GENES OF MITOCHONDRIAL ORIGIN IN THE GENUS *ENTAMOEBA*

CHRISTINA BAKATSELOU

THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

JANUARY 2002

London School of Hygiene and Tropical Medicine

UNIVERSITY OF LONDON

TO:

MY PARENTS AND MY SISTER

ABSTRACT

Entamoeba histolytica is the protozoan parasite that causes amoebic dysentery and amoebic liver abscesses in humans. For many years it was believed to be a primitive organism because it lacks many typical eukaryotic features including mitochondria. Recently, two genes that in other organisms encode proteins normally found in the mitochondrion have been isolated, giving evidence for the secondary loss of mitochondrial function in *E. histolytica.* These are the pyridine nucleotide transhydrogenase (PNT) and the mitochondrial chaperonin cpn60 genes. In this study we isolated and characterised a gene encoding a mitochondrial-type heat shock protein 70 from *E. histolytica.* cDNA and genomic library clones have been isolated and sequenced. Comparison with previously published sequences confirmed the assumption that *E. histolytica* comes from mitochondrion - bearing ancestors.

Southern blot hybridisation revealed there are two copies of the gene in the genome. Northern blot analysis revealed two transcripts hybridising to the mt-hsp70 probe that differ in length and which are induced by heat shock. In addition, an apparently non-coding, polyadenylated RNA that is also induced by heat shock is encoded immediately upstream of the mitochondrial-type hsp70 gene.

Expression analysis was also performed in four other *Entamoeba* species. Partial cpn60, PNT, and mt-hsp70 genes were isolated and the size of the mRNAs and their heat shock induction levels were investigated by hybridisation to these probes.

The similarity of the mt-hsp70 amino terminus to those of hydrogenosomal proteins in conjunction with the phylogenetic analyses suggests it is also likely to be targeted to the mitochondrion-derived organelle of *E. histolytica* known as the mitosome.

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to my supervisor Dr. Graham Clark for his guidance and support throughout the length of my PhD.

I would also like to thank Dr. Jorge Tovar (Royal Holloway University, London) for his help with various techniques in the laboratory and helpful advice at the beginning of the project.

I wish to express my sincere thanks to Dr. Andrew Roger (Dalhousie University, Canada) for performing the Puzzleboot analysis of the mt-hsp70 gene and for helpful discussions; Dr. Michael Duchêne (University of Vienna) and Dr. Egbert Tannich (Bernhard Nocht Institute) for providing us with the cDNA and the genomic libraries of *E. histolytica,* respectively; and Dr. Martin Embley (Natural History Museum, London) for helpful suggestions.

Special thanks also to Dmitry and to my cousin Stratos for helping me with the various computer programmes and giving me advice. A special thanks also to my flatmate TJ for letting me use his computer for the final changes.

I wish to thank my friends for believing in me and supporting me in various ways throughout the length of the PhD.

Finally, I am grateful to my family. Without their financial support and encouragement this PhD wouldn't be possible. They were always there when I needed them most.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 *ENTAMOEBA HISTOLYTICA*

Entamoeba histolytica is a protozoan parasite that causes amoebic dysentery and amoebic liver abscesses in humans. 40 million cases of disease are estimated to occur annually, with up to 100,000 deaths per year (1). It predominantly affects individuals of lower socio-economic status who live in developing countries as a result of unhygienic conditions. Patients among whom the mortality due to invasive amoebiasis is high include malnourished individuals, children younger than 1 year of age, pregnant women and individuals receiving corticosteroid therapy (2). This enteric protozoan is the third leading cause of morbidity and mortality due to parasitic disease after malaria and schistosomiasis (3).

E. histolytica can be found world-wide but disease occurs at a very low incidence in those regions that practice sanitary sewage treatment. Ineffective sewage treatment, use of untreated surface water for drinking purposes and also use of human waste as fertiliser are the usual scenarios for cases of amoebiasis. Transmission can also occur through faecal exposure during sexual contact (in which case not only cysts, but also trophozoites could prove infective), especially in homosexual males.

1.1.1 Classification

The genus *Entamoeba* is grouped according to the Protozoa classification by Cavalier-Smith in 1993 (4) as follows:

EMPIRE EUKARYOTA
Superkingdom: METAKARYOTA
Kingdom: Protozoa
Subkingdom: Dictyozoa
Branch: Bikonta
Infrakingdom: Neozoa
Parvakingdom: Entamoebia
Phylum: Entamoebia

However, its true relationships remain unknown at present.

1.1.2 Life cycle and pathology

E. histolytica is named for its lytic effect on tissue and is transmitted to humans primarily by ingestion of mature cysts (Figure 1.1, 1) by faecally contaminated hands, in water and occasionally by the foodborne route. The cyst when swallowed passes intact through the stomach into the small intestine. Excystation occurs as the cyst enters the large intestine (Figure 1.1, 2). A quadrinucleated amoeba emerges from the cyst (Figure 1.1, 3) and divides into four and then eight trophozoites by nuclear and cytoplasmic divisions. The trophozoites of the organism then move downstream to colonise the mucous layer of the large bowel, where they feed on bacteria and cellular debris and start multiplying by binary fission.

At this point the trophozoites may re-encyst (Figure 1.1, 4), a process that is apparently stimulated by luminal conditions that are less than ideal for the trophozoites. Nuclear division and metabolism continue within the cyst. The life cycle is complete with the excretion of the quadrinucleate cysts in the faeces, which may remain viable for weeks to months depending on the environmental conditions. Trophozoites can also be passed in diarrhoeal 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.

Continued growth of the trophozoites, which are the invasive form of the parasite, can lead to ulceration of the intestinal mucosa, and cause diarrhoea and severe intestinal cramps. Diarrhoea is followed by a condition referred to as amoebic dysentery, characterised by lower abdominal pain and the passage of intestinal exudates, blood, and mucus. The infection is occasionally complicated by intestinal perforation, while in some people an increased frequency of bowel movements with or without blood and mucus may occur. A chronic form amoebic colitis can produce symptoms similar to those of ulcerative colitis or other forms of inflammatory bowel disease that can make it easy to misdiagnose (5).

If amoebic dysentery is not treated, it may lead to haematogenous spread of the organisms to the liver via the portal venous system, resulting in extraintestinal lesions, mainly amoebic liver abscesses. Lesions occasionally occur in the lung, brain, or other organs. Growth in these tissues can cause abscesses and other tissue damage that can lead eventually to death if untreated. Cerebral amoebiasis is a rare condition but almost always fatal (6). Some people get abscesses without having had dysentery.

Asymptomatic infections with *E. histolytica* also occur. Patients colonised with this organism are at risk of developing invasive disease in the future (months to years) and should be treated. In some cases these infections are spontaneously lost. Extraintestinal

amoebiasis develops in 10% of individuals with symptomatic *E. histolytica* infections. The ratio between asymptomatic and symptomatic infections is poorly known but it seems that asymptomatic infections are ~10 fold more common than symptomatic (7).

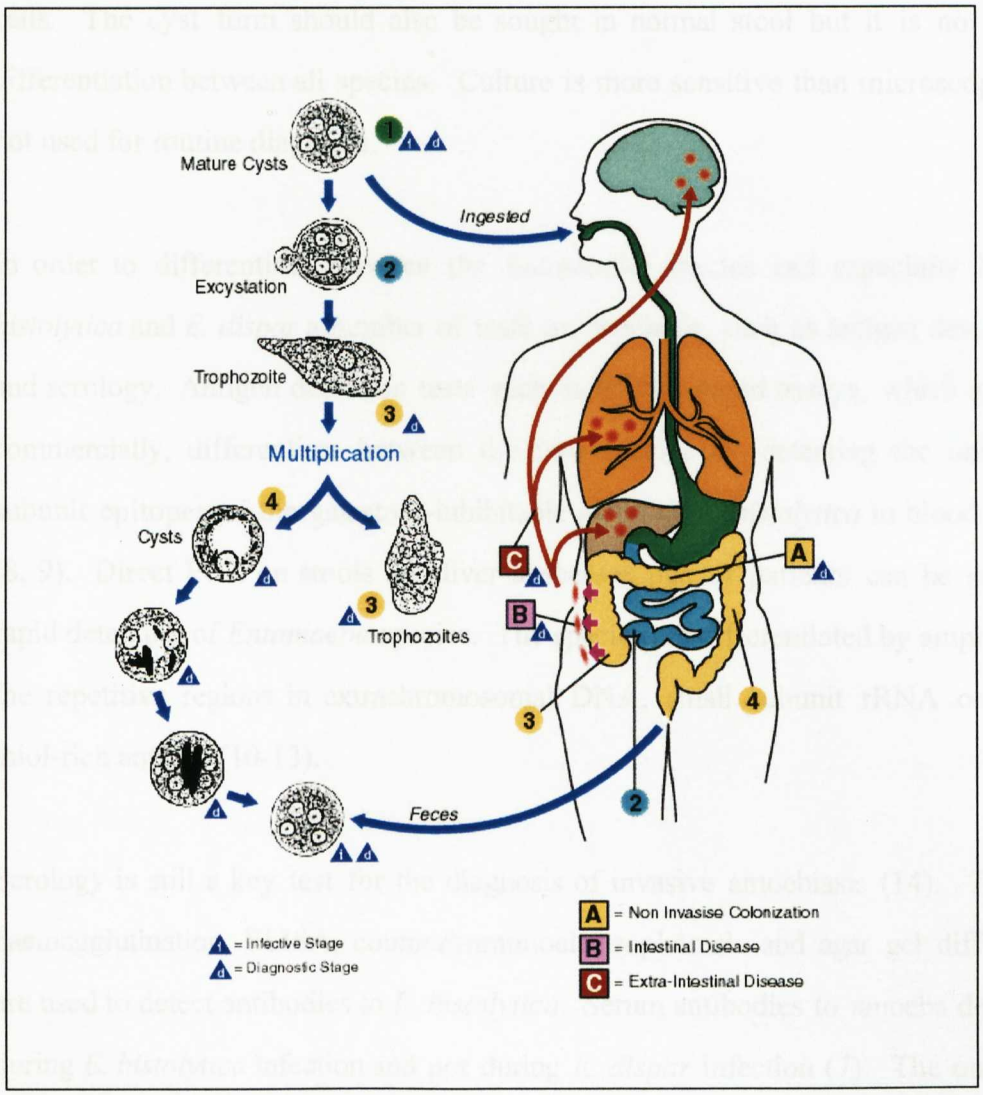


Figure 1.1. The life cycle of *Entamoeba histolytica*.

The diagram was obtained from the CDC website (http://www.dpd.cdc.gov/dpdx/HTML/Amebiasis.htm). The trophozoites may remain confined to the intestinal lumen (A: non-invasive 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: extra-intestinal disease), with resultant pathologic manifestations.

1.1.3 Diagnosis and treatment

Amoebiasis is easily diagnosed by microscopic examination of diarrhetic stool samples, liver abscess pus or colonic biopsies for the trophozoites of *E. histolytica,* which can be differentiated from non-pathogenic *Entamoeba* species by the presence of ingested red cells. The cyst form should also be sought in normal stool but it is not useful for differentiation between all species. Culture is more sensitive than microscopy but it is not used for routine diagnosis.

In order to differentiate between the *Entamoeba* species and especially between *E. histolytica* and *E. dispar* a number of tests are available, such as antigen detection, PCR and serology. Antigen detection tests such as ELISA-based assays, which are available commercially, differentiate between the two species by detecting the unique heavy subunit epitopes of the galactose-inhibitable lectin of *E. histolytica* in blood and faeces (8, 9). Direct PCR on stools and liver abscesses pus of patients can be used for the rapid detection of *Entamoeba* species. The species are differentiated by amplification of the repetitive regions in extrachromosomal DNA, small subunit rRNA or the 30kD thiol-rich antigen (10-13).

Serology is still a key test for the diagnosis of invasive amoebiasis (14). The indirect haemagglutination, ELISA, counterimmunoelectrophoresis and agar gel diffusion tests are used to detect antibodies to *E. histolytica.* Serum antibodies to amoeba develop only during *E. histolytica* infection and not during *E. dispar* infection (7). The only problem with serology testing is that patients may remain positive for years after an episode of amoebiasis (3) and therefore this method cannot differentiate acute from remote infections in areas of high endemicity.

Non-invasive imaging studies, such as ultrasonography, computerised tomography (CT) and MRI, have also been used for the diagnosis of amoebic liver abscesses.

Luminal infections can be treated with the drugs paromomycin, diiodohydroxyquin or diloxanide furoate. Metronidazole is useful against amoebas in the extraintestinal sites of infection but must be followed by a luminal agent to remove amoebas still present in the intestine. If these sites become secondarily infected with bacteria, antibiotics and drainage are necessary. Spontaneous cures do occur, implying that the host immune system may play some role in ending the infection. However, protective immunity is not afforded by primary infection because reinfection is not uncommon. Vaccines for amoebiasis or reliable forms of chemoprophylaxis are not available.

1.1.4 Morphology

The trophozoite or feeding form of *E. histolytica* is highly variable in shape and its form and motility are extremely sensitive to changes in temperature, pH, osmolarity, and redox potential (15). Locomotion is by means of a single well-defined pseudopodium without any apparent microfilaments. Microtubules are only observed during nuclear division. Amoebas move in one direction in a slug-like manner, without a clear differentiation between ectoplasm and endoplasm (16).

Rounded trophozoites measure between 10 and 40 μm (17). Amoebas obtained directly from liver or intestinal lesions are generally large, measuring about 20-40 μm in diameter, while those found in non-dysenteric stools or in cultures measure 7-30 μm (15). If present, vacuoles are usually small except when in axenic culture. The cytoplasm is rich in glycogen and may also contain bacteria and ingested erythrocytes when the trophozoites are isolated from symptomatic individuals. A single spherical nucleus is present, 4-7 μm in diameter that contains fine peripheral chromatin. The nucleus has a small, compact spherical karyosome, 0.5 μm in diameter usually centrally located that is

surrounded by an achromatic capsule-like structure (18). The function of the karyosome is unknown.

The cysts of *E. histolytica* are spherical, measuring 10 μm to 16 μm in diameter. Nuclear division takes place twice only during encystation, giving four nuclei in the mature form. The impervious cyst wall is composed of chitin, which is an N-acetyl-D-glucosamine polymer. Glycogen may be diffuse or absent in the mature cyst and clumped chromatoid masses may be present in early cysts. These chromatoid bodies are ribosomes in helical arrays that aggregate into large crystal inclusions. The structure of the cyst makes it resistant to the acidic pH of the stomach when ingested, and it can survive in a moist environment for several weeks.

1.1.5 History and Taxonomy

Amoebiasis seems to be an ancient disease. Sanskrit documents written more than 3000 years ago refer to cases of mucosanguineous diarrhoea, which were probably caused by amoebas. In 1875, Fedor Aleksandrovich Lösch in St. Petersburg, Russia, first discovered that dysentery could be caused by amoebic trophozoites, which he named 'Amoeba coli' (19). He even managed to reproduce the disease in dogs. A formal taxonomic description was given in 1903 by Fritz Schaudinn who gave to the pathogenic species the name *Entamoeba histolytica* (20).

In 1913, Walker and Sellards demonstrated in the Philippines that *Entamoeba coli* was non-pathogenic and *E. histolytica* was pathogenic, and also emphasised the role of cysts in the transmission of amoebiasis (21). During the next few years the life cycle of the parasite was studied and other *Entamoeba* species and other genera of amoebas were found living in the gastrointestinal tract. This continued until 1919 when Clifford Dobell recognised only three species of *Entamoeba* as being human parasites, *E.*

histolytica, E. coli and *E. gingivalis* (16). Oral infections were ascribed to *E. gingivalis*, species producing cysts with four nuclei to *E. histolytica* and species producing cysts with eight nuclei to *E. coli*. All the other described species were deemed to be synonymous with one of these three (22).

Von Prowazek, in 1912, had described an organism he named *Entamoeba hartmanni*, which was morphologically similar to *E. histolytica* but smaller in size (23). In the beginning it was believed to be synonymous with *E. histolytica* until the late 1950s when it was demonstrated that there are morphological differences between the two forms and therefore the two distinct species were recognised (24).

In 1925, Emile Brumpt proposed the establishment of a new non-pathogenic species of amoeba, which he called *E. dispar*, in contrast to the morphologically indistinguishable but pathogenic "*E. dysenteriae*" (i.e. *E. histolytica*) (25). Protozoologists did not accept this proposal because the two species were morphologically indistinguishable. This resulted in a great confusion in the next decades regarding the significance of the parasitological diagnosis of intestinal infection with *E. histolytica*.

The first biochemical evidence for subgroups within *E. histolytica* was in 1973 when it was reported that amoebas isolated from patients with disease showed different lectin agglutination characteristics to those isolated from individuals with asymptomatic infections (26). A few years later isoenzymes were shown to distinguish between the invasive and non-invasive *Entamoeba histolytica* (27). More electrophoretic patterns of isoenzymes (zymodemes) were analysed later (28, 29). Monoclonal antibodies were then described that recognised antigens specific for 'pathogenic' strains (30), and genomic DNA revealed differences using Southern blot analysis (31). All these studies clearly revealed genetic differences between *E. dispar* and *E. histolytica,* initially called non-pathogenic and pathogenic *E. histolytica* strains. Studies with ribosomal RNA gene

(rDNA) PCR amplification also supported the existence of separate species (10). The formal recognition of the two species occurred in 1993 by Diamond and Clark (17).

1.1.6 Metabolism

E. histolytica trophozoites are aerotolerant anaerobes. They are able to consume oxygen despite the lack of mitochondria. They have no cytochromes or a functional tricarboxylic acid cycle. Pyruvate is the only substrate that is decarboxylated to acetyl CoA.

Catalases, peroxidases, other heme-containing enzymes, glutathione and its related enzymes are also absent. Despite the lack of the above, electrons are transferred from reduced substances to molecular oxygen via a succession of carriers, including flavins and nonheme iron (32, 33). Studies have shown that they can grow under aerobic conditions with up to 5% oxygen. Below this concentration, the trophozoites are able to detoxify the products of oxygen reduction. It is possible that this metabolism allows the amoebas to survive the shift from the anaerobic environment of the intestinal lumen to the bloodstream and the organs where the oxygen supply is abundant (34).

Carbohydrates are the main source of energy for *E. histolytica.* Glucose is metabolised to pyruvate via the Embden-Meyerhof pathway, and fructose phosphate is phosphorylated prior to cleavage by enzymatic reactions unique to the amoebas. Inorganic pyrophosphate, an end product of metabolism, is used as an energy source replacing ATP in several glycolytic reactions. The principal end products of anaerobic carbohydrate metabolism are ethanol, CO_2 and acetate (34), whose proportions vary with the extent to which the organisms are deprived of oxygen. Lactate is not an end product and lactate dehydrogenase has not been reported.

Little is known about protein, lipid and nucleic acid metabolism (35). No studies on cyst metabolism have been done up to now as the induction of encystment *in vitro* in the absence of bacteria is not possible (36).

1.2 OTHER *ENTAMOEBA* SPECIES

The genus *Entamoeba* includes several species of human parasites: *E. histolytica, E. dispar, E. hartmanni, E. coli*, and *E. gingivalis*. Other non-human species of the genus *Entamoeba* include *E. polecki*, *E. chattoni*, *E. invadens*, *E. terrapinae* and *E. moshkovskii*. Only *E. histolytica* is an important cause of disease in humans (37), so it is vital to be able to distinguish this pathogenic species from the non-pathogens.

Morphological differences to describe each species of *Entamoeba* are not enough. The classical subdivision for *Entamoeba* species up to now was based on the number of nuclei their cysts produce when mature, i.e. one, four or eight. With the exception of *E. gingivalis*, which has no cysts, the rest are divided into 3 categories: Uninucleated cyst-producing species, which include *E. chattoni*, *E. polecki* and *E. bovis*; quadrinucleated cyst-producing species which include *E. histolytica*, *E. dispar*, *E. hartmanni*, *E. invadens, E. moshkovskii* and *E. terrapinae*; and octonucleated cyst-producing species which include *E. coli*, *E. gallinarum* and *E. muris*.

Restriction site polymorphism analysis of PCR amplified small subunit rRNA gene sequences (riboprinting) has been used to find out more about the relationship between the species (38, 39). Phylogenetic analyses of representative ribosomal RNA gene sequences (39, 40) proved the monophyletic nature of the cyst-producing groups. Silberman *et al* also confirmed the monophyletic nature of the genus *Entamoeba* (40). The distance among groups that produce cysts with eight, one and four nuclei has been found to be very large, thus the nuclei per cyst taxonomy is confirmed. *E. coli, E.*

gingivalis and *E. moshkovskii* showed intraspecific variation, while the morphologically indistinguishable *E. histolytica*, *E. dispar* and *E. moshkovskii* were found to be genetically similar to each other. *E. histolytica* and *E. dispar* have been demonstrated to be sibling species (38).

E. gingivalis is the only species that lives in the buccal cavity around the root of the teeth. It has no cyst form and is transmitted by mouth-to-mouth contact. The other species live in the large intestine of humans or other animals.

1.2.1 Uninucleated cyst-producing species

E. polecki is a common parasite in pigs but it is also found in monkeys, cattle, goats, sheep and dogs. It has been occasionally found to infect humans (41) but it is generally considered to be non pathogenic. Symptomatic cases can be difficult to treat. The cysts are 10-18 μm in diameter. Chromatoid bodies are abundant and pointed rather than rounded as in *E. histolytica*. Normal hosts of *E. chattoni* are non-human primates such as monkeys and human infection is mainly reported in zookeepers (42). Cysts are 11-13 μm in diameter. *E. bovis* is a parasite of cattle but it has not been found to cause disease.

Various uninucleated *Entamoeba* species have been found in a number of animals and given different names. However, these species cannot be distinguished from each other morphologically and therefore it is not possible to know if they are genetically distinct from each other and whether they occur in humans. A recent study on small subunit rRNA genes has shown that there are at least four genetic types of uninucleated cyst-producing *Entamoeba* species that infect humans, two of which were related to *E. polecki* and *E. chattoni* and two were previously unreported. Whether *E. chattoni*-like infections originated from contact with monkeys or whether the *E. polecki*-like

infections originated from pigs is not known (43) but until all the species involved can be identified the authors suggested naming the agent of all uninucleated human *Entamoeba* infections as '*E. polecki*-like'.

1.2.2 Quadrinucleated cyst-producing species

The species belonging to the quadrinucleate cyst grouping which are morphologically indistinguishable, i.e. *E. histolytica*, *E. dispar* and *E. moshkovskii*, are also genetically similar to each other (38). *E. histolytica* and *E. dispar* have been found to be sibling species, but *E. dispar* is 9 times more prevalent than *E. histolytica* and is believed to infect up to 10% of the world's population (44).

E. moshkovskii is a free-living amoeba first isolated from Moscow sewage. Since then it has been found in sewage worldwide as well as in sediments from freshwater rivers, lakes and brackish coastal splash pools. In 1961 it was identified in a stool of a patient in Laredo, Texas by Dreyer (45) and since then there have been a small number of cases where it has been shown to infect humans (38, 46). No other host has been found. Its cyst form has four nuclei, measures between 7-16 μm in diameter and it is morphologically very similar to *E. histolytica*. The difference is that it can grow both at room temperature and at 37 °C, is osmotolerant and is resistant to emetine (47, 48). *E. hartmanni* is also morphologically similar to *E. histolytica* but smaller. Its quadrinucleated cyst rarely reaches 10 μm in diameter (usually between 6-7 μm). It is not pathogenic and has a few distinct morphological features that distinguish it from *E. histolytica*.

E. invadens is a reptile parasite. It is pathogenic to lizards and snakes causing severe necrotic enterohepatitis but it does not harm turtles, which serve as carriers. Cysts usually measure between 11-20 μm in diameter. *E. invadens* has been used as a model

for encystation of amoebas as *E. histolytica* encystation does not occur in axenic media, making biochemical analysis impossible.

E. terrapinae was first isolated in 1930 from the turtle *Chrysemys elegans* by Sanders and Cleveland and it is not pathogenic to turtles or snakes (49). This amoeba does not undergo nuclear division after excysting. It undergoes two cytoplasmic divisions to give rise to binucleate amoebas then to four uninucleated trophozoites (50). Cyst size is the same as *E. histolytica* and *E. dispar* (10-16 μm in diameter).

E. gingivalis, although it does not form cysts, groups with the species that have quadrinucleate cysts indicating that it is probably a descendant of this group and lost the ability to encyst during evolution. Transmission occurs directly through saliva and it is the only species known to ingest white blood cells.

1.2.3 Octonucleate cyst-producing species

E. coli is the most common *Entamoeba* infection in humans. It is a harmless commensal and has the largest cysts of the genus ranging from 15-30 μm in diameter. In addition to the 8 nuclei in its cysts, which differentiates it from *E. histolytica*, *E. coli* also has chromatoid bodies with irregular splintered ends.

Other members of the octonucleate cyst group include *E. gallinarum* and *E. muris* which are found in the intestine of healthy birds and rodents respectively. No infection in humans has been reported.

It is likely that many more species exist. Lots have been described but have not been studied.

1.3 EVOLUTION OF *ENTAMOEBA HISTOLYTICA*

Until recently it was believed that *E. histolytica* was an archetypically primitive eukaryote (51) which belonged to an early branching eukaryotic lineage, as it lacks typical cytoplasmic organelles such as a structured cytoskeleton, mitochondria, peroxisomes, rough endoplasmic reticulum, Golgi apparatus, centrioles and microtubules (52). *E. histolytica* also has an unusual glycolytic metabolism (53) that resembles more closely anaerobic and microaerophilic bacteria than typical eukaryotes (34, 54).

The unusually structured chromatin, the presence of divergent histones, the rDNA on extrachromosomal circular plasmids, the rarity of introns and the extremely short 5′ untranslated regions of mRNAs also led some authors to consider *E. histolytica* as a representative of an ancestral eukaryote (54). This was also supported by molecular sequence comparisons using elongation factor (EF)-1α and EF-2 and β-tubulin which placed *E. histolytica* as a basal eukaryote, implying it diverged from other eukaryotes prior to the appearance of cell organelles (55-57).

However, ribosomal RNA-based phylogenetic trees have shown that *E. histolytica* branches more recently than several lineages with typical eukaryotic organelles and metabolism (39, 58). This led to the exclusion of Entamoebidae from the Archezoa and placed them in the kingdom Protozoa (4). It is possible that those typical cytoplasmic organelles were lost secondarily or 'downgraded' in the lineage leading to *E. histolytica* as a result of its parasitic lifestyle (59). 18S small subunit rRNA sequence alignments suggest that the genus *Entamoeba* may have evolved from a non-ciliated mycetozoan (related to *Dictyostelium*) although not enough evidence is available at present to verify this theory (4).

1.3.1 Secondary loss hypothesis and organelle origins

Mitochondria are believed to have evolved from an endosymbiosis that occurred over one billion years ago between a eukaryotic cell and a member of the α subdivision of purple bacteria (α-Proteobacterium) (60), based on phylogenetic analyses of the sequences of ribosomal RNA and heat-shock proteins (61-63). Previous inferences based on cytochrome c comparisons also support this view (64). With several genes, such as the cpn60, the closest known eubacterial relatives of the ancestor to mitochondria include members of the rickettsial subdivision of the α-Proteobacteria, which include genera such as *Rickettsia*, *Anaplasma* and *Ehrlichia* (65). *Rickettsia* seems to be the closest relative to mitochondria out of the three, possibly because it favours an intracellular lifestyle that identifies these bacteria as the sort of organism that might have initiated the endosymbiotic scenario leading to modern mitochondria. Both *Rickettsia* and mitochondria have genomes that are the products of several types of reductive evolution (58, 66), and probably share a proteobacterial ancestor and a similar evolutionary history. Transfer of genes from a mitochondrial ancestor to the nucleus of the host would both reduce the mitochondrial genome size and stabilise the symbiotic relationship.

Microsporidia (e.g. *Nosema locustae)*, Metamonada (e.g. *Giardia lamblia*), and Parabasala (e.g. *Trichomonas vaginalis*) were the three deepest branching eukaryotic lineages in early small subunit ribosomal RNA phylogenies, and they all lack mitochondria. The absence of mitochondria and plastids in these organisms was the reason for believing that they were primitively amitochondriate and diverged before the endosymbiosis. However the alternative scenario is that they secondarily lost these organelles.

It is logical that if any of these amitochondriates had lost this organelle, some mitochondrion-specific enzymes may have remained useful, despite the absence of a

sub-cellular compartment. The detection of those genes in the nucleus would provide clear evidence for the secondary loss hypothesis.

Recent studies have indeed revealed genes in amitochondriate eukaryotes that in other organisms code for proteins found in the mitochondrial compartment. This is an indication that the common ancestor of all the existing eukaryotic species could have been a mitochondriate eukaryote. Loss of mitochondria has now been demonstrated for *Entamoeba histolytica* (67), *Trichomonas vaginalis* (68-70), *Vairimorpha necatrix* (71), *Nosema locustae* (72), and *Giardia lamblia* (70, 73). Phylogenetic analyses of these genes place these organisms within the mitochondrial assemblage and the origin of the genes with α-Proteobacteria (74). This is evidence that the genes are of mitochondrial origin and were probably part of the genome of the ancestral mitochondria (67). Therefore, all eukaryotic lineages appear to have had a mitochondriate eukaryote as a common ancestor, and mitochondria evolved after an endosymbiosis of an aerobic bacterium with a eukaryotic cell (74, 75).

Trichomonas vaginalis, although it is amitochondriate and does not have peroxisomes, has another double-membrane bound organelle, the hydrogenosome. This organelle measures 0.5-1 μm in diameter, and its inner membrane lacks the cristae that are characteristic of mitochondria. It also does not have DNA, and has an analogue of pyruvate:ferredoxin oxidoreductase that metabolises pyruvate as a major substrate and oxidises it to acetyl-CoA, which is then converted to acetate and ADP. It also produces ATP anaerobically. It produces H_2 as the reduced end product of its energy metabolism (76). The lack of DNA is probably due to a complete transfer of the hydrogenosomal genes into the nucleus of an ancestral host cell. All the genes for known hydrogenosomal proteins are encoded in the nucleus (77). No hydrogenosome has been described in multicellular plants or animals (76).

Evidence for hydrogenosome origins also comes in part from the isolation and characterisation of genes encoding cpn60 and mitochondrial heat shock protein 70 (mt-hsp70) proteins targeted to the hydrogenosome. They are found to be most similar to those from mitochondria (68, 69).

Hydrogenosomes are present in a number of unrelated eukaryotic groups, including certain free-living ciliates, rumen-dwelling ciliates and chytrid fungi, leading to the conclusion that they have arisen independently on several occasions. Two theories exist for the evolution of the hydrogenosome. It was either converted from or replaced pre-existing mitochondria or peroxisomes, or it descended from an anaerobic bacterium that contained the hydrogenosomal enzymes (76, 78). After the endosymbiotic event the necessary genetic material of the symbiont was transferred to the nucleus and the rest was lost.

At present there are two possibilities over the origin of the mitochondrion and the hydrogenosome. Either the two organelles evolved from two independent symbioses involving α-Proteobacteria, or there was a single bacterial ancestor that subsequently diverged into the two organelles. The current theory of what gave rise to the mitochondria is based on the energy metabolism of eukaryotic cells, bacteria and the hydrogenosome and is known as the hydrogen hypothesis (79). This hypothesis is based on the assumption that a free-living eubacterial symbiont (α-Proteobacterium), which produced H_2 and CO_2, and a methanogenic archaebacterium merged together in a symbiotic event so that the host (the archaebacterium) could use the H_2 and CO_2 as sources of energy. When the pair was removed from the H_2-rich environmental source due to changes in the early Earth's atmosphere the host became completely dependent on the eubacterial symbiont. This subsequently led to the transfer of genes from the symbiont to the host, which finally gave rise to an organelle. This ancestral eukaryote gave rise to the three eukaryotic groups: the mitochondriate eukaryotes after loss or repression of amitochondriate energy metabolism; the Type II amitochondriate

eukaryote after loss of respiration which led to hydrogenosome containing cells; and the Type I amitochondriate eukaryotes which includes eukaryotes that lack all such organelles.

Recently, a study identified an inner-membrane protein (Hmp31) of the mitochondrial carrier family of proteins in the hydrogenosome of *T. vaginalis* that has been shown to be phylogenetically related to the mitochondrial ADP-ATP carrier (ACC) proteins (80). The ACC carrier is an integral inner-membrane component that exchanges ADP and ATP between the cytosol and the mitochondrion. In this study it was demonstrated that the hydrogenosomal ACC homologue can be targeted into the mitochondrial membranes using the same pathway that the mitochondrial ACC proteins are using. The yeast mitochondrial ACC was also successfully targeted into the membranes of the hydrogenosome. This indicates a common evolutionary origin for the membrane carriers and membrane protein-targeting machinery of hydrogenosomes and mitochondria. It is highly likely therefore that these two distinct organelles share a common ancestor that evolved into a progenitor organelle in the early eukaryotic cells and that ultimately gave rise to both of them. This also supports the hydrogen hypothesis for the origin of the eukaryotic cell (79).

1.3.2 Secondary loss hypothesis in *E. histolytica*

As in the case of *T. vaginalis* evidence of secondary loss of mitochondrial function in *Entamoeba histolytica* has been obtained by the cloning of two genes that encode proteins normally found in the mitochondrion, i.e. pyridine nucleotide transhydrogenase (PNT) and the mitochondrial chaperonin 60 (cpn60). The cpn60 is present only in organelles in eukaryotes and is homologous to the GroEL family of bacterial chaperonins. Both PNT and the cpn60 seem to encode signal peptides at the amino termini of the proteins (67, 73, 81).

Two independent studies published almost simultaneously described the existence of a small cylindrical organelle about 1-2 μm in size, named the mitosome (of putative mitochondrial origin (82)) or the crypton (of previously hidden of cryptic function (83)) which is believed to be a degenerate form of the ancestral mitochondrion in *E. histolytica*. The majority of the cells seemed to have one organelle but some seemed to have 2 or 3. A likely explanation for this was that those cells were ready to divide.

The cpn60 was found to have a recognisable import signal similar to those required for targeting to mitochondria and hydrogenosomes in other organisms. Both groups localised the cpn60 gene to the organelle where the import signal is cleaved and then visualised it by Western blotting and confocal fluorescence microscopy. The cpn60 also seemed to be a functional chaperone as it rescued a temperature-sensitive *Escherichia coli* GroEL mutant, which is deficient in chaperonin function (83). One group reported (84) that when the mitosome was stained with multiple DNA-binding fluorochromes it contained genomic DNA. However this is in conflict with other reports that did not find any cytoplasmic DNA. The organelle also appeared to be filled with electron-dense material, bound by a double membrane, which would have been expected, as it is likely to have descended from a common ancestor that gave rise to the mitochondria (84). The mitosome is morphologically very similar to the double membrane bound hydrogenosomes of *Trichomonas vaginalis* and *Nyctotherus ovalis* (a protozoan parasite of the cockroach hindgut). 'Mitochondrial bodies' of *E. histolytica* were detected years ago by vital staining and cytochemistry (85, 86). The relationship of the recently found organelle with the 'mitochondrial bodies' is not known.

It is likely that the hydrogenosome, the mitochondrion and the recently described organelle of *E. histolytica*, the mitosome, share a common ancestor.

1.4 HEAT SHOCK PROTEINS

The heat shock proteins (HSPs) or molecular chaperones are a highly conserved and abundant family of proteins, members of which have been identified in all organisms examined. These proteins are required for the correct folding, transport and degradation of other proteins *in vivo*. The expression of various heat shock proteins is induced by numerous cellular stresses, including elevated temperature, oxygen deprivation, infection with bacterial and viral agents, exposure to inhibitors of various metabolic processes, heavy metals, amino acid analogues, and oxidants (87-89).

The family of molecular chaperones is large and includes proteins of the hsp100, hsp90, hsp70, hsp60 (i.e. cpn60), hsp40 and hsp20-30 classes which are thought to be involved in protein biogenesis. The number indicates the approximate size of the protein in kilo Daltons (kDa).

Their role in the protecting the organism from heat and other stresses could have played a very important role in the survival of the early life form in the primitive environment with high ambient temperature and deficiency in oxygen (90). These proteins are usually induced under conditions of cellular stress, but still are essential and abundant even in unstressed cells. The proposed role for hsps in unstressed cells includes the modulation of protein-protein interactions (91).

1.4.1 Hsp70 proteins

Genes encoding 70 kDa heat shock proteins (Hsp70s) are present in the genome of most organisms, including archaea, eubacteria and eukaryotes, and the proteins are found in organelles such as mitochondria and chloroplasts. They are one of the most conserved proteins known (92-94).

Hsp70s participate in a number of cellular processes, such as the correct folding of mature proteins in the cytoplasm, without being components of the functional assembled structures. They also facilitate the translocation of newly synthesised polypeptides across membranes, from the cytosol into the mitochondrion and the endoplasmic reticulum, by binding during the initial stages of translocation (94-96).

During times of environmental stress proteins may unfold. The hsp70s then bind to the surfaces of those unfolded proteins exposed by the stress, thus preventing their aggregation and facilitating re-folding. The binding consumes ATP from hsp70, which triggers the conformational changes in the denatured protein allowing it to reassemble correctly.

In eukaryotes, hsp70s constitute a multigenic family, the members of which are localised in the cytosol, endoplasmic reticulum, mitochondria and chloroplasts (72, 94). The major eukaryotic hsp70 groups arose prior to the divergence of the earliest known eukaryotes, approximately 2 billion years ago (97). The hsp70s localised in the mitochondria are most similar to DnaK hsp70 homologues in purple bacteria, which is consistent with the proposed prokaryotic origin of these organelles.

The degree of conservation during evolution between the different hsp70s reflects the biological importance of this protein (87, 92). Proteins from the most distantly related species share at least 42% identity (92, 97). Eukaryotic proteins are between 50 and 98% identical to each other (95).

1.4.2 Cpn60 proteins

Cpn60s are a highly conserved group of chaperonin proteins found in both eukaryotes and prokaryotes. In eukaryotes they are localised in the mitochondria, chloroplasts and hydrogenosomes. The eukaryotic cpn60s are structurally and functionally related to the heat-inducible GroEL protein of *Escherichia coli*. This group which is also known as heat shock protein 60 (hsp60) includes the ribulose biphosphate carboxylase subunit-binding protein of the chloroplast stroma (rubisco).

The eukaryotic cpn60 is synthesised on cytosolic ribosomes before transport into the mitochondrion. This protein is essential for the *in vivo* folding and assembly of polypeptides after transport across membranes.

Cpn60s have played a very important role in the study of molecular evolution of the eukaryotic cell and the endosymbiosis theory. Phylogenetic analyses using the cpn60 protein sequences have been used to study the origin and evolution of mitochondria and chloroplasts (65). The detection of the cpn60 gene provides a clue to whether the organisms harbour(ed) a mitochondrial endosymbiont.

1.4.3 PNT proteins

Pyridine nucleotide transhydrogenase (PNT) is a highly conserved enzyme that has been located in the cytoplasmic membranes of bacteria and in the inner mitochondrial membranes of mammals. It is absent from some bacteria, *Saccharomyces cerevisiae* and probably from higher plants, but it is present in Cyanobacteria and the green alga *Acetabulum acetabularia* (98). In the inner mitochondrial membranes it is associated with the classical components of aerobic metabolism located there. PNT enzyme catalyses the reversible transfer of a hydride ion equivalent between NAD^+ and $NADP^+$

by the reaction: $NADH + NADP^+ + H^+_{out} \leftrightarrow NAD^+ + NADPH + H^+_{in}$* (98, 99). This reaction involves the movement of protons across the cell membrane in prokaryotes and mitochondria in eukaryotes.

Transhydrogenase consists of three components: dI, which has a binding site for NAD, dIII, which has a binding site for NADP and protrudes from the membrane, and dII which spans the membrane (Figure 1.2).

In bacteria, the dII component is always separated into dIIa and dIIb. In *Escherichia coli*, the enzyme consists of two subunits tightly bound to the cytoplasmic membrane: alpha or PntA (comprising dI-dIIa) and beta or PntB (comprising dIIb-dIII) subunits of molecular weights of 54,000 and 48,700 respectively (Figure 1.2) (100, 101). Quite different in subunit structure from that of *E. coli* is the PNT enzyme of *Rhodospirillum rubrum* (an α-Proteobacterium) which consists of a soluble peripheral protein factor having a molecular weight of about 70,000 and an integral membrane-bound component of unknown molecular weight. Its α subunit consists of two proteins AA and AB (Figure 1.2).

The mammalian mitochondrial transhydrogenase consists of a single polypeptide of molecular weight 97,000 to 120,000, which runs dI-dII-dIII, N-terminus to C-terminus (102, 103). The amino terminus of the bovine PNT is homologous to the bacterial α subunit, while the carboxyl terminus is homologous to the bacterial β subunit (104).

PNTs have also been found in a number of protozoan parasites such as *Eimeria tenella* (105, 106), *Plasmodium falciparum* and *Entamoeba histolytica* (67, 81). These are single polypeptides of which the order of components is dIIb-dIII-dI-dIIa. Their amino and

---

* NAD^+: nicotinamide adenine dinucleotide; $NADP^+$, nicotinamide adenine dinucleotide phosphate; NADH: reduced nicotinamide adenine dinucleotide; NADPH: reduced nicotinamide adenine dinucleotide phosphate.

carboxyl ends are homologous with the β and α subunits of bacterial PNTs respectively, thus the opposite to the α and β subunit order of the mammalian mitochondrial transhydrogenase. Quite interestingly, an extra segment is present between the dIII and dI components in *E. histolytica* which possibly serves as a linker (Figure 1.2). The transhydrogenase is believed to be very important in maintaining balance between reduced and oxidized forms of NADP and NADPH particularly in parasites growing under anaerobic conditions in the intestinal lumen.

E. histolytica transhydrogenase enzyme activity was first detected in 1976 by Harlow *et al* (99). It was later shown to be a single β-α fusion protein as the other protozoan parasites (67). Studies on the components of this transhydrogenase showed that it is possibly a proton-translocating transhydrogenase (98). The *E. histolytica* PNT is thought to be located in the mitosome, which possibly possess a proton-pumping electron transport chain that would drive transhydrogenation (98). The high degree of homology between the *E. histolytica* PNT and the other homologues is strong evidence for its mitochondrial origin, although the PNT gene is not well suited for phylogenetic analysis (67).

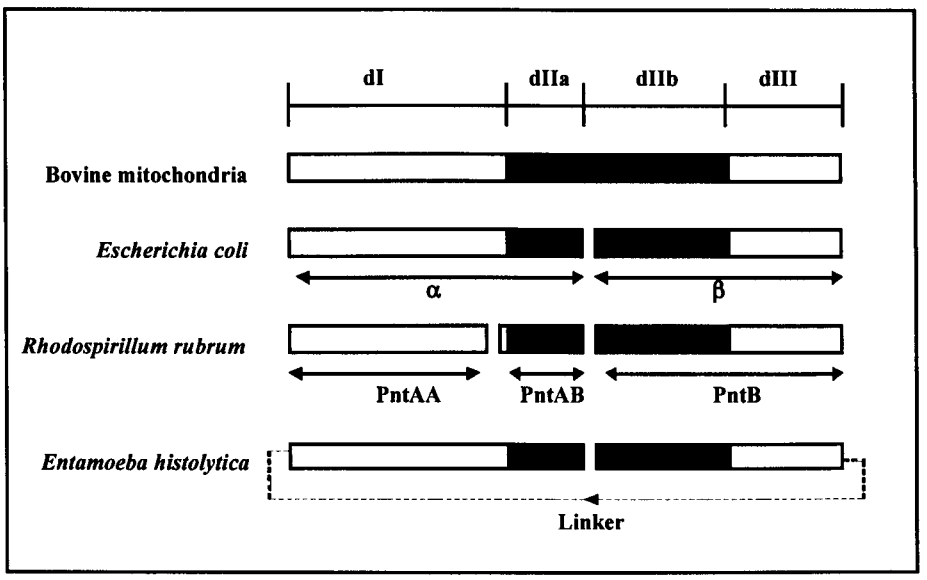


Figure 1.2. Subunit organisation of transhydrogenases.
Taken from Weston *et al* (98).

1.5 PURPOSE OF INVESTIGATION

If *E. histolytica* had a mitochondrial endosymbiont it is likely that more genes encoding 'mitochondrial' proteins will be found. This would support the secondary loss hypothesis for the organism, and indicate that the common ancestor of all the extant eukaryotic species could have been a mitochondriate eukaryote. It would also, hopefully, help identify the current function of the organelle.

The aims of this study were:

1. To isolate and characterise a gene encoding a mitochondrial-like Hsp70 homologue (mt-Hsp70) from *E. histolytica.*
2. Identify the putative targeting amino terminal signals of the mt-hsp70 protein.
3. Do some preliminary cell localisation studies by transfection into the organism.
4. Carry out heat shock experiments for *E. histolytica* and the other *Entamoeba* species and study the expression patterns of the mitochondrial-type genes that have been isolated up to now by Northern blot analysis.
5. Perform phylogenetic analyses of the mt-hsp70 of *E. histolytica* and the cpn60 and PNT genes isolated from other *Entamoeba* species to study the molecular evolution of these genes.

This work would contribute to finding the location of the mt-hsp70 protein in the amoeba and would enhance our understanding of the molecular evolution of the mitosome and the genus *Entamoeba.*

CHAPTER 2
MATERIALS AND METHODS

2.1 *ENTAMOEBA* ISOLATES

The *Entamoeba* isolates used in this study are given in table 2.1. Where a description of the isolate has been previously being published the citation is given. The American Type Culture Collection (ATCC, Rockville, MD) accession number is also given from where the isolates are available.

Table 2.1. *Entamoeba* isolates used in the study.

Species	Isolate designation	Host (county of origin)	Reference	ATCC accession no.
Entamoeba histolytica	HM-1:IMSS clone 9	Human (Mexico)	(107)	30459
Entamoeba dispar	SAW 760	Human (England)		50484
Entamoeba invadens	IP-1	Snake (*Natrix cyclopion*)	(108)	30994
Entamoeba moshkovskii	Laredo	Human (USA)	(45)	30042
Entamoeba terrapinae	M	Turtle (*Chrysemys picta*)	(108)	30043

2.2 *IN VITRO* CULTIVATION OF *ENTAMOEBA* SPECIES

Trophozoites of strain HM-1:IMSS clone 9 of *Entamoeba histolytica* were grown axenically in YI-S or LYI-S-2 medium with 15% adult bovine serum. *Entamoeba dispar* SAW 760 cultures were grown monoxenically in YI-S or LYI-S-2 with 10% adult bovine serum supplemented with *Crithidia fasciculata* ReF-1: PRR (ATCC 50083) grown in the same medium. The rest of the isolates were grown axenically in YI-S or LYI-S-2 with 10% bovine serum.

E. histolytica and *E. dispar* cells were cultured in 16 x 125 mm borosilicate glass screw cap culture tubes (Corning). The cultures were chilled for 5 minutes in an ice-water bath to release attached trophozoites from the tube and then inverted a few times to disperse the cells. The inoculum was passed aseptically to a culture tube containing 13 ml of freshly prepared medium. In the *E. dispar* cultures 0.1 ml of *Crithidia fasciculata* was added to each tube. The tubes were incubated at 36 °C, at 5° to the horizontal. The subculturing was done twice a week at 72-96 hour intervals.

E. moshkovskii Laredo, *E. terrapinae* M and *E. invadens* IP-1 cells were cultured as above except that the tubes were incubated at 22 °C at 5° to the horizontal. The subculturing was done every 2 weeks.

2.2.1 YI-S medium for *Entamoeba* species

This medium was developed by Diamond *et al* (109) as an alternative to TYI-S-33 (110), which is the most widely used medium. Due to difficulties in obtaining lots of casein digest peptone that would support adequate growth of *Entamoeba* cells, the casein was replaced by additional yeast extract.

The following ingredients are dissolved in 600 ml of reverse osmosis purified water (MilliQ II, Millipore): K_2HPO_4, 1 g; KH_2PO_4, 0.6 g; NaCl, 2.4 g; yeast extract, 30 g; glucose, 10 g; L-cysteine hydrochloride, 1 g; ascorbic acid, 0.2 g; ferric ammonium citrate (brown form), 22.8 mg. The pH was adjusted to 6.8 with 1 N NaOH (7.5 ml were used usually). Water was added to bring the final volume to 880 ml. The medium was aliquoted in units of 88 ml into PYREX® 100ml bottles and autoclaved for 15 minutes at 121 °C. The autoclaved medium was then stored at –20 °C for up to 6 months and thawed every time we needed to use it.

2.2.2 LYI-S-2 medium for *Entamoeba* species

The medium was developed by Clark and Diamond (unpublished), due to difficulties obtaining lots of yeast extract that could support adequate growth of *Entamoeba* cells. The only difference with the YI-S medium is that 5 g of the yeast extract are replaced by 5 g of neutralised liver digest (Oxoid).

2.2.3 Vitamin mixture #18 & bovine serum

The vitamin mixture preparation was described by Diamond *et al* (109).

Solution 1: Niacinamide, 45 mg; pyridoxal hydrochloride, 4mg; calcium pantothenate, 23 mg; thiamine hydrochloride, 5 mg; and vitamin B12, 1.2 mg were dissolved in distilled water to a final volume of 25 ml. Riboflavin, 7 mg was solubilised using minimal amounts of 0.1 N NaOH and dissolved in water to a final volume of 45 ml. Folic acid, 5.5 mg was also solubilised using minimal amounts of 0.1 N NaOH and dissolved in water to a final volume of 45 ml. d-biotin, 2 mg was dissolved in water to a final volume of 45 ml. The solutions were combined together.

Solution 2: DL-6,8 thioctic acid (oxidised), 1 mg was dissolved in 5 ml of 95% ethanol and 500 mg of Tween-80. Water was added to bring to a final volume of 20 ml.

Solutions 1 and 2, were combined and water was added to bring to final volume of 200 ml. The final solution was filter-sterilised through cellulose nitrate membrane filter (pore size, 0.22 μM; Nalgene), aliquoted into 15 ml glass tubes and stored at 4 °C. The final concentration of the vitamin mixture is 50x.

Adult bovine serum (Sigma-Aldrich) was heat inactivated for 30 minutes at 56 °C before being used. It could be stored at 4 °C for up to 6 months.

For one unit of medium (ca. 100 ml), 2 ml of vitamin mix and 15 or 10 ml bovine serum were added to the 88 ml of the YI-S broth. Any unused medium was then stored at 4 °C for up to 2 weeks.

2.3 DNA EXTRACTION

2.3.1 DNA purification from trophozoites of *Entamoeba* species

The method for the rapid isolation of *Entamoeba* DNA (111) uses cetyltrimethylammonium bromide (CTAB) for DNA purification. This is a cationic detergent that helps remove the polysaccharides from cells. These interfere with many enzymes during DNA analysis.

Culture tubes were chilled in ice for 5 minutes and then centrifuged at 275 x *g* for 3 minutes to pellet the *Entamoeba* cells. The pellet was washed with sterile phosphate buffer saline (PBS), transferred into a 1.5 ml microcentrifuge tube where cells were pelleted (≤ 50μl packed cell volume) and then dispersed in 0.25 ml lysis buffer (0.1 M EDTA, pH 8.0, 0.25% SDS). Proteinase K was added to a final concentration of 0.1 mg

ml^{-1} and the tube was incubated at 55 °C for 1 hour. 3.5 M NaCl was added a final concentration of 0.7 M and mixed. CTAB (10% in 0.7 M NaCl) was added to a final concentration of 1%, mixed and incubated at 65 °C for 15 minutes. The lysate was extracted with 1 volume of chloroform and centrifuged at 14,000 x *g* for 10 minutes. The supernatant was then extracted with an equal volume of phenol : chloroform and centrifuged again for 10 minutes. The supernatant was precipitated with 2.5 volumes of 100% ethanol for 10 minutes at room temperature and then centrifuged at 14,000 x *g* for 10 minutes. The pellet was washed with 70% ethanol, centrifuged for 5 minutes at 14,000 x *g*, air dried and finally resuspended in a small volume of sterile water or 10 mM Tris-Cl, pH 8.0; 1 mM EDTA.

After resuspending the DNA, the sample was passed through a microspin Sephacryl-400 column (Pharmacia Biotech) to clean it of certain compounds (e.g. salts) that can be inhibitory to PCR, restriction enzymes and other molecular reagents. 1 µl of RNase A (1 mg ml^{-1}) was added to the final solution to remove any RNA that may have been left.

2.3.2 DNA extraction from agarose gels

DNA samples were electrophoresed in an 1% agarose gel in 150 Volts and the gel was stained with 1 µg ml^{-1} ethidium bromide and photographed under the UV light. The band of interest was cut from the gel, crushed through an 1 ml syringe into an empty column (micro bio-spin chromatography columns, BioRad), and centrifuged at 14,000 x *g* for 5 minutes. The sample was then purified using an S-200 spin column and the DNA was either ethanol precipitated from the flow through liquid or used directly.

2.4 TRANSFORMATION OF BACTERIAL CELLS

Bacterial transformation reactions were performed after ligating PCR products into a pGEM® T-easy vector (Promega), either by heat shocking the cells to allow uptake of DNA, or by electroporation, a process that transiently permeabilises the cell membranes with an electric pulse, thus permitting cell uptake of a wide variety of biological molecules.

2.4.1 Transformation of cells by heat shock

E. coli JM109 and DH5α high efficiency competent cells (Promega, Life Technologies) were used for the transformation reactions. The cells were mixed with plasmid DNA, heat shocked at 42 °C for 45 seconds, mixed with SOC medium, incubated at 37 °C for 1.5 hours and then plated onto LB and ampicillin (100 μg ml^{-1}) plates with X-Gal (80 μg ml^{-1}) and IPTG (0.5 mM), to allow growth of the cells with plasmids. The plates were incubated at 37 °C overnight (112).

2.4.2 Electro-transformation

E. coli JM109 and DH5α were prepared for transformation as described by Dower *et al* (113). The cells were grown at 37 °C, with vigorous shaking in LB broth until they reached an ABS_{600} of 0.5 to 1. They were then harvested by chilling the flasks briefly on ice and centrifuging at 4,000 x *g* for 15 minutes at 4 °C. The cells were resuspended in cold 1 mM HEPES (pH 7.0), centrifuged as above, resuspended in cold 1 mM HEPES, centrifuged, resuspended in 10% glycerol, centrifuged, and finally resuspended in 10% glycerol (concentration of the cells should be between 2 - 4 x 10^{10} ml^{-1}). The concentrated suspension was aliquoted in 1.5 ml eppendorf tubes (40 μl in each) and stored at -70 °C.

The cells were then transformed by high voltage electroporation using the Gene Pulser® II from BioRad. Cells were mixed with the DNA and allowed to sit on ice for 1 minute. The pulse generator was set to the 25 μF capacitor, 1.8 kV, and 200 Ω in parallel with the sample buffer. The mixture was transferred to a cold 0.1 cm electroporation cuvette. One pulse was applied at the above settings that resulted in a discharge of 12.5 kV/cm with a time constant of 4.5 to 5 msec. SOC medium was added immediately to the cuvette and cells were resuspended with a pipette. The cell suspension was transferred to a polypropylene tube and incubated at 37 °C for 1 hour. The cells were then plated in LB agar with ampicillin plates and X-Gal/IPTG as above (section 2.4.1). The plates were incubated at 37 °C overnight to allow growth of colonies.

White colonies were picked the next day, allowed to grow in LB and ampicillin broth, before isolation of the recombinant plasmid DNA.

A heavy inoculum of cells that had grown in LB broth with ampicillin was mixed with equal volume of 50% glycerol and stored at -70 °C.

2.4.3 Isolation of plasmid DNA from recombinant cells

DNA from recombinant bacterial cells was extracted by the alkaline lysis method followed by adsorption of DNA onto silica in the presence of high salt and elution in low-salt buffer. The QIAprep Spin Miniprep, the Endofree Plasmid Maxi, and the Plasmid Maxi kit kits from QIAGEN were used for this purpose.

2.5 RESTRICTION ENZYME DIGESTION

Pure *Entamoeba* DNA or plasmid DNA was digested with the appropriate restriction endonucleases, obtained from Promega, Boehringer Mannheim and MBI Fermentas, under conditions recommended by the suppliers.

2.6 mRNA EXTRACTION

Poly A^+ RNA was isolated from axenically grown normal and heat shocked *Entamoeba* cells using the QuickPrep® Micro mRNA Purification Kit from Pharmacia Biotech.

Culture tubes were chilled in ice for 5 minutes and then centrifuged for 3 minutes at 275 x *g*. The pellet was washed with PBS and centrifuged three times then the manufacturer's protocol was followed. The kit uses guanidinium thiocyanate, which protects and denatures the mRNA, and Oligo (dT)-Cellulose tubes to which the poly A^+ RNA binds. The tubes are then washed sequentially with high-salt and low-salt buffer to clean the sample from any proteins, other nucleic acids and carbohydrates. The mRNA is then eluted through a polypropylene minicolumn in 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. The concentration of the polyadenylated RNA was determined by spectrophotometry ([RNA]= A_{260} x 40 μg ml^{-1}), precipitated in glycogen, potassium acetate solution and 95% ethanol and finally redissolved at 0.1 μg μl^{-1} in elution buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA).

2.7 TOTAL RNA EXTRACTION

Total RNA was isolated from axenically grown normal and heat shocked *Entamoeba* cells using a guanidinium method (114).

Cells were harvested and washed as above (section 2.3.1) before the protocol was followed. 0.25 ml of extraction buffer (mRNA purification kit, Amersham Pharmacia Biotech) per 50 µl of pelleted cells was used to denature the RNA. 0.1 ml of 2 M sodium acetate (pH 4.0), and was added and mixed. An equal volume of phenol: chlorophorm: isoamyl alcohol was added and incubated at 4 °C for 15 minutes. The suspension was centrifuged at 14,000 x *g* for 20 minutes at 4 °C. The aqueous phase was precipitated with 1 volume of 100% isopropanol for 30 minutes at -20 °C and then centrifuged at 10,000 x *g* for 10 minutes at 4 °C. 4 M LiCl was added to wash out the glycogen from the RNA pellet and insoluble RNA was pelletted by centrifuging for 10 minutes at 5,000 x *g*. RNA was dissolved in extraction buffer and precipitated again with equal volume of 100% isopropanol as described above. The RNA pellet was washed in 75% ethanol, incubated for 15 minutes at room temperature and centrifuged for 5 minutes at 10,000 x *g*. The pellet was then dried in a vacuum for 15 minutes, dissolved in 100 µl of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and incubated for 15 minutes at 55 °C.

The concentration of the total RNA was determined by spectrophotometry ([RNA]= A_{260} x 40 µg ml^{-1}), precipitated in 0.4 M sodium acetate and 3 volumes of 100% ethanol, and finally re-dissolved at 2 µg $µl^{-1}$ in elution buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA).

2.8 POLYMERASE CHAIN REACTION TECHNIQUES

2.8.1 Gene amplification

The reagents used for the PCR reaction were from Promega, MBI Fermentas or Bioline and used either *Taq* or *Pfu* DNA polymerase from the first two or BIOTAQ™ DNA polymerase for the latter source. Concentrations used were according to the suppliers' recommendations.

The PCR assay was performed in 50 or 100 µl reactions containing at a final concentration of 1 µM dNTPs, 100 pmol μl^{-1} of each primer, 1x PCR buffer, 1.5 mM $MgCl_2$ and 2.5 units μl^{-1} of the DNA polymerase, and ca. 5 ng of DNA. The reaction was overlaid with two drops of nuclease-free mineral oil to prevent condensation and evaporation and the tube was then placed into a thermocycler.

The thermal cycling programme consisted of 30 cycles of denaturation at 94 °C for 1 minute, annealing at primer-specific temperature for 1.5 minutes and elongation at 72 °C for 2 minutes. The reaction was finished with 1 cycle at 72 °C for 5 minutes.

For each of the *Entamoeba* species different annealing temperatures were used to get optimal results according to the Tm of the primers. Touchdown PCR was also used for some species to get the desired PCR product (115).

10 µl of the PCR product was then electrophoresed in a 1% agarose gel in 1x TBE buffer at 10 Volts cm^{-1}. The gel was then stained with 1 µg ml^{-1} ethidium bromide and photographed under the UV light.

The PCR product was cleaned using QIAquick PCR purification kit (QIAGEN) before being used for further investigations.

Positive PCR products were cloned into pGEM®-T Easy vector system II (Promega) following the manufacturers' protocols, and transformed into JM109 cells using heat shock as above (section 2.4.1). Transformed cells were plated, colonies selected and plasmid DNA isolated as above (section 2.4). The plasmid DNA was then used for sequencing to identify the origin of the clones.

2.8.2 Reverse transcription- polymerase chain reaction (RT-PCR)

RT-PCR was used for 5′ and 3′ RACE (Rapid amplification of cDNA ends) analysis. The Access RT-PCR System (Promega) and the Marathon™ cDNA Amplification (Clontech) kits were used according to the manufacturers' protocols. All the components that were required in the Marathon cDNA amplification kit were included except the ones used for the PCR reaction. For this purpose the same company supplied the Advantage® 2 PCR Enzyme System kit.

Poly A^+ RNA was isolated as described in section 2.6.

5 µl of the products were analysed in 1.2% agarose gel by electrophoresis at 10 Volts cm^{-1} in 1x TBE buffer and stained with ethidium bromide (1 µg ml^{-1}). The PCR products were then purified using the QIAquick PCR Purification kit by QIAGEN.

Positive PCR products were cloned into pGEM®-T Easy vector system II (Promega), transformed, plated and colonies selected as above (section 2.4).

2.8.3 Cycle sequencing

Three different cycle sequencing mixes were used; The Terminator Ready Reaction Mix kit (Perkin Elmer), the Thermo Sequenase II dye terminator cycle sequencing kit (Amersham Pharmacia Biotech) and the ABI PRISM BigDye™ Terminator Cycle sequencing Ready Reaction kit (PE Biosystems). The manufacturers' protocols were followed for the first two kits. In the BigDye kit, 4 µl of Terminator Ready Reaction Mix were used instead of 8 µl, supplemented with 1 µl of 10x KCl buffer. The amount of primer added was increased from 3.2 pmol to 6.4 pmol per reaction and the amount of DNA used was always 450 ng per reaction.

The primers that were used to obtain the complete sequence of the mt-hsp70 gene of *E. histolytica* and the partial sequence of the mt-hsp70 of the other *Entamoeba* species are given in Appendix A.

To remove excess dye-labelled dideoxynucleotide terminators following sequencing, the reactions were passed either through a microspin column S-200 (Pharmacia Biotech), or purified using the DyeEx™ Spin kit (QIAGEN) following the manufacturers' protocol. Alternatively they were precipitated using a modified method to the one suggested in the Amersham sequencing kit protocol.

20 µl from the termination reaction were removed (avoiding the oil), and transferred to a tube containing 7.5 M ammonium acetate (Sigma) to a final concentration of 2.5 M. 2.5 volumes of 95% ethanol (-20 °C) were added and the tube was left on ice for 15 minutes. It was then centrifuged for 15 minutes at 14,000 x *g* and the pellet was washed in 70% ethanol. The tube was centrifuged for 2 minutes at 14,000 x *g* again, and finally the pellet was vacuum-dried in a vacuum centrifuge. The sample was then ready for loading on the automated DNA sequencer ABI PRISM™ 377 (Perkin Elmer).

Sequences obtained were assembled using the Autoassembler™ DNA Sequence Assembly Software by Perkin Elmer. Derived amino acid sequences were compared with those previously published using the BLAST 2.0 programme (116) of the National Centre for Biotechnology Information (NCBI) databases.

2.9 LABELLING A PROBE

Partial cpn60, PNT, mt-hsp70 and β-actin genes were amplified by PCR (section 2.8.1) with gene specific primers (Appendix A) and labelled with radioactive [α-^{32}P]-dCTP to be used as a probe for detecting the gene or mRNA in the different molecular biology techniques that have been used. The template DNA was either total genomic DNA or a

recombinant plasmid containing cloned PCR amplified gene fragments, or cDNA and genomic clones isolated from DNA libraries.

The partial DNA probes amplified were random labelled with [α-^{32}P]-dCTP using the Rediprime system from Amersham or the Prime-a Gene® Labelling system from Promega. Both methods use a mixture of random sequence hexadeoxyribonucleotides to prime DNA synthesis on any denatured DNA template (117, 118). The sample was then purified using an S-200 microspin column (Pharmacia Biotech). The labelled probe was then transferred to the container with the prehybridised blot and mixed with the hybridisation solution.

The filters were prehybridised in a capped Pyrex glass tube (HYBAID™) containing 12 ml of 1M NaCl, 1% SDS, and 10% dextran sulfate. The tube was then placed into a rotisserie oven at 65 °C for the Southern blot and the phage lift filters. After a few hours the labelled probe was added and the filters were left to hybridise overnight.

The following day the filters were washed with vigorous agitation in 2x SSC/ 1% SDS, then in 0.5x SSC/ 0.25% SDS and 0.2x SSC/ 0.1% SDS for 30 minutes each at 65 °C. The filters were then wrapped in cling film and exposed to X-ray film at -70 °C overnight.

For the Northern blot filters, the prehybridisation/hybridisation solution was provided by the NorthernMax™ kit (Ambion Inc). The filter was prehybridised, hybridised at 42 °C with a ^{32}P labelled β-actin gene probe, exposed to X-ray film, then stripped by a 30 minute incubation in a boiling solution of 0.1% SDS in diethyl pyrocarbonate (DEPC) treated H_2O of the filter, and rehybridised to one of the other DNA probes (Appendix A, Table A.2). The filters were washed in low stringency solution (2x SSC/ 0.1% SDS) twice for 5 minutes and then in high stringency solution (0.1x SSC/ 0.1% SDS) twice for 15 minutes each at 42 °C. They were then exposed to X-ray film at -70 °C for about 2 weeks for the mt-hsp70 genes or less for the rest of the gene probes depending on the

amount of the radioactivity present on the filter. The actin was used as a non-induced control and the probe was stripped because of the relative abundance of actin compared to the other mRNAs (estimated to be > 100:1). Residual actin probe was still visible after stripping and exposure.

2.10 SOUTHERN BLOT ANALYSIS

Genomic *E. histolytica* DNA was digested with restriction endonucleases overnight. The samples were then loaded in a 0.8% agarose gel, electrophoresed in 1x TBE buffer at 8 Volts cm^{-1}, stained with 1 $\mu g\ ml^{-1}$ ethidium bromide and photographed under the UV light.

The gel was incubated in 0.25 M HCl for 10 minutes to increase the efficiency of transfer of the large fragments of DNA to the membrane by depurinating the strands. The DNA was then denatured for 30 minutes in 1.5 M NaCl/ 0.5 M NaOH, and finally neutralised in 1.5 M NaCl/ 1 M Tris pH 7.0 for 30 minutes. Incubations were done with continuous agitation.

The Southern blot apparatus was set up by using the upward capillary transfer method and the DNA was left to transfer to the nylon membrane Biodyne B (Life Technologies) overnight.

The next day the transfer apparatus was disassembled and the filter was removed and blotted dry on a Whatman 3MM filter paper. The DNA was then crosslinked to the filter using the Stratagene's UV Stratalinker™ 1800 for 30 seconds.

Prehybridisation and hybridisation of the filter was as described in section 2.9.

2.11 NORTHERN BLOT ANALYSIS

The NorthernMax™ kit from Ambion was used to size fractionate the mRNA and everything was made according to the manual's instructions. The mRNA samples and RNA size markers 0.28-6.58 kb (Promega) were loaded in a formaldehyde-containing 1% agarose mini-gel and electrophoresed at 10 Volts cm^{-1}. The markers were subsequently stained with ethidium bromide (1 $\mu g\ ml^{-1}$) and photographed under the UV light with a fluorescent ruler.

RNA was transferred to a nylon Biodyne B membrane (Life Technologies) using the downward capillary transfer method for 1.5 hours and then hybridised as described in section 2.9.

2.12 PULSED FIELD GEL ELECTROPHORESIS (PFGE)

2.12.1 Preparation of plug molds

E. histolytica HM-1:IMSS clone 9 cells were grown, harvested and washed as described above (section 2.3.1). Cells were then transferred into a 14 ml Falcon tube (Becton Dickinson) and washed with PBS. An equal volume of 2% low melting point (LMP) agarose was added, mixed gently and kept at 42 °C. 80 µl of the cell/agarose mixture was transferred into plug molds (BioRad) and left on ice for 15 minutes to allow the agarose to solidify. The plugs were then transferred into a 15-ml tube containing 10 ml of lysis buffer (0.5 M EDTA, pH 9.5; 1% sodium lauryl sarcosine) and 1 mg ml^{-1} proteinase K, and incubated at 50 °C for 48 hours. The plugs were then kept at 4 °C until used.

2.12.2 PFGE method

Each agarose plug was inserted into the sample well of a 1% agarose gel and the wells were sealed with 2% LMP agarose. The agarose was allowed to set and then run in 14 °C, 0.5x TBE, using a CHEF MAPPER™ (BioRad). Electrophoresis was at 6 Volts cm^{-1}, with a 120° angle and a variable linear ramping factor for switching times.

The gel was then stained for 45 minutes with 1 μg ml^{-1} ethidium bromide, destained in H_2O for 30 minutes and photographed under the UV light. The Southern blot was then set up as described in section 2.10.

To estimate the molecular weights of the DNA in the sample, a lambda ladder (BioRad) and Yeast chromosomal marker (BioRad) were used.

Different conditions were tried to get optimal results. The concentration of the buffer was changed to 0.4x TBE and finally to 0.25x TBE. The range of expected molecular sizes setting was varied between gels by altering the switching programme.

2.13 SCREENING OF LIBRARY & *IN VIVO* EXCISION OF THE PHAGEMID

Two types of libraries were screened in order to isolate clones of the mt-hsp70 gene of *E. histolytica*. The first was a cDNA library of strain SFL-3 that was made between an *Eco*RI and an *Xho*I site, and cloned into λZap II vector (provided by Dr Michael Duchêne, Vienna). The second one was a genomic DNA library of strain HM-1:IMSS that was made between two *Eco*RI sites and cloned also in a λZap II vector (119). The genomic library was a gift from Dr. Egbert Tannich (Hamburg).

For the screening of the two libraries the protocol from the Lambda Zap® II/*Eco*RI cloning kit by Stratagene was followed. The cloning kit uses *E. coli* XL-1 Blue cells for the blue/white colour selection. XL-1 cells infected with the library were plated on NZY agar plates overnight.

Two rounds of screening were done. In the first round plates with approximately 12,500 plaques per 90 mm plate, were chosen. In the second round 220 plaques per 90 mm plate or less, were chosen for screening. These were chilled for 2 hours at 4 °C and then the plaques were then transferred in duplicate to 82 mm nylon filters (Hybond-N+, Amersham Pharmacia Biotech). The filters were denatured (1.5 M NaCl/0.5 M NaOH) for 2 minutes, neutralised (1.5 M NaCl/ 0.5 M Tris-HCl, pH 8.0) for 5 minutes, rinsed for 30 seconds in 0.2 M Tris-HCl (pH 7.5)/ 2x SSC, blotted dry and finally crosslinked using the UV Stratalinker 1800 (Stratagene) for 30 seconds.

The filters were prehybridised/hybridised with mt-hsp70 probe at 65 °C in a rotisserie oven and exposed into an X-ray film at -70 °C, as described in section 2.9. Positive plaques were identified, extracted from the plates and transferred into sterile 1.5 ml micro-centrifuge tubes containing 500 µl SM buffer (For 1 litre: 5 g NaCl, 2 g $MgSO_4$, 50 ml 1 M Tris-HCl pH 7.5, 5 ml 2% gelatin) and 2-3 drops of chloroform to release the Lambda ZAP II phage particles. The tubes were stored at 4 °C.

The *in vivo* excision protocol using the ExAssist helper phage/XLOLR strain system (Stratagene) was used to excise the phagemids. The kit uses XL1-Blue- MRF′ and XLOLR cells for the excision.

The excised pBK-CMV phagemid, which was packaged as filamentous phage particles, was mixed with freshly prepared XLOLR *E. coli* cells and spread on LB-ampicillin (50 µg ml^{-1}) agar plates to produce colonies. These colonies contained the pBK-CMV double-stranded DNA phagemid with the cloned DNA insert.

Colonies were picked and grown in LB and ampicillin broth. The recombinant plasmid DNA was isolated as described in section 2.4.3 and sequenced by primer walking (section 2.8.4). The primers used to get the complete mt-hsp70 sequence of *E. histolytica* are given in Appendix A.

2.14 PHYLOGENETIC ANALYSIS

2.14.1 Sequence alignment of hsp70 proteins

The complete sequence of the mt-hsp70 gene of *E. histolytica* was aligned with mitochondrial, hydrogenosomal and related bacterial hsp70 homologues of 15 species.

The complete alignment of the hsp70 proteins can be seen in Appendix B. The areas that were not included in the phylogenetic analysis are the alignment positions: 1-61, 104-109, 150, 170-171, 203-205, 256-259, 282-285, 319-332, 375-394, 450, 487-488, 525 and 637-782. These positions were excluded because the alignment was not unambiguous.

Table 2.2. Organisms used for the sequence alignment and phylogenetic analysis of the hsp70 proteins.

Species name	Accession number in GenBank™
Mitosome:	
Entamoeba histolytica	AF214549 & AF214550
Hydrogenosome:	
Trichomonas vaginalis	AAB17251 & U70308
Mitochondria:	
Saccharomyces cerevisiae	M27229
Mus musculus	P38647
Leishmania major	X64137
Eimeria tenella	Z46965
Pisum sativum	X54739
α-Proteobacteria:	
Rhodopseudomonas sp.	D78133
Agrobacterium tumefaciens	X87113
Rhodobacter capsulatus	U57637
Rickettsia prowazekii	AJ235270
γ-Proteobacteria:	
Legionella pneumophila	D89498
Haemophilus ducreyi	U25996
Chlamydia:	
Chlamydia trachomatis	X52175
Spirochaete:	
Borrelia burgdorferi	X67646

2.14.2 Sequence alignment of cpn60 proteins

Partial sequences of the cpn60 gene of *E. histolytica*, *E. dispar*, *E. invadens*, *E. terrapinae* and *E. moshkovskii* Laredo were aligned with mitochondrial, hydrogenosomal and related bacterial cpn60 homologues of 21 species. The alignment of the cpn60 proteins can be seen in Appendix C.

Table 2.3. Organisms used for the sequence alignment and phylogenetic analysis of the cpn60 proteins.

Species name	Accession number in GenBank™
Mitosome:	
Entamoeba histolytica	AF029366
Entamoeba dispar	-
Entamoeba invadens	-
Entamoeba moshkovskii Laredo	-
Entamoeba terrapinae	-
Hydrogenosome:	
Trichomonas vaginalis	AAB39487
Mitochondria:	
Saccharomyces cerevisiae	P19882
Homo sapiens	NP_002147
Zea mays	P29185
Trypanosoma cruzi	Q95046
Trypanosoma brucei	Q37683
α-Proteobacteria:	
Bradyrhizobium japonicum	Z22603
Bartonella bacilliiformis	P35635
Agrobacterium tumefaciens	P30779

Table 2.3 Continued over page.

Table 2.3 Continued from previous page.

Species name	Accession number in GenBank™
α-Proteobacteria:	
Brucella abortus	L09273
Rhizobium meliloti	P35469
Rickettsiales:	
Cowdria ruminantium	U13638
Rickettsia tsutsugamushi	P16625
Ehrlichia chaffeensis	P42382
γ-Proteobacteria:	
Legionella micdadei	P26194
Haemophilus influenzae	P43733
Pseudomonas aeruginosa	P30718
Escherichia coli	P06139
Coxiella burnetti	P19421
β-Proteobacteria:	
Neisseria gonorrhoeae	P29842
Chlamydia:	
Chlamydia pneumoniae	P31681

2.14.3 Sequence alignment of PNT proteins

Partial sequences of the PNT gene of *E. histolytica*, *E. dispar*, *E. invadens*, *E. terrapinae* and *E. moshkovskii* Laredo were aligned with mitochondrial, and related bacterial PNT homologues of 11 species. The alignment of the PNT proteins can be seen in Appendix D.

Table 2.4. Organisms used for the sequence alignment and phylogenetic analysis of the PNT proteins.

Species name	Accession number in GenBank™
Mitosome:	
Entamoeba histolytica	AAC41577
Entamoeba dispar	-
Entamoeba invadens	-
Entamoeba moshkovskii Laredo	-
Entamoeba terrapinae	-
Mitochondria:	
Homo sapiens	GO2257
Eimeria tenella	T18520
Plasmodium falciparum	Genome project data
α-Proteobacteria:	
Rhodospirillum rubrum	AAC43257
Rickettsia prowazekii	NP_220468
γ-Proteobacteria:	
Escherichia coli	NP_288037
Haemophilus influenzae	AAC23010
Vibrio cholerae	NP_232954
β-Proteobacteria:	
Neisseria meningitidis MC58	NP_265026
Synechocystis sp. PCC 6803	BAA17540

Table 2.4. Continued over page.

Table 2.4. Continued from previous page.

Species name	Accession number in GenBank™
Mycobacteria:	
Mycobacterium tuberculosis	NP_214671

2.14.4 Maximum parsimony analysis

A phylogenetic tree was constructed with maximum parsimony (MP) using the PAUP* 4.0b2 programme (120). Unweighted MP analysis was carried out by 50 rounds of random stepwise addition heuristic searches with tree bisection reconnection (TBR) branch swapping. Bootstrap analysis was carried out on 500 bootstrap re-samplings. Gaps were treated as a 21st amino acid.

2.14.5 Distance analysis

The distance analysis phylogenetic tree was also constructed using the PAUP* 4.0b2 programme. Mean character difference with minimum evolution and TBR branch swapping was carried out. Topology support was assessed using 500 bootstrap re-samplings for the distance generated trees.

2.14.6 Maximum likelihood analysis

The maximum likelihood (ML) trees were generated using PUZZLE 4.0.2 programme (121). Amino acid frequencies, the proportion of invariable sites and the α parameter of the gamma distribution (8 categories) were estimated from the data set by maximum likelihood optimisation using the JTT model of amino acid substitution. Quartet puzzling (QP) trees were derived from 10000 puzzling steps. For maximum likelihood

distance bootstrap analysis of the mt-hsp70 sequence, the estimated PUZZLE parameters were used with Puzzleboot 1.0.3 (M.E. Holder and A.J. Roger, unpublished) to build bootstrapped distance matrices in conjunction with the PUZZLE program, and trees were generated using the FITCH programme with global rearrangements (122).

2.15 EXPRESSION VECTOR CONSTRUCTION

2.15.1 EhNEOplus vector with cpn60 lacking the amino terminus (NtΔ) experiments

Two primers complementary to each other were made based on the mt-hsp70 gene putative targeting sequence (mt-hsp70 TS): "MFVSQPARSTC..." to create an adaptor.

```
MTS-hsp70 I:         A TGT TTG TTT CAC AAC CAG CAA GAT CAA CTT GT AC
MTS-hsp70 II:  CA TGT ACA AAC AAA GTG TTG GTC GTT CTA GTT GAA
```

The method followed was by Broude *et al* (123). Equal volumes of each oligonucleotide was mixed in a 1.5 ml eppendorf tube in a total volume of 20 µl, and spun briefly. The sample was incubated in a 75 °C water bath for a few seconds until the tube reached that temperature. The water bath was then turned off and transferred into a 4 °C room. The pair of oligonucleotides was annealed by gradually cooling down the mixture in the water bath from 75 °C to 10 °C for approximately 1.5 hours. When the water bath temperature reached 10 °C the adaptor was ready to use. The sample was then stored at –20 °C until further use.

The plasmid DNA R.CPN60-NtΔ (construct B) containing the cpn60 gene without its amino-terminal sequence cloned into the EhNEOplus vector (82) was restriction digested with *Kpn*I in a total volume of 100 µl, at 37 °C overnight. The *Kpn*I enzyme

was then heat inactivated by a 70 °C incubation for 30 minutes. The sample was then passed through an S-200 microspin column.

50 pmol of the mt-hsp70 TS adaptor were ligated into the R.CPN60-NtΔ using 3 Weiss units of T4 ligation enzyme (Promega) in a 10 µl reaction volume. The ligation reaction was incubated at 4 °C overnight and then used directly for transformations into bacterial cells (section 2.4).

2.16 TRANSFECTION OF *E. HISTOLYTICA* TROPHOZOITES

Plasmid DNA was extracted as described in section 2.4.3 to be used for the transfection of *E. histolytica* trophozoites.

E. histolytica trophozoites in late logarithmic growth stage were harvested and washed twice in PBS as above (section 2.3.1), and then once in cytomix buffer (10 mM K_2HPO_4/KH_2PO_4 (pH 7.6), 120 mM KCl, 0.15 mM $CaCl_2$, 25 mM HEPES, 2 mM EGTA, and 5 mM $MgCl_2$). Pelleted cells were resuspended in 2.5 ml of cytomix buffer supplemented with 4 mM ATP and 10 mM glutathione. 0.8 ml of cells were transferred into a 0.4-cm gap electroporation cuvette (BioRad) together with 60 – 200 µg of plasmid DNA. The electroporation was performed with the Gene-Pulser® II (BioRad). Two consecutive pulses were applied, each with an exponential discharge of 3000 V/cm, with a 25 µF capacitance and 200 Ω in parallel, and a time constant of 0.74 msec (124, 125). Electroporated cells were transferred into LYI-S-2 medium with 15% bovine serum supplemented with 25 units ml^{-1} of penicillin and 25 µg ml^{-1} streptomycin solution and incubated at 36 °C, at 5° to the horizontal.

Drug selection was started after 48 hours using the antibiotic G418 at 10 μg ml^{-1} concentration. After another 48 hours the concentration was decreased to 3 μg ml^{-1}. When cell growth was stable, the G418 concentration was increased gradually up to 60 μg ml^{-1}. The cultures were re-fed with fresh medium regularly to eliminate cellular debris and dead cells.

CHAPTER 3

SEQUENCE ANALYSIS OF THE MT-HSP70 GENE OF *ENTAMOEBA HISTOLYTICA*

3.1 INTRODUCTION

The isolation and characterisation of a gene encoding a mitochondrial-type heat shock protein 70 homologue (mt-hsp70) from *E. histolytica* is reported in this study. Using degenerate primers based on conserved amino acid residues observed in a published alignment of hsp70 sequences (97), 700 bp fragments of *E. histolytica* hsp70 genes were amplified by PCR from total genomic DNA and cloned (126). Partial sequence analysis identified a mitochondrial-type hsp70 isoform as well as cytoplasmic and endoplasmic reticulum hsp70 isoforms. The plasmid DNA was then used to amplify a probe for the mt-hsp70 gene using the degenerate primers Deg 1 and Deg 2 (Appendix A, Table A2).

3.2 SCREENING OF LIBRARIES

3.2.1 cDNA library

A λZap II cDNA library was screened by hybridisation with the mt-hsp70 specific PCR fragment of *E. histolytica*. Positive plaques were isolated and insert DNA was sequenced by primer walking.

Only two cDNA plaques out of 60,000 screened were positive after the screening of the cDNA library with the partial mt-hsp70 gene of *E. histolytica*. Insert DNA was excised in the pBluescript vector, the plasmid DNA was isolated and then restriction digested with enzymes to find the size of the insert. An insert of approximately 1800 bp was present in both cDNA clones but the coding region was incomplete at the 5′ end. The cDNA sequence was deposited in GenBank™ with the accession number AF214550.

3.2.2 Genomic library

A λZap II genomic DNA library was screened by hybridisation with the mt-hsp70 specific PCR fragment. Positive plaques were isolated and insert DNA was sequenced by primer walking.

After two rounds of screening, six plaques were finally purified of which only three contained the 5′ end of the gene. The insert in each was approximately 1.3 kb, of which ca. 1 kb was upstream of the open reading frame. No significant open reading frames were identified in this upstream region and a database search did not find any close matches. The genomic DNA sequence was deposited in GenBank™ with the accession number AF214549.

A composite genomic DNA-cDNA sequence revealed an open reading frame of 1794 bp encoding a 597 amino acid mitochondrial-type hsp70 of *E. histolytica* with a theoretical isoelectric point (pI) of 9.24 (http://www.expasy.org/tools/pi_tool.html; (127)). The mt-hsp70 proteins also seem to have a hydrophilic carboxyl terminus, which agrees with the mitochondrial proteins of *Trypanosoma cruzi* and yeast (128). The diagram below shows a schematic representation of the mt-hsp70 gene that was sequenced (Figure 3.1).

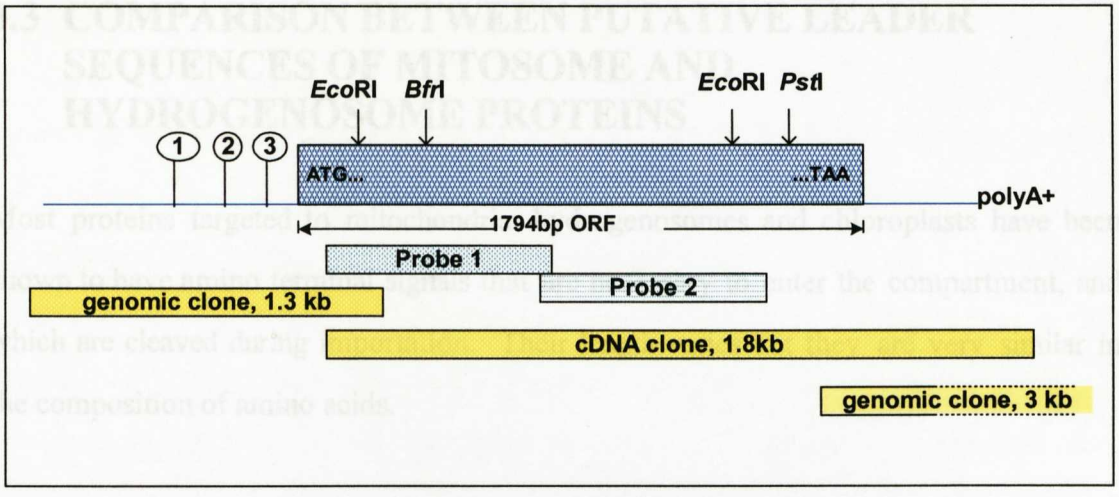


Figure 3.1. Schematic representation of the mt-hsp70 gene.

Probe 1 (position 37-701) was obtained with PCR amplification using the degenerate primers 1 and 2 on total genomic *E. histolytica* DNA, and probe 2 (positions 678-1460) by using the cDNA clone with primers 70.9 and 70.4. The circle numerals 1, 2 and 3 are representing the putative TATA box, the GAAC box and the initiator element respectively (section 3.4). The *Eco*RI enzyme cuts at bases 95 and 1367 of the mt-hsp70 gene. The *Bfr*I and *Pst*I enzymes cut at bases 364 and 1553 respectively.

The genomic DNA library was screened one more time with a 3′ end mt-hsp70 DNA probe to obtain the region 3′ to the end of the mt-hsp70 cDNA. After one round of screening, only one positive plaque was isolated out of the 50,000 plaques. After the second round of screening, eight plaques were isolated of which four had the 3′ end of the gene. The size of the insert was approximately 3 kb, of which ca. 2.8 kb was downstream of the open reading frame. The plasmid DNAs were partially sequenced (about 600 bp) from both directions to find the origin of the inserts. No significant open reading frames were identified in the sequenced part of the downstream region and a database search using the BLAST 2.0 programme of the NCBI databases did not find any close matches (116).

3.3 COMPARISON BETWEEN PUTATIVE LEADER SEQUENCES OF MITOSOME AND HYDROGENOSOME PROTEINS

Most proteins targeted to mitochondria, hydrogenosomes and chloroplasts have been shown to have amino terminal signals that are necessary to enter the compartment, and which are cleaved during importation. Their length varies but they are very similar in the composition of amino acids.

It seems that the amino terminus of the *E. histolytica* mt-hsp70 protein has greater similarity to certain *Trichomonas vaginalis* hydrogenosomal proteins than to the other mitochondrial-related proteins of *E. histolytica* identified so far (Figure 3.2). The amino terminal signal sequence of hydrogenosomal proteins is between 8-12 amino acids in length and its cleavage site is 2 amino acids to the carboxyl terminal side of an arginine (R) residue. For the cpn60 gene of *E. histolytica* it has been shown that the first 13 or 15 amino acids are necessary to get the protein into the mitosome (82, 83). Proteins of mitochondrial origin such as the pgp6, serine dehydratase and the transporter cpn10 have been identified in the *E. histolytica* genome project database and used for comparison with the cpn60 and PNT amino termini. The cpn60, the PNT and the pgp6 amino termini are very similar (3 serines near the amino terminus) but distinct from the amino terminus of the mt-hsp70. The mt-hsp70 shares only one serine with the PNT, the cpn60 and the pgp6 at position 4 (highlighted in blue in Figure 3.2), but shares both a serine (S) and an arginine with several *T. vaginalis* hydrogenosomal proteins. The mt-hsp70 amino terminus is also quite different from the cpn10, the transporter and the serine dehydratase amino termini. Serine dehydratase amino terminus has 3 threonines (T) which are very similar to the 3 serines of cpn60, PNT and pgp6 amino termini as both these two amino acids share quite similar properties. The cpn10, the pgp6 and the transporter proteins have multiple in frame methionines (M) in their amino termini.

This makes it quite difficult to be sure which one corresponds to the start of the coding region.

Studies have revealed that there might be a common evolutionary origin for protein importation into the mitochondrion, mitosome and hydrogenosome (82, 83, 129). Hydrogenosomal protein signal sequences targeted a reporter protein (dihydrofolate reductase) to the mitochondria of trypanosomes and yeast (80, 129), and a mitochondrial signal sequence from *Trypanosoma cruzi* targeted cpn60 to the *E. histolytica* mitosome (82). This suggests a homology of the importation apparatus and a common origin for the three organelles. The similarity of the mt-hsp70 amino terminus to those of hydrogenosomal proteins in conjunction with the phylogenetic analyses suggests it is also likely to be targeted to the mitosome.

If the mt-hsp70 gene is targeted in the mitosome in which the mitochondrial cpn60 homologue of *E. histolytica* is localised (82), this could add proof to the theory that there may be a common ancestor for the mitosome, the hydrogenosome and the mitochondrion.

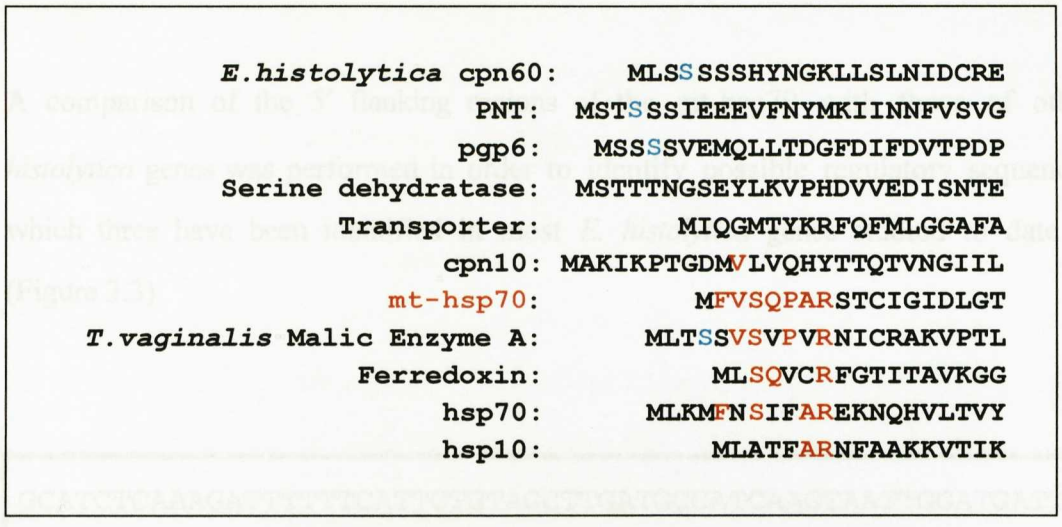


Figure 3.2. Alignment of the amino termini of the mitosome/hydrogenosome proteins.

The amino termini of the mt-hsp70 (this study), PNT (67), cpn60 (73), pgp6 ((130); GenBank™ Accession No. U01056), transporter (*E. histolytica* genome project; http://tigr.org/tdb/e2kl/ehal/), cpn10 (*E. histolytica* genome project) and serine dehydratase (*E. histolytica* genome project) of *E. histolytica*, and certain *Trichomonas vaginalis* hydrogenosomal proteins (131-133) are shown. Amino acid identities observed between the amino acid sequence of *E. histolytica* mt-hsp70 and the *T. vaginalis* hydrogenosomal proteins are given in red. The transporter, cpn10 and serine dehydratase proteins were identified in the *E. histolytica* genome project database by BLAST analysis using the *T. vaginalis* hydrogenosomal HMP31 (80), the cpn10 protein of *Bacillus halodurans* (134) and the *Saccharomyces cerevisiae* serine dehydratase (135) respectively. All positive hits were aligned and a consensus sequence was created.

3.4 PROMOTER ELEMENTS

A comparison of the 5′ flanking regions of the mt-hsp70 with those of other *E. histolytica* genes was performed in order to identify possible regulatory sequences, of which three have been identified in most *E. histolytica* genes studied to date (136) (Figure 3.3).

GCATCTCAAAGATTTTTTCATTCTGTAGCTTGATGCCATCAAGTAATTGGATGATTTCA
GGTGGTACTTCTCTCTTGCATTACTTTTCATTTTGATTTTAAGATTATCAAATCTTCTT
TTCAACT**CTATTTATCC**AAAAGTTCAACAA**AATGAAAG**TTATTTTTCTCTC**ACAG**
AATCTGAATTTTTAAC**ATG**

Figure 3.3. The region upstream of the mt-hsp70 gene open reading frame.

This part of the mt-hsp70 gene sequence was obtained after screening the genomic DNA library. The highlighted part represents the TATA box, the red part the conserved element 1 (CE1) and the bold part the conserved element 2 (CE2) also known as the 'GAAC' element. The start codon ATG of the mt-hsp70 gene is also given in bold.

The putative initiator element, conserved element 1 (CE1) is normally located between nucleotides –21 and –1 from the start codon and has the consensus sequence $A_{57\%}A_{46\%}A_{54\%}A_{46\%}A_{81\%}T_{48\%}T_{70\%}C_{97\%}A_{97\%}$, where the subscript indicates the percentage occurrence of each nucleotide. The putative CE1 in *E. histolytica* mt-hsp70 is located between nucleotides –20 and –12 from the start codon with sequence ACAGAATCT. The underlined bases correspond to those matching the consensus.

The second conserved element (CE2) is located upstream of CE1 between -30 and –14 from the start codon and its consensus sequence is $A_{61\%}A_{52\%}A_{61\%}G_{81\%}A_{97\%}A_{87\%}C_{58\%}T_{61\%}$. In the mt-hsp70 gene it may be located between nucleotides –41 and –34 from the start codon upstream of the CE1 with the sequence AATGAAAG, or alternatively it could be ACAGAATC, which overlaps with

CE1. This is also known as the 'GAAC' element. Studies have shown that this region acts in a position-independent manner due to its variable positioning in different core promoters (137). The role of the GAAC element may include stabilization of the pre-initiation complex via interactions with transcription factors or RNA polymerase II.

The third element (CE3) is located between bases –52 and –28 from the start codon, and has the consensus sequence $G_{32\%}T_{73\%}A_{76\%}T_{81\%}T_{92\%}T_{81\%}A_{92\%}A_{97\%}A_{89\%}G/C_{68\%}$. The closest match in the region of the mt-hsp70 gene is CTATTTATCC sequence between positions –64 and –55 from the start codon. This AT-rich region may function as an *E. histolytica* TATA sequence. Bruchhaus *et al* had found that this TATA box-like sequence is a target for protein-DNA interaction and it is commonly located approximately 30 nucleotides upstream of the transcription initiation site (138). Directed site mutagenesis of the putative TATA element decreased reporter gene expression by 48% (136), proving that this conserved element is a very important factor for gene transcription. The *E. histolytica* TATA box binding protein has recently been described (139).

The sequences of the putative promoter elements in the mt-hsp70 gene upstream region are not well conserved with those of other *E. histolytica* genes. Except for CE1, none are in the same relative positions as those found in the other *E. histolytica* genes. Whether this contributes to the low message abundance for the mt-hsp70 is not known.

Little is known about transcription regulation of protein-encoding genes in *E. histolytica*. Sequence comparison of the TATA and initiator elements of other eukaryotes gives little similarity, and comparison with promoter region of other protozoa shows significant divergence (137).

The characterisation of the conserved promoter elements is quite important for understanding the transcription and gene regulation in *E. histolytica*. Studies of promoter

structure and function with characterisation of the *hgl2* and *hgl5* gene promoters identified the three elements as regulators for full promoter activity *in vivo* (136, 137, 140). Mutation experiments of all three elements showed that they have an important role in determining the site of transcription initiation. Mutation of the TATA or the GAAC elements allowed other sites, normally not used, to become functional transcription start sites, indicating that they have a role in the transcription of *E. histolytica* genes. Mutations in the initiator element on the other hand resulted in no change in the site of transcription initiation (137). Mutation experiments need to be carried out in the mt-hsp70 putative promoter elements to verify if they are actual promoter sequences and their possible role in the transcription of this *E. histolytica* gene.

Purdy *et al* had also identified 5 AT-rich upstream regulatory elements (URE 1-5) in the *hgl5* gene promoter region that appear to be different from other protozoan and metazoan promoters (136). Mutations on the URE1, URE2, URE4 and URE5 elements decreased reporter gene expression while the mutation in URE3 increased expression. Recently *E. histolytica* sequence specific binding proteins have been identified that bind to the URE3 element (TATTCTATT) and the URE4 (AAAAATGAA) DNA motifs. The URE3-BP protein has specific URE3 binding ability (141) and the EhEBP1 and EhEBP2 have specific URE4 binding ability (142). URE3 sequence seems to be important for the transcription regulation of at least two *E. histolytica* genes: the *hgl5* and the *fdx* genes, while the URE4 seems to be important for *hgl5* regulation. Further characterisation of upstream regulatory elements and their cognate binding proteins will broaden the understanding in the transcriptional regulation in *E. histolytica*. In the upstream region of the mt-hsp70 gene there does not seem to be sequence similarity to either element URE3 or URE4.

A polyadenylation signal has also been described with the conserved sequence TA(A/T)TT and occurs 5-14 bases upstream from the mRNA poly A^+ tail (138). In the *E. histolytica* mt-hsp70 gene the sequence TAATT occurs approximately 20 bases

upstream the polyadenylation site. This putative polyadenylation signal overlaps the translation termination codon, something that also occurs in the cpn60 gene of *E. histolytica* (82).

3.5 UPSTREAM OF THE CODING REGION

The region upstream of open reading frame of the mt-hsp70 gene was sequenced from clones obtained after screening the genomic library (section 3.2.3). The sequence obtained was approximately 1.0 kb. No significant open reading frames were identified and BLAST searches did not find any close match.

Interestingly, this region is not particularly high in A+T composition (66.5%), but is very close to the 67.5% composition of the coding region of the mt-hsp70 gene. In general the *E. histolytica* genome is A-T rich, with the 67% within the coding regions and 78% overall (145, 146), indicating that most non-coding regions are very A+T rich.

The 'sense strand' of the genomic library untranslated region clone was 40.02% T and 26.48% A, in contrast to the coding region that was the exact opposite: 42.59% A and 24.92% T. This can be seen in the base composition graph (Figure 3.4). The significance of these results is not known.

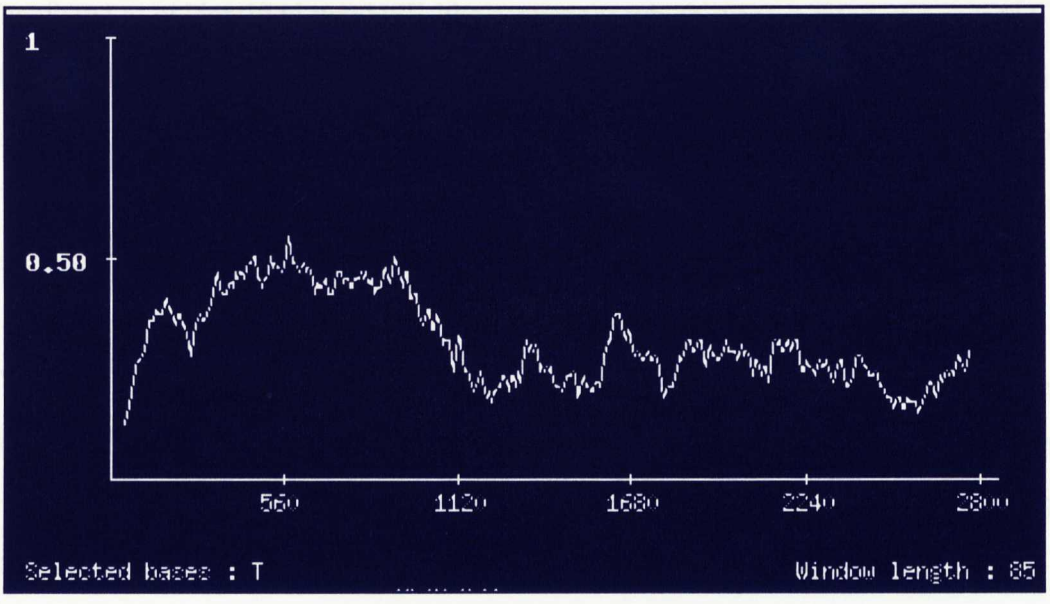


Figure 3.4. Graph showing the T-rich site of the upstream region of the mt-hsp70 of *E. histolytica* open reading frame.

The first area up to approximately 1110 bp represents the T-rich upstream region of the mt-hsp70 gene. The rest represents the coding region of the gene. The graph was obtained at http://abil.univ-mrs.fr/d_abim/riche-adn.html.

3.6 HEAT SHOCK ELEMENTS

Studies have shown that the expression of heat shock proteins is also regulated via heat shock elements (HSE), which are found in promoter regions of genes. Comparison of the HSE found in promoters of various heat shock protein genes of *Saccharomyces cerevisiae* with those of the mt-hsp70 gene of *E. histolytica* identified some possible HSE in the 5' flanking regions of the mt-hsp70 gene. Tachibana *et al* (143) have shown that the expression of *S. cerevisiae* MDJ1 is regulated via a novel non-consensus HSE which consists of at least three separate pentameric "nGAAn" motifs. "nTTCn (11 bp)-nGAAn-(5 bp)-nGAAn", where n is any nucleotide is the minimum consensus HSE. The HSE can still function with a 5-bp insertion between two repeating units if the spacing and orientation of the pentameric elements are maintained (144).

Tachibana *et al* identified two HSE-like sequences which were located in positions –235 and –212 of the hsp82 and YNL099W genes of the *S. cerevisiae* respectively: 'tTTCtaGAAcgccgtgGAAg' and 'gTTCccacaagtatgtGAAatcgtttGAAc'. These are very similar with the sequence 'tTTCtctcacaGAAtctGAAt' of the mt-hsp70 gene of *E. histolytica*. YNL099W is an unidentified open reading frame.

The HSE(-like) sequence 'gGAAttTTCcaGAAa' found at position –168 of the hsp60 gene of *S. cerevisiae* from the start codon , is very similar to the 'cGAAtaTTCtcTCAt' of the mt-hsp70 gene of *E. histolytica* which is located between nucleotides –827 and –813. The underlined bases correspond to those matching the consensus.

The HSE(-like) sequence "aGAAatTTCtaGAAtcTTGa" found at position –164 of the hsp10 gene of *S. cerevisiae* from the start codon is similar to the 'gGAAcatctTTCttGAAgat' and the 'aTTCcattGAAcTTGTTCt' sequences found on the mt-hsp70 gene of *E. histolytica* which are located between nucleotides –586 and –568, and –212 and –194 respectively.

The HSE(-like) sequences found in the mt-hsp70 gene upstream region are quite similar to those found in the *S. cerevisiae* heat shock proteins. A more extensive search should be done to see if these HSE are also present in the other mitochondrial genes of *E. histolytica* and also examine their role in the regulation of expression of these genes.

3.7 SOUTHERN BLOT ANALYSIS

Southern blotting suggests that there are two allelic forms of the mt-hsp70 gene in *E. histolytica.*

Table 3.1. Results of the hybridisation of the mt-hsp70 ORF probe with the Southern blot filter.

Restriction endonuclease	Restriction sites in the mt-hsp70 gene	Restriction site in the probe	Bands in the filter	Size of bands in kb
*Eco*RI	2	1	1	1.3
*Hind*III	-	-	1	>20
*Dde*I	-	-	2	6 and 3.5
*Taq*I	2	-	1	1.5
*Bfr*I	1	-	1	>20
*Bsp*LU11I	1	-	3	4 and 2.8

Southern blot hybridisation with the mt-hsp70 gene of *E. histolytica* gave two bands with some enzymes that do not cut within the coding region and one with others (Table 3.1, Figure 3.5). Genomic DNA digested with either *Bsp*LU11I or *Dde*I gave two bands although neither enzyme had a site in the probe (the third band in the *Bsp*LU11I was considered to be a result of partial indigestion). This result would normally indicate that there are two loci for the mt-hsp70 gene. However, hybridisation of the same filter with the upstream region probe gave the same two bands as with the mt-hsp70 coding region probe with *Dde*I (*Bsp*LU11I has a site adjacent to the start codon). Unless the upstream sequence is conserved between the two loci, this suggests that the two bands may correspond instead to two alleles of the mt-hsp70 gene that differ slightly (by a few bases) in the flanking region that happen to alter the restriction sites for these two enzymes. The genome of *E. histolytica* is thought to be at least tetraploid (147) and

alleles of certain genes differ in sequence (148). The situation in the case of the mt-hsp70 can not be fully resolved from these results. No sequence differences between PCR products and cDNAs were found.

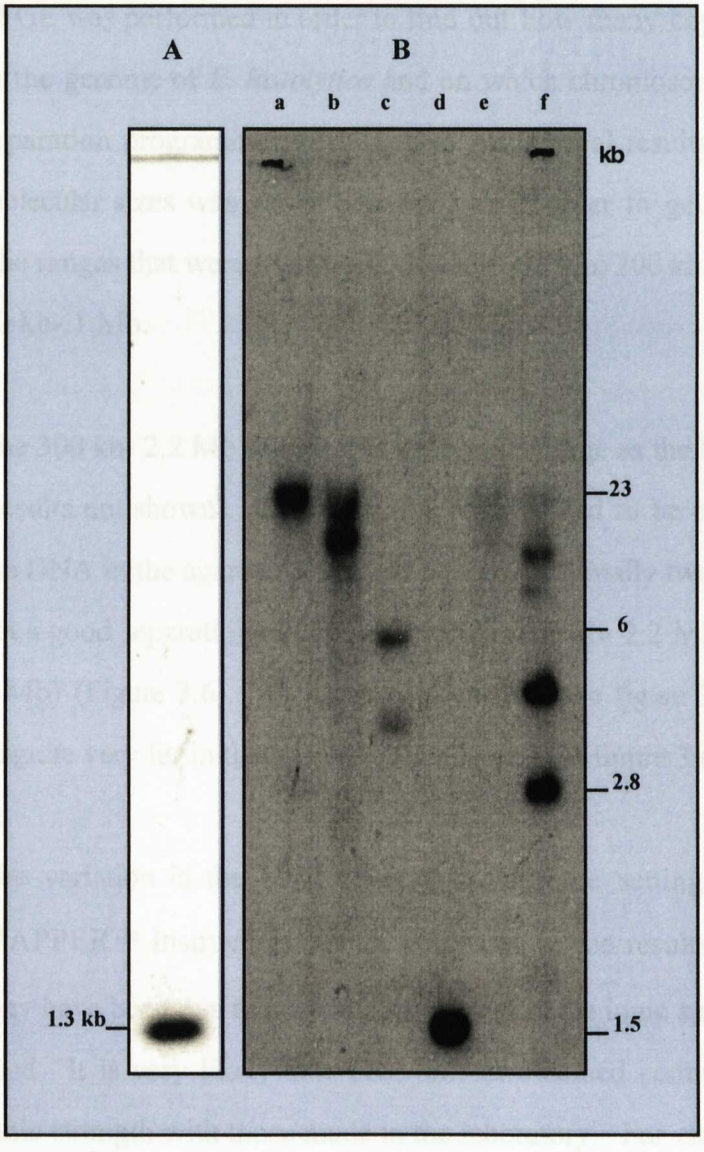


Figure 3.5. Southern blot analysis.

Genomic DNA was extracted from *E. histolytica* trophozoites, digested with restriction enzymes, run by electrophoresis into a 0.8% agarose gel and blotted into a nylon filter. The filter was hybridised with a specific mt-hsp70 probe (probe 1). The size marker used was λ DNA/*Hind*III.
A. *Eco*RI digested genomic DNA of *E. histolytica*.
B. Undigested DNA (lane a); partially undigested *Hind*III –cut DNA (lane b); *Dde*I –cut DNA (lane c); *Taq*I -cut DNA (land d); *Bfr*I –cut DNA (lane e); *Bsp*LU11I –cut DNA (lane f).

3.8 PULSED FIELD GEL ELECTROPHORESIS

Pulsed field gel electrophoresis (PFGE) was developed to separate large pieces of DNA by periodically re-orientating the current direction so that large pieces of DNA cannot

remain orientated "end-on" and separation by size is possible. PFGE has become a suitable method for chromosome separation and karyotype analysis in protozoan organisms due to the relatively small size of their chromosomes (149).

PFGE was performed in order to find out how many copies of the mt-hsp70 gene exist in the genome of *E. histolytica* and on which chromosomes they are located. Different separation programmes were used to get optimal results. The setting for the expected molecular sizes was varied between gels in order to get good separation of the bands. The ranges that were tried were: 300 kb- 2.2 Mb; 200 kb- 2.2 Mb; 100 kb- 2.2 Mb; and 50 kb- 1 Mb.

The 300 kb- 2.2 Mb setting was not a good range as the DNA in the gel migrated too far (results not shown). Therefore, the settings had to be changed so that the migration of the DNA in the agarose gel would be slower. Finally two different settings were used to get a good separation of the upper bands (100 kb- 2.2 Mb) and the lower bands (50 kb- 1 Mb) (Figure 3.6). As it can be seen from the figure 3.6 A, the lower bands did not migrate very far in the agarose gel in contrast to figure 3.6 B & C where they did.

The variation in the band separation with the settings recommended by the CHEF MAPPER™ instruction manual compared to the results obtained in these experiments may have been due to possible differences in the ionic strength between the TBE buffers used. It is very likely that TBE buffers obtained commercially have slightly different ionic strength with those made in the laboratory. For example the EDTA concentration we use is different. For this reason the concentration of the TBE buffer was lowered from 0.5x to 0.4x and finally to 0.25x to see if better band separation could be obtained and if migration of the bands would be slower.

Southern blot hybridisation with the mt-hsp70 gene of *E. histolytica* gave four bands at 2000 kb, 1300 kb, 450 kb and 310 kb (Figure 3.7). Comparison of these results with the

14 linkage groups found in *Entamoeba histolytica* (147) identified two linkage groups which the mt-hsp70 bands might be correlated to. The higher bands at 2000 kb and 1300 kb (Fig, 4.3) are closest to linkage group 12 where three bands are present at 1900 kb, 1850 kb and 1300 kb. The lower bands hybridised to the mt-hsp70 gene were 450 kb and 310 kb (Figure 3.7 B & C). Those two bands are closest to the ones in linkage group 1. This linkage group has 4 bands at 550 kb, 470 kb, 440 kb and 430 kb approximately. It is possible that the 2000 kb and the 450 kb bands are actually doublets but the resolution in the gel was not sharp enough to show good separation of the bands. It is also likely that the size estimates were not completely accurate which could explain why it was not easy to immediately identify the linkage groups. Another possibility is that the faint bands in the X-ray films are a result of cross-hybridisation of the mt-hsp70 probe with the cytoplasmic or the endoplasmic reticulum hsp70 genes of *E. histolytica.* The only way to find out if that is the case is to hybridise the blots with specific probes for the other hsp70 genes. This was not done in this study due to lack of time.

The finding of two different linkage groups for the mt-hsp70 gene of *E. histolytica* indicates that there are two copies of the mt-hsp70 gene each of them in a different locus.

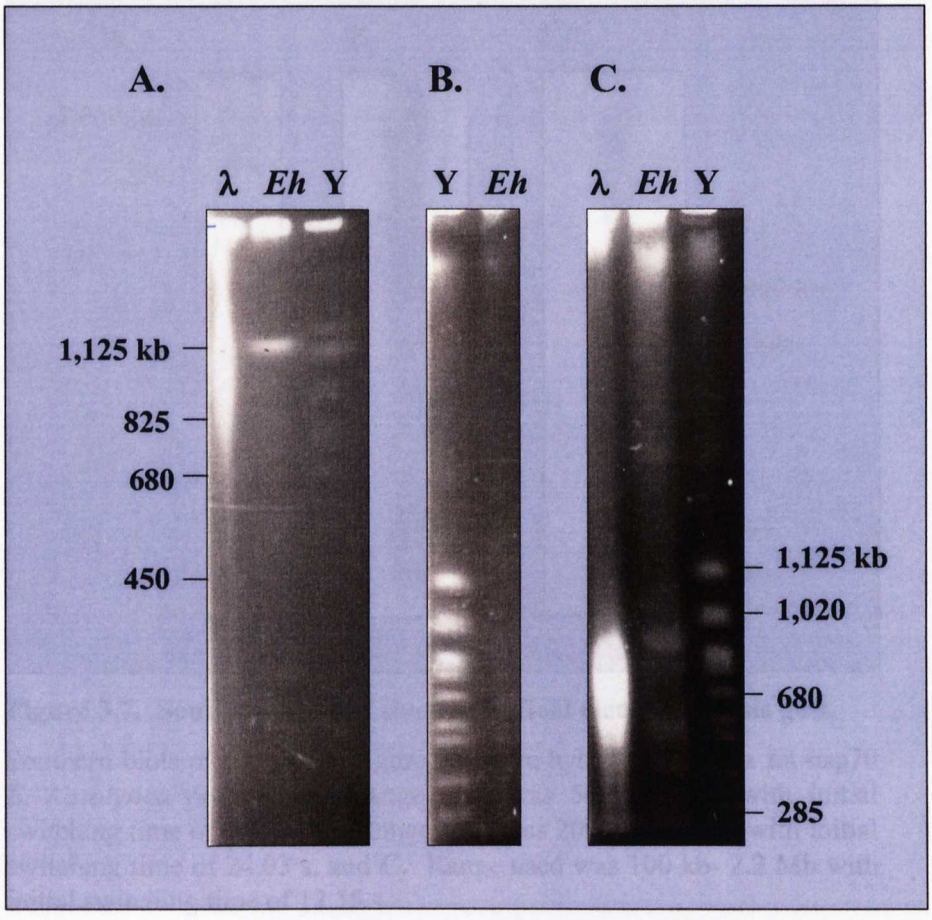


Figure 3.6. Separation of *E. histolytica* DNA by pulsed field gel electrophoresis.

E. histolytica chromosomal DNA was prepared in agarose plugs as described in section 2.12.1. DNA was electrophoresed in 0.25x TBE, at 14 °C at 6 Volts cm^{-1} with 120° angle and a variable linear ramping factor for switching times. For size markers, lambda DNA ladder and a yeast chromosomal marker were used. The gel was stained with ethidium bromide and visualised under the UV light. **A.** Range used was 50 kb-1 Mb with initial switching time of 6.75 s. **B.** Range used was 100 kb- 2.2 Mb with initial switching time of 12.55 s, and **C.** Range used was. 200 kb- 2.2 Mb with initial switching time of 24.03 s.

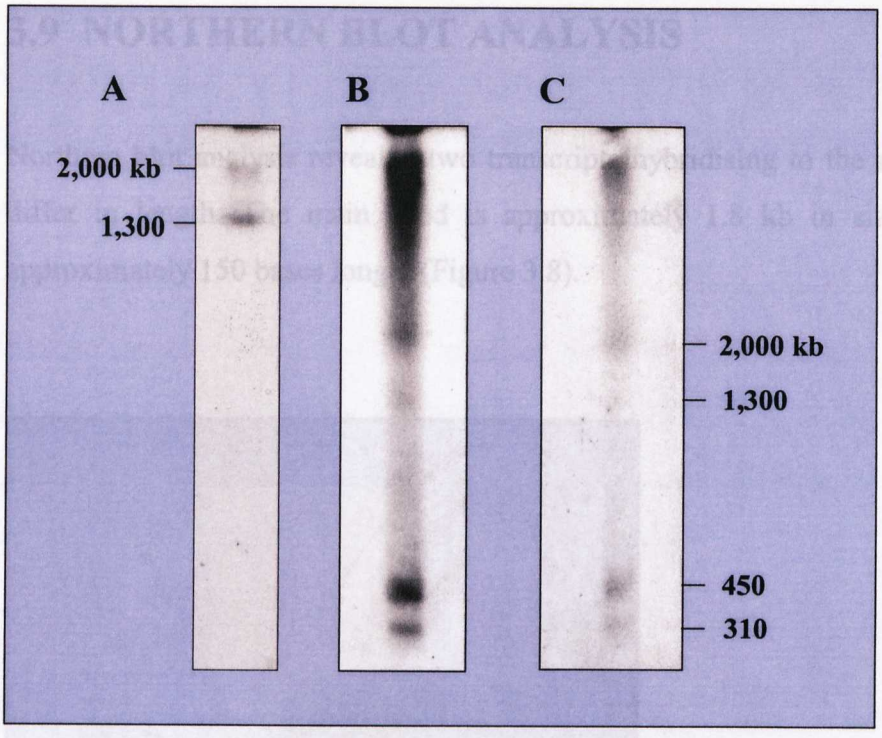


Figure 3.7. Southern blots of the pulsed field electrophoresis gels.

Southern blots of the gels in figure 3.6 were hybridised with a mt-hsp70 *E. histolytica* probe. **A.** Range used was 50 kb-1 Mb with initial switching time of 6.75 s. **B.** Range used was 200 kb- 2.2 Mb with initial switching time of 24.03 s, and **C.** Range used was 100 kb- 2.2 Mb with initial switching time of 12.55 s.

3.9 NORTHERN BLOT ANALYSIS

Northern blot analysis revealed two transcripts hybridising to the mt-hsp70 probe that differ in length. The main band is approximately 1.8 kb in size and the other is approximately 150 bases longer (Figure 3.8).

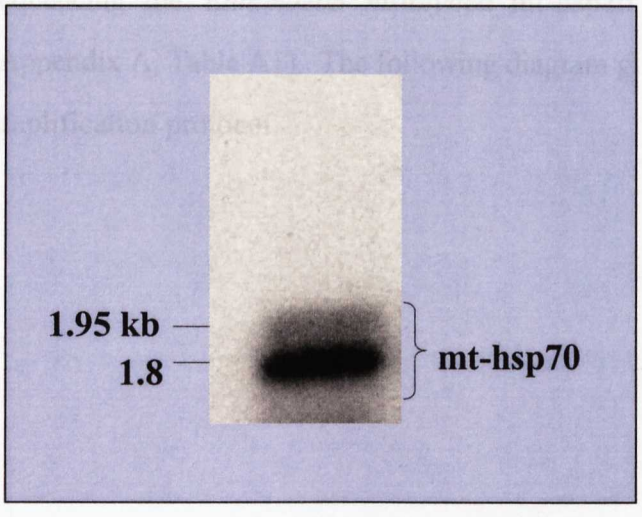


Figure 3.8. Northern blot of RNA extracted from non-heat shocked *E. histolytica* cells.

Poly A^+ RNA extracted from non-heat shocked trophozoites in the logarithmic phase of growth, separated in an agarose gel, blotted and hybridised to a mt-hsp70 coding region probe. The size of the mRNA was measured by comparison to an RNA marker (0.28-6.58 kb; Promega).

3.10 RAPID AMPLIFICATION OF cDNA ENDS

In order to explain the length difference between the mt-hsp70 mRNAs after hybridisation with a mt-specific hsp70 probe of *E. histolytica*, RACE techniques were employed to localise the length polymorphism to the 5′ or 3′ end of the messages.

The upstream primers that were used for the 3′ RACE reaction were 70.2, 70.5, 70.9, 70.18, 70.21 and 70.22 (Appendix A, Table A1). The downstream primer was the JTR1 oligonucleotide (5′-GAC TCG AGT CGA CAT CGA TTT TTT TTT TTT TTT TT-3′).

The gene specific primers (GSP) used for the 5′ RACE were also ones designed for sequencing the *Entamoeba histolytica* mt-hsp70 gene, i.e. 70.14, 70.19 and 70.20 (Appendix A, Table A1). The following diagram gives a schematic representation of the amplification protocol.

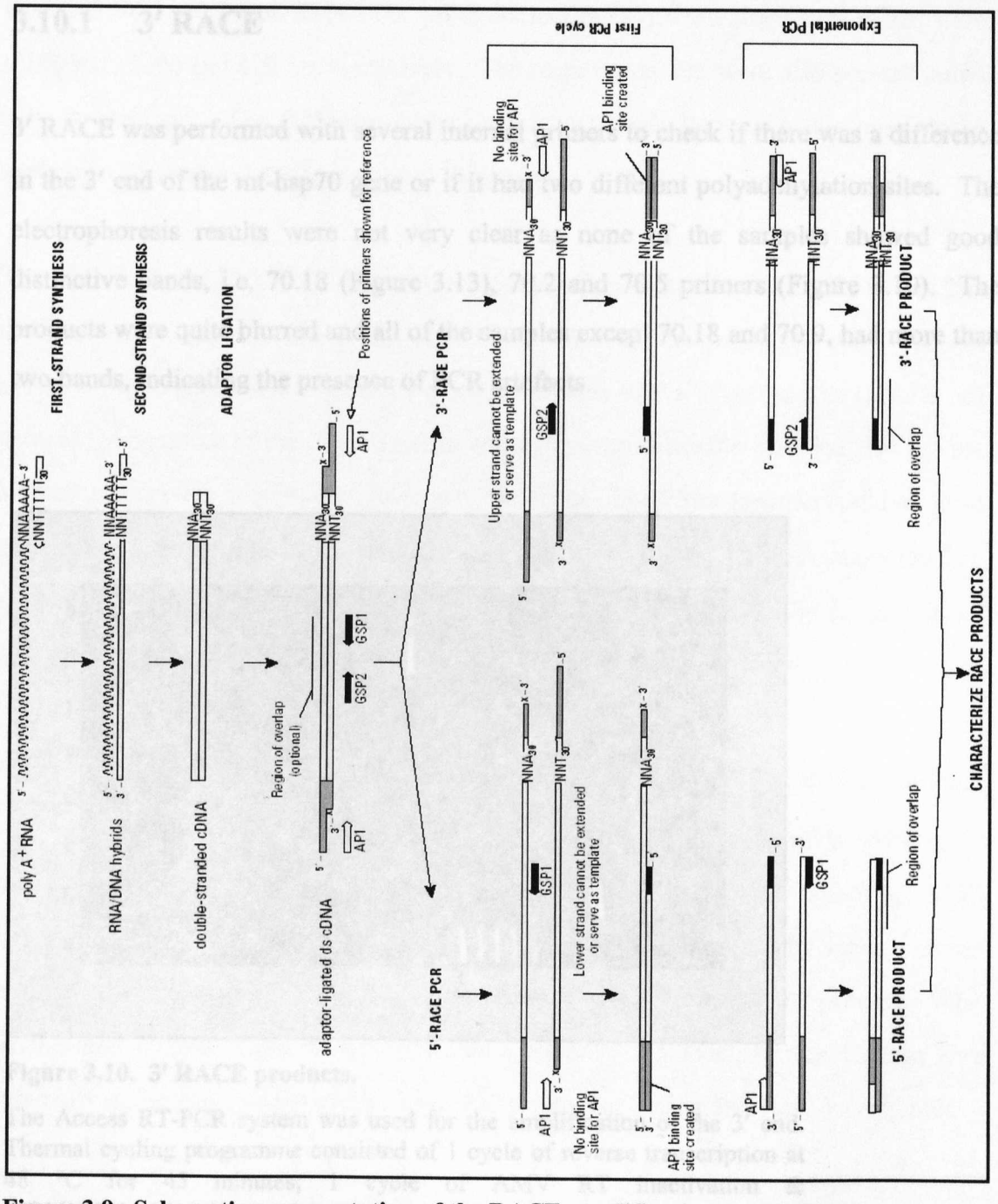


Figure 3.9. Schematic representation of the RACE amplification protocol.

Poly A^+ RNA is used for the 1st strand cDNA synthesis, which is followed by the second strand synthesis. The double stranded cDNA is then ligated to the Marathon cDNA adaptor and used for the 5′ and 3′ RACE reactions.

3.10.1 3′ RACE

3′ RACE was performed with several internal primers to check if there was a difference in the 3′ end of the mt-hsp70 gene or if it had two different polyadenylation sites. The electrophoresis results were not very clear as none of the samples showed good distinctive bands, i.e. 70.18 (Figure 3.13), 70.2 and 70.5 primers (Figure 3.10). The products were quite blurred and all of the samples except 70.18 and 70.9, had more than two bands, indicating the presence of PCR artefacts.

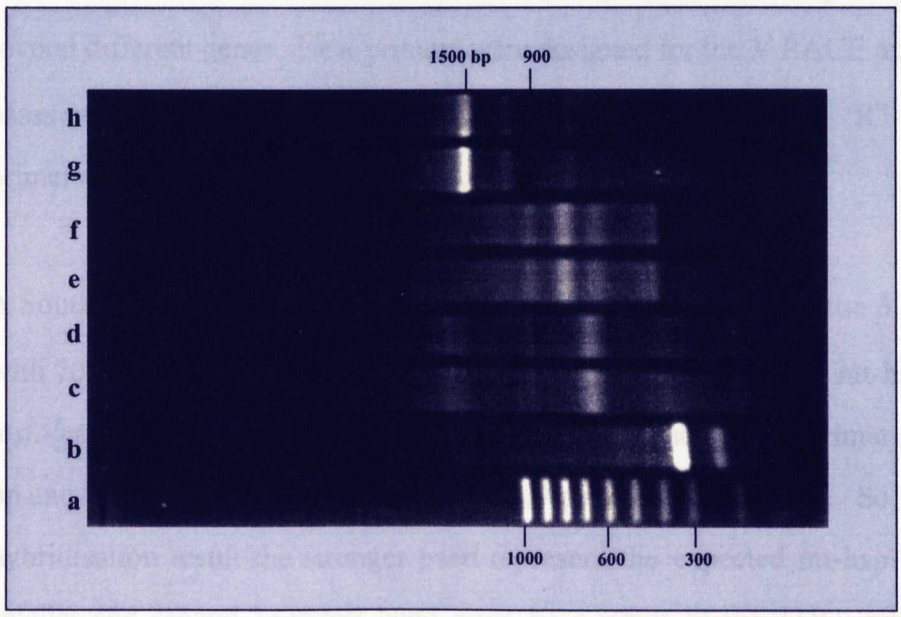


Figure 3.10. 3′ RACE products.

The Access RT-PCR system was used for the amplification of the 3′ end. Thermal cycling programme consisted of 1 cycle of reverse transcription at 48 °C for 45 minutes; 1 cycle of AMV RT inactivation & RNA/cDNA/primer extension at 94 °C for 2 minutes; 40 cycles of denaturation at 94 °C for 30 seconds, annealing at 55 °C for 1 minute, extension at 68 °C for 2 minutes; 1 cycle of final extension at 68 °C for 7 minutes and cooling to 4 °C. The PCR reactions were purified and 10 µl of each was analysed on 1% agarose gel: 100 bp DNA ladder (lane a); positive control with 326 and 200 bp products (lane b); 70.2 primer PCR product (lanes c and d); 70.5 primer PCR product (lane e and f); and 70.9 primer PCR product (lanes g and h). Lanes c, e and g have the PCR product before the purification, and lanes d, f and h after the purification.

The 70.2, 70.5, 70.9 and 70.18 PCR products had the expected size bands which were 1500, 900, 1200 and 490 bp respectively. The other bands that were also present could only be artefacts. The 70.5 sample had too many bands that were not possible to explain so no further work was done at that point.

The 70.2, 70.9 and 70.18 PCR products were restriction digested with *Eco*RI to confirm their mt-hsp70 origin. *Eco*RI cuts at position 1367 of the mt-hsp70 gene, but the products of the restriction digestion did not give bands of the expected size (results not shown). Alignment of the mitochondrial and cytoplasmic hsp70s revealed that all the above primers were in conserved regions of the genes. Therefore products could be from several different genes. New primers were designed for the 3′ RACE approximately 250 bases away from the stop codon. These were 70.21 and 70.22. RT-PCR with these primers did not give any clear results.

A Southern blot was performed to find the origin of the bands of the 3′ RACE products with 70.2, 70.5 and 70.9 primers. The blot was hybridised to the mt-hsp70 probe of *E. histolytica* (Figure 3.11). The PCR product size for each of the primers is 900 bp, 1500 bp and 1200 bp for the 70.5, 70.2 and 70.9 primers respectively. So by looking at the hybridisation result the stronger band represent the expected mt-hsp70 product. The smaller and second brightest band must be a result of the JTR1 primer binding also internally in the gene, which gives a truncated product of 500 bp less than the original.

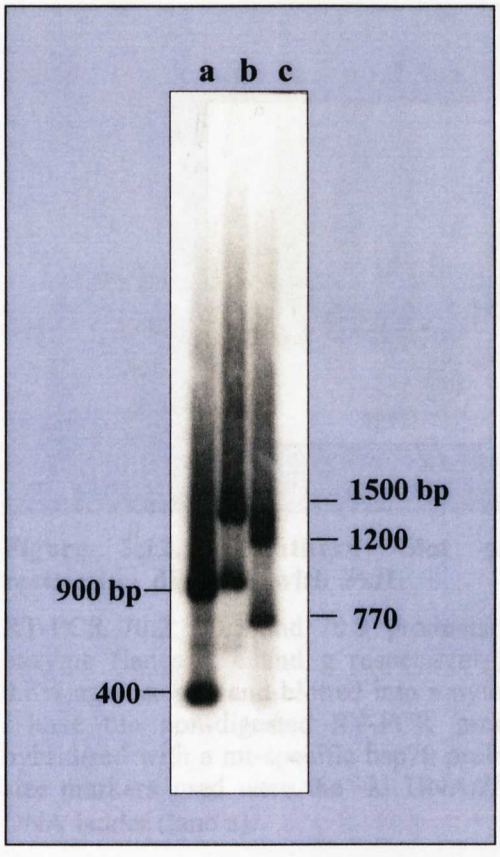


Figure 3.11. Southern blot of 3′ RACE products.

3′ RACE products were electrophoresed in a 0.8% agarose gel and blotted in a nylon filter. The filter was hybridised with a mt-specific hsp70 ORF probe of *E. histolytica*. Lane a: 70.5 product; Lane b: 70.2 product; Lane c: 70.9 product. The λ DNA/HindIII and 100 bp DNA ladders were used as size markers.

The PCR products were also digested with *Pst*I, which cuts near the 3′ end of the gene, to confirm the origin of the bands detected (Figure 3.12). Digested PCR products were approximately 300 bp less than the original. The *Pst*I site is 241 bp away from the poly A^+, which points out that the products are of mitochondrial origin. Only one band hybridised in each 3′ RACE reaction indicating that the mRNA length difference is not in the 3′ untranslated region of the mt-hsp70 transcript.

This led to test the possibility that the mt-hsp70 gene had two different initiation sites so 5′ RACE experiments were performed.

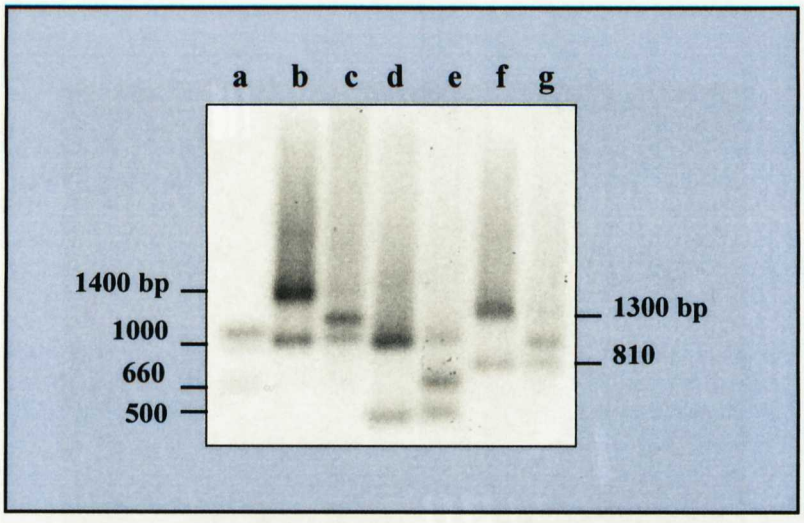


Figure 3.12. Southern blot of RT-PCR products restriction digested with *Pst*I.

RT-PCR 70.2, 70.5 and 70.9 products were digested with *Pst*I enzyme (lanes c, e and g respectively), electrophoresed in a 0.8% agarose gel, and blotted into a nylon filter. Lanes b, d and f have the non-digested RT-PCR products. The filter was hybridised with a mt-specific hsp70 probe of *E. histolytica*. The size markers used were the λ DNA/*Hind*III, and the 100 bp DNA ladder (lane a).

3.10.2 5′ RACE

5′ RACE was attempted but although mt-hsp70 specific products were obtained, none were full length and the possibility of 5′ end length variation could not be resolved unambiguously by this method.

Primer 70.14 was used for the amplification of the 5′ end of the mt-hsp70 gene. The expected size of the product was approximately 539 bp. Because the suggested PCR protocol did not give strong bands (94 °C for 1 minute; 25 cycles of denaturation at 94 °C for 30 seconds and annealing at 68 °C for 4 minutes), a modified version of the touchdown PCR programme was used to get good intensity, clear bands (Figure 3.13).

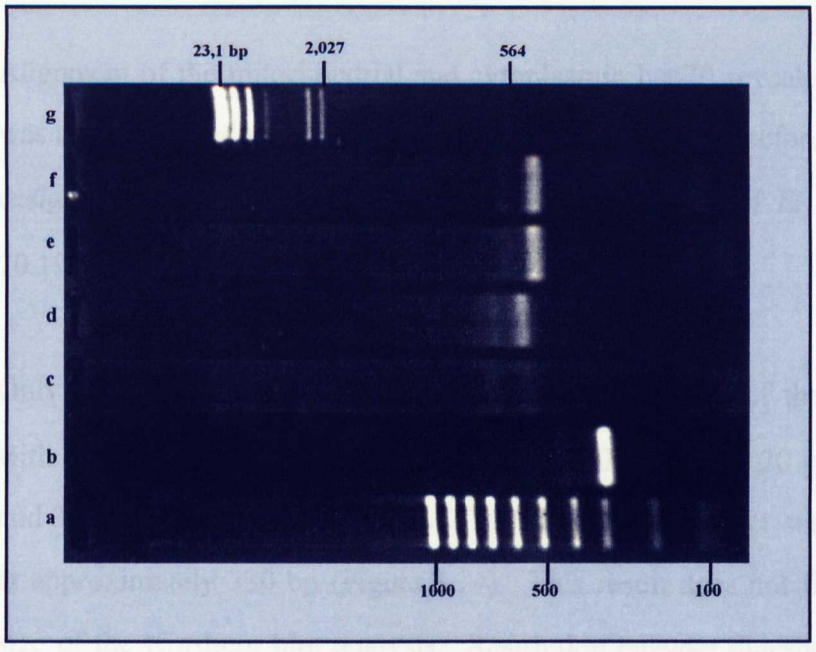


Figure 3.13. 5′ and 3′ RACE products.

A slight modification of the programme used in the manufacturer's protocol for the touchdown PCR was used for the RACE experiment. The annealing temperatures were reduced because the Tm of the GSP was between 60-65 °C. The thermal cycling programme consisted of 94 °C for 1 minute, 5 cycles of denaturation at 94 °C for 30 seconds, annealing at 66 °C for 4 minutes; 5 cycles at 94 °C for 30 seconds, 64 °C for 4 minutes; and 25 cycles at 94 °C for 30 seconds, 62 °C for 4 minutes. 10 µl of each sample was analysed on a 1% agarose gel: 100 bp DNA ladder (lane a); positive 300 bp control (lane b); 5′ RACE products with 70.14 primer (lanes c and d); 3′ RACE products with 70.18 primer (lanes e and f); λ DNA/*Hind*III ladder (lane g). Adaptor-ligated cDNA was used for the RACE either undiluted (lanes c & e) or 1:10 diluted (lanes d & f).

Two bands appeared to be present in the PCR product with approximately 100 bp difference. To find out if the product was a mt-hsp70 fragment, it was digested with *Eco*RI enzyme. Restriction digestion was not complete probably due to salt conditions that may have affected the enzyme, so the sample was cloned, and the plasmid DNA was extracted and sequenced. A BLAST® search of the sequence revealed that the cloned insert was a fragment of the cytoplasmic hsp70 gene of *Entamoeba histolytica* (150).

Alignment of the mitochondrial and cytoplasmic hsp70 revealed that the 70.14 primer was in a conserved region of the hsp70 gene family and therefore, new primers had to be designed to amplify the 5′ end of the mt-hsp70 mRNA of *E. histolytica*. These were 70.19 and 70.20 primers.

Only one band appeared in the 5′ RACE PCR products of the double stranded DNA with 70.19 and two with the 70.20 primer. 70.19 and 70.20 gave an expected 300 bp and 330 bp products respectively. The 70.20 PCR product seemed to have two bands at approximately 330 bp (Figure 3.14). This result does not fit with the difference in size of the Northern blot analysis. Restriction enzyme digestion of 5′ RACE product with *Eco*RI and *Taq*I gave the correct size bands indicating that the product was indeed of mt-hsp70 origin.

The 70.20 PCR product was cloned into a pGEM T-easy vector and plasmid DNA was isolated. Restriction enzyme digestion of the DNA samples with the *Eco*RI enzyme revealed the mt-hsp70 origin of the insert. The *Eco*RI enzyme cuts in position 94 of the mt-hsp70 gene. The PCR product was approximately 330 bp; so two fragments were present one at approximately 200 bp and the other at 100 bp. (Figure 3.15).

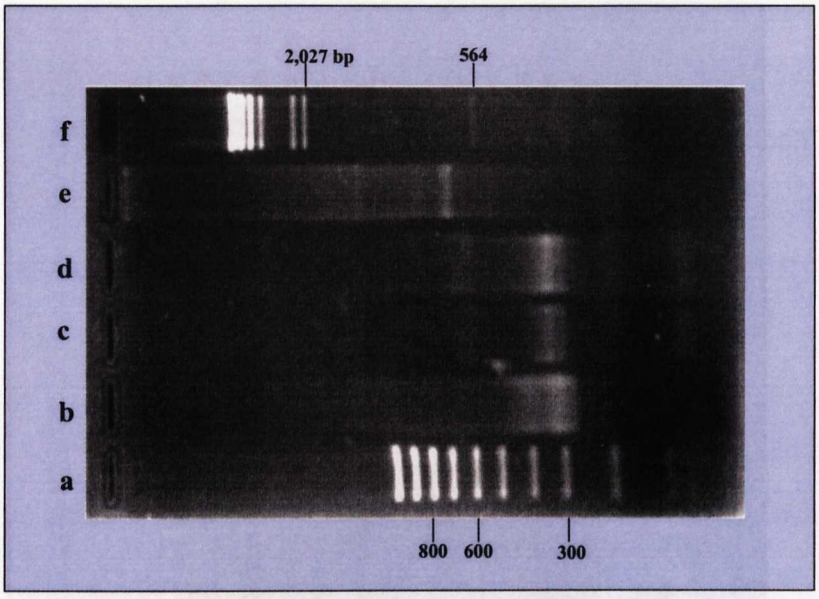


Figure 3.14. 5′ RACE products.

The thermal cycling programme for the reaction consisted of 94 °C for 1 minute; 5 cycles of denaturation at 94 °C for 30 seconds, annealing at 62 °C for 4 minutes; 5 cycles at 94 °C for 30 seconds, 60 °C for 4 minutes; and 25 cycles at 94 °C for 30 seconds, 58 °C for 4 minutes. 5 µl of each reaction was analysed in a 1.2% agarose gel: 100 bp DNA ladder (lane a); 70.19 primer PCR product (lane b); 70.20 PCR product (lanes c and d); 2.6 kb positive control (lane e); and λ DNA/ *Hind*III ladder (lane f). 1:10 diluted adaptor-ligated cDNA was used for lanes b & c, and 1:5 in lane d.

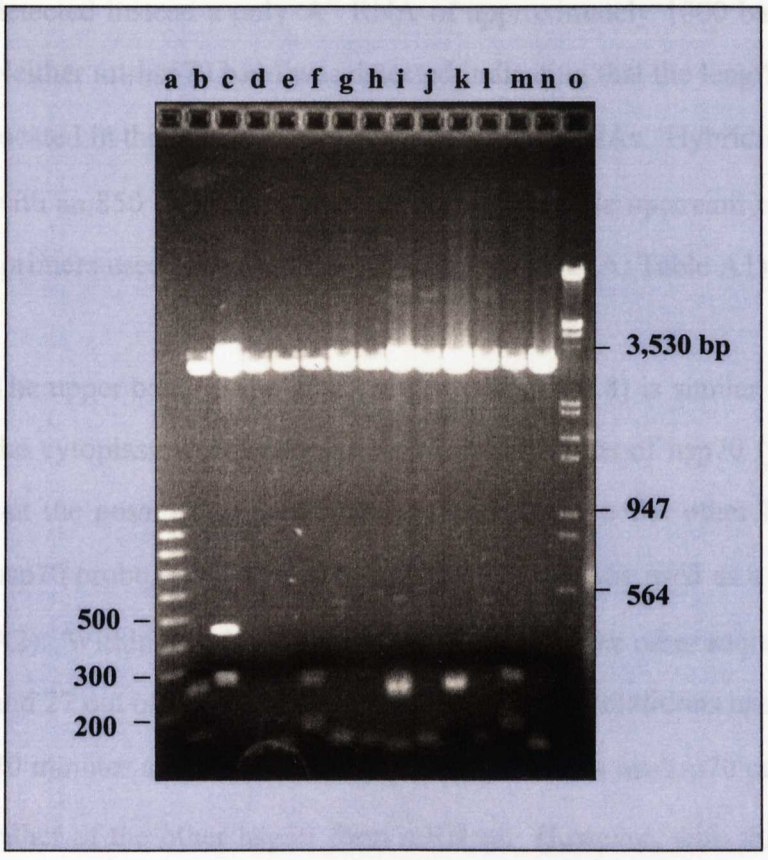


Figure 3.15. Restriction digestion of cloned 5′ RACE products with *Eco*RI.

Lanes i and k have the correct size fragments, one at 200 bp and the other at 100 bp approximately. Lane a has the 100 bp ladder and lane o the λ DNA/*Eco*RI-*Hind*III marker. The products were separated in a 1% agarose gel.

Sequencing with T7 and SP6 primers and then BLAST® search confirmed the mitochondrial origin of the hsp70 gene. The 5′end of the cDNA was not full length so the point of origin of the transcript could not be found.

A probe was therefore made covering the 300 bp region immediately upstream of the initiation codon (Appendix A, Table A2). As the 5′ untranslated regions of *E. histolytica* genes are usually very short it was expected the probe would hybridise only to the upper mt-hsp70 mRNA if the length difference was at the 5′ end of the message. However, Northern blot analysis of the mRNA samples using the upstream probe

detected instead a poly A^+ RNA of approximately 1000 bases in length (Figure 3.16). Neither mt-hsp70 band was detected indicating that the length variation is unlikely to be located in the 5′ untranslated region of the mRNAs. Hybridisation of the Northern blots with an 850 bp PCR product covering the whole upstream region gave the same results (primers used were 70.23 and 70.24; Appendix A, Table A1).

The upper band in the Northern blot (Figure 3.8) is similar in size to that expected for the cytoplasmic or endoplasmic reticulum forms of hsp70 (approx. 1950 bp). To rule out the possibly of weak cross-hybridisation to the other hsp70 mRNAs by the mt-hsp70 probe, the 3′ end of the mt-hsp70 gene was used as a probe (Appendix A, Table A2). Within this probe the closest matches to the other sequences were 16 out of 17 bp and 27 out of 36 bp. Under the stringent wash conditions used (0.2x SSC/ 0.1% SDS for 30 minutes at 65 °C) it is very unlikely that this mt-hsp70 probe will cross-hybridise to either of the other hsp70 form mRNAs. However, with this specific 3′ end probe the same two-band hybridisation pattern was obtained. If the probe is not cross-hybridising the only remaining possibility is variation in the length of the poly A^+ tract, which may affect message stability.

3.10.3 First strand cDNA synthesis

To completely eliminate the possibility of the length difference to be at the 5′ end of the gene 1st strand cDNA synthesis was done of the mt-hsp70 mRNA using the Marathon™ cDNA Amplification kit. The 1st strand cDNA sample and a *E. histolytica* mRNA sample were loaded in a formaldehyde-containing 1% agarose mini-gel and electrophoresed at 10 Volts cm^{-1}. RNA was transferred to a nylon membrane and then hybridised to a mt-hsp70 ORF probe of *E. histolytica* as described in section 2.9. After two weeks exposure in an X-ray film at -70 °C, only one band was present in the film,

but it could not be determined unambiguously to which sample it corresponded. The mRNA sample normally should have had two bands as seen before. If the cDNA sample had one band the difference would have been in the poly A^+ tail, if it had two bands the difference would have been at the 5′ end of the gene. Results were therefore inconclusive.

3.10.4 PCR amplification of the coding region

PCR amplification of the mt-hsp70 gene to get the DNA probes to use for hybridisation showed that no length variation in the regions amplified and no indication of an intron. Only one band was present in both of the products that were amplified. To rule out the possibility of the length difference being in the rest of the coding region a primer was designed starting at the ATG codon of the gene (70.25 primer) and used it with an internal ORF primer (70.4) for the amplification of the complete amino terminal region of the ORF on total genomic DNA. Only one band was present indicating that there is no length polymorphism in the ORF of the mt-hsp70 gene.

3.11 UPSTREAM REGION OF THE MT-HSP70 GENE

3.11.1 Northern blot analysis

Northern blot analysis of the mRNA samples using a probe immediately upstream of the initiation codon of the mt-hsp70 gene detected a poly A^+ RNA of approximately 1000 bases in length (Figure 3.16).

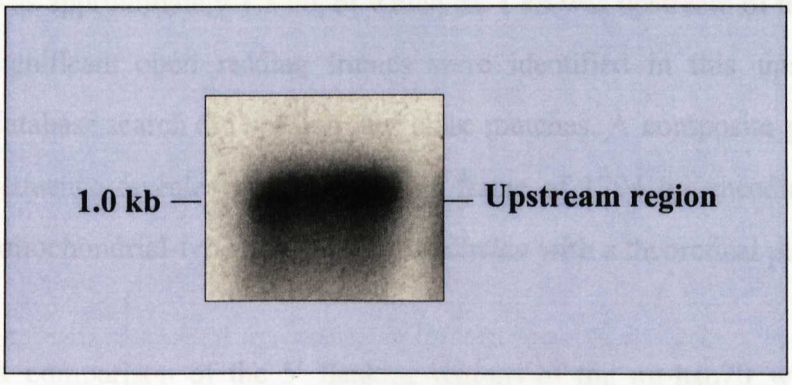


Figure 3.16. Northern blot of non-heated shocked *E. histolytica* trophozoites.

Poly A^+ RNA extracted from non-heat shocked trophozoites in the logarithmic phase of growth, separated in an agarose gel, blotted and hybridised to a probe immediately upstream of the initiation codon. The size of the mRNA was measured by comparison to an RNA marker (0.28-6.58 kb; Promega).

The upstream region thus seems to be transcribed despite the fact that no open reading frame was found in this region. Several classes of non-translated polyadenylated RNAs have been found in *E. histolytica* but those described to date have been much more abundant (151). It is possible that these transcripts function as RNA molecules. They could be regulatory RNAs having a role in transcription or translation or even having a structural role. The function of these molecules remains unknown and up to now in no other protozoa has the presence of such transcripts been reported (151).

3.12 SUMMARY & CONCLUSIONS

Two *E. histolytica* libraries were screened with a mt-hsp70 gene fragment: a cDNA library of strain SFL-3 and a genomic DNA library of strain HM-1: IMSS both in λZap II. Only two cDNA plaques out of 60,000 screened were positive indicating it is a very rare message. An insert of approximately 1800 bp was present in both cDNA clones but the coding region was incomplete at the 5′ end. Screening of the genomic library yielded 3 clones containing the 5′ end of the gene. The insert in each was approximately 1.3 kb, of which ca. 1 kb was upstream of the coding region. No significant open reading frames were identified in this upstream region and a database search did not find any close matches. A composite genomic DNA-cDNA sequence revealed an open reading frame of 1794 bp encoding a 597 amino acid mitochondrial-type hsp70 of *E. histolytica* with a theoretical pI of 9.24.

A comparison of the 5′ flanking regions of the mt-hsp70 with those of other *E. histolytica* genes to identify possible regulatory sequences did not reveal any close similarities. A putative polyadenylation signal occurs approximately 20 bases upstream of the polyadenylation site and overlaps the translation termination codon, something that also occurs in the cpn60 gene of *E. histolytica.*

Genomic library screening with a 3′ end mt-hsp70 DNA probe, gave an insert of 2.8 kb downstream of the open reading frame. No significant ORFs were identified and BLAST searches did not find any close database matches.

The identification of mitochondrion specific proteins such as cpn60 and PNT (67) and the mt-hsp70 reported here confirms that *E. histolytica* is not a surviving relic of pre-mitochondrial evolution, but that its ancestors once harboured mitochondria or an endosymbiont related to the progenitor of mitochondria.

Comparison between putative leader sequences of mitosome and hydrogenosome proteins revealed that there might be a homology of the importation apparatus and a common evolutionary origin for protein importation into the mitochondrion, mitosome and hydrogenosome. The amino terminus of the mt-hsp70 protein has greater similarity to certain *T. vaginalis* hydrogenosomal proteins than to the other mitochondrial-related proteins of *E. histolytica* identified so far. The similarity of the mt-hsp70 amino terminus to those of hydrogenosomal proteins in conjunction with the phylogenetic analyses suggests it is also likely to be targeted to the mitosome.

Southern blot analysis revealed that there are either two allelic forms or two copies of the mt-hsp70 gene in *E. histolytica.*

PFGE results also showed that there are two copies of the mt-hsp70 gene each of them in a different locus as two linkage groups were associated to the mt-hsp70 gene.

The mt-hsp70 gene appears to give rise to two mRNA transcripts differing in length by approximately 150 bases. In addition, a 1-kb apparently non-coding transcript was detected upstream of the mt-hsp70, despite the fact that no open reading frame was found in this region. cDNA library screening to find the start of this poly A^+ RNA was not successful.

Rapid amplification of cDNA ends was performed in order to explain the two bands of the Northern blot of the *E. histolytica* mRNA. The experiments were carried out to find out if there was any length polymorphism at either end of the mt-hsp70 mRNA.

Results obtained with the 5′ RACE did not show different initiation sites in the region upstream of the gene, as only one band was seen. The 5′ end that was finally obtained

by cloning was not full length and the point of origin of transcription could not be determined.

3′ RACE experiments did not give any clear results either. None of the primers used for the experiments gave distinctive bands. The products were blurred and had multiple bands. Further work confirmed the mt-hsp70 origin of certain bands detected and led to the conclusion that the mRNA length variation is not in the 3′ untranslated region of the mt-hsp70 transcript.

The possibility of weak cross-hybridisation of the cytoplasmic and endoplasmic reticulum forms to the mt-hsp70 probe was also ruled out by using as a probe the 3′ end of the mt-hsp70 gene. PCR amplification of the coding region also showed one band indicating that the polymorphism is not at the ORF either.

The problems encountered with the 5′ and 3′ RACE experiments could be because the message is quite rare and it is difficult for the primers to pick up the mt-hsp70 mRNA and amplify it. In some cases the primers amplified the cytoplasmic hsp70 mRNA which is more abundant (150). This was because hps70s are a highly conserved group of proteins and some of the primers that were used for both the 5′ and the 3′ RACE were in conserved regions of the mitochondrial and the cytoplasmic mt-hsp70 genes. New primers had to be designed from the non-conserved regions that gave the expected mt-hsp70 products.

The remaining explanation for the location of the variation is that the two transcripts differ in the length of the poly A^+ tract. Why this should be the case and how it would be regulated is unclear but it may have an effect on message stability.

CHAPTER 4

HEAT SHOCK AND NORTHERN BLOT ANALYSIS

4.1 INTRODUCTION

The expression of hsp70 and cpn60 proteins is traditionally thought to be induced by numerous cellular stresses, including elevated temperature, oxygen deprivation, and exposure to inhibitors of various metabolic processes (87-89). Heat shock studies on the *E. histolytica* cpn60 have given conflicting results. After incubating the trophozoites at 41 °C for various time periods Tovar *et al* showed that the gene is not activated by heat shock (82). In contrast to these results, Mai *et al* found that the cpn60 mRNA is induced after one hour of heat shock at 42 °C (83). So, it was decided to study the effect of heat shock expression of hsp70 and cpn60 and other mitochondrial protein genes not only in *E. histolytica* but also in other *Entamoeba* species.

4.2 PARTIAL DNA SEQUENCES USED AS PROBES FOR FILTER HYBRIDISATION

Partial cpn60, PNT, mt-hsp70, and β-actin gene sequences were amplified from *E. histolytica* by PCR (section 2.8.1) with gene specific primers and labelled with radioactive [α-^{32}P]-dCTP to be used as probes for detecting the mRNA in the Northern blot analyses (section 2.9). The sequences of the primers used can be found in

Appendix A, Tables A1 and A2. cDNA and/or genomic clones were used as templates for the PCR reactions where available.

A 700 bp mt-hsp70 fragment of *E. histolytica*, was amplified using specific DNA primers 70.4 and 70.9 on plasmid DNA obtained after screening the cDNA library (section 3.2.1).

Mt-hsp70s of *Trichomonas vaginalis*, *Pisum sativum*, *Eimeria tenella*, *Saccharomyces cerevisiae*, *Leishmania major* and *Entamoeba histolytica* and the cytoplasmic and endoplasmic reticulum hsp70 genes of *E. histolytica* were aligned (Figure 4.1) to design primers to amplify a 750 bp fragment of the mt-hsp70 gene from the other *Entamoeba* species. Degenerate primers D1.70 and D2.70 were designed based on amino acid residues conserved between the mt-hsp70s and not observed in the cytoplasmic and endoplasmic reticulum hsp70s of *E. histolytica*. The mt-hsp70 gene of *E. invadens* was the only one from the *Entamoeba* species that was not successfully amplified with these primers. For each of the *Entamoeba* species different amplification conditions were used to get optimal results (Figure 4.2 & 4.3).

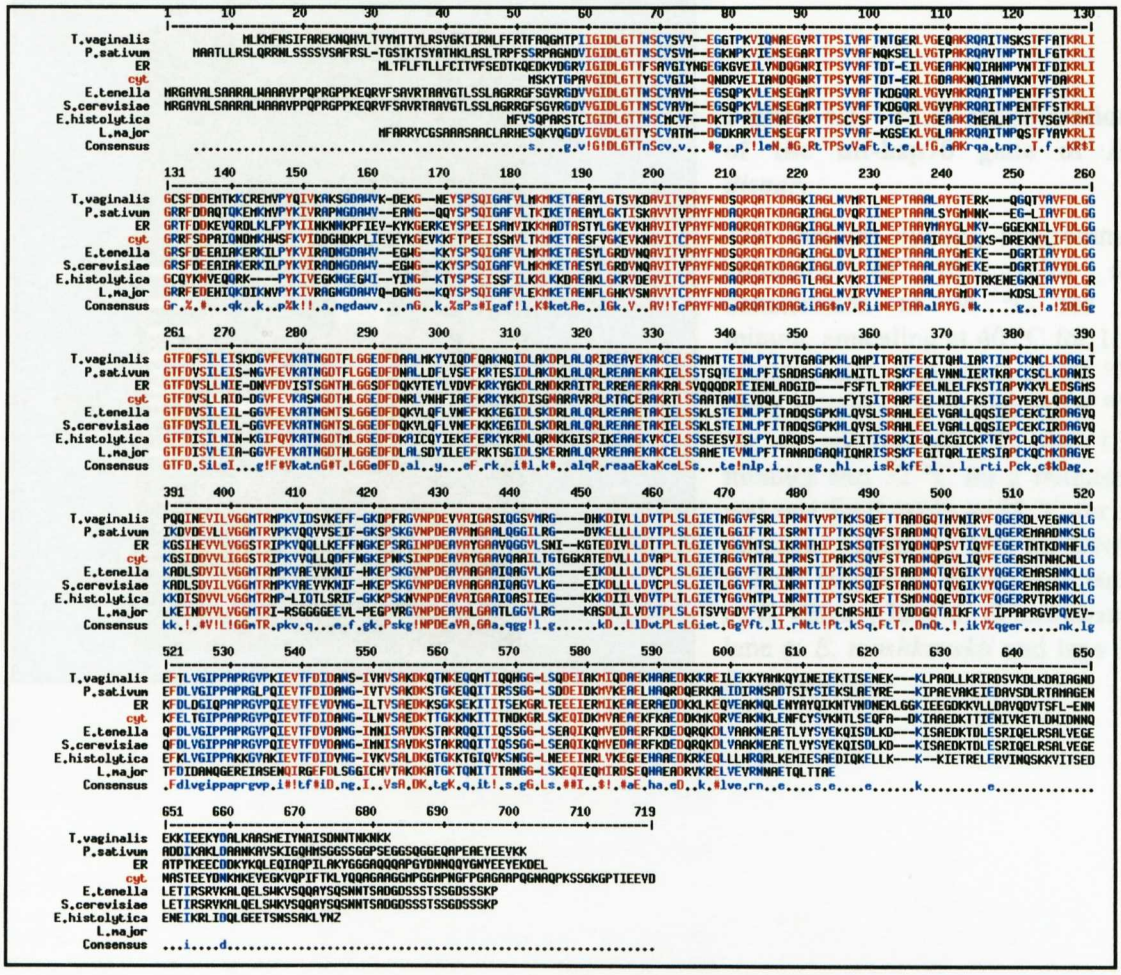
The PNT gene of *E. histolytica* was amplified by PCR using primers PNT-A and PNT-B on total genomic DNA giving a fragment of approximately 850 bp in size. The primers were designed based on the complete PNT gene sequence (67). In the rest of the *Entamoeba* species the PNT probe was amplified from a recombinant plasmid DNA that contained a ca. 450 bp PNT insert which had been previously amplified by PCR using degenerate primers PNT5 and PNT3 on total genomic DNA (152).

Degenerate primers cpn5 and cpn3 that were designed based on the alignment of various available cpn60 protein sequences, successfully amplified a 1 kb cpn60 gene fragment of *E. invadens*, *E. moshkovskii* and *E. terrapinae* using a plasmid DNA. The plasmid DNA contained a cpn60 insert that had been previously amplified by PCR using the

degenerate primers on total genomic DNA (153). Primers N-CPN60 and C-CPN60 based on the complete cpn60 gene sequence (82) were used for the amplification of a 500 bp cpn60 gene fragment of *E. histolytica*. An 800 bp *E. dispar* cpn60 probe was amplified using the degenerate primer cpn3 and the primer Disp-1, which was designed based on the partial *E. dispar* cpn60 sequence that was available.

The β-actin gene of *E. histolytica* was amplified using β-actin A and β-actin B genes on total genomic DNA giving a fragment of 500 bp approximately. This fragment was used as a probe for cross hybridisation in all the Northern blots of the *Entamoeba* species, except *E. invadens*. For the latter species a 290 bp *E. invadens* specific actin DNA probe was obtained by employing primers that were used for RT-PCR of *Entamoeba* species in another study (154).

A.



B.

	5′	3′
Consensus mt- hsp70	GVFEVKAT	VGIPPAPR
Eh ER	NVFDVIST	DGIQPAPR
Eh cytoplasmic	GVFEVKAS	TGIPPAPR

Figure 4.1. Alignment of Hsp70 genes of *Entamoeba histolytica* and other species.

A. The mitochondrial, cytoplasmic and endoplasmic reticulum hsp70s of *E. histolytica* were aligned using the multalin website (http://www.toulouse.inra.fr/multalin.html; (155)), with the mt-hsp70s of *Trichomonas vaginalis*, *Pisum sativum*, *Eimeria tenella*, *Saccharomyces cerevisiae*, and *Leishmania major*. Regions of conservation between the genes are given in red. D1.70 and D2.70 primers were designed from amino acid positions 274-281, and 525-532 respectively.

B. Alignment of the amino acid regions of the *E. histolytica* cytoplasmic and endoplasmic reticulum and the consensus mt-hsp70s where the degenerate primers were designed. The amino acids of the *E. histolytica* cytoplasmic and endoplasmic reticulum genes that differ from the consensus mt-hsp70 are given in red.

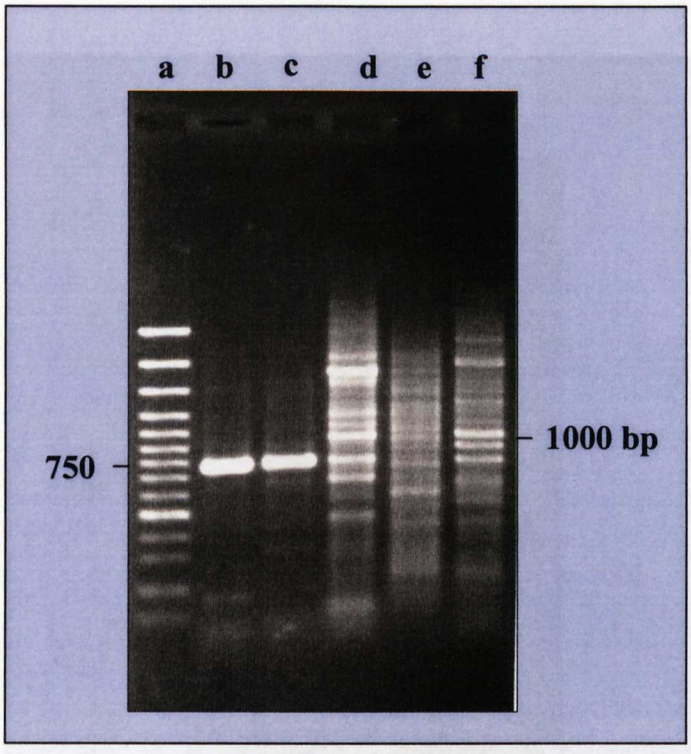


Figure 4.2. PCR amplification of the mt-hsp70 gene of *E. dispar*.

The thermal cycling programme consisted of 10 cycles of denaturation at 94 °C for 1 minute, annealing at 40 °C for 1.5 minutes and elongation at 72 °C for 2 minutes, 25 cycles at 94 °C for 1 minute, 48 °C for 1.5 minutes and 72 °C for 2 minutes, and one final cycle at 72 °C for 5 minutes. Lane a: 100 bp DNA ladder; lane b: *E. histolytica*; lane c: *E. dispar*; lane d: *E. invadens*; lane e: *E. moshkovskii* and lane f: *E. terrapinae*.

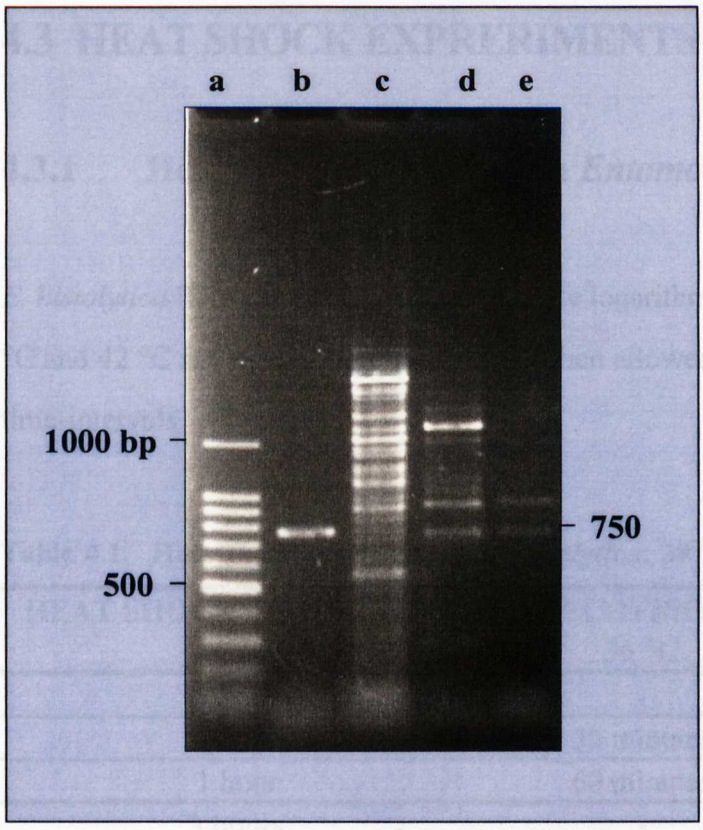


Figure 4.3. PCR amplification of *E. moshkovskii* and *E. terrapinae* mt-hsp70 gene.

Lane a: 100 bp ladder; lane b: *E. histolytica*; lane c: *E. invadens*; lane d: *E. moshkovskii*; and lane e: *E. terrapinae*. The thermal cycling programme consisted of 10 cycles of denaturation at 94 °C for 1 minute, annealing at 44 °C for 1.5 minutes and elongation at 72 °C for 2 minutes, 25 cycles at 94 °C for 1 minute, 48 °C for 1.5 minutes and 72 °C for 2 minutes, and one final cycle at 72 °C for 5 minutes. The product in lanes d and e was re-amplified and precipitated, separated in a 1% agarose gel and the band of the correct size (i.e. 750 bp approx.) cut out (section 2.3.2). After purification, the PCR product was ligated into pGEM®-T Easy vector and the identity of the clones confirmed by sequencing.

4.3 HEAT SHOCK EXPRERIMENTS

4.3.1 Heat shock experiments in *Entamoeba histolytica*

E. histolytica HM-1:IMSS clone 9 cells in late logarithmic stage were heat shocked at 39 °C and 42 °C for various time-periods and then allowed to recover at 36 °C for different time-intervals.

Table 4.1. Heat shock conditions - *E. histolytica*: 39 °C

HEAT SHOCK TIME at 39 °C	RECOVERY at 36 °C
1 hour	-
1 hour	30 minutes
1 hour	60 minutes
3 hours	-
3 hours	30 minutes
3 hours	60 minutes
3 hours	3 hours
3 hours	5 hours
3 hours	6 hours
3 hours	24 hours

Table 4.2. Heat shock conditions - *E. histolytica*: 42 °C

HEAT SHOCK TIME AT 42 °C	RECOVERY PERIOD AT 36 °C
1 hour	-
1 hour	1 hour
1 hour	3 hours
2 hours	-

4.3.2 Heat shock experiments in *E. dispar*

E. dispar SAW 760 cells in late logarithmic stage were heat shocked at 39 °C for various time-periods and then allowed to recover at 36 °C for different time-intervals.

Table 4.3. Heat shock conditions - *E. dispar* : 39 °C

HEAT SHOCK TIME at 39 °C	RECOVERY at 36 °C
1 hour	-
1 hour	1 hour
2 hours	-
3 hours	-
3 hours	1 hour
3 hours	3 hours
6 hours	-

4.3.3 Heat shock experiments in low-temperature *Entamoeba* species

E. invadens IP-1, *E. moshkovskii* Laredo and *E. terrapinae* M cells in late logarithmic stage were heat shocked at 37 °C for various time-periods and then allowed to recover at 22 °C for different time-intervals.

Table 4.4. Heat shock conditions - low temperature *Entamoeba* spp: 37 °C

HEAT SHOCK TIME at 37 °C	RECOVERY at 22 °C
1 hour	-
1 hour	1 hour
1 hour	3 hours
3 hours	-
3 hours	1 hour
3 hours	3 hours
3 hours	6 hours
6 hours	-

4.4 NORTHERN BLOT ANALYSIS & BAND QUANTITATION

After heat shock and recovery of the *Entamoeba* cells, mRNA was purified immediately as described (section 2.6), and separated by electrophoresis in a formaldehyde-containing agarose mini-gel. Northern blots were set up as described in section 2.11. As a control, mRNA of non-heat shocked *Entamoeba* cells was used. The Northern blot quantitation was done using the Phoretix 1D Advanced version 4.01 software (Linear Dynamics Ltd). The quantitative data are given below each band in the figures. Values varying by up to 2 fold were considered non-significant to take into account any measurement errors that may have occurred during quantitation of the mRNA samples.

4.4.1 Effects of heat shock on mitochondrial gene transcription of *E. histolytica*

Northern blot analysis had revealed two transcripts of the mt-hsp70 gene that differ in length. The main band is approximately 1.8 kb in size and the other is approximately 150 bases longer (Figures 4.4 & 4.5). The main mt-hsp70 message appears to be slightly induced during recovery from 3-hour heat shock at 39 °C when compared with the control β-actin mRNA, while the larger mRNA seems to be more strongly induced (Figure 4.4). The induced message level is highest after 24 hours of recovery (increased 18 fold; Figure 4.4, lane d). The actin is non-inducible, the difference in the intensity among the bands being due to the amount of the mRNA loaded.

Northern blot analysis after heat shock at 42 °C did not show any induction of expression in any of the two messages when compared to the non-induced β-actin control (Figure 4.5). It is possible that 1-hour heat shock at 42 °C is not enough to

induce any expression of the mRNA message. Another possibility could be that 42 °C is too high a temperature for heat shock induction to take place or that recovery was not long enough for it to be detected.

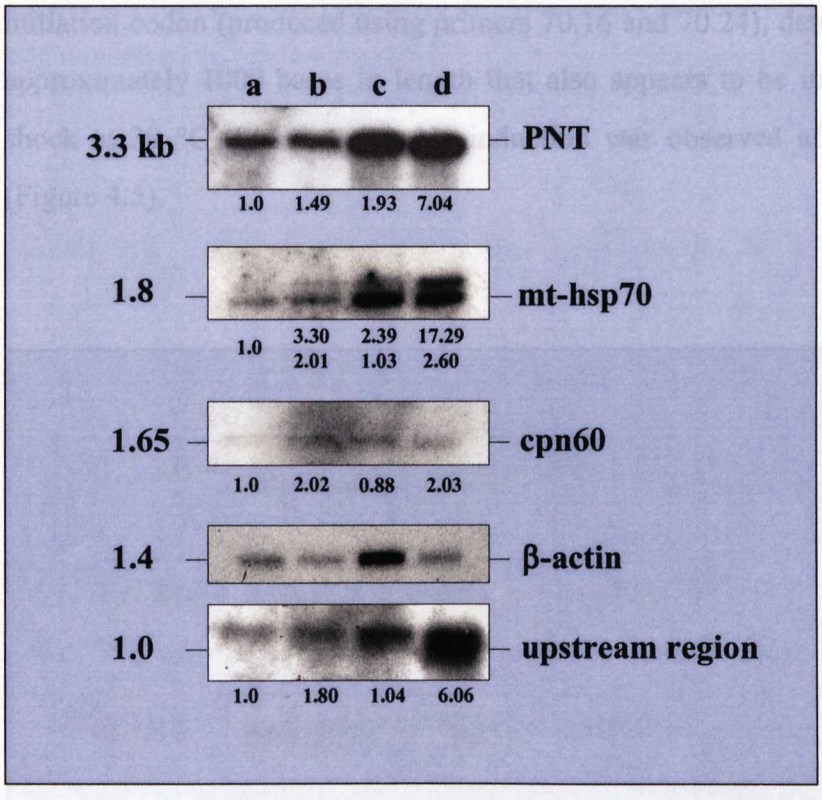


Figure 4.4. Northern blot of heat shocked *E. histolytica* cells at 39 °C.

Poly A^+ RNA extracted from trophozoites in the logarithmic phase of growth before heat shock (lane a), and following a 3 hour heat shock at 39 °C with recovery at 36 °C for 0 (lane b), 3 (lane c) and 24 hours (lane d). After electrophoresis and blotting, the filter was sequentially hybridised to five probes, which were randomly labelled with radioactive [α-^{32}P]-dCTP. After hybridisation, blots were exposed to X-ray film at –70 °C for up to 2 weeks, then stripped in boiling 0.1% SDS/DEPC-treated H_2O before re-hybridising to the other probes. The sizes of the mRNAs were measured by comparison to an RNA marker (0.28-6.58kb; Promega).

To verify that the upper band and the observed induction were not artefacts the same filter was hybridised to a probe for chaperonin cpn60, a gene previously shown not to be induced by heat shock (82). The result, which showed a single non-induced band for

cpn60, confirms that the mt-hsp70 probe is detecting two different transcripts at least one of which is induced following heat shock (Figure 4.4 & 4.5).

Hybridisation with a probe covering the 300 bp region immediately upstream of the initiation codon (produced using primers 70.16 and 70.24), detected a poly A^+ RNA of approximately 1000 bases in length that also appears to be induced by a 3-hour heat shock at 39 °C (Figure 4.4). No induction was observed after a 42 °C heat shock (Figure 4.5).

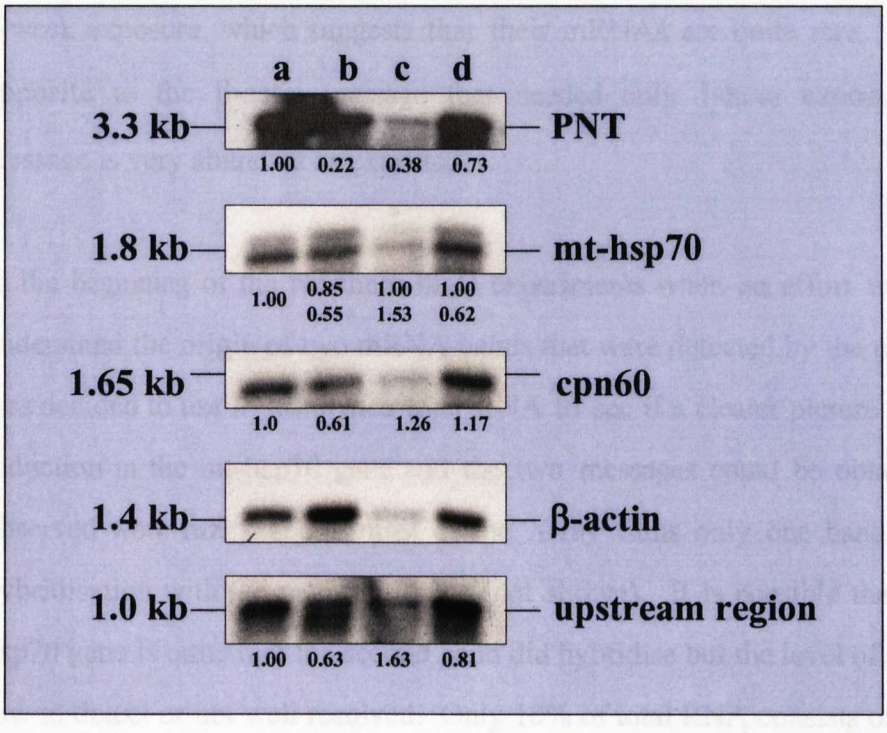


Figure 4.5. Northern blot of heat shocked *E. histolytica* cells at 42 °C.

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 1 hour heat shock at 42 °C with recovery at 36 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d). Blotting and hybridisation was as stated in figure 4.4.

Northern blot analysis with the PNT probe revealed a single band at 3.3 kb which seems to be induced after 3 hours of heat shock at 39 °C and highest after 24 hours of recovery (7 fold increase; Figure 4.4, lane d). The PNT mRNA level appears to decrease significantly after 1 hour of heat shock at 42 °C but has largely recovered to normal levels after 3 hours at 36 °C.

The relative exposure times of the X-ray films at –70 °C in the *E. histolytica* mitochondrial genes probes were the same except for the PNT where it took only an overnight exposure to see results. In contrast to that the cpn60 and mt-hsp70 required a 2-week exposure, which suggests that their mRNAs are quite rare. This is quite the opposite to the β-actin message that needed only 1-hour exposure indicating the message is very abundant as expected.

In the beginning of the Northern blots experiments when an effort was being made to understand the origin of two mRNA bands that were detected by the mt-hsp70 probe, it was decided to use *E. histolytica* total RNA to see if a clearer picture of the heat shock induction in the mt-hsp70 gene and the two messages could be obtained. The bands observed were fuzzy and in most of the X-ray films only one band was visible after hybridisation with the mt-hsp70 gene (not shown). It is possible that because the mt-hsp70 gene is quite rare the second band did hybridise but the level of transcript was too low to detect or not well resolved. Only 10% of total RNA consists of mRNA.

4.4.2 Effects of heat shock on mitochondrial gene transcription of *E. dispar*

Northern blot analysis of the *E. dispar* mRNA revealed one transcript hybridising to the mt-hsp70 probe of approximately 1.9 kb in size, which seems to be induced by 2-3 fold following a 3-hour heat shock (Figures 4.6 & 4.7). The mRNA band in the X-ray films

is very faint probably due to the amount of the mRNA being loaded. The *E. dispar* cultures used were monoxenic, which means some of the mRNA that was isolated belonged to *Crithidia fasciculata* and therefore less *E. dispar* mRNA would be present in the samples. After hybridisation with the mt-hsp70 probe the filters were exposed for 2 weeks at -70 °C, indicating that the mt-hsp70 is a rare message and therefore more mRNA would be needed to get stronger bands in the X-ray films. The solution could possibly be to grow more dense cultures of *E. dispar* for the mRNA purification.

As a non-induced control the β-actin gene of *E. histolytica* was used as a probe. The cross-hybridisation between the two species was successful giving good intensity bands (Figures 4.6 & 4.7).

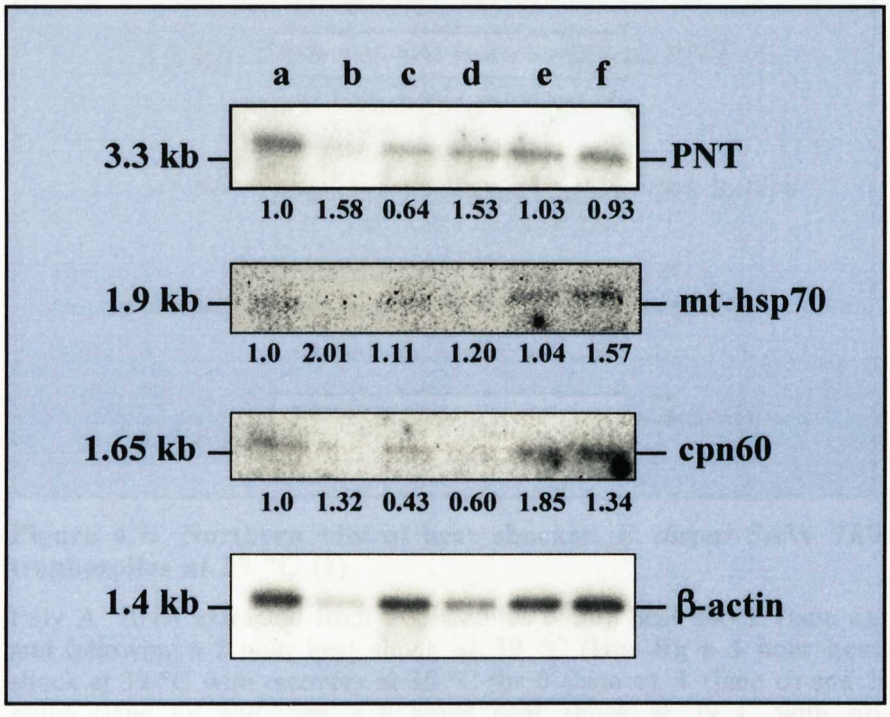


Figure 4.6. Northern blot of heat shocked *E. dispar* SAW 760 trophozoites at 39 °C (1).

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 1 hour heat shock at 39 °C with recovery at 36 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d); and then a 2 and a 6 hours heat shock at 39 °C with no recovery (lanes e and f respectively). Blotting and hybridisation was as stated in figure 4.4.

Hybridisation with the cpn60 gene gave one non-induced band at 1.65 kb approximately. After two unsuccessful hybridisations with the *E. dispar* PNT probe that did not reveal any band in the filter, the *E. histolytica* PNT gene was used. Cross hybridisation gave one band at 3.3 kb that seems to increase slightly after 3 hours of heat shock at 39 C with 3 hours of recovery (2.71 fold increase; Figure 4.7, lane e). This reading could be a bit misleading as the actin loading was lower than the rest in this lane.

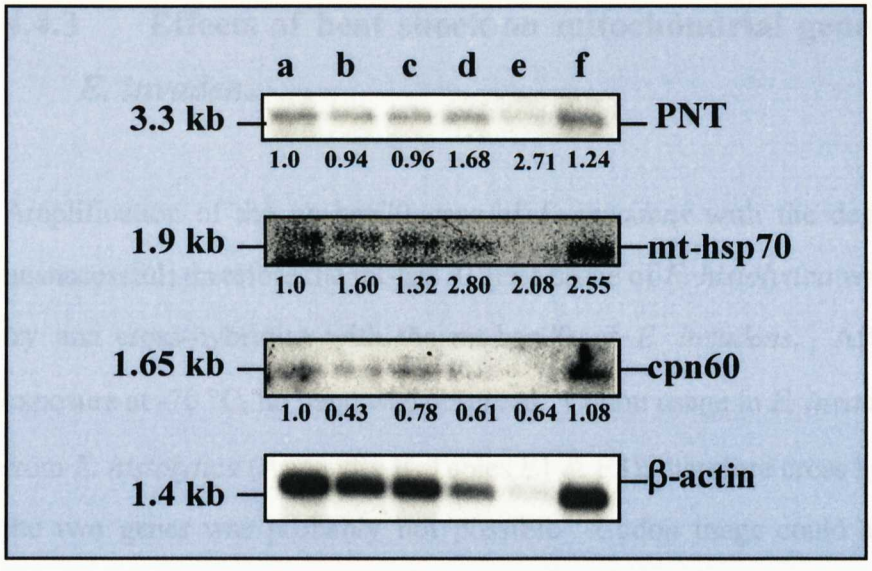


Figure 4.7. Northern blot of heat shocked *E. dispar* SAW 760 trophozoites at 39 °C (2).

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 2 hour heat shock at 39 °C (lane b); a 3 hour heat shock at 39 °C with recovery at 36 °C for 0 (lane c), 1 (lane d) and 3 hours (lane e); and then a 6 hours heat shock at 39 C with no recovery (lane f). Blotting and hybridisation was as stated in figure 4.4.

The sizes of the cpn60, PNT and β-actin mRNAs of *E. dispar* are the same as those of *E. histolytica*. Northern blot analysis with the mt-hsp70 probe revealed one transcript at 1.9 kb, which is 100 bp longer than the main mRNA transcript of the *E. histolytica* mt-hsp70. As expected the exposure time of the X-ray film for both the mt-hsp70 and cpn60 was 2 weeks indicating that the mRNA messages are very rare in *E. dispar*. This agrees with the *E. histolytica* results as well. The exposure time for the PNT was also 2 weeks probably because the *E. histolytica* PNT probe was used and due to low amounts of *E. dispar* mRNA loaded in the agarose gel.

4.4.3 Effects of heat shock on mitochondrial gene transcription of *E. invadens*

Amplification of the mt-hsp70 gene of *E. invadens* with the degenerate primers was unsuccessful; therefore the mt-hsp70 gene probe of *E. histolytica* was used as a probe to try and cross-hybridise with the mt-hsp70 of *E. invadens*. After 2 weeks of film exposure at -70 °C, no band was observed. Codon usage in *E. invadens* is quite different from *E. histolytica* (Appendix E, Tables E1 & E3); therefore cross hybridisation between the two genes was probably not possible. Codon usage could also explain the large number of non-specific amplified bands in the PCR of genomic DNA with the mt-hsp70 degenerate primers (Figure 4.3), if the degenerate primers amplified other genes of non mt-hsp70 origin.

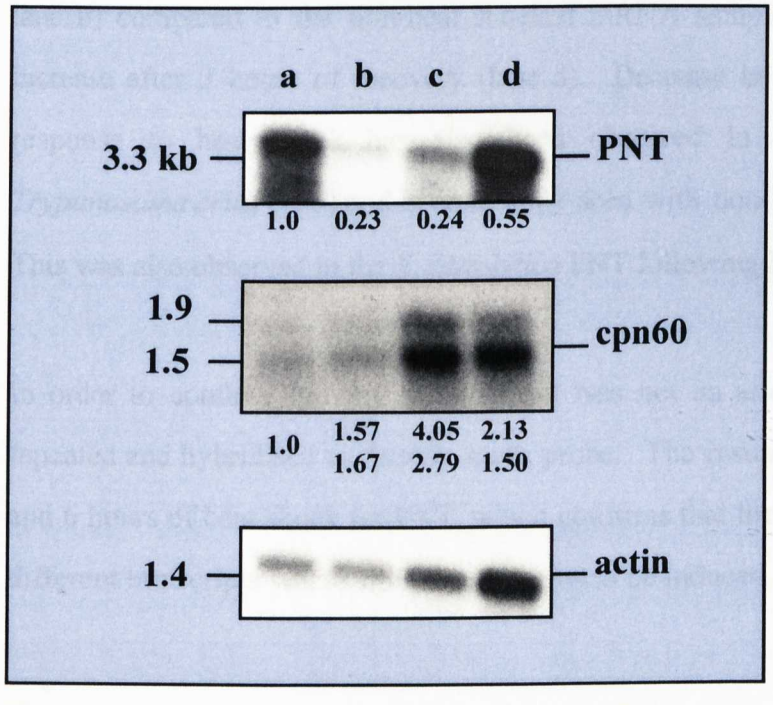


Figure 4.8. Northern blot of heat shocked *E. invadens* IP-1 trophozoites at 37 °C (1).

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 1 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d). Blotting and hybridisation was as stated in figure 4.4.

Northern blot analysis revealed two transcripts hybridising to the *E. histolytica* cpn60 probe that differ about 300 bp in length (Figures 4.8 & 4.9). The cpn60 message appears to be induced by 4 and 2.79 fold for the upper and lower bands respectively during recovery at 22 °C after 1 hour heat shock at 37 °C. After 3 hours of recovery the message seems to decrease (Figure 4.8). Hybridisation of the filter with the homologous *E. invadens* cpn60 probe did not give any results.

Hybridisation of the filter with the PNT probe revealed one band after 1 hour of heat shock but two bands after 3 hours of heat shock. The second message, which is about 4.1 kb, is induced and is most abundant after 6 hours of heat shock with no recovery (60 fold; Figure 4.9, lane e). This reading could be misleading because in the non-induced lane there was no band present. Non-heat shocked mRNA and all 1-hour heat shock samples gave only one band at 3.3 kb (Figure 4.8). Interestingly, the message seems to have decreased after 1 and 3 hours of heat shock with no recovery (Figures 4.8 & 4.9, lane b) compared to the non-heat shocked mRNA sample but levels have started to increase after 3 hours of recovery (lane d). Decrease in the levels of a transcript in response to heat shock has also been observed in the mt-hsp70 mRNAs of *Trypanosoma cruzi* (156) and is commonly seen with non-heat shock induced mRNAs. This was also observed in the *E. histolytica* PNT following heat shock at 42 °C.

In order to confirm that the second band was not an artefact the Northern blot was repeated and hybridised to freshly made probe. The results showed two bands after 3 and 6 hours of heat shock for PNT, which confirms that the PNT probe is detecting two different transcripts where at least one seems to be induced by heat shock.

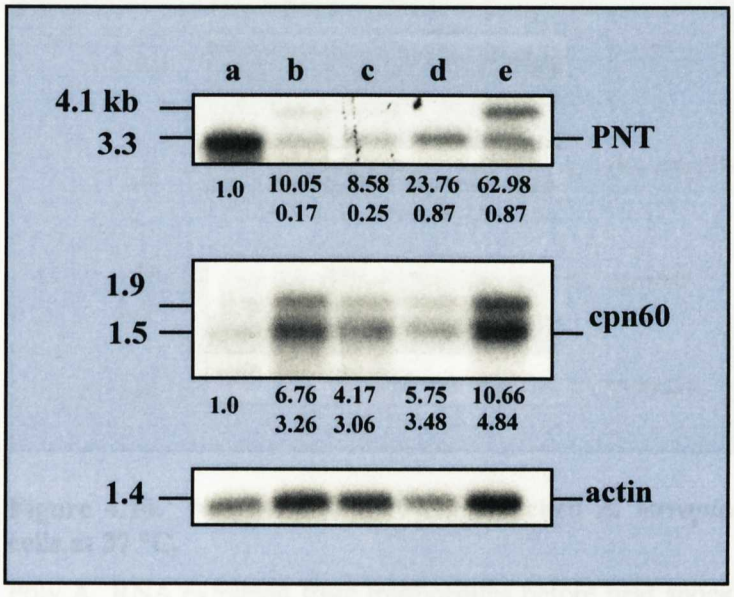


Figure 4.9. Northern blot of heat shocked *E. invadens* IP-1 trophozoites at 37 °C (2).

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 3 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d); and a 6 hour heat shock with no recovery (lane e). Blotting and hybridisation was as stated in figure 4.4.

4.4.4 Effects of heat shock on mitochondrial gene transcription of *E. terrapinae*

Northern blot analysis of *E. terrapinae* mRNA cross-hybridising to the β-actin *E. histolytica* probe showed nice sharp bands of the same intensity in all lanes indicating equal mRNA loading in the gel. Hybridisation to the PNT probe revealed one band of 3.3 kb in size, which seems to be strongly induced after 3 hours heat shock at 37 °C and 1 hour recovery at 22 °C and 6 hours of heat shock at 37 °C (9.7 and 9.2 fold; Figure 4.10, lanes f & g respectively). It is slightly less induced during recovery from a 1 hour heat shock (3 fold).

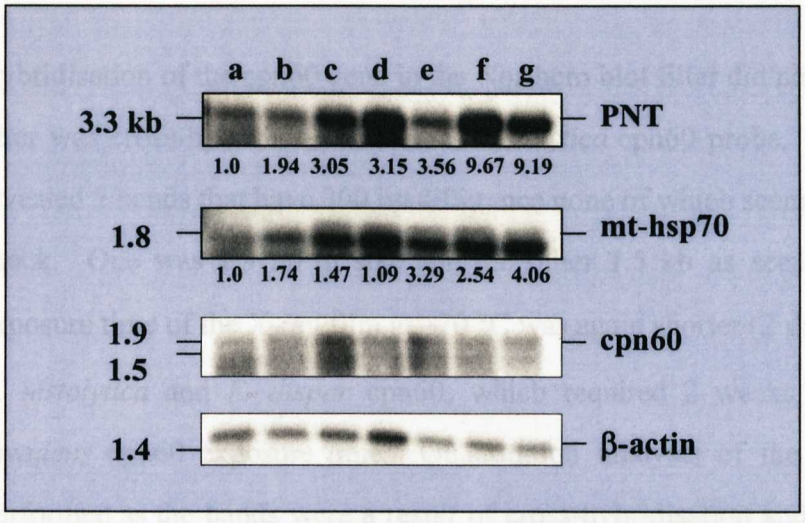


Figure 4.10. Northern blot of heat shocked *E. terrapinae* M cells at 37 °C.

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 1 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d); a 3 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane e), 1 (lane f), and a 6 hours heat shock at 22 °C with no recovery (lane g). Blotting and hybridisation was as stated in figure 4.4.

Northern blot analysis hybridising to the mt-hsp70 gene probe from the purified gel band (method in Figure 4.3), revealed 3 bands at 600 bp, 1.25 kb and 1.8 kb, of which the 600 bp was the most intense. After cloning and sequence verification of the PCR product then using it as a probe for hybridising to the mt-hsp70 mRNA, the result was one band at 1.8 kb (Figure 4.10). The three bands appearing in the film the first time were probably mRNAs belonging to other genes that were also present in the PCR product. In contrast to the *E. histolytica* and *E. dispar* mt-hsp70, the *E. terrapinae* mt-hsp70 required shorter exposure time of the X-ray film at –70 °C. It only took 1-½ days to get a clear result, which was not expected since the mt-hsp70 mRNA is not abundant in other *Entamoeba* species. Induction of the mt-hsp70 message was significantly higher after 3 hours of heat shock (3 fold increase) reaching its highest after 6 hours of heat shock (4 fold increase). These results are considered to be very reliable the loading of the mRNA was very consistent as it can be seen from the β-actin.

Hybridisation of the cpn60 gene in the Northern blot filter did not give any bands so the filter was cross-hybridised with an *E. histolytica* cpn60 probe. Northern blot analysis revealed 2 bands that have 300 bp difference none of which seems to be induced by heat shock. One was 1.9 kb in size and the other 1.5 kb as seen in *E. invadens*. The exposure time of the X-ray film at –70 °C was again shorter (2 days) than in the case of *E. histolytica* and *E. dispar* cpn60, which required 2 weeks, but similar to the *E. invadens* cpn60 exposure time. Quantitation analysis of the cpn60 bands was not performed as the bands were a result of cross-hybridisation and were quite fuzzy so it was not possible to get an accurate calculation.

4.4.5 Effects of heat shock on mitochondrial gene transcription of *E. moshkovskii*

Northern blot analysis of the *E. moshkovskii* mRNA did not give very clear bands (Figures 4.11 & 4.12). The bands were fuzzy and a smear was present in all of the samples when hybridised to any of the DNA probes. This could be because the trophozoites grow in large numbers and therefore there were too many cells for the mRNA purification to be optimal. A smear could also indicate that the migration of the mRNA samples was not normal and therefore size measurement would not be very accurate. It is also possible that poly A^+ tails are very variable in length in this species.

Hybridisation of the Northern blot filter with the purified mt-hsp70 PCR product from the gel band (method in Figure 4.3) gave a band of 1.5 kb. A sample of the gel band was cloned and sequenced. Sequence analysis revealed that the product was of mitochondrial hsp70 origin. The plasmid DNA was used as a PCR template and amplified with degenerate primers D1.70 and D2.70 to get the specific mt-hsp70 fragment, which was then used to re-hybridise the Northern blot filters. No band was visible after 2 weeks of

X-ray film exposure at -70 °C. It is likely that the hybridisation did not work either because the labelling of the probe was not successful or because there was something wrong with the PCR fragment. So it is possible that the unusually small mRNA detected is not actually mt-hsp70 but a second sequence present in the PCR product. The same explanation could be given for the β-actin mRNA in the *E. moshkovskii* blots, where after hybridisation with a heterologous *E. histolytica* probe a smaller than expected mRNA (1.25 kb) was present. This could also be explained by the apparent low quality and unusual migration of the mRNA samples which made accurate calculation of the message size quite difficult.

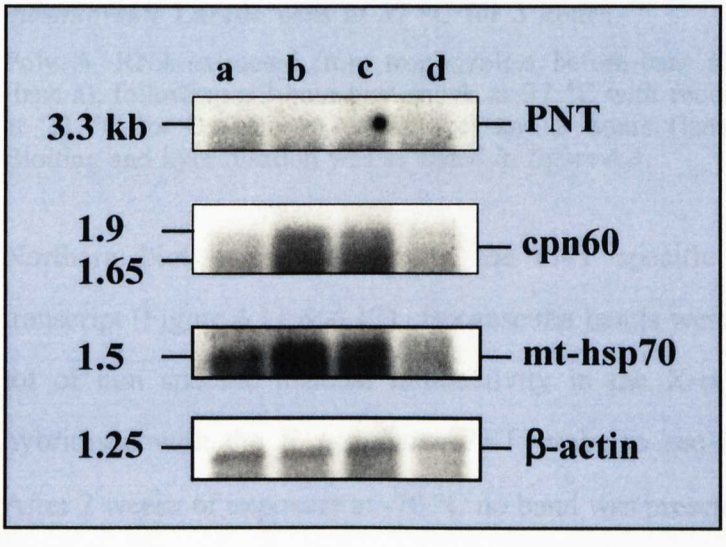


Figure 4.11. Northern blot of heat shocked *E. moshkovskii* Laredo cells at 37 °C for 1 hour.

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), and following a 1 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d). Blotting and hybridisation was as stated in figure 4.4.

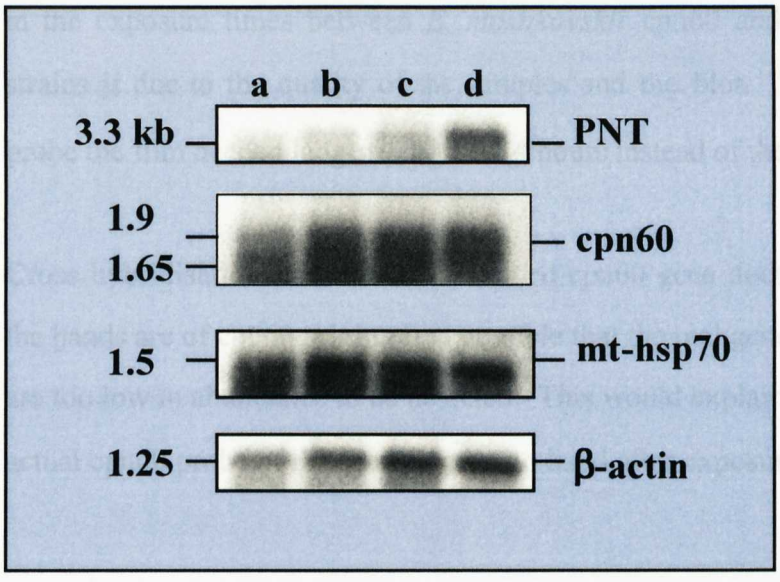


Figure 4.12. Northern blot of heat shocked *E. moshkovskii* Laredo cells at 37 °C for 3 hours.

Poly A^+ RNA extracted from trophozoites before heat shock (lane a), following a 3 hour heat shock at 37 °C with recovery at 22 °C for 0 (lane b), 1 (lane c) and 3 hours (lane d). Blotting and hybridisation was as stated in figure 4.4.

Northern blot hybridisation with the PNT specific probe gave a 3.3 kb mRNA transcript (Figure 4.11 & 4.12). Because the bands were not very clear and there was a lot of non specific residual radioactivity in the X-ray films, the filters were cross hybridised with the *E. histolytica* PNT probe to see if a clearer band would appear. After 2 weeks of exposure at -70 °C no band was present in the X-ray film. This could be due to the different codon usage between the two *Entamoeba* species (Appendix E, Table E1 & E4).

Hybridisation with the cpn60 probe did not give any clear bands either. Cross hybridisation with the *E. histolytica* cpn60 probe gave two bands at approximately 1.9 and 1.65 kb that seem to be constant during heat shock. The upper band agrees with both the *E. invadens* and *E. terrapinae* cpn60 results, and the lower band with the cpn60 single band results of *E. histolytica* and *E. dispar*. Exposure time of the X-ray film at -70 °C was 2 weeks, which is in contrast to the *E. invadens* and *E. terrapinae* but

agrees with the *E. histolytica* and *E. dispar* exposure times. It is possible the difference in the exposure times between *E. moshkovskii* cpn60 and the other low-temperature strains is due to the quality of the samples and the blot. Even with the β-actin gene probe the film needed longer exposure (4 hours instead of the usual 1 hour).

Cross hybridisation with the *E. histolytica* cpn60 gene does not necessarily mean that the bands are of cpn60 origin. It is possible that the real genes are not expressed or they are too low in abundance to be detected. This would explain why hybridisation with the actual cpn60 probe of the species gave no band after exposure at –70 °C for 2 weeks.

Table 4.5. Bands in the Northern blots of the *Entamoeba* species

	β-actin	Mt-hsp70	Cpn60	PNT
E. histolytica	1.4	1.95 1.8	1.65	3.3
E. dispar	1.4	1.9	1.65	3.3
E. invadens	1.4	-	1.9 1.5	4.1 3.3
E. terrapinae	1.4	1.8	1.9 1.5	3.3
E. moshkovskii	1.25	1.5	1.9 1.65	3.3

✶The band sizes are given in kb.
✶The mRNA transcripts that are induced during heat shock are given in red colour.

Table 4.6. Bands in the Northern blots using homologous DNA probes

	β-actin	Mt-hsp70	Cpn60	PNT
E. histolytica	1.4	1.95 1.8	1.65	3.3
E. dispar	-	1.9	1.65	ND
E. invadens	1.4	-	ND	4.1 3.3
E. terrapinae	-	1.8	ND	3.3
E. moshkovskii	-	ND	ND	3.3

★The band sizes are given in kb.
★The mRNA transcripts that are induced during heat shock are given in red colour.
★ND: no detected – a homologous probe was used but no band was seen.

4.5 SUMMARY & CONCLUSIONS

The mt-hsp70 gene of *E. histolytica* appears to give rise to two mRNA transcripts differing in length by approximately 150 bases the larger of which is clearly induced at 39 °C, reaching a maximum level after 24 hours recovery. This result contrasts with a previous heat shock result for the cpn60 gene of *E. histolytica* (82), which has been confirmed here, indicating that not all genes of mitochondrial origin are regulated in the same way. In addition, a 1-kb apparently non-coding transcript detected upstream of the mt-hsp70 gene is also induced.

The existence of two mt-hsp70 mRNAs has also been reported in *Trypanosoma cruzi* and *T. brucei*, two other protozoan species. In *Leishmania major* four mRNAs have been reported, suggesting that multiple, and perhaps functionally distinct mt-hsp70

proteins exist (96). In these species multiple genes appear. In *E. histolytica* it has recently been discovered that two mt-hsp70 genes are present (Chapter 4).

The rest of the *Entamoeba* species appeared to have only one mRNA transcript for the mt-hsp70 gene. The mRNA size is variable between the species as it can be seen from table 4.5. The *E. dispar* and *E. terrapinae* mt-hsp70s also seem to be induced by heat shock. *E. invadens* was the only organism for which the mt-hsp70 gene was not cloned after amplification using the mt-hsp70 degenerate primers. It is possible that one of the bands present in the PCR product (Figure 4.2, lane d) corresponded to the mt-hsp70 gene but because of the other non-specific bands it is difficult to know which one it could be.

E. histolytica and *E. dispar* cpn60 mRNAs had the same size (1.65 kb) in the Northern blot filters. After cross hybridisation of the rest of the *Entamoeba* species blots with the cpn60 gene of *E. histolytica* two mRNA transcripts were picked up at 1.9 and 1.5 kb. With the exception of *E. invadens* cpn60 mRNA, none seem to be induced by heat shock. It is likely that the cross hybridisation with the *E. histolytica* cpn60 probe was not specific and therefore these mRNAs are not of cpn60 origin. The fact that all the Northern blots of the low-temperature *Entamoeba* strains were negative after hybridisation to the homologous cpn60 gene could mean that either the cpn60 gene in these strains is not expressed under those conditions or that the level of transcript is too low to detect. The chances that the synthesis of all the homologous cpn60 probes did not work is quite slim.

As in the case of the cpn60 gene, using other heterologous gene probes to cross-hybridise did not always give clear results. In addition the *E. moshkovskii* β-actin mRNA was smaller than expected. This could be explained by the irregular migration of the mRNA samples and the fuzziness in the bands that made the size measurement inaccurate in these species. Northern blots for the rest of the *Entamoeba* species that

were cross-hybridised with the *E. histolytica* β-actin probe gave the expected 1.4 kb size mRNA. Using the PNT of *E. histolytica* probe for the *E. dispar* blots gave a band at 3.3 kb, also as expected.

The heat shock results for the mt-hsp70 gene of the *Entamoeba* species, except *E. histolytica*, are in contrast to the induction of this mRNA during times of stress observed for other hsp70 homologues (150). Nevertheless these results together with those for cpn60 are in agreement with the heat shock results for the chaperonin 60 protein found in *E. histolytica* (82) and in *Giardia lamblia* (73). Experiments have shown that the expression of those proteins is not responsive to stress.

In contrast to these results, Mai *et al* reported that the cpn60 mRNA is induced after one hour of heat shock at 42 °C (83). In this experiment they used the amoebapore A gene as control and RT-PCR to amplify the mRNA. This method is inadequate to address this question, because the amount of mRNA template present for the two genes comparable. cpn60 is a rare message, in contrast to the amoebapore A mRNA. So it is likely that the amoebapore product reached PCR saturation before the cpn60.

All the *Entamoeba* species appeared to have the same size PNT mRNA at 3.3 kb. In *E. histolytica*, *E. invadens* and *E. terrapinae* the mRNA seems to be induced after 3 hours of heat shock. *E. invadens* produced a second mRNA at 4.1 kb, which seemed to be at its highest level after 6 hours of heat shock at 37 °C. As far as we are aware nobody has ever looked at the effect of heat shock expression on PNT in other organisms so no comparison with other similar studies could be made. The PNT of *Entamoeba* species is a single polypeptide orientated in a β to α direction, which is why a single large band at 3.3 kb is seen in the Northern blots (98).

As it can be seen from these results, the response to heat shock was variable between genes and between species. In *E. histolytica* blots only the mt-hsp70 mRNA was

induced. In the rest of the species no differential induction of this gene was observed following heat shock. The PNT gene also seemed to be induced in *E. terrapinae* and *E. invadens*. The *E. moshkovskii* PNT was very faint even after 2 weeks of exposure at -70 °C so if there were any slight induction it would be difficult to spot. The PNT mRNA of *E. invadens*, and *E. histolytica* seems to decrease in abundance during heat shock and starts increasing again after one-hour recovery or more.

CHAPTER 5

PHYLOGENETIC ANALYSIS OF THE MITOCHONDRIAL GENES OF *ENTAMOEBA* SPECIES

5.1 INTRODUCTION

In *E. histolytica* several genes encoding proteins located in the mitochondrion of other eukaryotes have been found; these include the cpn60, PNT (67) and mt-hsp70 gene reported here. Phylogenetic analyses of the cpn60 and PNT genes support their mitochondrial origin (67) and the hypothesis of secondary mitochondrial loss in *E. histolytica.* Furthermore, the cpn60 gene has been localised to a sub-cellular compartment named the mitosome, which is believed to be a mitochondrial remnant (82).

As the existence of mitochondrial genes has been shown in other *Entamoeba* species (Chapter 4) it is very likely that the mitosome is also present in these species. Phylogenetic studies of the complete *E. histolytica* mt-hsp70 sequence and the partial cpn60 and PNT genes from other species of *Entamoeba* should reveal a close relationship to other mitochondrial homologues.

5.2 PHYLOGENETIC ANALYSIS

The complete sequence of the mt-hsp70 of *E. histolytica* was obtained after PCR with degenerate primers on total genomic DNA and library screening (Chapter 3). PCR with degenerate primers on total genomic *Entamoeba* DNA followed by cloning and DNA sequencing confirmed the existence of the cpn60, PNT and mt-hsp70 genes in other *Entamoeba* species (Chapter 4, section 4.2). The sequences available for the mt-hsp70 genes were not complete enough to be used for phylogenetic analysis.

5.2.1 Phylogenetic analysis of the *E. histolytica* mt-hsp70 gene

The complete inferred amino acid sequence of the mt-hsp70 of *E. histolytica* was aligned by eye to mitochondrial, hydrogenosomal and related bacterial hsp70 homologues from 15 species (Figure 5.1 & Appendix B) and subjected to a variety of phylogenetic analyses. The protein sequences were obtained from GenBank™. Ambiguous areas of the alignment were not included in the phylogenetic analysis, leaving 521 alignment positions in the final data set.

Overall the amino terminus of these sequences is highly conserved. The most conserved part, which is very close to the mature amino terminus, is the IGIDLGTT motif (Figure 5.1) from which the degenerate primer 1 was designed. The carboxyl terminus is not conserved so it was not used for the sequence alignment and the phylogenetic analysis.

Phylogenetic trees were obtained by maximum parsimony, distance, maximum likelihood, and quartet puzzling analyses and their topologies were very similar. A consensus tree was finally made (Figure 5.2).

Phylogenetic analysis of the *E. histolytica* mt-hsp70 gene strongly supports the view that it is of mitochondrial origin as it clusters with a group of sequences that in other

lineages are targeted to the mitochondrial or hydrogenosomal compartment. Bootstrap values supporting the mitochondrial origin of the *E. histolytica* sequence were 86% for maximum parsimony, 96% for distance, 94% for maximum likelihood and 87% for quartet puzzling analyses, all of which are considered significant. As is the case for many of its genes (67), the evolutionary rate for the *E. histolytica* mt-hsp70 is clearly very high as seen by its long-branch length in the consensus tree.

The branching order within the 'mitochondrial' clade was poorly resolved. There is no support for placing the *E. histolytica* sequence outside the 'mitochondrial' clade and little support for placing it as the deepest branch within this clade. The maximum support for the latter topology is 21% in Puzzleboot – for comparison, a clade formed of *Eimeria tenella* and *Pisum sativum* was also deepest 21% of the time in the same analysis.

The mitochondrial clade in the hsp70 phylogenetic tree forms a sister-group with the α-Proteobacteria, as is the case for many mitochondrial proteins, suggesting that the mitochondrial endosymbiosis occurred before the emergence of *E. histolytica.* The bootstrap values for this mitochondria/α-Proteobacteria sister grouping were 83% for maximum parsimony, 75% for distance, 74% for maximum likelihood and 79% for quartet puzzling analyses. These are reasonably significant and support previous studies indicating that mitochondria evolved from an endosymbiosis between an eukaryotic cell and an α-Proteobacterium (61-63, 66).

The identification of mitochondrion specific proteins in *E. histolytica* such as cpn60 and PNT (67) and the mt-hsp70 reported here confirms that it is not a surviving relic of pre-mitochondrial evolution, but that its ancestors once harboured mitochondria or an endosymbiont related to the progenitor of mitochondria. Most functions of this organelle have been lost secondarily during evolution, possibly as a result of parasitic life or due to adaptations in an anaerobic environment. The function of the remaining compartment, the mitosome, remains obscure.

This finding together with the studies made on the mt-hsp70 in other amitochondriate groups, such as *Trichomonas vaginalis* (69, 133), *Nosema locustae* (72), *Vairimorpha necatrix* (71) and the cpn60 of *Giardia lamblia* (73) and *T. vaginalis* (68, 75, 133), confirms the hypothesis that all these organisms lost their mitochondrion during evolution. Phylogenetic trees place these genes within the mitochondrial clade of the α-Proteobacteria. This indicates that the common ancestor of all the existing eukaryotic species could have been a mitochondriate eukaryote (61, 73, 75, 83, 133).

Comparison of a comparable cpn60 phylogenetic tree (68) with that of mt-hsp70 (Figure 5.2) reveals that the latter gives stronger support for the mitochondrial origins of the *E. histolytica* proteins. It is also interesting to note that while *Rickettsia* always groups with the mitochondrial clade in cpn60 trees, in the mt-hsp70 trees it clusters with the other α-Proteobacteria. This could mean that *Rickettsia* is not, in fact, more closely related to mitochondria than the other α-Proteobacteria as has been suggested (66). A consensus built from maximum likelihood analysis of many mitochondrial specific genes will be needed to resolve this issue.

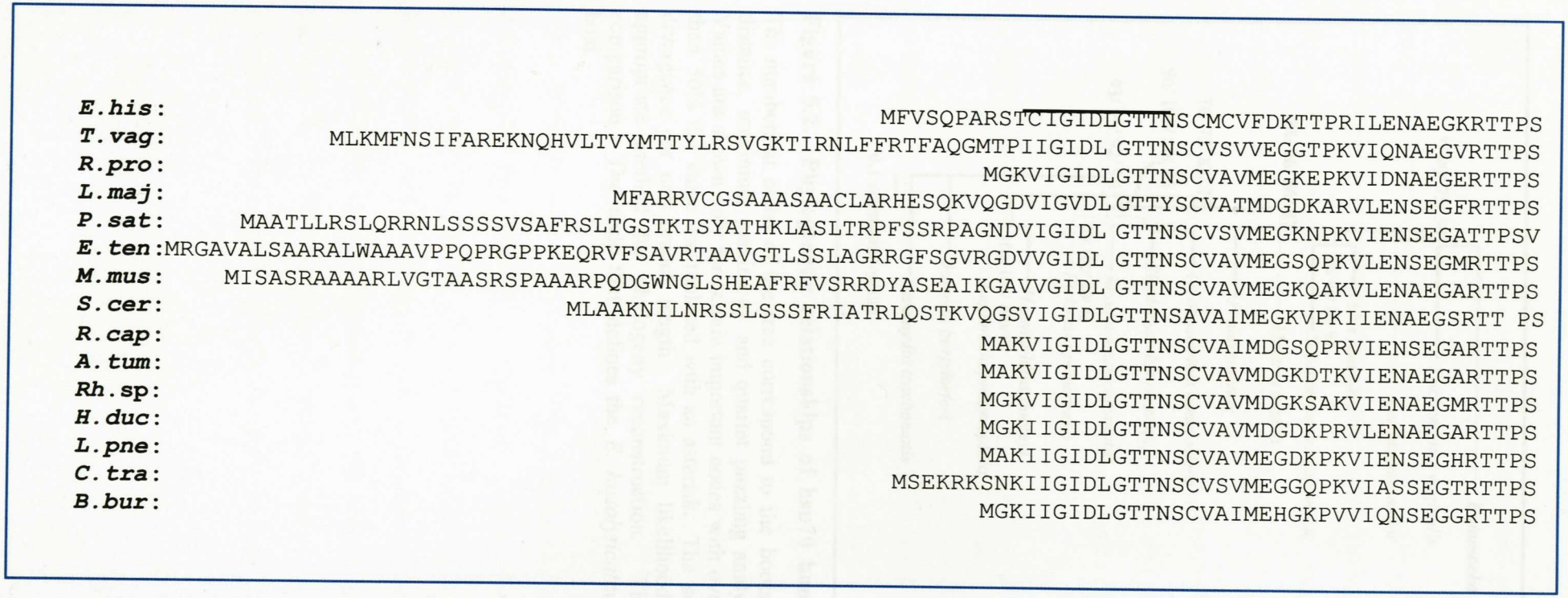


Figure. 5.1. Alignment of amino terminus of various hsp70 proteins
The underlined area is a highly conserved region of the mitochondrial hsp70 pr otein that was used to obtain degenerate primer 1. The amino terminus of the alignment is shown, in descending order, for the following s pecies: *Entamoeba histolytica, Trichomonas vaginalis, Rickettsia prowazekii, Leishmania major, Pisum sativum, Eimeria tenella, Mus musculus, Saccharomyces cerevisiae, Rhodobacter capsulatus, Agrobacterium tumefaciens, Rhodopseudomonas* sp., *Haemophilus ducreyi, Legionella pneumophila, Chlamydia trachomatis,* and *Borrelia burgdorferi.*

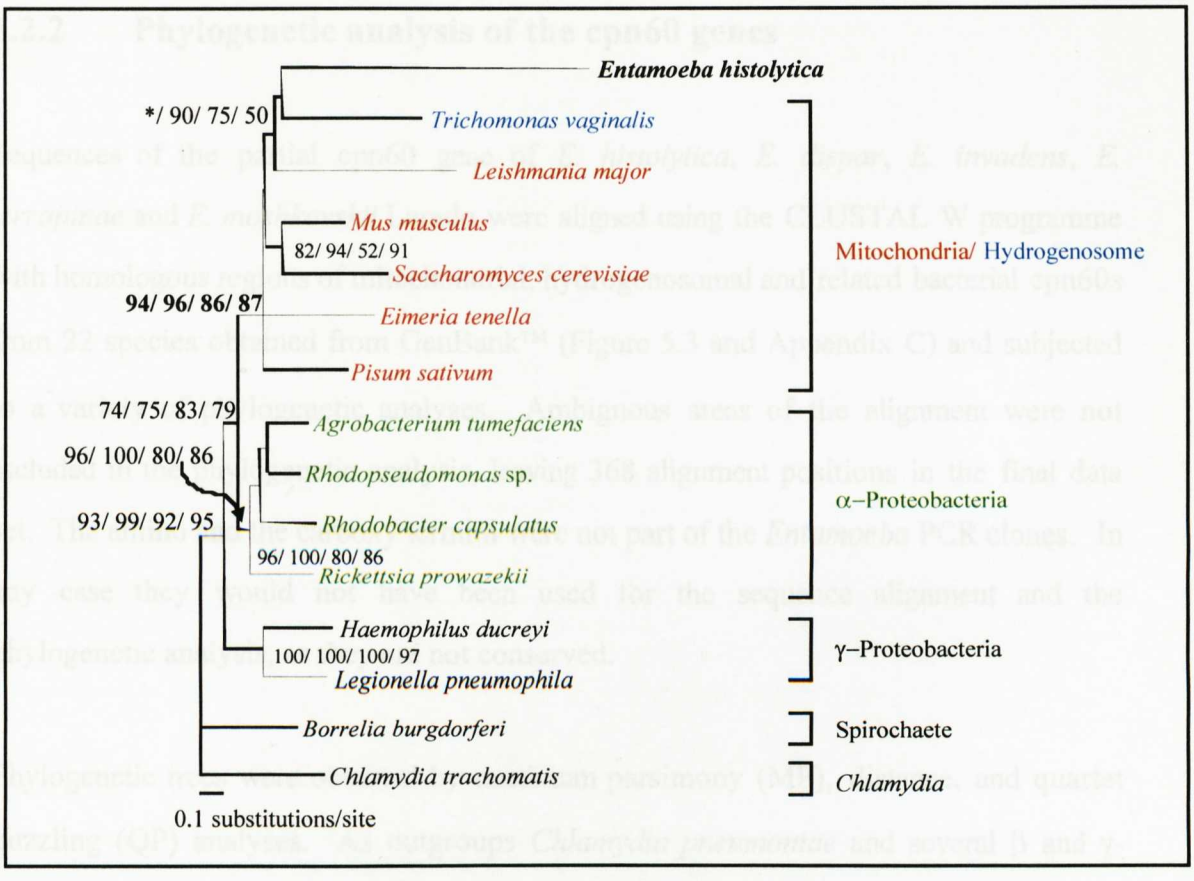


Figure 5.2. Phylogenetic relationships of hsp70 homologues.

The numbers at nodes of the tree correspond to the bootstrap values for maximum likelihood, distance, maximum parsimony and quartet puzzling analyses from left to right respectively. Values are shown only for certain important nodes with over 50% support. If a value was lower than 50% the value was replaced with an asterisk. The scale bar indicates estimated sequence divergence per unit branch length. Maximum likelihood is generally accepted as the most appropriate method for phylogeny reconstruction. The other methods are included for comparison. The node that defines the *E. histolytica*/mitochondrial relationship is given in bold.

5.2.2 Phylogenetic analysis of the cpn60 genes

Sequences of the partial cpn60 gene of *E. histolytica*, *E. dispar*, *E. invadens*, *E. terrapinae* and *E. moshkovskii* Laredo were aligned using the CLUSTAL W programme with homologous regions of mitochondrial, hydrogenosomal and related bacterial cpn60s from 22 species obtained from GenBank™ (Figure 5.3 and Appendix C) and subjected to a variety of phylogenetic analyses. Ambiguous areas of the alignment were not included in the phylogenetic analysis, leaving 368 alignment positions in the final data set. The amino and the carboxy termini were not part of the *Entamoeba* PCR clones. In any case they would not have been used for the sequence alignment and the phylogenetic analysis, as they are not conserved.

Phylogenetic trees were obtained by maximum parsimony (MP), distance, and quartet puzzling (QP) analyses. As outgroups *Chlamydia pneumoniae* and several β and γ-Proteobacteria have been used. A consensus tree was finally made (Figure 5.4).

In all the cpn60 trees the *Entamoeba* species were clustered together with 100% bootstrap support. The topology of the *Entamoeba* branches is similar to that reported in phylogenetic trees based on small subunit ribosomal RNA gene sequences (22, 40). *E. histolytica* and *E. dispar* form a sister group with high bootstrap support (100%) confirming the fact that they are the most closely related species in the genus. *E. invadens* branches first in all three trees followed by either *E. terrapinae* or *E. moshkovskii*. In the distance tree *E. terrapinae* is clustered together with *E. moshkovskii* with 81% bootstrap support. The branching order of these two species is not resolved. In contrast, in small subunit ribosomal RNA trees *E. moshkovskii* is closely related to *E. histolytica* and *E. dispar* with high bootstrap support.

	Cpn5	Cpn3
Entamoeba histolytica	TKDGVSVAKALTFSDNTLNVG...	...VCAVKAALAEGIVPGGG
Entamoeba dispar	TKDGVTVAKALTFSDNTLNVG...	...VYAVKAALAEGIVPGGG
Entamoeba moshkovskii (Lar)	TKDGVTVAKSLTFSDNTLNVG...	...VCAVKAALAEGIVPGGG
Entamoeba terrapinae	TKDGVTVAKAISFSDNTLNVG...	...VCAVKAALSEGIVPGGG
Entamoeba invadens	TKDGVTVAKAIEFSDKTLNAG...	...VCAVKAALSEGIVPGGG
Chlamydia pneumoniae	TKDGVTVAKEIELEDKHENMG...	...QHATIAAVEEGILPGGG
Escherichia coli	TKDGVSVAREIELEDKFENMG...	...LHATRAAVEEGVVAGGG
Haemophilus influenzae	TKDGVSVAREIELEDKFENMG...	...LHATRAAVEEGIVPGGG
Pseudomonas aeruginosa	TKDGVSVAKEIELKDKFENMG...	...LHATRAAVEEGVVPGGG
Neisseria gonorrhoeae	TKDGVTVAKEIELKDKFENMG...	...LHATRAAVEEGVVAGGG
Legionella micdadei	TKDGVSVAKEIEFENRFKNMG...	...LHATRAAVEEGIVAGGG
Coxiella burnetii	TKDGVSVAKEIELEDKFENMG...	...LHATRAAVEEGVVPGGG
Agrobacterium tumefaciens	TKDGVSVAKEIELEDKFENMG...	...LNATRAAVQEGIVPGGG
Rhizobium meliloti	TKDGVSVAKEIELEDKFENMG...	...LNATRAAVQEGIVPGGG
Brucella abortus	TKDGVSVAKEVELEDKFENMG...	...LNATRAAVEEGIVAGGG
Bartonella bacilliformis	TKDGVSVAKEIELELNKFENMG...	...LNATRAAVEEGIVPGGG
Bradyrhizobium japonicum	TKDGVAVAKEIELDDKFENMG...	...MHATRAAVEEGIVPGGG
Homo sapiens	TKDGVTVAKSIDLKDKYKNIG...	...LNATRAAVEEGIVLGGG
Zea mays	TKDGVTVAKSIEFKDRVKNVG...	...LNATKAAVEEGIVPGGG
Saccharomyces cerevisiae	TKDGVTVAKSIVLKDKFENMG...	...LNATRAAVEEGILPGGG
Trypanosoma cruzi	TKDGVTVAKAIEFKDPFENMG...	...LCSTRAAVQEGIVPGGG
Trypanosoma brucei	TKDGVTVAKSIEFKDPFENMG...	...LCSTRAAVQEGIVPGGG
Trichomonas vaginalis	TKDGVTVAKSIEFADKWHNMG...	...LNATRAAIEEGIVAGGG
Rickettisa tsutsugamushi	TKDGVSVAKAIQLKDKSLNVG...	...LHATRAAVEEGIVPGGG
Cowdria ruminantium	TKDGYKVMKSIKPEDPLALAI...	...LHATRAAVEEGVVPGGG
Ehrlichia chaffeensis	TKDGYKVIKSIKPEDPLALAI...	...LHATRAAVEEGVVPGGG

Figure 5.3. Part of the cpn60 protein alignment used for the phylogenetic analysis.

Partial cpn60 sequences of the *E. histolytica*, *E. dispar*, *E. moshkovskii*, *E. terrapinae* and *E. invadens* were obtained by PCR using degenerate primers on total genomic DNA (Chapter 5, section 5.3.1). The PCR products were cloned, sequenced to confirm their origin and then aligned using the CLUSTAL W website (http://www2.ebi.ac.uk/clustalw; (157)) with mitochondrial, hydrogenosomal and related bacterial homologues. The coloured amino acids represent regions of conserved areas between the cpn60 proteins from where the degenerate primers cpn5 and cpn3 were designed.

In both distance and QP analysis the *T. vaginalis* sequence forms a group with the *Entamoeba* clade indicating a possibly close relationship between the two organelles that have been connected by this cpn60 gene (hydrogenosome and mitosome). The bootstrap support is quite high at 59% and 90% respectively. The MP support for this group is low (21%). In the MP tree *T. vaginalis* is grouped with *Zea mays* mitochondrial cpn60 but only with 18% bootstrap support.

Phylogenetic analysis of the *Entamoeba* cpn60 genes supports the view that they are of mitochondrial origin as they cluster with a group of proteins that in other lineages are

targeted to the mitochondrial or hydrogenosomal compartment, suggesting that the mitochondrial endosymbiosis occurred before the emergence of the genus *Entamoeba.* Bootstrap values supporting the mitochondrial origin of the *Entamoeba* cpn60 sequences were 45% for MP, 43% for distance, and 83% for QP analyses. The weak support for this grouping in the distance and MP analyses has also been observed in other cpn60 trees (68). As is the case for many of its genes (67), the evolutionary rate for the *Entamoeba* cpn60 genes is clearly very high as seen by their long-branch lengths in the consensus tree.

In previously reported cpn60 trees (67, 68), the Rickettsiales group of the α-Proteobacteria (comprised of *Rickettsia tsutsugamushi*, *Cowdria ruminantium* and *Ehrlichia chaffeensis*) are always clustered with the mitochondrial clade indicating a close relationship between the two groups. *T. vaginalis* also formed a clade with the mitochondrial and *E. histolytica* cpn60 homologues (68). The same grouping is also observed in this study where the Rickettsiales are clustered with the mitochondria/mitosome/hydrogenosome group. This sister grouping was supported relatively well with 100%, 88% and 47% bootstrap support for the distance, MP and QP analyses respectively, indicating a possible closer relationship of *Rickettsia* to the mitochondria that to the other α-Proteobacteria (66). This comes in contrast to the mt-hsp70 phylogenetic tree of *E. histolytica* (Figure 5.2) where *Rickettsia* clustered with the other α-proteobacteria rather than the mitochondrial clade. As was suggested in chapter 3, a consensus built from maximum likelihood analysis of many mitochondrial specific genes will be needed to resolve this conflict.

The mitochondrial/hydrogenosome/*Rickettsia*/mitosome clade in the cpn60 phylogenetic tree forms a sister-group to the α-proteobacteria, as is the case for many mitochondrial proteins. The bootstrap values for this mitochondria/α-Proteobacteria sister grouping were 15% for MP, 48% for distance and 93% for QP analyses. With the exception of the MP analysis the other two are relatively significant and support previous studies

indicating that mitochondria evolved from an endosymbiosis between an eukaryotic cell and an α-Proteobacterium (61-63, 66).

The MP analysis gave in general very low bootstrap support, for the topology of all the different clades. This could be because MP analysis searches for the tree that requires the smallest number of evolutionary changes to explain the differences observed (158). The long-branch attraction artefact and also the rate of variation among sites can cause problems for MP analysis and result in low support.

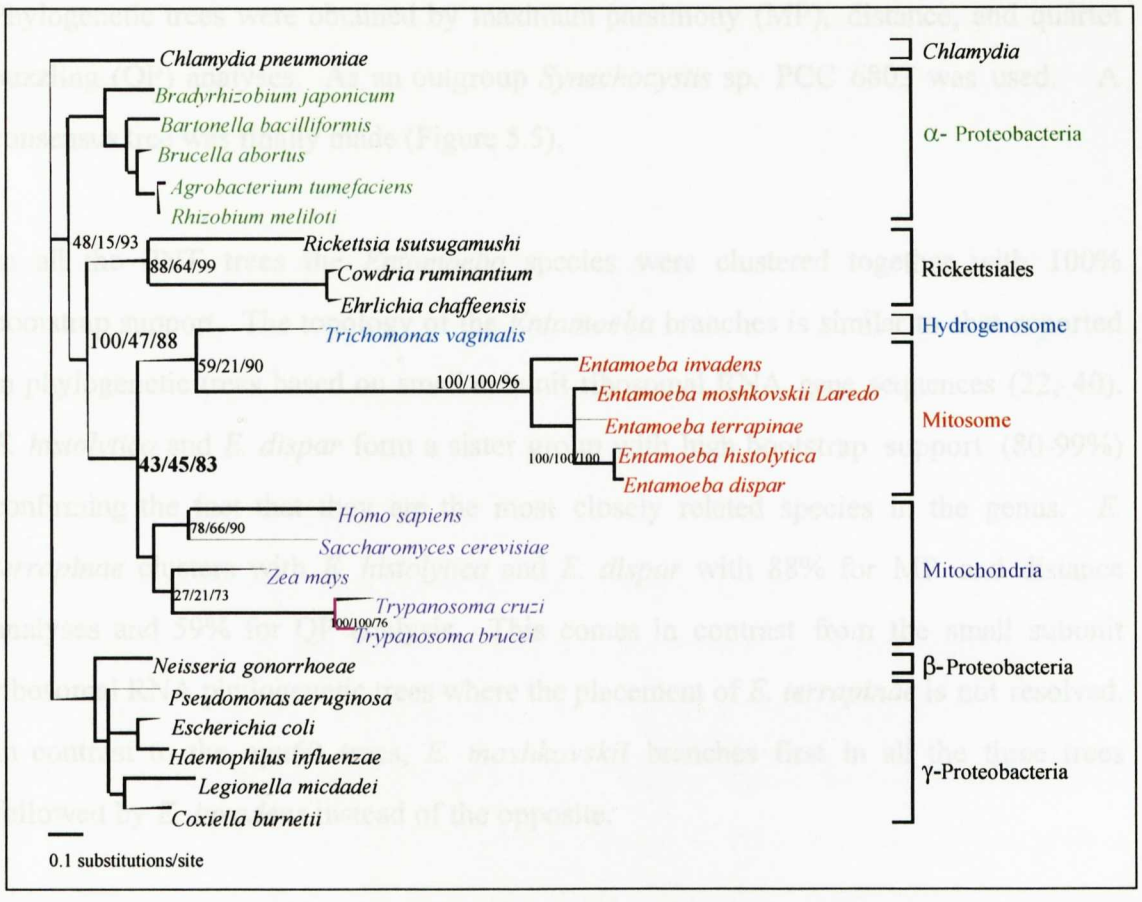


Figure 5.4. Phylogenetic relationships of the cpn60 homologues.

The numbers at nodes of the tree correspond to the bootstrap values for distance, maximum parsimony and quartet puzzling analyses from left to right respectively. Values are shown only for certain important nodes. The scale bar indicates estimated sequence divergence per unit branch length. The node that defines the *Entamoeba* species/mitochondrial relationship is given in bold.

5.2.3 Phylogenetic analysis of the PNT genes

Sequences of the partial β PNT fragment of *E. histolytica*, *E. dispar*, *E. invadens*, *E. terrapinae* and *E. moshkovskii* Laredo were aligned using the CLUSTAL W programme with amino acid sequences of the same β PNT fragment of mitochondrial and related bacterial PNT homologues from 11 species obtained from GenBank™ (Appendix D) and subjected to a variety of phylogenetic analyses. 142 alignment positions were included in the final data set out of the total 151 obtained after sequencing analysis.

Phylogenetic trees were obtained by maximum parsimony (MP), distance, and quartet puzzling (QP) analyses. As an outgroup *Synechocystis* sp. PCC 6803 was used. A consensus tree was finally made (Figure 5.5).

In all the PNT trees the *Entamoeba* species were clustered together with 100% bootstrap support. The topology of the *Entamoeba* branches is similar to that reported in phylogenetic trees based on small subunit ribosomal RNA gene sequences (22, 40). *E. histolytica* and *E. dispar* form a sister group with high bootstrap support (80-99%) confirming the fact that they are the most closely related species in the genus. *E. terrapinae* clusters with *E. histolytica* and *E. dispar* with 88% for MP and distance analyses and 59% for QP analysis. This comes in contrast from the small subunit ribosomal RNA phylogenetic trees where the placement of *E. terrapinae* is not resolved. In contrast to the cpn60 trees, *E. moshkovskii* branches first in all the three trees followed by *E. invadens* instead of the opposite.

In the QP tree the *Entamoeba* species cluster together with the mitochondria and the α-Proteobacteria group in relatively good bootstrap support of 61%. For the other trees the support for this grouping was very low at 21% and 8% for MP and distance analyses. With the exception of the distance tree where all the mitochondriates are grouped together (*P. falciparum, E. tenella* and *H. sapiens*) in the rest of the trees either the *P. falciparum* or *H. sapiens* was always branching somewhere else in the tree. The mitochondria and the α-Proteobacteria grouped together although this was not strongly supported. The bootstrap values for the MP, distance and QP were 8%, 17% and 52% respectively. With these values of bootstrap support for the *Entamoeba*, mitochondria and α-Proteobacteria grouping it is difficult to support the theory that the mitochondria have descended from the α-Proteobacteria and further more that the mitosome has mitochondrial origins.

In all three trees the *Rhodospirillum rubrum* forms a sister group with *Mycobacterium tuberculosis* with good bootstrap support of 78%, 88% and 68% for MP, distance and QP respectively.

The γ-Proteobacteria clustered together with good bootstrap support for distance and QP analyses of 81% and 72%. MP analysis was low at 37%. This group was clustered together with *Neisseria meningitidis* in high bootstrap support, which belongs to the β-Proteobacteria.

The fact that different groups of bacteria cluster together where in other trees normally they wouldn't means that the PNT does not give reliable trees and information. It seems that the length of the β fragment of the PNT gene used for the analysis was too small and the sequences were too similar to accurately resolve the relationships among the organisms. Also there are still a limited number of PNT sequences available from other organisms and especially from other amitochondriates. More sequences will be needed in order to fully resolve the phylogenetic relationships and the branching order of the PNT genes.

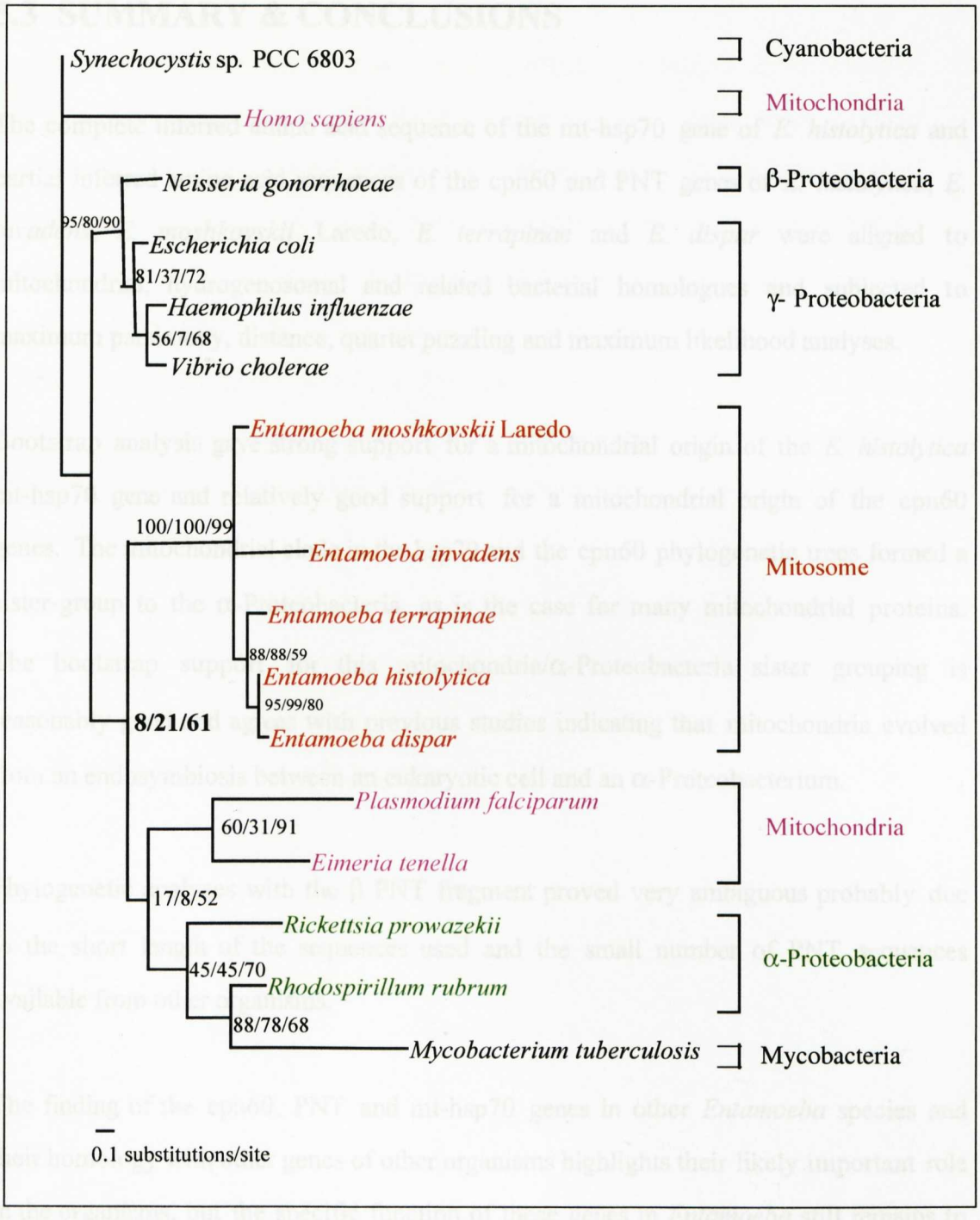


Figure 5.5. Phylogenetic relationships of the PNT homologues.

The numbers at nodes of the tree correspond to the bootstrap values for distance, maximum parsimony and quartet puzzling analyses from left to right respectively. Values are shown only for certain important nodes. The scale bar indicates estimated sequence divergence per unit branch length. The node that defines the *Entamoeba* species/mitochondrial relationship is given in bold.

5.3 SUMMARY & CONCLUSIONS

The complete inferred amino acid sequence of the mt-hsp70 gene of *E. histolytica* and partial inferred amino acid sequences of the cpn60 and PNT genes of *E. histolytica*, *E. invadens*, *E. moshkovskii* Laredo, *E. terrapinae* and *E. dispar* were aligned to mitochondrial, hydrogenosomal and related bacterial homologues and subjected to maximum parsimony, distance, quartet puzzling and maximum likelihood analyses.

Bootstrap analysis gave strong support for a mitochondrial origin of the *E. histolytica* mt-hsp70 gene and relatively good support for a mitochondrial origin of the cpn60 genes. The mitochondrial clade in the hsp70 and the cpn60 phylogenetic trees formed a sister-group to the α-Proteobacteria, as is the case for many mitochondrial proteins. The bootstrap support for this mitochondria/α-Proteobacteria sister grouping is reasonably good and agrees with previous studies indicating that mitochondria evolved from an endosymbiosis between an eukaryotic cell and an α-Proteobacterium.

Phylogenetic analyses with the β PNT fragment proved very ambiguous probably due to the short length of the sequences used and the small number of PNT sequences available from other organisms.

The finding of the cpn60, PNT and mt-hsp70 genes in other *Entamoeba* species and their homology with other genes of other organisms highlights their likely important role in the organisms, but the specific function of these genes in *Entamoeba* still remains to be discovered.

The identification of mitochondrion specific proteins such as cpn60, PNT, and the mt-hsp70 in the other species of the genus *Entamoeba* confirms that they are not surviving relics of pre-mitochondrial evolution, but that their common ancestors once harboured mitochondria or an endosymbiont related to the progenitor of mitochondria.

The presence of these mitochondrial genes in *E. dispar*, *E. invadens*, *E. moshkovskii* and *E. terrapinae* supports the idea that the mitosome may also be present in these species and that mitochondrial loss is not a recent event. Preliminary Western blotting experiments indicate that the mitosome might be present in *E. moshkovskii* (results not shown). More information is needed about the rest of the *Entamoeba* species. If the *Entamoeba* lineage has acquired these genes through lateral transfer recently then these genes would not have been expected to be present in all the species. It is still possible that the organelle may have been lost in some of these organisms as a result of adaptation to a different living environment and the proteins are now cytoplasmic.

CHAPTER 6

EPILOGUE

6.1 UPDATE ON THE MT-HSP70 GENE COPY NUMBER

Since the results about the gene copy number and linkage groups of the mt-hps70 gene of *E. histolytica* were obtained (Chapter 3) new information has emerged. Another mt-hsp70 sequence of *E. histolytica* HM-1:IMSS has been submitted into the GenBank with accession number AB056658. Sequence comparison with the mt-hsp70 found in this study revealed some different nucleotides between the two sequences. This could indicate that there are either two different variants of the mt-hsp70 gene of *E. histolytica* or that that there is a strain difference as most of our mt-hsp70 gene was obtained after screening a cDNA library of the SFL-3 strain and not the HM-1:IMSS. After searching the TIGR site of the *E. histolytica* genome project (http://www.tigr.org/tdb/e2kl/ehal/) with these two sequences it was discovered that both are present in the database. This means that there are two different versions of the mt-hsp70 gene that co-exist in the HM-1:IMSS strain. It is also very possible that both variants are present in other *E. histolytica* strains. The two bands in the Northern blots quite possibly correspond also to the two mt-hsp70 gene variants.

This new information confirms the results of the Southern blotting and the PFGE. The two bands in the Southern blot would thus correspond to the two variants of the mt-hsp70 gene. In the PFGE, two linkage groups were found each one corresponding to one of the two variants. So, the two copies of the mt-hsp70 gene correspond to two distinct loci.

6.2 THE *ENTAMOEBA HISTOLYTICA* GENOME PROJECT

As is the case for several other parasites, an *E. histolytica* genome project is currently underway through the combined efforts of The Institute for Genomics Research and the Sanger Centre. The target of the genome project is to obtain approximately 99% of the DNA sequence of this organism in three years.

This is very important not only because of the potential medical benefits for the estimated 40 million people infected per year, but also because of the evolutionary information it will produce. The data should provide insight into the correct placement of *E. histolytica* in phylogenetic trees and help clarify the mitochondrial history of the organism too.

One goal of current research on amitochondriate organisms such as *E. histolytica* (and others previously thought to be primitive amitochondriate) is to establish beyond doubt whether their current amitochondriate state represents an ancestral, "premitochondriate" feature or whether all of these organisms are descended from ancestors that once harboured mitochondria and lost them secondarily.

The one-by-one search for genes in these amitochondriate organisms that could have been part of the genome of the ancestral mitochondrion but subsequently transferred to the nuclear genome, such as the mt-hsp70 gene found in this study and the cpn60 and PNT found in previous studies (68), has been the most commonly used approach to date. As the genome project continues and more genes are sequenced more information is revealed about the evolution of the organism. Specifically, a transporter, the cpn10 and serine dehydratase proteins, which are thought to be linked with mitochondria, were identified in the *E. histolytica* genome project database in this study and compared with other genes of mitochondrial and hydrogenosomal origin (Chapter 3, Section 3.3).

The finding of more such genes in combination with sequence comparisons and phylogenetic reconstructions such as the ones done for the mt-hsp70, cpn60 and PNT of *E. histolytica* will provide additional evidence for the post mitochondrial nature of the organism (Chapter 5).

More specifically to the mitosome project, new data should appear that relate to this organelle. When the genome of *E. histolytica* is sequenced a clearer image of the protein composition of the mitosome should emerge, although its role in the cell will be harder to determine (159). The mitosome may have lost its bioenergetic functions but it seems to share a common evolutionary origin with the hydrogenosome and the mitochondrion. Studies on newly identified genes of mitochondrial origin should be carried out in order to find out which ones are really targeted to the mitosome and therefore further support the mitochondrial origin of this organelle.

Comparison of the genes found in *E. histolytica* with homologues already sequenced in other organisms may not only determine the function of these genes but also reveal information about the evolution of the organism and its relationship to prokaryotic and to other eukaryotic organisms. The knowledge of such phylogenetic relationships is essential to understanding the evolutionary history of *E. histolytica*.

The genome data will provide a large library of sequences that will speed up genetic studies in the organism. It will help researchers to find new genes by comparison with those already deposited in genome databases. This will add greatly to our knowledge of the metabolic pathways and cell biology of the organism and also possibly identify additional genes that influence pathogenicity. Other genes of interest may code for biological molecules that are potential candidates for diagnosis, therapy and drug targets, or that could lead to the development of vaccines. It is possible that some genes may encode functions unique to *E. histolytica*. The identification of these genes together with

bioinformatic analyses should be able to help in the development of new antiamoebic drugs that are targeted to key specific elements of the amoeba.

Little is known about the DNA organisation and gene expression in *E. histolytica*. Variations have been observed in the DNA content and organisation in the same strain under different growth conditions in culture (160) and tissue invasion (161). No telomeres or centromeres have been cloned and the presence of introns in genes initially was thought to be very rare (162-164). Recent evidence has revealed that introns are more common than was previously suggested (165). Analysis of the 5′ and 3′ untranslated regions of structural genes shows they are unusually short (138).

The genome project should shed some light on these and other aspects of the structure and organisation of *E. histolytica* genome. More specifically information should emerge about its size, ploidy, number of chromosomes and the DNA content of the nucleus, as there is not a universally accepted value for any of these. Recently, a linkage map has been established with 14 linkage groups and a likely functional ploidy of at least 4 (147).

6.3 LOCALISATION OF THE MITOCHONDRIAL PROTEINS

Recent studies have localised the cpn60 protein of *E. histolytica* to a previously undescribed organelle of possible mitochondrial origin named the mitosome or crypton (82, 83). It has been shown that the removal of the first 15 (82) or 13 (83) amino acids of the protein leads to a clear enrichment of the recombinant cpn60 in the cytoplasm. The addition of a functional mitochondrial targeting signal from *Trypanosoma cruzi* hsp70 to the truncated protein led to the accumulation of the recombinant cpn60 into the mitosome (82). The hsp70 proteins participate in a number of cellular processes such as the correct folding of mature proteins in the cytoplasm without being

components of the functional assembled structures. They also facilitate the translocation of newly synthesised polypeptides across membranes (into the mitochondrion and the endoplasmic reticulum) (95, 96). If the mt-hsp70 protein is localised in the mitosome this would be further evidence that the mitosome is a mitochondrial remnant. The investigation of the cellular localisation of the mt-hsp70 by either inserting a complete tagged mt-hsp70 coding region into an expression vector or by replacing the deleted 15 amino acids of cpn60 with the putative mt-hsp70 targeting sequence is an important task.

The cloning of the complete mt-hsp70 gene into the R.CPN60 vector (construct A; (82)) was unsuccessful so an adaptor containing the putative mitochondrial putative targeting sequence of the mt-hsp70 gene was designed and then successfully cloned into the R.CPN60 NtΔ plasmid (82). The plasmid was then successfully transfected into *E. histolytica* trophozoites (Figure 6.1). The transfected trophozoites were either cell fractionated or used for immunofluorescence microscopy.

Neither the cell fractionation nor the immunofluorescence gave any conclusive results. There was no obvious enrichment of the protein in the mitosomal fractions after cell fractionation or any subcellular localisation in the *E. histolytica* cells. They were two problems encountered in these experiments. The cell fractionation failed to separate the cell components and therefore the same cpn60 protein was present in all the fractions. The recombinant protein could not be detected in the Western blots. Even if the recombinant protein was not targeted to the mitosome, the native cpn60 should have been found in that fraction only. The absence of recombinant cpn60 could have been a result of homologous recombination in the vector, which removed the cpn60 gene. If the cpn60 gene was not present in the vector this would explain why the fluorescence microscopy did not work. There would be no tagged cpn60 protein for the antibodies to bind to.

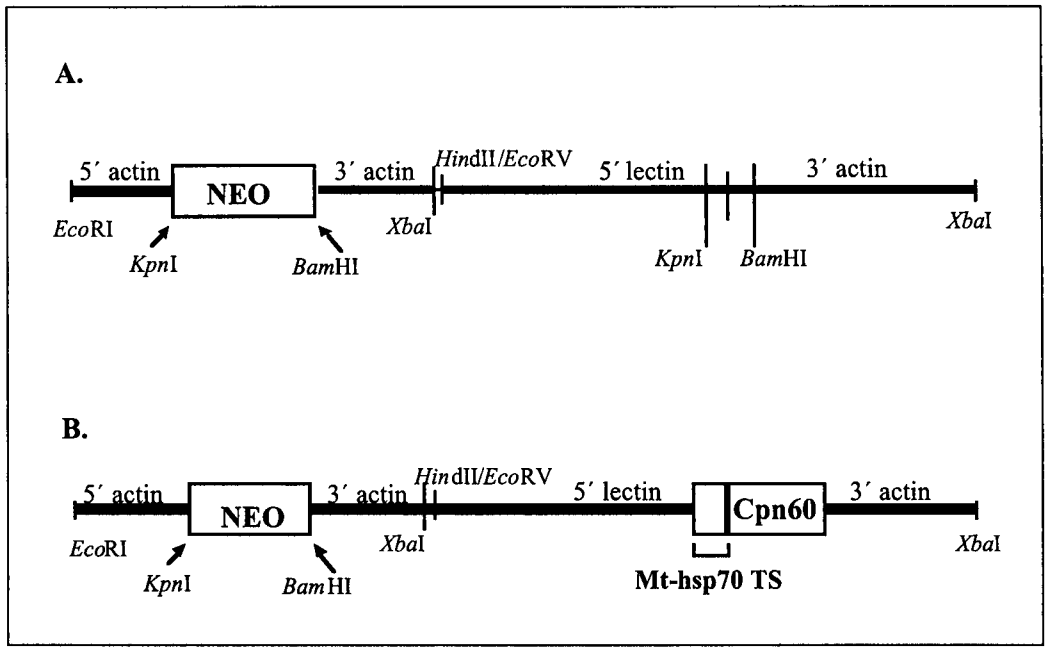


Figure 6.1. Schematic representation of the EhNEOplus vector constructs.

A. EhNEOplus vector.

This vector is a pBS derived plasmid containing the NEO coding region from ATG to the stop codon, flanked by untranslated 5′ and 3′ sequences of an *E. histolytica* actin gene (5′ actin & 3′ actin). The rest of the vector was inserted to the pEhNEO and contains the cloning site, which has *Kpn*I and *Bam*HI as restriction sites. The cloning site is flanked by 485 bp of untranslated 5′ sequence of an *E. histolytica* lectin gene (5′ lectin) and 600 bp of the untranslated 3′ sequence of an *E. histolytica* actin gene (3′ actin); (124, 125).

B. R.CPN60-NtΔ-EhmtHSP70 TS plasmid.

The coding region of the *E. histolytica* cpn60 gene without its first 15 amino acids was inserted into the EhNEOplus vector by using the *Kpn*I and the *Bam*HI sites (82). The mt-hsp70 targeting sequence was then inserted to the cpn60 gene to create the plasmid.

Preliminary Western blotting experiments were also performed on *E. invadens* and *E. moshkovskii* to see if the mitosome is present in these organisms. In *E. moshkovskii* a faint band seemed to be present in the mitosomal fraction at approximately 60 kDa, indicating that the mitosome might still be present in this organisms (results not shown). These results clearly need to be repeated in the future. More information is needed about the rest of the *Entamoeba* species. If the *Entamoeba* lineage has acquired these genes through lateral transfer recently then the genes could not have been expected to be present in all the species. It is still possible that the organelle may have been lost in

some of these organisms as a result of adaptation to a different living environment and the proteins are now cytoplasmic. Ideally specific antibodies should be raised for the *E. invadens*, *E. moshkovskii*, *E. terrapinae* and *E. dispar* cpn60s to find if the mitosomal compartment is still present in these species or it has been lost during evolution of the organisms.

6.4 FURTHER DIRECTIONS FOR THIS STUDY

In this study it has been shown that high temperature seems to induce both the mRNAs of the mt-hsp70 gene. Further expression studies should be performed in order to investigate the gene's response to a variety of other conditions. Nutritional factors are known to affect gene expression in other systems. The amoebas could be starved of sugar or/and serum, for example. It is known that the absence of serum in the medium blocks the *E. histolytica* cell cycle (166). The expression of genes could therefore be studied over the cell cycle by starving the organism of serum and then reintroducing it to synchronise the cells. Environmental factors such as pH change and elevated O_2 levels could also be studied. Elevated ethanol and acetate levels, two metabolic end products, may also have an effect on the expression of the mt-hsp70 and other genes.

The above experiments could be performed using DNA microarrays. Microarrays are an orderly arrangement of DNA samples that provides a medium for matching known and unknown DNA samples based on base-pairing rules and automating the process of identifying the unknowns. Usually DNA chips are fabricated by high-speed robotics, on glass or nylon substrates, for which probes with known identity are used to determine complementary binding, thus allowing massively parallel gene expression and gene discovery studies (167).

After the discovery of additional mitosome-related genes during the genome project, a more holistic study can be undertaken to evaluate the patterns of expression in different groups of genes including the ones studied in this project. Expression can be examined during different heat shock regimens and also under a variety of other conditions such as the ones mentioned above.

Unfortunately because we do not yet know the function of the mitosome we cannot predict what may initiate the expression of a mitosomal protein gene. The only way forward is to try different conditions and examine their effect on transcription of the genes. Hopefully the genome sequence will give us an insight into the organelle's function.

The experiments attempting to localise mt-hsp70 protein of *E. histolytica* to the mitosome need to be repeated, as they were not successful in this study.

The role of the amino-terminal extension could be investigated by deletion or mutagenesis to investigate whether it is required for the import of the protein into the mitosome. Localisation of the other proteins of putative mitochondrial origin should be performed not only in *E. histolytica* but also in the other *Entamoeba* species to see if a mitochondrial-type import mechanism is functional as was shown for the cpn60 (82). This would give further evidence for the mitochondrial origin of the mitosome. A transfection system does not exist for *Entamoeba* species (except *E. histolytica* and *E. dispar*). So a way to localise the other mitochondrial-type proteins would be to fractionate the cells and use antibodies that can detect the wild type protein. A heterologous antibody could be used as long as it cross-reacts with the protein. Ideally, epitopes contained in the proteins could be identified that will then be recognised by antibodies raised to the peptides. Immunofluorescence studies should also be used in order to localise the protein in the trophozoites.

Partial sequencing of the other genomes of the genus *Entamoeba* would allow comparisons between the species and could possibly point out important genetic differences between them. This would be most useful for *E. dispar* as it has identical morphology to *E. histolytica* and is also found in humans in quite high prevalence rates but without causing any disease. Phylogenetic studies have shown that they are sister species, very closely related genetically. Any differences between these two species could point out potential virulence genes in *E. histolytica*. A small number of genes that are specific to *E. histolytica* and are lacking in *E. dispar* have already been identified, such as the recently characterised membrane-bound cysteine proteinase and the *ariel* gene family (168).

The vast amount of sequence information that is currently being accumulated by the parasite genome projects creates the need for molecular techniques that can replace the traditional but time-consuming chromosomal gene 'knockout' methodologies, enabling rapid screening of identified open reading frames for functional significance (169). In the case of *E. histolytica* where, perhaps because of its unusual gene organisation, gene knockout and targeted gene deletion experiments have not been successful so far, the need for such methods is even greater to allow the further manipulation of the parasite.

Recently, a PNA oligomer was used as an antisense agent for the down-regulation of expression of genes *in vivo* in *E. histolytica* (170). PNA oligomers are polynucleotide analogues in which the entire phosphate-sugar backbone has been replaced by an isomorphic pseudo-peptidic chain to which the four bases are linked. They are uncharged, stable to cleavage by enzymes, and it has been shown they can down-regulate gene expression when used at micro-molar concentrations (171). RNA antisense methods have also been used for blocking gene expression in *E. histolytica* (172, 173). Another approach that could be used for the functional analysis of the genome is RNA-mediated interference (RNAi). RNAi transiently inhibits the activity of a gene by the introduction of double-stranded RNA of sequence specific to the targeted

gene. The open reading frames of two chromosomes of *Caenorhabditis elegans* have been analysed successfully with this method (174, 175), and RNAi was also used in generating knockout phenotypes for several genes of *Trypanosoma brucei* (176).

As the genome project is now underway, more information will emerge about the genome of *E. histolytica* that will help us understand this parasite and answer many questions about its structure, function and evolution.

.

REFERENCES

1. Anonymous. (1997) WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. *Epidemiol Bull PAHO* **18:** 13-14.

2. Ravdin JI. (1995) Amebiasis (Review). *Clin Infect Dis* **20:** 1453-1466.

3. Petri WA. (1996) Recent advances in amebiasis. *Crit Rev Clin Lab Sci* **33:** 1-37.

4. Cavalier-Smith T. (1993) Kingdom protozoa and its 18 phyla. *Microbiol Rev* **57:** 953-994.

5. Istre GR, Kreiss K, Hopkins R, et al. (1982) An outbreak of amebiasis spread by colonic irrigation at a chiropractic clinic. *N Engl J Med* **307:** 339-342.

6. Martínez-Palomo A, Ruíz-Palacios G. (1990) Amebiasis. In: Warren KS, Mahmoud AAF (eds.) *Tropical and Geographical Medicine.* McGraw Hill, New York, pp. 327-344.

7. Ravdin JI, Jackson TFHG, Petri WA. (1990) Association of serum antibodies to adherence lectin with invasive amebiasis and asymptomatic infection with pathogenic *Entamoeba histolytica. J Infect Dis* **162:** 768-772.

8. Haque R, Kress K, Wood S, et al. (1993) Diagnosis of pathogenic *Entamoeba histolytica* infection using a stool ELISA based on monoclonal antibodies to the galactose-specific adhesin. *J Infect Dis* **167:** 247-249.

9. Abd-Alla MD, Jackson TFHG, Gathiram V, El-Hawey AM, Ravdin JI. (1993) Differentiation of pathogenic *Entamoeba histolytica* infections from nonpathogenic infections by detection of galactose inhibitable adherence protein antigen in sera and feces. *J Clin Microbiol* **31:** 2845-2850.

10. Clark CG, Diamond LS. (1991) Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct. *Mol Biochem Parasitol* **47:** 297-302.

11. Tannich E, Burchard GD. (1991) Differentiation of pathogenic from nonpathogenic *Entamoeba histolytica* by restriction fragment analysis of a single gene amplified *in vitro. J Clin Microbiol* **29:** 250-255.

12. Tachibana H, Kobayashi S, Kato Y, Nagakura K, Kaneda Y, Takeuchi T. (1990) Identification of a pathogenic isolate-specific 30,000-M_r antigen of *Entamoeba histolytica* by using a monoclonal antibody. *Infect Immun* **58:** 955-960.

13. Tachibana H, Ihara S, Kobayashi S, Kaneda Y, Takeuchi T, Watanabe Y. (1991) Differences in genomic DNA sequences between pathogenic and nonpathogenic isolates of *Entamoeba histolytica* identified by polymerase chain reaction. *J Clin Microbiol* **29:** 2234-2239.

14. Reed SL. (2000) Clinical manifestations and diagnosis. In: Ravdin JI (ed.) *Amebiasis.* Imperial College Press, London, pp. 113-126.

15. Martínez-Palomo A. (1982) The biology of *Entamoeba histolytica.* In: Brown KN (ed.) *Tropical Medicine Research Studies Series.* John Wiley & Sons, Chichester.

16. Dobell C. (1919) The amoebae living in man. A zoological monograph. J. Bale, Sons, and Danielson, London.

17. Diamond LS, Clark CG. (1993) A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. *J Euk Microbiol* **40:** 340-344.

18. Clark CG, Espinosa Cantellano M, Bhattacharya A. (2000) *Entamoeba histolytica*:An overview of the biology of the organism. In: Ravdin JI (ed.) *Amebiasis. Series on Tropical Medicine: Science and Practice- Vol. 2.* Imperial College Press, London, pp. 1-45.

19. Lösch F. (1875) Massenhafte Entwicklung von Amöben im Dickdarm. *Virchow's Archiv* **65:** 196-211.

20. Schaudinn F. (1903) Untersuchungen über die Fortpflanzung einiger Rhizopoden. (Vorläufige Mittheilung). *Arbeit Kaiserlichen Gesundheitsamte* **19:** 547-576.

21. Walker EL, Sellards AW. (1913) Experimental entamoebic dysentery. *Phillipine J Sci B Trop Med* **8:** 253-331.

22. Clark CG. (2000) The evolution of *Entamoeba histolytica*, a cautionary tale. *Res Microbiol* **151:** 599-603.

23. von Prowazek S. (1912) Weiterer Beitrag zur Kenntnis der Entamöben. *Arch Protistenk* **26:** 241-249.

24. Burrows RB. (1957) *Endamoeba hartmanni. Am J Hyg* **65:** 172-188.

25. Brumpt E. (1925) Étude sommaire de l' "*Entamoeba dispar*" n. sp. Amibe à kystes quadrinucléés, parasite de l'homme. *Bull Acad Méd (Paris)* **94:** 943-952.

26. Martínez-Palomo A, González-Robles A, De la Torre M. (1973) Selective agglutination of pathogenic strains of *Entamoeba histolytica* induced con A. *Nature New Biol* **245:** 186-187.

27. Sargeaunt PG, Williams JE, Grene JD. (1978) The differentiation of invasive and non-invasive *Entamoeba histolytica* by isoenzyme electrophoresis. *Trans R Soc Trop Med Hyg* **72:** 519-521.

28. Sargeaunt PG, Williams JE, Bhojnani R, Kumate J, Jimenez E. (1982) A review of isoenzyme characterization of *Entamoeba histolytica* with particular reference to pathogenic and non-pathogenic stocks isolated in Mexico. *Arch Invest Méd (Méx)* **13 (suppl 3):** 89-94.

29. Sargeaunt PG. (1987) The reliability of *Entamoeba histolytica* zymodemes in clinical diagnosis. *Parasitol Today* **3:** 40-43.

30. Strachan WD, Spice WM, Chiodini PL, Moody AH, Ackers JP. (1988) Immunological differentiation of pathogenic and non-pathogenic isolates of *Entamoeba histolytica. Lancet* **i:** 561-563.

31. Tannich E, Horstmann RD, Knobloch J, Arnold HH. (1989) Genomic DNA differences between pathogenic and nonpathogenic *Entamoeba histolytica. Proc Natl Acad Sci USA* **86:** 5118-5122.

32. Weinbach EC, Diamond LS. (1974) *Entamoeba histolytica*: I. Aerobic metabolism. *Exp Parasitol* **35:** 232-243.

33. Weinbach EC, Clagget CE, Takeuchi T, Diamond LS. (1978) Biological oxidation and flavoprotein analysis in *Entamoeba histolytica. Arch Invest Med (Suppl 1)* **9:** 89-98.

34. Martínez-Palomo A. (1993) Parasitic amebas of the intestinal tract. In: Kreier JP, Baker JR (eds.) *Parasitic protozoa.* Academic Press, San Diego, CA, pp. 65-141.

35. Weinbach EC. (1988) Metabolism in *Entamoeba histolytica.* In: Ravdin JI (ed.) *Amebiasis: Human Infection by Entamoeba histolytica.* John Wiley & Sons, New York, p. 69.

36. Weinbach EC. (1981) Biochemistry of enteric parasitic protozoa. *Trends Biochem Sci*: 254-257.

37. Bruckner D. (1992) Amebiasis. *Clin Microbiol Rev* **5:** 356-369.

38. Clark CG, Diamond LS. (1997) Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. *J Euk Microbiol* **44:** 142-154.

39. Clark CG, Silberman JD, Diamond LS, Sogin ML. (1998) Molecular systematics of the intestinal amoebae. In: Coombs GH, Vickerman K, Sleigh MA, Warren A (eds.) *Evolutionary Relationships Among Protozoa.* Kluwer Academic Publishers, Dortrecht, The Netherlands, pp. 169-180.

40. Silberman JD, Clark CG, Diamond LS, Sogin ML. (1999) Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small subunit ribosomal RNA gene sequence analysis. *Mol Biol Evol* **16:** 1740-1751.

41. Chacín-Bonilla L. (1992) *Entamoeba polecki*: human infections in Venezuela. *Trans R Soc Trop Med Hyg* **86:** 634.

42. Sargeaunt PG, Patrick S, O'Keefe D. (1992) Human infections of *Entamoeba chattoni* masquerade as *Entamoeba histolytica. Trans R Soc Trop Med Hyg* **86:** 633-634.

43. Verweij JJ, Polderman AM, Clark CG. (2001) Genetic variation among human isolates of uninucleated cyst-producing *Entamoeba* species. *J Clin Microbiol* **39:** 1644-1646.

44. Markell EK, John DT, Krotoski WA. (1999) Lumen dwelling protozoa. In: *Markell and Voge's Medical Parasitology.* W.B. Saunders Company, pp. 24-55.

45. Dreyer DA. (1961) Growth of a strain of *Entamoeba histolytica* at room temperature. *Tex Rep Biol Med* **19:** 393-396.

46. Haque R, Ali IKM, Clark CG, Petri WA, Jr. (1998) A case report of *Entamoeba moshkovskii* infection in a Bangladeshi child. *Parasitol Int* **47:** 201-202.

47. Richards CS, Goldman M, Cannon LT. (1966) Cultivation of *Entamoeba histolytica* and *Entamoeba histolytica*-like strains at reduced temperature and behavior of the amebae in diluted media. *Am J Trop Med Hyg* **15:** 648-655.

48. Goldman M. (1960) *Entamoeba histolytica*-like amoebae occuring in man. *Bull WHO* **40**.

49. Sanders EP, Cleveland LR. (1930) The morphology and life-cycle of *Entamoeba terrapinae* spec. nov., from the terrapin *Chrysemys elegans*. *Arch Protistenk* **70:** 267-272.

50. Neal RA. (1966) Experimental studies on *Entamoeba* with reference to speciation. *Adv Parasitol* **4:** 1-51.

51. Reeves RE. (1984) Metabolism of *Entamoeba histolytica* Schaudinn, 1903. *Adv Parasitol* **23:** 105-142.

52. Martínez-Palomo A. (1986) Biology of *Entamoeba histolytica*. In: Martínez-Palomo A (ed.) *Amebiasis*. Elsevier, Amsterdam, pp. 12-43.

53. Müller M. (1988) Energy metabolism of protozoa without mitochondria. *Annu Rev Microbiol* **42:** 465-488.

54. Bakker-Grunwald T, Wöstmann C. (1993) *Entamoeba histolytica* as a model for the primitive eukaryotic cell. *Parasitol Today* **9:** 27-31.

55. Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T. (1992) Early divergences in the evolution of eukarytoes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. *J Mol Evol* **36:** 380-388.

56. Shirakura T, Hashimoto T, Nakamura Y, et al. (1994) Phylogenetic place of a mitochondria-lacking protozoan, *Entamoeba histolytica*, inferred from amino acid sequences of elongation factor II. *Jpn J Genet* **69:** 119-135.

57. Edlind TD, Li J, Visvesvara GS, Vodkin MH, McLaughlin GL, Katiyar SK. (1996) Phylogenetic analysis of β-tubulin sequences from amitochondrial protozoa. *Mol Phylogenet Evol* **5:** 359-367.

58. Sogin ML. (1991) Early evolution and the origin of eukaryotes. *Curr Opin Genet Dev* **1:** 457-463.

59. Boore JL, Fuerstenberg SI. (1999) *Entamoeba histolytica*: a derived, mitochondriate eukaryote? *Trends Microbiol* **7:** 426-428.

60. Palmer JD. (1997) The mitochondrion that time forgot. *Nature* **387:** 454-455.

61. Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. (1985) Mitochondrial origins. *Proc Natl Acad Sci USA* **82:** 4443-4447.

62. Gray MW, Cedergren R, Abel Y, Sankoff D. (1989) On the evolutionary origin of the plant mitochondrion and its genome. *Proc Natl Acad Sci USA* **86:** 2267-2271.

63. Gray MW, Burger G, Lang BF. (1999) Mitochondrial evolution. *Science* **283:** 1476-1481.

64. Schwartz RM, Dayhoff MO. (1978) Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts. *Science* **199:** 395-403.

65. Viale AM, Arakaki AK. (1994) The chaperone connection to the origins of the eukaryotic organelles. *FEBS letters* **341:** 146-151.

66. Andersson SG, Zomorodipour A, Andersson JO, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396:** 133-140.

67. Clark CG, Roger AJ. (1995) Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc Natl Acad Sci USA* **92:** 6518-6521.

68. Roger AJ, Clark CG, Doolittle WF. (1996) A possible mitochondrial gene in the early-branching amitochondriate protist *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* **93:** 14618-14622.

69. Germot A, Philippe H, Le Guyader H. (1996) Presence of a mitochondrial-type 70-kDa heat shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. *Proc Natl Acad Sci USA* **93:** 14614-14617.

70. Hashimoto T, Sanchez LB, Shirakura T, Müller M, Hasegawa M. (1998) Secondary absence of mitochondria in *Giardia lamblia* and *Trichomonas vaginalis* revealed by valyl-tRNA synthetase phylogeny. *Proc Natl Acad Sci USA* **95:** 6860-6865.

71. Hirt RP, Healy B, Vossbrinck CR, Canning EU, Embley TM. (1997) A mitochondrial Hsp70 orthologue in *Vairimorpha necatrix*: molecular evidence that microsporidia once contained mitochondria. *Curr Biol* **7:** 995-998.

72. Germot A, Philippe H, Le Guyader H. (1997) Evidence for loss of mitochondria in Microsporidia from a mitochondrial-type HSP70 in *Nosema locustae*. *Mol Biochem Parasitol* **87:** 159-168.

73. Roger AJ, Svärd SG, Tovar J, et al. (1998) A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: Evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. *Proc Natl Acad Sci USA* **95:** 229-234.

74. Sogin ML. (1997) History assignment: when was the mitochondrion founded? *Curr Opin Genet Dev* **7:** 792-799.

75. Horner DS, Hirt RP, Kilvington S, Lloyd D, Embley TM. (1996) Molecular data suggest an early acquisition of the mitochondrion endosymbiont. *Proc R Soc Lond B Biol Sci* **263:** 1053-1059.

76. Müller M. (1993) The hydrogenosome. *J Gen Microbiol* **139:** 2879-2889.

77. Plümper EP, Bradley PJ, Johnson PJ. (1998) Implications of protein import on the origin of hydrogenosomes. *Protist* **149:** 303-311.

78. Embley TM, Finlay BJ, Dyal PL, Hirt RP, Wilkinson M, Williams AG. (1995) Multiple origins of anaerobic ciliates with hydrogenosomes within the radiation of aerobic ciliates. *Proc R Soc Lond B* **262:** 87-93.

79. Martin M, Müller M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature* **392:** 37-41.

80. Dyall SD, Koehler CM, Delgadillo-Correa MG, et al. (2000) Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. *Mol Cell Biol* **20:** 2488-2497.

81. Yu Y, Samuelson J. (1994) Primary structure of an *Entamoeba histolytica* nicotinamide nucleotide transhydrogenase. *Mol Biochem Parasitol* **68:** 323-328.

82. Tovar J, Fischer A, Clark CG. (1999) The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. *Mol Microbiol* **32:** 1013-1021.

83. Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J. (1999) Hsp60 is targeted to a cryptic mitochondrion-derived organelle ("Crypton") in

the microaerophilic protozoan parasite *Entamoeba histolytica*. *Mol Cell Biol* **19:** 2198-2205.

84. Ghosh S, Field J, Rogers R, Hickman M, Samuelson J. (2000) The *Entamoeba histolytica* mitochondrion-derived organelle (crypton) contains double-stranded DNA and appears to be bound by a double membrane. *Infect and Immun* **68:** 4319-4322.

85. Causey D. (1925) Mitochondria and Golgi bodies in *Endamoeba gingivalis* (Gros) Brumpt. *Univ Calif Publ Zool* **28:** 1-18.

86. Ray HN, Sen Gupta PC. (1954) A cytochemical study of *Entamoeba histolytica*. *J Indian Med Assoc* **23:** 529-533.

87. Lindquist S, Craig EA. (1988) The heat-shock proteins. *Annu Rev Genet* **22:** 631-677.

88. Stuart RA, Cyr DM, Craig EA, Neupert W. (1994) Mitochondrial molecular chaperones: their role in protein translocation. *Trends Biochem Sci* **19:** 87-92.

89. Morimoto RJ, Kline MP, Bimston DN, Cotto JJ. (1997) The heat-shock response: Regulation and function of heat-shock proteins and molecular chaperones. *Essays in Biochemistry* **32:** 17-29.

90. Kasting JF. (1993) Earth's early atmosphere. *Science* **259:** 920-926.

91. Ellis RJ, Hemmingsen SM. (1989) Molecular chaperones:proteins essential for the biogenesis of some macromolecular structures. *TIBS* **14:** 339-342.

92. Gupta RS, Golding GB. (1993) Evolution of HSP70 gene and its implications regarding relationships between archaebacteria, eubacteria, and eukaryotes. *J Mol Evol* **37:** 573-582.

93. Gupta RS, Singh B. (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr Biol* **4:** 1104-1114.

94. Karlin S, Brocchieri L. (1998) Heat shock protein 70 family: multiple sequence comparisons, function,and evolution. *J Mol Evol* **47:** 565-577.

95. Craig EA, Gambill BD, Nelson RJ. (1993) Heat shock proteins: molecular chaperones of protein biogenesis. *Microbiol Rev* **57:** 402-414.

96. Klein KG, Olson CL, Donelson JE, Engman DM. (1995) Molecular comparison of the mitochondrial and cytoplasmic hsp70 of *Trypanosoma cruzi*, *Trypanosoma brucei* and *Leishmania major*. *J Euk Microbiol* **42:** 473-476.

97. Boorstein WR, Ziegelhoffer T, Craig EA. (1994) Molecular evolution of the HSP70 family. *J Mol Evol* **38:** 1-17.

98. Weston CJ, White SA, Jackson JB. (2001) The unusual transhydrogenase of *Entamoeba histolytica*. *FEBS Letts* **488:** 51-54.

99. Harlow DR, Weinbach EC, Diamond LS. (1976) Nicotinamide nucleotide transhydrogenase in *Entamoeba histolytica*, a protozoan lacking mitochondria. *Comp Biochem Physiol [B]* **53:** 141-144.

100. Clarke DM, Bragg PD. (1985) Cloning and expression of the transhydrogenase gene of *Escherichia coli*. *J Bacteriol* **162:** 367-373.

101. Clarke DM, Loo TW, Gilliam S, Bragg PD. (1986) Nucleotide sequence of the pntA and pntB genes encoding the pyridine nucleotide transhydrogenase of *Escherichia coli*. *Eur J Biochem* **158:** 647-653.

102. Anderson WM, Fisher RR. (1978) Purification and partial characterisation of bovine heart mitochondrial pyridine dinucleotide transhydrogenase. *Arch Biochem Biophys* **187:** 180-190.

103. Anderson WM, Fowler WT, Pennington RM, Fisher RR. (1981) Immunochemical characterisation and purification of bovine 'heart' mitochondrial pyridine dinucleotide transhydrogenase. *J Biol Chem* **256:** 1888-1895.

104. Yamaguchi M, Hatefi Y, Trach K, Hoch JA. (1988) The primary structure of the mitochondrial energy-linked nicotinamide nucleotide transhydrogenase deduced from the sequence of cDNA clones. *J Biol Chem* **263:** 2761-2767.

105. Kramer RA, Tomchak LA, McAndrew SJ, et al. (1993) An *Eimeria tenella* gene encoding a protein with homology to the nucleotide transhydrogenases of *Escherichia coli* and bovine mitochondria. *Mol Biochem Parasitol* **60:** 327-331.

106. Vermeulen AN, Kok JJ, van den Boogaart P, Dijkema R, Claessens JA. (1993) *Eimeria* refractile body proteins contain two potentially functional characteristics: transhydrogenase and carbohydrate transport. *FEMS Microbiol Lett* **110:** 223-229.

107. Diamond LS, Mattern CFT, Bartgis IL. (1972) Viruses of *Entamoeba histolytica* I. Identification of transmissible virus-like agents. *J Virol* **9:** 326-341.

108. Meerovitch E. (1958) A new host of *Entamoeba invadens* Rodhain 1934. *Can J Zool* **36:** 423-427.

109. Diamond LS, Clark CG, Cunnick CC. (1995) YI-S, a casein-free medium for axenic cultivation of *Entamoeba histolytica*, related *Entamoeba, Giardia intestinalis* and *Trichomonas vaginalis*. *J Euk Microbiol* **42:** 277-278.

110. Diamond LS, Harlow DR, Cunnick CC. (1978) A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans R Soc Trop Med Hyg* **72:** 431-432.

111. Clark CG. (1992) DNA purification from polysaccharide-rich cells. In: Lee JJ, Soldo AT (eds.) *Protocols in Protozoology.* Allen Press, Lawrence, Kansas, pp. D-3.1 - D-3.2.

112. Sambrook J, Fritsch EF, Maniatis T. (1989) Molecular cloning: a laboratory manual. New York Cold Spring Harbour Laboratory Press.

113. Dower WJ, Miller JF, Ragsdale CW. (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16:** 6127-6145.

114. Ausubel FM, Brent R, Kingston RE, et al. (1999) Short protocols in molecular biology. Wiley.

115. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19:** 4008.

116. Altschul SF, Madden TL, Schaffer AA, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-3402.

117. Feinberg AP, Vogelstein B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132:** 6-13.

118. Feinberg AP, Vogelstein B. (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal Biochem* **137:** 266-267.

119. Descoteaux S, Ayala P, Orozco E, Samuelson J. (1992) Primary sequences of 2 P-glycoprotein genes of *Entamoeba histolytica. Mol Biochem Parasitol* **54:** 201-212.

120. Swofford DL. (1998) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Assosiates, Sutherland, Massachusetts.

121. Strimmer K, Von Haeseler A. (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13:** 964-969.

122. Felsenstein J. (1993) PHYLIP, Phylogeny inference package. Version 3.75c. Univ. of Washington, Seattle.

123. Broude NE, Chandra A, Smith CL. (1997) Differential display of genome subsets containing specific interspersed repeats. *Proc Natl Acad Sci USA* **94:** 4548-4553.

124. Nickel R, Tannich E. (1994) Transfection and transient expression of chloramphenicol acetyltransferase gene in the protozoan parasite *Entamoeba histolytica. Proc Natl Acad Sci USA* **91:** 7095-7098.

125. Hamann L, Nickel R, Tannich E. (1995) Transfection and continuous expression of heterologous genes in the protozoan parasite *Entamoeba histolytica. Proc Natl Acad Sci USA* **92:** 8975-8979.

126. Kidgell C. (1997) The 70kDa heat shock proteins of *Entamoeba histolytica. MSc Thesis.* London School of Hygiene and Tropical Medicine, London, p. 59.

127. Wilkins MR, Gasteiger E, Bairoch A, et al. (1998) Protein Identification and Analysis Tools in the ExPASy Server in: 2-D Proteome Analysis Protocols. A.J. Link. Humana Press, New Jersey.

128. Engman DM, Fehr SC, Donelson JE. (1992) Specific functional domains of mitochondrial hsp70s suggested by sequence comparison of the trypanosome and yeast proteins. *Mol Biochem Parasitol* **51:** 153-155.

129. Haüsler TY, Stierhof D, Blattner J, Clayton C. (1997) Conservation of mitochondrial targetting sequence function in mitochondrial and hydrogenosomal proteins from the early- branching eukaryotes *Crithidia, Trypanosoma* and *Trichomonas. Eur J Cell Biol* **73:** 240-251.

130. Descoteaux S, Samuelson J. Cloning of a fourth *Entamoeba histolytica* P-glycoprotein gene. *GenBank.* Accesion No.U01056.

131. Hrdy I, Müller M. (1995) Primary structure of the hydrogenosomal malic enzyme of *Trichomonas vaginalis* and its relatioship to homologous enzymes. *J Euk Microbiol* **42:** 593-603.

132. Johnson PJ, d' Oliviera CE, Gorrell TE, Müller M. (1990) Molecular analysis of the hydrogenosomal ferredoxin of the anaerobic protist *Trichomonas vaginalis*. *Proc Natl Acad Sci USA* **87:** 6097-6101.

133. Bui ET, Bradley PJ, Johnson PJ. (1996) A common evolutionary origin for mitochondria and hydrogenosomes. *Proc Natl Acad Sci USA* **93:** 9651-9656.

134. Takami H, Nakasone K, Takaki Y, et al. (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* **28:** 4317-4331.

135. Bornaes C, Holmberg S, Petersen JGL. (1992) Serine and threonine catabolism in *Saccharomyces cerevisiae*: the CHA1 polypeptide is homologous with other serine and threonine dehydratases. *Genetics* **131:** 531-539.

136. Purdy JE, Pho LT, Mann BJ, Petri WAJ. (1996) Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Mol Biochem Parasit* **78:** 91-103.

137. Singh U, Rogers JB, Mann BJ, Petri WA, Jr. (1997) Transcription initiation is controlled by three core promoter elements in the *hgl5* gene of the protozoan parasite *Entamoeba histolytica*. *Proc Natl Acad Sci USA* **94:** 8812-8817.

138. Bruchhaus I, Leippe M, Lioutas C, Tannich E. (1993) Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. *DNA Cell Biol* **12:** 925-933.

139. Luna-Arias JP, Hernandez-Rivas R, de Dios-Bravo G, Garcia J, Mendoza L, Orozco E. (1999) The TATA-box binding protein of *Entamoeba histolytica*: cloning of the gene and location of the protein by immunofluorescence and confocal microscopy. *Microbiology* **145:** 33-40.

140. Buß H, Lioutas C, Dobinsky S, Nickel R, Tannich E. (1995) Analysis of the 170-kDa lectin gene promoter of *Entamoeba histolytica*. *Mol Biochem Parasitol* **72:** 1-10.

141. Gilchrist CA, Holm CF, Hughes MA, Schaenman JM, Mann BJ, Petri WA, Jr. (2001) Identification and characterization of an *Entamoeba histolytica* Upstream Regulatory element 3 sequence-specific DNA-binding protein containing EF-hand motifs. *J Biol Chem* **276:** 11838-11843.

142. Schaenman JM, Gilchrist CA, Mann BJ, Petri WA, Jr. (2001) Identification of two *Entamoeba histolytica* sequence-specific URE4 enhancer-binding proteins with homology to the RNA-binding motif RRM. *J Biol Chem* **276:** 1602-1609.

143. Tachibana T, Astumi S, Shioda R, Ueno M, Uritani M, Ushimaru T. (2002) A novel non-conventional heat shock element regulates expression of MDJ1 encoding a DnaJ homolog in *Saccharomyces cerevisiae. J Biol Chem* (in press).

144. Amin J, Ananthan J, Voellmy R. (1988) Key features of heat shock regulatory elements. *Mol Cell Biol* **8:** 3761-3769.

145. Gelderman AH, Bartgis IL, Keister DB, Diamond LS. (1971) A comparison of genome sizes and thermal denaturation-derived base composition of DNAs from several members of *Entamoeba* (*histolytica* group). *J Parasitol* **57:** 912-916.

146. Tannich E, Horstmann RD. (1992) Codon usage in pathogenic *Entamoeba histolytica. J Mol Evol* **34:** 272-273.

147. Willhoeft U, Tannich E. (1999) The electrophoretic karyotype of *Entamoeba histolytica. Mol Biochem Parasitol* **99:** 41-53.

148. Clark CG, Diamond LS. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp Parasitol* **77:** 450-455.

149. Qian F, Germino GG. (1999) Introduction to Pulse-Field Gel Electrophoresis. In: Hilderbrandt F, Igarashi P (eds.) *Techniques in Molecular Medicine.* Springer Lab Manual.

150. Ortner S, Plaimauer B, Binder M, Wiedermann G, Scheiner O, Duchêne M. (1992) Humoral immune response against a 70-kilodalton heat shock protein of *Entamoeba histolytica* in a group of patients with invasive amoebiasis. *Mol Biochem Parasitol* **54:** 175-183.

151. Bhattacharya A, Bhattacharya S, Ackers JP. (1999) Nontranslated polyadenylated ribonucleic acids from the protozoan parasite *Entamoeba histolytica. Curr Sci* **77:** 564-567.

152. Kadri AOA-H. (1999) Analysis of PNT genes among different *Entamoeba* species. *MSc Thesis.* London School of Hygiene and Tropical medicine, London, p. 51.

153. Beste DJV. (1999) The cpn60 gene of *Entamoeba* species *MSc Thesis.* London School of Hygiene and Tropical medicine, London, p. 58.

154. Field J, Van Dellen K, Ghosh SK, Samuelson J. (2000) Responses of *Entamoeba invadens* to heat shock and encystation are related. *J Euk Microbiol* **47:** 511-514.

155. Corpet F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl Acids Res* **16:** 10881-10890.

156. Engman DM, Sias SR, Gabe JD, Donelson JE, Dragon EA. (1989) Comparison of HSP70 genes from two strains of *Trypanosoma cruzi*. *Mol Biochem Parasitol* **37:** 285-287.

157. Thompson JD, Higgins DG, Gibson TJ. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680.

158. Li W-H, Sharp PM. (1991) Molecular Phylogeny. In: Li W-H, Sharp PM (eds.) *Fundamentals of molecular evolution.* Sinauer Associates, pp. 99-135.

159. Müller M. (2000) A mitochondrion in *Entamoeba histolytica*? *Parasitology Today* **16:** 368-369.

160. López-Revilla R, Gómez R. (1978) *Entamoeba histolytica, E. invadens,* and *E. moshkovskii*: fluctuations of the DNA content of axenic trophozoites. *Exp Parasitol* **44:** 243-248.

161. Mackenstedt U, Schmidt M, Raether W, Mehlhorn H, Uphoff M. (1990) Increase in DNA content of tissue stages of *Entamoeba histolytica* strain SFL 3. *Parasitol Res* **76:** 373-378.

162. Lohia A, Samuelson J. (1993) Cloning of the *Eh cdc2* gene from *Entamoeba histolytica* encoding a protein kinase p34^{cdc2} homologue. *Gene* **127:** 203-207.

163. Plaimauer B, Ortner S, Wiedermann G, Scheiner O, Duchêne M. (1994) An intron-containing gene coding for a novel 39-kilodalton antigen of *Entamoeba histolytica*. *Mol Biochem Parasitol* **66:** 181-185.

164. Urban B, Blasig C, Förster B, Hamelmann C, Horstmann RD. (1996) Putative serine/threonine protein kinase expressed in complement-resistant forms of *Entamoeba histolytica*. *Mol Biochem Parasitol* **80:** 171-178.

165. Willhoeft U, Campos-Gorgora E, Touzni S, Bruchhaus I, Tannich E. (2001) Introns of *Entamoeba histolytica* and *Entamoeba dispar*. *Protist* **152:** 149-156.

166. Gangopadhyay SS, Ray SS, Kennady K, Pande G, Lohia A. (1997) Heterogeneity of DNA content and expression of cell cycle genes in axenically growing *Entamoeba histolytica* HM1:IMSS clone A. *Mol Biochem Parasitol* **90:** 9-20.

167. Ekins R, Chu FW. (1999) Microarrays: their origins and applications. *Trends Biotech* **17:** 217-218.

168. Willhoeft U, Buß H, Tannich E. (1999) DNA sequences corresponding to the ariel gene family of *Entamoeba histolytica* are not present in *E. dispar*. *Parasitol Res* **85:** 787-789.

169. Ramakrishnan G, Petri W. (2001) Getting sense and finding function in protozoa. *Nature Biotechnol* **19:** 213-215.

170. Stock RJ, Olvera A, Sánchez R, et al. (2001) Inhibition of gene expression in *Entamoeba histolytica* with antisense peptide nucleic acid oligomers. *Nature Biotechnol* **19:** 231-234.

171. Soomets U, Hallbrink M, Langel U. (1999) Antisense properties of peptide nucleic acids. *Front Biosci* **4:** D782-786.

172. Alon RN, Bracha R, Mirelman D. (1997) Inhibition of expression of the lysine-rich 30kDa surface antigen of *Entamoeba dispar* by transcription of its antisense RNA. *Mol Biochem Parasitol* **90:** 193-201.

173. Ankri S, Stolarsky T, Mirelman D. (1998) Antisense inhibition of expression of cysteine proteinases does not affect *Entamoeba histolytica* cytopathic or haemolytic activity but inhibits phagocytosis. *Mol Microbiol* **28:** 777-785.

174. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408:** 325-330.

175. Gönczy P, Echeverrl C, Oegema K, et al. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408:** 331-336.

176. LaCount DJ, Bruse S, Hill KL, Donelson JE. (2000) Double-stranded RNA interference in *Trypanosoma brucei* using head-to-head promoters. *Mol Biochem Parasitol* **111:** 67-76.

177. Marck C. (1988) DNA Strider: a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Res.* **16:** 1829-1836.

APPENDICES

APPENDIX A

Table A1. Oligonucleotides used for sequencing the mitochondrial hsp70 gene

Name of primer	Sequence of oligonucleotide
pUC/M13 Forward (Promega)	GTT TTC CCA GTC ACG AC
pUC/M13 Reverse (Promega)	CAG GAA ACA GCT ATG AC
T3	AAT TAA CCC TCA CTA AAG GG
T7	CGG GAT ATC ACT CAG CAT AATG
SP6	TAT TTA GGT GAC ACT ATA G
70.1	TAG GGA AAA GGG TAG ATG
70.2	GAT GCA CAA CGA CAA GCT AC
70.3	ACA CAA GAC GGT GTT GTT CG
70.4	CCT TTG TCA AGT GCA GAG AC
70.5	GAT GTT GTA CTT GTT GGA GG
70.6	CTT GCT TGA ATT GCT GCT CC
70.7	TTC AAG TCG TTT GGT TGG TG
70.8	AAT GGT GGT GTG AGT GAG GG
70.9	GCT TGG AGG AGA AGA TTT TG
70.10	CCT CCA ACA AGT ACA ACA TC
70.11	GGG AGT GAG TGT GGT GGT AA
70.12	TTC AAG TCG TTT TGT TGG TG
70.13	GCA AAG AGA AGT ACC ACC TG
70.14	CCT GCA AGT GTT CCT GCA TC
70.15	GGA GAA CGT AGA GTT ACA AG
70.16	TTT GGG AGT GAG TGT GGT GG
70.17	CCA CCA CAC TCA CTC CCA AA

Table A1. Continued over page.

Table A1. Continued from previous page.

Name of primer	Sequence of oligonucleotide
70.18	GCA CTT GAC AAA GGA ACA GG
70.19	ACA GTG GTT GTT GGA TGT AG
70.20	CCA TCC TTC TCC ATT CTT TC
70.21	CGA CTA GTT AAG GAA GGA GA
70.22	GAA CAT GCT GCA GAA GAC AA
70.23	GCT AGA ACA ACT CCT TCA GT
70.24	TCA GAT TCT GTG AGA GAA AAA TAA C
70.25	ATG TTT GTC TCA CAA CCT GC

Table A2. Oligonucleotide used to obtain probes of the *Entamoeba* genes

Name of primer	Sequence of oligonucleotide 5'-3'	*Entamoeba* genes
β-actin A	GGA TTT GCT GGT GAT GAT	*E. histolytica* β-actin gene
β-actin B	TCA GTA AGA TCA CGT CCT	*E. histolytica* β-actin gene
Deg 1	GGN ATH GAY YTN GGN ACN AC	*E. histolytica* mt-hsp70 gene
Deg 2	GGR TCR AAR TCN TCN CCN CC	*E. histolytica* mt-hsp70 gene
70.4	CCT TTG TCA AGT GCA GAG AC	*E. histolytica* mt-hsp70 gene
70.9	GCT TGG AGG AGA AGA TTT TG	*E. histolytica* mt-hsp70 gene
N-CPN60	TAT AGG GAT CCG CAT GCT TTC ATC TTC AAG	*E. histolytica* cpn60 gene
C-CPN60	TAC TAG ATA TCA ATG TCT TCT TTT AGT GT	*E. histolytica* cpn60 gene
US-5	TTT GGG AGT GAG TGT GGT GG	*E. histolytica* mt-hsp70 upstream region of the gene
US-3	TCA GAT TCT GTG AGA GAA AAA TAA C	*E. histolytica* mt-hsp70 upstream region of the gene

Table A2. Continued over page

Table A2. Continued from previous page

Name of primer	Sequence of oligonucleotide 5'-3'	*Entamoeba* genes
D1.70	GGN ATH TTY CAR GTN AAR GCN AC	*E. dispar* SAW 760; *E .moshkovskii* Laredo; *E. terrapinae* M; *E. invadens* IP-1 mt-hsp70
D2.70	TTY TTN GCN GGN GGD ATN CCN AC	*E. dispar* SAW 760; *E .moshkovskii* Laredo; *E. terrapinae* M; *E. invadens* IP-1 mt-hsp70
Ei actin 1	TGG TGA TGA CGC ACC ACG CG	*E. invadens* actin gene
Ei actin 2	TTC AGT CAA GAG GAC TGG GTG	*E. invadens* actin gene
PNT 5	GGN GGN GCN GAY ATG CCN GTN GT	*E. dispar* SAW 760; *E .moshkovskii* Laredo; *E. invadens* IP-1; *E. terrapinae* M
PNT 3	CCN GGN CAN CKN CCN GCN CAN GG	*E. dispar* SAW 760; *E .moshkovskii* Laredo; *E. invadens* IP-1; *E. terrapinae* M PNT
Sp6	TAT TTA GGT GAC ACT ATA G	*E. dispar* SAW 760; *E. moshkovskii* Laredo; *E. terrapinae* M; *E. invadens* IP-1 PNT
T7	CGG GAT ATC ACT CAG CAT AATG	*E. dispar* SAW 760; *E. moshkovskii* Laredo; *E. terrapinae* M; *E. invadens* IP-1 PNT

Table A2. Continued over page.

Table A2. Continued from previous page

Name of primer	Sequence of oligonucleotide 5'-3'	*Entamoeba* genes
Cpn5	ACN AAR GAY GGN GTN ACN GTN GC	*E. moshkovskii* Laredo; *E. terrapinae* M cpn60
Cpn3	CCN CCN CCN GGN ACD ATN CCY TC	*E. moshkovskii* Laredo; *E. terrapinae* M, *E. dispar* SAW 760 cpn60
Disp-1	CAG TCG CAA GAG TAT CAG CA	*E. dispar* SAW 760 cpn60
70.21	CGA CTA GTT AAG GAA GGA GA	*E. histolytica* 3' end probe of mt-hsp70
3*Bgl*II	GCA GGG CAG ATC TAT TAT ATA ATT TAG CAG AAG	*E. histolytica* 3' end probe of mt-hsp70
PNT-A	GTA GGA CTT GCA GCA GTA TT	*E. histolytica* PNT probe
PNT-B	GGT AAT CTT CCT GCA ACT GG	*E. histolytica* PNT probe

Note:

Key to degenerate primers abbreviations: N = A + G + C + T
K = G + T
Y = C+ T
H = A + T + C
R = A + G
D = G + A + T

APPENDIX B

Complete alignment of hsp70 sequences (input data matrix)

```
1111111111222222222233333333334444444444555555555566666666667777777 7
Taxon/Node          12345678901234567890123456789012345678901234567890123456789012345678901234567
-------------------------------------------------------------------------------------------------
E.histolytica       ----------------------------------------------------MFVSQPARSTCIGIDLGTTNSCMCV
T.vaginalis         --------------MLKMFNSIFAREKNQHVLTVYMTTYLRSVGKTIRNLFFRTFAQGMTPIIGIDLGTTNSCVSV
P.sativum           --MAATLLRSLQRRNLSSSSVSAFR-----SLTGSTKTSYATHKLASLT--RPFSSRPAGNDVIGIDLGTTNSCVSV
E.tenella           --MRGAVALSAARALWAAAVPPQPRGPPKEQRVFSAVRTAAVGTLSSLAGRRGFSG--VRGDVVGIDLGTTNSCVAV
M.musculus          --MISASRAAAARLVGTAASRSPAAARPQDGWNGLSHEAFRFVSRRDYA------SEAIKGAVVGIDLGTTNSCVAV
S.cerevisiae        --------MLAAKNILNRSSLS-SSFRIAT----------RLQST----------KVQGSVIGIDLGTTNSAVAI
L.major             ---------MFARRVCGSAAASAACLARHE--------------SQ----------KVQGDVIGVDLGTTYSCVAT
R.prowazekii        ------------------------------------------------------MGKVIGIDLGTTNSCVAV
R.capsulatus        ------------------------------------------------------MAKVIGIDLGTTNSCVAI
A.tumefaciens       ------------------------------------------------------MAKVIGIDLGTTNSCVAV
Rhodopseudomonas sp.------------------------------------------------------MGKVIGIDLGTTNSCVAV
H.ducreyi           ------------------------------------------------------MGKIIGIDLGTTNSCVAV
L.pneumophila       ------------------------------------------------------MAKIIGIDLGTTNSCVAV
C.trachomatis       ----------------------------------------------------MSEKRKSNKIIGIDLGTTNS
B.burgdorferi       ------------------------------------------------------MGKIIGIDLGTTNSCVAI
```

Input data matrix (continued):

```
                                    1111111111111111111111111111111111111111111111111111111111
            7788888888889999999999000000000011111111112222222222333333333344444444445555 5
Taxon/Node  8901234567890123456789012345678901234567890123456789012345678901234567890123 4
-----------------------------------------------------------------------------------------------
E histoly     FDKTTPRILENAEGKRTTPSCVSFTP----TGILVGEAAKRMEALHPTTTVSGVKRMIGCQYK--NVEQQRK---PY
T vaginalis   VEGGTPKVIQNAEGVRTTPSIVAFTN---TGERLVGEQAKRQAITNSKSTFFATKRLIGCSFDDEMTKKCRE-MVPY
P sativum     MEGKNPKVIENSEGARTTPSVVAFNQ---KSELLVGTPAKRQAVTNPTNTLFGTKRLIGRRFDDAQTQKEMK-MVPY
E ten         MEGSQPKVLENSEGMRTTPSVVAFTK---DGQRLVGVVAKRQAITNPENTFFSTKRLIGRSFDEEAIAKERK-ILPY
M musc        MEGKQAKVLENAEGARTTPSVVAFTA---DGERLVGMPAKRQAVTNPNNTFYATKRLIGRRYDDPEVQKDTK-NVPF
Yeast         MEGKVPKIIENAEGSRTTPSVVAFTK---EGERLVGIPAKRQAVVNPENTLFATKRLIGRRFEDAEVQRDIK-QVPY
L major       MDGDKARVLENSEGFRTTPSVVAFKG----SEKLVGLAAKRQAITNPQSTFYAVKRLIGRRFEDEHIQKDIK-NVPY
R prowazekii  MEGKEPKVIDNAEGERTTPSIIAFA----NSERLVGQPAKRQAVTNPRNTIYAVKRLIGRNFTDPMVRKDQG-LVPY
R caps        MDGSQPRVIENSEGARTTPSIVAYTD----NERLVGQPAKRQAVTNPTNTVFAVKRLIGRRTTDAEVEKDKK-LVPY
A tumef       MDGKDTKVIENAEGARTTPSMVAFSD---DGERLVGQPAKRQAVTNPTNTLFAVKRLIGRRYEDPTVEKDKA-LVPF
Rhpseud       MDGKSAKVIENAEGMRTTPSIVAITD---DGERLVGQPAKRQAVTNPERTFFAVKRLIGRRYDDPMVEKDKG-LVPY
H ducreyi     MDGDKPRVLENAEGARTTPSIIAYT----DKETLVGQPAKRQAITNPKNTLFAIKRLIGRRFTDNEVQRDIE-IMPF
L pneum       MEGDKPKVIENSEGHRTTPSIVAFTD---DNEILVGQSAKRQSVTNPEKTLFAIKRLIGRRFDDPIVQKDIK-MVPY
C trach       MEGGQPKVIASSEGTRTTPSIVAFKG----GETLVGIPAKRQAVTNPEKTLASTKRFIGRKFS--EVESEIK-TVPY
B burg        MEHGKPVVIQNSEGGRTTPSIVAYTN---KGERLVGQVAKNQMVTNPENTIYSIKRFMGRRFE--EVASEIK-MVPY
```

Input data matrix (continued):

```
                111111111111111111111111111111111111111111112222222222222222222222222222222
                555556666666666777777777788888888889999999990000000000111111111122222222233
Taxon/Node      567890123456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------------
E histolyt      KIVEGKNGEGWIYIN--GKTYSPSEISSFILKKLKKDAEAKLGKRVDE---AVITCPAYFNDAQRQATKDAGTLAGL
T vaginalis     QIVKAKSGDAWVKDEK-GNEYSPSQIGAFVLMKMKETAEAYLGTSVKD---AVITVPAYFNDSQRQATKDAGKIAGL
P sativumm      KIVRAPNGDAWVEAN--GQQYSPSQIGAFVLTKIKETAEAYLGKTISK---AVVTVPAYFNDAQRQATKDAGRIAGL
E tenella       KVIRADNGDAWVEGW--GKKYSPSQIGAFVLMKMKETAESYLGRDVNQ---AVITVPAYFNDSQRQATKDAGKIAGL
M musculus      KIVRASNGDAWVEAH--GKLYSPSQIGAFVLMKMKETAENYLGHTAKN---AVITVPAYFNDSQRQATKDAGKIAGL
Yeast           KIVKHSNGDAWVEAR--GQTYSPAQIGGFVLNKMKETAEAYLGKPVKN---AVVTVPAYFNDSQRQATKDAGQISGL
L major         KIVRAGNGDAWVQDGN-GKQYSPSQIGAFVLEKMKETAENFLGHKVSN---AVVTCPAYFNDAQRQATKDAGTIAGL
R prowazekii    NIVKADNGDAWVEAD--NHKYSPSQISAFILQKMKETAENYLGEKVTQ---AVITVPAYFNDAQRQATKDAGKIAGL
R capsulatus    NIVDGGNGDAWVEVR--GEKFSPAQVSAVILQKMKETAESYLGETVTQ---AVITVPAYFNDAQRQATKDAGKIAGL
A tumef         EIVKGDNGDAWVKAQ--DKNYSPSQISAMILQKMKETAESYLGEKVEK---AVITVPAYFNDAQRQATKDAGRIAGL
Rhpseud         KIVKASNGDAWVEAD--GKTYSPSQVSAFILQKMKETAEAHLGQKVDQ---AVITVPAYFNDAQRQATKDAGKIAGL
H ducreyi       EIAKADNGDAWVNVK--GDKLAPPQISAEVLKKMKKTAEDFLGEPVTE---AVITVPAYFNDAQRQATKDAGRIAGL
L pneum         KIMKADNGDAWVRVK--DQDKAPPQISAEVLRKMKKTAEDYLGEEVKE---AVITVPAYFNDSQRQATKDAGRIAGL
C trach         KVAPNSKGDAVFDVE--QKLYTPEEIGAQILMKMKETAEAYLGETVTE---AVITVPAYFNDSQRASTKDAGRIAGL
B burg          KIEKGLNGDARVNISNIKKQMSPPEISAATLTKMKETAEAYLGEKVTE---AVITVPAYFNDAQRQATKDAGKIAGL
```

Input data matrix (continued):

```
                2222222222222222222222222222222222222222222222222222222222222333333333
                3333333344444444445555555555566666666667777777777888888888899999999990000000000
Taxon/Node      2345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------------------------
E histoly       KVKRIINEPTAAALAYGIDTRKEN--EGKNIAVYDLGRGTFDISILNINK----GIFQVKATNGDTMLGGEDFDKAI
T vaginalis     NVMRTLNEPTAAALAYGTE-RKQG----QTVAVFDLGGGTFDFSILEISK---DGVFEVKATNGDTFLGGEDFDA
P sativum       DVQRIINEPTAAALSYGMN-NKEG-----LIAVFDLGGGTFDVSILEISN----GVFEVKATNGDTFLGGEDFDNAL
E tenella       DVLRIINEPTAAALAYGME-KEDG----RTIAVYDLGGGTFDVSILEILG----GVFEVKATNGNTSLGGEDFDQKV
M musculus      NVLRVINEPTAAALAYGLD-KSED----KVIAVYDLGGGTFDISILEIQK----GVFEVKSTNGDTFLGGEDFDQAL
Yeast           NVLRVVNEPTAAALAYGLE-KSDS----KVVAVFDLGGGTFDISILDIDN----GVFEVKSTNGDTHLGGEDFDIYL
L major         NVIRVVNEPTAAALAYGMD-KTKD----SLIAVYDLGGGTFDISVLEIAG----GVFEVKATNGDTHLGGEDFDLAL
R prowazekii    EVLRIINEPTAAALAYGFE-KSSS----KTIAVYDLGGGTFDVSILEISD----GVFEVKSTNGDTFLGGEDFDTRI
R capsulatus    EVLRIINEPTAAALAYGLD-KKDS----KTIAVYDLGGGTFDISVLEIDD----GLFEVKSTNGDTFLGGEDFDMRI
A tumef         DVLRIINEPTAAALAYGLD-KKEG----KTIAVYDLGGGTFDISVLEIGD----GVFEVKSTNGDTFLGGEDFDMRL
Rhpseud         EVLRIINEPTAAALAYGLD-KAKT----GTIAVYDLGGGTFDVSILEIGD----GVFEVKSTNGDTFLGGEDFDMRL
H ducreyi       EVKRIINEPTAAALAYGLD-SKKE---NQIIAVYDLGGGTFDISIIEIDNFDGEQTFEVRATNGDTHLGGEDFDNRV
L pneum         EVKRIINEPTAAALAYGMD-KKRG---DSVIAVYDLGGGTFDISIIEIAEVDGEHQFEVLATNGDTFLGGEDFDLAL
C trach         DVKRIIPEPTAAALAYGID-KEGD----KKIAVFDLGGGTFDISILEIGD----GVFEVLSTNGDTHLGGDDFDGVI
B burg          EVKRIVNEPTAAALAYGIE-KKHE----EIVAVYDLGGGTFDISILELGD----GVFEVKSTNGDTHLGGDNFDDEI
```

Input data matrix (continued):

```
                3333333333333333333333333333333333333333333333333333333333333333333333333333333333
                0111111111122222222223333333333444444444455555555556666666666777777777788888888
Taxon/Node      9012345678901234567890123456789012345678901234567890123456789012345678901234567890
-------------------------------------------------------------------------------------------------
E histoly       CQYIEKEFER--------------KYKRNLQRNKKGISRIKEAAEKVKCELSSSEESVISLPYLDRQD---------
T vaginalis     MKYVIQDFQA--------------KNQIDLAKDPLALQRIREAVEKAKCELSSMMTTEINLPYITVTGAG--------
P sativum       LDFLVSEFKR--------------TESIDLAKDKLALQRLREAAEKAKIELSSTSQTEINLPFISADASG--------
E tenella       LQFLVNEFKK--------------KEGIDLSKDRLALQRLREAAETAKIELSSKLSTEINLPFITADQSG--------
M musculus      LRHIVKEFKR--------------ETGVDLTKDNMALQRVREAAEKAKCELSSSVQTDINLPYLTMDASG--------
Yeast           LREIVSRFKT--------------ETGIDLENDRMAIQRIREAAEKAKIELSSTVSTEINLPFITADASG--------
L major         SDYILEEFRK--------------TSGIDLSKERMALQRVREAAEKAKCELSSAMETEVNLPFITANADG--------
R prowazekii    LNHLIEVFKK--------------ESGIDLSKDPLALQRLKEAAEKAKKELSSTSTTDINLPYITADSTG--------
R caps          VNYLADEFKK--------------EHGVDLTKDKMALQRLKEAAEKAKIELSSASQTEINQPFISMNAATG-------
A tumef         VEYLAGEFKK--------------DQGIDLKNDKLALQRLKEAAEKAKIELSSSQQTEINLPFITADASG--------
Rhpseud         VNYLADEFQK--------------EQGIDLRKDKLALQRLKEAAEKAKIELSSTTQTEINLPFITADQSG--------
H ducreyi       INYLVEEFQK--------------EQGVDLRNDPMAMQRVKEAAEKAKIELSSAQETEVNLPYITADATG--------
F tul           MNYLIDEFKK--------------EQGIDLHNDKLALQRVREAAEKAKVELSPAQQTDVNLPYITADATG--------
L pneum         IEYLASEFKK--------------DTGIDLHNDPLALQRLKEAAEKAKIELSSAQQTDVNLPYITADASG--------
C trach         INWMLDEFKK--------------QEGIDLSKDNMALQRLKDAAEKAKIELSGVSSTEINQPFITIDANG--------
B burg          IKHLISEFKK--------------ESAIDLSNDKMALQRLKEAAEKAKIELSGAQEASINLPFITADANG--------
```

Input data matrix (continued):

```
                3333333333333344444444444444444444444444444444444444444444444444444444444444444444
                8888999999999900000000001111111111222222222233333333334444444444555555555566666
Taxon/Node      6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012
-------------------------------------------------------------------------------------------------------------------
E histoly       --------SLEITISRRKIEQLCKGICKRTEYPCLQCMKDAKLRKKDISDVVLVGGMTRMPLIQ-TLSRIFGKKPSK
T vaginalis     ------PKHLQMPITRATFEKITQHLIARTINPCKNCLKDAGLTPQQINEVILVGGMTRMPKVIDSVKEFFGKDPFR
P sativum       ------AKHLNITLTRSKFEALVNNLIERTKAPCKSCLKDANISIKDVDEVLLVGGMTRVPKVQQVVSEIFGKSPSK
E tenella       ------PKHLQVSLSRAHLEELVGALLQQSIEPCEKCIRDAGVQKADLSDVILVGGMTRMPKVAEVVKNIFHKEPSK
M musculus      ------PKHLNMKLTRAQFEGIVTDLIKRTIAPCQKAMQDAEVSKSDIGEVILVGGMTRMPKVQQTVQDLFGRAPSK
Yeast           ------PKHINMKFSRAQFETLTAPLVKRTVDPVKKALKDAGLSTSDISEVLLVGGMSRMPKVVETVKSLFGKDPSK
L major         ------AQHIQMRISRSKFEGITQRLIERSIAPCKQCMKDAGVELKEINDVVLVGGMTRIRSGGGGEEVLP-EGPVR
R prowazekii    ------PKHLNIKFTRAELEKLVDDLIEKTIEPCRQALKDAGFKPNDIQEVVLVGGMTRMPKVQEAVKKFFGREPHK
R capsulatus    -----VPLHMVMKLTRAKLESLVDDLIKASLKPCAAALKDAGVSKDEIDEVVLVGGMTRMPRVVEEVTKFFGKEPHK
A tumef         ------PKHLTPKLTRAKFESLVDDLVQRTVAPCKAALKDAGVTAAEIDEVVLVGGMSRMPKVREVVKQLFGKEPHK
Rhpseud         ------PKHLTMKLTRAKFEALVDDLVQKTIEPCRKALKDAGLTAGEISEVVLVGGMTRMPKVQEVVKQLFGKEPHK
H ducreyi       ------PKHLNIKVTRAKLEALVEDLVSRSLEPLKTALADAGLSVSEINDVILVGGQTRMPLVQKKVADFFGKEPRK
L pneum         ------PKHLNIKLTRAKLESLVEKLVERTIEPCKTALKDAGLTVSQINEVILVGGQTRMPLVQKTVEEFFGKEPRK
C trach         ------PKHLALTLTRAQFEHLASSLIERTKQPCAQALKDAKLSASDIDDVLLVGGMSRMPAVQAVVKEIFGKEPNK
B burg          ------PKHLQYTLTRAKFEQMVDHLVQKTKEPCLKAIKDAGLKASDINEVILVGGSTRIPAIQKIVKDIFGQDPNK
```

Input data matrix (continued):

```
                4444444444444444444444444444444444445555555555555555555555555555555555555555
                6666666777777777788888888889999999990000000000111111111122222222233333333333
Taxon/Node      3456789012345678901234567890123456789012345678901234567890123456789012345678923
-------------------------------------------------------------------------------------------------
E histoly       NVNPDEAVAIGAAIQASIIEGKKK--DIILVDVTPLTLGIETYGGVMTPLINRNTTIPTSVS-KEFTTSMDNQQEVD
T vaginalis     GVNPDEVVAIGASIQGSVMRGDHK--DIVLLDVTPLSLGIETMGGVFSRLIPRNTVVPTKKS-QEFTTAADGQTHVN
P sativum       GVNPDEAVAMGAALQGGILRGDVK--ELLLLDVTPLSLGIETLGGIFTRLISRNTTIPTKKS-QVFSTAADNQTQVG
E tenella       GVNPDEAVAAGAAIQAGVLKGEIK--DLLLLDVCPLSLGIETLGGVFTRLINRNTTIPTKKS-QIFSTAADNQTQVG
M musculus      AVNPDEAVAIGAAIQGGVLAGDVT--DVLLLDVTPLSLGIETLGGVFTKLINRNTTIPTKKS-QVFSTAADGQTQVE
Yeast           AVNPDEAVAIGAAVQGAVLSGEVT--DVLLLDVTPLSLGIETLGGVFTRLIPRNTTIPTKKS-QIFSTAAAGQTSVE
L major         GVNPDEAVALGAATLGGVLRGKAS--DLILVDVTPLSLGTSVVGDVFVPIIPKNTTIPCMRS-HIFTTVDDGQTAIK
R prowazekii    GVNPDEVVALGAAIQGGVLNKEVT--DILLLDVTPLSLGIETLGGVFTRLIDRNTTIPSKKS-QVFSTADDNQHAVT
R capsulatus    GVNPDEVVALGAAIQAGVLQGDVK--DVVLLDVTPLSLGIETLGGVFTRLIDRNTTIPTKKS-QIFSTAEDNQNAVT
A tumef         GVNPDEVVAMGAAIQAGVLQGDVK--DVLLLDVTPLSLGIETLGGVFTRLIDRNTTIPTKKS-QTFSTAEDNQSAVT
Rhpseud         GVNPDEVVAIGAAIQAGVLQGDVK--DVLLLDVTPLSLGIETLGGVFTRIIDRNTTIPTKKS-QVFSTAEDNQNAVT
H ducreyi       DVNPDEAVAIGAAVQGGVLAGDVT--DVLLLDVTPLSLGIETMGGVMTSLIEKNTTIPTKKS-QVFSTAEDNQSAVT
L pneum         DVNPDEAVAVGAAIQAAVLSGEVK--DILLLDVTPLSLGIETMGGVMTKLIEKNTTIPTKAT-QVFSTADDNQTA
C trach         GVNPDEVVAIGAAIQGGVLGGEVK--DVLLLDVIPLSLGIETLGGVMTPLVERNTTIPTQKK-QIFSTAADNQPAVT
B burg          GVNPDEAVAIGAAIQGGILTGETK--DMVLLDVTPLSLGIETLGGVMTKLIERNTTIPTKKS-QVFSTAADNQTSVD
```

Input data matrix (continued):

```
                5555555555555555555555555555555555555555555555555555555566666666666666666
                4444444444555555555566666666667777777777888888888899999999990000000000111111
Taxon/Node      0123456789012345678901234567890123456789012345678901234567890123456789012345
-------------------------------------------------------------------------------------------
E histoly       IKVFQGERRVTRKNKKLGEFKLVGIPPAKKGVAKIEVTFDIDVNGIVKVSALDKGTGKKTGIQVKSNGGLNEEEINR
T vaginalis     IRVFQGERDLVEGNKLLGEFTLVGIPPAPRGVPKIEVTFDIDANSIVHVSAKDKQTNKEQQMTIQQHGGLSQDEIAK
P sativum       IKVLQGEREMAADNKSLGEFDLVGIPPAPRGLPQIEVTFDIDANGIVTVSAKDKSTGKEQQITIRSSGGLSDDEIDK
E tenella       IKVYQGEREMASANKLLGQFDLVGIPPAPRGVPQIEVTFDVDANGIMNISAVDKSTAKRQQITIQSSGGLSEAQIKQ
M musculus      IKVCQGEREMAGDNKLLGQFTLIGIPPAPRGVPQIEVTFDIDANGIVHVSAKDKGTGREQQIVIQSSGGLSKDDIEN
Yeast           IRVFQGERELVRDNKLIGNFTLAGIPPAPKGVPQIEVTFDIDADGIINVSARDKATNKDSSITVAGSSGLSENEIEQ
L major         FKVFQGEREIASENQIRGEFDLSGIPPAPRGVPQVEVTFDIDANGICHVTAKDKATGKTQNITITANGGLSKEQIEQ
R prowazekii    IRVFQGEREMAKDNKLLGQFNLEGIPPAPRGVPQIEVTFDIDANGIVHVSAKDKASGKEQKVTIQASGGLSDAEIEQ
R capsulatus    IRVFQGEREMAADNKMLGMFNLENIPPAPRGVPQIEVTFDIDANGIVSVKAKDKGTGKEQQITIQASGGLSDDDIEK
A tumef         IRVSQGEREMAQDNKLLGQFDLVGLPPSPRAVPQIEVTFDIDANGIVQVSAKDKGTGKEQQIRIQASGGLSDADIEK
Rhpseud         IRVFQGEREMAADNKMLGQFDLMGIPPAPRGMPQIEVTFDIDANGIVNVSAKDKATGKEQQIRIQASGGLSDSEIDK
H ducreyi       IHVLQGERKRAADNKSLGQFNLEGINPAPRGMPQIEVTFDIDANGIINVSAKDKNTGKEQQIKIQASSGLSDEEVEQ
L pneum         VHVLQGEREQASANKSLGRFDLRDIPPAPRGVPQIEVTFDIDANGILNVSAKDKATGKAQSIVIKASSGLSEEEVAA
C trach         IVVLQGERPMAKDNKEIGRFDLTDIPPAPRGHPQIEVTFDIDANGILHVSAKDAASGREQKIRIEASSGLKEDEIQQ
B burg          IKVLQGEREMAAQNRILGNFILDGIPAAPRGVPQIEVSFDIDANGIVHVSAKDMGTGKEQKIRIESSSGLSESEIDR
```

Input data matrix (continued):

```
                6666666666666666666666666666666666666666666666666666666666666666666666666666
                1112222222222333333333344444444445555555555666666666677777777778888888888999
Taxon/Node      7890123456789012345678901234567890123456789012345678901234567890123456789012
-------------------------------------------------------------------------------------------
E histoly       LVKEGEEHAAEDKRKEQLLLHRQRLKEMIESAEDIQKELLKKKIETRELERVINQSKKVITSEDENEIKR--LIDQL
T vaginalis     MIQDAEKHAAEDKKKREILEKKYAMKQYINEIEKTISENEKKLPADLLKRIRDSVKDLKDAIAGND---------EK
P sativum       MVKEAELHAQRDQERKALIDIRNSADTSIYSIEKSLAEYREKIPAEVAKEIEDAVSDLRTAMAGEN---------AD
E tenella       MVEDAERFKDEDQRQKDLVAAKNEAETLVYSVEKQISDLKDKISAEDKTDLESRIQELRSALVEGE---------LE
M musculus      MVKNAEKYAEEDRRKKERVEAVNMAEGIIHDTETKMEEFKDQLPADECNKLKEEISKVRALLARKDS------ETGE
Yeast           MVNDAEKFKSQDEARKQAIETANKADQLANDTENSLKEFEGKVDKAEAQKVRDQITSLKELVARVQGGEE---VNAE
L major         MIRDSEQHAEADRVKRELVEVRNNAETQLTTAERQLGEWKYVSDAE-KENVKTLVAELRKAMENPN-------VAKD
R prowazekii    MVKDAEHNADEDKKRKELIETKNAADSLVYSTEKTLREYGDKLSSEEKGTVEEALTSLKAALESE---------DAS
R capsulatus    MIKDAEANAEADKKRKELVEAKNTGESLLHSTRKSLEEHGDKVDGSTVEMIELACNALEESLKSE---------DPG
A tumef         MVKDAEANAEADKKRRAGVEAKNQAESLIHSTEKSVKEYGDKVSETDRKAIEDAIASLKTAVEAAE-------PDAD
Rhpseud         MVKDAEANAAEDKKRREAVDAKNHADALVHSTEKALAEHGSKIDEGERRSIEDALSDLREALKGD---------DAE
H ducreyi       MVRDAEANAESDRQFEELVQTRNQADSIAHATRKQISEAGDALTAEDKAKIETALAELETAAKGE---------DKA
L pneum         MVKDAQSHAEEDKKFKEMAELRNQADSLIHSCEKSMKDLADELSEDEKRGIETAISELKEAVQGT---------DKA
C trach         MIRDAELHKEEDKQRKEASDVKNEADGMIFRAEKAVKDYHDKIPAELVKEIEEHIEKVRQAIKED--------ASTT
B burg          MVKDAEAHAEEDKKLKENIEAKNTANSLIYQTEKSLKEYSEKISSEDKEAIESKIKELKESLEKE---------DIS
```

Input data matrix (continued):

```
                6666667777777777777777777777777777777777777777777777777777777777777777
                9999990000000000111111111122222222223333333333444444444455555555556666666667
Taxon/Node      4567890123456789012345678901234567890123456789012345678901234567890123456789
-------------------------------------------------------------------------------------------
E histoly       G-EETSNSSAKLYN-----------------------------------------------------------
T vaginalis     KIEEKYDALKAASMEIYNAISDNN----------TNKNKK---------------------------------
P sativum       DIKAKLDAANKAVSKIGQHMSGGS----------SGGPSEG----------------------GSQGG-EQAPEAE
E tenella       TIRSRVKALQELSWKVSQQAYS-Q----------SNNTSAD----------------------GDSSS-TSSGDSS
M musculus      NIRQAASSLQQASLKLFEMAYKKM----------ASEREGS----------------------GSSGTGEQK----
Yeast           ELKTKTEELQTSSMKLFEQLYKND----------SNNNNN----------------------NNGNNAESG----
L major         DLAAATDKLQKAVMECGRTEYQQA----------AAANSGQ----------------------C-----------
R prowazekii    LIKEKTGNLTAANMKIGEAMYKT--------QTENQHSEANTV--------------------NNEKVVDADF---
R capsulatus    KIKGAVQNLTDAAMKLGEAIYKAQ-------ASEAGPASDDEDGPRS-----------------VDDDIVDADFEDM
A tumef         DIQAKTQTLMEVSMKLGQAIYEAQ-------QAEAGDASAE----------------------GKDDVVDADYEEI
Rhpseud         AIKTKSNTLAQASMKLGEAMYKQA-------EAGGAAQQAG----------------------KDDVVDAEFTEV
H ducreyi       EIEAKIEAVIKASEPLMQAAQAKA------QTNQAGEQQSS-A--------------------KDDSVVDAEFEEV
L pneum         RIEDKLKVLTDASAKMAERIYAKKSSEGQAAQGQTQSQESTKP--------------------AEEGVVDAEFE
C trach         AIKAASDELSTRMQKIGEAMQAQSASA---AASSAANAQGGPNINSED-------------------------
L inter         RIESAKASISKIATDIASKVYSQSAPG---AEQAAAGGPNGAGSND-----------QGNSTNNGEKVVDADYTVV
B burg          LIKSRTEELQKASYKIAEMMYKDSSQ-----QNANSQQENGP---------------QSNTSEEGK---EADYEVV
```

APPENDIX C

Alignment of cpn60 amino acid sequences used for the phylogenetic analyses.
The numbers on the top of the sequences represent the positions of the amino acids of the *E. histolytica* cpn60 protein.

```
                                                                                100
Entamoeba histolytica       TKDGVSVAKALTFSDNTLNVGGKIAKEVASKVNDRSGDGTTTATCLLRKVACEGVQAINTGLSGTDLLKGIS
Entamoeba dispar            TKDGVTVAKALTFSDNTLNVGGKLAKEVASKVNDRSGDGTTTATCLLRKVACEGVQAINTGLSSTDLLKGIS
Entamoeba moshkovskii(Lar)TKDGVTVAKSLTFSDNTLNVGAKMAKGVASKINDRSGDGTTTATCLLRKVACEGSQAVNTGLATSDLLKGIS
Entamoeba terrapinae        TKDGVTVAKAISFSDNTLNVGAKLAKEVASKINDKSGDGTTTATCLLRKVTSDGIQAVNTGLPTADLLRGIS
Entamoeba invadens          TKDGVTVAKAIEFSDKTLNAGAKIAKGVASKINDNSGDGTTTATCLLRKVVMEGVQAVNTGLPTPDLLSGIN
Chlamydia pneumoniae        TKDGVTVAKEIELEDKHENMGAQMVKEVASKTADKAGDGTTTATVLAEAIYSEGLRNVTAGANPMDLKRGID
Escherichia coli            TKDGVSVAREIELEDKFENMGAQMVKEVASKANDAAGDGTTTATVLAQAIITEGLKAVAAGMNPMDLKRGID
Haemophilus influenzae      TKDGVSVAREIELEDKFENMGAQMVKEVASKANDVAGDGTTTATVLAQSIVSEGLRAVAAGMNPMDLKRGID
Pseudomonas aeruginosa      TKDGVSVAKEIELKDKFENMGAQLVKDVASRPTDAAGDGTTTATVLAQAIVNEGLKAVAAGMNPMDLKRGID
Neisseria gonorrhoeae       TKDGVTVAKEIELKDKFENMGAQMVKEVASKTNDVAGDGTTTATVLAQSIVAEGIKAVTAGMNPTDLKRGID
Legionella micdadei         TKDGVSVAKEIEFENRFKNMGAQMVKEVAAKTSDTAGDGTTTATVLARAIVVEGHKAVAAGMNPMDLKRGID
Coxiella burnetii           TKDGVSVAKEIELEDKFENMGAQMVKEVASRTSDDAGDGTTTATVLAQAILVEGIKAVIAGMNPMDLKRGID
Agrobacterium tumefaciens TKDGVSVAKEIELEDKFENMGAQLVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGID
Rhizobium meliloti          TKDGVSVAKEIELEDKFENMGAQMVREVASKTNDIAGDGTTTATVLAQAIVREGAKAVAAGMNPMDLKRGID
Brucella abortus            TKDGVSVAKEVELEDKFENMGAQMLREVASKTNDTAGDGTTTATVLGQAIVQEGAKAVAAGMNPMDLKRGID
Bartonella bacilliformis    TKDGVSVAKEIELENKFENMGAQMLREVASKTNDIAGDGTTTATVLGQAIVQEGVKAVAASMNPMDLKRGID
Bradyrhizobium japonicum    TKDGVAVAKEIELDDKFENMGAQMVREVASKAADAAGDGTTTATVLAAAIVREGAKSVAAGMNPMDLKRGID
Homo sapiens                TKDGVTVAKSIDLKDKYKNIGAKLVQDVANNTNEEAGDGTTTATVLARSIAKEGFEKISKGANPVEIRRGVM
Zea mays                    TKDGVTVAKSIEFKDRVKNVGASLVKQVANATNDTAGDGTTCATVLTKAIFTEGCKSVAAGMNAMDLRRGIS
Saccharomyces cerevisiae    TKDGVTVAKSIVLKDKFENMGAKLLQEVASKTNEAAGDGTTSATVLGRAIFTESVKNVAAGCNPMDLRRGSQ
Trypanosoma cruzi           TKDGVTVAKAIEFKDPFENMGAQLVRQVCNKTNDLAGDGTTTSAVLVASVFSESLRCIATGTNPIDMKRGMD
Trypanosoma brucei          TKDGVTVAKSIEFKDPFENMGAQLVRQVCNKTNDLAGDGTTTSAVLVASIFSEGIKSIATGTNPIDMKRGMD
Trichomonas vaginalis       TKDGVTVAKSIEFADKWHNMGAQLVISVAQKTNDVAGDGTTTATLLTRELYKEALKALSSGLDPNEVRKGMT
Rickettisa tsutsugamushi    TKDGVSVAKAIQLKDKSLNVGAQFVISVASKTADVAGDGTTTATVIADAAVRELNKAEVAGIDIQEVRKGAE
Cowdria ruminantium         TKDGYKVMKSIKPEDPLALAIANIIAQSASQCNDKVGDGTTTCSILTAKVIEEVSKVKAAGADIICVREGVL
Ehrlichia chaffeensis       TKDGYKVIKSIKPEDPLALAIANIITQSASQCNDKVGDGTTTCSILTAKVIEEVSKAKAAGADIVCIKEGVL
```

```
                                                  149
Ehist   IAKDIVLKEITKQSKP-TLKEDIISVARVSANNDEKIGEMVGDIFGKIGRDGAVDIETGKGTKDI-VNIVEGMVLDQGFLSRYFTTDEKNTKVDIRNTDV
Edisp   IAKDIVLKEITKQSKP-TLKEDIISVARVSANNDEKIGEMVGDIFGKIGRDGAVDIETGKGTKDI-VNIVEGMVLDQGFLSRYFTTDEKNTKVDIRNTDV
Emosh   LAKEVVLQEIVKQSKP-TSKEDIVSVARVSANNDDKIGEMIGDIFGKIGKDGAVDIETGKGTKDV-VSVVEGMVLDQGFLSRYFTTDDKNTKVDIRNADV
Eterr   MAKESVLKEIVKQSKP-TSKQDIISVARVSANNDEKIGEMIGDIFGRIGKDGAVDIETGKGTKDN-VTVVEGMVLDQGFLSRYFTTDDKNTKVDIRNADV
Einvad  TAKKVVIDEIVKQSKP-ITKSEVVSVARVSANNDEKIGSMIGDIFSKIGKDGAVDIETGKAVKDR-ISVVEGMVLDQGYLSRYFTTDDKNTKVDIHNGDV
Cpneum  KAVKVVVDELKKISKPVQHHKEIAQVATISANNDSEIGNLIAEAMEKVGKNGSITVEEAKGF-ETVLDVVEGMNFNRGYLSSYFSTNPETQECVLEDALI
Ecoli   KAVTAAVEELKALSVPCSDSKAIAQVGTISANSDETVGKLIAEAMDKVGKEGVITVEDGTGL-QDELDVVEGMQFDRGYLSPYFINKPETGAVELESPFI
Haem    KAVAAVVEELKAISKPCETSKEIEQVGTISANSDETVGKLIAQAMEKVGKEGVITVEDGTGL-DDALDVVEGMQFDRGYLSPYFINKPEAGTVELDNPYI
Pseud   KPTVAIVAQLKELAKPWRDTKAIAQVGTISANSDESIGQIIAEAMEKVGKEGVITVEEGSGL-ENELSVVEGMQFDRGYLSPYFVNKPDTMAAELDSPLL
Ngono   KAVAALVEELKNIAKPCDTSKEIAQVGSISANSDEQVGAIIAEAMEKVGKEGVITVEDGKSL-ENELDVVEGMQFDRGYLSPYFINDAEKQIAGLDNPFV
Lmicd   KAVAAVTKKLQEMSKPCKDGKAIAQVGTISANSDQAIGSIIAEAMEKVGKEGVITVEDGNSL-ENELAVVEGMQFDRGYISPYFINNQQNMSAELEHPFI
Cox     KAVTAAVAELKKISKPCKDQKAIAQVGTISANSDKSIGDIIAEAMEKVGKEGVITVEDGSGL-ENALEVVEGMQFDRGYLSPYFINNQQNMSAELENPFI
Agro    LAVAEVVKDLQAKAKKINTSEEVAQVGTISANGERQIGLDIAEAMQRVGNEGVITVEEAKTA-ETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYI
Rhiz    LAVAEVVKDLLAKAKKINTSDEVAQVGTISANGEKQIGLDIAEAMQKVGNEGVITVEEAKTA-ETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAFI
Bruc    LAVNEVVAELLKKAKKINTSEEVAQVGTISANAEAEIGKMIAEAMQKVGNEGVITVEEAKTA-ETELEVVEGMQFDRGYLSPYFVTNPEKMVADLEDAYI
Barto   AAVEAVVADLFKKAKKIQTSEEIAQVATISANGAEDIGKMIADAMEKVGNEGVITVEEAKTA-ETELEVVEGMQFDRGYLSPYFVTNSEKMMVDLDDPYI
Brady   LAVEAVVADLQKNSKKVTSNDEIAQVGAISANGDQEIGKFLADAVKKVGNEGVITVEEAKSL-ETELDVVEGMQFDRGYISPYFVTNADKMRVEMDDAYI
Human   LAVDAVIAELKKQSKPVTTPEEIAQVATISANGDKEIGNIISDAMKKVGRKGVITVKDGKTL-NDELEIIEGMKFDRGYISPYFINTSKGQKCEFQDAYV
Zea     MAVDAVVTNLKGMARMISTSEEIAQVGTISANGEREIGELIAKAMEKVGKEGVITIADGNTL-YNELEVVEGMKLDRGYISPYFITNPKTQKCELEDPLI
Yeast   VAVEKVIEFLSANKKEITTSEEIAQVATISANGDSHVGKLLASAMEKVGKEGVITIREGRTL-EDELEVTEGMRFDRGFISPYFITDPKSSKVEFEKPLL
Tcruz   RAVGVILQSVAEQNRKVTSTENIVQVATISANGDEELGRLIGQAMEKVGKDGVITTQDGKTM-TTELEVVEGMSIDRGYISPYFVTDAKAQKAELEDAFV
Tbruc   RAVEVILKNIESQSRTVTNTENVVQVATISANGDVELGKLIGEAMEKVGKDGVITTQDGKTL-TTELEVVEGMSVDRGYISPYFVTDAKTQKAELEDAFV
Trich   LAVDAAVAEIKKLSRKVSSDSEIAQVATVSANGDHTIGELIAKAFKAVGQEGVITVQNGNSF-EHKLDVVEGMKIDRGYLSAFFMTNNKTMKCEYENPYI
Rick    KAVEAVIADVRKNSSPVKNEEEIAQVATVSSNGDREIGEKIANAMKQVGQEGVITVEDSKNF-NFEVEVVKGMRFDRGYISQYFATNREKMITEFENPYI
Cowdr   KAKEAVLEALKCMKREVLSEEEIAQVATISANGDKNIGTKIAQCVKEVGKDGVITVEESKGFKELDVEKTDGMQFDRGYLSPYFVTNSEKMLVEFENPYI
Ehrl    KAKEAVLEALMSMKREVLSEEEIAQVATISANGDKNIGSKIAQCVQEVGKDGVITVEESKGFKELDVEKTDGMQFDRGYLSPYFVTNSEKMLVEFENPYI
```

```
      218                                                                277
Ehist IVCDYKLSSSQSVVPLLELCLKRKRPLVVISDTIDGDALTTLVLNKLRG-LPIAAVRAPGFGETRKGILHDIGIITGATVISNEAGKKIEEVTEKDLGKIGHF
Edisp IVCDYKLNSSQNVVPLLELCVKRKRPLVIISDTIDGDALTTLVLNKLRG-LPIAAVRAPGFGETRKGILHDIGIITGATVISNEVGKKIGEVTEKDLGKIGHF
Emosh LVCDYKLDSSQSVVPLLELCVKRKRPLVIIADTIDGDALTTLVLNKLRG-LPISAVRAPGFGDTRKGILHDIGIITGATVISNDAGKKVEDVTEKDLGKIGHF
Eterr VVCDYKLSTSQNVVPLLQLCVKRKNPLVIIADTIDGDALTTLVLNKLRG-LPIAAVRAPGFGETRKGIIHDIGIITGATVISNDAGKKVEDIAEKQLGRIGHF
Einva IVCDAKLETSQQVVPLLQMCVKRKRPLIIIADTIDGDALTTLVLSKLRG-LPVAAVRAPGFGETRKGILQDIGIITGATVISNDTGNNTEKITESQLGKCGHF
Cpneu LIYDKKISGIKDFLPVLQQVAESGRPLLIIAEEIEGEALATLVVNRLRAGFRVCAVKAPGFGDRRKAMLEDIAILTGGQLVSEELGMKLENTTLAMLGKAKKV
Ecoli LLADKKISNIREMLPVLEAVAKAGKPLLIIAEDVEGEALATAVVNTIRGIVKVAAVKAPGFGDRRKAMLQDIATLTGGTVISEEIGMELEKATLEDLGQAKRV
Haem  ILVDKKISNIREILPVLEAVAKAGKPLLIIAEDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQDIAILTAGTVISEEIGMELEKATLEELGQAKRV
Pseud LLVDKKISNIREMLPVLEAVAKAGRPLLIVAEDVEGEALATLVVNNMRGIVKVAAVKAPGFGDRRKAMLQDIAILTGGTVISEEVGLSLEGATLEHLGKPKRV
Ngono LLFDKKISNIRDLLPVLEQVAKASRPLLIIAEDVEGEALATLVVNNIRGVLKTVAVKAPGFGDRRKAMLQDIAILTGAVVISEEVGLSLEKATLDDLGQAKRI
Lmicd LLVDKKISTIRDMLSVLEAVAKSGPSLLIIAEDVEGEALATLVVNNMRGIVKVCAVKAPGFGDRRKAMLQDIAILTAGEVISEEVGTSLESATLDSLGTAKRV
Cox   LLVDKKISNIRELIPLLENVAKSGRPLLVIAEDIEGEALATLVVNNIRGVVKVAAVKAPGFGDRRKAMLQDIAVLTGGKVISEEVGLSLEAASLDDLGSAKRV
Agro  LLHEKKLSNLQAMLPVLEAVVQTGKPLVIIAEDVEGERLATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTVISEDLGIKLESVTLDMLGKSKKV
Rhiz  LLHEKKLSNLQAMLPVLEAVVQTGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGTVISEDLGIKLESVTLDMLGRAKKV
Bruc  LLHEKKLSNLQALLPVLEAVVQTSKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAILTGGQVISEDLGIKLESVTLDMLGRAKKV
Barto LIHEKKLSNLQSLLPVLEAVAQSGKPLLIIAEDVEGEALATLVVNKLRGGLKIAAVKAPGFGDRRKAMLEDIAVLTSGQVISEDVGIKLENVTLEMLGRAKKV
Brady LINEKKLSSLNELLPLLEAVVQTGKPLVIVAEDVEGEALATLVVNRLRGGLKVAAVKAPGFGDRRKAMLQDIAILTGGQAISEDLGIKLENVTLNMLGRAKKV
Human LLSEKKISSIQSIVPALEIANAHRKPLVIIAEDVDGEALSTLVLNRLKVGLQVVAVKAPGFGDNRKNQLKDMAIATGGAVFGEELTLNLEDVQPHDLGKVGEV
Zea   LIHDKKVTNMHAVVKVLEMALKKQKPLLIVAEDVESEALGTLIINKLRAGIKVCAVKAPGFGENRKANLQDLAILTGGEVITEELGMNLENFEPHMLGTCKKV
Yeast LLSEKKISSIQDILPALEISNQSRRPLLIIAEDVDGEALAACILNKLRGQVKVCAVKAPGFGDNRKNTIGDIAVLTGGTVFTEELDLKPEQCTIENLGSCDSI
Tcruz LVSAKKVSSIHTILPALNHVVGTGRPLLIIADDVESEALTTMIFNKLQGKLKIACVKAPGFGDNKTAMMQDIAIFAGARLVGEELELDAENFDPAILGTVKKA
Tbruc LVSAKKLNNIHTILPVLNHVVRSGRPLLIIADDVESEALTTMIFNKLQGKLKIACVKAPGFGDNKAAMLQDIAIFSGACVVGEEVELDAEKFEASILGSVKKA
Trich LITDMKISSFATIAPALEACIHANRPLLIIADDVEGDALATLIVNKIRGGLKVCAVKAPGFGDNKKATLQDIATVTGGQVIAEELGLKLESIQLSQLGQCSKI
Rick  LLLDQKVSTVQPLVPVLEAVAHTGKPLVLIADDVDGEALTALILNNLKGSIKVVAVKAPGFGDRKKEMLEDIAILTNGEVITEQLGIKLEKVDTSKLGTANRV
Cowdr LLTEKKLNIIQPLLPILENIARSGRPLLIIAEDVEGEALSTLVLNKLRGGLHVAAVKAPGFGDRRKDMLGDIAILTGAKVINDELAIKMEDLTLCDLGTAKNI
Ehrl  LLTEKKLNIIQPILPILENVARSGRPLLIIAEDVEGEALSTLVLNKLRGGLHVAAVKAPGFGDRRKDMLGDIAILTGAKVISDDLAIKMEDLTLAELGTAKNI
```

```
                330                                                  380
Ehist VSTKDETIITGGAGSKAEVLARINELKNAKEVSDSSYEKEKLEGRIARLTGGVAVISVGGSSEAEVGERKDRIEDAVCAVKAALAEGIVPGGG
Edisp VSTKDETIITGGAGSKKEVLTRINELKKAKEISDSNYEKEKLEGRIARLTGGVAVISVGGSSEAEVGERKDRIEDAVYAVKAALAEGIVPGGG
Emosh LSSKDETVITGGAGSKAEVLARVAELKAAKEASDSTYEKEKLQNRIARLTGGVAVVSVGGSSETEVGERKDRNEDAVCAVKAALAEGIVPGGG
Eterr LSSQDETVITGGAGSKAEVVARVEELKAARDASDSTYEKEKLQGRIARLTGGVAVVSVGGSSEAEVGERKDRIEDAVCAVKAALSEGIVPGGG
Einva VSNQDESVITGGAGKKSDVLQRVDELREALRKSESLYEKEKLQGRIARLTGGVAVISVGGSSEAEVGERKDRIEDAVCAVKAALSEGIVPGGG
Cpneu IVTKEDTTIVEGLGNKPDIQARCDNIKKQIEDSTSDYDKEKLQERLAKLSGGVAVIRVGAATEIEMKEKKDRVDDAQHATIAAVEEGILPGGG
Ecoli VINKDTTTIIDGVGEEAAIQGRVAQIRQQIEEATSDYDREKLQERVAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVAGGG
Haem  VITKDNTTIIDGIGDEAQIKARVVQIRQQIEDSTSDYDKEKLQERVAKLAGGVAVIKVGAATEVAMKEKKDRVDDALHATRAAVEEGIVPGGG
Pseud VINKENTTIIDGAGVQADIEARVLQIRKQIEETTSDYDREKLQERLAKLAGGVAVIKVGAATEVEMKEKKAPVEDALHATRAAVEEGVVPGGG
Ngono EIGKENTTVIDGFGDAAQIEARVAEIRQQIETATSDYDKEKLQERVAKLAGGVAVIKVGAATEVEMKEKKDRVEDALHATRAAVEEGVVAGGG
Lmicd VVTKENTTIIDGEGKAADINARITQIRAQMEETTSDYDREKLQERVAKLAGGVAVIKLVLLP--NRMKRKARVEDALHATRAAVEEGIVAGGG
Cox   VVTKDDTTIIDGSGDAGDIKNRVEQIRKEIENSSSDYDKEKLQERLAKLAGGVAVIKVGAATEVEMKEKKARVEDALHATRAAVEEGVVPGGG
Agro  SISKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRIDDALNATRAAVQEGIVPGGG
Rhiz  SITKENTTIVDGAGQKSDIEGRVAQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRIDDALNATRAAVQEGIVPGGG
Bruc  SISKENTTIVDGAGQKAEIDARVGQIKQQIEETTSDYDREKLQERLAKLAGGVAVIRVGGATEVEVKEKKDRVDDALNATRAAVEEGIVAGGG
Barto HVSKETTTIVDGAGQKSEINARVSQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGSTEVEVKEKKDRVDDALNATRAAVEEGIVPGGG
Brady MIDKENTTIVNGAGKKADIEARVAQIKAQIEETTSDYDREKLQERLAKLAGGVAVIRVGGATEVEVKERKDRVDDAMHATRAAVEEGIVPGGG
Human IVTKDDAMLLKGKGDKAQIEKRIQEIIEQLDVTTSEYEKEKLNERLAKLSDGVAVLKVGGTSDVEVNEKKDRVTDALNATRAAVEEGIVLGGG
Zea   TVSKDDTVILDGAGDKKSIEERAEQIRSAIENSTSDYDKEKLQERLAKLSGGVAVLKIGGASEAEVGEKKDRVTDALNATKAAVEEGIVPGGG
Yeast TVTKEDTVILNGSGPKEAIQERIEQIKGSIDITTTNYEKEKLQERLAKLSGGVAVIRVGGASEVEVGEKKDRYDDALNATRAAVEEGILPGGG
Tcruz TITKDDTVLLNGGGESSMVKERVELLRGLIDGETSDYNREKLQERLAKLSGGVAVIKVGGGSEVEVNEKKDRITDALCSTRAAVQEGIVPGGG
Tbruc TITKDDTVLLNGGGDVAMMKERVDLVRGLIERETSDYNREKLQERLAKLSGGVAVIRVGGASEVEVNEKKDRITDALCSTRAAVQEGIVPGGG
Trich TVSKDDCIVMGGAGQKDSIKGRAEEIRKQLESTESKYEKDKLKERLAKLTGGVAVINVGGASEVEVNETKDLIDDALNATRAAIEEGIVAGGG
Rick  IVTKDHTTIVHDKNNEKKVNSRCEQIREAIKDTTSDYEKEKLQERLAKLRNGVAVLKVGGATEVEQKERKDRVEDALHATRAAVEEGIVPGGG
Cowdr RITKDTTTIIGSVDNCAHVQSRICQIRMQIDNSTSDYDKEKLQERLAKLSGGVAVLKVGGSSEVEVKERKDRVEDALHATRAAVEEGVVPGGG
Ehrl  RITKDTTTIIGSVDNSANVQSRINQIKMQIEASTSDYDKEKLRERLAKLSGGVAVLKVGGSSEVEVKERKDRVEDALHATRAAVEEGVVPGGG
```

APPENDIX D

Alignment of PNT amino acid sequences used for the phylogenetic analyses.

```
Entamoeba terrapinae          GGADMPVVVSMLNSYSGWATAAAGFLLNNYAMIVGGALIGSSGAILSYIMCKAMNRHFLSVIFGTFG-----E
Entamoeba moshkovskii (Lar)   GGADMPVVVSMLNSYSGWATAASGFLLNNYAMIVGGALIGSSGAILSYIMCKAMNRSFLSVIFGGFG------
Entamoeba moshkovskii FIC     GGADMPVVVSMLNSYSGWATAASGFLLNNYAMIVGGALIGSSGAILSYIMCKAMNRSFLSVIFGGFG------
Entamoeba histolytica         GGADMPVVVSMLNSYSGWATAASGFLLNNYAMIVGGALIGSSGAILSYIMCKAMNRSFMSVIFGGFG------
Entamoeba dispar              GGADMPVVVSMLNSYSGWATAASGFLLNNYAMIVGGALIGSSGAILSYIMCKAMNRSFMSVIFGGFG------
Entamoeba invadens            GGADMPVVISMLNSYSGWATAAAGFLLNNYAMIVSGALIGSSGAILSYIMCAAMNRGFLSVILGIGT-HKTET
Plasmodium falciparum         GAADMPVIISVLNSYSGFATAISGFLLHNNLLIISGALIGSSGAILSYIMCIGMNRDIFNIILGGWDDYENMG
Rhodospirillum                GGADMPVVISMLNSYSGWAAAGIGFTLGNPLLIIAGALVGSSGAILSYIMCKGMNRSIFNVILGGFG------
Escherichia coli              GGADMPVVVSMLNSYSGWAAAAAGFMLSNDLLIVTGALVGSSGAILSYIMCKAMNRSFISVIAGGFG------
Homo sapiens                  GGADMPVVITVLNSYSGWALCAEGFLLNNNLLTIVGALIGSSGAPLSYIMCVAMNRSLANVILGGYG------
Eimeria tenella               GGADMPVVISLLNSYSGVALAASGFMLDNNLLIIAGALIASSGAILSYIMCKGMNRSLWNVVLGGFE------
Haemophilus influenzae        GGADMPVVVSMLNSYSGWAAAAAGFILNNDLLIVTGALVGSSGAILSYIMCKAMNRSFVSVIAGGFG------
Mycobacterium tuberculosis    GGADMPVVISMLNAMTGLSAAAAGLALNNTAMIVAGMIVGASGSILTNLMAKAMNRSIPAIVAGGFG------
Rickettsia                    GSADMPVIVSMLNSYSGFATAGIGFTLSNSLLIITGSLIGSSGAILSYIMCKAMNRSLIKVIFGAFL------
Neisseria gonorrhoeae         GGADMPVVVSMLNSYSGWAAAAAGFMLSNDLLIVTGALVGSSGAILSYIMCKAMNRSFVSVIAGGFG------
Synechocy                     GGADMPVVISMLNSYSGWAAAAAGFMLSNDLLIITGALVGSSGAILSYIMCKAMNRSFISVILGGFG-----E
Vibrio cholera                GGADMPVVVSMLNSYSGWAAAAAGFMLANDLLIVTGALVGSSGAILSYIMCKAMNRSFISVIAGGFG------
```

```
E.terrapin    GAATKKSNTEDEGPKEINPIQAPELAKLLMESHNIAIVPGYGMAVAKAQHVVANLAEQLIKAGKQVRFIIHPVAGRLP
EmLaredo      ATPTVTSKHDDEGPREANPIQTQELAKLLIDAQNIAIVPGYGMAVAKAQHVVADLADQLIKSGKKVRFIIHPVAGRFP
EmFIC         ATPTVTSAHEDEGPKEVNPIQTQELAKLLIDAKNVAIIPGYGMAVAKAQHVVAELSNILTQAGKQVRFIIHPVAGRIP
E.histolyt    ATPSKTRNQEDEGPKEANTIQTPELAKLLMEAHNIAIVPGYGMAVAKAQHVVASLAEELIKAGKEVRFIIHPVAGRLP
E.dispar      ATTNKTKNQEDEGPKEANTIQTAELAKILMEAHNIAIVPGYGMAVAKAQHVVANLAEELIKAGKEVKFIIHPVAGRLP
E.invadens    AASPVTKNEKDTGNKEVNPIEVGEFAKLLVDSHKIAIVPGYGMAVAKAQHVVAELANMLTKAGKTVNFIIHPVAGRLL
P.falcipar    ESIYDQNFIEKKNKQTINSTTNKYVAENLINAKNIIIVPGYGTALSKCQRELAEICSILTSRNINVTFAIHPVAGRMP
Rhodospir     SEGGVAAAGGAAGDRSVKAGSAEDAAFIMKNASKVIIVPGYGMAVAQAQHALREMADVLKKEGVEVSYAIHPVAGRMP
Esch.coli     --TDGSSTGDDQEVGEHREITAEETAELLKNSHSVIITPGYGMAVAQAQYPVAEITEKLRARGINVRFGIHPVAGRLP
H.sapiens     -TTSTAGGKPMEISGTHTEINLDNAIDMIREANSIIFTPGYGLCAAKAQYPIADLVKMLTEQGKKVRFGIHPVAGRMP
E.tenella     ---EAEDVGAASPQGAVQQATADQVADELLAARKVLIVPGYGMAVARCQSELADIAKNLMNCGITVDFGIHPVAGRMP
Haemophil     ---NDVQVSSSEEQGEHRETTAEEVAELLKNASSVIITPGYGMAVAQAQYPVADITAKLRERGVNVRFGIHPVAGRLP
Mycobact      --GGGVAPSGGGDDKHVKATSAADAAIQMAYANQVIVVPGYGLAVAQAQHAVKDLATLLEDRGVPVKYAIHPVAGRMP
Rickettsia    -PPPTGINKDICDDKIAKISTPEEAAHLLLNASSVIIVPGYGMAVAQSQHSIKEMVDILERSDINVRFAIHPVAGRMP
Neisseria     -SDSGTLSSGSQEIGEYREVKAADIAEMLKGANNVIITPGYGMAVAQAQYPVAEITELLRKNGTEVRFGIHPVAGRLP
Synechocy     GSSAASKGPAQEVIGTVTKTSAEEVAEELLDAQSVIITPGYGMAVAQAQHPVAEITKLLKEKGVKVRFGIHPVAGRLP
Vibrio        ---QEVVISSDEEQGEHRETSAEEVAEMLKNSKSVIITPGYGMAVAQAQYPVYEITEKLRAQGVTVRFGIHPVAGRLP
```

APPENDIX E

The complete coding sequences of the PNT and cpn60 genes of each of the *Entamoeba* species were used to construct a codon usage table for each organism using the DNA Strider 1.2 programme (177). The tables were used for codon usage comparisons between the species. The complete up to date codon usage tables for the *E. histolytica*, *E. dispar* and *E. invadens* genes sequenced until now can be found at http://www.kazusa.or.jp/codon/.

Table E1. *Entamoeba histolytica* HM-1:IMSS codon usage based on the PNT and cpn60 genes.

TTT phe F	9	TCT ser S	7	TAT tyr Y	7	TGT cys C	6
TTC phe F	2	TCC ser S	-	TAC tyr Y	-	TGC cys C	-
TTA leu L	30	TCA ser S	20	TAA OCH Z	-	TGA OPA Z	-
TTG leu L	-	TCG ser S	-	TAG AMB Z	-	TGG trp W	1
CTT leu L	7	CCT pro P	2	CAT his H	5	CGT arg R	2
CTC leu L	-	CCC pro P	-	CAC his H	-	CGC arg R	-
CTA leu L	3	CCA pro P	10	CAA gln Q	7	CGA arg R	1
CTG leu L	-	CCG pro P	-	CAG gln Q	-	CGG arg R	-
ATT ile I	37	ACT thr T	9	AAT asn N	17	AGT ser S	7
ATC ile I	1	ACC thr T	-	AAC asn N	3	AGC ser S	1
ATA ile I	5	ACA thr T	24	AAA lys K	37	AGA arg R	17
ATG met M	10	ACG thr T	-	AAG lys K	5	AGG arg R	-
GTT val V	29	GCT ala A	21	GAT asp D	22	GGT gly G	11
GTC val V	-	GCC ala A	-	GAC asp D	6	GGC gly G	-
GTA val V	17	GCA ala A	28	GAA glu E	34	GGA gly G	37
GTG val V	-	GCG ala A	-	GAG glu E	-	GGG gly G	5

Table E2. ***Entamoeba dispar*** **SAW 760 codon usage table based on the PNT and cpn60 genes**

TTT phe F	10	TCT ser S	10	TAT tyr Y	7	TGT cys C	5
TTC phe F	2	TCC ser S	-	TAC tyr Y	1	TGC cys C	-
TTA leu L	27	TCA ser S	17	TAA OCH Z	-	TGA OPA Z	-
TTG leu L	1	TCG ser S	-	TAG AMB Z	-	TGG trp W	1
CTT leu L	6	CCT pro P	1	CAT his H	5	CGT arg R	1
CTC leu L	3	CCC pro P	-	CAC his H	-	CGC arg R	1
CTA leu L	3	CCA pro P	10	CAA gln Q	9	CGA arg R	-
CTG leu L	-	CCG pro P	1	CAG gln Q	-	CGG arg R	-
ATT ile I	40	ACT thr T	9	AAT asn N	19	AGT ser S	5
ATC ile I	-	ACC thr T	-	AAC asn N	2	AGC ser S	-
ATA ile I	5	ACA thr T	24	AAA lys K	40	AGA arg R	18
ATG met M	8	ACG thr T	2	AAG lys K	4	AGG arg R	-
GTT val V	28	GCT ala A	20	GAT asp D	29	GGT gly G	14
GTC val V	2	GCC ala A	3	GAC asp D	3	GGC gly G	2
GTA val V	17	GCA ala A	22	GAA glu E	28	GGA gly G	30
GTG val V	-	GCG ala A	-	GAG glu E	-	GGG gly G	7

Table E3. ***Entamoeba invadens*** **IP-1 codon usage table based on the PNT and cpn60 genes**

TTT phe F	6	TCT ser S	6	TAT tyr Y	3	TGT cys C	2
TTC phe F	3	TCC ser S	5	TAC tyr Y	4	TGC cys C	4
TTA leu L	11	TCA ser S	5	TAA OCH Z	-	TGA OPA Z	-
TTG leu L	12	TCG ser S	7	TAG AMB Z	-	TGG trp W	2
CTT leu L	7	CCT pro P	2	CAT his H	1	CGT arg R	2
CTC leu L	5	CCC pro P	-	CAC his H	5	CGC arg R	2
CTA leu L	-	CCA pro P	5	CAA gln Q	9	CGA arg R	4
CTG leu L	2	CCG pro P	6	CAG gln Q	3	CGG arg R	1
ATT ile I	19	ACT thr T	15	AAT asn N	6	AGT ser S	12
ATC ile I	16	ACC thr T	6	AAC asn N	16	AGC ser S	1
ATA ile I	6	ACA thr T	8	AAA lys K	23	AGA arg R	7
ATG met M	11	ACG thr T	6	AAG lys K	19	AGG arg R	2
GTT val V	23	GCT ala A	23	GAT asp D	14	GGT gly G	16
GTC val V	15	GCC ala A	7	GAC asp D	16	GGC gly G	7
GTA val V	4	GCA ala A	8	GAA glu E	14	GGA gly G	19
GTG val V	11	GCG ala A	12	GAG glu E	13	GGG gly G	9

Table E4. ***Entamoeba moshkovskii*** **Laredo codon usage table based on the PNT and cpn60 genes**

TTT phe F	10	TCT ser S	8	TAT tyr Y	7	TGT cys C	5
TTC phe F	7	TCC ser S	5	TAC tyr Y	1	TGC cys C	-
TTA leu L	15	TCA ser S	14	TAA OCH Z	-	TGA OPA Z	-
TTG leu L	9	TCG ser S	3	TAG AMB Z	-	TGG trp W	1
CTT leu L	10	CCT pro P	1	CAT his H	5	CGT arg R	-
CTC leu L	6	CCC pro P	1	CAC his H	3	CGC arg R	1
CTA leu L	7	CCA pro P	10	CAA gln Q	11	CGA arg R	1
CTG leu L	-	CCG pro P	1	CAG gln Q	2	CGG arg R	3
ATT ile I	27	ACT thr T	14	AAT asn N	10	AGT ser S	4
ATC ile I	5	ACC thr T	1	AAC asn N	11	AGC ser S	3
ATA ile I	4	ACA thr T	13	AAA lys K	34	AGA arg R	13
ATG met M	11	ACG thr T	3	AAG lys K	8	AGG arg R	2
GTT val V	30	GCT ala A	12	GAT asp D	24	GGT gly G	12
GTC val V	7	GCC ala A	9	GAC asp D	12	GGC gly G	4
GTA val V	11	GCA ala A	20	GAA glu E	18	GGA gly G	22
GTG val V	4	GCG ala A	5	GAG glu E	4	GGG gly G	9

Table E5. ***Entamoeba terrapinae*** **M codon usage table based on the PNT and cpn60 genes**

TTT phe F	5	TCT ser S	6	TAT tyr Y	3	TGT cys C	4
TTC phe F	6	TCC ser S	4	TAC tyr Y	4	TGC cys C	1
TTA leu L	4	TCA ser S	13	TAA OCH Z	-	TGA OPA Z	-
TTG leu L	17	TCG ser S	2	TAG AMB Z	-	TGG trp W	1
CTT leu L	8	CCT pro P	2	CAT his H	4	CGT arg R	5
CTC leu L	3	CCC pro P	1	CAC his H	2	CGC arg R	-
CTA leu L	-	CCA pro P	10	CAA gln Q	8	CGA arg R	1
CTG leu L	6	CCG pro P	1	CAG gln Q	5	CGG arg R	3
ATT ile I	23	ACT thr T	13	AAT asn N	12	AGT ser S	7
ATC ile I	18	ACC thr T	5	AAC asn N	9	AGC ser S	2
ATA ile I	-	ACA thr T	11	AAA lys K	24	AGA arg R	9
ATG met M	10	ACG thr T	4	AAG lys K	16	AGG arg R	3
GTT val V	25	GCT ala A	27	GAT asp D	22	GGT gly G	23
GTC val V	13	GCC ala A	8	GAC asp D	9	GGC gly G	2
GTA val V	4	GCA ala A	19	GAA glu E	14	GGA gly G	28
GTG val V	9	GCG ala A	1	GAG glu E	14	GGG gly G	2

ELSEVIER

Molecular and Biochemical Parasitology 110 (2000) 177–182

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

www.elsevier.com/locate/parasitology

Short communication

A mitochondrial-type hsp70 gene of *Entamoeba histolytica*☆

Christina Bakatselou, Claire Kidgell, C. Graham Clark *

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

Received 27 March 2000; received in revised form 5 May 2000; accepted 9 May 2000

Keywords: *Entamoeba histolytica*; Heat shock; Hsp70; Mitochondria; Phylogeny

Until recently it was believed that *Entamoeba histolytica* belonged to an early branching eukaryotic lineage, as it lacks organelles such as mitochondria, peroxisomes, rough endoplasmic reticulum, Golgi dictyosomes, and centrioles [1], and has an unusual glycolytic metabolism [2]. However, ribosomal RNA-based phylogenetic trees place *E. histolytica* on a branch arising more recently than several lineages with typical eukaryotic organelles and metabolism [3–5]. This indicates that the situation in *E. histolytica* represents a derived rather than a primitive state.

Evidence for secondary loss of mitochondrial function in *E. histolytica* has also been obtained through the analysis of two genes that encode proteins normally found in the mitochondrion: pyridine nucleotide transhydrogenase (PNT) and the mitochondrial chaperonin cpn60. The latter is present only in organelles in eukaryotes and is homologous to the GroEL family of bacterial chaperonins. Both *E. histolytica* genes appear to encode signal peptides at the amino termini of the protein [6,7], and the cpn60 has been localised to an organelle called the mitosome [8] (or 'Crypton' [9]). Organelle localisation of the protein is dependent on the putative signal peptide, which can be functionally replaced by an authentic mitochondrial targeting sequence [8]. Phylogenetic analysis of cpn60 also shows a clear link between the *E. histolytica* gene and the mitochondrial lineage [6,7].

Since the ancestors of *E. histolytica* appear to have had a mitochondrial endosymbiont it is likely that more 'mitochondrial' genes will be found in this organism. Demonstration of their presence would further support the secondary loss hypothesis, and reduce the likelihood of the *E. histolytica* cpn60 representing a separate gene acquisition. Here we report the isolation and characterisation of a gene encoding a

☆ *Note*: Nucleotide sequence data reported in this paper have been submitted to the GenBank™, EMBL and DDJB databases with the accession numbers AF214549 (genomic DNA) and AF214550 (cDNA).

* Corresponding author. Tel.: +44-207-9272351; fax: +44-207-6368739.

E-mail address: graham.clark@lshtm.ac.uk (C. Graham Clark).

PII: S0166-6851(00)00264-4

mitochondrial-type heat shock protein 70 homologue (mt-hsp70) from *E. histolytica*.

Using degenerate primers based on conserved amino acid residues observed in a published alignment of hsp70 sequences [10], 700 bp fragments of *E. histolytica* hsp70 genes were amplified by PCR from genomic DNA and cloned. Partial sequence analysis identified a mitochondrial-type hsp70 isoform as well as cytoplasmic and endoplasmic reticulum hsp70 isoforms.

Two *E. histolytica* libraries were screened with the mt-hsp70 gene fragment: a cDNA library of strain SFL-3 [11], and a genomic DNA library of strain HM-1:IMSS [12] both in λZap II. Only two cDNA plaques out of 60 000 screened were positive indicating it is a very rare message. An insert of approximately 1800 bp was present in both cDNA clones but the coding region was incomplete at the 5′ end. Screening of the genomic library yielded three clones containing the 5′ end of the gene. The insert in each was approximately 1.3 kb, of which ca. 1 kb was upstream of the coding region. No significant open reading frames were identified in this upstream region and a database search did not find any close matches. A composite genomic DNA–cDNA sequence revealed an open reading frame of 1791 bp encoding a 597 amino acid mitochondrial-type hsp70 of *E. histolytica*.

A comparison of the 5′ flanking regions of the mt-hsp70 with those of other *E. histolytica* genes to identify possible regulatory sequences [13,14] did not reveal any close similarities. A putative polyadenylation signal [15] occurs approximately 20 bases upstream of the polyadenylation site and overlaps the translation termination codon, something that also occurs in the cpn60 gene of *E. histolytica* [8].

Southern blot hybridisation with the mt-hsp70 gene as a probe gave two bands with some enzymes that do not cut within the coding region and one with others. This result would normally indicate that there are two loci for the mt-hsp70 gene. However, hybridisation with an upstream region probe gave the same bands as with the coding region probe. Unless the upstream sequence is conserved between the two loci, this suggests that the two bands may correspond instead to two alleles of the mt-hsp70 gene that differ slightly in the flanking region. The genome of *E. histolytica* is thought to be tetraploid [16] and alleles of certain genes differ in sequence [17]. The situation in the case of the mt-hsp70 has not been fully resolved. No sequence differences between PCR products and cDNAs have been found.

Northern blot analysis revealed two transcripts hybridising to the mt-hsp70 probe that differ in length. The main band is approximately 1.8 kb in size and the other is approximately 150 bases longer (Fig. 1A). The main mt-hsp70 message appears to be slightly induced during recovery from heat shock when compared with the control β-actin mRNA, while the larger mRNA seems to be more strongly induced (Fig. 1B). The induced message level is highest after 24 h of recovery (Fig. 1B, lane d). The actin is non-inducible, the difference in the intensity among the bands being due to the amount of the mRNA loaded.

To verify that the upper band and the observed induction were not artefacts, we hybridised the same filter with a probe for chaperonin cpn60, a gene previously shown not to be induced by heat shock [8]. The result, which showed a single non-induced band for cpn60, confirms that the mt-hsp70 probe is detecting two different transcripts at least one of which is induced following heat shock (Fig. 1B).

As no length variation was detected in the coding region using PCR, 3′ RACE was used to investigate the 3′ untranslated region. Only one product size was detected. The 5′ ends of the mRNAs were studied using a 300 bp probe located immediately upstream of the initiation codon. Neither mt-hsp70 RNA hybridised but the upstream probe detected instead a polyA$^+$ RNA of approximately 1000 bases in length that also appears to be induced by heat shock (Fig. 1B). The upstream region thus seems to be transcribed despite the fact that no open reading frame was found in this region. Several classes of non-translated polyadenylated RNAs have been found in this organism but those described to date have been much more abundant [18].

The upper band is similar in size to that expected for the cytoplasmic or endoplasmic reticu-

lum forms of hsp70 (approx. 1950 bp). To rule out the possibly of weak cross-hybridisation to the mt-hsp70 probe the 3′ end of the mt-hsp70 gene was used. Within this probe the closest matches to the other sequences were 16 out of 17 bp and 27 out of 36 bp. Under the stringent wash conditions used (0.2 × SSC/0.1% SDS for 30 min at 65°C) it is very unlikely that this mt-hsp70 probe will cross-hybridise to either of the other hsp70 form mRNAs. However, with this specific 3′ end probe the same two-band hybridisation pattern was obtained. If the probe is not cross hybridising the only remaining possibility is variation in the length of the polyA$^+$ tract, which may affect message stability.

The complete inferred amino acid sequence of the mt-hsp70 gene of *E. histolytica* was aligned by hand to mitochondrial, hydrogenosomal and related bacterial hsp70 homologues from 15 species and subjected to a variety of phylogenetic analyses. Ambiguous areas of the alignment were not included in the phylogenetic analysis, leaving 521 alignment positions in the final data set. Phylogenetic trees were obtained by maximum parsimony (MP), distance, quartet puzzling (QP) and maximum likelihood (ML) analyses and their topologies were very similar. Bootstrap support for the nodes of interest gives strong support for a mitochondrial origin of the *E. histolytica* mt-hsp70 gene (Fig. 2A). The branching order within the 'mitochondrial' clade was poorly resolved. However, there is no support for placing the *E. histolytica* sequence outside the 'mitochondrial' clade and little for placing it as the deepest branch within this clade. As is the case for many of its other genes, the evolutionary rate for the *E. histolytica* mt-hsp70 is clearly very high as seen by its long branch length in the tree (Fig. 2A). The mitochondrial clade in the hsp70 phylogenetic tree forms a sister-group to the α-proteobacteria, as is the case for many mitochondrial proteins. The

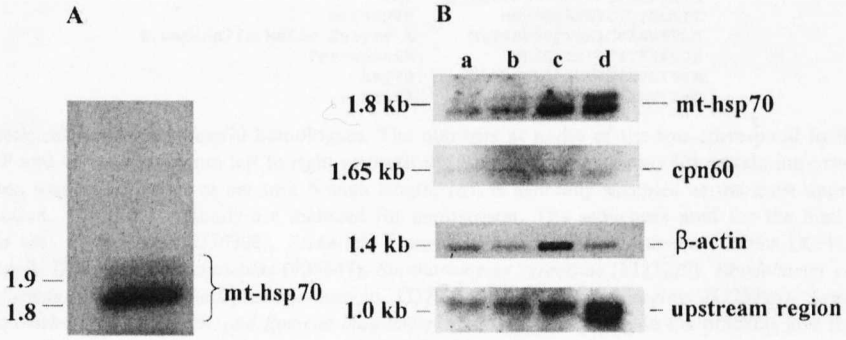
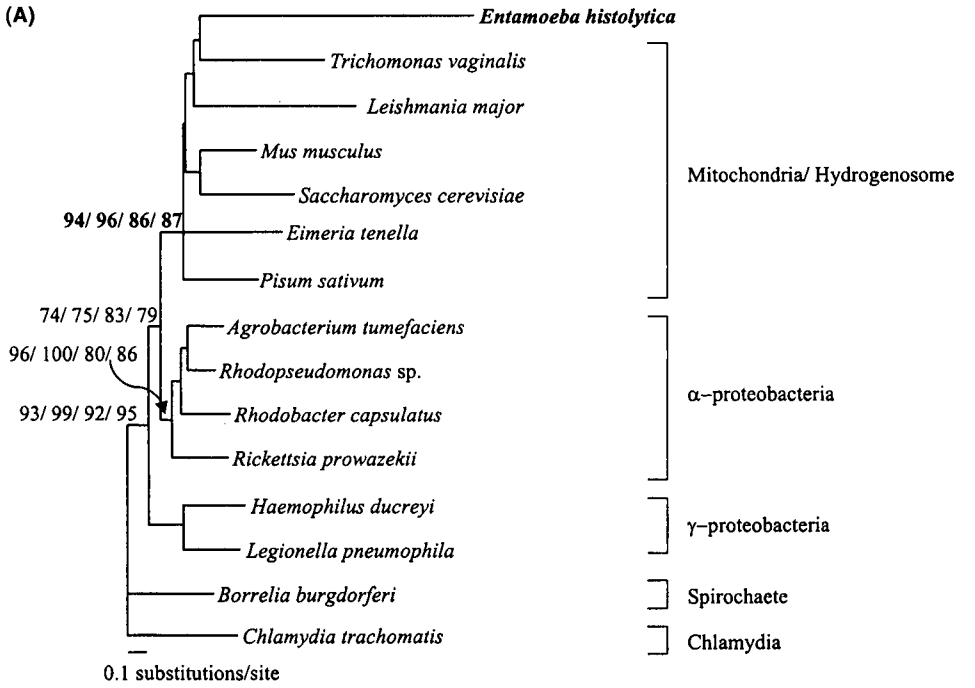


Fig. 1. Northern blot analysis of *E. histolytica* mt-hsp70. Trophozoites of *E. histolytica* strain HM-1: IMSS clone 9 were grown axenically in YI-S medium with 15% adult bovine serum at 36°C as described [24]. Polyadenylated RNA was extracted using the QuickPrep®Micro mRNA purification kit (Amersham Pharmacia Biotech). (A) PolyA$^+$ RNA was extracted from non-heat shocked trophozoites in the logarithmic phase of growth, separated in an agarose gel, blotted and hybridised to a mt-hsp70 coding region probe. (B) PolyA$^+$ RNA was extracted from trophozoites before heat shock (lane a), and following a 3 h heat shock at 39°C with recovery at 36°C for 0 (lane b), 3 (lane c) and 24 h (lane d). After gel electrophoresis and blotting, the filter was sequentially hybridised to four probes, which were randomly labelled with radioactive [α-^{32}P]-dCTP using the Rediprime system (Amersham Pharmacia Biotech). After hybridisation, blots were exposed to X-ray film at −70°C for up to 2 weeks, then stripped in boiling 0.1% SDS/DEPC-treated H_2O before rehybridising to the other probes. The following primer pairs were used to obtain probe DNAs: degenerate primers 1 (5′-GGN ATH GAY YTN GGN ACN AC-3′) and 2 (5′-GGR TCR AAR TCN TCN CCN CC-3′), or the mt-hsp70 specific primers 70.4 (5′-CCT TTG TCA AGT GCA GAG AC-3′) and 70.9 (5′-GCT TGG AGG AGA AGA TTT TG-3′) were used to amplify coding region probes from the mt-hsp70 cDNA; β-actin A (5′-GGA TTT GCT GGT GAT GAT-3′) and β-actin B (5′-TCA GTA AGA TCA CGT CCT-3′) for the β-actin gene; N-CPN60 (5′-TAT AGG GAT CCG CAT GCT TTC ATC TTC AAG-3′) and C-CPN60 (5′-TAC TAG ATA TCA ATG TCT TCT TTT AGT GT-3′) for the cpn60 gene; and US-5 (5′-TTT GGG AGT GAG TGT GGT GG-3′) and US-3 (5′-TCA GAT TCT GTG AGA GAA AAA TAA C-3′) for the upstream region of the mt-hsp70 gene. The sizes of the mRNAs were measured by comparison to an RNA marker (0.28–6.58kb; Promega).



(B)

E.histolytica cpn60	MLSSSSHYNGKLLSLNIDC
PNT	MSTSSSIEEEVFNYMKIINNFVSVG
mt-hsp70	MFVSQPARSTCIGIDLGTT
T.vaginalis Malic Enzyme A	MLTSSVSVPVRNICRAKVPTLK
Ferredoxin	MLSQVCRFGTITAVKGGV
hsp70	MLKMFNSIFAREKNQHVLTVYM
hsp10	MLATFARNFAAKKVTIKP

Fig. 2. (A) Phylogenetic relationships of hsp70 homologues. The numbers at nodes of the tree correspond to the bootstrap values for ML, distance, MP and QP analyses from left to right respectively. Values are shown only for certain important nodes. The scale bar indicates estimated sequence divergence per unit branch length. ML is generally accepted as the most appropriate method for phylogeny reconstruction. The other methods are included for comparison. The sequences used for the final alignment and the phylogenetic analysis are: *T. vaginalis* (U70308), *Rickettsia prowazekii* (AJ235270), *Leishmania major* (X64137), *Pisum sativum* (X54739), *Eimeria tenella* (Z46965), *Mus musculus* (P38647), *Saccharomyces cerevisiae* (M27229), *Rhodobacter capsulatus* (U57637), *Agrobacterium tumefaciens* (X87113), *Rhodopseudomonas* sp. (D78133), *Haemophilus ducreyi* (U25996), *Legionella pneumophila* (D89498), *Chlamydia trachomatis* (X52175), *and Borrelia burgdorferi* (X67646). The codes in the brackets give the accession number in GenBank. Amino acid frequencies, the proportion of invariable sites and the α parameter of the gamma distribution (eight categories) were estimated from the data set by ML optimisation using the JTT model of amino acid substitution using PUZZLE [25]. QP trees were derived from 10 000 puzzling steps. For ML distance bootstrap analysis (500 replicates), the estimated PUZZLE parameters were used with Puzzleboot 1.0.3 (M.E. Holder and A.J. Roger, unpublished) to build bootstrapped distance matrices in conjunction with the PUZZLE program, and trees were generated using the FITCH programme with global rearrangements [26]. Unweighted MP analysis was carried out by 50 rounds of random stepwise addition heuristic searches with tree bisection reconnection (TBR) branch swapping. Gaps were treated as a 21st amino acid. Distance was measured by mean character difference with minimum evolution and TBR branch swapping was carried out to generate the tree. Topology support was assessed using 500 bootstrap re-samplings for both MP and distance generated trees using PAUP [27]. (B) Alignment of the amino terminus of the mitosome/hydrogenosome proteins. The amino termini of the mt-hsp70 (this study), PNT [6], and cpn60 [7] of *E. histolytica* and certain *T. vaginalis* hydrogenosomal proteins [28–30] are shown. Amino acid identities observed between the amino acid sequence of *E. histolytica* mt-hsp70 and the *T. vaginalis* hydrogenosomal proteins are shaded.

bootstrap support for this mitochondria/α-proteobacteria sister grouping is reasonably good and agrees with previous studies indicating that mito chondria evolved from an endosymbiosis between an eukaryotic cell and an α-proteobacterium [19–21], although it is interesting to note that while *Rickettsia* always groups with the mitochondrial clade in cpn60 trees, in mt-hsp70 trees it clusters with the other α-proteobacteria. This could mean that *Rickettsia* is not, in fact, more closely related to mitochondria than other α-proteobacteria as has been suggested [21]. A consensus built from maximum likelihood analysis of many genes will be needed to resolve this issue.

The identification of mitochondrion specific proteins such as cpn60 and PNT [6] and the mt-hsp70 reported here confirms that *E. histolytica* is not a surviving relic of pre-mitochondrial evolution, but that its ancestors once harboured mitochondria or an endosymbiont related to the progenitor of mitochondria. Most functions of this organelle were lost secondarily during evolution, possibly as a result of parasitic life or adaptation to an anaerobic environment. The function of the remaining compartment, the mitosome, remains obscure.

The amino terminus of the mt-hsp70 protein has greater similarity to certain *Trichomonas vaginalis* hydrogenosomal proteins than to the other mitochondrial-related proteins of *E. histolytica* identified so far (Fig. 2B). Studies have revealed that there might be a common evolutionary origin for protein importation into the mitochondrion, mitosome and hydrogenosome [8,22]. Hydrogenosomal protein signal sequences targeted a reporter protein to the mitochondria of trypanosomes and yeast [22,23], and a mitochondrial signal sequence from *Trypanosoma cruzi* targeted cpn60 to the *E. histolytica* mitosome [8]. This suggests a homology of the importation apparatus and a common origin for the three organelles. The similarity of the mt-hsp70 amino terminus to those of hydrogenosomal proteins in conjunction with the phylogenetic analyses suggests it is also likely to be targeted to the mitosome. Work to confirm this is already in progress.

Acknowledgements

We would like to thank Dr Andrew Roger (Dalhousie University, Canada) for performing the Puzzleboot analysis and for helpful discussions; Dr Michael Duchêne (University of Vienna) and Dr Egbert Tannich (Bernhard Nocht Institute) for providing us with the cDNA and the genomic libraries of *E. histolytica*, respectively; Dr Martin Embley (Natural History Museum, London) for helpful suggestions; and Mehreen Zaki for reading the manuscript. This work was supported by the Wellcome Trust (grant number 048647).

References

[1] Martínez-Palomo A. Biology of *Entamoeba histolytica*. In: Martínez-Palomo A, editor. Amebiasis. Amsterdam: Elsevier, 1986:12–43.
[2] Müller M. Energy metabolism of protozoa without mitochondria. Annu Rev Microbiol 1988;42:465–88.
[3] Sogin ML. Early evolution and the origin of eukaryotes. Curr Opin Genet Dev 1991;1:457–63.
[4] Cavalier-Smith T. Kingdom protozoa and its 18 phyla. Microbiol Rev 1993;57:953–94.
[5] Silberman JD, Clark CG, Diamond LS, Sogin ML. Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small subunit ribosomal RNA gene sequence analysis. Mol Biol Evol 1999;16:1740–51.
[6] Clark CG, Roger AJ. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc Natl Acad Sci USA 1995;92:6518–21.
[7] Roger AJ, Svärd SG, Tovar J, et al. A mitochondrial-like chaperonin 60 gene in *Giardia lamblia*: evidence that diplomonads once harbored an endosymbiont related to the progenitor of mitochondria. Proc Natl Acad Sci USA 1998;95:229–34.
[8] Tovar J, Fischer A, Clark CG. The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. Mol Microbiol 1999;32:1013–21.
[9] Mai Z, Ghosh S, Frisardi M, Rosenthal B, Rogers R, Samuelson J. Hsp60 is targeted to a cryptic mitochondrion-derived organelle ('Crypton') in the microaerophilic protozoan parasite *Entamoeba histolytica*. Mol Cell Biol 1999;19:2198–205.
[10] Boorstein WR, Ziegelhoffer T, Craig EA. Molecular evolution of the HSP70 family. J Mol Evol 1994;38:1–17.
[11] Ortner S, Plaimauer B, Binder M, Wiedermann G, Scheiner O, Duchêne M. Humoral immune response against a 70-kilodalton heat shock protein of *Entamoeba*

histolytica in a group of patients with invasive amoebiasis. Mol Biochem Parasitol 1992;54:175–83.

[12] Descoteaux S, Ayala P, Orozco E, Samuelson J. Primary sequences of two P-glycoprotein genes of *Entamoeba histolytica*. Mol Biochem Parasitol 1992;54:201–11.

[13] Purdy JE, Pho LT, Mann BJ, Petri WA Jr. Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. Mol Biochem Parasitol 1996;78:91–103.

[14] Buß H, Lioutas C, Dobinsky S, Nickel R, Tannich E. Analysis of the 170-kDa lectin gene promoter of *Entamoeba histolytica*. Mol Biochem Parasitol 1995;72:1–10.

[15] Bruchhaus I, Leippe M, Lioutas C, Tannich E. Unusual gene organisation in the protozoan parasite *Entamoeba histolytica*. DNA Cell Biol 1993;12:925–33.

[16] Willhoeft U, Tannich E. The electrophoretic karyotype of *Entamoeba histolytica*. Mol Biochem Parasitol 1999;99:41–53.

[17] Clark CG, Diamond LS. *Entamoeba histolytica*: a method for isolate identification. Exp Parasitol 1993;77:450–5.

[18] Bhattacharya A, Bhattacharya S, Ackers JP. Nontranslated polyadenylated ribonucleic acids from the protozoan parasite *Entamoeba histolytica*. Curr Sci 1999;77:564–7.

[19] Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR. Mitochondrial origins. Proc Natl Acad Sci USA 1985;82:4443–7.

[20] Gray MW, Cedergren R, Abel Y, Sankoff D. On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci USA 1989;86:2267–71.

[21] Andersson SG, Zomorodipour A, Andersson JO, et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 1998;396:133–40.

[22] Haüsler T, Stierhof YD, Blattner J, Clayton C. Conservation of mitochondrial targeting sequence function in mitochondrial and hydrogenosomal proteins from the early-branching eukaryotes *Crithidia*, *Trypanosoma* and *Trichomonas*. Eur J Cell Biol 1997;73:240–51.

[23] Dyall SD, Koehler CM, Delgadillo-Correa MG, Bradley PJ, Plümper E, Leuenberger D, et al. Presence of a member of the mitochondrial carrier family in hydrogenosomes: conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. Mol Cell Biol 2000;20:2488–97.

[24] Diamond LS, Clark CG, Cunnick CC. YI-S, a casein-free medium for axenic cultivation of *Entamoeba histolytica*, related *Entamoeba*, *Giardia intestinalis* and *Trichomonas vaginalis*. J Euk Microbiol 1995;42:277–8.

[25] Strimmer K, Von Haeseler A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 1996;13:964–9.

[26] Felsenstein J. PHYLIP, Phylogeny inference package. Version 3.75c. University of Washington, Seattle, WA, 1993

[27] Swofford DL. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sutherland, MA, 1998.

[28] Hrdy I, Müller M. Primary structure of the hydrogenosomal malic enzyme of *Trichomonas vaginalis* and its relationship to homologous enzymes. J Euk Microbiol 1995;42:593–603.

[29] Johnson PJ, d'Oliviera CE, Gorrell TE, Müller M. Molecular analysis of the hydrogenosomal ferredoxin of the anaerobic protist *Trichomonas vaginalis*. Proc Natl Acad Sci USA 1990;87:6097–101.

[30] Bui ET, Bradley PJ, Johnson PJ. A common evolutionary origin for mitochondria and hydrogenosomes. Proc Natl Acad Sci USA 1996;93:9651–6.