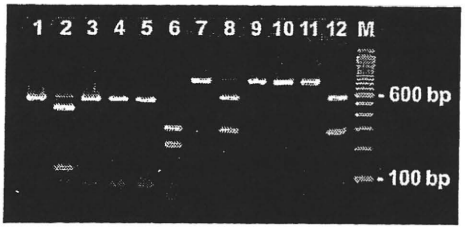

Fig. 1. RFLP patterns of PCR products amplified from the DNA of *P. westermani* metacercariae (lanes 1-3 for both the diploid and triploid forms; lanes 7-9 for the diploid form; lanes 10-12 for the triploid form) or *P. miyazakii* metacercariae (lanes 4-6). The ITS2 PCR products were untreated (lanes 1 and 4) or treated with endonucleases *Sna*BI (lanes 2 and 5) or *Bss*SI (lanes 3 and 6). The 16S rDNA PCR products were also untreated (lanes 7 and 10) or treated with endonucleases *Sna*BI (lanes 8 and 11) or *Bsr*DI (lanes 9 and 12). A 100-bp DNA ladder marker was used to estimate the size of the fragments.

sis.

PCR amplification with the primer pair 3S and A28 generated single 520-bp products from the metacercarial DNA samples. Electrophoresis of the restriction enzyme-digested products resulted in two species-specific RFLP patterns, as previously described (4). Species identification of the metacercariae was made based on the digestion patterns of amplification products. Products that were digested with *Sna*BI to produce 2 fragments (about 420 bp and 100 bp) but remained undigested with *Bss*SI were identified as those of *P. westermani*. Products that were undigested with *Sna*BI but were digested with *Bss*SI to produce 2 fragments (about 300 bp and 220 bp; Fig. 1) were identified as those of *P. miyazakii*.

DNA samples prepared from *P. westermani* metacercariae were further analyzed to determine the form, i.e., diploid or triploid. PCR amplification of mitochondrial DNA with the primer pair SP6-1 and T7-1 produced a single 840-bp product. Restriction digestion of PCR products was used to identify the diploid and triploid forms. Products that were digested with *Sna*BI to produce 2 fragments (about 550 bp and 290 bp) but remained undigested with *Bsr*DI were identified as those of the diploid form. Products that remained undigested with *Sna*BI but were digested with *Bsr*DI to produce 2 fragments (about 560 bp and 280 bp; Fig. 1) were identified as those of the triploid form. The species and forms identified by the RFLP analyses were verified by sequencing of the respective PCR products.

Consequently, as shown in Table 1, most of the metacercariae were identified as *P. miyazakii* (157 metacercariae from 36 positive crabs), while the others were *P. westermani* (3 metacercariae from 3 positive crabs and 9 metacercariae from 5 positive crabs were of the diploid and triploid forms, respectively). However, there were no mixed infections either with *P. miyazakii* and *P. westermani* (diploid and/or triploid forms) or with both forms of *P. westermani* in any crab examined in the present study.

Sawagani from Miyazaki Prefecture were also purchased at a retail fish market in Fukuoka City in April 2008 and were examined for *Paragonimus* metacercariae. *P. miyazakii* metacercariae (35 in total) were detected in 15 of 30 examined crabs. This finding implies that Sawagani with *Paragonimus* metacercariae that are responsible for human infections are likely also sold in retail fish markets in areas other than Tokyo.

The heat resistance of *P. westermani* metacercariae within the crab hosts was investigated almost a century ago (8). The Japanese mitten crab, *Eriocheir japonicus*, which played a major role as the second intermediate host in spreading the human infection of *P. westermani* at that time in Japan was investigated (*P. miyazakii* metacercariae have never been isolated from this crab species). It was shown that boiling infected crabs at 55°C for 5 min killed all the metacercariae (8). However, to the best of our knowledge, the conditions required to kill metacercariae of *P. westermani* and *P. miyazakii* in Sawagani have not yet been well examined, although we are currently investigating these conditions. Therefore, the implementation of a health education campaign is recommended throughout Japan to emphasize that Sawagani, even those sold at retail fish markets, are potential sources of lung fluke infection in humans. Special attention should be paid to ethnic dishes that are prepared with uncooked Sawagani.

This study was supported in part by grants for Emerging and Re-emerging Infectious Diseases from the Ministry of Health, Labour and Welfare of Japan (H18-Shinko-ippan-008 and H20-Shinko-ippan-016).

This article appeared in the Infectious Agents Surveillance Report (IASR), vol. 29, p. 284-285, 2008 in Japanese.

REFERENCES

1. Miyazaki, I. (1991): Paragonimiasis. p. 76-146. *In* I. Miyazaki (ed.), An Illustrated Book of Helminthic Zoonoses. International Medical Foundation of Japan, Tokyo.
2. Kawanaka, M., Arakawa, K., Morishima, Y., et al. (2004): Paragonimiasis cases among foreigners staying in Japan due to retaining their own dietary habit. Infect. Agents Surveillance Rep., 25, 121-122 (in Japanese).
3. Okuyama, S., Narisawa, E., Fujita, A., et al. (2007): A case of chronic paragonimiasis westermani, diagnosed by the detection of an egg in pleural effusion. Clin. Parasitol., 18, 35-37 (in Japanese).
4. Sugiyama, H., Morishima, Y., Kameoka, Y., et al. (2002): Polymerase chain reaction (PCR)-based molecular discrimination between *Paragonimus westermani* and *P. miyazakii* at the metacercarial stage. Mol. Cell. Probes, 16, 231-236.
5. Blair, D., Agatsuma, T., Watanobe, T., et al. (1997): Geographical genetic structure within the human lung fluke, *Paragonimus westermani*, detected from DNA sequences. Parasitology, 115, 411-417.
6. Agatsuma, T., Iwagami, M., Sato, Y., et al. (2003): The origin of the triploid in *Paragonimus westermani* on the basis of variable regions in the mitochondrial DNA. J. Helminthol., 77, 279-285.
7. Sato, H., Suzuki, K., Osanai, A., et al. (2006): *Paragonimus westermani* and some rare intestinal trematodes recovered from raccoon dogs (*Nyctereutes procyonoides viverrinus*) introduced recently on Yakushima Island, Japan. J. Vet. Med. Sci., 68, 681-687.
8. Nakagawa, K. (1917): Human pulmonary distomiasis caused by *Paragonimus westermani*. J. Exp. Med., 26, 297-323.

Genomes of Foodborne and Waterborne Pathogens
Edited by P. Fratamico, Y. Liu, and S. Kathariou

Chapter 23

Genomic and Postgenomic Approaches to Understanding the Pathogenesis of the Enteric Protozoan Parasite *Entamoeba histolytica*

KUMIKO NAKADA-TSUKUI AND TOMOYOSHI NOZAKI

EPIDEMIOLOGICAL OVERVIEW

Entamoeba histolytica is the causative agent of intestinal amebiasis and infects approximately 50 million people each year, causing 100,000 deaths annually (89). *E. histolytica* is a facultative pathogen that exhibits various infection outcomes, ranging from asymptomatic cases to more serious diseases. The most common manifestations of symptomatic amebic infections are dysentery and liver abscess, but infections of the lung, heart, and brain may also occur (55, 60). As with most diarrheal diseases, children in poor and developing countries suffer most from the morbidity and mortality of amebiasis. Cohort studies in these nations highlight the seriousness of amebiasis as a public health problem. In Dhaka, Bangladesh, where diarrhea is the leading cause of childhood death, 55% of preschool children studied prospectively had a new *E. histolytica* infection during the 2 years of follow-up. The annual incidence of amebic colitis is 2.2%, which is comparable to the rate of *Shigella* dysentery (56). In developed countries, amebiasis is most often found in men who have sex with men and in people who live in institutions for the mentally handicapped, as well as travelers and immigrants from areas where *Shigella* is endemic (76, 91, 112, 129). Furthermore, it was recently reported that the incidence of amebiasis among female commercial sex workers has increased and spread (91). The high amebiasis rates in people who engage in male homosexual or bisexual practices, and those who engage in the sex trade, suggest that amebiasis is a sexually transmitted disease in these countries.

LIFE CYCLE, SYMPTOMS, AND VIRULENCE MECHANISMS

E. histolytica has a simple life cycle consisting of two stages: a proliferative motile trophozoite form and an infective nonmotile dormant cyst stage. After ingestion of infectious cysts in food or water contaminated with fecal matter, excystation occurs in the lumen of the small intestine and then the resultant motile trophozoites colonize the colon. The trophozoites encyst and are excreted in the feces. Although the majority of infections remain asymptomatic, asymptomatic cyst carriers serve as the source of amebic infections in other individuals. About 10% of those infected develop amebic colitis characterized by subacute watery or bloody diarrhea with associated abdominal pains and weight loss (55, 60).

Several molecules that play fundamental roles in the virulence of this parasite have been identified. Among them are cell surface galactose/N-acetylgalactosamine-specific lectins, cysteine proteases (CPs), and amoebapores, which are involved in the interactions between the host cells and bacteria, the destruction of immune and nonimmune cells, and in the lysis of ingested microorganisms, respectively (46, 107, 150). Infections caused by *E. histolytica* trophozoites involve motility and phagocytosis, which are central to the pathogenicity of this parasite. Erythrophagocytosis is one of the diagnostic criteria used to differentiate *E. histolytica* from nonpathogenic species of amoeba such as *E. dispar*. Studies have shown that an *E. histolytica* cell line defective in phagocytosis exhibited attenuated virulence, suggesting the critical role of phagocytosis in disease causation (59, 96). Vesicular

Kumiko Nakada-Tsukui and Tomoyoshi Nozaki • Department of Parasitology, National Institute of Infectious Diseases, Tokyo 162-8640, Japan.

trafficking is also considered to play a pivotal role in the pathogenesis of *E. histolytica,* as it is closely linked to phagocytosis and the delivery of pathogenic factors (92). Elaborate regulation of membrane trafficking is essential for the targeting of molecules necessary for pathogenesis (e.g., lectin and CPs) to precise intracellular compartments such as the surface membrane and lysosomes. Regulation of such sophisticated trafficking is multifaceted and partially achieved by the well-studied Rab small GTPases. The diversification of CPs and Rab are discussed in this chapter, in the context of their diverse localization and function.

Encystation and excystation are fundamental processes required for the transmission and development of amebiasis; however, little is known about the molecular mechanisms involved in stage conversion. This is mainly due to the lack of in vitro encystation and excystation systems in *E. histolytica.* To overcome this problem, the related reptilian species *E. invadens* has been extensively used as a model organism to study the mechanisms behind encystation and excystation (43, 103).

MAJOR FINDINGS OF THE *E. HISTOLYTICA* GENOME

Overview of the *E. histolytica* Genome Project

The *E. histolytica* genome project was initiated in 2000 and continued at the Wellcome Trust Sanger Institute (United Kingdom), The Institute for Genomic Research, and the J. Craig Venter Institute (United States), with the support of the Wellcome Trust and the National Institute of Allergy and Infectious Diseases. The draft genome of the *E. histolytica* reference strain HM-1:IMSS (HM-1), isolated from a Mexican man with amebic dysentery in 1967, was published in 2005 (74).

The current genome assembly is approximately 23.7 million base pairs (Mbp) in size. Only the draft genome based on the assembly of the shotgun reads was published because of the highly repetitive nature and low GC content (24.1%) of the genome. The currently predicted genome size appears to be correct because this overall size is similar to the data predicted from pulse-field gel electrophoresis and reassociation kinetic experiments (49, 50, 148). Fourteen chromosomes ranging in sizes from 0.3 to 2.2 Mbp have been demonstrated by pulse-field gel electrophoresis, and the genome was predicted to be tetraploid (148). Neither a typical centromere nor telomeric repeats were found in the database; however, there is circumstantial evidence that the chromosome ends may contain arrays of transfer RNA (tRNA) genes (see section titled Insights Obtained from the Sequencing of Related *Entamoeba* Species).

Number of Genes and Gene Families

The current assembly predicts that the genome contains 8,160 genes (NCBI, GENBABK), almost 1.5-fold more than the number of genes in *Plasmodium falciparum* (5,268) or *Saccharomyces cerevisiae* (5,538), and close to that in *Dictyostelium discoideum* (12,500) (42, 48, 53). The average size of the protein coding genes is 1,167 bp, and approximately 49% of the genome consists of protein coding regions. The average length of the intergenic regions is 0.8 kb and the gene density in *E. histolytica* (1.9 kb/gene) is relatively high compared to other protists (e.g., 4.3 and 2.5 kb/gene in *P. falciparum* and *D. discoideum*, respectively). Approximately 32% of the predicted proteins showed no homology to proteins with known functions. The relatively large number of protein coding genes may reflect the complexity of *E. histolytica* biology and be due in part to the presence of large gene families (e.g., CPs and Rab small GTPases). Other expanded gene families appear to be associated with the ability of the parasite to sense and adapt to the environment within the human host and its ability to ingest and incorporate nutrients. A novel class of approximately 90 transmembrane serine/threonine kinases (TMKs), predicted to be involved in signal transduction, were discovered. These TMKs appear to be distributed in only a limited lineage of organisms such as plants, animals, and choanoflagellates (7, 74, 84). Conversely, the loss of genes is also evident as a consequence of parasitism, and this is most notable in the metabolic pathways such as amino acid and nucleic acid biosynthesis (9, 74).

Gene Structure and Size

The majority of *E. histolytica* genes consist of only a single exon, while as many as 25% of the genes appear to contain at least one intron and 6% contain two or more introns (28). This may indicate that mRNA splicing in *E. histolytica* is far less common compared to the related protist *D. discoideum* (42) or the malaria parasite *P. falciparum* (48). Genes in *E. histolytica* are surprisingly short because of the lack of introns, and the predicted lengths of the proteins they encode are also short. The average length of the predicted proteins in *E. histolytica* is 389 amino acids (aa), which is 129 aa and 372 aa shorter than those observed in *D. discoideum* and *P. falciparum,* respectively. Protein length distribution is most similar to that of the microsporidian *Encephalitozoon cuniculi,* which has a very compact genome of 3 Mbp and <2,000 genes. Since the average protein length is usually very well conserved among eukaryotes, the reason for the significantly shorter length of *E. histolytica* proteins is unclear. In bacteria, reduced average protein lengths

were suggested to reflect a reduced capacity for signaling (151); however, this does not appear to be the case for *E. histolytica* because it possesses an expanded TMK gene family that is assumed to be involved in intracellular signaling (7, 74, 84).

Common and Underrepresented Functional Domains

The three most common domains found in the *E. histolytica* proteome are the Trp-Asp (WD) domain, leucine-rich repeat domain, and protein kinase domain, which are also common in other organisms (28). The domains or domain-containing gene families that are unusually frequent in the genome may reflect the unusual aspects of this parasite. For instance, the Rab and Rho families, which are involved in the regulation of vesicle trafficking and cytoskeleton rearrangement, respectively, are among the most common domains in *E. histolytica*, while they are usually not among the 50 most common domains in other organisms.

Another notable feature of its domain organization is the presence of unusual proteins containing both Rho guanine nucleotide exchange factor and ADP ribosylation factor GTPase activating protein (Arf-GAP) domains, suggesting the direct interaction between the regulators of vesicle budding and cytoskeleton rearrangement (141). Some domains that are common in the majority of other organisms are observed to be rare or missing in *E. histolytica*. For example, most mitochondrial carrier domain proteins (66, 98) are not found in *E. histolytica*, and its mitochondrion-related organelle, the "mitosome," has evolved uniquely for its constituents and functions (85).

Transcriptional Machinery

Several unique features of the transcriptional machinery of *E. histolytica* were recently discovered. RNA polymerase II from *E. histolytica* is alpha-amanitin-resistant (73). The alpha-amanitin binding site in *E. histolytica* RNA polymerase II is highly divergent as observed in the alpha-amanitin resistant *Trichomonas vaginalis* (108). In addition, the carboxyl-terminal domain of the *E. histolytica* RNA polymerase II large subunit, which recruits various RNA processing/export and histone-modifying factors to the transcription complex and serves as a platform to couple mRNA metabolism and chromatin function to transcription (16, 26, 27, 102) is highly divergent.

The core promoter of *E. histolytica* has an unusual tripartite structure consisting of three conserved elements: a degenerated TATA box (GTATTTAAAG/C), a unique core promoter element GAAC (GAACT), and a putative initiator element (AAAAATTCA) (106, 123, 124, 125). It is predicted that there are unique DNA binding proteins for the preinitiation complex that recognizes these elements. *E. histolytica* possesses two TATA-binding-motif-containing proteins in addition to the TATA-binding protein (TBP), a subunit of the TFIID general transcription factor, which is required for the recognition of the core promoter (58). Based on these unique features of the *E. histolytica* transcription machinery, it is not surprising that from among the 14 evolutionally conserved subunits of TFIID, only 6 (TBP-associated factors 1, 5, 6, 10, 12, and 13) were conserved. Histone acetyltransferase (HAT) activity was previously reported in *E. histolytica*, and TBP-associated factors 5, 6, 10, and 12 are components of HAT complexes. However, not all of the components of the HAT complexes have been identified (109). Homologues of some of the other general transcription factors (TFII E, F, and H), but not the large and small subunits, have also been identified.

E. histolytica has homologues to approximately 80% of the *S. cerevisiae*-splicing machinery (61). *E. histolytica* mRNAs have short 5′ and 3′ untranslated regions and possess 5′ capping and a 3′ poly-A tail, as observed in other organisms (24, 70, 110, 140). However, only 8 of the 18 cleavage and polyadenylation specificity factor subunits in yeast have been identified in *E. histolytica*.

Translation

A general translation system is well conserved in *E. histolytica*, except that tRNA genes are present in various forms of arrays containing short tandem repeats (see the next section). All of the enzymes necessary for the tRNA splicing system, which differ from that of mRNA, were found as tRNA modification enzymes. Two tRNAs (Ile and Tyr) in *E. histolytica* are predicted to be spliced. The majority of ribosomal protein genes are well conserved, and only the gene for the large subunit protein L41 has not been identified. In eukaryotic translation systems, elongation factor (EF)-1 is activated upon GTP binding and forms a ternary complex, EF-1 α/β/γ or α/δ/γ, with aminoacyl tRNAs and ribosomes. *E. histolytica* lacks EF-1 δ, a protein involved in the exchange of GDP with the GTP of EF-1 α. In addition, eukaryotes typically have two polypeptide release factors, eRF1 and eRF3, both of which are present in *E. histolytica*.

UNEXPECTED FEATURES OF THE *E. HISTOLYTICA* GENOME

The *E. histolytica* genome presents a number of unexpected features (28, 74). This chapter will focus only on several important pathways closely related to

the parasite's molecular epidemiology, biology, and virulence. A more detailed description of its genome is available elsewhere (28).

tRNA Array

One of the most unusual structural features of the *E. histolytica* genome is the number and organization of tRNA genes (29). More than 10% of the shotgun sequence reads contained tRNA genes, which were (with a few exceptions) organized in linear arrays. Clark et al. identified 25 distinct arrays with unit sizes ranging from 0.5 to 1.8 kb (29). These arrayed tRNA genes were predicted to be functional because the 42 acceptor types were exclusively found in the arrays, but not elsewhere in the genome. Three of these arrays encode 5S RNA and one encodes what is thought to be a small nuclear RNA. It was estimated that there are about 4,500 tRNA genes in the genome. Southern blot analysis, using rare-cutting restriction endonucleases, suggested that the tRNA genes are located at the ends of the chromosomes and that these repeat units may perform a structural role (29, 148). As shown in *D. discoideum,* where it was demonstrated that ribosomal DNA functions as a telomere (42), the tRNA arrays in *E. histolytica* may also have the same function (28).

Regions containing tRNA are often linked with multiple short tandem repeats (STRs), and these STRs exhibit substantial variations between *E. histolytica* isolates. Recently, tRNA-linked STRs have been used extensively in the genotyping of isolates (5). Ali et al. reported that the genotypes of the isolates derived from amebic liver abscess, diarrhea/dysentery, and asymptomatic cases were different (2, 3), suggesting the presence of a link between the tRNA-linked STR pattern and the outcome of infection.

Expansion of Cysteine Proteases

E. histolytica shows an extraordinary capacity to invade and destroy human tissues, and its main lytic activity has been attributed to CPs. CPs are also important virulence factors of various infectious agents and are the main proteolytic enzymes in many protozoon parasites (83, 117). The importance of CPs in the pathogenicity of *E. histolytica* has been demonstrated in several in vivo and in vitro studies (10, 47, 62, 71, 77, 78, 111, 119, 127). For example, CP5 overexpression led to an increase in cytopathic activity, as measured by in vitro monolayer disruption, and a significant increase in amebic liver abscess formation was also observed in laboratory animals (135).

Homology searches using conserved active site regions revealed that the *E. histolytica* genome contains approximately 50 genes encoding CPs (28, 134). The majority of the CPs are structurally related to the C1 papain superfamily, while the others are more similar to the C2 (calpain-like CPs), C19 (ubiquitinyl hydrolase), C48 (Ulp1 peptidase), C54 (autophagin), and C65 (otubain) families. Phylogenetic analysis of the 37 C1-family members revealed that they represent three distinct clades (A, B, and C), consisting of 13, 11, and 13 members, respectively. Members of clades A and B correspond to two previously described subfamilies, designated as EhCP-A and EhCP-B, respectively (23). EhCP-A and EhCP-B subfamily members are classical pre- and pro-enzymes with an overall cathepsin L-like structure (15), as indicated by the presence of an ERFNIN motif in the pro region of at least 21 of the 24 EhCP-A and EhCP-B enzymes. Interestingly, biochemical studies with purified EhCP-A indicated a cathepsin B-like substrate specificity (119). This is likely to be due to the postulated S2 pocket that corresponds to residue 205 in papain (15). The EhCP-A and EhCP-B subfamilies differ in the length of their pro regions and catalytic domains, and have distinct sequence motifs in the amino-terminal regions of the mature enzymes (DWP vs. PCPN) (23).

Conversely, clade C (EhCP-C family) represents a new group. EhCP-C family members are not pre-pro enzymes, and these enzymes lack a hydrophobic signal sequence and an identifiable pro region. Instead, the enzymes have a hydrophobic region located 11 to 28 aa from the amino terminus. This region is predicted to form transmembrane helices, suggesting that these enzymes could be membrane associated. As no homologue of the EhCP-C subfamily has been found in other organisms, the specific roles of this group of CPs remain unknown.

Family C2 is a group of calpain-like peptidases. It contains several calcium-binding domains, and the enzymes in this family participate in various cellular processes, including remodeling of the cytoskeleton and membranes, signal transduction pathways, and apoptosis (65, 67, 101).

The members of family C54 are called autophagins (EhAUTO1–4). EhAUTO members show significant homology to Atg4, which is involved in the posttranslational modification of Atg8 in other organisms. Since Atg8, Atg3, and Atg7 are conserved in *E. histolytica,* it is conceivable that EhAUTO1–4 are involved in autophagy in *E. histolytica* (103, 104). Autophagy is a mechanism for the degradation of intracellular proteins and the removal of damaged organelles (64, 149).

The CPs that belong to families C19, C48, and C65 are known to be involved in the degradation of ubiquitin or the small ubiquitin-like modifier protein (SUMO). Family C19 consists of CPs that are ubiquitin-specific in humans (11, 12). Ulp1, which is synonymous to C48, is a member of a family of peptidases that controls the function of SUMO (126). CPs in family

C65 are called otubains, and these enzymes exhibit isopeptidase activity that releases ubiquitin or SUMO from polyubiquitin or poly-SUMO (14, 18, 44).

Expansion and Diversification of Rab Small GTPases

Small GTP-binding proteins are ubiquitous molecular switches found in all eukaryotes. These proteins are involved in various important cellular processes including cell proliferation, cytoskeletal assembly, and intracellular membrane trafficking. There are more than 100 proteins in the Raslike superfamily. Based on their structure and primary sequence, the members of this superfamily are classified into five families: Ras, Rho/Rac, Rab, Sar/Arf, and Ran (19, 130). Rab small GTPases constitute the largest group of this superfamily and are essential regulators of vesicular transport pathways (90). Compared to unicellular eukaryotes, multicellular organisms are expected to have a more complex set of these regulators. *Homo sapiens, Arabidopsis thaliana,* and *Drosophila melanogaster,* for example, have 60, 29, and 29 Rab genes, respectively, while *S. cerevisiae* has only 11 (100). Interestingly, the *E. histolytica* genome contains more than 90 Rab genes (115). About 75% of the genes are unique to *E. histolytica,* while close to 25%, including the genes for Rab 1, 2, 5, 7, 8, 11, and 21 subfamilies, have homology higher than 40% with known Rab families found in humans or yeast. Approximately 70% (64 genes) of amebic Rab genes contain introns, and 23% (22 genes) of amebic Rab genes contain 2–4 introns. Considering that only 19% and 6% of the *E. histolytica* genes have a single or multiple introns, respectively (28, 74), and that the average size of amebic Rab genes is 702 bp, which is shorter than the average size of all genes, amebic Rab genes are extremely intron rich.

Several reports demonstrated a direct or indirect role of individual Rab small GTPases in *E. histolytica.* Among the multiple EhRab7 isotypes, EhRab7A is involved in CP transport, most likely trafficking to the lysosome, while EhRab11B is involved in CP secretion (87, 88). EhRab5 and EhRab7A coordinately regulate the formation and maturation of the prephagosomal vacuole, a unique organelle in *E. histolytica* that is formed during phagocytosis, and is likely involved in the processing, activation, or storage of hydrolases that are transported to the phagosome (116). *E. histolytica*-specific EhRabA was initially suggested to be involved in its motility and polarization rather than in membrane trafficking (147). Recently, however, it has been shown to be involved in the transport of Gal/GalNac specific lectin (146).

In general, Rab small GTPases have unique functions via their specific interactions with effector molecules. Interestingly, the majority of the Rab effectors found in other organisms, such as Rabaptin-5 for Rab5 or Rabring7 for Rab7, are not conserved in *E. histolytica.* Instead, it utilizes a unique EhRab7A-binding partner called the "retromer" complex, which apparently consists of three components: EhVps26, EhVps29, and EhVps35. EhRab7A binds to EhVps26 directly via its unique carboxyl terminus and coordinately regulates CP transport (88). There is no precedent for the interaction between Rab7 and the retromer complex. Recently, it has been reported that Rab7 interacts with the retromer complex via Vps35 in mammals (113). It is of interest to investigate if other effectors are conserved in both lower and higher eukaryotes, considering that membrane trafficking appears to be highly divergent among eukaryotes.

INSIGHTS OBTAINED FROM THE SEQUENCING OF RELATED *ENTAMOEBA* SPECIES

Whole genome sequencing projects of other related *Entamoeba* species are currently under way. In this section, we focus on the insights obtained from comparing their genome information with that of *E. histolytica* to understand the conservation and unique evolution of the *Entamoeba* species. As representative cases, we only discuss the analysis of repetitive elements, CPs, and Rabs.

Other *Entamoeba* Species and Genome Projects

Whole genome sequencing data of *E. dispar, E. moshkovskii, E. invadens,* and *E. terrapinae* are partially available. *E. dispar* is a nonpathogenic sibling of *E. histolytica,* first reported in 1925, and often used to verify whether potential virulence factors discovered in *E. histolytica* are indeed involved in pathogenesis. *E. histolytica* and *E. dispar* are considered to be closely related species on the basis of extensive genetic, immunological, and biochemical analyses (37, 128, 132). *E. moshkovskii* was originally isolated from sewage samples in Moscow and was thought to be a free-living environmental strain (139); however, recent studies have shown the occasional detection of *E. moshkovskii* in humans (6, 32, 45, 57, 99, 131). *E. moshkovskii* trophozoites grow at room temperature, are osmotolerant, and resistant to emetine. These characteristics clearly distinguish *E. moshkovskii* from *E. histolytica* and *E. dispar* (30, 31). On the other hand, *E. invadens* and *E. terrapinae* are pathogenic and commensal parasites of reptiles (51) and do not cause disease in humans. *E. invadens* remains an important model for encystation and excystation because an in vitro

system to induce encystation in *E. histolytica* has not yet been established (43). The genome databases for *E. dispar, E. invadens, and E. histolytica* are available at http://pathema.jcvi.org/cgi-bin/Entamoeba/PathemaHomePage.cgi. The shotgun reads of *E. moshkovskii* and *E. terrapinae* are available at http://www.sanger.ac.uk/Projects/Protozoa/.

COMPARATIVE GENOMICS OF DIFFERENT *ENTAMOEBA* SPECIES

Comparative Genomics of Repetitive Element Among *E. histolytica, E. dispar, and E. invadens*

The *Entamoeba* genome is littered with transposable elements (TEs). TEs play an important role in nuclear architecture, genome stability, gene amplification, and altered gene regulation (13, 105, 121, 122). TEs are conventionally classified into two broad classes, I and II. Class I elements (or retrotransposons) are mobilized through an RNA intermediate, while class II elements (or DNA transposons) transpose directly via a DNA intermediate. Class I elements have two subclasses, the non-long terminal repeat (LTR) retrotransposons, which include long and short interspersed elements (LINEs and SINEs), and the LTR retrotransposons. There are three well-separated LINEs found in *E. histolytica* (EhLINE1–3) and *E. dispar* (EdLINE1–3), whereas *E. invadens* has only one EiLINE (13, 28, 75). Eh/EdLINE1–3 show 70% to 86% mutual identity (75, 121), suggesting that they may have emerged from the common ancestor of *E. histolytica* and *E. dispar* after the speciation of *E. invadens*. Phylogenetic analysis of the reverse transcriptase consensus sequence of LINEs from these three species also supported this scenario (75). After the speciation of *E. invadens*, a single LINE most likely duplicated giving rise to the two separate lineages of Eh/EdLINE1, Eh/EdLINE2, and Eh/EdLINE3 in the ancestral organism of *E. histolytica* and *E. dispar*. The fact that *E. invadens* predominantly possesses class II transposons, while *E. histolytica* and *E. dispar* genomes are rich in class I repeats also supports the difference in the evolution of transposons between these species (105).

Three SINEs were found in *E. histolytica* (EhSINE1–3) and *E. dispar* (EdSINE1–3), but not in *E. invadens* (13, 75). Shire and Ackers identified EdSINE1 as a homologue of EhSINE3. Although the origin of EdSINE1/EhSINE3 is not clear, dot plot alignments and phylogenic analysis of *E. dispar* SINE1 and *E. histolytica* SINE3 (122) indicated that these SINEs originated as a chimeric element in the common ancestor of *E. histolytica* and *E.dispar* (75).

LINEs and SINEs affect the human genome via multiple mechanisms such as spreading in the genome and shuffling the sequence via transduction (97). They affect gene expression by providing alternative promoters, splicing and polyadenylation sites, and by heterochromatinization (97). Microarray and Northern blot analyses show that the expression of LINEs and SINEs was altered in *E. dispar* and an *E. histolytica* nonvirulent variant, the Rahman strain (79). The expression levels of LINE1 and LINE3 were significantly lower in *E. dispar* and the *E. histolytica* Rahman strain than in the highly virulent *E. histolytica* HM-1: IMSS strain. Additionally, two SINEs were not transcribed in *E. dispar* (79). Notably, while *E. histolytica* isolates showed variations in EhLINE expression, virulent *E. histolytica* strains always exhibited higher expression levels of EhLINE1 and EhLINE3 than nonvirulent species and strains. SINEs also function as stress sensors in the silkworm (63), but it remains to be determined whether the *Entamoeba* SINEs are involved in this process.

E. histolytica possesses two class II TEs (EMULE and Hydargos) and two novel TEs (EhERE1 and 2) that do not belong to class I or class II (75). There are no homologues of EhERE2 in *E. dispar* and *E. invadens*. Since DNA repeats mediate genomic rearrangements in prokaryotic organisms, which alter the expression of virulence-associated genes (54), it is tempting to speculate that the acquisition and expansion of elements such as EhERE2 may play a role in the acquisition of pathogenicity traits in *E. histolytica*. For instance, such elements may promote genomic rearrangements or alter the function or expression of genes involved in processes such as cell attachment and evasion of the host immune response. The leucine-rich/BpsA-like gene family, for instance, which encodes for potential plasma membrane receptors and may interact with the host's fibronectin molecules (36), is frequently associated with TEs (41 out of 114 of these genes) (13, 75). This observation suggests the possible contribution of TEs to the pathogenicity of *E. histolytica*.

Comparative Genomics of Cysteine Proteases in *E. histolytica* and *E. invadens*

CPs are one of the most important class of determinants of virulence in *E. histolytica*. To understand if the repertoire of CPs is conserved among *Entamoeba* species, the *E. invadens* genome database was searched for CP genes that belong to the C1 family (134; Escueta-de Cadiz et al., unpublished data). All of the *E. invadens* C1 family CPs (EiCPs) were categorized into A, B, and C subfamilies based on their conserved domains (ERFNIN and DWR for CP-A, ERFNIN and PCNC for CP-B, and HSICP for CP-C) and also by phylogenetic analysis (Fig. 1). *E. invadens* has 19 CP-A, 17 CP-B, and 10 CP-C genes (Table 1). Some of the CP-A subfamilies have extensively diverged in AQ1

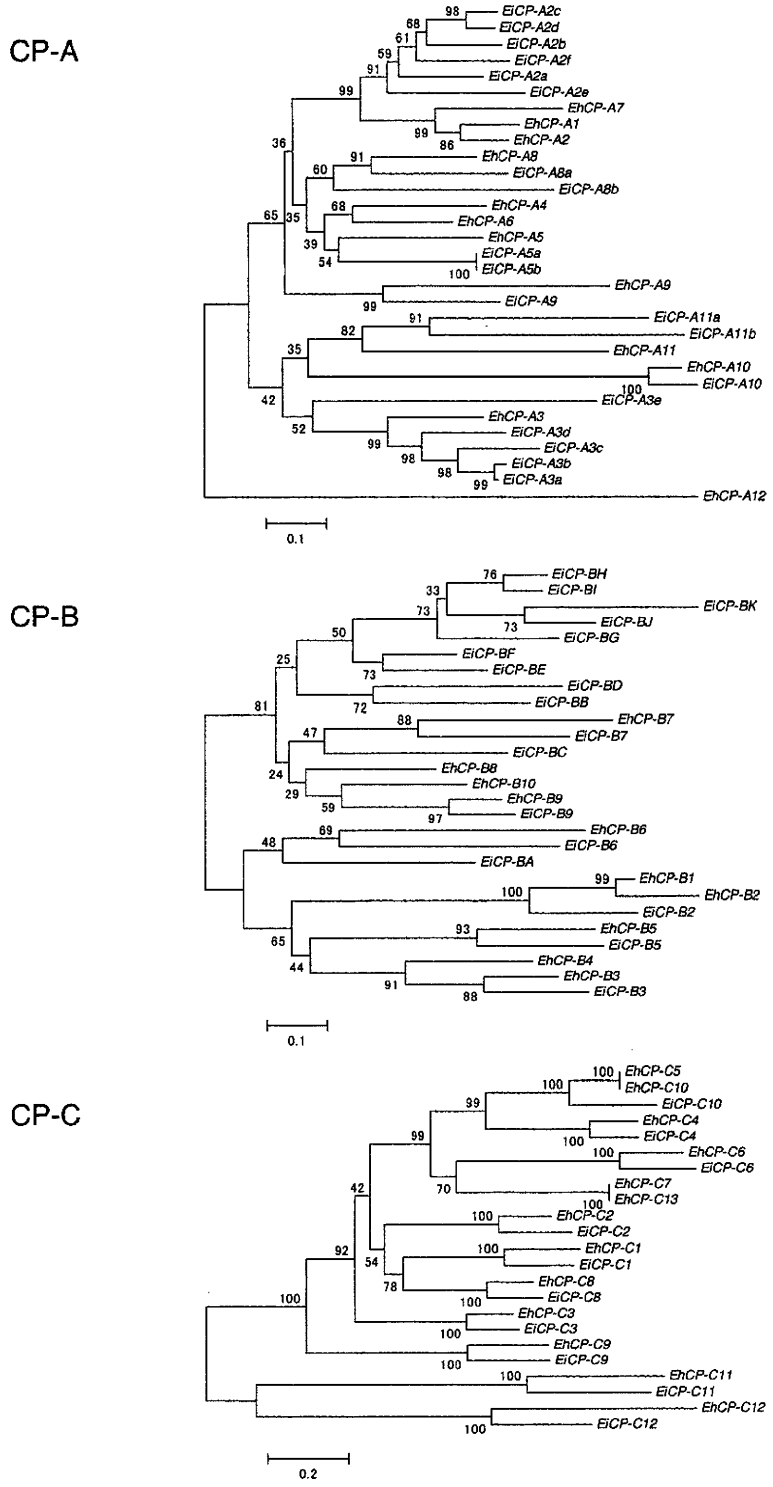

Figure 1. Phylogenetic analysis of CPs from *E. histolytica* and *E. invadens*. Phylogenetic analysis of CPs from the C1 family of *E. histolytica* and *E. invadens* was performed using CLUSTAL W. Trees were drawn using MEGA4. The consensus phylogenetic trees of the CP-A, CP-B, and CP-C families are shown. The numbers at the nodes represent the bootstrap values for 1,000 iterations shown in percentages. The scale bar indicates 0.1 or 0.2 substitutions at each amino acid position.

Table 1. CPs from *E. histolytica* and *E. invadens*[a]

E. histolytica		*E. invadens*		%identity to	Protein	Active	Conserved motifs		
EhCP	Accession No.	EiCP	Accession No.	Eh homologue	length	site			
CP-A							ERFNIN	DWR	RGD
EhCP-A1	XP_650156								
EhCP-A2	XP_650642	EiCP-A2a	EIN_036980	65(A2), 61(A1), 57(A7)	313	QCHN	ERFNVN	DLR	-
		EiCP-A2b	EIN_172350	60(A2), 55(A1), 54(A7)	319	QCHN	ERFNVN	DLR	-
		EiCP-A2c	EIN_168460	57(A2), 54(A1), 50(A7)	315	QCHN	ERFNVN	DLR	-
		EiCP-A2d	EIN_151430	57(A2), 53(A1), 51A7)	315	QCHN	ERFNVN	DLR	-
		EiCP-A2e	EIN_013830	55(A2), 53(A1), 51(A7)	304	QCHN	ERFNVN	DLR	-
		EiCP-A2f	EIN_105250	58(A2), 55(A1), 51(A7)	247	QCHN	-	DLR	-
EhCP-A3	XP_653254	EiCP-A3a	EIN_253190	70(A3)	306	QCHN	ERYNVT	DWR	-
		EiCP-A3b	EIN_085960	69(A3)	306	QCHN	ERYNVT	DWR	-
		EiCP-A3c	EIN_253450	65(A3)	306	QCHN	ERYNVT	DWR	-
		EiCP-A3d	EIN_120520	65(A3)	306	QCHN	ERFNVN	DWR	-
		EiCP-A3e	EIN_192250	41(A3)	1157	Q-HN	ERFNVQ	DYT	-
EhCP-A4	XP_656602								
EhCP-A5	XP_650937	EiCP-A5a	EIN_118870	63(A5), 54(A4), 55(A6)	314	QCHN	ERFNVN	DWR	-
		EiCP-A5b	EIN_184910	61(A5), 58(A4), 57(A6)	214	—HN	ERFNVN	-	-
EhCP-A6	XP_657364								
EhCP-A7	XP_648996								
EhCP-A8	XP_657446	EiCP-A8a	EIN_036350	57(A8)	319	QCHN	ERFNVN	DWR	-
		EiCP-A8b	EIN_224330	46(A8)	322	QCHN	ERFNVN	DWR	-
EhCP-A9	XP_655675	EiCP-A9	EIN_038430	42(A9)	323	QCHN	ERFNVN	DWR	-
EhCP-A10	XP_651147	EiCP-A10	EIN_241400	75(A10)	421	QCHN	ERFSIN	DFR	-
EhCP-A11	XP_651690	EiCP-A11a	EIN_218930	40(A11)	318	QCHN	ERFNVD	DMR	-
		EiCP-A11b	EIN_190570	36(A11)	328	QCHN	ERYNVS	DLR	-
EhCP-A12	XP_653823								
EhCP-A13	not annotated								
CP-B							ERFNIN	PCNC	RGD
EhCP-B1	XP_651581								
EhCP-B2	AAO03568	EiCP-B2	EIN_052180	40(B2)	399	QCHN	ERFNFN	PCNC	-
EhCP-B3	XP_656747	EiCP-B3	EIN_051660	51(B3)	493	QCHN	EKFNIN	PCNC	-
EhCP-B4	XP_648501								
EhCP-B5	XP_652671	EiCP-B5	EIN_176790	42(B3)	453	QCHN	ERFKIN	PCNC	-
EhCP-B6	XP_652465	EiCP-B6	EIN_292720	46(B6)	333	QCHN	-	PCNC	-
EhCP-B7	XP_650400	EiCP-B7	EIN_241340	41(B7)	733	-C-N	ERFSYN	PCNC	-
EhCP-B8	XP_651049								
EhCP-B9	XP_652993	EiCP-B9	EIN_152250	55(B9)	307	QC-N	ERFNIN	PCNC	-
EhCP-B10	XP_648306								
EhCP-B11	XP_648013								
		EiCP-BA	EIN_184830		365	QCHN	ERFNIN	PCNC	-
		EiCP-BB	EIN_199850		488	QCHN	ERFNIN	PCNC	-
		EiCP-BC	EIN_277820		506	QCHN	ERFEIN	PCNC	RGD
		EiCP-BD	EIN_050320		494	QCHN	NRFSIS	PCNC	-
		EiCP-BE	EIN_114990		644	QCHN	ERFTIN	PCNC	RGD
		EiCP-BF	EIN_103270		662	QCHN	QKFSIN	PCNC	RGD
		EiCP-BG	EIN_315880		762	QCHN	ERFKCN	PCNC	RGD
		EiCP-BH	EIN_140010		1006	QCHN	QRFSVN	PCNC	RGD
		EiCP-BI	EIN_245400		1041	QCHN	QRFSIN	PCNC	RGD
		EiCP-BJ	EIN_210450		187	—HN	-	-	RGD
		EiCP-BK	EIN_162840		149	—HN	-	-	-
CP-C							HSICP		RGD
EhCP-C1	XP_654453	EiCP-C1	EIN_161500	60(C1)	569	QCHN	HSICP		-
EhCP-C2	XP_656632	EiCP-C2	EIN_112080	68(C2)	566	QCHN	HSICP		RGD
EhCP-C3	XP_655128	EiCP-C3	EIN_135460	75(C3)	582	ECRN	HSLCP		-
EhCP-C4	XP_655800	EiCP-C4	EIN_108920	69(C4)	436	QCH-	-		-
EhCP-C5	XP_654800								
EhCP-C6	XP_651553	EiCP-C6	EIN_051150	64(C6)	552	QCH-	HSLCP		-
EhCP-C7	XP_657273								
EhCP-C8	XP_655479	EiCP-C8	EIN_039050	64(C8)	642	QCHN	HSICP		RGD
EhCP-C9	XP_655011	EiCP-C9	EIN_110280	68(C9)	535	QSHN	HSICP		-

Continued on following page

Table 1. *Continued*

E. histolytica EhCP	Accession No.	*E. invadens* EiCP	Accession No.	%identity to Eh homologue	Protein length	Active site	Conserved motifs	
CP-C							HSICP	RGD
EhCP-C10	XP_654829	EiCP-C10	EIN_240210	60(C10), 59.9(C5)	560	QCHN	HSLCP	-
EhCP-C11	XP_648083	EiCP-C11	EIN_248160	38(C10)	541	-	VSRCF	-
EhCP-C12	XP_650829	EiCP-C12	EIN_109530	39(C11)	497	-	ISYCG	-
EhCP-C13	XP_656556							

[a]The *E. invadens* genome database (http://pathema.jcvi.org/cgi-bin/Entamoeba/PathemaHomePage.cgi) was searched using *E. histolytica* CPs as query sequences. All potential *E. invadens* CPs were retrieved and analyzed with all *E. histolytica* CPs (Fig. 1). *E. invadens* CPs were annotated based on phylogenetic inferences and their percentage identity to the corresponding *E. histolytica* homologues. For the annotation of *E. histolytica* CPs, see (133).

E. invadens. EiCP-2A and EiCP-A3 consist of six and five members, respectively, while EiCP-A5, A8, and A11 have two isotypes each. EhCP-A1, A2, and A7 belong to a single well-supported clade, as inferred from phylogeny. EhCP-A4, A5, and A6 also show monophyly. While all of the CP-A family proteins present in *E. histolytica*, except for CP-A12 and A13, are conserved in *E. invadens*, a group of members that belong to CP-B and CP-C, CP-B1, B4, B8, B10, B11, C5, C7, and C13, are missing in *E. invadens*. Of the 11 CP-B members found in *E. histolytica*, 5 are not conserved in *E. invadens*, but it has instead 11 additional species-specific CP-B members, EiCP-BA to BK. The CP-C family is an *E. histolytica*-specific CP family uniquely found in this organism (28, 134). Of 13 *E. histolytica* CP-C family members, 10 are conserved in *E. invadens*. These data are consistent with the premise that the CP-A and CP-C members are involved in shared and housekeeping roles in *Entamoeba*, while CP-B members, particularly EiCP-BA to BK may have unique species-specific roles in *E. invadens*. It is tempting to speculate that some CP-B members may be involved in host-specificity.

Comparative Genomics of *E. histolytica* and *E. invadens* Rab Small GTPases

To understand the conservation and/or unique evolution of vesicular trafficking between *Entamoeba* species, the *E. invadens* genome database was thoroughly searched for Rab small GTPases. The *E. invadens* genome encodes 85 putative Rab genes, comparable to the number in *E. histolytica*. Putative *E. invadens* Rab genes were grouped based on their similarity to their *E. histolytica* homologues and phylogenetic inferences (Fig. 2 and Table 2). Similar to *E. histolytica*, the Rab1/8, 2, 5, 7, and 11 families were conserved in *E. invadens*, with the exception of Rab2C. In addition, Rab21 is not conserved in *E. invadens*. Among the Rab genes that were exclusively found in *E. histolytica*, but missing in other organisms (EhRabA–P and EhRabX1–36), 24 Rab genes including EhRabC6, D2, I2, M2, P2, X4~9, X13, X15, X18, X20, X21, X24, X26, X27, X28, X32, X33, X35, and X36 are not conserved in *E. invadens*. Conversely, some Rab subfamilies including EhRab5, 7G, C3, X11, X17, X22, X31, and X34 have expanded and consist of two to four members.

Seven EiRabs (EiRabZ1–7) that show low similarity to *E. histolytica* Rab genes were considered to have uniquely evolved in *E. invadens*. EiRabD1 is unusually long (744 aa) and is predicted to have an extension at its amino terminus. Only 54% of the *E. histolytica*-specific solitary Rab genes (EhRabA, B, H, and X1 to X36) that show low (< 40%) mutual identity to the Rab genes from human, yeast, and other Rab members of *E. histolytica* are conserved in *E. invadens*. Conversely, the majority (84%) of *E. histolytica* Rab genes that form subfamilies are conserved, suggesting shared house-keeping roles for these Rab subfamilies in *Entamoeba* species. Phagosome-associated Rab genes in *E. histolytica* that were previously demonstrated by proteomics (EhRab1, 5, 7A–E, 8, 11B–D, B, C13, and X17) are conserved in *E. invadens* (81, 94, 95). This suggests that the molecular mechanisms of vesicular trafficking involved in phagocytosis are conserved between these two *Entamoeba* species.

APPLICATIONS

After the *E. histolytica* genome was determined in 2005, a number of studies exploiting postgenomic "omics" approaches have been reported. These studies included comparisons of the transcriptomes between virulent *E. histolytica* strains, which included clinical isolates, and attenuated or avirulent *E. histolytica* strains. Transcriptional analyses of trophozoites derived from axenic cultures, animal intestines, and liver abscesses, as well as the heat shock responses in *E. histolytica*, have also been reported (34, 41, 52, 79, 80, 120, 134, 143, 144, 145). Furthermore, a number of proteomic studies have been published,

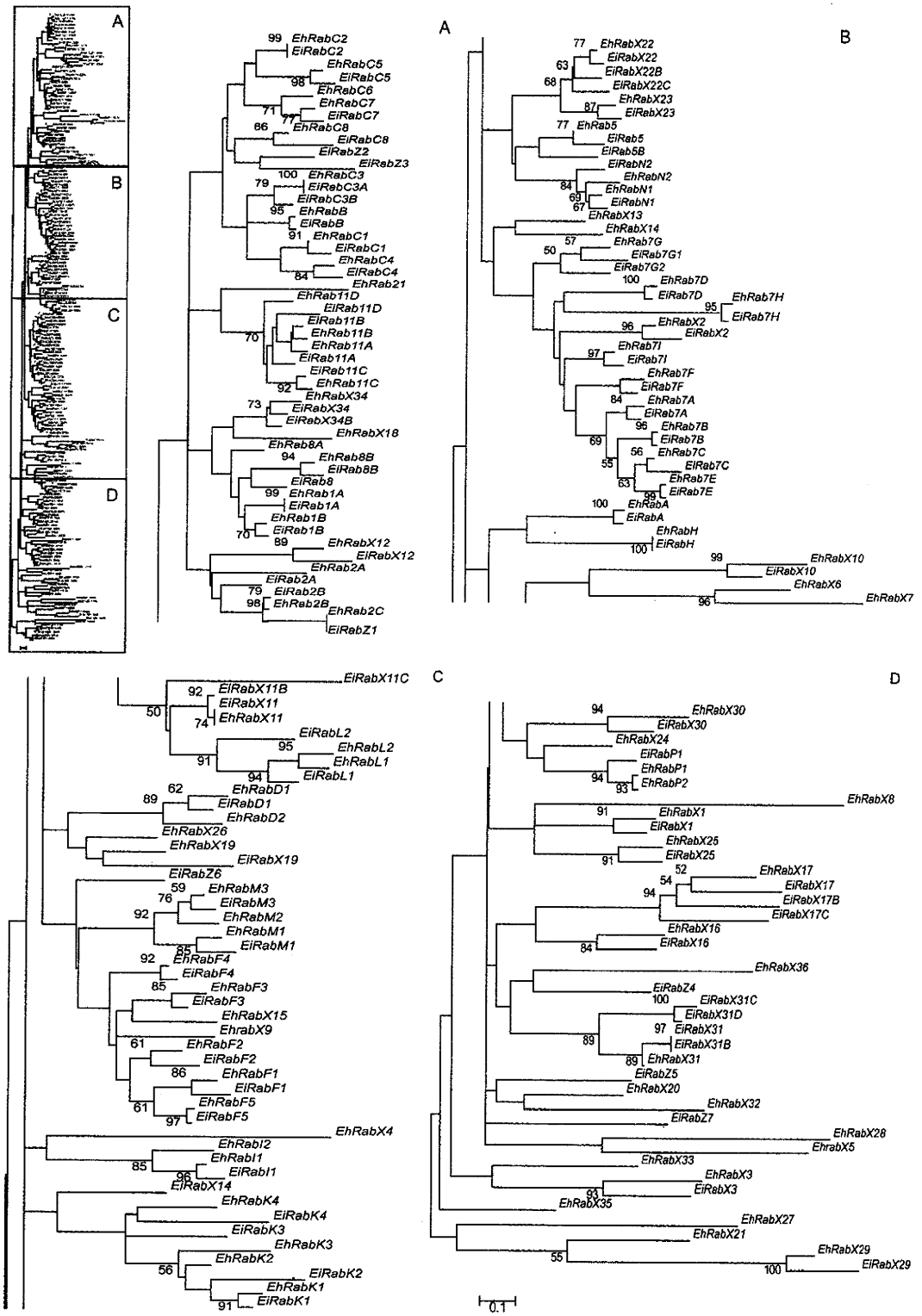

Figure 2. Phylogenetic analysis of Rab genes from *E. histolytica* and E. *invadens*. Phylogenetic analysis was performed as described in the legend of Fig. 1. An overview of the whole phylogenetic tree is shown on the left, while magnified portions (A–D) of the tree are shown separately on the right.

Table 2. Rab genes from *E. histolytica* and *E. invadens*[a]

E. histolytica		*E. invadens*		% Identity to	Protein	Missing regions	C-terminal
EhRab	Accession No.	EiRab	Accession No.	Eh homologue	length		peptides
EhRab1A	XP_651336	EiRab1A	EIN_104660	83	207		CXXX
EhRab1B	XP_649033	EiRab1B	EIN_033160	75	208		CXCX
EhRab2A	XP_649924	EiRab2A	EIN_200080	57	217		-
EhRab2B	XP_649335	EiRab2B	EIN_277910	84	211		-
EhRab2C	XP_656786						
EhRab5	XP_655377	EiRab5	EIN_277000	83	194		XXCC
		EiRab5B	EIN_051650	50	188		XCCC
EhRab7A	XP_649196	EiRab7A	EIN_112310	91	205		XXCC
EhRab7B	XP_656820	EiRab7B	EIN_202680	88	206		XXCC
EhRab7C	XP_652334	EiRab7C	EIN_094230	68	213		XXCC
EhRab7D	XP_651915	EiRab7D	EIN_133760	80	200		XXCC
EhRab7E	XP_651202	EiRab7E	EIN_148860	87	206		XXCC
EhRab7F	XP_650338	EiRab7F	EIN_235870	65	206		XXCC
EhRab7G	XP_656477	EiRab7G1	EIN_015310	62	197		XXCC
		EiRab7G2	EIN_299020	61	193		XXCC
EhRab7H	XP_653414	EiRab7H	EIN_288760	46	203		XXCC
EhRab7I	XP_649308	EiRab7I	EIN_196420	66	206		XCCX
EhRab8	XP_653051	EiRab8	EIN_252640	48	202		XXCC
EhRab8B	XP_652309	EiRab8B	EIN_149010	67	207		XXCC
EhRab11A	XP_647948	EiRab11A	EIN_165880	82	208		XXCC
EhRab11B	XP_652776	EiRab11B	EIN_108540	74	213		XXCC
EhRab11C	XP_649609	EiRab11C	EIN_014150	71	222		XXCC
EhRab11D	XP_652598	EiRab11D	EIN_050950	65	211		XXCC
EhRab21	XP_651927						
EhRabA	XP_652258	EiRabA	EIN_280980	85	198		XXXC
EhRabB	XP_652994	EiRabB	EIN_110330	69	198		XXCC
EhRabC1	XP_656355	EiRabC1	EIN_058650	81	198		XXCC
EhRabC2	XP_653593	EiRabC2	EIN_094390	81	206		XXCC
EhRabC3	XP_652352	EiRabC3A	EIN_093190	89	207		XXCC
		EiRabC3B	EIN_089560	54	213		XXCC
EhRabC4	XP_656897	EiRabC4	EIN_072390	57	200		XXCC
EhRabC5	XP_654231	EiRabC5	EIN_014860	73	204		XXCC
EhRabC6	XP_654710						
EhRabC7	XP_652882	EiRabC7	EIN_217920	70	201		XXCC
EhRabC8	XP_651035	EiRabC8	EIN_168870	61	190		XXCC
EhRabD1	XP_652887	EiRabD1	EIN_184580	61	744(190)		XXCC
EhRabD2	XP_655208						
EhRabF1	XP_651799	EiRabF1	EIN_239370	68	201		CXXX
EhRabF2	XP_651513	EiRabF2	EIN_106040	61	192		XXCC
EhRabF3	XP_655210	EiRabF3	EIN_112540	70	188		XXCC
EhRabF4	XP_654217	EiRabF4	EIN_239300	65	192		XXCC
EhRabF5	XP_656060	EiRabF5	EIN_228360	85	196		XXXC
EhRabH	XP_657074	EiRabH	EIN_156660	78	217		XXCC
EhRabI1	XP_655925	EiRabI1	EIN_014210	65	204		XXCC
EhRabI2	XP_654235						
EhRabK1	XP_652298	EiRabK1	EIN_311490	81	210		XXCC
EhRabK2	XP_649362	EiRabK2	EIN_243000	65	200		XXCC
EhRabK3	XP_651827	EiRabK3	EIN_028700	37	226		XXCC
EhRabK4	XP_648284	EiRabK4	EIN_059050	40	227		XXCC
EhRabL1	XP_651210	EiRabL1	EIN_081690	55	202	swithch I	XXCC
EhRabL2	XP_648952	EiRabL2	EIN_061000	52	210		XCCX
EhRabM1	XP_652833	EiRabM1	EIN_054310	73	213		XXCC
EhRabM2	XP_652253						
EhRabM3	XP_651723	EiRabM3	EIN_052610	79	174	switch I	XXCC
EhRabN1	XP_652702	EiRabN1	EIN_136950	71	201		CXXX
EhRabN2	XP_649426	EiRabN2	EIN_054510	68	199		XXXC
EhRabP1	XP_651771	EiRabP1	EIN_298880	59	205	switch I	CXXX
EhRabP2	XP_656067						
EhRabX1	XP_650791	EiRabX1	EIN_299880	49	167		-
EhRabX2	XP_650041	EiRabX2	EIN_134920	57	193		-
EhRabX3	XP_655050	EiRabX3	EIN_135850	47	340		-
EhRabX4	XP_651716						

Continued on following page

Table 2. *Continued*

E. histolytica EhRab	*E. histolytica* Accession No.	*E. invadens* EiRab	*E. invadens* Accession No.	% Identity to Eh homologue	Protein length	Missing regions	C-terminal peptides
EhRabX5	XP_657472						
EhRabX6	XP_652555						
EhRabX7	XP_654103						
EhRabX8	XP_649911						
EhRabX9	XP_654197						
EhRabX10	XP_656920	EiRabX10	EIN_160760	45	193	switch I	CXXX
EhRabX11	XP_655922	EiRabX11	EIN_014910	84	209		XXCC
		EiRabX11B	EIN_059070	77	209		XXCC
		EiRabX11C	EIN_092300	44	205		-
EhRabX12	BAD82860	EiRabX12	EIN_252860	48	209	switch I	XXCC
EhRabX13	XP_653656						
EhRabX14	XP_657126	EiRabX14	EIN_260390	36	228		XXCC
EhRabX15	XP_649287						
EhRabX16	XP_655529	EiRabX16	EIN_051300	47	214		XXCC
EhRabX17	XP_656536	EiRabX17	EIN_072650	68	197		CXXX
		EiRabX17B	EIN_212470	57	197	switch I	CXXX
		EiRabX17C	EIN_107380	51	208	switch I	CXXX
EhRabX18	XP_649285						
EhRabX19	XP_654968	EiRabX19	EIN_288570	46	226	box 2	XXCC
EhRabX20	XP_652547						
EhRabX21	XP_650747						
EhRabX22	XP_655103	EiRabX22	EIN_134610	68	197		XXCC
		E9RabX22B	EIN_075970	65	184	switch I	XXCC
		EiRabX22C	EIN_136400	62	197		XXCC
EhRabX23	XP_650671	EiRabX23	EIN_241550	76	202		XXCC
EhRabX24	XP_656866						
EhRabX25	XP_653064	EiRabX25	EIN_243640	64	189		XXCC
EhRabX26	XP_657341						
EhRabX27	XP_650814						
EhRabX28	XP_647919						
EhRabX29	XP_656310	EiRabX29	EIN_051680	64	198	box 2, switch I	CXXX
EhRabX30	XP_651921	EiRabX30	EIN_148610	54	210		CXXX
EhRabX31	XP_648905	EiRabX31	EIN_127810	70	189		XXCC
		EiRabX31B	EIN_262240	62	191		XXCC
		EiRabX31C	EIN_191990	42	170	box 1, switch I	XCXC
		EiRabX31D	EIN_197670	43	170	box 1, switch I	XCXC
EhRabX32	XP_651095						
EhRabX33	XP_655812						
EhRabX34	XP_650332	EiRabX34	EIN_281740	68	204	box 2, box 3	CXXX
		EiRabX34B	EIN_202640	63	200		CXXX
EhRabX35	XP_649164						
EhRabX36	XP_657040						
		EiRabZ1	EIN_050800		204		XXCC
		EiRabZ2	EIN_289320		200		
		EiRabZ3	EIN_192430		205		XXCC
		EiRabZ4	EIN_297180		205		XXCC
		EiRabZ5	EIN_039070		203		XXCC
		EiRabZ6	EIN_051060		187		XXCC
		EiRabZ7	EIN_058330		207		XXCC

[a]A systematic search was made to retrieve Rab genes from *E. histolytica* and *E. invadens*. All hits were analyzed as described in the legend of Table 1. The phylogenetic trees of *E. histolytica* and *E. invadens* Rab genes are shown in Fig. 2. *E. invadens* Rab genes were annotated based on phylogenetic inferences and their percentage identity to the corresponding *E. histolytica* homologues. For the annotation of *E. histolytica* Rab genes, see (114).

including comparisons of the protein profiles of whole *E. histolytica* trophozoites between wild-type and metronidazole-resistant strains, and between virulent and avirulent strains, and for organelle proteomes using isolated phagosomes (17, 33, 35, 68, 69, 81, 94, 95, 137).

In this section, new aspects in the molecular pathogenesis of amebiasis discovered through transcriptomics will be discussed. Insights into the developmental regulation of CPs and Rab small GTPases during encystation are also discussed as examples of transcriptomic approaches to under-

stand the biology and pathogenesis of this group of enteric protozoa.

Transcriptome Analysis of *E. histolytica* CP

Several transcriptomic studies using whole genome microarrays of *E. histolytica* demonstrated that EhCP-A1, A2, and A5 were the most abundant transcripts in axenically grown HM-1 trophozoites, while EhCP-A6, A10, A11, B2, C4, and CALP1 were only expressed at moderate levels (80, 134). Gilchrist et al. compared the transcriptomes of HM-1 trophozoites derived from the mouse cecum on day 1 or day 29 after inoculation with that of trophozoites cultivated in vitro (52). The expression of EhCP-A4 was induced by 28- to 35-fold between days 1 and 29, while the expression levels of EhCP-A1, A6, and A8 were induced by 2- to 9-fold. Among all of the CPs, EhCP-A1 mRNA was the most abundant in trophozoites derived from the intestine and was about 60 times higher than the levels of EhCP-A4, A6, and A8. In an independent experiment, EhCP-A6 was reported to be induced by heat stress (134, 145), and that EhCP-4A and 6A, which show high similarity to and form a monophyletic clade with EhCP-A5 (Fig. 1), were strongly induced by environmental changes occurring during *E. histolytica* infection (52, 134, 145). Weber et al. also demonstrated that when *E. histolytica* trophozoites were cultivated at 42°C for 4 h, EhCP-A6 was induced by 8.9-fold, while EhCP-4 was upregulated by 2.1-fold (145). There are, however, conflicting reports. One study showed that EhCP-A5 and A6 were induced after 4 h of cultivation at 42°C, while another study reported that C1 family CP genes were not induced after 1 h of cultivation at 42°C (80, 134). These data indicate that the expression of these genes may be affected by subtle differences in the experimental procedures and conditions used.

Ehrenkaufer et al. reported on the differences in transcriptomes between recent clinical isolates that partially encyst in regular xenic medium and laboratory strains that have apparently lost their ability to encyst (41). Among the genes that were potentially regulated developmentally, EhCP-A1 and A2 were predominantly expressed in the laboratory strains, while EhCP-A3, A4, A8, B1, B3, B8, B9, and B10 were expressed at a higher level in the recent clinical isolates. In future studies, there is a need to examine the differences in the transcriptomes of trophozoites derived from axenic and xenic cultures, and those obtained from host tissues and stool samples. However, since the amount of available clinical specimens is limited, microarray studies that use amplified cDNA appear to be a powerful tool to examine the transcriptome profiles of *E. histolytica* on a small scale.

Transcriptome Analysis of *E. histolytica* Rab

It was previously demonstrated by qRT-PCR that among the EhRab7 isotypes in the HM-1 strain, EhRab7H is the most abundantly transcribed, followed by EhRab7A and EhRab7E (114). EhRab7A facilitates CP transport, while the roles of EhRab7E and EhRab7H remain unknown (88). The basal expression levels of EhRab11 isotypes in HM-1 were also examined by qRT-PCR (87). The expression levels of the four family members varied, with EhRab11A having the highest expression and with the levels of expression decreasing in the order of 11B, 11C, and 11D. Since EhRab11A is recruited to the cell surface during iron or serum starvation, it was suggested to be involved in encystation (82). Conversely, the overexpression of EhRab11B caused an enhancement of the secretion of CPs (87), while the roles of EhRab11C and EhRab11D remain unknown.

There are several reports on the transcriptional regulation of Rab genes under stress conditions. MacFarlane et al. reported that incubation of trophozoites at 42°C for 1 h caused a 2-fold upregulation of the EhRab7A gene (80), while in another study, EhRabX14 expression was upregulated on days 1 and 29 in the mouse cecum (52). During host infection, amebic trophozoites are exposed to reactive oxygen and nitrogen species. The incubation of trophozoites with hydrogen peroxide or a nitric oxide donor (dipropylenetriamine-NONOate) caused a 3- to 4-fold upregulation of EhRabI1, and a 4- to 50-fold downregulation of EhRab7F. Since these Rab genes appeared to be regulated by hydrogen peroxide and nitric oxide in a similar way, a common regulator and pathway may be responsible for their identical responses against these stress factors. However, most of the Rab genes are not regulated in the same manner. EhRabM1, for example, was upregulated only by hydrogen peroxide, while EhRabX15, X32, and X35 were upregulated only by dipropylenetriamine-NONOate. Similarly, EhRabD2 was downregulated by oxidative stress, whereas EhRabX19, L1, and C2 were downregulated in nitric oxide-treated cells. Although the role of these Rab genes remains unknown, these findings indicate a possible link between oxidative/nitrosative stress responses and membrane trafficking.

Several EhRab genes have been suggested to play a role in the stage conversion of *E. histolytica* (41). Ehrenkaufer et al. (41) analyzed the transcriptome of recent clinical isolates and a laboratory *E. histolytica*

strain and demonstrated that EhRabM1 and EhRabN1 were upregulated in the recent clinical isolates by 6.7- and 4.1-fold, respectively. These data suggest that these Rab genes may be upregulated in the cyst stage. In contrast, EhRab5, 7D, 11A, 11B,11D, B, C1, C5, C6, D2, H, K2, and M2 were all expressed at higher levels in a laboratory strain that does not encyst compared to clinical isolates, suggesting that these Rab genes are upregulated in the trophozoite stage. It is worth noting that EhRabM1 is upregulated in the cyst stage and by oxidative stress, which may suggest a common role for EhRabM1 in the response to stress and differentiation. Conversely, EdRab5, X6, and X13 were downregulated in the nonpathogenic *E. dispar* (79). Since EhRab5 plays a role in prephagosomal vacuole formation (116) in *E. histolytica*, the repression of EdRab5 expression in *E. dispar* may indicate a decrease in phagocytosis and endocytosis in *E. dispar*.

AQ2

NEW INSIGHTS INTO THE MOLECULAR EVENTS THAT OCCUR DURING ENCYSTATION

The expression profiles of all the genes in *E. invadens* were recently analyzed using a custom-made full genome microarray of *E. invadens* on an Affymetrix platform (Eh_Eia520620F_Ei) (Escueta-de Cadiz et al., unpublished). The array contained 12,385 independent probe sets, each of which had 11 probe pairs. The probes were designed based on the genome sequence of the *E. invadens* reference strain IP-1, available from the Pathema Bioinformatics Resource Center (data release 6.0). Encystation was induced using 47% LG medium with reduced osmolarity and no glucose (118). Transcriptomes were examined at various time points up to 120 h after induction and the percentage of encystation was determined to be 1.9, 14.8, 49.7, and 86.4% at 0, 24, 48, and 120 h, respectively.

AQ1

Identification of Encystation-associated CPs

A massive degradation of cellular components in trophozoites must occur during encystation. This degenerative process is partly mediated by CPs. As described above, *E. histolytica* and *E. invadens* have large gene families of CPs; however, the majority of CPs are not expressed in the *E. histolytica* trophozoite stage when cultured in enriched regular medium (23, 134). The transcriptome of *E. invadens* during encystation has revealed several unique features of the kinetics of CPs (Fig. 3, upper panel). The kinetics of the steady-state level of mRNA of CPs from the CP-A, B, and C families were categorized into three groups: CP genes dominantly expressed in the cyst stage (>2-fold higher at 120 h than at 0 h after the induction of encystation); CP genes dominantly expressed in the trophozoite stage (>2-fold higher at 0 than at 120 h); and genes constitutively expressed (<2-fold change between 0 and 120 h). In general, among the EiCP-A, B, and C families, the members of the EiCP-A family show the highest expression in both trophozoites and cysts. About 50% of EiCP-A proteins were expressed constitutively (EiCP-A2a, A2e, A2f, A3c, A3d, A5a, A5b, and A10), while the rest were expressed preferentially in trophozoites (EiCP-2b, A2c, A2d, A3a, A3b, A8a, A8b, A11a, and A11b). Two EiCP-A members were categorized into the cyst-specific group, but their expression level was relatively low. Among the 9 trophozoite-dominant EiCP-A members, EiCP-A2c, A3a, and A3b showed strong induction 8 h after encystation, while 8 of the 10 EiCP-C genes were trophozoite specific, and their expression levels were gradually reduced during encystation.

The expression level of the EiCP-B family members was generally low to intermediate in trophozoites; however, the expression of EiCP-B2, B6, B9, and BA was strongly induced during encystation or remained upregulated in the cyst stage. This suggests that members of the EiCP-C family mainly function in trophozoites, while members of the EiCP-A family function in the trophozoite and cyst stages. In contrast, EiCP-B family members appear to have a role in developmental conversion. It was previously reported that EiCP-B9 was induced at 24 h and repressed at 48 h of encystation (39), and it was also identified as a cyst-specific CP using transcriptome analysis of *E. histolytica* clinical isolates (41). The developmentally regulated CPs appear to be, at least in part, shared by *E. histolytica* and *E. invadens*. In *Giardia lamblia*, one CP has been shown to be involved in the processing of a cyst wall protein (138). Further investigation needs to be conducted on the physiological substrates and functions of the CP-A and CP-B families to cast light on their roles in the encystation of *E. histolytica* and *E. invadens*.

Identification of Encystation-associated Rab Genes

During encystation, the drastic turnover of cellular components occurs in a rapid and regulated manner, involving the de novo synthesis of new proteins and degradation of undesired proteins. To support this massive reconfiguration of cellular components, the amoeba is expected to have an encystation-specific system of membrane trafficking. To gain insights into the mechanisms of encystation-specific

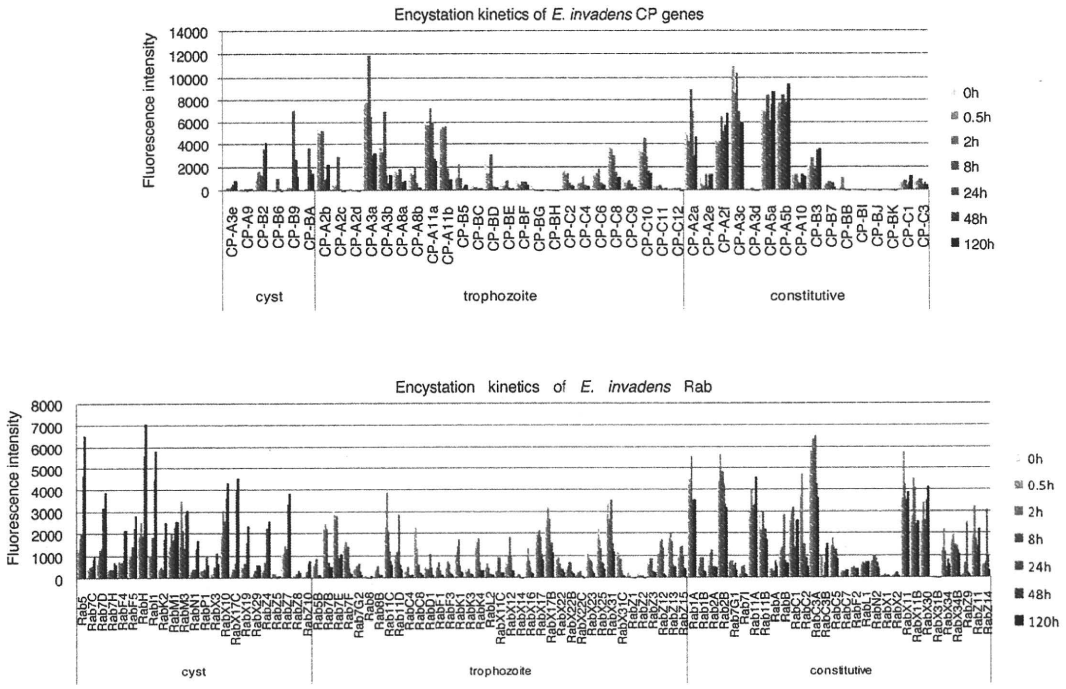

Figure 3. Kinetics of CP and Rab gene expression during encystation in *E. invadens*. *E. invadens* trophozoites were cultured in encystation medium. Total RNA was extracted at 0, 0.5, 2, 8, 24, 48, and 120 h, and subjected to transcriptome analysis using *E. invadens* DNA microarrays. The normalized relative fluorescence intensity of each gene is shown.

membrane trafficking, we also investigated the transcriptomic changes of Rab genes during the encystation of *E. invadens* (Fig. 3, lower panel). The kinetics of the steady-state levels of mRNA for individual Rab genes were categorized into cyst-specific, trophozoite-specific, or constitutive genes, using the same criteria described above for the CP genes. As shown in Fig. 3, 23 cyst-dominant, 36 trophozoite-dominant, and 31 constitutive Rab genes were identified. Some of these genes were either induced or repressed in a time-specific fashion. For instance, EiRab11C, EiRabM3, and EiRabX16 showed their highest level of expression at 2 h, while EiRabX12 expression level peaked at 8 h. EiRab11D, EiRabB, EiRabZ2, and EiRabZ5 expression level peaked at 24 h, while EiRabX14 expression level peaked at 48 h. The expression of EiRabC3b was markedly reduced at 2 to 8 h, while another isotype, EiRabC2, showed a sharp peak at 0.5 to 2 h. Other isotypes that belonged to a subfamily also showed different kinetics during encystation. The expression of EiRab7C, 7D, and 7H, for instance, was upregulated at 48 h, while that of EiRab7B, 7E, 7F, and 7G2 was repressed at 24 h of encystation. Furthermore, the expression of EiRab7G1 was constant, whereas that of EiRab7I was induced from 2 to 24 h, but was repressed at 48 h of encystation. EhRab7B is involved in lysosome biogenesis (114), while EhRab7A is involved in the targeting of hydrolases to lysosomes (88); therefore, it is conceivable that the expression of EiRab7B is repressed in cysts because lysosome function is assumed to be less active in dormant cysts, where endocytosis and phagocytosis are likely to be repressed. EiRab5A was strongly induced during encystation, while the expression of its isotype, EiRab5B, was repressed after 8 h of encystation. EiRab11A and EiRab11B showed similar constitutive patterns of expression, while the expression levels of EiRab11C and EiRab11D were reduced in cysts. In addition, EiRab11C and EiRab11D showed marked sharp peaks at 2 and 24 h, respectively. Although EhRab11A was shown to be translocated to the cell periphery by serum or iron starvation, its constitutive expression pattern does not support its involvement in encystation (82). Furthermore, although *E. histolytica* Rab11B has been demonstrated to be involved in the secretion of CPs (87), its constitutive pattern of expression suggests a general role throughout the life cycle.

FUTURE IMPACT OF GENOMIC AND POSTGENOMIC APPROACHES

Future Impact of Genomics

Despite the optimistic presumption that genomic information would reveal all the secrets of *E. histolytica*, we have learned that its genome shows an unexpected level of complexity and has raised more questions than answers. The fact remains, however, that researchers generally design and perform experiments based on genome-derived information; thus, genome information remains undoubtedly important.

There is a need to obtain the entire genome sequences of other *E. histolytica* clinical isolates that show different clinical presentations and with distinct geographic origins. Such information would help investigators understand the genetic diversity of *E. histolytica* as a species and the geographic distribution of genetic traits. It would also facilitate the discovery of factors that determine the severity of infection and tissue tropisms, and it may also help with understanding the role of transposable elements in the pathogenicity of *E. histolytica*.

Genomic approaches may be undertaken to analyze the genetic variations observed in the human host. It was previously shown that HLA class II alleles were related to the susceptibility of the host to amebiasis (38). It has recently been suggested that polymorphisms in leptin receptors have a close association with susceptibility to amebiasis (Petri et al., unpublished). Genomic studies on the pathogen and host should also clarify the important question of why only 5% to 10% of infected individuals develop disease. The use of new-generation sequencing technology should also enable metagenomic studies of bacterial flora from cases of intestinal amebiasis (colitis and dysentery), amebic liver abscesses, and asymptomatic cases. Such studies may answer the question of whether bacterial flora affects the outcome of *E. histolytica* infection.

AQ1

Future Impact of Postgenomics

As described here, using CP and Rab genes as examples, transcriptomic and proteomic approaches have become applicable and affordable methods to analyze the molecular mechanisms of virulence and parasitism in pathogenic amoeba. The recent developments in metabolomics should also provide us with a comprehensive understanding of metabolic fluxes and key metabolites that other "omics" studies are not able to reveal. Because of space restrictions, only the transcriptional changes in the CP and Rab genes are described here. There are, however, numerous other *Entamoeba* genes involved in various biological processes whose expression is modulated during encystation. For instance, we have discovered a number of stage-specific myb transcription factors in *E. invadens* that are expressed within a very narrow time range. This finding was in good accordance with a recent report on several myb transcription factors implicated in the encystation of *E. histolytica* (40). These cyst-specific *E. histolytica* myb genes were discovered by comparing the transcriptomes of *E. histolytica* clinical isolates that occasionally encyst in vitro and attenuated encystation-defective strains. It would be impossible, however, to find tightly regulated time-dependent genes without performing kinetic experiments. Thus, the transcriptomics of *E. invadens* remains an important encystation/excystation model, until in vitro stage conversion is accomplished with *E. histolytica*.

An important but not yet fully exploited application of transcriptomics is the discovery of the roles of metabolic enzymes whose functions are currently unknown. We should be able to understand a specific role of individual enzymes by comparing the transcriptomes of a parental line and cell lines in which the expression of the gene under investigation is repressed by RNA interference (1, 72, 142) or gene silencing (8, 20, 21, 22, 25, 86, 150). Such attempts are under way to characterize the role of methionine gamma-lyase, a unique amino acid metabolizing enzyme (4, 93, 136). Transcriptomics can also be used to unravel the action of drugs currently in clinical use or under development. The discovery of new metabolic pathways through transcriptomics should also provide opportunities to identify new drug targets.

Acknowledgements. We thank Aleyla Escueta-de Caz and Afzal Husain for their help in the annotation of the CP and Rab genes from *E. invadens* and for the transcriptome analyses. We also appreciate the helpful comments and personal assistance of Yumiko Saito-Nakano regarding the Rab genes from *E. invadens*, Gil Penuliar for proofreading the manuscript, and Lis Caler, Bioinformatics Resource Center, J. Craig Venter Institute for sharing unpublished information of the *E. invadens* genome.

REFERENCES

1. **Abed, M., and S. Ankri.** 2005. Molecular characterization of *Entamoeba histolytica* RNase III and AGO2, two RNA interference hallmark proteins. *Exp. Parasitol.* 110:265–269.
2. **Ali, I. K., S. Solaymani-Mohammadi, J. Akhter, S. Roy, C. Gorrini, A. Calderaro, S. K. Parker, R. Haque, W. A. Petri, and C. G. Clark.** 2008. Tissue Invasion by *Entamoeba histolytica:* Evidence of genetic selection and/or DNA reorganization events in organ tropism. *PLoS Negl. Trop. Dis.* 2:e219.
3. **Ali, I. K., U. Mondal, S. Roy, R. Haque, W. A. Petri, Jr., and C. G. Clark.** 2007. Evidence for a link between parasite genotype and outcome of infection with *Entamoeba histolytica*. *J. Clin. Microbiol.* 45:285–289.
4. **Ali, V., and T. Nozaki.** 2007. Current therapeutics, their problems, and sulfur-containing-amino-acid metabolism as a novel

target against infections by "amitochondriate" protozoan parasites. *Clin. Microbiol. Rev.* **20**:164–187.
5. **Ali, I. K., M. Zaki, and C. G. Clark.** 2005. Use of PCR amplification of tRNA gene-linked short tandem repeats for genotyping *Entamoeba histolytica*. *J. Clin. Microbiol.* **43**:5842–5847.
6. **Ali, I. K., M. B. Hossain, S. Roy, P. F. Ayeh-Kumi, W. A. Petri, Jr., R. Haque, and C. G. Clark.** 2003. *Entamoeba moshkovskii* infections in children, Bangladesh. *Emerg. Infect. Dis.* **9**: 580–584.
7. **Anamika, K., A. Bhattacharya, and N. Srinivasan.** 2008. Analysis of the protein kinome of *Entamoeba histolytica*. *Proteins* **71**:995–1006.
8. **Anbar, M., R. Bracha, Y. Nuchamowitz, Y. Li, A. Florentin, and D. Mirelman.** 2005. Involvement of a short interspersed element in epigenetic transcriptional silencing of the amoebapore gene in *Entamoeba histolytica*. *Eukaryot. Cell* **4**:1775–1784.
9. **Anderson, I. J., and B. J. Loftus.** 2005. *Entamoeba histolytica:* observations on metabolism based on the genome sequence. *Exp. Parasitol.* **110**:173–177.
10. **Ankri, S., T. Stolarsky, R. Bracha, F. Padilla-Vaca, and D. Mirelman.** 1999. Antisense inhibition of expression of cysteine proteinases affects *Entamoeba histolytica*-induced formation of liver abscess in hamsters. *Infect. Immun.* **67**:421–422.
11. **Baek, K. H.** 2006. Cytokine-regulated protein degradation by the ubiquitination system. *Curr. Protein. Pept. Sci.* **7**: 171–177.
12. **Baek, K. H.** 2003. Conjugation and deconjugation of ubiquitin regulating the destiny of proteins. *Exp. Mol. Med.* **35**:1–7.
13. **Bakre, A. A., K. Rawal, R. Ramaswamy, A. Bhattacharya, and S. Bhattacharya.** 2005. The LINEs and SINEs of *Entamoeba histolytica:* comparative analysis and genomic distribution. *Exp. Parasitol.* **110**:207–213.
14. **Balakirev, M. Y., S. O. Tcherniuk, M. Jaquinod, and J. Chroboczek.** 2003. Otubains: a new family of cysteine proteases in the ubiquitin pathway. *EMBO Rep.* **4**:517–522.
15. **Barrett, A. J.** 1998. Cysteine peptidase, p. 543–798. *In* A. J. Barrett, N. D. Rawlings, and J. F. Woessner (ed.), *Handbook of Proteolytic Enzymes*, Academic Press, San Diego, CA.
16. **Bentley, D. L.** 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* **17**:251–256.
17. **Boettner, D. R., C. D. Huston, A. S. Linford, S. N. Buss, E. Houpt, N. E. Sherman, and W. A. Petri, Jr.** 2008. *Entamoeba histolytica* phagocytosis of human erythrocytes involves PATMK, a member of the transmembrane kinase family. *PLoS Pathog.* **4**:e8.
18. **Borodovsky, A., H. Ovaa, N. Kolli, T. Gan-Erdene, K. D. Wilkinson, H. L. Ploegh, and B. M. Kessler.** 2002. Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem. Biol.* **9**:1149–1159.
19. **Bourne, H. R., D. A. Sanders, and F. McCormick.** 1990. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**:125–132.
20. **Bracha, R., Y. Nuchamowitz, N. Wender, and D. Mirelman.** 2007. Transcriptional gene silencing reveals two distinct groups of *Entamoeba histolytica* Gal/GalNAc-lectin light subunits. *Eukaryot. Cell* **6**:1758–1765.
21. **Bracha, R., Y. Nuchamowitz, M. Anbar, and D. Mirelman.** 2006. Transcriptional silencing of multiple genes in trophozoites of *Entamoeba histolytica*. *PLoS Pathog.* **2**:e48.
22. **Bracha, R., Y. Nuchamowitz, and D. Mirelman.** 2003. Transcriptional silencing of an amoebapore gene in *Entamoeba histolytica*: molecular analysis and effect on pathogenicity. *Eukaryot. Cell* **2**:295–305.
23. **Bruchhaus, I., B. J. Loftus, N. Hall, and E. Tannich.** 2003. The intestinal protozoan parasite *Entamoeba histolytica* contains 20 cysteine protease genes, of which only a small subset is expressed during in vitro cultivation. *Eukaryot. Cell* **2**:501–509.
24. **Bruchhaus, I., M. Leippe, C. Lioutas, and E. Tannich.** 1993. Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. *DNA Cell Biol.* **2**:925–933.
25. **Bujanover, S., U. Katz, R. Bracha, and D. Mirelman.** 2003. A virulence attenuated amoebapore-less mutant of *Entamoeba histolytica* and its interaction with host cells. *Int. J. Parasitol.* **33**:1655–1663.
26. **Buratowski, S.** 2005. Connections between mRNA 3′ end processing and transcription termination. *Curr. Opin. Cell Biol.* **17**:2572–2561.
27. **Cho, E. J.** 2007. RNA polymerase II carboxy-terminal domain with multiple connections. *Exp. Mol. Med.* **39**:247–254.
28. **Clark, C. G., U. C. Alsmark, M. Tazreiter, Y. Saito-Nakano, V. Ali, S. Marion, C. Weber, C. Mukherjee, I. Bruchhaus, E. Tannich, M. Leippe, T. Sicheritz-Ponten, P. G. Foster, J. Samuelson, J. Noël, R. P. Hirt, T. M. Embley, C. A. Gilchrist, B. J. Mann, U. Singh, J. P. Ackers, S. Bhattacharya, A. Bhattacharya, A. Lohia, N. Guillén, M. Duchêne, T. Nozaki, and N. Hall.** 2007. Structure and content of the *Entamoeba histolytica* genome. *Adv. Parasitol.* **65**:51–190.
29. **Clark, C. G., I. K. Ali, M. Zaki, B. J. Loftus, and N. Hall.** 2006. Unique organization of tRNA genes in *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* **146**:24–29.
30. **Clark, C. G., and L. S. Diamond.** 1997. Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. *J. Eukaryot. Microbiol.* **44**:142–154.
31. **Clark, C. G., and L. S. Diamond.** 1991. Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct. *Mol. Biochem. Parasitol.* **49**:297–302.
32. **Clark, C. G., and L. S. Diamond.** 1991. The Laredo strain and other "*Entamoeba histolytica*-like" amoebae are *Entamoeba moshkovskii*. *Mol. Biochem. Parasitol.* **46**:11–18.
33. **Davis, P. H., M. Chen, X. Zhang, C. G. Clark, R. R. Townsend, and S. L. Stanley.** 2009. Proteomic comparison of *Entamoeba histolytica* and *Entamoeba dispar* and the Role of *E. histolytica* alcohol dehydrogenase 3 in virulence. *PLoS Negl. Trop. Dis.* **3**:e415.
34. **Davis, P. H., J. Schulze, and S. L. Stanley, 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. *Mol. Biochem. Parasitol.* **151**:118–128.
35. **Davis, P. H., X. Zhang, J. Guo, R. R. Townsend, and S. L. Stanley, Jr.** 2006. Comparative proteomic analysis of two *Entamoeba histolytica* strains with different virulence phenotypes identifies peroxiredoxin as an important component of amebic virulence. *Mol. Microbiol.* **61**:1523–1532.
36. **Davis, P. H., Z. Zhang, M. Chen, X. Zhang, S. Chakraborty, and S. L. Stanley, Jr.** 2006. Identification of a family of BspA like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats. *Mol. Biochem. Parasitol.* **145**:111–116.
37. **Diamond, L. S., and C. G. Clark.** 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J. Eukaryot. Microbiol.* **40**:340–344.
38. **Duggal, P., R. Haque, S. Roy, D. Mondal, R. B. Sack, B. M. Farr, T. H. Beaty, and W. A. Petri., Jr.** 2004. Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. *J. Infect. Dis.* **189**: 520–526.
39. **Ebert, F., A. Bachmann, K. Nakada-Tsukui, I. Hennings, B. Drescher, T. Nozaki, E. Tannich, and I. Bruchhaus.** 2008. An *Entamoeba* cysteine peptidase specifically expressed during encystation. *Parasitol. Int.* **7**:521–524.

40. **Ehrenkaufer, G. M., J. A. Hackney, and U. Singh.** 2009. A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in *Entamoeba histolytica*. *Cell Microbiol.* 11:898–910
41. **Ehrenkaufer, G. M., R. Haque, J. A. Hackney, D. J. Eichinger, and U. Singh.** 2007. Identification of developmentally regulated genes in *Entamoeba histolytica*: insights into mechanisms of stage conversion in a protozoan parasite. *Cell Microbiol.* 9:1426–1444.
42. **Eichinger, L., J. A. Pachebat, G. Glöckner, M. A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabbinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzer, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell, and A. Kuspa.** 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* 435:43–57.
43. **Eichinger, D.** 1997. Encystation of entamoeba parasites. *Bioessays* 19:633–639.
44. **Evans, P. C., T. S. Smith, M. J. Lai, M. G. Williams, D. F. Burke, K. Heyninck, M. M, Kreike, R. Beyaert, T. L. Blundell, and P. J. Kilshaw.** 2003. A novel type of deubiquitinating enzyme. *J. Biol. Chem.* 278:23180–23186.
45. **Fotedar, R., D. Stark, D. Marriott, J. Ellis, and J. Harkness.** 2008. *Entamoeba moshkovskii* infections in Sydney, Australia. *Eur. J. Clin. Microbiol. Infect. Dis.* 27:133–137.
46. **Frederick, J. R., and W. A. Petri, Jr.** 2005. Roles for the galactose-/N-acetylgalactosamine-binding lectin of *Entamoeba* in parasite virulence and differentiation. *Glycobiology* 15:53R–59R.
47. **Gadasi, H., and E. Kessler.** 1983. Correlation of virulence and collagenolytic activity in *Entamoeba histolytica*. *Infect. Immun.* 39:528–5231.
48. **Gardner, M. J., N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M. S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, and B. Barrell.** 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 19:498–511.
49. **Gelderman, A. H., I. L. Bartgis, D. B. Keister, and L. S. Diamond.** 1971. A comparison of genome sizes and thermal-denaturation-derived base composition of DNAs from several members of *Entamoeba* (*histolytica* group). *J. Parasitol.* 57: 912–916.
50. **Gelderman, A. H., D. B. Keister, I. L. Bartgis, and L. S. Diamond.** 1971. Characterization of the deoxyribonucleic acid of representative strains of *Entamoeba histolytica*, *E. histolytica*-like amebae, and *E. moshkovskii*. *J. Parasitol.* 57:906–911.
51. **Ghadirian, E., and E. Meerovitch.** 1984. Lectin-induced agglutination of trophozoites of different species and strains of *Entamoeba*. *Z. Parasitenkd.* 70:147–52.
52. **Gilchrist, C. A., E. Houpt, N. Trapaidze, Z. Fei, O. Crasta, A. Asgharpour, C. Evans, S. Martino-Catt, D. J. Baba, S. Stroup, S. Hamano,, G. Ehrenkaufer, M. Okada, U. Singh, T. Nozaki, B. J. Mann, and W. A. Petri, Jr.** 2006. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Mol. Biochem. Parasitol.* 147:163–176.
53. **Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver.** 1996. Life with 6000 genes. *Science* 274:546, 563–567.
54. **Hacker, J., U. Hentschel, and U. Dobrindt.** 2003. Prokaryotic chromosomes and disease. *Science.* 301:790–793.
55. **Haque, R., C. D. Huston, M. Hughes, E. Houpt, and W. A. Petri, Jr.** 2003. Amebiasis. *N. Engl. J. Med.* 348:1565–1573.
56. **Haque, R., P. Duggal, I. M. Ali, M. B. Hossain, D. Mondal, R. B. Sack, Farr, B. M., T. H. Beaty, and W. A. Petri, Jr.** 2002. Innate and acquired resistance to amebiasis in Bangladeshi children. *J. Infect. Dis.* 186:547–552.
57. **Haque, R., I. K. Ali, S. Akther, and W. A. Petri, Jr.** 1998. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. *J. Clin. Microbiol.* 36:449–452.
58. **Hernández, R., J. P. Luna-Arias, and E. Orozco.** 1997. Comparison of the *Entamoeba histolytica* TATA-binding protein (TBP) structure with other TBP. *Arch. Med. Res.* 28:43–45.
59. **Hirata, K. K., X. Que, S. G. Melendez-Lopez, A. Debnath, S. Myers, D. S. Herdman, E. Orozco, A. Bhattacharya, J. H. McKerrow, and S. L. Reed.** 2007. A phagocytosis mutant of *Entamoeba histolytica* is less virulent due to deficient proteinase expression and release. *Exp. Parasitol.* 115192–199.
60. **Huston, C. D.** 2004. Parasite and host contributions to the pathogenesis of amebic colitis. *Trends Parasitol.* 20:23–26.
61. **Jurica, M. S., and M. J. Moore.** 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell.* 12:5–14.
62. **Keene, W. E., M. E. Hidalgo, E. Orozco, and J. H. McKerrow.** 1990. *Entamoeba histolytica*: correlation of the cytopathic effect of virulent trophozoites with secretion of a cysteine proteinase. *Exp. Parasitol.* 71:199–206.
63. **Kimura, R. H., P. V. Choudary, K. K. Stone, and C. W. Schmid.** 2001. Stress induction of Bm1 RNA in silkworm larvae: SINEs, an unusual class of stress genes. *Cell Stress Chaperones* 6:263–272.
64. **Klionsky, D. J.** 2007. Autophagy: from phenomenology to molecular understanding in less than a decade. *Nat. Rev. Mol. Cell Biol.* 8:931–937.
65. **Kuchay, S. M, and A. H. Chishti.** 2007. Calpain-mediated regulation of platelet signaling pathways. *Curr. Opin. Hematol.* 14:2492–2954.
66. **Kunji, E. R.** 2004. The role and structure of mitochondrial carriers. *FEBS Lett.* 564:239–244.
67. **Lebart, M. C., and Y. Benyamin.** 2006. Calpain involvement in the remodeling of cytoskeletal anchorage complexes. *FEBS J.* 273:3415–3426.
68. **Leitsch, D., D. Kolarich, I. B. Wilson, F. Altmann, and M. Duchêne.** 2007. Nitroimidazole action in *Entamoeba histolytica*: a central role for thioredoxin reductase. *PLoS Biol.* 5:e211.
69. **Leitsch, D., C. Radauer, K. Paschinger, I. B. Wilson, H. Breiteneder, O. Scheiner, and M. Duchêne.** 2005. *Entamoeba histolytica*: analysis of the trophozoite proteome by two-dimensional polyacrylamide gel electrophoresis. *Exp. Parasitol.* 110:191–195.

70. **Li, Y., Z. Y. Chen, W. Wang, C. C. Baker, and R. M. Krug.** 2001. The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNAs in vivo. *RNA.* 7:920–923.
71. **Li, E., W. G. Yang, T. Zhang, and S. L. Stanley, Jr.** 1995. Interaction of laminin with *Entamoeba histolytica* cysteine proteinases and its effect on amebic pathogenesis. *Infect. Immun.* 63: 4150–4153.
72. **Linford, A. S., H. Moreno, K. R. Good, H. Zhang, U. Singh, and W. A. Petri, Jr.** Short hairpin RNA-mediated knockdown of protein expression in *Entamoeba histolytica.* BMC *Microbiol.* 2009 Feb 17;9:38.
73. **Lioutas, C., and E. Tannich.** 1995. Transcription of protein-coding genes in *Entamoeba histolytica* is insensitive to high concentrations of alpha-amanitin. *Mol. Biochem. Parasitol.* 73: 259–261.
74. **Loftus, B., I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabbinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillén, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, Jr., C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser, and N. Hall.** 2005. The genome of the protist parasite *Entamoeba histolytica. Nature* 433:865–858.
75. **Lorenzi, H., M. Thiagarajan, B. Haas, J. Wortman, N. Hall, and E. Caler.** 2008. Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. *BMC Genomics* 9:595.
76. **Lowther, S. A., M. S. Dworkin, and D. L. Hanson.** 2000. *Entamoeba histolytica*/*Entamoeba dispar* infections in human immunodeficiency virus-infected patients in the United States. *Clin. Infect. Dis.* 30:955–959.
77. **Luaces, A. L., and A. J. Barrett.** 1988. Affinity purification and biochemical characterization of histolysin, the major cysteine proteinase of *Entamoeba histolytica. Biochem. J.* 250:903–909.
78. **Lushbaugh, W. B., A. F. Hofbauer, and F. E. Pittman.** 1985. *Entamoeba histolytica*: purification of cathepsin B. *Exp. Parasitol.* 59:328–336.
79. **MacFarlane, R. C., and U. Singh.** 2006. Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. *Infect. Immun.* 74:340–351.
80. **MacFarlane, R., D. Bhattacharya, and U. Singh.** 2005. Genomic DNA microarrays for *Entamoeba histolytica*: applications for use in expression profiling and strain genotyping. *Exp. Parasitol.* 110:196–202.
81. **Marion, S., C. Laurent, and N. Guillén.** 2005. Signalization and cytoskeleton activity through myosin IB during the early steps of phagocytosis in *Entamoeba histolytica*: a proteomic approach. *Cell Microbiol.* 7:1504–1518.
82. **McGugan, G. C. Jr., and L. A. Temesvari.** 2003. Characterization of a Rab11-like GTPase, EhRab11, of *Entamoeba histolytica. Mol. Biochem. Parasitol.* 129:137–146.
83. **McKerrow, J. H., C. Caffrey, B. Kelly, P. Loke, and M. Sajid.** 2006. Proteases in parasitic diseases. *Annu. Rev. Pathol.* 1:497–536.
84. **Mehra, A., J. Fredrick, W. A. Petri, Jr., S. Bhattacharya, and A. Bhattacharya.** 2006. Expression and function of a family of transmembrane kinases from the protozoan parasite *Entamoeba histolytica. Infect. Immun.* 74:5341–4351.
85. **Mi-ichi, F., Yousuf, M. A., Nakada-Tsukui, K., and Nozaki, T.** 2009. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway. *Proc. Natl. Acad. Sci. USA* 106: 21731–21736.
86. **Mirelman, D., M. Anbar, and R. Bracha.** 2008. Epigenetic transcriptional gene silencing in *Entamoeba histolytica. IUBMB Life* 60:598–604.
87. **Mitra, B. N., Y. Saito-Nakano, K. Nakada-Tsukui, D. Sato, and Nozaki, T.** 2007. Rab11B small GTPase regulates secretion of cysteine proteases in the enteric protozoan parasite *Entamoeba histolytica. Cell Microbiol.* 9:2112–2125.
88. **Nakada-Tsukui, K., Y. Saito-Nakano, V. Ali, and T. Nozaki.** 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. Mol. Biol. Cell.* 16:5294–5303.
89. **World Health Organization.** 1997. WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. Mexico City, Mexico January 28–29, 1997. *Epidemiol. Bull.* 18: 13–14.
90. **Novick, P., and M. Zerial.** 1997. The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* 9:496–504.
91. **Nozaki, T., S. Kobayashi, T. Takeuchi, and A. Haghighi.** 2006. Diversity of clinical isolates of *Entamoeba histolytica* in Japan. *Arch. Med. Res.* 37:277–279.
92. **Nozaki, T., and K. Nakada-Tsukui.** 2006. Membrane trafficking as a virulence mechanism of the enteric protozoan parasite *Entamoeba histolytica. Parasitol. Res.* 98:179–183.
93. **Nozaki, T., V. Ali, and M. Tokoro.** 2005. Sulfur-containing amino acid metabolism in parasitic protozoa. *Adv. Parasitol.* 60:1–99.
94. **Okada, M., C. D. Huston, M. Oue, B. J. Mann, W. A. Petri, Jr., K. Kita, and T. Nozaki.** 2006. Kinetics and strain variation of phagosome proteins of *Entamoeba histolytica* by proteomic analysis. *Mol. Biochem. Parasitol.* 145:171–183.
95. **Okada, M., C. D. Huston, B. J. Mann, W. A. Petri, Jr., K. Kita, and T. Nozaki.** 2005. Proteomic analysis of phagocytosis in the enteric protozoan parasite *Entamoeba histolytica. Eukaryot. Cell* 4:827–831.
96. **Orozco, E., G. Guarneros, A. Martinez-Palomo, and T. Sánchez.** 1983. *Entamoeba histolytica.* Phagocytosis as a virulence factor. *J. Exp. Med.* 158:1511–1521.
97. **Ostertag, E. M., and H. H. Kazazian, Jr.** 2001. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35: 501–538.
98. **Palmieri, F., G. Agrimi, E. Blanco, A. Castegna, M. A. Di Noia, V. Iacobazzi, F. M. Lasorsa, C. M. Marobbio, L. Palmieri, P. Scarcia, S. Todisco, A. Vozza, and J. Walker.** 2006. Identification of mitochondrial carriers in *Saccharomyces cerevisiae* by transport assay of reconstituted recombinant proteins. *Biochim. Biophys. Acta.* 1757:1249–1262.
99. **Parija, S. C., and K. Khairnar.** 2005. *Entamoeba moshkovskii* and *Entamoeba dispar*-associated infections in pondicherry, India. *J. Health Popul. Nutr.* 23:292–295.
100. **Pereira-Leal, J. B., and M. C. Seabra.** 2001. Evolution of the Rab family of small GTP-binding proteins. *J. Mol. Biol.* 313:889–901.
101. **Perrin, B. J., and A. Huttenlocher.** 2002. Calpain. *Int. J. Biochem. Cell Biol.* 34:722–725.
102. **Phatnani, H. P., and A. L. Greenleaf.** 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20:2922–2936.
103. **Picazarri, K., K. Nakada-Tsukui, and T. Nozaki.** 2008. Autophagy during proliferation and encystation in the protozoan parasite *Entamoeba invadens. Infect. Immun.* 76:278–288.
104. **Picazarri, K., K. Nakada-Tsukui, D. Sato, and T. Nozaki.** 2008. Analysis of autophagy in the enteric protozoan parasite *Entamoeba. Methods Enzymol.* 451:359–371.