

- 1 Todd, J., Post-Beittenmiller, D., and Jaworski, J. G. (1999). KCS1 encodes a  
2 fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis  
3 in *Arabidopsis thaliana*. *Plant Journal* **17**, 119-30.
- 4 Tokoro, M., Asai, T., Kobayashi, S., Takeuchi, T., and Nozaki, T. (2003).  
5 Identification and characterization of two isoenzymes of methionine S-  
6 lyase from *Entamoeba histolytica*: a key enzyme of sulfur-amino acid  
7 degradation in an anaerobic parasitic protist that lacks forward and  
8 reverse trans-sulfuration pathways. *Journal of Biological Chemistry*  
9 **278**, 42717 - 42727.
- 10 Toonen, R. F., and Verhage, M. (2003). Vesicle trafficking: pleasure and pain  
11 from SM genes. *Trends in Cell Biology* **13**, 177-86.
- 12 Torii, S., Takeuchi, T., Nagamatsu, S., and Izumi, T. (2004). Rab27 effector  
13 granophilin promotes the plasma membrane targeting of insulin  
14 granules via interaction with syntaxin 1a. *Journal of Biological*  
15 *Chemistry* **279**, 22532-8.
- 16 Touz, M. C., Nores, M. J., Slavin, I., Carmona, C., Conrad, J. T., Mowatt, M.  
17 R., Nash, T. E., Coronel, C. E., and Lujan, H. D. (2002). The activity of  
18 a developmentally regulated cysteine proteinase is required for cyst wall  
19 formation in the primitive eukaryote *Giardia lamblia*. *Journal of*  
20 *Biological Chemistry* **277**, 8474-81.
- 21 Tovar, J., Fischer, A., and Clark, C. G. (1999). The mitosome, a novel organelle  
22 related to mitochondria in the amitochondriate parasite *Entamoeba*  
23 *histolytica*. *Molecular Microbiology* **32**, 1013-1021.
- 24 Trilla, J. A., Duran, A., and Roncero, C. (1999). Chs7p, a new protein involved  
25 in the control of protein export from the endoplasmic reticulum that is  
26 specifically engaged in the regulation of chitin synthesis in  
27 *Saccharomyces cerevisiae*. *Journal of Cell Biology* **145**, 1153-63.
- 28 Trombetta, E. S., and Parodi, A. J. (2003). Quality control and protein folding  
29 in the secretory pathway. *Annu Rev Cell Dev Biol* **19**, 649-76.
- 30 Tye, B. K. (1999). MCM proteins in DNA replication. *Annual Review of*  
31 *Biochemistry* **68**, 649-86.

- 1 Uemura, T., Ueda, T., Ohniwa, R. L., Nakano, A., Takeyasu, K., and Sato, M.  
2 H. (2004). Systematic analysis of SNARE molecules in *Arabidopsis*:  
3 dissection of the post-Golgi network in plant cells. *Cell Structure and*  
4 *Function* **29**, 49-65.
- 5 Uwanogho, D. A., Hardcastle, Z., Balogh, P., Mirza, G., Thornburg, K. L.,  
6 Ragoussis, J., and Sharpe, P. T. (1999). Molecular cloning,  
7 chromosomal mapping, and developmental expression of a novel  
8 protein tyrosine phosphatase-like gene. *Genomics* **62**, 406-16.
- 9 Van Dellen, K., Field, J., Wang, Z., Loftus, B., and Samuelson, J. (2002a).  
10 LINEs and SINE-like elements of the protist *Entamoeba histolytica*.  
11 *Gene* **297**, 229-39.
- 12 Van Dellen, K., Ghosh, S. K., Robbins, P. W., Loftus, B., and Samuelson, J.  
13 (2002b). *Entamoeba histolytica* lectins contain unique 6-Cys or 8-Cys  
14 chitin-binding domains. *Infection and Immunity* **70**, 3259-63.
- 15 Van Dellen, K. L., Bulik, D. A., Specht, C. A., Robbins, P. W., and Samuelson,  
16 J. C. (2006a). Heterologous expression of an *Entamoeba histolytica*  
17 chitin synthase in *Saccharomyces cerevisiae*. *Eukaryotic Cell* **5**, 203-6.
- 18 Van Dellen, K. L., Chatterjee, A., Ratner, D. M., Magnelli, P. E., Cipollo, J. F.,  
19 Steffen, M., Robbins, P. W., and Samuelson, J. (2006b). Unique  
20 posttranslational modifications of chitin-binding lectins of *Entamoeba*  
21 *invadens* cyst walls *Eukaryotic Cell* **5**, 836-848.
- 22 van der Giezen, M., Cox, S., and Tovar, J. (2004). The iron-sulfur cluster  
23 assembly genes *iscS* and *iscU* of *Entamoeba histolytica* were acquired  
24 by horizontal gene transfer. *BMC Evolutionary Biology* **4**, 7.
- 25 Van Mullem, V., Wery, M., Werner, M., Vandenhoute, J., and Thuriaux, P.  
26 (2002). The Rpb9 subunit of RNA polymerase II binds transcription  
27 factor TFIIE and interferes with the SAGA and elongator histone  
28 acetyltransferases. *Journal of Biological Chemistry* **277**, 10220-5.
- 29 Vanacová, S., Liston, D. R., Tachezy, J., and Johnson, P. J. (2003). Molecular  
30 biology of the amitochondriate parasites, *Giardia intestinalis*,

- 1           *Entamoeba histolytica* and *Trichomonas vaginalis*. *International*  
2           *Journal for Parasitology* **33**, 235-55.
- 3 Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S., and Inze, D.  
4           (2002). Genome-wide analysis of core cell cycle genes in *Arabidopsis*.  
5           *Plant Cell* **14**, 903-16.
- 6 Varela-Gomez, M., Moreno-Sanchez, R., Pardo, J. P., and Perez-Montfort, R.  
7           (2004). Kinetic mechanism and metabolic role of pyruvate phosphate  
8           dikinase from *Entamoeba histolytica*. *Journal of Biological Chemistry*  
9           **279**, 54124-30.
- 10 Vargas-Villarreal, J., Olvera-Rodriguez, A., Mata-Cardenas, B. D., Martinez-  
11           Rogriguez, H. G., Said-Fernandez, S., and Alagon-Cano, A. (1998).  
12           Isolation of an *Entamoeba histolytica* intracellular alkaline  
13           phospholipase A2. *Parasitology Research* **84**, 310-4.
- 14 Vargas, M., Sansonetti, P., and Guillen, N. (1996). Identification and cellular  
15           localization of the actin-binding protein ABP-120 from *Entamoeba*  
16           *histolytica*. *Molecular Microbiology* **22**, 849-57.
- 17 Vartiainen, M. K., and Machesky, L. M. (2004). The WASP -Arp2/3 pathway:  
18           genetic insights. *Current Opinion in Cell Biology* **16**, 174-81.
- 19 Vats, D., Vishwakarma, R. A., Bhattacharya, S., and Bhattacharya, A. (2005).  
20           Reduction of cell surface glycosylphosphatidylinositol conjugates in  
21           *Entamoeba histolytica* by antisense blocking of *E. histolytica* GlcNAc-  
22           phosphatidylinositol deacetylase expression: effect on cell proliferation,  
23           endocytosis, and adhesion to target cells. *Infection and Immunity* **73**,  
24           8381-92.
- 25 Vayssie, L., Vargas, M., Weber, C., and Guillen, N. (2004). Double-stranded  
26           RNA mediates homology-dependent gene silencing of gamma-tubulin  
27           in the human parasite *Entamoeba histolytica*. *Molecular and*  
28           *Biochemical Parasitology* **138**, 21-8
- 29 Villalobo, E., Wender, N., and Mirelman, D. (2005). *Entamoeba histolytica*:  
30           molecular characterization of an aldose 1-epimerase (mutarotase).  
31           *Experimental Parasitology* **110**, 298-302.

- 1 Vishwakarma, R. A., Anand, M. T., Arya, R., Vats, D., and Bhattacharya, A.  
2 (2006). Glycosylated inositol phospholipid from *Entamoeba histolytica*:  
3 identification and structural characterization. *Molecular and*  
4 *Biochemical Parasitology* **145**, 121-4.
- 5 Volpicelli-Daley, L. A., Li, Y., Zhang, C. J., and Kahn, R. A. (2005). Isoform-  
6 selective effects of the depletion of ADP-ribosylation factors 1-5 on  
7 membrane traffic. *Molecular Biology of the Cell* **16**, 4495-508.
- 8 Vowels, J. J., and Payne, G. S. (1998). A dileucine-like sorting signal directs  
9 transport into an AP-3-dependent, clathrin-independent pathway to the  
10 yeast vacuole. *EMBO Journal* **17**, 2482-93.
- 11 Walenta, J. H., Didier, A. J., Liu, X., and Kramer, H. (2001). The Golgi-  
12 associated hook3 protein is a member of a novel family of microtubule-  
13 binding proteins. *Journal of Cell Biology* **152**, 923-34.
- 14 Wang, C. C., and Cheng, H. W. (1984). The deoxyribonucleoside  
15 phosphotransferase of *Trichomonas vaginalis*. A potential target for  
16 anti-trichomonal chemotherapy. *Journal of Experimental Medicine*  
17 **160**, 987-1000.
- 18 Wang, Z., Samuelson, J., Clark, C. G., Eichinger, D., Paul, J., Van Dellen, K.,  
19 Hall, N., Anderson, I., and Loftus, B. (2003). Gene discovery in the  
20 *Entamoeba invadens* genome. *Molecular and Biochemical Parasitology*  
21 **129**, 23-31.
- 22 Wassmann, C., Hellberg, A., Tannich, E., and Bruchhaus, I. (1999).  
23 Metronidazole resistance in the protozoan parasite *Entamoeba*  
24 *histolytica* is associated with increased expression of iron-containing  
25 superoxide dismutase and peroxiredoxin and decreased expression of  
26 ferredoxin 1 and flavin reductase. *Journal of Biological Chemistry* **274**,  
27 26051-6.
- 28 Watson, W. H., Yang, X., Choi, Y. E., Jones, D. P., and Kehrer, J. P. (2004).  
29 Thioredoxin and its role in toxicology. *Toxicological Sciences* **78**, 3-14.
- 30 Weber, C., Guigon, G., Bouchier, C., Frangeul, L., Moreira, S., Sismeiro, O.,  
31 Gouyette, C., Mirelman, D., Coppee, J. Y., and Guillen, N. (2006).

- 1           Stress by heat shock induces massive down regulation of genes and  
2           allows differential allelic expression of the Gal/GalNAc lectin in  
3           *Entamoeba histolytica*. *Eukaryotic Cell* **5**, 871-5.
- 4   Wegmann, D., Hess, P., Baier, C., Wieland, F. T., and Reinhard, C. (2004).  
5           Novel isotypic gamma/zeta subunits reveal three coatomer complexes in  
6           mammals. *Molecular and Cellular Biology* **24**, 1070-80.
- 7   Weinbach, E. C., and Diamond, L. S. (1974). *Entamoeba histolytica*. I. Aerobic  
8           metabolism. *Experimental Parasitology* **35**, 232-43.
- 9   Weinberg, M. V., Jenney, F. E., Jr., Cui, X., and Adams, M. W. (2004).  
10           Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus*  
11           is a rubredoxin-dependent, iron-containing peroxidase. *Journal of*  
12           *Bacteriology* **186**, 7888-95.
- 13   Wennerberg, K., Rossman, K. L., and Der, C. J. (2005). The Ras superfamily at  
14           a glance. *Journal of Cell Science* **118**, 843-6.
- 15   Wera, S., and Hemmings, B. A. (1995). Serine/threonine protein phosphatases.  
16           *Biochemical Journal* **311**, 17-29.
- 17   Werries, E., and Thurn, I. (1989). Breakdown of glucopolysaccharides in  
18           *Entamoeba histolytica* by phosphorylase. *Journal of Protozoology* **36**,  
19           607-12.
- 20   Werries, E., Franz, A., and Geisemeyer, S. (1990). Detection of glycogen  
21           debranching system in trophozoites of *Entamoeba histolytica*. *Journal*  
22           *of Protozoology* **37**, 576-80.
- 23   West, C. M. (2003). Comparative analysis of spore coat formation, structure,  
24           and function in *Dictyostelium*. *International Review of Cytology* **222**,  
25           237-93.
- 26   West, C. M., van der Wel, H., Coutinho, P. M., and Henrissat, B. (2005).  
27           Glycosyltransferase genomics in *Dictyostelium*, pp. 235-264. In W. F.  
28           Loomis, and A. Kuspa (Eds). *Dictyostelium Genomics*, Horizon  
29           Bioscience, Norfolk, UK.

- 1 Whitney, J. A., Gomez, M., Sheff, D., Kreis, T. E., and Mellman, I. (1995).  
2 Cytoplasmic coat proteins involved in endosome function. *Cell* **83**, 703-  
3 13.
- 4 WHO (1998). *The world health report 1998. Life in the 21st century: a vision*  
5 *for all*. World Health Organization. Geneva, Switzerland.
- 6 Willhoeft, U., and Tannich, E. (1999). The electrophoretic karyotype of  
7 *Entamoeba histolytica*. *Molecular and Biochemical Parasitology* **99**,  
8 41-53.
- 9 Willhoeft, U., Buss, H., and Tannich, E. (1999). Analysis of cDNA expressed  
10 sequence tags from *Entamoeba histolytica*: identification of two highly  
11 abundant polyadenylated transcripts with no overt open reading frames.  
12 *Protist* **150**, 61-70.
- 13 Willhoeft, U., Buss, H., and Tannich, E. (2002). The abundant polyadenylated  
14 transcript 2 DNA sequence of the pathogenic protozoan parasite  
15 *Entamoeba histolytica* represents a nonautonomous non-long-terminal-  
16 repeat retrotransposon-like element which is absent in the closely  
17 related nonpathogenic species *Entamoeba dispar*. *Infection and*  
18 *Immunity* **70**, 6798-804.
- 19 Winder, S. J., and Ayscough, K. R. (2005). Actin-binding proteins. *Journal of*  
20 *Cell Science* **118**, 651-4.
- 21 Winkelmann, J., Leippe, M., and Bruhn, H. (2006). A novel saposin-like  
22 protein of *Entamoeba histolytica* with membrane-fusogenic activity.  
23 *Molecular and Biochemical Parasitology* **147**, 85-94.
- 24 Wittenberg, C., Sugimoto, K., and Reed, S. I. (1990). G1-specific cyclins of *S.*  
25 *cerevisiae*: cell cycle periodicity, regulation by mating pheromone, and  
26 association with the p34CDC28 protein kinase. *Cell* **62**, 225-37.
- 27 Wohlschlegel, J. A., Dhar, S. K., Prokhorova, T. A., Dutta, A., and Walter, J. C.  
28 (2002). *Xenopus* Mcm10 binds to origins of DNA replication after  
29 Mcm2-7 and stimulates origin binding of Cdc45. *Molecular Cell* **9**, 233-  
30 40.

- 1 Wood, Z. A., Poole, L. B., and Karplus, P. A. (2003a). Peroxiredoxin evolution  
2 and the regulation of hydrogen peroxide signaling. *Science* **300**, 650-3.
- 3 Wood, Z. A., Schroder, E., Robin Harris, J., and Poole, L. B. (2003b).  
4 Structure, mechanism and regulation of peroxiredoxins. *Trends in*  
5 *Biochemical Sciences* **28**, 32-40.
- 6 Woods, S. A., Schwartzbach, S. D., and Guest, J. R. (1988). Two biochemically  
7 distinct classes of fumarase in *Escherichia coli*. *Biochimica et*  
8 *Biophysica Acta* **954**, 14-26.
- 9 Wöstmann, C., Tannich, E., and Bakker-Grunwald, T. (1992). Ubiquitin of  
10 *Entamoeba histolytica* deviates in 6 amino acid residues from the  
11 consensus of all other known ubiquitins. *FEBS Letters* **308**, 54-58.
- 12 Wu, G., and Müller, M. (2003). Glycogen phosphorylase sequences from the  
13 amitochondriate protists, *Trichomonas vaginalis*, *Mastigamoeba*  
14 *balamuthi*, *Entamoeba histolytica* and *Giardia intestinalis*. *Journal of*  
15 *Eukaryotic Microbiology* **50**, 366-72.
- 16 Xu, Y., Moseley, J. B., Sagot, I., Poy, F., Pellman, D., Goode, B. L., and Eck,  
17 M. J. (2004). Crystal structures of a Formin Homology-2 domain reveal  
18 a tethered dimer architecture. *Cell* **116**, 711-23.
- 19 Yadava, N., Chandok, M. R., Prasad, J., Bhattacharya, S., Sopory, S. K., and  
20 Bhattacharya, A. (1997). Characterization of EhCaBP, a calcium  
21 binding protein of *Entamoeba histolytica* and its binding proteins.  
22 *Molecular and Biochemical Parasitology* **84**, 69-82.
- 23 Yamawaki-Kataoka, Y., Tamaoki, T., Choe, H. R., Tanaka, H., and Kataoka, T.  
24 (1989). Adenylate cyclases in yeast: a comparison of the genes from  
25 *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*.  
26 *Proceedings of the National Academy of Sciences USA* **86**, 5693-7.
- 27 Yang, W., Li, E., Kairong, T., and Stanley, S. L., Jr. (1994). *Entamoeba*  
28 *histolytica* has an alcohol dehydrogenase homologous to the  
29 multifunctional *adhE* gene product of *Escherichia coli*. *Molecular and*  
30 *Biochemical Parasitology* **64**, 253-260.

- 1 Yang, B., Steegmaier, M., Gonzalez, L. C., Jr., and Scheller, R. H. (2000).  
2 nSec1 binds a closed conformation of syntaxin1A. *Journal of Cell*  
3 *Biology* **148**, 247-52.
- 4 Yasui, K., Ishiguro, A., and Ishihama, A. (1998). Location of subunit-subunit  
5 contact sites on RNA polymerase II subunit 3 from the fission yeast  
6 *Schizosaccharomyces pombe*. *Biochemistry* **37**, 5542-8.
- 7 Yeo, M., Lin, P. S., Dahmus, M. E., and Gill, G. N. (2003). A novel RNA  
8 polymerase II C-terminal domain phosphatase that preferentially  
9 dephosphorylates serine 5. *Journal of Biological Chemistry* **278**, 26078-  
10 85.
- 11 Yin, Q. Y., de Groot, P. W., Dekker, H. L., de Jong, L., Klis, F. M., and de  
12 Koster, C. G. (2005). Comprehensive proteomic analysis of  
13 *Saccharomyces cerevisiae* cell walls: identification of proteins  
14 covalently attached via glycosylphosphatidylinositol remnants or mild  
15 alkali-sensitive linkages. *Journal of Biological Chemistry* **280**, 20894-  
16 901.
- 17 Yoshihisa, T., Barlowe, C., and Schekman, R. (1993). Requirement for a  
18 GTPase-activating protein in vesicle budding from the endoplasmic  
19 reticulum. *Science* **259**, 1466-8
- 20 Young, R. A. (1991). RNA polymerase II. *Annual Review of Biochemistry* **60**,  
21 689-715.
- 22 Yu, X., and Warner, J. R. (2001). Expression of a micro-protein. *Journal of*  
23 *Biological Chemistry* **276**, 33821-5.
- 24 Zerial, M., and McBride, H. (2001). Rab proteins as membrane organizers.  
25 *Nature Reviews. Molecular and Cellular Biology* **2**, 107-17.
- 26 Zhang, J. (2000). Protein-length distributions for the three domains of life.  
27 *Trends in Genetics* **16**, 107-9.
- 28 Zhang, Z. Y., and VanEtten, R. L. (1991). Pre-steady-state and steady-state  
29 kinetic analysis of the low molecular weight phosphotyrosyl protein  
30 phosphatase from bovine heart. *Journal of Biological Chemistry* **266**,  
31 1516-25.



- 1 Zhang, W. W., Shen, P. S., Descoteaux, S., and Samuelson, J. (1994). Cloning  
2 and expression of the gene for an NADP<sup>+</sup>-dependent aldehyde  
3 dehydrogenase of *Entamoeba histolytica*. *Molecular and Biochemical*  
4 *Parasitology* **63**, 157-61.
- 5 Zhang, Z., Wang, L., Seydel, K. B., Li, E., Ankri, S., Mirelman, D., and  
6 Stanley, S. L., Jr. (2000). *Entamoeba histolytica* cysteine proteinases  
7 with interleukin-1 beta converting enzyme (ICE) activity cause  
8 intestinal inflammation and tissue damage in amoebiasis. *Molecular*  
9 *Microbiology* **37**, 542-8.
- 10 Zhang, Z., Duchene, M., and Stanley, S. L., Jr. (2002). A monoclonal antibody  
11 to the amebic lipophosphoglycan-proteophosphoglycan antigens can  
12 prevent disease in human intestinal xenografts infected with *Entamoeba*  
13 *histolytica*. *Infection and Immunity* **70**, 5873-6.
- 14 Zuo, X., and Coombs, G. H. (1995). Amino acid consumption by the parasitic,  
15 amoeboid protists *Entamoeba histolytica* and *E. invadens*. *FEMS*  
16 *Microbiology Letters* **130**, 253-258.
- 17

1 *Table 1. Genome summary statistics for selected single celled*  
 2 *organisms with sequenced genomes.*

3

	<i>Entamoeba histolytica</i>	<i>Plasmodium falciparum</i>	<i>Dictyostelium discoideum</i>	<i>Saccharomyces cerevisiae</i>	<i>Encephalitozoon cuniculi</i>
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

4

1 *Table 2. Summary properties of the repeated DNAs.*

2 (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) <sup>a</sup>
Ehssp1	0.9-1.1	Not Determined	306 (9)	1.5 (9)

3

4 (b)

Family 16	GTAATGAATATAYAACTAAGAATTTTCATTTAAAATGRATATG
Family 17	CAACAAATAAATRGKTTCAATAAAATA

5

6 (a) References for data: (1): (Van Dellen *et al.*, 2002a), (2): (Bakre *et al.*,  
7 2005), (3): This analysis, (4) Bakre and Bhattacharya, unpublished  
8 observations; (5): (Burch *et al.*, 1991), (6): (Cruz-Reyes *et al.*, 1995), (7): Shire  
9 and Ackers, submitted, (8): (Davis *et al.*, 2006), (9): (Satish *et al.*, 2003).<sup>a</sup> -  
10 although no transcript was detected the protein has been demonstrated on the  
11 cell surface and in Western blots using antibodies (Davis *et al.*, 2006).

- 1 (b) Consensus sequences of Family 16 and 17 repeats. Standard abbreviation
- 2 for degenerate sequence positions are used: R= purine, Y= pyrimidine, K= G or
- 3 T.

1 **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	249	1	139	1	287	2	414	1	1137	3	694	1	719	2
WD domain, G-beta repeat														
LRR_1	131	2	40	2	55	12	43	17	3793	2	494	5	372	4
Leucine Rich Repeat														
Pkinase	95	3	27	5	78	8	116	2	839	4	405	8	225	7
Protein kinase domain														
HEAT	70	4	13	15	44	17	114	3	220	17	162	26	108	12
HEAT repeat														
efhand	58	5	7	28	80	7	29	25	422	8	213	20	153	9
EF hand														
RRM_1	57	6	30	3	95	6	86	6	375	10	223	19	134	10
RNA recognition motif.														
Ras	46	7	9	22	13	44	25	28	78	68	66	76	126	11
Ras family														
TPR_1	42	8	23	7	48	15	103	4	334	12	180	22	168	8
Tetratricopeptide repeat														
Ank	34	9	6	34	55	12	61	9	431	6	629	2	446	3
Ankyrin repeat														
PUF	33	10	8	23	15	34	51	13	142	32	75	68	34	62
Pumilio-family RNA binding repeat														
RhoGAP	27	11	2	118	1	520	11	80	9	559	31	138	45	39
RhoGAP domain														
Myb_DNA-binding	22	12	15	12	10	62	21	34	424	7	30	141	55	26
Myb-like DNA-binding domain														
RhoGEF	22	12	1	230	0	1215	3	366	0	2581	34	130	47	37
RhoGEF domain														
Helicase_C	20	14	28	4	64	11	74	8	150	31	98	49	84	20
Helicase conserved C-terminal domain														
DEAD	20	14	22	9	49	14	59	10	103	50	76	67	48	35
DEAD/DEAH box helicase														
PH	19	16	1	230	5	123	25	28	22	255	77	63	94	16
PH domain														
Metallophos	19	16	6	34	16	32	21	34	66	83	78	62	31	67
Calcineurin-like phosphoesterase														
Gelsolin	18	18	2	118	2	295	4	255	33	169	12	323	29	68
Gelsolin repeat														
LIM	17	19	0	703	0	1215	8	116	16	341	103	47	56	25
LIM domain														
CH	16	20	4	54	1	520	7	137	26	211	57	87	49	33
Calponin homology (CH) domain														
Filamin	16	20	0	703	1	520	0	1842	2	1450	55	91	10	203
Filamin/ABP280 repeat														

18 Columns labeled “#” give the total number of occurrences of a particular domain. Columns labeled “Rank” give the ranking of the domain where

19 the most common domain is ranked 1. The organisms shown are *Entamoeba histolytica* (EH), *Encephalitozoon cuniculi* (EC), *Plasmodium*

20 *falciparum* (PF), *Arabidopsis thaliana* (AT), *Saccharomyces cerevisiae* (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
1	EhCP-A1	EhCP1	XP_650156	315 (13,80,222)	QCHN	ERFNIN, DWR	
2	EhCP-A2	EhCP2	XP_650642	315 (13,80,222)	QCHN	ERFNIN, DWR	
3	EhCP-A3	EhCP3	XP_653254	308 (13,79,216)	QCHN	ERFNIN, DWR	
4	EhCP-A4	EhCP4	XP_656602	311 (20,73,218)	QCHN	ERFNIN, DWR	
5	EhCP-A5	EhCP5	XP_650937	318 (20,72,225)	QCHN	ERFNIN, DWR, RGD	Degenerate in <i>E. dispar</i>
6	EhCP-A6	EhCP6	XP_657364	320 (17,79,224)	QCHN	ERFNIN, DWR	
7	EhCP-A7	EhCP8	XP_648996	315 (13,80,222)	QCHN	ERFNIN, DWR	
8	EhCP-A8	EhCP9	XP_657446	317 (15,82,220)	QCHN	ERFNIN, DWR	
9	EhCP-A9	EhCP10	XP_655675	297 (17,90,190)	QCHN	ERFNIN, DWR	
10	EhCP-A10	EhCP17	XP_651147	420 (18,148,254)	QCHN	ERFNIN, DWR	
11	EhCP-A11	EhCP19	XP_651690	324 (17,79,228)	QCIN <sup>a</sup>	ERFNIN, DWR	
12	EhCP-A12	new	XP_653823	317 (14,83,220)	(d)	ERFNIN, DWR	
13	EhCP-B1	EhCP7	XP_651581	426 (15,106,305)	QCHN	ERFNIN, PCNC	hydrophobic C-terminus
14	EhCP-B2	EhCP11	AAO03568	431 (15,106,310)	QCHS <sup>a</sup>	ERFNIN, PCNC	GPI cleavage site
15	EhCP-B3	EhCP12	XP_656747	474 (16,107,351)	QCHN	ERFNIN, PCNC	TMH:444-466 aa
16	EhCP-B4	EhCP13	XP_648501	379 (16,105,258)	QCHN	ERFNIN, PCNC	TMH or GPI cleavage site
17	EhCP-B5	EhCP14	XP_652671	434 (12,108,314)	QCHN	ERFNIN, PCNC	GPI cleavage site
18	EhCP-B6	EhCP15	XP_652465	300 (14,55,231)	QCHN	PCNC	hydrophobic C-terminus
19	EhCP-B7	EhCP16	XP_650400	650 (18,144,488)	QCHN	ERFNIN, PCNC	hydrophobic C-terminus, Cys-rich profile
20	EhCP-B8	EhCP18	XP_651049	473 (15,105,353)	QCHN	ERFNIN, PCNC, RGD	GPI cleavage site
21	EhCP-B9	EhCP112	XP_652993	446 (19,112,315)	QCHN	ERFNIN, PCNC, RGD	hydrophobic C-terminus, Cys-rich profile
22	EhCP-B10	new	XP_648306	372 (b)	QCHN	ERFNIN, PCNC, RGD	hydrophobic C-terminus
23	EhCP-B11	new	XP_648013	133 (b)	Q???	PCNC	hydrophobic C-terminus
24	EhCP-C1	new	XP_654453	586 (c)	QCIN <sup>a</sup>	HS(X) <sub>6</sub> ICP	TMH:12-34
25	EhCP-C2	new	XP_656632	567 (c)	QCHN	HS(X) <sub>6</sub> ICP	TMH:27-49
26	EhCP-C3	new	XP_655128	572 (c)	QCHN	HS(X) <sub>6</sub> LCP	TMH:17-39
27	EhCP-C4	new	XP_655800	502 (c)	QCHN	LT(X) <sub>6</sub> LCP	
28	EhCP-C5	new	XP_654800	557 (c)	QCHN	IS(X) <sub>6</sub> ICP	TMH:20-42
29	EhCP-C6	new	XP_651553	557 (c)	QCHD <sup>a</sup>	HS(X) <sub>6</sub> LCA	TMH:14-36
30	EhCP-C7	new	XP_657273	595 (c)	QCHN	IS(X) <sub>6</sub> LCP	TMH:19-41
31	EhCP-C8	new	XP_655479	627 (c)	QCHN	IS(X) <sub>6</sub> ICP	TMH:29-51

1	EhCP-C9	new	XP_655011	518 (c)	(d)	HS(X) <sub>6</sub> ICP	TMH:12-34
2	EhCP-C10	new	XP_654829	530 (c)	QCHN	IS(X) <sub>6</sub> ICP	TMH:15-37
3	EhCP-C11	new	XP_648083	526 (c)	(d)	HS(X) <sub>6</sub> ICP	TMH:20-42
4	EhCP-C12	new	XP_650829	473 (c)	(d)	MS(X) <sub>6</sub> LCG	TMH:26-48 & 449-471
5	EhCP-C13	new	XP_656556	564 (c)	QCHN	Vs(X) <sub>6</sub> RCG	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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18



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 <sup>g</sup>	Similar to (aa sequence identity, %) <sup>f</sup>	Homologous proteins in other organisms (aa sequence identity, %)
	entire	signal peptide (predicted) <sup>a</sup>	proform/mature <sup>c</sup>	position aa residues				
Amoebapore A	98	21 <sup>b</sup>	77	22-98	amoebapore A precursor ACCESSION_XP_653265 sporein-like protein ACCESSION_XP_653836		Amoebapore A Acc:AAA29111 (100%) Amoebapore A Acc:AAA29111 (64%)	Disparpore A Acc:AAA18632 <i>Ediipar</i> (94%) Disparpore A Acc:AAA18632 <i>Ediipar</i> (85%)
SAPLIP 1	92	15	77	16-92				
Amoebapore B	96	19 <sup>b</sup>	77	20-96	Pore-forming peptide amoebapore B precursor (PFLAPP) ACCESSION_Q24824		Amoebapore B Acc:CAA54226 (100%)	Disparpore B Acc:AAF04195 <i>Ediipar</i> (90%)
Amoebapore C	101	24 <sup>b</sup>	77	25-101	amoebapore C ACCESSION_XP_653829		Amoebapore C Acc:CAA54225 (100%)	Disparpore C Acc:AAF04196 <i>Ediipar</i> (85%)
SAPLIP 2	153	15	138	71-153	hypothetical protein ACCESSION_XP_656037			
SAPLIP 3	94	16	78	18-94	hypothetical protein ACCESSION_XP_656082		Amoebapore A Acc:AAA29111 (90%)	Invapore X Acc:AAF03381 <i>E.mutans</i> (67%)
SAPLIP 4	96	17	79	18-96	hypothetical protein ACCESSION_XP_652159 AND hypothetical protein ACCESSION_XP_652303		Amoebapore C Acc:CAA54225 (27%)	Disparpore C Acc:AAF04196 <i>Ediipar</i> (90%)
SAPLIP 5	1026	18	1008	946-1026	chromosome partition protein ACCESSION_XP_655789			
SAPLIP 6	92	15	77	14-92	hypothetical protein ACCESSION_XP_655820			
SAPLIP 7	926	17	909	855-926	conserved hypothetical protein ACCESSION_XP_656441			
SAPLIP 8	980	15	965	902-980	hypothetical protein ACCESSION_XP_656913			
SAPLIP 9	140	15	125	61-140	hypothetical protein ACCESSION_XP_656976			
SAPLIP 10	657	16	641	577-657		Genomic survey sequence ACCESSION_AZ687176		
SAPLIP 11	693	17	676	615-693 <sup>d</sup>		Genomic survey sequence ACCESSION_AZ692183		
SAPLIP 12	873	16	857	798-873	hypothetical protein ACCESSION_XP_652721			
SAPLIP 13	1009	no signal peptide predicted	1009	931-1005	hypothetical protein ACCESSION_XP_655089			
SAPLIP 14	915	17	898	834-915		Genomic survey sequence ACCESSION_AZ690015		

1 SAPLIPs were named according to the similarity of their SAPLIP domain to amoebapore A:

2 <sup>a</sup> by the programme SignalP and manually corrected if predicted cleavage site is within the SAPLIP domain

3 <sup>b</sup> verified by experimental data

4 <sup>c</sup> with the exception of amoebapores it is not possible to decide whether proteins are further processed

- 1 <sup>d</sup> identified manually
- 2 <sup>e</sup> extracted from InterPro databases
- 3 <sup>f</sup> if no similarity is reported, there is none outside of the SAPLIP domain
- 4 <sup>g</sup> sequences only found in GSS section of GenBank with given identifier

1 **Table 7. The number of genes encoding representative proteins involved in vesicular trafficking in *E. histolytica*.**

Protein	<i>E. histolytica</i>	<i>S. cerevisiae</i>	<i>C. elegans</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>	<i>A. thaliana</i>	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

2 **References (1), Bock et al. (2001); (2), Wennerberg et al. (2005); (3), Pasqualato et al. (2002); (4), Pereira-Leal and Seabra (2001); (5), Burri**  
 3 **and Lithgow (2004); (6), Uemura et al. (2004); (7), Sanderfoot et al. (2000); (8) Boehm et al. (2001).**

1 Table 8. Reassessment of the 96 candidate LGT cases identified in the original genome publication.

Acc. <sup>a</sup> RefSeq	Acc. <sup>a</sup> EHL <sup>b</sup>	Top Prokaryotic blast hit	PL <sup>c</sup>	%ID <sup>d</sup>	Top Eukaryotic blast hit	EL <sup>c</sup>	%ID <sup>d</sup>	PE-score <sup>e</sup>	EE-score <sup>f</sup>	P/E Ratio <sup>g</sup>
<b>41 LGT cases that remain strongly supported according to our criteria</b>										
EAL43201	XP_648590.1 487	<i>Treponema denticola</i>	507	57	<i>Trichomonas vaginalis</i>	398	43	1.00E-167	5.00E-88	2.00E-80
EAL43619	XP_649008.1 621	<i>Vibrio vulnificus</i>	673	41	<i>Saccharomyces cerevisiae</i>	664	40	1.00E-132	1.00E-125	1.00E-07
EAL43678	XP_649067.1 538	<i>Fusobacterium nucleatum</i> <i>Mannheimia</i>	562	47	<i>Trichomonas vaginalis</i>	477	34	1.00E-135	2.00E-60	5.00E-76
EAL43850	XP_649240.1 880	<i>succiniciproducens</i>	898	63	<i>Mastigamoeba balamuthi</i>	882	45	0	0	N/A
EAL44182	XP_649570.1 260	<i>Bacteroides thetaiotaomicron</i>	273	34	<i>Yarrowia lipolytica</i>	298	29	2.00E-35	4.00E-10	5.00E-26
EAL44226	XP_649612.1 262	<i>Bacteroides thetaiotaomicron</i>	267	28	<i>Tetrahymena thermophila</i>	1476	30	2.00E-25	0.11	1.82E-24
EAL44778	XP_650165.1 188	<i>Bacteroides thetaiotaomicron</i>	188	43	<i>Neurospora crassa</i>	546	34	8.00E-41	1.8	4.44E-41
EAL45076	XP_650453.1 358	<i>Bacteroides fragilis</i>	362	46	<i>Trichomonas vaginalis</i>	562	22	1.00E-87	0.24	4.17E-87
EAL45145	XP_650531.1 825	<i>Staphylococcus aureus</i>	1036	30	<i>Trichomonas vaginalis</i>	2468	20	3.00E-59	0.016	1.88E-57
EAL45220	XP_650606.1 479	<i>Clostridium tetani</i>	471	45	<i>Arabidopsis thaliana</i>	581	31	1.00E-114	1.00E-54	1.00E-60
EAL44744	XP_650131.1 160	<i>Bacteroides fragilis</i>	424	41	<i>Yarrowia lipolytica</i>	169	31	3.00E-24	7.00E-11	4.29E-14
EAL46110	XP_651498.1 157	<i>Bacteroides fragilis</i>	166	49	<i>Arabidopsis thaliana</i>	627	35	5.00E-35	3.2	1.56E-35
EAL45378	XP_650765.1 311	<i>Haloarcula marismortui</i>	299	43	<i>Leishmania major</i>	411	43	3.00E-54	1.00E-32	3.00E-22
EAL45618	XP_651004.1 159	<i>Bacteroides thetaiotaomicron</i>	157	46	<i>Plasmodium vivax</i>	1275	33	2.00E-28	0.69	2.90E-28
EAL46311	XP_651697.1 248	<i>Synechococcus elongates</i>	270	36	<i>Trichomonas vaginalis</i>	3075	18	1.00E-30	0.38	2.63E-30
EAL46679	XP_652065.1 218	<i>Methanosarcina mazei</i>	230	37	<i>Canidia glabrata</i>	461	24	8.00E-31	0.079	1.01E-29
EAL46975	XP_652361.1 370	<i>Bordetella bronchiseptica</i>	368	46	<i>Cryptococcus neoformans</i>	372	40	8.00E-83	3.00E-71	2.67E-12
EAL47525	XP_652912.1 380	<i>Clostridium perfringens</i>	296	23	<i>Plasmodium falciparum</i>	390	34	2.00E-13	1.3	1.54E-13
EAL47905	XP_653291.1 227	<i>Clostridium perfringens</i>	259	33	<i>Tetrahymena thermophila</i>	1425	24	4.00E-19	0.32	1.25E-18
EAL48587	XP_653973.1 425	<i>Clostridium perfringens</i>	442	60	<i>Yarrowia lipolytica</i>	572	37	1.00E-149	9.00E-57	1.11E-93
EAL48979	XP_654365.1 732	<i>Desulfovibrio vulgaris</i>	740	40	<i>Cryptococcus neoformans</i>	735	28	1.00E-135	3.00E-64	3.33E-72
EAL49084	XP_654474.1 350	<i>Thermotoga neapolitana</i>	241	29	<i>Anopheles gambiae</i>	784	40	1.00E-24	5.00E-06	2.00E-19
EAL49209	XP_654596.1 247	<i>Methanococcus jannaschii</i>	243	38	<i>Thalassiosira pseudonana</i>	269	22	7.00E-43	0.0002	3.50E-39
EAL49277	XP_654665.1 737	<i>Bacteroides fragilis</i>	781	31	<i>Cryptococcus neoformans</i>	935	24	1.00E-111	6.00E-44	1.67E-68
EAL49613	XP_654999.1 168	<i>Bacteroides thetaiotaomicron</i>	237	34	<i>Tetrahymena thermophila</i>	487	38	1.00E-16	6.00E-06	1.67E-11
EAL49813	XP_655200.1 186	<i>Sulfolobus solfataricus</i>	200	31	<i>P. brasiliensis</i>	257	26	2.00E-13	0.47	4.26E-13
EAL49869	XP_655257.1 390	<i>Escherichia coli</i>	407	56	<i>Ashbya gossypii</i>	490	39	1.00E-124	8.00E-73	1.25E-52
EAL50263	XP_655646.1 390	<i>Campylobacter jejuni</i>	408	48	<i>Yarrowia lipolytica</i>	428	38	1.00E-98	3.00E-60	3.33E-39
EAL50440	XP_655826.1 344	<i>Porphyromonas gingivalis</i>	491	54	<i>Rhizopus oryzae</i>	510	40	1.00E-101	2.00E-67	5.00E-35
EAL50508	XP_655888.1 348	<i>Bacillus anthracis</i>	340	55	<i>Mus musculus</i>	168	40	1.00E-106	2.00E-18	5.00E-89
EAL50603	XP_655988.1 567	<i>Wolinella succinogenes</i>	622	45	<i>Trichomonas vaginalis</i>	632	39	1.00E-141	2.00E-99	5.00E-43
EAL50801	XP_656185.1 499	<i>Bacteroides thetaiotaomicron</i>	513	52	<i>Trichomonas vaginalis</i>	514	28	1.00E-145	3.00E-40	106
EAL50992	XP_656375.1 140	<i>Archaeoglobus fulgidus</i>	184	40	<i>Trichomonas vaginalis</i>	195	46	1.00E-27	0.018	5.56E-26