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 microbial 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).

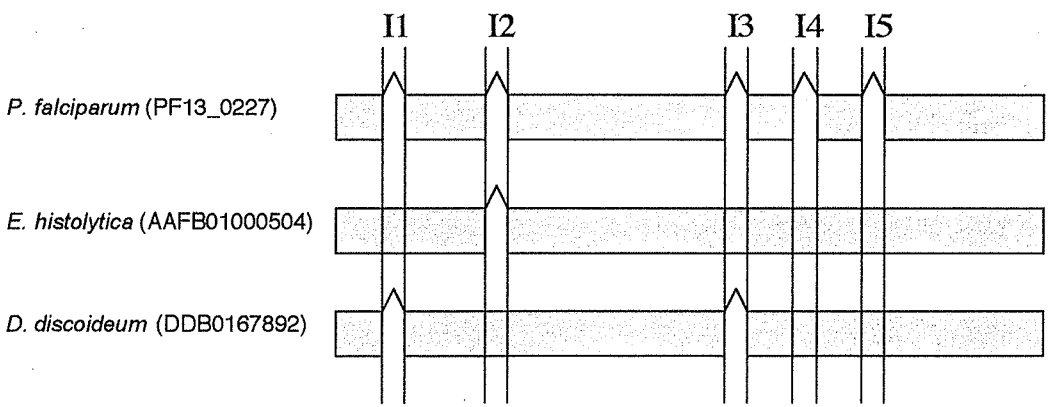
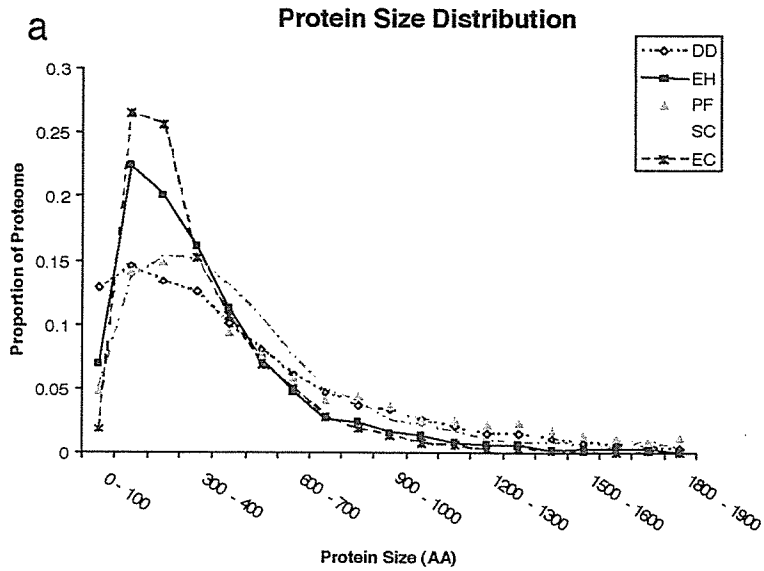

I1
I2
I3
I4
I5
P. falciparum (PF13_0227)
E. histolytica (AAFB01000504)
D. discoideum (DDB0167892)

Fig. 1

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)

b

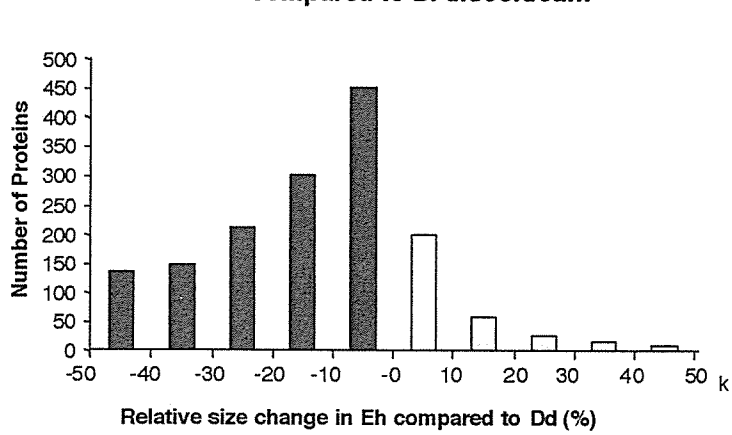

Degree of reduction in length of proteins in E. histolytica compared to D. discoideum
Number of Proteins
500
450
400
350
300
250
200
150
100
50
0
-50
-40
-30
-20
-10
-0
10
20
30
40
50
k
Relative size change in Eh compared to Dd (%)

Fig. 2

1
2

1

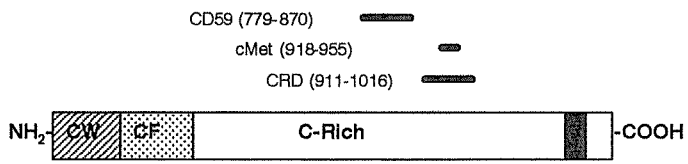


Fig. 3

2

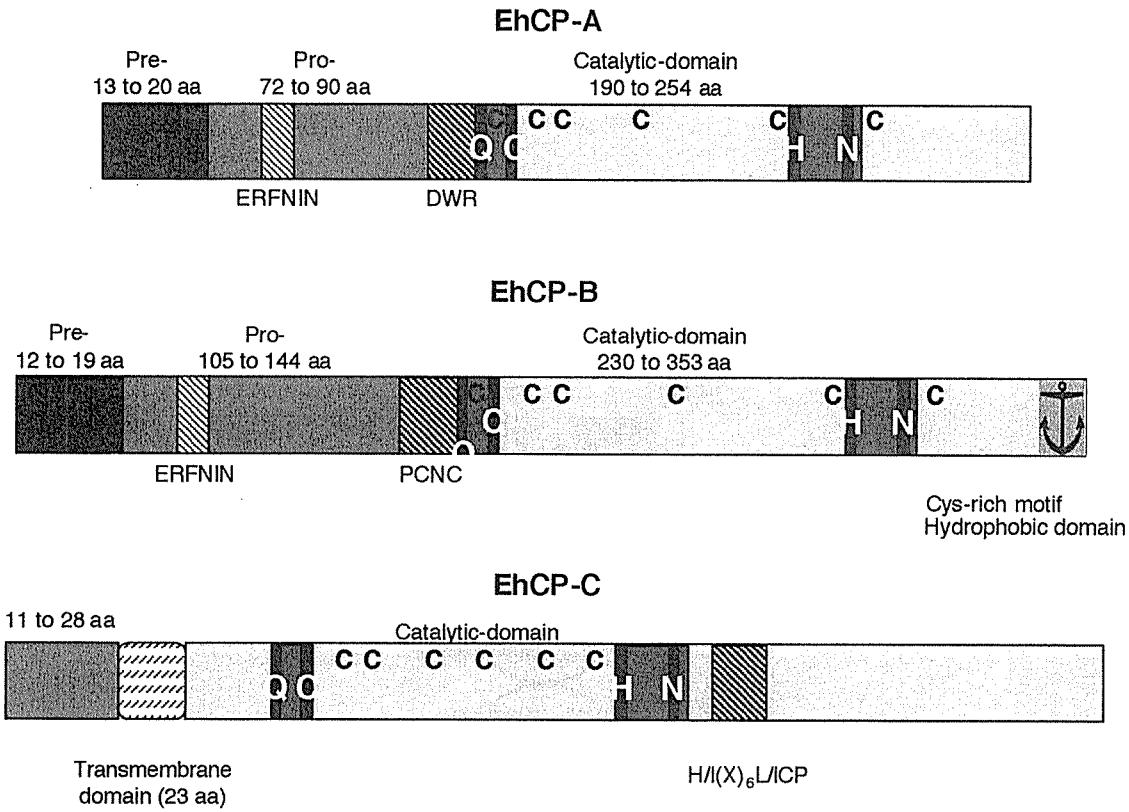

EhCP-A
Pre-
13 to 20 aa
Pro-
72 to 90 aa
Catalytic-domain
190 to 254 aa
C C
C
C
C
Q
H
N
ERFNIN
DWR
EhCP-B
Pre-
12 to 19 aa
Pro-
105 to 144 aa
Catalytic-domain
230 to 353 aa
C C
C
C
C
Q
H
N
ERFNIN
PCNC
Cys-rich motif
Hydrophobic domain
EhCP-C
11 to 28 aa
Catalytic-domain
C C
C
C
C
C
Q
Q
H
N
Transmembrane
domain (23 aa)
H/I(X)6L/ICP

Fig. 4

1

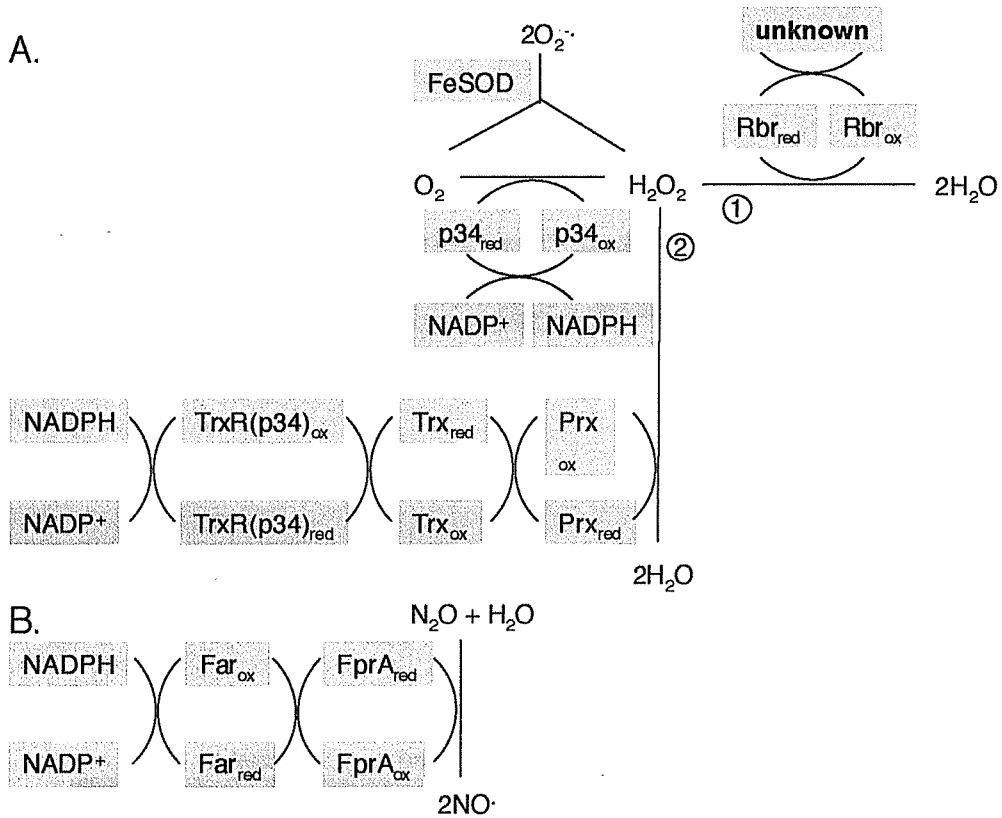

A.
$2O_2^{\cdot-}$
FeSOD
unknown
Rbr_{red}
Rbr_{ox}
O_2
H_2O_2
$2H_2O$
①
$p34_{red}$
$p34_{ox}$
②
$NADP^+$
NADPH
NADPH
$TrxR(p34)_{ox}$
Trx_{red}
Prx_{ox}
$NADP^+$
$TrxR(p34)_{red}$
Trx_{ox}
Prx_{red}
$2H_2O$
B.
$N_2O + H_2O$
NADPH
Far_{ox}
$FprA_{red}$
$NADP^+$
Far_{red}
$FprA_{ox}$
$2NO^{\cdot}$

Fig. 5

1

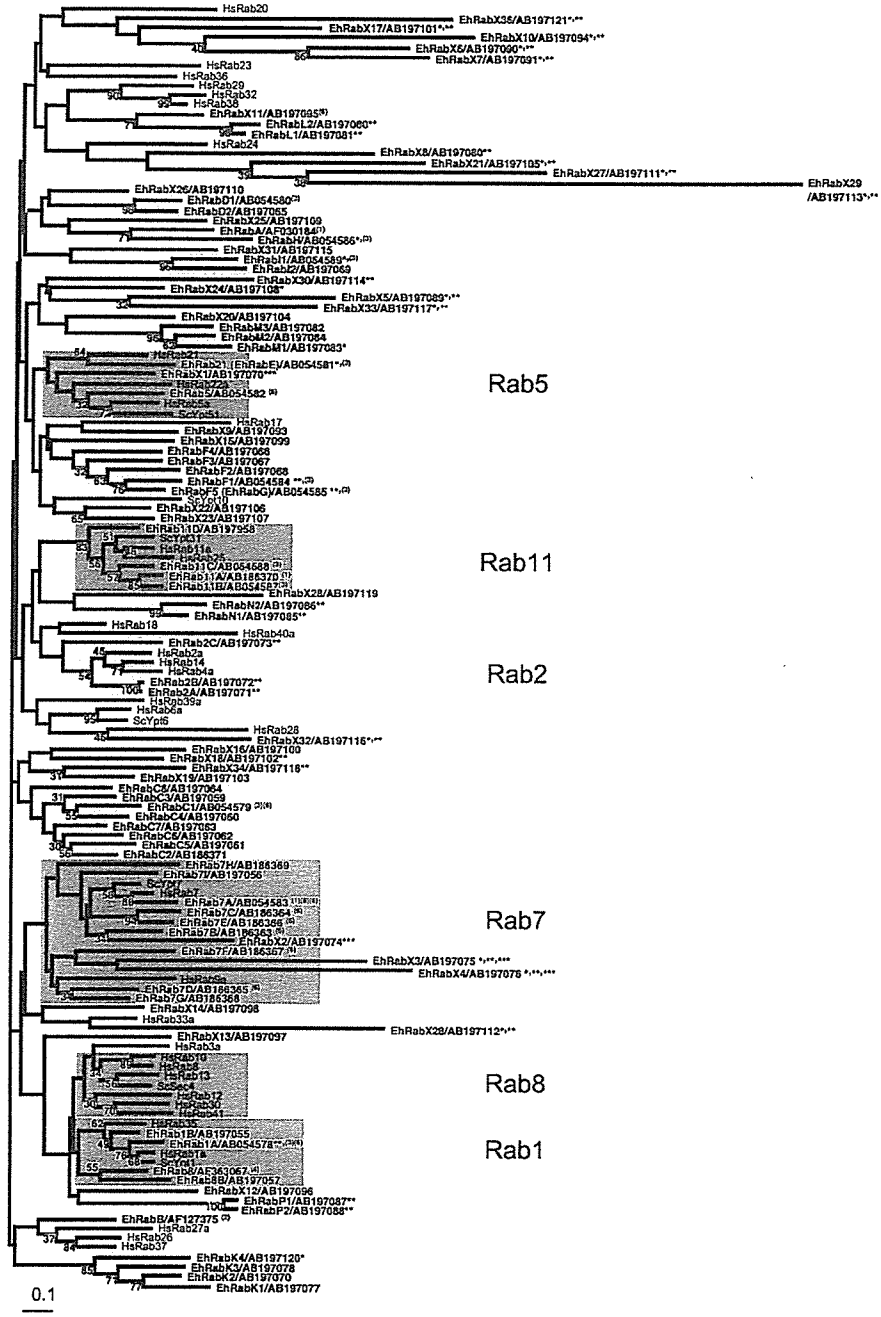
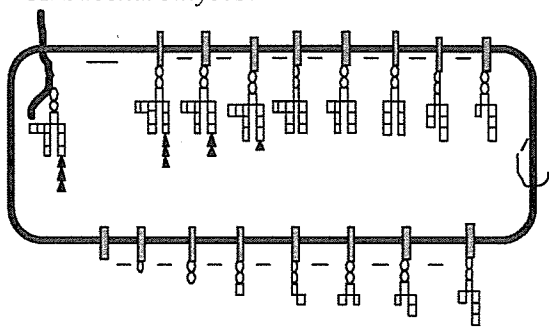
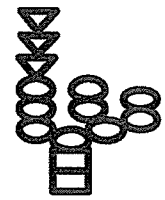

HsRab20
EhRabX35/AB197121*,**
EhRabX17/AB197101*,**
EhRabX10/AB197094*,**
EhRabX6/AB197090*,**
EhRabX7/AB197091*,**
HsRab23
HsRab36
HsRab29
HsRab32
HsRab38
EhRabX11/AB197085
EhRabL2/AB197080**
EhRabL1/AB197081**
HsRab24
EhRabX8/AB197086**
EhRabX21/AB197105*,**
EhRabX27/AB197111*,**
EhRabX29 /AB197113*,**
EhRabX26/AB197110
EhRabD1/AB054580
EhRabD2/AB197065
EhRabX25/AB197109
EhRabA/AF030184
EhRabH/AB054586*,
EhRabX31/AB197115
EhRabI1/AB054589*,
EhRabI2/AB197069
EhRabX30/AB197114**
EhRabX24/AB197108*
EhRabX5/AB197089*,**
EhRabX33/AB197117*,**
EhRabX20/AB197104
EhRabM3/AB197082
EhRabM2/AB197064
EhRabM1/AB197083*
HsRab21
EhRab21 (EhRabE)/AB054581*,
EhRabX1/AB197070***
HsRab22a
EhRab5/AB054582
HsRab5a
ScYpt51
HsRab17
EhRabX9/AB197093
EhRabX15/AB197099
EhRabF4/AB197066
EhRabF3/AB197067
EhRabF2/AB197068
EhRabF1/AB054584 **,
EhRabF5 (EhRabG)/AB054585 **,
ScYpt52
EhRabX22/AB197106
EhRabX23/AB197107
EhRab11D/AB197958
ScYPT31
HsRab11a
HsRab25
EhRab11C/AB054588
EhRab11A/AB186370
EhRab11B/AB054587
EhRabX28/AB197119
EhRabN2/AB197086**
EhRabN1/AB197085**
HsRab18
HsRab40a
EhRab2C/AB197073**
HsRab2a
HsRab14
HsRab4a
EhRab2B/AB197072**
EhRab2A/AB197071**
HsRgp39a
HsRab6a
ScYpt6
HsRab28
EhRabX32/AB197116*,**
EhRabX16/AB197100
EhRabX18/AB197102**
EhRabX34/AB197118**
EhRabX19/AB197103
EhRabC6/AB197064
EhRabC3/AB197059
EhRabC1/AB054579
EhRabC4/AB197060
EhRabC7/AB197063
EhRabC5/AB197062
EhRabC5/AB197061
EhRabC2/AB186371
EhRab7H/AB186369
EhRab7f/AB197056
ScYPT7
HsRab7
EhRab7A/AB054583
EhRab7C/AB186364
EhRab7E/AB186366
EhRab7B/AB186363
EhRabX2/AB197074***
EhRab7F/AB186367
EhRabX3/AB197075 *,**,***
EhRabX4/AB197076 *,**,***
HsRab9a
EhRab7D/AB186365
EhRab7G/AB186368
EhRabX14/AB197098
HsRab33a
EhRabX28/AB197112*,**
EhRabX13/AB197097
HsRab3a
HsRab10
HsRab8
HsRab13
ScSec4
HsRab12
HsRab30
HsRab41
HsRab15
EhRab1B/AB197055
EhRab1A/AB054578
HsRab1a
ScYPT1
EhRab6/AB363067
EhRab6B/AB197057
EhRabX12/AB197096
EhRabP1/AB197087**
EhRabP2/AB197088**
EhRabB/AF127375
HsRab27a
HsRab26
HsRab37
EhRabK4/AB197120*
EhRabK3/AB197078
EhRabK2/AB197070
EhRabK1/AB197077
Rab5
Rab11
Rab2
Rab7
Rab8
Rab1
0.1

Fig. 6

1
2

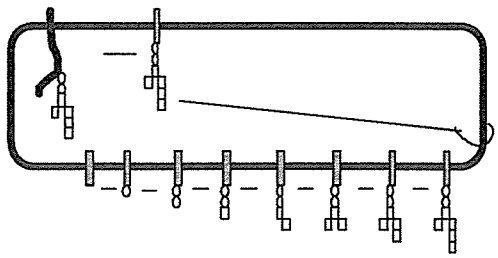
A. *Saccharomyces*:

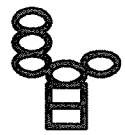


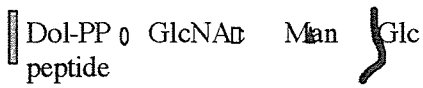


Glc_3Man_9GlcN

B. *Entamoeba*:







1

2 Fig. 7

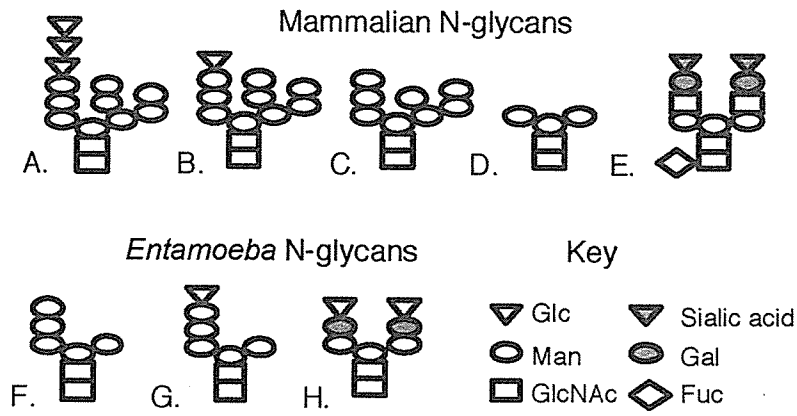

Mammalian N-glycans
A.
B.
C.
D.
E.
Entamoeba N-glycans
F.
G.
H.
Key
Glc
Sialic acid
Man
Gal
GlcNAc
Fuc

Fig. 8

1
2

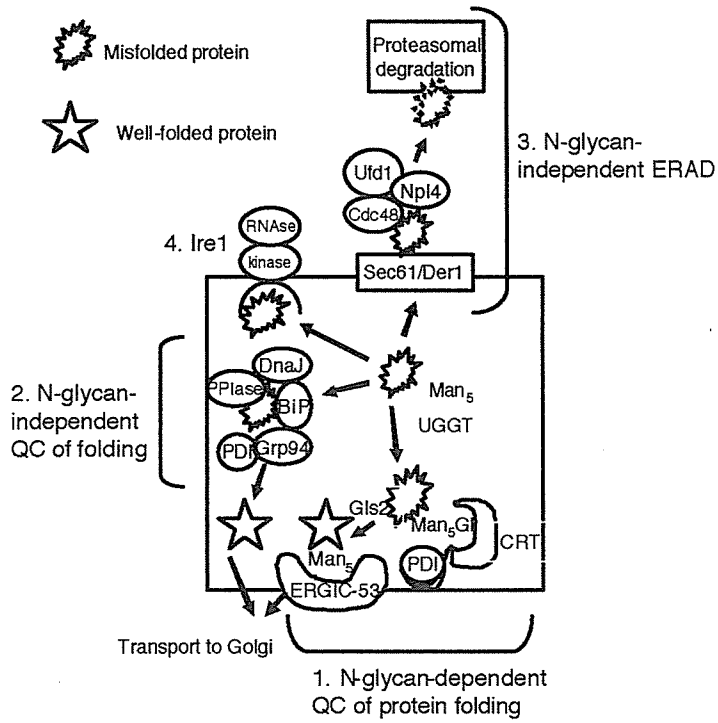

Misfolded protein
Well-folded protein
Proteasomal degradation
3. N-glycan-independent ERAD
Ufd1
Npl4
Cdc48
4. Ire1
RNAse
kinase
Sec61/Der1
2. N-glycan-independent QC of folding
DnaJ
PPIase
BiP
PDI
Grp94
Man5
UGGT
Gls2
Man5GlcNAc2
CRT
Man5
PDI
ERGIC-53
Transport to Golgi
1. N-glycan-dependent QC of protein folding

Fig. 9

1

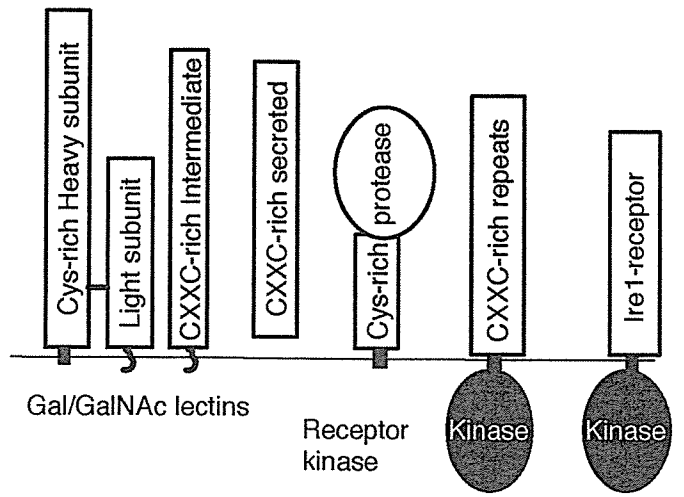

Cys-rich Heavy subunit
Light subunit
CXXC-rich Intermediate
CXXC-rich secreted
Cys-rich protease
CXXC-rich repeats
Ire1-receptor
Gal/GalNAc lectins
Receptor kinase
Kinase
Kinase

Fig. 10

1

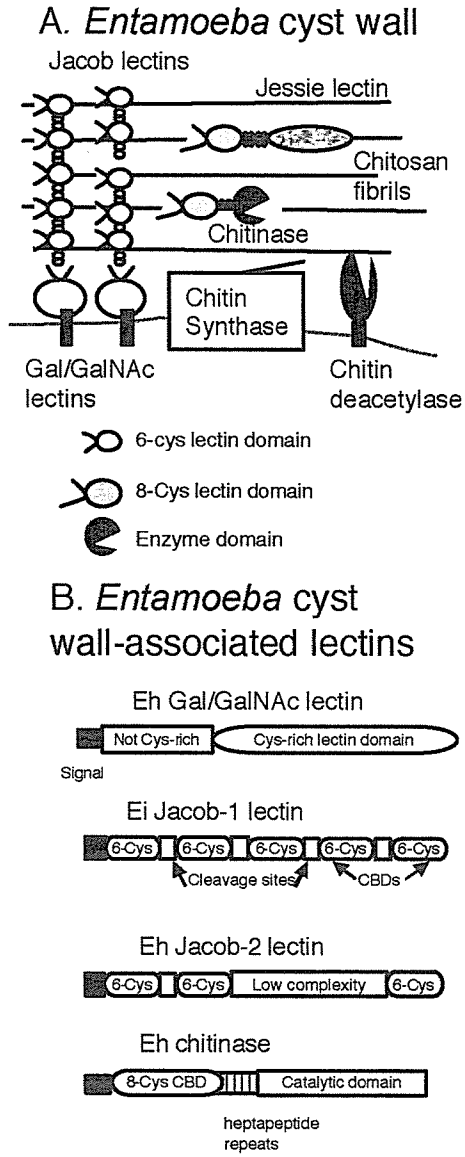

A. Entamoeba cyst wall
Jacob lectins
Jessie lectin
Chitosan fibrils
Chitinase
Chitin Synthase
Gal/GalNAc lectins
Chitin deacetylase
6-cys lectin domain
8-Cys lectin domain
Enzyme domain
B. Entamoeba cyst wall-associated lectins
Eh Gal/GalNAc lectin
Not Cys-rich
Cys-rich lectin domain
Signal
Ei Jacob-1 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

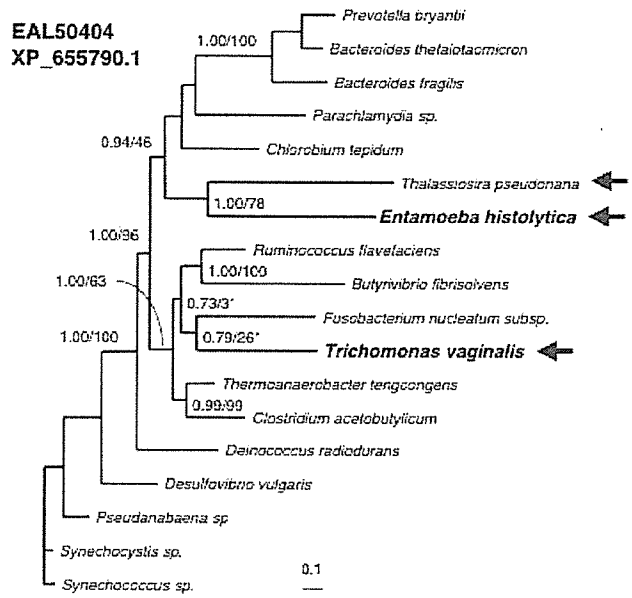

EAL50404
XP_655790.1
Prevotella bryantii
1.00/100
Bacteroides thetaiotaomicron
Bacteroides fragilis
Parachlamydia sp.
0.94/46
Chlorobium tepidum
Thalassiosira pseudonana
1.00/78
Entamoeba histolytica
1.00/96
Ruminococcus flavefaciens
1.00/63
1.00/100
Butyrivibrio fibrisolvens
0.73/3*
Fusobacterium nucleatum 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

1
2