1. Introduction
Parasites assigned to the genus *Entamoeba* are unicellular eukaryotes and parasitise few classes of vertebrates, invertebrates and possibly other unicellular eukaryotes also. Taxonomical classification of *Entamoeba* and other related protozoa is shown in Table 1. All species have a simple life cycle that usually consists of an infective cyst stage and a multiplying trophozoite stage. Transmission of the infection occurs via ingestion of cysts through faecally contaminated food or water. Humans can be host to at least six species of *Entamoeba* in addition to several amoebae belonging to other genera (Fig. 1), (Table 2). Differentiation between the *Entamoeba* species is based on a number of factors and characteristics of trophozoites and cysts: size, number of nuclei and general morphology are the most important factors. Other characteristics such as motility of trophozoites are also useful. At the present time there is only one of the *Entamoeba* species, *Entamoeba histolytica* pathogenic to human. *E. histolytica* is a microaerophilic, enteric protozoan parasite that infects human causing amoebiasis. With an estimated 40 million cases worldwide and 100,000 deaths, it is responsible for a great deal of suffering, particularly in the developing countries.

### 1.1. History

*E. histolytica* was first identified by Feder Losch in 1875 while working with human gut infecting organisms (Losch, 1875). However, it was only later in 1903 that a formal classification of this organism was done by Fritz Schaudinn (Schaudinn et al., 1903). Clifford Dobell further shed light on the unusual properties of this parasite. He classified the related organisms into three species based on the number of nuclei present in the cyst – *E. histolytica*, *E. coli* and *E. gingivalis* (Dobell, 1919).

### 1.2. *E. histolytica* VS *E. dispar*

*E. histolytica* and *E. dispar* are two morphologically similar but genetically distinct species which infect humans. One of them is pathogenic (*E. histolytica*) while the other is nonpathogenic (*E. dispar*). Until the late 1970s both of them were
<table>
<thead>
<tr>
<th>PHYLUM</th>
<th>SUBPHYLUM</th>
<th>CLASS</th>
<th>ORDER</th>
<th>GENERA</th>
</tr>
</thead>
<tbody>
<tr>
<td>SARCOMASTIGOPHORA</td>
<td>MASTIGOPHORA</td>
<td>ZOOMASTIGOPHOREA</td>
<td>RETORTAMONADIDA</td>
<td>CHILOMASTIX</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TRICHOMONADIDA</td>
<td>DIENTAMOEBA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DIPLOMONADIDA</td>
<td>GIARDIA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KINETOPLASTIDA</td>
<td>LEISHMANIA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TRYPANOSOMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AMOEBIDAE</td>
<td>ENDOLIMAX</td>
</tr>
<tr>
<td></td>
<td>SARCODINA</td>
<td>LOBOSEA</td>
<td></td>
<td>ENTAMOEBA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IODAMOEBA</td>
</tr>
</tbody>
</table>

Table 1: Classification of *Entamoeba* and other protozoa
**AMEBAE**

<table>
<thead>
<tr>
<th></th>
<th>Entamoeba histolytica</th>
<th>Entamoeba hartmanni</th>
<th>Entamoeba coli</th>
<th>Entamoeba polecki</th>
<th>Endolimax nana</th>
<th>Iodamoeba bütschli</th>
<th>Dientamoeba fragilis 1</th>
</tr>
</thead>
</table>

1 Rare, probably of animal origin  
2 Flagellate

**Scale:** 0 5 10 μm  
Adapted from Brooke and Melvin, 1964

**Fig. 1:** Protozoan found in human stool specimens (source: http://www.dpd.cdc.gov/dpdx/Default.htm)
<table>
<thead>
<tr>
<th>Species</th>
<th>Size (Length)</th>
<th>Motility</th>
<th>Number</th>
<th>Peripheral Chromatin Description</th>
<th>Karyosomal Chromatin Description</th>
<th>Appearance Description</th>
<th>Inclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entamoeba hartmanni</td>
<td>5-12 mm. Usual range, 8-10 mm.</td>
<td>Usually nonprogressive but may be progressive occasionally.</td>
<td>One, (Not visible in unstained preparations).</td>
<td>Similar to E. histolytica.</td>
<td>Small, discrete, often eccentric.</td>
<td>Finely granular.</td>
<td>Bacteria.</td>
</tr>
<tr>
<td>Entamoeba coli</td>
<td>15-50 mm. Usual range, 20-25 mm.</td>
<td>Sluggish, nonprogressive, with blunt pseudopods.</td>
<td>One, (Often visible in unstained preparations)</td>
<td>Coarse granules, irregular in size and distribution.</td>
<td>Large, discrete, usually eccentric.</td>
<td>Coarse, often vacuolated</td>
<td>Bacteria, yeasts, other materials</td>
</tr>
<tr>
<td>Entamoeba polecki</td>
<td>10-25 mm. Usual range, 15-20 mm.</td>
<td>Similar to E. coli. Occasionally, in diarrheaic specimens, motility may be progressive.</td>
<td>One, (May be slightly visible in unstained preparations)</td>
<td>Usually fine granules evenly distributed. Occasionally granules may be irregularly arranged. Chromatin sometimes in plaques or crescents.</td>
<td>Small, discrete, eccentric. Occasionally large, diffuse or irregular.</td>
<td>Coarsely, granular, may resemble E. coli. Contains numerous vacuoles.</td>
<td>Bacteria, yeasts.</td>
</tr>
<tr>
<td>Endolimax nana</td>
<td>6-12 mm. Usual range, 8-10 mm.</td>
<td>Sluggish, usually nonprogressive with blunt pseudopods.</td>
<td>One, (Visible occasionally in unstained preparations)</td>
<td>None.</td>
<td>Large, irregularly shaped, blot-like.</td>
<td>Granular, vacuolated.</td>
<td>Bacteria.</td>
</tr>
<tr>
<td>Iodamoeba buetschli</td>
<td>8-20 mm. Usual range, 12-15 mm.</td>
<td>Sluggish, usually nonprogressive</td>
<td>One, (Not usually visible in unstained preparations)</td>
<td>None.</td>
<td>Large, usually central. Surrounded by refractile, achromatic granules. These granules are often not distinct even in stained slides.</td>
<td>Coarsely granular, vacuolated</td>
<td>Bacteria, yeasts, or other material.</td>
</tr>
<tr>
<td>Dientamoeba fragilis</td>
<td>5-15 mm. Usual range, 9-12 mm.</td>
<td>Pseudopods are angular, serrated, or broad lobed, and hyaline, almost transparent.</td>
<td>Two, (In approximately 20% of organisms only 1 nucleus is present.) Nuclei invisible in unstained preparations.</td>
<td>None.</td>
<td>Large cluster of 4-8 granules.</td>
<td>Finely, granular.</td>
<td>Bacteria: occasionally red blood cells.</td>
</tr>
</tbody>
</table>

Table 2. Differential Morphology of Protozoa found in human stool specimens (source: http://www.dpd.cdc.gov/dpdx/Default.htm)
Introduction

considered to be same species. Although the two species hypothesis was first proposed in 1925 by Brumpt, these two organisms were classified under *E. histolytica*, due to their close similarity. Substantial biochemical and molecular data has accumulated over the last 2-3 decades, which indicated that the nonpathogenic isolates of *E. histolytica* were genetically distinct from the pathogenic isolates (Diamond and Clark, 1991; Diamond and Clark, 1993). Some of the first noted differences were the zymodeme patterns, in vitro growth characteristics, agglutination with concanavalin A (conA), and resistance to complement lysis. Isoenzyme analysis revealed different zymodemes for the pathogenic and non-pathogenic isolates (Sargeaunt et al., 1987; Sargeaunt et al., 1994 and Strachan et al., 1988). Pathogenic strains have the ability to grow in axenic cultures (without bacteria) whereas nonpathogenic strains required bacteria for in vitro growth. The ConA agglutination (Martinez-Palamo et al., 1973) and complement resistance implied that the outer surfaces of the pathogenic and nonpathogenic strains were different. Several antigenic differences in surface proteins have been noted and nonpathogenic isolates have been reported to lack a particular glycoconjugate on their surface (Bhattacharya et al, 2000).

Analysis of DNA and sequencing of several genes also revealed genotypic differences between the pathogenic and non-pathogenic isolates (Tannich et al., 1989). The most striking variation at nucleotide level is the 2.2% difference between the ribosomal RNA gene sequence of pathogenic and non-pathogenic isolates (Clark and Diamond, 1991). Slight morphological differences between the pathogenic and nonpathogenic isolates were also noted (Espinosa-Cantellano et al, 1998).

Finally, after around 70 years, *E. dispar* was formally separated from *E.histolytica* and this classification was universally adopted in the field of amebiasis research (Diamond and Clark, 1993). The acceptance of *E. histolytica* and *E. dispar* as distinct species has had major impact on our views of amebiasis, in particular its clinical management.
Main biological features that distinguish *E. histolytica* from *E. dispar* are as follows

- Isoenzyme patterns, particularly hexokinase.
- Specific epitopes, recognized by reaction with several monoclonal antibodies.
- Sequence differences in the rDNA episome.
- Significant (2–18%) sequence differences between homologous genes.
- A small number of genes, for example *ariel* (Willhoeft *et al.* 1999a) and *cp5* (Bruchhaus *et al.* 1996; Willhoeft *et al.* 1999b) appear so far to be unique to *E. histolytica*.
- An abundant polyadenylated transcript 2 (ehapt2) is absent in *E. dispar* (Willhoeft *et al.*, 2002).
- It has proved much easier to adapt *E. histolytica* to axenic growth. Axenic culture of *E. dispar* proved extremely difficult and has so far been achieved for only one strain (Clark 1995).
- Scanning electron microscopic examination of axenic cultures of both species shows significant differences—particularly in the appearance of the surface (Clark *et al.* 2000). This may be linked to the apparent lack of surface lipophosphoglycan (LPG) from *E. dispar* (Bhattacharya *et al.* 2000).

### 1.3. Life cycle

There are two morphologically distinguishable stages in the life cycle of *Entamoeba* the infective non motile dormant form known as the cyst and the non infective, motile and dividing form known as trophozoite (Fig. 2). Infection in human starts with the ingestion of cysts present in the food and water contaminated with fecal matter. The cysts pass through the gastrointestinal tract into ileocecal region, where under favorable conditions, they excyst to produce eight trophozoites per cyst. The trophozoites migrate to the colon and continue to divide further and colonize the host tissues, causing dysentery. They are however incapable of causing fresh infection in another individual. Under certain
Fig. 2: Life cycle of E. histolytica

Cysts are passed in feces 1. Infection by Entamoeba histolytica occurs by ingestion of mature cysts 2 in fecally contaminated food, water, or hands. Excystation 3 occurs in the small intestine and trophozoites 4 are released, which migrate to the large intestine. The trophozoites multiply by binary fission and produce cysts 5, which are passed in the feces 1. Because of the protection conferred by their walls, the cysts can survive days to weeks in the external environment and are responsible for transmission. (Trophozoites can also be passed in diarrheal stools, but are rapidly destroyed once outside the body, and if ingested would not survive exposure to the gastric environment.) In many cases, the trophozoites remain confined to the intestinal lumen (A: noninvasive infection) of individuals who are asymptomatic carriers, passing cysts in their stool. In some patients the trophozoites invade the intestinal mucosa (B: intestinal disease), or, through the bloodstream, extraintestinal sites such as the liver, brain, and lungs (C: extraintestinal disease), with resultant pathologic manifestations.
conditions the trophozoites penetrate the intestinal mucosa and give rise to ulcers. They can further move on to the blood stream, from where they travel to various organs like the liver and brain causing abscesses, which are fatal. Some of the trophozoites in the colon encyst to produce quadrinucleated cysts, which are released with the faeces. Cysts have hard outer wall made up of chitin and can survive in the outside environment in a dormant state.

Within the genus *Entamoeba*, different species produce cysts with one, four, or eight nuclei. There is no cyst stage in few species, such as *E. gingivalis*, which implies that direct transmission must be taking place through saliva (Clark et al., 2000). In *E. histolytica* the mature cyst contains four nuclei. The process of excystation has been studied in vitro (Marinets et al., 1997) and in monkeys with identical results (Espinosa-Cantellano and Martinez-Palomino, 1991). The first stage involves movement of the amoeba within the cyst wall. This is followed by the thinning of the wall at one position through which the trophozoite emerges, usually after numerous extensions and retractions of pseudopodia, leaving behind an empty cell wall. The multinucleated trophozoite may start feeding even before fully emerging from the cyst. The four nucleated amoeba undergoes three rounds of cytokinesis and one round of nuclear division to give rise to eight daughter amoebae (Marinets et al., 1997).

### 1.4. Cellular Organization

*E. histolytica* was for many years used to illustrate the probable structure of an ancestral eukaryote (Meza, 1992; Bakker-Grunwald and Worstmann, 1993). Electron microscopy revealed the absence of many characteristic organelles of eukaryotes—mitochondria, Golgi apparatus, rough endoplasmic reticulum, microtubular cytoskeleton (Martinez-Palomino, 1986).

The following description of *E. histolytica*, closely mirrors that of the Diamond and Clark, but none of the essential features recorded has changed since the description of (Dobell in 1919) with the exception of the size ranges. It is also based on organisms isolated directly from the host; *E. histolytica* in axenic culture has quite different appearance.
1.4.1. Trophozoite
The Trophozoites of *E. histolytica* is a highly dynamic and pleomorphic cell, whose form and motility are greatly affected by changes in temperature, pH, osmolarity and redox potential. *E. histolytica* trophozoites in axenic culture tend to be rounder in shape. The diameter of the cell varies between 20\(\mu\)m-60\(\mu\)m. Locomotion is by means of a single well defined pseudopodium (Espinosa-Cantelloano et al., 1998) often extended eruptively, without clear differentiation between ectoplasm and endoplasm. There is normally a single nucleus. The cytoplasm often has ingested red blood cells when the trophozoites are isolated from symptomatic individuals. Sometimes leukocytes and bacteria are visible. The cytoplasm is rich in glycogen and has ribosomes arranged in helices which aggregate. There are no classical mitochondria, rough endoplasmic reticulum or Golgi apparatus visible.

1.4.2. Cyst
*E. histolytica* cysts are round or slightly oval hyaline bodies 10\(\mu\)m-16\(\mu\)m in diameter. They are surrounded by a refractile wall, 125-150nm thick, which appears to be made of fibrillar material forming a tight mesh that may give rise to several lamellae. In *E. invadens* the cyst wall has been shown to contain chitin (Vega et al., 1997). When stained with iron hematoxylin, the cyst cytoplasm appears vacuolated with numerous glycogen deposits that decrease in size and number as the cyst matures. Chromatid bodies, which are aggregated ribosomes, can be identified inside the cytoplasm as rod shaped structures with blunt or rounded ends. Cyst nuclei are morphologically similar to those of trophozoites, but are smaller in the mature cyst.

1.4.3. Nuclear structure and chromatin
The nucleus of *E. histolytica* is 4\(\mu\)m-7\(\mu\)m in diameter. The nuclear membrane appears as double membrane, interrupted by numerous nuclear pores approximately 65nm in diameter. In *E. histolytica* the nuclear periphery is coarse and thick. Chromatin clumps are usually uniform in size and evenly distributed
inside the nuclear membrane although in some cells the chromatin appears concentrated on one side as a crescent shaped mass. The karyosome is a small spherical mass approximately 0.5 μm in diameter, located in the central part of the nucleus. In situ hybridization experiments have identified the ribosomal RNA genes of *E. histolytica* in the peripheral chromatin (Zurita et al., 1991). The karyosome is therefore not equivalent to the nucleolus of other eukaryotes. Intranuclear bodies 0.1-0.2 μm in diameter are frequently observed but their nature and function are unknown. The genes encoding histones H1 (Scharfetter et al., 1997), H2B, (Sanchez et al., 1994), H3 (Fodinger et al., 1993) and H4 (Binder et al., 1995) have been identified and partially characterized in *E. histolytica*. The presence of a homologue of H2A has also been shown by EST analysis (Willhoeft et al., 1999). The nucleotide sequence of these genes was found to be different from other eukaryotes.

The chromatin structure of *E. histolytica* is not clear although some form of nucleosomal structure is thought to be present (Torres-Guerrero et al., 1991). *E. histolytica* chromatin does not condense and this may be either due to the divergent nature of the histones (Scharfetter et al., 1997; Sanchez et al., 1994) or to some property of the non-histone chromatin proteins. Nuclear division proceeds without dissolution of the nuclear membrane and involves the participation of a thick microtubule spindle. Individual microtubules can be identified in cryofixed and cyrosubstituted samples (Gonzalez-Robels, 1997). A gene encoding gamma tubulin, a protein specifically associated with spindle formation, has been identified (Ray et al., 1997). With DNA binding fluorescent dyes, six DNA containing plates can be seen by light microscopy in dividing nuclei, possibly corresponding to some of the organism’s chromosomes (Arguello et al., 1992). The doubling time during exponential growth in axenic culture has been calculated to be 7-8 hours (Dvorak et al., 1995) but little is known about the molecular mechanisms controlling growth and proliferation of the parasite. Several homologues of genes involved in regulation of the cell division cycle and signal transduction pathways in other eukaryotic cells have been isolated from *E. histolytica*; Eh cdc2 (Lohia and Samuelson, 1993a), Eh ras, Eh rap (Shen et al.,
1.5. Pathogenesis

The pathogenesis of infection by *E. histolytica* is governed at several levels, chief among them are:

1. Adherence of trophozoite to the target cell,
2. Lysis of the target cell and
3. Phagocytosis of target cell.

Several molecules which are involved in this process have been identified. Recently different technical advances, like development of transfection system to introduce genes into trophozoites are helping to understand the mechanism of pathogenesis in amoebiasis. With the wealth of information from the ongoing genome sequencing efforts of *E. histolytica* by TIGR and Sanger center, a genomics approach to identify novel virulence determinants in the parasite and their role in causing invasive disease has been undertaken. The interaction between the parasite and the host cells is multifactorial. In the first step the amoeba recognize the host cells by using a number of surface adhesins. In the second stage various effector molecules cause damage to the cells. Our knowledge about *E. histolytica* pathogenesis is still incomplete. However three pathogenic factors have been extensively studied and characterized at molecular level.

**A. Gal/GalNAc** lectin borne on surface mediates adherence of the trophozoite to human colonic glycoproteins (Chadee et al., 1987), human colonic epithelium (Ravdin et al., 1985), human neutrophils and erythrocytes (Ravdin and Guerrant, 1981; Guerrant et al., 1981; Burchard and Bilke, 1992), a variety of cell culture lines and to bacteria. The mucin layer of the human colon may be the first natural target for binding by trophozoites in the colon. The mucin layer may protect the host from contact dependent cytolysis by neutralizing the binding epitopes on the lectin, while simultaneously providing a site of attachment for the colonizing...
parasite. Penetration through the mucin layer permits the trophozoite to invade and attack host tissue. The Gal/GalNAc lectin is also important in the cytolytic function following adherence. The related avirulent *E. dispar* also expresses a similar lectin on its surface; however, there are differences in epitope specificity, and the levels of Gal/GalNac inhibitable adherence appear to be lower.

**B. Amoebapore** (Leippe et al., 1991) are a family of small proteins contained in cytoplasmic granules in the trophozoite. Trophozoites are able to depolarize target cells by the insertion of these pore forming proteins that form ion channels in lipid membranes. In addition to eukaryotic cell targets, amoebapores are effective in forming pores in bacterial membranes.

**C. Cysteine proteases** are a key virulence factor of *E. histolytica* and play a role in intestinal invasion by degrading the extracellular matrix and circumventing the host immune response through cleavage of secretory immunoglobulin A (sIgA), IgG, and activation of alternate complement pathway. From *E. histolytica* genome database total 20 full-length genes were identified which show 10 to 86% sequence identity (Bruchhaus et al., 2003). Southern blot analysis revealed that orthologous sequences for all of the newly identified proteases are present in *E. dispar*. Majority of the various cysteine protease genes are not expressed in *E. histolytica* or *E. dispar* trophozoites during in vitro cultivation. Therefore, it is likely that at least some of these enzymes are required for infection of the human host and/or for completion of the parasite life cycle. Evidence supporting the role of the extracellular cysteine proteinases of *E. histolytica* as virulence factors include the production and extracellular release of 10- to 1,000-fold more cysteine proteinase from lysates of *E. histolytica* cells than from lysates of noninvasive *E. dispar* (Reed et al., 1989). Cysteine proteinases purified from axenized *E. histolytica* cleave collagen, elastin, fibrinogen and laminin, elements of the extracellular matrix that trophozoites must penetrate to cause invasive disease (Keene et al., 1986; Luaces et al., 1988). Cysteine proteinases are responsible for the detachment of tissue culture monolayers, the most widely used assay for amebic toxins and other virulence factors. The cytopathic effect on fibroblast monolayers with clinical *E. histolytica* strains was completely inhibited by Z-Phe-
Arg-CH2F, a specific, irreversible cysteine proteinase inhibitor that is not toxic to host cells (Reed et al., 1993). In vivo studies by Stanley and Mirelman's groups demonstrated that inhibition of cysteine proteinase activity with inhibitors or an antisense construct significantly decreased liver abscess formation in SCID mice (Stanley et al., 1995) and hamsters (Ankri et al., 1999). Taken together, the data supporting a key role of cysteine proteinases in virulence are extremely strong.

1.6. The genome

1.6.1. Size of the genome

Estimation of genome size of *E. histolytica* by different methods has failed to give comparable values due to inherent complexities of the organism in terms of uncertain level of ploidy, presence of multinucleated cells and poorly demarcated cell division cycle. DNA renaturation kinetics experiments (Gelderman et al., 1971a, Gelderman et al., 1971b) indicated genome size of $10^7$bp, which was also found by pulse field gel electrophoresis (PFGE) data (Willhoeft and Tannich et al., 1999). Flow cytometry (Dvorak et al., 1995) and diphenyl amine reactions put the figures at $10^8$bp. This difference may be due to various factors, including repetitive DNA, which is about 20% of the total genome of *E. histolytica*, the level of ploidy and multinucleate cells that are often seen in axenic cultures.

1.6.2. Chromosomal organization

*Entamoeba* has both linear chromosomes as well as circular plasmid DNA. The presence of linear chromosomes has been shown by Bal 31 susceptibility of the bands separated on PFGE (Bagchi et al., 1999) and direct visualization by FISH (Willhoeft and Tannich 2000). Telomeres and centromeres have not been identified in *Entamoeba*. Tannich's group have shown by extensive hybridization studies of Pulsed Field Gel Electrophoretically separated chromosomes, that there are 14 linkage groups and ploidy could be four (Willhoeft et al., 1999). *E. histolytica* chromosomes separated by PFGE under different pulse and run conditions and by ethidium bromide staining have shown the presence of 17 bands of *E. histolytica* strain HK9 and 14 bands of strain HM1:IMSS ranging in
sizes from 300-1920 and 400-1840 kb respectively (Bagchi et al., 1999; Bhattacharya et al., 2000).

There are a number of extrachromosomal plasmids present in *Entamoeba* (Dhar et al., 1995). Among the circular episomes those that carry the rRNA transcription units have been extensively studied (Bhattacharya et al., 1989). They can be broadly divided into those that contain one or two transcription units of rDNA (Sehgal et al., 1994).

1.6.3. Physical organization of the genome

Eukaryotic genome is generally organized in the form of highly condensed structures with nucleosomes being the basic unit. This allows the chromosomes to condense at metaphase and facilitate analysis visually by light microscopy. However, *Entamoeba* genome does not condense to allow visual observation during mitosis and it is not yet clear how the chromosomes are organized in nucleosomal structures. Electron microscopic studies of chromatin spread revealed different levels of chromatin organization. A nucleosome like particle of 10 nm diameter was seen in the most extended fibers, with a variable internucleosomal length. Progressive digestion of the chromatin with micrococcal nuclease did not produce the characteristic ladder as expected for a typical nucleosomal organization (Torres-Guerrero, 1991). It is speculated that there are two chromatin structures in *Entamoeba* – an active structure devoid of typical nucleosomes and an inactive structure that gives a nucleosomal ladder. Only a small portion of the chromatin is micrococcal nuclease resistant. Proteins involved in the formation of nucleosomal structure, such as histones have not been fully characterized in *Entamoeba*.

1.6.4. Codon usage

Analysis of the codon usage data has immense importance in understanding of the molecular organization of a genome. All codons are not equally used by all organisms. Codon usage is non random and species specific (Grantham et al., 1981). Studies have shown that diverse patterns of codon usage may arise from
compositional constraints of the genome as observed in the case of extremely GC or AT rich organisms. The *Entamoeba* genome has a G+C content of 22.4%. In general, most strains and species are comparable except *E. moshkovoskii* strain Laredo, which was reported to have about 10% higher G+C content compared with the other species of *Entamoeba* (Tannich and Horstmann 1992). Analysis of 50,651 codons showed that *E. histolytica* genes have a preference for NNU (43.1%) and NNA (42.3%) codons. CGG and CGC (arginine) are underrepresented. The codon usage pattern was found to be distinctly different in genes that are highly expressed as compared to the ones that are not highly expressed (Ghosh et al., 2000). C ending codons are significantly higher in the putatively highly expressed genes suggesting that C ending codons are translationally optimal in this organism. In putatively low expressing genes A and G or T coding codons are predominant.

1.6.5. Complete sequencing of the genome

A large scale sequencing project of *E. histolytica* genome has been completed recently by National Institute of Allergy and Infectious Diseases (NIAID) at The Institute of Genomic Research (TIGR) in collaboration with Wellcome trust Sanger Institute to undertake whole genome shotgun sequencing of *E. histolytica* strain HM1:IMSS, analyze and annotate the data and provide equal access to the sequence information and analysis. Genome analysis has been done on a 12.5-fold coverage genome assembly consisting of 23,751,783 base pairs distributed among 888 scaffolds. Analysis suggests presence of 9,938 predicted genes, average 1.17 kilobases in size and comprise 49% of the genome (Loftus et. al., 2005). One-quarter of *E. histolytica* genes are predicted to contain introns, with 6% of genes containing multiple introns. No homologues could be identified for a third of predicted proteins (31.8%) from the public databases.

A variety of metabolic adaptations in *E. histolytica* is shared with two other amitochondrial protist pathogens: *Giardia lamblia* and *Trichomonas vaginalis*. These adaptations include reduction or elimination of most mitochondrial metabolic pathways and the use of oxidative stress enzymes generally
associated with anaerobic prokaryotes. The haploid genome of HM1 is < 20 Mb in 14 chromosomes; and intergenic sequences are short. Phylogenomic analysis identifies evidence for lateral gene transfer of bacterial genes into the *E. histolytica* genome, the effects of which centre on expanding aspects of *E. histolytica*’s metabolic repertoire. The presence of these genes and the potential for novel metabolic pathways in *E. histolytica* may allow for the development of new chemotherapeutic agents. The genome encodes a large number of novel receptor kinases and contains expansions of a variety of gene families, including those associated with virulence. Additional genome features include an abundance of tandemly repeated transfer-RNA-containing arrays, which may have a structural function in the genome.

### 1.7. Repetitive DNA in *Entamoeba* genome

Several repetitive sequences has been identified and characterized. The 24.5kb rDNA plasmid constitutes one of the major repeated DNA in *Entamoeba*. One of the rDNA EhR1 has been fully sequenced and characterized (Sehgal et al., 1994; Bhattacharya et al., 1998). Other circular DNA, of unknown function have also been reported (Dhar et al., 1995; Wilhoeft et al., 2000). Apart from extrachromosomal ribosomal DNA, many chromosomal repeats exist in *Entamoeba*. *E. histolytica* transcribes different unusual transcripts from repetitive DNA, some of these -Tr, UEE1 and SINE1 that are polyadenylated but do not have extensive ORF and are unlikely to code for proteins. These may function as RNA molecules as suggested by presence of extensive secondary structures. It has been speculated that these may be involved in regulation of transcription (Ackers et al., 1999). As is seen in other parasites, some of the protein coding genes of *E. histolytica* also exhibit tandem repeats that tend to show intraspecies repeat variation. Polymorphic genes like SREHP (Stanley et al., 1990) and chitinase (De La Vega et al., 1997) encoding genes are some of them. Microsatellite loci are also present in *Entamoeba* genome (Zaki and Clark, 2001), which have been used as markers to study polymorphism among species and strains of *E. histolytica*. 
1.7.1. Extrachromosomal rDNA

The *E. histolytica* ribosomal DNA are present in 200 copies per genome and are carried exclusively on circular DNA molecules. This was reported independently by two groups (Bhattacharya et al., 1989; Huber et al., 1989). There may be one or two copies of rRNA transcription unit located on each circle. The nucleotide sequence of one such rDNA circle belonging to the strain HM1:IMSS, has been determined (Sehgal et al., 1994). It measures 24.5kb and carries two copies of the ribosomal RNA transcription units arranged as inverted repeats. This molecule is named as EhR1 and complete map with sequence details has been worked out (Fig. 3A). According to EcoR1 restriction sites circle has been devided into HMe, HMd and HMg regions. The two transcription units of EhR1 are separated by a 3.7 kb downstream intergenic spacer region and a 9.2 kb upstream spacer. Which is composed of DraI, Scal, PvuI, HinfI, and AvaiI repeats. Some stocks of the strain HM-1:IMSS contain a variant rDNA episome EhR2 (Fig. 3A), which harbours a single rDNA unit (Ghosh et al., 2001). The single rDNA unit in EhR2 corresponds to rDNA1 of EhR1 rDNAII along with the adjuscent HMg region is missing. The HMd downstream of rDNA1, which contains the Dra1 and Sca1 repeats, and the upstream HMe region are present in EhR2.

The rDNA circle for other strains (200:NIH, HK-9 and Rahman) of *E. histolytica* and Laredo strain of *E. moshkovskii* have also been characterized to some extent. The 200:NIH circle is very similar to EhR1 and differs only in the number of repeats in intergenic spacers. In contrast to HK-9, Rahman and Laredo circles are strikingly different in that they contain only one rRNA unit per circle. Not much is known about the origin and maintenance of these circles but they have been used extensively for diagnostics purposes as molecular markers.

1.7.2. *E. histolytica* long interspersed nuclear elements (EhLINEs)

In a study of repetitive DNA sequence in *E. histolytica*, the first retotransposon like element EhRLE1 was identified in this organism (Sharma et al., 2001). Whole genome shotgun sequencing of *E. histolytica* revealed three families of
Fig. 3: rDNA episomes present in *E. histolytica* strains: rRNA transcription units are represented as filled arrows in each rDNA circle. Number of transcription units may vary among strains. Restriction enzyme sites indicated are EcoRI (E) and Hind III (H). (A) Strain HM-1:IMSS has two different types of rDNA circle (i) EhR1 and (ii) EhR2 having two and one transcriptional units respectively. EcoRI fragments have been labeled as HMe, HMd and HMg. Various families of short tandem repeats in the region upstream and downstream are marked as PvuII Scal, HinfI, AvaII 74bp and DraI. In EhR2 second rDNA unit and HMg region have been lost possibly due to recombination. (B) other strains 200:NIH (i) has two rRNA transcription units, while Rahman (ii), and HK-9 (iii) have only one rRNA transcription unit. Sizes of each rDNA is mentioned in the circle.
Introduction

non long terminal repeat (LTR) retrotransposon or long interspersed nuclear element (LINEs) (Van Dellen et al., 2002). The EhLINEs account for about 6% of the *E. histolytica* genome as deduced from database analysis. Most copies of each element are truncated at the 5'- or 3'- end or at both ends. The 4.8 kb EhLINEs each have a either one or two ORF with a putative nucleic acid binding motif (CCHC) and restriction enzyme like endonuclease domain located downstream of reverse transcriptase (RT) domain. Phylogenetic analysis of the RT domain placed the EhLINEs in the R4 clade of non LTR elements, with its closest relative being the R4 element of *Ascaris lumbricoides* and the Dong element of *Bombyx mori*. EhLINE1 had a length of 4804 bp and is the most abundant EhLINE having 140 copies per genome. All copies show sequence variation with respect to one another (2-4%). A complete ORF is missing in most copies. The 5' and 3' UTR are short (14 and 17 nucleotides respectively).

EhLINE1 shares a common 3' end with a highly transcribed 0.55kb short interspersed nuclear element (EhSINE1) (Fig. 4A). Similarly EhLINE2 shares a common 3' end with a highly transcribed 0.65kb EhSINE2. this feature is generally seen in partner LINES/SINEs (Boeke, 1997; Okada et al., 1997). Discovery of a EhLINE-encoded endonuclease (EN) activity which could nick a natural target site of EhSINE insertion, provided evidence that EhSINE1 could utilize the EhLINE1 machinery for its own transposition (Mandal et al., 2004). EhLINE1 and EhSINE1 probes (Bagchi et al., 1999) showed that these elements reside on all chromosomal bands, do not seem to be telomeric, and might be dispersed in the *E. histolytica* genome. However, all the elements seemed to insert in AT-rich sequences, with a clear preponderance of T-residues in a 50-nt stretch upstream of the site of insertion of each element (Bakre et al., 2005).

1.7.3. *E. histolytica* short interspersed nuclear elements (EhSINEs)

EhSINEs are nonautonomous, non-LTR, retroposons without ORFs and so lack the machinery necessary to replicate themselves. An Eh SINE was first identified as a short repetitive sequence of 550bp, named interspersed element (IE) or *E. histolytica* abundant polyadenylated transcript 2 (Ehapt2), reported by two groups
It appears frequently in *E. histolytica* cDNA libraries, but all copies sequenced so far lack an ORF. Subsequent genome analysis showed that in *E. histolytica* there are three types of SINEs EhSINE1, EhSINE2 and EhSINE3 (Bakre et al., 2005). Copy numbers of these elements are 272 (EhSINE1), 117 (EhSINE2) and only single copy in case of EhSINE3 (Bakre et al., 2005). The sequence identity between different copies is of the order of 95%. A poly A⁺ transcript corresponding to the element is detected in northern blots. The EhSINEs like the EhLINEs are widely distributed in genome and frequently seen close to protein coding genes. EhSINEs shares a common 100bp sequence at the 5' end and have common 3' end (70bp) to their corresponding EhLINE pair (Van Dellen et al., 2002) (Fig. 4). Eh SINE1 is absent in the closely related but nonpathogenic species *E. dispar* (Willhoeft and Tannich, 2002). In *E. histolytica* it has also observed in various genomic locations e.g. upstream of gene for poreforming peptide (Bruchhaus et al., 1993) and as an insertion into an EhLINE1 family sequence (EhRLE5) (Sharma et al., 2001).

**1.7.4. Tr region**

Circular ribosomal DNA molecule EhR1 in strain HM1:IMSS has two transcription units separated by upstream and downstream intergenic spacer region. Upstream region contains a sequence coding for 0.7 kb polyadenylated RNA (Tr) which does not code for protein. An interesting feature of the transcript is the presence of 26bp repeats. There are 9 of these near perfect repeats in the Tr of *E. histolytica* strain HM1:IMSS. The number of these repeats varies from strain to strain and it may range from eight to twenty (Clark and Diamond, 1993). Moreover, there are also strains of *E. histolytica* which do not have Tr. HK9 and Rahman are two such strains (Sehgal et al., 1994). The absence of an ORF of reasonable length suggests that Tr may function through its RNA. It could either have a regulatory role in transcription or translation or may have a structural role.
Fig. 4: Pairwise comparison between the nontranslated polyadenylated EhLINEs and EhSINEs from *Entamoeba histolytica*: (A) (i) Sequence similarity of EhLINE1 and EhSINE1 existing at the 3'-end of each of the element (indicated in blue). (ii) Sequence alignment of the 74 nucleotides shared by the 3' ends of EhLINE1 and EhSINE1. (Source: Bhattacharya et al., 2003) (B) (i) Sequence similarity of EhSINEs existing at 5' end of EhSINEs (indicated in blue). (ii) Sequence alignment of 100 nucleotide at 5' end shared by EhSINEs.
1.8. Epidemiology

*E. histolytica* infect 500 million individuals and is responsible for around 100,000 deaths per annum worldwide. This figure places it second only to malaria in terms of mortality caused due to protozoan parasite (WHO, 1997). The World Health Organisation has recommended that *E. histolytica* should be specifically identified and if present should be treated. Amoebiasis is a serious health hazard, particularly in developing countries like India where poor sanitary conditions prevail. Sporadic incidences of this disease occur in developed countries too. This disease mostly spreads through contaminated food and water. The parasite usually thrives in the intestine, but during the invasive form of the disease, due to unknown reasons it penetrates the intestinal mucosa and enters the bloodstream from where it goes on to infect the vital organs, mainly the liver and sometimes the brain, which could be fatal if not diagnosed timely.

The realization that *E. histolytica* and *E. dispar* are morphologically identical species with remarkably different physiological and pathological characteristics has impacted on all aspects of amebiasis but notably on the epidemiology. Therefore, reliable epidemiologic data are essential to estimate the burden of disease and to formulate policy to control amebiasis.

1.8.1. Studies done worldwide

The true epidemiology of *E. histolytica* infection remains unknown because much of the earlier published epidemiological data fail to distinguish between *E. histolytica* and *E. dispar*. Most of the individuals who were previously believed to have asymptomatic infection with *E. histolytica* actually carry *E. dispar*, which has never been shown to cause invasive human disease (Diamond et al., 1993). Worldwide studies indicate that prevalence of *E. histolytica* ranges between 10 to 30% in different parts of the world. Study conducted in Bangladesh on children (Haque et al., 2003) indicated the prevalence to be 9%. In Ecuador in the village of Barbon (Gatti et al., 2002) shows the infection rate was 27% of which 70% showed patterns of *E. dispar*. Another longitudinal study in central Vietnam (Blessman et al., 2003) indicated the prevalence of 11.2% and an annual new
infection rate of 4.1% in the population. Thus geographical distribution of this parasite is not constant.

Furthermore, only ~10% of individuals who become infected with *E. histolytica* actually develop invasive disease (Gathiram et al., 1987). The best current estimate is that *E. histolytica* causes between 34 million and 50 million symptomatic infections (i.e. amoebic colitis or liver abscess) each year (Walsh, 1987). The occurrence of amoebic liver abscess is probably 5–50 times less common than that of diarrhoea (Walsh, 1987). Although *E. histolytica* infection occurs worldwide, morbidity and mortality are greatest in Central and South America, Africa and the Indian subcontinent (Petri, 1996). In Dhaka, Bangladesh, where diarrhoeal diseases are the leading cause of death in children younger than six years of age, ~50% of children have serological evidence of exposure to *E. histolytica* by five years of age (Haque et al., 1999).

### 1.8.2. Indian scenario

Sporadic studies have been performed in India, but detailed analysis, especially after the separation of *E. dispar* from *E.histolytica* has not been reported. Therefore it becomes extremely important to characterize this parasite in Indian population with the aim of better management of amoebiasis. A study was conducted on the parasitic cause of diarrhea in Delhi by direct smear microscopy by Kaur et al., (2002) indicated *E. histolytica* in 14% cases. In another comparative study of the intestinal parasites prevalent among children living in rural and urban Chennai (Fernandez et al., 2002) with stool sample microscopy reveals Giardia in 15% and *E. histolytica* in 4% cases of rural location in contrast, 22% Giardia and 10.6% *E. histolytica* in urban cases. Molecular probes and PCR technique have also been tried in some studies. In one of such report (Prakash et al., 2000) PCR analyses of cysteine protease gene of symptomatic intestinal and ALA cases have been compared with stool microscopy. The clinical isolates negative for *E. histolytica* by stool microscopy demonstrated the presence of cysteine proteinase gene by PCR amplification. Also the gene copy number was shown to increase in ALA samples compared with intestinal cases.
Riboprinting (PCR+RFLP) of rRNA genes has also been tried (Mukhopadhyay et al., 2002) and detected *E. histolytica* in all of the samples and mixed infection of *E. dispar* in few of them. A new Sau3A site having a discriminatory value was identified in isolates from India. In a study from Vohra et al., 1989, it was observed that certain non-pathogenic zymodemes were potentially pathogenic for the guinea-pigs.

1.9. Methods used in *Entamoeba* typing so far

1.9.1. Microscopy

Historically, light microscopy has been the method of choice to diagnose amoebiasis. Although the presence of haematophagous amoebic trophozoites (i.e. trophozoites that have ingested red blood cells) in a stool sample strongly suggests *E. histolytica* infection, such a finding is rarely seen (Gonzalez-Ruiz et al., 1994). In the absence of haematophagous trophozoites, the sensitivity of microscopy is limited by its inability to distinguish between samples infected with *E. histolytica* and those infected with *E. dispar* (which may be ~10 times more common). Confusion between *E. histolytica*, other non-pathogenic amoebae (such as *Endolimax nana*), and white blood cells (leucocytes) in the faeces frequently results in overdiagnosis of amoebiasis (Krogstad et al., 1978). Delays in the processing of stool samples affect the sensitivity of light microscopy.

1.9.2. Isoenzyme analysis

This technique, which was amongst the earliest to suggest that “pathogenic" and "non-pathogenic" *E. histolytica* were in fact two separate species, had until recently been applied to more specimens than any other and thus deserved its reputation as the “gold standard" against which newer methods need to be validated (Sargeaunt, 1988). In practice, however, it has several disadvantages: culture of microscopically cyst-positive faeces is by no means always successful (Sehgal et al., 1995); it may take seven to fourteen days to grow enough trophozoites to prepare the lysate for analysis and the process itself is somewhat cumbersome. Even though the original procedure involving starch-gel
electrophoresis of four enzymes is often simplified to merely examining hexokinase mobility in agarose mini-gels (Strachan et al., 1988), isoenzyme electrophoresis is now rarely used in clinical diagnosis. Also it now appears that isoenzyme patterns are not fixed (Blanc and Sargeaunt, 1991). Therefore many zymodeme assignments are unreliable.

1.9.3. Antigen detection

Although only a few *E. histolytica* genes appear to be totally absent from *E. dispar*, almost all homologous proteins contain amino-acid substitutions which result in the expression of species specific epitopes. Detection of these epitopes with monoclonal antibodies is the basis of a number of quick and convenient diagnostic methods. The target molecule which has been most intensively studied is the heavy subunit of the galactose/N-acetyl-galactosamine inhibitable lectin; of six monoclonal antibodies raised against this molecule (Petri et al., 1990b), only two reacted with *E. histolytica* and *E. dispar* while the other four reacted only with *E. histolytica* (Petri et al., 1990a). These antibodies are the basis of two kits manufactured by TechLab Inc. (www.Techlab.com), one of which (based on one of the non-specific monoclonal antibodies) identifies *E. histolytica* *E. dispar* while the other, based on a specific antibody can identify *E. histolytica*. Sequential application of these two kits can specifically identify both species although they cannot distinguish mixed infection with *E. histolytica* and *E. dispar* from infection with *E. histolytica* alone. Evaluation in Bangladesh shows clearly that these kits are more sensitive and specific than either wet-film microscopy or culture (Haque et al., 1998; 2000). Because the gold-standard test requires cultivation of the organism which is known not to be 100% sensitive it is difficult to assess whether the kits produce false-positive results, but PCR suggests that most culture or microscopy negative but antigen positive samples are true positives (Haque et al., 1998). Other workers have used similar monoclonal antibodies with equal success for species specific diagnosis in Egypt (Abd-Alla et al., 2000b; Abd-Alla and Ravdin, 2002). Other kits are available from Cellabs in Australia (http://www.cellabs.com.au/) (detects *E. histolytica* and *E.

1.9.4. DNA blotting
The use of species-specific DNA probes to hybridise with unamplified DNA isolated from fecal samples has the great attraction, particularly if radioactive labels are used. Dot blot was found to be more sensitive as compared to microscopy and isozyme analysis. The method has been applied successfully with probe of different genomic region e.g. P145/B133 from rDNA (Bracha et al., 1990; Agarwal et al 1998), and tandem repeat probes (Samuelson et al 1989) but not widely taken up, because it may cross hybridise sometimes. The polymerase chain reaction (PCR) method, however, has been widely tested for amoebiasis.

1.9.5. PCR based methods
The ability of the PCR to specifically amplify small amounts of pathogen DNA has revolutionised the diagnosis of many infectious diseases, and the numerous sequence differences between homologous genes in E. histolytica and E. dispar make it a natural candidate for identifying these two species. A number of methods have been published (Tannich and Burchard, 1991; Acuna-Soto et al., 1993; Katzwinkel-Wladarsch et al., 1994; Rivera et al., 1996; Britten et al 1997; Troll et al., 1997; Evangelopoulos et al., 2000; Verweij et al., 2000). Most, but not all, rely on amplifying unique regions of the SSUrRNA episome, its high copy number providing increased sensitivity. Many polymorphic protein coding genes like SREHP (Stanley et al., 1990) and chitinase (De, la, Vega et al., 1997) have also been used. Although in the original procedures the product was often detected by gel electrophoresis followed by ethidium bromide staining, colorimetric detection using specific probes has been employed with rDNA
probes (Britten et al., 1997). PCR has also proved capable of detecting *E. histolytica* DNA in liver abscess (Tachibana et al., 1992; Britten et al., 1997; Zaman et al., 2000).

In *in vitro* testing with cultured trophozoites PCR was about one hundred times more sensitive than antigen detection (Mirelman et al., 1997). However, it is important to be aware of the disadvantages of the method. Firstly, three separate steps are required – DNA extraction, amplification and product detection. Secondly, as with all PCR-based methods, great care has to be taken to eliminate the risk of false positives due to contamination from product prepared earlier.

Our present knowledge in intraspecies variation in *E. histolytica* is limited. Non-coding regions, including microsatellite loci in the vicinity of tRNA genes have been used to design primers for strain identification in *E. histolytica* (Zaki and Clark, 2001). However the method requires more than one set of primers to distinguish strains, and is unable to differentiate very closely related strains like HM1:IMSS and HK9. Therefore, there is a need to generate better molecular markers for strain differentiation in *E. histolytica*.

### 1.10. Transposon display

Transposon display (TD) is a modification of the AFLP technique (Vos et al., 1995) that uses a primer anchored in a transposon to simultaneously detect several hundred markers in the genome (Casa et al., 2000; Casa et al., 2004). TD involves the amplification of sequences flanking the transposon by ligation-mediated PCR (Fig. 5), so the resulting fragments are locus-specific and can be analyzed by polyacrylamide gel electrophoresis. TD has been used to study the behavior and stability of transposable elements, including retrotransposons, in various organisms, especially plants (Flavell et al., 1998; Purugganan and Wessler 1995).

Available techniques to study polymorphism have their own advantages and disadvantages. RFLPs are reproducible but difficult to perform and/or yield a low density of markers. RAPDs are straightforward but lack reproducibility. AFLPs
Fig. 5: Schematic representation of transposon display procedure: Genomic DNA was digested with Vsp1 enzyme (frequent cutter), ligated with adapters. Preamplification was done to enrich Eh SINE1 population, finally selective amplification was done with radiolabelled primer to amplify radiolabelled Eh SINE1 and upstream fragments. Amplified product was run in denaturing urea PAGE and autoradiographed.
yield a high number of markers but cannot be used in mixed culture DNA samples. TD has several potential advantages over these.

- It is a sensitive and reliable technique for strain differentiation. The technique is cost effective, as limited numbers of primers are required to give several bands.

- It can be very useful to differentiate strains from mixed culture when coupled with a particular class of transposable elements.

- Apart from strain differentiation this technique can also be used to understand transposition mechanism and behavior.
1.11. Aim and scope of work

As mentioned earlier, the delineation of former *E. histolytica* into two genetically distinct species, the invasive *E. histolytica* and the noninvasive *E. dispar*, has had major impact on our views of amebiasis, in particular its epidemiology and clinical management. Many techniques have been developed for species and strain identification but there is need for a reliable method which can accurately detect and differentiate *E. histolytica* and *E. dispar* directly from clinical isolates, without cultivation. The varied organ tropisms and clinical presentations of infection by *E. histolytica* raise the question of the role of parasite genetic diversity in virulence. We do not know how many strains of the parasite exist in our population. Since all molecular biology studies with *E. histolytica* have been done with this parasite grown axenically, therefore, we do not know whether some properties of the parasite are actually adaptive response to growth in culture. For example certain regions of the rDNA circle are lost in some strains of *E. histolytica* presumably by recombination (Ghosh et al., 2000). We do not know whether this phenomenon occurs in natural isolates also. In India, we are well poised to ask some of these questions because we have access to the parasite in natural host.

This study specifically deals with the following objectives:

1. To design molecular probes for detection and differentiation of *E. histolytica* and *E. dispar* from stool samples of asymptomatic and symptomatic individuals and pus samples from amoebic liver abscess patients.

2. To develop a method for *E. histolytica* strain differentiation.

3. To study the genetic variation in *E. histolytica* strains from axenic cultures based on insertion of LINEs/SINEs.

4. To establish a correlation, between the virulence and genetic diversity of *E. histolytica* strains.