

# Evolutionary genomics of *Entamoeba*

Gareth D. Weedall\*, Neil Hall

*Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, UK*

Received 16 December 2010; accepted 17 December 2010

Available online 1 February 2011

## Abstract

*Entamoeba histolytica* is a human pathogen that causes amoebic dysentery and leads to significant morbidity and mortality worldwide. Understanding the genome and evolution of the parasite will help explain how, when and why it causes disease. Here we review current knowledge about the evolutionary genomics of *Entamoeba*: how differences between the genomes of different species may help explain different phenotypes, and how variation among *E. histolytica* parasites reveals patterns of population structure. The imminent expansion of the amount genome data will greatly improve our knowledge of the genus and of pathogenic species within it.

© 2011 Institut Pasteur. Published by Elsevier Masson SAS. Open access under [CC BY license](#).

**Keywords:** *Entamoeba*; Genetic diversity; Genomics; Molecular epidemiology

## 1. Introduction

*Entamoeba histolytica* is a parasite of the human large intestine, commonly contracted by ingesting contaminated water or food. The parasite has a two-stage life cycle in which the infective stage in the environment is the cyst and the motile stage within the host is the trophozoite. Infection with the parasite is endemic in many parts of the world where sanitation infrastructure is poor; in other places infection tends to be restricted to certain groups, such as residents in institutions for the mentally handicapped and men who have sex with men (Haghighi et al., 2002, 2003; Rivera et al., 2006). The global prevalence of infection was estimated in 1986 to be 10% of the world's population (Walsh, 1986). Of these, 90% were estimated to be asymptomatic carriers and 10% to develop symptoms of invasive amoebiasis. Amoebiasis results from invasion of the gut wall, leading to diarrhoea and dysentery (bloody stools), and in some cases to colonisation of organs (commonly the liver) and production of abscesses. The global prevalence estimate was made prior to the re-description of *E. histolytica* in 1993 that separated it into two species (Diamond and Clark, 1993): the potentially virulent *E. histolytica* and the

avirulent *Entamoeba dispar*. Despite this change, invasive amoebiasis still appears to be a relatively rare outcome of *E. histolytica* infection.

Understanding what determines the outcome of infection, and the nature of amoebic virulence more generally, motivates a substantial body of *Entamoeba* research. As part of this effort, the genome sequences of a number of *Entamoeba* species have been determined. These offer a fascinating insight into the evolution of these organisms. Here we review a number of notable features of *Entamoeba* genomes, from the point of view both of the evolution of different species lineages and of genetic diversity among *E. histolytica* populations.

## 2. Whole-genome sequences of *Entamoeba* species

The *Entamoeba* genus contains many species infecting a wide range of hosts. The simplest morphological feature used to distinguish species is the number of nuclei in the cysts, commonly 1, 4 or 8, although some species like the oral parasite *Entamoeba gingivalis* do not form cysts. The phylogeny of the genus shows often large evolutionary distances between *Entamoeba* species. Fig. 1 shows a phylogeny of the *Entamoeba* genus, and indicates species for which genome sequence data are, or are soon to be, available.

\* Corresponding author.

E-mail address: G.D.Weedall@liverpool.ac.uk (G.D. Weedall).

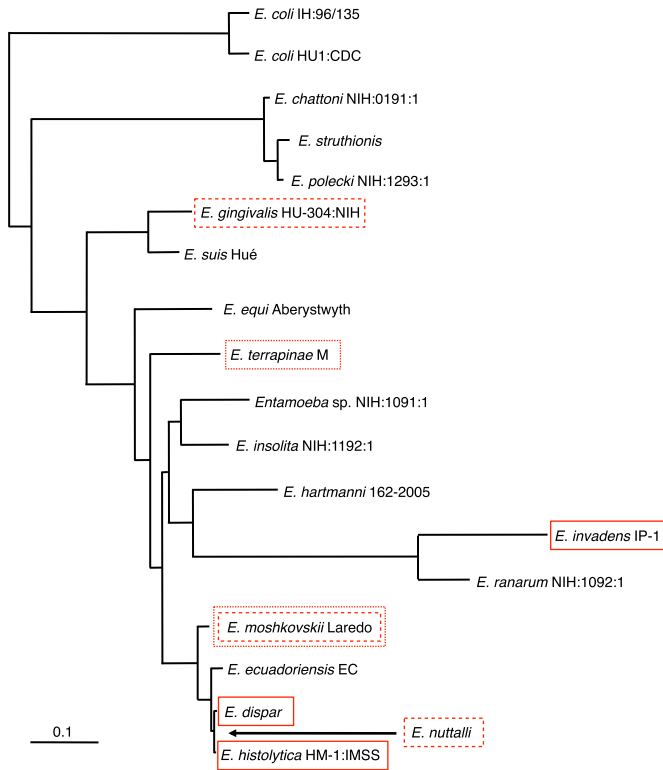


Fig. 1. Phylogeny of *Entamoeba*, adapted from Clark et al. (2006b). Phylogeny based on small subunit rRNA genes. Red boxes indicate species for which genomes have been sequenced. Red dotted boxes indicate species for which there is low coverage shotgun sequence data. Red dashed boxes indicate species due to be sequenced.

Table 1 shows details of the *Entamoeba* genomes sequenced and assembled so far. The most up-to-date annotations and many genome-level datasets are presented at the amoebaDB website (Aurrecochea et al., 2011; [www.amoebadb.org](http://www.amoebadb.org)).

The draft genome sequence of *E. histolytica* (strain HM1:IMSS) was published and analysed in 2005 (Loftus et al., 2005) and was subsequently re-assembled and re-annotated (Lorenzi et al., 2010; Clark et al., 2007). The genome assembly consists of 20,800,560 base pairs of DNA in 1496 scaffolds. The genome is very AT-rich (approximately 75% AT) and quite gene-rich: around half of all assembled sequence is predicted to be coding sequence, with 8333 annotated genes.

*E. dispar* is the closest described relative of *E. histolytica* and is morphologically identical, being designated a separate species in 1993 (Diamond and Clark, 1993). It is not known to be virulent, but rather to live as a commensal in the gut. The

Table 1  
Statistics describing sequenced *Entamoeba* genomes.<sup>a</sup>

	<i>E. histolytica</i>	<i>E. dispar</i>	<i>E. invadens</i>
Num. of scaffolds	1496	3312	1149
Num. of megabases	20.80	22.96	40.89
N50 <sup>b</sup>	49118	27840	243235
Coverage <sup>c</sup>	8x	4.32x	2.8x
Num. of genes	8333	8748	11549

<sup>a</sup> Numbers represent data in AmoebaDB version 1.1.

<sup>b</sup> 50% of bases are in scaffolds of this size or greater.

<sup>c</sup> As stated in genome project information in GenBank.

genome assembly is of a similar size to that of *E. histolytica*, consisting of 22,955,291 bp of DNA in 3312 scaffolds. Its AT content is also similar to *E. histolytica* (approximately 76.5% AT) and a similar proportion is predicted to be coding sequence, with 8749 annotated genes.

*Entamoeba invadens* is a parasite of reptiles and, although only distantly related to *E. histolytica*, is an important model for the encystation process, since it can be induced to encyst in axenic laboratory culture, while *E. histolytica* cannot. The genome appears to be larger than that of *E. histolytica* or *E. dispar*, consisting of 40,888,805 base pairs of DNA in 1149 scaffolds. It is also slightly less AT-rich (approximately 70% AT). Approximately 38% of all the sequence is predicted to be a coding sequence, with 11,549 annotated genes. In addition to these genome assemblies, low coverage shotgun sequencing projects have been carried out on the genomes of a further two *Entamoeba* species: *E. terrapinae* and *Entamoeba moshkovskii*. Sequencing reads are publicly available.

Next-generation sequencing platforms allow rapid sequencing of entire genomes. *E. moshkovskii*, a putative free-living species which has nonetheless been shown to infect humans (Ali et al., 2003), has been sequenced by our group (manuscript in preparation). The oral parasite *E. gingivalis*, which has not been shown to form cysts, will also be sequenced. *Entamoeba nuttali*, a pathogen of monkeys and apparently more closely related to *E. histolytica* than *E. dispar* (Tachibana et al., 2007), will be sequenced by colleagues at the J. Craig Venter Institute (Dr. Elisabet Caler, personal communication). In addition to these species, our group and colleagues at the JCVI are in the process of sequencing multiple strains of *E. histolytica* to assay intraspecies genomic diversity. All of these data will be made publicly available. As the number of sequenced genomes increases, our understanding of the evolutionary processes shaping these genomes will improve.

### 3. Structure and organization of the genome

The content of the *E. histolytica* genome has been reviewed extensively elsewhere (Clark et al., 2007). A number of interesting evolutionary features of the genome have been highlighted, not least the significant number of genes (at least 68) that appear to have been gained by horizontal gene transfer from bacteria (Loftus et al., 2005; Clark et al., 2007; Alsmark et al., 2009). These genes tend to be involved in metabolic processes characteristic of the anaerobic lifestyle of the organisms (Rosenthal et al., 1997; Field et al., 2000; Andersson et al., 2006), and the majority of transfers appear to have been ancient, as orthologues are found in both *E. histolytica* and the highly divergent *E. invadens* (Roy et al., 2006).

Much remains unknown about the large-scale structure and architecture of *Entamoeba* genomes. For instance, neither the natural ploidy nor the haploid number of chromosomes is known, although there are estimates of both. Hybridisation of gene markers to pulsed-field gels identified linkage groups and a haploid number of 14 chromosomes (Willhoeft and Tannich, 1999). Suggestions of tetraploidy and diploidy have been

advanced: tetraploidy based on labelling patterns in PFGE (Willhoeft and Tannich, 1999) and diploidy based on fluorescence in situ hybridisation to putative single-copy genes (Ghosh et al., 2000). Ploidy appears to be variable even within a cell lineage under different growth conditions (Mukherjee et al., 2008), although whether this results from growth in laboratory culture only or also occurs in nature is not known.

Studies of the molecular karyotype of *E. histolytica* show complex patterns of differences in chromosome sizes between strains and a mixture of linear and circular DNA molecules (Willhoeft and Tannich, 1999; Riveron et al., 2000). A number of circular DNA structures occur in *E. histolytica* (Dhar et al., 1995; Lioutas et al., 1995; Báez-Camargo et al., 1996; Riveron et al., 2000). The rRNA genes occur on one such molecule that exists in multiple copies per nucleus (Bhattacharya et al., 1998). Lorenzi et al. (2010) suggested that segmental duplications detected in the genome assembly might, in fact, represent some of these circular molecules (Lorenzi et al., 2010). These structures could be important for parasite phenotypes and their presence raises many intriguing questions. For instance, is their copy number different from that of the 'core' chromosomes and do they segregate in the same way?

An unusual feature of the genomic organisation of *E. histolytica* is that its tRNA genes occur in arrays of tandemly duplicated combinations of genes separated by DNA that may contain short tandem repeats (Clark et al., 2006a; Tawari et al., 2008). It has been suggested that these could act as telomeres, as a telomeric sequence has not been identified in *Entamoeba* genomes. The arrangement of tRNA genes appears to be quite variable and to have changed in the evolution of different species lineages. Fig. 2 shows examples of different arrangements of tRNA genes in different species. The DNA between the tRNA genes in a number of arrays is highly variable between isolates of *E. histolytica* (and *E. dispar*) and they have been developed for use as population genetic markers (Ali et al., 2005). However, the pattern of variable short tandem repeats seen in *E. histolytica* is not seen in *E. moshkovskii*, where these regions are variable but not repetitive (Tawari et al., 2008).

#### 4. Transposable elements, synteny and genomic rearrangements

In parasitic protozoa such as *Plasmodium*, large-scale genome architecture appears to be quite stable over large evolutionary distances. Distant relatives in the genus (the human parasite *Plasmodium falciparum* and the rodent parasites *Plasmodium yoelii*, *Plasmodium chabaudi* and *Plasmodium berghei*) generally possess a shared 'core genome' covering the central region of each chromosome, while the subtelomeric chromosomal regions tend to be more variable and species-specific and contain members of large gene families (Kooij et al., 2005). A similar distribution of variability is seen among different *P. falciparum* isolates (Volkman et al., 2007).

Such a pattern cannot be seen in *Entamoeba*, as the karyotype is not yet well enough resolved and its complexity

suggests a degree of genomic plasticity within *E. histolytica*. Such variability of chromosome size has been described in other protozoa (Adam, 1992; Blaineau et al., 1991; Melville et al., 1999) and genomic plasticity and instability may be an important feature of the evolution of *Entamoeba*, both within and between species. Genome rearrangement associated with invasive disease has been suggested as one possible explanation for the different tRNA-STR genotypes detected in liver abscess and stool-derived parasites from the same infected person (Ali et al., 2007).

Transposons and repetitive DNA, which are present in abundance in *Entamoeba*, may facilitate genome rearrangements. A comprehensive study of the repetitive elements of three *Entamoeba* genomes (Lorenzi et al., 2008) found hundreds of copies of LINE and SINE elements, as well as *Entamoeba*-specific repeats. These *Entamoeba*-specific ERE1 and ERE2 sequences represent a large proportion of the genome of *E. histolytica*. The ERE2 sequence may be unique to *E. histolytica*, as it was found in neither *E. dispar* nor *E. invadens* (Lorenzi et al., 2008). LINE and SINE elements are class I transposons (propagated via reverse transcription). Class II transposons (DNA transposons) have also been detected in *Entamoeba* species, and though they are rare in *E. histolytica* or *E. dispar*, they appear to be much more prevalent in *E. invadens* and *E. moshkovskii* (Pritham et al., 2005). These results suggest expansion and contraction of the number of transposable elements in different lineages, with likely consequences for genome rearrangement. Fig. 3 shows an example of a break in synteny between *E. histolytica* and *E. dispar* across a chromosome region containing repetitive elements.

Comparisons of the genomes of *E. histolytica* and *E. dispar* show that transposons have been active since these species diverged (Willhoeft et al., 2002; Shire and Ackers, 2007; Lorenzi et al., 2008). Active transposition may still be occurring. Huntley et al. (2010) identified a number of putative recent transpositions of *Eh*SINE1 elements in the HM1:IMSS genome (Huntley et al., 2010).

#### 5. Gene families and diversity

Possession of large gene families often indicates the importance and complexity of particular processes. *E. histolytica* contains a number of large multi-gene families (Lorenzi et al., 2010; see (Fig. 4a)). A large gene family encodes a group of AIG1-like GTPases (Lorenzi et al., 2010). Their precise function is unknown, but differential expression suggests they may be associated with virulence and/or adaptation to the intestinal environment (Davis et al., 2007; Biller et al., 2010; Gilchrist et al., 2006). Members of the AIG1-like family, among a number of gene families, often occur near to transposons (Lorenzi et al., 2010). It remains to be seen whether gene duplication and subsequent growth of gene families is promoted by the proximity of these elements. Another large gene family encodes proteins homologous to a bacterial fibronectin-binding protein (BspA of *Bacteroides forsythus*), which encodes a large number (75–116) of

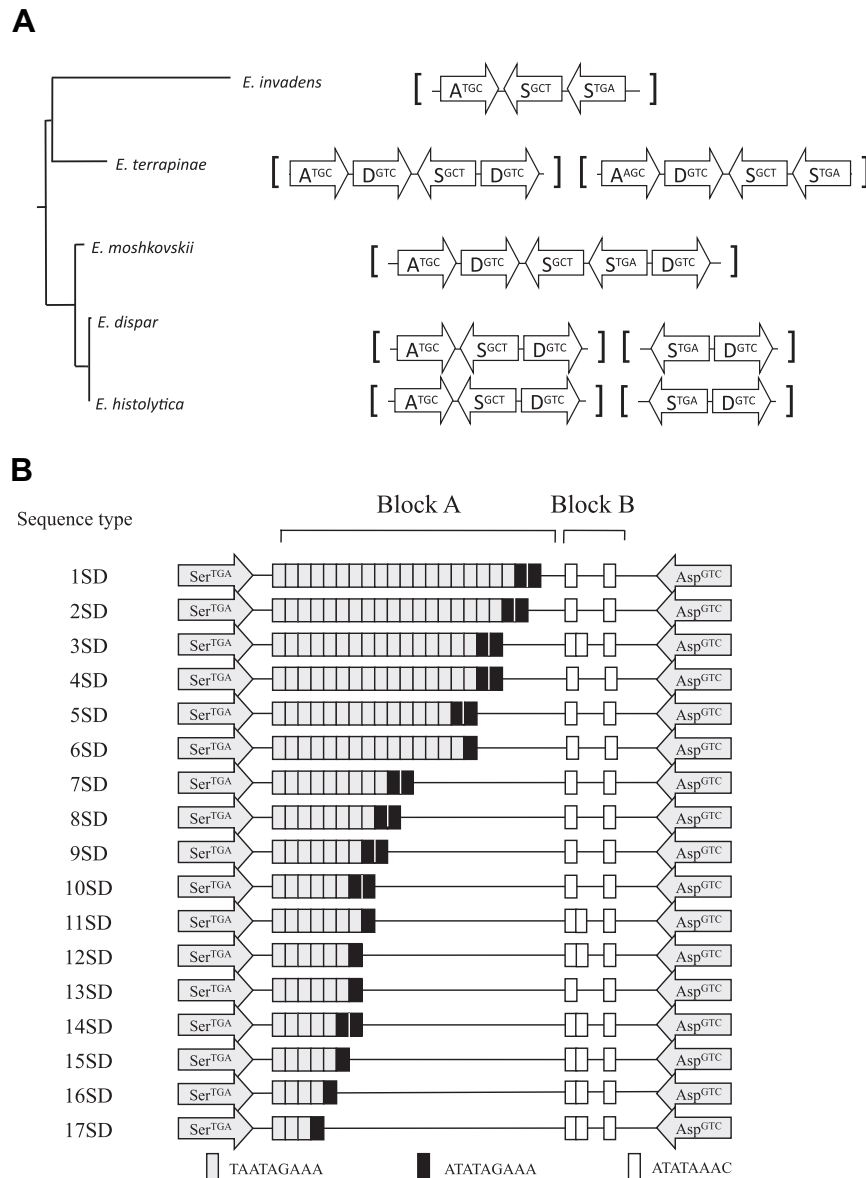


Fig. 2. Diversity of tRNA gene arrays, from Tawari et al. (2008). (A) Rearrangements of tRNA gene array units containing alanine, serine and aspartic acid tRNA genes, among *Entamoeba* species. Genes are represented by arrows containing single-letter amino acid codes (A = alanine, S = serine, D = aspartic acid) and superscripted anticodons. (B) Schematic representation of polymorphism in the intergenic DNA between SD array units, comprising serine (Ser<sup>TGA</sup>) and aspartic acid (Asp<sup>GTC</sup>), among *E. histolytica* isolates.

proteins containing leucine-rich repeats (Davis et al., 2006; Lorenzi et al., 2010). At least one member of this family is expressed at the parasite surface (Davis et al., 2006). A survey of *E. invadens* sequence reads indicated the presence of multiple copies of these leucine-rich repeat-containing genes (Wang et al., 2003). Differential gene expression within gene families occurs, although it is unclear whether gene expression is controlled so that single gene family members are expressed at any one time, as seen in important *Trypanosoma* and *Plasmodium* protein families.

In common with the protozoan parasite *Trichomonas vaginalis* (Lal et al., 2005) and the free-living *Tetrahymena thermophyla* (Bright et al., 2010), *Entamoebae* encode a very large number of Rab GTPases. These genes control vesicular

trafficking in the cell and the size of the gene family points to the importance and complexity of these processes in *Entamoeba*. In *E. histolytica*, 102 Rab GTPases, forming over 16 subfamilies, have been annotated (Saito-Nakano et al., 2005; Nakada-Tsukui et al., 2010). A comparison of the Rab GTPases of *E. histolytica* and *E. invadens* showed a general pattern of conservation of orthologous genes between the two species (Nakada-Tsukui et al., 2010; see Fig. 4b). This indicates that the expansion of the gene family largely occurred prior to the divergence of the two species, and suggests that complex Rab GTPase-controlled vesicular trafficking is an important feature of the genus and its machinery is conserved.

Important differences are seen between species' gene family repertoires. Families encoding heavy- and light-chain

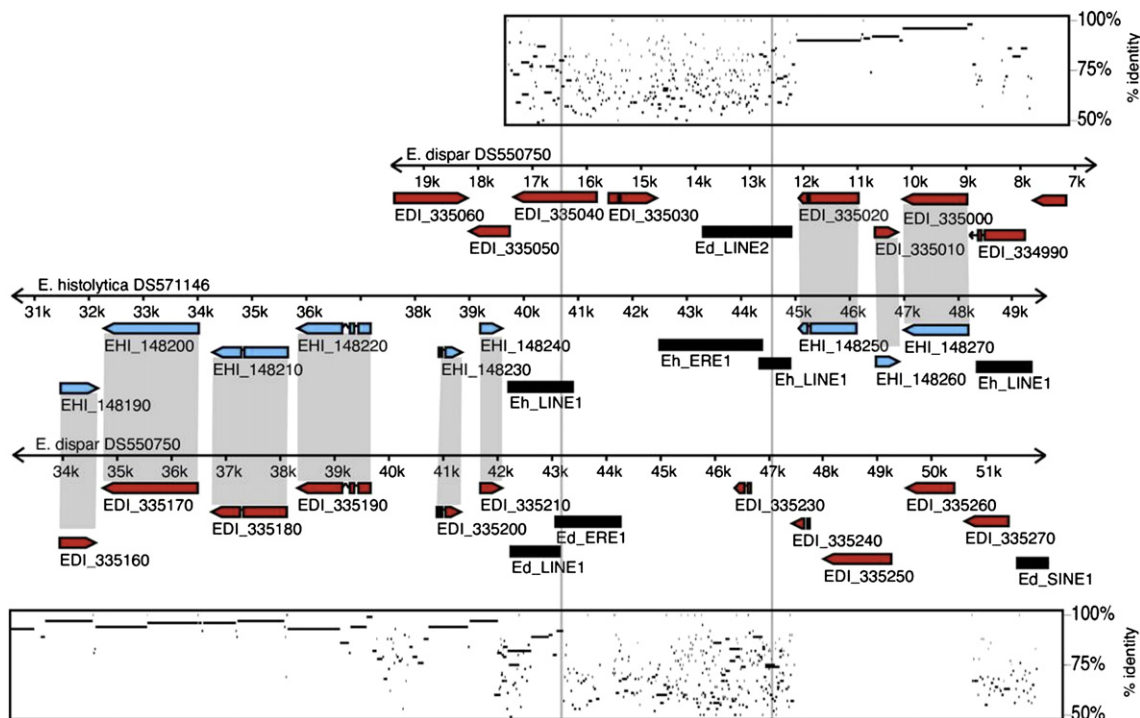


Fig. 3. Example of genome rearrangement between *E. histolytica* and *E. dispar* associated with the presence of transposons, from Lorenzi et al. (2008). Synteny (represented by grey bars connecting orthologous *E. histolytica* and *E. dispar* genes) switches to a different *E. dispar* scaffold across the region containing LINE, SINE and ERE elements. This is also shown by percentage sequence identity plots which show that scaffolds lose significant sequence identity between species across the LINE, SINE and ERE elements.

subunits of the virulence factor Gal/GalNAc lectin occur in multiple *Entamoeba* species, but the Gal/GalNAc lectin intermediate-chain subunit genes have not been detected in species other than *E. histolytica* and *E. dispar* (Clark et al., 2007). The cysteine protease family occurs in both *E. histolytica* and *E. dispar* (Bruchhaus et al., 1996), but the key virulence factor cysteine protease-5 is a pseudogene in *E. dispar* (Willhoeft et al., 1999b). Southern blot evidence indicates that the ariel surface proteins in *E. histolytica* are not present, or are highly divergent, in *E. dispar* (Willhoeft et al., 1999a).

## 6. Genetic diversity and population structure within *E. histolytica*

The *E. histolytica* genome does not appear to contain microsatellites. Therefore, measurement of genetic diversity and estimation of population structures has relied upon other genetic markers, among them genes containing polymorphic internal repeat regions such as that encoding the serine-rich *Entamoeba Histolytica/Dispar* Protein (SREHP/SREDP) (Ayeh-Kumi et al., 2001; Haghghi et al., 2002, 2003, 2008; Simonishvili et al., 2005; Rivera et al., 2006; Samie et al., 2003) and chitinase (Haghghi et al., 2002, 2003). The tRNA-STR loci of *E. histolytica* have proved to be useful population genetic markers (Zaki and Clark, 2001; Zaki et al., 2003; Ali et al., 2005, 2007; Escueta-de Cadiz et al., 2010) and have been used to identify genotypes associated with different clinical manifestations (Ali et al., 2007; Escueta-de Cadiz

et al., 2010). Studies of genetic diversity based upon variation in repetitive DNA (tRNA-STR, SREHP, chitinase) often indicate very high levels of diversity circulating in populations of *E. histolytica* (Ayeh-Kumi et al., 2001; Haghghi et al., 2002; Zaki et al., 2003; Ali et al., 2007; Samie et al., 2008; Escueta-de Cadiz et al., 2010) and *E. dispar* (Haghghi et al., 2008; Mojarad et al., 2009).

In contrast, studies of single nucleotide polymorphism suggest more limited diversity. Comparative genomic hybridisation studies of *E. histolytica* and *E. dispar* strains (Fig. 5) suggest that genome-wide diversity among *E. histolytica* strains is rather low (Shah et al., 2005). Sequence analysis of defined regions supports this (Bhattacharya et al., 2005; Ghosh et al., 2000). Such low single nucleotide diversity suggests a relatively recent common ancestor for *E. histolytica*. This is not inconsistent with a more rapid mutation mechanism leading to diverse repetitive regions. Such a model of population history has been proposed for *E. histolytica* (Ghosh et al., 2000).

Patterns of polymorphism often reflect population structures. In Japan, diversity among parasites infecting men who have sex with men is high, while diversity is much more limited among parasites infecting residents of institutions (Haghghi et al., 2002). Similar low diversity among parasites infecting residents of institutions was seen in the Philippines, where clear population structuring was observed within and between locations (Rivera et al., 2006). In South Africa, genotypes clustered within households but showed extensive diversity among different households (Zaki et al., 2003).

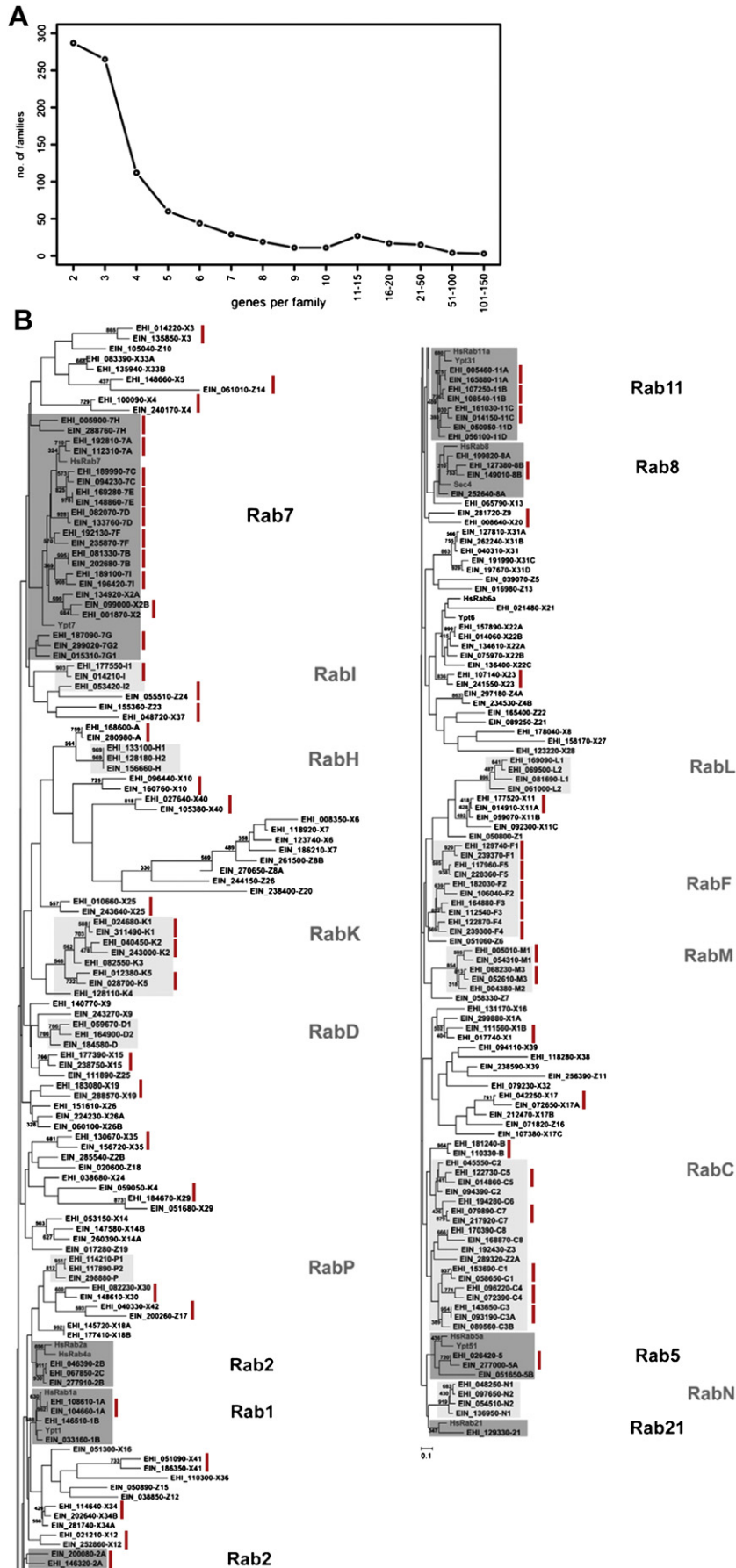


Fig. 4. Large gene families in *Entamoeba*. (A) Size distribution of *E. histolytica* multi-gene families, showing a number of very large families, from Lorenzi et al. (2010). (B) Phylogeny of Rab GTPase genes in *E. histolytica* and *E. invadens*, adapted from Nakada-Tsukui et al. (2010). Putative orthologous gene pairs are indicated with red bars, and their frequency indicates an ancient radiation of Rab GTPase genes.

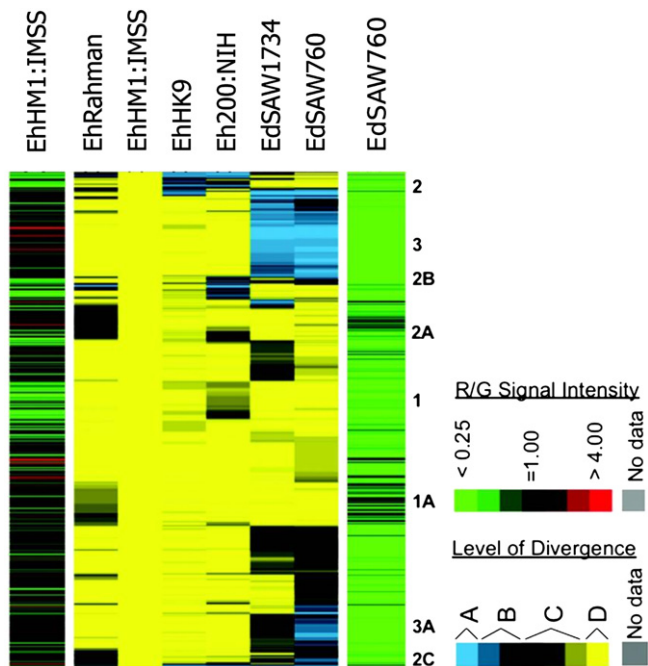


Fig. 5. Genome-scale genetic diversity in *E. histolytica* and divergence from *E. dispar*, from Shah et al. (2005). The array was made from an *Eh*HM1:IMSS clone library. Relative hybridisation to the array was colour-coded to represent divergence: A = blue = absent or highly divergent; B = significantly divergent; C = moderately divergent; D = yellow = highly conserved; grey = missing data (see Shah et al., 2005 for definitions). The genomic abundance for each clone is represented by a green to red scale (green for low genomic abundance,  $R/G < 0.25$ , red for high genomic abundance,  $R/G > 4.0$ ) in the leftmost and in the rightmost columns, for *E. histolytica* (HM1:IMSS) and *E. dispar* (SAW760), respectively. Labels to the right of the figure indicate patterns of variation: 1 and 1A = divergent in one *Eh* strain; 2, 2A, 2B, and 2C = divergent in *Eh* and *Ed* strains; 3 and 3A = divergent in both *Ed* strains but conserved in all *Eh* strains.

Repetitive DNA markers appear to be stable enough to link closely related parasites, recently transmitted among members of a household, an institution, or recent sexual partners (Salit et al., 2009). However, the variability of these markers, as indicated by the extremely extensive diversity they show in endemic populations, suggests that larger-scale, longer-term population structure may be undetectable using them, and that SNP markers may be preferable in some situations.

An important unanswered question about the population structure of *Entamoeba* is whether they are predominantly sexual or clonal. The *E. histolytica* genome project revealed a complement of genes necessary for meiosis, pointing to the possibility of sex in natural populations (Ramesh et al., 2005; Logsdon, 2008; Loftus et al., 2005; Stanley, 2005). This is significant because sexual populations can exchange genes such as virulence and drug resistance genes, generating selectively advantageous genotypes that can spread rapidly through populations. Genetic exchange has been demonstrated in other species of protozoa previously believed to be clonal. *Giardia lamblia* (Poxleitner et al., 2008; Cooper et al., 2007), *Leishmania major* (Akopyants et al., 2009), *Trypanosoma brucei* (Gibson et al., 2008) and *Trypanosoma cruzi* (Gaunt et al., 2003) all show evidence of sex. Determining patterns of linkage

disequilibrium between genetic markers in endemic *E. histolytica* populations (specifically, does linkage disequilibrium decrease with increasing distance between genetic markers as in obligately sexual protozoa such as *P. falciparum* (Conway et al., 1999)?) should allow us to answer the question definitively.

## 7. Concluding remarks

Many questions remain concerning the evolution of *Entamoeba* species, related to the complex architecture of the genome and to the structure of *Entamoeba* populations. The ability to rapidly generate whole genome sequences may help to answer some of these questions, both by allowing comparative analyses of genomes to be undertaken within and between species and by identifying genetic markers for use in molecular epidemiological studies. The challenge is to translate these new data into more effective interventions against the disease.

## Acknowledgements

We thank journals and colleagues for permission to reprint their data.

## References

- Adam, R.D., 1992. Chromosome-size variation in *Giardia lamblia*: the role of rDNA repeats. *Nucleic Acids Res.* 20, 3057–3061.
- Akopyants, N.S., Kimblin, N., Secundino, N., Patrick, R., Peters, N., Lawyer, P., Dobson, D.E., Beverley, S.M., Sacks, D.L., 2009. Demonstration of genetic exchange during cyclical development of *Leishmania* in the sand fly vector. *Science* 324, 265–268.
- Ali, I.K., Hossain, M.B., Roy, S., Ayeh-Kumi, P.F., Petri Jr., W.A., Haque, R., Clark, C.G., 2003. *Entamoeba moshkovskii* infections in children, Bangladesh. *Emerg. Infect. Dis.* 9, 580–584.
- Ali, I.K., Zaki, M., Clark, C.G., 2005. Use of PCR amplification of tRNA gene-linked short tandem repeats for genotyping *Entamoeba histolytica*. *J. Clin. Microbiol.* 43, 5842–5847.
- Ali, I.K., Mondal, U., Roy, S., Haque, R., Petri Jr., W.A., Clark, C.G., 2007. Evidence for a link between parasite genotype and outcome of infection with *Entamoeba histolytica*. *J. Clin. Microbiol.* 45, 285–289.
- Alsmark, U.C., Sicheritz-Ponten, T., Foster, P.G., Hirt, R.P., Embley, T.M., 2009. Horizontal gene transfer in eukaryotic parasites: a case study of *Entamoeba histolytica* and *Trichomonas vaginalis*. *Methods Mol. Biol.* 532, 489–500.
- Andersson, J.O., Hirt, R.P., Foster, P.G., Roger, A.J., 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *BMC Evol. Biol.* 6, 27.
- Aurrecochea, C., Barreto, A., Brestelli, J., Brunk, B.P., Caler, E.V., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., Heiges, M., Iodice, J., Kissinger, J.C., Kraemer, E.T., Li, W., Nayak, V., Pennington, C., Pinney, D.F., Pitts, B., Roos, D.S., Srinivasamoorthy, G., Stoeckert Jr., C.J., Treatman, C., Wang, H., 2011. AmoebaDB and MicrosporidiaDB: functional genomic resources for Amoebozoa and Microsporidia species. *Nucleic Acids Res.* 39 (Database issue), D612–D619.
- Ayeh-Kumi, P.F., Ali, I.M., Lockhart, L.A., Gilchrist, C.A., Petri Jr., W.A., Haque, R., 2001. *Entamoeba histolytica*: genetic diversity of clinical isolates from Bangladesh as demonstrated by polymorphisms in the serine-rich gene. *Exp. Parasitol.* 99, 80–88.
- Báez-Camargo, M., Riverón, A.M., Delgadillo, D.M., Flores, E., Sánchez, T., García-Rivera, G., Orozco, E., 1996. *Entamoeba histolytica*: gene linkage groups and relevant features of its karyotype. *Mol. Gen. Genet.* 253, 289–296.

- Bhattacharya, S., Som, I., Bhattacharya, A., 1998. The ribosomal DNA plasmids of *Entamoeba*. *Parasitol. Today* 14 (5), 181–185.
- Bhattacharya, D., Haque, R., Singh, U., 2005. Coding and noncoding genomic regions of *Entamoeba histolytica* have significantly different rates of sequence polymorphisms: implications for epidemiological studies. *J. Clin. Microbiol.* 43, 4815–4819.
- Billler, L., Davis, P.H., Tillack, M., Matthiesen, J., Lotter, H., Stanley Jr., S.L., Tannich, E., Bruchhaus, I., 2010. Differences in the transcriptome signatures of two genetically related *Entamoeba histolytica* cell lines derived from the same isolate with different pathogenic properties. *BMC Genomics* 11, 63.
- Blaineau, C., Bastien, P., Rioux, J.A., Roizès, G., Pagès, M., 1991. Long-range restriction maps of size-variable homologous chromosomes in *Leishmania infantum*. *Mol. Biochem. Parasitol.* 46, 292–302.
- Bright, L.J., Kambesis, N., Nelson, S.B., Jeong, B., Turkewitz, A.P., 2010. Comprehensive analysis reveals dynamic and evolutionary plasticity of Rab GTPases and membrane traffic in *Tetrahymena thermophila*. *PLoS Genet.* 6, e1001155.
- Bruchhaus, I., Jacobs, T., Leippe, M., Tannich, E., 1996. *Entamoeba histolytica* and *Entamoeba dispar*: differences in numbers and expression of cysteine proteinase genes. *Mol. Microbiol.* 22, 255–263.
- Clark, C.G., Ali, I.K., Zaki, M., Loftus, B.J., Hall, N., 2006a. Unique organisation of tRNA genes in *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 146, 24–29.
- Clark, C.G., Kaffashian, F., Tawari, B., Windsor, J.J., Twigg-Flesner, A., Davies-Morel, M.C., Blessmann, J., Ebert, F., Peschel, B., Le Van, A., Jackson, C.J., Macfarlane, L., Tannich, E., 2006b. New insights into the phylogeny of *Entamoeba* species provided by analysis of four new small-subunit rRNA genes. *Int. J. Syst. Evol. Microbiol.* 56, 2235–2239.
- Clark, C.G., Alsmark, U.C., Tazreiter, M., Saito-Nakano, Y., Ali, V., Marion, S., Weber, C., Mukherjee, C., Bruchhaus, I., Tannich, E., Leippe, M., Sichert-Ponten, T., Foster, P.G., Samuelson, J., Noël, C.J., Hirt, R.P., Embley, T.M., Gilchrist, C.A., Mann, B.J., Singh, U., Ackers, J.P., Bhattacharya, S., Bhattacharya, A., Lohia, A., Guillén, N., Duchêne, M., Nozaki, T., Hall, N., 2007. Structure and content of the *Entamoeba histolytica* genome. *Adv. Parasitol.* 65, 51–190.
- Conway, D.J., Roper, C., Oduola, A.M., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F., Greenwood, B.M., 1999. High recombination rate in natural populations of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 96, 4506–4511.
- Cooper, M.A., Adam, R.D., Worobey, M., Sterling, C.R., 2007. Population genetics provides evidence for recombination in *Giardia*. *Curr. Biol.* 17, 1984–1988.
- Davis, P.H., Zhang, Z., Chen, M., Zhang, X., Chakraborty, S., Stanley Jr., S.L., 2006. Identification of a family of BspA like surface proteins of *Entamoeba histolytica* with novel leucine rich repeats. *Mol. Biochem. Parasitol.* 145, 111–116.
- Davis, P.H., Schulze, J., Stanley Jr., S.L., 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.
- Dhar, S.K., Choudhury, N.R., Bhattacharaya, A., Bhattacharya, S., 1995. A multitude of circular DNAs exist in the nucleus of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 70, 203–206.
- Diamond, L.S., Clark, C.G., 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J. Eukaryot. Microbiol.* 40, 340–344.
- Escueta-de Cadiz, A., Kobayashi, S., Takeuchi, T., Tachibana, H., Nozaki, T., 2010. Identification of an avirulent *Entamoeba histolytica* strain with unique tRNA-linked short tandem repeat markers. *Parasitol. Int.* 59, 75–81.
- Field, J., Rosenthal, B., Samuelson, J., 2000. Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.* 38, 446–455.
- Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., Miles, M.A., 2003. Mechanism of genetic exchange in American trypanosomes. *Nature* 421, 936–939.
- Gibson, W., Peacock, L., Ferris, V., Williams, K., Bailey, M., 2008. The use of yellow fluorescent hybrids to indicate mating in *Trypanosoma brucei*. *Parasit. Vectors* 1, 4.
- Ghosh, S., Frisardi, M., Ramirez-Avila, L., Descoteaux, S., Sturm-Ramirez, K., Newton-Sanchez, O.A., Santos-Preciado, J.I., Ganguly, C., Lohia, A., Reed, S., Samuelson, J., 2000. Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (Demographic sweep) and transcontinental spread of diploid parasites. *J. Clin. Microbiol.* 38, 3815–3821.
- Gilchrist, C.A., Houpt, E., Trapaidze, N., Fei, Z., Crasta, O., Asgharpour, A., Evans, C., Martino-Catt, S., Baba, D.J., Stroup, S., Hamano, S., Ehrenkauf, G., Okada, M., Singh, U., Nozaki, T., Mann, B.J., Petri Jr., W. A., 2006. Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Mol. Biochem. Parasitol.* 147, 163–176.
- Haghighi, A., Kobayashi, S., Takeuchi, T., Masuda, G., Nozaki, T., 2002. Remarkable genetic polymorphism among *Entamoeba histolytica* isolates from a limited geographic area. *J. Clin. Microbiol.* 40, 4081–4090.
- Haghighi, A., Kobayashi, S., Takeuchi, T., Thammapalerd, N., Nozaki, T., 2003. Geographic diversity among genotypes of *Entamoeba histolytica* field isolates. *J. Clin. Microbiol.* 41, 3748–3756.
- Haghighi, A., Rasti, S., Mojarad, E.N., Kazemi, B., Bandehpour, M., Nochi, Z., Hooshyar, H., Rezaian, M., 2008. *Entamoeba dispar*: genetic diversity of Iranian isolates based on serine-rich *Entamoeba dispar* protein gene. *Pak. J. Biol. Sci.* 11, 2613–2618.
- Huntley, D.M., Pandis, I., Butcher, S.A., Ackers, J.P., 2010. Bioinformatic analysis of *Entamoeba histolytica* SINE1 elements. *BMC Genomics* 11, 321.
- Kooij, T.W., Carlton, J.M., Bidwell, S.L., Hall, N., Ramesar, J., Janse, C.J., Waters, A.P., 2005. A *Plasmodium* whole-genome synteny map: indels and synteny breakpoints as foci for species-specific genes. *PLoS Pathog.* 1, e44.
- Lal, K., Field, M.C., Carlton, J.M., Warwicker, J., Hirt, R.P., 2005. Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. *Mol. Biochem. Parasitol.* 143, 226–235.
- Lioutas, C., Schmetz, C., Tannich, E., 1995. Identification of various circular DNA molecules in *Entamoeba histolytica*. *Exp. Parasitol.* 80, 349–352.
- Loftus, B., Anderson, I., Davies, R., Alsmark, U.C., Samuelson, J., Amedeo, P., Roncaglia, P., Berriman, M., Hirt, R.P., Mann, B.J., Nozaki, T., Suh, B., Pop, M., Duchene, M., Ackers, J., Tannich, E., Leippe, M., Hofer, M., Bruchhaus, I., Willhoeft, U., Bhattacharya, A., Chillingworth, T., Churcher, C., Hance, Z., Harris, B., Harris, D., Jagels, K., Moule, S., Mungall, K., Ormond, D., Squares, R., Whitehead, S., Quail, M.A., Rabinowitsch, E., Norbertczak, H., Price, C., Wang, Z., Guillén, N., Gilchrist, C., Stroup, S.E., Bhattacharya, S., Lohia, A., Foster, P.G., Sichert-Ponten, T., Weber, C., Singh, U., Mukherjee, C., El-Sayed, N.M., Petri Jr., W.A., Clark, C.G., Embley, T.M., Barrell, B., Fraser, C.M., Hall, N., 2005. The genome of the protist parasite *Entamoeba histolytica*. *Nature* 433, 865–868.
- Logsdon Jr., J.M., 2008. Evolutionary genetics: sex happens in *Giardia*. *Curr. Biol.* 18, 66–68.
- Lorenzi, H., Thiagarajan, M., Haas, B., Wortman, J., Hall, N., Caler, E., 2008. Genome wide survey, discovery and evolution of repetitive elements in three *Entamoeba* species. *BMC Genomics* 9, 595.
- Lorenzi, H.A., Puiu, D., Miller, J.R., Brinkac, L.M., Amedeo, P., Hall, N., Caler, E.V., 2010. New assembly, reannotation and analysis of the *Entamoeba histolytica* genome reveal new genomic features and protein content information. *PLoS Negl. Trop. Dis.* 4, e716.
- Melville, S.E., Gerrard, C.S., Blackwell, J.M., 1999. Multiple causes of size variation in the diploid megabase chromosomes of African trypanosomes. *Chromosome Res.* 7, 191–203.
- Mukherjee, C., Clark, C.G., Lohia, A., 2008. *Entamoeba* shows reversible variation in ploidy under different growth conditions and between life cycle phases. *PLoS Negl. Trop. Dis.* 2, e281.
- Mojarad, E.N., Haghighi, A., Kazemi, B., Rostami Nejad, M., Abadi, A., Zali, M.R., 2009. High genetic diversity among Iranian *Entamoeba dispar* isolates based on the noncoding short tandem repeat locus D-A. *Acta Trop.* 111, 133–136.
- Nakada-Tsukui, K., Saito-Nakano, Y., Husain, A., Nozaki, T., 2010. Conservation and function of Rab small GTPases in *Entamoeba*: annotation of *E. invadens* Rab and its use for the understanding of *Entamoeba* biology. *Exp. Parasitol.* 126, 337–347.



- Pritham, E.J., Feschotte, C., Wessler, S.R., 2005. Unexpected diversity and differential success of DNA transposons in four species of *entamoeba* protozoans. *Mol. Biol. Evol.* 22, 1751–1763.
- Poxleitner, M.K., Carpenter, M.L., Mancuso, J.J., Wang, C.J., Dawson, S.C., Cande, W.Z., 2008. Evidence for karyogamy and exchange of genetic material in the binucleate intestinal parasite *Giardia intestinalis*. *Science* 319, 1530–1533.
- Ramesh, M.A., Malik, S.B., Logsdon Jr., J.M., 2005. A phylogenomic inventory of meiotic genes; evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Curr. Biol.* 15, 185–191.
- Riveron, A.M., Lopez-Canovas, L., Baez-Camargo, M., Flores, E., Perez-Perez, G., Luna-Arias, J.P., Orozco, E., 2000. Circular and linear DNA molecules in the *Entamoeba histolytica* complex molecular karyotype. *Eur. Biophys. J.* 29, 48–56.
- Rivera, W.L., Santos, S.R., Kanbara, H., 2006. Prevalence and genetic diversity of *Entamoeba histolytica* in an institution for the mentally retarded in the Philippines. *Parasitol. Res.* 98, 106–110.
- Rosenthal, B., Mai, Z., Caplivski, D., Ghosh, S., de la Vega, H., Graf, T., Samuelson, J., 1997. Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. *J. Bacteriol.* 179, 3736–3745.
- Roy, S.W., Irimia, M., Penny, D., 2006. Very little intron gain in *Entamoeba histolytica* genes laterally transferred from prokaryotes. *Mol. Biol. Evol.* 23, 1824–1827.
- Saito-Nakano, Y., Loftus, B.J., Hall, N., Nozaki, T., 2005. The diversity of Rab GTPases in *Entamoeba histolytica*. *Exp. Parasitol.* 110, 244–252.
- Salit, I.E., Khairnar, K., Gough, K., Pillai, D.R., 2009. A possible cluster of sexually transmitted *Entamoeba histolytica*: genetic analysis of a highly virulent strain. *Clin. Infect. Dis.* 49 (3), 346–353.
- Samie, A., Obi, C.L., Bessong, P.O., Houpt, E., Stroup, S., Njayou, M., Sabeta, C., Mduluzi, T., Guerrant, R.L., 2008. *Entamoeba histolytica*: genetic diversity of African strains based on the polymorphism of the serine-rich protein gene. *Exp. Parasitol.* 118, 354–361.
- Shah, P.H., MacFarlane, R.C., Bhattacharya, D., Matese, J.C., Demeter, J., Stroup, S.E., Singh, U., 2005. Comparative genomic hybridizations of *Entamoeba* strains reveal unique genetic fingerprints that correlate with virulence. *Eukaryot. Cell.* 4, 504–515.
- Shire, A.M., Ackers, J.P., 2007. SINE elements of *Entamoeba dispar*. *Mol. Biochem. Parasitol.* 152, 47–52.
- Simonishvili, S., Tsanava, S., Sanadze, K., Chlikadze, R., Miskalishvili, A., Lomkatsi, N., Imnadze, P., Petri Jr., W.A., Trapaidze, N., 2005. *Entamoeba histolytica*: the serine-rich gene polymorphism-based genetic variability of clinical isolates from Georgia. *Exp. Parasitol.* 110, 313–317.
- Stanley Jr., S.L., 2005. The *Entamoeba histolytica* genome: something old, something new, something borrowed and sex too? *Trends Parasitol.* 21, 451–453.
- Tachibana, H., Yanagi, T., Pandey, K., Cheng, X.J., Kobayashi, S., Sherchand, J.B., Kanbara, H., 2007. An *Entamoeba* sp. strain isolated from rhesus monkey is virulent but genetically different from *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 153, 7–14.
- Tawari, B., Ali, I.K., Scott, C., Quail, M.A., Berriman, M., Hall, N., Clark, C.G., 2008. Patterns of evolution in the unique tRNA gene arrays of the genus *Entamoeba*. *Mol. Biol. Evol.* 25, 187–198.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner Jr., D.A., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., Wirth, D.F., 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* 39, 113–119.
- Walsh, J.A., 1986. Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Rev. Infect. Dis.* 8, 228–238.
- Wang, Z., Samuelson, J., Clark, C.G., Eichinger, D., Paul, J., Van Dellen, K., Hall, N., Anderson, I., Loftus, B., 2003. Gene discovery in the *Entamoeba invadens* genome. *Mol. Biochem. Parasitol.* 129, 23–31.
- Willhoeft, U., Buss, H., Tannich, E., 1999a. DNA sequences corresponding to the ariel gene family of *Entamoeba histolytica* are not present in *E. dispar*. *Parasitol. Res.* 85, 787–789.
- Willhoeft, U., Tannich, E., 1999. The electrophoretic karyotype of *Entamoeba histolytica*. *Mol. Biochem. Parasitol.* 99, 41–53.
- Willhoeft, U., Hamann, L., Tannich, E., 1999b. A DNA sequence corresponding to the gene encoding cysteine proteinase 5 in *Entamoeba histolytica* is present and positionally conserved but highly degenerated in *Entamoeba dispar*. *Infect. Immun.* 67, 5925–5929.
- Willhoeft, U., Buss, H., Tannich, E., 2002. The abundant polyadenylated transcript 2 DNA sequence of the pathogenic protozoan parasite *Entamoeba histolytica* represents a nonautonomous non-long-terminal-repeat retrotransposon-like element which is absent in the closely related nonpathogenic species *Entamoeba dispar*. *Infect. Immun.* 70, 6798–6804.
- Zaki, M., Clark, C.G., 2001. Isolation and characterization of polymorphic DNA from *Entamoeba histolytica*. *J. Clin. Microbiol.* 39, 897–905.
- Zaki, M., Reddy, S.G., Jackson, T.F., Ravdin, J.I., Clark, C.G., 2003. Genotyping of *Entamoeba* species in South Africa: diversity, stability, and transmission patterns within families. *J. Infect. Dis.* 187, 1860–1869.