DEVELOPMENT AND TESTING OF MARKERS FOR GENOTYPING OF *ENTAMOEBA HISTOLYTICA*

A thesis submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy of the University of London

2005

Ibne Karim Md. Ali

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

ABSTRACT

Only one in ten *Entamoeba histolytica* infections is invasive but they are responsible for an annual death toll of up to 100,000 people. A key question in amoebiasis is, therefore, what is responsible for the variable outcome of infection. To investigate whether it is linked to the genotype of the infecting strain, we developed 6 pairs of species-specific primers for genotyping after investigating 46 potentially polymorphic short tandem repeat loci adjacent to tRNA genes that were identified during the *E. histolytica* genome project. We tested the primers using *E. histolytica* samples from Bangladesh and 11 other countries. Results revealed that the genotypes present in 3 different clinical populations - asymptomatic, diarrhoeal/dysenteric and liver abscess - were different from each other. A few individual genotypes also showed links to the outcome of infection although their sample coverage was low and therefore they had little predictive value. Although the loci used as polymorphic markers are unlikely to be directly responsible for the outcome of infection, the results do suggest that parasite genetic factors are at least partly responsible.

Our multilocus genotyping method is simple and reliable as it amplifies DNA extracted from axenic or xenic culture, stool samples, and liver abscess pus samples. We believe that these markers will help in studying the patterns of transmission of this important disease and the epidemiological links between individual infections.

LISTS OF ABBREVIATIONS USED

ALA	Amoebic liver abscess
CTAB	Cetyl trimethylammonium bromide
DLV	Double locus variant
DNA	Deoxynucleic acid
EDTA	Ethylene diamine tetraacetic acid
ELISA	Enzyme linked immunosorbent assay
Gal/GalNAc	Galactose/N-acetyl-galactoseamine
ICDDR,B	Int'l Centre for Diarrhoeal Disease Research, Bangladesh
LSHTM	London School of Hygiene and Tropical Medicine
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
SDS	Sodium dodecyl sulphate
SLV	Single locus variant
SREHP	Serine rich *Entamoeba histolytica* protein
SSG	Strain specific gene
SSU-rDNA	Small subunit - ribosomal DNA
STR	Short tandem repeat
tRNA	Transfer RNA

In specimen IDs:

MS	Monthly sample
DS	Diarrhoeal/dysenteric sample
FS	Family member sample
LA	Liver abscess sample

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First of all I would like to thank the almighty Allah for giving me the opportunity and strength to complete this study.

I am indebted to my supervisor Dr. Graham Clark for his excellent suggestions and invaluable guidance throughout this work. His supervision was absolutely 'outstanding'.

I am very much grateful to Dr. Rashidul Haque, Scientist and Head of Parasitology Laboratory of ICDDR,B, and Prof. William A. Petri, Jr. for providing me the samples of *E. histolytica*, *E. dispar* and *E. moshkovskii* from their Mirpur field project.

I would like to thank Dr. David Conway of LSHTM for his invaluable suggestions for the genotype analysis.

I would also like to thank Ms. Blessing Tawari, Research Assistant for her cooperation in the lab work. I would like to thank Dr. Mehreen Zaki, now a postdoctoral fellow at Birmingham University for providing me help and suggestions at the beginning of this study.

I would like to convey my special thanks to my parents for their unconditional support and encouragement throughout my research career.

Finally I would like to thank the Commonwealth Scholarship Commission and the British Council, UK for their financial support.

CHAPTER 1

INTRODUCTION

Entamoeba histolytica is an intestinal protozoan parasite and the causative agent of amoebiasis. According to the most cited reference (Walsh, 1986) one-tenth of the world population is infected with *E. histolytica*, resulting in up to 100,000 deaths worldwide each year (Anonymous, 1997; Petri et al, 2000). However, this estimate predates the formal separation of pathogenic *E. histolytica* from non-pathogenic *E. dispar* (Diamond and Clark, 1993) and is now being reassessed. Current data indicate that *E. dispar* is perhaps ten times more common than *E. histolytica* world-wide (Gathiram and Jackson, 1985), but local prevalence may vary significantly. However it is already clear that not all *E. histolytica* infections lead to disease in the host but that at most one in ten *E. histolytica* infections progresses to development of clinical symptoms (Gathiram and Jackson, 1987). Nevertheless, *E. histolytica* remains a significant source of morbidity and mortality in developing countries (Stanley, 2003). For example, the annual incidence of amoebic dysentery in preschool children in Bangladesh is 2.2% compared to 5.3% for *Shigella* dysentery. In that country one in every 30 children dies of diarrhoea or dysentery before reaching his or her fifth birthday (Haque et al, 2003). In Hue City, Vietnam the annual incidence of amoebic liver abscess is 21 cases per 100,000 inhabitants (Blessmann et al, 2002).

1.1. *Entamoeba histolytica* and *Entamoeba dispar*

There are many species in the genus *Entamoeba* and at least six of them infect humans, namely *Entamoeba histolytica, E. dispar, E. coli, E. moshkovskii, E. hartmanni,* and *E. gingivalis*. In addition there are very rare case reports of *E. polecki* and *E. chattoni* infections in humans (Chacin-Bonilla, 1992; Kuroki et al, 1989; Sargeaunt et al, 1992; Verweij et al, 2001) though they are normally associated with pigs and non-human primates, respectively. It is only *E. histolytica* that causes disease in humans. This organism

is the fourth leading cause of mortality due to a protozoan infection after malaria, Chagas' disease, and leishmaniasis and the third cause of morbidity among protozoa after malaria and trichomoniasis (Anonymous, 1998).

The correct taxonomic classification of *Entamoeba* species is not very clear. The following is one the taxonomic classification of *Entamoeba histolytica* that has been widely used (Levine et al, 1980); however, many of the taxonomic names may not represent monophyletic evolutionary groups :

Kingdom	: Protista
Subkingdom	: Protozoa
Phylum	: Sarcomastigophora
Subphylum	: Sarcodina
Class	: Lobosea
Order	: Amoebida
Family	: Endamoebidae
Genus	: *Entamoeba*
Species	: *E. histolytica*

1.1.1. History

This brief history of *E. histolytica* is based on: Martinez-Palomo (1993) and Petri (1996).

In 1875, Fedor Aleksandrovich Lösch first identified amoebae in clinical samples from a case of fatal dysentery, reproduced the disease in dogs, and suggested the name *Amoeba coli* for this species. In 1890, Osler first positively diagnosed a case of amoebic liver abscess. One year later Councilman and Lafleur, at Johns Hopkins Hospital, confirmed the pathological role of amoebae through studies on patients with dysentery and liver abscesses and introduced the terms "amoebic dysentery" and "amoebic liver abscess". The organism was formally named *Entamoeba histolytica* by Schaudinn in 1903. In 1919, Dobell reviewed all the published species descriptions and concluded that there was only one

species of *Entamoeba* in humans that produced quadranucleated cysts, and retained Schaudinn's name *E. histolytica* for it.

As early as 1925, Brumpt suggested the existence of two morphologically identical species, one being pathogenic and the other not, and proposed the name *Entamoeba dispar* for the nonpathogenic species. Following much debate *E. dispar* was finally separated from *E. histolytica* (Diamond and Clark, 1993) and accepted as a distinct species.

1.1.2. The life cycle

Dobell described the life cycle of *E. histolytica* in 1928. It consists of four consecutive stages, namely, trophozoite, precyst, cyst, and metacyst. Infection is acquired by the ingestion of the mature cyst form of the amoeba which is, in contrast to the trophozoite form, resistant to the acidic pH of the stomach. Excystation occurs as it enters the large bowel, with the emerging amoeba dividing into four and then eight metacystic trophozoites. Trophozoites have the ability to colonize and / or invade the large bowel and may encyst. Cysts do not develop within tissues (Petri, 1996).

Upon excretion, cyst forms of the amoeba remain viable for weeks to months depending on environmental conditions but excreted trophozoites degenerate rapidly outside the body. Infection can result from the ingestion of a single cyst (Ravdin, 1995), however a higher inoculum results in a shorter incubation period of a few days for the onset of disease rather than the more usual few weeks. The duration of infection varies from subject to subject and cysts have been demonstrated in untreated persons for as long as two years (Martinez-Palomo, 1993). A schematic representation of the life cycle of *E. histolytica* is shown in Figure 1.1.

1.1.3. Morphology

Under normal light microscopy the morphology of *E. histolytica* and *E. dispar* trophozoites and cysts is considered to be indistinguishable (Diamond and Clark, 1993). The trophozoite

form of both species is highly dynamic and pleomorphic (i.e. many-formed) with an average diameter of 25 μm (range 10-60 μm). It has a single nucleus of 3-5 μm in diameter, which contains fine peripheral chromatin and a central karyosome (dark-spot-like in appearance). The nucleus has no fixed position in the cytoplasm, and also moves freely. Motility and pseudopod formation are rapid and unidirectional, although movement rarely occurs in a straight line.

The mature cyst is round to oval shaped and contains 4 nuclei (i.e. is quadranucleate). Its average diameter is 12 μm (range 10-16 μm). Generally, glycogen and chromatoid bodies (aggregated crystals of ribosomes) are seen in the immature precystic stage but disappear as the cyst matures. The nucleus is smaller than in the trophozoite and also contains a smaller centrally located karyosome .
In general, in axenic culture *E. dispar* trophozoites are elongated while *E. histolytica* trophozoites are more round shaped. Also the cell surface of *E. histolytica* is rough with numerous small circular openings in contrast to that of *E. dispar* which are smoother in appearance.

1.1.4. Cultivation

In 1925, Boeck and Drbohlav first successfully cultivated *E. histolytica* in vitro by using Locke's-egg-serum medium. However the most frequently used xenic culture media (in the presence of a mixed bacterial flora) are those of Balamuth (1946), Jones (1946), and Robinson (1968). These media are not specific for *E. histolytica* and *E. dispar* and support the growth of other *Entamoeba* species. The first axenic cultivation of *E. histolytica* was achieved by Diamond (1961, 1968, 1978). Clark (1995) first established *E. dispar* in axenic culture after much difficulty. Clark and Diamond (2002) have provided a succinct account of the history of intestinal amoeba cultivation.

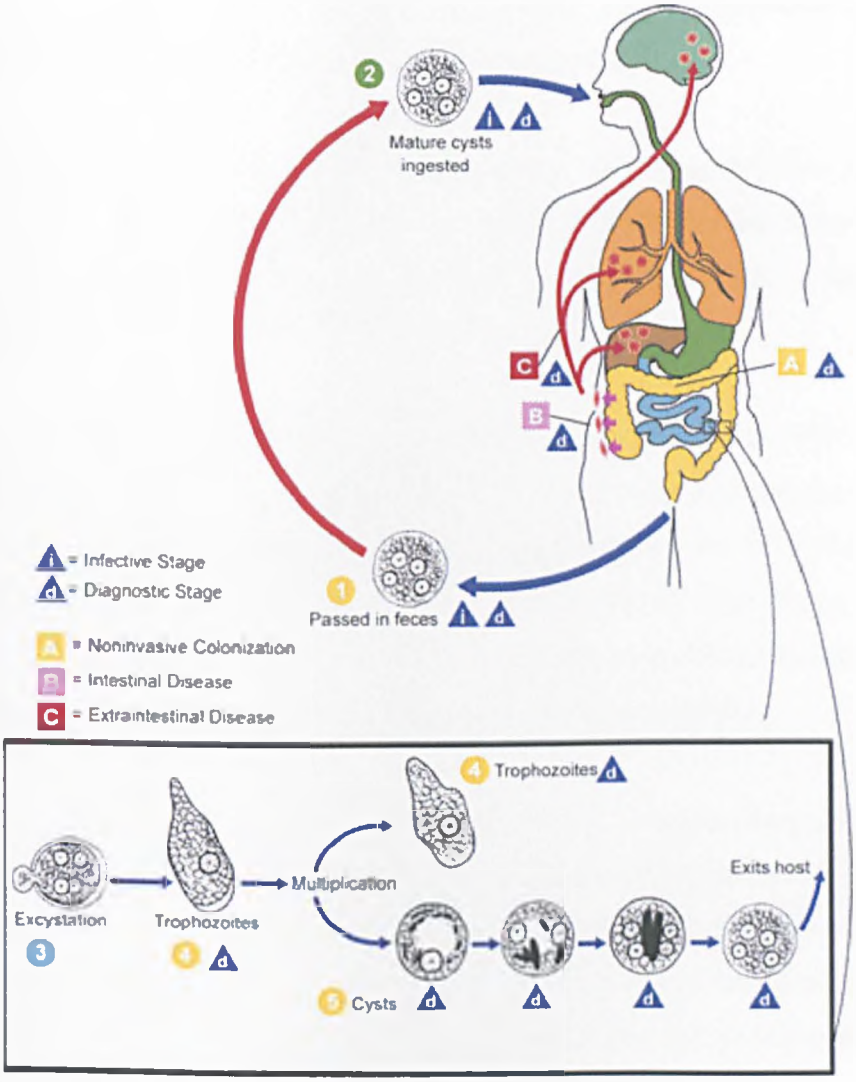


Figure 1.1. Life cycle of *Entamoeba histolytica* (taken from: www.dpd.cdc.gov/dp-dx/HTML/Amebiasis.htm).

1.1.5. Clinical manifestations

Most *E. histolytica* infections remain asymptomatic. Invasive intestinal infection or amoebic colitis is marked by loose stools containing blood and mucus. Complications of amoebic colitis include perforation and secondary bacterial infection of ulcers and are

associated with very high mortality rates. Occasionally, chronic ulceration results in amoeboma formation (a tumour-like growth due to amoeba).

The most common form of extraintestinal amoebiasis is the amoebic liver abscess, which unless properly diagnosed and promptly treated is a potentially lethal disease. In areas of endemicity, amoebic liver abscess is suspected in a patient with fever, weight loss, and right upper quadrant pain and tenderness (Haque et al, 2000).

There are sporadic case reports of cerebral amoebiasis (Hughes et al, 1975; Sundaram et al, 2004; Schmutzhard et al, 1986) which has an abrupt onset and rapid progression, resulting almost always in death. Infections of kidney (Ramakrishnan et al, 1971), heart (Mehta et al, 1968), lung (Hara et al, 2004; Lichtenstein et al, 2005; Snyders and Welke, 1989; Teramoto et al, 2001), appendices (Casavilca et al, 2000; Ramdial et al, 2002), and skin (Kroft et al, 2005; Magana et al, 2004) by *E. histolytica* have also been reported.

Although *E. dispar* is considered a harmless commensal in human hosts, there are a few reports that *E. dispar* is capable of producing variable focal intestinal lesions in animals (Chadee et al, 1985; Espinosa-Cantellano et al, 1997; Vohra et al, 1989). Also there is some evidence that pathological changes may occur in some humans (McMillan et al, 1984) due to *E. dispar* infection. All these findings indicate that there may be some strains of *E. dispar* that are capable of producing mild disease in hosts.

1.1.5. Treatment

E. histolytica infections can have various manifestations in humans. Infected persons may remain asymptomatic throughout before clearing the amoebae spontaneously. Alternatively, infection may cause invasive disease at intestinal or extraintestinal sites. From the treatment point of view a decision has been made that *E. histolytica* infections, whether symptomatic or not, must be treated (Anonymous, 1997). On the other hand, *E. dispar* infection (and perhaps *E. moshkovskii* infection as well) does not require any

treatment but should be monitored closely for the possibility of becoming co-infected with *E. histolytica*. This depends on being able to distinguish the species, however.

The treatment of amoebiasis has two fundamental goals: one, to eliminate the passage of cysts or trophozoites from the bowel lumen, and two, to cure invasive disease at both intestinal and extraintestinal sites. Diloxanide furoate (Furamide) is currently considered the drug of choice for eradicating intralumenal cyst passage and has an 85% success rate over a 10-day course (3 doses of 500 mg per day). Treatment is considered successful if the stool remains free of cysts or trophozoites for one month. Diloxanide is relatively nontoxic with only occasional mild gastrointestinal symptoms (McAuley et al, 1992). Another lumenal agent is paromomycin (Humatin). Unlike diloxanide it is not absorbed from the gastrointestinal tract and therefore is safe in pregnancy. Its side effects can include diarrhoea and abdominal pain (McAuley and Juranek, 1992). However, a recent study has compared the efficacy of both drugs and has shown that paromomycin had a much higher cure rate than diloxanide furoate (85% versus 51%) in the treatment of asymptomatic *E. histolytica* carriers (Blessmann and Tannich, 2002).

Metronidazole (Flagyl) is the preferred drug for patients with acute and chronic invasive intestinal disease. Metronidazole is a 5-nitroimidazole derivative with anti-microbial activity against some anaerobic bacteria as well as protozoa such as *E. histolytica* and *Giardia intestinalis*. Investigations in many controlled studies have shown that metronidazole has a cure rate of over 90% (Samuelson et al, 1992; Simjee et al, 1985). Depending on the amount of drug (750 mg to 2.4 g) metronidazole can be used for 3 to 10 days to achieve success (Briggs et al, 1994). Although multi-drug resistance genes and drug-efflux mechanisms do exist in *E. histolytica*, resistance to metronidazole has not yet been a clinical problem. Since this drug cannot eliminate the intestinal infection, a lumenal agent must also be administered with or following it. Common side effects of metronidazole include gastrointestinal (metallic taste, abdominal discomfort) and central nervous system (headache, dizziness, and drowsiness) problems (Roe, 1985). Tinidazole, another 5-nitroimidazole, is another drug of choice outside of the USA and Canada for invasive amoebiasis and has fewer side effects.

For the treatment of extraintestinal lesions like amoebic liver abscess, the drug of choice is again metronidazole: a single dose of 2.5 g for mild to moderate disease is enough followed by an agent to eradicate intralumenal infection. An alternative drug for amoebic liver abscess is chloroquine, which is an amoebicide in hepatic tissue alone.

1.1.6. Vaccine against amoebiasis

Researchers are using molecular biology techniques to try and develop a safe, reliable and efficient vaccine to prevent amoebiasis. Studies with the gerbil model of amoebic liver abscess have identified recombinant *E. histolytica* antigens that can protect animals from amoebic liver abscess (Zhang and Stanley, 1996). Successful oral administration of the vaccine has been achieved in the gerbils, with induction of mucosal immune responses. However, there are no data available on the efficacy of this candidate vaccine in preventing intestinal amoebiasis. This is a problem mainly because there is no reliable animal model for intestinal amoebiasis. However an earlier study showed that guinea pigs infected intracaecally with a noninvasive strain of *E. histolytica* could be protected against subsequent challenge with a virulent strain (Jain et al, 1980). Several other studies have shown that rats, hamsters, or monkeys vaccinated with either intact amoebae or various forms of complex amoebic antigens could be protected from amoebic liver abscess formation (Ghadirian and Meerovitch, 1978; Ghadirian et al, 1980; Jimenez Cardoso et al, 1989; Martinez-Reyes et al, 1980; Sepulveda, 1980). The recent demonstration that the recombinant amoebic SREHP antigen can be mucosally immunogenic when delivered by attenuated *Vibrio cholerae* or *Shigella typhi* vectors provides additional evidence that the combination vaccine can be effective against multiple enteric pathogens (Zhang et al, 1995; Zhang and Stanley, 1996). Cheng and Tachibana (2001) and Lotter et al (2004) have reported protection of hamsters and gerbils, respectively, from amoebic liver abscess by immunization with surface lectin antigens of *E. histolytica*. Houpt et al (2004) have reported successful prevention of intestinal infection by *E. histolytica* using the galactose/N-acetyl-galactosamine (Gal/GalNAc) lectin vaccine in a mouse model.

1.1.7. Epidemiology

The worldwide prevalence of *E. histolytica* and *E. dispar* as separate species is not well studied. The prevalence of amoebic infection is high in the Indian subcontinent, Africa, the Far East, and areas of South and Central America. In developing countries it depends largely on cultural habits, age, level of sanitation, crowding and socio-economic status. In developed countries, the infection is mostly due to *E. dispar* and is confined to certain group of people: immigrants from or travelers to areas of endemicity, homosexual males, patients infected with immunodeficiency virus, and institutionalized populations (Petri, 1996). In Japan, however, carriage of *E. histolytica* is more common (Martinez-Palomo, 1993). Case studies of patients with amoebic colitis in Natal, South Africa, showed that there was a peak incidence of infection among children <14 years of age and a second increase in infection in adults >40 years old (Gathiram and Jackson, 1985). Acuno-Soto et al (2000) after reviewing all the published reports from 1929 to 1997 found that the male to female ratios for invasive intestinal amoebiasis and asymptomatic carriage were 3.2:1 and 1:1, respectively.

Some examples of the prevalence of *E. histolytica* and *E. dispar* infection using different diagnostic methods are provided below:

Using microscopy:

- Thailand - 7.1% *E. histolytica* / *E. dispar* among the mentally handicapped (Sirivichayakul et al, 2003).
- Lebanon - 2.3% *E. histolytica* / *E. dispar* in adults (Saab et al, 2004).
- Gaza Strip, Palestine - 7.0% *E. histolytica* / *E. dispar* among 6 to 11 year-old children (Astal, 2004).
- Bat Dambang, Cambodia - 0.8% *E. histolytica* / *E. dispar* among school children (Park et, 2004)
- Korea - 1.8% *E. histolytica* / *E. dispar* among handicapped people (Lee et al 2000).

Using isoenzyme analysis:

- South Africa – 1.0% *E. histolytica* and 9.0% *E. dispar* (Gathiram and Jackson, 1985).
- Ecuador - *E. dispar* 4-times more prevalent than *E. histolytica* (Gatti et al, 2002).

Using antibody detecting ELISA tests:

- Mexico - 8.4% of the population seropositive (Caballero-Salcedo et al, 1994).

Using antigen detecting ELISA tests:

- Bangladesh - 5% *E. histolytica* and 13% *E. dispar* among 2-5 year old children (Haque et al, 1999) and a 12-month follow-up revealed that only about 3% of *E. histolytica* infected children developed symptoms attributable to amoebic dysentery (Haque et al, 2001).

Using species-specific PCR tests:

- Ethiopia - Only *E. dispar* among school children, prisoners and HIV patients (Kebede et al, 2003, 2004).
- Japan - 56% of homosexual males carry *E. histolytica* (Ohnishi et al, 2004).
- Vietnam - 11.2% *E. histolytica* in asymptomatic adults and a 15-month follow-up of 43 asymptomatic *E. histolytica* carriers revealed that none developed invasive intestinal disease, although one developed an amoebic liver abscess (Blessmann et al, 2003). Previously they showed the rate of amoebic liver abscess infection in this community was 21 cases per 100,000 inhabitants (Blessmann et al, 2002).
- Iran - *E. dispar* 10-15 times more prevalent than *E. histolytica* (Hooshyar et al, 2004).

The association of amoebasis with HIV-positive individuals is not clear. A recent case-control study at Taiwan with 951-HIV infected individuals, 429 HIV uninfected but with gastrointestinal symptoms and 178 HIV uninfected healthy individuals shows that HIV infected individuals are at higher risk for invasive amoebiasis. Over 3% of HIV infected individuals had *E. histolytica* infections compared to only 1.4% in uninfected healthy individuals as detected by PCR and antigen detection ELISA (Hung et al, 2005). In contrast, Mohandas et al (2002) observed that only 2 of 120 HIV seropositive persons were positive for *E. histolytica/E. dispar* by microscopy in India. Since the information on infections with *E. histolytica/E. dispar* in a control group was not available it is not possible to predict whether HIV-positive individuals were at higher risk in India. However, two separate studies with children in India found a high prevalence of *E. histolytica/E. dispar* by microscopy. One, a microscopy investigation of stool specimens of school children from Delhi, found that 11% of them were positive for *E. histolytica / E. dispar* (Kaur et al, 2002). The second, a study carried out in rural and urban locations in and around Chennai detected 4% and 10.6% of *E. histolytica / E. dispar*, respectively, among school children by microscopy (Fernandez et al, 2002). Considering these study results it seems more likely that HIV-infected individuals are not at a higher risk for infection with *E. histolytica / E. dispar* in India.

1.1.8. Diagnosis

From the epidemiological point of view specific detection of *E. histolytica* and *E. dispar* is necessary to know the true prevalence of each species. Most of the epidemiological data for intestinal amoebiasis are based on the diagnosis of these two species in stool specimens by light microscopy. However, identification of *E. dispar* and *E. histolytica* by microscopy is not precise. Microscopy is unable to differentiate *E. histolytica* from *E. dispar*, is at best 60% sensitive, and can be confounded by false-positive results due to misidentification of macrophages and non-pathogenic species of *Entamoeba*. Culture is more sensitive than microscopy (Gonzalez-Ruiz et al, 1994), and isoenzyme analysis of cultured amoebae enables the differentiation of *E. histolytica* from *E. dispar*. However, culture and isoenzyme analysis requires a week to complete and is negative in many microscopy-

positive samples due to delays in sample processing or due to the initiation of antiamoebic therapy prior to stool collection (Gonzalez-Ruiz et al, 1994; Haque et al, 1997; Strachan et al, 1988). New approaches have been used to detect *E. histolytica* and *E. dispar* based on antigen detection in stool specimens (Abd-Alla et al, 1993; Haque et al, 1993; Haque et al, 1995; Haque et al, 1997; Strachan et al, 1988). Although the detection of amoebic antigen in stool specimens is a simple and reliable method, there is one limitation: it cannot detect a mixed infection of both species simply because there is no *E. dispar* specific monoclonal antibody available at present. There are many reports on the successful species specific PCR amplification of *E. histolytica* and *E. dispar* DNA (Britten et al, 1997; el-Hamshary et al, 2004; Katzwinkel-Wladarsch et al, 1994; Newton-Sanchez et al, 1997; Romero et al, 1992; Sharma et al, 2003; Tannich and Burchard, 1991). Paglia et al (2004) reported the development of a diagnostic nested PCR to detect *E. histolytica* and *E. dispar* from both fresh and formalin-fixed stool samples. Recently Blessmann et al (2002) and Roy et al (2005) described two very fast and highly sensitive Real-Time PCRs for the detection and differentiation of *E. histolytica* and *E. dispar* in faecal samples. Serological tests, particularly in areas of endemicity, can provide little information, as they cannot distinguish between acute and past infections.

Furrows et al (2004) tried to assess the efficiency of 3 diagnostic methods – a commercial antigen detecting ELISA (TechLab, Inc.), a polymerase chain reaction-solution hybridization enzyme linked immunoassay (PCR-SHELA), and a Real-Time PCR (Artus) – for diagnosis of *E. histolytica* infection using 101 clinical samples. They concluded that all three methods performed adequately and choice of method should depend on the budget and timeframe of the study.

The diagnosis of amoebic liver abscess is sometimes difficult since its clinical manifestations are highly variable. Imaging techniques such as ultrasound, computed tomography, and magnetic resonance have excellent sensitivity for the detection of a liver abscess arising from any cause but cannot distinguish amoebic abscesses from pyogenic (bacterial) abscesses or necrotic tumors. Some studies have reported that amoebic liver abscess patients do not have concomitent amoebic colitis and therefore stool microscopy or

antigen detection in stool specimens is not helpful for diagnosis: less than 10% patients have identifiable amoebae in stool (Haque et al, 2000; Katzenstein et al, 1982). Serological tests demonstrate the presence of antiamoebic antibodies in serum and are positive for most patients with amoebic liver abscess. Several groups have reported the detection of amoebic antigen in the serum of amoebic liver abscess patients (Abd-Alla et al, 1993; Haque et al, 2000; Karki and Parija, 1999). For example, Abd-Alla and colleagues (1993) detected the Gal/GalNAc lectin, a major surface antigenic molecule in *E. histolytica*, in the sera of 75% of South African patients with amoebic liver abscess. In Dhaka, Bangladesh, the TechLab *E. histolytica* II test detected Gal/GalNAc lectin in the sera of 96% of amoebic liver abscess patients tested prior to treatment with the antiamoebic drug metronidazole.

In conclusion, the accurate diagnosis of *E. histolytica* and *E. dispar* is important for two reasons, one, to understand the worldwide distribution of the two species separately, and two, to prevent unnecessary chemotherapy in patients infected with *E. dispar* but wrongly identified as being infected with *E. histolytica*.

1.2. *Entamoeba moshkovskii*

Entamoeba moshkovskii, which is mainly considered to be a free-living amoeba, is indistinguishable in both its cyst and trophozoite forms from *E. histolytica* and *E. dispar*. Both *E. polecki* (associated with pigs) and *E. chattoni* (associated with non-human primates) which occasionally infect humans, have uninucleate cysts. *E. coli* and *E. hartmanni* have 8 and 4 nuclei each, respectively, but the size of the *E. hartmanni* cysts and trophozoites is much smaller than *E. histolytica* and *E. dispar*. *E. gingivalis*, a parasite of human oral cavity, is a non-encysting species. As a result, as far as diagnosis of amoebiasis is concerned none of these species is a problem. *E. moshkovskii* has so far rarely been shown to infect humans but appears to be ubiquitous in anoxic sediments. Although the early isolations of this species were from sewage, *E. moshkovskii* can also be found in environments ranging from clean riverine sediments to brackish coastal pools (Clark and Diamond, 1997). It is osmotolerant, can be cultured at room temperature, and is resistant to emetine, all characteristics that distinguish it from *E. histolytica* and *E. dispar* (Clark and

Diamond, 1997; Dreyer, 1961; Entner and Most, 1965; Richards et al, 1966). Human isolates of *E. moshkovskii* to date have come from North America, Italy, South Africa and Bangladesh, and they have never been associated with disease (Clark and Diamond, 1997; Haque et al, 1998a).

The morphological resemblance of the apparently innocuous *E. moshkovskii* to the disease-causing *E. histolytica* makes it important to differentiate the two species. In the clinical setting, for example, an *E. moshkovskii* infected patient could be diagnosed as infected with *E. histolytica* and be treated unnecessarily with anti amoebic chemotherapy. Most studies that have investigated the prevalence of *E. histolytica* and *E. dispar* have not considered the possible presence of *E. moshkovskii*. This is partly due to a lack of tools to detect *E. moshkovskii*. The importance of this species is therefore unclear at present.

1.3. Inter- and intra-species genetic variation

There have been several studies of inter- and intra-species variation in *E. histolytica* and *E. dispar*. Sargeaunt et al (1978) first introduced isoenzyme electrophoresis to study polymorphism in intestinal protozoan parasites in humans. After studying approximately 10,000 intestinal amoebic isolates Sargeaunt (1988) identified 20 different isoenzyme patterns (known as 'zymodemes'): 9 proved to be those of *E. histolytica* (pathogenic zymodemes, PZ) and while the rest were *E. dispar* (non-pathogenic zymodemes, NPZ). However, recent studies suggest that there are only 4 principal zymodemes (those found in axenic cultures) that are reliable for isolate typing. These are *E. histolytica* zymodemes II, XIV, and XIX, and *E. dispar* zymodeme I (Jackson and Suprasad, 1997). The remaining zymodemes appear to be due to bacterial enzymes (present in xenic cultures).

A few PCR-based DNA typing methods have been reported for *E. histolytica*, making use of repetitive elements contained within both protein coding genes and non-coding DNAs. In *E. histolytica* the rRNA genes are present on a highly repetitive extrachromosomal circular DNA, about 25 kb in size and 200 copies per genome (Bhattacharya et al, 1988, 1989; Huber et al, 1989). The strain specific gene (SSG) (Burch et al, 1991) or *Tr* (Sehgal

et al, 1994) has been described in the upstream region of the rDNA circle, for which a noncoding transcript has been detected. The SSG contains tandem repeats and PCR amplification shows variations in the number of repeat units among *E. histolytica* strains (Clark and Diamond, 1993). In some *E. histolytica* strains this gene is completely absent. *E. dispar* strains do not possess this gene.

Polymorphism has been investigated in the repeat-containing protein-coding chitinase gene in *E. histolytica.* The chitinase gene, expressed only in the cyst form of the amoeba, shows only very limited diversity (de la Vega et al, 1997a,b; Ghosh et al, 2000; Haghighi et al, 2002, 2003). Similarly the Gal/GalNAc lectin, a major amoebic virulence protein, showed only very limited diversity among strains of *E. histolytica* (Beck et al, 2002). Perhaps the most widely studied protein-coding gene is the serine-rich *E. histolytica* protein (SREHP) gene, which encodes an immunodominant surface antigen containing tandem repeats of related dodeca- and octa-peptides (Kohler and Tannich, 1993; Stanley et al, 1990). Repeat number, sequence and restriction site-derived variations have been observed among strains of *E. histolytica* (Ayeh-Kumi et al, 2001; Clark and Diamond, 1993; Ghosh et al, 2000; Haghighi et al, 2002 and 2003). A recent study in Bangladesh with *E. histolytica* DNA from clinical strains showed extensive SREHP polymorphism, and the polymorphic patterns observed in amoebic liver abscess strains were different from those of intestinal strains (Ayeh-Kumi et al, 2001). Homologous genes in *E. dispar* have also shown substantial polymorphism (E. Paez, 1997, unpublished MSc thesis).

Besides the rDNA and protein coding genes mentioned above, several classes of tandemly repeated or dispersed multicopy DNAs located elsewhere in the genome have been identified (Azam et al, 1996; Cruz-Reyes et al, 1995; Huang et al, 1997; Lohia et al, 1990; Michel et al, 1992; Mittal et al, 1994). However, for the most part, the polymorphisms in these repetitive DNAs have not been tested.

Polymorphism has been investigated using random amplified polymorphic DNA (RAPD)-PCR in a few north Indian isolates (Prakash et al, 2002). Although this method showed a good degree of polymorphism, each isolate invariably produced multiple products in PCR,

which is not ideal for large scale surveys of genetic epidemiology as the analysis of results will be difficult. Also, RAPD may amplify DNA from host and bacteria and therefore is problematic.

Recently Shah et al (2005) used a comparative genomic hybridization approach to investigate the genotypic differences among 4 reference strains of *E. histolytica* and two of *E. dispar*. Between *E. histolytica* and *E. dispar* strains they identified a number of genes restricted to *E. histolytica* only. One of the drawbacks of this approach is that it requires a large amount of genomic DNA (5 to 10 μg), which has so far restricted its use to laboratory strains only.

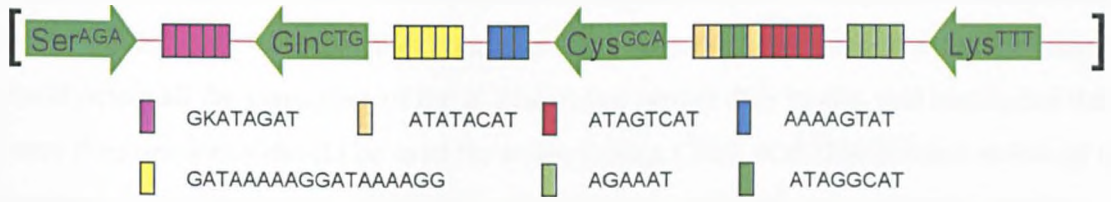
1.3.1. Transfer RNA (tRNA) genes in *E. histolytica*

One of the striking findings in the *E. histolytica* genome sequence is the abundance and unique organisation of the tRNA genes in the genome (Clark et al, 2005). Over 10% of all the sequence reads contained tRNA genes. tRNA genes are 71-87 base pairs long and rich in Gs and Cs. Almost all of them are organised in tandem arrays with intergenic regions being rich in A+T (about 80%) and containing short tandem repeats (STRs). A total of 25 distinct arrays were identified containing 1 to 5 tRNA genes in each. The unit lengths of arrays vary between 490 bp and 1775 bp. In 4 arrays, G+C rich regions other than tRNA genes were identified. Three of these have been identified as the previously unknown 5S RNA genes in *E. histolytica*, while the fourth is a suspected small nuclear RNA-encoding gene (Banerjee and Lohia, 2003). Outside the coding regions, apart from being A+T rich, there are no significant sequence similarities between the arrays. However there exists a compositional bias, i.e. one strand contains on an average about 70% pyrimidine. An example of a tRNA gene unit in *E. histolytica* containing four genes is shown in Figure 1.2.

1.3.2. Polymorphism in repeat loci containing tRNA genes

A number of non-coding loci in *E. histolytica* and *E. dispar* have been identified by different groups of researchers and each locus contains tandem repeat units. Three of these

(A)



(B)

Figure 1.2. tRNA gene organisation in *E. histolytica*. (A) Schematic representation of a 4 tRNA gene array unit [SQCK]. Arrows represent the tRNA genes. STRs are found in intergenic regions and are represented by coloured boxes. (B) 3 tandem units of tRNA array [SQCK] are shown.

repeat loci were originally identified separately by other groups but their potential as polymorphic markers was investigated by Zaki and Clark (2001): locus 3-4, a 978 bp element described by Michel et al (1992) (GenBank accession number M77091); locus 9-4, a 931 bp DNA element isolated by Rosales-Encina and Eichinger (GenBank accession number AF265348); locus 16-17, a 964 bp element reported by Huang et al (1997). There was close similarity between locus 3-4 and locus 9-4. In addition, Zaki and Clark (2001) also isolated and tested locus 1-2 (GenBank accession numbers AF276055-9) and locus 5-6 (GenBank accession numbers AF276060-5). All these polymorphic loci were later found to contain genes for tRNA (Clark, unpublished data).

Diversity in these loci among *E. histolytica* strains is mainly due to varying numbers of STRs (Zaki and Clark, 2001). Both 1-2 and 5-6 contain a single STR block consisting of two types of related sequences arranged in tandem. In contrast, loci 3-4, 9-4, and 16-17 contain two or more STR blocks with more than one type of related repeat sequence arranged in tandem. However, these loci varied not only in the number but also in the sequence and arrangement of STRs when the comparison was made between *E. histolytica* and *E. dispar* (Zaki et al, 2002). Haghighi et al (2002 and 2003) compared the polymorphic

potential of loci 1-2 and 5-6 with those of the protein-coding chitinase and SREHP genes using samples from institutionalised and homosexual people from Japan and clinical and reference strains from some other countries. They observed that none of these loci alone could detect all the genotypes of the *E. histolytica* strains they tested, and concluded that more than one locus should be used for strain typing. Clark et al (2002) have reviewed the present status of genetic diversity in *E. histolytica* and concluded that for strain typing amplification by PCR directly from stool DNA would be the method of choice over the alternative - labourious amoeba cultivation and sequencing of DNA. They also emphasized the need for using multiple loci for reliable typing of amoeba strains.

What determines the outcome of an *E. histolytica* infection is still a mystery, but one possibility is that it is linked to the genotype of the infecting strain. To investigate this relationship a simple, sensitive and reliable method for strain identification is required.

There are many more STRs in the flanking regions of tRNA genes in *E. histolytica* that still need to be tested for their usefulness in detecting polymorphism in *E. histolytica*. It is at this point that the work described in this thesis began.

1.4. AIMS & OBJECTIVES

The key objective of this study is to develop a reliable multilocus genotyping system for *E. histolytica* after screening all tRNA-linked STR sequences identified during the genome project.

The system developed will then be used to genotype clinical strains of *E. histolytica* from Bangladesh and other parts of the world to investigate a possible association of genotype and clinical outcome of infection.

The sequence diversity among clinical strains of *E. histolytica* at two of the STR loci will be investigated to help understand the basis of the polymorphism observed.

Prevalence of *E. moshkovskii* and polymorphism in *E. dispar* strains will also be investigated. The specificity of the genotyping system will be tested with both *E. dispar* and *E. moshkovskii* DNAs.

Finally, attempts will be made to estimate the copy number of the tRNA-linked loci in the genome.

CHAPTER 2

GENERAL METHODS

2.1. Axenic isolates of *E. histolytica*, *E. dispar* and *E. moshkovskii*

In this thesis, 'species' mainly refers to one of the three morphologically identical species of *Entamoeba – E. histolytica / E. dispar / E. moshkovskii*. When samples are from stool and culture the organisms are referred to as a 'strain' and 'isolate', respectively. 'Strain' is also used as a general term where more than one type of sample is being discussed. Based on polymorphisms at one or more loci, individual strains or isolates are assigned to a specific 'genotype'. The term 'clone' is used only in the context of PCR products from that have been inserted into recombinant DNA vectors and not to cell lines.

E. histolytica HM-1:IMSS, *E. moshkovskii* Laredo and FIC are maintained in axenic medium and *E. dispar* SAW760 is maintained in both axenic and monoxenic medium in our laboratory. 7 other axenic isolates of *E. histolytica* were maintained elsewhere in the London School of Hygiene and Tropical Medicine (LSHTM). These 8 *E. histolytica* and one *E. dispar* were used in the initial testing of primers. Information about the patients' history (except for *E. moshkovskii* FIC which was isolated from a sewage sample) is provided in the relevant Chapters.

Axenic isolates of *E. histolytica* were maintained in LYI-S-2 (Liver extract Yeast extract Iron-Serum) medium (Clark and Diamond, 2002). This medium is an alternative to the more widely used medium TYI-S-33 for axenic cultivation of *E. histolytica*. LYI-S-2 medium differs from TYI-S-33 in that it contains liver extract instead of casein. LYI-S-2 consists of a nutrient broth, LYI, which is enriched with the addition of a vitamin mix and 15% heat inactivated adult bovine serum. Sub-culturing was done every 72 and 96 hours (usually Monday and Friday of the week) and cultures were incubated at 36°C. *Entamoeba moshkovskii* isolates Laredo and FIC were maintained axenically in LYI-S-2 medium

(Clark and Diamond, 2002) with 10% adult bovine serum at room temperature and subcultured every 2 weeks. The monoxenic culture of *E. dispar* was maintained in LYI-S-2 medium supplemented with 10% heat inactivated bovine serum and *Crithidia fasciculata*, and incubated at 36°C. Subcultures were performed at 72 hour and 96 hour intervals.

In the study of genotyping we used many xenic culture and amoebic liver abscess (ALA) pus DNA samples which will be discussed in the relevant sections.

2.2. Isolation of DNA

2.2.1. CTAB method: This method was employed for the isolation of DNA from all axenic and monoxenic isolates and pus specimens as well as from some xenic isolates of *E. histolytica* and *E. dispar* (modified from Clark and Diamond, 1991; Clark, 1992):

Lysate preparation:

A) From axenic and monoxenic cultures: Cultures were chilled for 5 minutes in an ice-water bath and sedimented by centrifugation at 275 x g for 5 minutes. The sedimented trophozoites were washed with 1 X phosphate buffered saline (PBS) to remove traces of medium and the washed cell pellets were resuspended in 250 µl of lysis buffer (0.25% SDS, 0.1 M EDTA, pH 8).

B) From xenic cultures: Sediment from a 72 hour culture in Robinson's medium was transferred to a 1.7 ml microcentrifuge tube and the trophozoites were pelleted by brief centrifugation in a microcentrifuge, washed with 1 X PBS to remove traces of medium, and resuspended in 250 µl of lysis buffer.

C) From lyophilized ALA pus specimens: Approximately 0.05-0.1 g of lyophilized pus was scraped into a microcentrifuge tube and re-constituted with 250 µl of lysis buffer.

Lysates are stable for months at 4°C.

The rest of the DNA isolation was same regardless of the type of specimen from which the lysate was prepared:

1. Proteinase K was added at a concentration of 0.1 mg/ml of lysate and incubated at 55°C for 20 minutes in a waterbath to get rid of all proteins including DNA-degrading DNase enzymes.
2. 75 µl of 3.5 M NaCl and 42 µl of 10% cetyl trimethylammonium bromide (CTAB) (prepared in 0.7 M NaCl) were added (final concentration of CTAB in lysis buffer is 1% in the presence of 0.7 M NaCl) and incubated at 65°C for 10 minutes. At this high salt concentration CTAB binds to carbohydrate molecules (that are abundent in these amoebas) and forms a CTAB-carbohydrate complex.
3. At room temperature, an equal volume (approximately 400 µl) of chloroform was added, mixed well by inversion, and centrifuged at 14,000 x g for 8 minutes. At this stage the CTAB-carbohydrate complex forms an insoluble interface, leaving DNA in the aqueous supernatant.
4. The clear supernatant was transferred to a new microcentrifuge tube and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) by inversion and centrifuging at 14,000 x g for 8 minutes.
5. The clear supernatant was transferred to a new microcentrifuge tube and DNA was precipitated by adding 2.5 volumes of absolute alcohol, mixed and left to stand for 5 minutes, and centrifuged at 14,000 x g for 10 minutes.
6. The supernatant was removed carefully without disturbing the DNA pellet and the pellet was washed in 70% ethanol, centrifuged at 14,000 x g for 5 minutes, air-dried (about 1 hour) and dissolved in 50 µl deionised autoclaved water.
7. The resuspended DNA was passed over a Microspin™ Sephacryl S-200 HR column (Amersham Pharmacia Biotech Inc) to remove non-polysaccharide PCR inhibitors and salts.
8. RNA was removed by the addition of RNase at a concentration of 0.05 µg/ml of DNA solution.

2.2.2. GlassMax method: This method was employed at International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) for the isolation of DNA from some of the xenic isolates of *E. histolytica* and *E. dispar* grown in Robinson's medium as well as from some stool specimens as previously described (Haque et al, 1998b; Katzwinkel-Wladarsch et al, 1994).

2.2.3. Qiagen method: The QIAamp DNA Stool Mini Kit was used according to the manufacturer's instructions (Qiagen, UK) to isolate DNA at ICDDR,B and in Vietnam for some of the stool samples.

2.3. Polymerase Chain Reaction (PCR)

2.3.1. Design of primers

Basically two different types of primers were used in the polymorphism study, 'tRNA primers' designed from tRNA gene sequences of *E. histolytica* HM-1:IMSS to amplify *E. histolytica* as well as *E. dispar* DNA, and 'specific primers' designed after the comparison of homologous sequences from *E. histolytica* and *E. dispar*.

The following factors were considered when designing a primer:

1. It should comprise of about 20-30 nucleotides.
2. It should contain about 50% G+C. For primers with low G+C content, a longer primer was chosen so as to avoid a low melting temperature.
3. Sequences with long runs (i.e. more than three or four) of a single nucleotide were avoided, when possible.
4. Primers with secondary structure were avoided.
5. Complementarity between the two primers of a pair and self complementarity were avoided.

For the tRNA primers both primers of a pair were located within the tRNA sequence(s). The sequences of all primers will be provided in the appropriate Chapters.

2.3.2. PCR amplification

All the PCR reagents used in this study were from BioLine (UK). The total volume of PCR mix used for one reaction was 30 µl. In general, 1 µl of DNA was used for each PCR. The final concentration of components of each PCR reaction mix is given below:

Buffer	1X (BioLine NH_4 buffer)
$MgCl_2$	1.5 mM
dNTPs	2.50 µM each
Taq DNA polymerase	1.5 U
Primer	2 µM each

The conditions for all PCRs for the study of polymorphism were similar except for the annealing temperatures. Standard cycling conditions consisting of 35 cycles were as in Table 2.1.

Table 2.1. Thermal cycling conditions.

Cycling step	Temperature and time	No. of cycles
Denaturation	94°C for 30 seconds	35
Annealing	V°C for for 30 seconds	
Extension (synthesis)	72°C for 30 seconds	
Final extension	72°C for 5 minutes	1

V stands for variable temperature for different primers, which derived from the Tm value of the primers (mostly Tm minus 5°C)

2.3.3. Nested PCR

Occasionally a nested PCR was necessary to achieve success in amplification particularly when the target DNA amount was very low as in a few stool and liver abscess DNA samples. In the nested PCR 1 µl of the tRNA primer-derived product was used as the template DNA. However all other PCR parameters remained exactly the same except that an increase of about 3 degrees in annealing temperature (Tm minus 2°C) was used.

2.3.4. Visualization of PCR products by gel electrophoresis

All PCR products were separated in 1.5% Bio-Gene or 1.8% 3:1 Nu-Sieve (FMC) agarose gels in 1X Tris-Borate-EDTA buffer and visualized after staining with ethidium bromide (1.0 µg/ml; Sigma). Gels were documented by photography and fragment sizes were estimated using the 100 bp DNA ladder (Promega, UK and Seegene, UK) run on the same gels.

2.4. Cloning of amplified products and plasmid isolation

2.4.1. Ligation

The cloning of amplified products was achieved by the use of the pGEM-T Easy Vector (Promega) and XL1-Blue Competent Cells (Stratagene) according to the manufacturers' instructions with proportionately reduced volumes. 1.0 µl of the PCR product was ligated to 0.5 µl of vector in a total volume of 5.0 µl. Then 2.0 µl of the ligation reaction was used to transform 35 µl of the XL-1 cells by heat-shock (45 seconds in a 42°C water-bath followed by immediate return to the ice for 2 minutes). The cells were then allowed to recover by the addition of 650 µl of chilled SOC medium (2% tryptone, 2% yeast extract, 10 mM NaCl, 0.25 mM KCl, and 20 mM glucose) and incubation for one hour at 37°C with shaking at 200-250 rpm. The transformed cells were then evenly spread on the surface of a LB agar plate containing ampicillin, IPTG (isopropyl ß-D-thiogalactopyranoside) and X-Gal (5-bromo-4-chloro-3-indolyl-ß–D-galactoside) and incubated overnight at 37°C. The

growth of blue and white colonies was observed the next morning. A white colony should contain the right insert, because a successful cloning interrupts the coding sequence of the enzyme ß-galactosidase that is responsible for the blue appearance of colonies. IPTG served as the inducer while X-gal served as the substrate for the enzyme ß-galactosidase. White colonies were picked and grown in 5 ml LB broth (Sigma) overnight at 37°C with shaking at 200-250 rpm.

2.4.2. Isolation of plasmid DNA by QIAprep Spin Mini Kit protocol

200 µl of LB broth containing bacterial cells was preserved at - 20°C after mixing with an equal volume of 50% glycerol for future use. The isolation of plasmid DNA from the remainder of the culture was according to the QIAprep Spin Mini Kit manufacturer's instructions, which is basically an alkaline lysis procedure.

2.4.3. *Eco*RI digestion of plasmid DNA

The isolated plasmid DNA was digested with restriction enzyme *Eco*RI (Promega) according to the manufacturer's instructions to reveal the size of the insert DNA and the digested products were run in a 1.2 % agarose gel (Bio-Gene, UK). Colonies with an appropriate insert were identified and selected for further analysis.

2.4.4. Rapid screening of plasmid by Sekar analysis

An alternative method for identifying recombinant plasmids was carried out using a screening method described by Sekar, 1987. This screening protocol was followed when 4 white colonies failed to reveal the insert of interest with the conventional method as described above.

1. 10-20 white colonies from LB agar plates were inoculated into 200 µl of LB broth supplemented with 100 µg/ml of ampicillin and incubated at 37°C for 5 hours on a shaker.

2. A small volume of each LB cell suspension was frozen at -20°C until needed.
3. The remaining cells were harvested by centrifugation and resuspended in 10 µl of Protoplasting buffer (30 mM Tris-HCl, pH 8.0, 5 mM EDTA, 50 mM NaCl, 20% Sucrose, 50 µg/ml RNase A, 50 µg/ml Lysozyme).
4. A 0.8% agarose gel in 1 X TBE was prepared with the addition of 0.05% SDS.
5. Gel wells were pre-loaded with 20 µl of Lysis buffer (1 X TBE, 2% SDS, 5% Sucrose and 0.04% bromophenol blue).
6. The protoplast suspension containing cells was then loaded into the wells.
7. Electrophoresis was carried out in 1 X TBE, initially for 15 minutes at 30 V and then 2 hours at 120 V.
8. Gels were stained with ethidium bromide (1.0 µg/ml) and photographed under UV light.

An 'empty' plasmid DNA was used as a size marker. Those plasmids that showed slower migration on the gel, possibly due to possession of additional 'insert DNA', were grown in LB broth, subjected to plasmid DNA isolation by the QIAprep Spin Miniprep Kit (Qiagen) as outlined above, and the size of the insert determined by *Eco*RI digestion.

2.5. Sequencing

Sequencing of the cloned insert was achieved using the ABI Prism® BigDye™ terminator cycle sequencing ready reaction kit according to the manufacturer's instructions except at half volumes. The reaction contained one of the two M13 primers shown in Table 2.2, and was prepared as follows:

BigDye reaction mix (v 3.1)	1µl
M13 primer (1.6 pmol/µl)	1µl
5X sequencing buffer	2 µl
Deionized distilled water	4 µl
Plasmid DNA	2 µl

This mixture was then overlaid with 1 drop of mineral oil and subjected to 30 cycles in a thermocycler each consisting of 96°C for 30 sec, 50°C for 15 sec, and 60°C for 4 min. The cycle sequencing products were precipitated using 75% ice-cold isopropanol (final concentration 60%), washed in 75% isopropanol, and dried in a heating block at 94°C for 2 minutes. The samples were then stored at -20°C until used. The reactions were analysed on a DNA sequencer ABI PRISM™ 377 (Perkin Elmer) by the LSHTM core facility and the output was examined for accuracy using Chromas software.

Table 2.2. Sequencing Primers.

Primer Name	Primer Sequence (5′ to 3′)
M13 forward	GTTTTCCCAGTCACGAC
M13 reverse	CAGGAAACAGCTATGAC

The direct sequencing of PCR products was achieved mostly by running the PCR products in a 1.5% agarose gel. The appropriate PCR product was excised from the gel and purified using the QIAquick® gel extraction kit (Qiagen, UK) according to the manufacturer's instructions. Depending on the concentration (as judged by the relative intensities of products on the gel) 1-6 µl of the gel purified product was used in a total volume of 10 µl with one of the corresponding species-specific primers, if possible, or one of the tRNA primers used in amplification, as above. Sometimes PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, UK) according to the manufacturer's instructions and used directly for sequencing as above.

The resulting sequences were assembled by either using the Multalin program (Corpet, 1988) (http://www.prodes.toulouse.inra.fr/multalin/multalin.html), or manually by eye.

2.6. Probe labeling

PCR product to be labeled was first purified using the QIAquick® PCR purification kit (Qiagen, UK) according to the manufacturer's instructions.

The labeling of the PCR product was achieved using the Prime-a-Gene® Labelling System (Promega, UK) according to the manufacturer's instructions. To the heat denatured DNA template we added nuclease free water, labeling buffer containing random primers, a mixture of unlabelled dNTPs and nuclease free BSA (bovine serum albumin). This was followed by addition of [α-^{32}P]-dCTP and DNA polymerase I (Klenow fragment) and incubation for 1 hour at room temperature. The reaction was terminated by the addition of EDTA at a final concentration of 20 mM and boiling. This labeled probe was then passed through a Sephacryl S-200 column (Amersham Bioscience, UK) to remove any unincorporated nucleotides.

CHAPTER 3

DEVELOPMENT OF A GENOTYPING METHOD

3.1. Introduction

A few PCR-based DNA typing methods have been reported for *E. histolytica*, making use of repetitive elements contained within both protein coding genes and non-coding DNAs (discussed in Section 1.3). However no single polymorphic locus can be used to detect all genotypes of *E. histolytica* and the need for using more than one locus for strain typing has been emphasized (Haghighi et al, 2002; Zaki and Clark, 2002).

To develop an optimal strain typing method, we decided to investigate the polymorphic potential of all tRNA-flanked STRs in *E. histolytica* using the raw sequences from the *E. histolytica* genome project, and then design and test species-specific primers for the most promising loci. We believed that the resulting method could give us the tools necessary to investigate the role of parasite genotypes in the outcome of infection with *E. histolytica* as well as to address other unanswered questions surrounding the epidemiology of this parasite.

3.2. *E. histolytica* and *E. dispar* samples

Eight axenic isolates of *E. histolytica* and 1 of *E. dispar* were used in this study for initial testing and selection of primers. Information on the patients' diagnoses and the origins of these is provided in Table 3.1.

In addition, DNA samples from 9 *E. histolytica* and 9 *E. dispar* isolates maintained xenically in Robinson's medium were received from ICDDR,B along with 2 DNA samples isolated directly from the aspirated pus of amoebic liver abscess (ALA) patients from

Bangladesh. These were used for subsequent testing of selected primers. The specificity of species-specific primers was tested using two DNA samples from reference isolates of *E. moshkovskii* (Laredo and FIC) and two *E. histolytica* positive and 2 *E. dispar* positive faecal DNA samples received from Vietnam and Bangladesh, respectively.

Table 3.1. Information on axenic isolates.

Species	Isolates	Clinical diagnosis	Suspected country of origin	Year of isolation
E. histolytica	HM-1:IMSS	Amoebic dysentery	Mexico	1967
E. histolytica	200:NIH	Amoebic dysentery	Isolated in the USA	1949
E. histolytica	IULA:0593:2	Diarrhoea, nausea, tenesmus, fever	Venezuela	1993
E. histolytica	H-303:NIH	Amoebic dysentery	Patient from Vietnam treated in the USA	1972
E. histolytica	HB-301:NIH	Amoebic dysentery	Burma	1960
E. histolytica	IULA:1092:1	Diarrhoea, nausea, tenesmus, fever	Venezuela	1992
E. histolytica	HK-9	Amoebic dysentery	Korea	1951
E. histolytica	Rahman	Asymptomatic	Sailor from India isolated in England	1972
E. dispar	SAW760	Asymptomatic	England	1979

3.3. Isolation of DNA

The CTAB DNA extraction method described in Section 2.2.1 was employed for all axenic and xenic isolates, pus specimens, and the two stool specimens from Bangladesh.

3.4. PCR amplification

In PCR amplifications two different groups of primers were used, 'tRNA primers' designed from tRNA gene sequences of *E. histolytica* HM-1:IMSS that amplify both *E. histolytica* and *E. dispar* DNA (Appendix I) and 'specific primers' designed after the comparison of homologous sequences from *E. histolytica* and *E. dispar* (Table 3.2) to amplify DNA from one species only. Naming of the STRs and primers was based on the single letter amino acid code for the relevant tRNA genes (or 5S RNA gene) flanking the STR being amplified.

The thermal cycler settings for all PCR reactions were the same as described in Section 2.3.2 except for the annealing temperatures.

Table 3.2. Selected tRNA primers and the species-specific primers.

A) Selected tRNA primers.

Primer name	Sequence (5' to 3')	AT
A-L5 A-L3	GGATCGATACCCCTCATCTCCA CGCATCTTGCGATAGCCGAG	64
D-A5 (EhR1)$ D-A3 (EhR2)$	CTGGTTAGTATCTTCGCCTGT CTTACACCCCCATTAACAAT	56
N-K5 N-K3	CGAACGGCTGTTAACCGTTA TTCCTAGCTCAGTCGGTAGA	55
R-R5 R-R3	AGCATCAGCCTTCTAAGCTG CTTCCGACTGAGCTAACAAG	55
S^{TGA}-D5 S^{TGA}-D3	CTCTGGATGCGTAGGTTCAA GTATCTTCGCCTGTCACGTG	58
S-Q5 S-Q3	GTGGTCTAAGGCGTGTGACT GAGATTCTGGTTCTTAGGACCC	56

B) *E. histolytica* specific primers.

Primer name	Sequence (5' to 3')	AT
AL-H5 AL-C	CATCTCCATTATTATCTAGATATCCTTTATTACT GGCACGAATGCTTTGATATATAA	57
DA-H5 (Hsp1)$ DA-H3 (Hsp2)$	GAGTTCTCTTTTTATACTTTTATATGTT ATTAACAATAAAGAGGGAGGT	50
NK2-H5 NK2-H3	GAAGCGTCTTTTTTACTATTAGTG GGCGTATTTTTAGAATAGGATAAG	59
RR-H5 RR-H3	GCGCCTTTTTATTCAATATACTCC GGATGAAGATATCTTCACAGGG	57
S^{TGA}D-H5 S^{TGA}D-H3	AAATCCTGCCACTGTCGTAA AATCCCCGTTGAAGAGTTCT	58
SQ5 SQ-H3	GTGGTCTAAGGCGTGTGACT GTGGGACCACTTTTTATACCTA	55

C) *E. dispar* specific primers.

Primer name	Sequence (5' to 3')	AT
AL-D5 AL-C	CATCTCCATTATTATGTATCTATTTATCTATTTA GGCACGAATGCTTTGATATATAA	60
DA-D5 (Dsp1)$ DA-D3 (Dsp2)$	TTGAAGAGTTCACTTTTTATACTATA TAACAATAAAGGGGAGGG	50
NK-D5 NK-D3	GCGGAAGCGTCTTTTTTACTATTATTA GGGCGTATTTTTATAATAAGAATAGGTGG	58
RR-D5 RR-D3	CATGAGGCGCCTTTTTATCA AGGGGATGATGATATTGAACACACTC	58
S^{TGA}D-D5 S^{TGA}D-D3	AAATCCTGCCACTGTCGTAC AATCCCCGTTGAAGAGTTCA	63
SQ-D5 SQ-D3	AGCATAGGTGTATAGGTGTTAGG GGTGGGACCATATTTTATACTTC	53

AT = Annealing temperature; suffixes: H = *E. histolytica*; D = *E. dispar*; C = Common (used in both *E. histolytica* and *E. dispar*); $ = These primers were previously described by Zaki et al (2002) using different names given inside the brackets.

3.5. Results

3.5.1. PCR product length polymorphism

In over one-quarter (13/46) of the STRs, amplification produced multiple products in one or more *E. histolytica* isolates (Table 3.3). Some STRs using tRNA primers show multiple bands because there is length variation between units in the same array, but in other cases the multiple bands have other origins. For example, STRs C-K, K-S, N-K, and K-N give 2 or more bands in HM-1:IMSS because the corresponding tRNA gene pairs are found in two distinct arrays. Similarly, V-5 amplification gives two bands because the primers amplify sequences from both [V5] and [VME5]. A second band is sometimes produced where the product is actually >1 array unit (e.g., T-R-T-R is amplified instead of T-R). These artifacts have been omitted from Table 3.3. Comparing the *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 results, we observed a single product at 40 STRs in these two species, and of which 36 STRs are homologous.

Table 3.3. Screening and selection of polymorphic STRs.

STR locus	*E. histolytica*		Degree of polymorphism	*E. dispar* SAW760		Further Investigation
	No. bands	Size range (bp)		No. bands	Size range (bp)	
A-A	1 (1)	580-600	+/-	1	750	No
A-L	1 (1)	580-620	+	1	520	Yes
L-A	1 (1)	610-640	+/-	1	720	No
A-S	1 (1)	500-520	+/-	1	610	No
S^{GCT}-D	1 (1)	320-350	+/-	1	340	No
D-A	1 (1)	390-420	+	1	500	Yes
G-G^{GCC}	1 (1)	810	-	1	815	No
G-G^{TCC}	1 (1)	450	-	1	550	No
H-H	1 (1)	630-650	+/-	1	650	No
L-S	1 (1)	500-550	+	1	600	No
S-L	1 (1)	500-530	+/-	2	410-570	No
L-T	1 (1)	390-410	+/-	1	430	No
T-L	1 (1)	580-600	+/-	1	600	No
M-R	1 (1)	480-580	+/-	1	480	No
R-M	1 (1)	630 or 520	+/-	1	520	No
N-K	1-2 (2)	540-830	+++	1	610	Yes
K-N	1-2 (2)	510-700	++	1	700	No
P-P	1 (1)	720-800	+/-	1	810	No
R-5	1-2 (1)	360-410	+	2**	350-380	No
5-R	1 (1)	440-500	+	1	420	No
R-R	1 (1)	690-730	+/-	1	690	Yes
R-T	1 (1)	450¥	+/-	1	600¥	No
T-R	1 (1)	570-600	+	1	350	No
S^{TGA}-D	1 (1)	275-310	+	1	270	Yes
D-S	1 (1)	540-570	+/-	1	630	No
S-P	1 (1)	490-530	++	1	590	No
P-C	1 (1)	410	-	1	450	No
C-K	2-3 (3)	520-600	+++	3	490-630	No
K-S	2 (2)	400-580	-	2	490-550	No
S-Q	1 (1)	390-460	+++	1	450	Yes
Q-C	1 (1)	230	-	1	200	No
T-Q	1 (1)	340	-	1	420	No
Q-T	1 (1)	500	-	1	600	No
T-X	1 (1)	590	-	1	700	No
X-T	1 (1)	630-650	+/-	1	750	No
V-5	2 (2)	510-530	+/-	1	510	No
5-V	1-2 (1)	280-350	+/-	2	310-480	No
V-F	1-3 (1)	450-600	++	2	550-750	No
F-V	1-2 (1)	330-400	+	1	820	No
V-M	1 (1)	220	-	1	220	No
M-E	1 (1)	550-710	++	1	600	No
E-V	1 (1)	600	+/-	1	800	No
W-I	1-2 (1)	530-800	+++	1	580	No
I-W	1-3 (2)	340-510	+++	1	520	No
Y-E	1-2 (1)	470-530	+/-	1	550	No
E-Y	1-2 (1)	630-680	+	1*	620*	No

All the size-ranges of the PCR products are approximate. Degree of polymorphism: '-' = none; '+/-' = little; '+' = some; '++' = high; '+++' = extensive. The rating is subjective.
No. bands: The numbers inside the brackets represent the number of band(s) observed in *E. histolytica* HM-1:IMSS. *A very faint band was observed. **Two very faint bands were observed. ¥ Another band at around 1000 bp and 1200 bp of unknown origin was observed in *E. histolytica* isolates and *E. dispar* SAW760, respectively.

3.5.2. Selection of polymorphic loci

The selection of polymorphic loci for further investigation was based on: 1) the degree of polymorphism they showed; 2) the number of amplification products they produced (loci that gave a single product were selected); 3) the relative success of PCR amplification in different DNA sample types; and 4) our ability to design species-specific primers. On the basis of these criteria we selected 6 out of the 46 loci for further investigation (Table 3.3).

The only exception to these criteria was N-K, for which we observed two bands in all isolates except two from Venezuela, which gave a single product (Figure 3.1). The smaller of the two *E. histolytica* HM-1:IMSS products (designated N-K1) was present and the same size in all others but absent from the Venezuelan isolates. Following cloning and sequencing of both PCR products from 2 isolates (HM-1:IMSS and H-303:NIH) we found the STRs and flanking sequences to be different. The primers were therefore amplifying two distinct STRs. The larger product is polymorphic and appears to be present in all isolates. We named this STR locus N-K2 and designed primers specific for this sequence. Unlike N-K, PCR amplification with N-K2 primers gave a single product.

Although the STR loci C-K, E-Y, F-V, I-W, K-N, and W-I showed substantial polymorphism, a majority of isolates gave multiple PCR products (Figure 3.2). Because multiple banding might complicate the future typing of strains, we decided not to investigate these loci further. Likewise, the species specific primer pair of STR locus R-T was producing multiple products with stool DNA and was excluded from further study. Similarly, although the loci M-E and S-P showed high polymorphism and produced a single product in a majority of the axenic *E. histolytica* isolates, we chose not to pursue these loci any further. We were unable to design reliable species-specific primers for locus M-E, despite repeated efforts, and we found amplification of locus S-P to be unreliable even using DNA from xenic cultures. For example, an initial investigation using 76 xenic culture DNAs showed that S-P PCR was only about 76% successful in comparison with 100% PCR success observed at loci A-L and S-D. It is possible that the copy number of

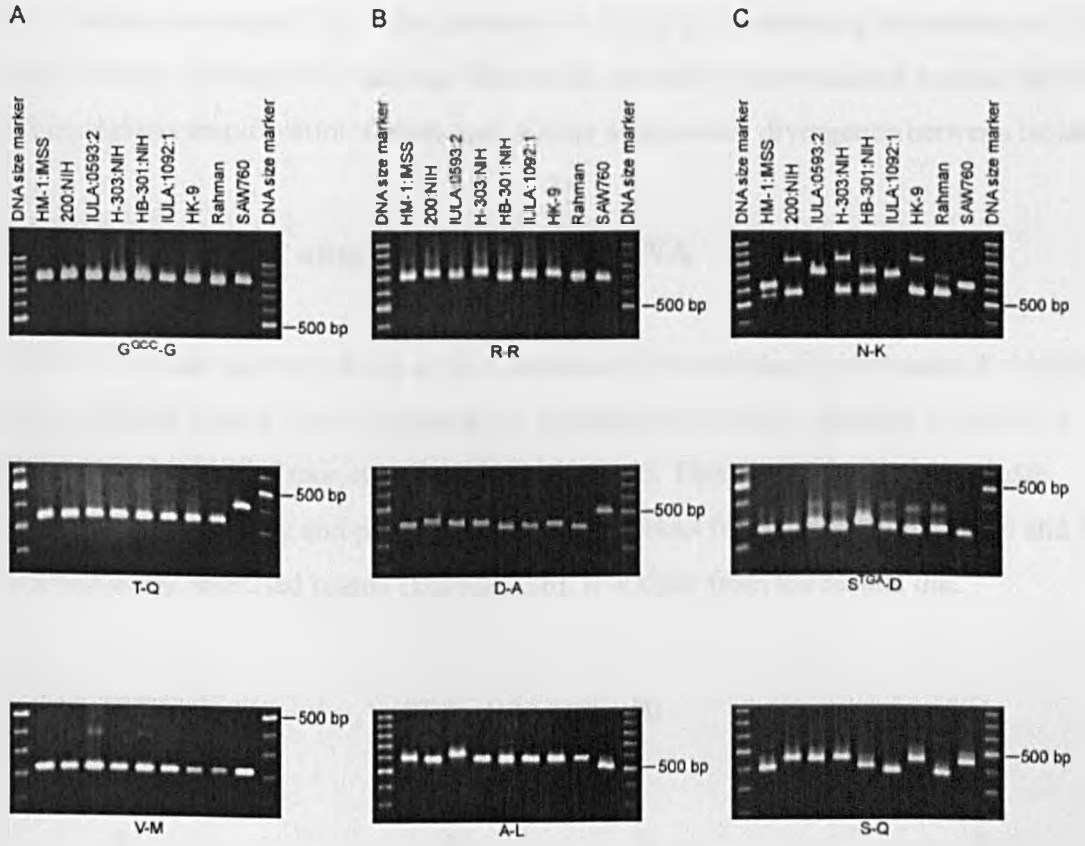


Figure 3.1. Examples of polymorphism observed in tRNA-linked STRs of *E. histolytica*. Panel A, 3 STRs showing little or no PCR product size polymorphism. Panel B, 3 of the selected STRs showing low to moderate polymorphism. Panel C, the remaining 3 selected STRs showing moderate to high polymorphism. All samples are *E. histolytica* except for *E. dispar* SAW760.

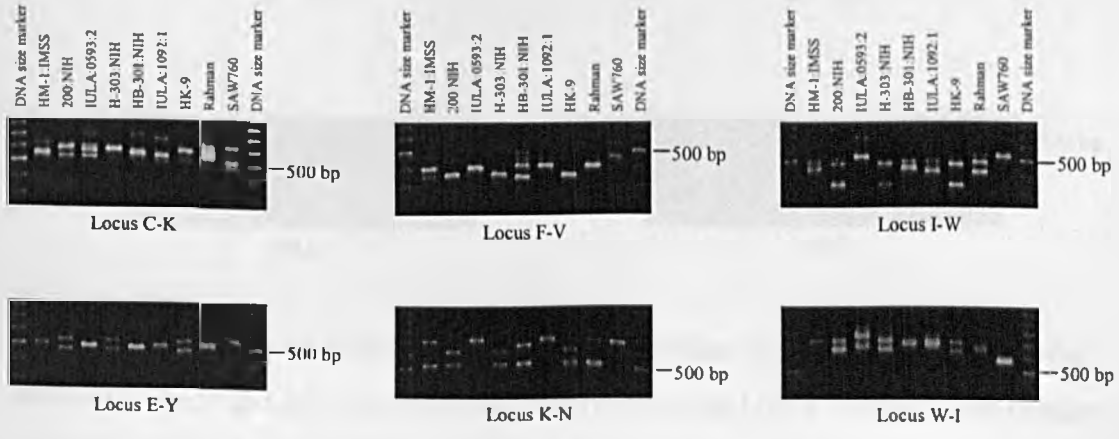


Figure 3.2. Example of STR loci showing polymorphism but not selected because they produced multiple bands in PCR amplifications.

these tandemly arranged loci in the genome is a factor in determining the success of PCR amplification. Alternatively, because the species-specific primers extend beyond the tRNA-coding regions amplification failure may be due to sequence divergence between isolates.

3.6. Polymorphism analysis using xenic DNA

PCR product size polymorphism at the 6 selected STRs was then tested using *E. histolytica* DNAs isolated from 9 xenic cultures and 2 lyophilised ALA pus samples. Similarly 9 DNAs from xenic *E. dispar* cultures were also tested. This allowed us to investigate amplification efficiency and polymorphism using DNAs from other sample types and a geographically restricted region (Bangladesh). It is clear from the results that

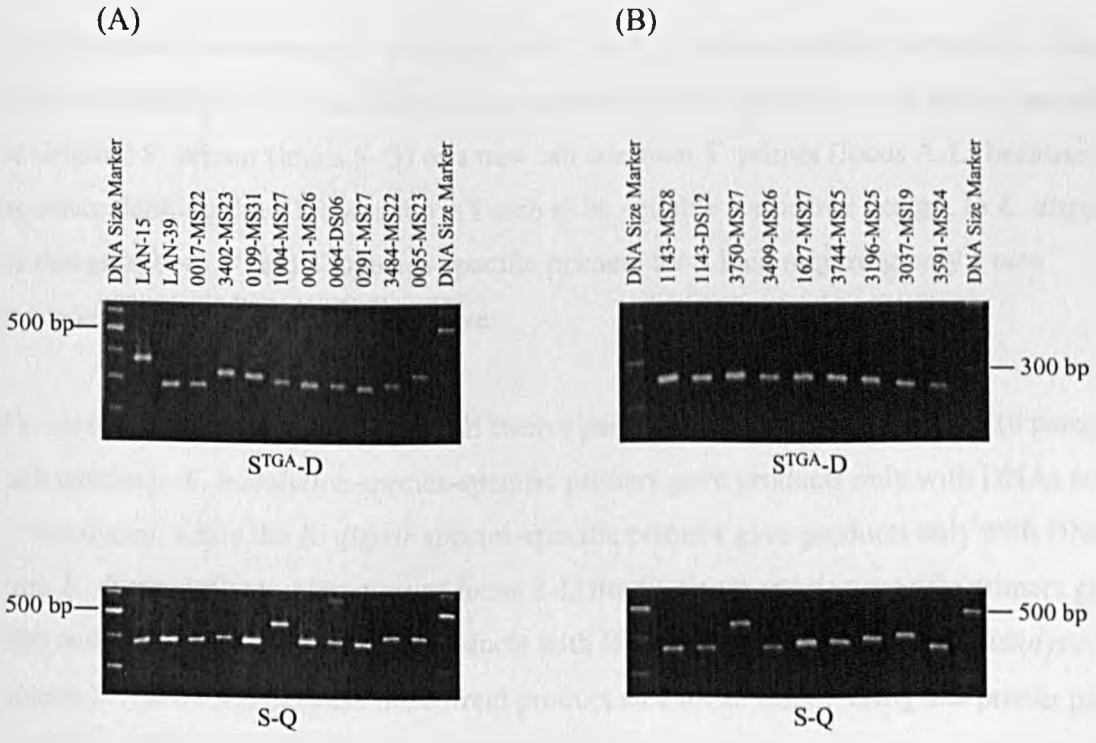


Figure 3.3. Examples of PCR product size polymorphism with (A) *E. histolytica* liver abscess (LAN-15 & LAN-39) and xenic culture DNAs and (B) *E. dispar* xenic culture DNAs at 2 of the selected STR loci (S^{TGA}-D and S-Q) using tRNA primers.

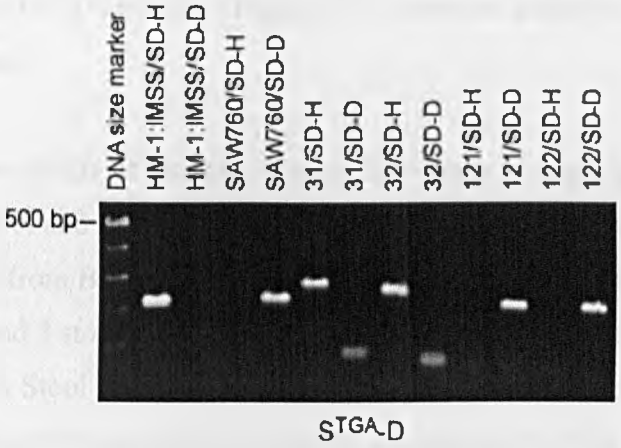
all STRs amplified well and showed polymorphism, with loci S-D and S-Q showing the most variation in *E. histolytica* (Figure 3.3A). However locus S-D showed almost no detectable polymorphism in *E. dispar* strains (Figure 3.3B).

3.7. Design and testing of the species-specific primers

In order to design species specific primers we sequenced the PCR products of *E. dispar* SAW760 for all 6 selected loci. We also sequenced PCR products from a second isolate of *E. histolytica* to ensure that the specific primer sequences were conserved within the species. All species-specific primers were tested using two control DNAs, from axenic *E. histolytica* HM-1:IMSS and *E. dispar* SAW760, and with DNA isolated from 2 xenic *E. histolytica* and 2 xenic *E. dispar* cultures (Figure 3.4).

In *E. histolytica*, we were able to design both 5′ and 3′ species-specific primers for 4 loci. For the remaining 2 loci, we designed one species-specific primer, but we had to use either the original 5′ primer (locus S-Q) or a new but common 5′ primer (locus A-L) because the sequence flanking the STR was too AT-rich to be suitable for primer design. In *E. dispar*, we designed both 5′ and 3′ species-specific primers for 5 loci, requiring only a new common primer for locus A-L as above.

We observed specific products with all twelve pairs of species-specific primers (6 pairs for each species): *E. histolytica* species-specific primers gave products only with DNAs from *E. histolytica*, while the *E. dispar* species-specific primers gave products only with DNA from *E. dispar* isolates. However, at locus S-D the *E. dispar* species-specific primers gave faint and small (less than 100 bp) products with DNAs from the two xenic *E. histolytica* isolates (Figure 3.4). Because the correct product size for *E. dispar* using this primer pair ranges from 200 to 220 bp, we do not think that these smaller non-specific products will interfere with species identification.



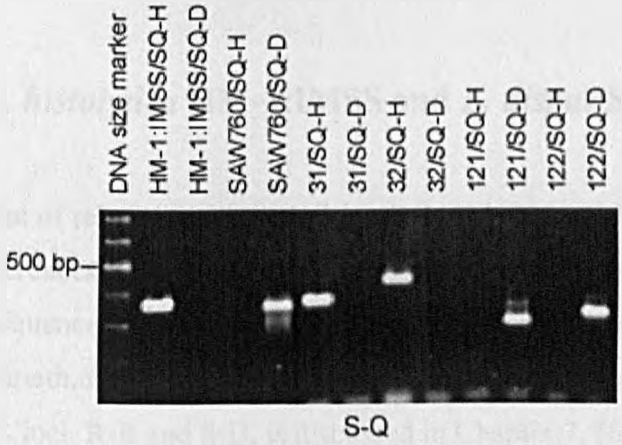


Figure 3.4. Species-specific primer testing. Each of the DNA samples was amplified using both the *E. histolytica*- (suffix –H) and the *E. dispar*-specific (suffix –D) primers for 2 of the selected STRs (Table 3.2). *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 were used as controls. 31 and 32 are known *E. histolytica* isolates while 121 and 122 are known *E. dispar* isolates from Bangladesh and the DNA was extracted from xenic cultures.

We also tested our species-specific primers with two *E. moshkovskii* reference DNAs: one was *E. moshkovskii* Laredo, an organism isolated from a human infection and the other was the *E. moshkovskii* FIC, an organism isolated from an environmental sample. No PCR amplification was observed with these *E. moshkovskii* DNA samples using the *E. histolytica* and *E. dispar* species-specific primers, except for 2 faint bands both with *E. moshkovskii* FIC DNA, one with the *E. histolytica*-specific S-P primers and the other with

the *E. dispar* specific W-I primers (Figure 3.5). However these two loci were not among the final 6 selected.

We then tested the ability of the species specific primers to amplify *E. histolytica* DNA extracted from stool and liver abscess pus samples. For this, we used 2 ALA and 2 stool samples obtained from Bangladesh, which were extracted by the modified CTAB method (Section 2.2.1), and 2 stool DNA samples from Vietnam, which had been extracted using the QIAamp DNA Stool Mini Kit (Section 2.2.3). Both the liver abscess and the stool DNA samples, regardless of the country of origin or the method of DNA isolation, were successfully amplified at all selected loci (Figure 3.6).

3.8. STRs in *E. histolytica* HM-1:IMSS and *E. dispar* SAW760

From the alignment of relevant sequences from *E. histolytica* and *E. dispar* isolates we examined the differences between the two species. We looked for short tandem repeats (STRs) in these sequences either by eye or by using Tandem Repeats Finder software (http://tandem.biomath.mssm.edu/trf.html). The sequence diversity among *E. histolytica* strains at two STR loci, R-R and S-D, is discussed in Chapter 7. Here we compared the extent of divergence between *E. histolytica* HM-1:IMSS and *E. dispar* SAW760.

Schematic representations of a selection of STRs are provided in Figure 3.7. With the exception of locus S-D, in which we could not detect STRs in *E. dispar* SAW760 all other loci contained STRs in both *E. histolytica* and *E. dispar*.

In some loci we found more repetitive DNA elements in one species than the other. For examples, in loci A-L and W-I, we detected more types of STR in *E. histolytica* HM-1:IMSS compared to *E. dispar* SAW760 whereas in locus N-K we found more STRs in *E. dispar* SAW760 compared to either N-K1 or N-K2 in *E. histolytica* HM-1:IMSS. In all loci the STR sequences themselves are completely different in these two species and even the flanking sequences of the STRs are very different from each other.

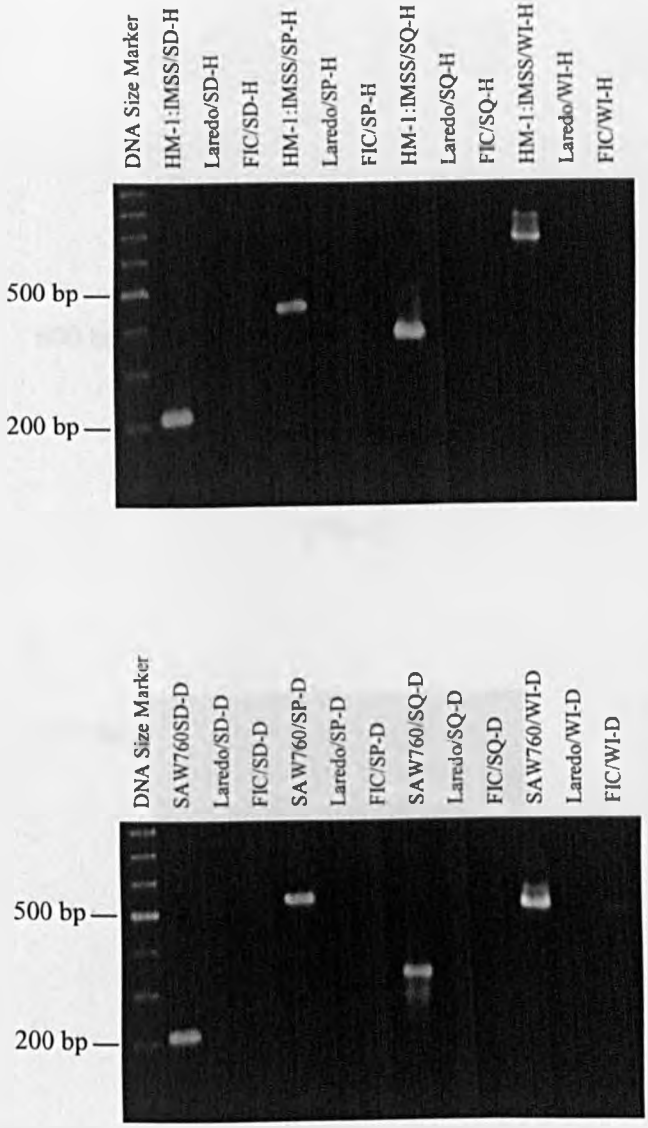


Figure 3.5. Specificity of *E. histolytica* and *E. dispar* species-specific primers tested with *E. moshkovskii* DNA. Laredo is a reference human isolate and FIC is an environmental isolate of *E. moshkovskii*. *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 were used as the positive controls. Each of the DNA samples was amplified using both the *E. histolytica*-(suffix –H) and the *E. dispar*-specific (suffix –D) primers. Although the STR loci S-P and W-I are shown in the Figure they were ultimately not selected for further analysis.

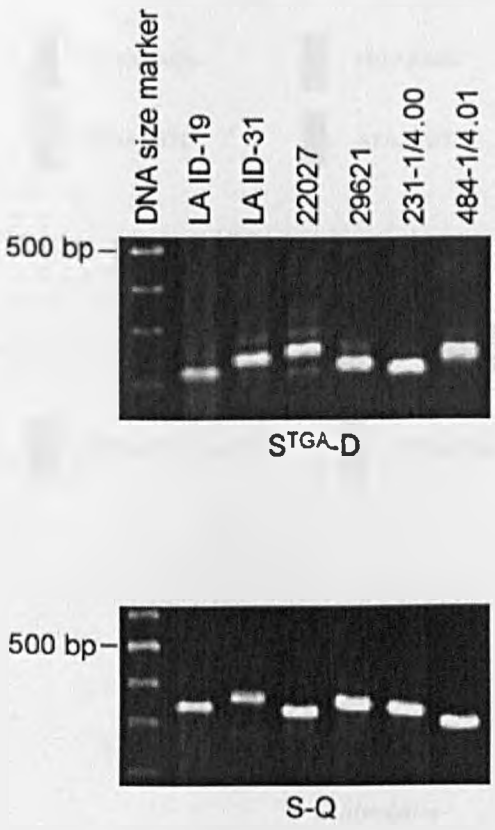
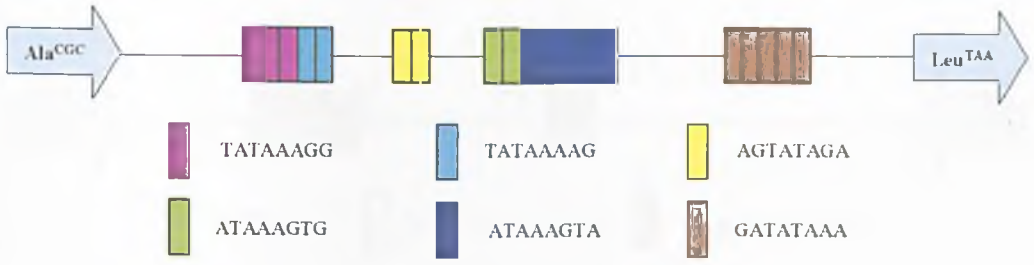


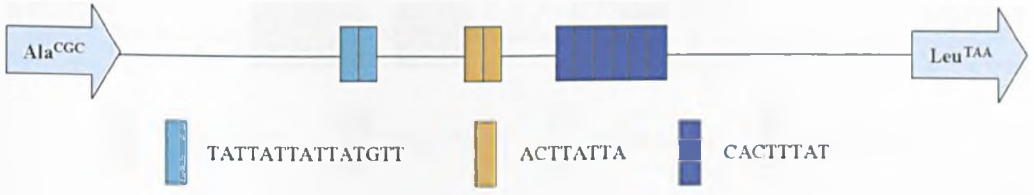
Figure 3.6. STR amplification using DNA extracted from faeces and liver abscess pus. *E. histolytica* species-specific primers (Table 3.2) were used. LAID-19 and LAID-31 are DNA samples extracted from liver abscess pus. 22027 and 29621 are faecal DNAs from Bangladesh isolated by the modified CTAB method. 231-1/4.00 and 484-1/4.01 are faecal DNAs from Vietnam isolated using the QIAamp DNA Stool Mini Kit.

A-L

E. histolytica

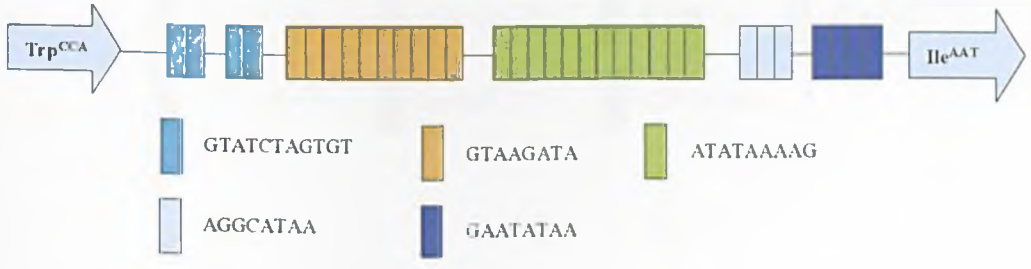


E. dispar

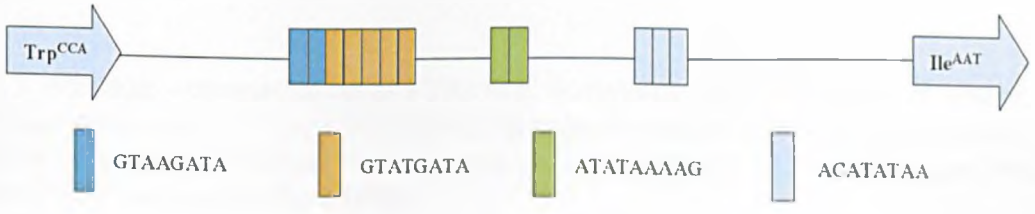


W-I

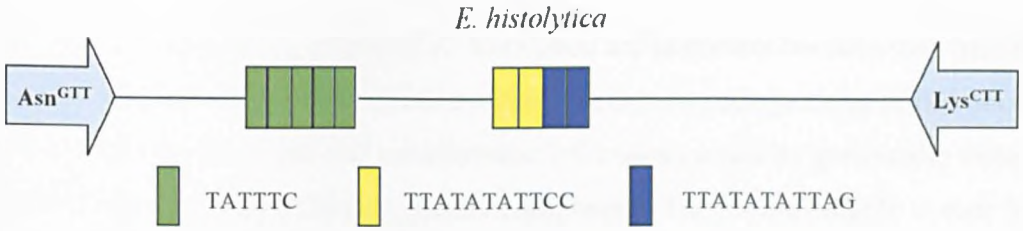
E. histolytica



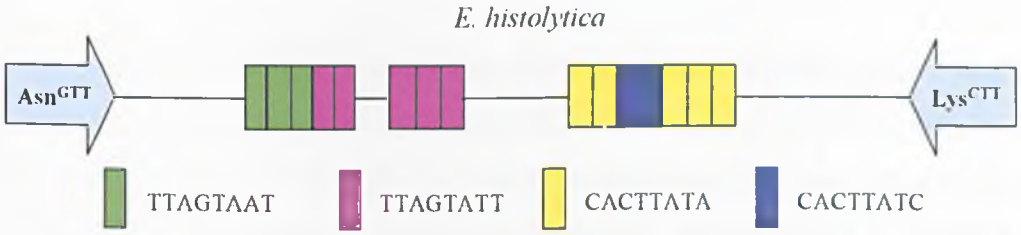
E. dispar



N-K1



N-K2



N-K

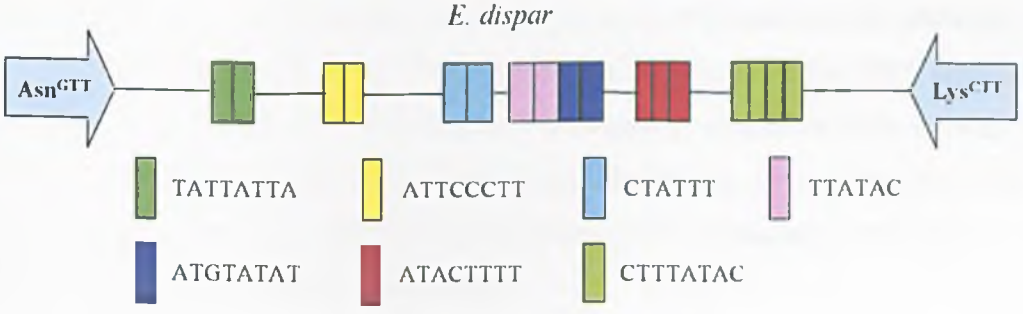


Figure 3.7. Schematic representation of STRs in *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 at 3 loci – A-L, N-K and W-I. The STR colours indicate different sequences in the same locus but are not the same between species or loci. Both STR loci N-K1 and N-K2 are shown for *E. histolytica* HM-1:IMSS.

3.9. Discussion

Tools that allow us to identify strains of *E. histolytica* are important because they may help to address some of the unanswered questions surrounding the pathogenicity of this parasite. For example, are asymptomatic and symptomatic infections caused by genetically distinct strains and does organ tropism have a genetic component? The tools available to date for identification of individual strains have been limited (Section 1.3).

The STR-containing loci originally studied by Zaki and Clark (2001) have shown promise as strain identification markers (Haghighi et al, 2002, 2003; Zaki et al 2002, 2003a, b). In the present study, we have developed an extended set of polymorphic markers making use of the presence of STRs adjacent to most tRNA genes in *E. histolytica* and *E. dispar*, a unique feature of these species (Loftus et al, 2005). The STRs selected for our optimal panel of markers were chosen on the basis of their reliability, sensitivity and degree of polymorphism detected in PCR. We have also designed *E. histolytica* and *E. dispar* species-specific primer pairs for all the selected loci, which will eliminate the potential problems caused by mixed infections. The primers have been tested using DNA samples from diverse geographic locations, and they have successfully amplified DNA isolated directly from liver abscess pus and stool samples, eliminating the need for culture of the amoebae. The DNA can be prepared for analysis using either a commercial kit or a modification of an existing purification method.

Zaki et al (2003b) have shown that amplification of tRNA loci using DNA prepared from cultures reflects the patterns seen with DNA extracted from the original stool sample, indicating that while culture is not necessary it is not misleading either. Recently, some of the STR loci were used to investigate the genotypes of *E. histolytica* and *E. dispar* in South Africa (Zaki et al 2003a). Among other things this study showed that the genotypes remained stable over time in the same infection. We have also observed no changes in the patterns obtained at any of the STR loci using DNAs extracted from *E. histolytica* HM-1:IMSS or *E. dispar* SAW760 at various times over the past several years. The markers therefore appear to be sufficiently stable to use for our intended purposes.

We and others have observed a considerable degree of STR-length polymorphism not only among axenic isolates of *E. histolytica* but also among isolates isolated from a restricted geographic location (Haghighi et al, 2002, 2003; Zaki et al, 2003a; and Chapters 4 and 5). Comparison of nucleotide sequences among *E. histolytica* isolates at many of the loci studied here revealed that the differences in PCR product sizes are due mainly to variable number of STRs, in agreement with earlier observations at two loci (Haghighi et al, 2003; Zaki and Clark 2001). Although the sequences of almost all the tRNA genes themselves are identical between *E. histolytica* and *E. dispar*, the STR regions flanked by them are completely different in sequence.

Since the genotyping of strains from restricted geographic locations has detected a high degree of polymorphism (Haghighi et al, 2002; Zaki et al, 2003a), it is very surprising that the PCR product sizes for 3 axenic isolates of *E. histolytica* (200:NIH, H-303:NIH, and HK-9) are identical at all 46 loci tested in this study. It is most likely that at some point the cultures of these isolates have been mislabeled. One of the unexpected applications of our markers may therefore be to verify the identity of laboratory isolates and detect such mix-ups.

We use our markers to investigate the possibility of a correlation between parasite genotype and clinical outcome of infection using samples from Bangladesh and other countries in Chapters 4 and 5, respectively. We believe that these markers should also prove useful in studying the patterns of transmission of this important disease and epidemiological links between individual infections.

CHAPTER 4

GENOTYPING OF BANGLADESHI *E. HISTOLYTICA* AND *E. DISPAR* STRAINS

4.1. Introduction

As few as one in ten *E. histolytica* infections result in signs and symptoms of intestinal or extraintestinal disorders in humans. What contributes to the variable virulence in this parasite is not clear. As a result, we still do not know whether a strain that is responsible for an asymptomatic infection in one individual has the potential to become invasive and to cause disease in another. In other words, are some *E. histolytica* strains intrinsically avirulent? It is possible that the genetic make up of the parasite is the determining factor for the clinical outcome in the infected individuals.

Duggal et al (2004) recently reported a potential association between a HLA class II allele and protection against *E. histolytica* infections in a cohort of Bangladeshi children, which they studied for 3 years. However, in the absence of appropriate knowledge about the *E. histolytica* genotypes and their levels of virulence in the host, it is difficult to know whether these study children were exposed to comparable *E. histolytica* strains. Therefore, we believe there is a need to investigate the existence of an association between parasite genotype and the clinical outcome of *E. histolytica* infections.

Genetic variation has previously been studied in two tRNA-gene linked loci in *E. histolytica* and *E. dispar* (Zaki and Clark, 2001; Zaki et al, 2002 and 2003a,b). Haghighi et al (2002) compared the polymorphic potential of these loci, 1-2 and 5-6, with those of protein-coding chitinase and SREHP genes. SREHP showed the most polymorphic patterns while chitinase was the least polymorphic.

However no association between the parasite genotypes and the outcome of infections has been reported for *E. histolytica.* We therefore decided to investigate the genotypes of

clinical strains of *E. histolytica* from Bangladesh using our multilocus genotyping system (Chapter 3) to see whether these show an association with the clinical outcome of infection.

4.2. Materials and methods

4.2.1. *E. histolytica* samples

Three clinical groups of samples from Bangladeshi *E. histolytica* infections, i. e., asymptomatic, diarrhoeal or dysenteric, and amoebic liver abscess (ALA), were collected and studied here. All the asymptomatic and a majority of the diarrhoeal/dysenteric samples were received or collected from an ongoing field project in Mirpur.

Dr. Rashidul Haque, ICDDR,B, Bangladesh, and Prof. William A. Petri, Jr., University of Virginia, USA, are jointly studying preschool children (2–5 years old) for immunity to amoebiasis through a field based project in Mirpur, about 5 miles from Dhaka city, the capital of Bangladesh. The inhabitants of Mirpur are of Bihari ethnic origin and settled there after the liberation of Bangladesh from Pakistan in 1971. The area is densely populated and has poor sanitary and hygienic conditions. They do not have water supplied directly to their houses neither do they have proper sanitary latrines in their house. They bring water from common water supply sources to their houses. There are several community latrines in the area which they use. Waste disposal is through Dhaka City Corporation but is maintained very poorly in this area. Most of these study children live in slums (in Figure 4.1, this is indicated by the household dots falling outside the small squares), but a small number of them live in small plots in the same area (Figure 4.1). Stool specimens are obtained every month (designated by a MS prefix; MS stands for monthly stool) for detection of *E. histolytica* infection by antigen capture ELISA and culture. If a study child develops diarrhoea/dysentery then its parents contact the project personnel at the field clinic and children and their parents are visited and interviewed every other day by health care workers. When diarrhoeal disease is detected, a stool sample (designated by a DS prefix; DS stands for diarrhoeal stool) is obtained and studied for enteropathogens, and the child is examined. All these samples are transported to the laboratory for processing within 2–4 h.

In addition to the Mirpur project, diarrhoeal samples were also obtained from patients who visited the outpatient department of the ICDDR,B hospitals.

ALA patients were admitted to various clinics and hospitals in Dhaka city other than ICDDR,B hospital. The aspirated pus from the liver abscess (designated by a LA prefix) was sent to ICDDR,B for diagnostic purposes.

A total of 161 *E. histolytica* positive samples, as diagnosed by a stool ELISA and / or SSU rRNA gene amplification, were used in this study. They were received or collected in 4 batches. The first batch of samples was received as DNAs purified from xenic culture. The second batch of samples was collected by me during my field trip to Bangladesh in April to July 2003. These were either stool or amoebic liver abscess pus samples. They were brought to LSHTM from Bangladesh in the form of lysates. A third batch containing more stool and liver abscess pus samples was received from ICDDR,B during December 2003 to February 2004. The fourth batch of samples was received as xenic culture DNAs during the third quarter of 2004.

These 161 samples include a total of 66 samples from 43 asymptomatic children including one from a child's parent from Mirpur, 52 diarrhoeal/dysenteric samples from both 38 children from Mirpur and 6 patients from the ICDDR,B hospital, and 43 liver abscess pus samples from 43 ALA patients. Only 9 of the 52 diarrhoeal/dysenteric samples were collected during an episode of dysentery (as opposed to diarrhoea). Eleven children were common in both asymptomatic and diarrhoeal/dysenteric groups.

In total, 70 children from Mirpur, both asymptomatic and with diarrhoea/dysentery, provided 112 samples, including 23 children who gave more than one (follow-up) samples.

4.2.2. *E. dispar* samples

A total of 68 *E. dispar* positive samples from 42 children, as diagnosed by a stool ELISA and / or SSU rRNA gene amplification, were received in the form of xenic culture DNA

from Bangladesh. All but one of these samples was from 2-5 year old children from the same on-going Mirpur project. This includes 61 samples from 37 children with no diarrhoea/dysentery, while 7 stool samples were collected during an episode of diarrhoea from 6 children. The exception is from a parent of a child, ID 1369-FS7.2; prefix FS stands for family sample. One child (3744) is represented in both the asymptomatic and diarrhoeal/dysenteric group.

4.2.3. Isolation of DNA and PCR amplification

DNA from the culture, stool or ALA pus lysates was isolated as already described (Section 2.2). PCR amplification was performed using the *E. histolytica*-specific primer pairs at 6 loci given in Table 3.2. Occasionally PCR was performed using the tRNA primer first and this was then used in a nested PCR with the corresponding *E. histolytica*-specific primers.

4.3. Data analysis

4.3.1. Genotype assignment

Each unique amplification pattern for a given locus was assigned a different number. Using the combination of numbers for a particular isolate at all 6 loci, it was assigned a genotype number. Similarly, results from various combinations of 5 loci (6 combinations), 4 loci (10 combinations) and 3 loci (20 combinations) were also used to assign a genotype specific for each combination.

4.3.2. Statistical analysis

4.3.2.1. Genotype distribution among clinical groups

Statistical analysis of the genotype distribution among 3 clinical population groups - asymptomatic, diarrhoeal/dysenteric and liver abscess - was done using a web based version of the analytical program GENEPOP (http://wbiomed.curtin.edu.au/genepop/). The

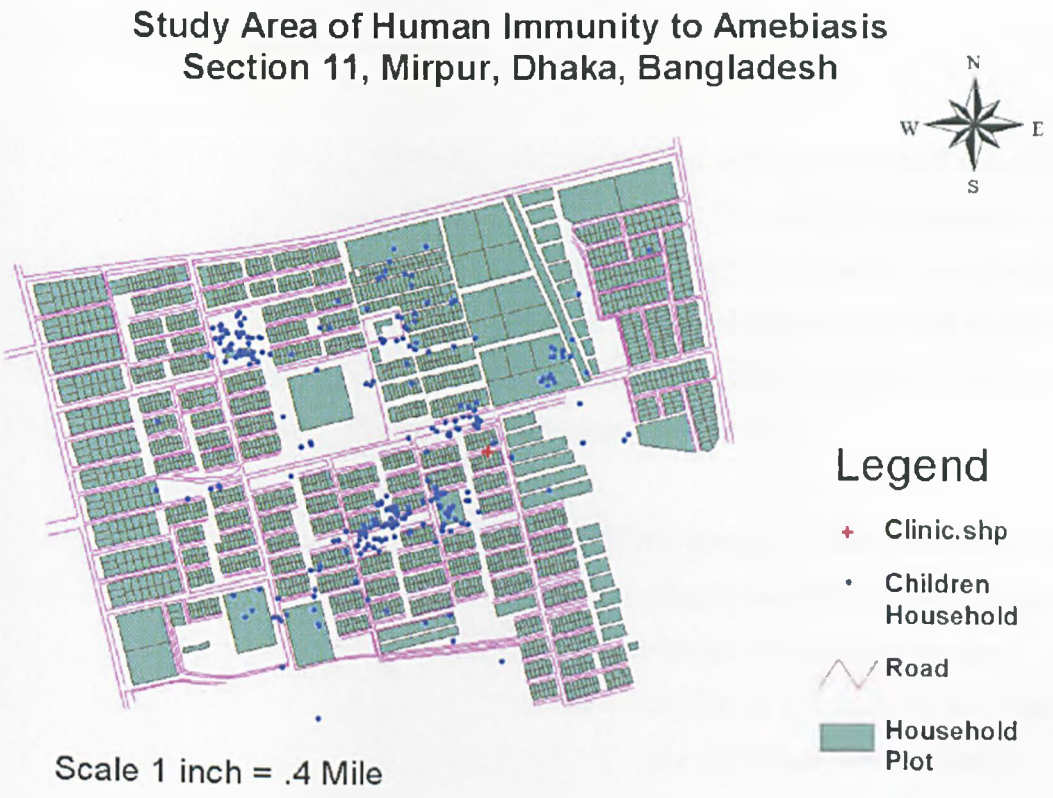


Figure 4.1. Map of the Mirpur study area. The map shows the distribution of households of all children enrolled in the 'Human Immunity to Amoebiasis' study, courtesy of Dr. Rashidul Haque, ICDDR,B, Dhaka, Bangladesh. Big squares indicate primarily school, market, playgrounds, ponds, etc, while small squares indicate housing plots, however some of the slum houses were located inside big squares (personal communication with Dr. Rashidul Haque).

software takes each PCR pattern of a particular locus into account while doing the analysis. The data were 'diploidised' before analysis of genotypic differentiation among populations. An unbiased estimate of the P-value of a log-likelihood (G) based exact test is performed.

The association of particular genotypes with clinical outcomes was analysed using the Chi square test (Epi Info, version 6.04). A P value of <0.05 was considered significant. The total number of predicted genotypes was calculated using the Chao Estimator equation (Colwell and Coddington, 1994).

4.3.2.2. Relationship among genotypes

The relationship among various genotypes was investigated using the eBURST program which is available at http://eburst.mlst.net (Feil et al, 2004). The eBURST algorithm was originally devised as a way of displaying the relationships between closely-related isolates of a bacterial species or populations. The procedure was developed for use with the data produced by multilocus sequence typing, but using a suitable criterion for the definition of groups of related genotypes, it can be used with other types of data.

The first step in eBURST analysis is to define meaningful groups in the experimental data. In our case, we used the 6 locus genotype data and investigated the relationship among the genotypes. In our definition, two genotypes will only be linked to each other by the eBURST diagram if they are identical at all but one locus, that is, at 5 out of 6 loci. This type of variant is called a single locus variant (SLV). If the difference between two genotypes is at two loci then each one is a double locus variant (DLV) to the other. The program automatically chooses a founder genotype, which by definition has the maximum number of SLVs and DLVs. However, it is possible to change the founder genotype assigned by the program to a new one if the bootstrap support for the new one is strong. eBURST clusters related genotypes into groups containing two or more genotypes each. The major group contains the maximum number of genotypes and is represented in the centre of the eBURST diagram while the minor groups are often placed around the periphery of the diagram. The genotypes that do not have SLVs or DLVs are termed singletons and are displayed in the eBURST diagram as unlinked dots. The numbers shown in the eBURST diagram are the genotype numbers and the size of a dot representing a genotype is proportional to the number of samples with this genotype. eBURST diagrams were also generated using 5 locus, 4 locus and 3 locus data sets.

4.4. Results

4.4.1. PCR amplification

PCR amplification was successful for all samples regardless of the origin of the DNA or the method of isolation of the DNA. However, a nested PCR using the tRNA primer-derived PCR product as the template was employed for a few stool and liver abscess DNA samples. This was necessary because the amount of target DNA in these samples seemed to be too low to be amplified by a simple PCR (data not shown).

4.4.2. Genotypes based on PCR amplification patterns at 6 loci

In 9 out of 161 samples, we observed a second faint band in PCR amplifications of some loci (most of these are shown in Figure 4.2), but for simplicity, we considered only the higher intensity band for genotype assignment.

In order to do the most meaningful analysis of association between genotype and clinical outcome of infection, we analyzed the genotypes of all Mirpur samples first since they were from a limited geographic area. We then compared these genotypes with those from liver abscess samples coming from different parts of Dhaka and often outside of Dhaka. We did not include the 6 ICDDR,B hospital samples as the number in this group was very low (only 6) and the geographic origin of these patients was unknown. As a result, they did not fit in with the Mirpur samples.

If the same genotype was identified in both asymptomatic and diarrhoeal samples from the same child it was considered to be 'diarrhoea-causing' and we removed it from the asymptomatic group. The total number of individuals, therefore, became 36 in the asymptomatic group. Two asymptomatic children showed 2 genotypes each in their follow-up samples: IDs 0291 and 1325. The total number of individuals in the diarrhoeal group of Mirpur children was 38, and again in this group follow-up samples of some children

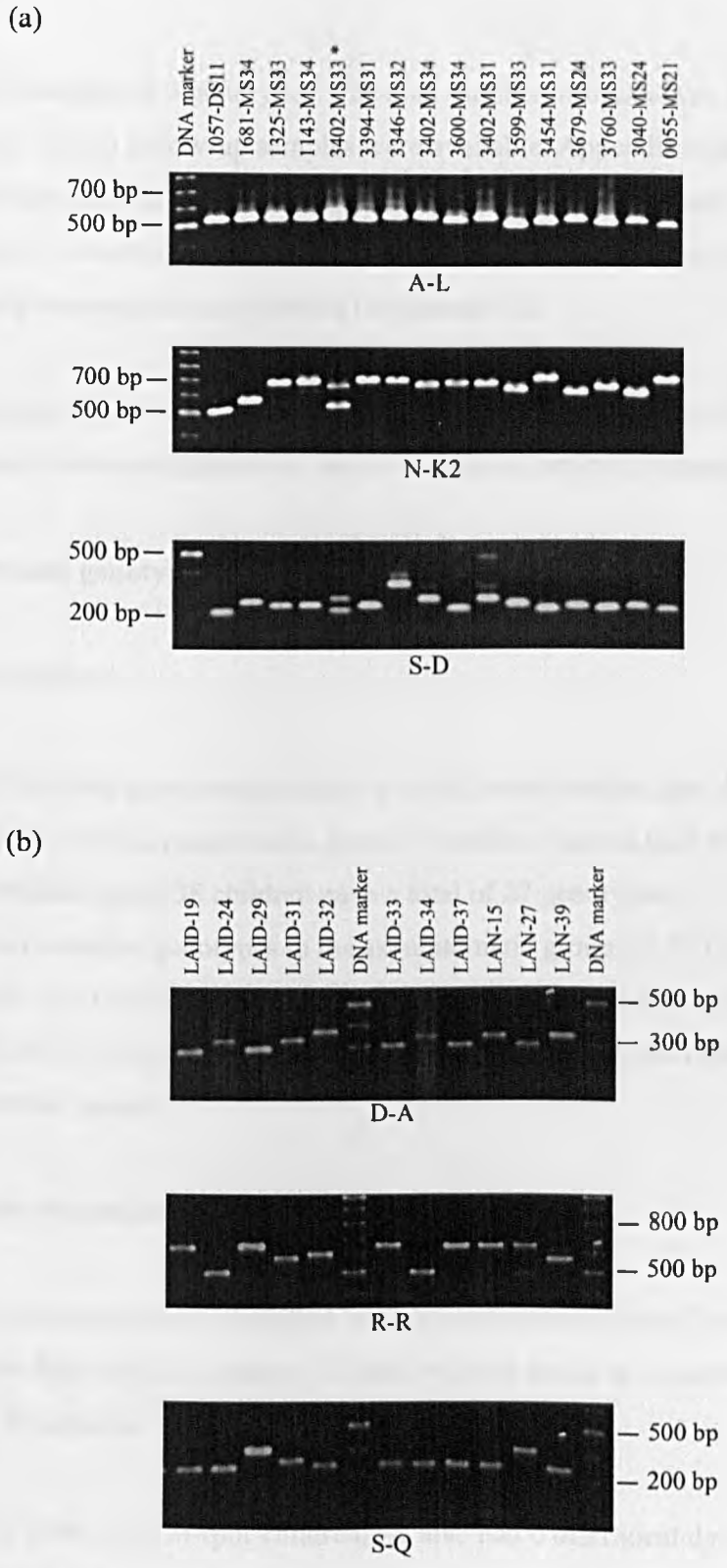


Figure 4.2. Gel electrophoresis patterns with a selection of *E. histolytica* samples. (a) Mirpur samples at STR loci A-L, N-K2 and S-D, (b) liver abscess samples at STR loci D-A, R-R and S-Q. * = indicates the transiently infected sample showing double bands at loci N-K2 and S-D.

while ID 3570 displayed 3 genotypes. The total number of individuals in the liver abscess group was 43, and no follow-up samples were available. Appendix II shows all the genotypes observed in asymptomatic, diarrhoeal/dysenteric (from both Mirpur and ICDDR,B) and liver abscess groups. Where multiple follow-up samples showed the same genotype only one sample was included in Appendix II.

Across all samples, a total of 26, 33 and 41 genotypes were identified among asymptomatic, diarrhoeal/dysenteric, and liver abscess patients, respectively.

4.4.2.1. Six-locus genotypes

(a) Mirpur children

A total of 70 children gave asymptomatic and diarrhoeal samples and these produced a total of 48 genotypes. In the asymptomatic group, 36 children gave a total of 26 genotypes, while in diarrhoeal group 38 children gave a total of 27 genotypes.
The two most common genotypes in the asymptomatic group are 51 (4 children), 13 and 19 (3 children each). In the diarrhoeal group, the two most common genotypes are 66 (10 patients), 50 and 51 (3 patients each; note that genotype 51 was also found in 4 children in the asymptomatic group).

(b) Liver abscess patients

A total of 41 genotypes were identified in 43 liver abscess patients. Only 2 genotypes were found in more than one ALA patient - 23 and 39, both found in 2 patients each.

(c) ICDDR,B hospital

In addition to those from Mirpur children, we also had 6 diarrhoeal/dysenteric samples from the ICDDR,B hospital, and all of these showed distinct genotypes (Appendix II).

4.4.2.2. Spatial distribution of Mirpur children's households infected by the two most common genotypes

The two most common genotypes using 6 loci were 66 and 51 (Appendix II). The location of households of the children infected with these genotypes in Mirpur was investigated.

Figure 4.3a shows the approximate household locations for the 10 children infected by genotype 66. These households were dispersed across the Mirpur study area, though half of the children seem to have lived close to each other. This probably suggests that this genotype has a widespread distribution in Mirpur. In contrast, the households of the 7 children infected by genotype 51 were clustered in a small area (Figure 4.3b), suggesting that they might have acquired the infection from a common source.

4.5. Genotype distribution among clinical groups

Using the GENEPOP software we analyzed the distribution of genotypes among clinical groups.

In Mirpur samples, using the PCR patterns at all 6 loci we found that the genotypes showed a statistically significant difference between the asymptomatic and diarrhoeal/dysenteric groups (Table 4.1). The individual loci contributing towards this difference were D-A and N-K2. In particular, two PCR patterns in locus D-A stand out– #2 found in 16 of 38 asymptomatic samples compared to only 4 of 44 diarrhoeal/dysenteric samples, and #3 found in 38 of 44 diarrhoeal/dysenteric samples compared to only 19 of 38 asymptomatic samples (Appendix II). Similarly, two PCR patterns in locus N-K2 – #8 found in 23 of 44 diarrhoeal/dysenteric samples compared to only 3 of 38 asymptomatic samples, and #9 found in 21 of 38 asymptomatic samples compared to only 9 of 44 diarrhoeal/dysenteric samples (Appendix II) – appear to be contributing the most to this result.

Across all Bangladeshi samples, using all the PCR patterns at 6 loci we found that the overall difference of genotypes among the 3 pairs of populations was statistically

significant (Table 4.1). Between the asymptomatic and diarrhoeal/dysenteric populations the loci contributing the most towards this were again D-A and N-K2 (Appendix II and Table 4.1). Between the diarrhoeal/dysenteric and liver abscess populations the loci contributing most towards this significant difference were A-L (PCR patterns 2 and 3), N-K2 (PCR patterns 4 and 9) and S-Q (PCR patterns 3, 4 and 5). Between the liver abscess and asymptomatic populations the loci contributing most towards this significant difference were also A-L (PCR patterns 1 and 2), N-K2 (PCR patterns 4, 6, 8 and 9) and S-Q (PCR patterns 3 and 4).

It was not proven that the amoeba was responsible for the diarrhoea/dysentery in that population. However if they were not, we would expect that the genotypes detected in asymptomatic and diarrhoeal/dysenteric groups would show no significance difference in the GENEPOP analysis. Since the genotypes showed significant difference in distribution, we believe that the amoeba had a significant impact on the outcome of the infection.

Table 4.1 Statistical (GENEPOP) analysis of genotype distribution among clinical groups.

Analysis	Population pair	No. of samples	Chi^2	Overall P values	Loci showing significant differences (P values)
Mirpur 6-locus	A & D	A = 38 D = 44	41.531	<0.0001	Locus D-A (0.0005) Locus N-K2 (0.0001)
Bangladesh 6-locus	A & D	A = 38 D = 50	Infinity	<0.0001	Locus D-A (0.0172) Locus N-K2 (<0.0001)
	D & L	D = 50 L = 43	Infinity	<0.0001	Locus A-L (0.0002) Locus N-K2 (0.0124) Locus S-Q (<0.0001)
	L & A	L = 43 A = 38	Infinity	<0.0001	Locus A-L (0.0224) Locus N-K2 (<0.0001) Locus S-Q (<0.0001)

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

a)

b)

re 4.3. Spatial distribution of *E. histolytica* infected children's households in Mirpur. ouseholds of children infected by a) genotype 66 and b) genotype 51 are identified.

4.6. Association between individual genotypes and clinical outcome of infections using 6 loci

Since the number of total genotypes is very high and the number of samples covered by a single genotype was low, we only tested two of the most common genotypes (66 and 51) to investigate whether they show an association with clinical outcome of infection (Table 4.2). Genotype 66 showed an association with diarrhoeal/dysenteric infection in both Mirpur and across all Bangladeshi samples (P = 0.0030 in Mirpur and overall P = 0.0003 in across all Bangladesh), while genotype 51 showed no significant association with type of infection in Mirpur or across all Bangladeshi samples. However we should be careful in interpreting our data, because although genotype 66 showed a statistically significant clinical association with diarrhoeal/dysenteric samples it was found only in 10 of the 44 (22.72%) such samples in Bangladesh.

4.7. The predicted total number of genotypes using 6 loci

The Chao estimator equation was used to calculate the predicted total number of genotypes (Colwell and Coddington, 1994) in the population. The equation used was

$S_1 = S_{obs} + (a^2/2b)$

Where, S_{obs} is the number of genotypes observed

a is the number of genotypes observed just once

b is the number of genotypes observed just twice

a) Mirpur samples

From Appendix II, we can see that, $S_{obs} = 48$, $a = 33$ and $b = 9$

So, the total predicted number of genotypes in Mirpur children is, $S_1 = 108.5$ more than twice the number detected in this study.

Table 4.2. Statistical (GENEPOP) analysis of the most common 6-locus individual genotypes among clinical groups.

Analysis	Genotype	Population pair with corresponding frequency of the genotype	Chi^2	P values
Mirpur 6-locus	66	A = 0/36 D = 10/38	8.82	0.0030
	51	A = 4/36 D = 3/38	0.01	0.9401
Bangladesh 6-locus	66	A = 0/36 D = 10/44	7.39	0.0066
		D = 10/44 L = 1/43	8.20	0.0042
		L = 1/43 A = 0/36	-	-
		Overall	16.11	0.0003
	51	A = 4/36 D = 3/44	0.08	0.7807
		D = 3/44 L = 0/43	1.33	0.2481
		L = 0/43 A = 4/36	2.99	0.0840
		Overall	4.67	0.0968

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess, - = no value for this could be calculated as the sample size was too small. P = <0.05 indicates that the association is statistically significant.

b) All Bangladeshi samples

From Appendix II, we can see that, $S_{obs} = 93$, a = 75 and b = 12

So, the total predicted number of genotypes in all Bangladesh is, $S_1 = 327.4$. This is probably an underestimate since sampling across Bangladesh was not comprehensive.

4.8. The association among genotypes by eBURST

4.8.1. eBURST with genotypes from Mirpur children

eBURST clustered 29 of 48 genotypes into 3 minor groups (containing 2 to 3 genotypes in each group) and a major group at the centre of the diagram linking 22 genotypes. The remaining 19 genotypes were represented as singletons (Figure 4.4) as they did not have any SLVs or DLVs. These singletons can be divided into three categories -10 singletons are unique to the asymptomatic children, 7 singletons are unique to the diarrhoeal children, and the remaining 2 singletons are shared between the two clinical types. Those clusters that seemed to contain genotypes of a particular clinical type are circled in the diagram (Figure 4.4).

One of the three smaller groups (pink circle in Figure 4.4) contained 2 genotypes and comprised asymptomatic children only. The remaining two small groups contained a mixture of clinical samples (Appendix II). In the large group, there were three clinical clusters (inside the blue circles) – two of these contained only diarrhoeal/dysenteric samples while in the remaining one, comprising 17 samples, all but one of the samples were from children with diarrhoea/dysentery. This latter cluster contains genotype 66 shown above to be significantly linked to diarrhoea/dysentery.

4.8.2. eBURST with all samples from Bangladesh

The main purpose of investigating an eBURST association between genotypes and clinical outcomes was to see whether genotypes from a particular sample type were related to each other. For example, we wanted to see whether genotypes from asymptomatic children or from symptomatic patients tend to form mutually exclusive clusters in the eBURST diagram. Even though the liver abscess patients were from a wide geographic area we decided to investigate the relationships among genotypes across all samples from Bangladesh.

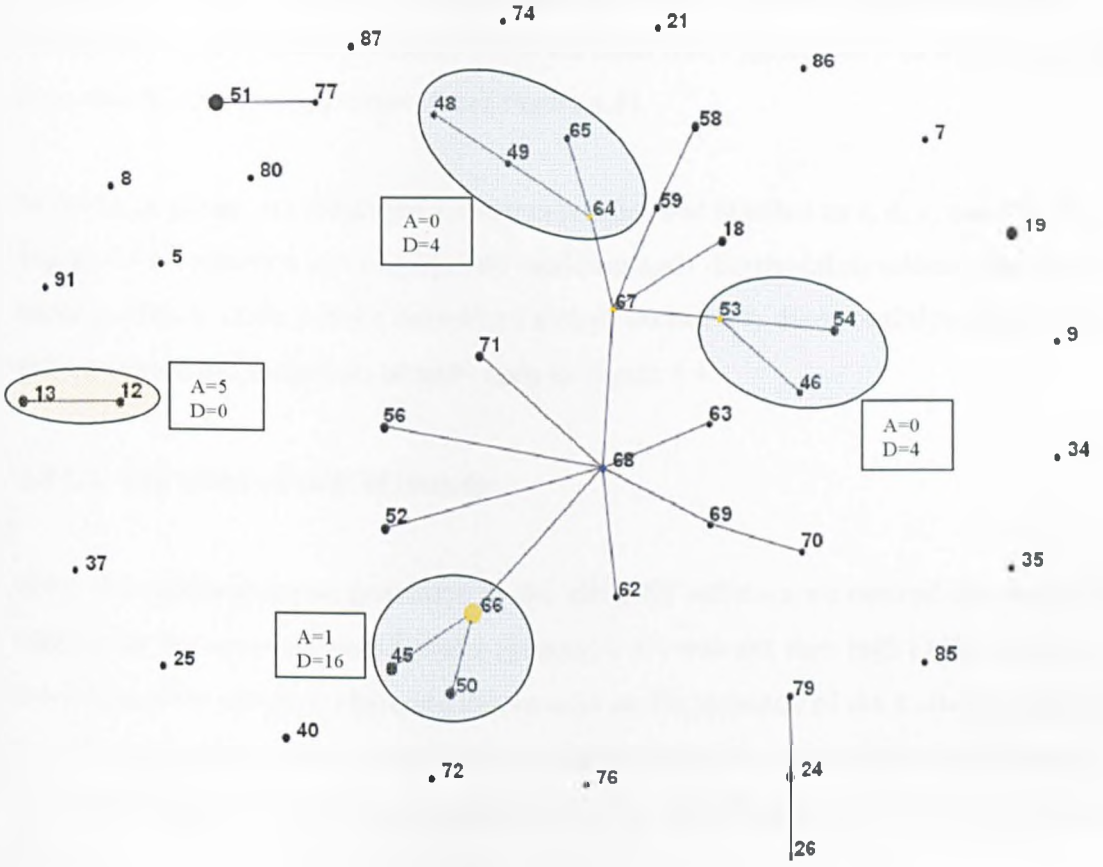


Figure 4.4. eBURST relationships among the 6-locus genotypes in Mirpur children. The number of children found in the respective clusters is given in the square boxes: A = asymptomatic and D = diarrhoeal/dysenteric.

eBURST clustered 59 of 93 6 locus genotypes into 7 minor groups containing 2 to 6 genotypes in each (shown along the periphery, Figure 4.5) and a major group with 35 genotypes (shown at the centre of the diagram, Figure 4.5) with the auto-assigned founder genotype 67, which had 8 SLVs and 13 DLVs. The remaining 34 genotypes had no SLVs or DLVs and are represented as singletons surrounding the major group.

Five of the 7 small groups comprised a mixture of genotypes of all 3 sample types. Of the 2 other small groups – a and b (circled in the Figure 4.5) - one had 4 genotypes seen exclusively in liver abscess strains, while the other had 4 genotypes 3 of which were from liver abscess strains (Appendix II and Figure 4.5).

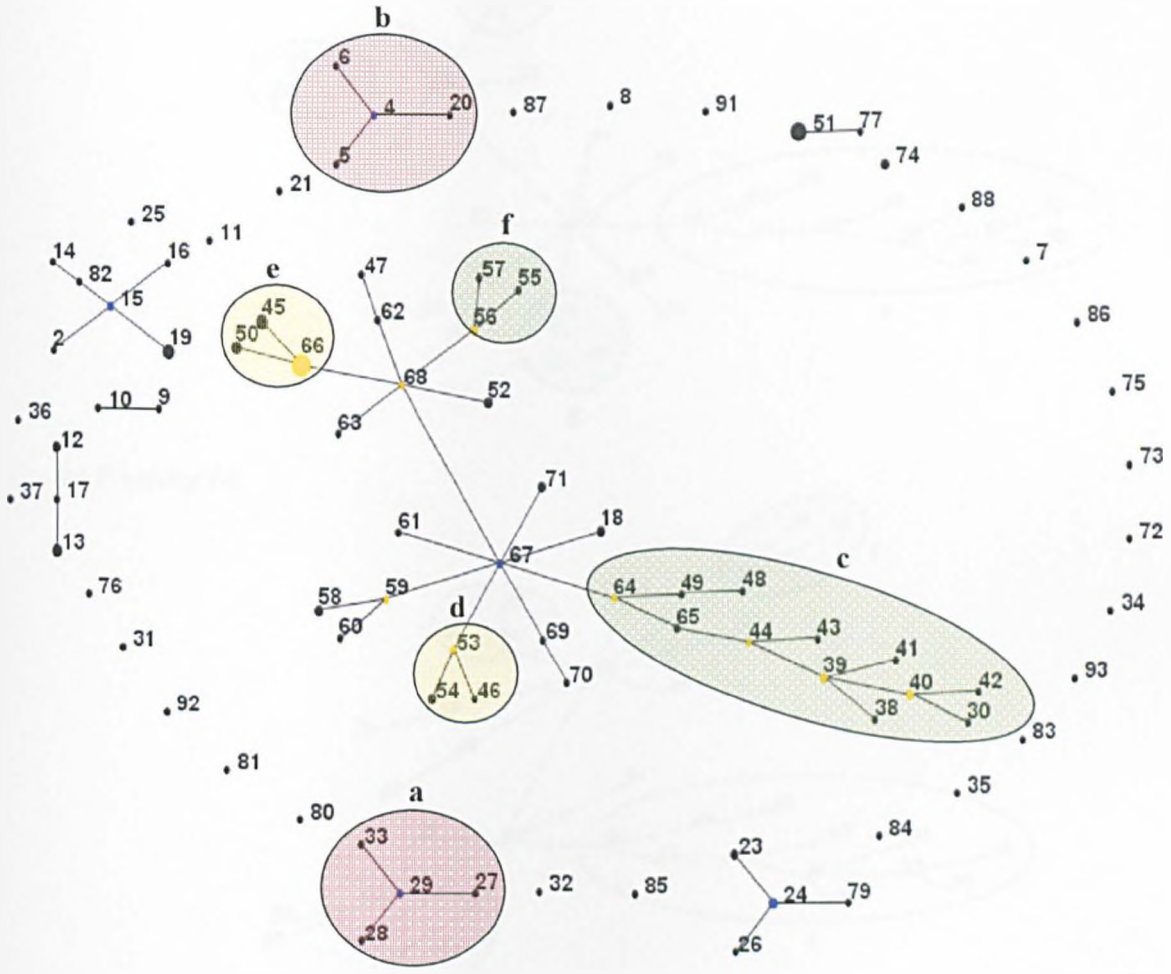
In the large group, we identified 4 clusters (circled and labelled as c, d, e, and f in the Figure 4.5). Clusters c and f comprised predominantly diarrhoeal/dysenteric and liver abscess strains, while d and e comprised almost exclusively diarrhoeal/dysenteric strains and corresponded to clusters already seen in Figure 4.4.

4.8.2.1. The effect of shift of founder

From the statistical report generated by the eBURST software we noticed that the bootstrap support for the auto-assigned founder (genotype 67) was not very high (31%), so, we investigated the effect of changing the founder on the structure of the 4 clusters (c, d, e and f) in the large group. There were 5 other candidate founder genotypes that had similar bootstrap supports (15-38%): genotypes 68 (38%), 64 (25%), 66 (16%), 53 (15%) and 56 (20%).

Figure 4.6 (a-f) shows the effect of shifting the founder on those clusters. The change affected only the clusters in the major group since all the new founders were also located in this group.

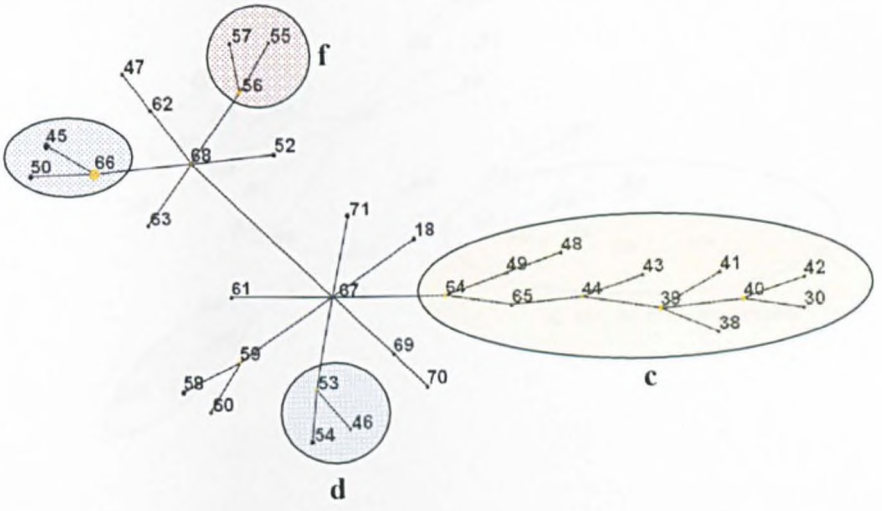
If the new founder genotype was not part of one of the 4 clusters, then the shift had no effect on these (Figures 4.6a and 4.6b). However, if the founder was one of the genotypes in a cluster, then that cluster was affected (Figures 4.6c, d, e and f). For example, when the founder was genotype 64, 3 of the 12 genotypes in cluster c including 64 moved to another branch (Figure 4.6c), and when the founder was genotype 66, the cluster e (Figure 4.6d) no longer existed. We conclude that the 4 clusters are relatively stable and independent of choice of founder.



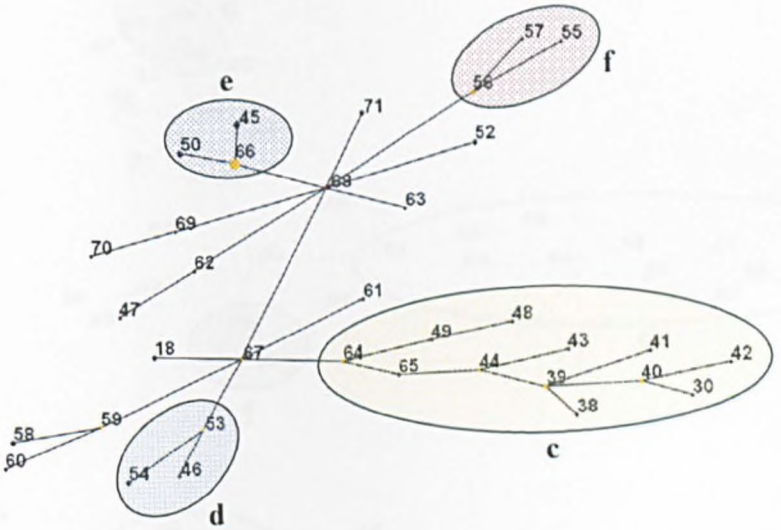
Clinical type	Cluster						
	a	b	c	d	e	f	Total
A	0	0	2	0	1	0	3
D	0	1	5	4	16	2	28
L	4	3	7	0	1	2	17
Total	4	4	14	4	18	4	48

Figure 4.5. eBURST diagram using all Bangladeshi *E. histolytica* strains at 6 loci. All the clusters with genotypes of clinical relevance are circled (a to f). Table shows the clinical distribution of samples in these clusters.

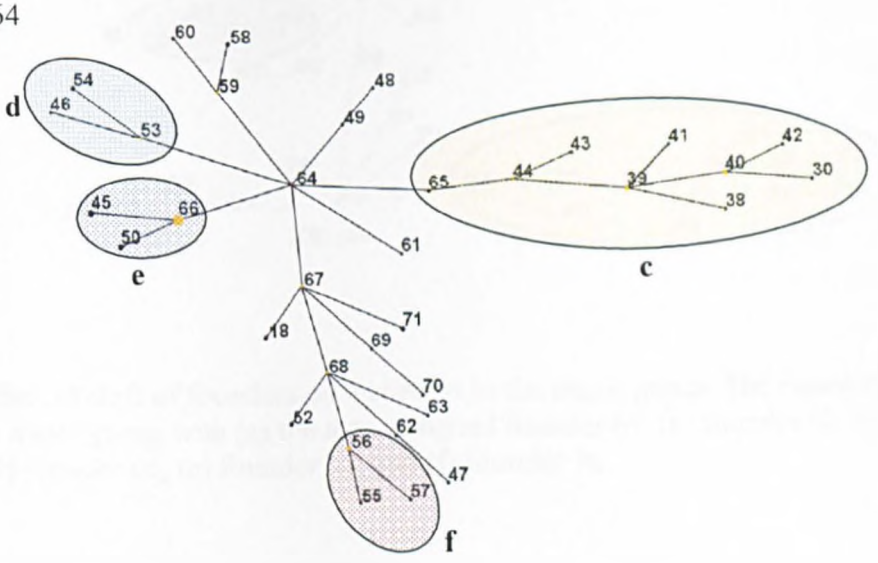
a) Founder 67



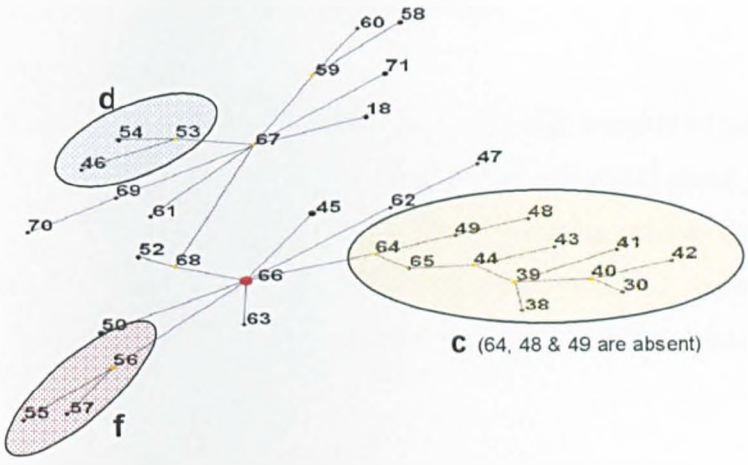
b) Founder 68



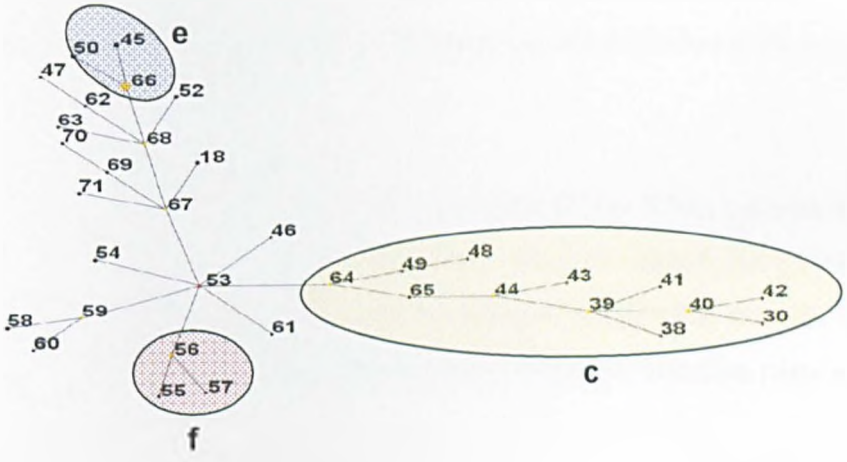
c) Founder 64



d) Founder 66



e) Founder 53



f) Founder 56

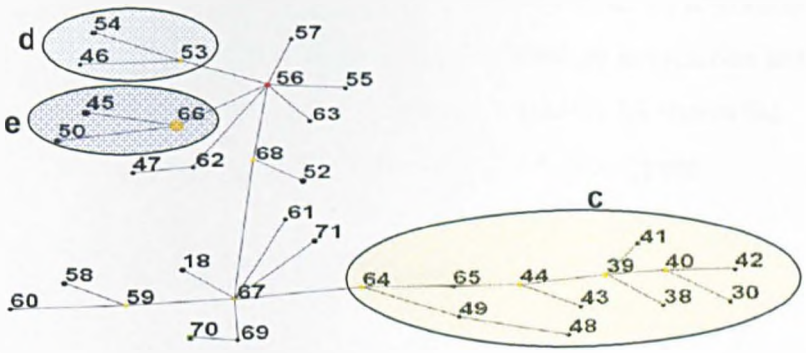


Figure 4.6. Effect of shift of founders on 4 clusters in the major group. The structure of clusters in the major group with (a) the auto-assigned founder 67, (b) founder 68, (c) founder 64, (d) founder 66, (e) founder 53, and (f) founder 56.

4.9. Genotypes using various combinations of loci

Because by using all 6 loci for genotyping we obtained a very high number of genotypes, 93 out of 119 individuals, and no obvious strong association of individual genotypes with clinical outcomes, we decided to investigate this association using the various combinations of 5, 4 and 3 loci to see if a subset of loci showed a stronger link. Table 4.3 summarizes the number of genotypes obtained using various combinations of loci with their clinical distributions.

From Table 4.3 we can see that the ranges of genotype numbers using various combinations of loci are 74 to 91, 54 to 72, and 30 to 67 for 5, 4 and 3 loci, respectively, across all samples. Among the clinical groups, the liver abscess samples always produced the largest number of genotypes.

The total number of genotypes obtained using any combination of 5 or 4 loci was again very high. For example, the minimum number of genotypes obtained using 5 loci (ADRSQ) was 68 and the minimum number of genotypes obtained using 4 loci (DARS) was 48. We did not find any association between genotype and clinical outcome of infection (data not shown).

However, the set of 3 loci – D-A, A-L and R-R (DAR) - which produced the minimum number of genotypes with 3 loci (30, Table 4.3) appeared to show an association with the clinical outcome of disease so we investigated this further. Appendix III shows the genotypes obtained for all samples from Bangladesh using DAR genotyping.

4.9.1. DAR genotype distribution among clinical groups

As for the 6-locus results, in Mirpur samples using all the PCR patterns at DAR loci we also found a statistically significant difference between asymptomatic and diarrhoeal/dysenteric groups (Table 4.4) and again the individual locus that contributed significantly towards this result was D-A.

Table 4.3. Summary of genotype numbers using various combinations of loci.

No. of loci used	Genotyping name	No. of genotypes			Total genotypes (n=123)
		Asymptomatic (n=36)	Diarrhoeal (n=44)	Liver abscess (n=43)	
6	ADNRSQ	26	33	41	93
5	ADNRQ	23	29	38	79
	ADNRS	26	31	36	80
	ADNSQ	25	29	39	85
	ADRSQ	24	25	40	74
	ANRSQ	26	33	41	91
	DNRSQ	25	32	40	87
4	ARSQ	20	25	38	62
	DANQ	21	26	34	65
	DANR	19	25	31	58
	DANS	24	23	32	64
	DARQ	19	19	32	56
	DARS	23	22	32	54
	DASQ	21	22	35	63
	DNRQ	22	28	36	72
	DNRS	25	29	33	71
	DRSQ	22	24	39	67
3	ANQ	18	23	30	52
	ANR	15	19	23	41
	ANS	20	20	26	50
	ARQ	24	17	27	42
	ARS	16	18	26	39
	ASQ	16	18	31	46
	DAN	17	17	24	39
	DAQ	14	16	25	41
	DAR	14	15	19	30
	DAS	17	16	26	35
	DNQ	18	24	29	54
	DNR	18	22	26	48
	DNS	21	21	27	53
	DRQ	17	17	29	45
	DRS	20	19	26	41
	DSQ	19	21	30	52
	NRQ	18	26	33	59
	NRS	20	26	30	56
	NSQ	21	26	34	67
	RSQ	18	24	36	54

n = number of total individuals; abbreviated form of locus names have been used in the naming of genotyping: A = A-L, D = D-A, N = N-K2, R = R-R, S = S-D, and Q = S-Q.

Using DAR across all Bangladeshi samples we found an overall statistically significant difference between the diarrhoeal/dysenteric and liver abscess populations only (Table 4.4) and the only locus contributing significantly towards this result was A-L. In contrast using all 6 loci we previously found significant differences among all 3 pairs of clinical populations. Other loci – N-K2 and S-Q - contributed significantly towards these differences. Although overall the differences in DAR analysis were not statistically

significant between the asymptomatic and diarrhoeal/dysenteric populations and between the asymptomatic and liver abscess populations, locus D-A and locus A-L did show a statistically significant difference in these 2 pairs of populations, respectively.

Table 4.4. Statistical analysis of DAR genotype distribution among clinical groups.

Analysis	Population pair	No. of samples	Chi^2	Overall P values	Loci showing significant differences (P values)
Mirpur DAR	A & D	A = 37 D = 42	18.175	0.0058	Locus D-A (0.0011)
Bangladesh DAR	A & D	A = 37 D = 48	10.084	0.1212	Locus D-A (0.0200)
	D & L	D = 48 L = 43	24.450	0.0004	Locus A-L (0.0004)
	L & A	L = 43 A = 37	10.522	0.1043	Locus A-L (0.0264)

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

4.9.2. Association between individual DAR genotypes and clinical outcome of infections

We tested a few of the most common DAR genotypes (18, 6 and 17) to investigate whether they showed an association with clinical outcome of infection (Table 4.5). Genotypes 18 and 6 showed an association with diarrhoeal/dysenteric and asymptomatic infections, respectively, both in Mirpur and across all Bangladeshi samples. However genotype 17 (the second most common in Mirpur with genotype 6 found in 12 children each) showed no significant association with the type of infection. Although genotypes 18 and 6 showed statistically significant clinical association with diarrhoeal/dysenteric and asymptomatic infections they covered only 50% and 28.6% of the samples in the corresponding clinical populations, respectively.

Table 4.5. Statistical analysis of the most common individual DAR genotypes among clinical groups.

Analysis	Genotype	Population pair with corresponding frequency of the genotype	Chi^2	P values
Mirpur DAR	18	A = 6/35 D = 22/38	12.80	0.0003
	6	A = 10/35 D = 2/38	7.21	0.0073
	17	A = 5/35 D = 7/38	0.23	0.6339
Bangladesh DAR	18	A = 6/35 D = 22/44	9.20	0.0024
		D = 22/44 L = 5/43	14.96	0.0001
		L = 5/43 A = 6/35	0.14	0.7121
		Overall	18.67	<0.0001
	6	A = 10/35 D = 4/44	5.07	0.0243
		D = 4/44 L = 4/43	0.11	0.7362
		L = 4/43 A = 10/35	4.86	0.0274
		Overall	7.45	0.0241

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

4.9.3. The predicted total number of DAR genotypes

a) In Mirpur samples:

From Appendix III, we can see that, $S_{obs} = 20$, a = 10 and b = 5

So, the total number of predicted DAR genotypes in Mirpur children is, $S_1 = 30$

b) Across all Bangladeshi samples:

From Appendix III, we can see that, $S_{obs} = 30$, a = 10 and b = 9

So, the total number of predicted DAR genotypes in Bangladesh is, $S_1 = 35.6$ indicating that most of the variants had been detected. However sampling is not comprehensive.

4.10. Association between genotypes using a selection of 5 loci, 4 loci and 3 loci genotyping data

Although we did not see an obvious and strong association between an individual genotype and clinical outcome of infection using any combination of 5, 4 and 3 loci, except for DAR (where a few individual genotypes showed a good association), we were, however, interested to investigate the relationship among genotypes for a selection of these combinations using the eBURST program. We investigated this for the maximum and minimum numbers of genotypes obtained (Table 4.3): for 5 loci – 91(ANRSQ) and 74 (ADRSQ), for 4 loci - 72 (DNRQ) and 54 (DARS), and for 3 loci - 67 (NSQ) and 30 genotypes (DAR) and the results are provided in Appendix IV.

In summary, eBURST showed better association among genotypes when a greater number of loci was used. As we decreased the number of loci, it became easier for the eBURST to find genotypes that differed by one locus and as a result it was linking different types of clinical samples into the same clusters.

4.11. *E. histolytica* follow-up samples

Follow-up samples were available from 23 children (Table 4.6). We were interested in investigating the stability of genotype patterns in these infections.

A maximum number of 7 follow-up samples were available from one child (3402) over a period of 11 months and all showed the same genotype, except for one sample (MS33) which seemed to have a mixture of two genotypes. However the new genotype was not found in the subsequent follow-up sample (MS34) suggesting that it was a transient infection. Six and 5 follow-up samples were available from two children – 0015 and 1049, respectively. Both these children carried the same genotype throughout the follow-up

period. Two and 3 follow-up samples were available from 13 and 6 children, respectively (Table 4.6).

From Figure 4.10 we can see that both 6-locus and DAR genotypes changed in 7 children. In 13 children all the corresponding follow-up samples from the same child showed same genotypes. In 3 children DAR genotypes remained the same while the 6-locus genotypes changed.

4.11.1. Number of loci changed during the new infection

We observed a total of 12 6-locus genotype changes in the follow-up samples of 10 children. Comparing the patterns between two consecutive follow-up samples, we observed a total of 37 PCR pattern changes involving all 6 loci. Only one pattern changed at locus A-L, which is the least polymorphic of the 6 loci. The maximum number of pattern changes (9) occurred at locus S-Q, one of the most polymorphic loci. On average 3.08 loci changed when a follow-up genotype was different.

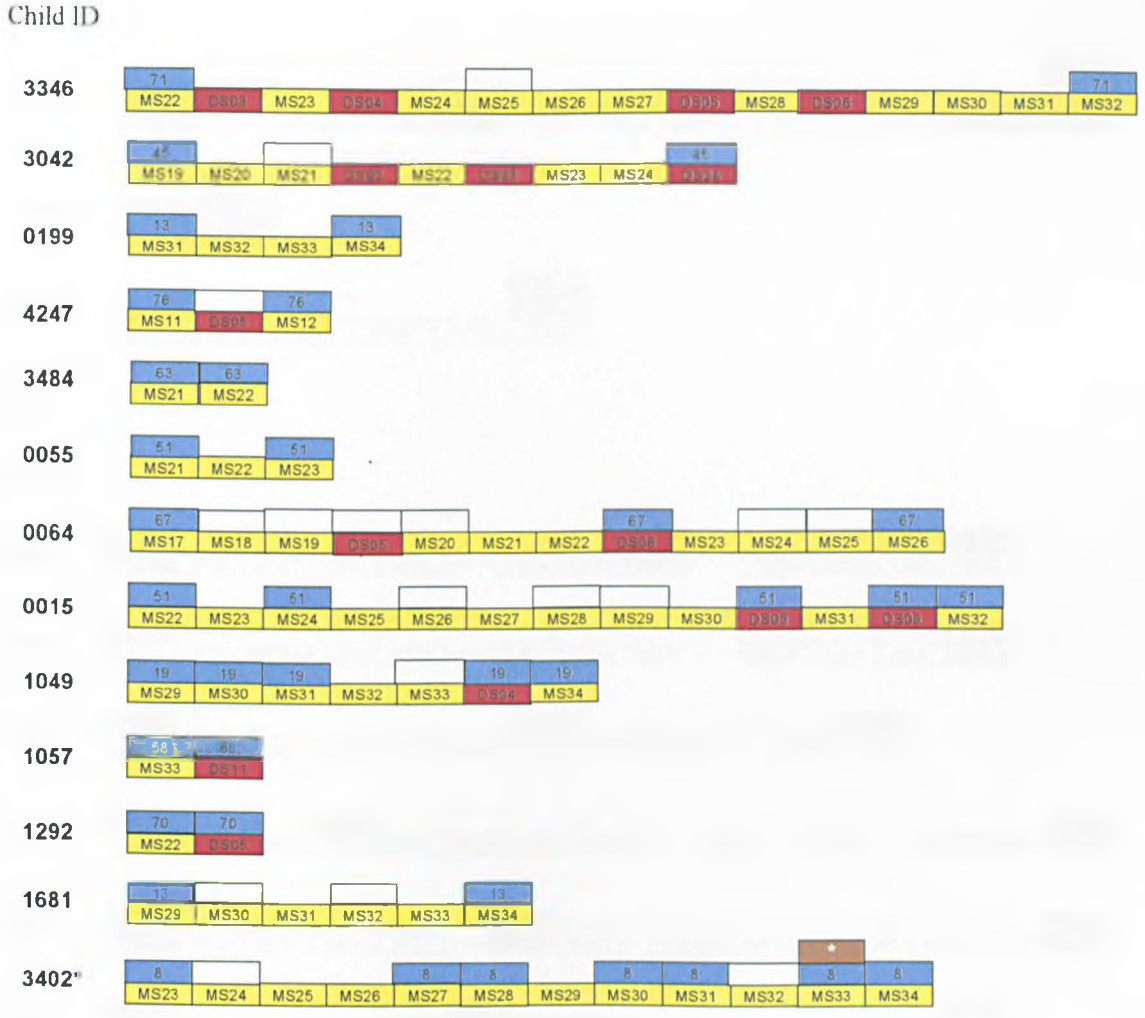
In most cases we ran all the follow-up samples from a particular child next to each other in the gel to detect the differences. In two children (1186 and 3679) we noticed that the PCR patterns of the follow-up samples changed at only one locus each. However, in both children the changes were easy to detect. In child 1186, the changed N-K2 patterns were 1 (500 bp) and 8 (630 bp) and in child 3679 the changed R-R patterns were 1 (480 bp) and 4 (590 bp). In conclusion even when changes occurred in only 1 or 2 PCR patterns in a few follow-up samples, they were real. In most cases the PCR pattern changes were extensive and loci with more polymorphism seemed to be affected more frequently.

Table 4.6. Follow-up samples in *E. histolytica* infected children.

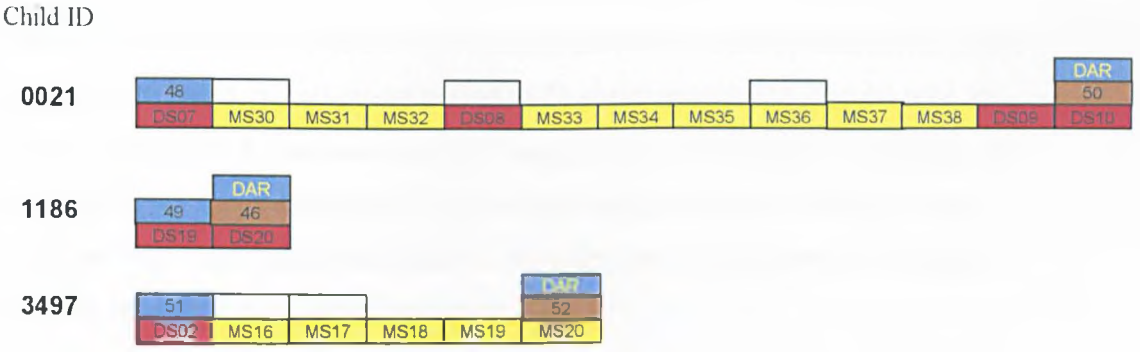
ID	Clinical status	D-A	A-L	R-R	N-K2	S-D	S-Q	Genotype
0015-DS04	D	3	3	3	9	3	5	51
0015-DS05	D	3	3	3	9	3	5	51
0015-MS22	A	3	3	3	9	3	5	51
0015-MS24	A	3	3	3	9	3	5	51
0015-MS27	A	3	3	3	9	3	5	51
0015-MS32	A	3	3	3	9	3	5	51
0021-DS07	D	3	3	3	6	3	3	48
0021-DS10	D	3	3	3	8	4	3	50
0027-DS10	D	3	3	4	8	4	3	66
0027-DS13	D	3	3	4	8	4	3	66
0027-DS17	D	3	3	3	9	3	5	51
0055-MS21	A	3	3	3	9	3	5	51
0055-MS23	A	3	3	3	9	3	5	51
0064-DS06	D	3	3	4	9	3	3	67
0064-MS17	A	3	3	4	9	3	3	67
0064-MS26	A	3	3	4	9	3	3	67
0199-MS31	A	2	3	4	6	6	3	13
0199-MS34	A	2	3	4	6	6	3	13
0291-MS32	A	3	3	4	3	2	3	58
0291-MS42	A	3	3	5	1	2	2	72
1036-DS01	D	3	3	4	4	4	3	62
1036-DS06	D	3	4	3	9	3	5	77
1036-MS59	A	2	4	5	9	3	2	24
1049-DS04	D	2	3	4	9	4	8	19
1049-MS29	A	2	3	4	9	4	8	19
1049-MS30	A	2	3	4	9	4	8	19
1049-MS31	A	2	3	4	9	4	8	19
1049-MS34	A	2	3	4	9	4	8	19
1057-DS11	D	3	3	4	3	2	3	58
1057-MS33	A	3	3	4	3	2	3	58
1186-DS19	D	3	3	3	8	3	3	49
1186-DS20	D	3	3	3	1	3	3	46
1292-DS05	D	3	3	4	9	6	4	70
1292-MS22	A	3	3	4	9	6	4	70
1325-MS22	A	3	3	3	9	4	3	52
1325-MS28	A	2	3	4	9	4	8	19
1325-MS33	A	2	3	4	9	4	8	19
1681-MS29	A	2	3	4	6	6	3	13
1681-MS34	A	2	3	4	6	6	3	13
3042-DS04	D	3	3	1	8	4	3	45
3042-MS19	A	3	3	1	8	4	3	45
3346-MS22	A	3	3	4	9	10	3	71
3346-MS32	A	3	3	4	9	10	3	71
3402-MS23	A	2	3	3	9	6	3	8
3402-MS27	A	2	3	3	9	6	3	8
3402-MS28	A	2	3	3	9	6	3	8
3402-MS30	A	2	3	3	9	6	3	8
3402-MS31	A	2	3	3	9	6	3	8
3402-MS33*	A	2	3	3	9	6	3	8
3402-MS34	A	2	3	3	9	6	3	8
3484-MS21	A	3	3	4	6	4	3	63
3484-MS22	A	3	3	4	6	4	3	63
3497-DS02	D	3	3	3	9	3	5	51
3497-MS20	A	3	3	3	9	4	3	52
3570-DS04	D	2	3	1	8	3	5	5
3570-DS14	D	4	3	3	8	4	2	86
3570-DS28	D	3	3	4	1	3	4	54
3645-DS14	D	4	2	1	4	5	4	80
3645-MS31	A	3	2	1	6	4	3	34
3679-DS15	D	3	3	4	8	4	3	66
3679-MS21	A	3	3	1	8	4	3	45
3679-MS24	A	3	3	1	8	4	3	45
4247-DS05	D	3	4	3	5	4	5	76
4247-MS11	A	3	4	3	5	4	5	76
4247-MS12	A	3	4	3	5	4	5	76

*This sample showed a transient infection with another genotype of *E. histolytica* that was not found in the next follow-up sample (MS34).

(a)


Child ID
3346
3042
0199
4247
3484
0055
0064
0015
1049
1057
1292
1681
3402*

(b)



(c)

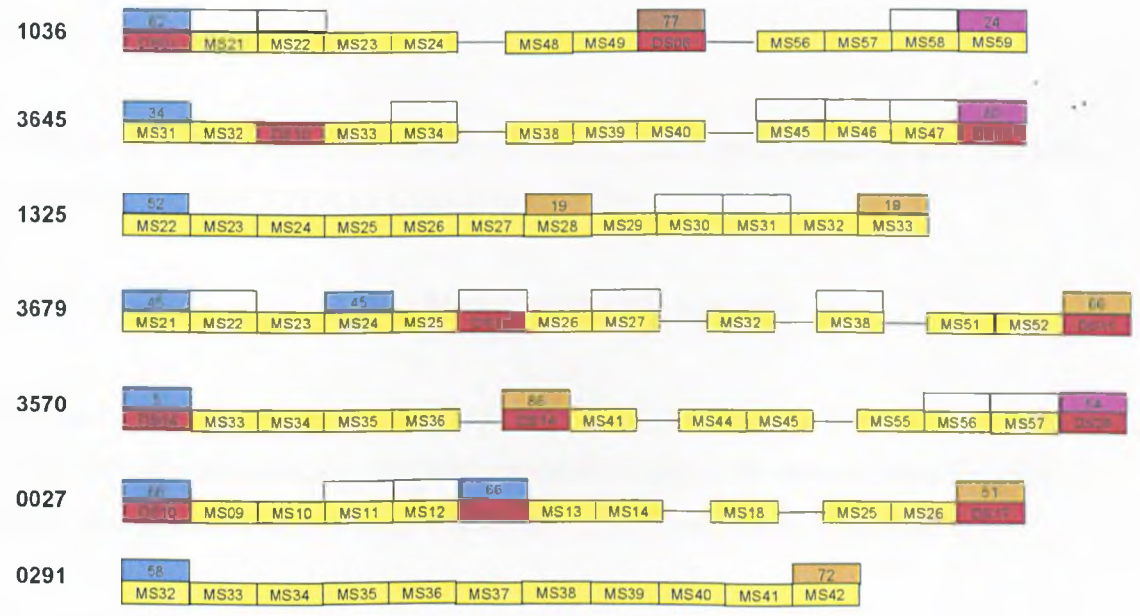


Figure 4.10. Genotypes identified in the follow-up samples from children infected with *E. histolytica*. Samples were collected monthly except when diarrhoea was present. Upper coloured bars represent the samples tested, white bars indicate that the sample was positive for *E. histolytica* but genotyping was not done, and changes in genotype are indicated by different colours and corresponding 6-locus genotype numbers. Lower bars represent asymptomatic (yellow) or diarrhoeal (red) status of the child. (a) Genotypes remained the same, (b) DAR genotypes remained the same but 6-locus genotypes changed from the earlier one(s), and (c) both 6-locus and DAR genotypes changed. *This sample was found to be transiently infected by a second strain of *E. histolytica* (child 3402).

4.11.2. Rate of re-infection

We tried to calculate the rate of re-infection from the available data on 23 follow-up children. The total follow-up period in 23 children was 223 months with an average period of 9.7 months per child (ranged from 1-month to 38-months). In this period we noticed a total of 12 genotype changes (in 10 children; 2 being infected with 2 new genotypes each) or re-infections, which gives an average of 18.6 months for a new infection to take place.

4.12. Results for *E. dispar* samples

4.12.1. PCR amplification

PCR amplification was successful for all samples using the *E. dispar* species specific primer pairs (Table 3.2) at all 6 loci (Figure 4.11)

4.12.2. Genotypes based on six PCR amplification patterns

Unlike *E. histolytica*, many *E. dispar* isolates produced multiple bands at STR locus N-K. Appendix V shows all the genotypes observed. If follow-up samples from the same individual shared the same pattern only one was included in the Appendix.

A total of 32 genotypes were observed among the 68 *E. dispar* isolates. 37 asymptomatic children produced 29 genotypes. Children 1805 and 3585 showed two genotypes each. Seven samples that were isolated during an episode of diarrhoea/dysentery (from 6 children including one who gave 2 diarrhoeal/dysenteric samples) showed 6 different patterns. Only one of these 7 samples had a concomitant detectable bacterial infection (data not shown). The most common genotypes were 4, 22 and 24, found in 3 children each. In addition, another six genotypes were found in 2 children each (Appendix V).

4.12.3. DAR genotypes with *E. dispar* samples

Three locus DAR genotyping produced 16 genotypes across all *E. dispar* strains. The diarrhoeal/dysenteric samples showed 4 genotypes (data not shown).

4.13. The predicted total number of genotypes

Using the Chao estimator the total number of predicted *E. dispar* genotypes in Mirpur children using 6 loci was about 85, and using DAR was 48.

4.14. Follow-up samples in *E. dispar*

Multiple samples were received from 15 children including 1 present in both the asymptomatic and diarrhoeal/dysenteric groups. Only 2 children showed different *E. dispar* genotypes in follow-up samples (Table 4.7 and Figure 4.12).

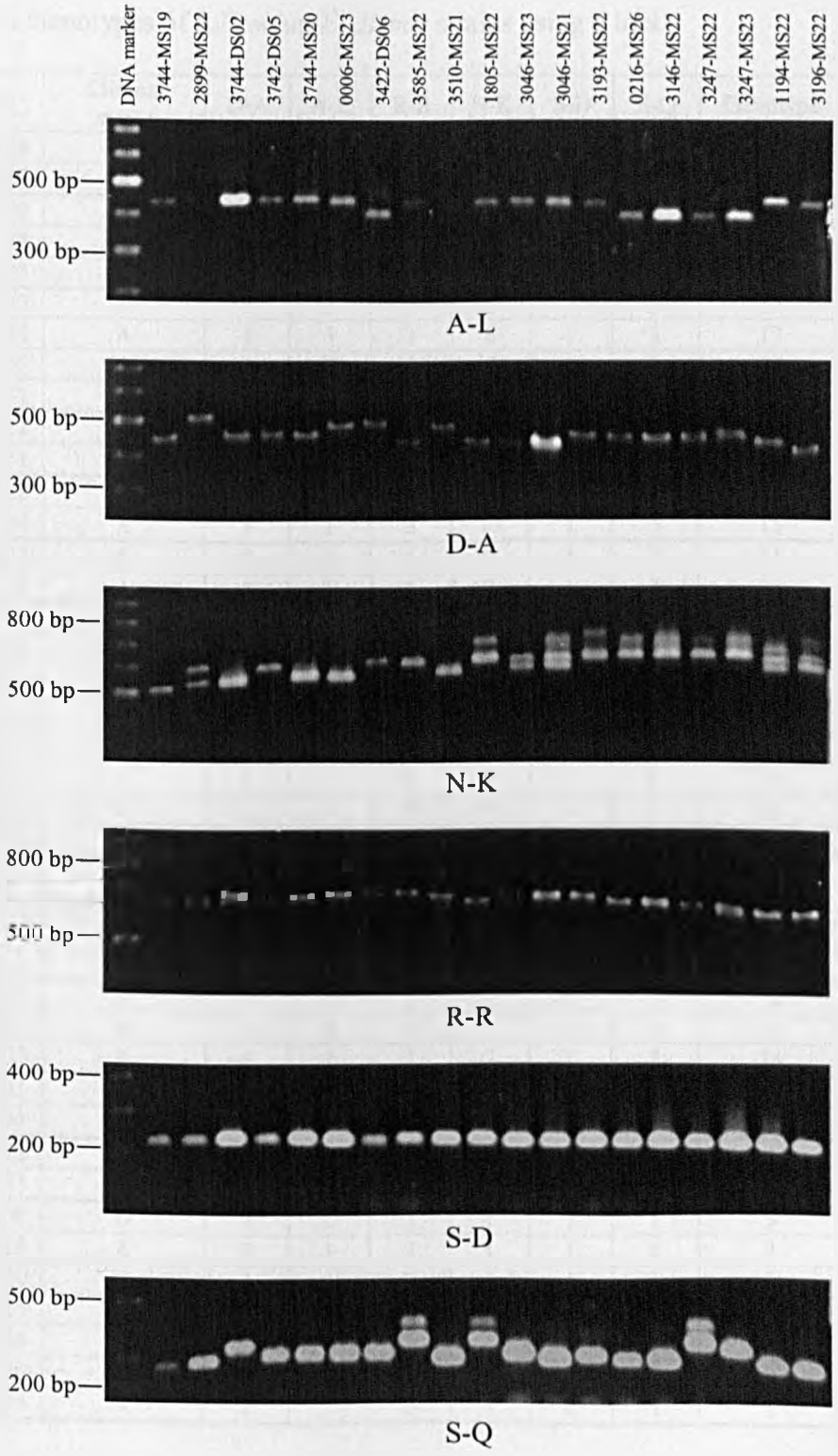


Figure 4.11. Gel electrophoresis patterns for a selection of *E. dispar* samples from Mirpur at the selected STR loci.

Table 4.7. Genotypes of follow-up *E. dispar* strains using 6 loci.

Child ID	Clinical status	D-A	A-L	R-R	N-K	S-D	S-Q	Genotype
1008-MS24	A	3	1	3	1	1	1	12
1008-MS25	A	3	1	3	1	1	1	12
1071-MS30	A	3	1	3	9	1	6	13
1071-MS32	A	3	1	3	9	1	6	13
1071-MS33	A	3	1	3	9	1	6	13
1194-MS22	A	4	3	3	8	1	3	17
1194-MS24	A	4	3	3	8	1	3	17
1194-MS25	A	4	3	3	8	1	3	17
1194-MS31	A	4	3	3	8	1	3	17
1627-MS22	A	5	2	3	10	1	1	26
1627-MS31	A	5	2	3	10	1	1	26
1805-MS22	A	2	1	1	5	1	6	6
1805-MS34	A	3	1	3	13	1	5	15
2899-MS22	A	8	1	1	2	1	2	31
2899-MS34	A	8	1	1	2	1	2	31
3046-MS21	A	2	1	3	6	1	1	7
3046-MS23	A	2	1	3	6	1	1	7
3146-MS22	A	5	2	3	5	1	3	22
3146-MS33	A	5	2	3	5	1	3	22
3146-MS35	A	5	2	3	5	1	3	22
3182-MS31	A	5	1	3	9	1	1	20
3182-MS32	A	5	1	3	9	1	1	20
3182-MS33	A	5	1	3	9	1	1	20
3193-MS22	A	5	1	3	7	1	1	19
3193-MS33	A	5	1	3	7	1	1	19
3247-MS22	A	5	2	3	5	1	3	22
3247-MS23	A	5	2	3	5	1	3	22
3499-MS26	A	5	2	3	7	1	3	24
3499-MS31	A	5	2	3	7	1	3	24
3499-MS33	A	5	2	3	7	1	3	24
3570-DS03	D	1	1	1	9	1	7	1
3570-DS04	D	1	1	1	9	1	7	1
3585-MS22	A	2	1	3	3	1	1	4
3585-MS33	A	5	2	3	7	1	3	24
3744-DS02	D	2	1	3	1	1	1	3
3744-MS19	A	2	1	3	1	1	1	3
3744-MS20	A	2	1	3	1	1	1	3
3744-MS25	A	2	1	3	1	1	1	3
3744-MS26	A	2	1	3	1	1	1	3
3744-MS30	A	2	1	3	1	1	1	3
3744-MS33	A	2	1	3	1	1	1	3

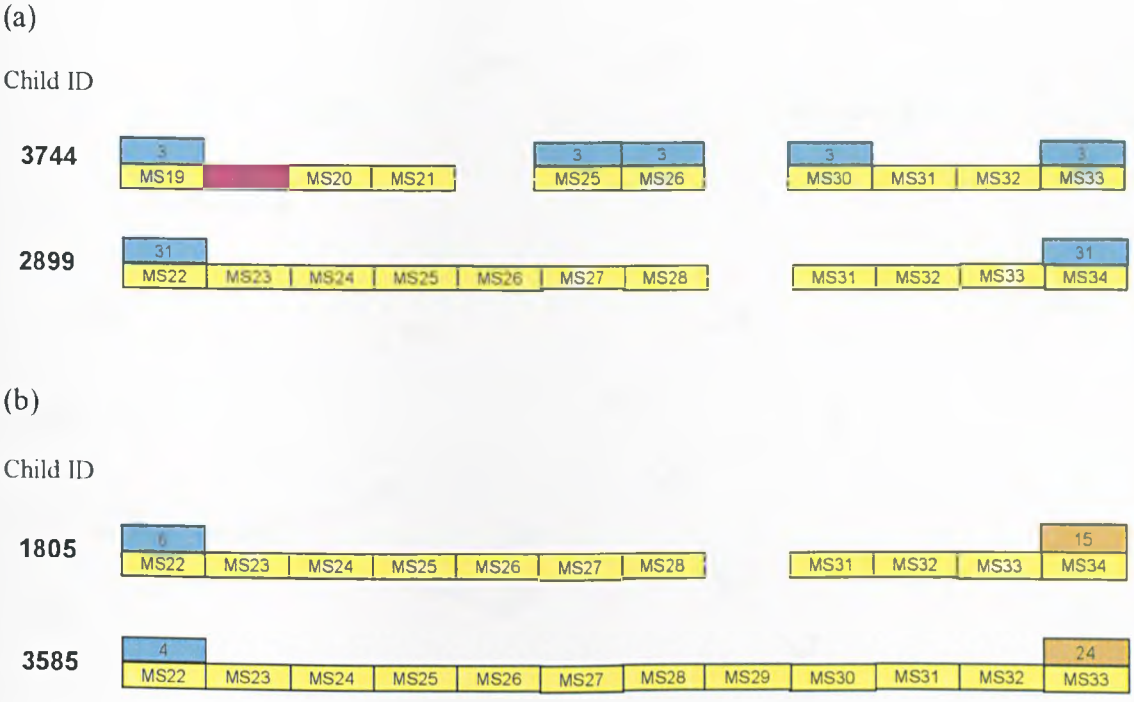


Figure 4.12. Genotypes identified in the follow-up samples from children infected with *E. dispar*. Samples were collected monthly except when diarrhoea was present. Upper coloured bars represent the genotypes of samples tested, and changes in genotype are indicated by different colours and corresponding 6-locus genotype numbers. Lower bars represent asymptomatic (yellow) and diarrhoeal (red) status of the child. (a) Genotypes remained the same (only two children are shown as representative of 13 children in this group) and (b) both 6-locus and DAR genotypes changed in 2 children.

4.15. eBURST analysis with *E. dispar* genotypes

The relationship between *E. dispar* genotypes was investigated using eBURST. eBURST detected 4 groups: 2 contained only two genotypes each, a 3rd group contained 5 genotypes and the 4th (major) group contained 8 genotypes (Figure 4.13). In the major group, the founder genotype is 23, which has 3 SLVs (genotypes 19, 24 and 26) and 4 DLVs (genotypes 20, 21, 22 and 25). Interestingly, no diarrhoeal samples were found in this group. In the 3rd group, the founder genotype is 3, which has 3 SLVs (genotypes 4, 7and 12) and one DLV (genotype 5). It is noticeable that 3 out of 6 diarrhoeal samples were found in this small group. Fifteen genotypes were identified by eBURST as singletons.

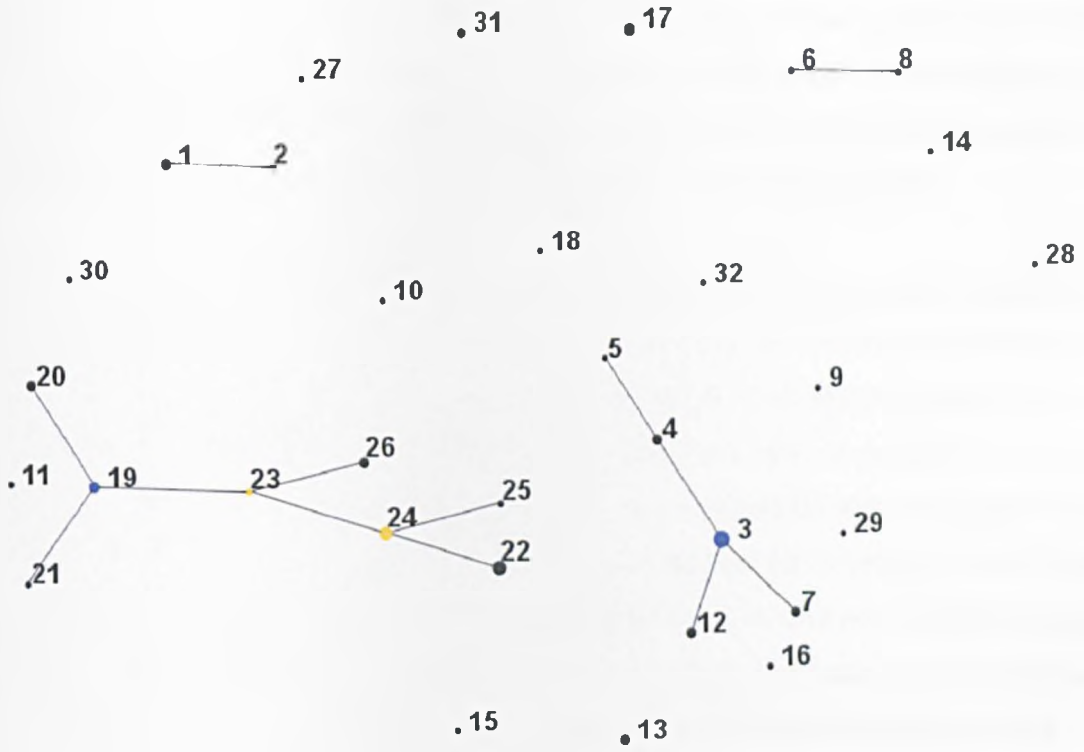


Figure 4.13. eBURST with Bangladeshi *E. dispar* isolates.

4.16. Discussion

Our main objectives in this study were to evaluate our newly developed and species-specific tRNA-linked repetitive DNA markers (discussed in the previous Chapter) and to genotype clinical strains of *E. histolytica* from Bangladesh. The reason for doing this genotyping was to see if genotypes correlate with the outcome of infection. We were successful in genotyping all available Bangladeshi strains of *E. histolytica*. PCR amplification for genotyping was successful using culture, stool and liver abscess DNA samples. Most importantly we found evidence of an association between genotypes and the clinical outcome of infection using both 6-locus and 3-locus (DAR) genotyping. eBURST analysis using six loci showed that many of the genotypes detected predominantly in

diarrhoeal/dysenteric and liver abscess samples seemed to be related to each other as they were found to cluster in the eBURST diagram. We observed some clusters in the eBURST diagram that had clinical relevance. Most of the time symptomatic strains clustered in the same group. However, these clusters represented only a minority of the total samples in three populations. As a result the predictive value of these clusters is low.

However we have some limitations in this study: (a) All our asymptomatic samples and most of our diarrhoeal/dysenteric samples are from preschool children from Mirpur and as a result we do not know whether they are representative of all age group and strains from all parts of Bangladesh. (b) We judged the PCR fragment sizes in the gel by eye; as a result, some of these sizes could be slightly different from those which were assigned. However, we tried to be as accurate as possible. (c) Since we had only 9 samples that were collected during an episode of dysentery (as opposed to diarrhoea) we did not establish a separate group (from the diarrhoeal group) to analyze these. However larger numbers of dysenteric samples may have helped us to find a possible association between genotypes and dysentery as a clinical outcome of infection. (d) Our genotyping is based on PCR fragment size differences in tRNA-linked loci. The same product size at a particular locus does not necessarily mean the same sequence and that is what we observed in our DNA sequence diversity investigation (Chapter 7). This might have some impact on the observed association between genotypes and clinical outcomes. We noticed that sometimes two strains with apparently identical genotype were linked to different clinical outcome in the infected individuals. It is possible that these are actually not same genotype. However, we would also expect to find some infections in asymptomatic individuals that would eventually give rise to invasive disease. So mixed clusters are to be expected.

Most of the diarrhoeal/dysenteric samples were subjected to bacteriological tests for pathogens, however no viral tests were done (personal communication with Dr. Rashidul Haque). Tests were not done for 9 of 52 samples in the *E. histolytica* group. Of the remaining 43 samples, potential diarrhoea/dysentery-causing bacterial pathogens were detected in 12 (27.9%) samples only (data not shown). In the 3-locus DAR analysis, genotype 18 showed a statistically significant association with diarrhoeal/dysenteric

infection, and we found that 6 of 20 (30%) samples with this genotype, in which bacterial tests were done, contained a concomitant bacterial pathogen infection. Since the overall rate of concomitant bacterial infection in this sample group is very similar to that found for genotype 18, we do not think that the associated bacterial infections would have influenced our results. In the *E. dispar* group only one of the 7 diarrhoeal/dysenteric samples had a concomitant bacterial infection, so the cause of diarrhoea/dysentery in these children was not clear. It could be of viral or other unknown etiological origin.

Clark and Diamond (1993) observed 16 different genotypes using the combined results of SREHP/*Alu*I restriction digestion and SSG amplification in 18 isolates of *E. histolytica* isolated from diverse geographic location. We detected 93 genotypes in clinical samples from 123 individuals using 6 loci, which is comparable in terms of the diversity of genotypes detected. Our results also showed similar trends to those in Ayeh-Kumi et al (2001) who studied clinical samples from Bangladesh using a nested SREHP PCR and found 25 genotypes in 42 intestinal isolates and 9 genotypes in 12 amoebic liver abscess strains. However, 8 out of 9 liver abscess genotypes were found to be unique among ALA patients investigated. We also noticed that most of our liver abscess genotypes were unique and not found among the intestinal strains.

Haghighi et al (2003) investigated the genotypes of 79 clinical *E. histolytica* isolates from Japan (40, mainly from mentally retarded patients and homosexual males), Thailand (27), Bangladesh (5), and other countries (7, mainly reference isolates) using 4 loci – two tRNA-linked loci, chitinase and SREHP – but failed to find an association between the parasite genotype and the clinical outcome of disease in the host. There are several possibilities for this apparent lack of association – i) the sample size was too small for such highly diverse geographic samples, ii) they did not use the proper analytical approaches, and iii) perhaps the polymorphic loci they used were not appropriate.

There are several positive aspects to our genotyping method

1. It amplifies DNA extracted from culture, stool and liver abscess pus.

2. If a nested PCR is required, particularly when the amount of DNA is too low in the biological samples, e.g., stool or liver abscess samples, it can be performed using the tRNA primers in the 1st PCR, and the species-specific primers in the 2nd amplification.
3. The amplifications are species-specific and as a result not influenced by the presence of other *Entamoeba* DNA.
4. Unlike SREHP, it does not require a restriction endonuclease digest to detect all genotypes.
5. These PCRs are fast and take less than two hours to complete.

We investigated the relationships between genotypes by eBURST rather than by phylogenetic trees because eBURST identifies groups of related genotypes in the population on the basis of a simple evolutionary model in which an ancestral genotype (i.e. the founding genotype in our case) is linked to other closely related genotypes to produce a cluster. While phylogenetic trees could be generated, the data from our 6 loci may only be considered as 6 separate characters (like 6 nucleotide or amino acid positions in an alignment) which are not adequate for this approach.

eBURST analysis of Mirpur samples (Figure 4.4) produced 4 clinically relevant clusters – 1 with only asymptomatic samples and 3 others predominantly with diarrhoeal/dysenteric samples. When all other samples from Bangladesh (that is the ICDDR,B and the liver abscess samples) were included with the Mirpur samples, the previously mentioned asymptomatic cluster lost its integrity by incorporating a diarrhoeal/dysenteric sample from ICDDR,B of unknown geographic origin, although the other clusters continued to show the link to symptoms (Figure 4.5). This indicates that the symptomatic samples appear to be more closely linked to each other as liver abscess samples often tended to cluster with diarrhoeal/dysenteric samples in eBURST. The clustering may indicate that invasive strains arose a limited number of times.

We found that the 6-locus genotype 66 showed a statistically significant association ($P<0.05$) with diarrhoeal/dysenteric samples in both Mirpur and across all Bangladeshi

samples. Since the total number of individuals sharing any other single genotype was very small, none showed any significant associations.

In DAR genotyping, the total number of individuals with the same genotype was relatively high. We found an association between 2 individual genotypes and the clinical outcome of infection – genotype 6 is linked with the asymptomatic group and genotype 18 with the diarrhoeal/dysenteric group. It is worth mentioning that DAR genotype 18 essentially contained all the individuals belonging to 6-locus genotype 66 in addition to other individuals. In the liver abscess group we did not find any association simply because no individual genotype contained more than 5 of the 43 liver abscess samples. Genotype 14, which was found in 2 asymptomatic and 5 liver abscess samples (and not found in any of the diarrhoeal/dysenteric group) failed to show a statistically significant association with clinical outcome ($P = 0.066$). There are two explanations for why liver abscess strains produce a greater number of genotypes. One, the liver abscess strains were from a much greater geographic area and two, it is possible that liver abscess strains are more divergent compared to asymptomatic and diarrhoeal/dysenteric strains. However, which if either of these was contributing more towards the diversity we observed is impossible to tell from this study. Only a properly designed case control based genotyping study with 3 types of clinical samples from same geographic location can give a clear answer.

We noticed some differences between *E. histolytica* and *E. dispar* in their polymorphism at the 6 loci studied here. First, although locus S-D was highly polymorphic among *E. histolytica* strains, it was non-polymorphic in *E. dispar*. The S-D sequence from *E. dispar* SAW760 does not contain any STRs and this is the likely explanation for the lack of variation. Second, multiple bands were quite common for locus N-K in *E. dispar*, but we did not see this in *E. histolytica*.

The total number of genotypes observed using 6-locus and DAR for Mirpur children were 48 and 20 for the 70 *E. histolytica* infected children (i.e. an average of 1.46 and 3.5 children shared the same genotype on average, respectively) and 32 and 16 for 42 *E. dispar* infected children (i.e. 1.31 and 2.63 children shared the same genotype, respectively). So this

suggests that a comparable number of genetically distinct *E. histolytica* and *E. dispar* strains circulate in Mirpur study area.

However, the Chao estimation for total predicted genotypes did not fully support the above findings as it predicted more DAR genotypes for *E. dispar* than *E. histolytica* (48 versus 30, respectively), but fewer 6-locus genotypes for *E. dispar* than *E. histolytica* (85 versus 109, respectively). The latter difference may be due to the lack of polymorphism at locus S-D in *E. dispar.*

In the follow-up samples we noticed that infection with a new genotype is more common in *E. histolytica* than *E. dispar* since 10 of 23 *E. histolytica* follow-up children seemed to have become infected with at least one other genotype compared to only 2 of 15 children in *E. dispar* group.

GENEPOP analysis using all 6-locus data showed that all 3 pairs of clinical populations are different from each other in Mirpur as well as across all Bangladesh. However a similar analysis with the 3-locus DAR data showed that the some of these significant differences among clinical populations do not persist any more. This is because the data from the other 3 loci not included in DAR were contributing more towards the significant difference seen in 6-locus data. We conclude that overall genotypic difference detected among the various clinical populations increases with an increase in the number of loci and therefore for population based analysis it would be more appropriate to use the greater number of polymorphic loci.

We observed a very large number of genotypes in Bangladeshi strains of *E. histolytica* – 93 and 30 in 119 individuals using 6 and 3-locus genotyping, respectively. This result was not a big surprise as Haghighi et al (2002 and 2003), Ayeh-Kumi et al (2001) and Zaki et al (2003a) also observed a high degree of polymorphism in *E. histolytica*. However too many genotypes can be a problem when looking for a strong association between strain genotype and clinical outcome.

Nevertheless, we found that our overall genotype results using 6 loci showed a statistically significant difference between the 3 sample populations, suggesting that the overall genetic make up of the strains identified in the clinical populations is distinct. Other approaches need to be implemented to locate exactly what is causing the variable outcome of diseases in the hosts as it is unlikely that these tRNA linked STR loci are directly responsible in any way.

We do not know whether the observed association between some of the individual genotypes in *E. histolytica* and clinical outcome of infection using 6-locus and DAR genotypes in Bangladesh also exist in other parts of the world. To address this question we extended our analysis to include a large number of non-Bangladeshi *E. histolytica* samples.

CHAPTER 5

DAR GENOTYPING OF NON-BANGLADESHI STRAINS OF *E. HISTOLYTICA*

5.1. Introduction

In the previous Chapter we investigated the genotypes in clinical samples of *E. histolytica* from Bangladesh in which all the asymptomatic samples were from children from Mirpur, diarrhoeal/dysenteric samples were also predominantly from Mirpur, and liver abscess samples were from a much broader area. We observed a high degree of polymorphism in this restricted area using 6 locus tRNA-linked genotyping. In addition to finding genetic diversity, we found an association between some of the individual genotypes of the *E. histolytica* strain and the clinical outcome of disease using the 3-locus DAR genotyping. Moreover we noticed an overall difference in genotypes between asymptomatic and diarrhoeal/dysenteric samples in Mirpur as well as between diarrhoeal/dysenteric and liver abscess samples from across Bangladesh.

In this Chapter we extended DAR genotyping to some of the available clinical strains of *E. histolytica* from other parts of the world to find out whether the genotypes that were found to be associated with asymptomatic or symptomatic infections in Bangladesh were also found in similar clinical samples in other countries, whether genotypes that were common in Bangladesh were also common in other countries, and whether there was an overall difference between genotypes in samples from different clinical groups from around the world.

5.2. *E. histolytica* isolates

A total of 87 *E. histolytica* DNA samples were available for investigation from 11 different countries. Appendix VI provides some background information on these samples.

All 27 Vietnamese stool DNA samples and all 11 South African culture DNA samples were from asymptomatic adults. In contrast, all 9 Georgian culture DNA samples and 5 of 6 Turkish stool DNA samples were from symptomatic diarrhoeal/dysenteric patients. Culture DNA samples were available from 26 individuals from Japan – 8 of them were asymptomatic, 7 had only diarrhoea/dysentery, 2 had only liver abscess, another 2 patients had both diarrhoea/dysentery and liver abscess (samples J7 and J8: cultures were established from stool specimens), while the clinical information was not available for the remaining 7 DNA samples. Apart from the 4 liver abscess samples from Japan (mentioned above) only one more sample in this category was available, from Korea.

In summary 50 asymptomatic, 25 diarrhoeal/dysenteric, 3 liver abscess, 2 diarrhoea/dysenteric plus liver abscess, and 7 samples with unavailable clinical information were studied here (Appendix VI).

5.2.1. 3-locus PCR amplification

E. histolytica species-specific primers for 3 loci – D-A, A-L and R-R (DAR) - were used to amplify the corresponding DNA fragments as described in Chapters 3 and 4.

5.3. Results

5.3.1. PCR amplifications

PCR amplification was successful for all samples, although occasionally a nested PCR was necessary (data not shown). However, in 12 samples we observed the presence of other fainter bands in one of the 3 loci which we ignored for two reasons: one, in most cases we thought the faint band was non specific and two, for the simplicity of the analysis. Figure 5.1 shows PCR product size variations in a selection of samples.

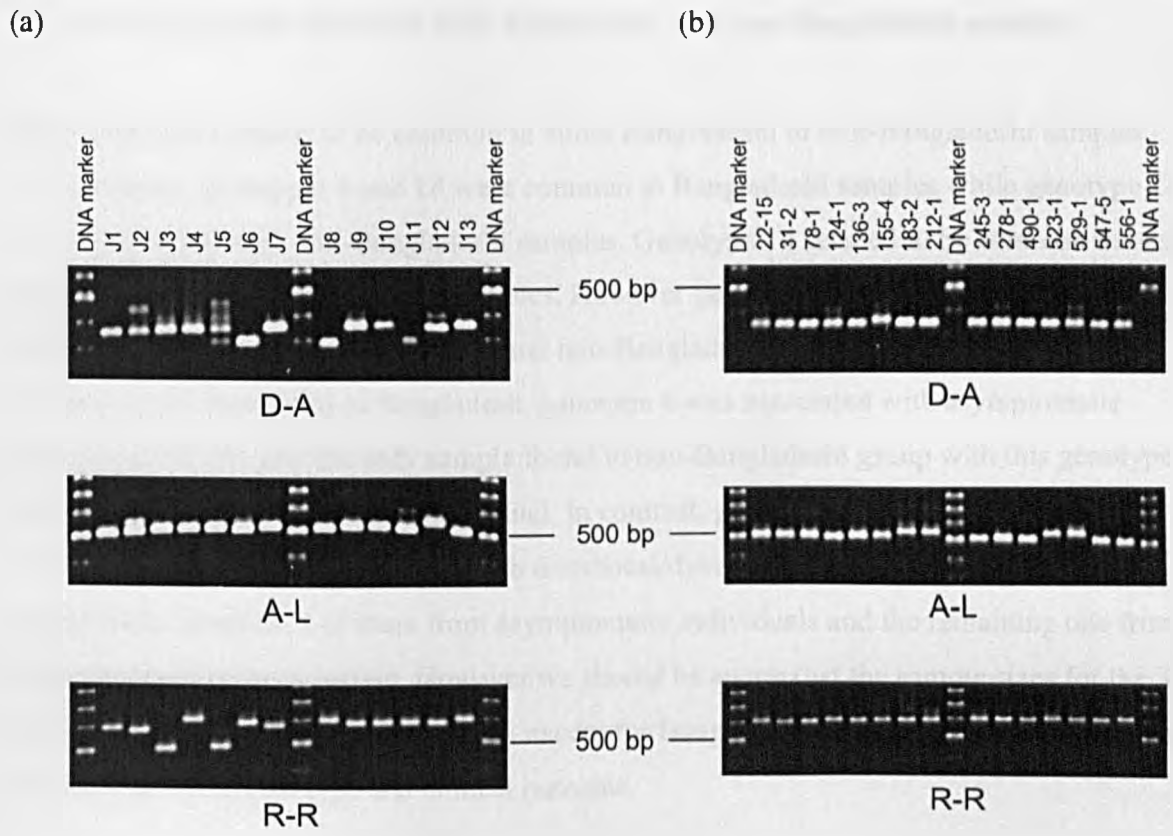


Figure 5.1. DAR PCR patterns of a selection of non-Bangladeshi samples. (a) 13 Japanese samples and (b) 15 Vietnamese samples.

5.3.2. DAR genotypes

In order to facilitate an easy comparison between DAR genotypes that were detected in Bangladesh and ones detected in other countries we kept the same genotype numbers 1 to 30 (corresponding to Bangladeshi DAR genotype numbers) while 31 to 46 were new genotypes detected only in the non-Bangladeshi survey. A total of 31 DAR genotypes were detected across all non-Bangladeshi samples (Appendix VII). However, 15 of 31 genotypes were previously detected in Bangladeshi samples (Chapter 4) and the 16 new genotypes (genotypes 31 to 46) comprised less than one-third of all non-Bangladeshi samples (27 of 87 samples, Appendix VII). The two most common non-Bangladeshi genotypes were 15 and 17, found in 23 and 7 individuals, respectively.

5.3.3. DAR genotypes found in both Bangladeshi and non-Bangladeshi samples

Some genotypes appear to be common in either Bangladeshi or non-Bangladeshi samples. For examples, genotypes 6 and 18 were common in Bangladeshi samples while genotype 21 was more common in non-Bangladeshi samples. Genotype 14 seemed to be common in both Bangladeshi and non-Bangladeshi samples. However genotypes 2, 5, 11, 20, 24, 27 and 30 seemed to be rare in both Bangladeshi and non-Bangladeshi samples.

We previously found that in Bangladesh, genotype 6 was associated with asymptomatic infection ($P<0.05$), and the only sample found in non-Bangladeshi group with this genotype was also from an asymptomatic individual. In contrast, genotype 18, which showed association in Bangladeshi samples with diarrhoeal/dysenteric patients, was found in 5 non-Bangladeshi samples, 4 of them from asymptomatic individuals and the remaining one from a diarrhoeal/dysenteric patient. However we should be aware that the sample sizes for the 3 clinical types in any individual country, except for Bangladesh, were too small to investigate the association of genotype and clinical outcome.

5.4. DAR genotype distribution among clinical groups

Statistical analysis on genotype distribution among 3 clinical groups was done using the web based analytical program GENEPOP (http://wbiomed.curtin.edu.au/genepop/). Since the sample size of the liver abscess group in the non-Bangladeshi sample was too small (3 samples) we compared the asymptomatic and diarrhoeal/dysenteric populations only.

In non-Bangladeshi samples, using all three PCR patterns, we found a significant overall difference ($P<0.05$) between the asymptomatic and diarrhoeal/dysenteric groups (Table 5.2). Two loci contributed significantly towards this difference, A-L and R-R. In particular, two PCR patterns in locus A-L – #2 and #4 – were found in 29 and 8 of 50 asymptomatic samples compared to only 7 and 1 of 27 diarrhoeal/dysenteric samples, respectively. Two PCR patterns in locus R-R, #3 and #4, were found in 16 and 6 of 27 diarrhoeal/dysenteric samples compared to 8 and 40 of 50 asymptomatic samples, also showing significant differences (Table 5.2 and Appendix VII). In contrast across all Bangladeshi samples there was no significant overall difference in DAR genotypes between the asymptomatic and

Table 5.1. Clinical distribution of the *E. histolytica* DAR genotypes from Bangladesh and other parts of the world.

DAR genotype #	Strain origin	Clinical status				Frequency in	
		A	D	L	NA	Individual country	Total
2	Bangladesh	0	0	1	0	1	2
	Japan	0	1*	0	0	1	
5	Bangladesh	2	0	0	0	2	3
	South Africa	1	0	0	0	1	
6	Bangladesh	10	4	4	0	18	19
	Japan	1	0	0	0	1	
11	Bangladesh	0	1	0	0	1	3
	Georgia	0	1	0	0	1	
	Turkey	0	1	0	0	1	
14	Bangladesh	2	0	5	0	7	12
	Burma	0	2	0	0	2	
	South Africa	3	0	0	0	3	
15	Bangladesh	0	0	2	0	2	25
	India	1	0	0	0	1	
	Japan	1	0	1	4	6	
	South Africa	1	0	0	0	1	
	Vietnam	15	0	0	0	15	
17	Bangladesh	5	7	1	0	13	20
	Georgia	0	3	0	0	3	
	Turkey	1	3	0	0	4	
18	Bangladesh	6	22	5	0	33	38
	Mexico	0	1	0	0	1	
	Vietnam	4	0	0	0	4	
19	Bangladesh	1	1	3	0	5	6
	Georgia	0	1	0	0	1	
20	Bangladesh	0	2	0	0	2	3
	Turkey	0	1	0	0	1	
21	Bangladesh	0	0	1	0	1	7
	Vietnam	6	0	0	0	6	
24	Bangladesh	0	0	2	0	2	4
	Korea	0	0	1	0	1	
	Japan	0	1	0	0	1	
27	Bangladesh	1	0	1	0	2	3
	South Africa	1	0	0	0	1	
29	Bangladesh	1	1	0	0	2	5
	Japan	0	2*	0	1	3	
30	Bangladesh	0	0	2	0	2	4
	South Africa	2	0	0	0	2	
Total		65	55	29	5	154	154

A = asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess, and NA = not available. *One of the 2 patients in genotype 29 and only Japanese patient in genotype 2 also had liver abscess, but we considered them in this clinical group since the DNA source was cultures isolated from stool samples.

Table 5.2. Statistical analysis of genotype distribution among clinical groups.

Analysis	Population pair	No. of samples	Chi^2	Overall P values	Loci showing significant differences (P values)
Non-Bangladeshi	A & D	A = 50 D = 27	Infinity	<0.0001	Locus A-L (0.0055) Locus R-R (<0.0001)
Combined	A & D	A = 87 D = 75	25.366	0.0003	Locus A-L (0.0007)
	D & L	D = 75 L = 46	23.954	0.0005	Locus D-A (0.0199) Locus A-L (0.0050)
	L & A	L = 46 A = 87	16.786	0.0101	Locus R-R (0.0035)

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

diarrhoeal/dysenteric groups, although locus D-A on its own was still significant (Table 4.4). In addition, the overall significant difference in DAR genotypes between asymptomatic and diarrhoeal/dysenteric groups in Mirpur children was due only to significant contribution by locus D-A compared to A-L and R-R in the non-Bangladeshi samples. The reason for different loci being significant in different parts of the world between the same pair of sample populations is not clear.

When all sample data available were combined we found an overall significant difference among all 3 clinical populations (Table 5.2). Between asymptomatic and diarrhoeal/dysenteric population the only locus contributing towards this significant difference was A-L. Between diarrhoeal/dysenteric and liver abscess population the loci contributing towards this significant difference were D-A and A-L. Between liver abscess and asymptomatic population the only locus contributing towards this significant difference was R-R.

5.5. Association of individual non-Bangladeshi DAR genotypes with clinical outcome of infection

We also investigated whether some of the more common DAR genotypes showed an association with clinical outcome of infection. Since the number of liver abscess samples in

non-Bangladeshi samples was low (only 3 samples with 3 different genotypes) we did not include them in the 'non-Bangladeshi'analysis (although they were included in the combined analysis). We also ignored the 7 samples for which clinical information was not available. The two Japanese samples (J7 and J8) that were from individuals with both diarrhoea/dysentery and liver abscess were included in diarrhoea/dysentery group since the origin of DNA was stool samples, not liver abscess pus. Across all non-Bangladeshi samples - we detected 16 DAR genotypes in 50 asymptomatic individuals and 18 genotypes in 27 diarrhoeal/dysenteric patients.

In non-Bangladeshi samples genotype 15 showed a strong association with asymptomatic infection (P = 0.0001) but was only found in 18 of 50 non-Bangladeshi asymptomatic samples, showing almost no predictive value. Similarly although the genotype 17 showed statistically significant association (P = 0.0114) with diarrhoeal/dysenteric samples, it was found only in 6 of 27 samples of this clinical type, showing no predictive value at all.

When we combined the DAR results of both non-Bangladeshi and Bangladeshi samples, we found that the two most common genotypes now were 18 and 6. Genotype 18 showed a statistically significant association with diarrhoeal/dysenteric infection while genotype 6 showed no association with any particular outcome of infection (Table 5.3).

We were also interested to know whether the association we observed for genotypes 15 and 17 in non-Bangladeshi samples alone still existed. In fact, they maintained the similar clinical associations even after all samples were combined (Table 5.3).

5.6. Association between DAR genotypes by eBURST analysis

From eBURST analysis of all non-Bangladeshi samples we could detect 3 clusters that had some clinical relevance (Figure 5.2). Cluster a contained 5 samples, 3 of which were diarrhoeal/dysenteric samples. Cluster b contained almost exclusively asymptomatic strains, and cluster c contained predominantly diarrhoeal/dysenteric samples.

Table 5.3. Statistical analysis of the most common individual DAR genotypes among clinical groups.

Analysis	Genotype	Population pair with corresponding frequency of the genotype	Chi^2	P values
Non-Bangladeshi samples	15	A = 18/50 D = 0/27	14.59	0.0001
	17	A = 1/50 D = 6/27	6.40	0.0114
Across all samples	18	A = 10/85 D = 23/71	9.87	0.0017
		D = 23/71 L = 5/46	7.10	0.0077
		L = 5/46 A = 10/85	0.02	0.8780
		Overall	13.24	0.0013
	6	A = 11/85 D = 4/71	2.38	0.1231
		D = 4/71 L = 4/46	0.07	0.7902
		L = 4/46 A = 11/85	0.530	0.4664
		Overall	2.46	0.2924
	15	A = 18/85 D = 0/71	17.00	<0.0001
		D = 0/71 L = 3/46	2.50	0.1138
		L = 3/46 A = 18/85	4.76	0.0291
		Overall	19.58	<0.0001
	17	A = 6/85 D = 13/71	4.58	0.0324
		D = 13/71 L = 1/46	6.90	0.0086
		L = 1/46 A = 6/85	0.61	0.4356
		Overall	9.48	0.0088

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

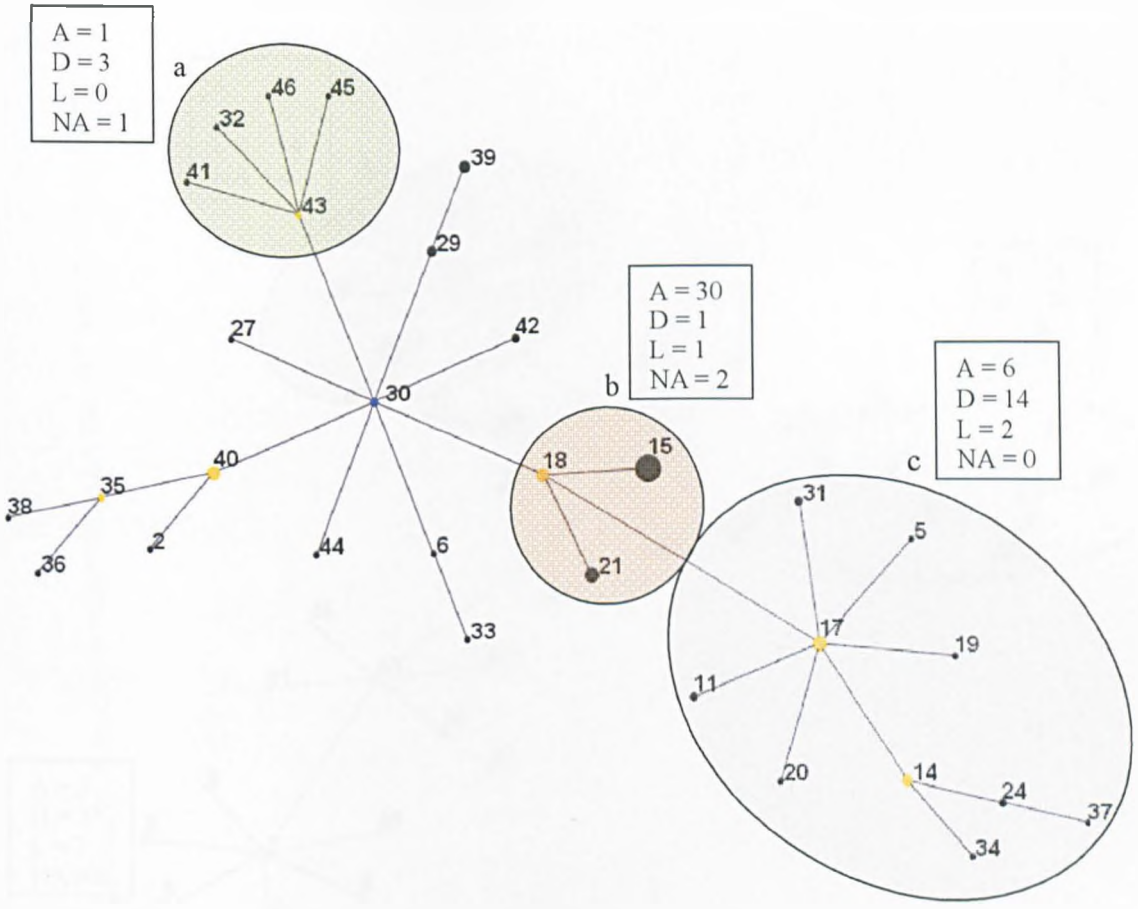


Figure 5.2. Non-Bangladeshi eBURST with genotype 30 as the founder (bootstrap support for this was 32% compared to next highest was 26% for 17).

eBURST analysis of combined non-Bangladeshi and Bangladeshi samples detected 3 clusters that had some clinical relevance (Figure 5.3). Cluster a contained 4 samples, 2 of which were found in liver abscess samples and the remaining 2 (asterisk in the Figure 5.3) were from the stool of patients who had diarrhoea/dysentery as well as liver abscess. Cluster b contained a relatively higher proportion of asymptomatic strains, and cluster c contained higher proportion of diarrhoeal/dysenteric samples.

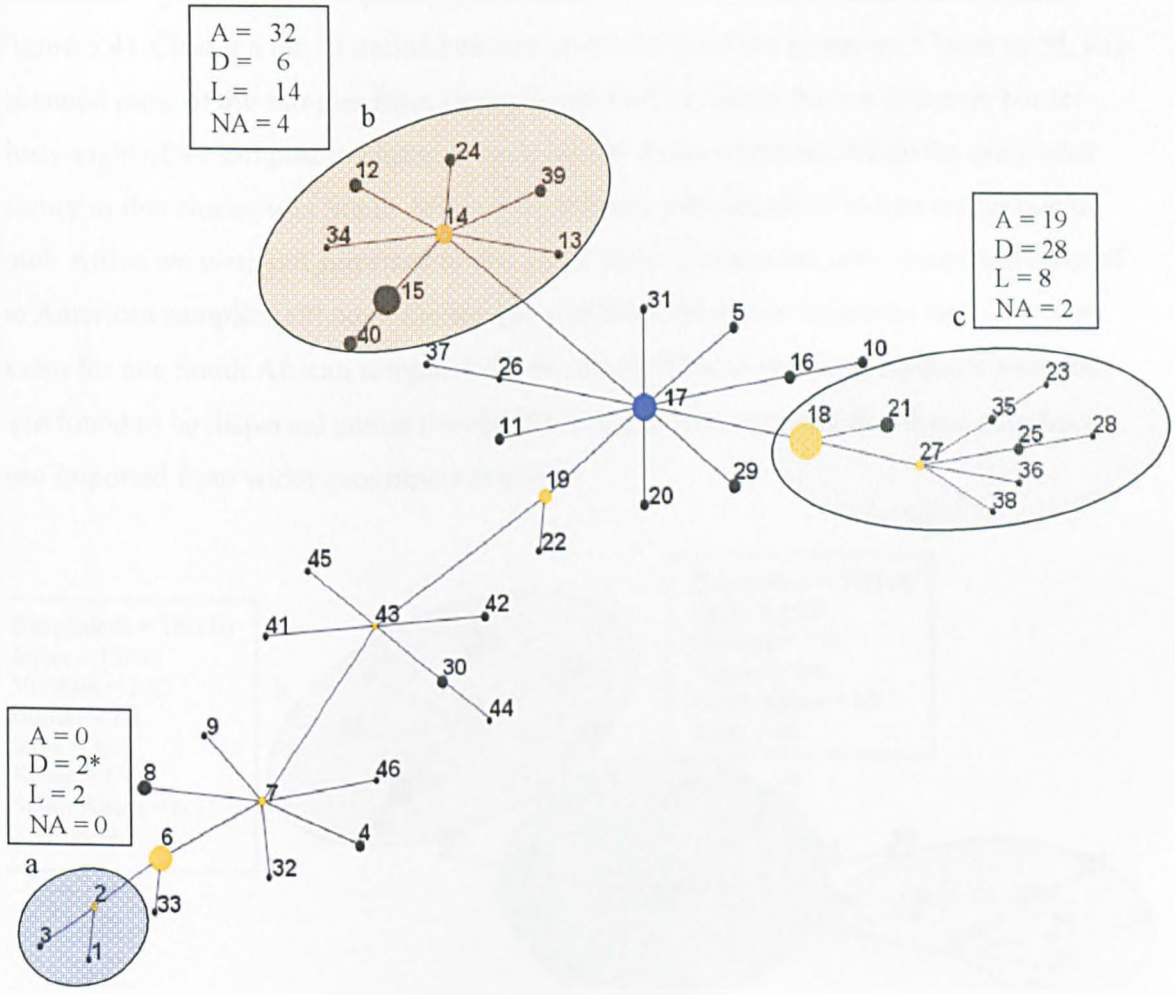


Figure 5.3. eBURST with combined non-Bangladeshi and all Bangladeshi samples using DAR genotyping. Bootstrap support for the auto assigned founder genotype 17 - was 31% compared to next highest value for genotype 14 which was only 16%; *indicates that in cluster a one sample was from a liver abscess patient who also had diarrhoea/dysentery.

The geographic distribution of genotypes using all samples revealed that there exists some evidence of a geographic component to the distribution of DAR genotypes world-wide (Figure 5.4). Cluster a (an ill defined cluster centred around the genotype 17 and its SLVs) contained most of the samples from Georgia and Turkey, which have a common border. Thirty-eight of 44 samples in cluster b were from 6 Asian countries, while the only other country in this cluster was South Africa. Since many individuals of Indian origin live in South Africa we were not surprised to see this. Cluster c contained only Asian and most of the American samples (although the sample size from American countries was very low) except for one South African sample. Like the South African samples, Japanese samples were found to be dispersed across the eBURST diagram, suggesting that these may have been imported from wider geographic regions.

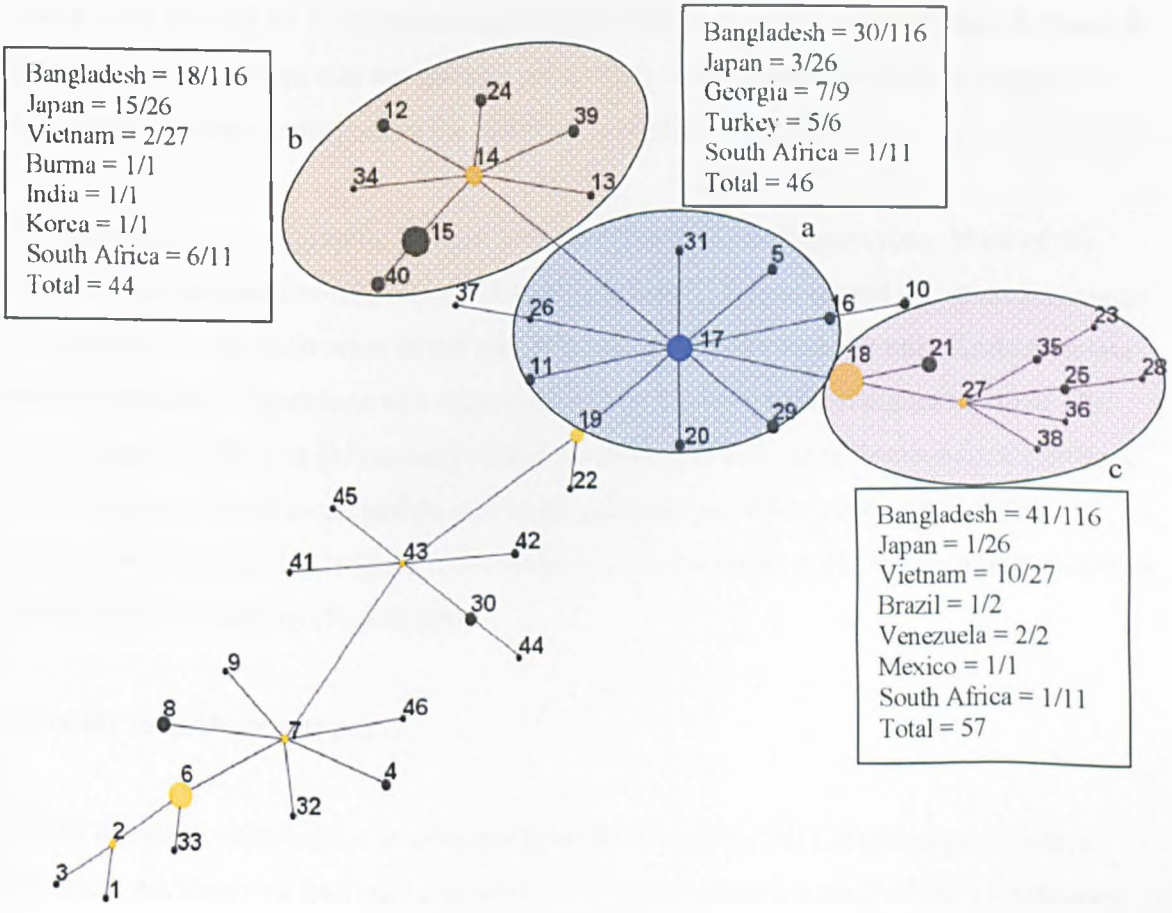


Figure 5.4. Geographic distribution of non-Bangladeshi and Bangladeshi DAR genotypes using eBURST.

5.7. Discussion

Three main aims of this study were to investigate – i) diversity of DAR genotypes in other countries, ii) geographic links to global distribution of these genotypes, and iii) any association between genotypes and clinical outcomes in other countries. We were mostly successful in achieving these objectives although we had some limitations. These were of two kinds – sample based limitations and time based limitations. Sample based limitations include: a) from 6 out of 11 countries we had only one or two samples, so it was impossible to know the extent of diversity in these countries, b) the number of liver abscess samples was very small, so it was impossible to make any conclusion about the non-Bangladeshi liver abscess genotypes, c) in 7 Japanese samples the clinical information was not available, which reduced our ability to analyze any associations. We had one experimental limitation – we did not attempt PCR amplification of the remaining 3 of the 6-loci (N-K2, S-D and S-Q) simply because there was not enough time, otherwise we may have seen a better association between genotypes in the eBURST analysis.

We observed some geographic links in the distribution of DAR genotypes. Most of the isolates from two neighboring countries - Georgia and Turkey - shared the same genotypes or occurred close to each other in the eBURST diagram. Likewise the only Korean isolate showed an identical genotype to a Japanese isolate. We also noticed contradictions. The only isolate from Burma (Myanmar) shared its genotype with three South African isolates. Similarly, the only Mexican isolate shared its genotype with 4 Vietnamese isolates. However we noticed that isolates from several Asian countries clustered almost exclusively in the eBURST diagram (Figure 5.4).

Diversity in DAR genotypes

One of our main objectives was to investigate the diversity of DAR genotypes in other countries. Although we had only one or two samples available from 6 of the 11 countries six or more samples were available for each of the remaining 5 countries. The diversity of DAR genotypes varied from one country to another (Table 5.4). For example, the diversity

was high among Georgian isolates but was much lower among the Vietnamese isolates. The samples from Georgia were drawn from a large region as patients attended the Institute of Parasitology and Tropical Medicine or the Hospital of Infectious Diseases from a variety of locations in the country, so we were not surprised to see a high degree of polymorphism in these samples. Likewise, the Japanese samples showed a high degree of polymorphism and these were from 5 different cities. In contrast, the samples from Vietnam were from a restricted geographic region and showed less diversity. The Vietnamese samples were from the Phu Cat commune of Hué City, which is more like the Mirpur study area in Bangladesh in size, and showed the least polymorphism. In contrast, samples from the restricted geographic region of Sanliurfa, Turkey, showed high polymorphism. The 11 South African samples are from a relatively restricted geographic region but show a high degree of diversity. However the 11 South African samples in this study are not random, but actually represent a much larger sample. We selected these samples for study based on the previous results of Zaki et al (2003a), where they were shown to be all different using 11 pairs of primers. In general, genotype diversity seems to be linked to the size of the geographic area from which the samples came.

We found association between some of the individual DAR genotypes and clinical outcome of infection in the non-Bangladeshi samples. Genotypes 15 and 17 showed association with non-Bangladeshi asymptomatic and diarrhoeal/dysenteric samples, respectively (not seen in Bangladeshi samples). This association was maintained even when we combined both non-Bangladeshi and Bangladeshi samples, suggesting that this association has a worldwide distribution. Likewise, genotype 18 showed a significant association with diarrhoeal/dysenteric infection not only in Bangladesh but also in the combined sample population. In contrast, genotype 6 which had shown an association with asymptomatic infection in Bangladeshi samples lost its statistical significance in the combined samples. However, we should be careful in interpreting our data. A greater number of non-Bangladeshi clinical samples would give us a better view of the clinical associations.

Table 5.4. Country wise diversity in DAR genotypes.

Strain origin	Sample area	No. of samples	No. of genotypes	Sample:genotype ratio
Turkey	Sanliurfa region of Turkey	6	3	2
Georgia	Attended one of 2 hospitals – [1]IPTM or [2]HID	9	6	1.5
South Africa	Periurban area surrounding the city of Durban	11	7	1.6
Japan	Tokyo, Kyoto, Okayama, Kanagawa, and Shizuoka	26	13	2
Vietnam	Phu Cat commune of Hué City	27	5	5.4

[1]IPTM = Institute of Parasitology and Tropical Medicine, [2]HID = Hospital of Infectious Diseases.

We were also interested to know how frequent the common Bangladeshi DAR genotypes were globally and vice versa. The first and third most common genotypes in Bangladesh (18 and 17 – found in 33 and 13 individuals, respectively) were found to be 5th and 2nd most common in the non-Bangladeshi samples (found in 5 and 7 individuals, respectively). However, we also noticed contradictions. For examples, the most common non-Bangladeshi genotype 15 (found in 23 individuals from 4 countries) was found only in 2 liver abscess patients in Bangladesh, while the second most common Bangladeshi genotype (i.e. genotype 6; found in 18 individuals) was found in only one asymptomatic individual from Japan (who had a history of voluntary work in Ghana).

Although the links between genotypes as revealed by eBURST analysis of non-Bangladeshi samples was not very strong, we were not too surprised. In Chapter 4 using Bangladeshi samples we noted that eBURST showed better clinical association between genotypes when the number of loci was higher and in fact we did not see any significant clustering of genotypes using DAR.

Perhaps the most exciting finding of this study was the detection of differences in DAR genotypes among the three clinical populations in non-Bangladeshi as well as in the combined samples. Although the large total number of genotypes means there is a very low predictive value for any individual genotype, the overall differences in the DAR genotypes among the clinical groups suggests that a genetic component of the parasite is linked to the outcome of infections worldwide.

Genotyping using all 6 loci and more clinical samples from countries other than Bangladesh will provide a better picture of the possible links between genotypes and the global distribution of *E. histolytica* strains.

CHAPTER 6

ENTAMOEBA MOSHKOVSKII INFECTIONS

6.1. Introduction

The free-living amoeba *Entamoeba moshkovskii* has so far been reported to infect humans very rarely. Its specific identification in stool specimens by microscopy is impossible because it is morphologically indistinguishable from the disease-causing *E. histolytica* and the non-pathogenic commensal *E. dispar*. However unlike *E. histolytica* and *E. dispar* there is no simple molecular technique available to detect *E. moshkovskii*. This has contributed to the fact that no study has investigated the prevalence of this species in a human population and thus raises the question about how precise the epidemiological data are on *E. histolytica* or *E. dispar* when the detection technique used was microscopy.

In this study we developed a simple and specific diagnostic technique and investigated the prevalence of *E. moshkovskii* in the Mirpur population. In addition we designed *E. moshkovskii* specific primers for the tRNA-linked STR R-R, which showed polymorphism in *E. histolytica* and *E. dispar* strains, in an attempt to study the diversity of *E. moshkovskii* strains from Bangladesh, Georgia and Peru.

6.2. Materials and Methods

6.2.1. *E. moshkovskii* samples

6.2.1.1. Stool specimens from Bangladesh

Faecal specimens included in this study were from 109 pre-school children aged 2-5 years from Mirpur, and were provided by Dr. Rashidul Haque, ICDDR,B. Based on PCR results on stool DNA samples, thirty-nine were *E. dispar* positive, 17 were *E. histolytica* positive,

and 1 was positive for both *E. histolytica* and *E. dispar*. Of the 52 samples negative by stool PCR, 18 were eventually found positive for *E. histolytica* and/or *E. dispar* either by PCR from culture DNA or by antigen detection tests performed on stool specimens and the remaining 34 samples were negative by all methods. Only 4 of the samples were from children with diarrhoea.

The TechLab (Blacksburg, Va.) *Entamoeba* test (designed to detect but not differentiate *E. histolytica* and *E. dispar* antigen in stool specimens) and TechLab *E. histolytica* test (designed to detect specifically *E. histolytica* in stool specimens) were performed on stool specimens from Bangladesh according to the manufacturer's instructions (Haque et al, 1993).

All stool samples from Bangladesh were cultured for *Entamoeba* species in Robinson's medium (Robinson, 1968) within 2-4 hours of collection and hexokinase isoenzyme analysis was performed when possible as previously described (Haque et al, 1993).

Stool DNA was isolated using a modified version of the method of Katzwinkel-Wladarsch et al (1994) as previously described (Haque et al, 1998b). Culture DNA was extracted by the CTAB method described earlier (Section 2.2.1).

6.2.1.2. Culture samples from Georgia and Peru

Nine culture DNA samples from Georgia were tested. These were maintained in Robinson's medium in Dr. Nino Trapaidze's laboratory at the Institute of Parasitology and Tropical Medicine, Georgia. Initially these samples were positive for *E. histolytica* by the TechLab antigen detection test and all of these were from diarrhoeal patients (data not shown). DNA was purified by the CTAB extraction method (Section 2.2.1).

One xenic culture lysate was received from Dr. William Roldan, University of San Marcos, Lima, Peru. This culture was originally established in Pavlova's medium from a stool

sample of a child and was growing at room temperature. DNA was purified by the CTAB extraction method (Section 2.2.1).

6.2.1.3. Reference isolates

E. moshkovskii Laredo and FIC were maintained axenically as described (Section 2.1). Laredo is a human isolate and FIC is an environmental isolate. *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 were used as controls.

6.3. Small subunit rRNA gene amplification

Based on the sequence of the small subunit rRNA gene (SSU-rDNA) of *E. moshkovskii* Laredo (GenBank accession number AF 149906), a nested set of primers, designated Em-1/Em-2, and nEm-1/nEm-2, was designed and used to detect *E. moshkovskii* in stool DNA (Table 6.1). In the initial PCR (total vol. 25 µl) 1.0 µl of stool or culture DNA was used. Thermal cycler conditions were - 30 cycles each consisting of 92°C for 1 min, 55°C for 1 min and 72°C for 1 min followed by a final extension of 7 min at 72°C. In the nested PCR, 1.0 µl of the first PCR product was used as the template DNA and the annealing temperature was raised to 62°C, leaving the other parameters of the amplification cycle unchanged. *E. moshkovskii*-specific nested SSU-rDNA gene amplification products were digested with restriction endonuclease *Xho* I for 1 hour at 37°C according to the manufacturer's instructions (Invitrogen,) to verify species identity.

Part of the SSU-rDNA gene was amplified from the Georgian and Peruvian culture DNA samples with the primer pair Entam 1 and Entam 2 described by Verweij et al (2001) and sequenced (Section 2.5).

6.4. R-R amplification

Although the R-R primer sequences were designed from *E. histolytica* HM-1:IMSS sequence, they also amplified *E. moshkovskii* DNA. The R-R amplification products from

E. moshkovskii Laredo, *E. moshkovskii* MS15-3646 (one of the infections detected) and *E. dispar* SAW760 were cloned into the pGEM-T Easy vector (Promega) and sequenced (MWG Biotech Ltd, UK). From the sequence results, an *E. moshkovskii*-specific primer pair, EmR-1 & EmR-2, was designed to amplify an *E. moshkovskii* R-R product specifically (Table 6.1). PCR amplification was performed at an annealing temperature of 58°C as described above for rRNA gene amplification.

Some parts of this study were carried out at ICDDR,B. This included the culture of Bangladeshi *Entamoeba* species in Robinson medium, the isoenzyme analysis, the TechLab stool ELISA tests, and diagnosis of *E. moshkovskii* by a specific nested SSU rRNA gene PCR. All the work on the Georgian and Peruvian strains and the R-R analysis of all samples was performed in London.

Table 6.1. Primers for *E. moshkovskii* analysis.

Primer name	Primer sequence (5' to 3')	Annealing temperature (°C)
Em-1	CTCTTCAGGGGGACTGCG	55
Em-2	TCGTTAGTTTCATTACCT	
nEm-1	GAATAAGGATGGTATGAC	62
nEm-2	AAGTGGAGTTAACCACCT	
Entam 1	GTTGATCCTGCCAGTATTATATG	57
Entam 2	CACTATTGGAGCTGGAATTAC	
EmR-1	GGCGCCTTTTTTACTTTATGG	58
EmR-2	GCTAACAAGGCCAATCGATAAA	

6.5. Results

6.5.1. Culture and Isoenzyme analysis in Bangladeshi samples

All 109 stool specimens were inoculated into Robinson's medium for growth of *Entamoeba* species. Incubation led to growth of *E. histolytica*/*E. dispar*/*E. moshkovskii* in 33 and *E. coli* in 8 cultures (no growth of *Endolimax nana* or *E. hartmanni* was observed). Hexokinase isoenzyme analysis was possible only for 10 cultures; 4 of them showed the

band pattern of *E. histolytica*, 5 *E. dispar* and 1 showed the band pattern of *E. dispar* with an extra band just behind the faster moving band, perhaps indicating a mixed culture.

6.5.2. Detection of *E. moshkovskii* by nested PCR in Bangladeshi samples

The reference isolate *E. moshkovskii* Laredo gave the expected band at ca. 260 bp with the *E. moshkovskii*-specific SSU-rDNA nested primers while control *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 DNAs were negative. Twenty-three out of 109 (21.1%) stool DNA samples were positive by nested PCR for *E. moshkovskii* (Figure 6.1). Seventeen of these *E. moshkovskii* positive samples were also positive for *E. histolytica* and/or *E. dispar* by either stool SSU rDNA PCR (13/17) or using the TechLab *Entamoeba* or *E. histolytica* tests (15/17) (Table 6.2). One of the 4 children with diarrhoea was positive for *E. moshkovskii* and was co-infected with *E. dispar*. The cause of his diarrhoea remained undetermined.

From a comparison of SSU-rDNA sequences from *E. moshkovskii*, *E. histolytica* and *E. dispar*, it was found that the restriction endonuclease *Xho* I cuts exclusively in the *E. moshkovskii* specific 258-bp nested PCR product to produce 236 bp and 22 bp fragments. Products from all 23 positive stool samples and the Laredo isolate showed the presence of this site (Figure 6.1).

6.5.3. Amplification of locus R-R and sequence analysis

In an attempt to detect polymorphism among the *E. moshkovskii* samples, we studied STR locus R-R, which shows polymorphism in *E. histolytica* and *E. dispar*. The R-R primers amplify *E. histolytica*, *E. dispar* and *E. moshkovskii* DNA. The sizes of the PCR products from *E. histolytica* HM-1:IMSS, *E. dispar* SAW760 and *E. moshkovskii* Laredo were 586 bp, 586 bp and 323 bp, respectively. We never observed a band in the 250-350 bp region in any of our *E. histolytica* and *E. dispar* strains with these primers (Chapters 4 and 5). Because 17 out of 23 *E. moshkovskii* positive samples were also positive for *E. histolytica* and/or *E. dispar* (by SSU-rDNA PCR or TechLab ELISA), we ignored products in the

Table 6.2. Nested SSU rDNA PCR for *E. histolytica* and *E. dispar* and stool antigen detection test results of the 23 *E. moshkovskii* positive samples from Bangladesh.

Samples	Stool antigen detection test results	SSU rRNA gene PCR for *E. histolytica* and/or *E. dispar*	
		Stool DNA	Culture DNA
1[$]	*E. histolytica*	*E. dispar*	Mixed
2	*E. dispar*	0	NC
3	*E. dispar*	*E. dispar*	NC
4	*E. histolytica*	0	NC
5	*E. dispar*	*E. dispar*	*E. dispar*
6[$]	*E. dispar*	Mixed	*E. dispar*
7	*E. dispar*	*E. dispar*	NC
8	*E. dispar*	*E. dispar*	NC
9	*E. dispar*	*E. dispar*	NC
10	0	*E. dispar*	NC
11	*E. dispar*	*E. dispar*	NC
12	*E. dispar*	0	NC
13[$$]	*E. dispar*	*E. histolytica*	NC
14	0	0	*E. dispar*
15	*E. dispar*	0	*E. dispar*
16	0	0	*E. dispar*
17	*E. histolytica*	0	NC
18	0	0	NC
19	0	0	NC
20	0	0	NC
21	0	0	NC
22	0	0	NC
23	0	0	NC

NC = no culture.; 0 = negative.

NB: All stool antigen tests that are positive for *E. histolytica* can also be mixed as no specific *E. dispar* antigen test exists.

$: It is most likely that subjects 1 and 6 were mixed infections with *E. histolytica* and *E. dispar*, where *E. histolytica* was much lower in number in the stool specimen. For subject 1, SSU rDNA PCR failed to detect *E. histolytica*, though both species grew in the culture. For subject 6, although SSU rDNA PCR could detect *E. histolytica* in stool DNA, the *E. histolytica* antigen detection test failed to detect *E. histolytica* and only *E. dispar* survived in the culture.

$$: The stool specimen of subject 13 was marginally negative by the *E. histolytica* antigen detection test (OD value was 0.13 where the cut off value for a positive result was 0.15).

500-600 bp region (assuming that they derived from *E. histolytica* and/or *E. dispar* DNA) and considered a sample positive for *E. moshkovskii* when it produced a band at around 300

bp. By this criterion, we found 18 out of 23 samples positive for *E. moshkovskii* (data not shown) and they showed slight PCR product size variation. The PCR products from one stool sample, *E. moshkovskii* Laredo and *E. dispar* SAW760 were cloned, sequenced and aligned with that of *E. histolytica* HM-1:IMSS, and *E. moshkovskii* specific primers (EmR-1 & EmR-2) were designed (Figure 6.2). In addition to significant PCR product size differences, analysis clearly shows that the *E. moshkovskii* sequence is completely different from those of *E. histolytica* and *E. dispar* and, unlike the *E. histolytica* and *E. dispar* sequences, it contains no short tandem repeat sequences (Figure 6.2C).

The EmR primers amplified the expected 265 bp fragment from *E. moshkovskii* Laredo DNA and did not amplify *E. histolytica* HM-1:IMSS or *E. dispar* SAW760 DNA. However, they successfully amplified 10 *E. moshkovskii* positive stool DNA samples out of a possible 23. The most likely explanation for why they did not amplify the other 13 *E. moshkovskii* DNA samples is that they differed in sequence in the primer-binding regions. Although the PCR product size of these 10 isolates was slightly different from that of Laredo, they were very similar in size to each other (Figure 6.3a). The DNA of the previously reported human *E. moshkovskii* ICDDRB:717 isolated from the same geographical location (Haque et al, 1998a) also gave a product of same size (Figure 6.3a). The EmR primers successfully amplified DNA from environmental *E. moshkovskii* isolate FIC but its product size was quite different from the Bangladeshi human isolates of *E. moshkovskii* (Figure 6.3a).

Both Georgian and Peruvian samples showed EmR amplification products similar in size to *E. moshkovskii* FIC (Figure 6.3b). Sequencing of PCR products revealed that the sequences were identical to that of *E. moshkovskii* FIC (data not shown).

6.5.4. rRNA gene amplification in Georgian and Peruvian samples

We used a pair of primers (Entam 1 and Entam 2) that were designed to amplify a part of the rRNA gene from all *Entamoeba* species (586 bp in *E. moshkovskii* Laredo and 584 bp in *E. moshkovskii* FIC). This primer pair amplified 5 of the 9 Georgian DNAs and the

Peruvian DNA (data not shown). Size differences in the amplified products were not obvious. We sequenced two Georgian and the Peruvian PCR products. Alignment of these sequences showed that the Georgian isolates were of *E. moshkovskii* Laredo type while the Peruvian isolate was of *E. moshkovskii* FIC type (Figure 6.4).

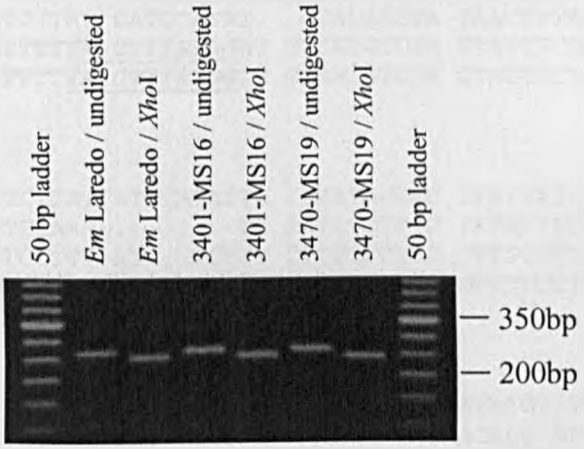


Figure 6.1. *E. moshkovskii* specific nested SSU rDNA PCR products followed by *Xho* I digestions. Data provided by Mohammad Bakhtiar Hossain, ICDDR,B.

A) 5'-end R-R sequence alignment:

```
            1                                                        50
HM1-IMSS    AGCATCAGCC TTCTAAGCTG AGGGTCGCAG GTTCGAGCCC TGCATGAGGC
SAW760      AGCATCAGCC TTCTAAGCTG AGGGTCGCAG GTTCGAGCCC TGCATGAGGC
Laredo      AGCATCAGCC TTCTAAGCTG AGGGTCGCAG GTTCGAGCCC TGCATGAGGC
3646-MS15   TTCATCAGCC TTCTAAGCTG AGGGTCGCAG GTTCGAGCCC TGCATGAGGC

            51                                                      100
HM1-IMSS    GCCTTTTTAT TCAATATACT CCTATACCTA TCACATCTTT ATACACTCTA
SAW760      GCCTTTTTAT CATCCATAT. .CTATAACTA TAACTATACC TATAACTATA
Laredo      GCCTTTTTTA CTTTATGGAT GTTATGTCTA CTATCTCTA. CTCGAATCTT
3646-MS15   GCCTTTTTTA CTTTATGGTT GTTATGTCTA CTATCTCTA. CTCGAATCTT

            101                                                     150
HM1-IMSS    TGTTTCTTAT ATGTATATTA CTTATACTAC TTATTATCTT ATATGTTTAT
SAW760      CCTATCCAAC ........TT TATACTCT.C TATGTTTCTT ATATGTATAT
Laredo      CGTGTCTTTT AT...TCTTT CTTTCTCTTC CTTTTCTTTT TTCCTCTTCA
3646-MS15   CGTGTCTTTT AT...TCTTT CTTTCTCTTC CTTTTCTTTA TTTCTCTTCA

            151                                                     200
HM1-IMSS    ATGTATATAT CACTATATGT TTATATGTTT ATATGTTTAT ATGTTTATAT
SAW760      TACTTATTTC TACTACTTAT TATCAACTAT ACACCTATCC TATACTATAT
Laredo      GAATTATTTT GAATATCTAT CTCTATATTT C.CTGCACTC GCTTTTATT.
3646-MS15   AAATTATTTT GAATATCTAT CTCTATTTTT C.CTGCACTC CCTTTTATT.
```

B) 3'-end R-R sequence alignment:

```
            1                                                        50
HM1-IMSS    .........T TTTATTATGT TCTTATGTTC TTATTTCTTT ATTTTATTAT
SAW760      ATTTTATTAT CCTATTATGT TATTATGTTA TTATGTTCTT ATTTTATTAT
Laredo      .......... .......... .......... ...CATTTTC ATCTTCTTAC
3646-MS15   .......... .......... .......... ...CATTTTC ATCTTCTTAC

            51                                                      100
HM1-IMSS    GTTCTTATGC TATTATTTCT TTATTTTATT ATGTTATTAT GTCCCTGTGA
SAW760      GTTCTTATGT TCTTATGTTA TTATGTTATT ATGTTATTAT GAGTGTGTTC
Laredo      TCTCTTATAC CTATTCCTTT CTTCCCCTCT CTCTTCCTTT CTTCTTTCTC
3646-MS15   TCTCTTATCC CTATTCCTTT CTTCC..... .......TTT CTTCTTTCTC

            101                                                     150
HM1-IMSS    AGATATCTTC ATCCCTTACC TATTTATACT ATAACCGATT GGCCTTGTTA
SAW760      A.ATATCATC ATCCCTTACC TATTTCTACT ATAACCGATT GGCCTTGTTA
Laredo      ACCTAGGATC GTACTCTTGC ATCTTTTTCT TTTATCGATT GGCCTTGTTA
3646-MS15   ACCTAGGATC GTACTCTTGC ATCTTTTTCT TTTATCGATT GGCCTTGTTA

            151         164
HM1-IMSS    GCTCAGTCGG AAG
SAW760      GCTCAGTCGG AAG
Laredo      GCTCAGTCGG AAG
3646-MS15   GCTCAGTCGG AAG
```

C)

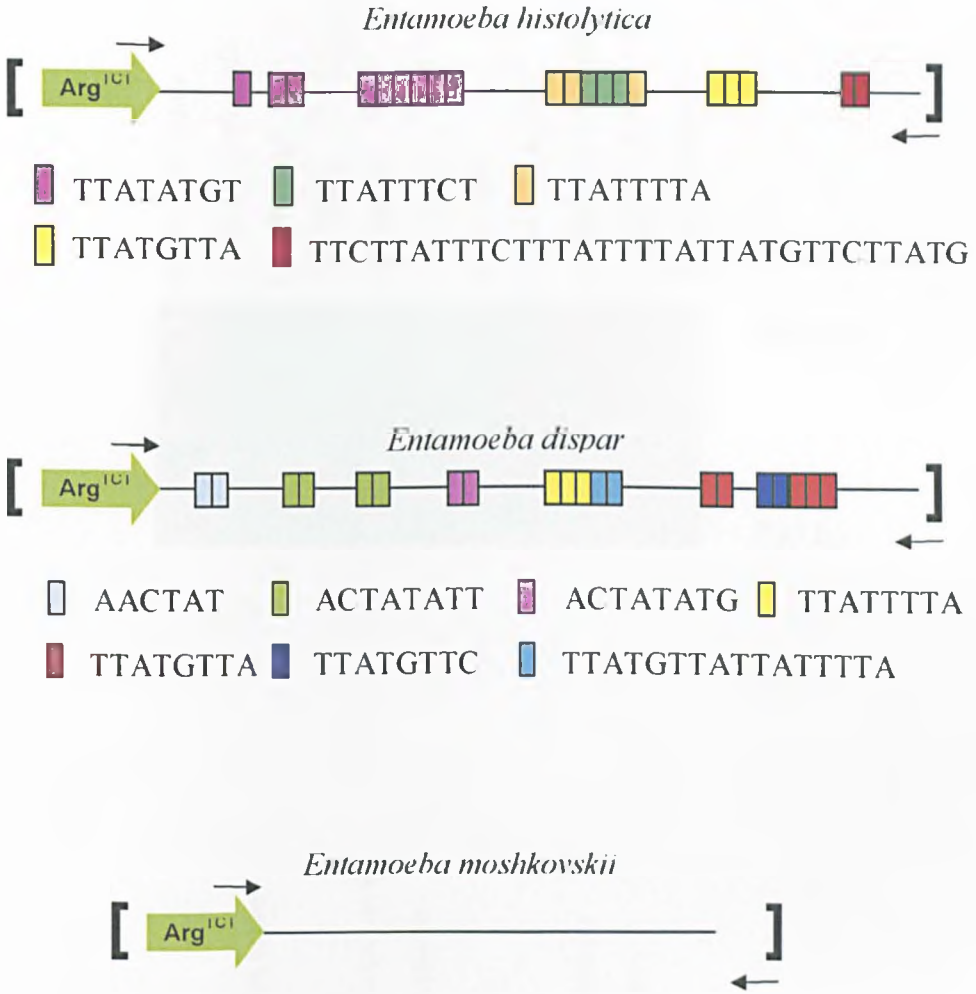
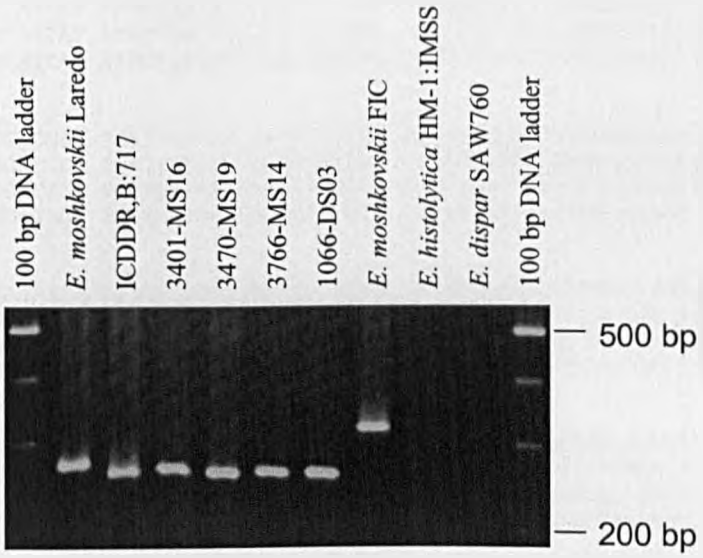


Figure 6.2. STR locus R-R. R-R sequences from *E. histolytica* HM-1:IMSS, *E. dispar* SAW760 (GenBank Accession number AF 525284), *E. moshkovskii* Laredo (GenBank Accession number AF 525285) and 3646-MS15 (GenBank Accession number AF525286) were aligned at the 5' (A) and 3' (B) ends to design *E. moshkovskii*-specific primers. The EmR primer sequences are shown in italic and bold with *E. moshkovskii* specific positions underlined. (C) Schematic representation of R-R from *E. histolytica* HM-1:IMSS, *E. dispar* SAW760, and *E. moshkovskii*. Locations of the primers used in PCR amplification are indicated by small arrows, the tRNA genes are indicated by large arrows and the short tandem repeats by shaded boxes. Not to scale.

(a)



(b)

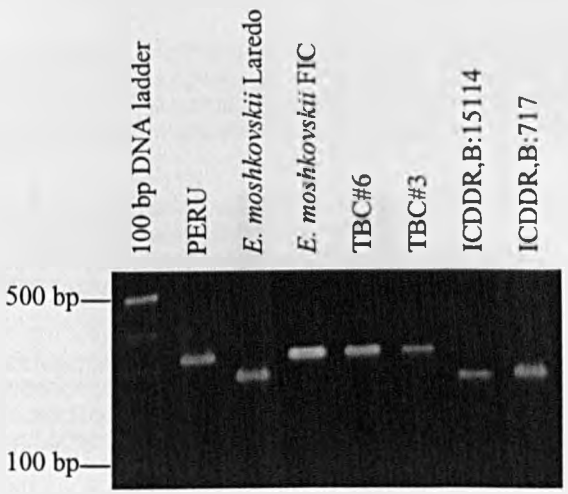


Figure 6.3. *E. moshkovskii* specific EmR PCR amplification. (a) Amplifications with a few Bangladeshi stool DNA samples and reference DNA samples. DNA samples from *E. histolytica* HM-1:IMSS and *E. dispar* SAW760 were used as negative controls. (b) Amplifications with culture DNA samples from Peru (PERU), Georgia (TBC#6 and TBC#3), Bangladesh (ICDDR,B:15114 and ICDDR,B:717) and two reference isolates of *E. moshkovskii* (Laredo and FIC).

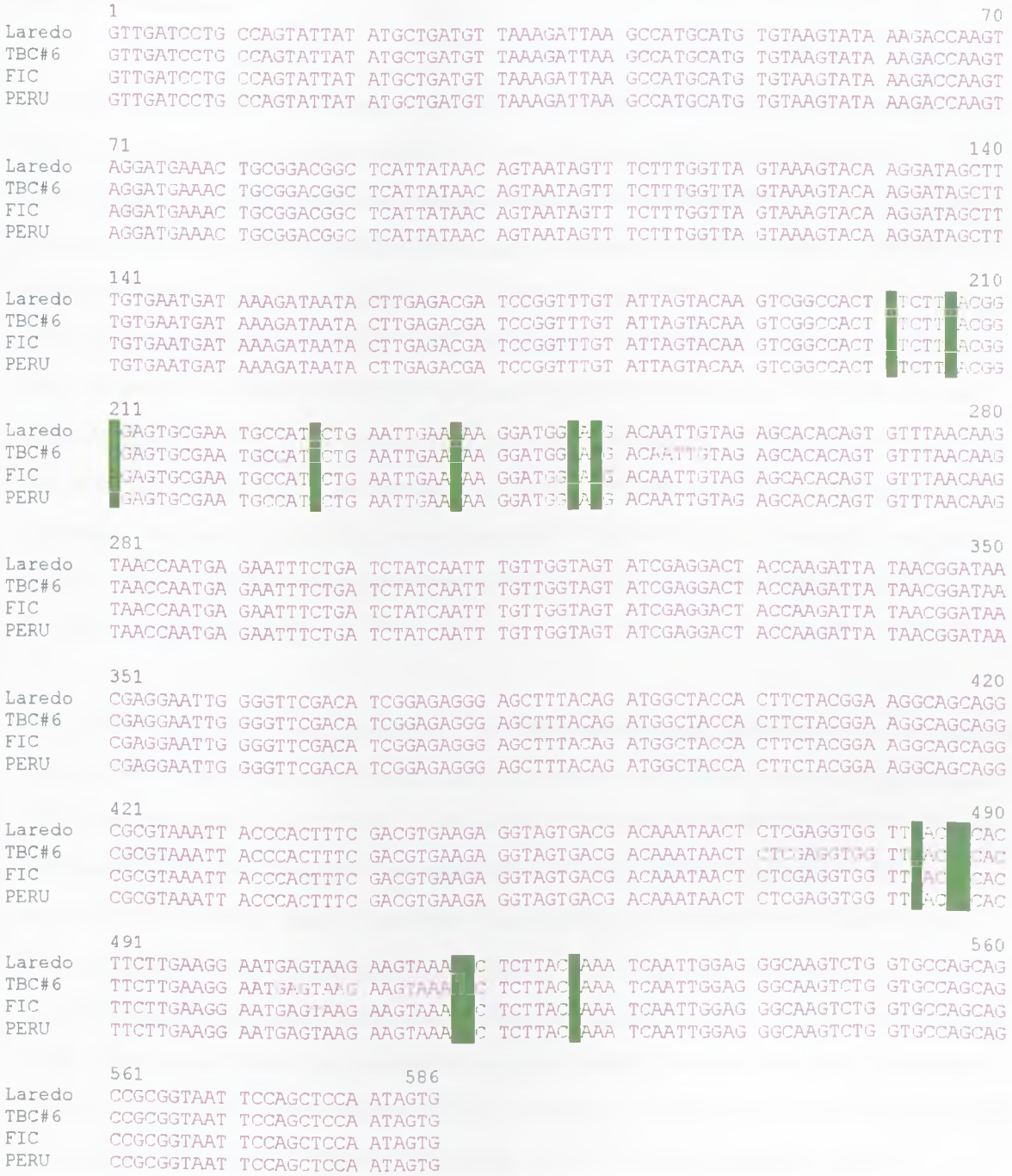


Figure 6.4. Alignment of rRNA gene (Entam 1/2 products) sequences from *E. moshkovskii* strains. Only one of the two Georgian sequences was shown here as representative of both as they were identical. All the nucleotide positions that showed differences are shaded.

6.6. Discussion

The main objectives of this study were to develop simple tools to identify *E. moshkovskii* and to investigate its prevalence and diversity in humans. We were successful in developing a simple diagnostic technique - a nested SSU-rDNA PCR with verification by restriction endonuclease digestion. We chose to use nested PCR to detect *E. moshkovskii* infections because previous experience in this study area showed it to be much more efficient in amplifying stool DNA (Ayeh-Kumi et al, 2001). Our attempt to produce a species specific polymorphic marker was not completely successful. The EmR primers failed to amplify 13 of 23 *E. moshkovskii* containing samples, probably because of sequence differences in primer binding sites. However, the R-R primers, which were originally designed to amplify *E. histolytica* and *E. dispar* DNA, did amplify a majority of the *E. moshkovskii* samples producing a product distinct in size from those of *E. histolytica* and *E. dispar.*

The present study has some limitations. Our Bangladeshi study subjects were 2-5 year old children, so we do not know whether they represent a true picture across all age groups. All previous human isolates of *E. moshkovskii* have belonged to ribodeme 2 (Clark and Diamond, 1991). Our attempts to perform riboprinting on these infections were unsuccessful, probably due to the size of the amplification target (ca 1.95 kb). Even if PCR had been successful, the presence of mixed infections with other eukaryotes would have prevented successful typing because the primers used amplify the SSU-rRNA gene of all eukaryotes. We did not attempt to sequence other *E. moshkovskii* R-R sequences simply because we did not find any STRs in sequences from *E. moshkovskii* Laredo and a Bangladeshi strain, and STRs are responsible for size variation in other species.

This study is unique and has several important findings. The overall *E. moshkovskii* prevalence in Mirpur (21%) suggests that *E. moshkovskii* infection is common among these children. It is also worth noting that *E. dispar* infected children were almost twice as likely to have a mixed infection with *E. moshkovskii* (35%) compared to those with (18%) or without *E. histolytica* (18%) infections. None of the 6 children with *E. moshkovskii* monoinfections had diarrhoea or dysentery which suggests that *E. moshkovskii* is a non-invasive parasite. It is possible that the high prevalence of *E.*

moshkovskii infection remained unnoticed over the years because most of them (74%) were mixed infections with *E. histolytica* and/or *E. dispar*. The high prevalence of *E. moshkovskii* revealed in this study population gives us an indication that perhaps humans are a true host for this putatively free-living amoeba, and are not just transiently infected. It may also explain some of the microscopy positive/antigen negative results obtained using the *Entamoeba* test kit (Haque et al, 1995). Likewise it may also explain the overdiagnosis of *E. histolytica/E. dispar* in Ethiopia using the microscopy (39%) compared to specific PCR (9%) detection (Kebede et al, 2003). Likewise it gives a possible reason for why these authors failed to detect 27.8% (30 of 108) of the *E. histolytica/E. dispar* microscopy positive samples using a specific real-time PCR in a separate study (Kebede et al, 2004).

It is difficult to interpret the results of *E. moshkovskii* infections in Georgian samples. This is because they were detected in culture samples, and these cultures were originally positive for *E. histolytica* (and 4 of the 5 *E. moshkovskii* positive cultures were still positive for *E. histolytica* as well). Laboratory personnel were unaware of the presence of *E. moshkovskii* in their cultures. Whether all 5 *E. moshkovskii* infections were real or just contamination from a single specimen is impossible to tell. It is unfortunate that we did not have the corresponding stool samples or stool DNA, which could provide better information on this. However, since we found a high mixed infection rate of *E. moshkovskii* in Bangladeshi children, we would not be surprised if all these Georgian infections were real. If these were indeed real infections, it is worth noting that these 5 *E. moshkovskii* were detected in 3 adults and 2 children, suggesting that infection is not only confined in children.

Based on the rRNA gene sequencing results, we found that the Georgian *E. moshkovskii* are of *E. moshkovskii* Laredo type (ribodeme II), which was previously found in all other human infections. In contrast, the Peruvian isolate was of *E. moshkovskii* FIC type (ribodeme III), which was so far believed to be a free-living environmental type of *E. moshkovskii*.

Based on locus R-R amplification and sequence results, we found that Bangladeshi *E. moshkovskii* resembled *E. moshkovskii* Laredo, while both Georgian and Peruvian isolates resembled *E. moshkovskii* FIC. This clearly indicates that the Laredo type of *E.*

moshkovskii (based on rRNA gene sequences) may have FIC type R-R sequence, and vice versa.

The prevalence and diversity of *E. moshkovskii* needs to be investigated in other parts of the world. In addition future epidemiological studies of *E. histolytica* infection should include tools to diagnose all three of these species individually, simultaneously and accurately.

CHAPTER 7

DNA SEQUENCE DIVERSITY IN *E. HISTOLYTICA* STRAINS

7.1. Introduction

With the development of new and simple techniques for DNA sequencing, the sequencing of a particular gene of interest or fragment of DNA of a pathogen of clinical importance is getting more and more popular. For example, DNA sequence diversity has been widely studied in malarial parasites (Da Silveira et al, 2001; Hoffmann et al, 2001; Leclerc et al, 2004; Rayner et al, 2004; Safitri et al, 2003; Tonon et al, 2004), in *Cryptosporidium parvum* (Gibbons-Matthews and Prescott, 2003; Strong et al, 2000), and in *Trypanosoma cruzi* (Machado and Ayala, 2002).

In Chapters 4 and 5 we investigated the PCR product based genotypes of *E. histolytica* strains from Bangladesh and other parts of the world and tried to establish a clinical association. However, very little is known about sequence diversity in this parasite. Ghosh et al, (2000) sequenced the serine rich *Entamoeba histolytica* protein (SREHP), chitinase, superoxide dismutase, and actin 3 genes in a few reference *E. histolytica* isolates. Ayeh-Kumi et al (2001) sequenced part of the SREHP gene in a few clinical isolates from Bangladesh. Zaki and Clark (2001) sequenced some of the reference isolates at tRNA linked STR loci 1-2 (now known as locus D-A) and 5-6 (now known as locus I-W). Haghighi et al (2002 and 2003) sequenced a number of clinical samples mainly from Japan, Thailand and Bangladesh at 4 loci: STR loci 1-2 and 5-6, SREHP, and chitinase. However, no association of sequence type with clinical outcome was detected.

In this Chapter our objectives were to investigate the sequence diversity of *E. histolytica* strains at two STR loci, namely R-R and S-D, in Bangladesh and also in a global context. We were particularly interested to see if there were any associations between particular sequence types and the clinical history of the strain. We were also interested to see whether some mutations were more common in particular types of strain than others.

7.2. Samples

We sequenced locus R-R in a total of 84 Bangladeshi clinical samples including 22 asymptomatic strains, 19 diarrhoeal or dysenteric strains, and 43 liver abscess strains and we sequenced locus S-D in a total of 80 Bangladeshi clinical samples including 19 asymptomatic strains, 17 diarrhoeal or dysenteric strains, and 44 liver abscess strains.

In addition we sequenced locus R-R in a total of 52 non-Bangladeshi clinical samples from 11 countries including 39 asymptomatic strains, 9 diarrhoeal or dysenteric strains, 2 liver abscess strains, and 2 samples for which the clinical information was not available.

We sequenced locus S-D in a total of 48 non-Bangladeshi clinical samples from 11 countries including 25 asymptomatic strains, 18 diarrhoeal or dysenteric strains, 1 liver abscess strain, and 4 samples for which the clinical information was not available.

7.3. Results

7.3.1. STR-based sequence types in locus R-R

Across all 136 Bangladeshi and non-Bangladeshi samples a total of 12 distinct R-R sequence types were detected. These contained either 5 (first 9 types) or 4 (types 10-12) distinct STR sequences (Figure 7.1). Four out of the 5 STRs were each 8 bases long and the 5th one (missing in the last 3 types) comprised 32 bases. The copy number in two STR blocks (Blocks 1 and 4) is constant across all samples (Figure 7.1). For the most variable STR block, the repeat copy number ranged from 3 to 7 (Block 2, Figure 7.1).

a) In Bangladesh:

Eleven out of the 12 sequence types were represented in Bangladeshi clinical samples; only sequence type 7 was not found (Figure 7.1 and Table 7.1). Five sequence types (1, 2, 3, 9, and 11) were represented by only one sample each, and all of these were from symptomatic patients. The three most common sequence types (5, 6 and 10) included

78.6% (66/84) of all samples distributed across clinical types - asymptomatic (A) = 86.4% (19/22), diarrhoeal/dysenteric (D) = 73.7% (14/19), and liver abscess (L) = 76.7% (33/43) (Table 7.1). While sequence type 6 was distributed reasonably equally across all clinical types – A = 31.8%, D = 26.3% and L = 23.3% - sequence type 5 was found in more A samples compared to D or L samples – A = 45.4%, D = 10.5% and L = 30.2%, and sequence type 10 was more common in D samples compared to A or L samples – A = 9.1%, D = 36.8% and L = 23.3%. Three sequence types (9, 11, and 12) were unique to Bangladeshi samples.

b) In other parts of the world:

Non-Bangladeshi samples included 9 sequence types (Table 7.1). Some sequence types were more common than the others. Five (1, 3, 7, 8, and 10) were represented by only one or two samples each. The two most common sequence types (5 and 6) included 67.3% (35/52) of all samples including 74.3% (29/39) of all asymptomatic samples. Since the numbers of other sample types – diarrhoeal/dysenteric (9/52) and liver abscess (2/52) - were low, relative prevalence numbers would be misleading. Only 1 sequence type (7) was unique to the non-Bangladeshi samples (seen in two Japanese individuals - one asymptomatic and one diarrhoeal patient).

7.3.2. Analysis of R-R sequence type among clinical groups

We analyzed whether the different sequence types show a statistically significant association among 3 clinical types of samples. Table 7.2 summarizes all the results. In Bangladeshi clinical samples we found only one statistically significant difference, which was between the asymptomatic and diarrhoeal/dysenteric groups (Table 7.2). This is due to sequence types #5 - found in 10 of 22 asymptomatic samples compared to only 2 of 19 diarrhoeal/dysenteric samples - and #10 – found in 7 of 19 diarrhoeal/dysenteric samples compared to only 2 of 22 asymptomatic samples.

In non-Bangladeshi samples we did not find any statistically significant association probably because of small sample sizes in diarrhoeal/dysenteric (9 samples) and liver abscess (2 samples) groups (Table 7.2).

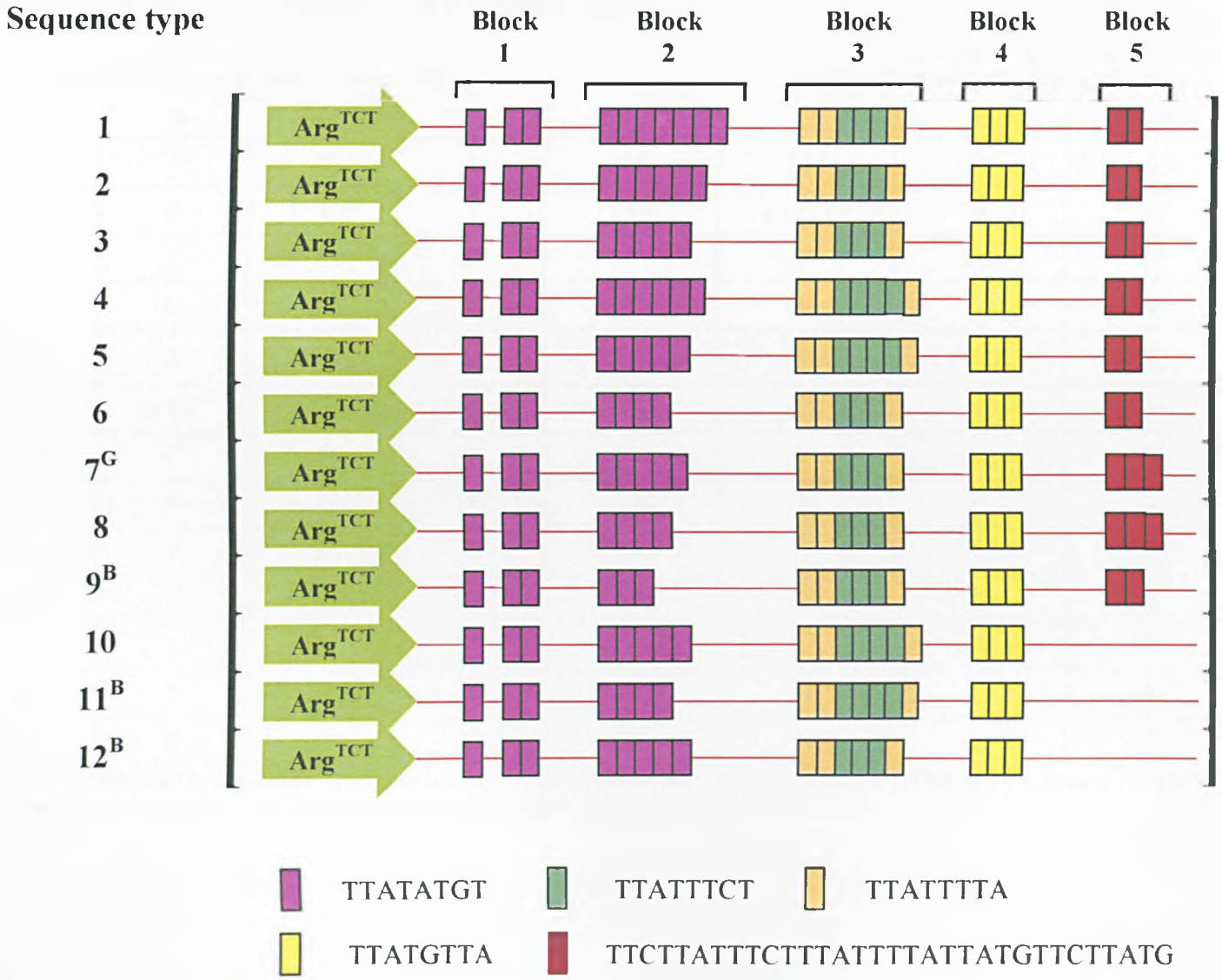


Figure 7.1. Schematic representation of locus R-R sequence types. In sequence types, superscript 'G' indicates that this type was found only in non-Bangladeshi samples and absent in Bangladeshi samples, and superscript 'B' indicates that these types were found only in Bangladeshi samples and absent in the others.

Table 7.1. Clinical and geographic distribution of R-R sequence types.

Seq. type	Distribution among Bangladeshi strains				Distribution among global strains				
	A	D	L	Total	A	D	L	N/A	Total
1	0	1	0	1	1 B	1 M	0	0	2
2	0	0	1	1	2 VT	1 VZ, 1 J	0	0	4
3	0	1	0	1	0	0	1 K	1 J	2
4	2	0	2	4	5 VT	0	0	0	5
5	10	2	13	25	1 I, 1 J, 1 GER, 4 SA, 5 VT	1 VZ	1 GER	0	14
6	7	5	10	22	10 SA, 7 G	4 T	0	0	21
7	0	0	0	0	0	0	0	1 J	1
8	1	0	4	5	1 B	0	0	0	1
9	0	0	1	1	0	0	0	0	0
10	2	7	10	19	1 J	1 J	0	0	2
11	0	1	0	1	0	0	0	0	0
12	0	2	2	4	0	0	0	0	0
Total	22	19	43	84	39	9	2	2	52

A = asymptomatic, D = diarrhoeal / dysenteric, L = liver abscess, N/A = not available, SA = South Africa, VT = Vietnam, J = Japan, T = Turkey, I = India, B = Brazil, GER = Germany, G = Georgia, VZ = Venezuela, M = Mexico, and K = Korea.

In combined samples we found statistically significant difference between asymptomatic and other two types of clinical samples. This is due to sequence types 5 and 6, which were more common in asymptomatic samples than in other two clinical types, and 10, which was more common in liver abscess and diarrhoeal/dysenteric samples than in asymptomatic samples (Table 7.1).

Table 7.2. Statistical analysis of R-R sequence type among clinical groups.

Sequences	Population pair	No. of samples	Chi^2	Overall P values
Bangladeshi	A & D	A = 22 D = 19	8.996	0.0111
	D & L	D = 19 L = 43	4.603	0.1001
	L & A	L = 43 A = 22	1.038	0.5951
Non-Bangladeshi	A & D	A = 39 D = 9	2.447	0.2941
	D & L	D = 9 L = 2	2.331	0.3118
	L & A	L = 2 A = 39	4.323	0.1152
Combined	A & D	A = 61 D = 28	18.230	0.0001
	D & L	D = 28 L = 45	4.644	0.0981
	L & A	L = 45 A = 61	7.628	0.0221

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

7.4. Mutations in R-R sequences

After aligning all the Bangladeshi and non-Bangladeshi R-R sequences using a multiple sequence alignment software (http://prodes.toulouse.inra.fr/multalin/multalin.html) (Corpet, 1988) we identified point mutations in comparison to a consensus sequence (Table 7.3; alignment not shown). These mutations were identified in 11 asymptomatic samples (9 from Bangladesh, 1 from Brazil, and 1 from Japan), 12 diarrhoeal/dysenteric samples (10 from Bangladesh, 1 from Turkey, and 1 from Japan), 15 liver abscess samples (14 from Bangladesh and 1 from Korea) (Table 7.3) and in one Japanese sample for which the clinical information was not available. A total of 94 mutations were detected with a distribution among clinical sample types: A = 19, D = 39, L = 34,

and N/A = 2. All these mutations seem to be base substitutions except for a deletion at position 403 which occurred in 3 samples, 1 from each of the 3 clinical types (Table 7.3 and Figure 7.2).

7.4.1. Mutations in terms of positions in R-R sequences

Table 7.3 shows all the positions where mutations could be detected in relation to the hypothetical consensus sequence (Figure 7.2). The 94 mutations were dispersed among 68 positions in the consensus sequence. The frequency of a mutation at a particular position ranged from 1 to 9. In fact, for 59 out of 68 positions, the mutation was observed in only one sample. At 4 positions mutations were observed in 2 samples and in 2 positions mutations were observed in 3 samples (Table 7.3).

Four out of 5 mutations at position 35 (substitution of G for A) occurred in asymptomatic samples (the 5th one was a diarrhoeal sample). Two particular positions were found to involve only liver abscess samples except one diarrhoeal/dysenteric sample - position 303 affecting 9 out of 15 ALA samples and position 570 affecting 6 out of 15 ALA samples (4 samples were in common).

7.4.2. Mutations in repetitive and non-repetitive sequences in R-R

In the hypothetical consensus sequence for locus R-R (Figure 7.2) we found that about 39.4% of the sequence (248 bases out of total 629 bases) is made up of the STRs and contained 21 of 68 mutation sites (30.9%) (Table 7.4). Of the two most significant positions, 303 and 570, where we observed 15 mutations almost exclusively in 11 liver abscess samples, one was located in the non-repetitive region (position 303) while the other located in a STR region (position 570). Similarly for the other mutations we did not find any association with their occurrence in repetitive or non-repetitive regions in the sequence (data not shown). Overall the mutations are spread quite evenly across the sequence.

7.4.3. Type of mutations in R-R sequences

In R-R, we observed base substitution mutations only, except at one position where a base deletion occurred in 3 samples (Table 7.3). DNA substitution mutations are of two types: transitions and transversions. Transitions are interchanges of purines (A↔G) or of pyrimdines (C↔T), which therefore involve bases of similar shape. Transversions are interchanges between purine and pyrmidine bases, which therefore involve exchange of one-ring and two-ring structures. Although there are twice as many possible transversions, transition mutations are more common than transversions.

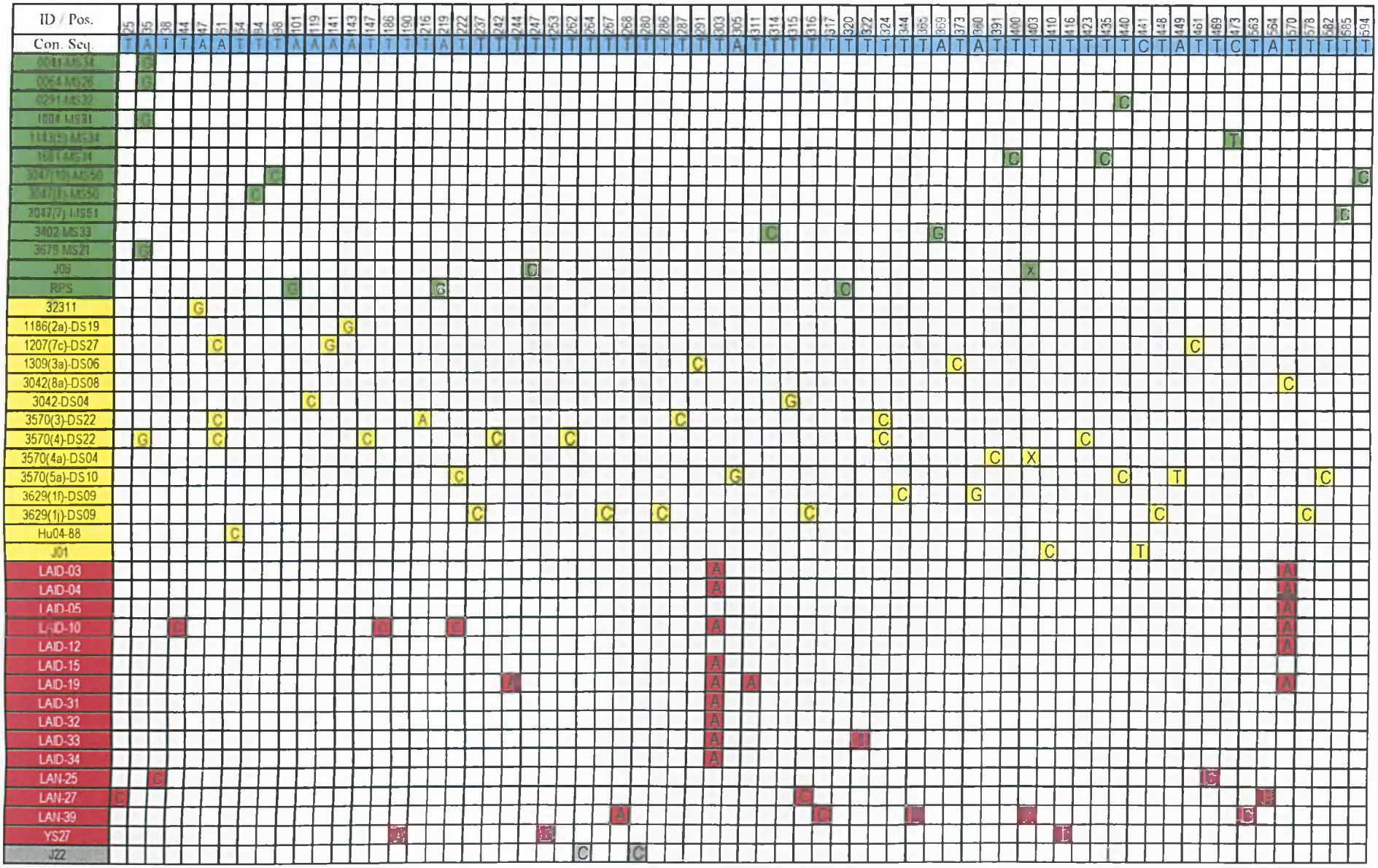
Although the predominant type of substitution mutation in asymptomatic (18 out of 18 or 100%) and diarrhoeal/dysenteric (32 out of 38 or 84.2%) samples is transition, the situation is exactly opposite in the liver abscess samples where transversion (16 out 33 or 57.6%) is the predominant type of mutation (Table 7.5).

7.4.4. Multiple clones sequenced in R-R

PCR products were sometimes cloned and sequenced for two reasons. Firstly, occasionally we experienced difficulties in obtaining clean sequences from direct sequencing of the PCR products. Secondly, we wanted to investigate the sequence variation among clones from same PCR products. In locus R-R, we sequenced 2 or more clones from 10 PCR products. Except for one sample (in two clones; data not shown), we detected sequence variation. A summary of sequence differences among clones is given in Table 7.6.

From the Table 7.6, we can see that in 5 out of 9 samples, the differences among cloned sequences were due to point mutations only at 1 (4 samples) or more (1 sample) nucleotide positions. Although the positions were variable, in all cases except one it was a 'C' substitution for 'T' (Table 7.3). In one (LAN-15) of the remaining 4 samples, the variation was due only to a variable number of block 5 repeats. The remaining 3 samples contained both point mutations (mainly 'C' substitution for 'T') and variation in repeat regions. No particular positions or STRs were involved at significantly higher frequencies.

Table 7.3. Positions of all mutations detected in R-R sequences. All the positions correspond to the hypothetical consensus sequence.



X = deletion

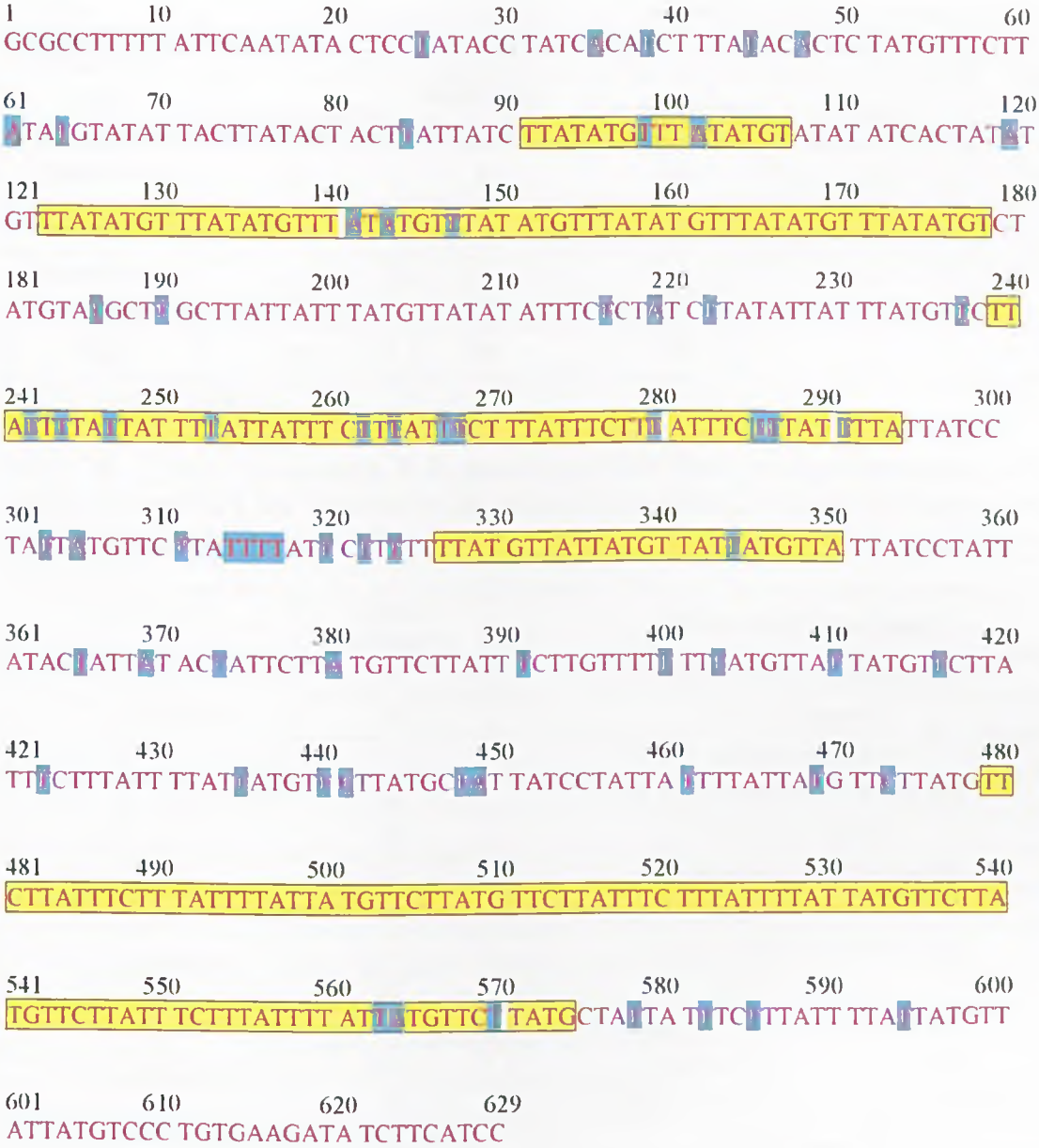


Figure 7.2. The hypothetical consensus R-R sequence with mutation locations. The positions of mutations are shown in highlighted with blue. Sequences inside yellow boxes are STRs. This hypothetical consensus sequence does not exist in any of our test isolates.

Table 7.4. Proportions of repetitive and non-repetitive sequences and positions of mutations in the hypothetical consensus R-R sequences.

Sequence type	Base pairs (%)	Number of mutation positions (%)
Non-repetitive	381 (60.6)	47 (69.1)
Repetitive	248 (39.4)	21 (30.9)
Total	629	68

Table 7.5. Type of mutations in R-R sequences.

Mutation type	Clinical types				Total
	Asymptomatic	Diarrhoeal / dysenteric	Liver abscess	*N/A	
Transitions	18	32	14	2	66
Transversions	0	6	19	0	25
Insertion / Deletion	1	1	1	0	3
Total	19	39	34	2	94

* N/A = not available.

Table 7.6. Type of variations in R-R cloned sequences from the same individuals. All mutation positions mentioned were relative to the consensus sequence. Only the mutations that differed between clones are mentioned.

ID	Clone identification	Differences between clones
1186-DS19	2a	A point mutation at position 143
	2b	No point mutation
	2c	No point mutation
3570-DS10	5a	A point mutation at position 440
	5b	No point mutation
3042-DS08	8a	A point mutation at position 570
	8c	No point mutation
1143-MS34	5	A point mutation at position 473
	9	No point mutation
3047-MS50	7	A point mutation at position 585
	8	A point mutation at position 84
	10	Two point mutations at positions 98 and 594
	12	No point mutation
LAN-15	4i	Two repeats in block 5 (Figure 7.1)
	4j	Three repeats in block 5 (Figure 7.1)
3629-DS09	1f	Point mutations at positions 344 and 380, and 7 repeats in block 2 and 5 repeats* in block 3
	1j	Point mutations at positions 237, 267, 286, 316, 448, and 578, and 5 repeats in block 2 and 6 repeats* in block 3.
3570-DS04	4a	A point mutation at position 391 and 6 repeats* in block 3
	4b	5 repeats* in block 3.
3570-DS22	3	Point mutations at positions 216 and 287 and 2 repeats in block 5
	4	Point mutations at positions 35, 147, 242, 262, and 423 and block 5 repeats were absent

*STR sequence: TTATTTCT

7.5. STR-based sequence types in locus S-D

Across all 128 samples a total of 14 distinct S-D sequence types were detected containing either 3 (first 8 types) or 2 (types 9-14) distinct STRs (Figure 7.3). Two of the STRs comprised 9 bases each while the 3rd comprised 8 bases and was present in a non-tandem fashion in some types (block 2 in Figure 7.3). The most variable STR type was repeated between 3 and 19 times (block 1 in Figure 7.3).

a) In Bangladesh:

Ten out of the 14 sequence types were represented in Bangladeshi clinical samples (Figure 7.3 and Table 7.7). Some sequence types are more common than the others. Five sequence types (1, 2, 3, 11, and 14) were represented by only one or two samples each, and all of these except one (11) were from symptomatic patients. The three most common sequence types, 7, 10 and 12, covered about 75.0% (60/80) of all samples with a fairly equal prevalence among clinical types - asymptomatic (A) = 84.2% (16/19), diarrhoeal/dysenteric (D) = 70.6% (12/17), and liver abscess (L) = 72.7% (32/44) (Table 7.7). Sequence types 6 and 12 seemed to occur in more L samples compared to A or D samples – A = 10.5%, D = 11.8% and L = 29.5%. A similar association for either A or D samples was not found.

b) In other parts of the world:

Non-Bangladeshi samples displayed 9 sequence types (Figure 7.3 and Table 7.7). Some sequence types were more common than the others. Five sequence types (4, 5, 6, 9, and 13) were represented by only one or two samples each. The two most common sequence types, 10 and 12, covered 70.8% (34/48) of all samples including 83.3% (15/18) of all D samples. Four sequence types (4, 8, 9, and 13) were unique to the non-Bangladeshi samples, and were not seen in any of the Bangladeshi samples.

7.5.1. Analysis of S-D sequence type among clinical groups

We analyzed whether the different sequence types show a statistically significant association among the 3 types of samples. Table 7.8 summarizes all the analysis results. In Bangladeshi clinical samples we did not find any statistically significant association (Table 7.8).

In non-Bangladeshi samples we found only one statistically significant difference, which was between asymptomatic and diarrhoeal/dysenteric groups (Table 7.8). This is due to sequence types 10 - found in 13 of 18 diarrhoeal/dysenteric samples compared to only 4 of 25 asymptomatic samples, and 12 – found in 10 of 25 asymptomatic samples compared to only 2 of 18 diarrhoeal/dysenteric samples. Since the number of sample in the liver abscess group was too small (only one) it was not possible to find any statistically significant association with this group (Table 7.8).

In the combined samples we noticed that the differences among clinical population pairs were borderline negative – suggesting an weak signal for the presence of distinctive sequence types among 3 groups (Table 7.8).

7.6. Mutations in S-D sequences

After aligning all the S-D sequences as before (Section 7.4) we identfied point mutations in comparison to a consensus sequence (Table 7.9; alignment not shown). These mutations were identified in 8 asymptomatic, 8 diarrhoeal/dysenteric, and 8 liver abscess samples, mostly from Bangladesh. In the diarrhoeal/dysenteric samples, 2 follow-up samples from a single patient (ID 3570) showed differences, while in another patient (ID 22027) 2 cloned sequences showed differences.

7.6.1. Mutations in terms of positions in S-D sequences

Table 7.9 shows all the positions of the mutations that could be detected in relation to the hypothetical consensus sequence (Figure 7.4). The 65 mutations were dispersed among 35 positions in the consensus sequence. The frequency of a mutation at a particular position ranged from 1 to 6. In fact, for 19 out of 35 positions, a mutation was observed in only one

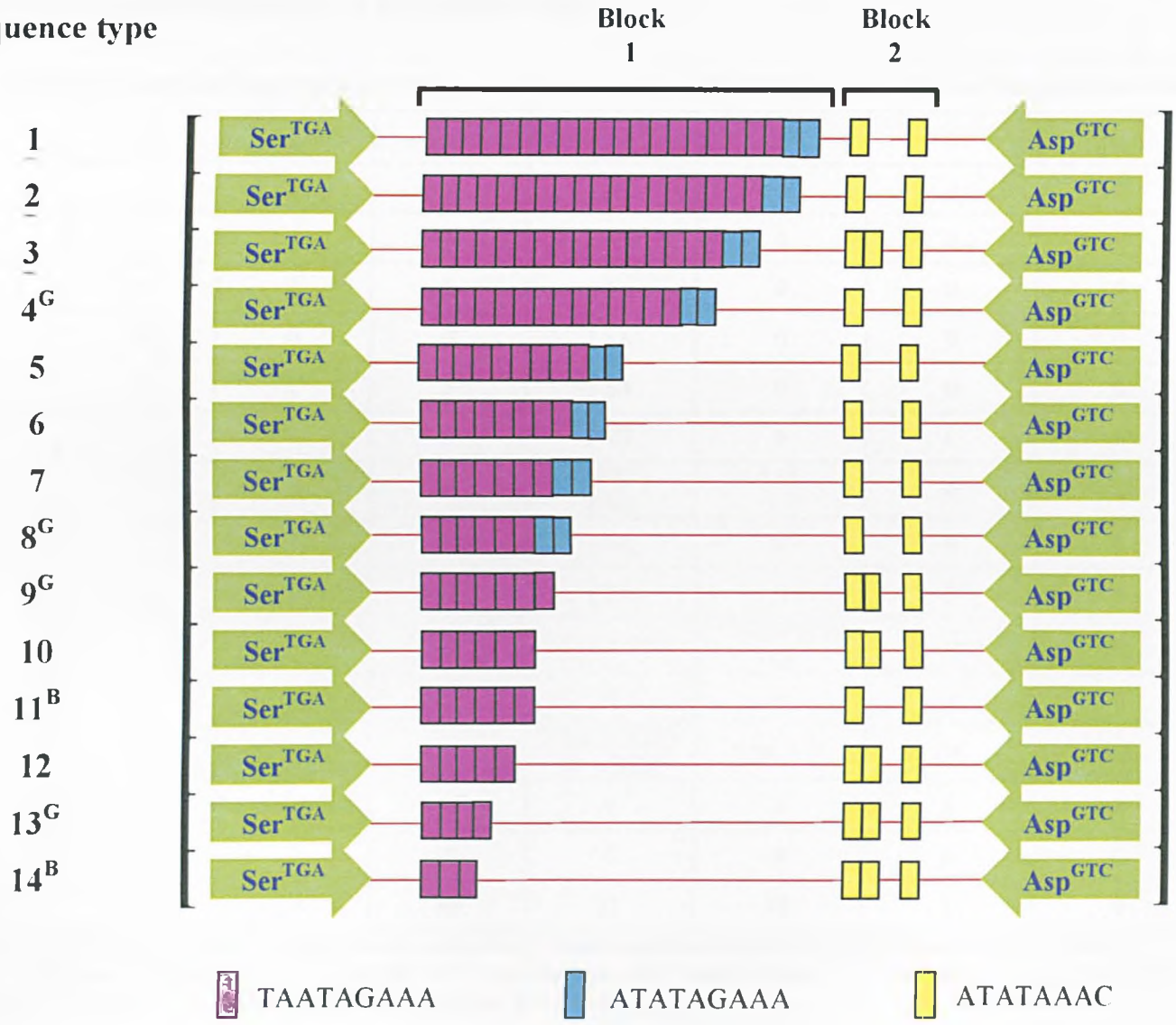


Figure 7.3. Schematic representation of locus S-D sequence types. *Superscript 'G' indicates that these types were found only in non-Bangladeshi samples, and superscript 'B' indicates that these types were found only in Bangladeshi samples.

Table 7.7. Clinical and geographic distribution of S-D sequence types.

Seq. type	Distribution among Bangladeshi strains				Distribution among non-Bangladeshi strains				
	A	D	L	Total	A	D	L	N/A	Total
1	0	1	1	2	0	0	0	0	0
2	0	0	1	1	0	0	0	0	0
3	0	0	1	1	0	0	0	0	0
4	0	0	0	0	1 SA	0	0	0	1
5	2	2	1	5	1 SA	0	0	0	1
6	0	1	7	8	2 VT	0	0	0	2
7	7	4	14	25	1 J, 1 SA	1 G, 1 J	0	0	4
8	0	0	0	0	3 SA	0	0	0	3
9	0	0	0	0	2 SA	0	0	0	2
10	7	7	12	26	1 T, 1 I, 2 SA	2 G, 9 T, 2 VZ	0	0	17
11	1	1	0	2	0	0	0	0	0
12	2	1	6	9	2 SA, 5 VT, 2 B, 1 GER	1 M, 1 J*	1 K	4 J	17
13	0	0	0	0	0	1	0	0	1
14	0	0	1	1	0	0	0	0	0
Total	19	17	44	80	25	18	1	4	48

NB. *This Japanese isolate (J7) was from the stool specimen of an amoebic dysenteric patient who also developed amoebic liver abscess.
A = asymptomatic, D = diarrhoeal / dysenteric, L = liver abscess, N/A = not available, SA = South Africa, VT = Vietnam, J = Japan, T = Turkey, I = India, B = Brazil, GER = Germany, G = Georgia, VZ = Venezuela, M = Mexico, and K = Korea.

Table 7.8. Statistical analysis of S-D sequence type among clinical groups.

Analysis	Population pair	No. of samples	Chi^2	Overall P values
Bangladesh	A & D	A = 19 D = 17	0.137	0.9340
	D & L	D = 17 L = 44	1.239	0.5382
	L & A	L = 44 A = 19	2.593	0.2735
Non-Bangladeshi	A & D	A = 25 D = 18	12.194	0.0023
	D & L	D = 18 L = 1	2.322	0.3132
	L & A	L = 1 A = 25	0.000	1.000
Combined	A & D	A = 44 D = 35	5.461	0.0652
	D & L	D = 35 L = 45	4.999	0.0821
	L & A	L = 45 A = 44	5.586	0.0612

A = Asymptomatic, D = diarrhoeal/dysenteric, L = liver abscess.

sample. At 2 positions mutations were observed in 5 samples and in one position mutations were observed in 6 samples (Table 7.9).

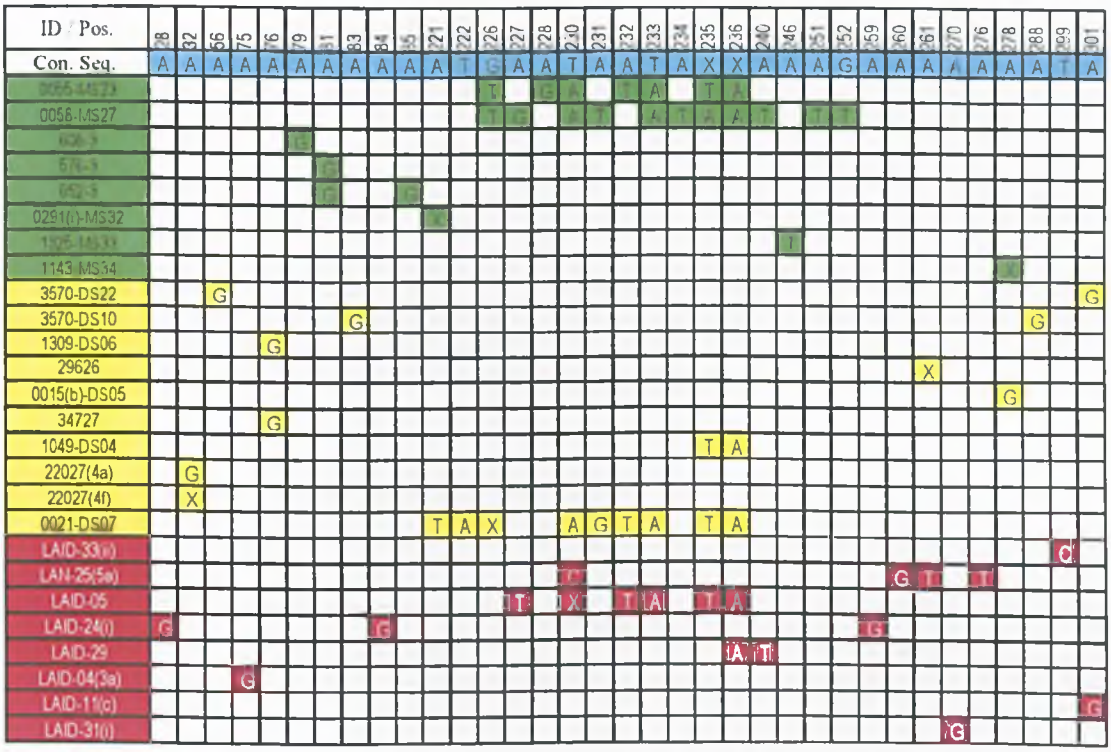
7.6.2. Mutations in repetitive and non-repetitive sequences in S-D

About 40% of the bases in hypothetical consensus sequence are in a non-repetitive region but these contained over 57% of all mutations (Table 7.10). However, we did not notice any clinical association of the distribution of mutations between repetitive and non-repetitive regions.

7.6.3. Type of mutations in S-D sequences

Unlike R-R, in S-D base insertion and deletion is relatively common – 17 out of 65 (26.2%) mutations are of this type. These included a total of 11 insertion mutations that occurred in two positions in 6 samples, and 6 separate deletions of a single base that occurred in 6 different samples (Tables 7.9 and 7.11). Although the most common type of

Table 7.9. Mutations detected in S-D sequences.



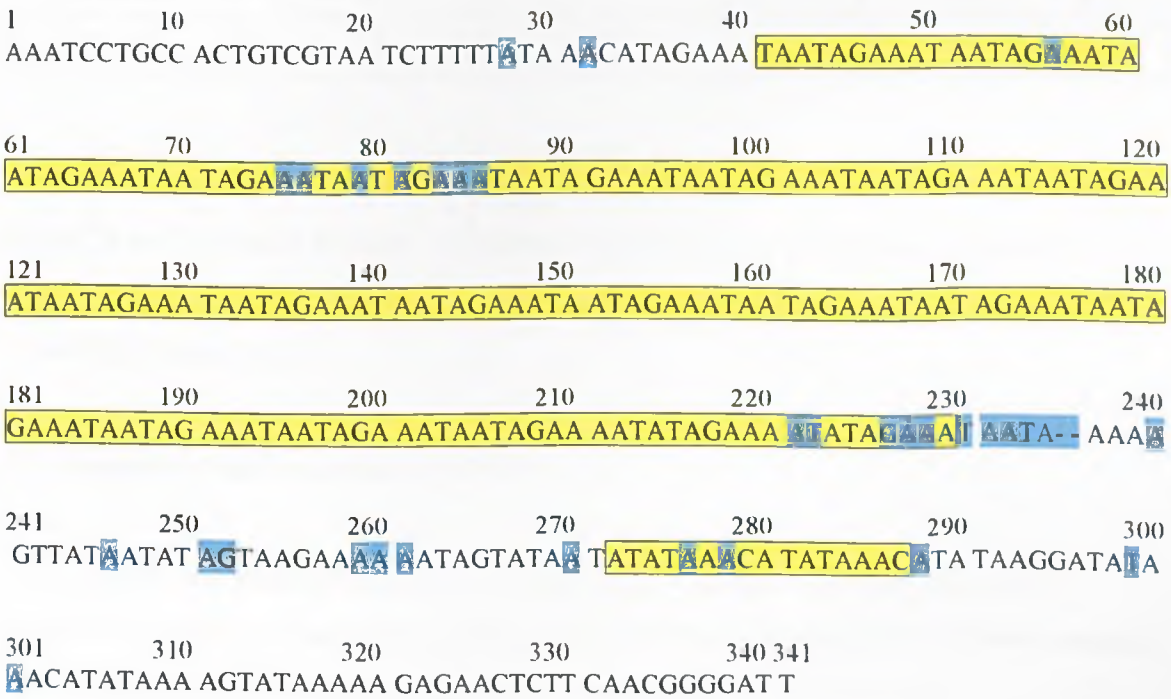


Figure 7.4. Hypothetical consensus sequence for S-D with mutation locations. Sequences inside yellow boxes are parts of STRs. The positions of mutations are shaded blue.

Table 7.10. Proportions of repetitive and non-repetitive DNA in S-D sequence with corresponding proportions of mutations.

Sequence type	Base pairs (%)	No. of mutations (%)
Non-repetitive	*136 (39.9)	**20 (57.1)
Repetitive	205 (60.1)	15 (42.9)
Total	341	35

NB. *The actual number of nucleotides in the non-repeat region was in fact 134 since in 2 positions there were no nucleotides. **The number of mutations considered in terms of positions involved. Insertion / deletion of two or more bases could have occurred at the same time, and therefore could be considered as a single event. Moreover, the frequency of mutations at a particular position was not considered.

Table 7.11. Type of mutations in S-D sequences.

Mutation type	Clinical types			Total
	Asymptomatic	Diarrhoeal/ dysenteric	Liver abscess	
Transitions	6	8	9	22
Transversions	13	6	6	25
Insertion	4	4	3	11
Deletion	2	3	1	6
Total	25	21	19	65

substitution mutation in diarrhoeal/dysenteric and liver abscess samples is a transition (8 out of 14 or 57.1% and 9 out of 15 or 60.0%, respectively), the situation is exactly opposite in the asymptomatic samples in which transversion is predominant (13 out of 19 or 68.4%) (Table 7.11).

7.7. Multiple clones sequenced in S-D

In locus S-D, we sequenced 2 or more clones from the same PCR products for 19 samples. We detected variation in one or more cloned sequences for 10 of these samples. A summary of sequence differences among clones is given in Table 7.12.

In 6 out of 10 samples, the differences between cloned sequences were due only to point mutations. Clones from 3 samples showed differences due to variable number

of STRs, while the remaining one showed differences due to both point mutations and copy number of a STR (Table 7.12).

Table 7.12. Type of variations in S-D cloned sequences from the same individuals. All mutation positions mentioned were in comparison with the consensus sequence position. Only the mutations that differed between clones are mentioned here.

ID	Clone identification	Differences between clones
22027	4a	A point mutation at position 32
	4e	No mutations
	4f	A point mutation at position 32 (but different from 4a)
	4g	No mutations
LAID-33	i	No mutations
	ii	A point mutation at position 299
LAID-11	a	No mutations
	b	No mutations
	c	A point mutation at position 301
0015-DS05	a	No mutations
	b	A point mutation at position 278
	c	No mutations
	d	No mutations
LAID-24	i	No mutations
	ii	3 point mutations at positions 28, 84 and 259
LAN-25	5a	Point mutations at positions 230, 260, 261 and 276
	5e	No mutations
	5g	No mutations
LAID-03	2c	9 repeats of type 1 STR* in block 1
	2d	8 repeats of type 1 STR* in block 1
	2e	9 repeats of type 1 STR* in block 1
	2f	9 repeats of type 1 STR* in block 1
LAN-33	a	9 repeats total in block 1 and 2 copies in block 2
	b	**7 repeats total in block 1 and 3 copies in block 2
LAID-04	3a	18 repeats of type 1 STR* in block 1 and a mutation at position 75
	3e	18 repeats of type 1 STR* in block 1
	3g	19 repeats of type 1 STR* in block 1
LAID-32	i	17 repeats of type 1 STR* in block 1
	ii	19 repeats of type 1 STR* in block 1

*STR sequence: TAATAGAAA, **Missing one of each STR type.

7.8. Discussion

Our main objectives in this study were: a) to investigate the nature of sequence diversity in two tRNA linked STR loci using samples of *E. histolytica* from Bangladesh and other parts of the world, b) to investigate any association of a particular sequence type with the clinical outcome of infection, and c) to investigate whether some mutations are more common in particular type of clinical samples. We were successful in achieving most of these. However, we had some limitations in this study. For examples, we had relatively few clinical samples from other parts of the world, 52 and 48 samples from 11 countries other than Bangladesh in locus R-R and locus S-D, respectively, an average of fewer than 5 samples from each country. In particular we had an extremely small number of non-Bangladeshi liver abscess samples (only 2 sequences for locus R-R and 1 for locus S-D). Moreover, clinical information for some of the Japanese samples was not available. We therefore were unable to find out whether an association between the outcome of an infection and a particular sequence type is the same in both Bangladesh and other parts of the world.

The reasons for selecting loci R-R and S-D for sequencing were because R-R was one of the largest and S-D one of the smallest of the 6 selected loci in our genotyping system. There are not many reports on sequencing of polymorphic DNA from isolates of *E. histolytica*. The only extensive study was carried out by Haghighi et al (2002 and 2003). They studied 74 samples mainly from Japan, Bangladesh and Thailand plus a few reference isolates. They sequenced two tRNA linked loci, namely locus D-A (formerly known as locus 1-2) and locus I-W (formerly known as locus 5-6), and two protein coding loci, chitinase and SREHP. In terms of polymorphism based on sequence types, they found that chitinase was the least polymorphic locus (9 types) while the SREHP was the most polymorphic (37 types). Locus D-A showed 13 types and locus I-W showed 15 types.

Total number of sequence types observed in our study was comparable to the above studies. We found 12 and 14 sequence types for locus R-R and locus S-D from 136 and 128 clinical samples, respectively.

Some of the main findings on sequence types are:

Locus R-R:

- In Bangladesh, 3 sequence types (9, 11 and 12) were found only in symptomatic samples (3 diarrhoeal/dysenteric and 3 liver abscess). These were not found in any other countries, suggesting that this is a unique Bangladeshi but relatively rare type that may be linked to symptomatic infection. However, since the number of samples with these types is small, we should be careful in over interpreting these data.
- Sequence types 5 and 6 seem to be common in asymptomatic samples worldwide. In Bangladesh 77.2% (17/22) and in other countries 74.4% (29/39) of the asymptomatic samples were of these types.
- Sequence type 10 seems to be common in symptomatic samples in Bangladesh. In 89.5% of cases (17/19) it was found in either diarrhoeal/dysenteric or liver abscess samples. Outside of Bangladesh, it was found only in Japan infecting 1 asymptomatic individual and one diarrhoeal/dysenteric patient (Table 7.1).
- Sequence type 1 which was found in reference isolate HM-1:IMSS from Mexico in a diarrhoeal/dysenteric patient, was also found in one of the two Brazilian samples (in an asymptomatic individual) and only one of the 84 Bangladeshi samples (in a diarrhoeal/dysenteric patient) suggesting that this type is rare but occurs worldwide.
- Three sequence types were found in 4 or more countries: type 2 (found in Bangladesh, Japan, Vietnam and Venezuela), type 5 (found in Bangladesh, India, Japan, Vietnam, Germany, South Africa and Venezuela), and type 6 (found in Bangladesh, Thailand, Germany and South Africa).

Locus S-D:

- Two sequence types were common in Bangladesh, type 7 and type 10. Only type 10 was also common in other countries, e.g., India, Thailand, Germany, Venezuela and South Africa infecting 17 individuals, while type 7, although found in

3 other countries (Japan, Georgia and South Africa), was found in only 4 individuals (Table 7.7).

- Type 12, which was the third most common type in Bangladesh (found in 9 samples) was also found in samples from 7 other countries (Japan, Korea, Vietnam, Germany, Venezuela, Brazil and South Africa). This is the most widely distributed S-D type in terms of the number of countries in which it was found.

Some findings based on mutation data:

- In locus R-R, mutations were detected in 33 Bangladeshi samples and in 5 samples from 4 other countries (Japan, Korea, Turkey and Brazil). A total of 94 mutations were detected at 68 positions of the hypothetical consensus R-R sequence.
- In locus S-D, mutations were detected in 22 Bangladeshi samples and in other 3 samples, all from Vietnam. A total of 65 mutations were detected at 35 positions of the consensus sequence.
- The length of the hypothetical consensus sequences in R-R is 629 bp and in S-D is 341. So, in terms of total number of mutations detected in two loci, 94 in locus R-R and 65 in S-D, it seems that S-D is more susceptible to mutation than R-R.
- In locus R-R, two different mutations seemed to be linked with liver abscess samples. Fifteen point mutations affecting positions 303 and 570 (with relation to the hypothetical consensus sequence) were found only in liver abscess samples, except for one diarrhoeal sample (Table 7.1). Many of these were A/T transversions which occurred 19 out of 21 times in liver abscess samples (Table 7.5). However, only 11 out of 45 liver abscess samples showed these mutations overall.
- In S-D, the situation is different. Transversions are more common in asymptomatic samples (68.4%) compared to diarrhoeal/dysenteric (42.9%) and liver abscess (40.0%) samples. Overall no strong mutational links to any of the clinical sample type was observed in locus S-D sequences.
- Insertion or deletion mutations appeared to be more common in S-D sequences than in R-R: 26.2% (17 out of 65 mutations) of all mutations in S-D were of this type compared to only 3.2% (3 out of 94 mutations) in R-R (Tables 7.4 and 7.9).

We cloned some of the PCR products into vectors and then sequenced them. We did this for two reasons. One, we wanted to know if there were any differences in the clones coming from the same PCR products, and two, direct sequencing of some of the PCR products was not possible for a few samples, as the sequences could not be read in the chromatogram (data not shown). In locus R-R, variation among cloned sequences was observed in 9 out of 10 samples. However, variation in cloned sequences was relatively less common in S-D sequences as only 10 out of 19 samples showed variations among the clones. Both point mutations and STR number variation were observed in both loci.

It was very surprising to find such a high degree of variation between cloned sequences from the same PCR product. The difference in sizes of some clones was sufficient to be detected in the gel electrophoresis, however, we never noticed additional bands, suggesting that their relative frequencies in the genome are low and we observed only the major products in the gel. It is unlikely that the mutations were introduced by Taq polymerase during PCR amplification, since the maximum sizes of these loci were only about 630 bases for R-R and 340 bases for S-D. Moreover, when we repeated the sequencing we observed the same results. Point mutations and size differences were also evident in raw genome sequences for tRNA-linked array units from isolate HM-1:IMSS, which were generated directly from genomic DNA and no amplification was involved. For example the STR regions in the arrays [YE], [VF] and [WI] show length variation, although the sequences are very closely related (Clark et al, 2005). However, the significance of this clonal difference is not clear.

In summary, we found some evidence that certain sequence types were more frequent in certain sample types. Likewise, certain mutations also showed a link to certain sample types. Overall, however, no clear predictive value could be attached to any particular feature of the sequences generated here.

As expected the PCR product size variation observed by gel electrophoresis was largely due to variation in STR number. However not all PCR products of the same size had the same sequence and not all STRs showed variation in number. The factors that determine the length and organization of these loci are still not clear. We believe

that sequencing of more clinical samples from other parts of the world other than Bangladesh may confirm some of these findings.

CHAPTER 8

DETERMINATION OF COPY NUMBER OF STR LOCI IN THE *E. HISTOLYTICA* GENOME

8.1. Introduction

Repetitive DNA sequences have been utilised in the study of *E. histolytica* polymorphism for about the last two decades. Some of these DNAs code for proteins and some do not. The most polymorphic locus in *E. histolytica* is the serine rich *Entamoeba histolytica* protein (SREHP) gene (Stanley et al, 1990; Kohler and Tannich, 1993; Clark and Diamond, 1993; Ghosh et al, 2000; Ayeh-Kumi et al, 2001; Haghighi et al, 2002 and 2003). The chitinase gene (de la Vega et al, 1997a,b; and Haghighi et al, 2002 and 2003) is also polymorphic. Non-coding DNA sequences, for example, the strain specific gene (SSG) (Burch et al, 1991), also known as *Tr* (Sehgal et al, 1994), have also been used to study polymorphism in *E. histolytica* (Clark and Diamond, 1993). The repeated nature of the DNA in all these loci has been well characterised, and the number of these genes in the genome has been established. SREHP and chitinase are both single copy chromosomal genes while the extra-chromosomal SSG is maintained on the ribosomal episome at about 200 copies per genome.

We have been investigating the polymorphism of STRs linked to tRNA genes in *E. histolytica* for the study of genetic diversity (Chapters 4 and 5). We have also investigated in detail the sequences of two tRNA linked STRs, namely R-R and S-D (Chapter 7). However the copy number of these loci in the genome is still unknown.

From the raw sequence data from the *E. histolytica* genome sequencing it was impossible to estimate accurately the copy number of these tRNA arrays in the genome. This is mainly because the two genome sequencing projects (The Institute of Genome Research, USA and The Wellcome Trust Sanger Institute, UK) used 4 different types of libraries, 2 with short inserts of about 2 kb each and 2 with relatively larger inserts of about 8 kb each. The cloning vectors were also propagated in different bacterial hosts. As a result, although tRNA containing sequences were abundant in all cases, their relative numbers differed significantly between these different libraries. Therefore we

have undertaken a small scale investigation to estimate the copy numbers of these tRNA loci in the *E. histolytica* HM-1:IMSS genome.

8.2. Materials and methods

We used a quantitative hybridisation method to determine the copy number of a selection of the tRNA arrays. The principle steps involved in this experiment are outlined in Figure 8.1 and are described briefly below.

8.2.1. Amplification of STR locus to be estimated

The STR locus for copy number estimation was first amplified using species-specific primers, where possible, or tRNA primers. The amplified products were then purified to get rid of primers using the Qiagen PCR purification kit according to the manufacturer's instructions.

8.2.2. Quantification of amplified products

We used Hyperladder II (Promega, UK) as a standard to quantify the PCR products (since its fragment sizes and the corresponding amounts of DNA were known) after running them together in a 1.5% agarose gel (Figure 8.2). The levels of fluorescence produced by the PCR products were compared to those produced by the fragments of the standard DNA using the Phoretix software (Molecular Dynamics/Amersham Bioscience).

8.2.3. Counting of *E. histolytica* trophozoites and isolation of DNA

E. histolytica isolate HM-1:IMSS was the source of genomic DNA. Before extracting the genomic DNA, the *E. histolytica* trophozoites were counted using a haemocytometer. The genomic DNA was isolated using the CTAB method (Section 2.2.1).

8.2.4. Blotting of PCR products and genomic DNA

A range of dilutions of both PCR products (10^6 – 10^4 copies) and genomic DNA (10^6 – $5x10^4$ cells) were boiled and then fixed to a nylon membrane using a slot blot apparatus. In addition, 10^7 copies of PCR products representing all STR regions were included on a separate control membrane to evaluate cross-hybridisation.

8.2.5. Hybridisation with STR locus specific probes

The membranes were hybridised with a probe prepared using the corresponding locus-specific PCR product. This was produced either using the corresponding tRNA primers or, where possible, using species specific primers in order to minimise the inclusion of tRNA gene regions; this is because certain tRNA genes show a significant degree of sequence similarity.

8.2.6. Exposure to PhosphorImager screen

After hybridisation with a ^{32}P labelled specific probe, the membranes were exposed to a PhosphorImager screen and the results were quantified using ImageQuant 5.0 software (Amersham Biosciences).

8.3. Results

8.3.1. Quantification of amplified products

Pooled PCR products of different sizes were separated in a gel containing different amounts of the Hyperladder II standard (Figure 8.2). An example of the calculated copy number of PCR products is given in the Table 8.1.

8.3.2. Cross hybridisation

In addition to blotting a known amount of PCR product and genomic DNA, we also included all STR locus-specific PCR products on a separate membrane (including the test locus), approximately 10^7 copies of each, to evaluate any cross hybridisation. The

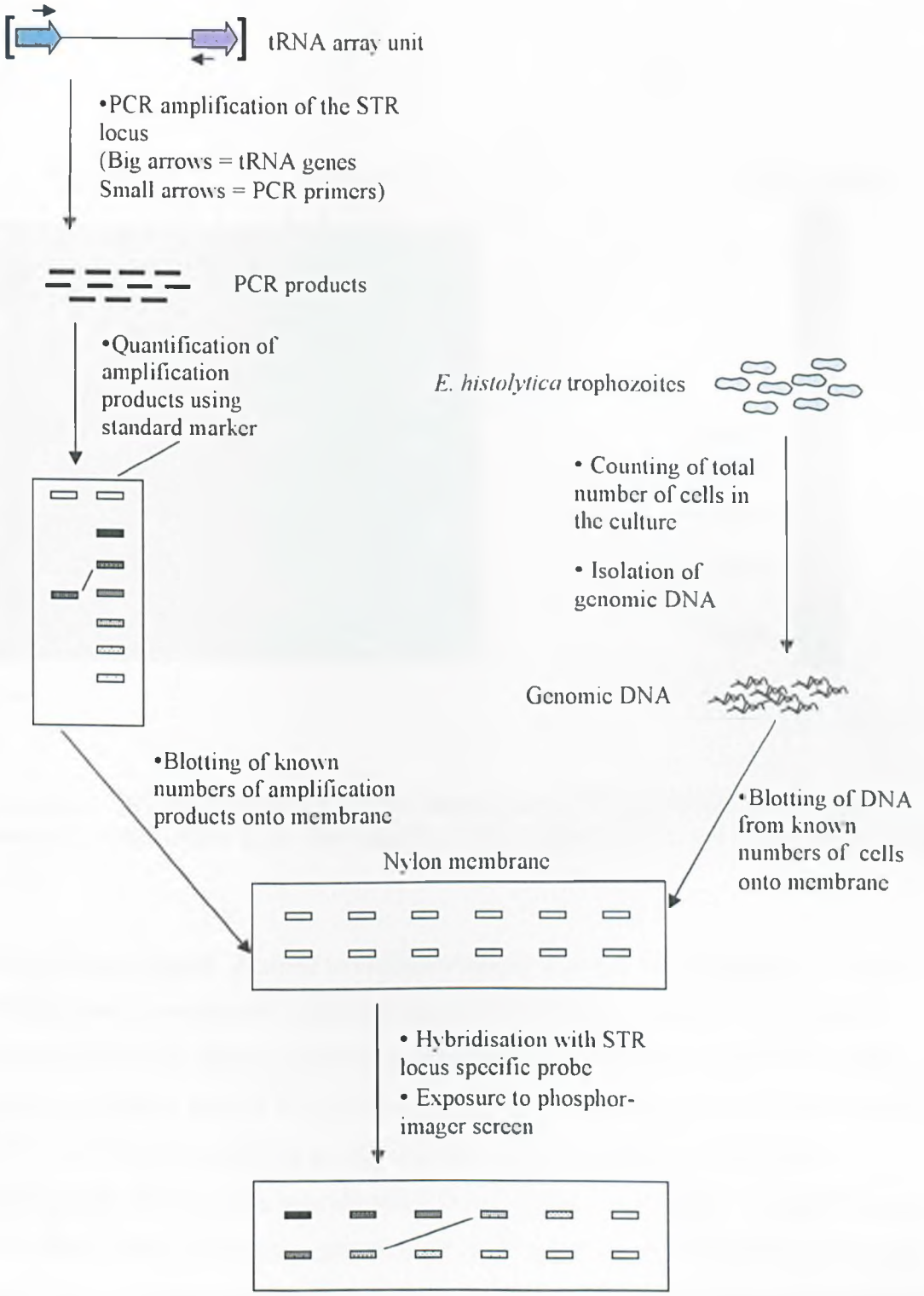


Figure 8.1. Principal steps involved in the determination of STR locus copy number.

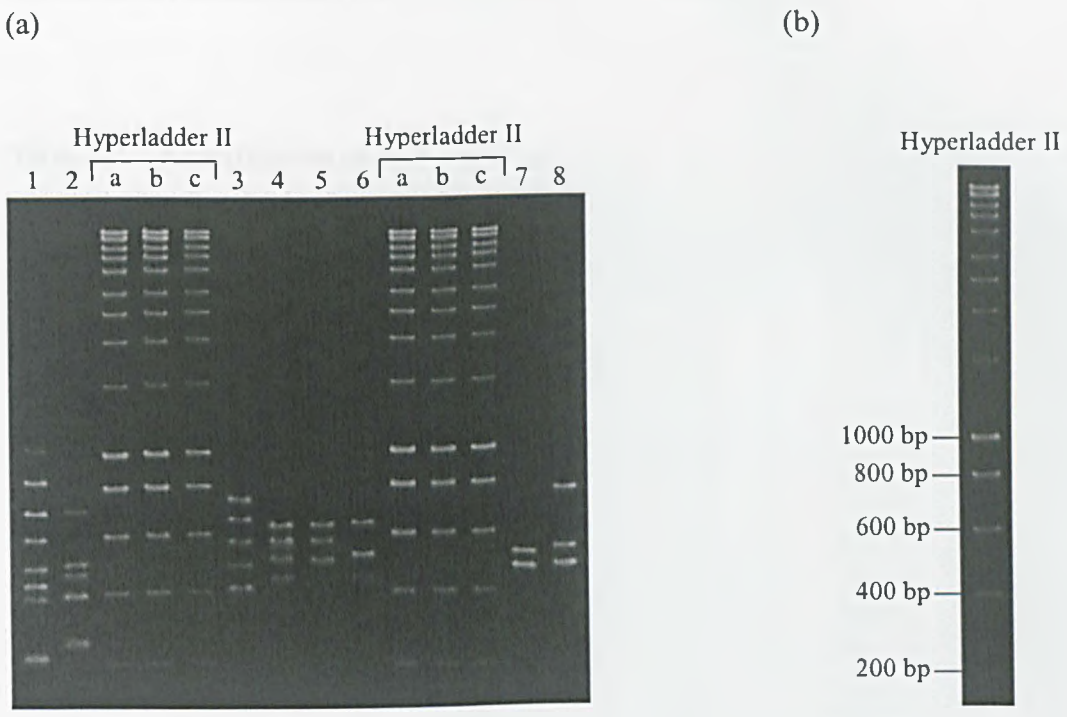


Figure 8.2. Gel electrophoresis of STR-locus specific PCR products (a) with the standard, Hyperladder II (b). The specific PCR products shown are identified in Table 8.1.

hybridisation signals obtained in PhosphorImager analysis were measured as relative (%) signals in comparison to the test locus. For example, using the *E. histolytica* species-specific S-D probe on the control membrane, we observed that 80% of the signal was due to locus S-D hybridisation and the remaining 20% was due to locus D-S^{TGA} (10.4%) and locus D-A (9.6%) products. In this case there is an obvious explanations for the cross hybridisation. D-S^{TGA} gave 10.4% signal because the same two tRNA genes are present - both loci (D-S^{TGA} and S-D) are derived from the same array, [SD]. On the other hand, although S-D and D-A are found in two different arrays ([SD] and [ASD], respectively) the Asp tRNA genes are identical in sequence and this is probably the basis for the 9.6% of cross hybridization signal seen. However, in many cases, we could not find any obvious explanations for the sometimes high degree of cross hybridisation observed at other loci (data not shown). We tolerated up to 30% of

cross hybridisation for any particular locus and the Table 8.3 shows the cross-hybridisation values observed for some loci.

Table 8.1. Quantification of PCR products.

Gel lanes	STR locus	Neat volume applied (µl)	Amount (ng)	PCR product conc. (ng/µl)	PCR product size (bp)	Copy numbers per µl of PCR product (x 10^{11})
1	R-5	1.5	69	46	838	0.49
1	R-R	1	89	89	686	1.18
1	E-V	2	66	33	581	0.52
1	S-L	3	61	20.3	480	0.38
1	*L-S	1	68	22.6	426	0.48
1	T-L	1.5	53	35.3	373	0.86
1	Q-C	1	86	86	222	3.52
2	I-W	5	28	5.6	702	0.07
2	V-F	1	52	52	492	0.96
2	G-G^{TCC}	1	45	45	461	0.89
2	S-Q	1	71	71	392	1.64
2	S^{TGA}-D	1	64	64	257	2.26
3	P-P	0.5	51	102	738	1.25
3	K-N2	0.5	45	90	642	1.27
3	L-T	1.5	48	32	561	0.52
3	T-Q	2	23	11.5	473	0.22
3	D-A	1	48	48	402	1.08
4	H-H	1	52	52	631	0.75
4	R-T	1	72	72	559	1.17
4	E-Y	1	46	46	492	0.85
4	N-K1	3	27	9	422	0.19
5	T-X	0.5	48	96	620	1.40
5	A-A	0.5	45	90	559	1.46
5	S-A	0.5	42	84	490	1.56
6	M-R	0.5	50	100	610	1.49
6	V-5	1.5	55	36.6	510	0.65
6	T-R	5	28	5.6	441	0.12
7	*L-A	0.5	67	134	522	2.31
7	S-P	0.5	76	152	473	2.92
8	G-G^{GCC}	1	97	97	810	1.08
8	D-S	1	99	99	550	1.63
8	N-K2	1	98	98	489	1.82

NB. The formula used for the calculation of copy number of PCR product was:
1 µg of a 1000 bp DNA = 9.1 x 10^{11} molecules.
* The amplification was done using species specific primers.

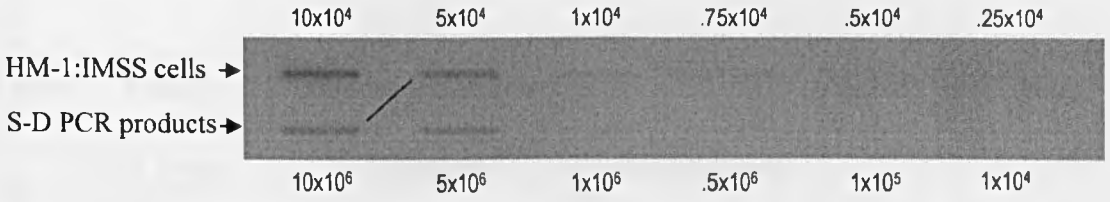


Figure 8.3. Signals from PhosphorImager screen for genomic DNA and locus S-D PCR products. The inserted line indicates the similar intensity bands.

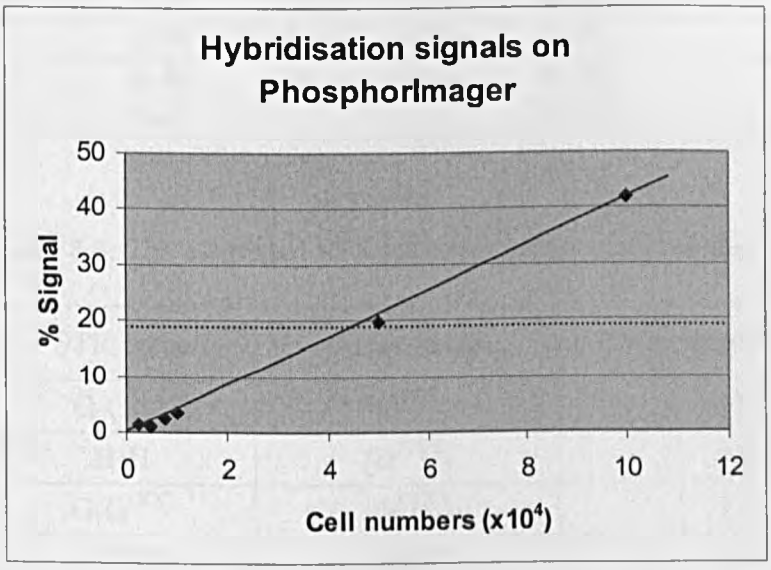


Figure 8.4. Relative (%) hybridisation signals on PhosphorImager with locus S-D probe. The signals from genomic DNA from various cell numbers are showing linear correlation except for two lowest numbers of cells. The horizontal dotted line indicates the signal from PCR products used for the estimation of copy number of the STR locus S-D (18.7%, Figure 8.3), which fell in the linear range of the signals produced by cells.

Table 8.2. Relative (%) signals due to genomic DNAs from HM-1:IMSS cells and locus S-D PCR products at various dilutions. All signals should add up 100%.

10×10^4 cells* 41.8%	5×10^4 cells 19.8%	1×10^4 cells 3.8%	0.75×10^4 cells 2.7%	0.50×10^4 cells 1.1%	0.25×10^4 cells 1.0%
S-D 10×10^6 copies 18.7%	S-D 5×10^6 copies 8.8%	S-D 1×10^6 copies 1.7%	S-D 0.5×10^6 copies 0.3%	S-D 1×10^5 copies 0.3%	S-D 1×10^4 copies 0.0%

*Indicates the number cells from which genomic DNA was isolated.

Table 8.3. Relative (%) signals on PhosphorImager screen due to cross hybridisation.

STR locus	Signal due to cross-hybridisation
$G\text{-}G^{TCC}$	7.7%
H-H	15.2%
$G\text{-}G^{GCC}$	30.0%
S-D	20.0%
T-Q	28.0%
R-T	20.0%
V-F	2.0%
A-A	30.0%
L-S	30.0%
I-W	5.0%
R-R	28.0%
S-Q	25.0%

Table 8.4. The estimated STR locus (or array) copy number.

STR locus	Array name	Copy number per cell
$G\text{-}G^{TCC}$	$[G^{TCC}]$	134
H-H	$[H^{GTG}]$	126
$G\text{-}G^{GCC}$	$[G^{GCC}]$	120
S^{TGA}-D	[SD]	264
Q-T	[TQ]	125
T-R	[RT]	80
V-F	[VF]	110
A-A	$[A^{AGC}]$	70
L-S	[LS]	81
W-I	[WI]	209
R-R	$[R^{TCT}]$	92
S-Q	[SQCK]	147

8.4. The estimation of copy number of the arrays

An example of the relative hybridisation signals for genomic DNA originally isolated from a variable but known numbers of cells are plotted in Figure 8.4. Except for the signals from low number of cells (e.g., 0.25×10^4 and 0.50×10^4 cells), cell numbers

gave rise to linear hybridisation signals. From the relative (%) hybridisation signals due to i) genomic DNA, ii) STR locus-derived PCR product, and iii) cross-hybridisation on the control membrane, we estimated the copy number of the STR locus (or array) in the genome. For an example, for the STR locus S-D, the 80% specific signal (because the cross-hybridisation was 20%, Table 8.3) shows that 5×10^4 cells contain approximately 10×10^6 copies of S-D (Figure 8.3). Taking into account only the specific signal, each cell contains 264 copies of S-D. So the copy number of STR locus S-D (which is same as array [S-D]) is approximately 264 per genome (Table 8.4). The Table 8.4 shows the results of the estimated copy numbers for a selection of loci.

8.5. Discussion

The overall prevalence of tRNA containing sequences as well as of a particular array in the two *E. histolytica* sequencing projects differ significantly (data not shown). This is likely to be because different sized inserts and different vectors were used. In addition, the cloned fragments were also propagated in different bacterial hosts. So the variation in prevalence may be due to variable stability of the array units in the different libraries. We have independently seen this when we cloned part of the [SD] array from different strains as PCR products and noticed that clones in only one orientation were ever recovered. So it is impossible to estimate the actual copy number of these arrays from the genome sequence data. Therefore it was decided to investigate the true number of these array units in the *E. histolytica* genome.

There were several limitations in our experimental procedures. One, during the genomic DNA isolation there will be some loss of DNA, but it is impossible to know the exact amount of this loss. Two, in the cross hybridisation control we included all the STR regions adjacent to tRNA genes. However, any cross hybridisation outside of these STR regions, although highly unlikely, was not considered. Three, not all the probes were produced using species-specific primers; this was simply because we did not have species-specific primers available for all STR loci. Otherwise, we might have prevented some undesired cross hybridisation due to sequence similarities among tRNA genes. However it is worth mentioning that most of this cross hybridisation was random, no sequence relevance was detected (data not shown).

Nevertheless, we have some evidence that our experimental results are reliable. Firstly, for one array ([VF]) the predicted array length based on pulsed-field gel electrophoresis using a frequent cutting restriction endonuclease that does not cut in the array unit (Zaki, unpublished) matches well with the copy number estimated by this quantitative hybridization (100 copies versus 110 copies, respectively). Secondly, taking an average of about 130 copies (Table 8.4) and an average array unit length of 975 bp (Clark et al, 2005) for the 25 arrays gives a value of 12.1% of the estimated 26.13 Mb genome (including tRNA-linked sequences) being made up of the tRNA arrays, which is close to their percentage representation in the total raw sequence data, which is just under 11% (B. Loftus, personal communication and Loftus et al, 2005).

Based on the available experimental data the projected copy number of tRNA genes in the *E. histolytica* genome is approximately 6500, using a value of 2 tRNA genes per array (since the total tRNA genes in 25 arrays was 49) with an average of 130 copies each. This is over 36 times the number found in the *Schizosaccharomyces pombe* genome and about 10 times the number in the human genome (including pseudogenes). The significance of this high copy number in *E. histolytica* is unclear. In addition, the difference among the copy number of different tRNA genes is just over 3.8-fold (70 to 264 copies) in *E. histolytica*, compared to a projected 6-fold (1 to 6 copies) difference among *Trypanosoma brucei* tRNA genes (Tan et al, 2002).

Preliminary analysis of tRNA gene organization in other *Entamoeba* species has revealed that arrays are present, although their organization varies (Tawari and Clark, unpublished). Clustering of tRNA genes is known from *T. brucei* (Tan et al, 2002) but no tandem arrays or associated STRs have been observed. No evidence for clustering of tRNA genes is found in *Dictyostelium* (Glockner et al, 2002), a distant relative of *Entamoeba* (Bapteste et al, 2002).

PCR amplification success at a particular locus in *E. histolytica* may be directly linked to its copy number in the genome. As a result, PCR amplification of the multi-copy STR loci was found to be more successful than amplification of the single-copy SREHP gene in stool DNA and liver abscess DNA (Roy S, personal communication). Moreover, in

the same sample sets, the amplification of STR locus S-D (about 264 copies per genome) was more successful than locus R-R (about 92 copies per genome).

A comparison of tRNA gene copy numbers in a selection of *E. histolytica* isolates may provide information on whether the copy numbers of these tRNA genes are variable among strains. Similarly, a comparison could also be made with non-pathogenic *Entamoeba* species that infect humans, like *E. dispar* or *E. moshkovskii*, to investigate whether there are quantitative differences in these arrays between them.

CHAPTER 9

COMPARISON OF INTESTINAL AND LIVER ABSCESS STRAINS OF *E. HISTOLYTICA* FROM SAME PATIENTS

9.1. Introduction

It is well established that variable clinical outcomes result from *E. histolytica* infections. Most infections remain asymptomatic, some develop intestinal diseases like diarrhoea and dysentery, and only a few develop extra-intestinal complications like liver abscess. In Chapters 4 and 5 we investigated the association between the genotypes of strains and the clinical outcomes of infection. We found that associations do exist, although we also noticed contradictions, for example, a genotype which was identified as linked to asymptomatic infection in Bangladesh was found in diarrhoea or liver abscess cases in other countries, and vice-versa.

Liver abscess is the most common extra-intestinal infection of *E. histolytica*, and is 10 times more common in men than in women but is a rare disease in children (Haque et al, 2003). There are variable reports on the proportion of amoebic liver abscess patients who had concomitant presence of *E. histolytica* in stool specimens, ranging from less than 10% (Katzenstein et al, 1982) to as high as 70% (Irusen et al, 1992). However, in *E. histolytica*, there have been no reports comparing the genotypes of strains identified in stool with those identified in the liver abscess of the same patients. We therefore were interested in whether the *E. histolytica* strain that was detected in the stool sample, that is 'the intestinal strain', was different from that detected in the liver abscess pus sample of the same ALA patient.

9.2. Samples

Eleven amoebic liver abscess (ALA) patients provided both liver abscess and stool specimens during 1998-99 (Haque et al, 2000). They were admitted to different private and public hospitals and clinics in Dhaka city with symptoms of ALA.

9.3. Results

9.3.1. PCR amplifications and gel electrophoresis results at STR loci

At first, a simple PCR amplification was attempted at all 6 loci using the tRNA primers for all 11 pairs of samples. However the success of PCR was very low, especially with stool DNAs (data not shown), and therefore we decided to use nested PCR (Section 2.3.3) using the first PCR product as the template for the nested PCR with species-specific primers. However PCR still was not successful for all loci. Locus S-Q showed multiple band patterns for almost all of these samples for unknown reasons. In most cases, amplification of the stool derived DNA was unsuccessful, while the corresponding liver abscess DNA was amplified successfully. PCR results for 5 loci, omitting locus S-Q, are shown in Table 9.1. The STR PCRs for D-A, A-L and S-D were the most successful, while the R-R and N-K2 PCRs were least successful.

The combined gel electrophoresis patterns for all 5 loci showed that the intestinal and the corresponding liver abscess patterns were different in all 11 pairs of samples (Figures 9.1 and 9.2, Table 9.1). However, the degree of difference was variable. Only one pair of samples, LAN-11, showed PCR pattern variation in only one locus (N-K2) out of 5, whereas the remaining pairs showed much greater differences. Four pairs of samples showed variation at 2 out of 3 loci, another 4 pairs showed variation at all 3 positive loci and the remaining 2 pairs (LAN-47 and LAN-61) showed variation at all 4 positive loci.

Of the seven pairs of products selected for sequencing 4 pairs of products were different in size from each other, 2 pairs appeared to be the same and for the remaining pair it was difficult to be sure whether there was a size difference (Figure 9.2b, lanes 1 and 2).

9.4. Sequence results of PCR amplified products

We have been able to sequence 7 pairs of PCR products obtained from these samples, in which one product of the pair was derived from the LA DNA and the other from the

stool DNA of the same patient. Table 9.2 summarizes a comparison between PCR patterns and sequence results.

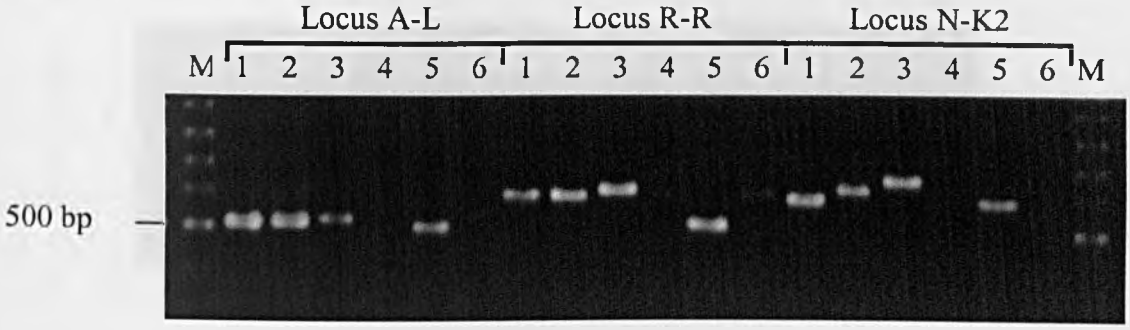
Sequence results verified all the corresponding gel observations, and in addition clearly revealed that the pair of samples that were too close in size to be differentiated by gel electrophoresis were indeed different in sequence (Figure 9.3f). In this pair, we found that the size difference in STR locus D-A from the liver abscess DNA and the stool DNA was 9 base pairs.

Table 9.1. PCR patterns with DNA samples from liver abscess and stool specimens of same ALA patients.

Patient ID	$Gel lane serial	DNA source	PCR patterns: same (s) or different (d)					
			A-L	D-A	R-R	N-K2	S-D	Overall
LAN-11	F1-1	LA	s	s	s	d	s	D
	F1-2	ST						
LAN-15	F2-1	LA	d	d?(seq)	-	-	d	D
	F2-2	ST						
LAN-25	F1-3	LA	*	s	d?	*	d	D
	F1-4	ST						
LAN-33	F2-3	LA	d (seq)	d	-	-	s (seq)	D
	F2-4	ST						
LAN-37	F2-5	LA	d	d	-	-	d (seq)	D
	F2-6	ST						
LAN-42	F2-7	LA	d	s (seq)	-	d[£]	d	D
	F2-8	ST						
LAN-46	F2-9	LA	d	d	-	-	d	D
	F2-10	ST						
LAN-47	F1-5	LA	d	d	d	*	d	D
	F1-6	ST						
LAN-51	F2-11	LA	d	s	-	-	d	D
	F2-12	ST						
LAN-61	F2-13	LA	d	d (seq)	-	d[£]	d	D
	F2-14	ST						
LAN-62	F2-15	LA	d	d (seq)	-	-	d	D
	F2-16	ST						

NB. $F1 = Figure 9.1; $F2 = Figure 9.2; LA = liver abscess; ST = stool; * = stool DNA did not amplify; '£' = not shown in the Figure; '–'= negative for both samples; d? = difference was unclear; (seq) = these products have been sequenced.

(a)



(b)

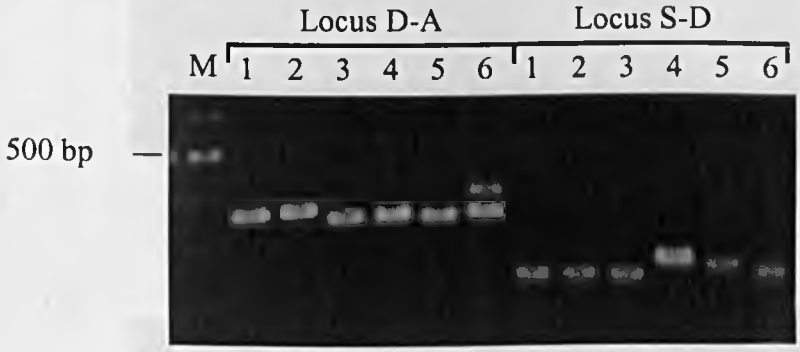


Figure 9.1. PCR fragment size polymorphism at 5 STR loci using DNA samples from liver abscess (odd numbered lanes) and stool specimens (even numbered lanes) of 3 amoebic liver abscess patients: LAN-11 (1 & 2), LAN-25 (3 & 4) and LAN-47 (5 & 6).

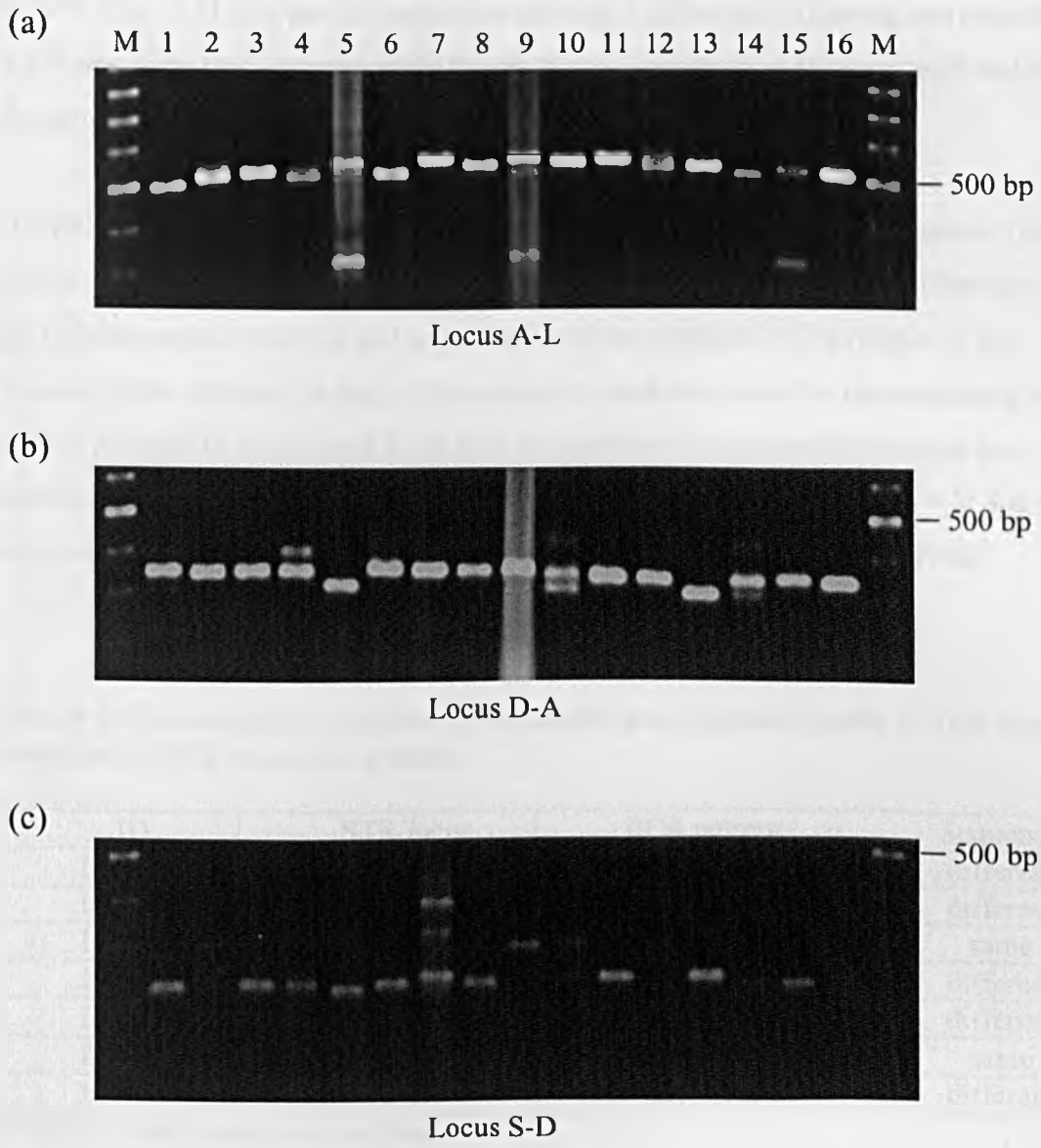


Figure 9.2. PCR fragment size polymorphism at 3 STR loci using DNA samples from liver abscess (odd numbered lanes) and stool (even numbered lanes) specimens of 8 amoebic liver abscess patients: LAN-15 (1 & 2), LAN-33 (3 & 4), LAN-37 (5 & 6), LAN-42 (7 & 8), LAN-46 (9 & 10), LAN-51 (11 & 12), LAN-61 (13 & 14) and LAN-62 (15 & 16).

In STR locus A-L, we sequenced only one pair, and these showed variation only in the number of STRs except for one position where an additional nucleotide 'A' was found in the liver abscess sequence. A schematic representation of STRs is provided in Figure 9.3a and this also highlights the only base insertion or deletion seen in the chromatograms of the two sequences.

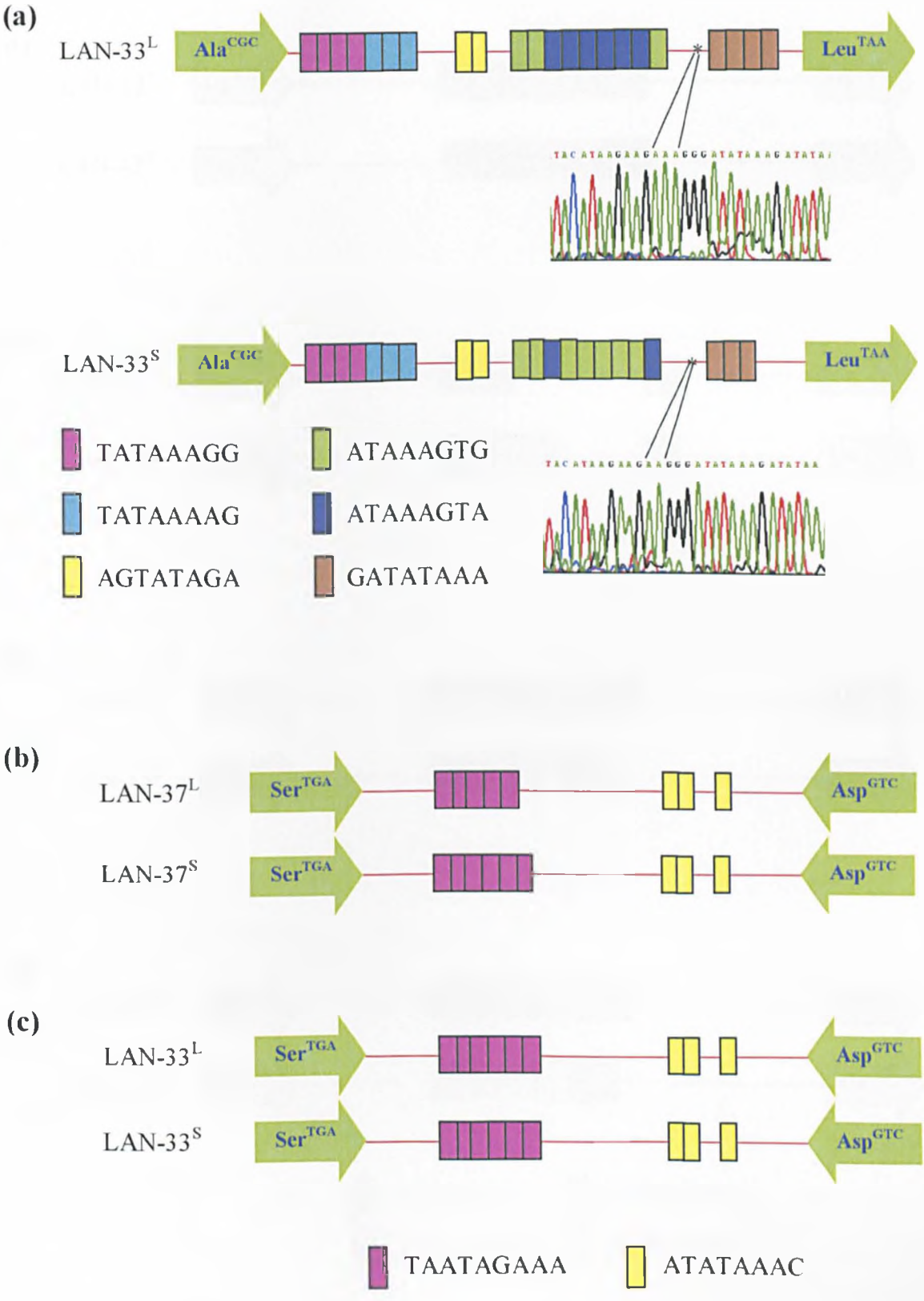
At STR locus S-D, one pair of sequences showed a difference in having one extra STR of 9 bases in the stool sample while the other pair was identical (Figures 9.3b and 9.3c, respectively).

At STR locus D-A, 3 out of 4 pairs of samples showed differences in sequence. There were no sequence differences in LAN-42 (Figure 9.3d). In LAN-61, the difference is due to both variable number and organisation of two types of STRs (Figure 9.3e). However, most striking finding of this sequence analysis is that for the remaining two pairs of samples (LAN-15 and LAN-62), the patterns of sequence differences are identical between the two liver abscess and two intestinal strains (Figures 9.3f and g). In other words LAN-15^{L} and LAN-62^{L} are identical as are LAN-15^{S} and LAN-62^{S}.

Table 9.2. The comparison between PCR patterns and sequence results of liver abscess versus stool DNA from ALA patients.

ID	STR locus	PCR pattern	Sequence
LAN-33	A-L	different	different
LAN-15	D-A	different?	different
LAN-42	D-A	same	same
LAN-61	D-A	different	different
LAN-62	D-A	different	different
LAN-33	S-D	same	same
LAN-37	S-D	different	different

different? = difference was unclear.


(a)
LAN-33L
AlaCGC
LeuTAA
LAN-33S
AlaCGC
LeuTAA
TATAAAGG
TATAAAAG
AGTATAGA
ATAAAGTG
ATAAAGTA
GATATAAA
(b)
LAN-37L
SerTGA
AspGTC
LAN-37S
SerTGA
AspGTC
(c)
LAN-33L
SerTGA
AspGTC
LAN-33S
SerTGA
AspGTC
TAATAGAAA
ATATAAAC

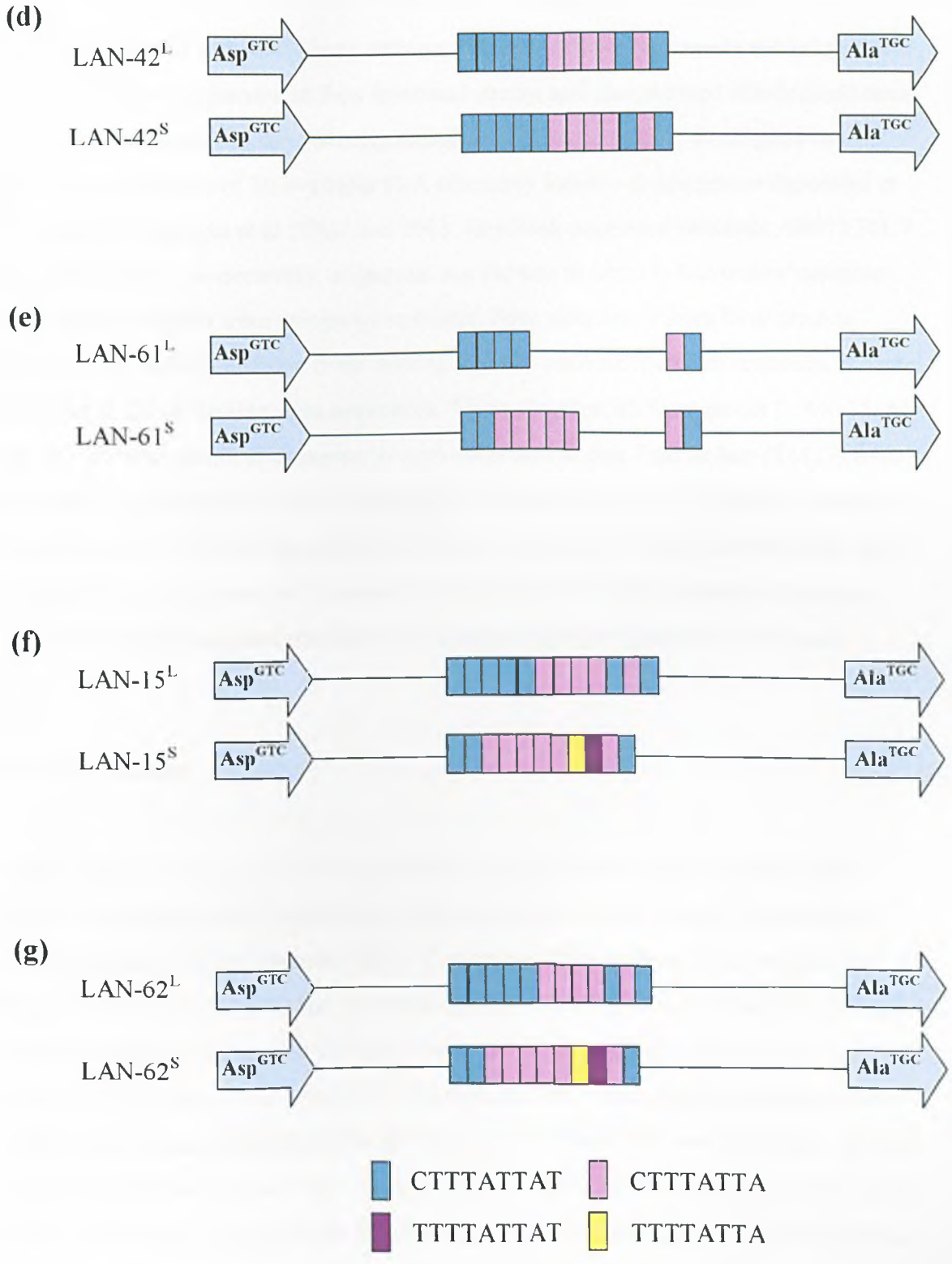


Figure 9.3 (a-g). Schematic representation of sequences of PCR products from liver (L) abscess DNA and stool (S) DNA from the same ALA patients.

9.5. Locus D-A sequence comparison with 10 GenBank sequences

Since we observed something very striking in locus D-A sequences - two ALA patients showed identical sequences in their intestinal strains and also showed identical identical but distinct sequences in their corresponding liver abscess strains, we aligned our sequences with those of 10 available D-A (formerly locus 1-2) sequences deposited in GenBank by Haghighi et al (2002 and 2003; GenBank accession numbers AB075701-7 and AB096653-5, respectively; alignment not shown) in order to know how common these sequence types were compared to others. Four samples, 3 from liver abscess (LAN-15, 42 and 62) and one from stool (LAN-42), showed identical sequence to each other and to 26 of the Haghighi sequences. Three samples, all from stools (LAN-15, 61 and 62), showed identical sequence to each other and to one Thai isolate (TM27) from an amoebic colitis patient. The remaining D-A sequence (LAN-61^{L}) showed identical sequence to two isolates, one isolated from an asymptomatic Ghanaian individual and the other from a Japanese ALA patient's stool who had amoebic colitis at the same time. This finding suggests that all these sequence types are globally distributed.

9.6. Discussion

In this study, 11 pairs of *E. histolytica* DNA samples isolated from stool and liver abscess of the same ALA patient were analysed genotypically in order to determine whether strains that had metastasized to liver tissue differed from those in intestinal sites. We believe that this is the first investigation of this type in *E. histolytica* and the results we found were very interesting: the intestinal strain was different from the liver abscess strain in all 11 ALA patients investigated. This observation was supported by another: in the same samples, although the nested SREHP PCR was successful for only three pairs, all three revealed that the intestinal and liver abscess strains were different from each other (S. Roy, ICDDR,B, personal communication). There are two possible explanations for these findings. First, the *E. histolytica* intestinal population contains multiple genotypes, and only a subpopulation has migrated to the liver via blood circulation and caused a liver abscess. It must be a minor subpopulation since the 'liver abscess' bands are not visible in the stool sample patterns. However we do not know

why only a minor variant in each of eleven patients' infections would show a preference for the liver tissues, while the major population remained in the intestine. Second, there may be DNA reorganisation or recombination events taking place when the amoebae migrate from the intestine to the liver. As more than one locus was usually observed to differ this recombination would have to be extensive and widespread in the genome.

In order to genotype these strains we used all 6 STR-locus based PCRs, but some amplifications were more successful than others. For example, D-A and S-D PCRs were successful for all 22 samples but R-R and N-K2 were unsuccessful for a majority of the samples. In addition, locus S-Q showed some unusual multiple band patterns with these samples. There are two possible explanations for this. One, these samples had been preserved in freezers for the last 5-6 years and may have undergone repeated freezing and thawing, which resulted in random degradation of the DNA. Two, it is possible that the variable success in PCR amplifications is a direct consequence of the variable copy number of these loci in the genome. For example, the most successful PCR, S-D, has about three times as many copies as R-R in the genome (approximately 264 copies versus 92 copies, respectively; Table 8.4, Clark et al, 2005). In general, PCR amplification was more successful in liver abscess DNA compared to corresponding stool DNA. This is possibly because the number of parasites in the stool samples was fewer than in the corresponding liver abscess samples.

One of the reasons for sequencing some of the PCR products was to better understand the nature of the differences between intestinal and liver abscess strains in the sample pairs. Although the overall degree of difference we observed was high, in two particular ALA patients (LAN-15 and LAN-62) we noticed something very striking. In these two patients, their intestinal strains showed identical sequences in locus D-A, and both liver abscess strains showed identical sequences to each other but different from the intestinal ones, suggesting that during the metastasis from intestine to the liver the same event (either subpopulation selection or DNA reorganisation) took place in two unrelated ALA patients. However the following facts strongly argue against the two patients being infected by the same mixture of *E. histolytica*:

One, there was a more than 7 month gap between the two sample collection dates of LAN-15 (1.4.1999) and LAN-62 (18.11.99), so there is almost no chance that the two

patients were infected from the same source with the same mixture of *E. histolytica* strains.

Two, there was no identity in PCR product sizes between these two pairs of samples at locus A-L (Figure 9.2a, lanes 1&2 and 15&16) or locus S-D (Figure 9.2c, lanes 1&2 and 15&16), which clearly suggests that all 4 of these strains of *E. histolytica* are genetically different from each other.

We were then interested to see how frequently we might expect the two sequence types (the intestinal type and the liver abscess type) to be found if they were in 2 unrelated samples. In order to investigate this we compared our sequence types with those observed by Haghighi et al (2002 and 2003). They detected 13 sequence types from 74 patients in locus D-A (formerly known as locus 1-2). If we assume that their result is also a reflection of sequence types in Bangladeshi strains then we can see that the frequency of the LAN-15 and LAN-62 liver abscess type, (i.e. type D of Haghighi et al, 2002 and 2003) was 26 out of 74. Similarly, the frequency of the LAN-15 and LAN-62 intestinal type, (i.e. type K of Haghighi et al, 2002 and 2003) was only 1 out of 74. Therefore, the probability of these two types being found in two samples is 26/74 x 1/74 or 0.0047 (i.e. less than a 0.5% chance). This suggests that it is very unlikely that this identical change of DNA sequence in this locus has happened simply by chance. However, we failed to find a similar change of DNA in other loci, based on comparison of gel patterns at other STR loci for these two patients. So, it seems more likely that the DNA reorganisation during metastasis is somewhat random - it does not affect all loci in the same way – if that is indeed what is occurring.

In *Leishmania* (*Viannia*), a very similar investigation using genotypic analysis of kinetoplast DNA RFLP patterns in ten pairs of strains isolated from mucosal and cutaneous lesions of the same patients found that each of the 10 cutaneous strains was different from their counterpart mucosal strains. The authors suggested that the population that had reached the mucosal tissue may have been a subpopulation of the original inoculum (Cuervo et al, 2004). However, some information on the sequences of the relevant loci could add some vital support to their conclusion. Nevertheless, their results show a striking parallel to those found here for *E. histolytica*.

In a study by Haque et al (2000), lectin antigens were detected in the serum of 32 out of 98 ALA patients in Bangladesh. This indicates it might be possible to amplify *E. histolytica* DNA from the whole blood of ALA patients. It might be possible to investigate the genotypes of strains in the stool, blood and liver abscess of the same ALA patient. Such an investigation could provide vital additional information on whether the observed differences in the genotypes of intestinal and liver abscess strains in this study are the consequence of DNA recombination or a variable intestinal population.

Finally, in the absence of a suitable animal model for the production of ALA following oral or caecum inoculation of amoebae, we believe that it will be very difficult to prove what is actually happening during the metastasis of the amoebae from intestine to the liver, unless we see many more recurrent patterns of DNA differences similar to those we observed for patients LAN-15 and LAN-62 in locus D-A. We also believe that it is worth investigating whether this observation is common in pairs of samples from ALA patients in other parts of the world.

CHAPTER 10

CONCLUDING REMARKS AND FUTURE PROSPECTS

Despite an annual death toll of up to 100,000 people from *E. histolytica* infection, the epidemiology and virulence pattern of *E. histolytica* is not very clear at present. In this study we wanted to address some of the important gaps surrounding the organism's genetic epidemiology: one, how many genetically distinct *E. histolytica* strains circulate in a geographically restricted population?, two, is the asymptomatic or symptomatic outcome of an *E. histolytica* infection linked to the genotype of the infecting amoeba?, and three, do *E. histolytica* strains show organ tropism?

Firstly, we have investigated the organism's genetic diversity not only in samples from a restricted area in Bangladesh but also in samples from some other countries using the multilocus genotyping system we developed. In general, the diversity among strains was high, increased as we studied more loci, and liver abscess strains showed the most variation.

Secondly, the analysis of results clearly showed that the *E. histolytica* genotypes detected in 3 clinical populations – asymptomatic, diarrhoeal/dysenteric and liver abscess – are indeed different from each other, suggesting that the outcome of infection is somehow linked to the genotype of the infecting strain. In addition a few individual genotypes showed a specific association with the outcome of infection, although the number of samples covered by these genotypes is not very high and therefore their predictive value is low.

Thirdly, it is not clear from our results whether *E. histolytica* strains show organ tropism although we noticed that the strains identified in stool samples and the corresponding liver abscess pus samples from the same ALA patient were different. The exact basis of this observed difference is not clear. If *E. histolytica* in the gut is a polyclonal population, then we do not know why it is always only a minor population that possesses the ability to migrate to the liver. Alternatively if DNA reorganization events are taking place during metastasis of the amoeba, then we do not know why it occurs or what the trigger is.

Unique organization and abundance of tRNA genes are two distinctive features of *E. histolytica*. About one-tenth of all genomic sequences contain tRNA genes and they are organised in 25 distinct units that are repeated tandemly in the genome. The intergenic STRs have not been identified in any other species except for *E. dispar*. So it is unlikely that similar tRNA-gene linked genotyping approaches may be applied to other *Entamoeba* species, thus making this a novel approach unique to *E. histolytica* and *E. dispar*.

The advantage of our genotyping system over the most widely used polymorphic marker in *E. histolytica* – SREHP - is that it does not require any restriction endonuclease digestion to reveal all the variants and produces a single major product for each locus in the amplification. In addition, our tRNA-gene linked arrays are multicopy compared to single copy SREHP gene, and this presumably makes it easier to detect the amoeba in biological samples, e.g., stool and liver abscess pus. Most importantly our genotyping has shown some links between genotype and outcome of infection which no other typing methods in *E. histolytica* have ever shown.

In some protozoan parasites microsatellite markers have been used successfully in the genotyping of species and strains. For example, the *Plasmodium falciparum* genome has an abundance of microsatellite loci: 1 such locus is found in every 2-3 kb of genome. They have been used to analyse traits that are inherited in progeny of an experimental cross, and to analyse the genetic structure of parasite populations around the world (Machado et al, 2004). Microsatellites have also been used in genotyping of species of *Cryptosporidium* (Widmer et al, 2004), *Toxoplasma* (Ajzenberg et al, 2005; Blackston et al, 2001), *Leishmania* (Bulle et al, 2002; Schwenkenbecher et al, 2004) and *Giardia* (Huetink et al, 2001). However, despite experimental efforts by Zaki and Clark (2001) no microsatellite was isolated from *E. histolytica* and the genome sequence revealed that they were almost absent from this species. Since the tandem repeat DNA element is short in microsatellites, e.g., 2-3 bases, it often requires special equipment such as sequencing gels to detect the size variation among samples. In our tRNA based genotyping system, the STRs are often 8 bases or longer, making it easier to detect the size variation in a conventional gel.

As mentioned above the most exciting finding of our study is that the genotypes of *E. histolytica* found in asymptomatic, diarrhoeal/dysenteric and liver abscess samples are significantly different from each other. All asymptomatic and a majority of diarrhoeal/dysenteric samples were from children of a similar age from Mirpur. So, the observed differences in these two sample groups are more like those that could be expected from a case-control study. However the liver abscess samples were from adult patients and from a much larger area of Bangladesh. As a result there is a possibility that some of the observed differences in genotypes between liver abscess and other two clinical groups are due to geographical variation. So we believe that more appropriate samples, e.g., 3 types of clinical samples from individuals from same age-group and living in the same geographic area, would be more informative.

Although our tRNA-linked genotyping data found (i) significant differences in genotypes between clinical populations, (ii) an association of a few individual *E. histolytica* genotypes with the clinical outcome of infections, and (iii) that transversion mutations are more common in locus R-R of liver abscess strains, it is highly unlikely that these tRNA genes or arrays are directly involved in the virulence process of the parasite. However the findings suggest that these tRNA-gene arrays are somehow linked to factors that are responsible for determining the outcome of infection. One possibility is that these arrays and the factors determining the outcome may be physically linked to each other, for example they may occur on the same chromosome in the genome. As a result, investigation can be narrowed down to those corresponding chromosomes to search for factors influencing the outcome of infection. Other approaches such as microarray-based comparative genomic or mRNA hybridisation could then be used to identify such determining factors.

We observed that some of the tRNA-linked loci showed little or no polymorphism in both *E. histolytica* and *E. dispar* strains and these were not selected for genotyping. Two such loci, E-V and X-T, however, showed significant PCR product size differences between the two species (Figure 10.1). Although locus E-V showed some variation among *E. dispar* isolates the product size difference with *E. histolytica* isolates is always about 100 bp. Similarly there is about a 100 bp size difference between *E. histolytica* and *E. dispar* locus X-T products. So, either of these two PCRs could potentially be used as a

diagnostic tool to detect *E. histolytica* and *E. dispar* simultaneously in a single PCR reaction.

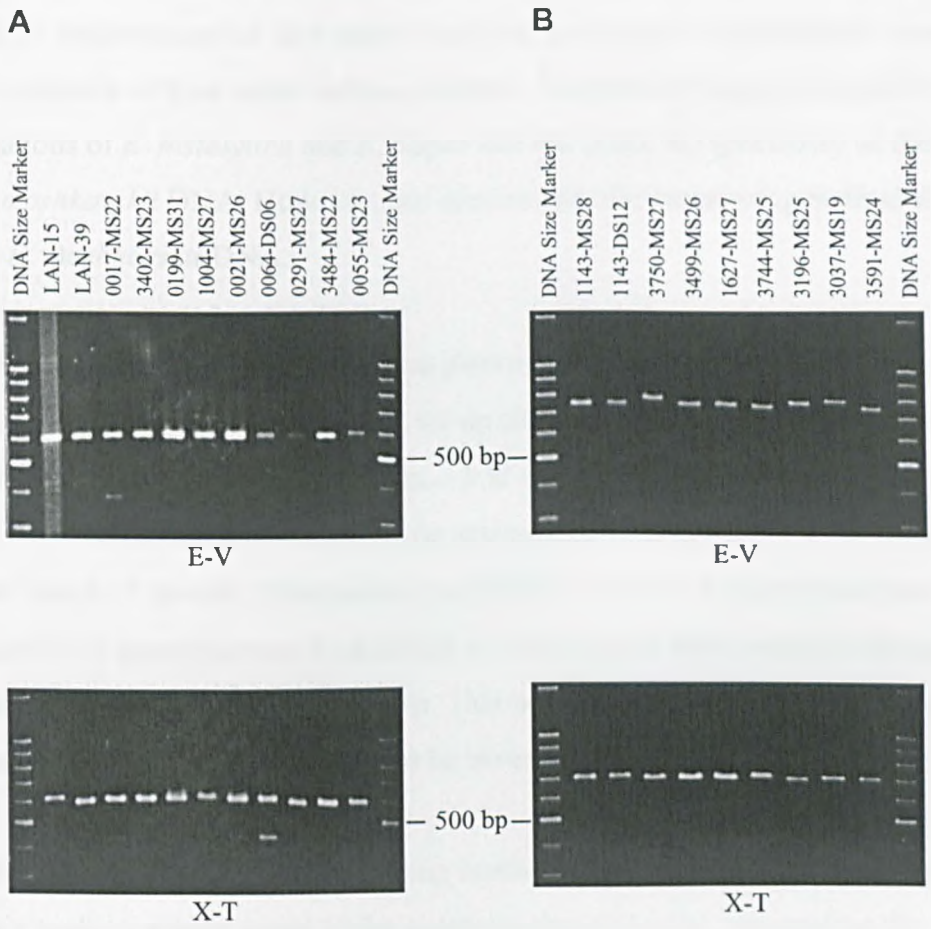


Figure 10.1. PCR product size variation between *E. histolytica* and *E. dispar* isolates. (A) 11 *E. histolytica* isolates showing almost no PCR product size differences at either locus E-V or locus X-T. (B) 9 *E. dispar* isolates are showing some PCR product size variation at locus E-V but not locus X-T.

Multiple strains of *E. histolytica* infecting the same individual are probably a rare event as we detected only one such case in a Mirpur child. In consecutive months, follow-up samples of this child revealed that it had gained and lost a transient infection. This finding probably contradicts the hypothesis that symptomatic infections of *E. histolytica* may be linked to multiple strains being involved in the infection process (Zaki et al, 2002), as the child that showed the multiple strains remained asymptomatic throughout.

One of the extraordinary findings of this study is the high prevalence of *E. moshkovskii* in Mirpur children. Most but not all of these *E. moshkovskii* infections were identified in association with an *E. histolytica* or *E. dispar* infection. If this is true for other parts of the world, then epidemiological data which were and are based on microscopic identification of these parasites will be under serious question. Furthermore many of the DNA based investigations of *E. histolytica* and *E. dispar* did not check the specificity of their methods with *E. moshkovskii* DNA. However, our species-specific genotyping method should not amplify *E. moshkovskii* DNA.

The follow-up samples in this study have shown that these tRNA-linked markers are generally stable in the same individual, for up to 11 months. A longitudinal study in Vietnam using 2 of the tRNA-linked loci – I-W and N-K - also found that the patterns mostly remained genetically stable during asymptomatic infection, over an observation period of about 15 months (Blessmann et al, 2003). However it is not clear how such a large number of genotypes can be detected in a small geographic area like Mirpur despite this observed genetic stability in the host. This seems to be a contradiction. How the genotype diversity is generated needs to be investigated.

We believe that the multilocus genotyping method described in this study will provide the necessary tools to answer many of the questions that still exist surrounding the epidemiology and variable virulence of *E. histolytica* infection worldwide, e.g., are any *E. histolytica* strains intrinsically invasive for all individuals?, do invasive *E. histolytica* strains differ in their capacity to produce disease?, do *E. histolytica* strains show variablity in their capacity to elicit an immune response in the host?, are epidemics / outbreaks of *E. histolytica* ever caused by a single strain of *E. histolytica*? To address such questions will require significant additional work and samples that are not currently available.

BIBLIOGRAPHY

1. **Abd-Alla, M. D., T. F. H. G. Jackson, V. Gathiram, A. M. El-Hawey, and J. I. Ravdin.** 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.
2. **Acuña-Soto, R., J. H. Maguire, and D. F. Wirth.** 2000. Gender distribution in asymptomatic and invasive amebiasis. Am J Gastroenterol **98**:1277-1283.
3. **Ajzenberg, D., A. Dumetre, and M. Darde.** 2005. Multiplex PCR for Typing Strains of *Toxoplasma gondii*. J Clin Microbiol **43**:1940-1943.
4. **Anonymous.** 1997. WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. Epidemiol Bull PAHO **18**:13-14.
5. **Anonymous.** 1998. The World Health Report 1998. Life in the 21st century: a vision for all. World Health Organization, Geneva, Switzerland.
6. **Astal, Z.** 2004. Epidemiological survey of the prevalence of parasites among children in Khan Younis governorate, Palestine. Parasitol Res **94**:449-51.
7. **Ayeh-Kumi, P. F., I. K. M. Ali, L. A. Lockhart, C. A. Gilchrist, W. A. Petri, Jr., and R. Haque.** 2001. *Entamoeba histolytica*: genetic diversity of clinical isolates from Bangladesh as demonstrated by polymorphisms in the serine-rich gene. Exp Parasitol **99**:80-88.
8. **Azam, A., J. Paul, D. Sehgal, J. Prasad, S. Bhattacharya, and A. Bhattacharya.** 1996. Identification of novel genes from *Entamoeba histolytica* by expressed sequence tag analysis. Gene **181**:113-116.
9. **Balamuth, W.** 1946. Improved egg yolk medium for cultivation of *Entamoeba histolytica* and other intestinal protozoa. Am J Clin Pathol **16**:380-384.
10. **Banerjee, S., and A. Lohia.** 2003. Molecular analysis of repetitive DNA elements from *Entamoeba histolytica*, which encode small RNAs and contain matrix/scaffold attachment recognition sequences. Mol Biochem Parasitol **126**:35-42.
11. **Bapteste, E., H. Brinkmann, J. A. Lee, D. V. Moore, C. W. Sensen, P. Gordon, L. Duruflé, T. Gaasterland, P. Lopez, M. Müller, and H. Philippe.** 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. Proc Natl Acad Sci USA **99**:1414-1419.

12. **Beck, D. L., M. Tanyuksel, A. J. Mackey, R. Haque, N. Trapaidze, W. R. Pearson, B. Loftus, and W. A. Petri, Jr.** 2002. *Entamoeba histolytica*: sequence conservation of the Gal/GalNAc lectin from clinical isolates. Exp Parasitol **101**:157-163.

13. **Bhattacharya, S., A. Bhattacharya, and L. S. Diamond.** 1988. Comparison of repeated DNA from strains of *Entamoeba histolytica* and other *Entamoeba*. Mol Biochem Parasitol **27**:257-262.

14. **Bhattacharya, S., A. Bhattacharya, L. S. Diamond, and A. T. Soldo.** 1989. Circular DNA of *Entamoeba histolytica* encodes ribosomal RNA. J Protozool **36**:455-458.

15. **Blackston, C. R., J. P. Dubey, E. Dotson, C. Su, P. Thulliez, D. Sibley, and T. Lehmann.** 2001. High-resolution typing of *Toxoplasma gondii* using microsatellite loci. J Parasitol **87**:1472-5.

16. **Blessmann, J., and E. Tannich.** 2002. Treatment of asymptomatic intestinal *Entamoeba histolytica* infection. N Engl J Med **347**:1384.

17. **Blessmann, J., P. Van Linh, P. A. Nu, H. D. Thi, B. Muller-Myhsok, H. Buss, and E. Tannich.** 2002. Epidemiology of amebiasis in a region of high incidence of amebic liver abscess in central Vietnam. Am J Trop Med Hyg **66**:578-83.

18. **Blessmann, J., I. K. M. Ali, P. A. Ton Nu, B. T. Dinh, T. Q. Ngo Viet, A. Le Van, C. G. Clark, and E. Tannich.** 2003. Longitudinal study of intestinal *Entamoeba histolytica* infections in asymptomatic adult carriers. J Clin Microbiol **41**:4745-4750.

19. **Briggs, G. G., R. K. Freeman, and S. J. Yaffe.** 1994. Drugs in pregnancy and lactation. A reference guide to fetal and neonatal risk., 4th ed. Williams and Wilkins., Baltimore.

20. **Britten, D., S. N. Wilson, R. McNerney, A. H. Moody, P. L. Chiodini, and J. P. Ackers.** 1997. An improved colorimetric PCR-based method for detection and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* in feces. J Clin Microbiol **35**:1108-1111.

21. **Bulle, B., L. Millon, J. M. Bart, M. Gallego, F. Gambarelli, M. Portus, L. Schnur, C. L. Jaffe, S. Fernandez-Barredo, J. M. Alunda, and R. Piarroux.** 2002. Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis. J Clin Microbiol **40**:3391-7.

22. **Burch, D. J., E. Li, S. Reed, T. F. H. G. Jackson, and S. L. Stanley, Jr.** 1991.

Isolation of a strain-specific *Entamoeba histolytica* cDNA clone. J Clin Microbiol **29:**696-701.

23. **Caballero-Salcedo, A., M. Viveros-Rogel, B. Salvatierra, R. Tapia-Conyer, J. Sepúlveda-Amor, G. Gutiérrez, and L. Ortíz-Ortíz.** 1994. Seroepidemiology of amebiasis in Mexico. Am J Trop Med Hyg **50:**412-419.

24. **Casavilca Zambrano, S., V. Gomez Anchante, and E. Cisneros Gallegos.** 2000. Complicated amoebic appendicitis. Report of a case (in Spanish). Rev Gastroenterol Peru **20:**296-299.

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

26. **Chadee, K., J. M. Smith, and E. Meerovitch.** 1985. *Entamoeba histolytica*: Electrophoretic isoenzyme patterns of strains and their virulence in the cecum of gerbils (*Meriones unguiculatus*). Am J Trop Med Hyg **34:**870-878.

27. **Cheng, X. J., and H. Tachibana.** 2001. Protection of hamsters from amebic liver abscess formation by immunization with the 150- and 170-kDa surface antigens of *Entamoeba histolytica.* Parasitol Res **87:**126-30.

28. **Clark, C. G., and L. S. Diamond.** 1991. The Laredo strain and other *Entamoeba histolytica* –like amoebae are *Entamoeba moshkovskii.* Mol Biochem Parasitol **46:**11-18.

29. **Clark, C. G.** 1992. DNA purification from polysaccharide-rich cells, p. D-3.1 - D-3.2. *In* J. J. Lee and A. T. Soldo (ed.), Protocols in Protozoology, vol. 1. Allen Press, Lawrence, Kansas.

30. **Clark, C. G., and L. S. Diamond.** 1993. *Entamoeba histolytica*: a method for isolate identification. Exp Parasitol **77:**450-455.

31. **Clark, C. G.** 1995. Axenic cultivation of *Entamoeba dispar* Brumpt 1925, *Entamoeba insolita* Geiman and Wichterman 1937 and *Entamoeba ranarum* Grassi 1879. J Euk Microbiol **42:**590-593.

32. **Clark, C. G., and L. S. Diamond.** 1997. Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. J Euk Microbiol **44:**142-154.

33. **Clark, C. G., M. Zaki, and I. K. M. Ali.** 2002. Genetic diversity in *Entamoeba histolytica.* J Biosci **27:**603-607.

34. **Clark, C. G., and L. S. Diamond.** 2002. Methods for cultivation of luminal parasitic protists of clinical importance. Clin Microbiol Rev **15:**329-341.

35. **Clark, C. G., I. K. M. Ali, M. Zaki, B. J. Loftus, and N. Hall.** 2005. Unique organization of tRNA genes in the protistan parasite *Entamoeba histolytica*. Submitted.

36. **Colwell, R. K., and J. A. Coddington.** 1994. Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Lond B Biol Sci **345:**101-18.

37. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res **16:**10881-90.

38. **Cruz-Reyes, J., T. ur-Rehman, W. M. Spice, and J. P. Ackers.** 1995. A novel transcribed repeat element from *Entamoeba histolytica*. Gene **166:**183-184.

39. **Cuervo, P., E. Cupolillo, N. Nehme, V. Hernandez, N. Saravia, and O. Fernandes.** 2004. *Leishmania* (*Viannia*): genetic analysis of cutaneous and mucosal strains isolated from the same patient. Exp Parasitol **108:**59-66.

40. **Da Silveira, L. A., W. L. Ribeiro, K. Kirchgatter, G. Wunderlich, H. Matsuoka, K. Tanabe, and M. U. Ferreira.** 2001. Sequence diversity and linkage disequilibrium within the merozoite surface protein-1 (Msp-1) locus of *Plasmodium falciparum*: a longitudinal study in Brazil. J Eukaryot Microbiol **48:**433-9.

41. **de la Vega, H., C. A. Specht, C. E. Semino, P. W. Robbins, D. Eichinger, D. Caplivski, S. Ghosh, and J. Samuelson.** 1997a. Cloning and expression of chintinases of *Entamoebae*. Mol Biochem Parasitol **85:**139-147.

42. **de la Vega, H., C. A. Specht, A. Chau, C. E. Semino, P. W. Robbins, D. Eichinger, D. Caplivski, S. Ghosh, and J. Samuelson.** 1997b. Cyst-specific exochitinases of *Entamoebae* contain unique hydrophilic repeats at their amino termini. Arch Med Res **28:**S143-6.

43. **Diamond, L. S.** 1961. Axenic cultivation of *Entamoeba histolytica*. Science **134:**336-337.

44. **Diamond, L. S.** 1968. Techniques of axenic cultivation of *Entamoeba histolytica* Schaudinn, 1903 and *E. histolytica*-like amebae. J Parasitol **54:**1047-1056.

45. **Diamond, L. S., C. F. T. Mattern, and I. L. Bartgis.** 1972. Viruses of *Entamoeba histolytica* I. Identification of transmissible virus-like agents. J Virol **9:**326-341.

46. **Diamond, L. S., D. R. Harlow, and C. C. Cunnick.** 1978. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. Trans R Soc Trop Med Hyg **72:**431-432.

47. **Diamond, L. S., and C. G. Clark.** 1993. A redescription of *Entamoeba histolytica*

Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. J Euk Microbiol **40:**340-344.

48. **Dreyer, D. A.** 1961. Growth of a strain of *Entamoeba histolytica* at room temperature. Tex Rep Biol Med **19:**393-396.

49. **Duggal, P., R. Haque, S. Roy, D. Mondal, R. B. Sack, B. M. Farr, T. H. Beaty, and W. A. Petri, Jr.** 2004. Influence of human leukocyte antigen class II alleles on susceptibility to *Entamoeba histolytica* infection in Bangladeshi children. J Infect Dis **189:**520-6.

50. **el-Hamshary, E. M., K. A. el-Shewy, M. M. Hezagy, and H. Zakaria.** 2004b. Selective identification of the pathogenic *E. histolytica* in fresh stool samples using polymerase chain reaction (PCR). J Egypt Soc Parasitol **34:**611-20.

51. **Entner, N., and H. Most.** 1965. Genetics of *Entamoeba*: characterization of two new parasitic strains that grow at room temperature (and at 37°C). J Protozool **12:**10-13.

52. **Espinosa Cantellano, M., G. Castanon Gutierrez, and A. Martinez-Palomo.** 1997. In vivo pathogenesis of *Entamoeba dispar*. Arch Med Res **28 Spec No:**204-6.

53. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **186:**1518-30.

54. **Fernandez, M. C., S. Verghese, R. Bhuvaneswari, S. J. Elizabeth, T. Mathew, A. Anitha, and A. K. Chitra.** 2002. A comparative study of the intestinal parasites prevalent among children living in rural and urban settings in and around Chennai. J Commun Dis **34:**35-9.

55. **Furrows, S. J., A. H. Moody, and P. L. Chiodini.** 2004. Comparison of PCR and antigen detection methods for diagnosis of *Entamoeba histolytica* infection. J Clin Pathol **57:**1264-6.

56. **Garcia Uribe, J. A., A. Padua Gabriel, O. Quintana, A. Garcia Vazquez, and G. de la Escosura.** 1991. 3 cases of hematogenous lung abscess of amebic origin. Rev Invest Clin **43:**264-8.

57. **Gathiram, V., and T. F. H. G. Jackson.** 1985. Frequency distribution of *Entamoeba histolytica* zymodemes in a rural South African population. Lancet **i:**719-721.

58. **Gathiram, V., and T. F. H. G. Jackson.** 1987. A longitudinal study of

asymptomatic carriers of pathogenic zymodemes of *Entamoeba histolytica*. South Afr Med J **72**:669-672.

59. **Gatti, S., G. Swierczynski, F. Robinson, M. Anselmi, J. Corrales, J. Moreira, G. Montalvo, A. Bruno, R. Maserati, Z. Bisoffi, and M. Scaglia.** 2002. Amebic infections due to the *Entamoeba histolytica-Entamoeba dispar* complex: a study of the incidence in a remote rural area of Ecuador. Am J Trop Med Hyg **67**:123-127.

60. **Ghadirian, E., and E. Meerovitch.** 1978. Vaccination against hepatic amebiasis in hamsters. J Parasitol **64**:742-743.

61. **Ghadirian, E., E. Meerovitch, and D. P. Hartmann.** 1980. Protection against amebic liver abscess in hamsters by means of immunization with amebic antigen and some of its fractions. Am J Trop Med Hyg **29**:779-784.

62. **Ghosh, S., M. Frisardi, L. Ramirez-Avila, S. Descoteaux, K. Sturm-Ramirez, O. A. Newton-Sanchez, J. I. Santos-Preciado, C. Ganguly, A. Lohia, S. Reed, and J. Samuelson.** 2000. Molecular epidemiology of *Entamoeba* spp.: evidence of a bottleneck (Demographic sweep) and transcontinental spread of diploid parasites. J Clin Microbiol **38**:3815-3821.

63. **Gibbons-Matthews, C., and A. M. Prescott.** 2003. Intra-isolate variation of *Cryptosporidium parvum* small subunit ribosomal RNA genes from human hosts in England. Parasitol Res **90**:439-44.

64. **Glockner, G., L. Eichinger, K. Szafranski, J. A. Pachebat, A. T. Bankier, P. H. Dear, R. Lehmann, C. Baumgart, G. Parra, J. F. Abril, R. Guigo, K. Kumpf, B. Tunggal, E. Cox, M. A. Quail, M. Platzer, A. Rosenthal, and A. A. Noegel.** 2002. Sequence and analysis of chromosome 2 of *Dictyostelium discoideum*. Nature **418**:79-85.

65. **Gonzalez-Ruiz, A., R. Haque, T. Rehman, A. Aguirre, A. Hall, F. Guhl, D. C. Warhurst, and M. A. Miles.** 1994. Diagnosis of amebic dysentery by detection of *Entamoeba histolytica* fecal antigen by an invasive strain-specific, monoclonal antibody-based enzyme-linked immunosorbent assay. J Clin Microbiol **32**:964-970.

66. **Haghighi, A., S. Kobayashi, T. Takeuchi, G. Masuda, and T. Nozaki.** 2002. Remarkable genetic polymorphism among *Entamoeba histolytica* isolates from a limited geographic area. J Clin Microbiol **40**:4081-4090.

67. **Haghighi, A., S. Kobayashi, T. Takeuchi, N. Thammapalerd, and T. Nozaki.** 2003. Geographic diversity among genotypes of *Entamoeba histolytica* field isolates. J Clin Microbiol **41**:3748-3756.

68. **Haque, R., K. Kress, S. Wood, T. F. H. G. Jackson, D. Lyerly, T. Wilkins, and W. A. Petri, Jr.** 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.

69. **Haque, R., L. M. Neville, P. Hahn, and W. A. Petri, Jr.** 1995. Rapid diagnosis of *Entamoeba* infection by using *Entamoeba* and *Entamoeba histolytica* stool antigen detection kits. J Clin Microbiol **33:**2558-2561.

70. **Haque, R., A. S. G. Faruque, P. Hahn, D. M. Lyerly, and W. A. Petri, Jr.** 1997. *Entamoeba histolytica* and *Entamoeba dispar* infection in children in Bangladesh. J Infect Dis **175:**734-736.

71. **Haque, R., I. K. M. Ali, C. G. Clark, and W. A. Petri, Jr.** 1998a. A case report of *Entamoeba moshkovskii* infection in a Bangladeshi child. Parasitol Int **47:**201-202.

72. **Haque, R., I. K. M. Ali, S. Akther, and W. A. Petri, Jr.** 1998b. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J Clin Microbiol **36:**449-52.

73. **Haque, R., I. K. M. Ali, and W. A. Petri, Jr.** 1999. Prevalence and immune response to *Entamoeba histolytica* infection in preschool children in Bangladesh. Am J Trop Med Hyg **60:**1031-1034.

74. **Haque, R., N. U. Mollah, I. K. M. Ali, K. Alam, A. Eubanks, D. Lyerly, and W. A. Petri, Jr.** 2000. Diagnosis of amebic liver abscess and intestinal infection with the TechLab *Entamoeba histolytica* II antigen detection and antibody tests. J Clin Microbiol **38:**3235-3239.

75. **Haque, R., I. K. M. Ali, R. B. Sack, B. M. Farr, G. Ramakrishnan, and W. A. Petri, Jr.** 2001. Amebiasis and mucosal IgA antibody against the *Entamoeba histolytica* adherence lectin in Bangladeshi children. J Infect Dis **183:**1787-1793.

76. **Haque, R., C. D. Huston, M. Hughes, E. Houpt, and W. A. Petri, Jr.** 2003. Amebiasis. N Engl J Med **348:**1565-1573.

77. **Hara, A., Y. Hirose, H. Mori, H. Iwao, T. Kato, and Y. Kusuhara.** 2004. Cytopathologic and genetic diagnosis of pulmonary amebiasis: a case report. Acta Cytol **48:**547-50.

78. **Hoffmann, E. H., L. A. da Silveira, R. Tonhosolo, F. J. Pereira, W. L. Ribeiro, A. P. Tonon, F. Kawamoto, and M. U. Ferreira.** 2001. Geographical patterns of allelic diversity in the *Plasmodium falciparum* malaria-vaccine candidate,

merozoite surface protein-2. Ann Trop Med Parasitol **95:**117-32.

79. **Hooshyar, H., M. Rezaian, B. Kazemi, M. Jeddi-Tehrani, and S. Solaymani-Mohammadi.** 2004. The distribution of *Entamoeba histolytica* and *Entamoeba dispar* in northern, central, and southern Iran. Parasitol Res **94:**96-100.

80. **Houpt, E., L. Barroso, L. Lockhart, R. Wright, C. Cramer, D. Lyerly, and W. A. Petri, Jr.** 2004. Prevention of intestinal amebiasis by vaccination with the *Entamoeba histolytica* Gal/GalNac lectin. Vaccine **22:**611-7.

81. **Huang, M., K. P. Chang, and R. A. Albach.** 1997. A 964 bp repetitive DNA in *Entamoeba histolytica* is associated with linear "chromosomal" DNAs of variable sizes. Arch Med Res **28:**S1-S4.

82. **Huber, M., B. Koller, C. Gitler, D. Mirelman, M. Revel, S. Rozenblatt, and L. Garfinkel.** 1989. *Entamoeba histolytica* ribosomal RNA genes are carried on palindromic circular DNA molecules. Mol Biochem Parasitol **32:**285-296.

83. **Huetink, R. E., J. W. van der Giessen, J. P. Noordhuizen, and H. W. Ploeger.** 2001. Epidemiology of *Cryptosporidium* spp. and *Giardia duodenalis* on a dairy farm. Vet Parasitol **102:**53-67.

84. **Hughes, F. B., S. T. Faehnle, and J. L. Simon.** 1975. Multiple cerebral abscesses complicating hepatopulmonary amebiasis. J Pediatr **86:**95-6.

85. **Hung, C. C., H. Y. Deng, W. H. Hsiao, S. M. Hsieh, C. F. Hsiao, M. Y. Chen, S. C. Chang, and K. E. Su.** 2005. Invasive amebiasis as an emerging parasitic disease in patients with human immunodeficiency virus type 1 infection in Taiwan. Arch Intern Med **165:**409-15.

86. **Irusen, E. M., T. F. H. G. Jackson, and A. E. Simjee.** 1992. Asymptomatic intestinal colonization by pathogenic *Entamoeba histolytica* in amebic liver abscess - prevalence, response to therapy, and pathogenic potential. Clin Infect Dis **14:**889-893.

87. **Jackson, T. F. H. G., and Suprasad.** 1997. Zymodeme stability of *Entamoeba histolytica* and *E. dispar*. Arch Med Res **28:**S304-S305.

88. **Jain, P., S. Sawhney, and V. K. Vinayak.** 1980. Experimental amoebic infection in guinea-pigs immunized with low grade amoebic infection. Trans R Soc Trop Med Hyg **74:**347-50.

89. **Jimenez Cardoso, J. M., E. Jimenez, M. de Jesus Bernal, and J. Kumate.** 1989. Induction of protective antiamebic immunity in hamsters with heterologous antigens. Rev Invest Clin **41:**133-8.

90. **Jones, W. R.** 1946. The experimental infection of rats with *Entamoeba histolytica*; with a method for evaluating the anti-amoebic properties of new compounds. Ann Trop Med Parasitol **40:**130-140.

91. **Karki, B. M. S., and S. C. Parija.** 1999. Co-agglutination test for the detection of circulating antigen in amebic liver abscess. Am J Trop Med Hyg **60:**498-501.

92. **Katzenstein, D., V. Rickerson, and A. Braude.** 1982. New concepts of amebic liver abscess derived from hepatic imaging, serodiagnosis, and hepatic enzymes in 67 consecutive cases in San Diego. Medicine (Baltimore) **61:**237-46.

93. **Katzwinkel-Wladarsch, S., T. Loscher, and H. Rinder.** 1994. Direct amplification and differentiation of pathogenic and nonpathogenic *Entamoeba histolytica* DNA from stool specimens. Am J Trop Med Hyg **51:**115-118.

94. **Kaur, R., D. Rawat, M. Kakkar, B. Uppal, and V. K. Sharma.** 2002. Intestinal parasites in children with diarrhea in Delhi, India. Southeast Asian J Trop Med Public Health **33:**725-9.

95. **Kebede, A., J. Verweij, W. Dorigo-Zetsma, E. Sanders, T. Messele, L. van Lieshout, B. Petros, and T. Polderman.** 2003. Overdiagnosis of amoebiasis in the absence of *Entamoeba histolytica* among patients presenting with diarrhoea in Wonji and Akaki, Ethiopia. Trans R Soc Trop Med Hyg **97:**305-7.

96. **Kebede, A., J. J. Verweij, T. Endeshaw, T. Messele, G. Tasew, B. Petros, and A. M. Polderman.** 2004. The use of real-time PCR to identify *Entamoeba histolytica* and *E. dispar* infections in prisoners and primary-school children in Ethiopia. Ann Trop Med Parasitol **98:**43-8.

97. **Köhler, S., and E. Tannich.** 1993. A family of transcripts (K2) of *Entamoeba histolytica* contains polymorphic repetitive regions with highly conserved elements. Mol Biochem Parasitol **59:**49-58.

98. **Kroft, E. B., A. Warris, L. E. Jansen, and R. van Crevel.** 2005. A Dutchman from Mali with a perianal ulcer caused by cutaneous amoebiasis. Ned Tijdschr Geneeskd **149:**308-11.

99. **Kuroki, T., S. Yamai, and T. Koyama.** 1989. *Entamoeba polecki* infection in a Southeast Asian refugee in Japan. Jpn J Med Sci Biol **42:**25-9.

100. **Leclerc, M. C., M. Menegon, A. Cligny, J. L. Noyer, S. Mammadov, N. Aliyev, E. Gasimov, G. Majori, and C. Severini.** 2004. Genetic diversity of *Plasmodium vivax* isolates from Azerbaijan. Malar J **3:**40.

101. **Lee, J., G. M. Park, D. H. Lee, S. J. Park, and T. S. Yong.** 2000. Intestinal

parasite infections at an institution for the handicapped in Korea. Korean J Parasitol **38:**179-181.

101a. **Levine, N.D., J.O. Corliss, F.E.G. Cox, G. Deroux, J. Grain, B.M. Honigberg, G.F. Leedale, A.R. Loeblich 3rd, J. Lom, D. Lynn, E.G. Merinfeld, F.C. Page, G. Poljansky, V. Sprague, J. Vavra, and F.G. Wallace.** 1980 A newly revised classification of the protozoa. J Protozool **27**:37-58.

102. **Lichtenstein, A., A. T. Kondo, G. S. Visvesvara, A. Fernandez, E. F. Paiva, T. Mauad, M. Dolhnikoff, and M. A. Martins.** 2005. Pulmonary amoebiasis presenting as superior vena cava syndrome. Thorax **60:**350-2.

103. **Loftus, B., I. Anderson, R. Davies, U. C. Alsmark, J. Samuelson, P. Amedeo, P. Roncaglia, M. Berriman, R. P. Hirt, B. J. Mann, T. Nozaki, B. Suh, M. Pop, M. Duchene, J. Ackers, E. Tannich, M. Leippe, M. Hofer, I. Bruchhaus, U. Willhoeft, A. Bhattacharya, T. Chillingworth, C. Churcher, Z. Hance, B. Harris, D. Harris, K. Jagels, S. Moule, K. Mungall, D. Ormond, R. Squares, S. Whitehead, M. A. Quail, E. Rabbinowitsch, H. Norbertczak, C. Price, Z. Wang, N. Guillen, C. Gilchrist, S. E. Stroup, S. Bhattacharya, A. Lohia, P. G. Foster, T. Sicheritz-Ponten, C. Weber, U. Singh, C. Mukherjee, N. M. El-Sayed, W. A. Petri, Jr., C. G. Clark, T. M. Embley, B. Barrell, C. M. Fraser, and N. Hall.** 2005. The genome of the protist parasite *Entamoeba histolytica.* Nature **433:**865-8.

104. **Lohia, A., N. Haider, and B. B. Biswas.** 1990. Characterisation of a repetitive DNA family from *Entamoeba histolytica* containing Saccharomyces cerevisiae ARS consensus sequences. Gene **96:**197-203.

105. **Lotter, H., H. Russmann, J. Heesemann, and E. Tannich.** 2004. Oral vaccination with recombinant *Yersinia enterocolitica* expressing hybrid type III proteins protects gerbils from amebic liver abscess. Infect Immun **72:**7318-21.

106. **Machado, C. A., and F. J. Ayala.** 2002. Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of *Trypanosoma cruzi.* Mol Biochem Parasitol **121:**33-47.

107. **Machado, R. L., M. M. Povoa, V. S. Calvosa, M. U. Ferreira, A. R. Rossit, E. J. dos Santos, and D. J. Conway.** 2004. Genetic structure of *Plasmodium falciparum* populations in the Brazilian Amazon region. J Infect Dis **190:**1547-55.

108. **Magana, M., M. L. Magana, A. Alcantara, and M. A. Perez-Martin.** 2004. Histopathology of cutaneous amebiasis. Am J Dermatopathol **26:**280-4.

109. **Martínez-Palomo, A.** 1993. Parasitic amebas of the intestinal tract, p. 65-141. *In* J. P. Kreier and J. R. Baker (ed.), Parasitic protozoa, 2nd ed, vol. 3. Academic Press, San Diego, CA.

110. **Martinez-Reyes, R., M. Gonzalez-Pacheco, I. Gomez-Reyes, M. de la Torre, and B. Sepulveda.** 1980. Induction of immunity to amebiasis in subhuman primates by lysosomal antigen of *Entamoeba histolytica*. VI. Intrahepatic inoculation of virulent *E. histolytica* in monkeys. Arch Invest Med (Mex) **11**:267-73.

111. **McAuley, J. B., B. L. Herwaldt, S. L. Stokes, J. A. Becher, J. M. Roberts, M. K. Michelson, and D. D. Juranek.** 1992. Diloxanide furoate for treating asymptomatic *Entamoeba histolytica* cyst passers - 14 Years' experience in the United States. Clin Infect Dis **15**:464-468.

112. **McAuley, J. B., and D. D. Juranek.** 1992. Paromomycin in the treatment of mild-to-moderate intestinal amebiasis. Clin Infect Dis **15**:551-552.

113. **McMillan, A., H. M. Gilmour, G. McNeillage, and G. R. Scott.** 1984. Amoebiasis in homosexual men. Gut **25**:356-360.

114. **Mehta, A. B., B. C. Mehta, S. L. Balse, and J. C. Patel.** 1968. Amoebic abscess of myocardium; a case report. Ind J Med Sci **22**:720-2.

115. **Michel, B., A. Alagon, P. Lizardi, and M. Zurita.** 1992. Characterization of a repetitive DNA element from *Entamoeba histolytica*. Mol Biochem Parasitol **51**:165-168.

116. **Mittal, V., A. Bhattacharya, and S. Bhattacharya.** 1994. Isolation and characterization of a species-specific multicopy DNA sequence from *Entamoeba histolytica*. Parasitology **108**:237-244.

117. **Mohandas, R. Sehgal, A. Sud, and N. Malla.** 2002. Prevalence of intestinal parasitic pathogens in HIV-seropositive individuals in Northern India. Jpn J Infect Dis **55**:83-4.

118. **Newton-Sanchez, O. A., K. Sturm-Ramirez, J. L. Romero-Zamora, J. I. Santos-Preciado, and J. Samuelson.** 1997. High rate of occult infection with *Entamoeba histolytica* among non-dysenteric Mexican children. Arch Med Res **28**:311-3.

119. **Ohnishi, K., Y. Kato, A. Imamura, M. Fukayama, T. Tsunoda, Y. Sakaue, M. Sakamoto, and H. Sagara.** 2004. Present characteristics of symptomatic *Entamoeba histolytica* infection in the big cities of Japan. Epidemiol Infect **132**:57-

60.

120. **Paglia, M. G., and P. Visca.** 2004. An improved PCR-based method for detection and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* in formalin-fixed stools. Acta Trop **92:**273-7.

121. **Park, S. J., T. S. Yong, J. K. Chang, K. Im, and C. T. Soh.** 1999. Characterization of YS-27, an axenic Korean strain of *Entamoeba histolytic*a. Korean J Parasitol **37:**59-62.

122. **Park, S. K., D. H. Kim, Y. K. Deung, H. J. Kim, E. J. Yang, S. J. Lim, Y. S. Ryang, D. Jin, and K. J. Lee.** 2004. Status of intestinal parasite infections among children in Bat Dambang, Cambodia. Korean J Parasitol **42:**201-3.

123. **Petri, W. A., Jr.** 1996. Recent advances in amebiasis. Crit Rev Clin Lab Sci **33:**1-37.

124. **Petri, W. A., Jr.** 2000. Protozoan parasites that infect the gastrointestinal tract. Curr Opin Gastroenterol **16:**18-23.

125. **Prakash, A., A. Chakraborti, R. C. Mahajan, and N. K. Ganguly.** 2002. DNA polymorphism in North Indian isolates of *Entamoeba histolytica* detected by PCR fingerprinting. Parasitol Res **88:**126-129.

126. **Ramakrishnan, A. S., P. M. Ratnasabapathy, R. Natanasabapathy, L. Ananthakrishnan, and A. R. Balakrishnan.** 1971. An amoebic liver abscess presenting as a space occupying lesion of the right kidney. Am Surg **37:**756-8.

127. **Ramdial, P. K., T. E. Madiba, S. Kharwa, B. Clarke, and B. Zulu.** 2002. Isolated amoebic appendicitis. Virchows Arch **441:**63-68.

128. **Ravdin, J. I.** 1995. Amebiasis (Review). Clin Infect Dis **20:**1453-1466.

129. **Rayner, J. C., C. S. Huber, D. Feldman, P. Ingravallo, M. R. Galinski, and J. W. Barnwell.** 2004. *Plasmodium vivax* merozoite surface protein PvMSP-3 beta is radically polymorphic through mutation and large insertions and deletions. Infect Genet Evol **4:**309-19.

130. **Richards, C. S., M. Goldman, and L. T. Cannon.** 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.

131. **Robinson, G. L.** 1968. The laboratory diagnosis of human parasitic amoebae. Trans R Soc Trop Med Hyg **62:**285-294.

132. **Romero, J. L., S. Descoteaux, S. Reed, E. Orozco, J. Santos, and J. Samuelson.**

1992. Use of polymerase chain reaction and nonradioactive DNA probes to diagnose *Entamoeba histolytica* in clinical samples. Arch Med Res **23:**277-279.

133. **Roy, S., M. Kabir, D. Mondal, I. K. M. Ali, W. A. Petri, Jr., and R. Haque.** 2005. Real-Time-PCR assay for diagnosis of *Entamoeba histolytica* infection. J Clin Microbiol **43:**000-000.

134. **Saab, B. R., U. Musharrafieh, N. T. Nassar, M. Khogali, and G. F. Araj.** 2004. Intestinal parasites among presumably healthy individuals in Lebanon. Saudi Med J **25:**34-7.

135. **Safitri, I., A. Jalloh, I. S. Tantular, S. Pusarawati, T. T. Win, Q. Liu, M. U. Ferreira, Y. P. Dachlan, T. Horii, and F. Kawamoto.** 2003. Sequence diversity in the amino-terminal region of the malaria-vaccine candidate serine repeat antigen in natural *Plasmodium falciparum* populations. Parasitol Int **52:**117-31.

136. **Samuelson, J. C., A. Burke, and J. M. Courval.** 1992. Susceptibility of an emetine-resistant mutant of *Entamoeba histolytica* to multiple drugs and to channel blockers. Antimicrob Agents Chemother **36:**2392-2397.

137. **Sargeaunt, P. G., J. E. Williams, and J. D. Grene.** 1978. The differentiation of invasive and non-invasive *Entamoeba histolytica* by isoenzyme electrophoresis. Trans R Soc Trop Med Hyg **72:**519-521.

138. **Sargeaunt, P. G., T. F. H. G. Jackson, S. R. Wiffen, and R. Bhojnani.** 1988. Biological evidence of genetic exchange in *Entamoeba histolytica*. Trans R Soc Trop Med Hyg **82:**862-867.

139. **Sargeaunt, P. G., S. Patrick, and D. O'Keefe.** 1992. Human infections of *Entamoeba chattoni* masquerade as *Entamoeba histolytica*. Trans R Soc Trop Med Hyg **86:**633-634.

140. **Schmutzhard, E., U. Mayr, E. Rumpl, M. Prugger, and P. Pohl.** 1986. Secondary cerebral amoebiasis due to infection with *Entamoeba histolytica*. A case report with computer tomographic findings. Eur Neurol **25:**161-5.

141. **Schwenkenbecher, J. M., C. Frohlich, F. Gehre, L. F. Schnur, and G. Schonian.** 2004. Evolution and conservation of microsatellite markers for *Leishmania tropica*. Infect Genet Evol **4:**99-105.

142. **Sehgal, D., V. Mittal, S. Ramachandran, S. K. Dhar, A. Bhattacharya, and S. Bhattacharya.** 1994. Nucleotide sequence organization and analysis of the nuclear ribosomal DNA circle of the protozoan parasite *Entamoeba histolytica*. Mol Biochem Parasitol **67:**205-214.

143. **Sekar, V.** 1987. A Rapid Screening Procedure for the Identification of Recombinant Bacterial Clones. BioTechniques **5:**11-13.

144. **Sepulveda, B.** 1980. Induction of immunity to amebiasis in subhuman primates by lysosomal antigen of *Entamoeba histolytica.* VII. Comments and conclusions. In Spanish. Arch Invest Med (Mex) **11:**275-6.

145. **Shah, P. H., R. C. MacFarlane, D. Bhattacharya, J. C. Matese, J. Demeter, S. E. Stroup, and U. Singh.** 2005. Comparative genomic hybridizations of *Entamoeba* strains reveal unique genetic fingerprints that correlate with virulence. Eukaryot Cell **4:**504-15.

146. **Sharma, M. P., and V. Ahuja.** 2003. Amebiasis (Letter). N Engl J Med **349:**307-308.

147. **Sibley, L. D., D. G. Mordue, C. Su, P. M. Robben, and D. K. Howe.** 2002. Genetic approaches to studying virulence and pathogenesis in *Toxoplasma gondii.* Philos Trans R Soc Lond B Biol Sci **357:**81-8.

148. **Silva, E. F., M. A. Gomes, and A. M. Barral de Martinez.** 1997. *Entamoeba histolytica*: axenization and characterization of isolated samples from symptomatic and asymptomatic patients from different regions of Brazil. Arch Med Res **28 Spec No:**288-9.

149. **Simjee, A. E., V. Gathiram, T. F. H. G. Jackson, and B. F. Khan.** 1985. A comparative trial of metronidazole v. tinidazole in the treatment of amoebic liver abscess. S Afr Med J **68:**923-4.

150. **Sirivichayakul, C., C. Pojjaroen-anant, P. Wisetsing, C. Siripanth, P. Chanthavanich, and K. Pengsaa.** 2003. Prevalence of intestinal parasitic infection among Thai people with mental handicaps. Southeast Asian J Trop Med Public Health **34:**259-63.

151. **Snyders, F. A., and H. G. Welke.** 1989. Amoebic lung abscesses. A case report. S Afr Med J **75:**344-6.

152. **Stanley, S. L., Jr., A. Becker, C. Kunz-Jenkins, L. Foster, and E. Li.** 1990. Cloning and expression of a membrane antigen of *Entamoeba histolytica* possessing multiple tandem repeats. Proc Natl Acad Sci USA **87:**4976-4980.

153. **Stanley, S. L., Jr.** 2003. Amoebiasis. Lancet **361:**1025-1034.

154. **Strachan, W. D., W. M. Spice, P. L. Chiodini, A. H. Moody, and J. P. Ackers.** 1988. Immunological differentiation of pathogenic and non-pathogenic isolates of *Entamoeba histolytica.* Lancet **i:**561-563.

155. **Strong, W. B., J. Gut, and R. G. Nelson.** 2000. Cloning and sequence analysis of a highly polymorphic *Cryptosporidium parvum* gene encoding a 60-kilodalton glycoprotein and characterization of its 15- and 45-kilodalton zoite surface antigen products. Infect Immun **68:**4117-34.

156. **Tan, T. H., R. Pach, A. Crausaz, A. Ivens, and A. Schneider.** 2002. tRNAs in *Trypanosoma brucei*: genomic organization, expression, and mitochondrial import. Mol Cell Biol **22:**3707-17.

157. **Tannich, E., and G. D. Burchard.** 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.

158. **Teramoto, K., N. Yamashita, M. Kuwabara, T. Hanawa, T. Matsui, and Y. Matsubara.** 2001. An amebic lung abscess: report of a case. Surg Today **31:**820-822.

159. **Tonon, A. P., E. H. Hoffmann, L. A. Silveira, A. G. Ribeiro, C. R. Goncalves, P. E. Ribolla, G. Wunderlich, and M. U. Ferreira.** 2004. *Plasmodium falciparum*: sequence diversity and antibody recognition of the Merozoite surface protein-2 (MSP-2) in Brazilian Amazonia. Exp Parasitol **108:**114-25.

160. **Urdaneta, H., M. Rondon, M. Muñoz, and M. Hernández.** 1995. Isolation and axenization of two *Entamoeba histolytica* strains. G E N **49:**23-28.

161. **Verweij, J. J., A. M. Polderman, and C. G. Clark.** 2001. Genetic variation among human isolates of uninucleated cyst-producing *Entamoeba* species. J Clin Microbiol **39:**1644-1646.

162. **Vohra, H., H. S. Bhatti, N. K. Ganguly, and R. C. Mahajan.** 1989. Virulence of pathogenic and non-pathogenic zymodemes of *Entamoeba histolytica* (Indian strains) in guinea pigs. Trans R Soc Trop Med Hyg **83:**648-650.

163. **Walsh, J. A.** 1986. Problems in recognition and diagnosis of amebiasis: Estimation of the global magnitude of morbidity and mortality. Rev Infect Dis **8:**228-238.

164. **Wang, X. Y., X. L. Shi, and B. L. Hao.** 2004. The tRNA and rRNA genes in the Oryza sativa genome. Yi Chuan Xue Bao **31:**871-7.

165. **Widmer, G., X. Feng, and S. Tanriverdi.** 2004. Genotyping of *Cryptosporidium parvum* with microsatellite markers. Methods Mol Biol **268:**177-87.

166. **Zaki, M., and C. G. Clark.** 2001. Isolation and characterization of polymorphic DNA from *Entamoeba histolytica*. J Clin Microbiol **39:**897-905.

167. **Zaki, M., P. Meelu, W. Sun, and C. G. Clark.** 2002. Simultaneous differentiation and typing of *Entamoeba histolytica* and *Entamoeba dispar*. J Clin Microbiol **40:**1271-1276.

168. **Zaki, M., S. G. Reddy, T. F. H. G. Jackson, J. I. Ravdin, and C. G. Clark.** 2003a. Genotyping of *Entamoeba* species in South Africa: Diversity, stability and transmission patterns within families. J Infect Dis **187:**1860-1869.

169. **Zaki, M., J. J. Verweij, and C. G. Clark.** 2003b. *Entamoeba histolytica*: direct PCR-based typing of strains using faecal DNA. Exp Parasitol **104:**77-80.

170. **Zhang, T., E. Li, and S. L. Stanley, Jr.** 1995. Oral immunization with the dodecapeptide repeat of the serine-rich *Entamoeba histolytica* protein (SREHP) fused to the cholera toxin B subunit induces a mucosal and systemic anti-SREHP antibody response. Infect Immun **63:**1349-1355.

171. **Zhang, T., and S. L. Stanley, Jr.** 1996. Oral immunization with an attenuated vaccine strain of *Salmonella typhimurium* expressing the serine-rich *Entamoeba histolytica* protein induces an antiamebic immune response and protects gerbils from amebic liver abscess. Infect Immun **64:**1526-1531.

APPENDIX I

Appendix I. tRNA primers.

Array name	STR name§	Primer name§	Sequence (5' to 3')	AT
[A^{AGC}]	A-A	A-A5	CATTCGCTTAGCATGCAGGG	54
		A-A3	GCTCTCCATCTGAGCTACAT	
[ALL]	A-L	A-L5	GGATCGATACCCCTCATCTCCA	64
		A-L3	CGCATCTTGCGATAGCCGAG	
	L-A	L-A5	TCTAAGGGGGCCGATTTAAGCT	62
		L-A3	CTTCGCATGCGAAGCGAA	
[ASD]	A-S	A-S5	GGGGGTGTAGCTCAGATGGTA	59
		A-S3	CTGCTAAGGAGTTTCCATTTGG	
	S^{GCT}-D	S^{GCT}-D5	CCAAATGGAAACTCCTTAGCAG	58
		S^{GCT}-D3	CACGTGACAGGCGAAGATAC	
	D-A	D-A5 (EhR1)$	CTGGTTAGTATCTTCGCCTGT	56
		D-A3 (EhR2)$	CTTACACCCCCATTAACAAT	
[G^{GCC}]	G^{GCC}-G	G^{GCC}-G5	TTCGATTCCCGGCAAGTGCA	59
		G^{GCC}-G3	GGCAGGGATACGTCATACCA	
[G^{TCC}]	G^{TCC}-G	G^{TCC}-G5	GTTCGAGTCCCGGCAAAT	62
		G^{TCC}-G3	TGCCCTTAGACCACAAATGC	
[H^{GTG}]	H-H	H-H5	GATTCCTGGTGTGAGCATC	56
		H-H3	GCCACAATCTGATGTACTACCTC	
[LS]	L-S	L-S5	TACGGATCTCTTCGGAGGCG	61
		L-S3	CACACGCCTTAGACCACTCG	
	S-L	S-L5	CGAGTGGTCTAAGGCGTGTG	62
		S-L3	AGACCACTCGGCCACTCTTG	
[LT]	L-T	L-T5	CCCACCTCTTGCATTTTA	55
		L-T3	CCATCTATGCTAAGAGGCC	
	T-L	T-L5	AGGCCTCTTAGCATAGATGG	54
		T-L3	CCTCCGAAGAGAACAGTACC	
[MR]	M-R	M-R5	TACGCTCTACCAACTGAGCTAC	56
		M-R3	TCGATCCCTGCTATGCTC	
	R-M	R-M5	CAATGGAACCGCATCAGACT	58
		R-M3	TGAGTATCGAACTCAAGACC	
[NK]	N-K	N-K5	CGAACGGCTGTTAACCGTTA	55
		N-K3	TTCCTAGCTCAGTCGGTAGA	
	K-N	K-N5	TCTACCGACTGAGCTAGGAAGG	60
		K-N3	TTCGCTCTGCCGACTGAG	
[P^{TGG}]	P-P	P-P5	GTGGTAAAATACCCGCTTTGGGTGCG	59
		P-P3	GGAATCGAACCCGGGCCACC	
[R5]	R-5	R-55	CTGGGGATTGCAGGTTCGATC	64
		R-53	TCACACAGATGGGAACGTCAGG	
	5-R	5-R5	CCTGACGTTCCCATCTGTGTGA	64
		5-R3	CGTTTCCATTGCGCCACA	
[R^{TCT}]	R-R	R-R5	AGCATCAGCCTTCTAAGCTG	55
		R-R3	CTTCCGACTGAGCTAACAAG	
[RT]	R-T	R-T5	CCTAAGCAGAAGATCGTGGGTTCG	64
		R-T3	GAATCGAACCCACGACCTTTCC	
	T-R	T-R5	GGGTTCGATTCCCGCAGAGG	65

		T-R3	CGAACCCACGATCTTCTGCTTAGG	
[SD]	S^{TGA}-D	S^{TGA}-D5	CTCTGGATGCGTAGGTTCAA	58
		S^{TGA}-D3	GTATCTTCGCCTGTCACGTG	
	D-S	D-S5	CACGTGACAGGCGAAGATAC	58
		D-S3	TTGAACCTACGCATCCAGAG	
[SPPCK] & [SQCK]	S-P	S-P5	TCTGGAGGCGTAGGTTCGAA	61
		S-P3	GTGGCATAATACTTGCTTCGGG	
	P-C	P-C5	ATCGAACCCAGGCCACTTGC	63
		P-C3	GCGGTAGAGCATAGGACTGCAG	
	C-K	C-K5	AATCGGATTTGAACCAATGA	54
		C-K3	TCCTAGCTCAGTCGGTAGAG	
	K-S	K-S5	CCCACGACCTTACGGTTAAA	59
		K-S3	AACGGCAGGATTCGAACCTA	
	S-Q	S-Q5	GTGGTCTAAGGCGTGTGACT	56
		S-Q3	GAGATTCTGGTTCTTAGGACCC	
	Q-C	Q-C5	CTGCCACTACACCATGGAAC	56
		Q-C3	GTGGTCATTGGTTCAAATCC	
[TQ]	T-Q	T-Q5	GGTCCCATGACCTTCTCTTT	56
		T-Q3	ATTCGAACTGGGGTCCTAAG	
	Q-T	Q-T5	TAGGACCCCAGTTCGAATCT	57
		Q-T3	AAAGAGAAGGTCATGGGACC	
[TX]	T-X	T-X5	GGCGGTTTAGCTTAGCGGTAGA	62
		T-X3	CTCACGCAGTACCCTGTTCGT	
	X-T	X-T5	ACGAACAGGGTACTGCGTGAG	62
		X-T3	GGCGACTATGGGGATCGAAC	
[V5]	V-5	V-55	CACATCACTTTTACGTAGTGAGG	55
		V-53	ATAATGGTACTGATCAGGCTCT	
	5-V	5-V5	AGAGCCTGATCAGTACCATTAT	55
		5-V3	AACCAACTACACCATGAAACC	
[VF]	V-F	V-F5	GGTTTCATGGTGTAGTTGGT	55
		V-F3	GAACTTCAGTCTAACGCTGC	
	F-V	F-V5	TAAAAGGGGGAAGAATAGGAA	55
		F-V3	GGTTTCATGGTGTAGTTGGT	
[VME5]	V-M	V-M5	CACTTTCACGTAGTGAGGGTCC	59
		V-M3	CAGGTCGGTTGATCGAAACT	
	M-E	M-E5	AGTTATGAGCCTGCCGCG	62
		M-E3	GTAACATGCAAGCCTCTCACG	
	E-V	E-V5	GTGAGAGGCTTGCATGTTACC	59
		E-V3	GGACCCTCACTACGTGAAAGTG	
[WI]	W-I	W-I5	GCCGGTCAGTGGTTCAATCC	63
		W-I3	GCGACCCTGGCGTTATTAGC	
	I-W	I-W5 (R6A)$	CTCAGTCGGTAGAGCATGGT	56
		I-W3 (R5A)$	CTAAAGCCCCCTTCTTCTATAATT	
[YE]	Y-E	Y-E5	CCTTAGGTCACTGGTTCGAA	55
		Y-E3	CACAAACGTGAAAGGCTTGT	
	E-Y	E-Y5	TACATAAGTCGTGGTAAAGAGAAG	52
		E-Y3	CTACATCTACAGTCCTCCGCT	

§ = The anticodon sequence of the tRNA is used in the identification where necessary; AT = Annealing temperature; X = G+C rich region of unknown function; $ = these primers were previously described by Zaki et al (2002 and 2003a) using the different names given inside the brackets; ¥ = this primer is adjacent to but not within the tRNA indicated.

APPENDIX II

Appendix II. Genotypes of all Bangladeshi *E. histolytica* using 6 loci.

Serial	ID	Clinical status	*E. histolytica* specific PCR at STR locus						Genotype
			D-A	A-L	R-R	N-K2	S-D	S-Q	
1	1354-MS69	A	2	3	3	9	4	2	7
2	*3402-MS34	A	2	3	3	9	6	3	8
3	3346-FS7.2	A	2	3	4	4	6	3	12
4	3792-MS20	A	2	3	4	4	6	3	12
5	0199-MS34	A	2	3	4	6	6	3	13
6	1296-MS34	A	2	3	4	6	6	3	13
7	1681-MS34	A	2	3	4	6	6	3	13
8	0041-MS34	A	2	3	4	9	3	3	18
9	3600-MS34	A	2	3	4	9	3	3	18
10	1004-MS31	A	2	3	4	9	4	8	19
11	1325-MS33	A	2	3	4	9	4	8	19
12	3394-MS31	A	2	3	4	9	4	8	19
13	1036-MS59	A	2	4	5	9	3	2	24
14	1631-MS61	A	2	4	5	9	3	2	24
15	4309-MS10	A	2	4	5	9	9	5	25
16	4198-MS10	A	2	5	5	9	3	2	26
17	3645-MS31	A	3	2	1	6	4	3	34
18	3599-MS33	A	3	2	1	7	4	4	35
19	4194-MS10	A	3	2	2	11	5	4	37
20	0024-MS51	A	3	2	3	8	3	6	40
21	3046-MS50	A	3	2	3	8	3	6	40
22	3679-MS24	A	3	3	1	8	4	3	45
23	0055-MS23	A	3	3	3	9	3	5	51
24	0058-MS27	A	3	3	3	9	3	5	51
25	0868-FS7.7	A	3	3	3	9	3	5	51
26	3454-MS31	A	3	3	3	9	3	5	51
27	1325-MS22	A	3	3	3	9	4	3	52
28	3497-MS20	A	3	3	3	9	4	3	52
29	0291-MS32	A	3	3	4	3	2	3	58
30	3433-MS21	A	3	3	4	3	3	3	59
31	3484-MS22	A	3	3	4	6	4	3	63
32	1143-MS34	A	3	3	4	9	4	3	68
33	3193-MS20	A	3	3	4	9	6	3	69
34	3346-MS32	A	3	3	4	9	10	3	71
35	0291-MS42	A	3	3	5	1	2	2	72
36	3040-MS24	A	4	3	1	7	4	3	85
37	3222-MS19	A	4	3	4	5	3	3	87
38	3760-MS33	A	5	3	3	9	3	3	91
39	3570-DS04	D	2	3	1	8	3	5	5
40	3191-DS12	D	2	3	4	3	6	1	9
41	**34727	D	2	3	4	8	4	8	15
42	**22027	D	2	3	4	8	6	3	17
43	1049-DS04	D	2	3	4	9	4	8	19
44	0017-DS01	D	2	3	5	8	4	3	21
45	**29621	D	3	1	3	8	3	6	30
46	**2470300	D	3	2	2	5	5	7	36
47	3042-DS04	D	3	3	1	8	4	3	45
48	3646-DS09	D	3	3	1	8	4	3	45
49	4204-DS01	D	3	3	1	8	4	3	45

50	1186-DS20	D	3	3	3	1	3	3	46
51	0021-DS07	D	3	3	3	6	3	3	48
52	1186-DS19	D	3	3	3	8	3	3	49
53	0021-DS10	D	3	3	3	8	4	3	50
54	1767-DS20	D	3	3	3	8	4	3	50
55	4196-DS02	D	3	3	3	8	4	3	50
56	0015-DS05	D	3	3	3	9	3	5	51
57	0027-DS17	D	3	3	3	9	3	5	51
58	3497-DS02	D	3	3	3	9	3	5	51
59	5027-DS02	D	3	3	4	1	3	3	53
60	3570-DS28	D	3	3	4	1	3	4	54
61	5039-DS01	D	3	3	4	1	3	4	54
62	2977-DS02	D	3	3	4	1	4	3	56
63	5017-DS02	D	3	3	4	1	4	3	56
64	1057-DS11	D	3	3	4	3	2	3	58
65	1036-DS01	D	3	3	4	4	4	3	62
66	0883-DS04	D	3	3	4	8	3	3	64
67	4283-DS01	D	3	3	4	8	3	4	65
68	0027-DS13	D	3	3	4	8	4	3	66
69	0139-DS02	D	3	3	4	8	4	3	66
70	1048-DS01	D	3	3	4	8	4	3	66
71	3679-DS15	D	3	3	4	8	4	3	66
72	3742-DS02	D	3	3	4	8	4	3	66
73	4045-DS04	D	3	3	4	8	4	3	66
74	4130-DS01	D	3	3	4	8	4	3	66
75	4280-DS01	D	3	3	4	8	4	3	66
76	4281-DS01	D	3	3	4	8	4	3	66
77	5014-DS01	D	3	3	4	8	4	3	66
78	0064-DS06	D	3	3	4	9	3	3	67
79	1292-DS05	D	3	3	4	9	6	4	70
80	3047-DS08	D	3	3	4	9	10	3	71
81	5028-DS03	D	3	3	5	8	4	2	74
82	4247-DS05	D	3	4	3	5	4	5	76
83	1036-DS06	D	3	4	3	9	3	5	77
84	0868-DS04	D	3	4	5	9	3	2	79
85	3645-DS14	D	4	2	1	4	5	4	80
86	3570-DS14	D	4	3	3	8	4	2	86
87	**34614	D	4	4	1	4	4	5	89
88	**32311	D	5	3	3	3	11	2	90
89	LAID-02	L	2	2	1	5	2	6	1
90	LAID-41	L	2	2	4	8	4	8	2
91	LAID-39	L	2	2	6	8	9	4	3
92	LAID-33	L	2	3	1	8	3	4	4
93	LAID-06	L	2	3	1	8	4	4	6
94	LAID-03	L	2	3	4	3	6	4	10
95	LAN-67	L	2	3	4	4	4	4	11
96	LAID-29	L	2	3	4	8	4	6	14
97	LAN-27	L	2	3	4	8	4	9	16
98	LAID-37	L	2	3	5	8	3	4	20
99	LAID-11	L	2	4	5	4	1	4	22
100	LAID-19	L	2	4	5	8	3	2	23
101	LAID-38	L	2	4	5	8	3	2	23
102	LAID-25	L	3	1	1	4	5	4	27
103	LAN-50	L	3	1	1	5	5	3	28
104	LAP-01	L	3	1	1	5	5	4	29
105	LAID-08	L	3	2	1	4	6	5	31
106	LAID-31	L	3	2	1	4	7	4	32
107	LAN-47	L	3	2	1	5	5	4	33

108	LAN-48	L	3	2	3	8	2	4	38
109	LAN-62	L	3	2	3	8	3	4	39
110	LAN-68	L	3	2	3	8	3	4	39
111	LAID-12	L	3	2	3	8	4	4	41
112	LAID-05	L	3	2	3	8	4	6	42
113	LAN-58	L	3	2	4	8	3	1	43
114	LAN-25	L	3	2	4	8	3	4	44
115	LAN-51	L	3	3	3	4	4	3	47
116	LAID-24	L	3	3	4	1	4	2	55
117	LAID-10	L	3	3	4	1	4	4	57
118	LAID-13	L	3	3	4	3	6	3	60
119	LAN-39	L	3	3	4	4	3	3	61
120	LAID-14	L	3	3	4	8	4	3	66
121	LAID-34	L	3	3	5	1	5	4	73
122	LAN-42	L	3	3	5	8	4	2	74
123	LAN-15	L	3	3	5	8	10	4	75
124	LAN-33	L	3	4	4	2	4	11	78
125	LAP-02	L	4	2	3	4	3	4	81
126	LAID-40	L	4	2	3	5	3	10	82
127	LAN-57	L	4	3	1	1	4	2	83
128	LAID-18	L	4	3	1	5	3	4	84
129	LAID-36	L	4	3	4	8	3	10	88
130	LAID-01	L	5	3	4	3	2	5	92
131	LAID-32	L	5	3	4	6	8	3	93

A = asymptomatic; D = diarrhoeal / dysenteric; and L = amoebic liver abscess.

*Seven follow-up samples were available from this child. All these showed genotype 7 as of MS34 sample, but MS33 seemed to be transiently infected by another strain which had a unique genotype. The transient strain showed PCR patterns: A-L = 3, D-A =3, R-R = 3, N-K2 = 3, S-D = 3, and S-Q = 3, but this transient genotype was not considered in the analysis for simplicity. **These samples were collected from the ICDDR,B hospital.

The actual band sizes (in base pairs):

A-L	D-A	N-K2	R-R	S-D	S-Q
1=480	1=290	1=500	1=480	1=180	1=320
2=500	2=300	2=530	2=500	2=200	2=330
3=510	3=330	3=540	3=570	3=210	3=340
4=530	4=340	4=560	4=590	4=220	4=350
5=540	5=350	5=580	5=600	5=230	5=360
	6=390	6=600	6=680	6=240	6=370
		7=620		7=280	7=400
		8=630		8=290	8=410
		9=650		9=300	9=420
		10=690		10=330	10=490
		11=700		11=350	11=500

APPENDIX III

Appendix III. DAR genotypes of all Bangladeshi *E. histolytica*.

Serial	ID	Clinical status	*E. histolytica* specific PCR at STR locus			DAR genotype
			D-A	A-L	R-R	
1	1354-MS69	A	2	3	3	5
2	3402-MS34(7)	A	2	3	3	5
3	0041-MS34	A	2	3	4	6
4	0199-MS34(2)	A	2	3	4	6
5	1004-MS31	A	2	3	4	6
6	1296-MS34	A	2	3	4	6
7	1325-MS33(2)	A	2	3	4	6
8	1681-MS34(2)	A	2	3	4	6
9	3346-FS7.2	A	2	3	4	6
10	3394-MS31	A	2	3	4	6
11	3600-MS34	A	2	3	4	6
12	3792-MS20	A	2	3	4	6
13	1036-MS59	A	2	4	5	8
14	1631-MS61	A	2	4	5	8
15	4309-MS10	A	2	4	5	8
16	4198-MS10	A	2	5	5	9
17	3599-MS33	A	3	2	1	12
18	3645-MS31	A	3	2	1	12
19	4194-MS10	A	3	2	2	13
20	0024-MS51	A	3	2	3	14
21	3046-MS50	A	3	2	3	14
22	3679-MS24(2)	A	3	3	1	16
23	0055-MS23(2)	A	3	3	3	17
24	0058-MS27	A	3	3	3	17
25	0868-FS7.7	A	3	3	3	17
26	1325-MS22	A	3	3	3	17
27	3454-MS31	A	3	3	3	17
28	0291-MS32	A	3	3	4	18
29	1143-MS34	A	3	3	4	18
30	3193-MS20	A	3	3	4	18
31	3346-MS32(2)	A	3	3	4	18
32	3433-MS21	A	3	3	4	18
33	3484-MS22(2)	A	3	3	4	18
34	0291-MS42	A	3	3	5	19
35	3040-MS24	A	4	3	1	25
36	3222-MS19	A	4	3	4	27
37	3760-MS33	A	5	3	3	29
38	3570-DS04	D	2	3	1	4
39	22027	D	2	3	4	6
40	34727	D	2	3	4	6
41	1049-DS04(5)	D	2	3	4	6
42	3191-DS12	D	2	3	4	6
43	0017-DS01	D	2	3	5	7
44	29621	D	3	1	3	11
45	2470300	D	3	2	2	13
46	3042-DS04(2)	D	3	3	1	16
47	3646-DS09	D	3	3	1	16
48	4204-DS01	D	3	3	1	16
49	0015-DS05(6)	D	3	3	3	17
50	0021-DS10(2)	D	3	3	3	17

51	0027-DS17	D	3	3	3	17
52	1186-DS20(2)	D	3	3	3	17
53	1767-DS20	D	3	3	3	17
54	3497-DS02(2)	D	3	3	3	17
55	4196-DS02	D	3	3	3	17
56	0027-DS13(2)	D	3	3	4	18
57	0064-DS06(3)	D	3	3	4	18
58	0139-DS02	D	3	3	4	18
59	0883-DS04	D	3	3	4	18
60	1036-DS01	D	3	3	4	18
61	1048-DS01	D	3	3	4	18
62	1057-DS11(2)	D	3	3	4	18
63	1292-DS05(2)	D	3	3	4	18
64	2977-DS02	D	3	3	4	18
65	3047-DS08	D	3	3	4	18
66	3570-DS28	D	3	3	4	18
67	3679-DS15	D	3	3	4	18
68	3742-DS02	D	3	3	4	18
69	4045-DS04	D	3	3	4	18
70	4130-DS01	D	3	3	4	18
71	4280-DS01	D	3	3	4	18
72	4281-DS01	D	3	3	4	18
73	4283-DS01	D	3	3	4	18
74	5014-DS01	D	3	3	4	18
75	5017-DS02	D	3	3	4	18
76	5027-DS02	D	3	3	4	18
77	5039-DS01	D	3	3	4	18
78	5028-DS03	D	3	3	5	19
79	1036-DS06	D	3	4	3	20
80	4247-DS05(3)	D	3	4	3	20
81	0868-DS04	D	3	4	5	22
82	3645-DS14	D	4	2	1	23
83	3570-DS14	D	4	3	3	26
84	34614	D	4	4	1	28
85	32311	D	5	3	3	29
86	LAID-02	L	2	2	1	1
87	LAID-41	L	2	2	4	2
88	LAID-39	L	2	2	6	3
89	LAID-06	L	2	3	1	4
90	LAID-33	L	2	3	1	4
91	LAID-03	L	2	3	4	6
92	LAID-29	L	2	3	4	6
93	LAN-27	L	2	3	4	6
94	LAN-67	L	2	3	4	6
95	LAID-37	L	2	3	5	7
96	LAID-11	L	2	4	5	8
97	LAID-19	L	2	4	5	8
98	LAID-38	L	2	4	5	8
99	LAID-25	L	3	1	1	10
100	LAN-50	L	3	1	1	10
101	LAP-01	L	3	1	1	10
102	LAID-08	L	3	2	1	12
103	LAID-31	L	3	2	1	12
104	LAN-47	L	3	2	1	12
105	LAID-05	L	3	2	3	14
106	LAID-12	L	3	2	3	14
107	LAN-48	L	3	2	3	14
108	LAN-62	L	3	2	3	14
109	LAN-68	L	3	2	3	14
110	LAN-25	L	3	2	4	15

111	LAN-58	L	3	2	4	15
112	LAN-51	L	3	3	3	17
113	LAID-10	L	3	3	4	18
114	LAID-13	L	3	3	4	18
115	LAID-14	L	3	3	4	18
116	LAID-24	L	3	3	4	18
117	LAN-39	L	3	3	4	18
118	LAID-34	L	3	3	5	19
119	LAN-15	L	3	3	5	19
120	LAN-42	L	3	3	5	19
121	LAN-33	L	3	4	4	21
122	LAID-40	L	4	2	3	24
123	LAP-02	L	4	2	3	24
124	LAID-18	L	4	3	1	25
125	LAN-57	L	4	3	1	25
126	LAID-36	L	4	3	4	27
127	LAID-01	L	5	3	4	30
128	LAID-32	L	5	3	4	30

A = asymptomatic; D = diarrhoeal/dysenteric; and L = liver abscess (amoebic).
The numbers inside the brackets indicate the number of follow-up samples showing same genotypes.
Three samples (3497-MS20, 0021-DS07, and 1186-DS19) which were present in the Appendix II for 6-locus genotypes are not shown in this Appendix because follow-up samples showed identical genotypes with these 3 loci, but their 6-locus genotypes were different.

APPENDIX IV

Appendix IV. eBURST association between genotypes using 5, 4 and 3 loci.

5 locus maximum genotypes

The combined results of loci A-L, N-K2, R-R, S-D and S-Q (ANRSQ) produced the 5 locus maximum number of genotypes (91) across all Bangladeshi samples.

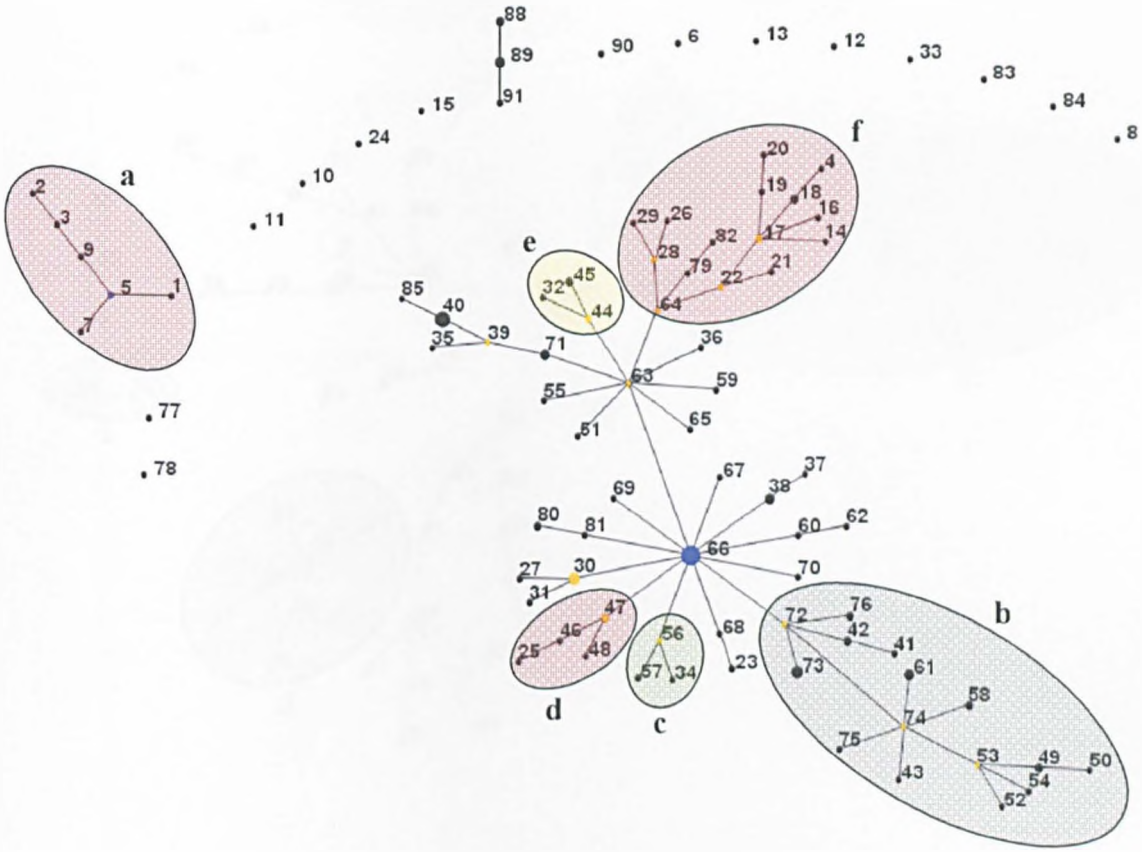
eBURST produced 2 minor groups – one with 3 genotypes (88, 89 and 91) and the other with 6 genotypes (cluster a in the Figure A1a), and one major group with 68 genotypes (Figure A1a). We detected 6 clusters – one 'a' already mentioned, and 5 others (b, c, d, e and f in the Figure A1a) which seemed to link genotypes of clinical importance. Clusters a, c, d, and f contained predominantly liver abscess strains, cluster e contained only diarrhoeal/dysenteric strains, and cluster b contained predominantly asymptomatic strains (Figure A1a).

5 locus minimum genotypes

The combined results of loci A-L, D-A, R-R, S-D and S-Q (ADRSQ) produced the 5 locus minimum number of genotypes (74) across all Bangladeshi samples.

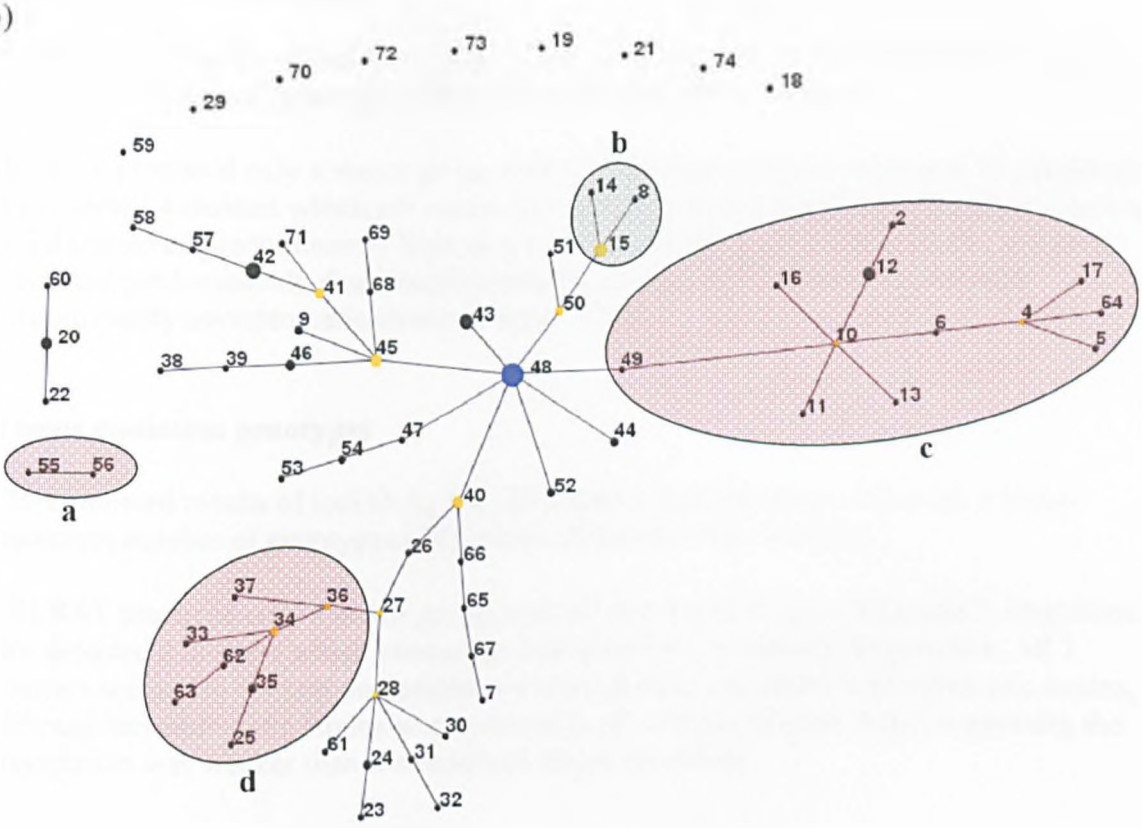
eBURST produced 2 minor groups – one with 2 genotypes (55 and 56, cluster a in the Figure A1b) and the other with 3 genotypes (20, 22 and 60), and one major group with 60 genotypes (Figure A1b). We detected 4 clusters – one 'a' already mentioned, and 3 others (b, c, and d in the Figure A1b) which seemed to link genotypes of clinical importance. Cluster a contained 2 liver abscess strains only, c and d contained predominantly liver abscess strains, while cluster b contained predominantly asymptomatic strains (Figure A1b).

a)



Clinical type	Cluster						
	a	b	c	d	e	f	Total
A	0	16	0	0	0	2	18
D	1	4	1	1	4	3	14
L	5	3	2	3	0	11	24
Total	6	23	3	4	4	16	56

b)



Clinical type	Cluster				
	a	b	c	d	Total
A	0	6	3	2	11
D	0	1	3	1	5
L	2	0	10	6	18
Total	2	7	16	9	34

Figure A1. eBURST for maximum and minimum numbers of genotypes using 5 loci. Association between genotypes observed in eBURST diagrams using 5-loci a) 'ANRSQ' (91 genotypes) and b) 'ADRSQ' (74 genotypes).

4 locus maximum genotypes

The combined results of loci D-A, N-K2, R-R and S-Q (DNRQ) produced the 4 locus maximum number of genotypes (72) across all Bangladeshi samples.

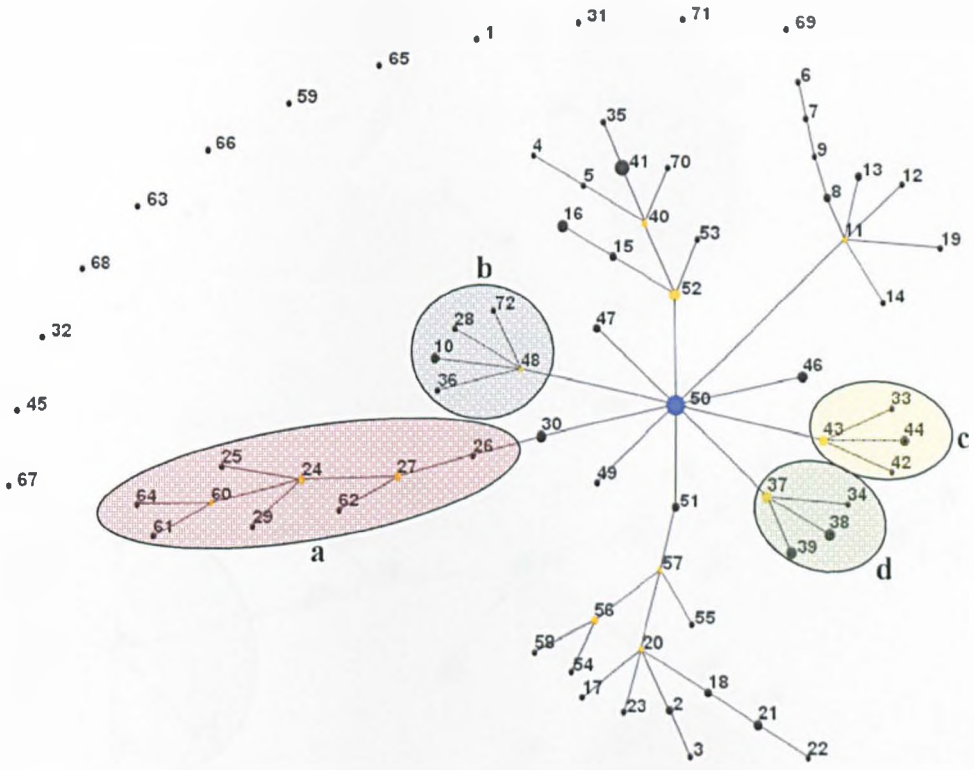
eBURST produced only a major group with 60 genotypes (Figure A2a) and 12 singletons. We detected 4 clusters which seemed to link genotypes of clinical importance. Clusters a, and d contained predominantly liver abscess and diarrhoeal/dysenteric strains, cluster c contained predominantly diarrhoeal/dysenteric strains, while cluster b contained predominantly asymptomatic strains (Figure A2a).

4 locus minimum genotypes

The combined results of loci D-A, A-L, R-R and S-D (DARS) produced the 4 locus minimum number of genotypes (54) across all Bangladeshi samples.

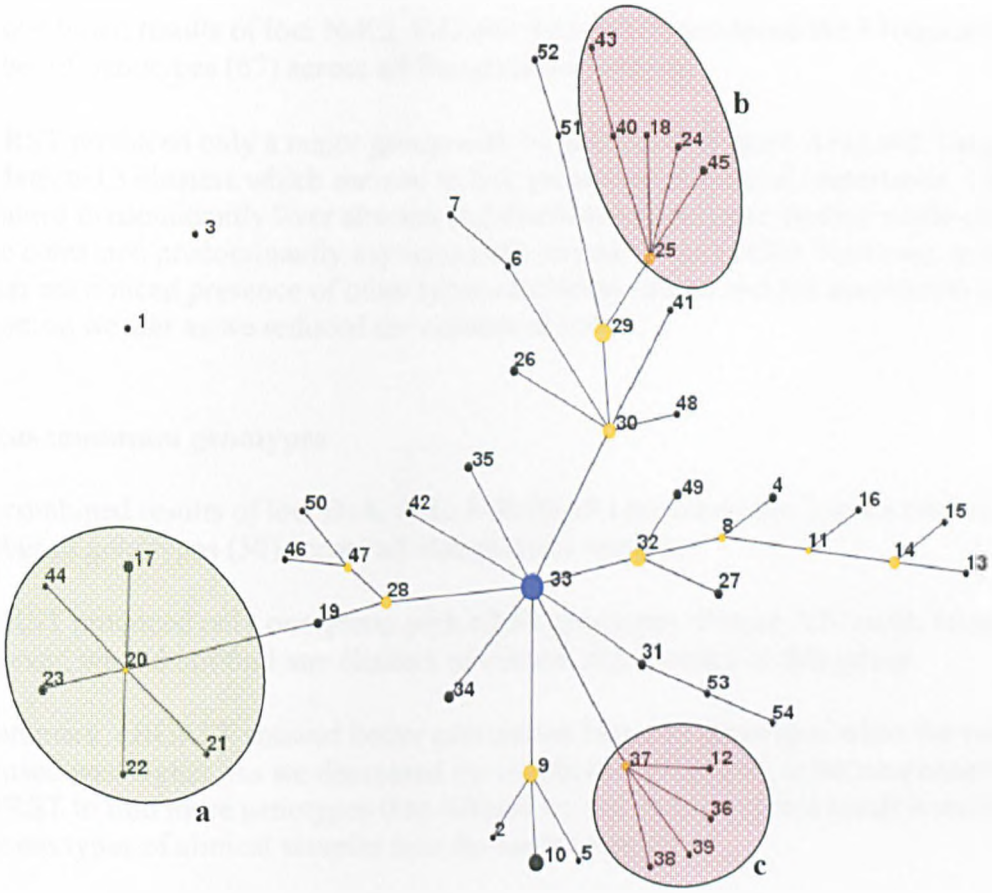
eBURST produced only a major group with 52 genotypes (Figure A2b) and 2 singletons. We detected 3 clusters which seemed to link genotypes of clinical importance. All 3 clusters seemed to contain predominantly liver abscess and diarrhoeal/dysenteric strains, although asymptomatic strains were present in all of these (Figure A2b), suggesting the association was weaker than we observed above for others.

a)



Clinical type	Cluster				
	a	b	c	d	Total
A	1	5	0	2	8
D	2	1	6	5	14
L	8	1	2	6	17
Total	11	7	8	13	39

b)



Clinical type	Cluster			
	a	b	c	Total
A	1	2	1	4
D	2	3	2	7
L	6	5	3	14
Total	9	10	6	25

Figure A2. eBURST for maximum and minimum numbers of genotypes using 4 loci. Distribution and association of genotypes observed in eBURST diagrams using 4-locus a) 'DNRQ' (72 genotypes) and b) 'DARS' (54 genotypes).

3 locus maximum genotypes

The combined results of loci N-K2, S-D and S-Q (NSQ) produced the 3 locus maximum number of genotypes (67) across all Bangladeshi samples.

eBURST produced only a major group with 64 genotypes (Figure A3a) and 3 singletons. We detected 3 clusters which seemed to link genotypes of clinical importance. Cluster a contained predominantly liver abscess and diarrhoeal/dysenteric strains, while clusters b and c contained predominantly asymptomatic strains (Figure A3a). However, in each cluster we noticed presence of other types of clinical strains and the association seemed to be getting weaker as we reduced the number of loci.

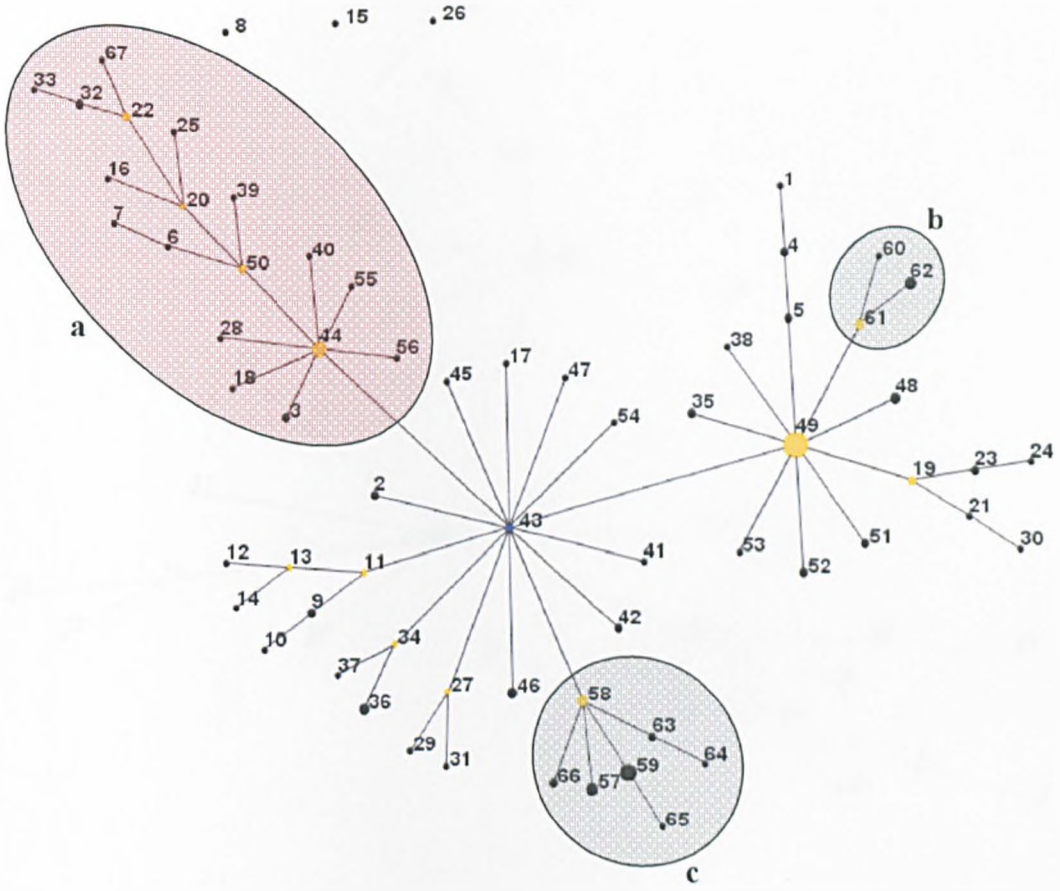
3 locus minimum genotypes

The combined results of loci D-A, A-L, R-R (DAR) produced the 3 locus minimum number of genotypes (30) across all Bangladeshi samples.

eBURST produced only one group with all 30 genotypes (Figure A3b) with no singletons. However, we did not find any clusters of clinical significance in this group.

In summary, eBURST showed better association between genotypes when the number of loci used was higher. As we decreased the number of genotypes, it became easier for the eBURST to find more genotypes that differed by one locus and as a result it was linking different types of clinical samples into the same clusters.

a)



Clinical type	Cluster			
	a	b	c	Total
A	2	7	13	22
D	5	1	8	14
L	20	0	0	20
Total	27	8	21	56

b)

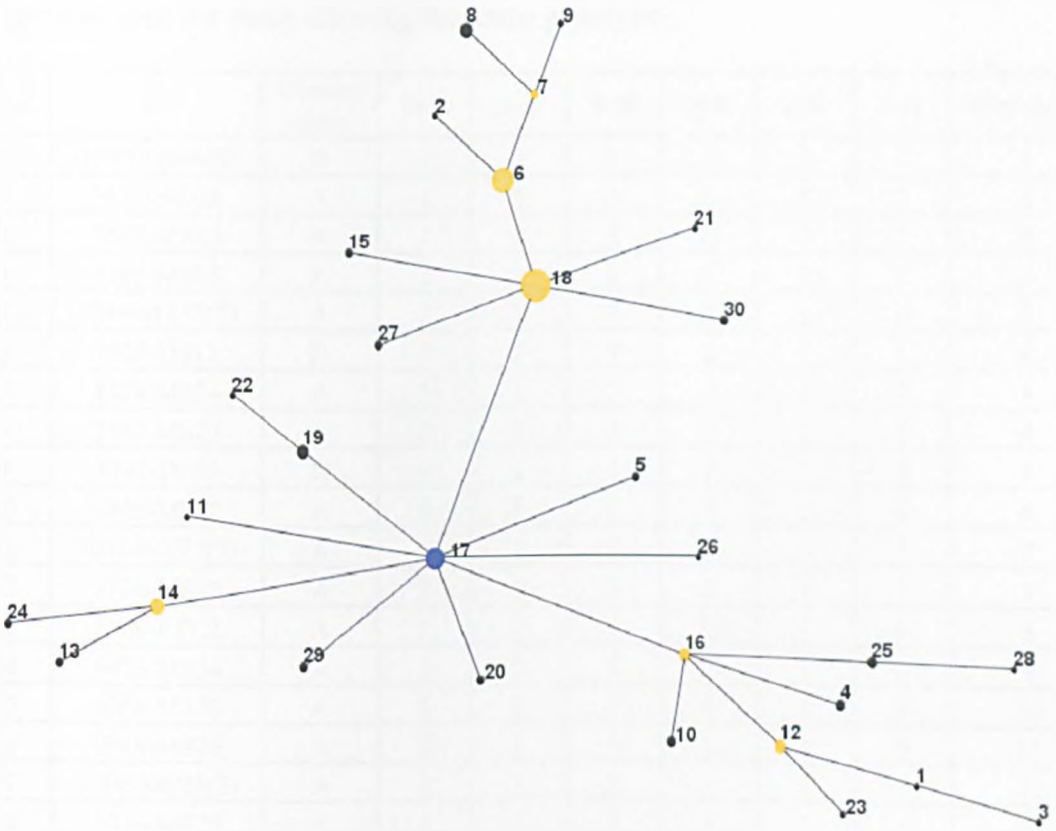


Figure A3. eBURST for maximum and minimum numbers of genotypes using 3 loci. Association of genotypes observed in eBURST diagrams using 3-loci a) 'NSQ' (67 genotypes) and b) 'DAR' (30 genotypes). No clusters of clinical significance were detected in DAR eBURST.

APPENDIX V

Appendix V. Genotypes of *E. dispar* strains using 6 loci. One representative follow-up sample was used for those showing the same genotype.

Serial	ID	Clinical status	D-A	A-L	R-R	N-K	S-D	S-Q	Genotype
1	3570-DS04(2)	D	1	1	1	9	1	7	1
2	3629-MS32	A	1	1	1	9	1	7	1
3	3498-MS33	A	1	1	1	12	1	7	2
4	3303-MS35	A	2	1	3	1	1	1	3
5	3744-MS33(7)	A	2	1	3	1	1	1	3
6	1028-DS11	D	2	1	3	3	1	1	4
7	1324-MS34	A	2	1	3	3	1	1	4
8	3585-MS22	A	2	1	3	3	1	1	4
9	3742-DS05	D	2	1	3	3	1	3	5
10	1805-MS22	A	2	1	1	5	1	6	6
11	3046-MS23(2)	A	2	1	3	6	1	1	7
12	3196-MS22	A	2	1	3	6	1	1	7
13	1369-FS7.2	A	2	1	1	11	1	6	8
14	1475-MS34	A	2	1	3	15	1	2	9
15	1066-MS23	A	2	2	3	9	1	1	10
16	1048-MS28	A	3	1	1	1	1	3	11
17	1008-MS25(2)	A	3	1	3	1	1	1	12
18	1036-MS29	A	3	1	3	1	1	1	12
19	1071-MS33(3)	A	3	1	3	9	1	6	13
20	1631-MS28	A	3	1	4	10	1	7	14
21	1805-MS34	A	3	1	3	13	1	5	15
22	3177-MS29	A	4	1	3	9	1	2	16
23	1194-MS31(4)	A	4	3	3	8	1	3	17
24	0165-MS20	A	5	1	4	1	1	1	18
25	3175-MS34	A	5	1	3	7	1	1	19
26	3193-MS33(2)	A	5	1	3	7	1	1	19
27	3182-MS33(3)	A	5	1	3	9	1	1	20
28	0017-MS35	A	5	1	3	14	1	1	21
29	0216-MS26	A	5	2	3	5	1	3	22
30	3146-MS35(3)	A	5	2	3	5	1	3	22
31	3247-MS23(2)	A	5	2	3	5	1	3	22
32	1293-MS36	A	5	2	3	7	1	1	23
33	3319-MS35	A	5	2	3	7	1	3	24
34	3499-MS33(3)	A	5	2	3	7	1	3	24
35	3585-MS33	A	5	2	3	7	1	3	24
36	3320-MS24	A	5	2	3	9	1	3	25
37	1627-MS31(2)	A	5	2	3	10	1	1	26
38	3240-MS30	A	5	2	3	10	1	1	26
39	1467-MS34	A	6	1	1	9	1	5	27
40	1309-DS08	D	6	1	2	10	1	1	28

41	3510-MS21	A	7	1	3	1	1	5	29
42	0006-MS23	A	7	1	3	4	1	4	30
43	2899-MS34(2)	A	8	1	1	2	1	2	31
44	3432-DS06	D	8	2	3	3	1	3	32

A = Asymptomatic, D = diarrhoeal/dysenteric.

Actual band sizes (in base pairs):

D-A	A-L	R-R	N-K	S-D	S-Q
1 = 420	1 = 430	1 = 600	1 = 500	1 = 210	1 = 270
2 = 430	2 = 400	2 = 610	2 = 500+560		2 = 290
3 = 450	3 = 435	3 = 620	3 = 550		3 = 260
4 = 460	4 = 500	4 = 630	4 = 490		4 = 350
5 = 470			5 = 550+620		5 = 320
6 = 480			6 = 510+540		6 = 310
7 = 490			7 = 550+630		7 = 320+400
8 = 500			8 = 510+540+580		
			9 = 530		
			10 = 530+620		
			11 = 530+570+620		
			12 = 480+500+550+610		
			13 = 500+510+535		
			14 = 480+510+610		
			15 = 510+540		

APPENDIX VI

Appendix VI. DNA samples from 11 countries used for genotyping.

Strain origin	Sample ID	Clinical status	DNA source	No. of samples	Reference
Brazil	462	A	AC	2	Silva et al, 1997
	RPS				
Burma/Myanmar	HB-301:NIH	D	AC	1	Diamond 1968
Georgia	TBC#05	D	XC	9	Trapaidze, unpublished
	TBC#06				
	TBC#07				
	TBC#08				
	TBC#09				
	TBC#10				
	TBC#23				
	TBC#24				
	TBC#27				
India	Rahman	A	AC	1	Sargeaunt et al, 1980
Japan	J1	D	XC	26	Haghighi et al, 2002 and 2003
	J2	D	AC		
	J3	D	AC		
	J4	NA	XC		
	J5	A	AC		
	[1]J6	A	XC		
	J7	D+L	XC		
	J8	D+L	XC		
	J9	NA	XC		
	J10	D	XC		
	J11	L	XC		
	[2]J12	A	XC		
	J13	A	XC		
	J14	A	XC		
	J15	D	XC		
	J16	D	XC		
	[3]J17	L	XC		
	J18	NA	XC		
	J19	NA	XC		
	J20	NA	XC		
	J21	NA	XC		
	J22	NA	XC		
	J23	D	XC		
	J24	A	XC		
	J25	A	XC		
	J26	A	XC		
Korea	YS-27	L	AC	1	Park et al, 1999
Mexico	HM-1:IMSS	D	AC	1	Diamond et al, 1972
South Africa	SA11	A	XC	11	Zaki et al, 2003a
	SA27				
	SA47				
	SA51				
	SA52				
	SA65				
	SA70				
	SA115				

	SA119				
	SA153				
	SA154				
Turkey	Hu 04-08	D	SS	6	Tanyüksel, unpublished
	Hu 04-39	D			
	Hu 04-64	D			
	Hu 04-81	D			
	Hu 04-87	D			
	Hu 04-96	A			
Venezuela	IULA:1092:1	D	AC	2	Urdaneta et al,1995
	IULA:0593:2	D			
Vietnam	22-15	A	SS	27	Blessmann et al, 2003
	31-2				
	78-1				
	124-1				
	136-3				
	155-4				
	183-2				
	212-1				
	245-3				
	278-1				
	490-1				
	523-1				
	529-1				
	547-5				
	556-1				
	557-1				
	566-2				
	576-3				
	585-1				
	587-1				
	608-3				
	614-3				
	631-1				
	649-1				
	652-3				
	659-4				
	697-3				

A = Asymptomatic, AC = Axenic culture, D = Diarrhoea/dysentery, L = Liver abscess, XC = Xenic culture, NA = Not available, SS = Stool specimen, [1] = this individual had history of voluntary work in Ghana, [2] = this individual had history of voluntary work in Cambodia, [3] = this individual was a Thai national, but a migrant worker at Japan.

APPENDIX VII

Appendix VII. DAR genotypes of all non-Bangladeshi *E. histolytica*.

ID	Clinical status	Strain origin	*E. histolytica* specific PCR			Genotype
			D-A	A-L	R-R	
J8	D+L	Japan	2	2	4	2
SA47	A	South Africa	2	3	3	5
J6	A	Japan / Ghana	2	3	4	6
Hu 04-08	D	Turkey	3	1	3	11
TBC#08	D	Georgia	3	1	3	11
SA11	A	South Africa	3	2	3	14
SA115*	A	South Africa	3	2	3	14
SA119*	A	South Africa	3	2	3	14
HB-301:NIH	D	Burma	3	2	3	14
124-1	A	Vietnam	3	2	4	15
136-3	A	Vietnam	3	2	4	15
22-15	A	Vietnam	3	2	4	15
245-3	A	Vietnam	3	2	4	15
278-1	A	Vietnam	3	2	4	15
31-2	A	Vietnam	3	2	4	15
490-1	A	Vietnam	3	2	4	15
556-1	A	Vietnam	3	2	4	15
557-1	A	Vietnam	3	2	4	15
566-2	A	Vietnam	3	2	4	15
587-1	A	Vietnam	3	2	4	15
649-1	A	Vietnam	3	2	4	15
659-4	A	Vietnam	3	2	4	15
697-3	A	Vietnam	3	2	4	15
78-1	A	Vietnam	3	2	4	15
J24	A	Japan	3	2	4	15
J25	A	Japan	3	2	4	15
J26	A	Japan	3	2	4	15
Rahman	A	India	3	2	4	15
SA154	A	South Africa	3	2	4	15
J17	L	Japan / Thailand	3	2	4	15
J18	NA	Japan	3	2	4	15
J21	NA	Japan	3	2	4	15
Hu 04-96	A	Turkey	3	3	3	17
Hu 04-64	D	Turkey	3	3	3	17
Hu 04-81	D	Turkey	3	3	3	17
Hu 04-87*	D	Turkey	3	3	3	17
TBC#05	D	Georgia	3	3	3	17
TBC#09	D	Georgia	3	3	3	17
TBC#10	D	Georgia	3	3	3	17
547-5	A	Vietnam	3	3	4	18
576-3	A	Vietnam	3	3	4	18
585-1	A	Vietnam	3	3	4	18
608-3	A	Vietnam	3	3	4	18
HM-1:IMSS	D	Mexico	3	3	4	18
TBC#23	D	Georgia	3	3	5	19
Hu 04-39	D	Turkey	3	4	3	20
183-2	A	Vietnam	3	4	4	21
212-1	A	Vietnam	3	4	4	21

523-1	A	Vietnam	3	4	4	21
529-1	A	Vietnam	3	4	4	21
631-1	A	Vietnam	3	4	4	21
652-3	A	Vietnam	3	4	4	21
J16	D	Japan	4	2	3	24
YS-27	L	Korea	4	2	3	24
SA65	A	South Africa	4	3	4	27
J10	D	Japan	5	3	3	29
J7	D+L	Japan	5	3	3	29
J9	NA	Japan	5	3	3	29
SA27	A	South Africa	5	3	4	30
SA52*	A	South Africa	5	3	4	30
TBC#06	D	Georgia	1	3	3	31
TBC#07	D	Georgia	1	3	3	31
TBC#27*	D	Georgia	1	3	5	32
SA70	A	South Africa	2	1	4	33
J11	L	Japan	2	2	3	34
IULA:1092:1	D	Venezuela	4	2	4	35
J19	NA	Japan	4	2	4	35
# 462	A	Brazil	4	4	4	36
J12	A	Japan / Cambodia	4	5	3	37
IULA:0593:2	D	Venezuela	4	5	4	38
SA153	A	South Africa	5	2	3	39
SA51*	A	South Africa	5	2	3	39
J2*	D	Japan	5	2	3	39
J22*	NA	Japan	5	2	3	39
155-4	A	Vietnam	5	2	4	40
J13	A	Japan	5	2	4	40
J14	A	Japan	5	2	4	40
J1	D	Japan	5	2	4	40
J15*	D	Japan	5	2	4	40
J20	NA	Japan	5	2	4	40
RPS	A	Brazil	5	2	5	41
J5*	A	Japan	5	3	1	42
J3	D	Japan	5	3	1	42
J4	NA	Japan	5	3	5	43
614-3	A	Vietnam	5	4	4	44
J23*	D	Japan	5	5	5	45
TBC#24*	D	Georgia	6	3	5	46

A = Asymptomatic, D = Diarrhoeal/dysenteric, L = Liver abscess, *these showed multiple bands in one of the 3 loci but only the higher intensity band was considered for genotyping for simplicity.