Cloning and expression of the gene for an $NADP^+$-dependent aldehyde dehydrogenase of *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **63**, 157-61.

Zhang, Z., Wang, L., Seydel, K. B., Li, E., Ankri, S., Mirelman, D., and Stanley, S. L., Jr. (2000). *Entamoeba histolytica* cysteine proteinases with interleukin-1 beta converting enzyme (ICE) activity cause intestinal inflammation and tissue damage in amoebiasis. *Molecular Microbiology* **37**, 542-8.

Zhang, Z., Duchene, M., and Stanley, S. L., Jr. (2002). A monoclonal antibody to the amebic lipophosphoglycan-proteophosphoglycan antigens can prevent disease in human intestinal xenografts infected with *Entamoeba histolytica*. *Infection and Immunity* **70**, 5873-6.

Zuo, X., and Coombs, G. H. (1995). Amino acid consumption by the parasitic, amoeboid protists *Entamoeba histolytica* and *E. invadens*. *FEMS Microbiology Letters* **130**, 253-258.

1

Table 1. **Genome summary statistics for selected single celled organisms with sequenced genomes**.

	Entamoeba histolytica	*Plasmodium falciparum*	*Dictyostelium discoideum*	*Saccharomyces cerevisiae*	*Encephalitozoon cuniculi*
Genome Size (Mb)	**23.7**	22.8	33.8	12.5	2.5
G+C content (%)	**24.1**	19.4	22.5	38	45.5
Gene number	**9,938**	5268	12,500	5538	1,997
Av. gene size (bp)	**1167**	2,534	1,756	1428	1077
% coding DNA	**49.2**	52.6	ND	70.5	ND
Av. protein size (aa)	**389**	761	518	475	359
Av. intergenic dist. (kb)	**0.8**	1.7	0.8	0.6	0.1
Gene density (kb per gene)	**1.9**	4.3	2.5	2.2kb	1.1
% Genes with introns	**25.2**	54	69	5	<1
Av. intron size (bp)	**102.1**	179	146	ND	-
Av. number of introns/gene	**1.5**	2.6	1.9	1	1

Table 2. **Summary properties of the repeated DNAs**.

(a)

Type	**Size (kb)**	**Estimated copy no. from genome sequence (Ref)**	**Estimated copy no. per haploid genome from hybridisation (Ref)**	**Transcript size (kb)(Ref)**
EhLINE1	4.8	142 (1) 409; 49 full-length(2)	140 (3)	No full-length transcript (4)
EhLINE2	4.72	79 (1) 290; 56 full-length (2)	Not Determined	Not Determined
EhLINE3	4.81	12 (1) 52; 3 full-length(2)	Not Determined	Not Determined
EhSINE1	0.5-0.6	219 (1) 272; 81 full-length(2) 214; >90 full-length (3)	500	0.7 (6)
EhSINE2	0.65	120 (1) 117; 62 full-length(2) 122; ~50 full-length (3)	Not Determined	0.75 (7)
EhSINE3	0.58	1 (1,2)	Not Determined	Not detected (3)
Tr	0.7	1 per rDNA episome (5)	Not Determined	0.7 (5)
BspA-like	0.96	77 (8)	190 (3)	Not detected(3) [a]
Ehssp1	0.9-1.1	Not Determined	306 (9)	1.5 (9)

(b)

Family 16	GTAATGAATATAYAACTAAGAATTTCATTTAAAATGRATATG
Family 17	CAACAAATAAATRGKTTCAATAAAATA

(a) References for data: (1): (Van Dellen *et al.*, 2002a), (2): (Bakre *et al.*, 2005), (3): This analysis, (4) Bakre and Bhattacharya, unpublished observations; (5): (Burch *et al.*, 1991), (6): (Cruz-Reyes *et al.*, 1995), (7): Shire and Ackers, submitted, (8): (Davis *et al.*, 2006), (9): (Satish *et al.*, 2003). [a] - although no transcript was detected the protein has been demonstrated on the cell surface and in Western blots using antibodies (Davis *et al.*, 2006).

(b) Consensus sequences of Family 16 and 17 repeats. Standard abbreviation for degenerate sequence positions are used: R= purine, Y= pyrimidine, K= G or T.

Table 3. **Number and ranking of Pfam domains across different genomes**.

		EH		EC		PF		SC		AT		CE		DD	
		#	Rank	#	Rank	#	Rank	#	Rank	#	Rank	#	Rank	#	Rank
WD40	WD domain, G-beta repeat	249	1	139	1	287	2	414	1	1137	3	694	1	719	2
LRR_1	Leucine Rich Repeat	131	2	40	2	55	12	43	17	3793	2	494	5	372	4
Pkinase	Protein kinase domain	95	3	27	5	78	8	116	2	839	4	405	8	225	7
HEAT	HEAT repeat	70	4	13	15	44	17	114	3	220	17	162	26	108	12
efhand	EF hand	58	5	7	28	80	7	29	25	422	8	213	20	153	9
RRM_1	RNA recognition motif.	57	6	30	3	95	6	86	6	375	10	223	19	134	10
Ras	Ras family	46	7	9	22	13	44	25	28	78	68	66	76	126	11
TPR_1	Tetratricopeptide repeat	42	8	23	7	48	15	103	4	334	12	180	22	168	8
Ank	Ankyrin repeat	34	9	6	34	55	12	61	9	431	6	629	2	446	3
PUF	Pumilio-family RNA binding repeat	33	10	8	23	15	34	51	13	142	32	75	68	34	62
RhoGAP	RhoGAP domain	27	11	2	118	1	520	11	80	9	559	31	138	45	39
Myb_DNA-binding	Myb-like DNA-binding domain	22	12	15	12	10	62	21	34	424	7	30	141	55	26
RhoGEF	RhoGEF domain	22	12	1	230	0	1215	3	366	0	2581	34	130	47	37
Helicase_C	Helicase conserved C-terminal domain	20	14	28	4	64	11	74	8	150	31	98	49	84	20
DEAD	DEAD/DEAH box helicase	20	14	22	9	49	14	59	10	103	50	76	67	48	35
PH	PH domain	19	16	1	230	5	123	25	28	22	255	77	63	94	16
Metallophos	Calcineurin-like phosphoesterase	19	16	6	34	16	32	21	34	66	83	78	62	31	67
Gelsolin	Gelsolin repeat	18	18	2	118	2	295	4	255	33	169	12	323	29	68
LIM	LIM domain	17	19	0	703	0	1215	8	116	16	341	103	47	56	25
CH	Calponin homology (CH) domain	16	20	4	54	1	520	7	137	26	211	57	87	49	33
Filamin	Filamin/ABP280 repeat	16	20	0	703	1	520	0	1842	2	1450	55	91	10	203

Columns labeled "#" give the total number of occurrences of a particular domain. Columns labeled "Rank" give the ranking of the domain where the most common domain is ranked 1. The organisms shown are *Entamoeba histolytica* (EH), *Encephalitozoon cuniculi* (EC), *Plasmodium falciparum* (PF), *Arabidopsis thaliana* (AT), *Saccharomyces cerevisae* (SC), *Dictyostelium discoideum* (DD)..

Table 4. **Family C1-like cysteine endopeptidases of *E. histolytica*.**

Protein Name	Previous designation	Accession No.	Protein length Total (pre,pro,mature)	Active site residues	Conserved motifs	Remarks
EhCP-A1	EhCP1	XP_650156	315 (13,80,222)	QCHN	ERFNIN, DWR	
EhCP-A2	EhCP2	XP_650642	315 (13,80,222)	QCHN	ERFNIN, DWR	
EhCP-A3	EhCP3	XP_653254	308 (13,79,216)	QCHN	ERFNIN, DWR	
EhCP-A4	EhCP4	XP_656602	311 (20,73,218)	QCHN	ERFNIN, DWR	
EhCP-A5	EhCP5	XP_650937	318 (20,72,225)	QCHN	ERFNIN, DWR, RGD	Degenerate in *E. dispar*
EhCP-A6	EhCP6	XP_657364	320 (17,79,224)	QCHN	ERFNIN, DWR	
EhCP-A7	EhCP8	XP_648996	315 (13,80,222)	QCHN	ERFNIN, DWR	
EhCP-A8	EhCP9	XP_657446	317 (15,82,220)	QCHN	ERFNIN, DWR	
EhCP-A9	EhCP10	XP_655675	297 (17,90,190)	QCHN	ERFNIN, DWR	
EhCP-A10	EhCP17	XP_651147	420 (18,148,254)	QCHN	ERFNIN, DWR	
EhCP-A11	EhCP19	XP_651690	324 (17,79,228)	QC IN [a]	ERFNIN, DWR	
EhCP-A12	new	XP_653823	317 (14,83,220)	(d)	ERFNIN, DWR	
EhCP-B1	EhCP7	XP_651581	426 (15,106,305)	QCHN	ERFNIN, PCNC	hydrophobic C-terminus
EhCP-B2	EhCP11	AAO03568	431 (15,106,310)	QCHS [a]	ERFNIN, PCNC	GPI cleavage site
EhCP-B3	EhCP12	XP_656747	474 (16,107,351)	QCHN	ERFNIN, PCNC	TMH:444-466 aa
EhCP-B4	EhCP13	XP_648501	379 (16,105,258)	QCHN	ERFNIN, PCNC	TMH or GPI cleavage site
EhCP-B5	EhCP14	XP_652671	434 (12,108,314)	QCHN	ERFNIN, PCNC	GPI cleavage site
EhCP-B6	EhCP15	XP_652465	300 (14,55,231)	QCHN	PCNC	hydrophobic C-terminus
EhCP-B7	EhCP16	XP_650400	650 (18,144,488)	QCHN	ERFNIN, PCNC	hydrophobic C-terminus, Cys-rich profile
EhCP-B8	EhCP18	XP_651049	473 (15,105,353)	QCHN	ERFNIN, PCNC, RGD	GPI cleavage site
EhCP-B9	EhCP112	XP_652993	446 (19,112,315)	QCHN	ERFNIN, PCNC, RGD	hydrophobic C-terminus, Cys-rich profile
EhCP-B10	new	XP_648306	372 (b)	QCHN	ERFNIN, PCNC, RGD	hydrophobic C-terminus
EhCP-B11	new	XP_648013	133 (b)	Q ? ? ?	PCNC	
EhCP-C1	new	XP_654453	586 (c)	QCIN [a]	$HS(X)_6ICP$	TMH:12-34
EhCP-C2	new	XP_656632	567 (c)	QCHN	$HS(X)_6ICP$	TMH:27-49
EhCP-C3	new	XP_655128	572 (c)	QCHN	$HS(X)_6LCP$	TMH:17-39
EhCP-C4	new	XP_655800	502 (c)	QCHN	$LT(X)_6LCP$	
EhCP-C5	new	XP_654800	557 (c)	QCHN	$IS(X)_6ICP$	TMH:20-42
EhCP-C6	new	XP_651553	557 (c)	QCHD [a]	$HS(X)_6LCA$	TMH:14-36
EhCP-C7	new	XP_657273	595 (c)	QCHN	$IS(X)_6LCP$	TMH:19-41
EhCP-C8	new	XP_655479	627 (c)	QCHN	$IS(X)_6ICP$	TMH:29-51

1	EhCP-C9	new	XP_655011	518 (c)	(d)	$HS(X)_6ICP$	TMH:12-34
2	EhCP-C10	new	XP_654829	530 (c)	QCHN	$IS(X)_6ICP$	TMH:15-37
3	EhCP-C11	new	XP_648083	526 (c)	(d)	$HS(X)_6ICP$	TMH:20-42
4	EhCP-C12	new	XP_650829	473 (c)	(d)	$MS(X)_6LCG$	TMH:26-48 & 449-471
5	EhCP-C13	new	XP_656556	564 (c)	QCHN	$VS(X)_6RCG$	TMH:21-43
6							

7 a - active sites that lack the canonical motif QCHN; b - incomplete sequence; c - cleavage sites to be determined; d - not conserved

Table 5. **Family C2-, C19-, C54-, and C65-like Cysteine endopeptidases of *E. histolytica***

Name	Homology	Family	ProteinID	Protein length	Active site
EhCALP1	Calpain-like	C2	XP_649922	591 aa	not cons.
EhCALP2	Calpain-like	C2	XP_657312	473 aa	QCHN
EhUBHY	Ubiquitin Hydrolase-like	C19	XP_657356	444 aa	NDTN
EhAUTO1	Autophagin-like	C54	XP_651386	325 aa	YCHS
EhAUTO2	Autophagin-like	C54	XP_653798	364 aa	YCHD
EhAUTO3	Autophagin-like	C54	XP_652043	364 aa	YCHD
EhAUTO4	Autophagin-like	C54	XP_656724	348 aa	YCHD
EhOTU	Otubain-like	C65	XP_654013	259 aa	DCH

Table 6. **Attributes of the identified SAPLIPs of *E. histolytica***

Name	Size, aa			SAPLIP domain		Identical to annotated protein	SAPLIP domain can be found within this sequence[g]	Similar to (aa sequence identity, %)[f]	Homologous proteins in other organisms (aa sequence identity; %)
	entire	signal peptide (predicted)[a]	proform / mature[c]	position aa residues	similarity[e] name, Acc. no.				
Amoebapore A	98	21[b]	77	22 -98	SAPOSIN B, IPR 008139	amoebapore A precursor ACCESSION XP_653265		Amoebapore A Acc. AAA29111 (100%)	Disparpore A Acc. AAA18632 *E.dispar* (94%)
SAPLIP 1	92	15	77	16 - 92	SAPOSIN B, IPR 008139	saposin-like protein ACCESSION XP_655836		Amoebapore A Acc. AAA29111 (64%)	Disparpore A Acc. AAA18632 *E.dispar* (68%)
Amoebapore B	96	19[b]	77	20 - 96	SAPOSIN B, IPR 008139	Pore-forming peptide amoebapore B precursor (EH-APP) ACCESSION Q24824		Amoebapore B Acc.CAA54226 (100%)	Disparpore B Acc.AAF04195 *E.dispar* (90%)
Amoebapore C	101	24[b]	77	25 - 101	SAPOSIN B, IPR 008139	amoebapore C ACCESSION XP_656029		Amoebapore C Acc. CAA54225 (100%)	Disparpore C Acc.AAF04196 *E.dispar* (88%)
SAPLIP 2	153	15	138	71 - 153	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_656037		-	-
SAPLIP 3	94	16	78	18 - 94	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_656682		Amoebapore A Acc. AAA29111 (30%)	Invapore X Acc. AAP80381 *E.invadens* (67%)
SAPLIP 4	96	17	79	18 - 96	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_652159 AND hypothetical protein ACCESSION XP_652303		Amoebapore C Acc.CAA54225 (27%)	Disparpore C Acc.AAF04196 *E.dispar* (30%)
SAPLIP 5	1026	18	1008	946 - 1026	SAPOSIN B, IPR 008139	chromosome partition protein ACCESSION XP_655789		-	-
SAPLIP 6	92	15	77	14 - 92	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_655820		-	-
SAPLIP 7	926	17	909	855 - 926	SAPOSIN B, IPR 008139	conserved hypothetical protein ACCESSION XP_656441		-	-
SAPLIP 8	980	15	965	902 - 980	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_656913		-	-
SAPLIP 9	140	15	125	61 - 140	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_650376		-	-
SAPLIP 10	657	16	641	577 - 657	SAPOSIN B, IPR 008139	-	Genomic survey sequence ACCESSION AZ687176	-	-
SAPLIP 11	693	17	676	615 - 693[d]	-	-	Genomic survey sequence ACCESSION AZ692153	-	-
SAPLIP 12	873	16	857	793 - 873	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_652721		-	-
SAPLIP 13	1009	no signal peptide predicted	1009	931 - 1005	SAPOSIN B, IPR 008139	hypothetical protein ACCESSION XP_655089		-	-
SAPLIP 14	915	17	898	834 - 915	SAPOSIN B, IPR 008139	-	Genomic survey sequence ACCESSION AZ690015	-	-

SAPLIPs were named according to the similarity of their SAPLIP domain to amoebapore A:

[a] by the programme SignalP and manually corrected if predicted cleavage site is within the SAPLIP domain

[b] verified by experimental data

[c] with the exception of amoebapores it is not possible to decide whether proteins are further processed

[d] identified manually

[e] extracted from InterPro databases

[f] if no similarity is reported, there is none outside of the SAPLIP domain

[g] sequences only found in GSS section of GenBank with given identifier

Table 7. **The number of genes encoding representative proteins involved in vesicular trafficking in *E. histolytica*.**

Protein	*E. histolytica*	*S. cerevisiae*	*C. elegans*	*D. melanogaster*	*H. sapiens*	*A. thaliana*	References
Sar1	1	1	1	1	2	4	1,2
COPII	9	6	5	4	9	12	1
Arf	10	6	11	11	25	17	3
COPI	11	7	7	7	9	9	1
AP-1		5	7	5	8	9	
AP-2		4	5	5	5	6	
AP-3		4	4	4	7	4	
AP-4		0	0	0	4	4	
AP total	18	13	16	14	24	23	1
Rab	91	11	29	26	60	57	1,4
Qa	8	7	9	7	12	18	
Qb	10(b+c)	6	7	5	9	11	
Qc		8	4	5	8	8	
R	10	5	6	5	9	14	
SNARE total	28	24	23	20	35	54	1,5,6
NSF	1	1	1	2	1	1	1,7
SNAP	1	1	1	3	1	3	
Sec1	5	4	6	5	7	6	8

References (1), Bock *et al.* (2001); (2), Wennerberg *et al.* (2005); (3), Pasqualato *et al.* (2002); (4), Pereira-Leal and Seabra (2001); (5), Burri and Lithgow (2004); (6), Uemura *et al.* (2004); (7), Sanderfoot *et al.* (2000); (8) Boehm *et al.* (2001).

Table 8. Reassessment of the 96 candidate LGT cases identified in the original genome publication.

Acc.[a] RefSeq	Acc.[a]	EhL[b]	Top Prokaryotic blast hit	PL[c]	%ID[d]	Top Eukayotic blast hit	EL[c]	%ID[d]	PE-score[e]	EE-score[f]	P/E Ratio[g]
41 LGT cases that remain strongly supported according to our criteria											
EAL43201	XP_648590.1	487	*Treponema denticola*	507	57	*Trichomonas vaginalis*	398	43	1.00E-167	5.00E-88	2.00E-80
EAL43619	XP_649008.1	621	*Vibrio vulnificus*	673	41	*Saccharomyces cerevisiae*	664	40	1.00E-132	1.00E-125	1.00E-07
EAL43678	XP_649067.1	538	*Fusobacterium nucleatum*	562	47	*Trichomonas vaginalis*	477	34	1.00E-135	2.00E-60	5.00E-76
EAL43850	XP_649240.1	880	*Mannheimia succiniciproducens*	898	63	*Mastigamoeba balamuthi*	882	45	0	0	N/A
EAL44182	XP_649570.1	260	*Bacteroides thetaiotaomicron*	273	34	*Yarrowia lipolytica*	298	29	2.00E-35	4.00E-10	5.00E-26
EAL44226	XP_649612.1	262	*Bacteroides thetaiotaomicron*	267	28	*Tetrahymena thermophila*	1476	30	2.00E-25	0.11	1.82E-24
EAL44778	XP_650165.1	188	*Bacteroides thetaiotaomicron*	188	43	*Neurospora crassa*	546	34	8.00E-41	1.8	4.44E-41
EAL45076	XP_650453.1	358	*Bacteroides fragilis*	362	46	*Trichomonas vaginalis*	562	22	1.00E-87	0.24	4.17E-87
EAL45145	XP_650531.1	825	*Staphylococcus aureus*	1036	30	*Trichomonas vaginalis*	2468	20	3.00E-59	0.016	1.88E-57
EAL45220	XP_650606.1	479	*Clostridium tetani*	471	45	*Arabidopsis thaliana*	581	31	1.00E-114	1.00E-54	1.00E-60
EAL44744	XP_650131.1	160	*Bacteroides fragilis*	424	41	*Yarrowia lipolytica*	169	31	3.00E-24	7.00E-11	4.29E-14
EAL46110	XP_651498.1	157	*Bacteroides fragilis*	166	49	*Arabidopsis thaliana*	627	35	5.00E-35	3.2	1.56E-35
EAL45378	XP_650765.1	311	*Haloarcula marismortui*	299	43	*Leishmania major*	411	43	3.00E-54	1.00E-32	3.00E-22
EAL45618	XP_651004.1	159	*Bacteroides thetaiotaomicron*	157	46	*Plasmodium vivax*	1275	33	2.00E-28	0.69	2.90E-28
EAL46311	XP_651697.1	248	*Synechococcus elongates*	270	36	*Trichomonas vaginalis*	3075	18	1.00E-30	0.38	2.63E-30
EAL46679	XP_652065.1	218	*Methanosarcina mazei*	230	37	*Candida glabrata*	461	24	8.00E-31	0.079	1.01E-29
EAL46975	XP_652361.1	370	*Bordetella bronchiseptica*	368	46	*Cryptococcus neoformans*	372	40	8.00E-83	3.00E-71	2.67E-12
EAL47525	XP_652912.1	380	*Clostridium perfringens*	296	23	*Plasmodium falciparum*	390	34	2.00E-13	1.3	1.54E-13
EAL47905	XP_653291.1	227	*Clostridium perfringens*	259	33	*Tetrahymena thermophila*	1425	24	4.00E-19	0.32	1.25E-18
EAL48587	XP_653973.1	425	*Desulfovibrio vulgaris*	442	60	*Yarrowia lipolytica*	572	37	1.00E-149	9.00E-57	1.11E-93
EAL48979	XP_654365.1	732	*Thermotoga neapolitana*	740	40	*Cryptococcus neoformans*	735	28	1.00E-135	3.00E-64	3.33E-72
EAL49084	XP_654474.1	350	*Methanococcus jannaschii*	241	29	*Anopheles gambiae*	784	40	1.00E-24	5.00E-06	2.00E-19
EAL49209	XP_654596.1	247	*Bacteroides fragilis*	243	38	*Thalassiosira pseudonana*	269	22	7.00E-43	0.0002	3.50E-39
EAL49277	XP_654665.1	737	*Bacteroides thetaiotaomicron*	781	31	*Cryptococcus neoformans*	935	24	1.00E-111	6.00E-44	1.67E-68
EAL49613	XP_654999.1	168	*Sulfolobus solfataricus*	237	34	*Tetrahymena thermophila*	487	38	1.00E-16	6.00E-06	1.67E-11
EAL49813	XP_655200.1	186	*Escherichia coli*	200	31	*P. brasiliensis*	257	26	2.00E-13	0.47	4.26E-13
EAL49869	XP_655257.1	390	*Campylobacter jejuni*	407	56	*Ashbya gossypii*	490	39	1.00E-124	8.00E-73	1.25E-52
EAL50263	XP_655646.1	390	*Porphyromonas gingivalis*	408	48	*Yarrowia lipolytica*	428	38	1.00E-98	3.00E-60	3.33E-39
EAL50440	XP_655826.1	344	*Bacillus anthracis*	491	54	*Rhizopus oryzae*	510	40	1.00E-101	2.00E-67	5.00E-35
EAL50508	XP_655888.1	348	*Wolinella succinogenes*	340	55	*Mus musculus*	168	40	1.00E-106	2.00E-18	5.00E-89
EAL50603	XP_655988.1	567	*Bacteroides thetaiotaomicron*	622	45	*Trichomonas vaginalis*	632	39	1.00E-141	2.00E-99	5.00E-43
EAL50801	XP_656185.1	499	*Bacteroides thetaiotaomicron*	513	52	*Trichomonas vaginalis*	514	28	1.00E-145	3.00E-40	3.33E-106
EAL50992	XP_656375.1	140	*Archaeoglobus fulgidus*	184	40	*Trichomonas vaginalis*	195	46	1.00E-27	0.018	5.56E-26

EAL50997	XP_656380.1	656	*Bacteroides thetaiotaomicron*	718	53	*Cryptococcus neoformans*	770	32	0	2.00E-69	0.00E+00
EAL51149	XP_656535.1	343	*Bacteroides fragilis*	359	43	*Pichia ofunaensis*	378	34	8.00E-84	1.00E-53	8.00E-31
EAL51236	XP_656622.1	259	*Symbiobacterium thermophilum*	274	45	*Oryza sativa*	315	21	3.00E-51	0.003	1.00E-48
EAL51348	XP_656749.1	171	*Methanopyrus kandleri*	204	37	*Tetrahymena thermophila*	2872	22	3.00E-21	0.1	3.00E-20
EAL51525	XP_656903.1	316	*Bacteroides thetaiotaomicron*	300	29	*Candida boidinii*	314	32	8.00E-27	0.0007	1.14E-23
EAL51565	XP_656946.1	415	*Clostridium perfringens*	900	43	*Trichomonas vaginalis*	897	40	1.00E-89	5.00E-81	2.00E-09
EAL51925	XP_657304.1	448	*T. tengcongensis*	481	43	*Giardia lamblia*	937	33	3.00E-96	2.00E-60	1.50E-36
EAL52001	XP_657387.1	303	*Oceanobacillus iheyensis*	306	27				2.00E-15	0.00E+00	
27 LGT cases that are more weakly supported than before according to our criteria											
EAL45152	XP_650539.1	122	*Shewanella oneidensis*	132	34	*Trypanosoma bruzeii*	385	24	5.00E-10	6.6	7.58E-11
EAL43347	XP_648734.1	848	*Burkholderia pseudomallei*	779	38	*Plasmodium falciparum*	2463	32	1.00E-136	4.00E-44	2.50E-93
EAL44257	XP_649643.1	407	*Clostridium acetobutylicum*	406	25	*Homo sapiens*	468	24	6.00E-23	1.00E-14	6.00E-09
EAL45586	XP_650972.1	460	*Clostridium tetani*	476	47	*Xenopus laevis*	513	38	1.00E-116	5.00E-84	2.00E-33
EAL46313	XP_651699.1	118	*Prochlorococcus marinus*	163	42	*Hordeum vulgare*	223	22	2.00E-21	1.4	1.43E-21
EAL46399	XP_651785.1	218	*Clostridium perfringens*	235	65	*Trypanosoma bruzeii*	295	52	3.00E-73	9.00E-54	3.33E-20
EAL46421	XP_651808.1	205	*Clostridium acetobutylicum*	230	40	*Arabidopsis thaliana*	241	33	7.00E-34	6.00E-12	1.17E-22
EAL46701	XP_652087.1	294	*Bacteroides fragilis*	308	45	*Thalassiosira pseudonana*	348	27	4.00E-63	1.00E-14	4.00E-49
EAL46757	XP_652143.1	95	*Lactococcus lactis*	103	31	*Tetrahymena thermophila*	112	32	3.00E-09	1.00E-07	3.00E-02
EAL46858	XP_652245.1	192	*Pseudomonas aeruginosa*	195	41	*Caenorhabditis briggsae*	229	40	6.00E-36	2.00E-17	3.00E-19
EAL47026	XP_652397.1	164	*Bacillus subtilis*	181	30	*Trichomonas vaginalis*	182	26	3.00E-10	2.00E-08	1.50E-02
EAL47464	XP_652839.1	504	*Treponema denticola*	509	39	*Piromyces sp.*	555	27	5.00E-88	2.00E-30	2.50E-58
EAL47648	XP_653034.1	259	*Methanosarcina mazei*	272	36	*Arabidopsis thaliana*	345	25	2.00E-39	3.00E-11	6.67E-29
EAL47787	XP_653173.1	546	*Spirochaeta thermophila*	571	56	*Solanum tuberosum*	552	46	1.00E-175	1.00E-135	1.00E-40
EAL48186	XP_653572.1	232	*Bacillus cereus*	279	34	*Thalassiosira pseudonana*	271	32	2.00E-10	2.00E-08	1.00E-02
EAL49309	XP_654698.1	358	*Methanosarcina mazei*	379	42	*Leishmania major*	373	31	5.00E-77	9.00E-44	5.56E-34
EAL48568	XP_653954.1	113	*Chlamydia pneumoniae*	271	38	*Debaryomyces hansenii*	699	38	5.00E-14	7.00E-16	7.14E+01
EAL48767	XP_654156.1	165	*Bacteroides fragilis*	177	40	*Trichomonas vaginalis*	189	28	7.00E-28	2.00E-05	3.50E-23
EAL48783	XP_654172.1	217	*Pseudomonas putida*	225	46	*Giardia lamblia*	239	35	2.00E-43	7.00E-24	2.86E-20
EAL49703	XP_655090.1	396	*Clostridium acetobutylicum*	398	34	*Tetrahymena thermophila*	445	29	4.00E-64	3.00E-44	1.33E-20
EAL49996	XP_655383.1	358	*Bacteroides thetaiotaomicron*	368	60	*Brachydanio rerio*	367	43	1.00E-121	5.00E-76	2.00E-46
EAL50325	XP_655711.1	447	*Clostridium tetani*	448	30	*Trichomonas vaginalis*	871	29	4.00E-46	1.00E-37	4.00E-09
EAL50521	XP_655905.1	285	*Streptococcus agalactiae*	323	29	*Leishmania major*	452	24	2.00E-22	3.00E-06	6.67E-17
EAL50620	XP_656005.1	261	*Wolinella succinogenes*	655	27	*Trichomonas vaginalis*	261	28	6.00E-21	1.00E-06	6.00E-15
EAL50838	XP_656225.1	299	*Anabaena* sp.	287	27	*Trichomonas vaginalis*	336	29	4.00E-15	0.0009	4.44E-12
EAL50986	XP_656369.1	219	*Bacteroides thetaiotaomicron*	240	31	*Xenopus laevis*	309	29	2.00E-20	1.00E-12	2.00E-08
EAL52121	XP_657511.1	220	*T. tengcongensis*	222	36	*Caenorhabditis elegans*	255	26	1.00E-30	1.00E-07	1.00E-23
14 cases where increased sampling has weakened that case for LGT											
EAL42539	XP_647925.1	213	*Bacteroides thetaiotaomicron*	319	47	*Entodinium caudatum*	411	43	3.00E-53	1.00E-32	3.00E-21
EAL42738	XP_648124.1	313	*Campylobacter jejuni*	324	40	*Trichomonas vaginalis*	313	36	1.00E-63	4.00E-42	2.50E-22
EAL44270	XP_649657.1	179	*Methanococcus maripaludis*	193	37	*Anopheles gambiae*	186	21	2.00E-27	2.00E-09	1.00E-18

EAL44593	XP_649979.1	220	*Vibrio vulnificus*	244	24	*Trichomonas vaginalis*	238	21	0.0002	2.6	7.69E-05
EAL45320	XP_650707.1	154	*Geobacillus kaustophilus*	183	53	*Thalassiosira pseudonana*	182	43	8.00E-38	2.00E-32	4.00E-06
EAL45332	XP_650718.1	392	*Methanosarcina acetivorans*	420	48	*Trichomonas vaginalis*	396	47	8.00E-99	2.00E-93	4.00E-06
EAL45528	XP_650913.1	349	*Sulfolobus acidocaldarius*	343	28	*Cyanophora paradoxa*	313	27	1.00E-24	5.00E-17	2.00E-08
EAL45907	XP_651293.1	380	*Streptomyces coelicolor*	603	32	*Dictyostelium discoideum*	457	30	2.00E-39	2.00E-35	1.00E-04
EAL46026	XP_651412.1	176	*Bacteroides fragilis*	184	51	*Tetrahymena thermophila*	323	32	2.00E-44	8.00E-08	2.50E-37
EAL46116	XP_651488.1	662	*Bacillus clausii*	684	48	*Solanum tuberosum*	761	48	0	1.00E-172	0.00E+00
EAL46656	XP_652044.1	419	*Dictyoglomus thermophilum*	579	30	*S. pombe*	493	41	2.00E-35	2.00E-19	1.00E-16
EAL50605	XP_655990.1	392	*Thermotoga maritima*	417	38	*Cryptococcus neoformans*	445	30	2.00E-69	1.00E-33	2.00E-36
EAL51270	XP_656656.1	251	*Porphyromonas gingivalis*	261	50	*Anopheles gambiae*	272	39	6.00E-53	1.00E-35	6.00E-18
EAL52102	XP_657492.1	345	*Bacteroides thetaiotaomicron*	358	54	*Thalassiosira pseudonana*	354	47	1.00E-105	7.00E-86	1.43E-20
Nine cases where *Entamoeba* is now recovered with a recently sequenced gene from another microbi al eukaryote											
EAL44213	XP_649600.1	710	*Bdellovibrio bacteriovorus*	698	37	*Trichomonas vaginalis*	713	35	1.00E-127	1.00E-127	1.00E+00
EAL44435	XP_649823.1	250	*Bacteroides fragilis*	395	40	*Trichomonas vaginalis*	395	33	1.00E-43	3.00E-35	3.33E-09
EAL44766	XP_650152.1	401	*Porphyromonas gingivalis*	419	36	*Trichomonas vaginalis*	445	32	3.00E-65	1.00E-51	3.00E-14
EAL47785	XP_653171.1	234	*Bacillus anthracis*	242	32	*Trichomonas vaginalis*	256	39	2.00E-30	3.00E-33	6.67E+02
EAL47859	XP_653246.1	337	*Clostridium acetobutylicum*	322	50	*C. reinhardtii*	352	44	9.00E-74	0	N/A
EAL49158	XP_654544.1	397	*T. tengcongensis*	412	49	*Trichomonas vaginalis*	416	46	1.00E-100	4.00E-99	2.50E-02
EAL49488	XP_654874.1	320	*Geobacter sulfurreducens*	336	34	*Leishmania major*	357	31	1.00E-38	4.00E-30	2.50E-09
EAL49791	XP_655177.1	164	*Oceanobacillus iheyensis*	177	42	*Thalassiosira pseudonana*	96	38	8.00E-30	6.00E-09	1.33E-21
EAL50404	XP_655790.1	718	*T. tengcongensis*	717	37	*Trichomonas vaginalis*	721	34	1.00E-139	1.00E-118	1.00E-21
Five cases where vertical inheritance is now the simplest explanation for the new tree											
EAL44346	XP_649732.1	314	*Oceanobacillus iheyensis*	239	47	*Dictyostelium discoideum*	278	65	1.00E-52	3.00E-95	3.33E+42
EAL45466	XP_650849.1	209	*Agrobacterium tumefaciens*	254	31	*Thalassiosira pseudonana*	227	35	3.00E-23	1.00E-27	3.00E+04
EAL45548	XP_650934.1	259	*Bacillus cereus (strain ZK)*	233	29	*Candida glabrata*	270	30	7.00E-06	5.00E-05	1.40E-01
EAL45595	XP_650981.1	284	*Pyrobaculum aerophilum*	293	27	*Ashbya gossypii*	343	27	1.00E-23	7.00E-16	1.43E-08
EAL50185	XP_655571.1	186	*Aeropyrum pernix*	192	31	*Thalassiosira pseudonana*	149	30	4.00E-13	5.00E-06	8.00E-08

All 96 trees reanalysed here can be downloaded (in pdf format) from the following web site: http://www.ncl.ac.uk/microbial_eukaryotes/

[a]GenBank accession numbers and RefSeq accession numbers, respectively, for the 96 original candidates LGT identified by phylogenetic analysis (Loftus *et al.*, 2005)

[b]EhL, the length of the *E. histolytica* protein

[c]PL/EL, the protein length of the prokaryotic or eukaryotic top BlastP hit, respectively

[d]%ID, the percent identity between the *E. histolytica* protein and the top prokaryotic or eukaryotic protein in BlastP alignments (in respective columns)

[e]PE-score, the e-score of the top prokaryotic hit

[f]EE-score, the e-score of the top eukaryotic hit

[g]P/E Ratio, the e-score ratio between the top prokaryotic hit and top eukaryotic hit

Abbreviated taxon names (to fit the columns):

Chlamydomonas reinhardtii: C. reinhardtii; Paracoccidioides brasiliensis: P. brasiliensis; *Schizosaccharomyces pombe: S. pombe*; *Thermoanaerobacter tengcongensis: T. tengcongensis*

Table 9. **Examples of microarray-detected transcriptional changes in some gene families and the conditions tested**

Gene family	Total number of genes in gene family	Number of genes transcriptionally regulated under condition tested	
		Heat shock[a] (1,131 genes on array)	**Host colonisation and invasion**[b] (9,435 genes on array)
Cysteine proteinases	29[c]	2 upregulated (CPs 6, 4); 7 down-regulated (CPs 1, 2, 3, 8, 13, 17,	21 genes on array; 4 up-regulated (CPs 1, 9, 4, 6);1 down-regulated (CP8)
Lectin (Heavy, Light, and Intermediate subunits)	12	1 up-regulated (Hgl-2); 5 down-regulated (Lgl-1 and 3, Igl 1 and 2, Hgl-3)	No change in heavy or intermediate subunits; Light subunit lgl2 and lgl3 down-regulated)
Amoebapore	3	1 down-regulated (amoebapore C)	No substantial changes
Transmembrane receptor kinases	>80	NA	6 up-regulated (TMKs 69, 53, 95, 105, 63, 56) 2 down-regulated (TMKs 03 and 17)
AIG-1 (similar to plant antibacterial proteins)	15	NA	5 up-regulated at day 1; 6 down-regulated at day 29 (all non-overlapping)

[a] Adapted from Weber *et al.* (2006); [b] Adapted from Gilchrist *et al.* (2006);
[c]Number of cysteine proteinase gene families in genome annotation at time studies were performed

Figure Legends

Figure 1.

Positions of introns in the vacuolar ATPase subunit D gene in *P. falciparum, D. discoideum,* and *E. histolytica*

Figure 2. **Comparison of protein sizes in *E. histolytica* and *D. discoideum.***

a: The graph shows the distribution of predicted amino acid length across sequenced genomes from single celled eukaryotes: *D. discoideum* (DD) *Encephalitozoon cuniculi* (EC), *P. falciparum* (PF), *E. histolytica* (EH), and *S. cerevisiae* (SC). *E. histolytica* and *E. cuniculi* have a distribution that is skewed toward smaller proteins relative to the other species.

b: The histogram displays the degree of size change of genes in *E. histolytica* relative to *D. discoideum* when comparing orthologous genes identified by reciprocal best blast hits. The black bars show genes that are smaller in *E. histolytica* where as the grey bars are smaller in *D. discoideum.*

Figure 3. **Domain diagram of the Hgl subunit of the Gal/GalNAc lectin**. **CW**-Cysteine-Tryptophan region; **CF**- Cysteine free region; **C-Rich**- Cysteine rich region. The black vertical box near the carboxy-terminus of the protein represents the single transmembrane domain. The horizontal black bars above the diagram indicate the location of a carbohydrate recognition domain (**CRD**), the region with similarity to the hepatic growth factor receptor, **c-Met**, and the region that has similarity to the **CD59**, the membrane inhibitor of the complement membrane attack complex. The numbers in parentheses indicate the location of these regions in the Hgl1 isoform (Mann *et al.*, 1991), where the methionine of the immature protein is residue 1.

Figure 4. **Structural domains of the 3 different types of family C1-like cysteine endopeptidases EhCP-A, EhCP-B and EhCP-C.** Shown are the location and length of domains specific for each the 3 types as well as the conserved active site and cysteine residue

Figure 5. **Predicted antioxidant system of *Entamoeba histolytica*.** A. Superoxide radical anions are detoxified by an iron-containing superoxide dismutase (FeSOD).

Molecular oxygen is reduced to hydrogen peroxide by a NADPH:flavin oxidoreductase (thioredoxin reductase, p34). Hydrogen peroxide is converted to water by rubrerythrin (Rbr). The nature of its redox partner is unknown. Hydrogen peroxide can also be converted to water via a classical thioredoxin redox system consisting of thioredoxin reductase (TrxR, p34), thioredoxin (Trx) and peroxiredoxin (Prx). B. Nitric oxide is reduced by an A-type flavoprotein (FprA) to nitrous oxide and water. For this reaction FprA receives electrons from NADH oxidase (Far).

Figure 6. **A phylogenetic tree of Rab proteins from *Entamoeba histolytica*, human, and yeast.** The number on the nodes represent the bootstrap proportions (%) of 1000 pseudo samples; only bootstrap proportions >30% are shown. *E. histolytica* Rab proteins are indicated in bold. Tentative subfamilies that revealed significant similarity (>40% identity) to their human or yeast counterpart are shaded dark, while *Entamoeba*-specific subfamilies have light shading. The scale bar indicates 0.1 substitutions at each amino acid position. *: *Eh* Rab proteins that lack the conserved effector region, switch regions, or GTP-binding boxes. **: *Eh* Rab proteins that possess a non-conventional carboxyl-terminus or lack carboxyl-terminal cysteines. ***: Rab proteins that were not classified as isotypes based on <40% identity to other members of the subfamily. References on tree: (1), Temesvari *et al.* (1999); (2), Rodríguez *et al.* (2000); (3), Saito-Nakano *et al.* (2001); (4), Juarez *et al.* (2001); (5), Saito-Nakano *et al.* (2004); and (6), Okada *et al.* (2005).

Figure 7. **Synthesis of N-glycan precursors by *S. cerevisiae* (A) and *E. histolytica* (B).** The N-glycan precursor of *S. cerevisiae* contains 14 sugars ($Glc_3Man_9GlcNAc_2$), each of which is added by a specific enzyme. The *E. histolytica* N-glycan precursor contains just seven sugars ($Man_5GlcNAc_2$), as the protist is missing enzymes that add mannose and glucose in the lumen of the ER. The figure is redrawn from Figure 1 of Samuelson *et al.* (2005). Glc = Glucose; GlcNAc = N-acetyl glucosamine; Man = Mannose.

Sialic acid

Figure 8. **Selected N-glycans of mammals (A-E) and *Entamoeba* (F-H).** Precursors transferred to nascent peptide (A and F). Glycosylated products involved in N-glycan-associated QC of protein folding (B and G). Mannosidase product involved in

N-glycan-associated protein degradation (mammals only) (C). Trimmed product that is building block for complex N-glycans (mammals and *Entamoeba*) (D). Complex N-glycans made in the Golgi (E and H). Glc = Glucose; GlcNAc = N-acetyl glucosamine; Man = Mannose; Gal = Galactose; Fuc = Fucose.

Figure 9. **Model of quality control of protein folding in *Entamoeba.*** 1. N-glycan-dependent QC of protein folding. 2. N-glycan-independent QC of protein folding. 3. N-glycan-independent ERAD. 4. Ire1 and unfolded protein response (see text for details).

Figure 10. **Structure of cysteine-rich plasma membrane proteins of *E. histolytica*.** These proteins include the various subunits of the Gal/GalNAc lectin, a cysteine protease, and numerous receptor kinases. Ire1, which is involved in the unfolded protein response, is also a receptor kianse but has no Cys-rich domain.

Figure 11. **Model for the *Entamoeba* cyst wall derived primarily from experiments with *E. invadens*.** A. The cyst wall consists of chitosan fibrils, which are made by chitin synthase and chitin deacetylase. Wall proteins include Jacob lectins with tandem arrays of 6-Cys chitin-binding domains (CBDs), as well as chitinase and Jessie lectins that have a single 8-Cys CBD. The Gal/GalNAc lectin in the plasma membrane binds sugars on the Jacob and Jessie lectins. B. Structures of representative lectins illustrated in A.

Figure 12. **Phylogenetic relationships of *E. histolytica* glutamine synthase**. The gene encoding glutamine synthase (EC 6.3.1.2) is now shared by *E. histolytica* and the diatom *Thalassiosira*. This gene is mainly restricted to prokaryotic genomes (eukaryotes are highlighted by arrows). *T. vaginalis* also contains a homologue but in this case it clusters weakly with *Fusobacterium*. The scale bar represents 10% of inferred sequence divergence. Both the GenBank and RefSeq accession numbers are given for the *E. histolytica* entry.

Figure 13. **Phylogenetic relationships of *E. histolytica* tryptophanase.** This tree suggests that the *E. histolytica* gene encoding a tryptophanase was acquired by LGT from a relative of the anaerobic bacterium *Fusobacterium*. In contrast, the *T.*

vaginalis gene appears to have a separate origin with a LGT from a relative of the anaerobic *Bacteroides* group. The scale bar represents 10% of inferred sequence divergence. Both the GenBank and RefSeq accession numbers are given for the *E. histolytica* entry. The EC number is also shown.

Figure 14. **Pie chart of functional categories for the 68 strongest LGT cases.** The cases are those discussed in the text and listed in Table 8. Most entries encode metabolic enzymes (KEGG annotation).

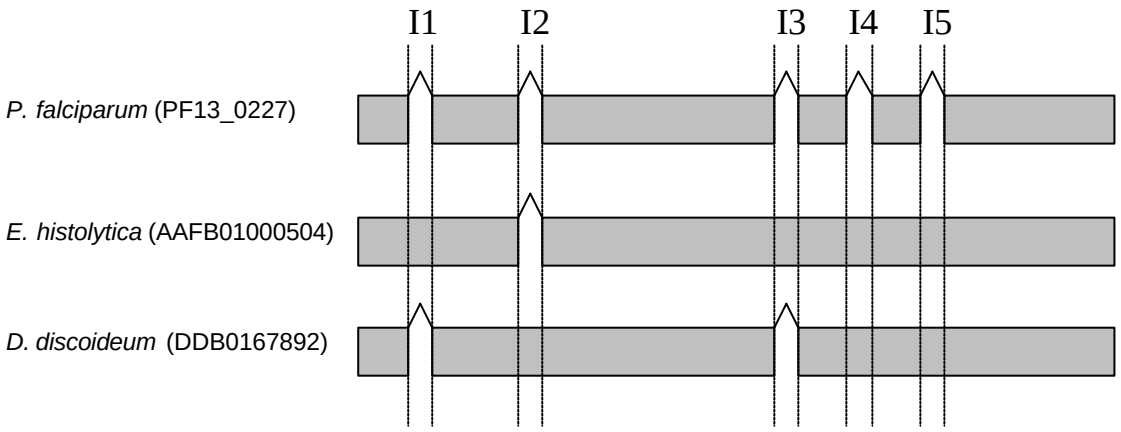
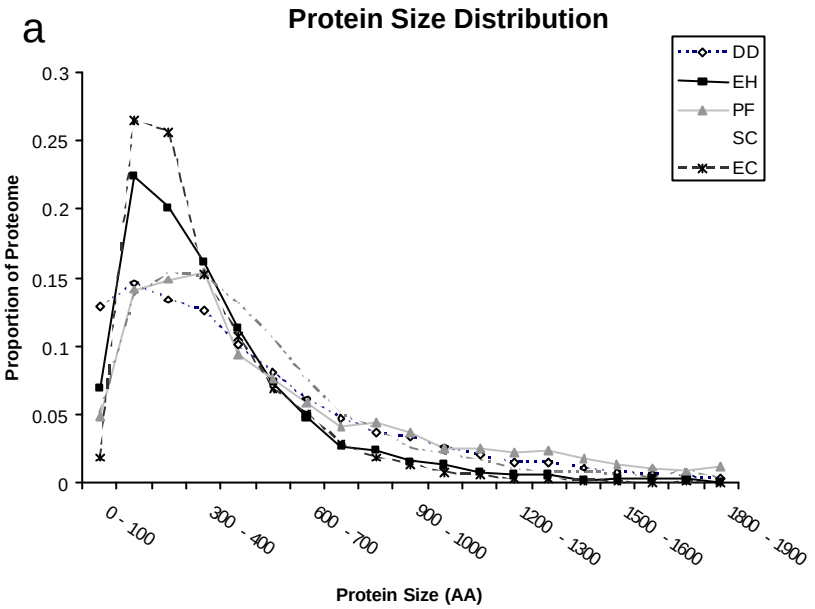

I1
I2
I3
I4
I5
P. falciparum (PF13_0227)
E. histolytica (AAFB01000504)
D. discoideum (DDB0167892)

Fig. 1


a
Protein Size Distribution
DD
EH
PF
SC
EC
Proportion of Proteome
0.3
0.25
0.2
0.15
0.1
0.05
0
0 - 100
300 - 400
600 - 700
900 - 1000
1200 - 1300
1500 - 1600
1800 - 1900
Protein Size (AA)

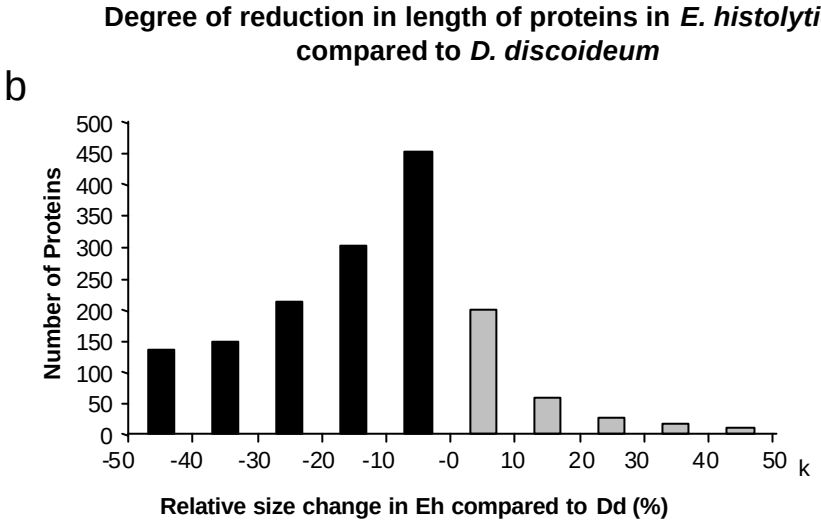

Degree of reduction in length of proteins in E. histolytica compared to D. discoideum
b
Number of Proteins
500
450
400
350
300
250
200
150
100
50
0
-50
-40
-30
-20
-10
-0
10
20
30
40
50
k
Relative size change in Eh compared to Dd (%)

Fig. 2

1
2

1

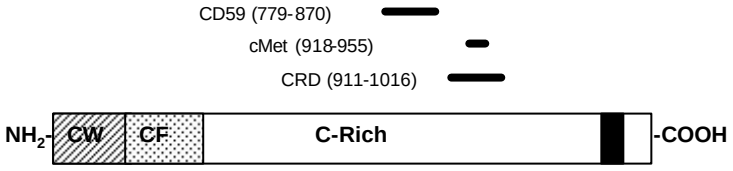


Fig. 3

2

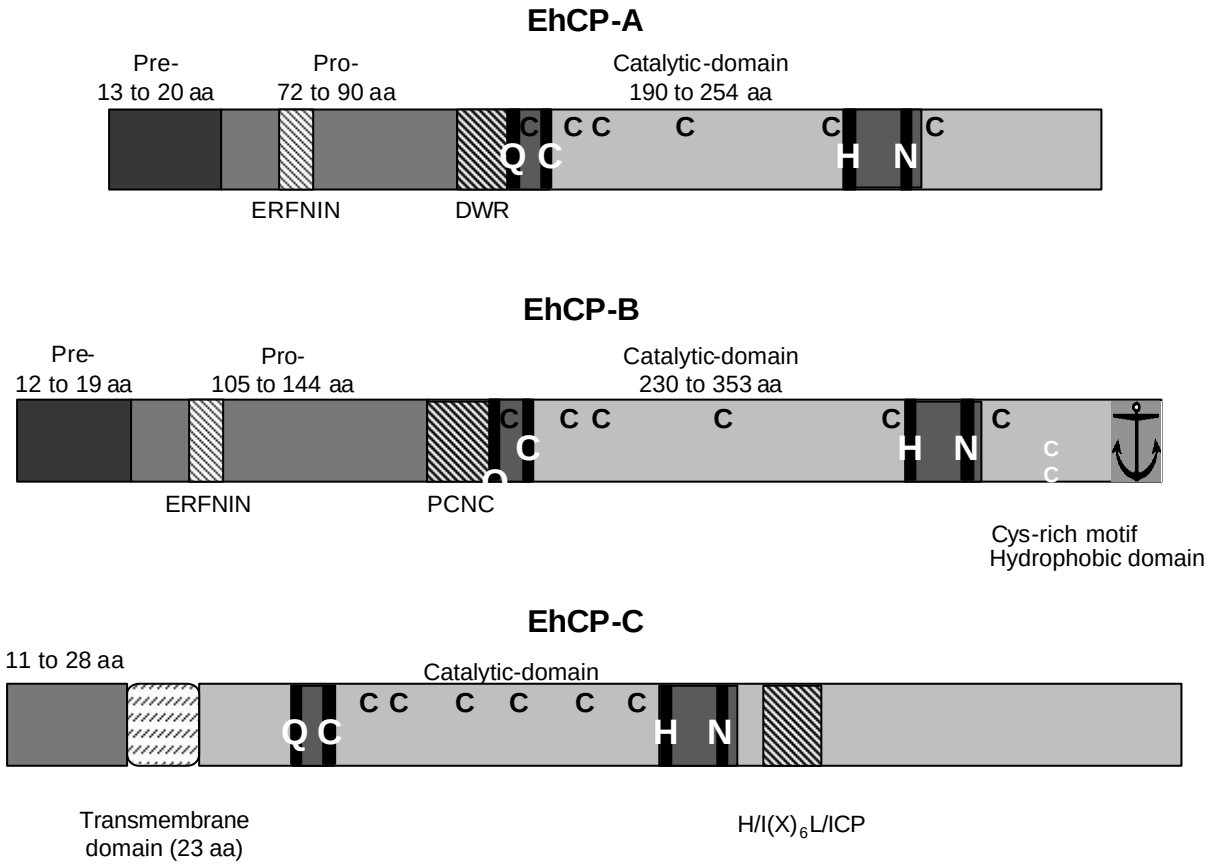

EhCP-A
Pre-
13 to 20 aa
Pro-
72 to 90 aa
Catalytic-domain
190 to 254 aa
ERFNIN
DWR
EhCP-B
Pre-
12 to 19 aa
Pro-
105 to 144 aa
Catalytic-domain
230 to 353 aa
ERFNIN
PCNC
Cys-rich motif
Hydrophobic domain
EhCP-C
11 to 28 aa
Catalytic-domain
Transmembrane
domain (23 aa)
H/I(X)6L/ICP

Fig. 4

1

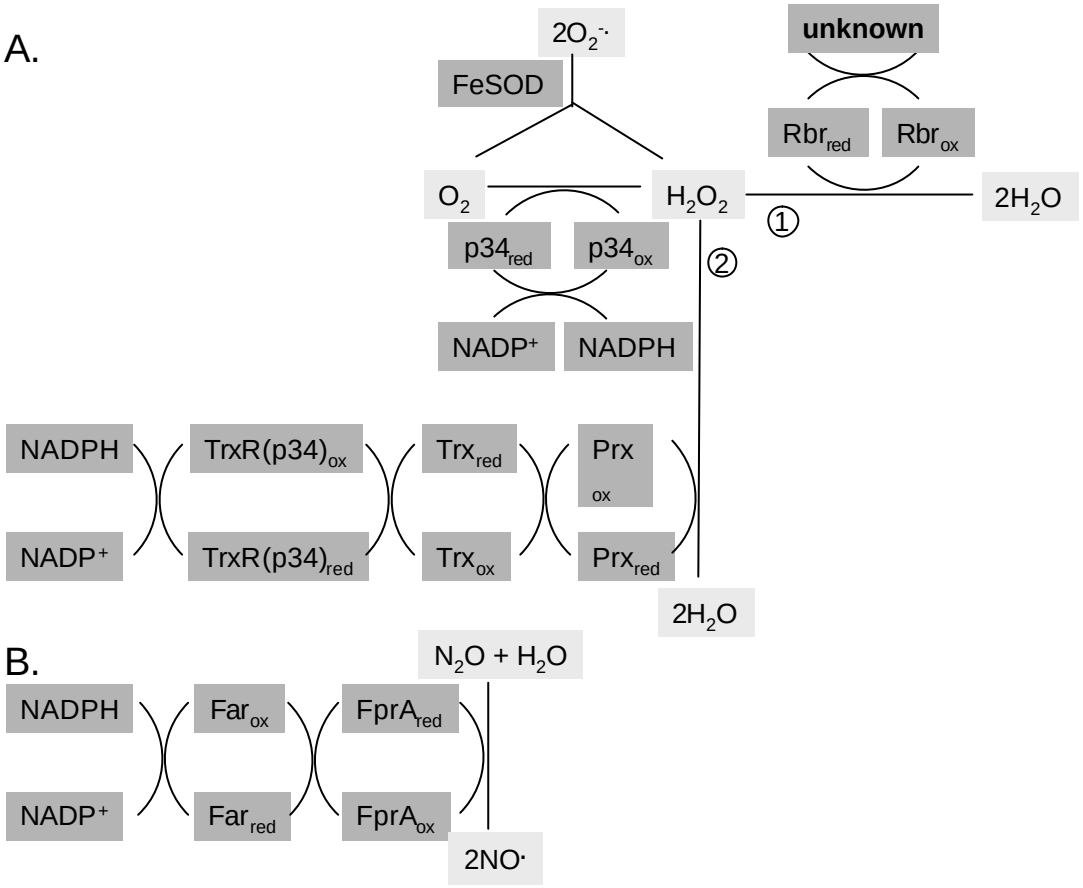

A.
$2O_2^{-\cdot}$
FeSOD
unknown
Rbr_{red}
Rbr_{ox}
O_2
H_2O_2
$2H_2O$
①
$p34_{red}$
$p34_{ox}$
②
$NADP^+$
NADPH
NADPH
$TrxR(p34)_{ox}$
Trx_{red}
Prx_{ox}
$NADP^+$
$TrxR(p34)_{red}$
Trx_{ox}
Prx_{red}
$2H_2O$
B.
$N_2O + H_2O$
NADPH
Far_{ox}
$FprA_{red}$
$NADP^+$
Far_{red}
$FprA_{ox}$
$2NO\cdot$

Fig. 5

1

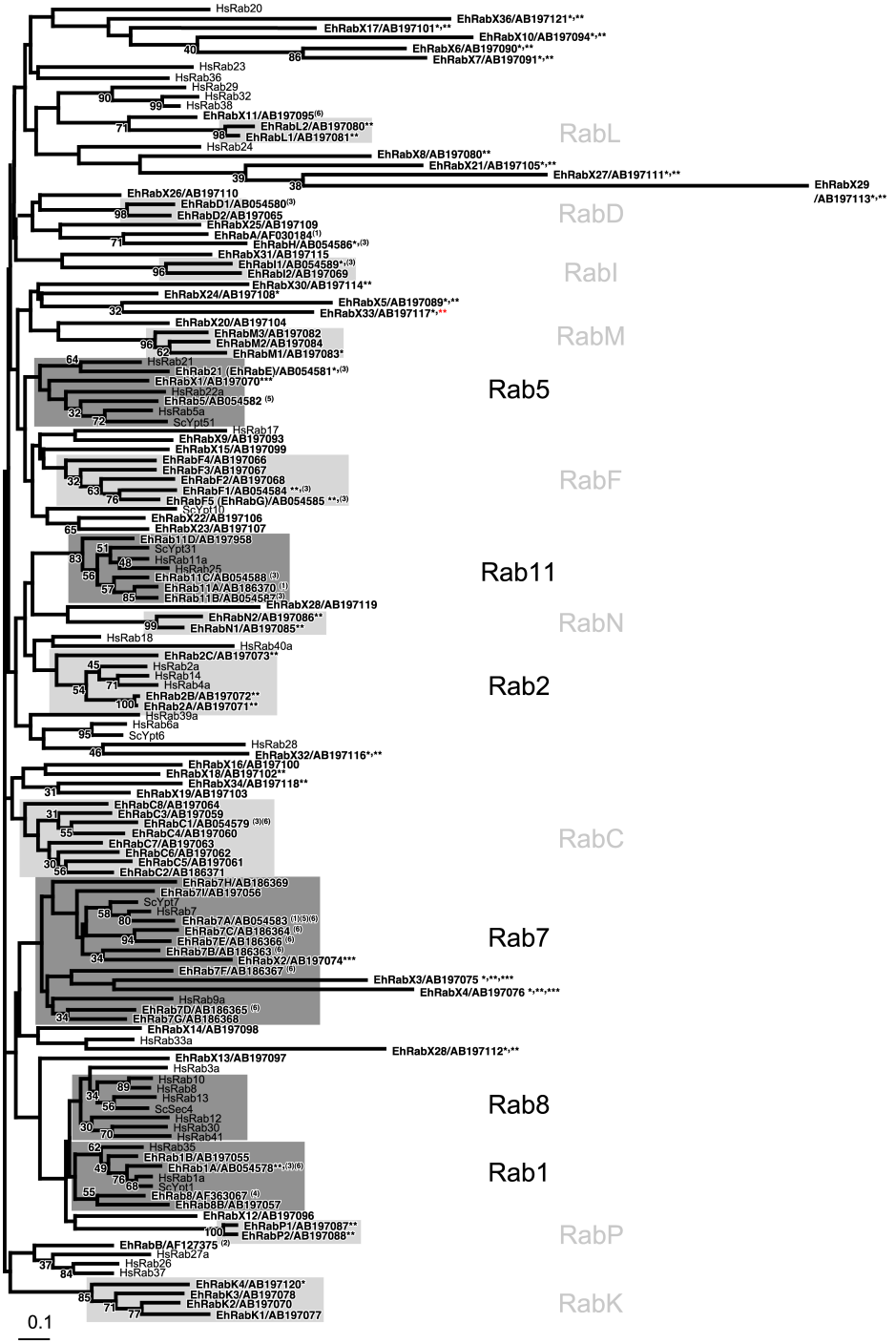
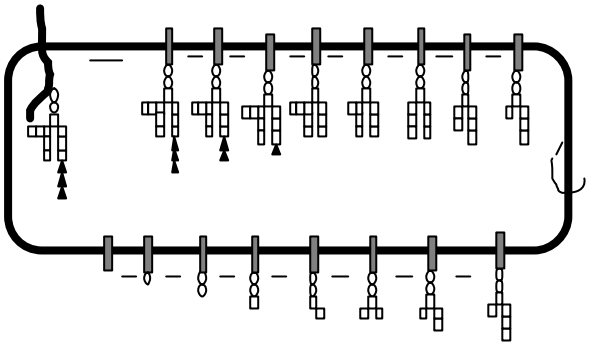
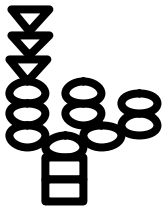

HsRab20
EhRabX36/AB197121*,**
EhRabX17/AB197101*,**
EhRabX10/AB197094*,**
EhRabX6/AB197090*,**
EhRabX7/AB197091*,**
HsRab23
HsRab36
HsRab29
HsRab32
HsRab38
EhRabX11/AB197095(6)
EhRabL2/AB197080**
EhRabL1/AB197081**
RabL
HsRab24
EhRabX8/AB197080**
EhRabX21/AB197105*,**
EhRabX27/AB197111*,**
EhRabX29 /AB197113*,**
EhRabX26/AB197110
EhRabD1/AB054580(3)
EhRabD2/AB197065
RabD
EhRabX25/AB197109
EhRabA/AF030184(1)
EhRabH/AB054586*,(3)
EhRabX31/AB197115
EhRabI1/AB054589*,(3)
EhRabI2/AB197069
RabI
EhRabX30/AB197114**
EhRabX24/AB197108*
EhRabX5/AB197089*,**
EhRabX33/AB197117*,**
EhRabX20/AB197104
EhRabM3/AB197082
EhRabM2/AB197084
EhRabM1/AB197083*
RabM
HsRab21
EhRab21 (EhRabE)/AB054581*,(3)
EhRabX1/AB197070***
HsRab22a
EhRab5/AB054582 (5)
HsRab5a
ScYpt51
Rab5
HsRab17
EhRabX9/AB197093
EhRabX15/AB197099
EhRabF4/AB197066
EhRabF3/AB197067
EhRabF2/AB197068
EhRabF1/AB054584 **,(3)
EhRabF5 (EhRabG)/AB054585 **,(3)
ScYpt10
RabF
EhRabX22/AB197106
EhRabX23/AB197107
EhRab11D/AB197958
ScYpt31
HsRab11a
HsRab25
EhRab11C/AB054588 (3)
EhRab11A/AB186370 (1)
EhRab11B/AB054587 (3)
Rab11
EhRabX28/AB197119
EhRabN2/AB197086**
EhRabN1/AB197085**
RabN
HsRab18
HsRab40a
EhRab2C/AB197073**
HsRab2a
HsRab14
HsRab4a
EhRab2B/AB197072**
EhRab2A/AB197071**
Rab2
HsRab39a
HsRab6a
ScYpt6
HsRab28
EhRabX32/AB197116*,**
EhRabX16/AB197100
EhRabX18/AB197102**
EhRabX34/AB197118**
EhRabX19/AB197103
EhRabC8/AB197064
EhRabC3/AB197059
EhRabC1/AB054579 (3)(6)
EhRabC4/AB197060
EhRabC7/AB197063
EhRabC6/AB197062
EhRabC5/AB197061
EhRabC2/AB186371
RabC
EhRab7H/AB186369
EhRab7I/AB197056
ScYpt7
HsRab7
EhRab7A/AB054583 (1)(5)(6)
EhRab7C/AB186364 (5)
EhRab7E/AB186366 (6)
EhRab7B/AB186363 (6)
EhRabX2/AB197074***
EhRab7F/AB186367 (6)
EhRabX3/AB197075 *,**,***
EhRabX4/AB197076 *,**,***
HsRab9a
EhRab7D/AB186365 (6)
EhRab7G/AB186368
Rab7
EhRabX14/AB197098
HsRab33a
EhRabX28/AB197112*,**
EhRabX13/AB197097
HsRab3a
HsRab10
HsRab6
HsRab13
ScSec4
HsRab12
HsRab30
HsRab41
Rab8
HsRab35
EhRab1B/AB197055
EhRab1A/AB054578***,(3)(6)
HsRab1a
ScYpt1
EhRab8/AF363067 (4)
EhRab8B/AB197057
Rab1
EhRabX12/AB197096
EhRabP1/AB197087**
EhRabP2/AB197088**
RabP
EhRabB/AF127375 (2)
HsRab27a
HsRab26
HsRab37
EhRabK4/AB197120*
EhRabK3/AB197078
EhRabK2/AB197070
EhRabK1/AB197077
RabK
0.1

Fig. 6

1
2

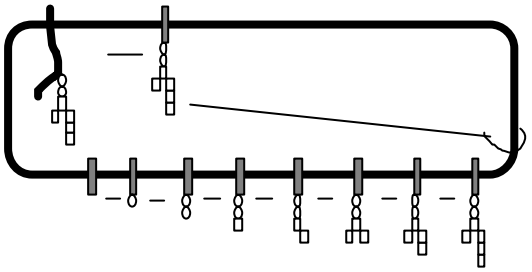
A. *Saccharomyces*:





Glc_3Man_9GlcN

B. *Entamoeba*:



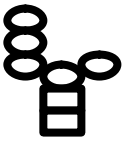


Fig. 7

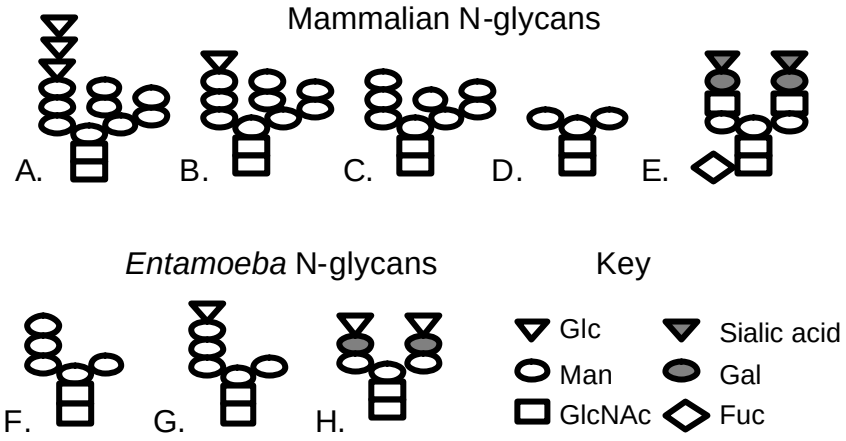

Mammalian N-glycans
A.
B.
C.
D.
E.
Entamoeba N-glycans
F.
G.
H.
Key
Glc
Man
GlcNAc
Sialic acid
Gal
Fuc

Fig. 8

1
2

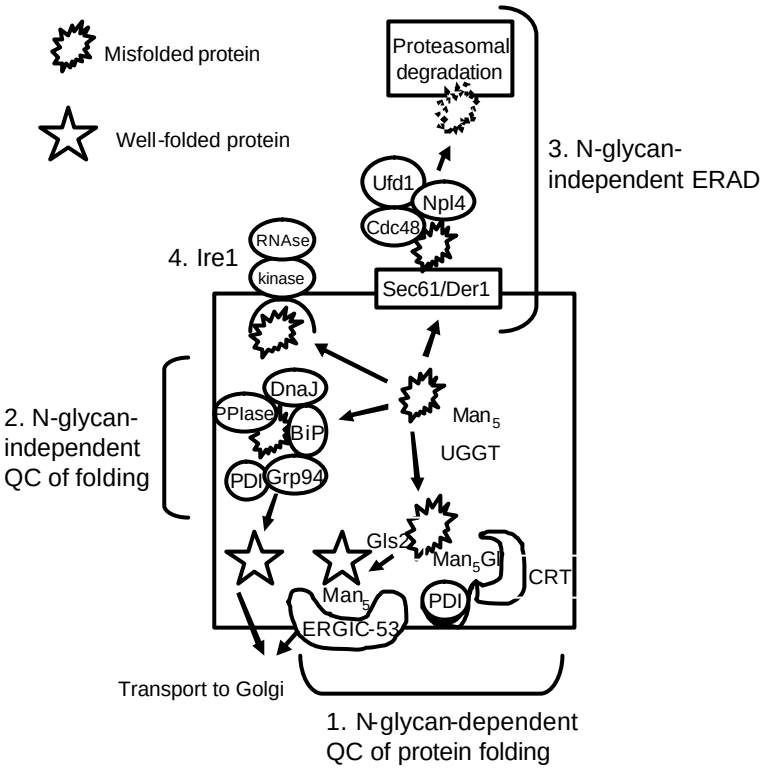

Misfolded protein
Well-folded protein
Proteasomal degradation
3. N-glycan-independent ERAD
Ufd1
Npl4
Cdc48
4. Ire1
RNAse
kinase
Sec61/Der1
2. N-glycan-independent QC of folding
DnaJ
PPIase
BiP
PDI
Grp94
Man5
UGGT
Gls2
Man5Gl
CRT
Man5
PDI
ERGIC-53
Transport to Golgi
1. N-glycan-dependent QC of protein folding

Fig. 9

1

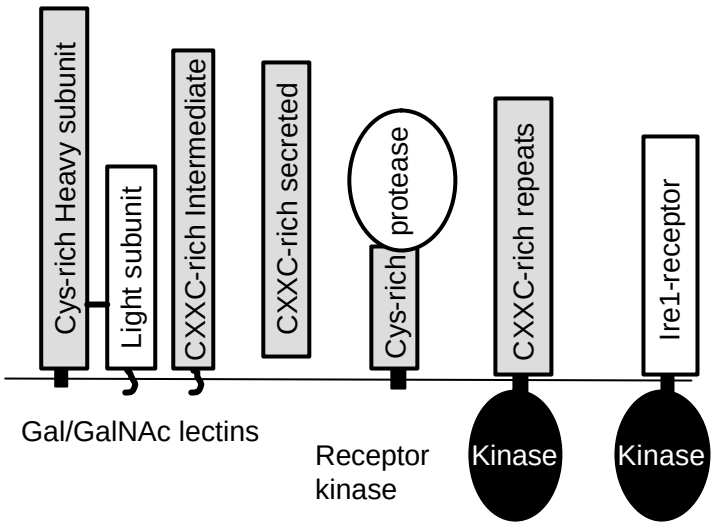

Cys-rich Heavy subunit
Light subunit
CXXC-rich Intermediate
CXXC-rich secreted
Cys-rich protease
CXXC-rich repeats
Ire1-receptor
Gal/GalNAc lectins
Receptor kinase
Kinase
Kinase

Fig. 10

1

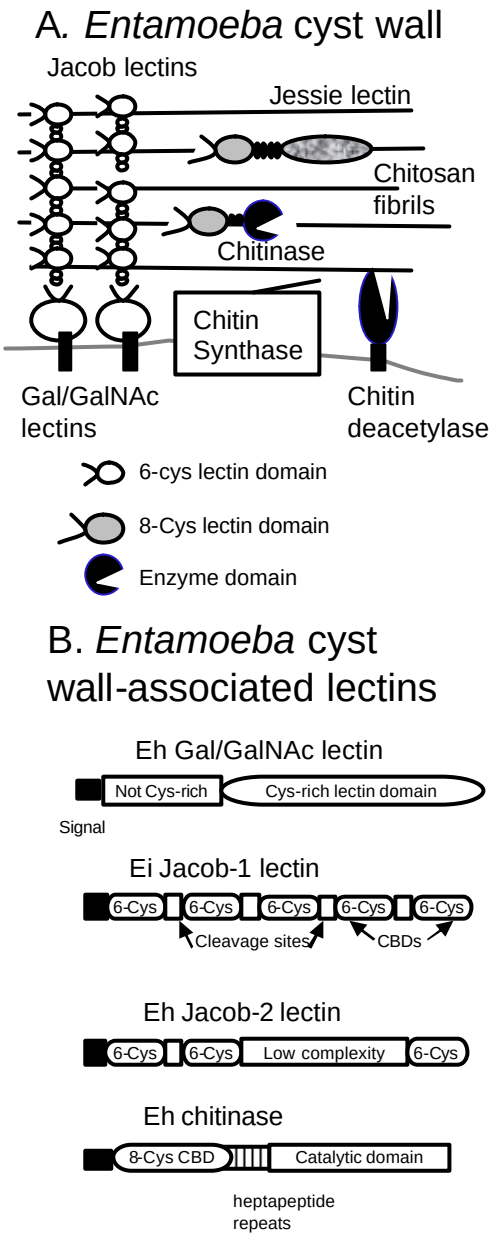

A. Entamoeba cyst wall
Jacob lectins
Jessie lectin
Chitosan fibrils
Chitinase
Chitin Synthase
Gal/GalNAc lectins
Chitin deacetylase
6-cys lectin domain
8-Cys lectin domain
Enzyme domain
B. Entamoeba cyst wall-associated lectins
Eh Gal/GalNAc lectin
Not Cys-rich
Cys-rich lectin domain
Signal
Ei Jacob-1 lectin
6-Cys
Cleavage sites
CBDs
Eh Jacob-2 lectin
Low complexity
Eh chitinase
8-Cys CBD
Catalytic domain
heptapeptide repeats

Fig. 11

1
2

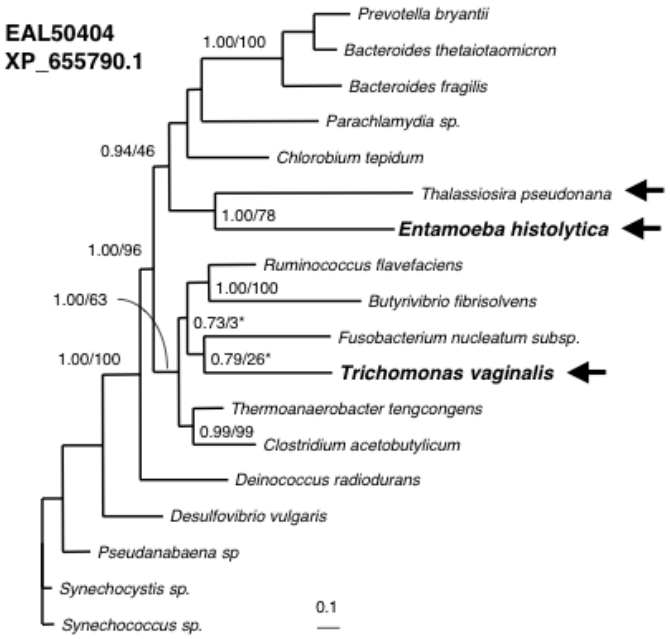

EAL50404
XP_655790.1
Prevotella bryantii
1.00/100
Bacteroides thetaiotaomicron
Bacteroides fragilis
Parachlamydia sp.
0.94/46
Chlorobium tepidum
Thalassiosira pseudonana
1.00/78
Entamoeba histolytica
1.00/96
Ruminococcus flavefaciens
1.00/100
1.00/63
Butyrivibrio fibrisolvens
0.73/3*
Fusobacterium nucleatum subsp.
1.00/100
0.79/26*
Trichomonas vaginalis
Thermoanaerobacter tengcongens
0.99/99
Clostridium acetobutylicum
Deinococcus radiodurans
Desulfovibrio vulgaris
Pseudanabaena sp
Synechocystis sp.
Synechococcus sp.
0.1

Fig. 12

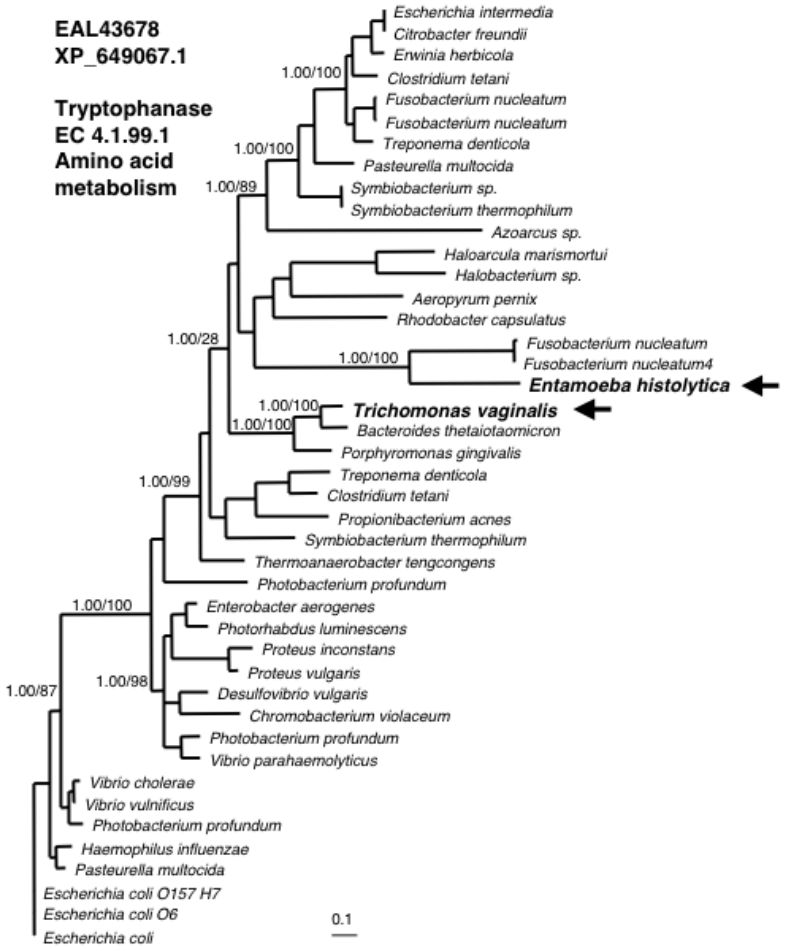

EAL43678
XP_649067.1
Tryptophanase
EC 4.1.99.1
Amino acid
metabolism
Escherichia intermedia
Citrobacter freundii
Erwinia herbicola
Clostridium tetani
Fusobacterium nucleatum
Fusobacterium nucleatum
Treponema denticola
Pasteurella multocida
Symbiobacterium sp.
Symbiobacterium thermophilum
Azoarcus sp.
Haloarcula marismortui
Halobacterium sp.
Aeropyrum pernix
Rhodobacter capsulatus
Fusobacterium nucleatum
Fusobacterium nucleatum4
Entamoeba histolytica
Trichomonas vaginalis
Bacteroides thetaiotaomicron
Porphyromonas gingivalis
Treponema denticola
Clostridium tetani
Propionibacterium acnes
Symbiobacterium thermophilum
Thermoanaerobacter tengcongens
Photobacterium profundum
Enterobacter aerogenes
Photorhabdus luminescens
Proteus inconstans
Proteus vulgaris
Desulfovibrio vulgaris
Chromobacterium violaceum
Photobacterium profundum
Vibrio parahaemolyticus
Vibrio cholerae
Vibrio vulnificus
Photobacterium profundum
Haemophilus influenzae
Pasteurella multocida
Escherichia coli O157 H7
Escherichia coli O6
Escherichia coli
1.00/100
1.00/100
1.00/89
1.00/28
1.00/100
1.00/100
1.00/100
1.00/99
1.00/100
1.00/98
1.00/87
0.1

Fig. 13

1
2

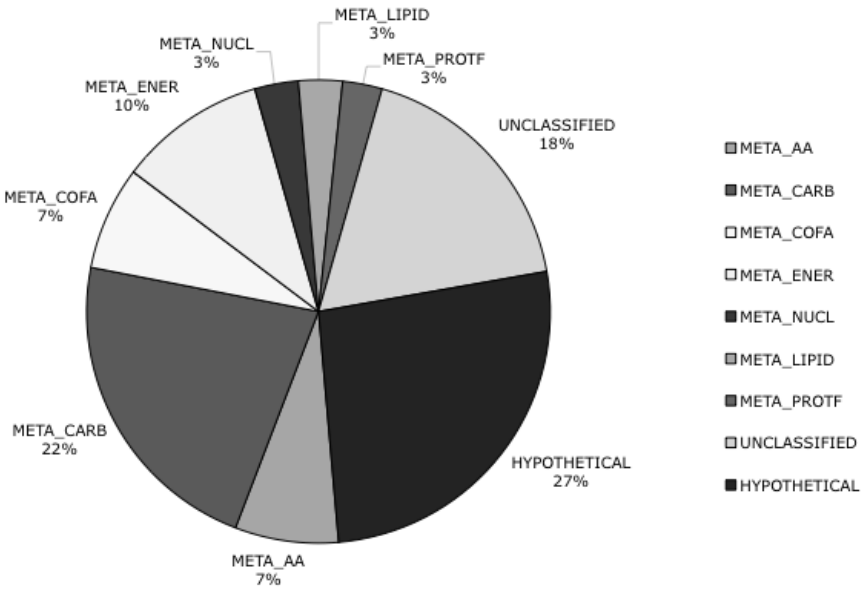

META_LIPID
3%
META_NUCL
3%
META_PROTF
3%
META_ENER
10%
UNCLASSIFIED
18%
META_COFA
7%
META_CARB
22%
HYPOTHETICAL
27%
META_AA
7%
META_AA
META_CARB
META_COFA
META_ENER
META_NUCL
META_LIPID
META_PROTF
UNCLASSIFIED
HYPOTHETICAL

Fig. 14

1