CHARACTERISATION OF POLYMORPHIC DNA AND ITS APPLICATION TO TYPING OF *Entamoeba histolytica* AND *Entamoeba dispar*

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

2002

Mehreen Zaki

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

ABSTRACT

A key question in amoebiasis is whether the variable symptoms of amoebic infections are a reflection of different strains of *E. histolytica*. To address this, tools that allow typing of *E. histolytica* isolates are needed and this objective formed the basis of the present study. A method for PCR-based DNA typing of *E. histolytica* isolates has been developed using multiple loci with internal short tandem repeats (STRs) as the polymorphic markers. It has been shown that *E. dispar* isolates can also be typed by this approach and species-specific primers have been developed for two loci.

E. histolytica and *E. dispar* samples from a wide geographic range were studied to validate the general utility of these loci. Results revealed that *E. histolytica* is genetically highly variable. This was evident in all the communities studied. *E. dispar* also displays intra-species variation. The patterns seen for individual strains of both species were stable over time in the same infection. With few exceptions a single *E. histolytica* or *E. dispar* strain was identified in samples from infected family groups and outbreaks. Our results show both the existence of mixed species infection as well as the possibility of co-infection with different strains of the same species.

The genomic organisation of these repeat loci was also assessed. Our results show that they are likely multicopy and are as a rule organised into long tandem arrays. Analysis of the array length shows that they constitute a significant part of the chromosome on which they are carried.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost I wish to express my deepest gratitude to my supervisor Dr. C. Graham Clark for the support, encouragement and unstinting patience he has shown me throughout the course of this research. His advice and guidance have been invaluable in helping me achieve the objectives we set out to accompolish.

I would like to thank the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom (CVCP) for the Overseas Research Student (ORS) award and the Aga Khan University for the Faculty Development Award (FDA) which made it financially possible for me to pursue this degree. I am also grateful to the Aga Khan University for granting me leave of absence for the period required to undertake these studies.

I am grateful to Dr. Rehana Vahidy who supervised and guided me through my 'first' masters research project and in the process instilled in me her enthusiasm for the subject. I would also like to extend my sincere thanks and appreciation to Drs. Viqar Zaman and Rabia Hussain for their mentorship and encouragement. They have been instrumental in helping me embark on my academic career. I would in particular like to thank Dr. M. Asim Beg for having introduced and directed me to Dr. C. Graham Clark and The London School of Hygiene and Tropical Medicine. My presence here in London and this work would certainly not have been possible without his efforts.

I wish to express my sincere thanks to all those who provided me with the samples and relevant information which have formed such a significant part of my research. The list is long and it is not possible for me to mention each of them by name, but I am indebted to all of them for sharing their valuable time and suggestions so generously with me.

I am indebted to all my friends, relatives and colleagues (the world over) for their overwhelming love, support and encouragement. Recalling all of them by name and

deed(s) would require a volume in itself and is not possible here, but their concern, advice and generosity are definitely deserving of a special mention.

Finally, I wish to extend a special note of thank you to my parents and family. The love, concern and support they have shown me has, as always, been boundless during this time. Their contributions towards my achieving this goal have been varied and countless and can not be expressed in words.

CHAPTER 1

INTRODUCTION

Amoebiasis is the infection of the human gastrointestinal tract with the protozoan parasite *Entamoeba histolytica*. This small (10 to 40μm) and relatively fragile protozoan lacks most eukaryotic organelles e.g. a structured cytoskeleton, the Golgi complex, endoplasmic reticulum and mitochondria (Clark *et al*, 2000). It is however capable of colonising and invading the intestinal mucosa and virtually every other tissue and organ, most commonly the liver, wherein lies its medical importance. On a global scale, amoebiasis probably comes third among parasitic causes of death, behind only malaria and schistosomiasis.

The parasite has a cosmopolitan distribution. On the other hand, invasive amoebiasis is a major health problem in areas of Africa, Asia and Latin America where inadequate sanitary conditions prevail (Walsh, 1986). A major puzzle, however, is the variable expression of the parasite's pathogenic potential. Estimates of the worldwide distribution of amoebiasis suggested that only a small fraction of the total number of infected individuals have or develop invasive intestinal or extraintestinal disease. The combined use of biochemical, immunological and molecular techniques has allowed scientists to determine that there are in fact two morphologically identical species (Diamond and Clark, 1993). One is an invasive pathogen (*E. histolytica*) exhibiting varying degrees of virulence and the other is non-invasive (*E. dispar*) having the capacity of producing at most superficial erosion of the colonic mucosa.

The acceptance of *E. dispar* as a distinct species has had profound implications for the epidemiology of amoebiasis. Most of the asymptomatic infections worldwide are now attributed to this non-invasive amoeba. However, fewer than 10% of individuals who are infected with *E. histolytica* (in its new sense) develop invasive disease. A major area of debate in recent years has been whether the differences in expression of disease are a function of host factors or of parasite factors.

At present there are no simple, rapid and low cost laboratory tests available for the differentiation of *E. histolytica* and *E. dispar* infections. Better knowledge of amoebiasis prevalence, transmission and epidemiology requires improved diagnostic and surveillance tools and a clear understanding of the biological and epidemiological significance of strain differences. The development of such valuable tools for use in clinical laboratories and large-scale epidemiological surveys has been made a priority (Anonymous, 1997a).

In the following sections I have endeavoured to give some account of our current understanding of the overall biology and genome organisation of *E. histolytica*. This is followed by a review of literature pertaining to its epidemiology so as to highlight key unanswered questions. Finally, pitfalls of the epidemiological indices currently in place and emerging trends in molecular epidemiology of amoebiasis are discussed. Although *E. histolytica* is the key focus of much of the information provided and discussed, due reference has been made to *E. dispar*, its biology and role in the confusing and conflicting history of amoebiasis epidemiology.

1.1. *Entamoeba histolytica* AN OVERVIEW

1.1.1. HISTORICAL PREVIEW

Brief accounts summarising the history of *E. histolytica* and amoebiasis can be found in a number of articles and the following information has been derived from some of them (Martínez-Báez, 1986; Martínez-Palomo, 1993; Petri, 1996). It was in 1875 that Fedor Aleksandrovich Lösch first described the clinical and autopsy findings of a case of fatal dysentery, identified the amoebas and reproduced the disease in dogs, thus almost fulfilling Kochs postulates. He proposed the name Amoeba coli for this new species. Stephanos Kartulis in 1887 and Sir William Osler in 1890, made positive diagnosis of amoebic liver abscess. A year later Councilman and Lafleur, at Johns Hopkins hospital, confirmed the pathological role of amoebas through studies on patients with dysentery and hepatic abscesses and created the terms "amoebic dysentery" and "amoebic liver abscess". The term amoebic dysentery is now considered incorrect and in more recent literature (Ravdin, 1995) has been replaced by the term "amoebic colitis". The organism was formally named *E. histolytica* by Schaudinn in 1903.

In the following fifteen years more than a dozen species names were introduced for organisms that were morphologically very similar to each other and to *E. histolytica.* In 1919, Clifford Dobell reviewed all the published species descriptions and concluded that there was only one species of *Entamoeba* that produced quadranucleated cysts and retained Schaudinn's name *E. histolytica* for it.

As early as 1925, Emile Brumpt suggested the existence of two morphologically identical quadranucleated cyst producing species, one being pathogenic and the other not, and suggested the name *E. dispar* for the non-pathogenic species. Following much debate *E. dispar* was finally separated from *E. histolytica* (Diamond and Clark, 1993).

1.1.2. LIFE CYCLE

The life cycle of *E. histolytica* in humans has not been studied. The best description available is based on studies carried out by Dobell in 1928, in which he used a strain of *E. histolytica* recovered from a monkey. However, most of his observations have since been corroborated and are widely accepted (Martínez-Palomo, 1993). The complete life cycle consists of four consecutive stages; namely, cyst, metacyst, trophozoite and precyst, forms.

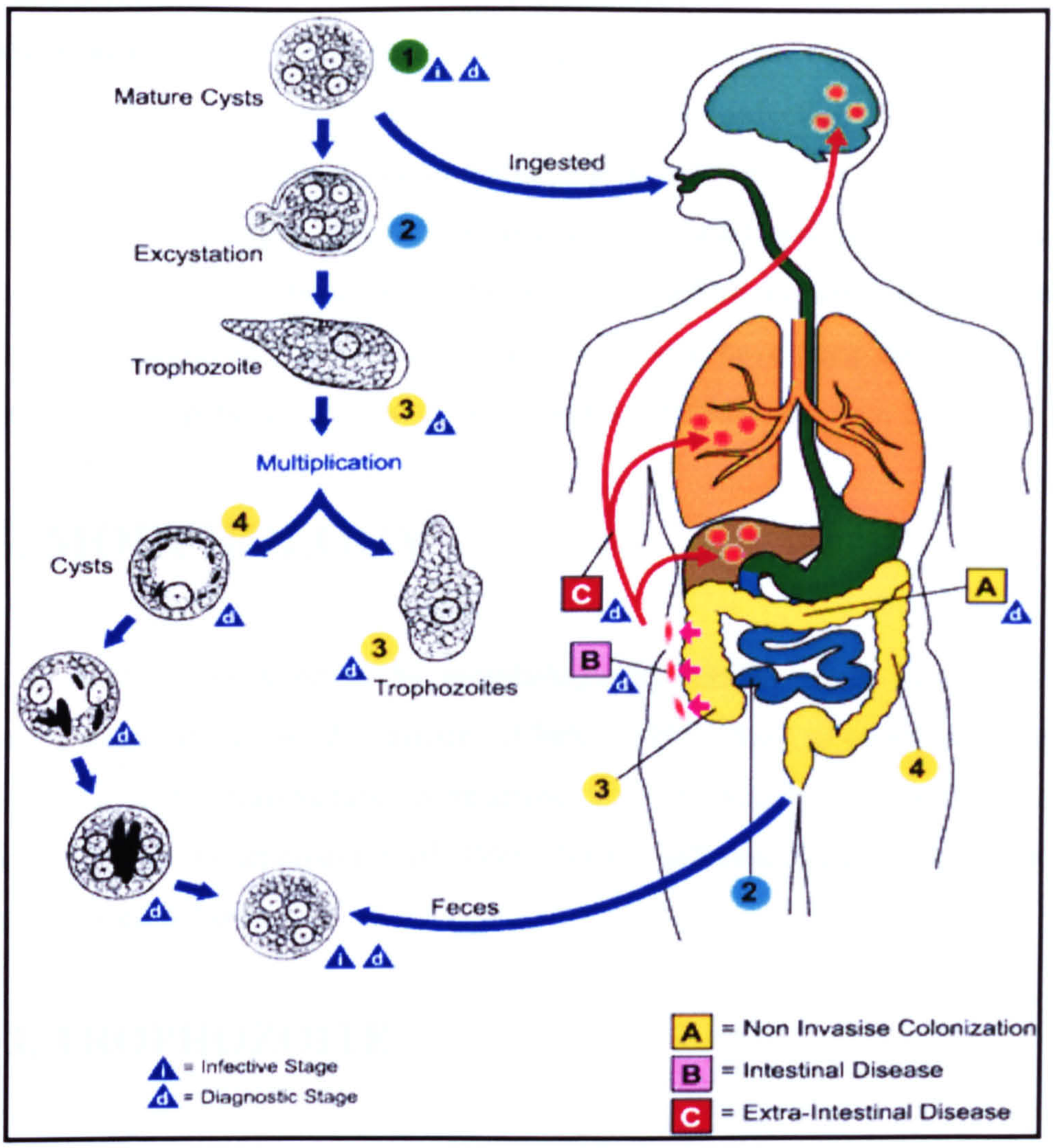


Fig. 1. Schematic Representation of Life Cycle of *Entamoeba histolytica*. Taken from the CDC website (http://www.dpd.cdc.gov/dpdx/HTML/Amebiasis.htm).

Infection is acquired by ingestion of the mature quadranucleate cyst which is resistant to the acidic pH of the stomach. Excystation occurs in the small or large bowel, with division of the emerging amoeba into four and then eight metacystic trophozoites. Trophozoites have the ability to colonise and/ or invade the large bowel, where they multiply by binary fission. Encystation may take place if luminal conditions are less than ideal for the trophozoites, resulting in four nucleated cysts after two successive nuclear divisions. Cysts do not develop within tissues.

Upon excretion cysts remain viable for weeks to months depending on environmental conditions. Infection is not transmitted by trophozoites (which may be excreted during episodes of acute colitis), because of their rapid degeneration outside the body and their destruction in normal gastric contents with a low pH. Infection resulting from the ingestion of as little as a single cyst in contaminated food or water has been reported (Ravdin, 1995), and larger inocula are usually associated with shorter incubation periods of days instead of the usual one to two weeks before the onset of symptomatic disease. The duration of infection is variable and cysts have been demonstrated in faeces of untreated persons for as long as two years (Martínez-Palomo, 1993).

1.1.3. MORPHOLOGY

E. dispar is still considered to be morphologically similar to *E. histolytica* but successful axenization of *E. dispar* (Clark, 1995) has allowed a detailed morphological and ultrastructural comparison with *E. histolytica* under identical conditions (Espinosa-Cantellano *et al*, 1998). Subtle differences have been observed and are discussed bellow.

1.1.3.i. TROPHOZOITE

The trophozoite is a highly dynamic and pleomorphic cell whose form and motility are strongly affected by changes in temperature, pH, osmolarity and redox potential. In general, both light and scanning electron microscopy have revealed that, in axenic cultures at least, *E. dispar* trophozoites are elongated with a prominent, broad, anterior pseudopod and a distinct posterior uroid. The uroid appears as a tail formed

of irregular folds of the membrane and fine processes called filopodia. In contrast, *E. histolytica* trophozoites from axenic cultures tend to be rounder in shape, have several small pseudopodia extending from different parts of the cell and a uroid is less commonly observed. Motility and pseudopodia formation are rapid and movement rarely occurs in a straight line. The diameter ranges widely, from 10 to 60 μm (average 25 μm), not only due to the pleomorphism of the parasite but also to the feeding conditions. Amoebae obtained directly from intestinal or liver lesions are generally larger (20 to 40 μm) than those found in non-dysenteric stools or in cultures (10 to 30 μm).

The cytoplasm is characterised by the absence of most of the organelles found in other eukaryotes. The clear ectoplasm surrounds a granular endoplasm, which contains abundant vacuoles. Some of these have been identified as phagocytic, macro- and micropinocytic vacuoles, lysosomes and food vacuoles. The latter may be filled with starch and ingested bacteria in xenic cultures and erythrocytes in dysenteric stools. In axenic cultures the cell surface of *E. histolytica* is rough with numerous circular openings that correspond to the mouths of micropinocytic vesicles. In contrast the cell surface of *E. dispar* trophozoites is smoother in appearance. The distribution of vacuoles differs between the two species. In *E. dispar* the cytoplasm appears patchy with vacuoles being concentrated in some regions, while in *E. histolytica* the vacuoles are larger and are uniformly distributed. A prominent feature of most *E. histolytica* and *E. dispar* trophozoites is the existence of cytoplasmic areas containing large deposits of glycogen granules interspersed with ribosomal bundles.

There is a single nucleus of 4 to 7 μm in diameter with a small, spherical, central karyosome. In light microscopic analysis the *E. dispar* nucleus has a thin peripheral rim made up of regularly placed dense circular granules, whereas in *E. histolytica* the nuclear periphery is coarser and thicker. The chromatin clumps are usually uniform in size and evenly distributed inside the nuclear membrane although in some cells the chromatin may be concentrated on one side as a crescentic mass. The nucleus has no fixed position in the cytoplasm, but moves freely.

1.1.3.ii. CYST

The mature *E. histolytica* cyst is a quadranucleate, round to oval structure. The average diameter is 12 μm (range 8 to 20 μm). It has a rigid wall made up of chitin and glycoproteins that protect the amoeba outside the human body. Generally, glycogen deposits and chromatoid bodies are seen in the immature precystic stage but disappear as the cyst matures, as do food vacuoles. A chromatoid body is a large crystalline inclusion that results from the aggregation of ribosomes. These are arranged in helical arrays in the trophozoite. The helices are ordered in a hexagonal packing pattern during encystation to form these classical inclusions which appear as rods with blunt or rounded ends. There are one to four small nuclei each containing a small and usually centrally located karyosome and the nuclear membrane is uniformly lined with peripheral chromatin.

1.1.4. CULTIVATION

Clark and Diamond (2002) have provided a succinct account of the history of *E. histolytica* cultivation. Boeck and Drbohlav in 1925, were the first to successfully cultivate *E. histolytica* by using the diphasic Locke's egg serum medium, a modification of which (Locke-Egg (LE)) is still in use today. Several monophasic media have been developed e.g. the egg yolk infusion medium defined by Balamuth in 1946 and Diamonds TYSGM-9. At present the most widely used media for xenic cultivation (i.e. growth of parasite in the presence of undefined flora) are the diphasic LE and Robinson's medium and the monophasic TYSGM-9. These media also support the growth of other *Entamoeba* species with varying degrees of success.

The first axenic cultivation (i.e. growth of parasite in the absence of any other metabolising cells) of *E. histolytica* was achieved by Diamond in 1961 using a complex serum-enriched diphasic medium. Successive improvements have resulted in TYI-S-33 which is currently the most widely used medium for axenic cultivation of *E. histolytica*. A major drawback of TYI-S-33 however, is the presence of casein, which varies from lot to lot in its ability to support good growth of *E. histolytica*, and is not

readily available. More recently, YI-S has been described as an alternative to TYI-S-33 (Diamond *et al*, 1995) and overcomes inherent problems of the latter by replacing casein with increased amounts of yeast extract.

Currently, monoxenic cultivation (i.e. growth of parasite in the presence of a single species of associate) is only used as a transition between xenic and axenic cultures. Though specific media for monoxenic cultivation do exist, axenic culture media can work just as well in some cases at least (Clark and Diamond, 2002).

1.1.5. PATHOGENESIS

The majority of what we know about the host-parasite interactions that take place in amoebiasis is based on *in vitro* studies. Experimental liver abscesses in hamsters and gerbils have been produced with axenic cultures of *E. histolytica* (Martínez-Palomo, 1993). At present, however, there is no experimental animal model that reproduces the invasive intestinal amoebic lesions seen in human intestinal amoebiasis (Espinosa-Cantellano and Martínez-Palomo, 2000). In the past decade, application of molecular biological techniques has led to rapid progress towards the identification of several amoebic and host related factors associated with virulence (Petri *et al*, 1994; Gilchrist and Petri, 1999; Espinosa-Cantellano and Martínez-Palomo, 2000). For descriptive purposes Martínez-Palomo *et al* (1985) have divided this host-parasite interplay into four stages: adhesion, cytolysis following contact, phagocytosis and intracellular degradation.

Among the host factors implicated in determining whether *E. histolytica* infection will result in colonisation and invasion are high caecal total ammonia (NH_3) and acid (H^+) concentrations (Leitch, 1988). It has also been suggested that lowering of the redox potential in the parasite promotes colonisation and invasion (Bracha and Mirelman, 1984). *E. histolytica* trophozoites survive within their human hosts by feeding on bacteria and cellular debris. Lowering of the redox potential is favoured by the anaerobic environment of the gut and by the ability of ingested bacteria to act as broad range scavengers of oxygen molecules.

Killing of target bacteria is reported to occur only on direct contact mediated by the parasites' galactose-inhibitable lectin (Gal/GalNAc lectin) which binds to the galactose and N-acetyl D-galactosamine moieties in bacterial cell walls (Petri *et al*, 1989). This protein also mediates adherence to a variety of other targets, including human erythrocytes, neutrophils, colonic mucins and epithelial cells (Huston and Petri, 1998). The intestinal mucus blanket is the first barrier and colonic mucin glycoproteins can act as an important host defence by binding to the parasites' Gal/GalNAc lectins i.e. they act as alternative target sites, thus preventing amoebic attachment to and cytolysis of host epithelial cells (Chadee *et al*, 1987). However, it has also been suggested that colonic mucins may in fact facilitate colonisation by virtue of the same adherence ability. Keller *et al* (1992) have reported that *E. histolytica* trophozoites in contact with human colonic cells induce fast release of both pre-formed and newly synthesised mucins. Such sustained hypersecretion may contribute to mucin depletion and alteration of the protective mucus blanket, hence facilitating amoebic invasion of the underlying epithelial layer. On the other hand parasite adherence to colonic epithelium can be blocked by destruction of the galactose-inhibitable lectin by pancreatic proteases, bile salts and bacterial glycosidases (Petri *et al*, 1994).

It has also been postulated that penetration of the mucus blanket can be brought about by mechanical amoeboid movement (Martínez-Palomo *et al*, 1985). In one type of interaction, the amoebas establish a hit-and-run pattern in which, after maintaining contact for a few minutes with a given epithelial cell, an amoeba moves on to another cell resulting in dislodgement of epithelial cells from the substrate. The second, less common, mechanism of mechanical lytic action involves "pinching off". Following adhesion, the trophozoites detach from the target cell body taking with them the membrane at the attachment site. This creates a hole in the plasma membrane of the cell resulting in lysis and subsequent phagocytosis of epithelial cells.

The Gal/GalNAc lectin is a multifunctional membrane protein and in addition to its role in amoebic adherence it has been shown to participate in cytolytic events, interact with the cytoskeleton, and appears to mediate amoebic resistance to complement-mediated lysis. Participation in the former has been evidenced by Saffer and Petri (1991), who demonstrated that monoclonal antibodies directed against the galactose

lectin decreased cytotoxicity, in addition to inhibiting adherence. A cell-signalling role for the lectin has been suggested by the observation that rapid polymerisation of actin occurs upon adhesion of amoeba trophozoites to target cells and that this polymerisation can be inhibited by pre-incubation with galactose (Gilchrist and Petri, 1999). The researchers postulate that such alterations in the cytoskeletal structure could, in turn, initiate other signalling cascades.

The existence of another amoebic protein, the "amoebapore", has also been demonstrated (Leippe, 1997). This protein, for which three isoforms (A, B and C) are known, is stored within granules found in large numbers in the amoeba's cytoplasm and is likely released only after direct contact of the amoeba with its target cell. Although isoform A is the most abundant and C the least, amoebapore C has a markedly higher cytolytic activity than A or B. Insertion and lateral diffusion of the amoebapore through target membranes results in the formation of water filled channels which may allow ions and other small molecules (e.g. toxins) to pass through. This changes the target cell's internal environment, which ultimately results in its lysis. Based on this proposed mode of action, two key functions have been suggested for these molecules. The primary function appears to be intracellular destruction of phagocytosed bacteria, the amoeba's main source of nutrients. Alternatively, an appropriate stimulus, e.g. cell to cell contact between the amoeba and target cell, may trigger granular exocytosis into the confined environment of the contact zone. High concentrations of amoebopores may be reached in the small inter-cellular space, sufficient to induce target cell death, as suggested by the swelling and massive surface blebbing seen in target cells minutes after the primary contact has been established with the surface lectin.

The major proteolytic enzymes released by *E. histolytica* are cysteine proteinases (Que and Reed, 1997). Unlike amoebapores, these enzymes do not kill target cells in a contact dependant manner and are secreted spontaneously. Six cysteine proteinase genes have been identified. The expression of one of these, a major neutral cysteine (thiol) proteinase, has been correlated to virulence. This major proteinase is also the only one localised on the amoeba surface, while the remaining are found in cytoplasmic granules. *In vitro* the major proteinase has been shown to mimic the cytopathic effect of the whole trophozoite. This neutral enzyme can degrade basement

membrane components (e.g. collagen) and cell anchoring proteins (e.g. fibronectin and laminin). Cysteine proteinases also appear to have a role in amoebic liver abscess formation as evidenced by a reduction in abscess formation when trophozoites were treated with specific protease inhibitors prior to inoculation into animals. The neutral proteinase can also activate the alternative complement pathway by cleavage of component C3, resulting in generation of anaphylatoxins C3a and C5a, which are crucial components of the host inflammatory response. However, the same proteinase can limit the inflammatory reaction by degrading and inactivating these anaphylatoxins. The remaining portions, C3b and C5b, participate in activation of the late acting components leading to formation of the membrane attack complex on the amoebic plasma membrane. Formation of this complex can, however, be blocked by the Gal/GalNAc lectin which binds to the terminal complement components C8 and C9 and hence prevents assembly of the complement membrane attack complex (Braga *et al*, 1992).

After contact dependant cytolysis of the target cell has occurred the amoebas ingest the lysed cell, although living cells can also be engulfed, following which the efficient cytoplasmic machinery rapidly degrades the ingested material. The ability of *E. histolytica* to phagocytize seems to be related to their virulence since virulent amoebas display active erythrophagocytosis while those of low virulence ingest few red cells. In fact, amoebas rendered defective in their phagocytic capacity lose their virulence (Martínez-Palomo, 1993).

1.1.6. PATHOLOGY

In humans, typically, the colonic lesions present either as non-specific thickening of the mucosa or as the classic flask-shaped amoebic ulcers. Colonic and caecal mucosa are invaded by *E. histolytica* trophozoites and cellular infiltration occurs around the invading amoeba. There is rapid lysis of the inflammatory cells and tissue necrosis. Ulceration may deepen and progress under the mucosa to form "flask ulcers" which extend into the submucosa and produce microhaemorrhages, which produce the red cells found in amoebas in stool specimens. Initially the ulcers are superficial with normal mucosa between the sites of invasion and a necrotic base. Further progression

of the lesions may produce loss of mucosa and submucosa covering the muscle layers and eventually ulceration may lead to rupture of the serosa (Espinosa-Cantellano and Martínez-Palomo, 2000).

Complications of intestinal amoebiasis include intestinal perforation, direct extension of ulcers to the skin, and dissemination of amoebas through the portal circulation to extraintestinal sites (Reed, 1992; Ravdin, 1995). Haematogenous dissemination of amoebae to extraintestinal sites most frequently involves the liver, though lungs, pericardium, brain and skin can be involved too. Interestingly, the presence and extent of liver involvement bears no relationship to the severity of intestinal disease and liver disease has been known to occur independent of an obvious intestinal infection. Characteristically, a human amoebic liver abscess consists of areas in which the parenchyma has been completely replaced by semisolid or liquid material composed of necrotic matter containing a few cells. Amoebas tend to be located at the periphery of the abscess. Liver abscesses may heal or rupture with further dissemination of infection. If and when properly treated, invasive amoebic lesions in the large intestine, liver or skin of humans almost invariably heal without the formation of scar tissue, a phenomenon for which no adequate explanation exists (Espinosa-Cantellano and Martínez-Palomo, 2000).

1.1.7. HUMORAL & CELLULAR IMMUNITY

A prompt local secretory response followed by an equally rapid systemic antibody response follows intestinal invasion by *E. histolytica*. IgA, IgG and IgM coproantibodies have been found by indirect haemagglutination in persons with active amoebiasis. Circulating antibodies to *E. histolytica* can be demonstrated as early as one week after the onset of invasive amoebiasis in both man and experimental animals. All immunoglobulin classes are involved but there seems to be a predominance of IgG2 antibodies (Salata and Ravdin, 1986; Martínez-Palomo, 1993). However, many individuals have recurrent intestinal amoebic infection despite having elevated titres of antiamoebic antibodies. In *in vitro* studies, virulent amoebas, by virtue of their major cysteine proteinase, are resistant to complement-mediated antibody lysis and can degrade both serum and secretory IgA as well as circulating

IgG antibodies (Que and Reed, 1997). Taken together these observations have led to a consensus among scientists that circulating antiamoebic antibodies are not protective against intestinal and even perhaps extraintestinal amoebiasis.

Studies involving immunosuppression of experimental animals support the existence of cellular, rather than humoral, immunity against extraintestinal amoebiasis. Polymorphonuclear (PMN) neutrophils are rapidly recruited and activated in response to proinflammatory chemokines, e.g. cytokines, and cleavage products of complement (Espinosa-Cantellano and Martínez-Palomo, 2000). The Gal/GalNAc lectin has been shown to induce *in vitro* production of T-cell derived cytokines, which in turn activate macrophages (Campbell and Chadee, 1997). There is both *in vivo* and *in vitro* evidence to suggest that activated macrophages show potent amoebicidal activity. T-cells may also be directly cytotoxic to amoebic trophozoites in a contact dependant manner mediated by the surface lectin.

However, downregulation of the host's cellular immune responses following amoebic invasion has been demonstrated. The exact mechanisms involved in downregulation are only partially understood but it has been shown that virulent amoebas can lyse neutrophils through a contact-dependant, lectin mediated mechanism (Salata and Ravdin, 1986). Both *in vivo* and *in vitro* studies have suggested that lysis of neutrophils initiates release of toxic neutrophil enzymes which may then contribute to the tissue necrosis that is seen with invasive amoebiasis. It has been demonstrated that amoebae are capable of altering key macrophage accessory and effector cell functions thereby inhibiting amoebicidal activity. These include inhibition of antigen presenting cell activity, reduced T-cell proliferation and activation, and decreased macrophage responsiveness to T-cell derived activating cytokines (Campbell and Chadee, 1997).

Clinical studies by Deleon in 1970 have however shown that recurrences of amoebic liver abscess are rare in humans (cited in Campbell and Chadee, 1997). Furthermore, vaccination of animals with purified or recombinant amoebic molecules results in a level of protection against challenge infection (Petri and Ravdin, 1991; Zhang *et al*, 1994). Taken together these data suggest that prior sensitisation of the immune system might be sufficient to overcome the ability of amoebic trophozoites to downregulate macrophage and T-cell functions.

1.1.8. CLINICAL SYNDROMES

The term 'amoebiasis' includes all cases of human infection with *E. histolytica*. The clinical presentations of intestinal *E. histolytica* infection are variable (Reed, 1992). One of the most common forms of intestinal infections is 'asymptomatic cyst passage or 'luminal amoebiasis', with the organism acting as a harmless commensal. Occasionally 'symptomatic non-invasive infections' may occur and exhibit non-specific gastrointestinal symptoms (Ravdin, 1995). 'Invasive intestinal infections' generally manifest as 'amoebic colitis' which is one of the major intestinal syndromes and is marked by loose stools containing blood and mucus. Characteristically, the patient has several evacuations (three to five) per day and fever, and general systemic manifestations are generally absent.

Complications of amoebic colitis, such as 'fulminating amoebic colitis', are uncommon but occur most often in children. These are marked by necrotic ulcerous lesions extending over large areas. Evacuations are frequent (twenty or more per day) and occasionally contain blood alone, accompanied by signs of systemic involvement e.g. fever, dehydration and shock. Perforation and secondary bacterial infection of ulcers may occur and is especially common in children. As might be expected, concurrent liver abscess is common. Another unusual complication of severe colitis is 'toxic megacolon', which is marked by diffuse distension of the colon and is usually associated with the administration of steroids (Reed, 1992). Occasionally, chronic ulceration can result in amoeboma formation. Amoebomas are pseudotumoural lesions that result from necrosis, inflammation and oedema of the mucosa and submucosa of the colon. They usually occur singly, though occasionally multiple masses have been detected. Less frequent complications include perianal cutaneous amoebiasis (Martínez-Palomo and Espinosa-Cantellano, 1998).

The most common form of 'extraintestinal amoebiasis' is the amoebic liver abscess, which unless properly diagnosed and promptly treated is potentially lethal. This condition results from the migration of *E. histolytica* trophozoites from the colon to the liver via the portal circulation. Both intestinal and extraintestinal localisation can occur at the same time, but they are usually manifested separately. The time lapse

between penetration of the mucosa of the large intestine and damage to the liver is unknown, however concurrent amoebic colitis has been found in only one-third of hepatic abscess cases (Martínez-Palomo and Espinosa-Cantellano, 1998). Amoebic liver abscesses can occur in all age groups but predominate in adults aged twenty to sixty years. There is a marked preference for the right lobe of the liver and it is at least three times more frequent in males than in females. The onset is usually abrupt and is accompanied by pain and systemic symptoms. Direct or haematogenous spread from the liver abscess can result in 'pleuropulmonary amoebiasis' and involvement of the peritoneum and/ or pericardium. 'Cerebral amoebiasis' though very rare has an abrupt onset and rapid progression, resulting almost always in death.

1.1.9. DIAGNOSIS

In light of the acceptance of the two species, diagnosis of amoebic infection requires specific detection of *E. histolytica* and its differentiation from *E. dispar*. This is important not only for appropriate treatment to be administered and to reduce unnecessary drug prescription but is also necessary for carrying out reliable epidemiological surveys. There are several inherent problems associated with the variety of diagnostic methods currently in use. A detailed review of these methods, their advantages and disadvantages is the focus of section 1.4 (this chapter).

Briefly, diagnosis of invasive intestinal amoebiasis still relies on the microscopic detection of haematophagous trophozoites of *E. histolytica* in stools (González-Ruiz *et al*, 1994b). Problems in diagnosis arise when cysts alone occur in stools of healthy or diarrhoreic individuals, as microscopy does not distinguish between cysts of *E. histolytica* and *E. dispar*. Colonoscopy and scraping or biopsy of the ulcer edge is the most definitive means of diagnosis for amoebic colitis and examination of specimens from this procedure reveals the diagnosis in 85% of cases (Ravdin, 1995; Walsh, 1986). However, invasive procedures such as endoscopy should be performed with great care, as there is risk of intestinal perforation. Culture techniques have been used in conjunction with isoenzyme electrophoresis to differentiate infections caused by *E. histolytica* and *E. dispar* (Sargeaunt *et al*, 1978), but at present this technique is not used for routine laboratory diagnosis.

Serological tests provide a valuable means of indirect diagnosis when direct demonstration or speciation of the organism is difficult. Serology is an especially useful diagnostic test for amoebic liver abscesses, where the inability to detect amoebas in abscess fluid is well known (Healy, 1986). The use of serological tests as an adjunct diagnostic tool is also justified in invasive intestinal disease where the use of substances that interfere with stool examination for parasites is suspected e.g. antidiarrhoeal agents, antiamoebic drugs, antibiotics etc. However, currently available serological tests do not differentiate present and past *E. histolytica* infections and so provide little aid in diagnosis of simple intestinal infections and in areas of endemicity (Petri, 1996). Non-invasive imaging studies have greatly improved diagnosis of amoebic liver abscesses, although differentiation of amoebic and bacterial abscesses is not always easy.

New technologies being developed for improved diagnosis of amoebiasis and specific detection of *E. histolytica* include antigen detection in stool and serum using monoclonal antibodies (Abd-Alla *et al*, 1993; Haque *et al*, 2000) and detection of parasite DNA in stool and liver abscess pus by nucleotide probes (Bracha *et al*, 1990) or PCR amplification (Verweij *et al*, 2000; Zaman *et al*, 2000). An enzyme-linked immunosorbent assay (ELISA) kit that employs monoclonal antibodies to the Gal/GalNAc lectin and can distinguish *E. histolytica* and *E. dispar* antigens directly from stool is now commercially available (Haque *et al*, 1998).

1.1.10. TREATMENT

The treatment of amoebic infection is a complex issue, as multiple drugs must be prescribed to eradicate the parasite from the bowel lumen and from the tissues. Additional complexities include the unavailability of certain drugs in many countries and multiple toxic effects of different drugs (Ravdin, 1995). People with invasive amoebiasis require prompt treatment. It was previously recommended that all asymptomatic carriers of *Entamoeba* should also be treated to prevent transmission and spread of potential pathogenic parasites. However, with the development of methods that can differentiate between *E. histolytica* and *E. dispar*, it is now recommended that *E. histolytica* carriers be identified specifically and treated

(Anonymous, 1997b). *E. dispar* carriers need only be monitored closely for the possibility of co-infection with *E. histolytica* (Martínez-Palomo and Espinosa-Cantellano, 1998).

Antiamoebic drugs may be classified in three groups: luminal, tissue and mixed amoebicides. The most frequently used drugs with luminal action are iodoquinol, diloxanide furoate and paromomycin and are thus the treatment of choice for asymptomatic cyst passers. Diloxanide furoate, although relatively nontoxic, is not easily available all over the world. Iodoquinol shows gastrointestinal toxicity and is in limited supply. Paromomycin is highly effective with few side effects and is considered safe for use in children and pregnant women (Ravdin, 1995). Amoebicides effective only in tissues are emetine hydrochloride and dehydroemetine. Emetines, which are given intramuscularly, have multiple adverse effects, including being toxic to the myocardium, and it is recommended that they be used only under extraordinary circumstances.

Drugs effective in both tissues and the intestinal lumen include metronidazole and other nitroimidazole derivatives such as tinidazole and ornidazole. These drugs have the additional advantage of being administered orally. Although there have been reports of their carcinogenic effects in rodents and mutagenic potential in bacteria, no such effect has been noted in humans. The occasional adverse effects noted in humans e.g. gastrointestinal upset, abdominal pain, nausea and vomiting, are more unpleasant than serious and therefore make these drugs the treatment of choice for invasive amoebiasis. Metronidazole is effective but relatively more toxic than some of the other nitroimidazoles available. This group of drugs is contraindicated in pregnant women and nursing mothers because of their ability to cross the placental barrier and their elimination in breast milk. Another drawback of metronidazole is that the drug is rapidly absorbed and therefore the concentrations reached in the intestine may not be as effective in eliminating cysts as the luminal agents described above (Martínez-Palomo, 1993). Therefore, to prevent recurrences and onward transmission, patients with invasive disease should be treated with luminal drugs following metronidazole therapy, since two-thirds of them will have asymptomatic intestinal colonisation as well (Martínez-Palomo and Espinosa-Cantellano, 1998). There is as yet no report of drug resistance in amoebiasis.

1.1.11. CONTROL

Numerous variables, including education, socio-economic level, method of excreta disposal, and water supply, interact to determine the extent of transmission. Thus, in aiming at the control of transmission of amoebiasis, there is a need to modify the economic, technical, educational and cultural patterns that facilitate the spread of amoebiasis within the given community (Martínez-Palomo and Martínez-Báez, 1983). The World Health Organisation, has recommended methods of control which are aimed at the improvement of environmental sanitation including water supply, food safety, and personal hygiene and health education to prevent faecal-oral transmission (Anonymous, 1985), as well as early, specific detection and treatment of cases of infection and carriers (Anonymous, 1997b).

The most common modes of transmission are ingestion of food and water contaminated with cysts and direct passage from person to person. Cyst passers are the main reservoirs of infection. Cysts can remain viable and infective for several days in faeces as well as in wet/ moist soil, water, seawater and sewage depending on the temperature. They are not killed by the quantity of chlorine normally used to purify water, therefore chlorination alone would not prevent spread of the organisms and potential epidemics due to faecal contamination of water. The safest and most practical method of rendering drinking water free of viable cysts is to boil it for 10 minutes, followed by adequate protection of the stored sterilised water (Martínez-Palomo, 1993). Cysts survive up to 45 minutes in faecal material lodged under fingernails, but are killed within 1 minute by desiccation on the surface of the hands. Thus, practices such as hand washing after defecation and before handling food and avoiding consumption of raw vegetables, fruits and exposed food bought from street vendors need to be continually reinforced through public health education in schools and the use of mass media.

There is at present no vaccine for amoebiasis. As *E. histolytica* only infects humans and some higher primates (Petri, 1996), with no apparent large animal reservoirs of disease, and in light of the limited data suggesting that acquired immunity to invasive disease does exist (section 1.1.7.), it should theoretically be possible to develop a

vaccine that protects against colonisation (Huston and Petri, 1998). Current anti-amoebic vaccine candidates include five highly conserved, immunogenic surface proteins. Three of these are the amoebic proteins thought to be critical to pathogenesis and virulence i.e. the Gal/GalNAc lectin, the cysteine proteinases and the amoebapore. Another key candidate is the serine-rich *E. histolytica* protein (SREHP or K2) with tandem 8- and 12-amino-acid repeats (Stanley *et al*, 1990; Köhler and Tannich, 1993). Lastly there is a 29-kDa cysteine-rich protein, which appears to be a thiol-dependant peroxidase (Huston and Petri, 1998). Immunogenicity of the amoebapore is not established and response to the 29-kDa protein appears to be present in liver abscesses cases only. Protection in animals and their potential as vaccine components has yet to be evaluated for cysteine proteinases and the amoebapores (Huston and Petri, 1998).

1.1.12. GENOME STRUCTURE

The *Entamoeba* genome is still not completely understood. Most of what we know today is based on studies carried out on *E. histolytica*. Recent studies have revealed several unusual features, but there is still uncertainty regarding the DNA content of the nucleus and exact ploidy, among other features (Clark *et al*, 2000; Bhattacharya *et al*, 2000).

Estimation of genome size by different methods has failed to give consistent values. However, determination of the electrophoretic karyotype of *E. histolytica* by pulsed-field gel electrophoresis (Willhoeft and Tannich, 1999) suggests that the size of the haploid genome is about 20 Mb. The G + C content of the genome is low, about 22.4%, while that of coding regions is approximately 33%. In general, most strains and species are comparable, except for *E. moshkovskii* strain Laredo which is reported to have about 10% higher G+C content than other species. Analysis of ca. 4500 codons used in *E. histolytica* genes showed a preference for A and T at the third position (Tannich and Horstmann, 1992; Char and Farthing, 1992). It appears that most genes lack introns, and when present they are small, typically less than 100 bp (Willhoeft *et al*, 2001). Intergenic regions are between 400 bp and 2.3 kb and suggest tight packing of genes (Bruchhaus *et al*, 1993).

Both linear chromosomes and a number of circular plasmid-like molecules are present in the *Entamoeba* genome. Depending on the *E. histolytica* isolate used for analysis (HM-1:IMSS, 200:NIH and HK-9), 31-35 chromosomes have been identified ranging in size from 0.3 to 2.2 Mb (Willhoeft and Tannich, 1999). 14 linkage groups were identified and the functional ploidy was estimated to be at least 4. The structure and organisation of *E. histolytica* chromosomes is not yet clear. The typical organisation of chromosomes in to nucleosomal structures has not been demonstrated. Although most of the histone genes have been identified, their nucleotide sequences were found to be quite divergent from their counterparts in other organisms. It is possible that these differences result in atypical interactions between histones and DNA resulting in alternative nucleosomal structures (Bhattacharya *et al*, 2000). Bagchi *et al* (1999) have shown that the chromosomes of *E. histolytica* are linear molecules. Distinct size variation has been seen among homologous chromosomes of different isolates (Willhoeft and Tannich, 1999). Several classes of dispersed and tandemly repeated-DNAs have been described (Lohia *et al*, 1990; Mittal *et al*, 1994; Cruz-Reyes *et al*, 1995) and there is evidence that at least some of them are present in *E. dispar* too (Huang *et al*, 1997; Shire and Ackers, 2000). Their functions have not yet been determined.

A variety of circular DNA molecules have also been described in *E. histolytica*. The most abundant and extensively studied circular DNAs are those that carry the rRNA genes (Bhattacharya *et al*, 1998). In fact these circular rDNA molecules have been demonstrated in all strains and species of *Entamoeba* studied. In *E. histolytica* the rRNA genes are estimated at about 200 copies and the complete sequence of one rDNA molecule has been published. It measures 24.5 kb in length and carries two copies of the ribosomal RNA gene as an inverted repeat. It encodes the small, large and 5.8S rRNAs but not the 5S rRNA. No proteins are encoded. Several classes of tandemly repeated-DNAs are also carried on the 24.5 kb circle, some are interrelated and one is transcribed. In some isolates, however, only one rDNA copy is present and certain repeated-DNA classes are also missing. Furthermore, some of the tandemly repeated regions are unstable and length variation due to changes in the number of repeats occurs quite frequently. *In situ* hybridisation indicates that the rDNA circles are located in the vicinity of the nuclear membrane and are not associated with the

chromosomes, and their segregation appears to precede the separation of chromosomes during nuclear division (Willhoeft and Tannich, 2000). No copy of the circular rDNA plasmid was found in any of the linear chromosomes (Bagchi *et al*, 1999).

Besides the 24.5 kb circles, several less abundant DNA circles in different size classes of ca. 5, 12 and 50 kb have also been reported in all strains and species of *Entamoeba* studied (Dhar *et al*, 1995; Lioutas *et al*, 1995). Their structure and function has not yet been studied.

1.2. *Entamoeba dispar* AN OVERVIEW

As mentioned earlier, *E. dispar* has been recognised as an independent species (Diamond and Clark, 1993; Anonymous, 1997b). To date every gene sequence that has been analysed and compared between the two species has been distinct. In fact, the estimated genetic distance between the two species is comparable to that seen between the same genes from humans and mice (Clark and Diamond, 1991b). However, for the most part observed differences between the two species are quantitative rather than qualitative and despite the degree of genetic difference between them, they are each other's closest relatives in the genus *Entamoeba* (Clark and Diamond, 1997).

No consistent morphological differences exist between the two species but differences have been reported in surface properties and ultrastructure (section 1.1.3.) (Espinosa-Cantellano *et al*, 1998). There is also a growing list of biological differences that distinguish the two species. Key among these is the fact that while *E. dispar* can grow just as well as *E. histolytica* in xenic cultures, the same is not true of monoxenic or axenic cultivation. Monoxenic cultures of *E. dispar* can be obtained but their growth is usually much inferior to *E. histolytica* under identical conditions. Similarly, axenic cultivation of *E. dispar* has proved very difficult (Clark, 1995; Kobayashi *et al*, 1998; Kobayashi *et al*, 2000). It has been proposed that these growth differences are in part due to the relative ability of the two organisms to obtain nutrients by pinocytosis rather than by phagocytosis (the method used in xenic culture and in the host). Scanning electron microscopic studies have shown that axenic *E. dispar* trophozoites have very few pinocytotic vesicle openings in the plasma membrane compared to *E. histolytica*, hence supporting this view (Espinosa-Cantellano *et al*, 1998).

The most obvious difference between the two species is the clinical outcome of infection with *E. histolytica* versus *E. dispar*. *E. histolytica* can cause invasive intestinal and extraintestinal disease while *E. dispar* can not. But how accurate is it to refer to *E. dispar* as 'non-pathogenic'? *E. dispar* has been shown to be capable of producing focal intestinal lesions in animals (Chadee *et al*, 1985; Vohra *et al*, 1989; Espinosa-Cantellano *et al*, 1997; Costa *et al*, 2000) and destroying epithelial cell

monolayers *in vitro* (Espinosa-Cantellano *et al*, 1998). There is also some evidence that pathological changes may occur in humans (McMillan *et al*, 1984) though invasive lesions and symptomatic infections have to date not been reported.

Interestingly, the three key factors which are thought to confer *E. histolytica* with its extraordinary capacity to destroy human tissues and cause invasive disease have also been reported in *E. dispar* (Tannich, 1998). Dodson *et al* (1997) have shown that under axenic conditions, *E. dispar* expresses a Gal/GalNAc lectin similar to that of *E. histolytica*, although the adherence and cytotoxicity to target cells, including neutrophils, is reduced in *E. dispar*. This has led to speculation that the main physiological role of the receptor may be in mediating amoebic colonisation of the intestine. Pore forming proteins similar to those found in *E. histolytica* have also been reported in *E. dispar* (Leippe *et al*, 1993). Genes for all three isoforms have been found and bear high similarity to those of *E. histolytica* (Tannich, 1998). Since both species display pore forming activity, their likely role is that of a normal intracellular component for killing ingested bacteria. However, at the protein level only amoebapores A and B have been detected in *E. dispar* lysates and in reduced concentrations compared to *E. histolytica*, while amoebapore C is virtually absent. This may explain the reduced capacity of *E. dispar* to destroy cells. Functional homologues to two of the six cysteine proteinase genes identified in *E. histolytica*, are missing in *E. dispar* (Bruchhaus *et al*, 1996). Their absence is of particular interest, since approximately seventy percent of the total cysteine proteinase activity seen in *E. histolytica* is the result of the expression of these two genes. Furthermore, one of these two genes is the major thiol proteinase localised on the *E. histolytica* trophozoite surface. Whether absence of this structurally and functionally unique molecule in *E. dispar* is in part responsible for the organism's inability to cause invasive disease remains to be seen.

A multicopy gene family, *ariel*, has been described in *E. histolytica* and encodes asparagine-rich *E. histolytica* antigens which have 80% sequence identity to the amoebic vaccine candidate SREHP (Mai and Samuelson, 1998). Although SREHP is present in both *E. histolytica* and *E. dispar*, DNA sequences corresponding to the *ariel* gene family have not been found in *E. dispar* (Willhoeft *et al*, 1999b). Whether *ariel* proteins have a role in determining pathogenicity remains to be seen.

There are additional differences at the biochemical, immunological and genetic level (Diamond and Clark, 1993). Based on these differences, numerous methods have been devised and are now available for distinguishing the two species. These differences, the methods employing them and their advantages and disadvantages in allowing more accurate diagnosis and data collection are reviewed in section 1.4.

1.3. EPIDEMIOLOGY OF AMOEBIASIS

Information on the geographical distribution of amoebic disease and the etiological agent *E. histolytica* is incomplete and unreliable. Surveyed populations differ as do sampling and laboratory techniques (Walsh, 1986) and differentiation of *E. histolytica* and *E. dispar* has only been attempted in a small number of surveys (Jackson, 2000).

Despite the inherent limitations certain generalisations can however be drawn from existing data (Jackson, 2000). Firstly, both *E. histolytica* and *E. dispar* have a world-wide distribution, being found in cold, temperate and tropical climates. Both species are more prevalent in disadvantaged communities and in tropical areas, where the prevalence of amoebic infection depends largely on factors such as social and cultural habits, poor sanitation, nutrition, socio-economic status, and crowding. Recognised high-risk areas for acquiring amoebiasis include Central and South America, South Africa, South and Southeast Asia, and parts of the Middle East, and invasive disease seems to be more common in these regions. In North America and Europe where prevalence rates are low reports of amoebic infection and disease come from studies of selected groups (Healy, 1986). These include members of extended families, immigrants from or travellers to endemic areas, children, institutionalised populations and male homosexuals. Individuals who are not necessarily at an increased risk for infection but among whom the severity of disease is high include users of corticosteroids, patients with malignancies, individuals suffering from malnutrition, pregnant women, the very young and the very old (Ravdin, 1995).

On the other hand, many of the results emerging from these studies present conflicting views and pose new questions. Cross-sectional assessments were made of two *Entamoeba* infected populations in South Africa. Both the communities studied were from the Cape, a region of South Africa considered by many to be a low risk area for amoebiasis. One of the study groups was from Langebaan, which lies on the West Coast of the Cape and boasts formal accommodation with water-borne sewage and reticulated clean water. In contrast, subjects from the rural community of Rawsonville, in the Boland, live on wine farms with minimal sanitation. *E. histolytica* infections were found in approximately 2% of subjects from Langebaan but none in

Rawsonville. Both the *E. histolytica* prevalence level and the seropositivity observed in Langebaan were comparable to those of the endemic Durban area. In contrast, subjects from Rawsonville had >50% prevalence of *E. dispar* and all were sero-negative, implying that the community was not exposed to *E. histolytica*.

What accounts for the variation seen in prevalence figures between adjacent communities within endemic areas, and why are there discrete pockets where *E. histolytica* infection occurs needs to be determined.

Prevalence of *E. histolytica* infection was determined for a population of pre-school children, aged 2-5 years, from an urban slum of Dhaka, Bangladesh (Haque *et al*, 1999). Using antigen detection tests, *E. histolytica* infection was detected in stools of approximately 5% of asymptomatic children. But when cumulative level of exposure to *E. histolytica* was measured by the presence of serum antibodies specific for the adherence lectin, almost 50% of the population demonstrated serological evidence of amoebiasis by five years of age. Seventeen of the *E. histolytica* infected, asymptomatic children were re-examined at six and twelve months. Antibodies to the lectin persisted in the sera of all seventeen children over one year of follow-up, but stool infection cleared without treatment in fifteen children and following anti-amoebic medication in the other two during the same period. In a longitudinal study of asymptomatic carriers of *E. histolytica* in Durban, South Africa, it was shown that majority of such infections result in self-cure, with only about 10% progressing to invasive amoebiasis (Gathiram and Jackson, 1987).

It is not yet certain why most *E. histolytica* infections do not progress to invasive disease. Whether this is a reflection of the variation in host susceptibility or differences in *E. histolytica* strains remains to be seen.

Over the past 20 years several epidemiological surveys have been undertaken in South Africa and the resultant sero-epidemiological findings have provided invaluable indices that may be used in the interpretation of epidemiological surveys in other parts of the world. One of the key observations made was the fact while *E. dispar* occurred more frequently in females of all age groups, asymptomatic infection with *E. histolytica* was equally common in both males and females. Invasive amoebiasis,

however, occurred more frequently in males than females (Gathiram and Jackson, 1985). Why females should be resistant to invasion by the parasite is unknown.

Another observation made by Jackson *et al* (1985) was that all individuals, both symptomatic and asymptomatic, infected with *E. histolytica* were seropositive, with the vast majority (94-100%) being strongly positive. In contrast, only 21% of individuals harbouring *E. dispar* were seropositive, with 3% strongly positive. This compared favourably with results observed in individuals from whom *Entamoeba* were not isolated, where seropositivity was 14-20% and strong positive responses were observed in 2-4%. Thus it appears that the serological responses observed in amoebiasis are attributable to *E. histolytica*, while *E. dispar* does not elicit a detectable serological response. Using a slightly modified version of the ELISA described by Ravdin *et al* (1990) for detection of serum anti-lectin antibodies, Haque *et al* (1999) also determined that antibodies to the lectin were present in the sera of all children who were colonised with *E. histolytica* at the time of serum collection. In contrast, children infected with *E. dispar* at the time of serum collection did not have higher rates of seropositivity than uninfected controls. Interestingly, using the original anti-lectin IgG ELISA method of Ravdin and colleagues, seropositivity rates in north-eastern Brazil were not found to be significantly different between individuals colonised with *E. dispar* or *E. histolytica* and those whose stools were *Entamoeba* negative (Braga *et al*, 1998).

Why is it that *E. histolytica* infections in certain parts of the world do not result in detectable antibodies to the *E. histolytica* lectin? Do these results reflect technical differences or does this mean there are fundamental differences in the parasite and/ or host in one or other region?

Sexual transmission of *Entamoeba* is well documented amongst homosexual men (Jackson, 2000). It appears that homosexual men in big cities of western countries such as New York, Toronto and London have a high prevalence of infection with *E. dispar*, based primarily on the absence of clinical symptoms and negative serology but also on a few reports employing isoenzyme characterisation. In Japan, however, amoebic infection was first seen in male homosexuals with invasive amoebiasis (Takeuchi *et al*, 1990). According to the authors more than 50 cases of invasive

disease had been detected among this community in Japan by 1990. More recently, a retrospective study of 28 symptomatic amoebic patients in an east-southeast area of Tokyo showed that 26 of the 27 male patients were Japanese, 48% of whom indicated that they engaged in homosexual or bisexual practices (Ohnishi and Murata, 1997). In 64% of cases the infection was apparently contracted in Japan.

What accounts for the distinct epidemiological differences in amoebic infection in male homosexual communities between Japan and the western countries is not certain.

In summary, it is clear that the epidemiology of amoebiasis as it stands at present is full of conflicting reports and paradoxes. Infection with and final outcome of disease may be dependent on a variety of factors related to the host, the parasite and the environment. With respect to the parasite it is now clear that for future epidemiological studies we must be able to distinguish not only between *E. histolytica* and *E. dispar*, but also between different isolates of the same species. Several methods are available for differentiating between the two species. The advent and use of molecular biology techniques has meant that we can now begin to delineate individual strains too. The methods currently available for identification, differentiation and typing of *E. histolytica* and *E. dispar* isolates, and their advantages and limitations are discussed in section 1.4.

1.4. EPIDEMIOLOGICAL INDICES

1.4.1. FREQUENCY OF AMOEBIASIS

A major limitation in the interpretation of epidemiological data and determination of amoebic disease frequency is the use of multiple parameters e.g. prevalence, incidence, morbidity and mortality, by different research groups. Even comparison between studies with a common parameter and the delineation of morbidity and mortality become difficult due to variation in case definitions used by different researchers and clinicians (Walsh, 1986). Thus, on the one hand terms without established definitions (e.g. mild, moderate or severe illness) are used to describe clinical disease, while on the other, intestinal disease alone may be described using a number of terms (e.g. dysentery, colitis, rectocolitis, etc).

Another important limitation is sample bias. Many parasitological and seroepidemiological studies survey a non-representative sample of the population and, although a large number of cases may be included, the prevalence data generated are biased by including those admitted to hospitals and healthcare. Data from developed countries may be skewed by the fact that majority of the studies are done in defined, high risk groups such as institutionalised persons, recent travellers or immigrants, and homosexuals (Healy, 1986). Likewise, the absence or scarcity of published reports from certain areas of the world expected to be at high risk (e.g. Afghanistan, Ethiopia and Somalia) is more likely due to lack of medical interest or opportunity rather than absence of the parasite.

Yet another limitation is that the groups studied may vary in age, sex, socio-economic status, nutrition, presence of underlying disease and other factors not adequately described in the report, yet known to be associated with susceptibility to infection or a higher risk of invasive disease (see above). Similarly, prevalence may change seasonally and there appears to be both acquisition and loss of the infection throughout the year such that individuals appear to "lose" the infection for many months and then become re-infected. Hence, in endemic areas the incidence of

infection could approach 100% but the prevalence, if measured at any one point in time, would be much lower.

In contrast to intestinal disease, the frequency of amoebic liver abscess is considered a reliable measure of the rates of liver infection, as liver infection can be identified clinically, in the laboratory or through post-mortem studies (Martínez-Palomo and Cantellano Espinosa, 1998).

1.4.2. STOOL MICROSCOPY

A serious problem in the taxonomy of amoebae has been, and still is, the scarcity of morphological features on which to base classification. Classically, species descriptions in *Entamoeba* have relied on only a few characters: 1. The host in which it was found; 2. The size of the cyst and trophozoite; 3. The size and appearance of the nucleus; 4. The appearance of the chromatoid bodies which appear during encystation; and 5. The number of nuclei in the mature cyst. Clearly these criteria are prone to observer bias. Thus, while on the one hand distinct species may be grouped together due to the lack of distinguishing features, the opposite may also be expected, that is, an invalid species may be created based on minor morphological or other variation. These limitations are illustrated by the re-description of *E. histolytica* and *E. dispar* (Diamond and Clark, 1993) and the re-classification of the Laredo strain and other *E. histolytica*-like amoebae as *Entamoeba moshkovskii* (Clark and Diamond, 1991a).

Nevertheless, the cornerstone in the diagnosis of intestinal amoebiasis is still the microscopic detection of cysts and/ or trophozoites of *E. histolytica* in stools or scrapings from the affected mucosa (Reed, 1992). It has been suggested that diagnosis of invasive intestinal disease can be made reliably when trophozoites containing ingested red blood cells are found in stools, especially when this occurs in association with gastrointestinal symptoms (González-Ruiz *et al*, 1994b). However problems arise when cysts alone occur in the stools of healthy or diarrhoeic individuals, since microscopy does not distinguish between cysts of *E. histolytica* and *E. dispar* (Walsh, 1986; Reed, 1992).

Even when trophozoites are present, several factors can affect the results of microscopic diagnosis and detection. Trophozoites autolyse readily soon after passage and may be rapidly killed by exposure to a variety of agents e.g. salt and other antidiarroheal agents, antibiotics, antiprotozoal agents, oily laxatives, watery, soapy or hypertonic enemas and barium sulphate, to name a few. Ideally, therefore, fresh stools should be examined within 20 minutes in wet mounts (Walsh, 1986; Reed, 1992). An even more reliable means of differential diagnosis, in patients with dysentery, is detection of motile trophozoites by examination of a fresh preparation of the stool, mucus or ulcer discharge collected by sigmoidoscopy and biopsy. However, patients undergoing these procedures have frequently been hospitalised and may therefore have been given one or more of the agents that interfere with the identification of these organisms in stool.

While diarrhoetic cases may be detected with only one or a few specimens, examination of at least three stool specimens on separate days is required to detect >80%-90% of infections of asymptomatic cyst passage due to the intermittent shedding of cysts (Walsh, 1986). Likewise, comparison of several methods has revealed that examination of concentrates of stool specimens detects 40%-50% more infections than does examination of direct smears with saline and iodine or with saline alone. Accuracy in the detection of trophozoites and cysts and hence efficacy of microscopic determination is also maximised by examining iodine, trichrome and/ or iron-hematoxylin stained preparations of stool samples (Walsh, 1986). The skill of the laboratory personnel is also a key feature. Although, trophozoites in fresh samples can easily be differentiated from leukocytes due to the latter's lack of motility, the same is not true of preserved specimens and even skilled laboratory personnel have mistaken macrophages containing ingested red blood cells for amoebas on examination of both fresh and preserved specimens. Taken together, variability in one or more of these factors would not only lead to an under- or over- estimate of the frequency of infection but would also make it difficult to compare surveys from different parts of the world.

1.4.3. CULTURE

Detection of amoebae in culture is considered to be more sensitive than direct microscopic examination. However, the culture of amoebas from stool samples is time consuming and is not routinely used in most laboratories (Reed, 1992). Furthermore, *in vitro* studies have shown that even minuscule amounts of *E. histolytica* can outgrow *E. dispar* in culture (Clark and Diamond, 1993b). A major drawback would be that mixed infections with both species could be overlooked. Haque *et al* (1998) have shown that mixed infections with both *E. histolytica* and *E. dispar* could only be detected after PCR was used on DNA extracted directly from stool. Isoenzyme analysis and antigen detection failed to identify such infections.

1.4.4. SEROLOGY

Serology for antibodies to *E. histolytica* can be a very useful diagnostic tool, but must be utilised with a complete understanding of the humoral immune response to the parasite. Serum antibodies to amoebae develop only during *E. histolytica* infection and not during *E. dispar* infection (Jackson *et al*, 1985). In general, serum antibodies to amoebae are detected in up to 95% of all patients who present with liver disease, while up to 80% of patients with invasive colitis will have circulating antibodies present (Reed, 1992; Ravdin, 1995). Antibodies usually appear in the blood approximately seven days after the onset of clinical symptoms and are generally found with greater frequency in the more severe forms of invasive disease (Martínez-Palomo, 1993). The titres then gradually decrease to low but still detectable levels by about nine to twelve months. Therefore absence of serum antibodies to *E. histolytica* (usually after one to two weeks of symptoms) could be considered strong evidence against a diagnosis of invasive amoebiasis of the colon or liver.

However, exceptions occur and higher levels of antiamoebic antibodies are not necessarily correlated with the severity of disease in man. Furthermore, a small proportion of patients may continue to have high titres for years after invasive amoebiasis is cured or when subclinical amoebic infection is controlled (Martínez-Palomo, 1993). The serological response in patients who are asymptomatic cyst

passers is highly variable and a response may be detected in as few as 0, to as many as 70% of the population, depending on the prevalence of the parasite and disease in certain countries and population groups. This accounts for the high prevalence of seropositivity or "background noise" seen in areas of endemicity (Healy, 1986). This failure to differentiate between present and past *E. histolytica* infections limits the use of serological tests in the diagnosis of amoebiasis in endemic settings. On the other hand, serology can be of great diagnostic value in countries where amoebiasis is not prevalent.

Virtually all known serological tests have been employed to detect antiamoebic antibody, including agar gel diffusion (AGD), immunofluorescent antibody (IFA), indirect haemagglutination (IHA), radioimmunoassay (RIA), counterimmunoelectrophoresis (CIE) and enzyme-linked immunosorbent assay (ELISA). The tests in ascending order of sensitivity are IHA, CIE, IFA, IHA+IFA and ELISA. For the most part these tests give virtually no false positive results with sera that have been collected very early in the infection. Furthermore, IHA and ELISA tests measure both present and past infections (IHA may be positive for more than ten years and ELISA for over three years) after clinical and parasitological cure has been achieved and in the absence of re-infection (Reed, 1992; Ravdin, 1995). Hence they may be useful for seroepidemiological studies (see above). In contrast, results of CIE and IFA become negative within six to twelve months of successful therapy i.e. they likely measure active or recent infections, and thus a positive result of one or more of these tests could be considered diagnostic (Reed, 1992; Ravdin, 1995).

However, taken together these facts indicate that the titres, as measured by the different tests, fall at different rates following treatment. Unequal rates of disappearance and the lack of titre correlation with severity of disease, in spite of fairly good positivity correlation, suggests that the various serological methods detect different families and levels of antibodies (Martínez-Palomo, 1986; Walsh, 1986). Hence, even though there may be no basic discrepancy between the results obtained from two or more tests for the same population, comparability between results is difficult because each technique measures a different parameter. The situation may be further complicated by the fact that even with the same method, different authors and

laboratories may use different titres as cut-off levels to separate a positive from a negative sample.

1.4.5. ISOENZYME ELECTROPHORESIS

Multi-locus enzyme electrophoresis or isoenzyme analysis is currently considered the gold standard for epidemiological studies, as it reliably differentiates between *Entamoeba histolytica* and *E. dispar* (Diamond and Clark, 1993; Jackson, 2000). Several enzymes have been found to distinguish between *E. histolytica* and *E. dispar* isolates (Sargeaunt *et al*, 1978; Sargeaunt, 1987a & b; Blanc, 1992) but the four enzyme system defined by Sargeaunt and co-workers, consisting of malic enzyme (ME), phosphoglucomutase (PGM), glucose phosphate isomerase (GPI) and hexokinase (HK) is used most commonly (Fig. 2).

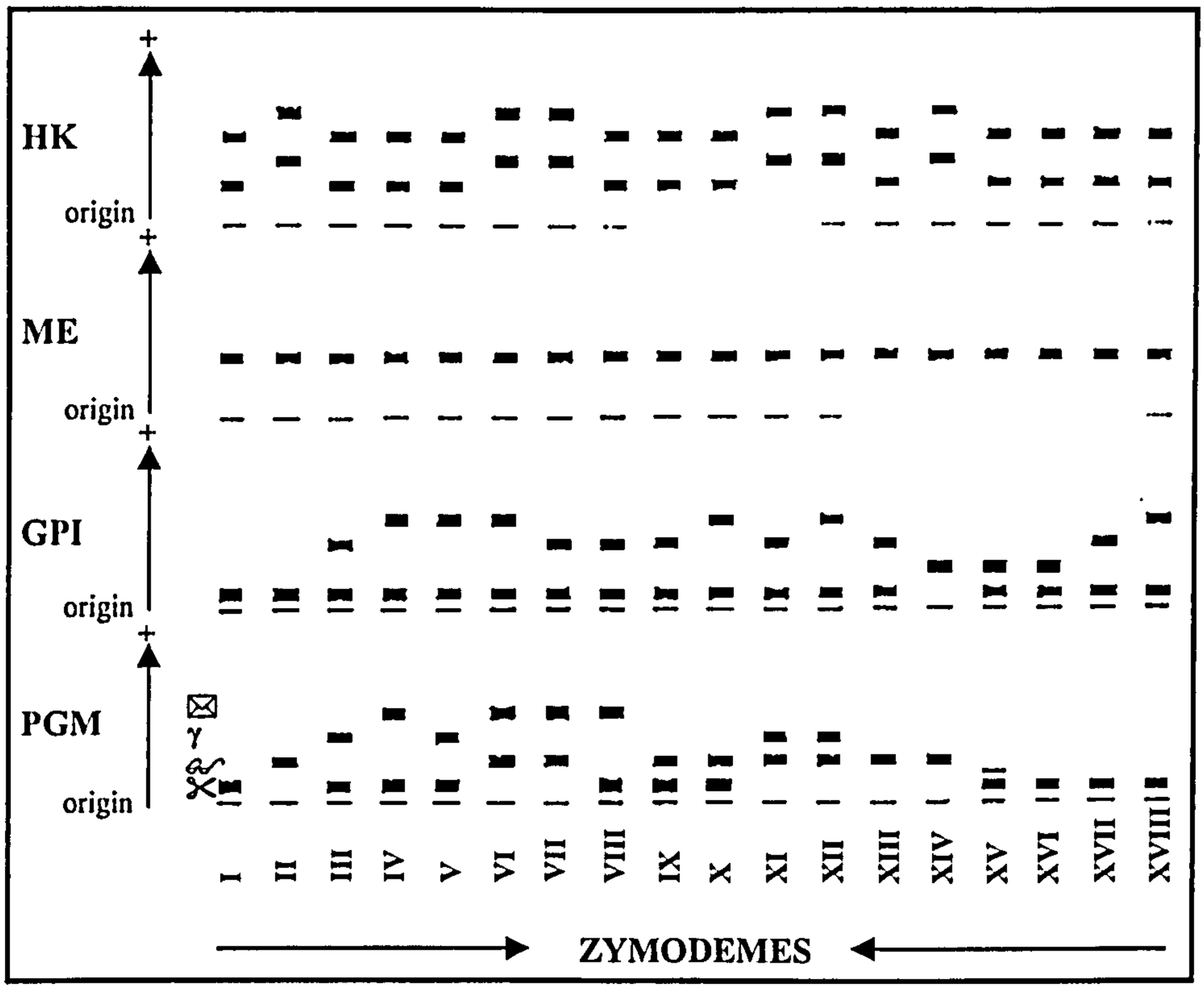


Fig. 2. Schematic Representation of Multilocus Enzyme Electrophoresis. Taken from Sargeaunt *et al* (1982).

ME migrates to the same position from the origin for both species and is therefore used to confirm their presence, whereas HK always shows double bands in either one of two positions. However, by examining the characteristic banding patterns of PGM and GPI simultaneously, more than twenty zymodemes (strains allocated on the basis of the particular isoenzyme patterns observed) can be recognised, of which nine belong to *E. histolytica* and the rest to *E. dispar*. Thus isoenzymes can be used as tools to monitor not only inter- but intra-species differences as well.

The stability and reliability of zymodeme patterns in differentiating between *E. histolytica* and *E. dispar* is well established (Jackson *et al*, 1992; Clark *et al*, 1992). No evidence has been seen of a change from a non-pathogenic to a pathogenic zymodeme and vice a versa even under the influence of stresses, such as adaptation to a new medium, exposure to high levels of antibiotics, alteration or loss of the bacterial flora, and exposure to irradiated bacteria.

However, isoenzyme analysis clearly has limitations in that patterns of only two enzymes, PGM (absence of α-band together with the presence of the β-band) and HK (advanced bands) are considered markers of *E. histolytica* and show consistently different patterns in starch gels for the two species (Sargeaunt, 1987a & b). In that too, zymodeme XIII is an exception as it is designated as *E. histolytica*, but does not conform to the classical HK pattern of this species. This raises doubts upon the earlier conclusion that both PGM and HK are markers of *E. histolytica*. It has also been indicated that the presence of γ and δ bands of GPI and PGM is dependent on specific culture conditions, namely the amount of starch in the medium (Blanc and Sargeaunt, 1991). It was later established that these variable or secondary bands disappeared upon axenization (Jackson *et al*, 1992) with conversion of *E. histolytica* zymodeme XI to *E. histolytica* zymodeme II and *E. dispar* zymodeme III to *E. dispar* zymodeme I through loss of the respective γ-bands.

Thus, while the differentiation of *E. histolytica* and *E. dispar* by isoenzyme electrophoresis appears to be invaluable (see above), there are several doubts as to the reliability of isoenzyme analysis in characterising isolates within these species. At present there are only four principle zymodemes (those found in axenic cultures), available for use as differentiating markers. These are *E. histolytica* zymodemes II,

XIV and XIX, and *E. dispar* zymodeme I (Jackson and Suparsad, 1997). Finally, the need for culture and large quantities of amoebae as well as the fact that even species differentiation may be suspect, because of possible selective growth of one or the other type of parasite from a mixed infection, limits the usefulness of isoenzyme analysis.

1.4.6. ANTIGEN CAPTURE TESTS

Enzyme linked immunosorbent assays (ELISAs) have proven to be a rapid and sensitive method for identification and diagnosis but have little role in intra-species differentiation. One of the earlier ELISAs was described by Grundy *et al* (1987), which employed polyclonal antiamoebic antibodies and could be used directly on stool samples. Several ELISAs based on antigen detection using monoclonal antibodies (mAbs) have been reported since. Strachan *et al* (1988) and González-Ruiz *et al* (1992) both developed promising *E. histolytica* specific mAb-based immunofluorescence tests. However the tests were limited by the fact that the mAbs were uncharacterised, the tests required culturing, and that imunofluorescence is largely a subjective method. Subsequently, González-Ruiz *et al* (1994a) were able to eliminate the culture step and demonstrated that the new faecal antigen capture test (FAC-ELISA) had 100% sensitivity, specificity and efficiency for the diagnosis of amoebic dysentery. A commercial ELISA, the ProSpecT ® EIA; Alexon Inc., Sunnyvale, CA 94089 USA, which also uses mAbs to detect *E. histolytical E. dispar* specific antigen in stool samples has been evaluated by various researchers (Jelinek *et al*, 1996; Ong *et al*, 1996). However, this test does not differentiate between *E. histolytica* and *E. dispar* and is therefore of limited clinical value.

The production of well characterised mAbs to the Gal/GalNAc adherence lectin has allowed development of highly specific stool antigen capture ELISAs (Petri *et al*, 1990; Abd-Alla *et al*, 1993; Haque *et al*, 1994). The first field trials with commercial antigen detection kits based on mAbs to the lectin were carried out in 1995 by Haque and co-workers and it was shown that the stool antigen assay was more rapid and as sensitive and specific as isoenzyme analysis. There are two kits available through TechLab, Inc., Blacksburg, VA., the *Entamoeba* test (designed to detect but not

differentiate the antigens of *E. histolytica* and *E. dispar* in stool specimens) and the *E. histolytica* test (designed to specifically detect *E. histolytica* in stool specimens). Identification of *E. dispar* can then be made by subtraction. Together, therefore, these tests have the advantage of being able to distinguish between the species. Although comparative results using trophozoites in culture clearly show that PCR is more sensitive and selective than the ELISA-based kits (including the TechLab kit) in the detection and characterisation of *E. histolytica* and *E. dispar* (Mirelman *et al*, 1997), field studies that compared PCR and the TechLab Kit directly on stool samples suggest that both these methods perform equally well (Huston *et al*, 1999).

Abd-Alla *et al* (1993; 2000) have demonstrated that ELISAs, based on epitope-specific monoclonal antibodies to the Gal/GalNAc lectin can also be used to detect antigen directly in serum and saliva samples. That the TechLab kit can detect lectin antigen directly in serum and liver abscess pus samples, as well as stool, has recently been demonstrated (Haque *et al*, 2000). However, the authors reported that although the test was able to detect serum antigen in almost all patients (96%) who had not received treatment with metronidazole, the majority (91%) of the serum lectin antigen-positive patients became negative within two weeks of starting metronidazole treatment. Similarly, all the liver abscess pus samples collected prior to treatment were positive for lectin antigen. Hence, the test is apparently only sensitive and useful prior to antiamoebic therapy.

1.4.7. NUCLEIC ACID ANALYSIS

Polymerase chain reaction (PCR) based studies have an advantage over isoenzyme and mAb analysis as they can amplify large amounts of a specific DNA sequence from the organism of interest using very little starting material (theoretically a single cell) in a very short time. Thus, by virtue of being highly specific, sensitive and rapid, PCR is fast becoming the epidemiological index tool of the future. Furthermore, PCR can be used for the detection of *E. histolytica* DNA isolated directly from stool (Acuña-Soto *et al*, 1993; Sanuki *et al*, 1997; Núñez *et al*, 2001) and liver abscess pus samples (Tachibana *et al*, 1992; Zengzhu *et al*, 1999). PCR has even been applied

directly to pus samples without the need for any pretreatment and DNA isolation steps (Zaman *et al*, 2000).

Species differentiation based on DNA amplification and its application in the field has been the focus of research of many groups. Several groups have described and used PCR differentiation based on amplification of repetitive regions in the extrachromosomal circular DNAs of *E. histolytica* and *E. dispar* (Garfinkel *et al*, 1989; Samuelson *et al*, 1989; Acuña-Soto *et al*, 1993; Aguirre *et al*, 1995; Britten *et al*, 1997; Verweij *et al*, 2000). In some cases the final product is detected using a colourimetric method based on an ELISA, thereby obviating the need for electrophoresis, ultra-violet light and photographic equipment (Aguirre *et al*, 1995; Britten *et al*, 1997; Verweij *et al*, 2000). Others have designed species-specific primers that amplify a region of the genes encoding the 30-kDa antigens of *E. histolytica* and *E. dispar* (Rivera *et al*, 1996; Sanuki *et al*, 1997). Tachibana *et al* (1991; 1992) have also amplified a region of the 30-kDa antigen from both *E. histolytica* and *E. dispar* followed by restriction-endonuclease digestion of the amplified product as a determinant of species.

PCR methods based on amplification of the small subunit rRNA (SSU-rRNA) from both species have also been employed (Clark and Diamond, 1991b; Katzwinkel-Wladarsch *et al*, 1994; Novati *et al*, 1996). In the first instance all three groups used a single primer pair to amplify products from both *E. histolytica* and *E. dispar*. Species differentiation was then done either by employing a second set of primers in a nested PCR (Katzwinkel-Wladarsch *et al*, 1994) or by taking advantage of the presence of a polymorphic restriction cleavage site within the amplified stretch (Clark and Diamond, 1991b; Novati *et al*, 1996). Sequence differences in the coding regions of the multicopy SSU-rRNA genes have also been used to design species-specific primer-pairs (Clark and Diamond, 1991b; Troll *et al*, 1997).

Primer pairs designed from cDNA clones representing single copy coding sequences have been used to distinguish between the two species (Tannich and Burchard, 1991). Restriction fragment analysis of the PCR product was then used for the differentiation. Gomes *et al* (1997) have used the same primers to amplify the single

copy DNA and then used the amplified band as template for the low-stringency single specific primer-PCR (LSSP-PCR).

Most of these methods are, however, limited by their use of two separate sets of primers or by the need for a second step that discriminates one species from other. Recently, Núñez *et al* (2001) have described a multiplex PCR methodology which allows simultaneous detection of both species using two pairs of species-specific primers in a single reaction mixture. However, a potential limitation of this method is that large differences in the length of target DNAs would favour amplification of the shorter target over the longer one (Núñez *et al*, 2001). A nested multiplex PCR (Evangelopoulos *et al*, 2000) offers a possible alternative.

A key advantage of using PCR technology is the reliability in clearly documenting mixed infections, where techniques such as microscopy and antigen detection tests have failed (Haque *et* al, 1998). In their study Haque *et al* (1998) found that, out of 12 mixed infections detected by the nested PCR performed directly on stool, 8 specimens had been identified as *E. histolytica* by antigen detection tests.

With one exception none of these methods however can detect intra-species variation. Clark and Diamond (1991a; 1991b) have described a technique known as 'riboprinting' which allows differentiation of the two species by examination of restriction enzyme site polymorphisms in PCR-amplified small subunit ribosomal RNA genes (see above). It seems that 'riboprinting' can also be used for intra-specific differentiation in some species (Clark and Diamond, 1997). There are four ribodemes (a population of organisms within a species that share the same riboprint pattern) in *E. coli*, two in *E. gingivalis* and six in *E. moshkovskii*. No variation in riboprint pattern was observed within either *E. histolytica* or *E. dispar* despite the large number of isolates tested.

1.5. REPEATED GENES & GENETIC VARIATION

1.5.1. CLASSIFICATION OF REPEATED-DNA & GENETIC VARIATION IN EUKARYOTES

Multicopy RNA encoding genes with identical or nearly identical sequences represent a major class of repeated-DNA. Examples include rRNA, 5S RNA and tRNA gene families (Long and Dawid, 1980). A wide variety of genomic organisation is displayed by this DNA class. The multiple copies may exist as integrated tandemly arranged units, each unit composed of a gene region and a noncoding spacer which separates one gene from the next. One or several clusters of these tandemly arranged genes may occur. This form of genomic arrangement is seen in the rDNA of *Leishmania* and many higher eukaryotes e.g. insects, amphibians and mammals, as well as in the 5SRNA of most eukaryotes. Alternatively the multicopy rDNA may occur as extrachromosomal molecules as seen in some unicellular eukaryotes e.g. *Paramecium* and *Naegleria*. The high copy number tRNA genes of eukaryotes also show multiple types of arrangements. They may exist as a few contiguous copies within a small DNA stretch that represents a single locus as seen in some *Drosophila* species. Alternatively, one or two tRNAs may lie within tandemly repeated-DNA fragments e.g. *Xenopus*, while the majority are found as dispersed individual transcription units e.g. yeast.

Copy numbers of some of these RNA coding genes are polymorphic among populations. Large variations in ribosomal gene numbers have been seen between closely related species and even between different strains of the same species e.g. in *Drosophila* and mouse.

Protein coding genes characterised by the presence of one or more blocks of tandemly arranged internal repeat units represent another class of repetitive DNA in some eukaryotes e.g. *Plasmodium* (Kemp *et al*, 1987). Variations in the length, sequence

and number of repeats occur among and within species and some of these internally repetitive protein coding genes have been used as genetic markers for work on population genetics and epidemiology of malaria (Paul *et al*, 1995).

Although repeated sequences occur in the RNA- and protein-coding DNA fraction of the genome, the majority of the repeated-DNA sequences in eukaryotes are present in the more substantial noncoding component and are usually referred to as 'repetitious DNA' (Darnell, 1990). Major classes of non-coding repetitious DNA include simple-sequence DNA, intermediate repeat DNA, and spacer or connecting DNA. The terms simple sequence DNA and intermediate repeat DNA are based on the relative reassociation kinetics of these DNA fractions. Simple sequence DNA reassociates almost immediately. Intermediate repeat DNA reassociates at an intermediate rate which falls in the range between that seen for simple sequence DNA and the slowest reassociating fraction which is composed primarily of single copy genes.

Most simple sequence DNA occurs in very long stretches of up to 10^5 to 10^6 base pairs in length, which are composed of short, 5 to 10 bp long tandemly repeated units, although tandem repeats of 20 to 200 nucleotides are also known to occur. Simple sequence DNA is often referred to as 'satellite DNA' because, by virtue of its unusual base composition and hence different buoyant density; it separates from the main DNA band as a 'satellite' band during equilibrium density-gradient centrifugation. However, not all simple sequence DNA separates from the bulk DNA and therefore the term simple sequence DNA is preferred to satellite DNA. Each animal and plant species has several types of simple sequence DNA characterised by different repeat units. In some species the repeating units may have very similar sequences, which suggests a common ancestral relationship between two or more simple sequence DNA types. However, although simple sequence DNA may be highly conserved in sequence, a variable number of repeats occurs due to unequal crossing over events resulting in length variations.

'Minisatellites', called thus to distinguish them from the more common 10^5 to 10^6 base pairs long satellites, are one type of simple sequence DNA. They exist as short sequences of about 15 to 100 bases that are repeated only 20 to 50 times in tandem, thus spanning a total stretch of only 1 to 5 kb. These loci are found scattered

throughout the genome. DNA fingerprinting is a commonly used technique that exploits the presence of minisatellites and the existence of length variations to differentiate between individuals. 'Microsatellites', which seldom exceed a total length of 100 bp, are another type of simple sequence repeat and are distinguished by shorter repeat units of about 1 to 10 base pairs (bp), although di-, tri- and tetranucleotide repeats are the most common. Like minisatellites, the number of repeat units in microsatellites shows high degrees of polymorphism between individuals. Intraspecific variations based on microsatellite markers have been described in unicellular eukaryotes e.g. *Trypanosoma* (Oliveira *et al*, 1998), and among higher eukaryotes e.g. fungi (Moon *et al*, 1999) and humans (Calafel *et al*, 1998).

Most, but not all, of simple sequence DNA is found near the centromeres, with the rest being found within the arms and telomeric regions of chromosomes. Although this type of DNA is generally not transcribed, it has been ascribed structural or organisational roles.

The intermediate repeats, which are sub-classified into two families, namely short interspersed elements (SINES) and long interspersed elements (LINES), do not exist in tandem repeats like simple sequence DNA but rather are interspersed at many places in the genome.

The noncoding spacers separating tandemly arranged rRNA and 5SRNA genes also show considerable variations in many eukaryotic organisms ranging from the protozoan *Leishmania* to higher eukaryotes e.g. humans. These variations are attributed to the fact that spacers are often internally repetitious. In fact, differences in the number of internal repeats between two or more spacers results in length heterogeneity which can even be seen between successive copies of repeated genes within a tandemly arranged cluster.

1.5.2. REPEATED-DNA & GENETIC VARIATION IN *E. histolytica*

It is estimated that repeated-DNA may compose about 20% of the *Entamoeba* genome (Bhattacharya *et al*, 2000) of which the ribosomal circles constitute the majority. As already mentioned (section 1.1.12.), in *E. histolytica* the rRNA genes are estimated at about 200 copies per genome and are located on circular, extrachromosomal DNA molecules, 24.5 kb in size (Bhattacharya *et al*, 1989; Huber *et al*, 1989). The complete sequence of one circle has been published (Sehgal *et al*, 1994) and it carries two copies of the rRNA cistron as an inverted repeat. However, the number of rRNA cistrons per circle is variable and the presence of only one rRNA gene copy has been reported in some *E. histolytica* strains (Rahman and HK-9) as well as strains of other *Entamoeba* species (*E. moshkovskii* Laredo).

In a recent study, nucleotide sequence analysis of the 5.8S rRNA gene and the flanking internal transcribed spacers (ITS-1 and -2) of *E. histolytica* strain HM-1:IMSS revealed that the sequences are fully conserved among the two transcription units and in fact among the multiple rRNA copies (Som *et al*, 2000). However, pronounced differences were present amongst the four *Entamoeba* species examined, with some variation also evident between different strains of *E. histolytica* and *E. dispar*.

The distribution of restriction enzyme sites in this repeated-DNA is also highly polymorphic. Using *Eco*RI, Bhattacharya *et al* (1988) have demonstrated restriction fragment length polymorphism (RFLP) between strains of *E. histolytica* as well as between different *Entamoeba* species, with few restriction fragments seen in common in the latter case.

I have already mentioned that in addition to ribosomal genes several classes of tandemly repeated-DNAs are carried on the extrachromosomal plasmid and that some of them exhibit length variation due to changes in the number of repeats (section 1.1.12.). These polymorphic fragments (e.g. *Pvu*I (Estevez *et al*, 1989), *Sca*I (Mittal *et al*, 1992b), *Dra*I (Mittal *et al*, 1992a) and *Hin*fI (Sehgal *et al*, 1993)) lie in the

noncoding regions downstream and upstream of the rDNA transcription units (Table 1). In Southern and dot blot hybridisation studies the *Dra*I, *Sca*I and *Pvu*I families appear to be specific for *E. histolytica* isolates only and were not detected in *E. dispar* or any of the other *Entamoeba* species tested (Estevez *et al*, 1989; Mittal *et al*, 1992a). Variations in the presence or absence of these repeats and restriction fragment length polymorphism have also been reported among strains from different *Entamoeba* species and even among clones of the same *E. histolytica* strain (HM-1:IMSS clones 3, 6 and 9) (Estevez *et al*, 1989; Bhattacharya *et al*, 1992). In fact, length polymorphism in the *Dra*I repeats were seen in the same clone (clone 6) over a three year period of continuous subculturing. These repeat families have been likened to the intergenic spacers separating successive rDNA units (as described above) (Sehgal *et al*, 1994).

A DNA stretch designated strain-specific gene (SSG) (Burch *et al*, 1991) or *Tr* (Sehgal *et* al, 1994) has been described in the upstream region of the HM-1:IMSS rDNA circle, for which a noncoding transcript has been detected. Dot blot hybridisation does not produce any signal with *E. dispar* DNA. A notable feature of this transcript is the presence of nine 26-nucleotide tandem repeats. PCR amplification of the repeat domains revealed considerable variation in the number of these repeats in fifteen of the eighteen *E. histolytica* strains tested (Clark and Diamond, 1993a). The number of repeats observed appeared to be stable under a variety of stresses such as long term *in vitro* cultivation, axenization, cloning and animal passage. However, the complete absence of the *Tr* transcript from certain strains of *E. histolytica* (HK-9, Rahman and DKB) was also reported. Interestingly enough it was observed that this absence correlated with the presence of only one rRNA cistron in these strains.

Loss of one of the two rDNA units from the circular episome of *E. histolytica* strain HM-1:IMSS has recently been reported (Gosh *et al*, 2001). The loss appears to be a result of intramolecular recombination between the noncoding repeats in the circle and is accompanied by loss of the SSG containing region. Since two rDNA units are also found in the *E. histolytica* strain HM-1:IMSS being used for genome sequencing (http://www.tigr.org/tdb/edb2/enta/htmls/), it is likely that these recombinational events are a relatively recent phenomenon. Hence, in view of these observations, the

SSG region may be unsuitable as a strain specific marker if it is subject to such frequent recombinational loss.

Table 1

RELATIVE DISTRIBUTION, SIZES & FREQUENCY OF OCCURRENCE OF VARIOUS REPETITIVE DNA CLASSES IN *E. histolytica*

Entamoeba strain	Repeat Loci	Length of one Repeat unit (bp)	Total number of Repeat units
Tandemly repeated families in the extrachromosomal rDNA circle			
Loci located upstream of the rDNA cistrons			
HM-1:IMSS clone6	*Pvu*I	145	11
HM-1:IMSS clone6	*Sca*I	144	6
HM-1:IMSS clone6	*Hin*fI	653	2
HM-1:IMSS clone6*	SSG/ Tr*	26	9
Loci located downstream of the rDNA cistrons			
HM-1:IMSS clone6	*Dra*I	170	10
HM-1:IMSS clone6	*Sca*I	144	7
Protein coding genes with internal tandem repeats			
HM-1:IMSS	SREHP	36 24	6 5
HM-1:IMSS	Chitinase	21	8

Note: "*" The repeat locus *Tr* was identified from strain HM-1:IMSS clone 6. The specific clone used for identifying the SSG locus has not been reported.

Protein encoding genes displaying tandemly repeated internal structures have also been found to exhibit significant polymorphism among *E. histolytica* isolates (Table 1). One such gene is the serine-rich *E. histolytica* protein (SREHP) which encodes an immunodominant surface antigen. It exhibits an internally repeated tandem array of related dodeca- and octa-peptides (Stanley *et al*, 1990; Köhler and Tannich, 1993). The gene is also present in *E. dispar*. There are differences in repeat number, sequence and arrangement, as well as sequence variations in repeat flanking regions

between *E. histolytica* and *E. dispar*. Both PCR fragment size variations and restriction enzyme site polymorphisms have been evidenced in the internal repeat regions among the eighteen different *E. histolytica* isolates tested (Clark and Diamond, 1993a). Moreover, these patterns of variation are stable under a range of stressful conditions (see above). Homologous genes in *E. dispar* also show similar patterns of both size and restriction site polymorphisms between different strains (Paez and Clark, unpublished data).

The second protein coding gene to demonstrate polymorphism is that encoding chitinase (de la Vega *et al*, 1997). Homologous chitinase genes have been described in *E. histolytica*, and other species tested including *E. dispar*. At the amino terminus of the protein are present tandemly arranged multiple copies of a degenerate 7-amino acid sequence. Both sequence and repeat number variations exist between *E. histolytica* and *E. dispar*. Furthermore, the repeat region shows considerable size polymorphism among *E. histolytica* and *E. dispar* isolates (Samuelson *et al*, 1997; Ghosh *et al*, 2000).

These and other protein coding genes are present on linear chromosomes, which also contain several classes of tandemly repeated or dispersed multicopy DNAs not associated with the extrachromosomal circular DNA repeats. Distinct size variation has been seen among homologous chromosomes of different isolates (Willhoeft and Tannich, 1999). Although the function and organisation of the aforementioned repetitive classes is not known, it is possible that the chromosomal size variation is due to deletion or expansion of these and other repeated sequences in the telomeric and subtelomeric regions, as observed for other eukaryotes (Darnell, 1990). For the most part, these repetitive DNA classes have not been investigated for their ability to reveal polymorphisms among *E. histolytica* isolates.

Comparison of the repetitive DNA family described by Lohia *et al* (1990) among several clinical isolates of *E. histolytica* indicated some degree of fragment size polymorphism in Southern blot hybridisation patterns. Sequence analysis of one of the members from the repetitive DNA family revealed the presence of several groups of tandemly arranged direct internal repeats.

A repetitive DNA element with two related types of tandemly arranged direct internal repeats has been reported by Michel *et al* (1992). This internally repetitive DNA itself appears to exists in tandem arrays and has significant similarity to the internally repetitive fragment described by Lohia *et al* (1990). Interestingly, dot blot hybridisation results suggest that this repetitive element is conserved among *Entamoeba* species.

Recently, Willhoeft and Tannich (2000) have defined a repeat element with two blocks of internal tandem repeats from *Entamoeba histolytica* strain HM-1:IMSS. The repeat blocks show remarkable similarity to those seen in the repeated-DNAs described by Lohia *et al* (1990) and Michel *et al* (1992).

Preliminary Southern hybridisation analysis of a tandemly repeated linear DNA characterised by Huang *et al* (1997) indicates that *E. histolytica* may contain more copies of this repeat than *E. dispar*. This suggests that the repeat may be used as a potential tool for distinguishing between the two species in clinical samples. This tandemly arranged DNA in turn contains several types of internal repeats which exist in tandem only or as both tandem and solitary copies.

The repetitive DNA element identified by Mittal *et al* (1994) displays marked differences in Southern hybridisation patterns between *E. histolytica* and *E. dispar* isolates. However no inter-strain differences could be detected in *E. histolytica*. Dot blot hybridisation gave no signal when DNA from other *Entamoeba* species was tested. This multicopy DNA does not display any tandem internal repeats.

Another repeat element, specific to *E. histolytica* only, is the 'interspersed element' (IE), which is reportedly present in 500 copies per cell (Cruz-Reyes *et al*, 1992 & 1995). Southern hybridisation patterns indicate that it has widespread genomic distribution in several strains of *E. histolytica*.

1.6. AIMS & OBJECTIVES

The key objective of this study is to identify multiple new polymorphic DNA markers and to gauge their usefulness for studying the molecular epidemiology of amoebiasis by testing them in a number of *E. histolytica* samples.

Attempts will be made to detect sequences corresponding to these markers in *E. dispar* so as to develop a method for simultaneous species and strain identification.

The general utility of these markers will be validated by examining samples from a wide geographic range including samples from regions of endemicity, from family groups and from amoebiasis outbreaks. Effects, if any, of using cultured versus faecal DNA, the methods of culturing and DNA extraction will be ascertained.

Finally, genomic organization of these markers will be assessed through a number of means to gather information on their copy number, arrangement and chromosomal distribution.

CHAPTER 2

GENERAL METHODS

2.1. *ENTAMOEBA* ISOLATES

In general, details of the strains and isolates used in this study, the various population groups they belong to and sources from which they were obtained are given in the relevant chapters.

The two reference strains being maintained by us are the axenic *E. histolytica* strain HM-1:IMSS clone 9 and the monoxenic *E. dispar* strain SAW760.

2.2. CULTIVATION

In our laboratory, axenic cultures of *E. histolytica* were grown in either YI-S (yeast extract, iron, serum) medium (Diamond *et al* 1995) or LYI-S-2 (liver digest, yeast extract, iron, serum) medium (Clark and Diamond, 2002) supplemented with 15% heat inactivated adult bovine serum (Sigma-Aldrich) before use (Appendix I). Cultures were maintained by sub-culturing at 72- and 96-hour intervals and incubated at 36°C.

Monoxenic cultures of *E. dispar* were maintained in YI-S or LYI-S-2 medium supplemented with 10% heat inactivated adult bovine serum and *Crithidia fasciculata*, at 36°C. Subcultures were also performed at 72- and 96-hour intervals.

Information on culture media used for the samples provided to us by different sources is given in the appropriate chapters.

2.3. ISOLATION OF DNA

DNA was isolated as described by Clark and Diamond (1991a) and Clark (1992).

Culture tubes containing 72-hour amoeba cultures were chilled in an ice-water bath and sedimented by centrifugation at 275 x g. The sedimented trophozoites were washed with sterile 1 X phosphate buffer saline (PBS) to remove traces of medium and the washed cell pellets resuspended in lysis buffer (0.1 M EDTA, pH 8.0/ 0.25% sodium dodecyl sulphate (SDS)). Proteinase K was added to 0.1mg ml^{-1} and the mixture incubated at 55°C for 1 hour. This was followed by addition of NaCl to 0.7 M and cetyl trimethylammonium bromide (CTAB) to 1% followed by incubation at 65°C for 15 minutes. Under these salt conditions CTAB binds the abundant carbohydrate in these cells. The lysate was then mixed with an equal volume of chloroform and centrifuged at 10,000 x g. Under these conditions the CTAB-carbohydrate complex forms an insoluble interface, leaving the DNA in the clear supernatant. DNA was then extracted with phenol: chloroform: isoamyl alcohol (25:24:1) and precipitated with 2.5 volumes of absolute ethanol. The precipitated DNA pellet was washed with 70% ethanol, air dried and dissolved in 10 mM Tris-Cl, pH 8.5 buffer.

Non-polysaccharide inhibitors and salts were removed by passing the resuspended DNA over a Microspin™ S-200 HR column (Amersham Pharmacia Biotech Inc). RNA was removed by the addition of RNase to 0.05 $\mu g\ ml^{-1}$.

2.4. ISOLATION OF RNA

Total RNA was isolated using a modification of the single-step acid guanidinium thiocyanate-phenol-chloroform method described by Chomczynski and Sacchi (1987). Culture tubes containing 72-hour old cultures were chilled, sedimented and washed with sterile 1 X PBS as above and the washed cell pellets were lysed in a buffered aqueous solution of guanidinium thiocyanate (GTC) and N-lauroyl sarcosine (Amersham Pharmacia Biotech). GTC is a very effective protein denaturant and

ensures rapid inactivation of endogenous RNases, thereby protecting the RNA. RNA in lysates was extracted with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) by first shaking the mixture vigorously for 10 seconds, chilling it on ice for 15 minutes and finally centrifugating it at 10,000 x g and 4°C for 20 minutes. RNA in the aqueous phase was precipitated with 2 volumes of isopropanol at -20°C for 1 hour followed by centrifugation at 10,000 x g for 20 minutes in the cold. The resulting RNA pellet was again dissolved in the GTC and N-lauroyl sarcosine solution and re-precipitated with 1 volume of isopropanol at -20°C for 1 hour. RNA was sedimented at 10,000 x g for 10 min at 4°C, resuspended in 75% ethanol, sedimented, vacuum dried for 15 minutes and dissolved in diethyl pyrocarbonate (DEPC) treated 10 mM Tris-HCl, pH 7.5/ 1 mM EDTA buffer. The concentration of RNA in the final solution was determined by spectrophotometry ([RNA] = A_{260} x 40 μg ml^{-1}) using a UV-Spectrophotometer (Shimadzu).

Polyadenylated RNA was isolated using the QuickPrep™ Micro mRNA Purification kit (Amersham Pharmacia Biotech). Briefly, the 1 X PBS washed and sedimented cells were disrupted in a buffered solution containing a high concentration of GTC. The mixture was then diluted with 2 volumes of a 10 mM Tris-HCl, pH 7.5/ 1 mM EDTA buffer (elution buffer) and sedimented by centrifugation at 10,000 x g for 1 minute. Dilution reduces the GTC concentration to a level low enough to allow efficient hydrogen bonding between poly(A) tracts on the mRNA molecules and Oligo(dT) attached to cellulose while still being high enough to maintain complete inhibition of RNases. The clear supernatant containing RNA was applied to an Oligo(dT)-Cellulose pellet in a microcentrifuge tube and gently mixed by inversion for 3 minutes to allow the poly(A)$^+$ RNA to bind to the cellulose matrix.

The Oligo(dT)-Cellulose was sedimented by centrifugation at 10,000x g for 10 seconds and the supernatant pipetted off. The pelleted material was washed a total of 5 times with 1 ml aliquots of high-salt buffer (10 mM Tris-HCl, pH 7.5; 1 mM EDTA; 0.5 M NaCl) followed by 2 washes with 1 ml aliquots of low-salt buffer (10 mM Tris-HCl, pH 7.5/ 1 mM EDTA/ 0.1 M NaCl). All washes were carried out at 10,000 x g for 10 seconds each. Washed pellets were resuspended in a small volume of low-salt buffer and the slurry applied to a MicroSpin column placed in a microcenrifuge tube. The column was washed three times with 0.5 ml aliquots of low-

salt buffer by centrifugation at 10,000 x g for 5 seconds each. Finally, the polyadenylated material was eluted with 0.2 ml of elution buffer pre-warmed at 65°C. RNA concentration in the final eluate was determined by spectrophotometry as described above.

Where required the RNA was concentrated by precipitation. This was achieved by the addition of 1/40 volume of glycogen (5-10 mg ml^{-1} in DEPC-treated water), 1/10 volume of potassium acetate (2.5 M, pH 5.0) and approximately 2 volumes of cold 95% ethanol. The mixture was incubated at -20°C for 30 minutes, pelleted by centrifugation at 4°C for 5 minutes and finally resuspended in the required amount of elution buffer.

2.5. ISOLATION OF REPEATED-DNA SEQUENCES

A nonradioactive method designed for rapid isolation of microsatellite sequences (Fischer and Bachmann, 1998; Oliveira *et al,* 1998) was adapted by us. The method is based on streptavidin-mediated affinity capture of single-stranded restriction fragments annealed to biotinylated microsatellite oligonucleotides followed by adapter-mediated PCR.

Genomic DNA (ca. 500 ng) was digested with blunt-end generating restriction enzymes, either *Alu*I or *Rsa*I (10U/ 20 µl reaction) (Helena Biosciences), for 2 hours followed by incubation at 65°C for 15 minutes to render the enzymes inactive. Samples were passed through S-200 columns to remove salts and other inhibitors.

A 5′-phosphorylated 24-mer (5′ pAGTCCGGATCCAAGCAAGAGCACA 3′) and a 20-mer (5′ CTCTTGCTTGGATCCGGACT 3′), with overlapping complementary sequences containing a *Bam*HI site, were used to generate an adapter molecule. 50 pmols of each oligonucleotide were mixed in a 20 µl reaction and annealed by cooling the mixture down from 70°C to 10°C over a period of not less than 1 hour. 2.5 pmols of this adapter molecule were then ligated to approximately 250 ng of digested DNA

in a 20 µl reaction. Ligation was carried out by incubating overnight in a 14°C water bath in the presence of T4 DNA ligase (Promega) at a final concentration of 3U μl^{-1}. Hybridisation of ligated constructs (equivalent to ca. 50 ng DNA) to biotinylated microsatellite oliginucleotides (20 pmols) (Table 2) was carried out in a 50 µl reaction in the presence of 1 X TE buffer (10 mM Tris-HCl, pH 8.0/ 1 mM EDTA) in a thermal cycler (Hybaid Omnigene) with denaturation at 95°C for 10 minutes and annealing at 60°C for 1 minute.

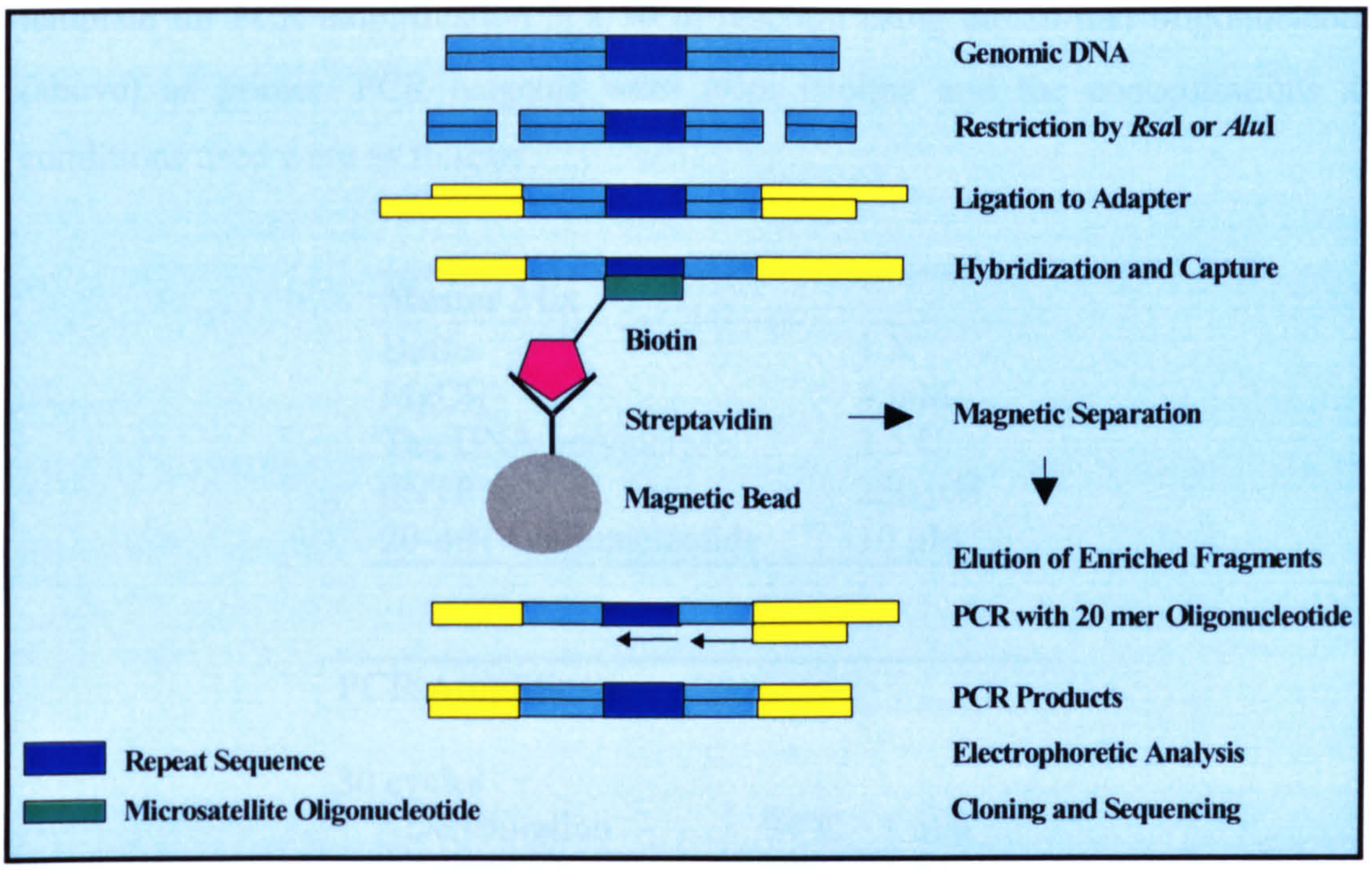


Fig. 3. Schematic Representation of Isolation of Repeated-DNA Sequences.

Table 2

BIOTINYLATED MICROSATELLITE OLIGONUCLEOTIDES

GATGATCCGACGCAT**(CA)12**
GATGATCCGACGCAT**(CT)12**
(CAA)12
(CTT)12
(CAT)12
(CTA)12
(TAA)12

These hybrids were subsequently bound to streptavidin coated magnetic beads (Dynbeads kilobaseBinder™ kit, Dynal) following the manufacturers instructions. Briefly, 100 µg of beads were washed and resuspended in the binding solution and mixed with the entire contents of the hybridisation mixture. This mixture was incubated at room temperature for 3 hours on a roller to keep the beads in suspension. Following incubation the Dynabeads-DNA complex was washed twice with the washing solution (10 mM Tris-HCl, pH 7.5/ 1 mM EDTA/ 2.0 M NaCl) and resuspended in 50 µl of TE buffer. 1/5 volume of this captured product was used as a template for PCR amplification in a 50 µl reaction using the 20-mer oligonucleotide (above) as primer. PCR reagents were from Bioline and the concentrations and conditions used were as follows;

Master Mix	
Buffer	1 X
$MgCl_2$	4 mM
Taq DNA polymerase	2.5 U
dNTP's	250 µM
20-mer Oligonucleotide	10 µM

PCR Amplification		
30 cycles		
Denaturation	94°C	1 min
Annealing	60°C	1 min
Extension	72°C	2 min
1 cycle		
Final extension	72°C	5 min

Amplification products were electrophoretically analysed (section 2.6). Ligated, but unselected, PCR amplified constructs of digested DNA and adapter (i.e. constructs that have not been subjected to hybridisation and capture) were used as controls.

Enriched PCR fragments were excised from the gels using a UV transilluminator and the DNA eluted by first passing the gel pieces through sterile, disposable 1 ml syringes (Becton Dickinson). The entire crushed contents were applied to empty Micro Bio-Spin® chromatography columns (Bio-Rad) which were centrifuged at

10,000 x g for 10 minutes. Eluted samples were analysed on agarose gels to check for purity of individual fragments.

2.6. ELECTROPHORETIC ANALYSES

Amplified products were analysed on 1.8% agarose (Gibco BRL or BioGene) gels in 1 X Tris-Borate-EDTA (TBE) buffer at 120 V, unless otherwise stated. The λ DNA/ *Hin*dIII and GeneRuler™ 100 bp DNA ladder (Helena Biosciences) or the 100 bp DNA ladder (Promega) were used as size markers. Gels were stained with ethidium bromide (1.0 μg ml^{-1}) and analysed under UV illumination.

2.7. CLONING

The pGEM® -T Easy Vector System (Promega) was used for cloning according to the manufacturers instructions (sections 3.3.1 & 3.3.4; section 5.3.2). Except for the eluates of enriched fragments (section 2.5) which were passed through S-200 columns before being cloned, all other PCR products in this study were cloned without further purification in ligation reactions which were incubated overnight at 4°C for maximum ligation efficiency.

Chemically treated, JM109 high efficiency (>10^8 cfu μg^{-1}) competent cells (Promega) were used for transformation. Cells were mixed with the ligation reactions and heat shocked for 45 seconds in a water bath at 42°C and then immediately chilled on ice for 2 minutes. Transformed cells were allowed to recover for 1 hour in SOC medium (2% tryptone, 2% yeast extract, 10 mM NaCl, 0.25 mM KCl, 20 mM Mg^{+2} and 20 mM glucose) at 37°C before plating onto 90 mm Luria-Bertani (LB) agar plates incorporating 100 μg ml^{-1} of ampicillin. The plates were spread with 100 μl of isopropyl β-D-thiogalactopyranoside (IPTG) (0.1 M) and 20 μl of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) (50 mg ml^{-1}) before application of transformed cells then incubated overnight at 37°C. IPTG and X-Gal serve as an indicator system for direct identification of recombinant clones by colour screening (blue-white screening). Successful cloning of an insert in the vector interrupts the coding

sequence of the enzyme ß-galactosidase resulting in white colonies. Experimental controls included those recommended by the manufacturers, namely: a positive control using the control insert DNA provided (4 ng μl^{-1}); a background control with no insert; and a transformation control with 0.1 ng of an uncut plasmid.

Alternatively (section 3.3.5; section 7.2.4), transformation was performed using Epicurian Coli® XL1-Blue competent cells (≥1.8 x 10^9 cfu μg^{-1}) (Stratagene) according to the manufacturer's instructions.

2.8. RAPID SIZING OF RECOMBINANT PLASMIDS

Rapid size identification of recombinant clones was carried out using a screening method described by Sekar (1987). Recombinant colonies (white colonies) from LB agar plates were inoculated into 200 µl of LB broth supplemented with 100 µg ml^{-1} of ampicillin and incubated at 37°C for 5 hours on a shaker. A small volume of LB cell suspension from each recombinant clone was mixed with glycerol to a final concentration of 25% and frozen at -20°C until required. The remaining cells were harvested by centrifugation and each sample 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^{-1} RNase A/ 50 µg ml^{-1} Lysozyme). 0.8% agarose gels were prepared in 1 X TBE with the addition of 0.05% SDS. Gel slots were preloaded with 20µl of Lysis buffer (1 X TBE, 2% SDS, 5% Sucrose and 0.04% bromophenol blue). The protoplast suspension was then loaded into the wells underneath the lysis solution. Electrophoresis was carried out in 1 X TBE initially at 30 V for 15 minutes and then at 120 V for an additional 2 hours. Gels were stained and photographed as before (section 2.6). Plasmid DNA in a size range similar to that of the expected product was used as size marker.

Where indicated (by results of the Sekar Analysis), recombinant cells from individual clones stored in glycerol were used to obtain plasmid DNA.

2.9. ISOLATION OF RECOMBINANT PLASMID DNA

The QIAprep Spin Miniprep Kit (Qiagen) was used for this purpose. Recombinant bacteria were inoculated into 5 ml of LB broth supplemented with 100 μg ml^{-1} ampicllin and incubated overnight at 37°C on a shaker. Cells were sedimented by centrifugation at 800 x g for 5 minutes. Cell pellets were resuspended in a buffer containing RNase A and lysed under alkaline conditions in a buffer containing NaOH and SDS, which solubilizes the cell membranes and denatures the chromosomal and plasmid DNAs and proteins. The lysate was then neutralised and adjusted to high-salt binding conditions to allow denatured proteins, chromosomal DNA, cellular debris and SDS to precipitate. Clear supernatant containing the plasmid DNA was then applied to the QIAprep columns (silica-gel membrane) which under the existing high salt conditions allows only plasmid DNA to be absorbed, while RNA, cellular proteins and metabolites are not retained. Columns were then subjected to a two-step wash procedure in order to remove endonucleases and salts. Finally, DNA was eluted in a low salt buffer (10 mM Tris-Cl, pH 8.0).

Isolated plasmid DNA samples were digested with *Eco*RI (10 U/ reaction) (Helena Biosciences) for two hours and analysed on 1% agarose gels (section 2.6).

2.10. SEQUENCING

The Thermo Sequenase II dye terminator cycle sequencing kit (Amersham Pharmacia Biotech) was used as per the manufacturer's instructions (sections 3.3.1 & 3.3.4; section 5.3.2). The primers used were T7 and SP6 (Promega) (Table 3) in a final concentration of 5 pmols/ reaction. Template DNA amounts were between 200-500 ng.

Table 3

SEQUENCING PRIMERS

Primer Name	Primer Sequence (5′ → 3′)
T7	CGG GAT ATC ACT CAG CAT AAT G
SP6	TAT TTA GGT GAC ACT ATA G

Thermal cycling conditions

30 cycles

96°C	30 sec
50°C	15 sec
60°C	1 min

The cycle sequencing product was then passed through an S-200 column to remove excess dye-labelled dideoxynucleotide terminators. Alternatively the precipitation protocol outlined in the Amersham kit was employed with a few modifications. Briefly, 20 µl of the cycle sequencing product was mixed with 7 µl of ammonium acetate (7.5 M) followed by the addition of 2.5 volumes of absolute ethanol (-20°C). The mixture was chilled on ice for a minimum of 15 minutes and then centrifuged at 10,000 x g for an additional 15 minutes. The supernatant was drawn off and the precipitate washed with 70% ethanol at 10,000 x g for 2 minutes. After drawing off the supernatant the precipitate was air or vacuum dried.

Alternatively, the ABI Prism® BigDye™ terminator cycle sequencing ready reaction kit was used according to the manufacturers instructions (section 3.3.5; sections 7.2.3 & 7.2.4). Template DNA concentrations were as stated above, but primers T7 and SP6 (Promega) (Table 3) were used in a final concentration of 3.2 pmols/ reaction.

Thermal cycling conditions

30 cycles

96°C	30 sec
50°C	15 sec
60°C	4 min

Cycle sequencing products were precipitated using the ethanol and ammonium acetate method described above.

Reactions were analysed on a DNA sequencer ABI PRISM™ 377 (Perkin Elmer) (sections 3.3.1 & 3.3.4; section 7.2.4). Alternatively plasmid DNAs were sent to MWG Biotech for sequencing (section 3.3.5; section 5.3.2; section 7.2.3).

The resultant sequences were assembled by one of three means, namely: the Autoassembler™ DNA Sequence Assembly software by Perkin Elmer; the Multalin program (Corpet, 1988) (http://www.prodes.toulouse.inra.fr/multalin/multalin.html); or by eye.

The sequences were compared to previously published ones using the Blast 2.0 programme (Altschul *et al*, 1990) of the National Centre for Biotechnology Information (NCBI) databases (http://www.ncbi.nlm.nih.gov/BLAST/).

2.11. PCR FRAGMENT SIZE POLYMORPHISM

Primers were designed from sequences flanking the repetitive domains selected for analysis. Table 4 lists the primers used for respective repeat loci.

Table 4

POLYMORPHIC LOCUS - SPECIFIC OLIGONUCLEOTIDE PRIMERS

Repeat Loci	Primer Name	Primer Sequence (5′ → 3′)
1-2	EhR1 (5′ primer)	CTG GTT AGT ATC TTC GCC TGT
	EhR2 (3′ primer)	CTT ACA CCC CCA TTA ACA AT
3-4	EhR3 (5′ primer)	GCT ATG GTC GGT ATC GAT ATC
	EhR4 (3′ primer)	CCT TAG GTC ACT GGT TCG AA
	EhR7 (5′ primer)	CTT TAC TTC TCT TTT ACC ACG
	EhR8 (3′ primer)	CGT GGT AAA AGA GAA GTA AAG
5-6	EhR5 (5′ primer)	CTA AAG CCC CCT TCT TCT AT
	EhR6 (3′ primer)	GTG CTA ATA ACG CCA GGG TC
	EhR5A (5′ primer)	CTA AAG CCC CCT TCT TCT ATA ATT
	EhR6A (3′ primer)	CTC AGT CGG TAG AGC ATG GT
9-4	EhR9 (5′ primer)	CTA CAT CTA CAG TCC TCC GCT
	EhR4 (3′ primer)	CCT TAG GTC ACT GGT TCG AA
	EhR10 (5′ primer)	CTT ACT TCT CTT TAC CAC GAC
	EhR11 (3′ primer)	GTC GTG GTA AAG AGA AGT AAG
16-17	EhR16 (5′ primer)	AAG CTT CCT TAG CTC AGC TG
	EhR17 (3′ primer)	TAA AAG GGG GAA GAA TAG GAA
	EhR18 (5′ primer)	GGT TTC ATG GTG TAG TTG GT
	EhR19 (3′ primer)	ACC AAC TAC ACC ATG AAA CC

Note: At locus 5-6 two primer pair sequences are listed (see legend to Fig. 5).

Genomic DNA of *Entamoeba* strains was used as templates for PCR amplification in 50 µl reactions. Except where noted (section 4.3.1), PCR reagents were from Bioline and the concentrations and conditions used were as follows:

Master Mix	
Buffer	1 X
$MgCl_2$	1.5 mM
Taq DNA polymerase	2.5 U
dNTP's	1.0 µM
Primer	2.0 µM

PCR Amplification

30 cycles

Denaturation	94°C	1 min
Annealing	Primer dependant (50°C - 56°C)	1 min 30 sec
Extension	72°C	2 min

1 cycle

Final extension	72°C	5 min

Amplified products were analysed as described (this chapter section 2.6).

2.12. PROBE LABELLING

PCR amplified products were generated from plasmids or genomic DNA. Prior to labelling, the PCR products were purified using the QIAquick PCR purification kit (Qiagen).

In most cases the fragments were labelled with [α-^{32}P]-dCTP using the Rediprime II random primer labelling system (Amersham Pharmacia Biotech). Briefly, the amplified DNA was denatured by boiling, snap cooled on ice and applied to the rediprime tube containing the complete labelling reaction to which the radioactive nucleotide was then added. Following a 10 minute incubation at 37°C, the labelled probe was passed through a S-200 column to remove unincorporated nucleotides and used for hybridisation.

In some instances however, the Prime-a-Gene® Labelling System (Promega) was used according to the manufacturers instructions (section 7.2.5 (Fig. 42; Table 23)). The denatured (boiled) template DNA was mixed in the first instance with nuclease free water, labelling buffer containing random primers, a mixture of unlabelled dNTPs and nuclease free BSA (bovine serum albumin) in that order. This was followed by addition of [α-^{32}P]-dCTP and finally the DNA Polymerase I, large (Klenow) fragment, was added and the whole incubated at room temperature for 1 hour. The

reaction was terminated by boiling and the addition of EDTA to 20 mM. Unincorporated nucleotides were removed as above.

2.13. NORTHERN HYBRIDISATION ANALYSIS

The NorthernMax™ kit (Ambion) was used for this purpose and is based on the use of formaldehyde for denaturing the RNA. 1% agarose gels were prepared in a formaldehyde containing buffer. Samples of polyadenylated RNA (2 µg) and total RNA (5.4 µg and 2.15 µg) as well as the RNA size marker (Promega) were prepared by mixing with 3 volumes of formaldehyde-containing loading buffer and incubating at 65°C for 15 minutes to denature any RNA secondary structures. Electrophoresis was carried out at 5V/ cm. Gels were stained and photographed as before (section 2.6). A downward capillary transfer system was set up using Biodyne®A membranes (Gibco BRL) membranes and transfer was allowed to continue for 2 hours. RNA was crosslinked to the membrane using Stratalinker™ UV Crosslinker (Stratagene) (120,000 µjoules).

The membranes were pre-hybridised in a formamide containing solution at 42°C for 30 minutes in a roller oven and hybridised with boiled labelled DNA probes (section 2.12) by incubating overnight at 42°C. Membranes were washed twice with a low stringency wash solution (equivalent to 2 X SSC) followed by two washes with a high stringency wash solution (equivalent to 0.1 X SSC) at 42°C, wrapped in cling film and exposed to X-ray film for autoradiography at -70°C.

2.14. SOUTHERN HYBRIDISATION ANALYSIS

Genomic DNA of strains HM-1:IMSS clone 9 was subjected to overnight digestion at 37°C with 10 U of each restriction enzymes (Gibco BRL or Helena Biosciences) in 20

µl reactions. Digested products were electrophoresed in 0.8% agarose gels (GibcoBRL) in 1 X TBE at 120 V and stained and photographed as before (section 2.6). The λ DNA/ *Hin*dIII and GeneRuler™ 100 bp DNA ladder (Helena Biosciences) were used as size markers. Prior to setting up capillary transfer the DNA was depurinated in a 0.25 M HCl solution for 10 minutes. This increases the efficiency of transfer of larger DNA fragments to the membrane. This was followed by denaturation in a 1.5 M NaCl/ 0.5 M NaOH solution for 30 minutes and finally the DNA was neutralised in a 1.5 M NaCl/ 1 M Tris-HCl, pH 7.5 solution for 30 minutes.

An upward capillary transfer system was set up using Biodyne®A membranes (Gibco BRL) and transfer was allowed to proceed overnight. The filter was blotted dry on a Whatman 3MM filter paper and DNA was crosslinked to the membrane using Stratalinker™ UV Crosslinker (Stratagene) (120,000 µjoules). The membranes were transferred to a HYBAID™ tube and prehybridized in 9 ml of 1 M NaCl, 1% SDS and 10% dextran sulphate, at 65°C for 2 - 4 hours in a roller oven. Hybridisation was carried out with boiled (denatured), labelled double stranded DNA probes (section 2.12) by incubating overnight at 65°C in a roller oven. Membranes were washed successively for 30 minutes each at 65°C with 2 X SSC/ 1% SDS, 0.5 X SSC/ 0.25% SDS and 0.2 X SSC/ 0.1% SDS, wrapped in cling film and exposed to X-ray film for autoradiography at -70°C.

2.15. PARTIAL RESTRICTION DIGESTS OF GENOMIC DNA

Genomic DNA of strain *E. histolytica* HM-1:IMSS clone 9 was digested with either *Alu*I or *Dde*I (Promega). To achieve partial digestion, enzymes were used at a final concentration of 0.1U/ 20 µl of reaction mixture. The tubes were incubated at 37°C and 5 µl each of the reaction mix was removed at intervals of 15, 30 and 60 minutes, with the last 5 µl being incubated for 2 hours. At each stage the enzymes were rendered inactive by addition of 0.5 µl of 0.1 M EDTA. Subsequently, reactions from all four time-points were pooled into one tube. Controls were generated by setting up

additional enzyme reactions with 10U of enzyme/ 20 µl reaction mix followed by overnight incubation at 37°C to achieve complete digestion.

Digested products were analysed and Southern hybridisation analysis was carried out as just described (section 2.14).

2.16. MINI LAMBDA DNA PREPARATION & AMPLIFICATION

A genomic lambda ZAP library derived from *E. histolytica* strain HM-1:IMSS was provided to us by Dr E. Tannich, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. Lambda phages were propagated on plates and purification of lambda DNA was carried out from phage lysates using the Lambda Mini Kit (Qiagen) as per the manufacturers instructions.

Briefly, the phage lysate was first adjusted to high salt conditions with buffers containing RNase A and DNase I, which digest bacterial RNA and chromosomal DNA respectively, followed by addition of polyethylene glycol (PEG) containing buffer to precipitate the phage particles. After centrifugation and removal of the supernatant the pellet was resuspended in a low salt buffer before being treated with buffers containing SDS, which denatures phage proteins and releases DNA, and potassium acetate, which helps remove the denatured proteins by formation of protein and potassium dodecyl sulfate complexes. Following centrifugation, the clear supernatant containing lambda DNA was applied to a pre-equilibrated Qiagen column. The column contains an anion-exchange resin, which under appropriate low salt and pH conditions allows lambda DNA to bind while the protein and carbohydrate impurities are removed with a medium salt wash buffer. Pure lambda DNA was then eluted in high salt buffer and subsequently concentrated and desalted by isopropanol precipitation. After a final wash with 70% ethanol to remove residual salt and replace isopropanol, the pellet was air dried and the DNA dissolved in 50 µl of TE buffer (10 mM Tris-Cl, pH 8.0/ 1 mM EDTA, pH 8.0).

DNA yield and quality was analysed by agarose gel electrophoresis on a 1% agarose gel at 120 V. Lambda DNA was used as template for PCR amplification with locus 1-2 (R1 & R2) and 5-6 (R5A & R6A) specific primers (Table 4) using the conditions described (section 2.11).

2.17. GENOMIC LIBRARY SCREENING

The genomic *E. histolytica* lambda ZAP library was screened by colony hybridisation according to the instructions (Stratagene). Duplicate replica filters representing ca. 60,000 (10,000/ plate) recombinants were denaturated in a 1.5 M NaCl/ 0.5 M NaOH solution for 2 minutes followed by neutralisation in a 1.5 M NaCl/ 0.5 M Tris-HCl, pH 8.0 solution for 5 minutes and finally rinsed in 0.2 M Tris-HCl, pH 7.5/ 2 X SSC for 30 seconds. The filters were blotted dry on Whatman 3MM filter paper and DNA was crosslinked to the membrane using Stratalinker™ UV Crosslinker (Stratagene) (120,000 μjoules) before being transferred to HYBAID™ tubes for prehybridization and hybridisation as described for Southern blotting (section 2.14).

The most promising clones were chosen for a second round of screening with ≤ 1000 recombinants/ plate. Hybridising plaques were isolated and phagemids released according to the instructions (Stratagene).

The excised phagemids were mixed with the *Escherichia coli* XLOLR host strain before being plated onto 90 mm Luria-Bertani (LB) agar plates containing 100 μg ml^{-1} of ampicillin and incubation overnight at 37°C. Individual colonies were selected, inoculated into 5 ml of LB broth supplemented with 100 μg ml^{-1} of ampicillin and recombinant plasmid DNA was isolated using the QIAprep Spin Miniprep Kit (Qiagen) kit (section 2.8). Isolated plasmid DNA samples were digested with EcoRI (10 U/ reaction) (Promega) for two hours and analysed on 1% agarose gels to select clones for sequencing.

2.18. PULSED - FIELD GEL ELECTROPHORESIS

For pulsed-field gel electrophoresis (PFGE) DNA was prepared in agarose blocks essentially as described by Clark *et al* (1990). Culture tubes containing 72-hour old cultures of *E. histolytica* strain HM-1:IMSS clone 9 were chilled, sedimented and washed with sterile 1 X PBS. The washed and harvested cell pellets (ca. 2 x 10^7 cells), were resuspended in a small volume (≤0.25 ml) of 1 X PBS and mixed with an equal volume of 2% low melting point (LMP) agarose (Gibco BRL) at 42°C. After gentle mixing, the slurry was rapidly pipetted into disposable CHEF plug molds (Bio-Rad) and cooled on ice for 10-15 minutes. The agarose blocks (4 x 10^6 cells each), were then placed in lysis buffer (1% sodium lauryl sarcosine/ 0.5 M EDTA, pH 9.5 with 1 mg/ ml proteinase K) at 50°C for 2 days and thereafter stored at 4°C until required.

PFGE was performed using a CHEF Mapper® XA system (Bio-Rad). One half of an agarose block was loaded per lane (2 x 10^6 cells). Electrophoresis was carried out in 1% agarose gels (BioGene), with 0.25 X or 0.5 X TBE running buffer at 14°C. Electrophoretic conditions included a separation range of 50 kb to 1.0 Mb, a 6V/ cm gradient, an angle of 120°, initial switch time of 6.75s, final switch time of 1m33.69s and a running time of 26.40 hrs. Size standards of *Saccharomyces cerevisiae* chromosomal DNA and λ DNA concatemers were obtained from Bio-Rad. Following electrophoresis, gels were stained with ethidium bromide (1.0 μg ml^{-1}) for 45 minutes followed by 30 minutes of de-staining in water before being analysed under UV illumination. Southern hybridisation analysis was carried out as already described (section 2.14).

2.19. RESTRICTION DIGESTION OF DNA IN AGAROSE

Agarose blocks were prepared as above. Following a 2 day lysis buffer treatment blocks were washed for 2 hours in 1 X TE (1 mM TrisCl, pH 7.5-8.0/ 1 mM EDTA,) with 1 mM phenyl methyl sulfonyl fluoride (PMSF) to inactivate the proteinase K, followed by three washes in 1 X TE for 2 hours each to remove the PMSF. All washes were done at room temperature and on a rocker. To remove excess carbohydrate, which could potentially reduce enzyme activity, the blocks were loaded in a 0.8% agarose gel (BioGene) and subjected to electrophoresis at 30 V for 15 minutes followed by an hour at 80 V. One half block was loaded per lane. Under these conditions the bulk of the carbohydrate remains in the wells along with some DNA, with majority of the DNA having migrated 1-1.5 cm into the gel as seen after ethidium bromide staining. To avoid possible interference of ethidium bromide in subsequent enzymatic reactions only the edge of the gel lane was stained. Realignment of the stained portion with the rest of the gel allowed the location of the unstained DNA to be identified, which was then excised in a small block of agarose.

The excised blocks were subjected to three 2 hour washes in 1 X TE to remove the TBE buffer. The blocks were then ready to be used for digestion. Before digestion blocks were equilibrated in the appropriate enzyme buffer for about 2 hours at 4°C. Each half block was weighed to obtain an approximate volume (weight in mg = volume in µl) and digested in a total volume of 0.2 ml of restriction enzyme buffer with BSA added to 0.25 mg ml^{-1} and 100U of *Eco*RI. The digest was allowed to proceed overnight at 37°C, followed by another overnight digestion with fresh enzyme mixture. EDTA, 10 mM final concentration, was used to stop the reaction.

PFGE was performed as in section 2.18 in a 0.25 X TBE running buffer with a separation range of 20-350 kb, an initial switch time of 2.98s, final switch time of 30.82s and a running time of 26.56 hrs. Gels were stained, photographed and subjected to Southern hybridisation analysis as described (section 2.14).

CHAPTER 3

ISOLATION & CHARACTERISATION OF POLYMORPHIC DNA FROM *E. histolytica*

3.1. INTRODUCTION

An important gap in our understanding of the epidemiology of amoebiasis is what determines the outcome of *E. histolytica* infections. Are the organisms that produce invasive, symptomatic disease genetically distinct from those that give rise to asymptomatic infections? Or do all *E. histolytica* isolates have the potential to become invasive? Do certain invasive isolates show tropism for specific organs, with some preferentially ending up in the intestinal wall while others reach extra-intestinal sites? To address the possibility of a relationship between parasite variation and infection outcome the ability to differentiate isolates of *E. histolytica* is necessary.

Our present knowledge of intra-species variation in *E. histolytica* is limited (reviewed in sections 1.4.5 & 1.5.2). At present, the most polymorphic genes of *E. histolytica* are two protein coding genes displaying internal tandem repeat structures (Clark and Diamond, 1993a; Ghosh *et al*, 2000). There is, however, a need for additional reliable polymorphic *E. histolytica* DNA markers.

In recent years the use of microsatellite locus analysis has gained considerable popularity as a tool for detecting intra- and inter-species variations in a number of organisms, including protozoan parasites such as *Trypanosoma* (Oliveira *et al*, 1998), *Leishmania* (Russel *et al*, 1999) and *Plasmodium* (Anderson *et al*, 1999). Using a method designed to isolate microsatellite loci we have obtained two new polymorphic DNAs containing tandemly repeated sequences from *E. histolytica*. Preliminary

characterisation of the two loci and the inter-strain variations they display is the focus of discussion in this chapter. In addition three other loci showing the presence of tandemly repeated sequences were studied for their potential as polymorphic markers for use in investigating the molecular epidemiology of *E. histolytica* and the results are also presented here.

3.2. *E. histolytica* ISOLATES

Except for strain HM-1:IMSS clone 9, the axenic isolates were provided by Dr. John Ackers (London School of Hygiene and Tropical Medicine) (Table 5). All axenic isolates were cultured in medium YI-S (sections 2.2 & Appendix I).

Xenic isolates were obtained from two sources (Table 5). Four samples were from Dr. Rashidul Haque of the International Centre for Diarrhoeal Disease Research, Bangladesh, via Dr. Aura Aguirre (London School of Hygiene and Tropical Medicine), while four others were provided by Dr. Terry Jackson of the Medical Research Council of South Africa, Durban. The South African isolates were from a patient who had recovered from amoebic liver abscess (37.0C) or close family contacts of such patients (i.e. individuals at high risk of acquiring amoebic infection) who were asymptomatic at the time. All the South African xenic strains were originally isolated in Robinson's medium (cited in Clark and Diamond (2002)) and were sent to us as lysates in lysis buffer (section 2.3). DNA was isolated as already described (section 2.3). Purified DNA of all four Bangladeshi isolates (extracted using phenol chloroform from cells cultured in Robinson medium) was provided by Dr. A. Aguirre.

Table 5

ORIGIN OF *E. histolytica* ISOLATES

Isolates	Origin	Culture Type	Clinical Information †
HM-1:IMSS clone 9	Mexico	Axenic	A
200:NIH	(uncertain)	Axenic	B
H-303:NIH	VietNam (?)	Axenic	C
IULA:0593:2	Venezuela	Axenic	D
IULA:1092:1	Venezuela	Axenic	D
8691	Bangladesh	Xenic	D
4530	Bangladesh	Xenic	D
1320300	Bangladesh	Xenic	D
48286	Bangladesh	Xenic	D
2596	South Africa	Xenic	E
26.253C	South Africa	Xenic	E
37.0C	South Africa	Xenic	F
39.384C	South Africa	Xenic	E

Note: "†" A: Isolated from rectal ulcer of patient with dysentery; **B**: Patient with amoebic dysentery; **C**: Patient with amoebic empyema and dysentery; **D**: Patient with amebic colitis; **E**: Patient asymptomatic; serology positive by antigen gel diffusion test; **F**: Convalescent amebic liver abscess patient; serology positive by antigen gel diffusion test.

3.3. RESULTS

3.3.1. ISOLATION OF REPEATED-DNA CONTAINING SEQUENCES FROM *E. histolytica* GENOME

To try and obtain DNA fragments containing microsatellites we employed a non-radioactive method based on affinity capture of single stranded restriction fragments annealed to biotinylated microsatellite oligonucleotides, with attachment to streptavidin coated magnetic beads (Dynal) followed by adapter mediated PCR (Fischer and Bachmann, 1998; Oliveira *et al*, 1998) (section 2.5). A total of twenty-two PCR fragments ranging in size from 250 bp to 700 bp were gel purified from the total amplification products of *Alu*I or *Rsa*I restriction fragments annealed to one of seven biotinylated oligonucleotides (Table 2). These fragments were chosen on the basis of their apparent enrichment when compared to control amplification products

and were cloned, sequenced (sections 2.7 & 2.10) and examined for the presence of microsatellites.

None of the products contained sequences corresponding to the microsatellite oligonucleotides used in their capture. Furthermore, the majority of the sequences did not reveal any tandemly repeated-DNAs. However, two clones (clone 1 and clone 4) derived from an approximately 480 bp fragment obtained from *Alu*I restriction fragments annealed to the $(CTT)_{12}$ oligonucleotide showed the presence of internal tandem repeats. The repeats seen in clones 1 and 4 were distinct. Two other clones (clone 1' and clone 5) derived from *Alu*I restriction fragments annealed to $(TAA)_{12}$ contained the same type of repeats as clone 4. Clones 1' and 5 contained fragments of approximately 480 bp and 450 bp respectively.

The fact that only four of the twenty-two (ca. 18%) enriched fragments isolated by us contained any repeats and the presence of identical sequences in three of them suggested that the enrichment process was inefficient. In the hope of maximising enrichment and increasing the probability of isolating the expected microsatellites a variety of stringent wash conditions (other than those recommended by Dynal, section 2.5) were tried. Modifications included the use of 6 X SSC buffer (150 mM NaCl/ 15 mM sodium citrate) at room temperature; 3X SSC with 0.5% SDS at 60°C; 2 X SSC with 0.1% SDS followed by 1X SSC at a variety of temperatures including room temperature, 37°C, 42°C, 45°C, 50°C, 55°C and 60°C and finally 0.5X SSC at 55°C and 60°C. We however failed to observe any improvement over the results we had already obtained.

Further analysis was carried out on clone 1, which represents Locus 5-6, and clone 4, which represents Locus 1-2.

3.3.2. CHARACTERISATION OF LOCI 1-2 & 5-6

The complete sequence of the locus 1-2 clone (Fig. 4A), not including the adapter sequence, is 402 bp long and contains a single repeat block with two related direct repeats arranged in tandem (Fig. 4B). In addition to the major repeat block, tandem duplications of 8 to 12 bp are also present in the flanking regions (not shown).

A

EhR1 →

```
1    CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG   50
51   TTGAAGAGTT CTCTTTTTAT ACTTTTATAT GTTTATATGT TTATATCCTT  100
101  ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA  150
151  TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTATCT TTATTATCTT  200
201  TATTATCTTT ATTATCTTTA TTATCTTTAT TACCTTTATT ACCTTTATTA  250
251  CCTTTATTAC CTTTATTATC TTTATTACCT TTATTATATC TATTCTCACT  300
301  TCCTATACGT ACTCTTTTTA CTACTCTTTT TACTACTCTT CTTACTATAC  350
351  CTCTTACTAC TCCTACTTTC ACCTCCCTCT TTATTGTTAA TGGGGGTGTA  400
401  AG                                                     450
```

← EhR2

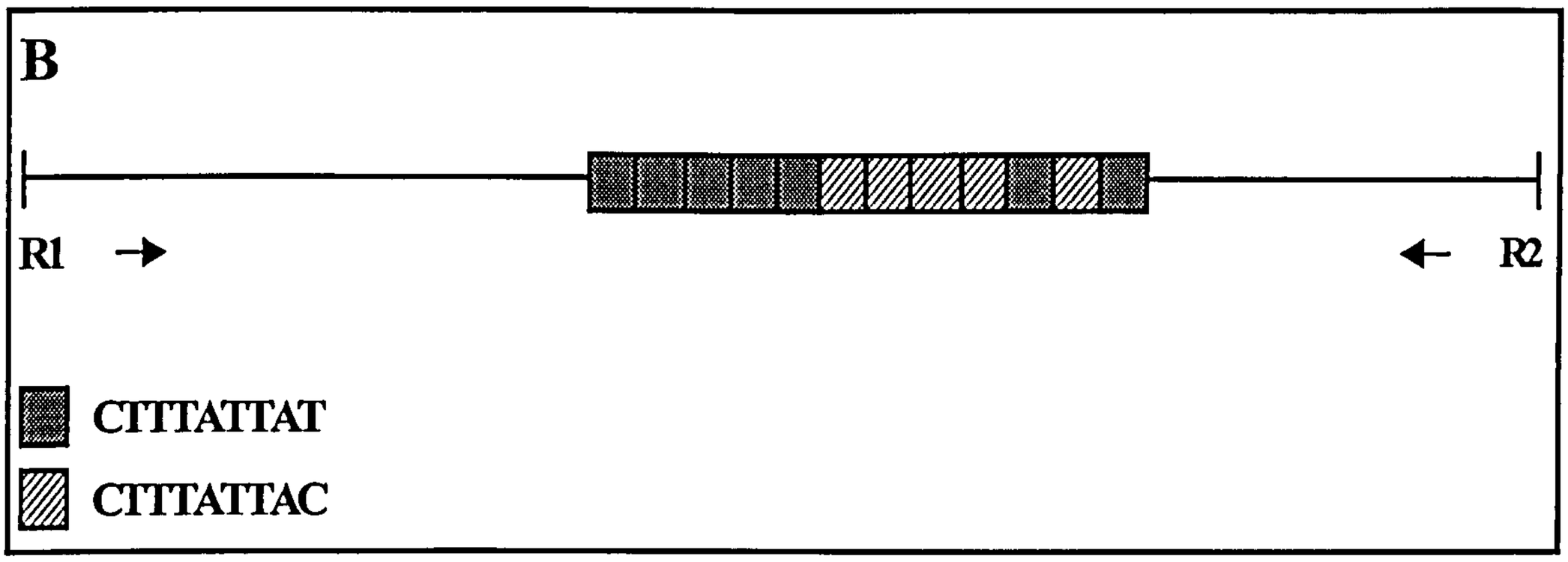


Fig. 4. Locus 1-2 from Strain HM-1:IMSS Clone 9. (A) Nucleotide Sequence. The main block of internal tandem repeats is in bold-face. Underlined regions indicate one of the two types of repeat units. Positions of amplification primers are highlighted. The nucleotide sequence data has been submitted to the GenBank database with accession number AF276055. **(B) Schematic Representation.** The two types of internal tandem repeats and their arrangement with respect to each other are shown. Tandem duplications in the flanking regions are not shown. Positions of amplification primers are indicated.

The complete sequence of the locus 5-6 clone (Fig. 5A), not including the adapter sequence, is 424 bp long and contains a single repeat block (Fig. 5B). As in locus 1-2, other tandem duplications in the regions flanking the repeat block are also evident (not shown).

A

EhR5/5A →

```
1     CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT   50
51    TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC  100
101   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA  150
151   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA  200
201   CTATGTATAT ATATATGTAT GTTTCTATGT ATGTTTCTAT GTATGTATAT  250
251   TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT ATGTATATTT  300
301   CTATGTATGT ATATTTCTAT GTATGTATAT TTCTATGTAC GTCTTTAACT  350
351   TTAAAAACAA TGCTCTCAGC AGGTTTCGAA CCCTGCGACC CTGGCGTTAT  400
401   TAGCACCATG CTCTACCGAC TGAG                              450
```

← EhR6/6A

B

R5 →
R5A →
←R6
←R6A

GTATGTTTCTAT

GTATGTATATTTCTAT

Fig. 5. Locus 5-6 from Strain HM-1:IMSS Clone 9. (A) Nucleotide Sequence. The main block of internal tandem repeats is in bold-face. Underlined regions indicate one of the two types of repeat units. Positions of amplification primers are highlighted. The nucleotide sequence data has been submitted to the GenBank database with accession number AF276060. **(B) Schematic Representation.** The two types of internal tandem repeats and their arrangement with respect to each other is shown. Tandem duplications in the flanking regions are not shown. Positions of amplification primers are indicated. Two primer pairs were designed for locus 5-6 (Table 4). Amplification products generated by primers R5 & R6 were cloned, sequenced and aligned for inter-strain nucleotide sequence comparisons (Fig. 7B; Appendix IV), while the primer pair R5A & R6A was used for studying inter-strain PCR product size polymorphisms (Fig. 6B).

Characteristic features of the repeat loci are summarised in Appendix II.

BLAST search results revealed no identity of either locus 1-2 or locus 5-6 to any previously reported *E. histolytica* sequences.

3.3.3. PCR PRODUCT SIZE POLYMORPHISMS AT LOCI 1-2 & 5-6

Primers were designed in the regions flanking the repeat blocks for both locus 1-2 and locus 5-6 and the PCR amplification products were analysed on 1.8% agarose gels to look for fragment size polymorphism among the thirteen *E. histolytica* isolates (Table 5).

Both locus-specific primer pairs were first tested with the five axenic strains. The primer pair originally designed for locus 5-6 (R5 & R6; Table 4) produced amplification products with all five axenic strains but we were unable to amplify any of the xenic *E. histolytica* isolates. Modifications of the annealing temperature or the amount of DNA used failed. New primers were designed (R5A & R6A; Table 4) and amplification was successfully achieved for all thirteen samples (Fig. 6B).

It is noteworthy that successful amplification depended greatly on the amount of DNA used in the PCR amplification reaction. Although an exact measure of the DNA concentration in each sample was not made and would have been especially difficult to determine for the xenic samples as the lysates also contained bacterial DNA, as a rule, 1µl of the extracted DNA was used directly in all PCR reactions. In some cases, however, the probability of obtaining amplification products was greatly improved by first diluting the DNA (1/ 10). It was observed that the optimal amount of DNA varied among loci even for the same strain (data not shown).

Amplification of locus 1-2 gave the expected product of about 400 bp in isolate HM-1:IMSS clone 9 (Fig. 6A). All the *E. histolytica* isolates gave a single major product but slight differences in size are evident among them. The four South African isolates gave the most variable patterns (right four lanes).

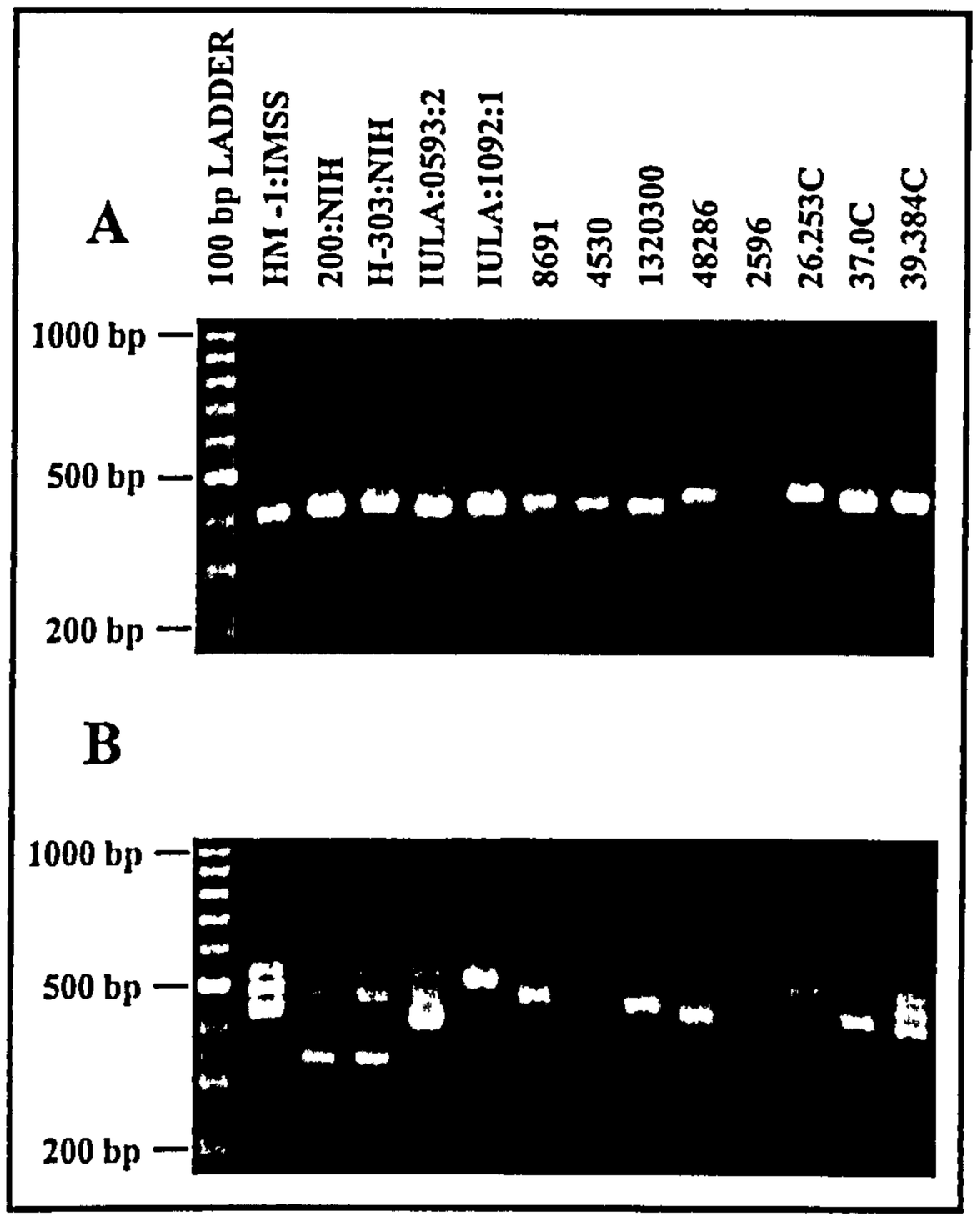
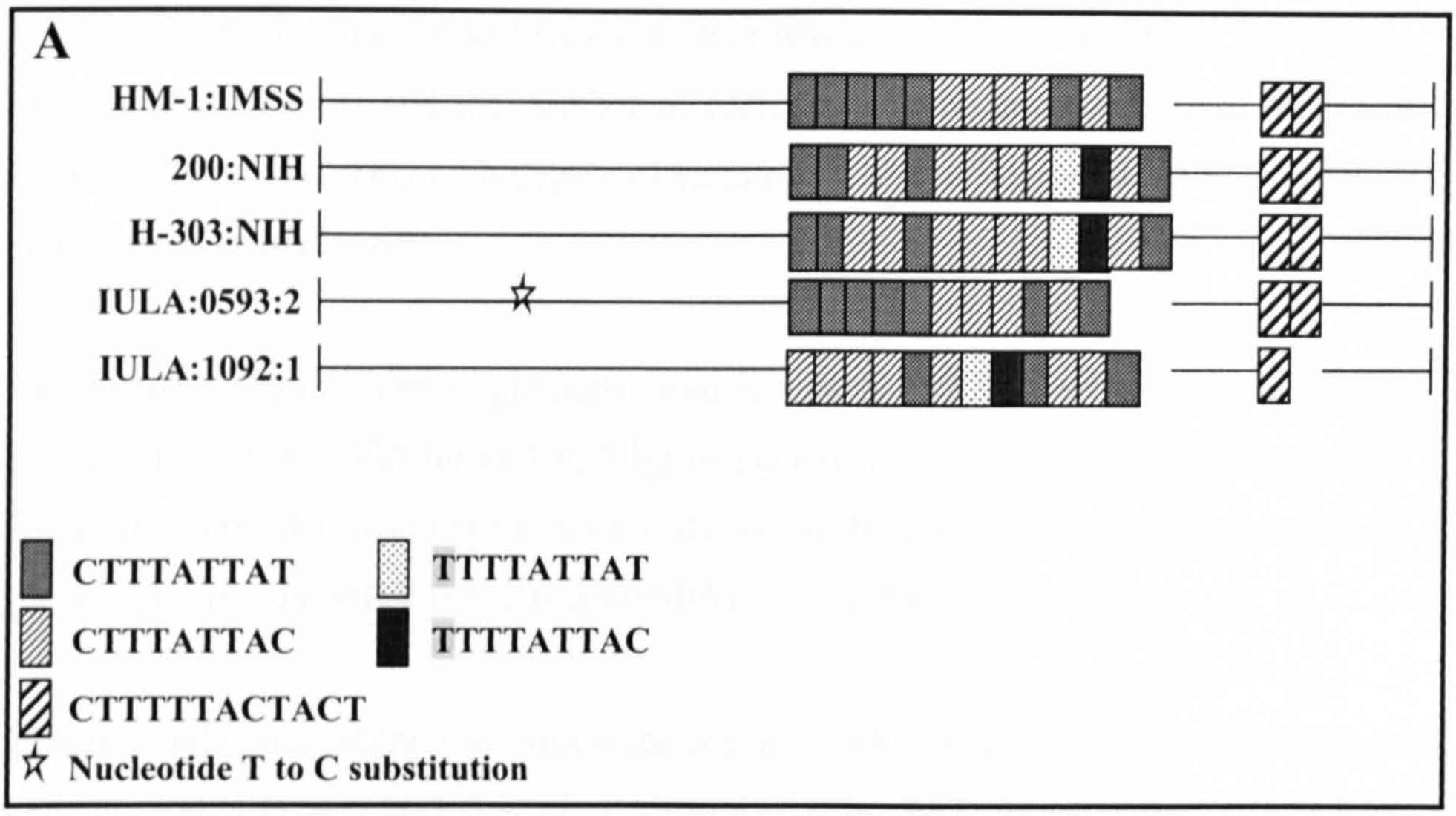


Fig. 6. Polymorphic DNA Analysis of *E. histolytica* Isolates at Loci 1-2 & 5-6. (A) Locus 1-2. Amplification products were generated using primers R1 & R2 at an annealing temperature of 53°C. **(B) Locus 5-6.** Amplification products were generated using primers R5A & R6A at an annealing temperature of 56°C.

Amplification of locus 5-6 gave the expected product of about 420 bp in isolate HM-1:IMSS clone 9 (Fig. 6B), but two additional bands of ca. 480 and 520 bp were also seen. This locus is highly polymorphic. Variation is seen in the total number of bands per isolate and their sizes.

3.3.4. NUCLEOTIDE SEQUENCE ANALYSIS & CHARACTERISATION OF THE OBSERVED SIZE POLYMORPHISM

In order to study the underlying nature of the observed size polymorphisms, the amplification products of all five axenic isolates at locus 1-2 and locus 5-6 were cloned and sequenced as previously described (sections 2.7 & 2.10). This analysis revealed differences in number and sequence of the repeat units as well as sequence variation in the regions flanking the repeat blocks. Complete multiple sequence alignment at locus 1-2 is shown in Appendix III and that for locus 5-6 is shown in Appendix IV. Schematic representations of these results are shown in Fig. 7A & 7B.



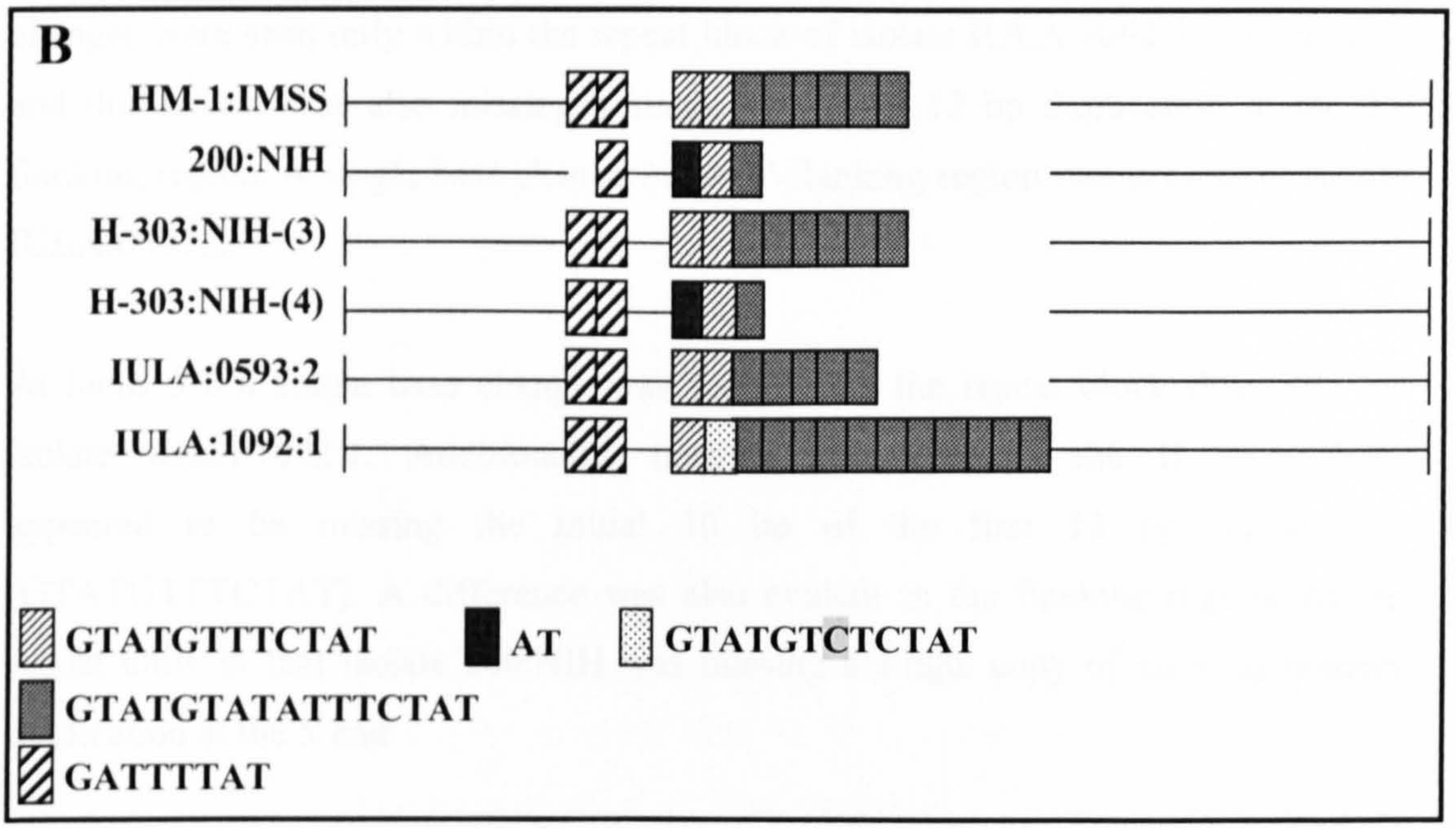


Fig 7. Schematic Representation of Locus Structure in Five Axenic Isolates of *E. histolytica*. (A) Locus 1-2. The nucleotide sequence data has been submitted to the GenBank database with accession numbers AF276055 to AF276059. **(B) Locus 5-6.** The nucleotide sequence data has been submitted to the GenBank database with accession numbers AF276060 to AF276065. Variations in number, sequence and arrangement of repeat units are shown. Gaps have been introduced to optimise alignment.

There was very little variation in the total number of repeat units among the five samples at locus 1-2 (Fig. 7A). This is consistent with the slight differences in PCR product size observed in Fig. 6A. However, considerable variation existed between the isolates in the relative numbers of repeat units i.e. CTTTATTAT versus

CTTTATTAC. In contrast to locus 1-2 there was considerable variation in both the total number of units and the number of each repeat type among the five strains at locus 5-6 (Fig. 7B). This high degree of variation is also reflected in the PCR product size comparison (Fig. 6B).

The locus 5-6 amplification products from isolate H-303:NIH revealed two distinct fragments of about 320 bp and 450 bp respectively. Cloning of the PCR products from this locus also resulted in two inserts which differed in size by about a 100 bp (designated H-303:NIH-(3) and H-303:NIH-(4); Fig. 7B).

Shared single base alterations within the repeat blocks were seen at two positions in isolates 200:NIH and H-303:NIH at locus 1-2 (Fig. 7A). Another two single base changes were seen only within the repeat block of isolate IULA:1092:1 at locus 1-2, and this isolate was also missing a single copy of a 12 bp duplication in the 3'-flanking region. A single base change in the 5'-flanking region was present in isolate IULA:0593:2.

At locus 5-6 a single base change was seen within the repeat block (Fig. 7B) for isolate IULA:1092:1. Additionally, both isolate 200:NIH and H-303:NIH-(4) appeared to be missing the initial 10 bp of the first 12 bp repeat unit (GTATGTTTCTAT). A difference was also evident in the flanking regions of the repeat units in that isolate 200:NIH was missing a single copy of an 8 bp tandem duplication at the 5′ end.

While it is possible that single nucleotide differences are PCR amplification artifacts, it is highly unlikely that the shared nucleotide differences among isolates and repeats or repeat number variations could have this origin.

3.3.5. CHARACTERISATION OF OTHER LOCI CONTAINING INTERNAL REPEATS

A number of other DNA elements containing internal tandem repeats have been reported in *E. histolytica*. No attempts have been made to study their potential for the detection of intra-species polymorphisms. Three of these internally repetitive loci were selected for study: a 978 bp element described by Michel *et al* (1992) (GenBank accession number M77091; our designation: locus 3-4), a 931 bp DNA element isolated by Rosales-Encina and Eichinger (personal communication; GenBank accession number AF265348; our designation: locus 9-4), and a 964 bp element reported by Huang *et al* (1997) (our designation: locus 16-17). Complete nucleotide sequences are given in Appendix V (locus 3-4), Appendix VI (locus 9-4) and Appendix VII (locus 16-17). Schematic representations of the repeat arrangements seen at these loci are given in Fig. 8A, 8B and 8C respectively. Characteristic features of the repeat loci are summarised in Appendix II.

Although all three loci have been reported as unique repeated-DNA sequences, there is a high degree of identity between the two major repeat blocks of locus 3-4 and those of locus 9-4. The ten CTATTATA tandem repeats of locus 9-4 (Fig. 8B) differ from the eleven CTTATTATA tandem repeats of locus 3-4 (Fig. 8A) only in the absence of a single nucleotide (T) at the second position of each unit. The repeat unit CTTTATTATTAT in locus 9-4 is identical to the 12 bp repeat units of locus 3-4 with the only difference being the total number of units seen, i.e. locus 3-4 has eight units while locus 9-4 has only seven (Fig. 8; Appendix IIB). In fact, this high degree of identity between the two loci is apparent in the repeat-flanking regions as well. The sequences from position 1 to 540 and 541 to 931 of locus 9-4 (Appendix VI) are very similar to the nucleotide stretches spanning positions 401 to 977 and 1 to 400 in locus 3-4, respectively (Appendix V). Despite the high degree of sequence identity however, there is enough sequence variation evident in the repeat and repeat-flanking regions to suggest that both are distinct loci.

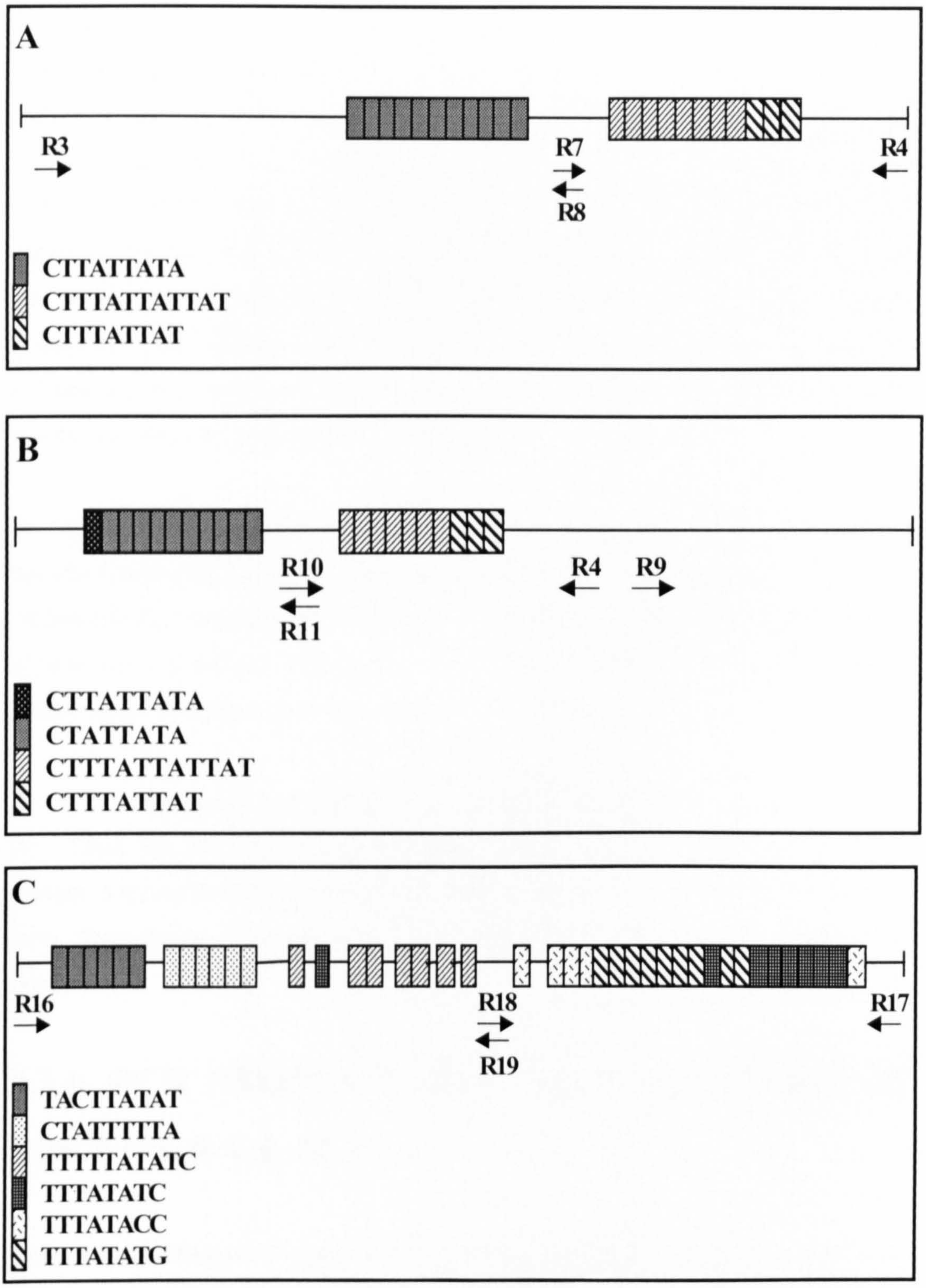


Fig 8. Schematic Representation of Repeat Arrangements at Three other Loci from Strain HM-1:IMSS Clone 9. (A) Locus 3-4. (B) Locus 9-4. (C) Locus 16-17. Only the major blocks of internal tandem repeats are shown. Positions of primers for whole and half-locus amplification are shown for all three loci. The unusual primer orientation at locus 9-4 **(B)** is further explained in Appendix VIII.

Having established that sequence diversity exists among different isolates in both the repeat and repeat-flanking regions of locus 1-2 and 5-6 and given that the locus 3-4 and 9-4 sequences have originated from two different isolates i.e. HK-9 and HM-1:IMSS respectively, we decided to clone and sequence loci 3-4 and 9-4 from the strain HM-1:IMSS clone 9 being maintained in our laboratory to see if the two are indeed unique loci or if they are in fact the same. PCR products of both were cloned and sequenced as previously described (sections 2.7 & 2.10) and the resulting sequences were aligned with existing ones. Nucleotide sequence comparisons between our two clones and previous sequences (Appendix IX) revealed differences in both the repeat and repeat-flanking regions.

On comparing the sequences of loci 1-2 and 5-6 with those of loci 3-4 and 9-4 we find that the repeat unit CTTTATTAT, which occurs a total of seven times in locus 1-2 (Appendix II), is identical to the three 9 bp units present in the second repeat blocks of both loci 3-4 and 9-4. The repeat units of locus 5-6 are however quite unique, as are the repeat flanking regions of both loci 1-2 and 5-6.

The nucleotide sequence of locus 16-17 is completely different from that of the other loci. There are six major types of internal repeats, with some being arranged in tandem only, while others exist as both tandem and solitary copies (Fig. 8C). Besides these, duplications of 5 to 8 bp are also seen interspersed among these repeats (not shown).

3.3.6. PCR PRODUCT SIZE POLYMORPHISM AT LOCI 3-4, 9-4 & 16-17

Primers were designed to amplify all three loci and the products were analysed on 1.8% agarose gels to look for size polymorphism among the thirteen *E. histolytica* isolates. In each case, primers were also designed in the regions between the main repeat blocks to look additionally for size variations in each half of the locus (Fig. 8A, 8B & 8C).

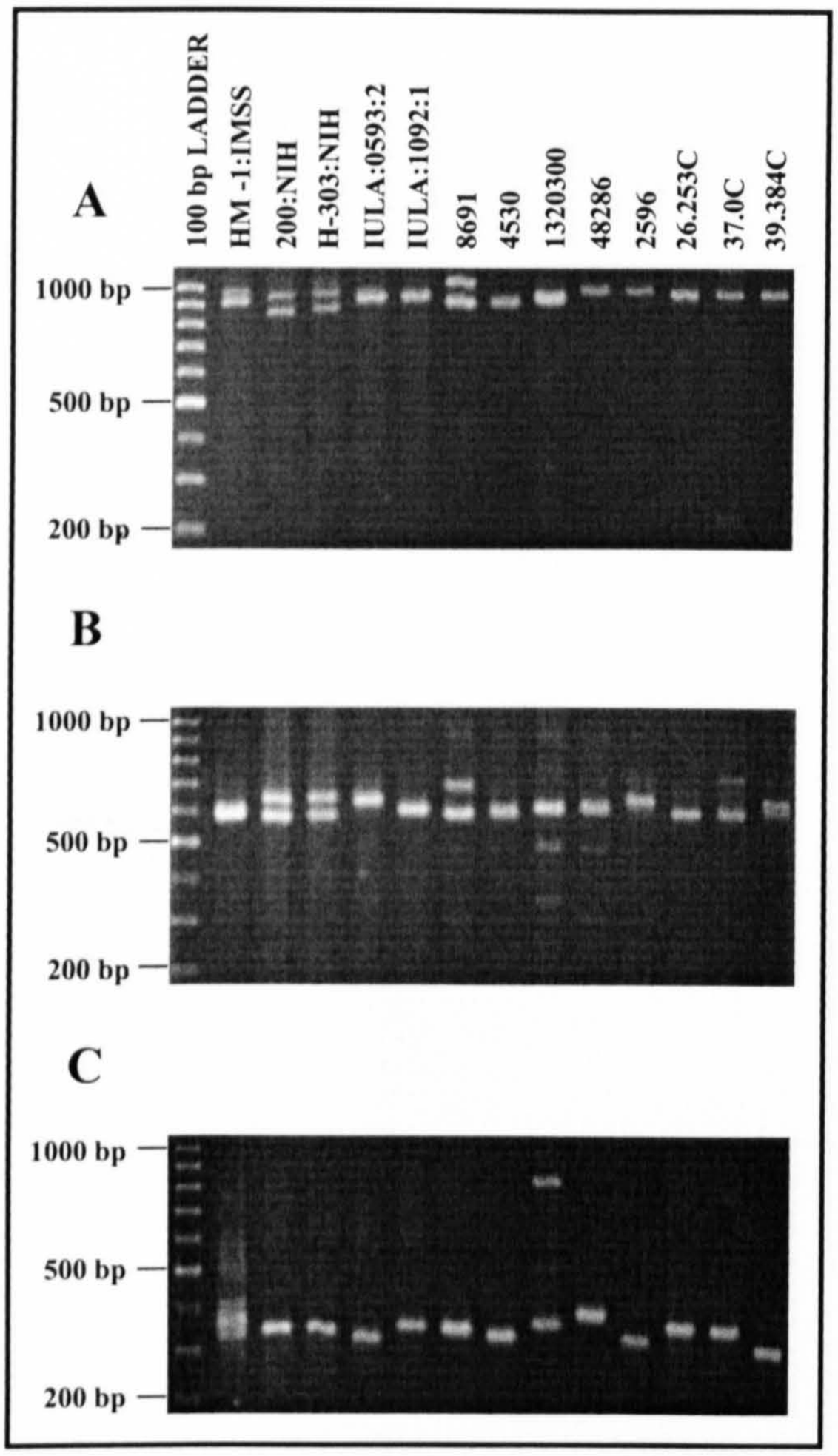


Fig. 9. Polymorphic DNA Analysis of *E. histolytica* Isolates at Locus 3-4 and Half Loci 3-8 & 7-4. (A) Locus 3-4. Amplification products were generated using primers R3 & R4 at an annealing temperature of 55°C. **(B) Half-locus 3-8.** Amplification products were generated using primers R3 & R8 at an annealing temperature of 50°C. **(C) Half-locus 7-4.** Amplification products were generated using primers R7 & R4 at an annealing temperature of 50°C.

Isolate HM-1:IMSS clone 9 gave a double amplification product of about 900 bp at locus 3-4 (Fig. 9A). Most *E. histolytica* isolates gave single major products with little size variation. Isolates 200:NIH and H-303:NIH, however, show a second band of equal intensity at ca. 850 bp, while isolates 8691 and 37.0C display a second band of ca. 1 kb. Amplification of the two half loci presented a very different pattern with much more variation seen than with the whole locus (Fig. 9B & 9C).

Amplification of locus 9-4 also gave two products between ca. 900 bp and 1 kb in isolate HM-1:IMSS clone 9 (Fig. 10A). Isolates 200:NIH and H-303:NIH again show two bands of equal intensity. As before, amplification of both half loci (Fig. 10B & 10C) produced a greater variety of banding patterns than seen at the whole locus.

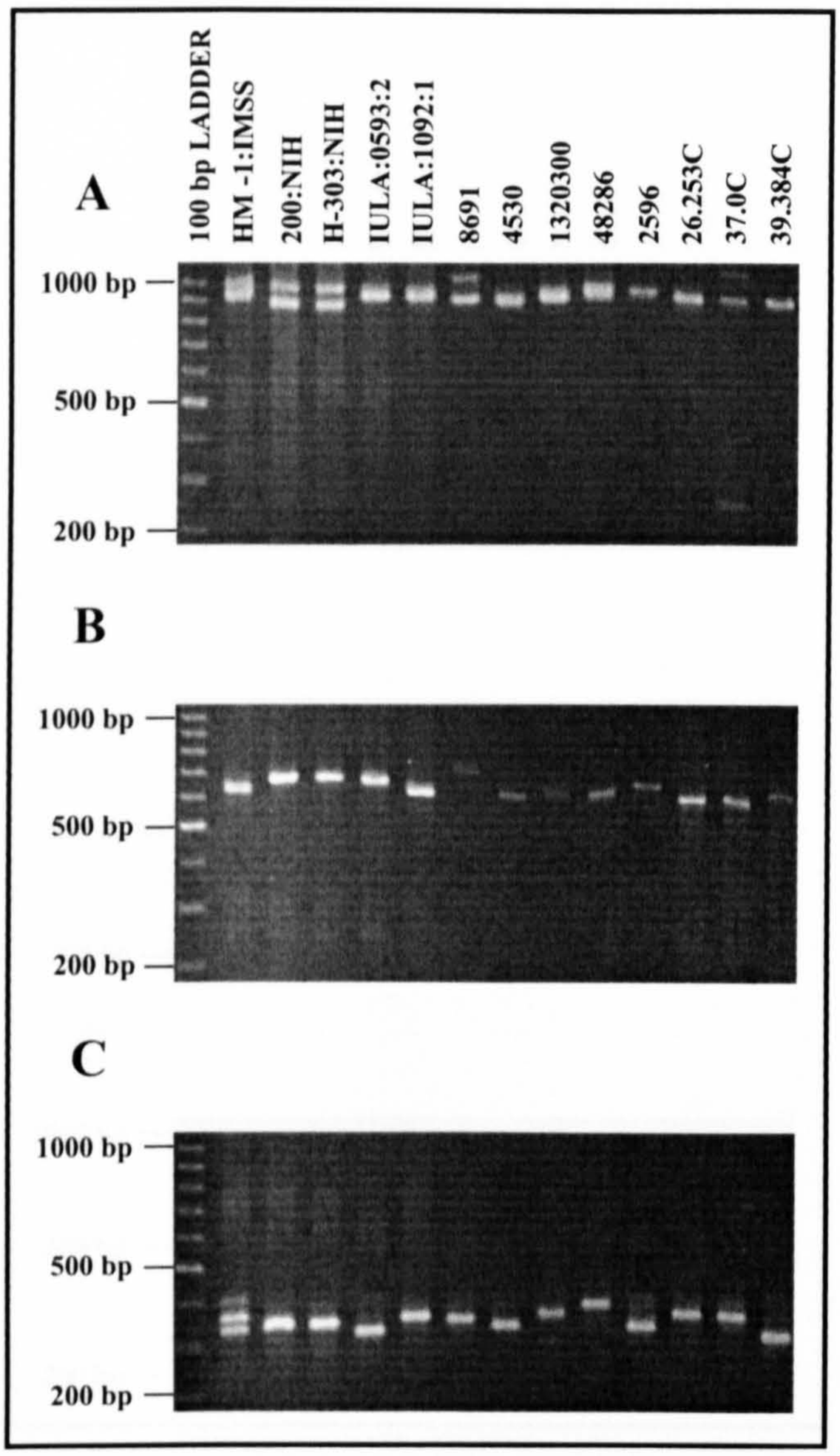


Fig. 10. Polymorphic DNA Analysis of *E. histolytica* Isolates at Locus 9-4 and Half Loci 9-11 & 10-4. (A) Locus 9-4. Amplification products were generated using primers R9 & R4 at an annealing temperature of 55°C. **(B) Half-locus 9-11.** Amplification products were generated using primers R9 & R11 at an annealing temperature of 50°C. **(C) Half-locus 10-4.** Amplification products were generated using primers R10 & R4 at an annealing temperature of 50°C.

Amplification of locus 16-17 gave the expected product of about 900 bp in isolate HM-1:IMSS clone 9 (Fig.11A). Many of the *E. histolytica* isolates gave single major products with little size variation among them, although isolate 4530 gave two clear products of equal intensity and certain others gave two bands very close in size. Amplification of the two half loci again produced a highly polymorphic array of bands (Fig. 11B & 11C).

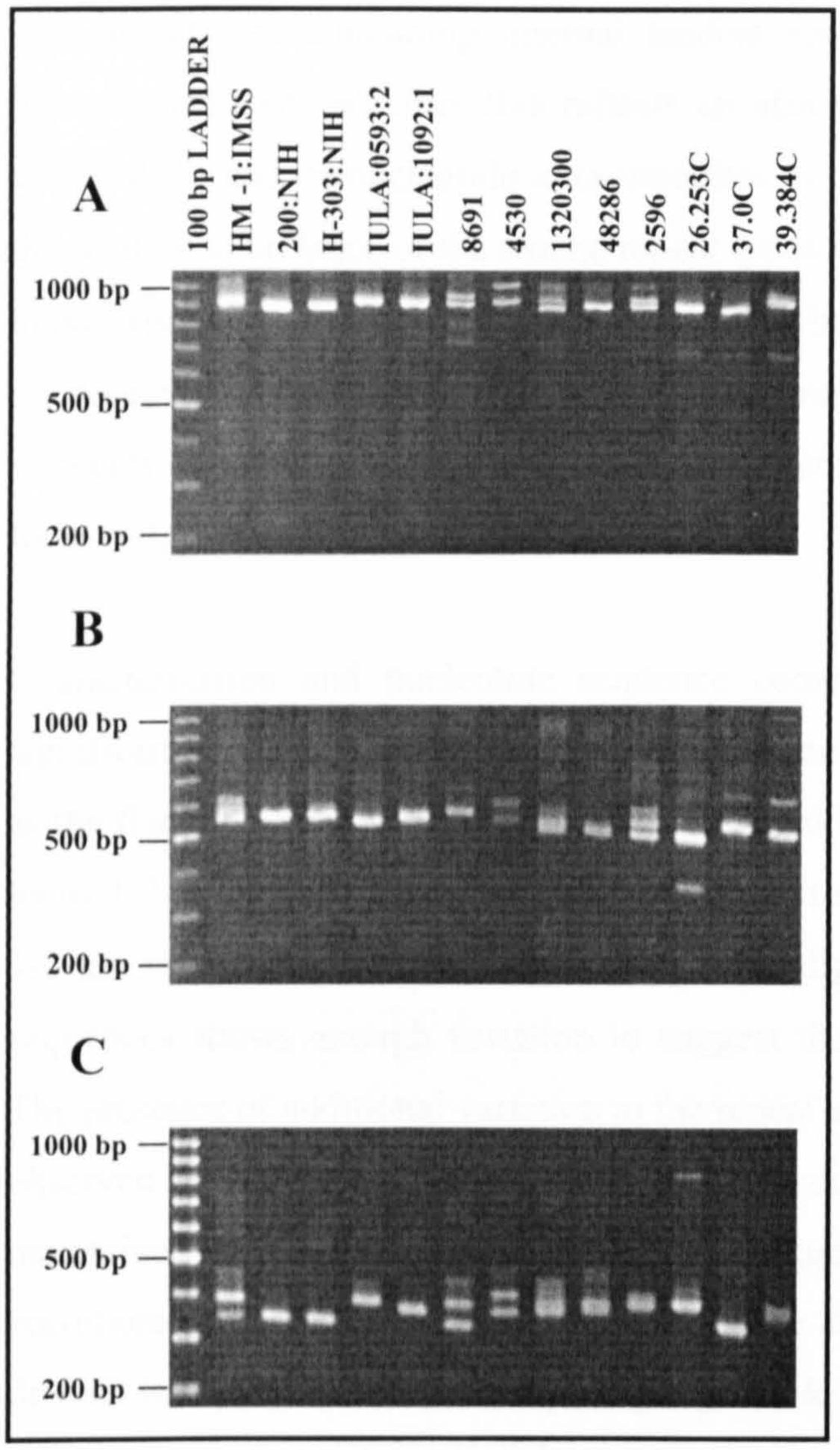


Fig. 11. Polymorphic DNA Analysis of *E. histolytica* Isolates at Locus 16-17 and Hlaf Loci 16-19 & 18-17. (A) Locus 16-17. Amplification products were generated using primers R16 & R17 at an annealing temperature of 55°C. **(B) Half-locus 16-19.** Amplification products were generated using primers R16 & R19 at an annealing temperature of 54°C. **(C) Half-locus 18-17.** Amplification products were generated using primers R18 & R17 at an annealing temperature of 54°C.

3.4. DISCUSSION

Attempts to clone microsatellite loci using a modification of a method that was successful in other organisms failed. In the study reported by Fischer and Bachmann (1998) most of the recombinant clones contained inserts of expected lengths and sequencing analysis revealed that twenty-nine of the forty-eight readable sequences contained microsatellites, in most cases corresponding to the oligonucleotides used as probes. Furthermore, no duplicate sequences were observed, indicating a high degree of microsatellite enrichment. Likewise, Oliveira *et al* (1998) have reported that 68% of the enriched clones sequenced by them showed the presence of the expected CA-repeats.

Two novel loci containing internal tandem repeats i.e. locus 1-2 and 5-6, were however isolated. Whether this reflects an absence or a reduced population of the classical di- and tri-nucleotide microsatellites in the *E. histolytica* genome or simply the relative abundance of the cloned repeat types remains to be seen. A $(GA)_{27}$ stretch in an expressed sequence tag has been reported by Azam *et al* (1996) suggesting that microsatellites do exist in this organism. Three other, previously reported DNA elements containing internal tandem repeats were also selected for analysis, namely locus 3-4, 9-4 and 16-17.

Characterisation and nucleotide sequence comparisons of the five loci revealed significant levels of identity between locus 3-4 and 9-4 in the repeat domains as well as the flanking regions. Similarity was also noted between some of the repeat units of locus 1-2 and those of loci 3-4 and 9-4 although the repeat-flanking regions of locus 1-2 are unique. However, complete sequence alignment of locus 3-4 and 9-4 DNA sequences shows enough variation to suggest that they are distinct (Appendix IX). The presence of additional variation in the repeat and repeat-flanking regions which is observed when locus 3-4 and 9-4 PCR products from the strain HM-1:IMSS maintained in our laboratory were cloned, sequenced and aligned with the former corroborates this view. Additionally, evidence that locus 3-4 and 9-4 are indeed distinct is clearly seen from comparing Fig. 9B & 10B. In fact the multiple alignment data suggests that sequence and repeat number variation may even exist at a given locus within the same strain e.g. locus 9-4 in strain HM-1:IMSS (Appendix IX) and locus 5-6 in strain H-303:NIH (Fig. 7B; Appendix IV).

A review of literature and search of the GenBank database revealed that the two other internal repeat containing DNA elements reported independently by Lohia *et al* (1990) and Willhoeft and Tannich (2000) also bear remarkable similarity in their repeat and repeat-flanking regions to loci 3-4 and 9-4. Once again complete sequence alignment of these two with the four already mentioned (Appendix IX) reveals enough differences to suggest that they are all related but distinct. Although the fragment reported by Lohia *et al* (1990) was obtained from strain 200:NIH suggesting that the observed differences may be due to strain variation, the DNA sequence described by Willhoeft and Tannich (2000) was also derived from strain HM-1:IMSS.

It is possible that these six DNA sequences are discrete members of the same family of repeated-DNA containing elements.

A similar situation is seen in other eukaryotes in case of the SINE and LINE families, where no two elements are exactly alike. However, the degree of sequence similarity among the different members within these families of repeated-DNA is significant enough to suggest a common ancestral relationship among them (Darnell, 1990).

Successful amplification of locus 9-4 using the primer orientation indicated in Fig. 8B and Appendix VI suggested that this element may be arranged in tandem arrays. This observation was subsequently confirmed for all five loci and the results are presented in Chapter 7 (section 7.2.5).

Size variations within the repeated domains were studied and all loci studied showed PCR product length polymorphism. The degree of diversity varied, with some loci showing more polymorphism than and thus having a greater potential for detecting inter-strain differences among *E. histolytica* isolates than others (i.e. locus 5-6 and the six half loci). Despite being from geographically restricted regions the Venezuelan, Bangladeshi and South African samples could be differentiated with ease. For the most part the variations in PCR product size appear to be a result of differences in the numbers of tandem repeat units. Finally while no single locus discriminated between all the samples, the collective use of multiple loci allowed differentiation of a majority of the *E. histolytica* isolates.

At most loci amplification resulted in two or three bands in at least some isolates. There are two possible explanations for this observation. The first reflects our understanding of the ploidy of the *Entamoeba* genome. Present evidence suggests that the *Entamoeba* genome is tetraploid with variations in length between homologous chromosomes (Willhoeft and Tannich, 1999). It is therefore possible that the multiple bands observed reflect polymorphism among homologous loci on allelic chromosomes. Multiple amplification products have also been reported for the SREHP gene in a number of *E. histolytica* isolates (Clark and Diamond, 1993a), a finding consistent with the isolation of cDNAs containing variable numbers of internal tandem repeats (Köhler and Tannich, 1993). The two cDNAs identified for

the SREHP gene differed in length by about 108 nucleotides, which is ample for the observation of size variation by PCR. On Southern blot the SREHP behaves like a single copy gene suggesting that the length differences are allelic variations.

Alternatively, the presence of multiple bands in this study could be explained by the existence of these repeat loci at multiple locations in the *Entamoeba* genome, each with a characteristic PCR product size. Some of the most abundant repeated-DNA species in eukaryotes i.e. mini- and microsatellite loci are found at multiple sites throughout the genome. Although their sequences are highly conserved, length variations due to unequal number of repeat units between two or more loci are a frequent occurrence (Darnell, 1990). The presence of multiple hybridisation bands on a blot is also thought to be indicative of a multicopy repeated-DNA element as already suggested for locus 3-4 and 16-17 by Michel *et al* (1992) and Huang *et al* (1997) respectively and as observed by us (section 7.2.5). Multicopy repeated-DNAs have also been reported by Mittal *et al* (1994) and Cruz-Reyes *et al* (1995). Unlike our loci however, these multicopy elements do not contain internal tandem repeats.

The observed size variation was further characterised by nucleotide sequence comparison of five axenic isolates at loci 1-2 and 5-6. Isolates 200:NIH and H-303:NIH are easily differentiated by PCR from most of the other isolates but cannot reliably be separated from each other by gel electrophoresis. However, DNA sequence comparison at locus 5-6 revealed that the two isolates can be distinguished by the absence of one copy of an 8 bp tandem duplication from strain 200:NIH a difference which is too small to be detected on conventional gel electrophoresis. Thus PCR product sizes alone may not discriminate among all distinct isolates, at least under these conditions. Given that locus 3-4 and 9-4, which appear to be related but discrete members of the same family, are probably multicopy elements, it is equally likely that the same is true of locus 5-6. As such, the sequence variation at locus 5-6 may reflect differences between copies of a family of related sequences and not variation between isolates. This could also potentially explain the presence of sequence and repeat number variation at a given locus within the same strain (see above).

Although there is no evidence that culture conditions or media have any effect on the polymorphisms studied it appears that variations in primer sequence and DNA concentration can greatly influence PCR amplification. Although we were able to amplify and subsequently sequence locus 5-6 PCR products from the five axenic strains with primers R5 and R6, the xenic isolates could only be amplified following primer sequence modifications i.e. R5A and R6A. Sequence variations in repeat-flanking regions have been shown for locus 1-2 and 5-6 among the small number of isolates examined (Fig. 7A & 7B) as well as between the locus 9-4 clone generated from strain HM-1:IMSS clone 9 in this study when compared to the sequence reported by Rosales-Encina and Eichinger (Appendix IX). The fact that these variations could be present between different *E. histolytica* strains or between different copies of a multigene family within the same strain suggests the need for caution when designing primers from a single sequence and interpreting negative results.

The most conceivable explanation for the observation that dilution of DNA increases the probability of obtaining amplification products from xenic isolates is the likely reduction in interference provided by the bacterial DNA, which is also present in those samples. Successful amplification currently appears to be reliant on the optimal amount of DNA used in the reaction, which in turn varies among loci. Not withstanding these limitations it is apparent that DNA isolated from both axenic and xenic cultures can be used. That DNA derived directly from stool samples can also prove a suitable template for PCR amplification at these loci is shown in Chapter 4.

In summary it appears from these results that a number of loci showing size polymorphism are present in *E. histolytica* and by virtue of their ability to discriminate among individual isolates, they appear to have potential as tools to answer many of the outstanding questions surrounding the epidemiology of this parasite. In the present work, all of the isolates came from individuals who had invasive disease or were likely infected by someone who had invasive disease. Whether invasive and non-invasive strains of *E. histolytica* exist remains to be established.

To try and address some of these questions and to determine the general utility of these polymorphic loci for investigating the molecular epidemiology of *E. histolytica* we next examined a larger sample from a broader geographic range.

CHAPTER 4

GLOBAL SURVEY OF INTER-STRAIN POLYMORPHISM IN *E. histolytica*

4.1. INTRODUCTION

Results from the previous chapter have shown that a number of repeated-DNA-containing loci showing PCR fragment size polymorphism are present in *E. histolytica*. However, the total number of samples studied was small and the different geographic regions were represented by only a few isolates.

Based on reported literature, prevalence of infection and incidence of disease varies between geographic locations (Walsh, 1986) and even between two close communities within an endemic area (Jackson *et al*, 2000). It appears that in endemic regions transmission of amoebae occurs between and across a variety of different social groups (Caballero-Salcedo *et al*, 1994). However, in North America (Healy, 1986), Europe (Gatti *et al*, 1995; Walderich *et al*, 1997) and Japan (Takeuchi *et al*, 1990; Ohnishi and Murata, 1997) amoebiasis is usually only seen in certain select population groups e.g. children, members of extended families, immigrants from or travellers to endemic areas, homosexuals and residents of institutions for the mentally handicapped. Furthermore, Japan is one of the few developed countries where many of the *E. histolytica* isolates identified are native to the region rather than being acquired through travel, emigration and/ or immigration.

This variation in disease burden is attributed to differences in transmission patterns, parasite virulence or host susceptibility, factors which may themselves be influenced by diverse ecological, socio-economic and cultural conditions.

To see if the polymorphic loci described in the previous chapter can be used for investigating the molecular epidemiology of *E. histolytica* in a larger sample from a wide geographic range and to gain better understanding of the polymorphic nature of *E. histolytica* isolates across the globe, DNAs were obtained from different sources and amplified by PCR to look for fragment size polymorphism at all eleven loci. The results are presented here.

4.2. MATERIALS & METHODS

4.2.1. *E. histolytica* ISOLATES

A total of eighty-four *E. histolytica* samples were studied. The various population groups and sources from which they were obtained are as follows:

GROUP 1

A total of ten axenic strains were obtained from different sources within the London School of Hygiene and Tropical Medicine (LSHTM) (Table 6A). Eight (#) are American Type Culture Collection (ATCC) strains (www.atcc.org) while the other two were originally isolated and characterised by Urdaneta *et al* (1995). Only one of these, strain HM-1:IMSS clone 9 (A), is being maintained by us. Four samples (B) were provided by Dr. J. Ackers and A. Shire, (LSHTM) while the remaining five samples (C) were obtained from Dr. D. Nolder (LSHTM). Where available, clinical, parasitological and sero-epidemiological information was obtained from relevant sources (cited above) and is presented in Appendix X.

Table 6A

E. histolytica ISOLATES: GROUP 1

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
1	HM-1:IMSS clone 9#	Mexico	Culture (axenic)	LYI-S-2	CTAB	A
2	200:NIH#	NA		YI-S		B
3	H-303:NIH#	VietNam (?)				
4	IULA:0593:2	Venezuela				
5	IULA:1092:1					
6	HK-9#	Korea		TYI-S-33		C
7	Rahman#	England/ India				
8	SD157#	USA				
9	DKB#	England				
10	HB-301:NIH#	Burma				

GROUP 2

DNA from ten Bangladeshi isolates (Table 6B), including the four already described in chapter 3, was obtained from Dr. A. Aguirre (LSHTM) (D).

Table 6B

E. histolytica ISOLATES: GROUP 2

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
11	8691	Bangladesh	DNA (xenic)	Robinson	Phenol Chloroform	D
12	4530					
13	1320300					
14	2701					
15	48286					
16	1363050					
17	1375975					
18	44298					
19	58996					
20	1353600					

The isolates were originally provided to her in 1995 by Dr. R. Haque of The International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). All samples were sent as culture pellets in dry ice and were from patients who presented at the ICDDR,B with diarrhoea (Appendix X). Dr. Aguirre further confirmed the species identification by PCR-SHELA (Aguirre *et al*, 1995; personal communication).

GROUP 3

Dr S. Kobayashi and Dr T. Takeuchi (Keio University School of Medicine, Japan) (E) provided us with a total of twenty-two lysates (Table 6C). Seventeen of these were isolated at the Keio University, while five (*) were originally isolated at the Tokai University, Japan, by Dr H. Tachibana's group. All except one individual are natives of Japan. Isolate J17 is from a Thai national who is a migrant worker in a Japanese bar. Two other individuals, J6 and J12, have a history of voluntary work in Ghana and Cambodia respectively (personal communication).

Table 6C

E. histolytica ISOLATES: GROUP 3

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
21	J1	Japan	Lysate (xenic)	Robinson	CTAB	E
22	J2		Lysate (axenic)	TYI-S-33		
23	J3					
24	J4		Lysate (xenic)	Robinson		
25	J5		Lysate (axenic)	TYI-S-33		
26	J6	Japan/ Ghana	Lysate (xenic)	Robinson		
27	J7	Japan				
28	J8					
29	J9					
30	J10					
31	J11					
32	J12	Japan/ Cambodia				
33	J13	Japan				
34	J14					
35	J15					
36	J16					
37	J17	Japan/ Thailand				
38	J18*	Japan				
39	J19*					
40	J20*					
41	J21*					
42	J22*					

GROUP 4

A single Korean isolate YS-27 (Table 6D; Chang *et al*, 1995), which has been well characterised (Park *et al*, 1999) was provided by Dr T.S. Yong of the Yonsei

University College of Medicine, Korea (F). Information given in Appendix X was confirmed by personal communication.

Table 6D

E. histolytica ISOLATES: GROUP 4

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
43	YS-27	Korea	Lysate (axenic)	TYI-S-33	CTAB	F

GROUP 5

A single DNA sample of an isolate from a male patient suffering from diarrohea and with history of recent travel to Malaysia and Australia (Table 6E), was provided by Dr A. Aguirre (LSHTM) (D). The faecal sample was originally received by J. Williams of the Diagnostic Parasitology Laboratory (LSHTM) cultured and subsequently given to Dr. Aguirre who characterised it as *E. histolytica* by PCR-SHELA (personal communication).

Table 6E

E. histolytica ISOLATES: GROUP 5

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
44	887C	Malaysia/ Australia	DNA (xenic)	Robinson	QIAamp Tissue DNA Extraction Kit	D

GROUP 6

DNA samples from ten Vietnamese individuals (Table 6F) were obtained courtesy of Dr E. Tannich of the Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany (G). These individuals were originally recruited as part of a large-scale epidemiological survey of amoebiais in a region of high incidence of amoebic liver abscess in Central Vietnam (Blessmann *et al*, 2002). Cultures were not established for any of the samples and hence zymodemes were not determined (Appendix X).

Table 6F

E. histolytica ISOLATES: GROUP 6

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
45	20-2	Vietnam	DNA (faecal)	Not Done	Qiagen Stool DNA Extraction Kit	G
46	26-6					
47	41-1					
48	46-5					
49	66-1					
50	72-1					
51	82-4					
52	84-3					
53	85-3					
54	148-7					

GROUP 7

J. J. Verweij of The Leiden University Medical Centre, The Netherlands (H), provided us with DNA of ten isolates (Table 6G).

Table 6G

E. histolytica ISOLATES: GROUP 7

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
55	25591C	South America	DNA (xenic)	Robinson	QIAamp Tissue DNA Extraction Kit	H
56	25591F		DNA (faecal)			
57	26825C	Rwanda/ Zaire	DNA (xenic)	Robinson		
58	26825F		DNA (faecal)			
59	27749C	Tibet/ Nepal	DNA (xenic)	Robinson		
60	27749F		DNA (faecal)			
61	28577C	Guinea	DNA (xenic)	Robinson		
62	28577F		DNA (faecal)			
63	28639C	NA	DNA (xenic)	Robinson		
64	28639F		DNA (faecal)			
65	29146C	NA	DNA (xenic)	Robinson		
66	29146F		DNA (faecal)			
67	29978C	West Africa	DNA (xenic)	Robinson		
68	29978F		DNA (faecal)			
69	30325C	Indonesia	DNA (xenic)	Robinson		
70	30325F		DNA (faecal)			
71	32083C	Nigeria	DNA (xenic)	Robinson		
72	32083F		DNA (faecal)			
73	32257C	Bangladesh	DNA (xenic)	Robinson		
74	32257F		DNA (faecal)			
75	1203	India	DNA (xenic)	Robinson	Phenol Chloroform	D
76	1205					

This particular laboratory acts as a reference centre and receives, among others, unpreserved faecal samples from all over the country suspected on the basis of microscopy or clinical signs to contain *E. histolytica/ E. dispar*. All ten samples were sent to us in duplicate with one tube containing DNA extracted directly from stool (suffix-F) while the other had culture-derived DNA (suffix-C) of the same isolate. Both culture and stool-derived DNA was extracted with the QIAamp kit (Table 6G). The samples were characterised as *E. histolytica* by PCR-SHELA as well as by the SREHP and SSG PCR (Clark and Diamond, 1993a; personal communication).

In addition we obtained two DNA samples (Table 6G) from Dr A. Aguirre who had been sent these isolates from the above mentioned reference lab in 1995 (D). The two samples came from members of a Dutch family among whom an outbreak of amoebiasis had been reported (Vreden *et al*, 2000).

GROUP 8

Dr A. Aguirre also gifted two DNA samples (Table 6H) that she had originally obtained from Drs M.E. Valls and T. Vinuesa at the Hospital Clinic i Provincial, Barcelona (D). One sample (822) which has been characterised as *E. histolytica* by PCR-SHELA (Aguirre *et al*, 1997) was from a native of Barcelona who had presented at the clinic with bloody diarrhoea (Appendix X) following a month long working trip to Vietnam. The other sample (2024) for which species identification has likewise been made (personal communication), was also from a native of Barcelona who presented with bloody diarrhoea after having been in India for a period of three months.

Table 6H

E. histolytica ISOLATES: GROUP 8

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
77	822	Vietnam	DNA (xenic)	Robinson	Phenol Chloroform	D
78	2024	India				

GROUP 9

Three of the four samples (Table 6I) provided by Dr. Terry Jackson of the Medical Research Council of South Africa, Durban (I) described in chapter 3 were also used (section 3.2).

Table 6I

E. histolytica ISOLATES: GROUP 9

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
79	37.0C	South Africa	Lysate (xenic)	Robinson	CTAB	I
80	39.384C					
81	2596					

GROUP 10

DNA from three axenized and characterised Brazilian strains (Table 6J; Silva *et al*, 1997) was provided by Dr M. A. Gomes, of the Universidade Federal de Minas Gerais, Brazil (J). The DNA was sent to us adsorbed on filter papers and eluted here.

Table 6J

E. histolytica ISOLATES: GROUP 10

No	Isolates	Strain Origin	DNA Origin	Culture Type	Extraction Method	Source
82	CSP	Brazil	DNA (axenic)	TYI-S-33	Phenol Chloroform	J
83	462					
84	RPS					

4.2.2. ISOLATION OF NUCLEIC ACIDS, PCR AMPLIFICATION & DATA ANALYSIS

DNA from all the samples sent to us as lysates was isolated as already described (section 2.3). PCR amplification was performed using the polymorphic locus -specific primer pairs given in Table 4. Primer 11, which is used to amplify the half-locus 9-11 was modified and the new primer 11A (Table 7) was used for the present analysis.

Table 7

MODIFIED POLYMORPHIC LOCUS-SPECIFIC OLIGONUCLEOTIDE PRIMER

Repeat Locus	Primer Name	Primer Sequence (5′ → 3′)
9-11	EhR11A (3′ primer)	TAC ATA AGT CGT GGT AAA GAG AAG

Amplification products were analysed using 1.8% agarose gels (BioGene) in 1X TBE at 120V.

Gels were documented by photography and fragment sizes were estimated using the DNA ladder run on the same gels. Once a size range had been determined for all the loci, the presence or absence of a given band was scored using a discrete two-character state i.e. 0 (absent) and 1 (present) (data not shown). Finally, using this information each unique pattern for a given isolate at a given locus was assigned a number i.e. genotype (Table 8).

4.3. RESULTS

4.3.1. PCR AMPLIFICATION

To begin with only the ten axenic samples were tested at all five loci and the six half-loci derived from them. Amplification was successful for all the samples except strains SD157 and HB-301:NIH which did not give any products at locus 9-11. Similarly although we were able to amplify the four Bangladeshi isolates originally tested (chapter 3), none of the remaining six isolates nor any of a randomly selected group of seventeen Japanese samples tested gave products at this locus. The primer R11 sequence was modified (Table 7) and following successful amplification of the aforementioned samples further analysis at locus 9-11 used this new R11A primer. Seventeen samples failed to amplify at locus 9-11 even after the modification (Table 8).

In all sixty-five of the eighty-four *E. histolytica* DNA samples included in this survey gave amplification products at all five loci and the six half-loci derived from them (Table 8). Nineteen samples (⊟) did not amplify with one or more of the locus-specific primer pairs, with two samples (82-4 & 84-3) not giving products at a majority of the loci.

A noteworthy observation is that while amplification of faecal DNA is possible with all or most of these polymorphic locus-specific primer-pairs, it appears to be dependant on the DNA extraction method and thermostable polymerase used. When attempting amplification of the samples sent to us from Holland (group 7) it was seen that although all ten culture derived DNA samples amplified well using the conditions previously described (section 2.11) none of the corresponding faecal DNA samples gave amplification products. Following personal communication with J. J. Verweij and based on the results recently published by their group (Verweij *et al*, 2000) we then attempted amplification of the faecal DNA using 2.5U of SuperTaq HC (HT Biotechnology, UK) and 2 µl of template DNA. All other conditions and concentrations remained as before. Results were obtained for all the samples at some (on average five to seven) but not all eleven of the loci (data not shown). Samples that failed to amplify were then tested using Hot Start Taq (DNAmp Ltd) along with a few randomly selected samples, which had given products with SuperTaq (HC). Not only was amplification successful in case of all the failed samples but results for the remaining samples that were tested, were comparable to those seen with SuperTaq (HC). An example of this can be seen in Fig. 12 which shows the amplification results of three samples with both SuperTaq (HC) and Hot Start Taq at locus 5-6. However, although we were eventually able to amplify almost all the samples at a majority of the loci (Table 8), product amount was in many cases low (Fig. 18).

Given the experience with this group of faecal DNAs, on first receiving the ten faecal DNA samples from Germany (group 6) a random group of five was chosen and amplification was attempted with both Hot Start Taq (DNAmp Ltd) and Taq polymerase (Bioline). Somewhat surprisingly, comparable results were obtained for all five samples with both types of polymerases (data not shown) and the remaining

samples were all amplified using Taq (Bioline). Furthermore, in contrast to the results seen for samples from Holland, the quality i.e. band intensity was comparable to that seen with most of the culture derived DNA samples (Fig. 17).

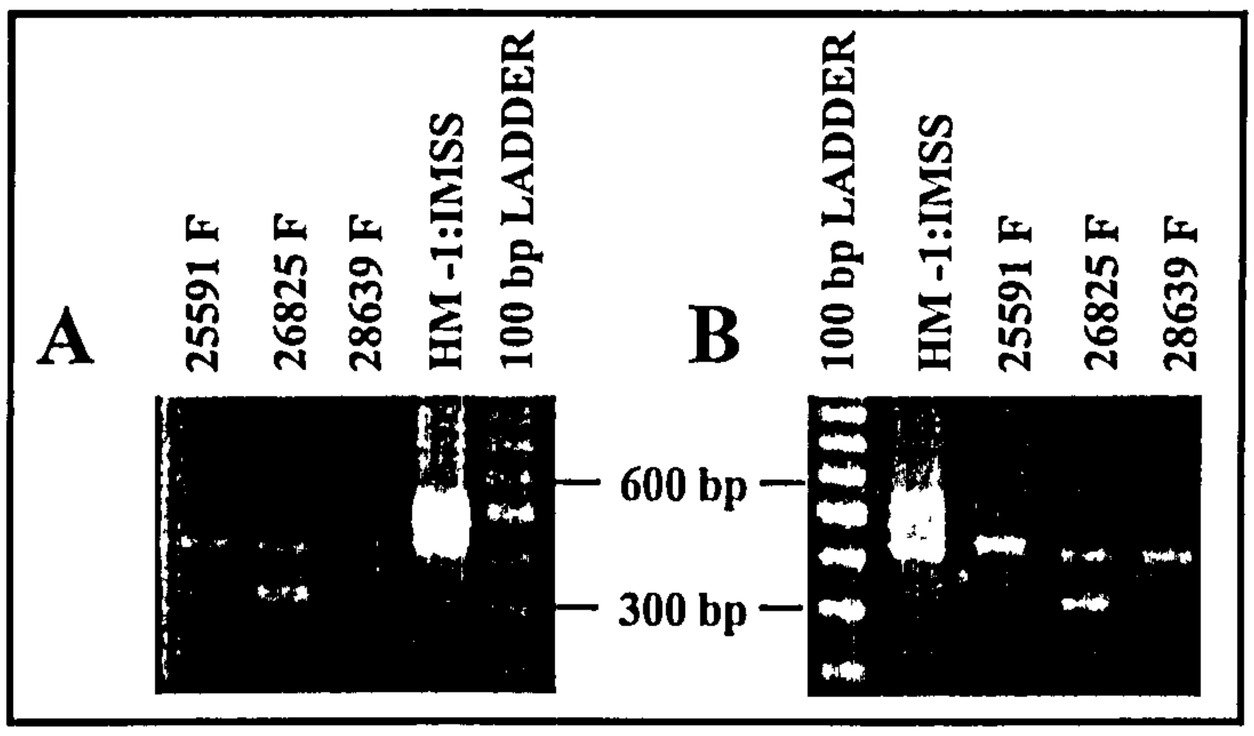


Fig. 12. Comparative Analysis of PCR Amplification with Faecal DNA at Locus 5-6. Amplification products were generated using primers R5A & R6A at an annealing temperature of 56°C. **(A)** Amplification with SuperTaq (HC) **(B)** Amplification with Hot Start Taq (DNAmp Ltd).

Table 8

GLOBAL SURVEY - *E. histolytica* GENOTYPES

No	Isolates	Locus 1-2	Locus 5-6	Locus 3-4	Locus 3-8	Locus 7-4	Locus 9-4	Locus 9-11	Locus 10-4	Locus 16-17	Locus 16-19	Locus 18-17
1	HM-1:IMSS	1	1	1	1	1	1	1	1	1	1	1
2	200:NIH	2	2	2	2	2	2	2	2	2	2	2
3	H-303:NIH	2	2	2	2	2	2	2	2	2	2	2
4	IULA:0593:2	1	34	3	4	21	7	2	17	11	3	7
5	IULA:1092:1	1	35	4	15	3	7	1	4	11	2	16
6	HK-9	2	2	2	2	2	2	2	2	2	2	2
7	Rahman	1	3	3	3	3	3	2	3	3	3	3
8	SD157	2	4	4	4	2	4	3	4	4	2	4
9	DKB	1	1	3	5	2	3	2	4	5	1	5
10	HB-301:NIH	1	5	3	1	4	3	2	5	6	2	5
11	8691	3	6	5	6	5	5	4	4	7	4	1
12	4530	3	7	5	5	5	6	1	4	8	1	6
13	1320300	3	4	3	4	6	7	5	6	3	3	7
14	2701	3	8	4	7	7	3	5	6	4	4	8
15	48286	4	4	6	8	8	3	5	7	3	2	9
16	1363050	4	9	6	8	7	3	6	8	4	4	1
17	1375975	4	10	4	9	6	4	7	8	9	5	9
18	44298	5	5	4	8	9	8	7	9	4	5	9
19	58996	6	11	4	10	6	4	4	8	7	1	10
20	1353600	3	12	7	11	10	9	7	8	4	6	11

21	J1	2	21	4	3	3	13	1	12	14	15	4
22	J2	2	21	4	14	6	3	8	4	11	1	19
23	J3	2	22	3	5	14	13	8	12	11	16	11
24	J4	2	1	4	4	15	3	[illegible]	2	11	17	1
25	J5	2	22	3	5	6	13	8	12	11	16	1
26	J6	11	23	3	4	3	13	9	12	11	1	4
27	J7	2	4	4	4	1	3	1	2	14	18	1
28	J8	11	5	3	5	9	3	8	4	11	2	20
29	J9	2	4	4	4	17	3	9	2	15	18	1
30	J10	2	24	4	4	17	3	[illegible]	2	15	19	1
31	J11	11	20	14	18	18	14	1	5	16	18	21
32	J12	2	20	14	3	14	15	8	12	16	20	14
33	J13	2	4	5	3	3	7	9	12	14	18	1
34	J14	2	4	5	3	3	7	9	12	14	18	1
35	J15	12	20	14	13	6	1	[illegible]	4	11	3	22
36	J16	12	20	14	13	6	1	9	4	16	16	14
37	J17	1	25	15	19	15	7	[illegible]	2	11	1	14
38	J18	1	25	15	3	15	13	[illegible]	2	11	1	14
39	J19	1	25	15	3	15	13	[illegible]	2	11	1	14
40	J20	1	25	15	3	15	13	[illegible]	2	11	1	14
41	J21	12	25	15	3	15	13	9	2	11	1	14
42	J22	1	25	15	3	15	13	9	2	11	1	14
43	YS-27	2	21	5	8	14	3	9	12	11	2	2
44	887C	1	26	16	20	14	1	7	1	11	3	16
45	020-2	1	6	12	1	12	19	2	12	14	2	7
46	026-6	1	13	12	1	12	19	2	12	11	25	11
47	41-1	1	3	14	5	12	19	1	12	2	26	2
48	46-5	1	6	12	1	12	19	2	12	14	2	11
49	66-1	1	13	12	1	12	19	2	12	11	2	7
50	72-1	1	3	14	5	12	7	1	12	2	11	2
51	82-4	[illegible]	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	[illegible]	2	11
52	84-3	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	4	[illegible]	2	11
53	85-3	1	31	3	5	12	7	1	12	2	25	2
54	148-7	1	31	3	5	12	7	1	12	14	27	22
55	25591C	1	15	11	15	12	12	1	12	2	2	11
56	25591F	1	15	11	15	12	12	1	12	2	2	11
57	26825C	1	16	3	16	12	1	8	13	11	2	15
58	26825F	1	16	3	16	12	1	8	13	11	2	15
59	27749C	1	17	11	15	12	12	1	12	11	9	16
60	27749F	1	17	11	15	12	12	1	12	11	9	16
61	28577C	10	18	11	15	12	12	1	12	11	10	2
62	28577F	10	18	11	15	12	12	1	12	11	10	2
63	28639C	1	4	11	15	12	12	[illegible]	12	2	3	2
64	28639F	1	4	11	15	12	12	[illegible]	12	2	3	2
65	29146C	1	16	3	17	12	7	1	12	2	11	2
66	29146F	1	16	3	17	12	7	1	12	2	11	2
67	29978C	1	19	11	15	12	7	1	12	11	2	16
68	29978F	1	19	11	15	12	7	1	12	11	2	16
69	30325C	1	4	12	17	13	1	1	14	12	12	16
70	30325F	1	4	12	17	13	1	1	14	12	12	16
71	32083C	10	16	13	9	12	3	[illegible]	2	11	2	16
72	32083F	10	16	13	9	12	3	[illegible]	2	11	2	16

73	32257C	1	20	11	17	12	7	▯	2	1	13	17
74	32257F	1	20	11	17	12	7	▯	2	1	13	17
75	1203	1	15	3	8	12	3	1	12	13	14	18
76	1205	1	15	3	8	12	3	1	12	13	14	18
77	822	7	13	8	12	11	10	▯	10	▯	7	11
78	2024	8	4	1	13	3	1	▯	11	▯	8	12
79	37.0C	1	28	17	6	14	18	10	4	17	21	2
80	39.384C	1	27	3	21	16	7	▯	12	11	21	20
81	2596	10	33	18	3	3	13	1	12	17	▯	20
82	CSP	9	8	4	5	5	6	1	10	10	2	13
83	462	1	1	9	3	1	1	2	1	10	2	2
84	RPS	3	14	10	14	3	11	4	12	10	2	14

Note: "▯" samples did not amplify and hence genotypes were not assigned.

4.3.2. PCR PRODUCT SIZE POLYMORPHISM

When the results depicted in Table 8 were collated it was apparent that the eighty-four samples included in this survey displayed a high degree of inter-strain polymorphism.

Table 9

GLOBAL SURVEY -

SUMMARY OF *E. histolytica* GENOTYPES

Loci	1-2	5-6	3-4	3-8	7-4	9-4	9-11	10-4	16-17	16-19	18-17
Total No of Genotypes	12	32	18	21	19	17	10	15	17	24	22
Group 1 (10 isolates)	2	7	4	6	5	5	3	6	7	3	7
Group 2 (10 isolates)	4	9	5	8	6	7	5	5	5	6	7
Group 3 (22 isolates)	4	9	5	7	9	6	3	4	5	9	8
Group 4 (1 isolate)	1	1	1	1	1	1	1	1	1	1	1
Group 5 (1 isolate)	1	1	1	1	1	1	1	1	1	1	1
Group 6 (10 isolates)	1	4	3	2	1	2	2	2	3	4	5
Group 7 (12 isolates)	2	7	4	5	2	4	2	4	5	8	6
Group 8 (2 isolates)	2	2	2	2	2	2	-	2	-	2	2
Group 9 (3 isolates)	2	3	3	3	3	3	2	2	2	1	2
Group 10 (3 isolates)	3	3	3	3	3	3	3	3	1	1	3

Table 9 gives a summary of the total number of different genotypes seen at each locus as well as the total number of genotypes seen within each of the ten major population groups. As observed previously (Chapter 3) some loci i.e. locus 5-6 and some half-loci, detected greater degrees of variation (Table 9) and the use of multiple loci allowed differentiation of a majority of the samples (Table 8).

GROUP 1

Analysis of the genotype data generated from the ten samples included in this group reveals two notable features.

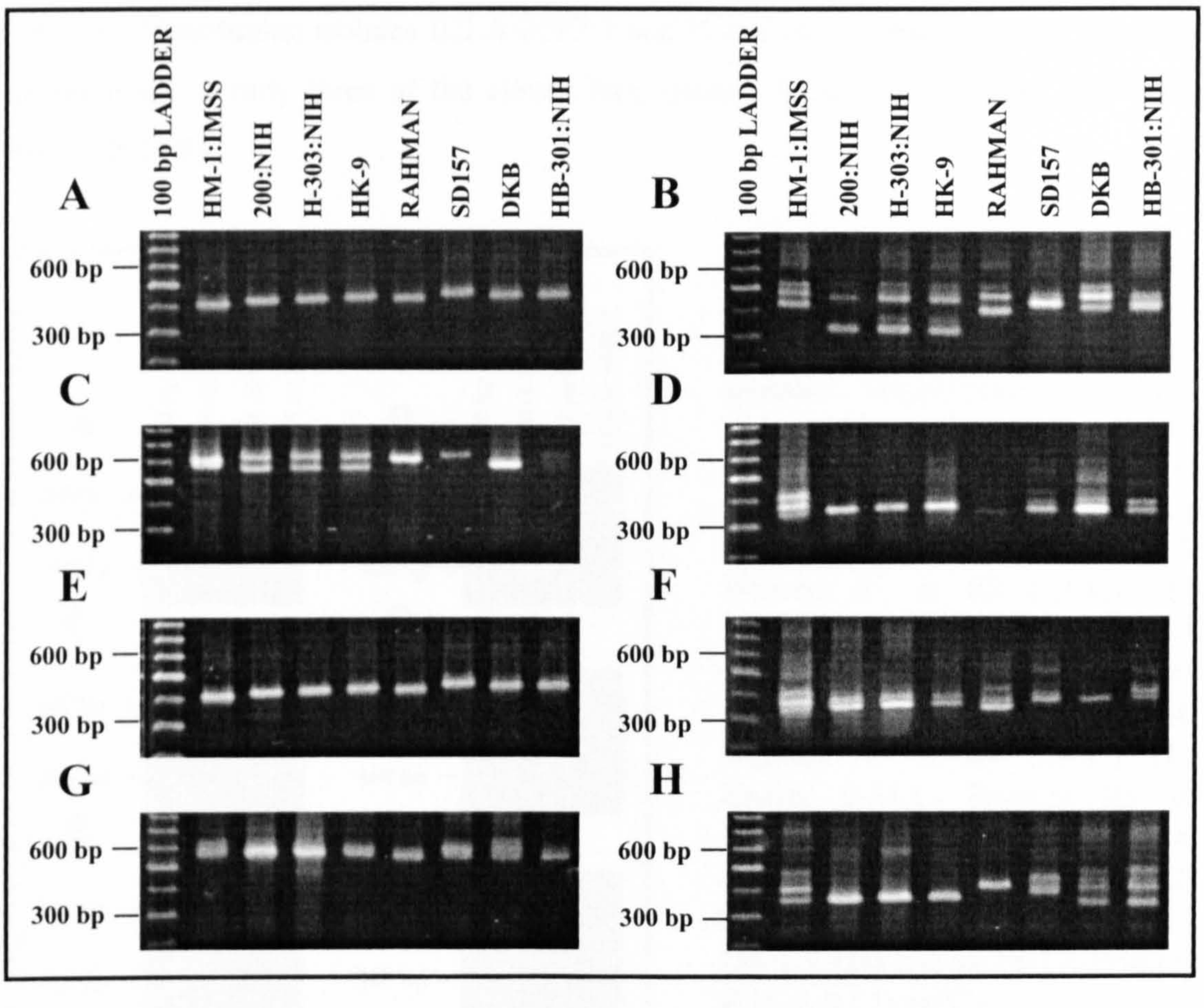


Fig. 13. Polymorphic DNA Analysis of Group 1 *E. histolytica* Isolates (ATCC strains). Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 9-11A.** Primers R9 & R11A (53°C). **(F) Locus 10-4.** Primers R10 & R4 (50°C). **(G) Locus 16-19.** Primers R16 & R19 (54°C). **(H) Locus 18-17.** Primers R18 & R17 (54°C).

First that the pattern of strain HM-1:IMSS is unique not only when compared to those of the other samples included in this group but in fact among all eighty-four samples in the survey (Fig. 13; Table 8). Secondly, the patterns observed for strains NIH:200, H-303:NIH and HK-9 are identical at all eleven loci, eight of which are shown in Fig. 13.

In contrast the four isolates Rahman, SD157, DKB and HB-301:NIH display very diverse patterns and can be distinguished from each other at three loci, namely locus 5-6, 3-8 and 16-17 (Fig. 13B & 13C; Table 8) and from the rest of the group at locus 16-17 (Table 8).

The two Venezuelan isolates IULA:0593:2 and IULA:1092:1 had identical genotype assignments at only three of the eleven loci, namely locus 1-2, 9-4 and 16-17 (Fig 14A; Table 8).

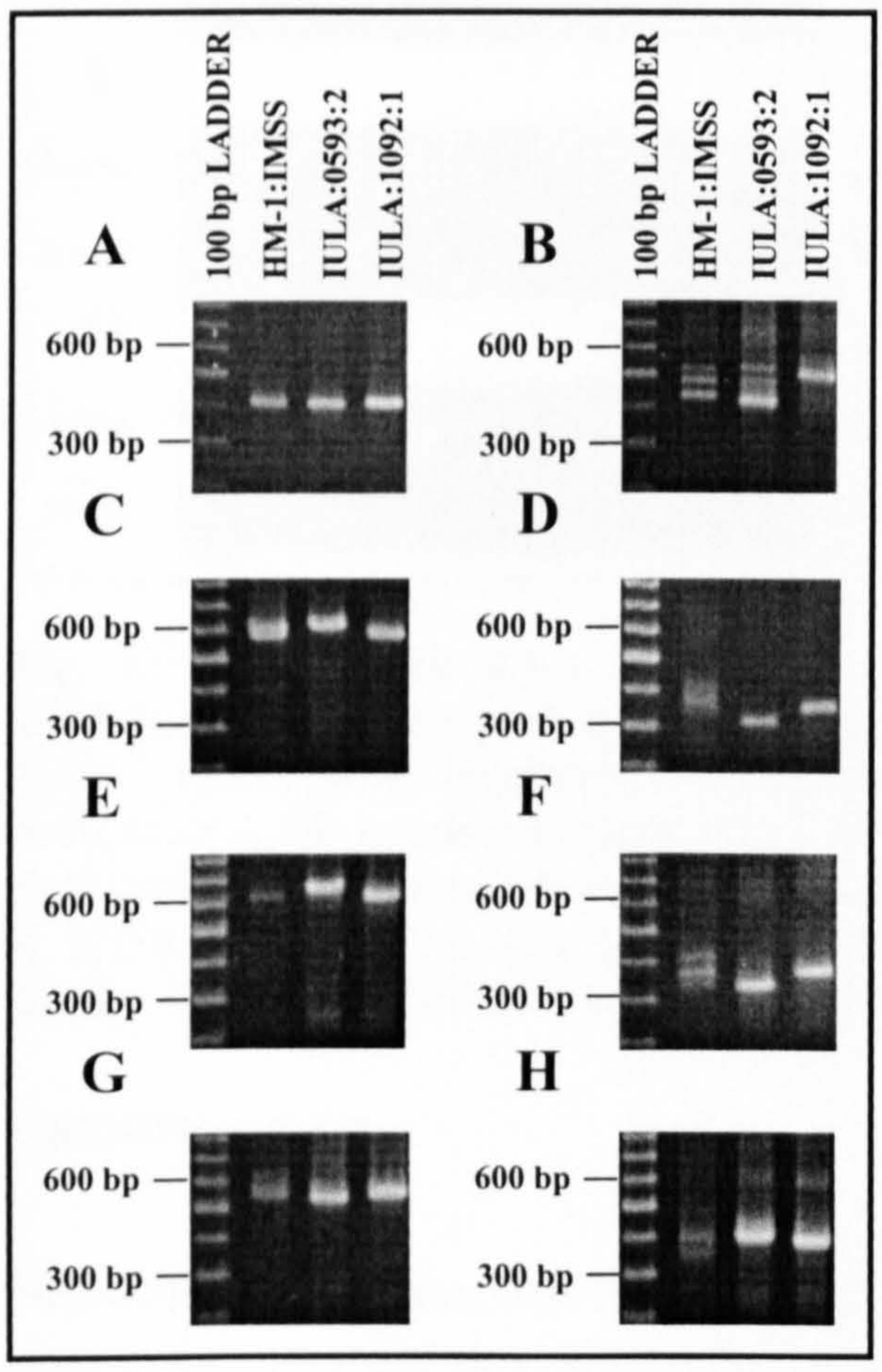


Fig. 14. Polymorphic DNA Analysis of Group 1 (Venezuela) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 9-11A.** Primers R9 & R11A (53°C). **(F) Locus 10-4.** Primers R10 & R4 (50°C). **(G) Locus 16-19.** Primers R16 & R19 (54°C). **(H) Locus 18-17.** Primers R18 & R17 (54°C).

GROUP 2

The ten Bangladeshi isolates displayed highly divergent patterns, with locus 1-2 showing only four genotypes and hence being the least divergent (Fig. 15A; Table 9). A total of nine distinct patterns are visible among the ten isolates at locus 5-6 (Fig 15B; Table 9).

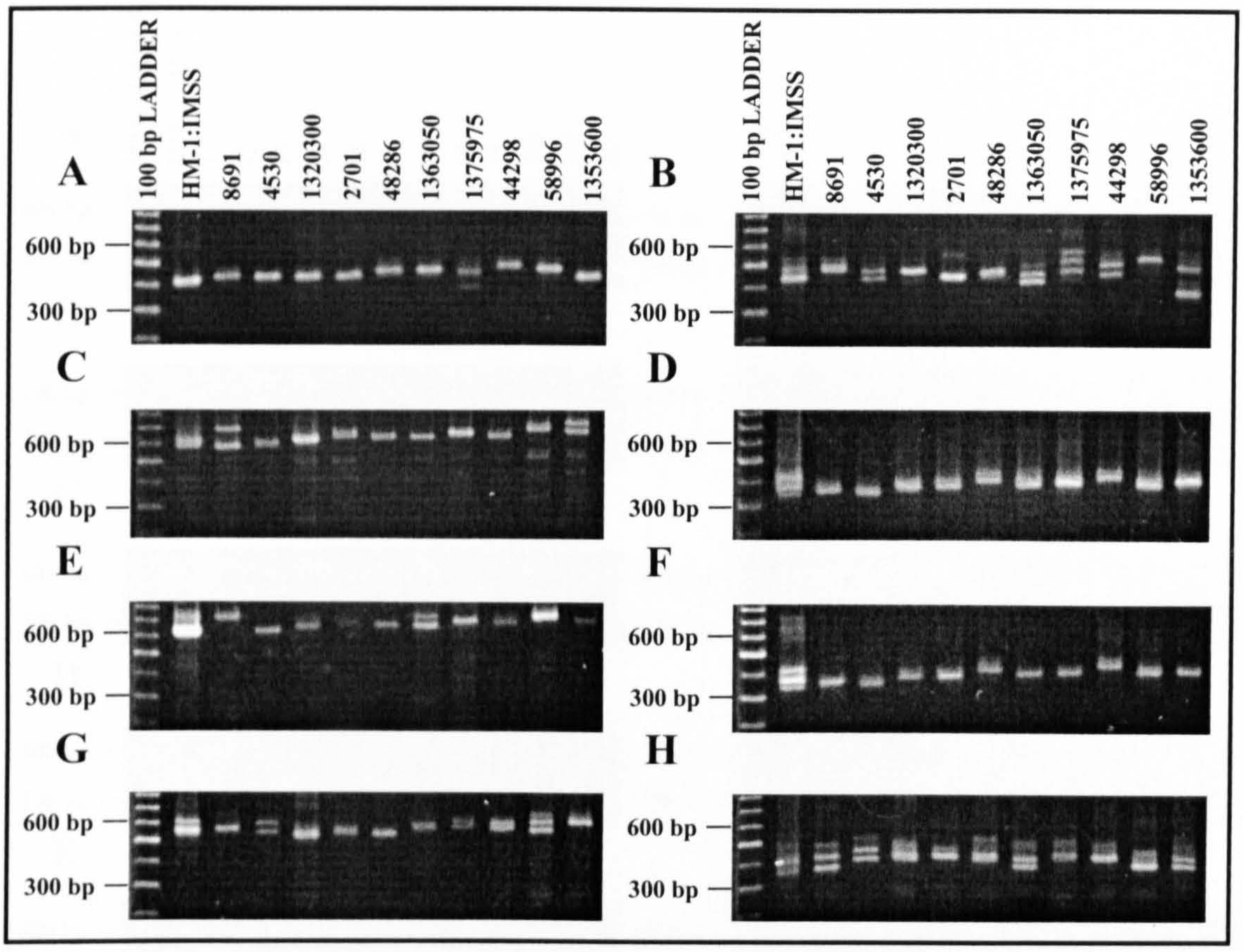


Fig. 15. Polymorphic DNA Analysis of Group 2 (Bangladesh) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 9-11A.** Primers R9 & R11A (53°C). **(F) Locus 10-4.** Primers R10 & R4 (50°C). **(G) Locus 16-19.** Primers R16 & R19 (54°C). **(H) Locus 18-17.** Primers R18 & R17 (54°C).

GROUPS 3, 4 & 5

Nine different genotypes were also seen among group 3 (Japanese samples) (Table 9). Two isolates (J13 and J14) shared genotype assignments at all eleven loci (Table 8), eight of which are shown in Fig. 16. Likewise, the five isolates J18 to J22 had

identical patterns at all the loci except at locus 1-2 where isolate J21 was different (Fig. 16A; Table 8). It was subsequently found that isolates J13 and J14 were from an outbreak that occurred at an institution for the mentally handicapped (personal communication). Similarly isolates J18 to J22 also came from an outbreak that occurred at a different institution for the mentally handicapped in the same district.

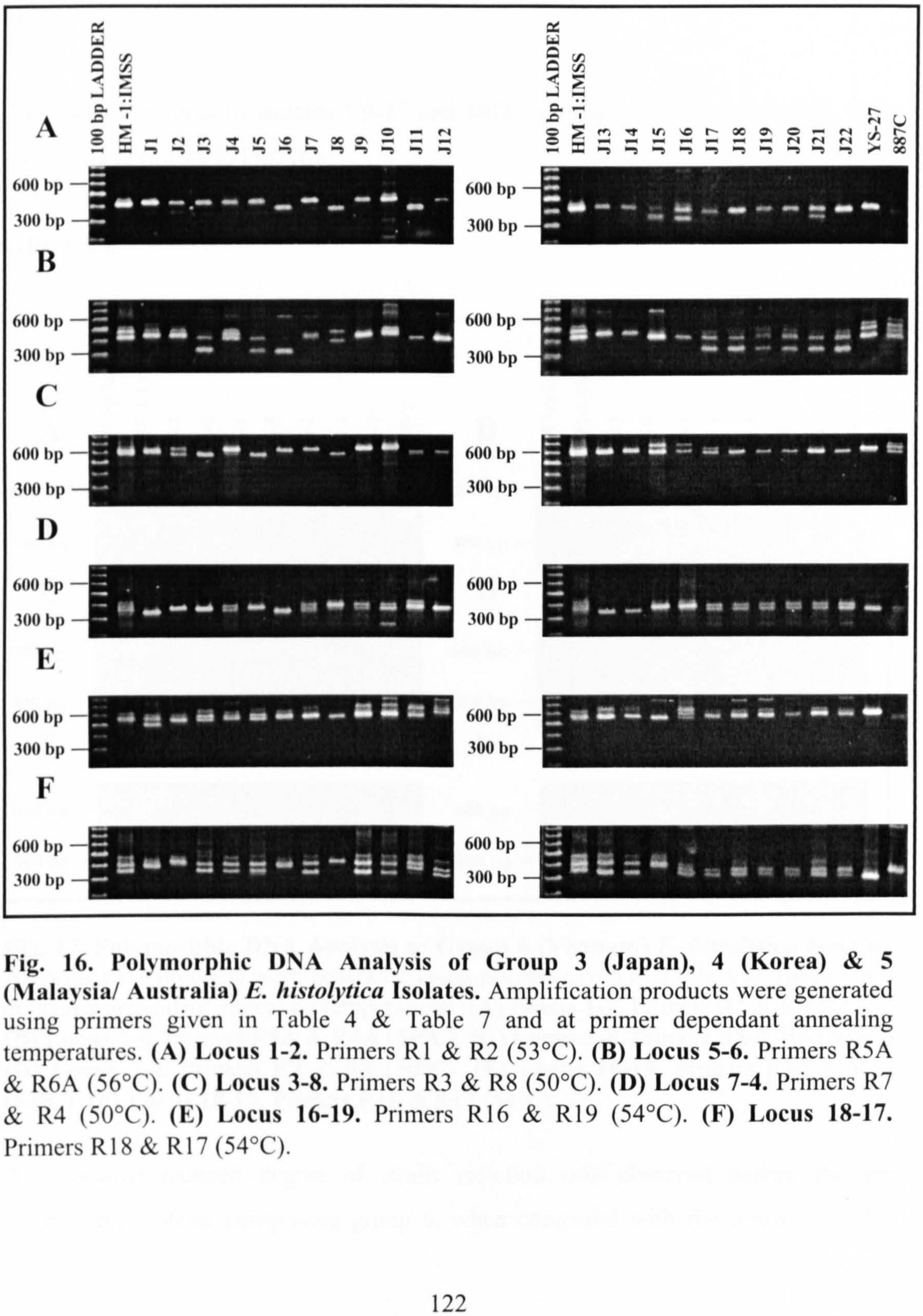


Fig. 16. Polymorphic DNA Analysis of Group 3 (Japan), 4 (Korea) & 5 (Malaysia/ Australia) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

Isolate J17, though apparently not related, showed remarkable similarity to samples J18 to J22 at nine of the eleven loci with differences only evident at locus 3-8 (Fig. 16C) and locus 9-4 (Table 8). Five of the samples, namely J1, J7, J10, J15 and J16, were obtained from homosexual males. Although isolate J7 and J10 share patterns at 6 of the eleven loci and isolates J15 and J16 at seven loci (Table 8; Fig. 16A-D & 16F) all five samples could be distinguished from each other at locus 16-19 (Table 8; Fig. 16E).

The overall patterns of isolates YS-27 and 887C, groups 4 and 5 respectively, were quite unique (Table 8; Fig. 16).

GROUP 6

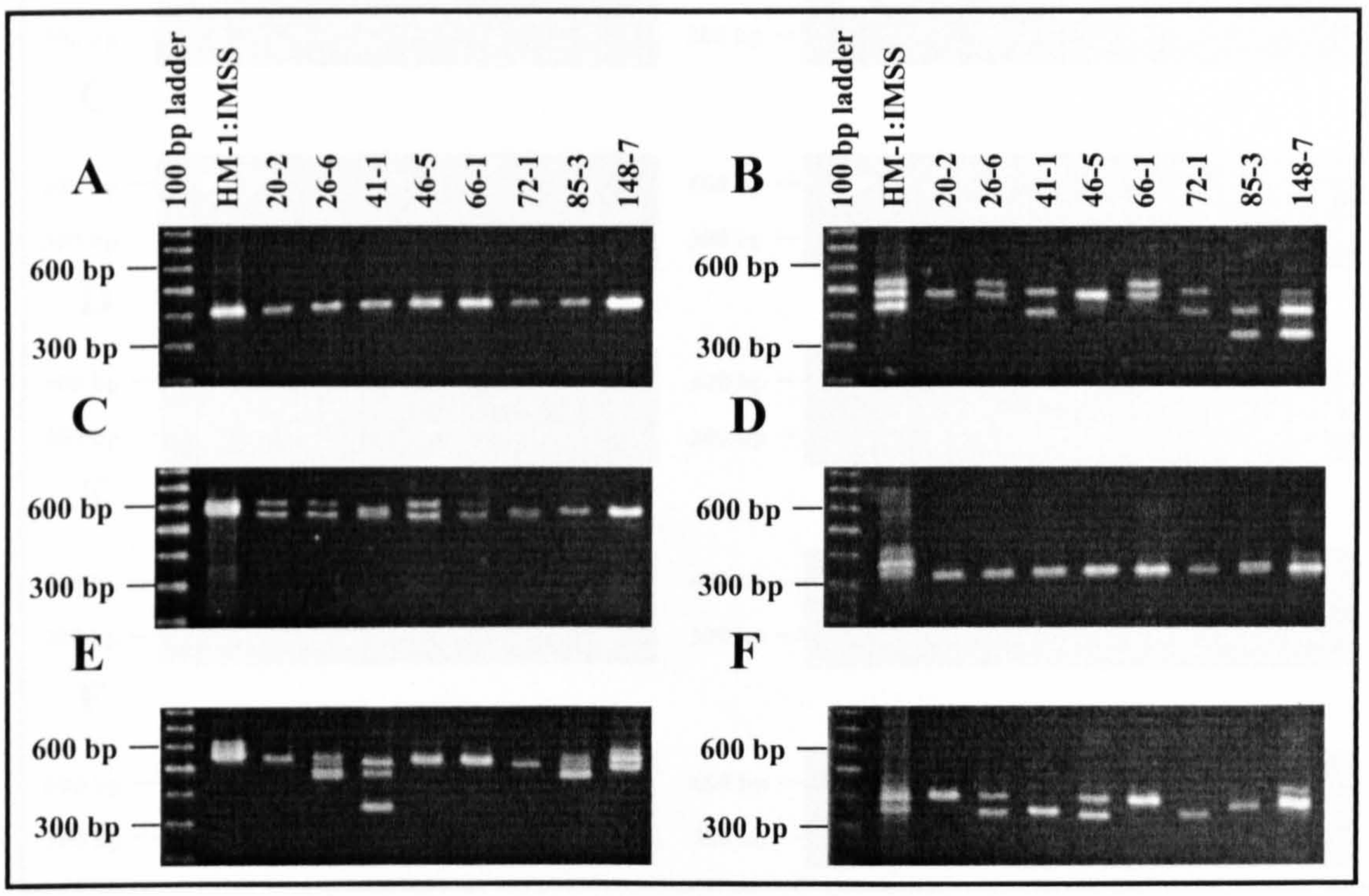


Fig. 17. Polymorphic DNA Analysis of Group 6 (Vietnam) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

A somewhat reduced degree of strain variation was observed among the ten Vietnamese isolates comprising group 6, when compared with the results seen for

samples in groups 2 and 3 (Table 9). All the isolates shared patterns at locus 1-2 and 7-4 (Fig. 17A & 17D; Table 8) with locus 5-6, 16-19 and 18-17 being the most polymorphic (Fig. 17E & 17F; Table 8).

GROUP 7

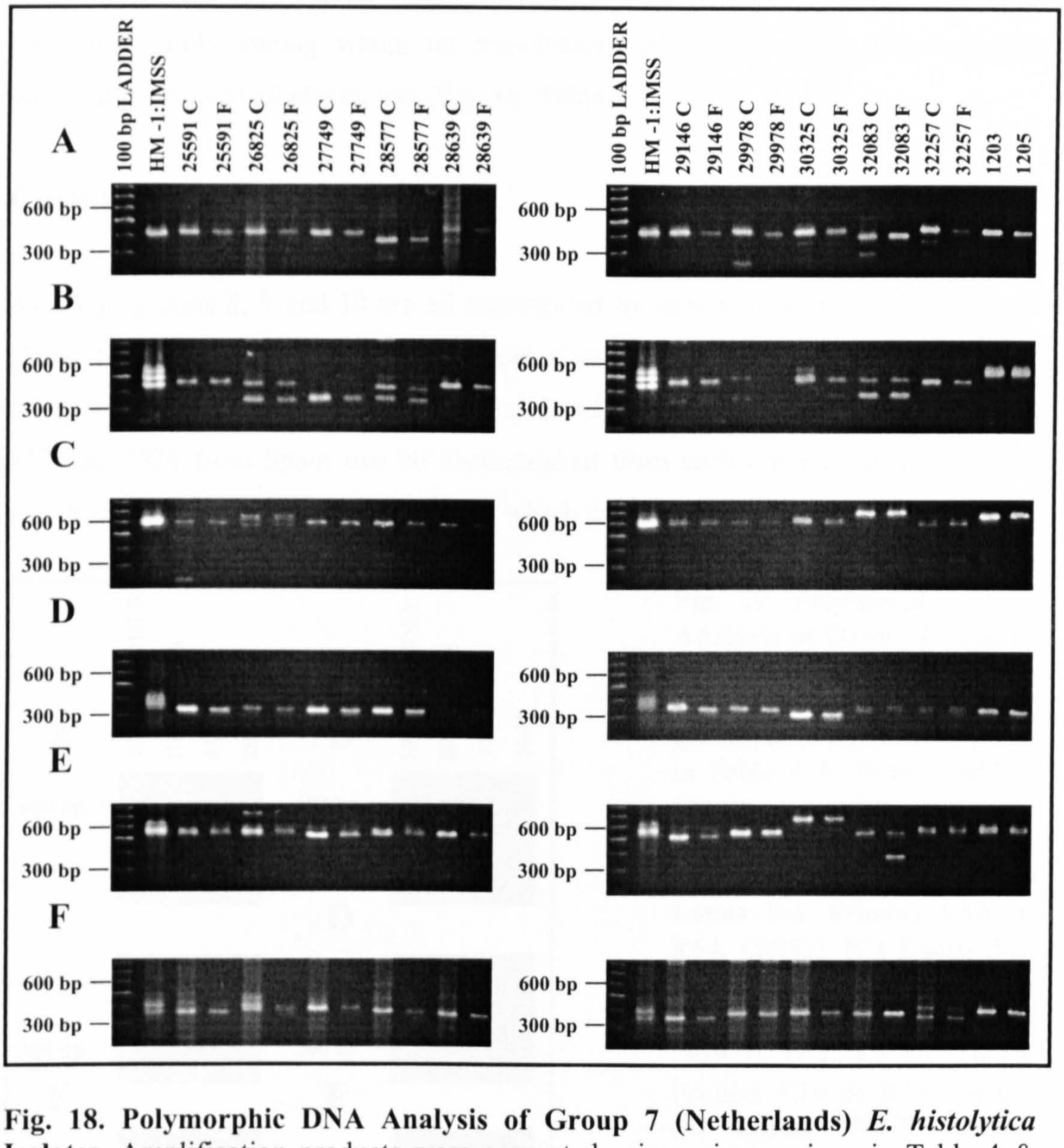


Fig. 18. Polymorphic DNA Analysis of Group 7 (Netherlands) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

Once again analysis of the ten isolates obtained from Holland revealed that locus 5-6 (Fig 18B) and some of the half loci, namely loci 16-19 (Fig. 18E) and 18-17 (Fig.

18F) were the most polymorphic when compared to the remaining loci (Table 8 & 9). It is noteworthy that based on the information provided (Table 6G) these infections appear to have originated from three different continents.

In all cases the results obtained with both culture derived and faecal DNA from the same infection were identical. Additionally, the two samples representing members of the same family among whom an amoebiasis outbreak has been reported gave identical patterns at all eleven loci (Fig. 18; Table 8).

GROUPS 8, 9 & 10

Although groups 8, 9 and 10 are all represented by only two or three samples each (Table 6H, I & J) they display high degrees of strain variation, comparable with those seen among the larger panels of Bangladeshi and Japanese isolates (Table 9). Isolates 822 and 2024 from Spain can be distinguished from each other at all nine loci at which amplification was achieved, six of which are shown (Fig. 19; Table 8).

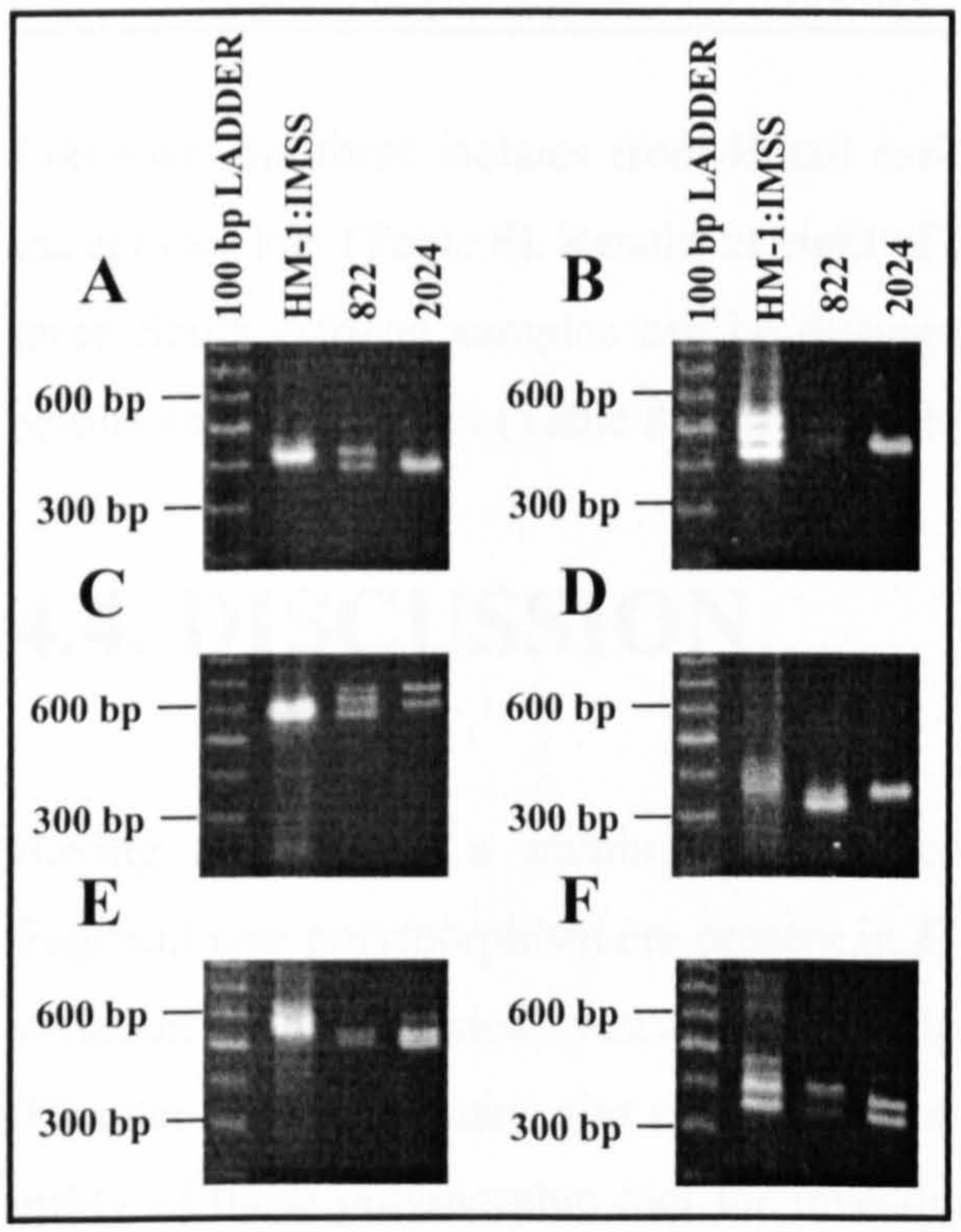


Fig. 19. Polymorphic DNA Analysis of Group 8 (Spain) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

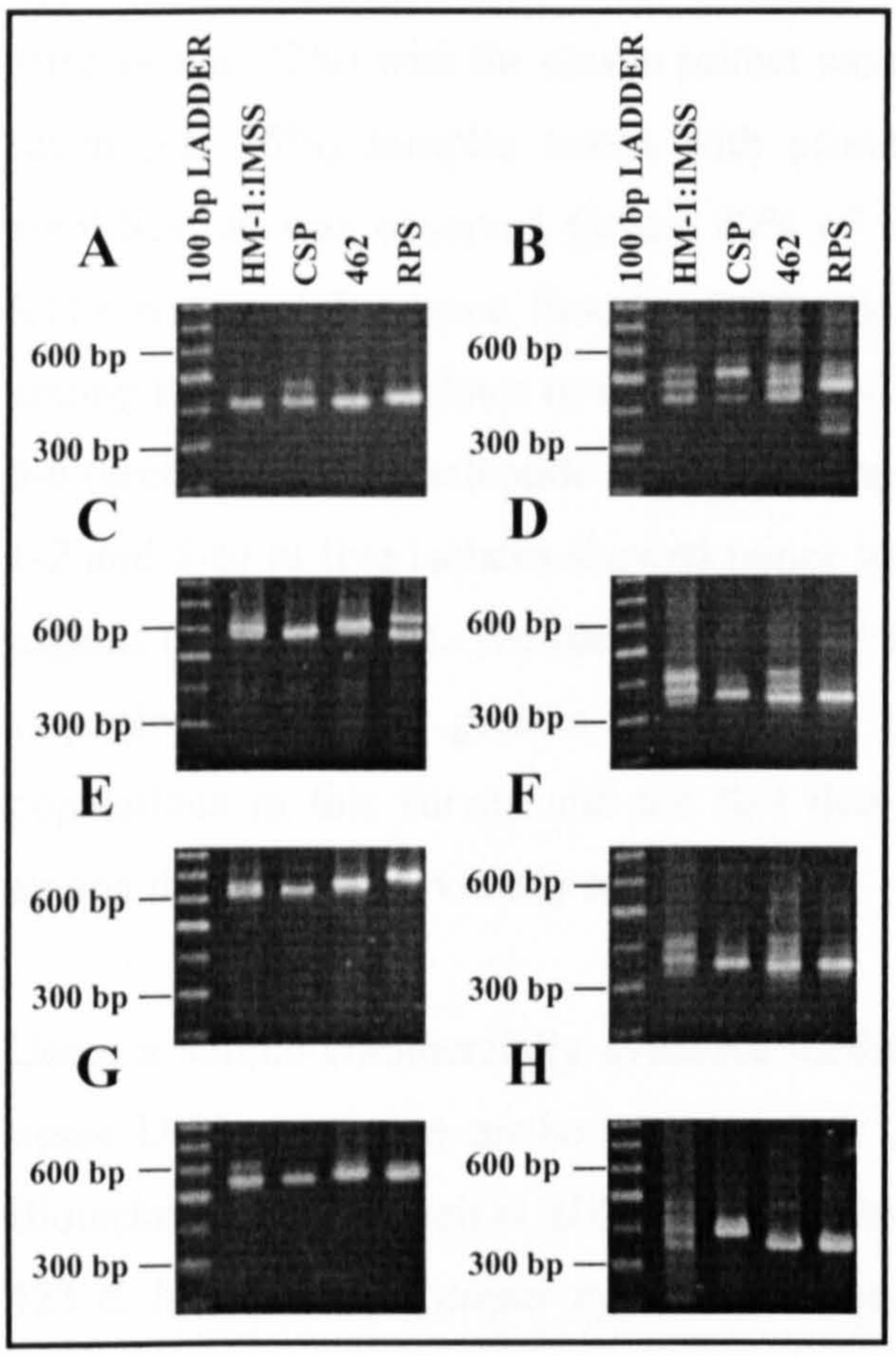


Fig. 20. Polymorphic DNA Analysis of Group 10 (Brazil) *E. histolytica* Isolates. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 9-11A.** Primers R9 & R11A (53°C). **(F) Locus 10-4.** Primers R10 & R4 (50°C). **(G) Locus 16-19.** Primers R16 & R19 (54°C). **(H) Locus 18-17.** Primers R18 & R17 (54°C).

Likewise, the three isolates from Brazil can be differentiated from each other at all except two loci (Table 8). Results at eight of the loci are shown in Fig. 20. Finally, all three South African samples can be distinguished from each other by virtue of the results seen at five loci (Table 8; Chapter 3, Fig. 6B, 9A, 9B, 9C & 10A).

4.4. DISCUSSION

Having shown that a number of repeated-DNA containing loci showing PCR fragment size polymorphism are present in *E. histolytica* and that significant levels of variation can be detected even among samples from geographically restricted regions (Chapter 3), the primary aim of the present analyses was to determine the general utility of these polymorphic loci for investigating the molecular epidemiology of *E. histolytica* in a larger sample.

Our results show that amplification was successful for a majority of the *E. histolytica* samples (ca. 77%) with the eleven primer pairs tested. While only four of the twenty-seven (ca. 15%) samples tested with primers R9 and R11 gave PCR products, amplification was observed for ca. 80% of the samples when the modified primer R11A was used. Evidence, based on PCR amplification, that sequence diversity exists among the thirteen isolates in the primer binding regions has also been seen at locus 5-6 (section 3.3.3). Nucleotide sequence comparison of products from two loci (locus 1-2 and 5-6) in five isolates showed minor sequence variation in the repeat-flanking regions (section 3.3.4). The results seen here for primer 11 are therefore perhaps not surprising given the greater number and diversity of samples and geographic populations in this survey and the fact that majority of them are not represented among the isolates previously sequenced.

Using a simple commercially available method for DNA isolation i.e. the QIAamp tissue DNA extraction method (QIAgen) in conjunction with SuperTaq (HC) (HT Biotechnology), Verweij *et al* (2000) were able to amplify approximately 98% of the 423 *E. histolytica/ E. dispar* cyst positive faecal samples analysed by PCR-SHELA. This method is based on amplification of a repetitive region in the extrachromosomal circular rDNA of *E. histolytica* and *E. dispar*. Only ca. 2% of the faecal DNA samples examined by them contained enough inhibitory factors to completely block amplification. In the present analysis, however, amplification was seen at only half of the loci for the ten faecal DNA samples using SuperTaq (Table 6G) but 100% success was achieved for all the failed samples when SuperTaq was replaced with Hot Start Taq. This suggests that the Hot Start polymerase may be better suited to overcome the effects of inhibitors in DNA extracted from stool by this method.

Amplification of the Vietnamese DNA samples obtained from Germany (Table 6F) was successful using a regular Taq polymerase (Bioline). Although these samples were also extracted using a commercially available kit, the method has been specifically modified for DNA extraction from stool samples and is clearly better equipped for the removal of inhibitors.

Not withstanding the problems encountered in amplifying faecal DNA samples it is apparent from the results presented here that amplification of these polymorphic loci

using DNA isolated directly from stool is possible. Furthermore, as previously observed (section 3.4), the culture media used appears to have no effect on the PCR patterns and the patterns obtained with both culture and faecal DNA from the same clinical samples were identical in all ten cases examined (Group 7; Fig 18).

Based on the high degree of inter-strain variation seen within the small group of samples previously analysed (Chapter 3), it was not surprising to find that between them the eighty-four samples studied in the present survey displayed extensive polymorphism (Table 9). Extensive diversity is seen in particular at locus 5-6 where nine different genotypes (Table 9) are evident among the ten Bangladeshi samples (Table 6B) and there were nine among the twenty-two samples obtained from Japan (Table 6C).

It was noteworthy that although the strain HM-1:IMSS clone 9 genotype at a given locus was shared by one or more isolate, the cumulative pattern displayed was unique among the eighty-four samples studied (Table 8). In contrast isolates 200:NIH, H-303:NIH and HK-9 have identical patterns at all eleven loci. This has already been observed for the first two isolates in Chapter 3 and although nucleotide sequence analysis at locus 5-6 (Fig. 7B; Appendix IV) revealed that the two could be distinguished, the sequence difference may be due to variation among different copies within a strain rather than between different strains (section 3.4). If this is true then it is possible that these three isolates are the same and have been misidentified. No other strains from different countries were found to be identical in this study.

Another interesting observation was the high degree of pattern identity seen between isolates J13 and J14 as well as isolates J18 to J22 which were from separate mass infection cases that occurred at institutions for the mentally handicapped (Fig. 16; Table 8). Several studies have reported on the presence of high rates of amoebic infection amongst severely disabled or mentally retarded, institutionalised individuals (Nagakura *et al*, 1990; Omar *et al*, 1991). It has been suggested that the abnormal hand-to-mouth and object-to-mouth behaviour exhibited by these patients is responsible for the high prevalence. In contrast Gatti *et al* (2000) have observed low infection rates in similar institutions believed to be due to relatively good housing facilities, provision of sanitary conditions, an adequate number of well trained staff,

and good control of the more susceptible subjects. However, in none of these studies could it be determined whether the strain was common among the different infected individuals. Given our results it appears that these polymorphic loci may be useful in tracking transmission of a given strain within a community.

An exception is seen at locus 1-2 where the PCR products for isolate J21 are unique (Fig. 16A) when compared to the those of isolates J18 to J20 and J22. Whether this reflects a different strain or a possible mixed infection with *E. dispar* can not be ascertained in the absence of additional data. Likewise it is difficult to interpret the finding that isolate J17 is identical to isolates J18 to J22 at nine of the eleven loci (Table 8).

Sexual transmission of *Entamoeba* is well documented amongst homosexual men (Jackson, 2000) and although homosexual men in the west have a high prevalence of infection with *E. dispar*, in Japan amoebic infection was first seen in male homosexuals with invasive amoebiasis (Takeuchi *et al*, 1990). Accordingly it was not surprising to find that five of the twenty-two samples received from Japan were from male homosexuals. Although a high degree of pattern identity was seen between samples J7 and J10 (six loci) and likewise between samples J15 and J16 (seven loci), the cumulative results do not suggest that any one *E. histolytica* strain is more prevalent in this community (Table 8). In fact all five isolates could be distinguished at locus 16-19 (Fig. 16E). In a recent study carried out in Japan using loci 1-2 and 5-6 and two polymorphic protein-coding genes i.e. the chitinase gene (Samuelson *et al*, 1997; Ghosh *et al*, 2000) and the gene encoding the serine-rich *E. histolytica* protein (SREHP) (Clark and Diamond, 1993a), it was shown that all nineteen isolates from male homosexuals were unique (Dr. T. Nozaki, personal communication).

Direct person to person transmission has also been implicated as a risk factor for the increased incidence of amoebiasis and the studies carried out by Gathiram and Jackson (1987) as well as those by Rivera *et al* (1998), have demonstrated family clustering of amoebic infections. Samples 1203 and 1205 (Table 6G) in this study, which are from members of a Dutch family among whom an outbreak of amoebiasis was reported (Vreden *et al*, 2000), gave identical genotypes at all eleven loci (Fig 16;

Table 8). These results again highlight the usefulness of these loci at tracking routes of disease transmission.

In summary, *E. histolytica* displays a highly polymorphic strain structure which is evident in all the communities studied here including endemic populations and among samples from individuals living in developed countries. There appears to be no obvious correlation between *E. histolytica* genotypes and different forms of clinical disease. However, it appears from these results that a single *E. histolytica* genotype is often implicated in samples from infected family groups and amoebiasis outbreaks and moreover that these loci can be used to detect common origins of infection by virtue of the shared polymorphisms they display. Subsets of isolates displaying identical polymorphic patterns, as seen by eye, also formed single clades following phylogenetic analysis with no additional, significant clusters being evident (Supplement, Backpocket).

To better establish that these loci can be used to track routes and sources of disease transmission we next wanted to test them with samples obtained from family-based groups in an endemic setting. However, it is now known that even in areas where invasive amoebiasis is common *E. dispar* is by far the more prevalent species (Gathiram and Jackson, 1985). Hence tools that allow accurate and simultaneous differentiation and typing of the two species are clearly needed. We therefore needed to determine whether sequences corresponding to these polymorphic loci were present in *E. dispar* as well.

CHAPTER 5

ISOLATION & CHARACTERISATION OF POLYMORPHIC DNA FROM *E. dispar*

5.1. INTRODUCTION

In the past decade development of tools that allow accurate differentiation of *E. histolytica* and *E. dispar* has been the focus of research of many groups and several methods based on DNA amplification have been reported (section 1.4.7). None of them can, however, detect intra-species variation.

Attempts were made to amplify DNA from *E. dispar* isolates using the polymorphic locus-specific primers described in previous chapters (Table 4; Chapter 3 & 4) to see if corresponding sequences and inter-strain polymorphism could be detected in this species as well. Sequences corresponding to the polymorphic loci were detected in *E. dispar* and the present chapter describes the inter-strain variations seen in this species as well as the design and testing of species-specific primers for two of the loci in both *E. histolytica* and *E. dispar*.

5.2. *ENTAMOEBA* ISOLATES

Table 10 summarises the *E. histolytica* and *E. dispar* samples used in this chapter. The monoxenic *E. dispar* strain SAW760 and the axenic *E. histolytica* strain HM-1:IMSS clone 9 are being maintained in our laboratory (section 2.2).

All the South African isolates were from individuals clinically classified as being asymptomatic and were serology negative using an agarose gel diffusion test, except

for isolate 62.628K which was listed as weakly positive (Appendix XI; T.F.H.G Jackson and S. Reddy, personal communication).

E. histolytica strains HM-1:IMSS, DKB, HB-301:NIH, IULA:0593:2, 887C, J2 and J3 are all from patients who presented with intestinal disease, while the other *E. histolytica* samples come from asymptomatic individuals.

Table 10

ORIGIN OF *ENTAMOEBA* ISOLATES

Isolates	Strain Origin	DNA Origin	*Entamoeba* Species	Source †
16.156N	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
22.211L	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
22.212M	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
29.284N	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
36.352L	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
41.410K	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
49.484L	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
49.485L	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
49.486L	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
50.503I	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
59.595K	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
62.628K	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
88.881H	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
94.943I	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
99.996G	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
11691	South Africa	Lysate (xenic)	*E. dispar*	A (a*)
A1	Columbia	DNA (xenic)	*E. dispar*	B (b*)
SAW 760	England	Culture (monoxenic)	*E. dispar*	C ($a^{\#}$)
HM-1:IMSS Clone 9	Mexico	Culture (axenic)	*E. histolytica*	C ($c^{\#}$)
DKB	England	Culture (axenic)	*E. histolytica*	D ($c^{\#}$)
Rahman	England/ India	Culture (axenic)	*E. histolytica*	D ($a^{\#}$)
HB-301:NIH	Burma	Culture (axenic)	*E. histolytica*	D ($c^{\#}$)
IULA:0593:2	Venezuela	Culture (axenic)	*E. histolytica*	E (c*)
887C	Malaysia/ Australia	DNA (xenic)	*E. histolytica*	F (d*)
RPS	Brazil	DNA (axenic)	*E. histolytica*	G (b*)
462	Brazil	DNA (axenic)	*E. histolytica*	G (b*)
J2	Japan	Lysate (axenic)	*E. histolytica*	H (e*)
J3	Japan	Lysate (axenic)	*E. histolytica*	H (e*)
J5	Japan	Lysate (axenic)	*E. histolytica*	H (b*)
66-1	Vietnam	DNA (faecal)	*E. histolytica*	I (f*)

Note: "†" A: T.F.H.G Jackson and S. Reddy, Medical Research Council Durban, South Africa. Robinson's medium. **B**: A. Aguirre, London School of Hygiene and Tropical Medicine. Robinson's medium. Purified DNA provided. **C**: This Laboratory. LYI-S-2 medium. **D**: D. Nolder, London School of Hygiene and Tropical Medicine. TYI-S-33 medium. **E**: J.P. Ackers and A. Shire, London School of Hygiene and Tropical Medicine. YI-S medium. **F**: J. Williams, London School of Hygiene and Tropical Medicine. Robinson's medium. Purified DNA provided. **G**: M. A. Gomes,

Universidade Federal de Minas Gerais, Brazil. TYI-S-33 medium. Purified DNA provided. **H:** S. Kobayashi and T. Takeuchi, Keio University, Japan. TYI-S-33 medium. **I:** E. Tannich, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. Purified DNA provided. Extracted using the Qiagen stool DNA extraction kit.

Clinical Diagnosis: (*) as provided by supplier of the sample except for (#) where the information was obtained from the catalogue of the American Type Culture Collection (www.atcc.org). **a:** Asymptomatic cyst passer. **b:** Asymptomatic. **c:** Amoebic dysentery. **d:** Diarrhoea. **e:** Colitis. **f:** Asymptomatic carrier.

5.3. RESULTS & DISCUSSION

5.3.1. PCR AMPLIFICATION OF REPEATED-DNA CONTAINING LOCI FROM *E. dispar*

The aim was to see if the intra-species PCR fragment size polymorphism described among *E. histolytica* isolates (Chapter 3 & 4) could be observed among *E. dispar* samples and also whether the primers used were species-specific or not. To this end, primers previously described from *E. histolytica* (Table 4) were used to attempt amplification of the corresponding loci in *E. dispar* strain SAW760. Amplification products were obtained at all loci and half loci and were compared to those of *E. histolytica* strain HM-1:IMSS (Fig. 21). Although the patterns appear in general to be similar for both strains, at loci 1-2, 3-4, 9-4 and 16-19 the SAW760 products are ca. 50 to100 bp larger.

Following successful amplification using SAW760, DNA from xenic isolates of *E. dispar* was tested. A total of one hundred and eleven South African isolates, characterised as *E. dispar* on the basis of zymodeme analysis, were available (Appendix XI). Initially a randomly selected group of some ten to twenty DNA samples were tested at all loci. Very few samples gave amplification products for loci 5-6, 3-8, 7-4, 9-11, 10-4, 16-17 and 18-17. At locus 5-6, nine of the initial set of twenty DNA samples tested gave amplification products but the approximately fifty samples tested thereafter all failed to amplify. In all, amplification at locus 5-6 was

attempted for seventy-seven of the one hundred and eleven samples before the analyses was discontinued. It was, however, notable that a high degree of PCR product size variation existed among the nine positive samples. At least four different patterns could be distinguished (data not shown). Amplifications at the remaining loci were either negative or gave non-specific banding patterns.

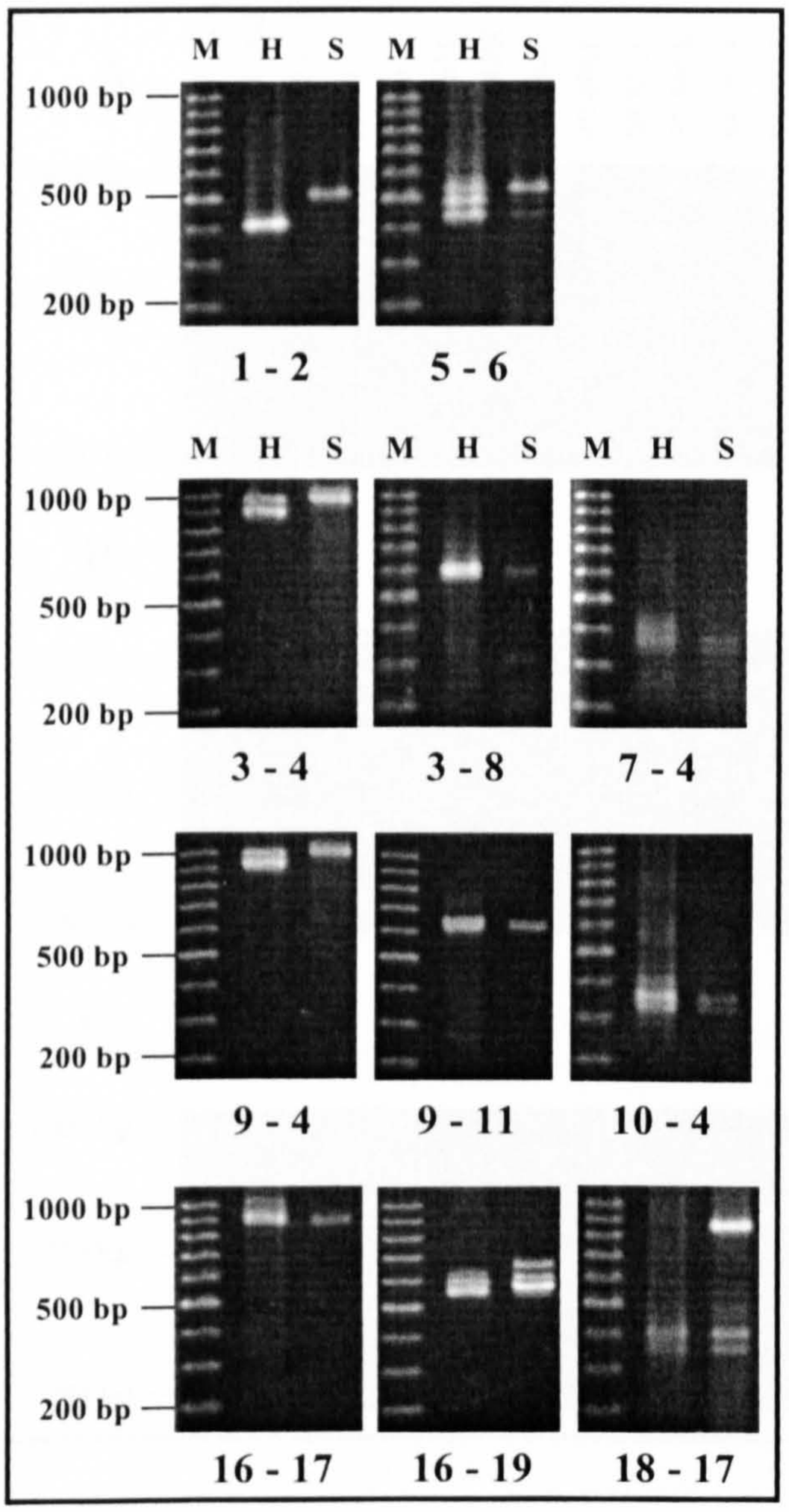


Fig. 21. Electrophoretic Comparison of Strain HM-1:IMSS Clone 9 & SAW 760 PCR Amplification Products at Eleven Loci. *E. histolytica* strain HM-1:IMSS **(H)** and *E. dispar* strain SAW760 **(S)**. The 100 bp DNA ladder is the size marker **(M)**. Amplification products were generated using primers given in Table 4 and at primer dependant annealing temperatures. **Locus 1-2.** Primers R1 & R2 (53°C); **Locus 5-6.** Primers R5A & R6A (56°C); **Locus 3-4.** Primers R3 & R4 (55°C); **Locus 3-8.** Primers R3 & R8 (50°C); **Locus 7-4.** Primers R7 & R4 (50°C); **Locus 9-4.** Primers R9 & R4 (55°C); **Locus 9-11.** Primers R9 & R11 (50°C); **Locus 10-4.** Primers R10 & R4 (50°C); **Locus 16-17.** Primers R16 & R17 (55°C); **Locus 16-19.** Primers R16 & R19 (54°C); **Locus 18-17.** Primers R18 & R17 (54°C).

Based on these results, PCR amplification was carried out on all one hundred and eleven DNA samples at loci 1-2, 3-4, 9-4 and half-locus 16-19. In the final analysis, DNA from eleven of the one hundred and eleven samples failed to amplify at any of these four loci. Each of the remaining one hundred samples produced amplification products with one or more of the four primer pairs, with 85 to 95% of the samples

being positive for any one of the four loci. The products displayed size polymorphism at all four loci. Results obtained for twelve representative samples at locus 1-2, 9-4 and 16-19 are shown in Fig. 22. The fragment sizes displayed by most of the isolates are similar to those seen with SAW760 at the same locus.

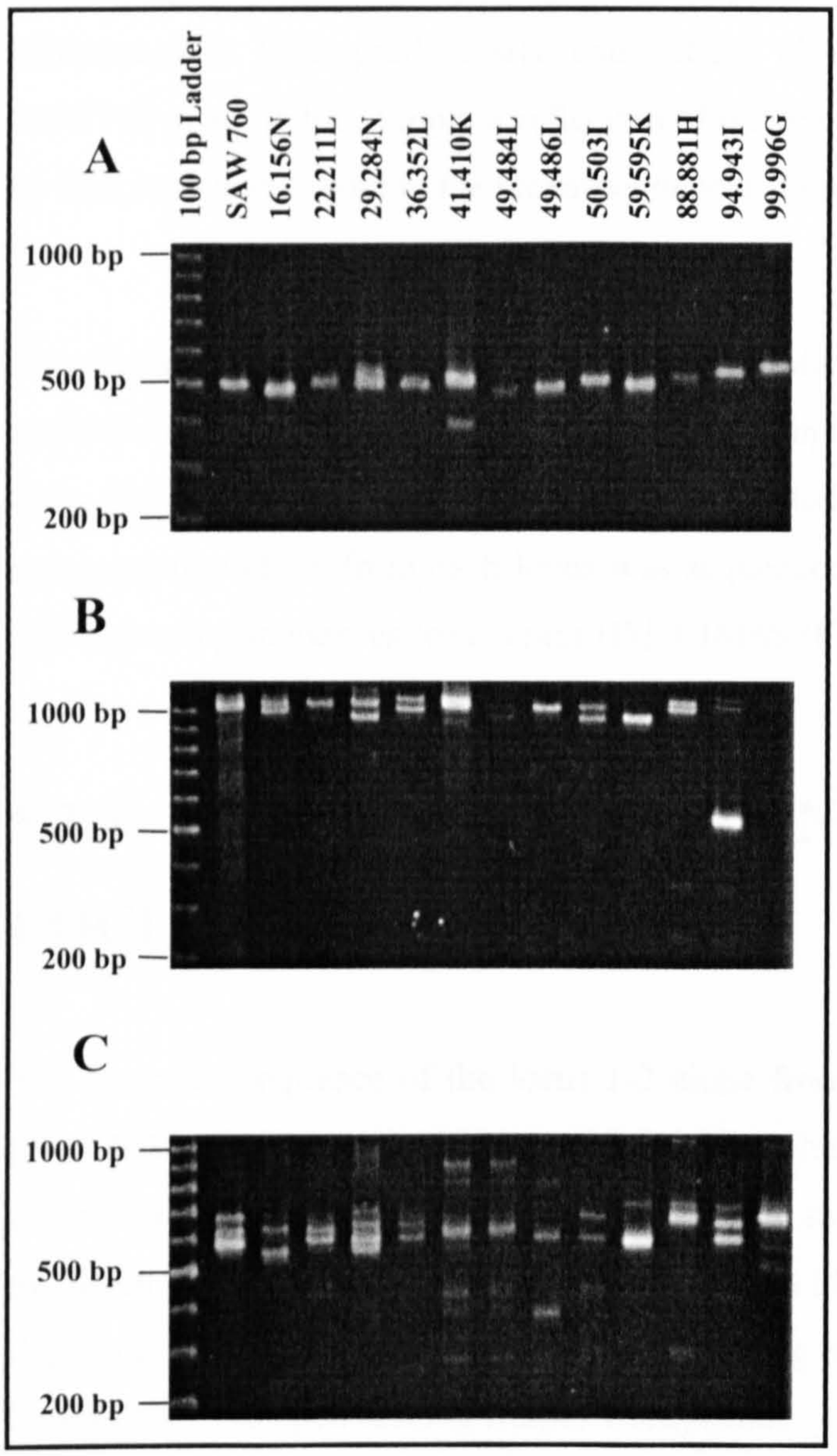


Fig. 22. Polymorphic DNA Analysis of *E. dispar* Isolates. (A) Locus 1-2. Amplification products were generated using primers R1 and R2 at an annealing temperature of 53°C. **(B) Locus 9-4.** Amplification products were generated using primers R9 and R4 at an annealing temperature of 55°C. **(C) Locus 16-19.** Amplification products were generated using primers R16 and R19 at an annealing temperature of 54°C.

5.3.2. ISOLATION OF REPEATED-DNA CONTAINING LOCI FROM *E. dispar*

The relative ease with which some of the primer pairs gave amplification products suggests that these oligonucleotides are derived from sequences which are conserved

among the sister species *E. histolytica* and *E. dispar*. On the other hand, the failure of other primer pairs suggests that these lie in regions of sequence divergence.

This raises a number of questions. What is the degree to which the sequences differ between species? Do the differences exist only in the repeat-flanking regions? Are the differences in PCR product size and pattern observed between HM-1:IMSS and SAW760 at some loci simply a reflection of different numbers of tandem repeat units or does variation extend to the repeat sequences themselves?

To address these questions the amplification products of strain SAW760 at loci 1-2 and 5-6 were cloned. A total of ten transformants in the case of locus 1-2 and eight for locus 5-6 containing cloned inserts in the expected size range were identified. One representative clone from each locus was sequenced (section 2.10) and compared to corresponding sequences from strain HM-1:IMSS (Fig. 4A & 5A).

5.3.3. NUCLEOTIDE SEQUENCE ANALYSIS AT LOCI 1-2 & 5-6

The complete sequence of the locus 1-2 clone from strain SAW760 is 495 bp long (Fig. 23A). There are two main repeat blocks which, between them, display seven distinct but highly related direct repeats arranged in tandem (Fig. 23B). Only one of the seven repeat types is represented in both repeat blocks. In addition, several tandem duplications of 5 to 9 bp are seen in the 5' flanking regions of the major repeat blocks (not shown). No open reading frames were found.

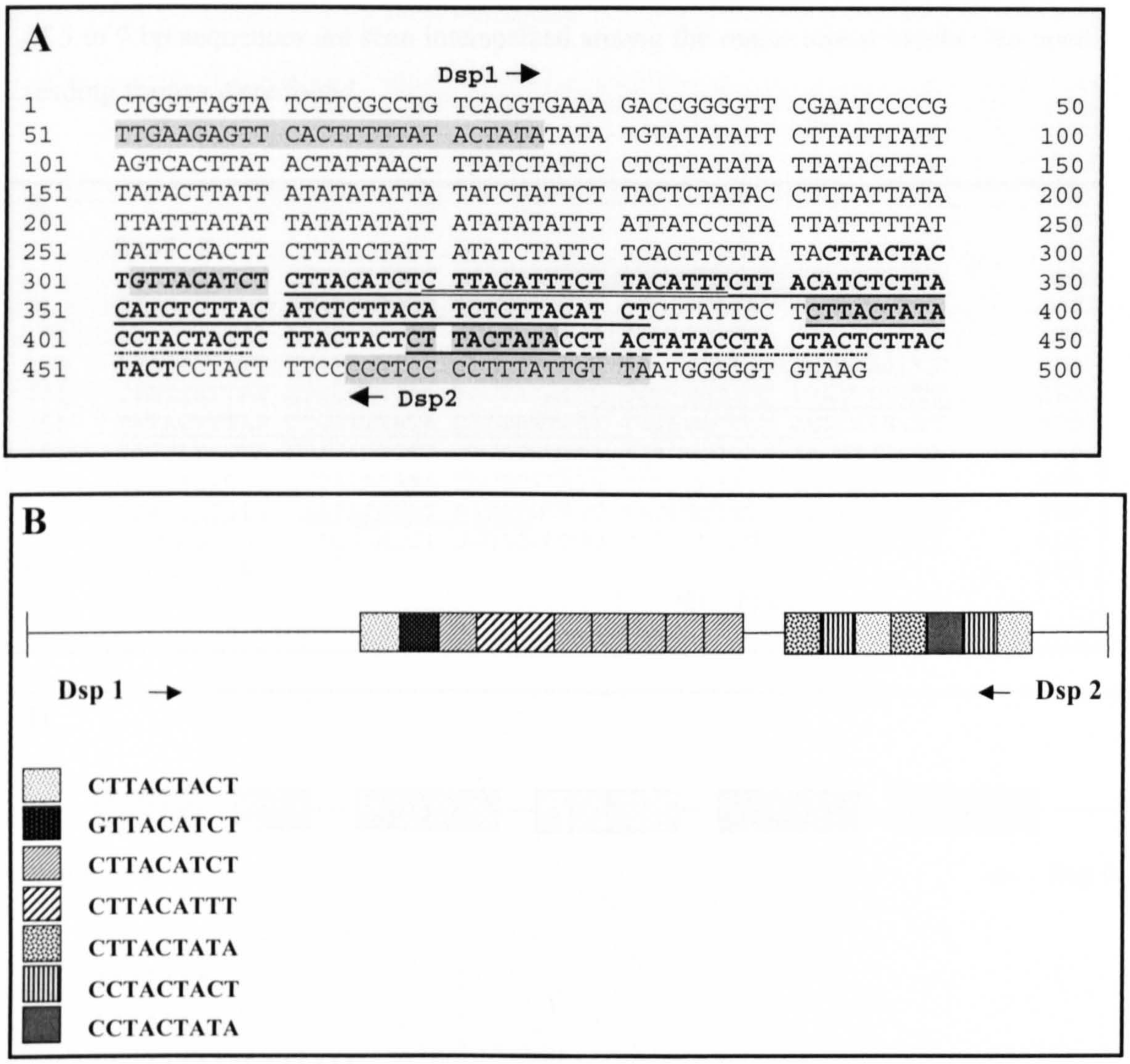


Fig. 23. Locus 1-2 from Strain SAW 760. (A) Nucleotide Sequence. The main blocks of internal tandem repeats are in bold-face. Underlined and/ or highlighted regions indicate the seven different types of repeat units. The nucleotide sequence data has been submitted to the GenBank database with accession number AY058216. **(B) Schematic Representation.** The different types of internal tandem repeats and their arrangements with respect to each other are shown. Tandem duplications in the flanking regions are not shown. Positions of species-specific amplification primers are indicated.

The complete sequence of the locus 5-6 clone from strain SAW760 is 510 bp long (Fig. 24A) and consists of six major types of tandemly arranged repeats organised into five main blocks (Fig. 24B). Block one consists of a set of three and two repeats of the same sequence separated by 41 bp in which a single copy of the same repeat is found (Fig. 24A, nucleotide position 90 to 96). In contrast to the interspersed arrangement of repeats seen at locus 1-2, four of the five blocks seen in locus 5-6

consist of one repeat unit type exclusively. Here too duplications and solitary copies of 5 to 9 bp sequences are seen interspersed among the major repeat blocks. No open reading frames were found.

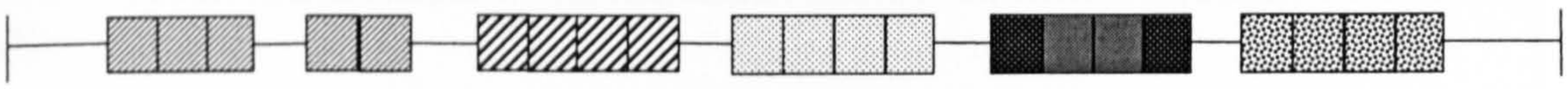
A

Dsp5 →

```
1     CTAAAGCCCC CTTCTTCTAT AATTAATCTA TACTATATTC TTTTTATGTA      50
51    CTTCCCTTAT ATATTATATA TTATATATTA TTATATTTAT TATATATATT     100
101   ATATGTATAT ATTCTACTTT ATATATTATA TATTATATCC TTTATATTCT     150
151   ATATTCTATT TTATATTCAT GACTTATACT TATACTTATA CTTATATCCT     200
201   TTATTATTAT ATTCCTTTAT TACTATATTC CTTTTATACT TTATTCCTTT     250
251   TATACTTTAT TCCTTTTATA CTTTATTCCT TTTATACTTT ATTCCTTTAT     300
301   TACTATATTA CTATATTCTT TTTATACTTT ATTCCTTTAT TACTTTATTA     350
351   CTTTATTCCT TTTATACAAA TACTCTCATG TATGTATATG TATAATATGT     400
401   ATAATATGTA TAATATGTAT AATATGTATG TATTTCTTTA AAAACAATGC     450
451   TCTCAGCAGG TTTCGAACCT GCGACCCTGG CGTTATTAGC ACCATGCTCT     500
501   ACCGACTGAG                                                550
```

← Dsp6

B



Dsp 5 → ← Dsp 6

TTATATA

CTTATA

CCTTTTATACTTTATT

CTTTATTC

CTTTATTA

GTATAATAT

Fig. 24. Locus 5-6 from Strain SAW 760. (A) Nucleotide Sequence. The main blocks of internal tandem repeats are in bold-face. Underlined and/ or highlighted regions indicate the six different types of repeat units. The nucleotide sequence data has been submitted to the GenBank database with accession number AY058217. **(B) Schematic Representation.** The different types of internal tandem repeats and their arrangements with respect to each other are shown. Tandem duplications in the flanking regions are not shown. Positions of species-specific amplification primers are indicated.

5.3.4. NUCLEOTIDE SEQUENCE COMPARISON AT LOCI 1-2 & 5-6

When the sequences of loci 1-2 and 5-6 from strain SAW760 (Fig. 23A & 24A) were compared to those of strain HM-1:IMSS (Fig. 4A & 5A), significant differences in the number, sequence and arrangement of the repeats between the two were revealed.

A

```
Ed:   1    CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCG      50
Eh:   1    CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCG      50

Ed:   51   TTGAAGAGTTCACTTTTTATACTAT-ATATATGTATAT------------      87
Eh:   51   TTGAAGAGTTCTCTTTTTATACTTTTATATGTTTATAT------------      88

Ed:   446  CTTACTACTCCTACTTTCCCCCTCCCCTTTATTGTTAATGGGGGTGTAAG      495
Eh:   353  CTTACTACTCCTACTTTCACCTCCCTCTTTATTGTTAATGGGGGTGTAAG      402
```

B

```
Ed:   1    CTAAAGCCCCCTTCTTCTATAATTAATCTATACTATATTCTTTTTATGTA      50
Eh:   1    CTAAAGCCCCCTTCTTCTATAATTTATATATTATTCTCTTTGAGACTTAT      50

Ed:   51   CTTCCCTTATATA-----------------------------------      63
Eh:   51   TTCTACTTTATTT-----------------------------------      63

Ed:   417  GTATAA---TATGTATGTATTT--CTTTAAAAACAATGCTCTCAGCAGGT      461
Eh:   325  GTATATTTCTATGTACGTCTTTAACTTTAAAAACAATGCTCTCAGCAGGT      374

Ed:   462  TTCGAACC-TGCGACCCTGGCGTTATTAGCACCATGCTCTACCGACTGAG      510
Eh:   375  TTCGAACCCTGCGACCCTGGCGTTATTAGCACCATGCTCTACCGACTGAG      424
```

Fig. 25. Alignment of Nucleotide Sequences from *E. histolytica* strain HM-1:IMSS (Eh) and *E. dispar* strain SAW760 (Ed). Only the repeat flanking region sequences are shown. Nucleotides belonging to a repeat unit are in bold-face. Single base differences are highlighted. Dashes indicate gaps introduced to optimise alignment. **(A) Alignment at Locus 1-2.** Underlined regions show locations of the two pairs of 5' and 3' species-specific primer sequences (Dsp1/ Hsp1 and Dsp2/ Hsp2) which were used for PCR amplification. **(B) Alignment at Locus 5-6.** Underlined regions show locations of the two pairs of 5' and 3' species-specific primer sequences (Dsp5/ Hsp5 and Dsp6/ Hsp6) which were used for PCR amplification.

Differences were also seen in the 5' and 3' repeat-flanking regions between the two species at both loci. The highest degree of sequence identity was seen between the

first sixty base pairs at the 5' end of locus 1-2 (100%) (Fig. 25A) and between the last seventy-five base pairs at the 3' end of locus 5-6 (98.6%) (Fig. 25B).

The significantly greater number of repeat units probably accounts for the larger products seen in *E. dispar*.

5.3.5. DESIGN & TESTING OF SPECIES-SPECIFIC PRIMERS

With the aim of amplifying species-specific products four sets of primers were designed in the repeat-flanking regions of both loci (Table 11; Fig. 25A & 25B).

Table 11

SPECIES-SPECIFIC OLIGONUCLEOTIDE PRIMERS

Primer	Primer Sequence (5' → 3')
Hsp1	GAG TTC TCT TTT TAT ACT TTT ATA TGT T
Hsp2	ATT AAC AAT AAA GAG GGA GGT
Hsp5	CTA TAA TTT ATA TAT TAT TCT CTT TGA GA
Hsp6	CAT TGT TTT TAA AGT TAA AGA CG
Dsp1	TTG AAG AGT TCA CTT TTT ATA CTA TA
Dsp2	TAA CAA TAA AGG GGA GGG
Dsp5	CTA TAC TAT ATT CTT TTT ATG TAC TTC CC
Dsp6	CTG AGA GCA TTG TTT TTA AAG AA

Amplification of strain HM-1:IMSS with the *E. histolytica* specific primers Hsp1 and Hsp2 gave the expected product of ca. 340 bp (Fig. 26A). No amplification was seen with strain SAW760. Similarly, strain SAW760 shows the expected product of ca. 430 bp with the *E. dispar* specific primers Dsp1 and Dsp2 and strain HM-1:IMSS failed to amplify with these primers. At locus 5-6 too, both the *E. histolytica* specific primers Hsp5 and Hsp6, as well as the *E. dispar* specific primers Dsp5 and Dsp6 gave the expected products of ca. 350 bp with strain HM-1:IMSS and 430 bp with strain SAW760 respectively and specifically (Fig. 26C).

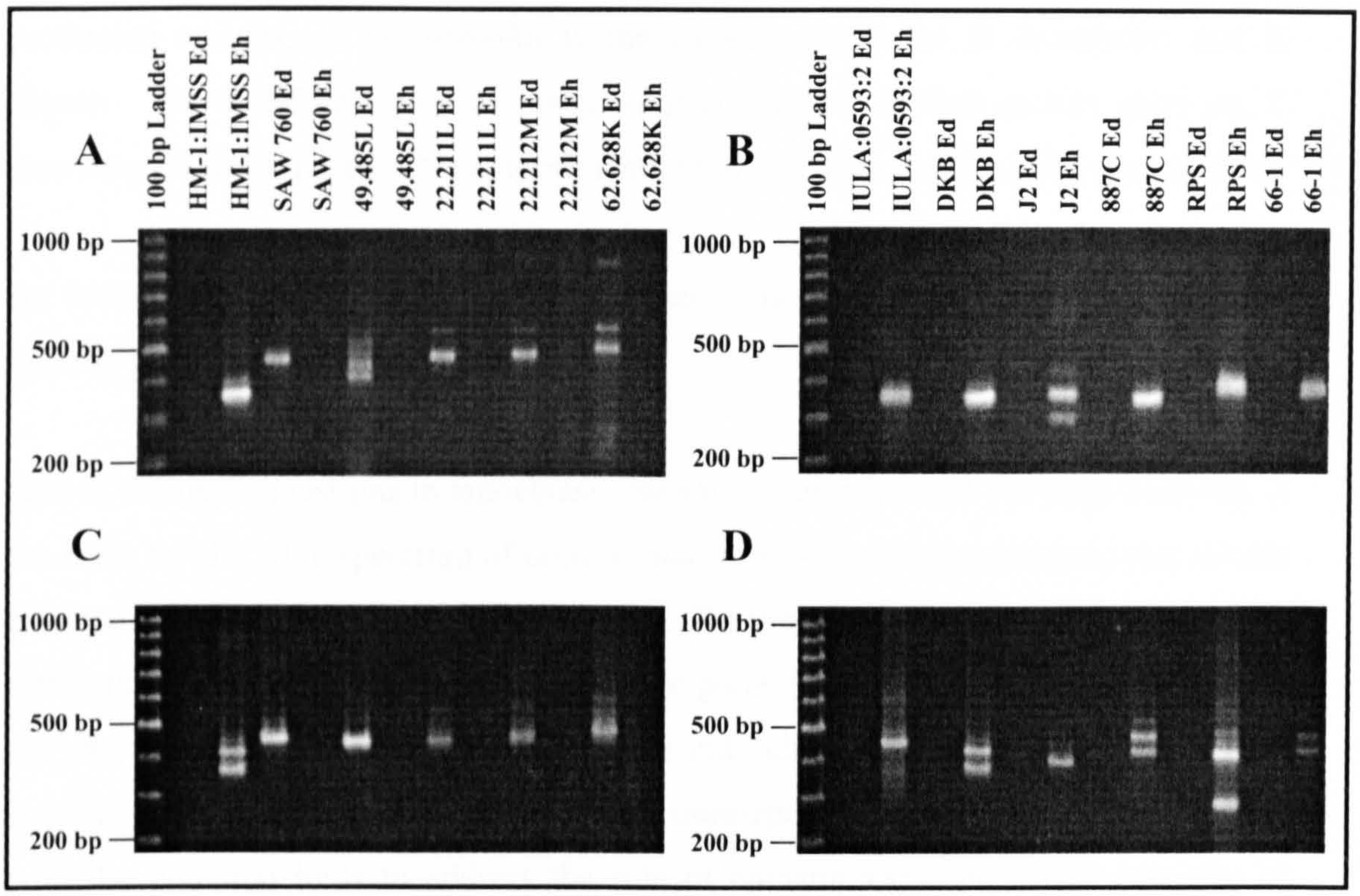


Fig. 26. Species-Specific DNA Analysis of *E. dispar* and *E. histolytica* Isolates. (A & B) Locus 1-2. Amplification products were generated using primers Dsp1 & 2 (Ed) and primers Hsp1 & 2 (Eh) at annealing temperatures of 50°C. **(C & D) Locus 5-6.** Amplification products were generated using primers Dsp5 & 6 (Ed) at an annealing temperature of 52°C and primers Hsp5 & 6 (Eh) at an annealing temperature of 48°C

Species-specificity of these primers was then tested on seventeen additional isolates, eleven of which had previously been characterised as *E. histolytica*. PCR amplification results for 10 of these isolates (Table 10) are shown for both locus 1-2 (Fig. 26A & B) and locus 5-6 (Fig. 26C & D). The presence or absence of PCR amplification products conformed to the original species classification in all cases.

The *E. histolytica* isolates tested came from a wide geographical range and those of *E. dispar* from three continents. All of them gave the expected PCR products with the appropriate species-specific 1-2 and 5-6 primers and there was no amplification with the other species-specific primers.

In their study Silberman *et al* (1999) used the small-subunit ribosomal RNA sequences to resolve relationships among the major lineages of genus *Entamoeba*. Their results demonstrated that species producing cysts with the same number of nuclei form monophyletic groups and that within the group of quadranucleated cyst

producing species, *E. moshkovskii* is the closest relative to *E. histolytica* and *E. dispar*. We therefore tested the species-specificity of all four primer pairs on *E. moshkovskii* strain Laredo. No amplification products were detected (data not shown).

As before, the source of the DNA, whether xenic or axenic culture, or extracted directly from stool was irrelevant.

One of the main questions in amoebiasis research, which has not yet been resolved, is the basis for the wide spectrum of clinical manifestations observed among individuals infected with *E. histolytica* and/ or *E. dispar*. The presence of both types of parasite or different strains of *E. histolytica* in the same patient could be one of the reasons for the varying signs and symptoms in infected individuals. The availability of species-specific markers that simultaneously detect intra-species polymorphisms provides us with the potential tools to address the role of parasite variation in the outcome of disease and to investigate patterns of transmission of both parasites.

CHAPTER 6

GENOTYPING OF *E. histolytica* & *E. dispar* IN SOUTH AFRICA

6.1. INTRODUCTION

Recently, a large longitudinal sero-epidemiological survey has been completed in South Africa (Jackson *et al*, 2000). Ninety-three convalescent amoebic liver abscess patients (index cases) and eight hundred and ninety-five of their close associates were recruited and followed for three years to monitor their infection patterns and serological responses. One early observation was that significant numbers of associates of the index cases were or became infected with *E. histolytica,* but most remained asymptomatic and never developed disease before eventually clearing the infection spontaneously. From isoenzyme analysis it was not possible to determine whether the index cases and associates were infected with the same or distinct strains of *E. histolytica*, due to the limited number of reliable zymodemes in the species (Jackson *et al*, 1992; Jackson and Suparsad, 1997).

Results of the global survey using the polymorphic loci (Chapter 4) suggested that a single *E. histolytica* genotype/ strain is often the source of amoebiasis outbreaks in family groups and closely associated individuals and moreover that these loci can help determine the common route and origin of infection by virtue of the shared polymorphisms they display. Furthermore, we have also shown that *E. dispar* strains can be typed by this approach and have developed species-specific primers for this purpose (Chapter 5). We here report the application of this typing method to a large set of samples acquired during the longitudinal survey in South Africa.

6.2. MATERIALS & METHODS

6.2.1. *ENTAMOEBA* ISOLATES

The monoxenic *E. dispar* strain SAW760 and the axenic *E. histolytica* strain HM-1:IMSS clone 9 were used as controls. All the other isolates used in this study were grown in Robinson's medium at the Medical Research Council, Durban, South Africa. These xenic cultures were maintained for short periods only and discontinued as soon as amoebae were sufficient in number to make a lysate. On average, cells were harvested and lysates prepared 5 to 6 days after cultures were initiated, by pelleting culture sediments and lysing the pellets (section 2.3).

A total of one hundred and fifty-eight xenic *Entamoeba* samples were used in this study. One hundred and fifty of these were isolated during the longitudinal sero-epidemiological study, including all forty-seven of the *E. histolytica* isolates (Appendix XII) and a majority (one hundred and three) of the *E. dispar* isolates (Appendix XI). The isolate numbering system is as follows. The study index cases and their families were numbered 1 through 100. Within each group, index cases were assigned the number '0', while family members and associates of the index case were identified with a personal number e.g. family 10 will have index case 10.0 and associates 10.101 through to 10.109. Letters were used to identify samples obtained at three-monthly visits to the laboratory e.g. A (first visit), B (second visit) three months later and so on. In some cases asymptomatic individuals with *E. histolytica* infection were also followed on a monthly basis. These visits within the three monthly interim periods were denoted by a second letter A or B. Thus samples from an individual who had completed three monthly visit J, would be designated JA and JB on the two successive monthly visits. This would be followed by the three monthly visit K and monthly visits KA and KB and so on. Some samples were available in duplicate or triplicate, in which case they were identified by further suffixes (i and ii). Eight *E. dispar* samples came from a cross-sectional survey of school children in Langebaan (Jackson *et al*, 2000).

Samples were subjected to PCR amplification and genotypes were assigned as before.

6.3. RESULTS

6.3.1. PCR PRODUCT SIZE POLYMORPHISM AMONG SOUTH AFRICAN *E. histolytica* ISOLATES

The forty-seven South African *E. histolytica* DNA samples, identified on the basis of zymodeme analysis, were amplified. Six of the forty-seven DNA samples failed to amplify at any locus and were therefore eliminated from the final analysis. Only the forty-one samples that worked are shown in Table 12. Amplification was successful for twenty-nine DNA samples at all the loci while the remaining twelve DNA samples amplified at most but not all of the loci.

Table 12

SOUTH AFRICAN SURVEY - *E. histolytica* GENOTYPES

No	Isolates	Locus 1-2	Locus 5-6	Locus 3-4	Locus 3-8	Locus 7-4	Locus 9-4	Locus 9-11	Locus 10-4	Locus 16-17	Locus 16-19	Locus 18-17
1	HM-1:IMSS	1	1	1	1	1	1	1	1	1	1	1
2	37.0C	1	28	17	6	14	18	10	4	17	21	2
3	39.0L	1	27	3	3	16	7	2	15	11	21	23
4	39.384C	1	27	3	21	16	7	-	12	11	21	20
5	39.384KB	1	27	3	3	16	7	2	15	11	21	23
6	2596	10	33	18	3	3	13	1	12	17	-	20
7	56.0L	1	28	4	5	3	7	1	12	11	3	24
8	63.0I	10	29	3	5	3	16	10	12	11	3	22
9	63.0Ii	10	29	3	5	3	16	10	12	11	3	22
10	63.0IA	10	29	3	5	3	16	10	12	11	3	22
11	63.0JA	10	29	3	5	3	16	10	12	11	3	22
12	63.0JB	13	29	3	5	3	16	10	12	11	3	22
13	63.0K	10	29	3	5	3	16	10	12	11	3	22
14	63.632K	10	29	3	5	3	16	10	12	11	3	22
15	63.635K	1	30	3	5	3	16	10	12	-	22	24
16	63.638L	4	29	17	3	3	17	10	12	13	23	11

17	73.733L	1	20	3	3	14	7	2	4	11	2	11
18	73.737I	1	20	3	3	14	7	2	4	11	2	11
19	73.737IA	1	20	3	3	14	7	2	4	□	2	11
20	73.737IB	1	20	3	3	14	7	2	4	11	2	11
21	73.737L	1	20	3	3	14	□	2	4	□	2	11
22	73.738HA	10	31	3	3	3	16	7	12	7	21	25
23	73.738HAi	10	31	3	3	3	16	7	12	7	21	25
24	73.740I	1	20	3	3	14	7	7	4	11	21	11
25	73.740Ii	1	20	□	3	14	□	7	4	□	□	11
26	85.857H	□	6	□	3	3	□	□	12	□	□	□
27	85.859H	3	6	3	3	3	7	7	12	□	2	20
28	87.0H	□	□	17	3	14	□	□	4	□	24	24
29	89.892I	3	30	3	5	14	7	1	4	11	24	24
30	90.901HA	1	20	□	3	19	□	2	16	11	10	24
31	90.901HAi	1	20	1	3	19	15	2	16	11	10	24
32	90.901HAii	1	20	1	3	19	15	2	16	11	10	24
33	90.901K	1	20	□	3	19	□	2	16	□	10	24
34	90.901KA	1	20	1	3	19	15	2	16	11	10	24
35	90.903G	1	20	1	3	19	15	2	16	11	10	24
36	90.903I	1	20	□	3	19	□	□	16	□	10	24
37	90.903K	1	20	1	3	19	15	2	16	11	10	24
38	90.906G	1	20	1	3	19	15	2	16	11	10	24
39	91.913G	1	4	1	3	20	15	1	16	3	10	22
40	91.913I	1	4	1	3	20	15	1	16	3	10	22
41	100.0F	3	32	18	21	14	13	1	4	11	3	7
42	100.1006G	3	28	18	21	3	13	2	12	11	1	25

Note: "□" samples did not amplify and hence genotypes were not assigned.

As expected all loci showed PCR product length polymorphism and the degree of variation detected differed among loci (Table 13).

Table 13

SOUTH AFRICAN SURVEY -

SUMMARY OF *E. histolytica* GENOTYPES

Loci	1-2	5-6	3-4	3-8	7-4	9-4	9-11	10-4	16-17	16-19	18-17
Total No of Genotypes	5	10	5	5	6	7	4	5	5	8	9

A total of twelve families are represented in the forty-one samples analysed. We therefore proceeded to examine polymorphisms within family groups and samples taken from the same individual on different visits. Twenty-seven of the forty-one isolates studied came from one of three families i.e. 63, 73 and 90.

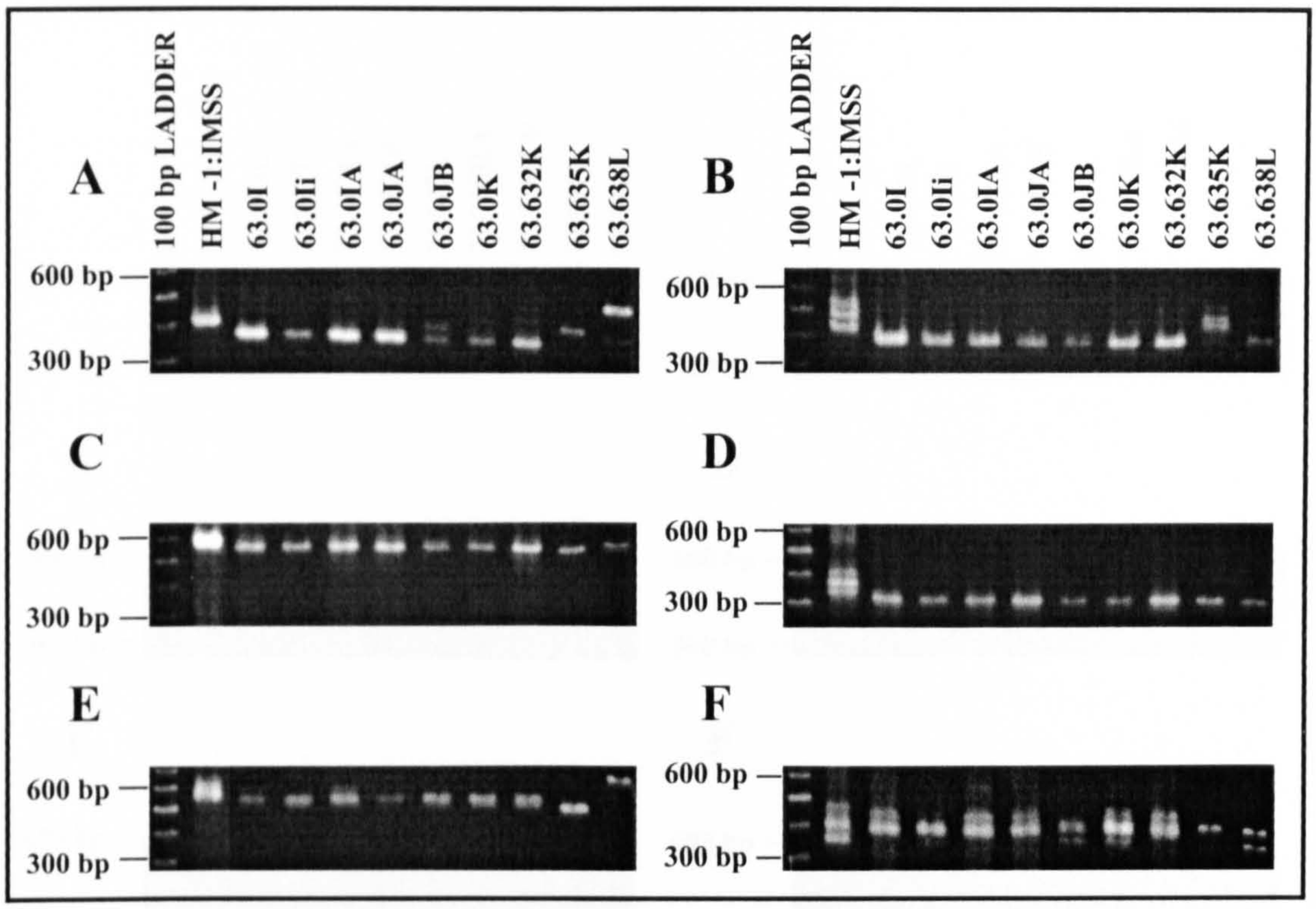


Fig 27. Polymorphic DNA Analysis of *E. histolytica* Isolates from Family 63. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

6.3.1.i. FAMILY 63

Family 63 is represented by a total of nine samples from four different individuals. Six of the DNA samples were from the index case (63.0) and corresponded to five different clinic visits I, IA, JA, JB and K while one sample was available in duplicate (63.0Ii). The remaining three samples are each from a unique individual in the family. All six samples belonging to index case 63.0 gave identical product sizes and patterns at all except one locus i.e. locus 1-2 (Fig. 27A). The results at six of the loci are shown. Patterns for samples 63.635K and 63.638L were quite different from the index case samples as well as from each other at several of the loci (Fig. 27; Table 12). In contrast, the PCR fragment sizes and patterns of sample 63.632K were identical to those of 63.0 at all loci (Fig. 27; Table 12).

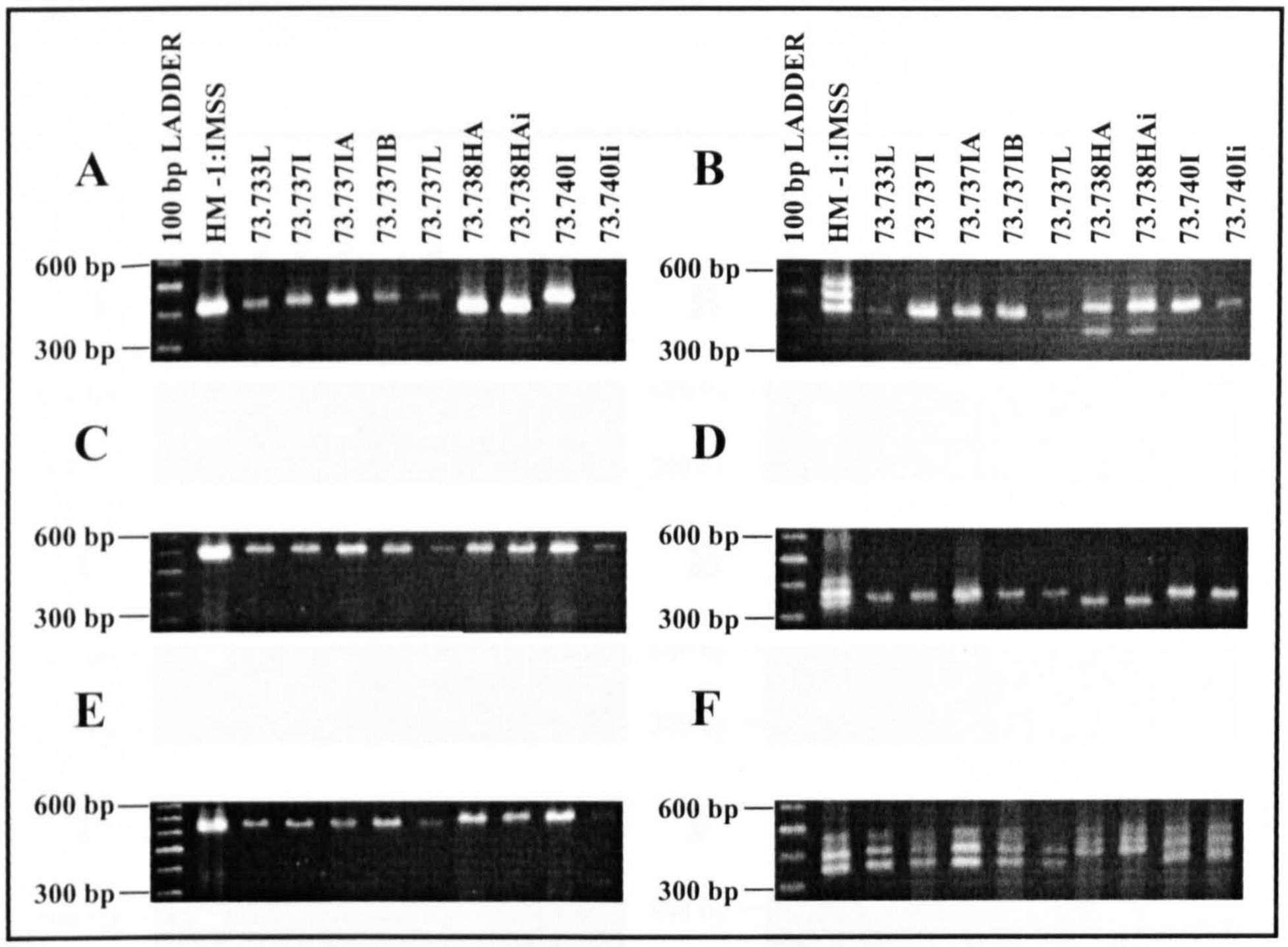


Fig 28. Polymorphic DNA Analysis of *E. histolytica* Isolates from Family 73. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

6.3.1.ii. FAMILY 73

Family 73 was also represented by nine DNA samples from four individuals. Four samples were from individual 73.737 and were collected at different clinic visits. Samples 73.738HA and 73.740I were available in duplicate. The final sample was from individual 73.733 at visit L. All four samples from individual 73.737 produced the same PCR product sizes and patterns at all the loci and results for six loci are shown (Fig. 28; Table 12). Once again one sample (73.733L) gave patterns identical to those of individual 73.737 at all loci. However, while the duplicate samples from individual 73.740 were identical, they differed from individuals 73.733 and 73.737 at locus 9-11 (Table 12) and locus 16-19 (Fig. 28E). The duplicate samples designated 73.738HA and HAi showed complete identity at all loci, but were distinct from the

rest of the family members at all except two loci i.e. locus 3-4 (Table 12) and locus 3-8 (Fig. 28C).

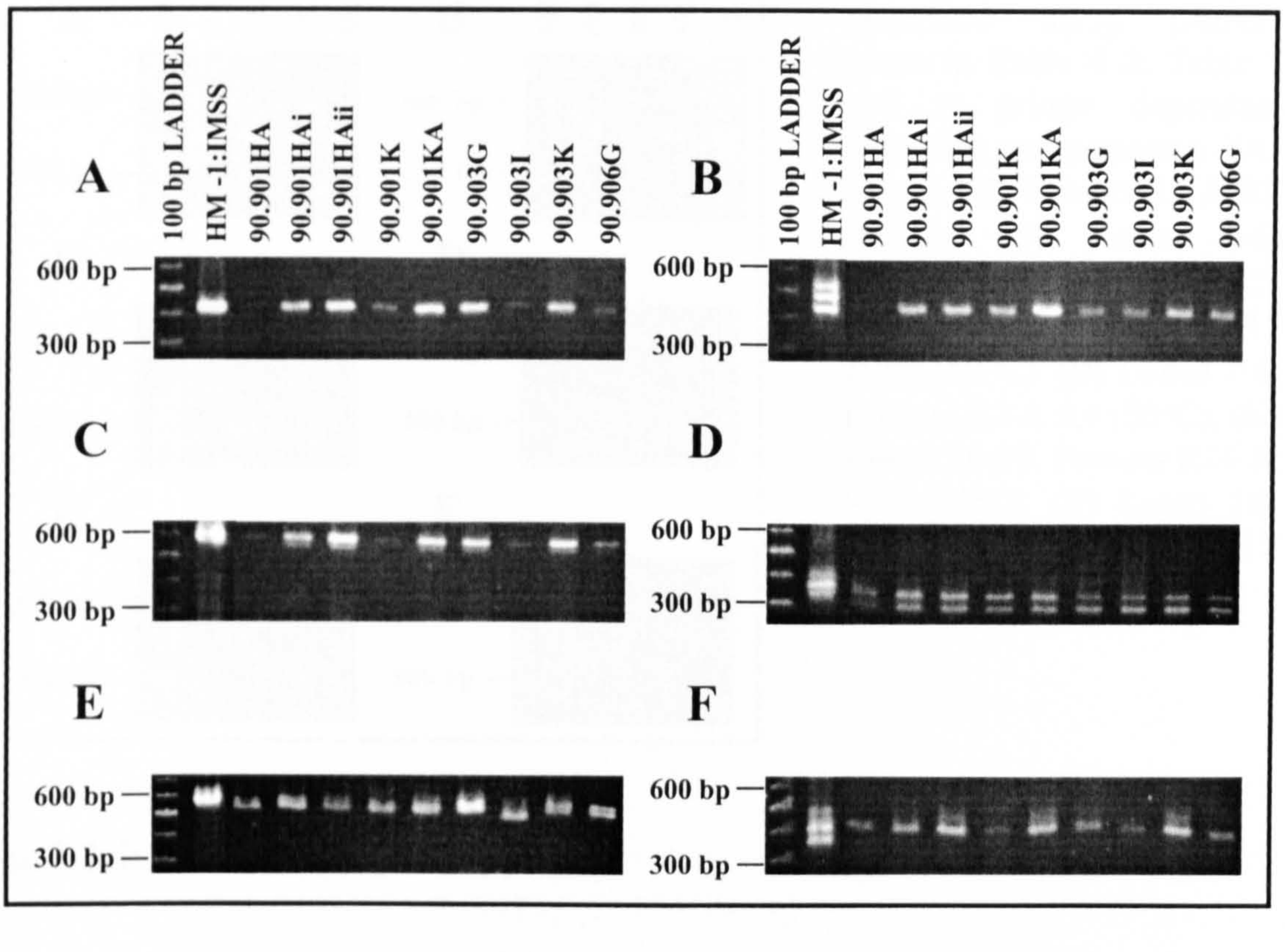


Fig 29. Polymorphic DNA Analysis of *E. histolytica* Isolates from Family 90. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

6.3.1.iii. FAMILY 90

Family 90 is also represented by nine samples from three individuals. Five of the nine samples came from individual 90.901, being the triplicate samples from visit HA as well as one sample each from visits K and KA. Three samples were from individual 90.903 on visits G, I and K while one sample was from individual 90.906 taken at visit G. PCR fragment sizes revealed that the patterns of all nine samples are identical at all the loci, six of which are depicted in Fig. 29.

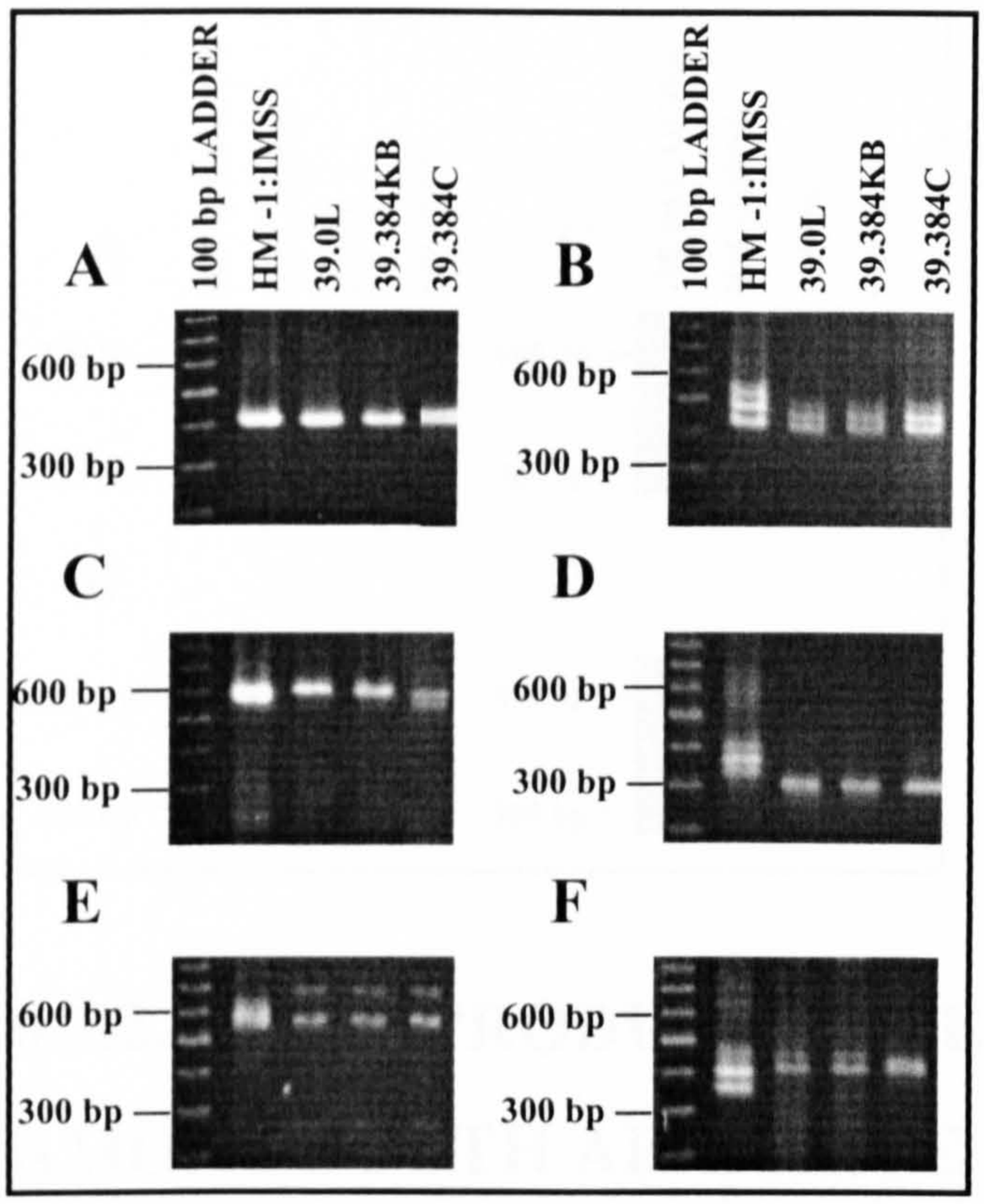


Fig 30. Polymorphic DNA Analysis of *E. histolytica* Isolates from Family 39. Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A) Locus 1-2.** Primers R1 & R2 (53°C). **(B) Locus 5-6.** Primers R5A & R6A (56°C). **(C) Locus 3-8.** Primers R3 & R8 (50°C). **(D) Locus 7-4.** Primers R7 & R4 (50°C). **(E) Locus 16-19.** Primers R16 & R19 (54°C). **(F) Locus 18-17.** Primers R18 & R17 (54°C).

6.3.1.iv. OTHER FAMILIES

Though small by comparison, family 39, which is represented by the index case sample 39.0L and two samples from individual 39.384, at visits C and KB, presented a very different picture. Differences were seen in the patterns between sample 39.384C and the others at loci 3-8 and 18-17 (Fig. 30C & F) and 10-4 (Table 12), while sample 39.384KB, which was collected from the same patient some two years later, produced a pattern identical to the sample from the index case collected one month later.

Three families were represented by only two samples each. The two members of family 100 (100.0F and 100.1006G) could be distinguished from each other at six of the loci (Fig. 31B; Table 12). The patterns were identical at the remaining five loci (Fig. 31A; Table 12). In contrast, the two samples corresponding to different individuals in family 85 (Table 12) and those from visits G and I of individual 91.913 (Fig. 31C & D) had identical patterns at all loci examined.

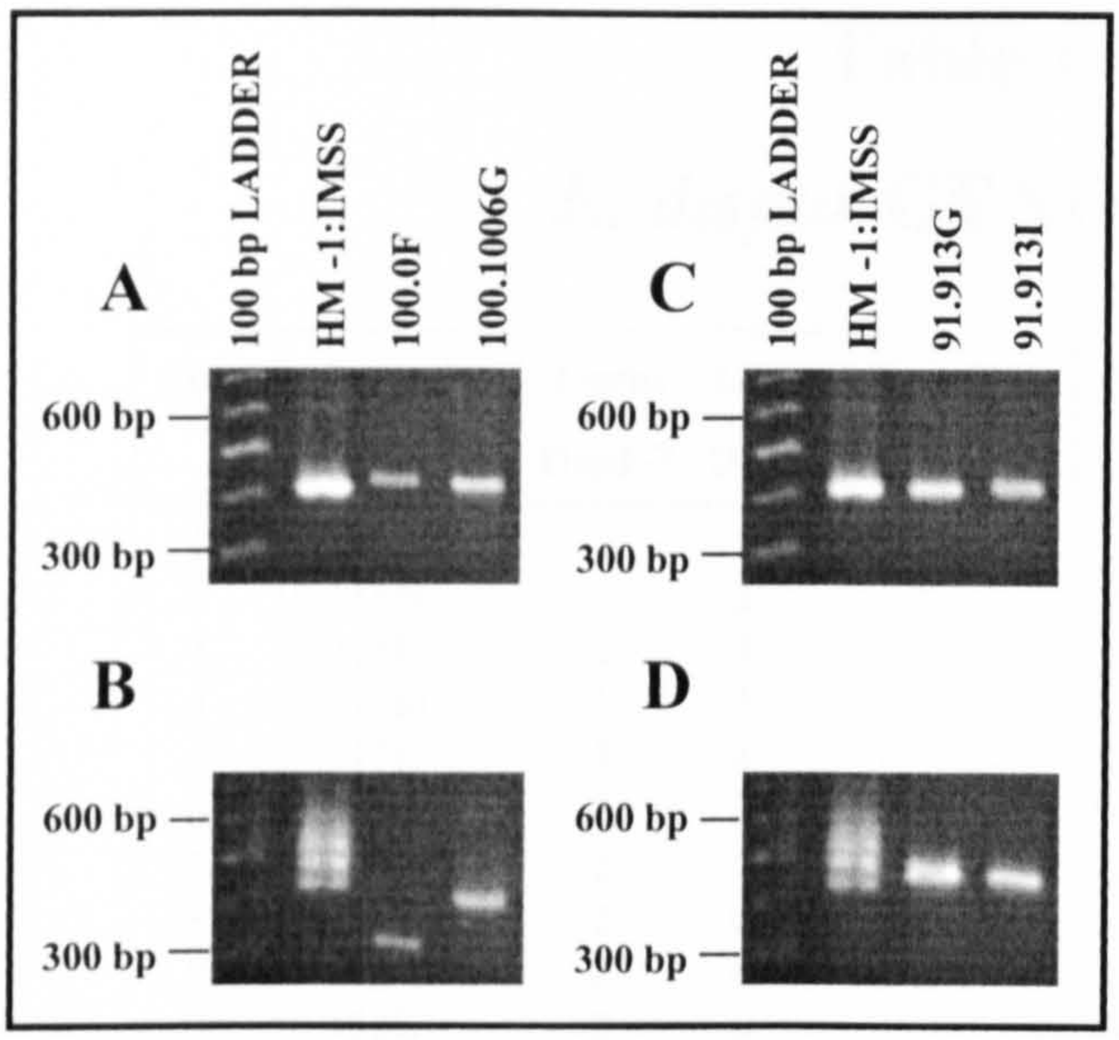


Fig 31. Polymorphic DNA Analysis of *E. histolytica* Isolates from Family 100 (A & B) and Family 91 (C & D). Amplification products were generated using primers given in Table 4 & Table 7 and at primer dependant annealing temperatures. **(A & C) Locus 1-2.** Primers R1 & R2 (53°C). **(B & D) Locus 5-6.** Primers R5A & R6A (56°C).

6.3.2. PCR PRODUCT SIZE POLYMORPHISM AMONG SOUTH AFRICAN *E. dispar* ISOLATES

A total of one hundred and eleven DNA samples, identified as *E. dispar* on the basis of zymodeme analysis were available. Nine failed to amplify with either of the two *E. dispar* specific primer pairs, namely Dsp 1-2 and Dsp 5-6 (Table 11). Four other samples gave amplification products at one but not both loci. These thirteen samples were subsequently excluded from the final analyses. Only the ninety-eight samples that amplified at both loci are shown in Table 14. A total of thirty-six families were represented by ninety samples (Table 14; Appendix XI). The remaining eight samples did not represent a defined family unit but rather a group of school children living in the same community.

Table 14

E. dispar GENOTYPES

No	Isolate	Locus Dsp1-2	Locus Dsp5-6
1	SAW760	1	1
2	16.156N	2	2
3	22.211L	1	3
4	22.212M	1	3
5	22.217L	3	4
6	27.270N	4	5
7	29.284M	2	6
8	29.284N	2	6
9	36.352N	2	6
10	36.352L	2	6
11	40.395M	5	7
12	40.397K	1	2
13	40.397L	1	2
14	40.397M	1	2
15	40.399M	6	2
16	41.410K	6	2
17	41.410M	6	2
18	44.440K	6	2
19	49.484L	1	8
20	49.486L	1	9
21	50.503I	1	10
22	52.521K	1	5
23	57.579K	1	3
24	58.590L	2	11
25	58.588L	2	11
26	59.596K	2	3
27	59.595K	2	7
28	60.0J	6	12
29	60.0Ji	6	12
30	60.607J	1	12
31	60.607Ji	1	12
32	61.612JB	6	2
33	61.612KA	6	7
34	62.628K	6	1
35	62.629I	1	2
36	62.629K	1	2
37	64.641H	1	13
38	64.648K	1	13
39	69.699L	1	7
40	72.721J	6	14
41	81.814G	6	12
42	81.815G	6	12
43	83.840G	6	15
44	84.843I	1	7
45	85.860H	2	2
46	86.861G	6	7
47	86.862G	1	7
48	86.863Gi	1	7
49	86.870J	7	9
50	87.874H	6	7
51	87.874I	6	7
52	87.878H	1	9
53	87.878I	1	9
54	87.880G	1	9
55	87.880Gi	1	9
56	87.880H	1	9
57	87.880I	1	9
58	88.881G	1	9
59	88.881H	1	9
60	89.891I	5	2
61	89.900K	6	7
62	90.902K	1	15
63	90.904G	1	15
64	90.905G	1	15
65	90.905I	1	15
66	90.907G	1	15
67	90.908GB	8	15
68	90.909HA	1	15
69	90.909IB	1	15
70	91.911I	1	15
71	91.915I	1	16
72	91.919I	1	2
73	92.926G	1	13
74	92 926Gi	1	13
75	92.926J	1	13
76	93.931G	6	3
77	93.934G	2	17
78	93.934H	2	17
79	93.934Hi	2	17
80	93.939G	1	5
81	94.942I	1	13
82	94.943G	1	13
83	94.943I	1	13
84	95.958I	2	5
85	97.0F	1	12
86	97.0Fi	1	12
87	97.0G	1	12
88	97.0Gi	1	12
89	99.993G	6	2
90	99.996F	6	8
91	99.996G	6	8
92	11581	6	2
93	11590	6	2
94	11593	6	2
95	11597	6	2
96	11598	6	2
97	11628	6	2
98	11663	6	3
99	11691	1	2

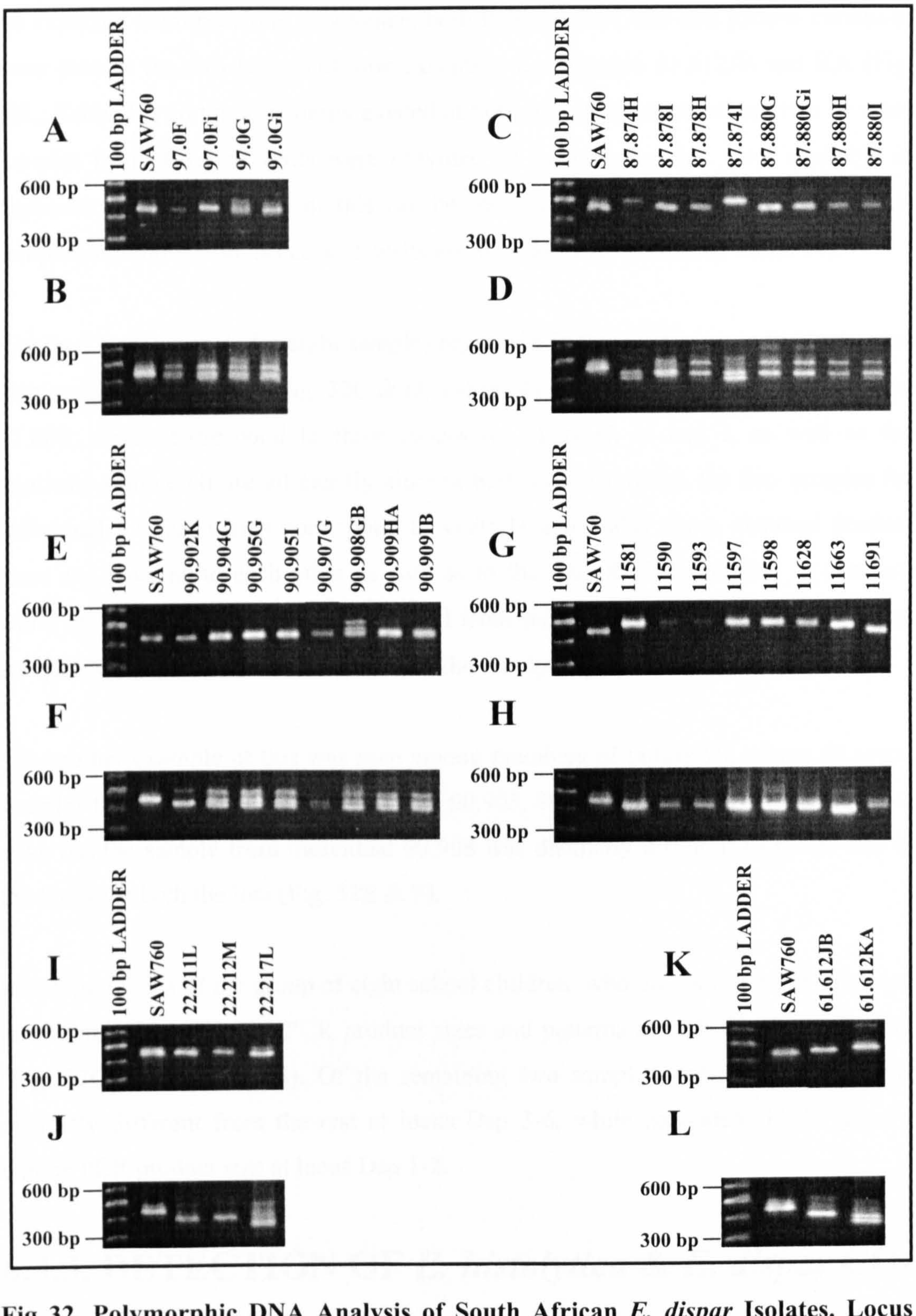


Fig 32. Polymorphic DNA Analysis of South African *E. dispar* Isolates. Locus Dsp 1-2 (A, C, E, G, I & K). Amplification products were generated using primers Dsp1 and Dsp2 at an annealing temperature of 50°C. **Locus Dsp 5-6 (B, D, F, H, J & L).** Amplification products were generated using primers Dsp5 and Dsp6 at an annealing temperature of 52°C.

As expected from previous experience, both PCR product size and pattern variations were seen at the two loci. With one exception i.e. samples 61.612JB and KA (Fig. 32L; Table 14) common patterns existed at both loci in all cases where two or more samples from different visits were provided or for samples that were available in duplicate. A clear example of this can be seen in family 97, where both duplicate samples and those from successive visits are all alike (Fig. 32A & B; Table 14).

Family 87, which includes eight samples representing three different individuals, also displays this feature well (Fig. 32C & D; Table 14). The four samples from individual 87.880, which correspond to three successive visits G, H and I, as well as the duplicate sample Gi are all exactly alike at both loci. Similarly, the two samples for individual 87.878, which correspond to visits H and I also show identical product sizes and patterns to each other as well as to the four 87.880 samples. In contrast, while the two successive samples H and I from individual 87.874 are identical, they are quite distinct from those of the rest of the family.

Yet another example of this was seen among members of family 90, where all seven samples from individuals 90.902, 90.904, 90.905, 90.907 and 90.909 (Table 14) were alike but the sample from individual 90.908 was distinctly different from the rest of the family at both the loci (Fig. 32E & F).

Interestingly, six of the group of eight school children, who do not represent a defined family unit, had identical PCR product sizes and patterns at both the *E. dispar* loci (Fig. 32G & H; Table 14). Of the remaining two samples, individual 11663 was distinctly different from the rest at locus Dsp 5-6, while individual 11691 gave a unique PCR product size at locus Dsp 1-2.

6.3.3. DETECTION OF *E. histolytica* & *E. dispar* CO-INFECTIONS

It has previously been mentioned that amplification at locus 5-6 was only successful in nine of seventy-seven South African *E. dispar* samples tested using the original primer pair R5A and R6A (section 5.3.1; Fig. 33A). In order to confirm the original

species designation of these nine samples we decided to test them using all four species-specific primer pairs (Table 11).

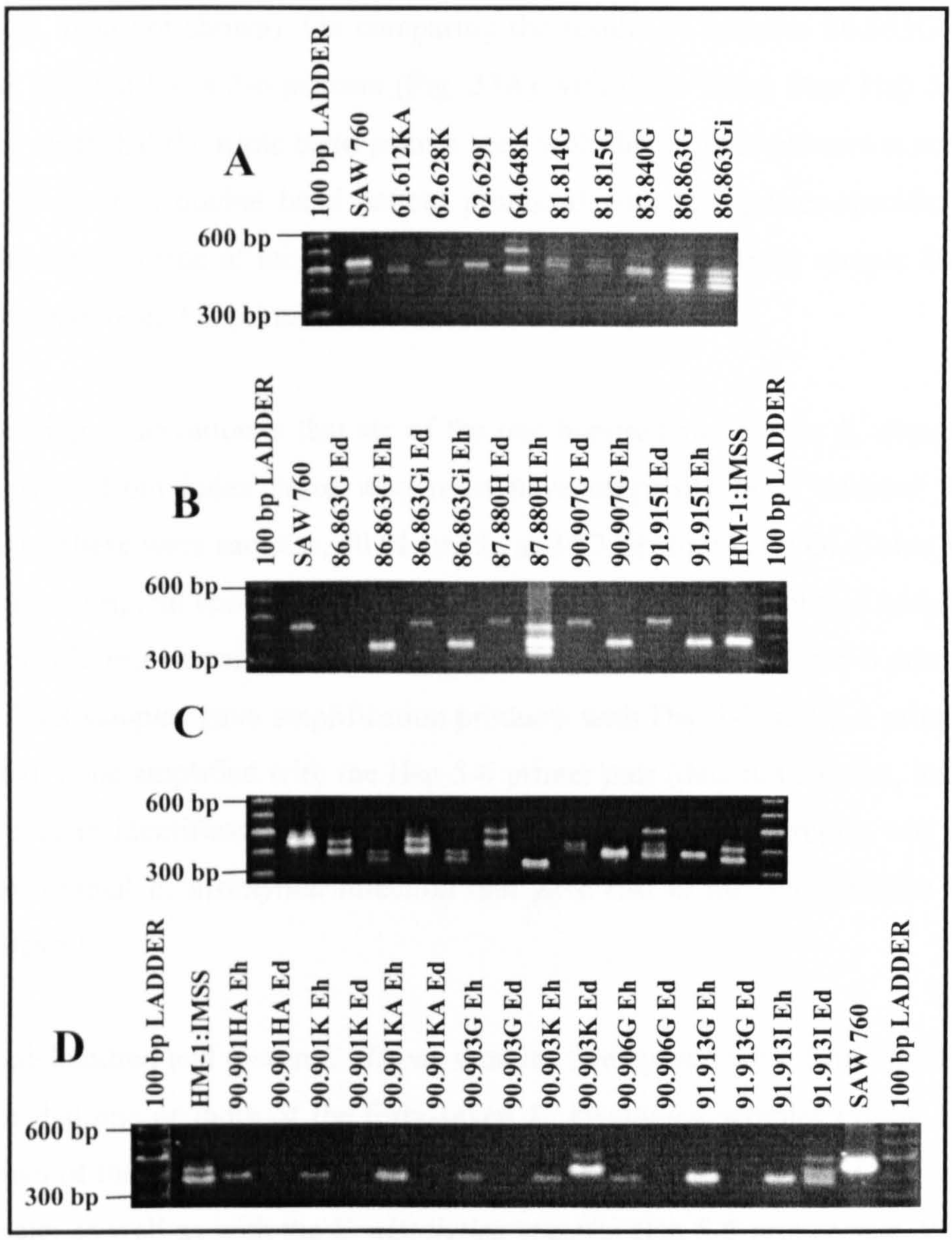


Fig. 33. Species-Specific DNA Analysis of South African *E. dispar* and *E. histolytica* Isolates. (A) Locus 5-6 was amplified using primers R5A & R6A at an annealing temperature of 56°C. **(B) Locus Dsp/ Hsp 1-2**. Amplification products were generated using primers Dsp1 & 2 (Ed) and primers Hsp1 & 2 (Eh) at annealing temperatures of 50°C. Sample 86.863G (Ed) did not amplify with the Dsp 1 & 2 primer pair. **(C) Locus Dsp/ Hsp 5-6**. Amplification products were generated using primers Dsp5 & 6 (Ed) at an annealing temperature of 52°C and primers Hsp5 & 6 (Eh) at an annealing temperature of 48°C. **(D) Locus Dsp/ Hsp 5-6**. Amplification products were generated as in **C.**

Sample, 86.863Gi gave amplification products with all four primer pairs (Fig. 33B) and 6C). (Sample 86.863G (Ed) failed to amplify with Dsp 1-2 (Fig. 33B). The remaining seven samples only amplified with the two *E. dispar* specific primer pairs (Table 14; data not shown). On comparing the results of samples 86.863G and Gi using the original locus 5-6 primers (Fig. 33A) with those using Dsp/ Hsp 5-6 (Fig. 33C) it is clear that the triple band pattern seen with the original primers is actually a mixture of the two double band pattern produced with the species-specific primer pairs. The same is true of the double band patterns produced from sample 86.863Gi by the original locus 1-2 primers (data not shown).

An interesting observation is that six of the one hundred and eleven *E. dispar* DNA samples were from index cases who must have originally been infected with *E. histolytica*. These were samples, 60.0J and Ji, and 97.0F, Fi, G and Gi (Table 14). To confirm their original species designation all six samples were amplified with both *E. dispar* specific primer pairs as well as the *E. histolytica* specific, Hsp 5-6 primer pair. While all six samples gave amplification products with Dsp 1-2 and 5-6 primer pairs (Table 14), none amplified with the Hsp 5-6 primer pair (data not shown), indicating that the species identification based on isoenzymes for these six samples was correct. Thus the original *E. histolytica* infection that gave rise to the liver abscess was no longer present.

Of the one hundred and eleven *E. dispar* samples twenty-two were from families that also provided one or more of the forty-seven *E. histolytica* samples (Table 15). All twenty-two of these *E. dispar* DNA samples were tested with both *E. dispar* specific primer pairs as well as with the *E. histolytica* specific Hsp 5-6 primer pair. While all samples gave amplification products with Dsp 1-2 and 5-6 primer pairs, three samples (87.880H, 90.907G and 91.915I) also amplified with the Hsp 5-6 primer pair (Fig. 33C) and subsequently with the Hsp 1-2 primer pair (Fig. 33B).

Due to amplification failure or inadequate DNA, only eight of the seventeen *E. histolytica* DNA samples which came from 'mixed species' families could be tested with both species-specific locus 5-6 primer pairs (Table 15; Fig. 33D). Six of these eight samples gave amplification products with primer pair Hsp 5-6 only. Samples 90.903K and 91.913I, however, also amplified with primer pair Dsp 5-6.

Table 15

FAMILIES WITH BOTH *E. histolytica* & *E. dispar*

Family Unit	Isolate Number	Classification (isoenzyme analysis)	Classification (species-specific locus analysis)
85	85.857H	*E. histolytica*	ND
	85.859H	*E. histolytica*	ND
	85.860H	*E. dispar*	*E. dispar*
87	87.0H	*E. histolytica*	ND
	87.874H	*E. dispar*	*E. dispar*
	87.874I	*E. dispar*	*E. dispar*
	87.878H	*E. dispar*	*E. dispar*
	87.878I	*E. dispar*	*E. dispar*
	87.880G	*E. dispar*	*E. dispar*
	87.880Gi	*E. dispar*	*E. dispar*
	87.880H	*E. dispar*	*E. histolytica* & *E. dispar*
	87.880I	*E. dispar*	*E. dispar*
89	89.891I	*E. dispar*	*E. dispar*
	89.892I	*E. histolytica*	ND
	89.900K	*E. dispar*	*E. dispar*
90	90.901HA	*E. histolytica*	ND
	90.901Hai	*E. histolytica*	ND
	90.901Haii	*E. histolytica*	*E. histolytica*
	90.901K	*E. histolytica*	*E. histolytica*
	90.901KA	*E. histolytica*	*E. histolytica*
	90.902K	*E. dispar*	*E. dispar*
	90.903G	*E. histolytica*	*E. histolytica*
	90.903I	*E. histolytica*	ND
	90.903K	*E. histolytica*	*E. histolytica* & *E. dispar*
	90.904G	*E. dispar*	*E. dispar*
	90.905G	*E. dispar*	*E. dispar*
	90.905I	*E. dispar*	*E. dispar*
	90.906G	*E. histolytica*	*E. histolytica*
	90.907G	*E. dispar*	*E. histolytica* & *E. dispar*
	90.908GB	*E. dispar*	*E. dispar*
	90.909HA	*E. dispar*	*E. dispar*
	90.909IB	*E. dispar*	*E. dispar*
91	91.911I	*E. dispar*	*E. dispar*
	91.913G	*E. histolytica*	*E. histolytica*
	91.913I	*E. histolytica*	*E. histolytica* & *E. dispar*
	91.915I	*E. dispar*	*E. histolytica* & *E. dispar*
	91.919I	*E. dispar*	*E. dispar*

Abbreviations; IC: Index Case (cured of amoebic liver abscess (ALA), now asymptomatic). NCA: Neighbour or Close Associate. ND: Not Done

In all, five out of thirty-seven samples originally identified as *E. dispar*, based on isoenzyme analysis, were shown to be co-infected with *E. histolytica* using the species-specific primer pairs. Likewise two out of eight samples originally identified as *E. histolytica* were found to be co-infected with *E. dispar* using the species-specific primers.

6.4. DISCUSSION

The aim of the present survey was to see if the polymorphic DNA analyses described thus far could prove useful in investigating patterns of transmission of both *E. histolytica* and *E. dispar* and for carrying out longitudinal molecular epidemiological studies in amoebiasis endemic areas.

Our results show that amplification was successful for a majority of the *E. histolytica* (ca. 62%) and *E. dispar* (ca. 83%) samples with the eleven and two primer pairs tested, respectively. The fact that some of the samples did not give amplification products at any of the loci suggests that the DNA might have been lost or degraded during the extraction procedure. A significant proportion of the *E. histolytica* (ca. 25%) and a few of the *E. dispar* (ca. 5%) samples gave amplification products at some but not all of the loci. There is evidence that this may be due to sequence diversity among isolates in the primer binding regions (sections 3.3.3 & 4.3.1). Nucleotide sequence comparison of products from two loci in five isolates showed minor sequence variation in the regions flanking the repeat blocks (section 3.3.4); but, none of the isolates sequenced were from South Africa.

It has previously been demonstrated that *E. histolytica* infections occur more frequently in family groups and their neighbours (Gathiram and Jackson, 1987). This phenomenon has also been reported by Rivera *et al* (1998) and in earlier studies where differentiation between *E. histolytica* and *E. dispar* was not possible (Spencer *et al*, 1977; Dykes *et al*, 1980).

Our results following analysis of a larger number of samples from both *Entamoeba* species suggest that *E. histolytica* and *E. dispar* are equally likely to be transmitted within family groups and closely associated individuals, which implies common mechanisms of transmission. In a given family most of the samples with distinct patterns were from individuals classified as neighbours or close associates (NCA) e.g. *E. histolytica* families 63, 73 and 100 (Table 16; Fig. 27, 28 & 31A & B) and *E. dispar* families 40, 91 and 93 (Table 17 and data not shown), while samples from blood relatives gave identical patterns in the majority of cases e.g. *E. dispar* families

64, 81 and 86 (Table 17 and data not shown). An exception was seen in family 87 (Table 17; Fig. 32C & D) where PCR product patterns observed for the index case's daughter (87.874) were completely different from those seen for the son (87.878), while the latter showed complete identity to NCA individual 87.880. Likewise, not all individuals described as NCAs displayed unique PCR patterns.

Table 16

SOUTH AFRICAN *E. histolytica* ISOLATES

Family Unit	Isolate #	Relationship	Dwelling
37	37.0C	IC	Suburb
39	39.0L	IC	Farm
	39.384C	Sister	Farm
	39.384KB	Sister	Farm
2596	2596	NA	IS
56	56.0L	IC	Suburb
63	63.0I	IC	Suburb
	63.0Ii	IC	Suburb
	63.0IA	IC	Suburb
	63.0JA	IC	Suburb
	63.0JB	IC	Suburb
	63.0K	IC	Suburb
	63.632K	NCA	Suburb
	63.635K	NCA	Suburb
	63.638L	NCA	Suburb
73	73.733L	NCA	IS
	73.737I	NCA	IS
	73.737IA	NCA	IS
	73.737IB	NCA	IS
	73.737L	NCA	IS
	73.738HA	NCA	IS
	73.738HAi	NCA	IS
	73.740I	NCA	IS
	73.740Ii	NCA	IS
85	85.857H	NCA	Farm
	85.859H	NCA	Farm
87	87.0H	IC	Farm
89	89.892I	Son	IS
90	90.901HA	NCA	Farm
	90.901HAi	NCA	Farm
	90.901HAii	NCA	Farm
	90.901K	NCA	Farm
	90.901KA	NCA	Farm
	90.903G	Wife	Farm
	90.903I	Wife	Farm
	90.903K	Wife	Farm
	90.906G	Son	Farm
91	91.913G	NCA	Farm
	91.913I	NCA	Farm

100	100.0F	IC	IS
	100.1006G	NCA	IS

Note: Only the forty-one samples shown in Table 12 are listed here. The complete list of all forty-seven isolates is given in Appendix XIV. **Abbreviations;** IC: Index Case (cured of amoebic liver abscess (ALA), now asymptomatic). NCA: Neighbour or Close Associate. IS: Informal Settlement.

Table 17

SOUTH AFRICAN *E. dispar* ISOLATES

Family Unit	Isolate #	Relationship	Dwelling
22	22.211L	NCA	IS
	22.212M	NCA	IS
	22.217L	NCA	IS
40	40.395M	NCA	IS
	40.397K	NCA	IS
	40.397L	NCA	IS
	40.397M	NCA	IS
	40.399M	NCA	IS
60	60.0J	IC	IS
	60.0Ji	IC	IS
61	61.612JB	None	IS
	61.612KA	None	IS
62	62.628K	NCA	Suburb
	62.629I	NCA	Suburb
64	64.641H	Mother	Farm
	64.648K	Daughter	Farm
81	81.814G	Sister	IS
	81.815G	Sister	IS
83	83.840G	None	Suburb
85	85.860H	NCA	Farm
86	86.861G	NCA	IS
	86.862G	Daughter	IS
	86.863G	Mother	IS
	86.863Gi	Mother	IS
	86.870J	NCA	IS
87	87.874H	Daughter	Farm
	87.874I	Daughter	Farm
	87.878H	Son	Farm
	87.878I	Son	Farm
	87.880G	NCA	Farm
	87.880Gi	NCA	Farm
	87.880H	NCA	Farm
	87.880I	NCA	Farm
89	89.891I	Mother	Suburb
	89.900K	NCA	IS

90	90.902K	NCA	Farm
	90.904G	NCA	Farm
	90.905G	NCA	Farm
	90.905I	NCA	Farm
	90.907G	NCA	Farm
	90.908GB	Husband	Farm
	90.909HA	NCA	Farm
	90.909IB	NCA	Farm
91	91.911I	NCA	Farm
	91.915I	NCA	Farm
	91.919I	NCA	Farm
93	93.931G	NCA	IS
	93.934G	NCA	IS
	93.934H	NCA	IS
	93.934Hi	NCA	IS
	93.939G	NCA	IS
97	97.0F	IC	IS
	97.0Fi	IC	IS
	97.0G	IC	IS
	97.0Gi	IC	IS
School Children from Langebaan	11581	Same School	Suburb
	11590	Same School	Suburb
	11593	Same School	Suburb
	11597	Same School	Suburb
	11598	Same School	Suburb
	11628	Same School	Suburb
	11663	Same School	Suburb
	11691	Same School	Suburb

Note: Only samples referred to in the text or those shown in the figures have been listed here. The complete list of all one hundred and eleven isolates is given in Appendix XIII. **Abbreviations;** IC: Index Case (cured of amoebic liver abscess (ALA), now asymptomatic). NCA: Neighbour or Close Associate. IS: Informal settlement.

These results are not surprising in view of the complex nature of family relationships that exist in this population. While recruiting subjects for the longitudinal study, attempts were made to enrol family members first, followed by others also living in the household but who were not necessarily blood relatives. It is not unusual for a number of unrelated individuals to share a household in these economically deprived areas. Where recruit numbers were inadequate, neighbours who had close contact with the index case were also enrolled. In almost all cases individuals classified as NCAs lived within a 30-50 metre radius of the index case and at recruitment all associates indicated that they had regular contact with the index case. Neighbours often share meals, latrines and drinking water sources. Furthermore, in addition to their contacts with the index case family, NCAs will have contact with other NCAs.

It is commonly accepted that the transmission of amoebic infection in certain regions of the world depends largely on factors such as quality of sanitation, availability of clean water and socio-economic status. Our results suggest that the likelihood of two or more individuals sharing PCR product patterns is the same irrespective of where they reside. Examples of this may be seen in *E. histolytica* family 63 (Table 16; Fig. 27) and *E. dispar* families 22 (Table 17; Fig. 32I & J) and 87 (Table 17; Fig. 32C & D), where two or more family members appear to be infected with a common strain whether they live in an informal settlement (family 22), where the households are usually 1-2 metres apart and a large number of individuals share a common water tap, on a farm (family 87), where rivers or bore-holes are the sources of water, or in the suburbs (family 63), where houses generally have their own taps. These findings imply that the occurrence of shared infection is not necessarily influenced by the source of water, living conditions and socio-economic status, nor whether it is *E. histolytica* or *E. dispar*. Person to person transmission seems the most likely underlying factor.

An interesting observation was that, with rare exceptions, all the samples resulting from different clinic visits by one individual showed identical PCR product sizes and patterns at all loci. This was true even of samples collected at intervals of six, nine and twelve months e.g. *E. histolytica* samples 90.903G to K (Fig. 29) and *E. dispar* samples 87.880G to I (Fig. 32C & D). These markers appear, therefore, to be stable and suitable for tracking transmission of a known strain within an individual, family unit or community.

The exceptions all have different implications. *E. histolytica* sample 63.0 at visit JB (Fig. 27) shows the presence of an extra band at locus 1-2 when compared to results for visits I, IA, JA and K, which are all identical. To confirm that the second band was not an amplification artefact, we amplified DNAs of samples 63.0JA and JB with the species-specific primer pair Hsp 1-2. The results were identical to those observed with the original locus 1-2 primers (data not shown). This appearance and subsequent loss of an additional product is suggestive of a transient mixed infection with a second strain of the same species. Individual 87.880 is infected with *E. dispar* at visits G, H and I (Fig. 32C & D) but shows evidence of a transient *E. histolytica* infection at visit H (Table 15; Fig. 33B & C). Patterns seen for individual 39.384 at visit C are

different from those observed at visit KB some two years later (Fig. 30C & F; Table 12). However the latter patterns share complete identity with index case sample 39.0L (Fig. 30; Table 12). The strain carried by 39.384 at visit C has been replaced by a distinct *E. histolytica* during the subsequent two years. However, whether this is the result of re-infection with the index case's strain can not be determined. Individual 61.612 also shows unique *E. dispar* patterns for the two visits JB and K represented in our study (Fig. 32K & L). This must reflect the loss of and rapid re-infection (within a month) with a different strain of the same species. How frequently such transient or replacement infections occur can not be deduced without complete sets of longitudinal samples from many individuals.

A drawback of the present survey is that typing was only performed on DNA derived from cultures. Our results following PCR amplification with species-specific primer pairs revealed that seven samples originally classified as either *E. histolytica* or *E. dispar* actually contained both (Table 15; Fig. 33). However, as not all samples were tested with the species-specific primer-pairs, this number represents the minimum. Furthermore, some mixed infections may well have been missed due to failure of one strain in a mixed infection to grow in culture. This potential problem could be eliminated in future studies by using DNA extracted directly from faeces. The use of PCR in documenting mixed infections from DNA extracted directly from faeces has been reported by others (Acuña-Soto *et al*, 1993; Haque *et al*, 1998). We have already shown that strain typing by amplification from faecal DNA is possible with our primers but is dependant on the DNA extraction method and thermostable polymerase used (Chapters 4 & 5).

In conclusion it appears that this PCR-based typing method has the potential to be used for studying the molecular epidemiology of *E. histolytica* and *E. dispar* in an endemic setting. The stability of the polymorphisms described means that the course of a single infection can be followed to determine its final outcome. Surprisingly no identifiable *E. histolytica* strain was detected in more than one family group indicating a remarkable degree of genetic diversity in this relatively restricted geographic area. This can also be seen in the phylogenetic analysis supplement. This observation will be particularly useful in allowing us to decide whether an *E. histolytica* strain infecting multiple individuals comes from the same source.

CHAPTER 7

GENOMIC ORGANISATION OF REPEAT LOCI

7.1. INTRODUCTION

Although the structure of the *E. histolytica* genome is still not well understood, various unusual features have been revealed. Both linear chromosomes and a number of circular plasmid-like molecules have been observed but information about their size and structure is incomplete and conflicting (Bhattacharya *et al*, 2000). Repeated-DNA sequences have been identified on both linear chromosomes (Mittal *et al*, 1994; Huang *et al*, 1997) and circular molecules (Sehgal *et al*, 1994). In fact it appears that up to 20% of the genome may be made up of repeated-DNA.

Repeated sequences play an important role in genetic recombination and are also known to be involved in gene regulation in many organisms (Darnell, 1990). Most of them are transcriptionally inactive. They are often located in centromeres and the subtelomeric regions of the chromosomes, thereby suggesting a structural or organisational role, and give rise to size variations in homologous chromosomes due to unequal crossing-over events. Localisation and mapping of repeat sequences in the genome can enhance our knowledge of the organism's genome structure and organisation.

A number of means have been employed to study the genomic organisation of the polymorphic repeat loci used in this study and the results are discussed in this chapter.

7.2. RESULTS

7.2.1. REPEAT LOCI & tRNA GENES

Although results of the Blast search revealed no homology between loci 1-2 and 5-6 and any of the previously reported *E. histolytica* repeat elements (section 3.3.2), interestingly, the 50 bp region between nucleotide positions 2 and 51 in locus 1-2 showed 86% identity to a $tRNA^{Asp}$ gene from *Drosophila* (Fig. 34A; Table 18) while a 43 bp region between nucleotide positions 382 and 424 in locus 5-6 had 93% identity to a $tRNA^{Ile}$ gene from *Dictyostelium discoidium* (Fig. 34B; Table 18).

```
              tggttagtatcttcgcctgtcacgtgaaagaccggggttcgaatccccgt
              ||||||||||||  ||||||||||| |  |||||||||||| | |||||||
cctcgatagtatagtggttagtatccccgcctgtcacgcgggagaccggggttcaattccccgtcgggga
```

Fig. 34A. Identity between *Drosophila* $tRNA^{Asp}$ Gene & Locus 1-2. The top strand shows the locus 1-2 nucleotide sequence between positions 2 and 51 with identity in 43 of the 50 base pairs. The total length of the *Drosophila* $tRNA^{Asp}$ gene (accession # X06213) is 72 bp. The position of the anticodon is given in bold-faced letters.

```
                       cctgcgaccctggcgttattagcaccatgctctaccgactgag
                       ||||||||| |||||||||||||||| ||||| |||||||||
tgctcgatgcagggatcgaacctgcgaccgtggcgttattagcaccacgctctgccgactgagctaatcgag
```

Fig. 34B. Identity between *Dictyostelium discoideum* $tRNA^{Ile}$ Gene & Locus 5-6. The top strand shows the nucleotide sequence between positions 382 and 424 with identity in 40 of the 43 bp. The total length of the *Dictyostelium discoideum* $tRNA^{Ile}$ gene (accession # X69659) is 73 bp. The position of the anticodon is given in bold-faced letters.

The presence of two tRNAs has been reported in locus 9-4 (accession # AF265348) (Table 18). Using the tRNA scan-SE v.1.11 (Lowe and Eddy, 1997) program (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/), we also detected tRNA genes in loci 3-4 and 16-17 (Table 18).

Table 18

CLASSIFICATION OF tRNA MOLECULES

Repeat Locus	Number of tRNA's	Type of tRNA (anticodon)	tRNA Position (bp)	tRNA Length (bp)	Anticodon Position (bp)	Intron Position (length) (bp)
1-2	1	Aspartic Acid (GTC)	2-51	50	20-22	None
5-6	1	Isoleucine (AAT)	382-424	43	399-401	None
3-4	1	Glutamic Acid (TTC)	429-508	80	473-475	None
	2	Leucine (TAG)	945-977 1-60	93	22-24	4-20 (17)
9-4	1	Glutamic Acid (TTC)	29-99	71	64-66	None
	2	Tyrosine (GTA)	508-593	86	558-560	544-556 (13)
16-17	1	PhenylAlanine (GAA)	3-75	73	37-39	None
	2	Valine (GAC)	516-589	74	550-552	None

Note: The position(s) (bp) are as seen in Fig. 4A (locus 1-2); Fig. 5A (locus 5-6) and Appendix V (locus 3-4); Appendix VI (locus 9-4); Appendix VII 16-17. The $tRNA^{Leu}$ position(s) in locus 3-4 reflect the tandem array nature of this locus (this chapter section 7.2.5).

Of particular interest is the presence of a 13 bp intron within the $tRNA^{Tyr}$ gene in locus 9-4 and a 17 bp long intron in the locus 3-4 $tRNA^{Leu}$ gene.

7.2.2. NORTHERN HYBRIDISATION ANALYSIS

Though all potential reading frames in all five loci had abundant stop codons, hybridisation to total RNA and poly A^+ RNA was carried out using all five locus-specific probes.

No signal was detected in the poly A^+ fraction with the loci 1-2 and 5-6 specific probes following a 24-hour exposure (Fig. 35A, lane 1). This supported the observed absence of any significant open reading frames in the respective nucleotide sequences. In fact no signal was detected in the poly A^+ fraction with the locus 1-2 specific probe even after a 6-day exposure (data not shown). However, after a 24-hour exposure hybridisation bands of under 280 bp were seen in the total RNA fraction (Fig. 35A,

lane 2 & 3). This size corresponds to the tRNA region in the ethidium bromide stained gel (data not shown).

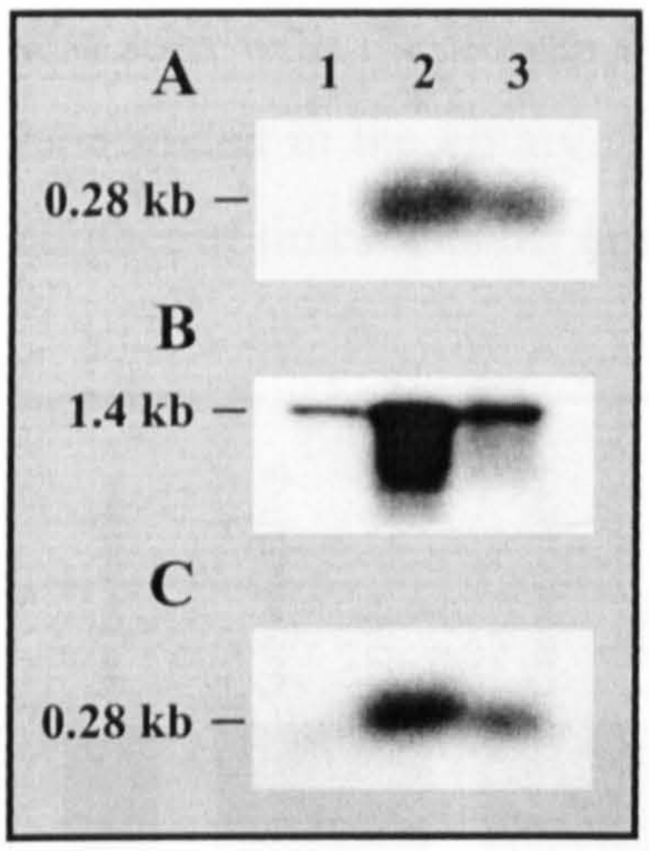


Fig. 35. Northern Hybridisation Analysis. Purified Poly A^+ RNA, 2µg (**lane 1**) and total RNA, 5.4 µg (**lane 2**); 2.15 µg (**lane 3**), were denatured with formaldehyde, electrophoresed on a 1% agarose gels at 5 V/cm and transferred to nylon membranes. The filters were hybridised with [α-^{32}P]-dCTP labelled locus 1-2 specific probe (**A**) (results with the locus 5-6 probe were identical); an Actin probe (**B**); a locus 3-4 specific probe (**C**) (results with locus 9-4 & 16-17 probes were identical). The size of the RNA was measured using an RNA marker (Promega).

Using a sample from the same poly A^+ RNA pool in a separate Northern hybridisation experiment, we were able to detect a strong signal within 4-hours for this fraction with a β-actin probe (Edman *et al*, 1987), thus confirming the presence of intact mRNA in the sample (Fig. 35B, lane 1).

Somewhat different results were seen with the locus 3-4, 9-4 and 16-17 specific-probes. All three probes produced weak signals with the poly A^+ fraction at or between 6- and 24-hours of exposure (Fig. 35C, lane 1). This is most likely due to contamination with small amounts of total RNA. Results with the total RNA fraction were the same as those observed with locus 1-2 and 5-6 probes (Fig. 35C, lanes 2 & 3).

7.2.3. IDENTIFICATION OF REPEAT-REGION FLANKING SEQUENCES

To obtain the sequences flanking both loci 1-2 and 5-6 with the hope of completing and hence confirming the presence of the putative tRNA gene sequences we decided to screen a genomic HM-1:IMSS lambda ZAP library.

To ascertain the presence of these loci in the library, purified total library DNA was used as template for PCR and the products were compared with those obtained from genomic HM-1: IMSS DNA. As is evident from the results (Fig. 36) no amplification was seen with the library DNA at locus 1-2, suggesting that this locus is no longer represented in the library. This is probably because the library has been amplified a number of times. Further analysis was carried out for locus 5-6 only.

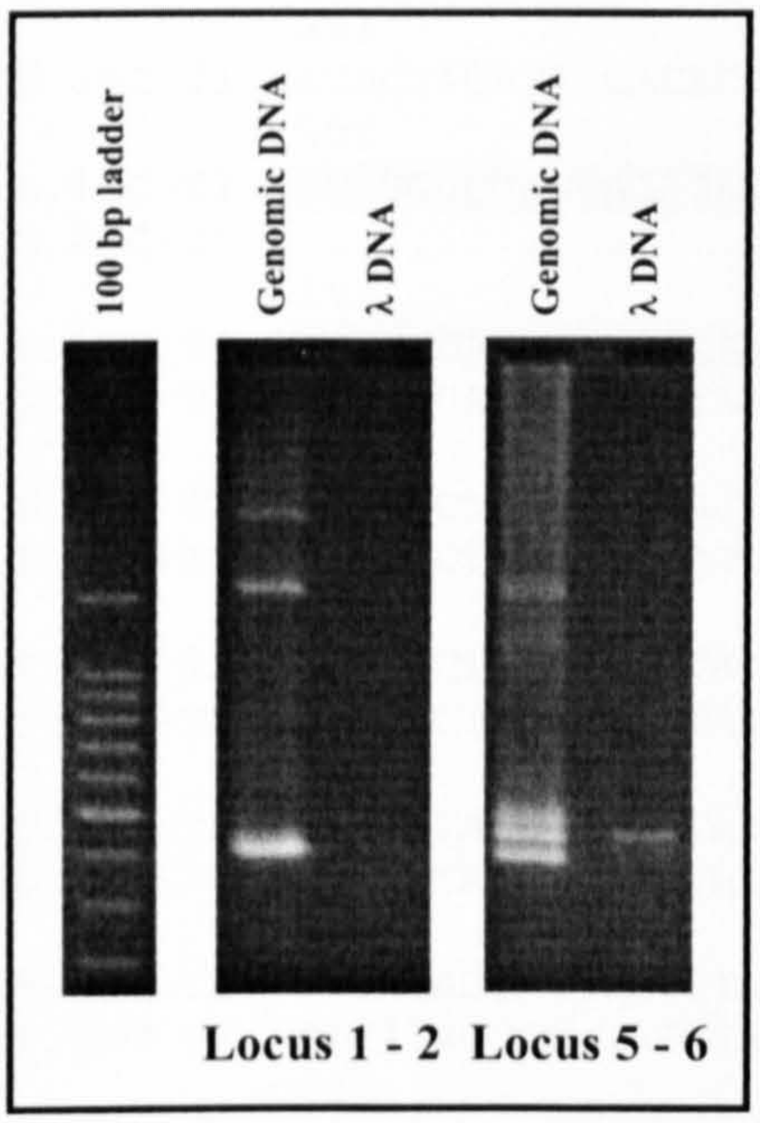


Fig. 36. Amplification of HM-1: IMSS Library DNA & Genomic DNA at Loci 1-2 & 5-6. Locus-specific primers (R1, R2, R5A & R6A) (Table 4) were used under previously described conditions (section 2.10). Amplified products were analysed on a 1.8% agarose gel at 120 V. The 100 bp DNA ladder (Promega) was used as a size marker.

Of the ca. 60,000 recombinants screened with a locus 5-6 specific probe in the first round, 15 putatively positive clones with a wide range of signal intensities were clearly visualised in autoradiographs after 7-days. Two of these were still positive in the second round of screening. One of these clones, designated 5.15C, was sequenced on both strands (Fig. 37A).

Analysis revealed that the complete sequence of the clone 5.15C fragment is ca. 969 bp long and contains two major blocks of repeats. One of the repeat blocks is identical to the one already described in locus 5-6 (Fig. 5A; Appendix IIB) while the second contains eleven tandem 8 bp units (CTTACTAT) (Appendix IID). As already mentioned for locus 5-6, in addition to the major repeat blocks there are several tandem and solitary duplications evident throughout the length of the 5.15C sequence.

```
          1                                                        50
5.15C-T3  TTTTATATCT TTTATATCTT TTATATATAT TTCTTCCTTT TTATACCTTT
          51                                                       100
5.15C-T3  TATATCCCTT TCCAGTGTTT ATATGTCTAT TTATACTACT TATGACTATA
          101                                                      150
5.15C-T3  TCTTACTATC TTACTATCTT ACTATCTTAC TATCTTACTA TCTTACTATC
          151                                                      200
5.15C-T3  TTACTATCTT ACTATCTTAC TATCTTACTA TCTTACTATA TTTCTCCCTT
          201                                                      250
5.15C-T3  ACTTATACAC TAGATATACA CTAGATACAC ACTAGATACA CACTAGATAT
          251                                                      300
5.15C-T3  ACACTAGATA CACACTAGAT ACCATCTATG ACTACTATAT GTATAAAAAG
          301                                                      350
5.15C-T3  TGAGGGCTAG AGGGATTGAA CCACTGACCG GCTGATCTGG AGTCAGCTGC
5.15C-T7  .......... .......... .......... .........G AGTCAGCTGC
          351                                                      400
5.15C-T3  TCTGCCACTG AGCTAAGCCC CCTTCTTCTA TAATTTATAT ATTATTCTCT
5.15C-T7  TCTGCCACTG AGCTAAGCCC CCTTCTTCTA TAATTTATAT ATTATTCTCT
          401                                                      450
5.15C-T3  TTAGACTTAT TCTACTTTAT TTCTTATACT TATTATATCT ACTTTAGATA
5.15C-T7  TTAGACTTAT TCTACTTTAT TTCTTATACT TATTATATCT ACTTTAGATA
          451                                                      500
5.15C-T3  CTTTACTTCA CTTTATATAT TCTACTTTAT ATTCCTGACT TATATCTTTT
5.15C-T7  CTTTACTTCA CTTTATATAT TCTACTTTAT ATTCCTGACT TATATCTTTT
          501                                                      550
5.15C-T3  ATGTTTATAT GTTTATATGA TTTTATGATT TTATGTTTAT ATTCTTCCTA
5.15C-T7  ATGTTTATAT GTTTATATGA TTTTATGATT TTATGTTTAT ATTCTTCCTA
          551                                                      600
5.15C-T3  TTTATACCAC TATGTATATA TATATGTATG TTTCTATGTA TGTTTCTATG
5.15C-T7  TTTATACCAC TATGTATATA TATATGTATG TTTCTATGTA TGTTTCTATG
          601                                                      650
5.15C-T3  TATGTATATT TCTATGTATG TATATTTCTA TGTATGTAT. ..........
5.15C-T7  TATGTATATT TCTATGTATG TATATTTCTA TGTATGTATA TTTCTATGTA
          651                                                      700
5.15C-T7  TGTATATTTC TATGTATGTA TATTTCTATG TATGTATATT TCTATGTACG
          701                                                      750
5.15C-T7  TCTTTAACTT TAAAAACAAT GCTCTCAGCA GGTTTCGAAC CTGCGACCCT
          751                                                      800
5.15C-T7  GGCGTTATTA GCACCATGCT CTACCGACTG AGCTATGAGA GCACCCACTT
          801                                                      850
5.15C-T7  TACCTTATAT CTATATATTA TATCTTTATA TTCTTATGTC TTTATATTCT
          851                                                      900
5.15C-T7  TATATTCTTA TATTCTTATA TTCTTATATA TTTATGTCTT TATGCCTTTA
          901                                                      950
5.15C-T7  TGCCTTTATG CCTTTATATG ACTACTTCTT CTTTATCTCA CTTATTATCT
          951
5.15C-T7  ATATCTTACT TATATCCTT
```

Fig. 37A. Nucleotide Sequence Alignment of 5.15C-T3 & T7 reads. Dots indicate ends of readable sequences. Locus 5-6 repeat regions are in bold with the two truncated repeats underlined. A second repeat block is shown (thick-underline). The two tRNA genes are highlighted.

Alignment of the 5.15C consensus sequence with that of locus 5-6 (Fig. 5A) revealed complete identity between the two in the repeat region (Fig. 37B). However, single nucleotide differences were evident in the repeat flanking regions.

```
              1                                                         350
5.15C consen  .......... .......... .......... .......... ..........
              351                                                       400
5.15C consen  TCTGCCACTG AGCTAA-GCC CCCTTCTTCT ATAATTTATA TATTATTCTC
   Locus 5-6  ---------- --CTAAAGCC CCCTTCTTCT ATAATTTATA TATTATTCTC
              401                                                       450
5.15C consen  TTT-AGACTT ATT-CTACTT TATTTCTTAT ACTTATTATA TCTACTTTAG
   Locus 5-6  TTTGAGACTT ATTTCTACTT TATTTCTTAT ACTTATTATA TCTACTTTAG
              451                                                       500
5.15C consen  ATACTTTACT TCACTTTATA TATTCTACTT TATATTCCTG ACTTATATCT
   Locus 5-6  ATACTTTACT TCACTTTATA TATTCTACTT TATATTCCTG ACTTATATCT
              501                                                       550
5.15C consen  TTTATGTTTA TATGTTTATA TGATTTTATG ATTTTATGTT TATATTCTTC
   Locus 5-6  TTTATGTTTA TATGTTTATA TGATTTTATG ATTTTATGTT TATATTCTTC
              551                                                       600
5.15C consen  CTATTTATAC CACTATGTAT ATATATATGT ATGTTTCTAT GTATGTTTCT
   Locus 5-6  CTATTTATAC CACTATGTAT ATATATATGT ATGTTTCTAT GTATGTTTCT
              601                                                       650
5.15C consen  ATGTATGTAT ATTTCTATGT ATGTATATTT CTATGTATGT ATATTTCTAT
   Locus 5-6  ATGTATGTAT ATTTCTATGT ATGTATATTT CTATGTATGT ATATTTCTAT
              651                                                       700
5.15C consen  GTATGTATAT TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT
   Locus 5-6  GTATGTATAT TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT
              701                                                       750
5.15C consen  ACGTCTTTAA CTTTAAAAAC AATGCTCTCA GCAGGTTTCG AACC-TGCGA
   Locus 5-6  ACGTCTTTAA CTTTAAAAAC AATGCTCTCA GCAGGTTTCG AACCCTGCGA
              751                                                       800
5.15C consen  CCCTGGCGTT ATTAGCACCA TGCTCTACCG ACTGAGCTAT GAGAGCACCC
   Locus 5-6  CCCTGGCGTT ATTAGCACCA TGCTCTACCG ACTGAG---- ----------
              801                                                       969
5.15C consen  .......... .......... .......... .......... ..........
```

Fig. 37B. Nucleotide Sequence Alignment of 5.15C-Consensus Sequence with Locus 5-6 Sequence. Dots indicate sequence data not shown. Dashes indicate start- and end-points of the locus 5-6 sequence. Gaps used to maximise alignment and single nucleotide differences have been highlighted. The locus 5-6 repeat region is in bold-face with the two truncated repeats underlined.

Using the consensus sequence of the alignment shown in Fig. 37A in the tRNA scan-SE program we detected the presence of two tRNA genes in the 5.15C sequence. One of these is the $tRNA^{Ile}$ gene for which partial sequence identity was seen with the locus 5-6 sequence (Fig. 34B; Table 18) and the other is a $tRNA^{Trp}$ gene (Fig. 37A; Table 19). Hence the overall structure of clone 5.15C fragment appears to be very similar to those seen for loci 3-4, 9-4 and 16-17.

Table 19

CLASSIFICATION OF tRNA MOLECULES IN CLONE 5.15C

Repeat Locus	Number of tRNA's	Type of tRNA (anticodon)	tRNA Position (bp)	tRna Length (bp)	Anticodon Position (bp)	Intron
5.15C-T3	1	Tryptophan (CCA)	301-372	72	338-340	None
5.15C-T7	1	Isoleucine (AAT)	720-792	73	757-759	None

Note: The positions (bp) are as seen in Fig. 37A.

We then used the consensus sequence of the alignment shown in Fig. 37A as search template to try and identify corresponding fragments in the TIGR genome sequencing database (http://www.tigr.org/tdb/e2k1/eha1/). Multiple alignment was carried out between our consensus sequence and some of the sequences identified from the database (data not shown) and the complete locus 5-6 unit length was found to be 1112 bp. The ca. 143 bp difference existed at the 5′ end and this stretch showed nine tandemly arranged copies of a 9 bp repeat unit (CTTTTATAT) (Appendix IID).

7.2.4. IDENTIFICATION OF LOCUS 1-2 REPEAT-REGION FLANKING SEQUENCES

Since genomic library screening would likely have failed to identify locus 1-2 flanking regions we decided to try another approach. Inverse primers (IP) were designed from the 5′ repeat-flanking region of locus 1-2 (Table 20) and were used to amplify genomic DNA of strain HM-1:IMSS clone 9 (Fig. 38).

As is evident from Fig. 38, two products of ca. 750 bp and 1200 bp were seen. We were not successful in cloning the ca. 1200 bp product, but all 18 of the ca. 750 bp fragment transformants selected randomly contained an insert (data not shown). Plasmid DNA was isolated from four of these and sequenced.

Table 20

LOCUS - SPECIFIC INVERSE OLIGONUCLEOTIDE PRIMERS

Repeat Locus	Primer Name	Primer Sequence ($5^{/} \rightarrow 3^{/}$)
1-2	IP1.1 ($5^{/}$ primer) IP1.2 ($3^{/}$ primer)	AAA GAC CGG GGT TCG AAT CC CAC GTG ACA GGC GAA GAT ACT

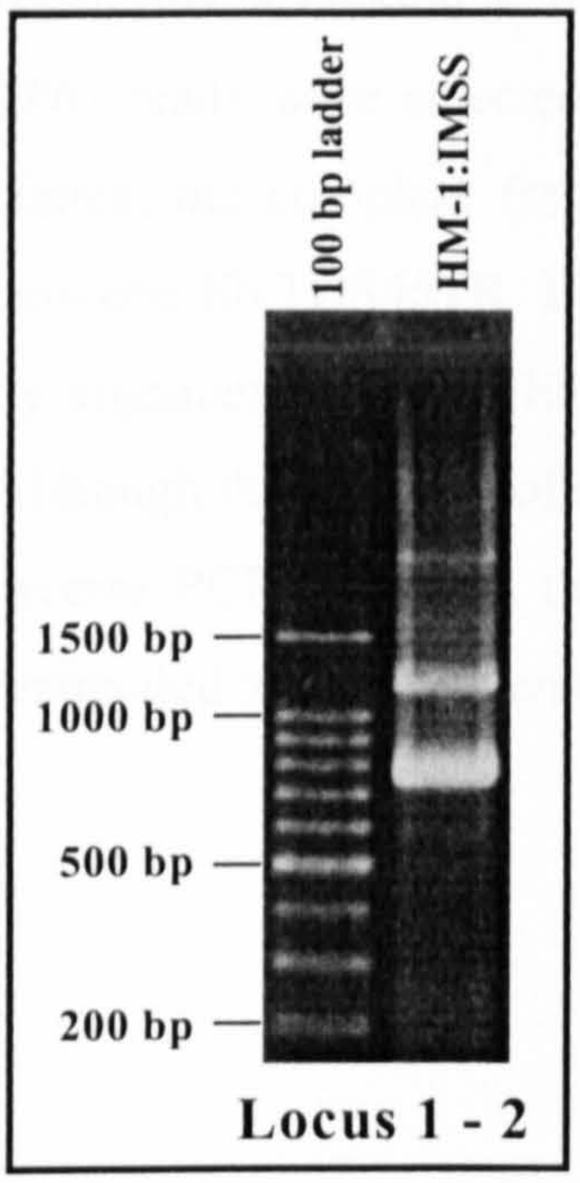


Fig. 38. Inverse PCR of HM-1: IMSS Genomic DNA at Locus 1-2. Locus-specific inverse primers (IP1.1 & IP1.2) (Table 20) were used under previously described conditions (section 2.10). Amplified products were analysed on a 1.8% agarose gel at 120 V. The 100 bp DNA ladder (Promega) was used as a size marker.

Sequencing was not entirely successful and the consensus sequence obtained after multiple alignment of the sequence data from all four clones was 614 bp long for the T7-reads but only 259 bp long for the SP6-reads. Our attempts to align the locus 1-2 sequence with those of the T7- and SP6- consensus reads revealed very poor matches (Fig. 39). There is essentially only a 2 bp overlap among the T7 and SP6 sequences between positions 489 and 490. Furthermore, there is little or no sequence identity between locus 1-2 and the T7-read beyond position 500. In fact this dissimilarity extends to the repeat unit type and number seen in locus 1-2 when compared with that observed in the T7-read (Fig. 39).

Using the T7-read sequence (Fig. 39) in the tRNA scan-SE program we detected the presence of a single $tRNA^{Ser}$ gene (Position 705-789). However, sequence identity between the locus 1-2 nucleotide stretch and the T7-read is negligible at these

positions and a tRNA gene structure could not be identified by the tRNA scan program in locus 1-2.

In an attempt to understand the observed sequence disparity we searched the TIGR genome sequencing database (http://www.tigr.org/tdb/e2k1/eha1/) using the locus 1-2, T7-read and SP6-read sequences. Two of the sequences (GenBank accession # ENTDS45TR and ENTBD56TF) identified by the locus 1-2 sequence and two (GenBank accession # ENTHN33TR and ENTJX69TF) identified by the T7- and SP6- reads were selected and aligned with the existing three sequences (Fig. 39). Hence, the complete fragment 1-2 length as suggested by the sequence alignment between ENTDS45TR, ENTBD56TF and locus 1-2 is 1061 bp while that suggested by alignment between ENTHN33TR, ENTJX69TF, T7- and SP6- reads is 1167 bp. Although the presence of a single band of about this size was expected following the inverse PCR (Fig. 38), the presence of two bands in this size range could not be reconciled with the inverse PCR results.

```
                1                                                   50
ENTHN33TR       GGGGCGGGGG GGCCCGCCAG GGCCCGCGGG GGGGAGGTTT TTAAGAAAGG
               51                                                  100
ENTHN33TR       CCGTGTGGCT GGAAAAGATA CCTTTTTGTA TATCCCCCTT TTACTTTTAT
              101                                                  150
ENTDS45TR       .......... .......... .GCCGCGGTG GGGCC----- GCCCCAGGGC
ENTHN33TR       ACTTTTATAC TTTTATACTT TTATACCTTT ATACCTTTAT ACCTTTATAC
              151                                                  200
ENTDS45TR       CG--ACCCGG TTAGGTCTTT TTTTTGGGTC TGTTTGCTGG GAGTAAATAT
ENTHN33TR       CTTTATACCT TTATATTATC TACTTCTATA TATATATTCC TTCTTATACT
              201                                                  250
ENTDS45TR       TGCAATCCCT TTATATATA- TCTCTTTATA ACT----TTT ATATTTATTC
ENTHN33TR       TTATCCACTT TTATATATAC TCCCTATTTT ACT----TAT ATCCTTATTA
ENTJX69TF       .......... .......... .CCAGGTGTG GCTGGAAAAA ATCCTTATTA
  Inv-SP6       .......... .......... ...CTATTTT ACT----TAT ATCCTTATTA
              251                                                  300
ENTDS45TR       TACCTTATTC TTATCTATAT GTTTATATGT TATTATGT-T ATTATGTTAT
ENTHN33TR       T-CCTTACAC TTTT-TATAC TTTTATTATA TATAATATAT ATTATATTTC
ENTJX69TF       TACCTTACAC TTTT-TATAC TTTTATTATA TATAATATAT ATTATATTTC
  Inv-SP6       T-CCTTACAC TTTT-TATAC TTTTATTATA TATAATATAT ATTATATTTC
              301                                                  350
ENTDS45TR       TATGTTTATA TGTTTACACC TTCATATGAC TACTTCCTTT ATACACTTTA
ENTBD56TF       .......... ......ATCT TTGGGCTGGA AAATTCCTTT -TACACTTTA
ENTHN33TR       TTATTTTCTA T-TTTCTATT TTCTATTTTC TATTTTCTAT TTTATATAT-
ENTJX69TF       TTATTTTCTA T-TTTCTATT TTCTATTTTC TATTTTCTAT TTTATATAT-
  Inv-SP6       TTATTTTCTA T-TTTCTATT TTCTATTTTC TATTTTCTAT TTTATATAT-
              351                                                  400
ENTDS45TR       CTTATTCTTT TA-T----GT TCCTTTTATA TCTTCTATAC ACTTCTTATT
ENTBD56TF       CTTATTCTTT TAAT----GT TCCTTTTATA TCTTCTATAC ACTTCTTATT
ENTHN33TR       CTTATCCTTT ATATCATATA TCATATATTA TTTCCTTTAT GTCCCTTTAT
ENTJX69TF       CTTATCCTTT ATATCATATA TCATATATTA TTTCCTTTAT GTCCCTTTAT
  Inv-SP6       CTTATCCTTT ATATCATATA TCATATATTA TTTCCTTTAT GTCCCTTTAT
              401                                                  450
ENTDS45TR       TATATAT-CT TTCCTCCCCC TTCTTAT--- ---------- -AGAAGTCTT
ENTBD56TF       TATATAT-CT TTCCTCCCCC TTCTTAT--- ---------- -AGAAGTCTT
ENTHN33TR       TATCTTTACA ATACTCCCCT TCCTTTTTAT ACTATTTTAT TAGAAGTCTT
ENTJX69TF       TATCTTTACA ATACTCCCCT TCCTTTTTAT ACTATTTTAT TAGAAGTCTT
  Inv-SP6       TATCTTTACA ATACTCCCCT TCCTTTTTAT ACTATTTTAT TAGAAGTCTT
              451                                                  500
ENTDS45TR       CAATAGTATA GCTGGTTAGT ATCTTCGCCT GTCACGTGAA AGACCGGGGT
ENTBD56TF       CAATAGTATA GCTGGTTAGT ATCTTCGCCT GTCACGTGAA AGACCGGGGT
Locus 1-2       .......... .CTGGTTAGT ATCTTCGCCT GTCACGTGAA AGACCGGGGT
ENTHN33TR       CAATAGTATA GCTGGTTAGT ATCTTCGCCT GTCACGTGAA AGACCGGGGT
ENTJX69TF       CAATAGTATA GCTGGTTAGT ATCTTCGCCT GTCACGTGAA AGACCGGGGT
  Inv-SP6       CAATAGTATA GCTGGTTAGT ATCTTCGCCT GTCACGTGAA ..........
   Inv-T7       .......... .......... .......... .........AA AGACCGGGGT
              501                                                  550
ENTDS45TR       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATG
ENTBD56TF       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATG
Locus 1-2       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATG
ENTHN33TR       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATC
ENTJX69TF       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATC
   Inv-T7       TCGAATCCCC GTTGAAGAGT TCTCTTTTTA TACTTTTATA TGTTTATATC
```

```
            551                                                       600
ENTDS45TR   TTTATATCCT TATTTATTAT TCTTTTATAT TCTTATCACT TCCTACTACT
ENTBD56TF   TTTATATCCT TATTTATTAT TCTTTTATAT TCTTATCACT TCCTACTACT
Locus 1-2   TTTATATCCT TATTTATTAT TCTTTTATAT TCTTATCACT TCCTACTACT
ENTHN33TR   CTTATATGTT TATATGTTTA TATATTATAC TATTTTTTCT TACTA----T
ENTJX69TF   CTTATATGTT TATATGTTTA TATATTATAC TATTTTTTCT TACTA----T
   Inv-T7   CTTATATGTT TATATGTTTA TATATTATAC TATTTTTTCT TACTA----T
            601                                                       650
ENTDS45TR   CTTATTTATT ATCCTTATTA TATCTATTCT TACTCCCTAT CTTTATTATC
ENTBD56TF   CTTATTTATT ATCCTTATTA TATCTATTCT TACTCCCTAT CTTTATTATC
Locus 1-2   CTTATTTATT ATCCTTATTA TATCTATTCT TACTCCCTAT CTTTATTATC
ENTHN33TR   ATTATAACTT TT---TATTA TTTCTATATT TTCTAT-TAT TTCTATTATT
ENTJX69TF   ATTATAACTT TT---TATTA TTTCTATATT TTCTAT-TAT TTCTATTATT
   Inv-T7   ATTATAACTT TT---TATTA TTTCTATATT TTCTAT-TAT TTCTATTATT
            651                                                       700
ENTDS45TR   TTTATTATCT TTATTATCTT TATTATCTTT ATTATCTTTA TTACCTTTAT
ENTBD56TF   TTTATTATCT TTATTATCTT TATTATCTTT ATTATCTTTA TTACCTTTAT
Locus 1-2   TTTATTATCT TTATTATCTT TATTATCTTT ATTATCTTTA TTACCTTTAT
ENTHN33TR   TCTATTATTT CTATTATTTC TATTATTTCT AT-GTTTATA AAAAGATTAC
ENTJX69TF   TCTATTATTT CTATTATTTC TATTATTTCT AT-GTTTATA AAAAGATTAC
   Inv-T7   TCTATTATTT CTATTATTTC TATTATTTCT AT-GTTTATA AAAAGATTAC
            701                                                       750
ENTDS45TR   TACCTTTATT ACCTTTATTA CCTTTATTAT CTTTATTACC TTTATTATAT
ENTBD56TF   TACCTTTATT ACCTTTATTA CCTTTATTAT CTTTATTACC TTTATTATAT
Locus 1-2   TACCTTTATT ACCTTTATTA CCTTTATTAT CTTTATTACC TTTATTATAT
ENTHN33TR   GACAGTGGCA GGATTTGA-A CCTACGCATC CAGAGGATAA TTGATT---T
ENTJX69TF   GACAGTGGCA GGATTTGA-A CCTACGCATC CAGAGGATAA TTGATT---T
   Inv-T7   GACAGTGGCA GGATTTGA-A CCTACGCATC CAGAGGATAA TTGATT---T
            751                                                       800
ENTDS45TR   CTATTCTCAC TTCCTATACG TACTCTTTTT ACTACTCTTT TTACTACTCT
ENTBD56TF   CTATTCTCAC TTCCTATACG TACTCTTTTT ACTACTCTTT TTACTACTCT
Locus 1-2   CTATTCTCAC TTCCTATACG TACTCTTTTT ACTACTCTTT TTACTACTCT
ENTHN33TR   CAAGTCAATC GCCTTAGAC- --CACTCGGC CACACTGTCA TTCTCCCTAT
ENTJX69TF   CAAGTCAATC GCCTTAGAC- --CACTCGGC CACACTGTCA TTCTCCCTAT
   Inv-T7   CAAGTCAATC GCCTTAGAC- --CACTCGGC CACACTGTCA TTCTCCCTAT
            801                                                       850
ENTDS45TR   TCTTACTATA CCTCTTACTA CTCCTACTTT CACCTCCCTC TTTATTGTTA
ENTBD56TF   TCTTACTATA CCTCTTACTA CTCCTACTTT CACCTCCCTC TTTATTGTTA
Locus 1-2   TCTTACTATA CCTCTTACTA CTCCTACTTT CACCTCCCTC TTTATTGTTA
ENTHN33TR   TCTTATTATA CACA--CCTT TATATCCTTT TATGTATATC TTTATATATC
ENTJX69TF   TCTTATTATA CACA--CCTT TATATCCTTT TATGTATATC TTTATATATC
   Inv-T7   TCTTATTATA CACA--CCTT TATATCCTTT TATGTATATC TTTATATATC
            851                                                       900
ENTDS45TR   ATGGGGGTGT AGCTCAGATG GTAGAGCACT CGCTTTGCAT GCGAGGGGCA
ENTBD56TF   ATGGGGGTGT AGCTCAGATG GTAGAGCACT CGCTTTGCAT GCGAGGGGCA
Locus 1-2   ATGGGGGTGT AAG....... .......... .......... ..........
ENTHN33TR   TTTATACT-T CTATCCTTTT ATGTATATCC CTTTATACCT TTATA-----
ENTJX69TF   TTTATACT-T CTATCCTTTT ATGTATATCC CTTTATACTT TTATA-----
   Inv-T7   TTTATACT-T CTATCCTTTT ATGTATATCC CTTTATACTT TTATA-----
            901                                                       950
ENTDS45TR   AGGGGATCGA TACCCCTCAC CTCCATTATT ATCTATTATT CCCTTTTATT
ENTBD56TF   AGGGGATCGA TACCCCTCAC CTCCATTATT ATCTATTATT CCCTTTTATT
ENTHN33TR   -------CTT TTATACTTTT ATACTTTTAT ACCC-TTATA CCTTTATACC
ENTJX69TF   -------CTT TTATACTTTT ATACTTTTAT ACCT-TTATA CCTTTATACC
   Inv-T7   -------CTT TTATACTTTT ATACTTTTAT ACCT-TTATA CCTTTATACC
            951                                                      1000
ENTDS45TR   CTTTTACTTT ATTTCTTATA TCTTATATTC TCTTTTATAT ACTTTAGTAT
ENTBD56TF   CTTTTACTTT ATTTCTTATA TCTTATATTC TCTTTTATAT ACTTTAGTAT
ENTHN33TR   TTTATACCTT TATACCTTTA TATTATCTAC TTCTATA-AT ATATTC--CC
ENTJX69TF   TTTATACCTT TATACCTTTA TATTATCTAC TTCTATATAT ATATTC--CT
   Inv-T7   TTTATACCTT TATACCTTTA TATTATCTAC TTCTATATAT ATATTC--CT
```

```
            1001                                                  1050
ENTDS45TR   TCTACTTCCC TATTCT-TAT TATATATAAT TTTTATGGAT ATATATATAT
ENTBD56TF   TCTACTTCCC TATTCT-TAT TATATATTAT TTTTATGTAT ATATATATAT
ENTJX69TF   TCTTATACTC TATCCACTTT TATATATACT CCCTAT-TTT ACTTATATCC
   Inv-T7   TCTTATACTT TATCCACTTT TATATATACT CCCTAT-TTT ACTTATATCC
            1051                                                  1100
ENTDS45TR   ATATTAATAT TCCCTTCTAT TTATTTCTTA TTCA...... ..........
ENTBD56TF   ATATTTATAT TTCCTTCTAT TTATTCTTTA TTCCCTTTTT ATATATGTGT
ENTJX69TF   TTATTATCCT TACA--CTTT TTATACTTTT ATTATATATA ATATATAT-T
   Inv-T7   TTATTATCCT TACA--CTTT TTATACTTTT ATTATATATA ATATATAT--
            1101                                                  1150
ENTBD56TF   ATACTTATAC TTATATATAT TATATTTATG TACCTTACTT ATATAAAAAT
ENTJX69TF   ATATTCCTTA TTTTCTATTT TCTATTTTC- -TATTTTCTA TTTTCCATTT
   Inv-T7   ATATTCCTTA TTTTCTATTT TCTATTCTC- -TATTTTCTA TTT.......
            1151                                                  1200
ENTBD56TF   TATATTCCTT CTTTTTATAT TCTTTATACT TTTATATTCT TTTCTAATTT
ENTJX69TF   TATATATCTA TCCCTTATAT CATATATCAT ATAATAAT-- TCCCTTATGT
            1201         1215
ENTBD56TF   TTATTCTTTA TACCT
ENTJX69TF   CCCTTTATTA A....
```

Fig. 39. Nucleotide Sequence Alignment of Locus 1-2 Sequence with Inverse T7- & SP6-reads. Dots indicate start- and end-points of readable sequences. Dashes have been used to maximise alignment. Regions of maximum identity are boxed. Bold-face and underlined bold-face regions show the two types of repeats as seen in locus 1-2 (Fig. 4A). Double underlined regions show the number and orientation of repeat units seen in the T7-read. ENTDS45TR and ENTBD56TF sequences were obtained from the TIGR genome sequencing database using locus 1-2 sequence (Fig. 4A) as the search template. ENTHN33TR and ENTJX69TF sequences were obtained from the TIGR genome sequencing database using the Inv-T7 and Inv-SP6 sequences as the search templates (highlighted pale grey). Repeat units seen in the 1167 bp alignment (highlighted pale grey) are indicated on the 33TR and/ or the 69TF strand (underline, dot dot dash, wave and double underline). The three tRNA genes are highlighted (dark grey).

However, a closer look at the 1167 bp sequence alignment revealed that the nucleotide stretch between positions 94 and 165 contains two types of related direct repeats arranged in tandem (Fig. 39, repeats shown on the 33TR strand & Fig. 40A), which are reiterated between nucleotide positions 888 and 972 (Fig. 39, repeats shown on the 69TF strand; Fig. 40A; Appendix IID).

Similarly tandem repeats were also seen between positions 308 and 343 (Fig. 39, repeats shown on the 33TR strand & Fig. 40A) which were reiterated between positions 1115 and 1144 (Fig. 39, repeats shown on the 69TF strand & Fig. 40A). This data therefore suggests that the 1167 bp alignment contains one and a half

repeats of a ca. 770 bp sequence, which coincides well with the inverse PCR results (Fig. 38).

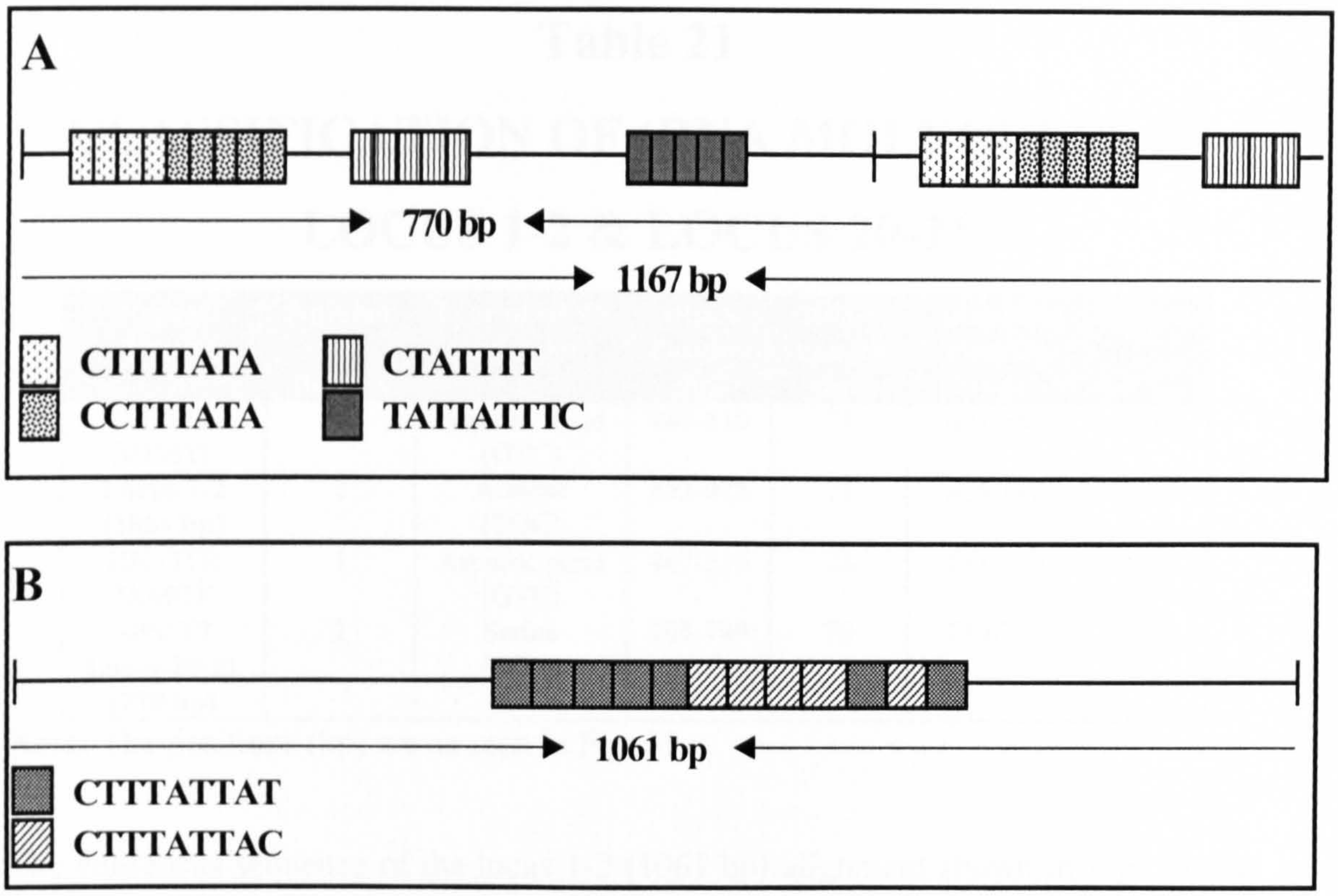


Fig. 40. Schematic Representations of Locus 1-2 & 20-21. (A) 770 bp Sequence (Locus 20-21). The repeat structure shown is derived from alignment of ENTHN33TR, ENTJX69TF, T7- and SP6- reads. Shown is the one(+) 770 bp unit. **(B) 1061 bp Sequence (Locus 1-2)**. The repeat structure shown is derived from alignment of ENTDS45TR, ENTBD56TF and locus 1-2 (Fig. 4A) sequences. Variations in number, sequence and arrangement of repeat units are shown.

Using the tRNA scan-SE program we detected the presence of two tRNA genes each in the four TIGR sequences (Table 21). In all cases one of the two genes is a $tRNA^{Asp}$ gene, as expected from the partial sequence identity seen using the locus 1-2 sequence (Fig. 34A; Table 18). However, the second gene identified is that of $tRNA^{Ala}$ (Fig. 39; Table 21) in the 1061 bp fragment and $tRNA^{Ser}$ in the 770 bp fragment. Hence it appears that the $tRNA^{Asp}$ gene that is characteristic of locus 1-2 occurs in two completely different types of arrays (see discussion). Thus while the 1061 bp sequence represents an almost complete locus 1-2 fragment, the 770 bp sequence is a distinct repeated-DNA element which we have called locus 20-21 and will refer to as such henceforth (Fig. 40A; Table 21). Schematic representations of the organisation

and arrangement of locus 1-2 and locus 20-21 repeats and associated tRNAs are shown in Appendix XVI.

Table 21

CLASSIFICATION OF tRNA MOLECULES IN LOCUS 1-2 & LOCUS 20-21

Repeat Locus	Number of tRNA's	Type of tRNA (anticodon)	tRNA Position (bp)	tRNA Length (bp)	Anticodon Position (bp)	Intron
DS45TR BD56TF	1	Aspartic Acid (GTC)	447-519	73	481-483	None
Locus 1-2 (1061 bp)	2	Alanine (TGC)	853-925	73	886-888	None
HN33TR JX69TF	1	Aspartic Acid (GTC)	447-519	73	481-483	None
SP6/ T7 **Locus 20-21 (770 bp)**	2	Serine (TGA)	705-789	79	750-752	None

Note: The positions (bp) are as seen in Fig. 39.

The consensus sequence of the locus 1-2 (1061 bp) alignment shown in Fig. 39 was then used as a search template to look for corresponding fragments in the TIGR genome sequencing database (http://www.tigr.org/tdb/e2k1/eha1/). Multiple alignment was carried out between our consensus sequence and some of the sequences identified from the database (data not shown) and the complete locus 1-2 unit length was actually found to be 1158 bp.

7.2.5. SOUTHERN HYBRIDISATION ANALYSIS

To assess the genomic distribution and copy number of the repeat loci, Southern hybridisation analysis was performed in the first instance using all five locus-specific probes and HM-1:IMSS clone 9 genomic DNA digested with five enzymes. Results seen with the locus 1-2 and 5-6 probes are shown (Fig. 41). Table 22 summarises the results seen on the filter with the five probes.

With the locus 1-2 specific probe, the band of ca. 400 bp in the *Alu*I lane was expected since the clone was obtained from an *Alu*I restriction fragment of about the same size (section 3.3.1). It was a surprise to find that the probe gave intense hybridisation signals at ≥23kb with the *Dra*I digested DNA as this enzyme cuts frequently in *E. histolytica* DNA and usually produces much smaller fragments as seen in the ethidium bromide stained gel. It is notable that *Alu*I, *Dde*I and *Rsa*I give major fragments of about the same size and that *Dra*I and *Eco*RI both give very large fragments.

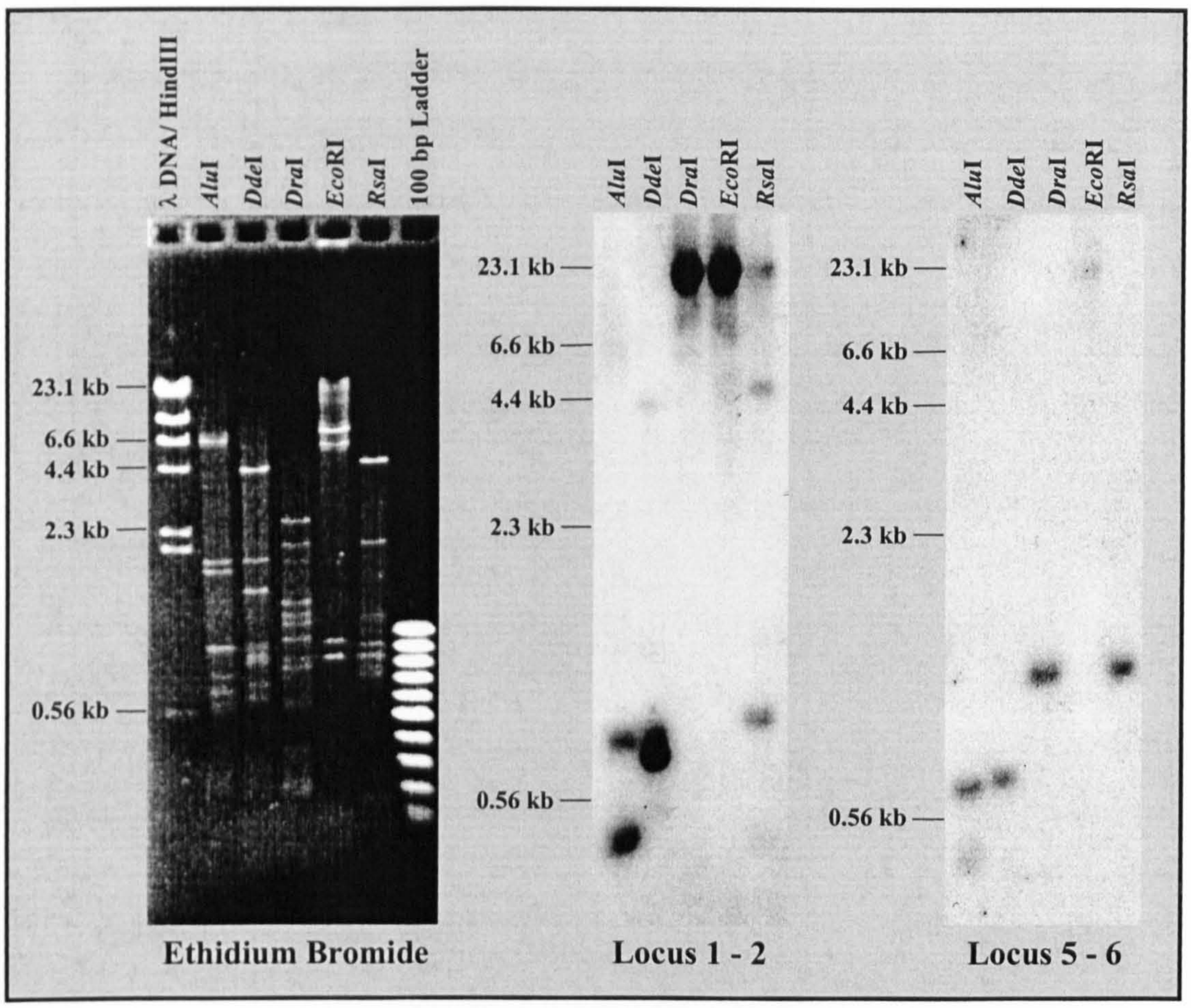


Fig. 41. Southern Hybridisation Analysis of Loci 1-2 & 5-6. Genomic DNA of *E. histolytica* strain HM-1:IMSS clone 9 was digested with restriction enzymes. The digested products were electrophretically separated in a 0.8% agarose gel, stained with ethidium bromide and blotted on to a nylon membrane. The membrane was hybridised with five [α-^{32}P]-dCTP labelled locus-specific probes. Hybridisation results with the locus 1-2 and locus 5-6 probes are shown. The λ DNA/ *Hin*dIII size marker was used to measure the sizes of the hybridisation products.

Table 22

SOUTHERN HYBRIDISATION ANALYSIS WITH FIVE LOCUS - SPECIFIC PROBES

Loci	Restriction Enzyme	Restriction Sites in the Probe	Number of Bands on the Filter	Size of Bands (kb)
Locus 1-2	*Alu*I	none	3	6.5 0.70 0.44
	*Dde*I	none	3	4.1 0.66 0.53
	*Dra*I	none	2	≥23 9.4
	*Eco*RI	none	2	≥23 7.8
	*Rsa*I	1	4	23 4.9 0.78 0.40
Locus 5-6	*Alu*I	none	2	0.62 0.40
	*Dde*I	2	1	0.66
	*Dra*I	1	1	1.0
	*Eco*RI	none	1	23
	*Rsa*I	1	1	1.0
Locus 3-4	*Alu*I	1	1	0.86
	*Dde*I	3	1	0.45
	*Dra*I	none	1	≥23
	*Eco*RI	none	3	3.0 1.85 0.9
	*Rsa*I	4	2	0.5 0.43
Locus 9-4 *	*Alu*I	1	1	0.86
	*Dde*I	3	1	0.45
	*Dra*I	none	1	≥23
	*Eco*RI	none	3	3.0 1.9 0.94
	*Rsa*I	4	2	0.54 0.43
Locus 16-17	*Alu*I	3	1	0.84
	*Dde*I	2	1	0.86
	*Dra*I	none	1	≥23
	*Eco*RI	none	1	≥23
	*Rsa*I	none	1	≥23

Note: "*" This information is based on the restriction map of sequence 9-4 as shown in Appendix VIII.

Similarly, a band of ca. 400 bp was expected in the *Alu*I lane with the locus 5-6 specific probe, as the clone was obtained from an *Alu*I restriction fragment of about this size (section 3.3.1). Once again hybridisation with this probe gave major fragments of about the same size with *Alu*I and *Dde*I and with *Dra*I and *Rsa*I. *Eco*RI again produces a very large fragment of ≥23kb.

From the sequences, bands of ca. 900 bp (Table 22) were expected in the *Eco*RI lanes with both locus 3-4 and 9-4 probes since in both cases *Eco*RI was the delimiting enzyme used to isolate the loci (Appendix IIA).

The presence of multiple hybridisation bands with some of the restriction enzymes in the absence of restriction sites in the probes suggests that these repeat loci may be multicopy. In addition, the large fragments generated by frequent cutters like *Dra*I and *Eco*RI and the presence of major bands of about the same size with two or more enzymes suggests that these loci are arranged in long tandem arrays. In order to confirm the tandem array arrangement we decided to see whether partial restriction digests generated ladder patterns.

To this end genomic DNA of strain *E. histolytica* HM-1:IMSS clone 9 was partially digested with either *Alu*I or *Dde*I and Southern hybridisation analysis was performed with all five locus-specific probes. Although we failed to achieve optimum digestion with *Alu*I, ladder patterns were observed with all five probes for the *Dde*I digested DNA. Results seen for two of the loci, locus 3-4 and 16-17, are shown in Fig. 42. Table 23 summarises the results seen on the filter with all five probes.

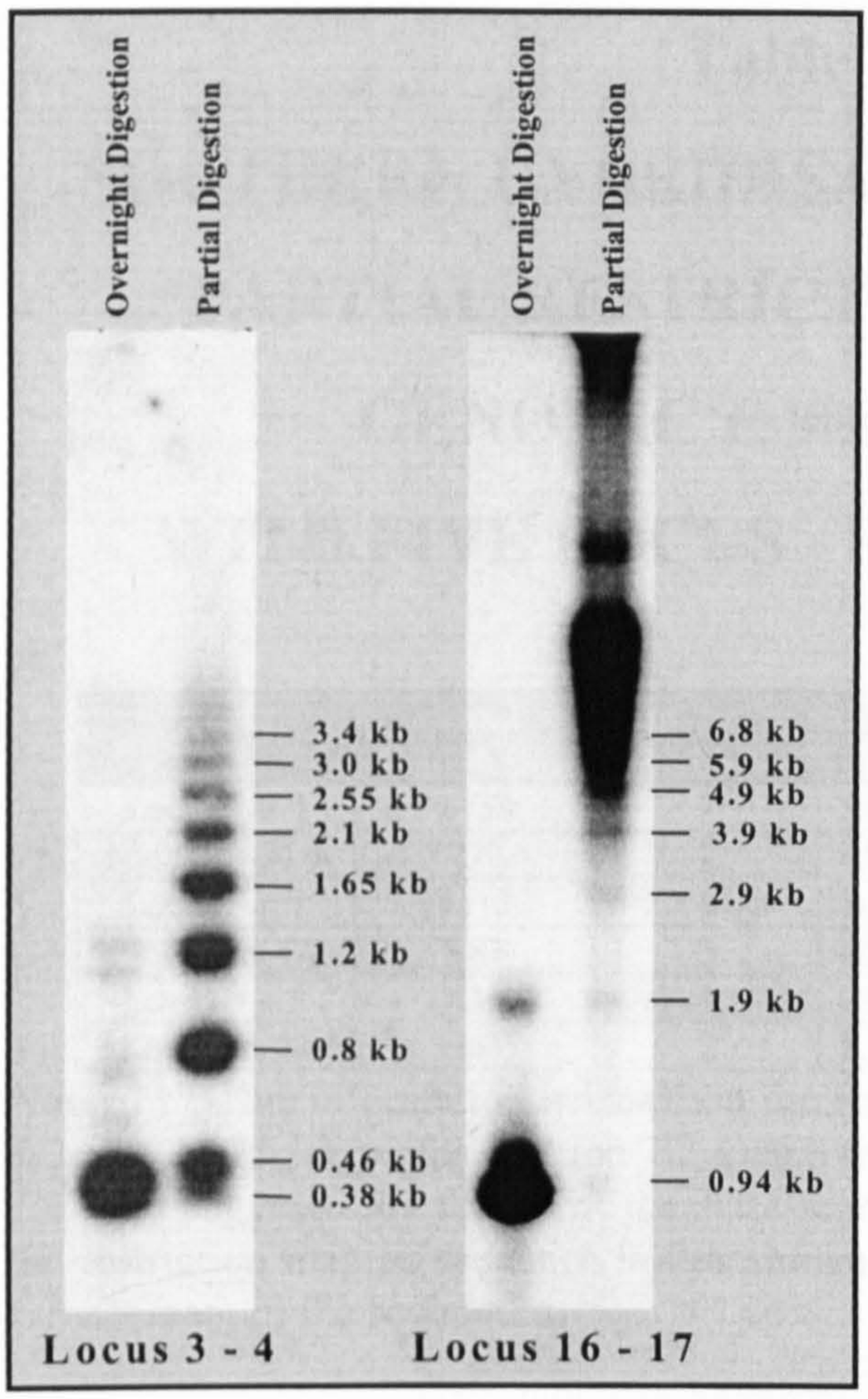


Fig. 42. Southern Hybridisation Analysis of Partial Restriction Digests. Genomic DNA of *E. histolytica* strain HM-1:IMSS clone 9 was digested with restriction enzymes. The digested products were electrophretically separated in a 0.8% agarose gel, stained with ethidium bromide and blotted on to a nylon membrane. The membrane was hybridised with five [α-^{32}P]-dCTP labelled locus-specific probes. λ DNA/ *Hin*dIII was used as size marker **Locus 3-4 & Locus 16-17.** Hybridisation results with *Dde*I digested DNA are shown.

Size measurement of the hybridising bands seen for locus 3-4 revealed the presence of approximately eight discrete bands with intervals of ca. 400-450 bp each (Fig. 42; Table 23B). Size differences between alternate bands were nearer ca. 850-900 bp each. These results correlate well with the fact that digestion of the locus 3-4 fragment with *Dde*I results in two major products of ca. 441 and 486 bp (Table 23A). Hence, if locus 3-4 is indeed arranged in long tandem arrays we would expect to see a ladder in which the band intervals were equal to multiples of ca. 450 bp, which is in fact the case.

By the same argument the results at locus 5-6, 9-4 and 16-17 are in full agreement with the sizes of the major *Dde*I restriction fragments generated (Table 23A & B). At locus 16-17 the single major restriction fragment (unit length 959 bp) is only 5 bp smaller than the full-length locus sequence and hence the observed ladder pattern with intervals of ca. 1 kb (Fig. 42) is to be expected.

Table 23

SOUTHERN HYBRIDISATION ANALYSIS OF PARTIAL RESTRICTION DIGESTS OF GENOMIC *E. histolytica* DNA WITH FIVE LOCUS - SPECIFIC PROBES

A

Loci	Total Length (bp)	Number of *Dde*I Sites	Site Position (nucleotide #)	Expected Fragment Lengths (bp)
Locus 1-2 †	1158	2	562, 1032	470, 687
Locus 20-21 †	770	1	654	770
Locus 5-6 †	1112	4	168, 222, 914, 919	54, 692, 5, 361
Locus 3-4	983	3	46, 532, 973	486, 441, 56
Locus 9-4 *	931	3	39, 511, 927	472, 416, 42
Locus 16-17	964	2	8, 13	5, 959

Note: "†" This information is based on the restriction map of the complete locus 1-2 & locus 20-21 fragments (section 7.2.4) and the complete locus 5-6 fragment (section 7.2.3) as suggested by the TIGR database searches. **"*"** This information is based on the restriction map of sequence 9-4 as shown in Appendix VIII. All assume tandem arrangement of the sequences.

B

	Locus 1-2 (kb)	Locus 5-6 (kb)	Locus 3-4 (kb)	Locus 9-4 (kb)	Locus 16-17 (kb)
***Dde*I Partial Digests**	3.2	3.2	3.4	4.8	6.8
	2.85	2.8	3.0	3.9	5.9
	2.5	2.3	2.55	3.4	4.9
	2.1	2.0	2.1	2.9	3.9
	1.65	1.65	1.65	2.4	2.9
	1.45	1.3	1.2	1.9	1.9
	1.3	1.0	0.80		0.94
	1.05	0.80	0.46		
	0.82	0.68	0.38		
	0.64				
	0.46				
	0.4				

C

	Locus 1-2 (kb)	Locus 5-6 (kb)	Locus 3-4 (kb)	Locus 9-4 (kb)	Locus 16-17 (kb)
***Dde*I Control**	0.67	0.62	1.25	0.94	1.9
	0.62		1.15	0.50	1.05
	0.41		0.74		0.88
			0.60		
			0.42		

Hybridisation with the locus 1-2 specific probe resulted in a large number and complex pattern of bands (Table 23B). Ladder patterns with intervals of ca. 687 bp and 470 bp or multiples thereof were expected. However, the presence of smaller bands in the range of 0.4 to 1.3 kb (Table 23B) was unexpected.

Results of Southern hybridisation analysis have suggested that these loci are multicopy (this chapter & this section). It is therefore possible that due to nucleotide sequence variation some copies of a repeat locus may contain additional *Dde*I sites (see discussion). It has been shown that sequence variation can occur within a given locus in the same strain (section 3.4). This could therefore explain the presence of smaller bands superimposed on the expected ladder pattern seen at locus 1-2.

It is not certain whether the extra band in the ca. 0.62 to 0.67 kb range observed with the overnight digested *Dde*I controls at locus 1-2 (Table 23C) is due to better resolution of DNA on the agarose gel as compared to before (Table 22) or reflects cross hybridisation due to the AT-rich nature of these loci. In case of the latter, variation in exposure time would make it possible for the extra/ spurious band to be visible in one instance but not the other. It is noteworthy that the signal for the ca. 0.67 kb band was weaker as compared to that of the ca. 0.62 kb band (data not shown).

As the ca. 100 bp stretch between nucleotide positions 461 and 549 (Fig. 39) is identical in both locus 1-2 and locus 20-21 it raises the possibility that the locus 1-2 specific probe may cross-hybridise with the locus 20-21 DNA repeat element. The fact that bands of ca. 770 bp are expected and can be seen with both *Alu*I and *Dde*I (deduced from the locus 20-21 restriction map (Table 23 & data not shown)) in the Southern hybridisation results (Fig. 41; Table 22) suggests that cross hybridisation is in fact taking place. While an *Alu*I fragment of about this size (ca. 756 bp) is also expected from locus 1-2 (data not shown), it is outside of the probe region and will not hybridise. The ca. 770 bp band seen must therefore be derived from locus 20-21. The size of the expected *Dde*I fragment from locus 1-2 is quite different from that of locus 20-21 i.e it is ca. 687 bp vs 770 bp (Table 23). This difference of ca. 100 bp is enough to have resulted in two bands on the Southern blot if cross-hybridisation was occurring, which is the case (Fig. 41; Table 22).

As in the case of the faint *Dde*I band seen for locus 1-2 all four of the extra bands seen with locus 3-4 (Table 22 & 23C) have significantly weaker signal intensities when compared to the expected band of ca. 0.42 kb (Fig. 42). This may reflect cross-hybridisation or incomplete digestion. However, the fact that these repeat loci are most likely multicopy and appear to be arranged in long tandem arrays gives rise to another possible explanation for the appearance of the faint bands. It is possible that these bands result from the probe hybridising to fragments immediately adjacent to the ends of the full array. Since each end fragment, is equivalent to a single copy gene, the signal would be expected to be faint.

At locus 16-17 the two bands of ca. 0.88 and 1.05 kb may again reflect differences in resolution of the ca. 0.86 kb band seen earlier (Table 22). Interestingly, while intensity of the ca. 1.9 kb band is again weaker than that of the other two it is identical in size to one of the bands seen in the ladder pattern generated by the partial digest (Table 23B & C; Fig. 42). This is suggestive of incomplete digestion of the control DNA sample. Likewise, at locus 9-4 while the band of ca. 0.50 kb is expected (Table 22 & 23C), the ca. 0.94 kb band which is of the same intensity (data not shown) reflects the size of the complete locus 9-4 fragment (Table 23A) hence suggesting incomplete digestion.

7.2.6. PULSED-FIELD GEL ELECTROPHORETIC ANALYSIS

Having ascertained that the repeat loci are most likely multicopy as well as being arranged in long tandem arrays we decided to try determining their chromosomal distribution and full tandem array lengths. For this purpose pulsed-field gel electrophoresis (PFGE) followed by Southern hybridisation was performed with both undigested and restriction digested genomic DNA of HM-1:IMSS clone 9. Results when undigested chromosomal DNA was hybridised to all five locus-specific probes are given in Table 24. Fig. 43 shows the results of both undigested and *Eco*RI digested chromosomal DNA obtained with the locus 16-17 specific probe.

Table 24

SOUTHERN HYBRIDISATION ANALYSIS OF PULSED-FIELD GEL ELECTROPHORESIS OF GENOMIC *E. histolytica* DNA WITH FIVE LOCUS - SPECIFIC PROBES.

Loci	Number of Bands on the filter	Size of Bands (kb)
Locus 1-2	1	>2,200
Locus 5-6	2	>2,200; 1,250
Locus 3-4	1	1,500
Locus 9-4	1	1,600
Locus 16-17	5	>2,200; 1,050; 680; 560; 400

As seen from the results in Table 24 all five probes hybridised to one or more undigested chromosomal DNA bands, with the locus 16-17 specific probe hybridising to as many as five discrete bands (Fig. 43).

At first glance we are unable to map the five loci to the fourteen linkage groups described by Willhoeft and Tannich, 1999 (Appendix XV). However, by assuming a certain degree of under or over estimation in our size measurements and in the absence of additional information on strains 200:NIH and HK-9 the locus 1-2 fragment may be mapped to linkage group 13 and by the same argument loci 3-4 and 9-4 can be assigned to linkage group 8. This, however, does not corroborate previously reported results. It has already been mentioned that both locus 3-4 and 9-4 share high degrees of sequence identity with each other as well as with the ca. 1 kb repeat element (designated M11) reported by Willhoeft and Tannich (2000) (section 3.4; Appendix IX). Using the M11 fragment as a probe the authors were able to map it to linkage groups 12 and 13 (Appendix XV).

Additionally, by presuming that complete separation had not occurred, the locus 5-6 fragment could be assigned to linkage group 12, while the ca. 2,200 and 1,050 kb bands seen at locus 16-17 could be mapped to linkage group 14. Alternatively, the 1,050 band could represent a mixture of the three bands seen in linkage group 4

(Appendix XV). In either case the remaining bands for locus 16-17 could not be assigned with confidence

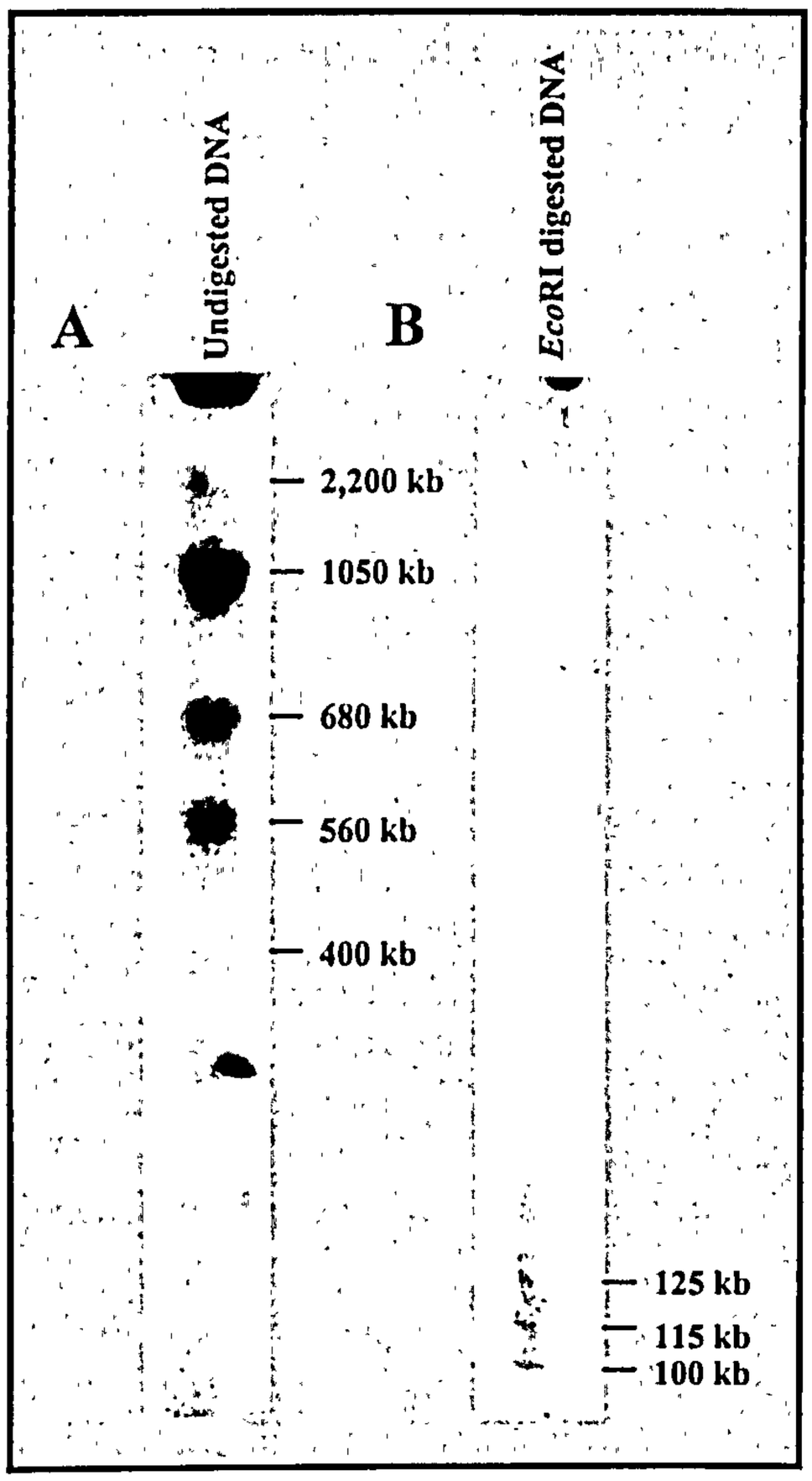


Fig. 43. Southern Hybridisation Analysis of Pulsed - Field Gel Electrophoresed DNA with a Locus 16-17 Specific Probe. **(A)** Undigested chromosomal DNA **(B)** *Eco*RI digested chromosomal DNA. Electrophoresis was carried out in 1% agarose gels with 0.5 X **(A)** or 0.25 X **(B)** TBE running buffer at 14°C. Separation parameters were the default values for the separation range of 50 kb-1.0Mb **(A)** and 20 kb-350 kb **(B)** Chromosomal DNA of *Saccharomyces cerevisiae* and λ DNA concatemers were used as size markers.

To see if we could make conclusive linkage group identification, we sent some of our locus 16-17 PCR product to Drs Willhoeft and Tannich to be used as a probe for their PFGE blots (Willhoeft and Tannich 1999). The locus 16-17 PCR product was gel purified and labelled by random priming and hybridisation was done at 55°C with non-stringent washes at 55°C in 2 X SSC followed by more stringent washes at 68°C in 2 X SSC buffer (personal communication).

Two clones of HM-1:IMSS and the original (wild type) strain of HM-1:IMSS all gave identical banding patterns with strong bands at 1,100, 1,050, 1,000 and 470 kb (data not shown). Weak bands that washed off at higher stringency appeared at >2,000, 770 and 700 kb. If we can assume that our earlier arguments regarding under or over estimation of sizes as well as incomplete electrophoretic separation are justified, a

possible correlation can be shown between our results and those obtained by Willhoeft and Tannich. Hence our 400, 560, 680 and 2,200 kb bands could in fact represent the 470, 700, 770 and >2,000 kb bands seen by Willhoeft and Tannich. The 1,050 band, which gives the strongest signal on our filter, could in fact represent a mixture of the 1,100, 1,050 and 1,000 kb bands seen in Germany.

Using additional information on strains 200:NIH and HK-9, the bands around 1,100, 1,050 and 1,000 kb could be assigned to linkage group 4. This was confirmed by hybridisation of a single copy cDNA reference probe *ehsod* (iron-containing superoxide dismutase; accession # M63816) to the same blot (personal communication; data not shown). However, the band at 470 kb with strain HM-1:IMSS and additional bands seen with strains 200:NIH and HK-9 (data not shown) could not be assigned. There are a number of possible explanations for this and these will be reviewed below (section 7.3).

Digestion of chromosomal DNA in agarose blocks with an enzyme that does not cut locus 16-17 suggests that the full tandem array length is between 100 and 125 kb (Fig. 43B). Hence, there may be an estimated 100 tandemly arranged copies of this repeat locus in a given array (unit length of locus 16-17 is 964 bp; Table 23A). Correlation between the three restriction fragments seen in Fig. 43B and one or other of the chromosomal bands seen in Fig. 43A is difficult. However, it is noteworthy that these arrays constitute a major part of the respective chromosome i.e. the 100 kb restriction fragment represents about 25% of the smallest (400 kb) chromosomal band and around 10% of the large (1,050 kb) band identified in the undigested chromosomal DNA.

7.3. DISCUSSION

Characterisation of all five repeat loci revealed some interesting features. A relatively early and unusual finding was the association of various tRNA genes with all three of the previously reported repeated-DNA containing elements, i.e. 3-4, 9-4 and 16-17 and the partial sequence identity to different tRNA species for both locus 1-2 and 5-6 which were isolated by us. The sizes of the total RNA hybridisation bands in northern

blot results with all five locus-specific probes also corresponded to the tRNA region of the ethidium bromide stained gel. The presence of abundant stop codons in all the potential reading frames of locus 1-2 and 5-6 and the absence of hybridisation signals in the respective polyadenylated RNA fractions strengthens the view that these repeated-DNA fragments probably do not code for any significant proteins.

Locus 1-2 and 5-6 repeat-flanking regions were isolated and characterised in an attempt to complete and confirm the presence of the putative tRNA gene sequences. Results revealed that both these loci are also associated with tRNA genes. There appear at present to be only three reports pertaining to the presence of tRNAs in *E. histolytica* (Hernandez-Velarde *et al*, 1980; Albach *et al*, 1984 and Hernandez *et al*, 1986). However none of these reports are related to the tRNA gene structure or arrangement.

Schematic representations of all the tRNA genes identified in this study and the repeat loci they are associated with are shown in Appendix XVI. In all cases the major blocks of tandem repeats seen at each of the loci (Fig. 4B, 5B & 8) appear to be located in the inter- tRNA spacer regions. Furthermore, as seen at loci 1-2 and 20-21, it appears that the same tRNA gene may be associated with more than one type of tandemly repeated-DNA.

The presence of repeated sequences of variable lengths in between coding regions has been observed in diverse organisms. Intergenic (IG) regions or spacers are often internally repetitious. Length heterogeneity apparently occurs due to varying numbers of repeated elements and may even be observed between different spacers in the same cluster of genes (Long and Dawid, 1980). The rRNA and 5S RNA genes of *Xenopus* and the rRNA genes of *Leishmania* and *Drosophila* display considerable spacer length variation.

The inter- tRNA gene spacers represented in our loci 1-2, 5-6, 3-4, 9-4 and 16-17 also show length variation due to differences in numbers of individual repeat units within the major repeat blocks as has been seen from PCR fragment size polymorphism analysis and nucleotide sequence comparisons (Chapter 3). Evidence also exists that sequence and repeat number variation occurs among different copies of an array unit

in the same strain (observed for locus 9-4 in strain HM-1:IMSS (section 3.4; Appendix IX) and for locus 5-6 in strain H-303:NIH (section 3.4; Fig. 7B).

The fact that all five repeat loci described by us were associated with tRNA genes, with the internal tandem repeat units located in the inter-gene spacer regions, seemed highly significant and raised the interesting possibility that other repeated sequences associated with tRNA genes could occur in the *Entamoeba* genome. To this end the data emerging from the *E. histolytica* genome project (http://www.tigr.org/tdb/e2k1/eha1/) were examined using the tRNA scan-SE program to look for the presence of additional tRNA genes and to see if their association with short tandem repeats (STRs) was a general phenomenon. Analysis of the first 49,000 sequences revealed that in fact 13% of all sequence reads contain tRNA genes (data not shown). Numerous reads had more than one tRNA gene, sometimes of the same isoacceptor type but more often distinct, and almost all the inter- tRNA spacer regions contained blocks of AT-rich STRs.

The presence of tRNA genes in repeat units has rarely been reported for other eukaryotes as well. In *Xenopus laevis* two $tRNA^{Met}$ genes lie within a tandemly repeated-DNA sequence separated by a spacer region (Clarkson *et al*, 1978). However, the tandemly repeated units are of identical length (3.18 kb) and there is no indication of short repeats within the spacers.

The presence of large fragments generated by frequent cutters like *Dra*I and *Eco*RI and the major bands of about the same size seen with two or more enzymes (Fig. 41; Table 22) confirmed that all these loci are likely arranged in long tandem arrays. This observation was further corroborated by the fact that in partial restriction digest experiments with frequent cutters ladder patterns were generated in all cases (Fig. 42; Table 23B). Analysis of the *E. histolytica* genome project data (see above) also revealed that these genes are usually organised into long tandem arrays. In fact only four of the forty-six tRNA isoacceptor types identified were not found in arrays and these four were present in only a few sequence reads. The size of the individual array units varies and from one to five tRNAs may be encoded by a single array unit. Furthermore, some array types display more than one unit length due to variable numbers of STRs in different units of the same array type i.e. locus length

polymorphism exists within a strain which corroborates the observations and inferences made earlier in this study (see above).

Attempts at ascertaining the chromosomal distribution of these loci were not entirely successful and conclusive linkage groups identifications based on a previously published scheme (Appendix XV; Willhoeft and Tannich, 1999) could not be made. Furthermore, efforts to get more reliable information using the PFGE blots previously generated by Drs Willhoeft and Tannich also revealed unexpected results. Only three of the four strong bands on the filter with the 16-17 probe could be assigned to one of the existing linkage groups. The linkage group proposed by them, i.e. number 4, consists of three closely spaced bands with differences of only ca. 50 kb between them and is one of the two possible assignments suggested by our results. Taken together these data indicate that the separation achieved by us may in fact have been insufficient and short of what was achieved by Willhoeft and Tannich in Germany on a different apparatus.

In both the results seen by us as well as those observed in Germany hybridisation with the locus 16-17 specific probe resulted in bands which could not be assigned to any of the existing linkage groups. There are a number of possible explanations for this. These bands may represent additional chromosomes related to linkage group 4 that were not detected with the probes used in the original mapping study. Alternatively they could be a subset of linkage group 1 chromosomes which were not previously identified as this linkage group was only tagged by three multicopy probes. This in turn would indicate that not all allelic chromosomes carry the same arrays. Finally they could represent a new and so far undetected linkage group.

Interestingly two of forty-six tRNA isoacceptor types identified contained introns ($tRNA^{IleTAT}$ (10 bp) and $tRNA^{TyrGTA}$ (13 bp)) (data not shown). Until recently it was believed that intron-containing genes are rare in *E. histolytica* as only five genes had been identified up until 1999 (Lohia and Samuelson, 1993; Plaimauer *et al*, 1994; Urban *et al*, 1996; Sánchez-López *et al*, 1998; Willhoeft *et al*, 1999a). Recently however nine additional intron-containing genes have since been identified indicating that introns are more common in this organism than previously suggested (Willhoeft *et al*, 2001). Three of the introns are present in genes encoding ribosomal proteins,

while the others include cysteine proteinase, chitin synthase and nuclear-binding protein genes.

However, these introns present in protein coding genes belong to a different class from those found in tRNA genes. Thus, the two intron sequences described here represent the first report of this class of introns from *E. histolytica*.

The presence of a 13 bp intron in the $tRNA^{Trp}$ gene from *D. discoideum* has been described (Peffley and Sogin, 1981). The intron is located one base pair away from the 3′ residue of the anticodon. The 13 bp long intron in the $tRNA^{Tyr}$ gene present in locus 9-4 and the 17 bp long intron in the $tRNA^{Leu}$ gene seen in locus 3-4 are also located one nucleotide away from the 3′ residues of the respective anticodons.

In conclusion it appears that the *E. histolytica* genome has several unusual features. The previously reported circular organisation of ribosomal RNA genes which make up about 15% of the total cell DNA (Bhattacharya *et al*, 1998), the presence of less abundant classes of nuclear DNA circles (Lioutas *et al*, 1995; Dhar *et al*, 1995) and identification of several families of transcribed genes that do not encode proteins (Bhattacharya *et al*, 1999) are some examples. In addition it now appears that the tRNA genes are also uniquely organised. It is likely that completion of the *Entamoeba* genome project will yield even more unusual features.

CHAPTER 8

CONCLUDING REMARKS & FUTURE PROSPECTS

Molecular based approaches have proven valuable for the study of epidemiology especially in the area of diagnosis and inter- and intra-species diversity (Thompson *et al*, 1998). PCR in particular is an important tool as it obviates the need to culture. The ability to detect and characterise genetic variability of infectious agents and in fact to discriminate between individual isolates or genotypes has allowed researchers to carry out molecular epidemiological studies on a variety of parasites, including *Giardia* (Hopkins *et al*, 1999), *Trypanosoma* (Macedo and Pena, 1998), *Leishmania* (Russell *et al*, 1999) and *Plasmodium* (Ranford-Cartwright *et al*, 1997). Scientists can now monitor patterns of disease transmission, determine the sources of an infection outbreak, deduce associations between a genotype and the course of clinical disease, detect infection of a single host with a mixture of genetically distinct strains, and even discriminate between recrudescence and re-infection.

The confusion that exists with regards to the epidemiology of amoebiasis has been reviewed in Chapter 1 and several unanswered questions highlighted. Chief among these is what determines the outcome of *E. histolytica* infections. The final outcome of an infection could be dependent on a variety of factors related to the host, the parasite and the environment. With respect to the parasite, attempts are underway to delineate individual strains so as to study the role of parasite variation in the outcome of disease (Clark and Diamond, 1993a; Ghosh *et al*, 2000; Som *et al*, 2000). The most polymorphic markers identified prior to this work were two protein coding genes, SREHP and chitinase (section 1.5.2). However, the SREHP gene is an immunodominant molecule that is a vaccine candidate (Huston and Petri, 1998) and hence selection rather than population history may be determining the observed polymorphism (Hastings, 1996). Clearly, therefore, additional polymorphic markers were needed and this objective formed the basis of the present study.

We originally set out to isolate and identify selectively 'neutral' microsatellite markers. Although we were unable to achieve this, a method for PCR-based DNA typing of individual *E. histolytica* isolates has been developed using multiple inter-tRNA gene spacers with internal blocks of AT-rich short tandem repeats (STRs) as the polymorphic markers (Chapter 3). Furthermore, it has been shown that *E. dispar* strains can also be typed by this approach and species-specific primers have been developed for this purpose at two of the loci (Chapter 5).

Results of surveys carried out using the polymorphic loci revealed that *E. histolytica* is genetically highly variable. This was evident in all the communities studied, including endemic populations as well as individuals living in developed countries (Chapters 4 & 6). *E. dispar* too displayed intra-species variation despite the fact that our analysis was limited by the use of only two loci (Chapter 6). The patterns seen for individual strains of both species were stable over a period of time in the same infection (longitudinal samples) suggesting that these markers are suitable for tracking transmission of a known strain within an individual, family unit or community. Despite this stability a remarkable degree of genetic diversity could be detected in a relatively restricted geographic area, an observation that will be particularly useful in allowing us to decide whether an *E. histolytica* strain infecting multiple individuals comes from the same source.

One of the main unsolved questions in amoebiasis research is the basis for the wide spectrum of clinical manifestations observed among individuals infected with *E. histolytica*. In the present study, no obvious correlation could be established between *E. histolytica* genotypes and different forms of clinical disease. The isolates studied in the global survey (Chapter 4) came from diverse geographic areas and with few exceptions (Appendix X) their hosts had been diagnosed with one or other form of invasive clinical disease. It was therefore difficult to ascertain whether the high degree of genetic diversity seen is a result of geographic variation or indicates a lack of association between polymorphism and the clinical form of disease. On the other hand, all the South African *E. histolytica* isolates (Chapter 6), were from individuals asymptomatic at the time of sample collection. Therefore there were no local 'invasive' isolates for comparison. To better address the possibility of a relationship

between parasite variation and infection outcome we need to undertake studies using these polymorphic DNA markers in a specific endemic area where individuals with both well characterised invasive disease and asymptomatic *E. histolytica* infections can be identified and recruited from the same population.

It is conceivable that the presence of both *E. histolytica* and *E. dispar* or different genotypes of *E. histolytica* in the same patient could be one of the reasons for the varying signs and symptoms in infected individuals. Our results showed both the existence of mixed species infection as well as the possibility of co-infection with different strains of the same species (Chapter 6). However, the frequency with which such infections occurred in a single host could not be deduced from the available samples and warrants further research.

Almost nothing is known about how competitive interactions within the host affects the subsequent transmission success of individual genotypes (Thompson *et al*, 1998). With respect to *E. histolytica* infections, for example, we do not know if the mixed populations observed in the same host consist of phenotypically distinct subgroups i.e. organisms that produce invasive disease and those that give rise to asymptomatic infections, nor do we know if a 'competitive dominance hierarchy' exists among such groups.

Studies on the molecular epidemiology of *Giardia* infections in Australian Aboriginal communities have also shown the presence of mixed infections with more than one genetic variant in the same host (Hopkins *et al*, 1999). Results of two samples collected from the same individual several months apart (longitudinal samples) revealed that these mixed genotype populations do not coexist over time and eventually one genotype predominates. This has also been observed by us (section 6.4).

Macedo and Pena (1998) have recently proposed that *T. cruzi* clones in the same host may compete fiercely for available resources. One would expect some clones to be eliminated by their inability to compete and propagate in the human environment or by the action of the host defense mechanisms. In addition, because of polymorphism,

different clones could present tissue tropism and hence escape the need to compete for the same resources.

The idea that mixed populations of *E. histolytica* genotypes competing for the same resources may promote tropism for specific organs is intriguing. An implication of this is that the intestinal parasite population would be distinct from that which infects extra-intestinal sites in the same individual. If true, genetic typing of strains carried out directly on faecal DNA would not show any significant correlation with clinical forms of disease.

Recently, Ayeh-Kumi *et al* (2001) investigated the genetic diversity of *E. histolytica* isolates from Bangladesh by analyzing polymorphism in the SREHP gene using DNA extracted from stool and liver aspirate pus. Their results show that the majority of *E. histolytica* from liver abscess pus display polymoprphisms which are not present in intestinal isolates from the same geographic area. This data is thus consistent with the existence of genetic differences between *E. histolytica* which cause intestinal infections and those which cause hepatic disease. In the present study we have shown that the source of the DNA, whether xenic or axenic culture, or extracted directly from stool was irrelevant. It will be interesting to see if the polymorphic loci described by us can be used to type strains directly from pus as comparison of pus and stool samples from the same individual would be of great interest.

It is clear from our results that the eleven loci studied differed in their diversity and it was only through the combined use of all of them that a majority of the isolates could be distinguished (Chapters 3, 4 & 6). It has been suggested that use of only a few loci results in ambiguity and that a large number of loci that represent a random sample of the genome, including both polymorphic and monomorphic loci, should be used to obtain a general picture of genetic differentiation among parasite subpopulations (Nei, 1973). Analysis of the data assembled from the *E. histolytica* genome project (Chapter 7) has revealed a large number of inter-tRNA gene spacers containing blocks of AT-rich STRs. In the loci studied by us, the STR sequences sizes and numbers are highly variable. In all there are an estimated forty more potentially polymorphic loci linked to tRNAs to be tested, with more anticipated before the genome project is completed.

Hence the number of loci available for isolate typing and the degree of diversity that can be detected will likely increase in the foreseeable future.

The genome project is providing information on one strain only, HM-1:IMSS. Sequence variation in repeat-flanking regions was seen even among the small number of isolates examined at two of the loci and it was observed that primer sequence modifications at locus 5-6 and 9-11 resulted in successful amplification of a greater number of samples (Chapters 3 & 4). There is, therefore, a need for caution when using primers designed from a single source and it is recommended for future studies that nucleotide sequence data be generated for a number of isolates from diverse geographical regions to avoid problems due to sequence diversity in the primer binding regions.

In addition to studying the repeat loci for their potential as polymorphic markers, their genomic organisation in *E. histolytica* was also studied (Chapter 7). Results of this analysis as well as examination of data emerging from the *E. histolytica* genome project revealed some unusual features, a few of which have been mentioned (Chapter 7). Perhaps the most interesting observation of all is the unique organisation of tRNA genes in this protozoan parasite. In most eukaryotes tRNA genes are usually not tightly clustered (Long and Dawid, 1980). However, the assembled genome project data has shown that the tRNA genes in *E. histolytica* are as a rule organised into long arrays. In all at least twenty-four distinct arrays, each with a characteristic but different tRNA gene content, can be identified in the data analysed so far. tRNA gene clustering on this scale is exceptional. Although attempts to study the chromosomal distribution of these arrays were not completely successful, pulsed-field gel electrophoresis of digested chromosomal DNA followed by Southern hybridisation has shown that these arrays constitute a significant part of any given chromosome.

In *E. histolytica* allelic chromosomes vary significantly in size between different isolates and also within isolates. However, as protein-encoding genes are clustered within certain conserved chromosomal regions, the observed size polymorphism is probably due to variation in the chromosome ends (Willhoeft and Tannich, 1999). In fact, chromosome organisation in several protozoan parasites seems to be guided by the same blueprints and chromosome compartmentalisation into conserved central

domains and polymorphic chromosome ends has been suggested (Lanzer *et al*, 1995). Extensive variations in length of allelic chromosomes have been reported for *Trypanosoma, Plasmodium, Leishmania* and *Giardia* and the size variations are primarily due to amplification or deletion of telomeric and subtelomeric regions. Sequences corresponding to telomeres have not yet been recognised in *Entamoeba*. It is possible that the tRNA arrays are located at the ends of *E. histolytica* chromosomes and are responsible for allelic size variation. If true, the number of arrays identified so far (twenty-four) and the estimated number of chromosome ends (twenty-eight) (Willhoeft and Tannich, 1999) suggests that a distinct array may be located at each end of most chromosomes. Thus these arrays may be fulfilling the role of classical telomeres in a manner similar to the telomeric retrotransposons of *Drosophila* (Pardue *et al*, 1996). The location of the arrays will only be fully resolved when the genome sequence is completed.

The presence of morphologically similar but genetically distinct species or a high level of population sub-structuring within a species can obscure the pattern of disease transmission as well as the identification of the causative agent of disease. The development and implementation of effective management policies, including strategies for the diagnosis, treatment, control and eradication of disease, are reliant upon sound epidemiological data. It is hoped that the multiple polymorphic loci described in this study will provide the tools necessary to answer many of the questions that still exist regarding amoebiasis epidemiology and thus allow more rational planning of interventions.

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. **Abd-Alla, M. D., T. F. H. G. Jackson, S. Reddy, and J. I. Ravdin.** 2000. Diagnosis of invasive amebiasis by enzyme-linked immunosorbent assay of saliva to detect amebic lectin antigen and anti-lectin immunoglobulin G antibodies. J. Clin. Microbiol. **38:**2344-2347.

3. **Acuña-Soto, R., J. Samuelson, P. de Girolami, L. Zarate, F. Millan-Velasco, G. Schoolnick, and D. Wirth.** 1993. Application of the polymerase chain reaction to the epidemiology of pathogenic and nonpathogenic *Entamoeba histolytica*. Am. J. Trop. Med. Hyg. **48:**58-70.

4. **Aguirre, A., D. C. Warhurst, F. Guhl, and I. A. Frame.** 1995. Polymerase chain reaction-solution hybridization enzyme-linked immunoassay (PCR-SHELA) for the differential diagnosis of pathogenic and non-pathogenic *Entamoeba histolytica*. Trans. R. Soc. Trop. Med. Hyg. **89:**187-188.

5. **Aguirre, A., S. Molina, C. Blotkamp, J. J. Verweij, T. Vinuesa, M. E. Valls, F. Guhl, A. M. Polderman, M. T. Jimenez De Anta, M. Corachan, A. Gonzalez-Ruiz, I. A. Frame, and D. C. Warhurst.** 1997. Diagnosis of *Entamoeba histolytica* and *Entamoeba dispar* in clinical specimens by PCR-SHELA. Arch. Med. Res. **28 (Suppl):**S282-S284.

6. **Albach, R. A., V. Prachayasittikul, and G. M. Heebner.** 1984. Isolation and characterization of RNA of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **12:**261-272.

7. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.

8. **Anderson, T. J. C., X. Z. Su, M. Bockarie, M. Lagog, and K. P. Day.** 1999. Twelve microsatellite markers for characterisation of *Plasmodium falciparum* from finger-prick blood samples. Parasitology. **119:**113-125.

9. **Anonymous.** 1985. Amoebiasis and its control. Bull. Wld. Hlth. Org. **63:**417-426.

10. **Anonymous.** 1997a. *Entamoeba* taxonomy. Bull. Wld. Hlth. Org. **75:**291-292.

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

12. **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.

13. **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.

14. **Bagchi, A., A. Bhattacharya, and S. Bhattacharya.** 1999. Lack of a chromosomal copy of the circular rDNA plasmid of *Entamoeba histolytica*. Int. J. Parasitol. **29:**1775-1783.

15. **Bhattacharya, A., S. Bhattacharya, and J. P. Ackers.** 1999. Nontranslated polyadenylated ribonucleic acids from the protozoan parasite *Entamoeba histolytica*. Curr. Sci. **77:**564-567.

16. **Bhattacharya, A., S. Satish, A. Bagchi, and S. Bhattacharya.** 2000. The genome of *Entamoeba histolytica*. Int. J. Parasitol. **30:**401-410.

17. **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.

18. **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.

19. **Bhattacharya, S., A. Bhattacharya, and L. S. Diamond.** 1992. *Entamoeba histolytica* extrachromosomal circular ribosomal DNA: analysis of clonal variation in a hypervariable region. Exp. Parasitol. **74:**200-204.

20. **Bhattacharya, S., I. Som, and A. Bhattacharya.** 1998. The ribosomal DNA plasmids of *Entamoeba*. Parasitol. Today. **14:**181-185.

21. **Blanc, D. S.** 1992. Determination of taxonomic status of pathogenic and nonpathogenic *Entamoeba histolytica* zymodemes using isoenzyme analysis. J. Protozool. **39:**471-479.

22. **Blanc, D. S., and P. G. Sargeaunt.** 1991. *Entamoeba histolytica* zymodemes: exhibition of γ and δ bands only of glucose phosphate isomerase and phosphoglucomutase may be influenced by starch content in the medium. Exp. Parasitol. **72:**87-90.

23. **Blessmann, J., P. V. Linh, T. N. P. Anh, D. T. Hao, H. Buss, and E. Tannich.** 2002. Epidemiology of amoebiasis in a region of high incidence of amoebic liver abscess in central Vietnam. Am. J. Trop. Med. Hyg. (in press).

24. **Bracha, R., and D. Mirelman.** 1984. Virulence of *Entamoeba histolytica* trophozoites. J. Exp. Med. **160:**353-368.

25. **Bracha, R., L. S. Diamond, J. P. Ackers, G. D. Burchard, and D. Mirelman.** 1990. Differentiation of clinical isolates of *Entamoeba histolytica* by using specific DNA probes. J. Clin. Microbiol. **28:**680-684.

26. **Braga, L. L., H. Ninomiya, J. J. McCoy, S. Eacker, T. Wiedmer, C. Pham, S. Wood, P. J. Sims, and W. A. Petri.** 1992. Inhibition of the complement membrane attack complex by the galactose-specific adhesin of *Entamoeba histolytica*. J. Clin. Invest. **90:**1131-1137.

27. **Braga, L. L., Y. Mendonca, C. A. Paiva, A. Sales, A. L. M. Cavalcante, and B. J. Mann.** 1998. Seropositivity for and intestinal colonization with *Entamoeba histolytica* and *Entamoeba dispar* in individuals in Northeastern Brazil. J. Clin. Microbiol. **36.:**3044-3045.

28. **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.

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

30. **Bruchhaus, I., T. Jacobs, M. Leippe, and E. Tannich.** 1996. *Entamoeba histolytica* and *Entamoeba dispar*: differences in numbers and expression of cysteine proteinase genes. Mol. Microbiol. **22:**255-263.

31. **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.

32. **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.

33. **Calafell, F., A. Shuster, W. C. Speed, J. R. Kidd, and K. K. Kidd.** 1998. Short tandem repeat polymorphism evolution in humans. Eur. J. Hum. Genet. **6:**38-49.

34. **Campbell, D., and K. Chadee.** 1997. Survival strategies of *Entamoeba histolytica*: modulation of cell-mediated immune responses. Parasitol. Today. **13:**184-190.

35. **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.

36. **Chadee, K., W. A. Jr. Petri, D. J. Innes, and J. I. Ravdin.** 1987. Rat and human colonic mucins bind to and inhibit adherence lectin of *Entamoeba histolytica*. J. Clin. Invest. **80:**1245-1254.

37. **Chang, J.-K., K.-i. Im, and C.-T. Soh.** 1995. Axenization of *Entamoeba histolytica*, a Korean strain YS-27. Korean J. Parasitol. **33:**387-390.

38. **Char, S., and M. J. G. Farthing.** 1992. Codon usage in *Entamoeba histolytica*. Int. J. Parasitol. **22:**381-383.

39. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:**156-159.

40. **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.

41. **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.

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

43. **Clark, C. G., and L. S. Diamond.** 1991b. Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct. Mol. Biochem. Parasitol. **49:**297-302.

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

45. **Clark, C. G., and L. S. Diamond.** 1993b. *Entamoeba histolytica*: an explanation for the reported conversion of "nonpathogenic" amebae to the "pathogenic" form. Exp. Parasitol. **77:**456-460.

46. **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.

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

48. **Clark, C. G., E. Y. Lai, C. Fulton, and G. A. M. Cross.** 1990. Electrophoretic karyotype and linkage groups of the amoeboflagellate *Naegleria gruberi*. J. Prozool. **37:** 400-408.

49. **Clark, C. G., C. C. Cunnick, and L. S. Diamond.** 1992. *Entamoeba histolytica*: is conversion of nonpathogenic amebae to the pathogenic form a real phenomenon? Exp. Parasitol. **74:**307-314.

50. **Clark, C. G., M. Espinosa Cantellano, and A. Bhattacharya.** 2000. *Entamoeba histolytica*: an overview of the biology of the organism., p. 1-45. *In* J. I. Ravdin (ed.), Amebiasis. Imperial College Press, London.

51. **Clarkson, S. G., V. Kurer, and H. O. Smith.** 1978. Sequence organisation of a cloned $tDNA_1^{met}$ fragment from *Xenopus laevis*. Cell. **14:**713-724.

52. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucl. Acids. Res. **16:**10881-10890.

53. **Costa, A. O., J. C. Viana, D. Assis, O. A. Rocha, and E. F. Silva.** 2000. Comparison of xenic and monoxenic *Entamoeba dispar* cultures using hepatic inoculation in hamster. Arch. Med. Res. **31:**S247-S248.

54. **Cruz-Reyes, J., and J. P. Ackers.** 1992. A DNA probe specific to pathogenic *Entamoeba histolytica*. Arch. Med. Res. **23:**271-275.

55. **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.

56. **Darnell, J. E.** 1990. Eukaryotic chromosomes and genes: molecular anatomy, p. 347-390. *In* J. Darnell, H. Lodish, and D. Baltimore (ed.), Molecular Cell Biology, Second ed. Scientific American Books., New York.

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

58. **Dhar, S. K., N. R. Choudhury, A. Bhattacharya, and S. Bhattacharya.** 1995. A multitude of circular DNAs exist in the nucleus of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **70:**203-206.

59. **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.

60. **Diamond, L. S., C. G. Clark, and C. C. Cunnick.** 1995. YI-S, a casein-free medium for axenic cultivation of *Entamoeba histolytica*, related *Entamoeba*, *Giardia intestinalis* and *Trichomonas vaginalis*. J. Euk. Microbiol. **42:**277-278.

61. **Dodson, J. M., C. G. Clark, L. A. Lockhart, B. M. Leo, J. W. Schroeder, and B. J. Mann.** 1997. Comparison of adherence, cytotoxicity, and Gal/GalNAc lectin gene structure in *Entamoeba histolytica* and *Entamoeba dispar*. Parasitol. Int. **46:**225-235.

62. **Dykes, A. C., T. K. Ruebush, L. Gorelkin, W. B. Lushbaugh, J. K. Upshur, and J. D. Cherry.** 1980. Extraintestinal amoebiasis in infancy: Report of three patients and epidemiological investigations of their families. Pediatrics. **65:**799-803.

63. **Edman, U., Meza, I. and Agabian, N.** 1987. Genomic and cDNA actin sequences from a virulent strain of *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA. **84:** 3024-3028.

64. **Espinosa-Cantellano, M., and A. Martínez-Palomo.** 2000. Pathogenesis of intestinal amebiasis: from molecules to disease. Clin. Microbiol. Rev. **13:**318-331.

65. **Espinosa-Cantellano, M., G. Castañon Gutierrez, and A. Martínez-Palomo.** 1997. *In vivo* pathogenesis of *Entamoeba dispar*. Arch. Med. Res. **28 (Suppl):**S204-S206.

66. **Espinosa-Cantellano, M., A. González-Robles, B. Chávez, G. Castañón, C. Argüello, A. Lázaro-Haller, and A. Martínez-Palomo.** 1998. *Entamoeba dispar*: ultrastructure, surface properties, and cytopathic effect. J. Euk. Microbiol. **45:**265-272.

67. **Estevez, M. L., J. A. Cruz, M. Zurita, J. Vargas-Villarreal, M. A. Meraz, R. Perez, A. Martínez-Palomo, A. Alagon, and P. M. Lizardi.** 1989. Molecular cloning and characterisation of ribosomal genes from *Entamoeba histolytica*. XI International seminar on amoebiasis, Mexico. Mexico City. IMSS:17.

68. **Evangelopoulos, A., G. Spanakos, E. Patsoula, N. Vakalis, and N. Legakis.** 2000. A nested, multiplex, PCR assay for the simultaneous detection and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* in faeces. Ann. Trop. Med. Parasitol. **94:**233-240.

69. **Fischer, D., and K. Bachmann.** 1998. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). Biotechniques. **24:**796-802.

70. **Garfinkel, L. I., M. Giladi, M. Huber, C. Gitler, D. Mirelman, M. Revel, and S. Rozenblatt.** 1989. DNA probes specific for *Entamoeba histolytica* possessing pathogenic and nonpathogenic zymodemes. Infect. Immun. **57:**926-931.

71. **Gathiram, V., and T. F. H. G. Jackson.** 1985. Frequency distribution of *Entamoeba histolytica* zymodemes in a rural South African population. The Lancet. **March 30:**719-721.

72. **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.

73. **Gatti, S., C. Cevini, A. Bruno, S. Novati, and M. Scaglia.** 1995. Transmission of *Entamoeba histolytica* within a family complex. Trans. R. Soc. Trop. Med. Hyg. **89:**403-405.

74. **Gatti, S., R. Lopes, C. Cevini, B. Ijaoba, A. Bruno, A. M. Bernuzzi, P. Lio, A. Monco, and M. Scaglia.** 2000. Intestinal parasitic infections in an institution for the mentally retarded. Ann. Trop. Med. Parasitol. **94:**453-460.

75. **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.

76. **Ghosh, S., M. Zaki, C. G. Clark, and S. Bhattacharya.** 2001. Recombinational loss of a ribosomal DNA unit from the circular episome of *Entamoeba histolytica* HM-1:IMSS. Mol. Biochem. Parasitol. **116:**105-108.

77. **Gilchrist, C. A., and W. A. Petri, Jr.** 1999. Virulence factors of *Entamoeba histolytica*. Curr. Opin. Microbiol. **2:**433-437.

78. **Gomes, M. A., E. F. Silva, A. M. Macedo, A. R. Vago, and M. N. Melo.** 1997. LSSP-PCR for characterization of strains of *Entamoeba histolytica* isolated in Brazil. Parasitology. **114:**517-520.

79. **González-Ruiz, A., R. Haque, T. Rehman, A. Aguirre, C. Jaramillo, G. Castañón, A. Hall, F. Guhl, G. Ruiz-Palacios, D. C. Warhurst, and M. A. Miles.** 1992. A monoclonal antibody for distinction of invasive and noninvasive clinical isolates of *Entamoeba histolytica*. J. Clin. Microbiol. **30:**2807-2813.

80. **González-Ruiz, A., R. Haque, T. Rehman, A. Aguirre, A. Hall, F. Guhl, D. C. Warhurst, and M. A. Miles.** 1994a. 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.

81. **González-Ruiz, A., R. Haque, A. Aguirre, G. Castañón, A. Hall, F. Guhl, G. Ruiz-Palacios, M. A. Miles, and D. C. Warhurst.** 1994b. Value of microscopy in the diagnosis of dysentery associated with invasive *Entamoeba histolytica*. J. Clin. Pathol. **47:**236-239.

82. **Grundy, M. S., A. Voller, and D. C. Warhurst.** 1987. An enzyme-linked immunosorbent assay for the detection of *Entamoeba histolytica* antigens in faecal material. Trans. R. Soc. Trop. Med. Hyg. **81:**627-632.

83. **Haque, R., L. M. Neville, S. Wood, and W. A. Petri, Jr.** 1994. Detection of *Entamoeba histolytica* and *E. dispar* directly in stool. Am. J. Trop. Med. Hyg. **50:**595-596.

84. **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.

85. **Haque, R., I. K. M. Ali, S. Akther, and W. A. Petri, Jr.** 1998. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J. Clin. Microbiol. **36:**449-452.

86. **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.

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

88. **Hastings, I. M.** 1996. Population genetics and the detection of immunogenic and drug-resistant loci in *Plasmodium*. Parasitology. **112:**155-164.

89. **Healy, G. R.** 1986. Immunologic tools in the diagnosis of amoebiasis: epidemiology in the United States. Rev. Infect. Dis. **8:**239-246.

90. **Hernández-Velarde, R., Q. F. B. G. Garduño-Rodríguez, and O. Muñoz.** 1980. Los acidos ribonucleicos de transferencia de *Entamoeba histolytica* I. analisis cromotografico de las especies isoaceptoras de tRNA de trofozoitos. Arch. Invest. Med. (Mex.). **11 (Suppl 1):**89-94.

91. **Hernández, R., G. Garduño, A. Valadez, and O. Muñoz.** 1986. Cambios en lisil-tRNA asociados al enquistamiento de *Entamoeba invadens*. Arch. Invest. Med. (Mex.). **17 (Suppl):**61-65.

92. **Hopkins, R. M., C. C. Constantine, D. A. Groth, J. D. Wetherall, J. A. Reynoldson, and R. C. A. Thompson.** 1999. PCR-based DNA fingerprinting of *Giardia duodenalis* isolates using the intergenic rDNA spacer. Parasitology. **118:**531-539.

93. **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 (Suppl):**S1-S4.

94. **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.

95. **Huston, C. D., and W. A. Petri, Jr.** 1998. Host-pathogen interaction in amebiasis and progress in vaccine development. Eur. J. Clin. Microbiol. Infect. Dis. **17:**601-614.

96. **Huston, C. D., R. Haque, and W. A. J. Petri.** 1999. Molecular-based diagnosis of *Entamoeba histolytica* infection. Exp. Rev. Mol. Med. **22 March:**http://www-ermm.cbcu.cam.ac.uk/99000599h.htm.

97. **Jackson, T. F. H. G.** 2000. Epidemiology, p. 47-63. *In* J. I. Ravdin (ed.), Amebiasis. Imperial College Press, London.

98. **Jackson, T. F. H. G., and S. Suparsad.** 1997. Zymodeme stability of *Entamoeba histolytica* and *E. dispar*. Arch. Med. Res. **28 (Suppl):**S304-S305.

99. **Jackson, T. F. H. G., V. Gathiram, and A. E. Simjee.** 1985. Seroepidemiological study of antibody responses to the zymodemes of *Entamoeba histolytica*. The Lancet. **March 30:**716-719.

100. **Jackson, T. F. H. G., V. Gathiram, S. Suparsad, and C. A. Anderson.** 1992. Stability of the zymodemes of *Entamoeba histolytica* in cultures. Arch. Med. Res. **23:**71.

101. **Jackson, T. F. H. G., S. Reddy, J. Fincham, M. D. Abd-Alla, S. Welles, and J. I. Ravdin.** 2000. A comparison of cross-sectional and longitudinal seroepidemiological assessments of *Entamoeba*-infected populations in South Africa. Arch. Med. Res. **31:**S36-37.

102. **Jelinek, T., G. Peyerl, T. Löscher, and H.-D. Nothdurft.** 1996. Evaluation of an antigen-capture enzyme immunoassay for detection of *Entamoeba histolytica* in stool samples. Eur. J. Clin. Microbiol. Infect. Dis. **15:**752-755.

103. **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.

104. **Keller, K., M. Olivier, and K. Chadee.** 1992. The fast release of mucin secretion from human colonic cells induced by *Entamoeba histolytica* is dependent on contact and protein kinase C activation. Arch. Med. Res. **23:**217-221.

105. **Kemp, D. J., R. L. Coppel, and R. F. Anders.** 1987. Repetitive proteins and genes of malaria. Ann. Rev. Microbiol. **41:**181-208.

106. **Kobayashi, S., E. Imai, H. Tachibana, T. Fujiwara, and T. Takeuchi.** 1998. *Entamoeba dispar*: cultivation with sterilized *Crithidia fasciculata*. J. Euk. Microbiol. **45:**3S-8S.

107. **Kobayashi, S., E. Imai, A. Haghighi, H. Tachibana, and T. Takeuchi.** 2000. Cultivation of *Entamoeba dispar*: growth promoting effect of ferredoxin. Arch. Med. Res. **31 (Suppl):**S210-S211.

108. **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.

109. **Lanzer, M., K. Fischer, and S. M. Le Blancq.** 1995. Parasitism and chromosome dynamics in protozoan parasites: is there a connection? Mol. Biochem. Parasitol **70:**1-8.

110. **Leippe, M.** 1997. Amoebapores. Parasitol. Today. **13:**178-183.

111. **Leippe, M., E. Bahr, E. Tannich, and R. D. Horstmann.** 1993. Comparison of pore-forming peptides from pathogenic and nonpathogenic *Entamoeba histolytica*. Mol. Biochem. Parasitol. **59:**101-110.

112. **Leitch, G. J.** 1988. Intestinal lumen and mucosal microclimate H^+ and NH_3 concentrations as factors in the etiology of experimental amebiasis. Am. J. Trop. Med. Hyg. **38:**87-229.

113. **Lioutas, C., C. Schmetz, and E. Tannich.** 1995. Identification of various circular DNA molecules in *Entamoeba histolytica*. Exp. Parasitol. **80:**349-352.

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

115. **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.

116. **Long, E. O., and I. B. Dawid.** 1980. Repeated genes in eukaryotes. Ann. Rev. Biochem. **49:**727-764.

117. **Lotter, H., E. Mannweiler, M. Schreiber, and E. Tannich.** 1992. Sensitive and specific serodiagnosis of invasive amebiasis by using a recombinant surface protein of pathogenic *Entamoeba histolytica*. J. Clin. Microbiol. **30:**3163-3167.

118. **Lowe, T. M., and S. R. Eddy.** 1997. tRNA scan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucl. Acids. Res. **25:**955-964.

119. **Macedo, A. M., and S. D. J. Pena.** 1998. Genetic variability of *Trypanosoma cruzi*: implications for the pathogenesis of Chagas disease. Parasitol. Today. **14:**119-124.

120. **Mai, Z., and J. Samuelson.** 1998. A new gene family (*ariel*) encodes asparagine-rich *Entamoeba histolytica* antigens, which resemble the amebic vaccine candidate serine-rich *E. histolytica* protein. Infect. Immun. **66:**353-355.

121. **Martínez-Báez, M.** 1986. Historical introduction, p. 587-601. *In* A. Martínez-Palomo (ed.), Human parasitic diseases. Amoebiasis, vol. 2. Elsevier Science Publishers B.V. (Biomedical Division)., Amsterdam.

122. **Martínez-Palomo, A.** 1986. Amoebiasis, p. 587-601. *In* H. M. Gilles (ed.), Clinics in Tropical Medicine and Communicable Diseases. Saunders, London.

123. **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.

124. **Martínez-Palomo, A., and M. Espinosa Cantellano.** 1998. Intestinal amoebae, p. 157-177. *In* F. E. G. Cox, J. P. Kreier, and D. Wakelin (ed.), Topley & Wilson's Microbiology and Microbial Infections, 9th ed, vol. 5. Arnold, New York.

125. **Martínez-Palomo, A., and M. Martínez-Báez.** 1983. Selective primary health care:strategies for control of disease in the developing world. X. Amoebiasis. Rev. Infect. Dis **5:**1093-1102.

126. **Martínez-Palomo, A., A. Gonzáles-Robles, B. Chávez, E. Orozco, S. Fernández-Castelo, and A. Cervantes.** 1985. Structural basis of the cytolytic mechanisms of *Entamoeba histolytica*. J. Protozool. **32:**166-175.

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

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

129. **Mirelman, D., Y. Nuchamowitz, and T. Stolarsky.** 1997. Comparison of use of enzyme-linked immunosorbent assay-based kits and PCR amplification of rRNA genes for simultaneous detection of *Entamoeba histolytica* and *E. dispar*. J. Clin. Microbiol. **35:**2405-2407.

130. **Mittal, V., A. Bhattacharya, and S. Bhattacharya.** 1992a. Organisation of repeated sequences in the region downstream to rRNA genes in the rDNA episome of *Entamoeba histolytica*. Arch. Med. Res. **23:**17-18.

131. **Mittal, V., D. Sehgal, A. Bhattacharya, and S. Bhattacharya.** 1992b. A second short repeat sequence detected downstream of rRNA genes in the *Entamoeba histolytica* rDNA episome. Mol. Biochem. Parasitol. **54:**97-100.

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

133. **Moon, C. D., B. A. Tapper, and B. Scott.** 1999. Identification of epichloe endophytes in planta by a microsatellite-based PCR fingerprinting assay with automated analysis. Appl. Environ. Microbiol. **65:**1268-1279.

134. **Nagakura, K., H. Tachibana, Y. Kaneda, H. Suzuki, K. Sasaoka, S. Kobayashi, and T. Takeuchi.** 1990. Amoebiasis in institutions for the mentally retarded in Kanagawa prefecture, Japan. Jpn. J. Med. Sci. Biol. **43:**123-131.

135. **Nei, M.** 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA. **70:**3321-3323.

136. **Novati, S., M. Sironi, S. Granata, A. Bruno, S. Gatti, M. Scaglia, and C. Bandi.** 1996. Direct sequencing of the PCR amplified SSU-rRNA gene of *Entamoeba dispar* and the design of primers for rapid differentiation from *Entamoeba histolytica*. Parasitology. **112:**363-369.

137. **Núñez, Y. O., M. A. Fernández, D. Torres-Núñez, J. A. Silva, I. Montano, J. L. Maestre, and L. Fonte.** 2001. Multiplex polymerase chain reaction amplification and dfferentiation of *Entamoeba histolytica* and *Entamoeba dispar* DNA from stool samples. Am. J. Trop. Med. Hyg. **64:**293-297.

138. **Ohnishi, K., and M. Murata.** 1997. Present characteristics of symptomatic amebiasis due to *Entamoeba histolytica* in the east-southeast area of Tokyo. Epidemiol. Infect. **119:**363-367.

139. **Oliveira, R. P., N. E. Broude, A. M. Macedo, C. R. Cantor, C. L. Smith, and S. D. J. Pena.** 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. USA. **95:**3776-3780.

140. **Omar, M. S., M. E. Al-Awad, and A. A. Al-Madani.** 1991. Giardiasis and amoebiasis in three Saudi closed communities. J. Trop. Med. Hyg. **94:**57-60.

141. **Ong, S.-J., M.-Y. Cheng, K.-H. Liu, and C.-B. Horng.** 1996. Use of ProSpecT® microplate enzyme immunoassay for the detection of pathogenic and non-pathogenic *Entamoeba histolytica* in faecal specimens. Trans. R. Soc. Trop. Med. Hyg. **90:**248-24.

142. **Pardue, M., O. N. Danilevskaya, K. Lowenhaupt, F. Slot, and K. L. Traverse.** 1996. *Drosophila* telomeres: new views on chromosome evolution. Trends. Genet. **12:**48-52.

143. **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 histolytica*. Korean J. Parasitol. **37:**59-62.

144. **Paul, R. E. L., M. J. Packer, M. Walmsley, M. Lagog, L. C. Ranford-Cartwright, R. Paru, and K. P. Day.** 1995. Mating patterns in malaria parasite populations of Papua New Guinea. Science. **269:**1709-1711.

145. **Peffley, D. M., and M. L. Sogin.** 1981. A putative $tRNA^{Trp}$ gene cloned from *Dictyostelium discoideum*: its nucleotide sequence and association with repetitive deoxyribonucleic acid. Biochemistry. **20:**4015-4021.

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

147. **Petri, W. A., Jr., M. D. Chapman, T. Snodgrass, B. J. Mann, J. Broman, and J. I. Ravdin.** 1989. Subunit structure of the galactose and N-acetyl-D-galactosamine-inhibitable adherence lectin of *Entamoeba histolytica*. J. Biol. Chem. **264:**3007-3012.

148. **Petri, W. A., Jr., T. F. H. G. Jackson, V. Gathiram, K. Kress, L. D. Saffer, T. L. Snodgrass, M. D. Chapman, Z. Keren, and D. Mirelman.** 1990. Pathogenic and nonpathogenic strains of *Entamoeba histolytica* can be differentiated by monoclonal antibodies to the galactose-specific adherence lectin. Infect. Immun. **58:**1802-1806.

149. **Petri, W. A., Jr., C. G. Clark, and L. S. Diamond.** 1994. Host-parasite relationships in amebiasis: conference report. J. Infect. Dis. **169:**483-484.

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

151. **Que, X. C., and S. L. Reed.** 1997. The role of extracellular cysteine proteinases in pathogenesis of *Entamoeba histolytica* invasion. Parasitol. Today. **13:**190-194.

152. **Ranford-Cartwright, L. C., J. Taylor, T. Umasunthar, L. H. Taylor, H. A. Babiker, B. Lell, J. R. Schmidt-Ott, L. G. Lehman, D. Walliker, and P. G. Kremsner.** 1997. Molecular analysis of recrudescent parasites in a *Plasmodium falciparum* drug efficacy trial in Gabon. Trans. R. Soc. Trop. Med. Hyg. **91:**719-724.

153. **Ravdin, J. I.** 1995. Amebiasis. Clin. Infect. Dis. **20:**1453-1466.

154. **Ravdin, J. I., T. F. H. G. Jackson, W. A. J. Petri, C. F. Murphy, B. L. P. Ungar, V. Gathiram, J. Skilogiannis, and A. E. Simjee.** 1990. Association of serum antibodies to adherence lectin with invasive amoebiasis and asymptomatic infection with pathogenic *Entamoeba histolytica.* J. Infect. Dis. **162:**768-772.

155. **Reed, S. L.** 1992. Amebiasis: an update. Clin. Infect. Dis. **14:**385-393.

156. **Rivera, W. L., H. Tachibana, M. R. A. Silva-Tahat, H. Uenura, and H. Kanbara.** 1996. Differentiation of *Entamoeba histolytica* and *E. dispar* DNA from cysts present in stool specimens by polymerase chain reaction: its field application in the Philippines. Parasitol. Res. **82:**585-589.

157. **Rivera, W. L., H. Tachibana, and H. Kanbera.** 1998. Field study on the distribution of *Entamoeba histolytica* and *Entamoeba dispar* in the northern Philippines as detected by the polymerase chain reaction. Am. J. Trop. Med. Hyg. **59:**916-921.

158. **Russell, R., M. P. Iribar, B. Lambson, S. Brewster, J. M. Blackwell, C. Dye, and J. W. Ajioka.** 1999. Intra and inter-specific microsatellite variation in the *Leishmania* subgenus *Viannia.* Mol. Biochem. Parasitol. **103:**71-77.

159. **Saffer, L. D., and W. A. Petri, Jr.** 1991. Role of the galactose lectin of *Entamoeba histolytica* in adherence-dependent killing of mammalian cells. Infec. Immun. **59:**4681-4683.

160. **Salata, R. A., and J. I. Ravdin.** 1986. Review of the human immune mechanisms directed against *Entamoeba histolytica.* Rev. Infect. Dis. **8:**261-272.

161. **Samuelson, J., R. Acuña-Soto, S. Reed, F. Biagi, and D. Wirth.** 1989. DNA hybridisation probe for clinical diagnosis of *Entamoeba histolytica*. J. Clin. Microbiol. **27:**671-676.

162. **Samuelson, J., D. Caplivski, K. Sturm-Ramirez, K. Kretzinger, S. Descoteaux, G. Hernandez, G. Vargas, R. Mammo, O. Newton, J. Santos, H. de la Vega, P. Robbins, C. Ganguly, and A. Lohia.** 1997. A proposal for a molecular biological system for classifying isolates of *Entamoeba histolytica* and *Entamoeba dispar*. Arch. Med. Res. **28 (Suppl):**S274-S2765.

163. **Sánchez-López, R., S. Gama-Castro, M. A. Ramos, E. Merino, P. M. Lizardi, and A. Alagón.** 1998. Cloning and expression of the *Entamoeba histolytica* ERD2 gene. Mol. Biochem. Parasitol. **92:**355-359.

164. **Sanuki, J., T. Asai, E. Okuzawa, S. Kobayashi, and T. Takeuchi.** 1997. Identification of *Entamoeba histolytica* and *Entamoeba dispar* cysts in stool by polymerase chain reaction. Parasitol. Res. **83:**96-98.

165. **Sargeaunt, P. G.** 1987a. The reliability of *Entamoeba histolytica* zymodemes in clinical diagnosis. Parasitol. Today. **3:**40-43.

166. **Sargeaunt, P. G.** 1987b. Zymodemes of Entamoeba histolytica. Parasitol. Today. **3:**158.

167. **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.

168. **Sargeaunt, P. G., T. F. H. G. Jackson, and A. E. Simjee.** 1982. Biochemical homogeniety of *Entamoeba histolytica* isolates, especially those from liver abscess. The Lancet. **June 19:**1386-1388.

169. **Sehgal, D., A. Bhattacharya, and S. Bhattacharya.** 1993. Analysis of a polymorphic locus present upstream of rDNA transcription units in the extrachromosomal circle of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **62:**129-130.

170. **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.

171. **Sekar, V.** 1987. A rapid screening procedure for the identification of recombinant bacterial clones. Biotechniques. **5:**11-13.

172. **Shire, A., and J. P. Ackers.** 2000. DNA sequences homologous to interspersed element of *Entamoeba histolytica* are present in *Entamoeba dispar*. Arch. Med. Res. **31 (Suppl):**S260-S262.

173. **Silberman, J. D., C. G. Clark, L. S. Diamond, and M. L. Sogin.** 1999. Phylogeny of the genera *Entamoeba* and *Endolimax* as deduced from small subunit ribosomal RNA gene sequence analysis. Mol. Biol. Evol. **16:**1740-175.

174. **Silva, E. F., M. A. Gomes, and A. M. Barral De Martinez.** 1997. *Entamoeba histolytica*: axenization and characterisation of isolated samples from symptomatic and asymptomatic patients from different regions of Brazil. Arch. Med. Res. **28 (Suppl).:**S288-S289.

175. **Som, I., A. Azam, A. Bhattacharya, and S. Bhattacharya.** 2000. Inter- and intra-strain variation in the 5.8S ribosomal RNA and internal transcribed spacer sequences of *Entamoeba histolytica* and comparison with *Entamoeba dispar*, *Entamoeba moshkovskii* and *Entamoeba invadens*. Int. J. Parasitol. **30:**723-728.

176. **Spencer, C. H., C. Muchnik, D. J. Sexton, P. Dodson, and K. W. Walls.** 1977. Endemic amoebiasis in an extended family. Am. J. Trop. Med. Hyg. **26:**628-635.

177. **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.

178. **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*. The Lancet **March 12:**561-563.

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

180. **Tachibana, H., S. Kobayashi, E. Okuzawa, and G. Masuda.** 1992. Detection of pathogenic *Entamoeba histolytica* DNA in liver abscess fluid by polymerase chain reaction. Int. J. Parasitol. **22:**1193-1196.

181. **Takeuchi, T., Y. Miyahira, S. Kobayashi, T. Nozaki, S. R. N. Motta, and J. Matsuda.** 1990. High seropositivity for *Entamoeba histolytica* infection in Japanese homosexual men: further evidence for the occurrence of pathogenic strains. Trans. R. Soc. Trop. Med. Hyg. **84:**250-251.

182. **Tannich, E.** 1998. *Entamoeba histolytica* and *E. dispar*: Comparison of molecules considered important for host tissue destruction. Trans. R. Soc. Trop. Med. Hyg. **92:**593-596.

183. **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.

184. **Tannich, E., and R. D. Horstmann.** 1992. Codon usage in pathogenic *Entamoeba histolytica*. J. Mol. Evol. **34:**272-273.

185. **Thompson, R. C. A., C. C. Constantine, and U. M. Morgan.** 1998. Overview and significance of molecular methods: what role for molecular epidemiology. Parasitology. **117:**S161-S175.

186. **Troll, H., H. Marti, and N. Weiss.** 1997. Simple differential detection of *Entamoeba histolytica* and *Entamoeba dispar* in fresh stool specimens by sodium acetate-acetic acid-formalin concentration and PCR. J. Clin. Microbiol. **35:**1701-1705.

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

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

189. **Verweij, J. J., J. Blotkamp, E. A. T. Brienen, A. Aguirre, and A. M. Polderman.** 2000. Differentiation of *Entamoeba histolytica* and *Entamoeba dispar* cysts using polymerase chain reaction on DNA isolated from faeces with spin columns. Eur. J. Clin. Microbiol. Infect. Dis. **19:**358-361.

190. **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.

191. **Vreden, S. G. S., L. G. Visser, J. J. Verweij, J. Blotkamp, P. C. Stuiver, A. Aguirre, and A. M. Polderman.** 2000. Outbreak of amoebiasis in a family in the Netherlands. Clin. Infect. Dis. **31:**1101-1104.

192. **Walderich, B., A. Weber, and J. Knobloch.** 1997. Differentiation of *Entamoeba histolytica* and *Entamoeba dispar* from German travelers and residents of endemic regions. Am. J. Trop. Med. Hyg. **57:**70-74.

193. **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.

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

195. **Willhoeft, U., and E. Tannich.** 2000. Fluorescence microscopy and fluorescence in situ hybridization of *Entamoeba histolytica* nuclei to analyse mitosis and the localization of repetitive DNA. Mol. Biochem. Parasitol. **105:**291-296.

196. **Willhoeft, U., L. Hamann, and E. Tannich.** 1999a. A DNA sequence corresponding to the gene encoding cysteine proteinase 5 in *Entamoeba histolytica* is present and positionally conserved but highly degenerated in *Entamoeba dispar*. Infect. Immun. **67:**5925-5929.

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

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

199. **Zaman, S., J. Khoo, S. Ng, R. Ahmed, M. Khan, R. Hussain, and V. Zaman.** 2000. Direct amplification of *Entamoeba histolytica* DNA from amoebic liver abscess pus using polymerase chain reaction. Parasitol. Res. **86:**724-728.

200. **Zengzhu, G., R. Bracha, Y. Nuchamowitz, W. Cheng-I, and D. Mirelman.** 1999. Analysis by enzyme-linked immunosorbent assay and PCR of human liver abscess aspirates from patients in China for *Entamoeba histolytica*. J. Clin. Microbiol. **37**:3034-3036.

APPENDIX I

YI-S MEDIUM

YI NUTRIENT BROTH

YI nutrient broth was prepared by dissolving the following ingredients in 600 ml of deionized distilled water (MilliQ II, Milipore): K_2HPO_4, 1 g; KH_2PO_4, 0.6 g; NaCl, 2 g; yeast extract, 30 g (Sigma Aldrich Y-1625); glucose, 10 g; L-cysteine hydrochloride, 1 g; ascorbic acid, 0.2 g; and ferric ammonium citrate (brown form), 2 ml of a 22.8 mg/ml stock solution. The pH was adjusted to 6.8 with 1N NaOH (approx. 7.5 ml are required to achieve this). The final volume was brought to 880 ml with deionized distilled water and the medium was aliquoted in units of 88 ml into 100 ml PYREX® bottles and autoclaved for 15 minutes at 121°C with 15 lbs. pressure. The sterile YI nutrient broth was then stored at -20°C for up to six months.

VITAMIN MIX #18

SOLUTION 1

Four individual solutions were prepared and combined together. Niacinamide, 45 mg; pyridoxal hydrochloride, 4 mg; calcium pantothenate, 23 mg; thiamine hydrochloride, 5 mg and vitamin B12, 1.2 mg were dissolved in a final volume of 25 ml distilled water. Riboflavin, 7 mg were dissolved in a minimum amount of 0.1 N NaOH and the final volume was brought to 45 ml with distilled water. Similarly 5.5 mg of folic acid were dissolved in a minimum amount of 0.1 N NaOH and the final volume brought to 45 ml with distilled water. Finally 2 mg of D-biotin were dissolved in water to a final volume of 45 ml.

SOLUTION 2

DL-6-8 thioctic acid (oxidized), 1 mg was dissolved in 5 ml of 95% ethanol. 500 mg of Tween-80 were added to this and the final volume brought to 30 ml with distilled water.

Solutions 1 and 2 were combined and the total volume raised to 200 ml with distilled water resulting in a final concentration of 50x. The final solution was filter sterilised through a 0.22 μm cellulose nitrate membrane filter (Nalgene), aliquoted into 15 ml sterile glass tubes and stored at 4°C for up to 6 months.

BOVINE SERUM

Adult bovine serum (Sigma Aldrich) was heat inactivated for 30 minutes at 56°C before use and was then stored at 4°C for up to 6 months.

COMPLETE MEDIUM

For one unit of complete working medium (approx. 100 ml), 2 ml of vitamin mix and 10 ml or 15 ml of bovine serum were added to 88 ml of YI nutrient broth just before use. Unused complete medium was stored at 4°C for up to 2 weeks.

LYI-S-2 MEDIUM

This medium is recommended when a given lot of yeast extract will support some growth, but poorly, of *E. histolytica*. The medium is identical to YI-S except that weight for weight it contains 0.5% of neutralised liver digest (Oxoid) and only 2.5% yeast extract.

APPENDIX II

CHARACTERISTIC FEATURES OF POLYMORPHIC LOCI FROM *E. histolytica*

A

Repeat Locus	Isolate	Restriction Enzyme	Total length of fragment (bp)
1-2	HM-1:IMSS (clone 9)	*Alu*I	402
5-6	HM-1:IMSS (clone 9)	*Alu*I	424
3-4	HK9	*Eco*RI	978
9-4	HM-1:IMSS	*Eco*RI	931
16-17	HM-1:IMSS	*Hin*dIII	964

Note: The restriction enzymes listed are the delimiting enzymes used to isolate the respective loci.

B

Repeat Locus	Repeat Block number	Repeat Unit type	Repeat Unit Sequence ✞	Unit Length (bp)
1-2 ✡	1	1	$(CTTTATTAT)_7$	9
		2	$(CTTTATTAC)_5$	9
5-6 ✡	1	1	$(GTATGTTTCTAT)_2$	12
		2	$(GTATGTATATTTCTAT)_6$	16
3-4	1	1	$(CTTATTATA)_{11}$	9
	2	1	$(CTTTATTATTAT)_8$	12
		2	$(CTTTATTAT)_3$	9
9-4	1	1	$(CTTATTATA)_1$	9
		2	$(CTATTATA)_{10}$	8
	2	1	$(CTTTATTATTAT)_7$	12
		2	$(CTTTATTAT)_3$	9
16-17	1	1	$(TACTTATAT)_6$	9
	2	1	$(CTATTTTTA)_6$	8
	3	1	$(TTTTTATATC)_7$	10
	4	1	$(TTTATACC)_5$	8
		2	$(TTTATATG)_9$	8
		3	$(TTTATATC)_8$	8

Note: "✡" The information provided is based on the sequences isolated and described in Chapter 3. "✞" The numbers given in subscript represent the frequency of repeat units within the repeat block. Both number and sequence of the repeat units described here are those seen in the respective loci from the isolates given in Table A.

C

Repeat Locus	Primer Name	Expected PCR fragment Size (bp) ✞
1-2	R1 R2	402
5-6	R5/R5A R6/R6A	424
3-4	R3/R4 R3/R8 R7/R4	929 625 325
9-4	R9/R4 R9/R11 R10/R4	922 632 311
16-17	R16/R17 R16/R19 R18/R17	900 535 449

Note: "✞" The expected PCR fragment sizes given here are those predicted for the isolates given in Table A. The expected PCR fragment size given at locus 5-6 is based on amplification with primer pair R5A and R6A (Table 4).

D

Repeat Locus	Repeat Block number	Repeat Unit type	Repeat Unit Sequence ✞	Unit Length (bp)
1-2	1	1 2	$(CTTTATTAT)_7$ $(CTTTATTAC)_5$	9 9
20-21	1	1 2	$(CTTTTATA)_4$ $(CCTTTATA)_5$	8 8
	2	1	$(CTATTTT)_5$	7
	3	1	$(TATTATTTC)_5$	9
5-6	1	1 2	$(GTATGTTTCTAT)_2$ $(GTATGTATATTTCTAT)_6$	12 16
	2	1	$(CTTACTAT)_{11}$	8
	3	3	$(CTTTTATAT)_9$	9

Note: The information provided is based on the complete sequences identified from the TIGR *E. histolytica* genome sequencing database (http://www.tigr.org/tdb/e2k1/eha1/) and described in Chapter 7; Locus 1-2 (zheng_36), Locus 20-21 (zheng_288) & Locus 5-6 (zheng_586). "✞" The numbers given in subscript represent the frequency of repeat units within the repeat block.

APPENDIX III

NUCLEOTIDE SEQUENCE ALIGNMENT OF FIVE AXENIC *E. histolytica* STRAINS AT LOCUS 1-2

```
                    1                                                  50
HM-1:IMSSclone9  CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG
      200:NIH    CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG
      H-303:NIH  CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG
      IULA:0593:2 CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG
      IULA:1092:1 CTGGTTAGTA TCTTCGCCTG TCACGTGAAA GACCGGGGTT CGAATCCCCG

                   51                                                 100
HM-1:IMSSclone9  TTGAAGAGTT CTCTTTTTAT ACTTTTATAT GTTTATATGT TTATATCCTT
      200:NIH    TTGAAGAGTT CTCTTTTTAT ACTTTTATAT GTTTATATGT TTATATCCTT
      H-303:NIH  TTGAAGAGTT CTCTTTTTAT ACTTTTATAT GTTTATATGT TTATATCCTT
      IULA:0593:2 TTGAAGAGTT CTCTTTTCAT ACTTTTATAT GTTTATATGT TTATATCCTT
      IULA:1092:1 TTGAAGAGTT CTCTTTTTAT ACTTTTATAT GTTTATATGT TTATATCCTT

                  101                                                 150
HM-1:IMSSclone9  ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA
      200:NIH    ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA
      H-303:NIH  ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA
      IULA:0593:2 ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA
      IULA:1092:1 ATTTATTATT CTTTTATATT CTTATCACTT CCTACTACTC TTATTTATTA

                  151                                                 200
HM-1:IMSSclone9  TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTATCT TTATTATCTT
      200:NIH    TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTATCT TTATTATCTT
      H-303:NIH  TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTATCT TTATTATCTT
      IULA:0593:2 TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTATCT TTATTATCTT
      IULA:1092:1 TCCTTATTAT ATCTATTCTT ACTCCCTATC TTTATTACCT TTATTACCTT

                  201                                                 250
HM-1:IMSSclone9  TATTATCTTT ATTATCTTTA TTATCTTTAT TACCTTTATT ACCTTTATTA
      200:NIH    TATTACCTTT ATTACCTTTA TTATCTTTAT TACCTTTATT ACCTTTATTA
      H-303:NIH  TATTACCTTT ATTACCTTTA TTATCTTTAT TACCTTTATT ACCTTTATTA
      IULA:0593:2 TATTATCTTT ATTATCTTTA TTATCTTTAT TACCTTTATT ACCTTTATTA
      IULA:1092:1 TATTACCTTT ATTACCTTTA TTATCTTTAT TACTTTTATT ACTTTTATTA

                  251                                                 300
HM-1:IMSSclone9  CCTTTATTAC CTTTATTATC TTTATTACCT TTATTAT--- ------ATCT
      200:NIH    CCTTTATTAC TTTTATTACT TTTATTATCT TTATTACCTT TATTATATCT
      H-303:NIH  CCTTTATTAC TTTTATTACT TTTATTATCT TTATTACCTT TATTATATCT
      IULA:0593:2 CCTTTATTAT CTTTATTACC TTTATTAT-- ---------- ------ATCT
      IULA:1092:1 TCTTTATTAT CTTTATTACC TTTATTACCT TTATTAT--- ------ATCT

                  301                                                 350
HM-1:IMSSclone9  ATTCTCACTT CCTATACGTA CTCTTTTTAC TACTCTTTTT ACTACTCTTC
      200:NIH    ATTCTCACTT CCTATACGTA CTCTTTTTAC TACTCTTTTT ACTACTCTTC
      H-303:NIH  ATTCTCACTT CCTATACGTA CTCTTTTTAC TACTCTTTTT ACTACTCTTC
      IULA:0593:2 ATTCTCACTT CCTATACGTA CTCTTTTTAC TACTCTTTTT ACTACTCTTC
      IULA:1092:1 ATTCTCACTT CCTATACGTA CTCTTTTTAC TACT------ ------CTTC
```

```
                    351                                                 400
HM-1:IMSSclone9  TTACTATACC TCTTACTACT CCTACTTTCA CCTCCCTCTT TATTGTTAAT
       200:NIH   TTACTATACC TCTTACTACT CCTACTTTCA CCTCCCTCTT TATTGTTAAT
     H-303:NIH   TTACTATACC TCTTACTACT CCTACTTTCA CCTCCCTCTT TATTGTTAAT
   IULA:0593:2   TTACTATACC TCTTACTACT CCTACTTTCA CCTCCCTCTT TATTGTTAAT
   IULA:1092:1   TTACTATACC TCTTACTACT CCTACTTTCA CCTCCCTCTT TATTGTTAAT

                    401
HM-1:IMSSclone9  GGGGGTGTAA G
       200:NIH   GGGGGTGTAA G
     H-303:NIH   GGGGGTGTAA G
   IULA:0593:2   GGGGGTGTAA G
   IULA:1092:1   GGGGGTGTAA G
```

Note: Two types of internal tandem repeat units are underlined. Single base substitutions are highlighted in bold-face letters. Dashes indicate gaps introduced to optimise alignment. The alignment is inclusive of both 5′ and 3′ primer sequences (EhR1 and EhR2) which were used for PCR amplification and detection of PCR fragment size variation.

APPENDIX IV

NUCLEOTIDE SEQUENCE ALIGNMENT OF FIVE AXENIC *E. histolytica* STRAINS AT LOCUS 5-6

```
                      1                                                  50
HM-1:IMSSclone9   CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT
       200:NIH    CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT
  H303:NIH-(3)    CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT
  H-303:NIH-(4)   CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT
    IULA:0593:2   .......... .......... .......... ......TCTT TGAGACTTAT
    IULA:1092:1   CTAAAGCCCC CTTCTTCTAT AATTTATATA TTATTCTCTT TGAGACTTAT

                     51                                                 100
HM-1:IMSSclone9   TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC
       200:NIH    TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC
  H303:NIH-(3)    TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC
  H-303:NIH-(4)   TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC
    IULA:0593:2   TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC
    IULA:1092:1   TTCTACTTTA TTTCTTATAC TTATTATATC TACTTTAGAT ACTTTACTTC

                    101                                                 150
HM-1:IMSSclone9   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA
       200:NIH    ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA
  H-303:NIH-(3)   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA
  H-303:NIH-(4)   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA
    IULA:0593:2   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA
    IULA:1092:1   ACTTTATATA TTCTACTTTA TATTCCTGAC TTATATCTTT TATGTTTATA

                    151                                                 200
HM-1:IMSSclone9   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA
       200:NIH    TGTTTATATG ATTTTAT--- -----GTTTA TATTCTTCCT ATTTATACCA
  H-303:NIH-(3)   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA
  H-303:NIH-(4)   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA
    IULA:0593:2   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA
    IULA:1092:1   TGTTTATATG ATTTTATGAT TTTATGTTTA TATTCTTCCT ATTTATACCA

                    201                                                 250
HM-1:IMSSclone9   CTATGTATAT ATATATGTAT GTTTCTATGT ATGTTTCTAT GTATGTATAT
       200:NIH    CTATGTATAT ATATAT---- ------ATGT ATGTTTCTAT GTATGTATAT
  H-303:NIH-(3)   CTATGTATAT ATATATGTAT GTTTCTATGT ATGTTTCTAT GTATGTATAT
  H-303:NIH-(4)   CTATGTATAT ATATAT---- ------ATGT ATGTTTCTAT GTATGTATAT
    IULA:0593:2   CTATGTATAT ATATATGTAT GTTTCTATGT ATGTTTCTAT GTATGTATAT
    IULA:1092:1   CTATGTATAT ATATATGTAT GTTTCTATGT ATGTCTCTAT GTATGTATAT

                    251                                                 300
HM-1:IMSSclone9   TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT ATGTATATTT
       200:NIH    TTCTAT---- ---------- ---------- ---------- ----------
  H-303:NIH-(3)   TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT ATGTATATTT
  H-303:NIH-(4)   TTCTAT---- ---------- ---------- ---------- ----------
    IULA:0593:2   TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT ATGTATATTT
    IULA:1092:1   TTCTATGTAT GTATATTTCT ATGTATGTAT ATTTCTATGT ATGTATATTT
```

```
                    301                                                 350
HM-1:IMSSclone9     CTATGTATGT ATATTTCTAT GTATGTATAT TTCTAT---- ----------
       200:NIH      ---------- ---------- ---------- ---------- ----------
 H-303:NIH-(3)      CTATGTATGT ATATTTCTAT GTATGTATAT TTCTAT---- ----------
 H-303:NIH-(4)      ---------- ---------- ---------- ---------- ----------
   IULA:0593:2      CTATGTATGT ATATTTCTAT ---------- ---------- ----------
   IULA:1092:1      CTATGTATGT ATATTTCTAT GTATGTATAT TTCTATGTAT GTATATTTCT

                    351                                                 400
HM-1:IMSSclone9     ---------- ---------- ---------- ---------- ----------
       200:NIH      ---------- ---------- ---------- ---------- ----------
 H-303:NIH-(3)      ---------- ---------- ---------- ---------- ----------
 H-303:NIH-(4)      ---------- ---------- ---------- ---------- ----------
   IULA:0593:2      ---------- ---------- ---------- ---------- ----------
   IULA:1092:1      ATGTATGTAT ATTTCTATGT ATGTATATTT CTATGTATGT ATATTTCTAT

                    401                                                 450
HM-1:IMSSclone9     ---------- ------GTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC
       200:NIH      ---------- ------GTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC
 H-303:NIH-(3)      ---------- ------GTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC
 H-303:NIH-(4)      ---------- ------GTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC
   IULA:0593:2      ---------- ------GTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC
   IULA:1092:1      GTATGTATAT TTCTATGTAC GTCTTTAACT TTAAAAACAA TGCTCTCAGC

                    451
HM-1:IMSSclone9     AGGTTTCGAA CCCTGCGACC CTGGCGTTAT TAGCAC
       200:NIH      AGGTTTCGAA CCCTGCGACC CTGGCGTTAT TAGCAC
 H-303:NIH-(3)      AGGTTTCGAA CCCTGCGACC CTGGCGTTAT TAGCAC
 H-303:NIH-(4)      AGGTTTCGAA CCCTGCGACC CTGGCGTTAT TAGCAC
   IULA:0593:2      AGGTTTCGAA CCCTGCGACC CTGGCGTTGT TAGCAC
   IULA:1092:1      AGGTTTCGAA CCCTGCGACC CTGGCGTTAT TAGCAC
```

Note: Two types of internal tandem repeats are underlined. Single base substitutions are highlighted in bold-face letters. Dots represent sequences that could not be read. Dashes indicate gaps introduced to optimise alignment. The alignment is inclusive of both 5′ and 3′ primer sequences (EhR5/ EhR5A and EhR6) but (EhR6A) is not shown (see legend to Fig. 5).

APPENDIX V

COMPLETE SEQUENCE OF LOCUS 3-4

```
1     GAATTCTATT CCTACATCTA CCTAGGTCCT CCCGATCTCT ACCAACTGAG      50
                        EhR3 →
51    CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTGC TTATTCTTTT     100

101   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTT     150

151   CTATATGTGC TTATATATCT TTCTATTCTT TATACCTATT CTTTATATAC     200

201   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT     250

251   GTATATATTT ATATTTATAT TTATATTATG TATCCCTTAT CTTTATATTA     300

301   TGTATATATT TTATATTATG TATATATTTT ATATTATGTG TAGTATATAT     350

351   TTATTTTATG TATATGTGTA GTGTACTATA TGTGTAGTGT ACTGATGTGT     400

401   ACTATATATA TATATTCTTG CTAAAAAATT CCATCTACGC CGGGAATCGA     450

451   ACCTAGTCAG GGGCACAAAC GTGAAGGGCT TGTATGTTAC CAACTACACC     500

501   ACAATGGATT TACCATGACT ATATTAGCCC CCTTAGACTT TATTCTTTAT     550

551   ATTCCTTATT ATACTTATTA TACTTATTAT ACTTATTATA CTTATTATAC     600

601   TTATTATACT TATTATACTT ATTATACTTA TTATACTTAT TATACTTATT     650
                              EhR7 →
651   ATACTTTACT TCTCTTTTAC CACGATTATG TATATTTATA TATGTTTCTA     700
          ← EhR8
701   TTGCTTTATA TTATTATTAT CTTTATTATT ATCTTTATTA TTATCTTTAT     750

751   TATTATCTTT ATTATTATCT TTATTATTAT CTTTATTATT ATCTTTATTA     800

801   TTATCTTTAT TATTATCTTT ATTATCTTTA TTATCTTTAT TATTTATGCT     850

851   TATGTACTTG ACTATATACT TATTTCTACT TATTTATACT CTTTTTCCTA     900

901   TACCTATACC TATCCCTATA CCTATCTATT AGTATAAAAA CTGATCCGAC     950
                  ← EhR4
951   CAACCGGATT CGAACCAGTG ACCTAAGGAA TTC                      1000
```

Note: Two blocks of internal tandem repeats are shown. *The first block (bold face)* has a 9 bp sequence (CTTATTATA) which is repeated eleven times in tandem. The second block (underlined bold-face) contains eight tandem 12 bp units (CTTTATTATTAT) followed by three truncated repeats (thick underlined bold-face) which are missing a second ATT stretch (CTTTATTAT). Positions of amplification primers are highlighted.

APPENDIX VI

COMPLETE SEQUENCE OF LOCUS 9-4

```
1     ACTATATATA TATATTCTTG CTAAAAAATT CCATCGCCGG GAATCGAACC     50

51    CGGGCACAAA CGTGAAAGGC TTGTATGTTA CCAACTACAC CACAATGGAT    100

101   TTACCATGAC TATATGCCCC TTAGACTTTA TTCTTTATAT TCTTATTATA    150

151   CTATTATACT ATTATACTAT TATACTATTA TACTATTATA CTATTATACT    200

201   ATTATACTAT TATACTATTA TACTATTATA CTTACTTCTC TTTACCACGA    250
      EhR10 →                                  ← EhR11
251   CTTATGTATA TTTATATATG TTTCTATTCC TTTATATTAT TATTATCTTT    300

301   ATTATTATCT TTATTATTAT CTTTATTATT ATCTTTATTA TTATCTTTAT    350

351   TATTATCTTT ATTATTATCT TTATTATTAT CTTTATTATC TTTATTATCT    400

401   TTATTATTTA TGCTTATGTA CTTGACTATA TACTTATTTC TACTTATTTA    450

451   TACTCTTTTC CTATACCTAT ACCTATCCCT ATACCTATCT ATTAGTATAA    500
                                    ← EhR4
501   AAACTGATCC GACCAACCGG ATTCGAACCA GTGACCTAAG GAATTCTATT    550
                          EhR9 →
551   CTACATCTAC AGTCCTCCGC TCTACCAACT GAGCTATGGT CGGTATCGAT    600

601   ATCACCTTTT TATATTCTTT TGTTATTCTT TTGATATCCT TTTCTTATTC    650

651   TTTTCTTATT CTTTTTATAT CCTTTTATAT TTCTATATGT GCTTATATAT    700

701   GTTTCTATTC TTTATACCTA TTCTTTATAT ACTTCTTATT ATCTCTTCCC    750

751   TGTTATTTCT CTATACCTTA TCTTTATATT ATGTATATAT TTATATTTAT    800

801   ATTTATATTA TGTATCCTTA TCTTATCTTT ATATTATGTA TATATTTATA    850

851   TTATGTATAT ATTTATATTA TGTGTAGTAT ATATTATTAT GTATATGTGT    900

901   AGTGTACTAT ATGTGTAGTG TACTTATGTG T                        950
```

Note: Two blocks of internal tandem repeats shown. The first block (bold-face) has an 8 bp sequence (CTATTATA) which is repeated ten times in tandem and is preceded at the 5′ end by a single 9 bp unit which differs in that it has an additional T (CTTATATA) (highlighted bold-face). The second block (underlined bold-face) contains seven tandem 12 bp units (CTTTATTATTAT) followed by three truncated repeats (thick underline bold-face) which are missing a second ATT stretch (CTTTATTAT). Positions of amplification primers are highlighted.

APPENDIX VII

COMPLETE SEQUENCE OF LOCUS 16-17

EhR16 →

1 AAGCTTCCTT AGCTCAGCTG GTAGCAGCGT TAGACTGAAG TTCTAAAGGT 50

51 CGTTGGTTCG AACCCGACAG AAGCATACTT CCCCTTTTTA CTCTATATAT 100

101 **TACTTATATT ACTTATATTA CTTATATTAC TTATATTACT TATATTACTT** 150

151 **ATATGTTTAT** ATGTTTATAT TCCTACTTAT A**CTATTTTTA CTATTTTTAC** 200

201 **TATTTTTACT ATTTTTACTA TTTTTACTAT TTTTA**CTACC TATACATACC 250

251 TTCTTATATT ACTATATCTT TATCTTACTA CCTTTTTATT ACTTCTATAT 300

301 CATATCTTAT GTATATATCA CTTTATCACT ATACTTATAC TTATCTTTAT 350

351 TATTATTTCC TTTTACTATA TCACTTTA**TT TTTATATC**TA **TTTATATC**TA 400

401 TTTC**TTTTTA TATCTTTTTA TATC**AGTC**TT TTTATATCTT TTTATATC**AG 450

451 TC**TTTTTATA TC**AGTC**TTTT TATATC**ACTC GATATACACC TTTCTGCCTT 500

EhR18 →

501 TTTCCTGTTT ATAAAGGTTT CATGGTGTAG TTGGTTATCA CATCATCTTG 550

← EhR19

551 ACGTGGTGAG GGTCCCCGGA TCGAACCCGG GTGAAACCAT TTTATACTTA 600

601 TTATCTCTAT ACCTTTATAC CTTTATATGT TTATCTCTTC ATCTC**TTTAT** 650

651 **ACC**TTTATAT ACTTCTATAT ATGTATATAT AT**TTTATACC TTTATACCTT** 700

701 **TATACCTTTA TATGTTTATA TGTTTATATG TTTATATGTT TATATGTTTA** 750

751 **TATGTTTATA TGTTTATATC TTTATATGTT TATATGTTTA TATCTTTATA** 800

801 **TCTTTATATC TTTATATCTT TATATCTTTA TATCTTTATA CC**TTTATATT 850

← EhR17

851 ATCTACTTCT ATATATATTC CTATTTATAT TCCTATTCTT CCCCCTTTTA 900

901 TATACTTATT ATTTCCTTAT ATACACCCTT ATATACTCCC TTTTTACTTC 950

951 CTTCTTTATC CTGT 1000

Note: Four blocks of internal tandem repeats are shown. The first block (bold-face) has six tandem repeat units of 9 bp each (TACTTATAT). The second repeat block (underlined bold-face) contains six tandem 8 bp units (CTATTTTTA). The third block (thick underline bold-face) has seven, 10 bp units both in tandem and as solitary copies (TTTTTATATC). The fourth block contains three types of repeat units which exist both as tandem and solitary copies; five, 8 bp units (TTTATACC) (double underline bold-face); nine, 8 bp units (TTTATATG) (dash underline bold-face); and eight, 8 bp units (TTTATATC) (dot-dash underline bold-face). Positions of amplification primers are highlighted.

APPENDIX VIII

MODIFIED SEQUENCE OF LOCUS 9-4

EhR9 →

```
1     GAATTCTATT CTACATCTAC AGTCCTCCGC TCTACCAACT GAGCTATGGT     50
51    CGGTATCGAT ATCACCTTTT TATATTCTTT TGTTATTCTT TTGATATCCT    100
101   TTTCTTATTC TTTTCTTATT CTTTTTATAT CCTTTTATAT TTCTATATGT    150
151   GCTTATATAT GTTTCTATTC TTTATACCTA TTCTTTATAT ACTTCTTATT    200
201   ATCTCTTCCC TGTTATTTCT CTATACCTTA TCTTTATATT ATGTATATAT    250
251   TTATATTTAT ATTTATATTA TGTATCCTTA TCTTATCTTT ATATTATGTA    300
301   TATATTTATA TTATGTATAT ATTTATATTA TGTGTAGTAT ATATTATTAT    350
351   GTATATGTGT AGTGTACTAT ATGTGTAGTG TACTTATGTG TACTATATAT    400
401   ATATATTCTT GCTAAAAAAT TCCATCGCCG GGAATCGAAC CCGGGCACAA    450
451   ACGTGAAAGG CTTGTATGTT ACCAACTACA CCACAATGGA TTTACCATGA    500
501   CTATATGCCC CTTAGACTTT ATTCTTTATA TTCTTATTAT ACTATTATAC    550
551   TATTATACTA TTATACTATT ATACTATTAT ACTATTATAC TATTATACTA    600
                                          EhR10 →
601   TTATACTATT ATACTATTAT ACTTACTTCT CTTTACCACG ACTTATGTAT    650
                  ← EhR11
651   ATTTATATAT GTTTCTATTC CTTTATATTA TTATTATCTT TATTATTATC    700
701   TTTATTATTA TCTTTATTAT TATCTTTATT ATTATCTTTA TTATTATCTT    750
751   TATTATTATC TTTATTATTA TCTTTATTAT CTTTATTATC TTTATTATTT    800
801   ATGCTTATGT ACTTGACTAT ATACTTATTT CTACTTATTT ATACTCTTTT    850
851   CCTATACCTA TACCTATCCC TATACCTATC TATTAGTATA AAAACTGATC    900
901   CGACCAACCG GATTCGAACC AGTGACCTAA GGAATTC                  950
                  ← EhR4
```

Note: The complete locus 9-4 sequence (Appendix VI) has been reoriented in the order dictated by locus 3-4 (Appendix V) (see section 3.3.5). Two blocks of internal tandem repeats are boxed. Positions of amplification primers are highlighted. Position 392 (**A**) is the original position 1 in the unmodified sequence shown in Appendix VI.

APPENDIX IX

MULTIPLE ALIGNMENT & SEQUENCE COMPARISON BETWEEN LOCUS 3-4 & MODIFIED LOCUS 9-4

```
                1                    primer 9                    50
  Tannich   GAATTCTATT -CTACATCTA C--AG-TCCT CC-G--CTCT ACCAACTGAG
Eichinger   GAATTCTATT -CTACATCTA C--AG-TCCT CC-G--CTCT ACCAACTGAG
   Michel   GAATTCTATT CCTACATCTA CCTAGGTCCT CCCGATCTCT ACCAACTGAG
HM-1: 3-4   .......... .......... .......... .......... .........G
HM-1: 9-4   .......... .CTACATCTA C--AG-TCCT CC-G--CTCT ACCAACTGAG

               51    primer 3                                    100
  Tannich   CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTG- TTATTCTTTT
Eichinger   CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTG- TTATTCTTTT
   Michel   CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTGC TTATTCTTTT
HM-1: 3-4   CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTG- TTATTCTTTT
HM-1: 9-4   CTATGGTCGG TATCGATATC ACCTTTTTAT ATTCTTTTG- TTATTCTTTT

              101                                               150
  Tannich   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTT
Eichinger   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTT
   Michel   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTT
HM-1: 3-4   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTT
HM-1: 9-4   GATATCCTTT TCTTATTCTT TTCTTATTCT TTTTATATCC TTTTATATTC

              151                                               200
  Tannich   CTATATGTGC TTATATATGT TTCTATTCTT TATACCTATT CTTTATATAC
Eichinger   CTATATGTGC TTATATATGT TTCTATTCTT TATACCTATT CTTTATATAC
   Michel   CTATATGTGC TTATATATCT TTCTATTCTT TATACCTATT CTTTATATAC
HM-1: 3-4   CTATATGTGC TTATATATGT TTCTATTCTT TATACCTATT CTTTATATAC
HM-1: 9-4   CTATATGTGC TTATATATGT TTCTATTCTT TATACCTATT CTTTATATAC

              201                                               250
  Tannich   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT
Eichinger   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT
   Michel   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT
HM-1: 3-4   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT
HM-1: 9-4   TTCTTATTAT CTCTTCCCTG TTATTTCTCT ATACCTTATC TTTATATTAT

              251                                               300
  Tannich   GTATATATTT ATATTTATAT TTATATTATG TATCCCTTAT CTTTAT----
Eichinger   GTATATATTT ATATTTATAT TTATATTATG TATCC-TTAT CTT-ATCTTT
   Michel   GTATATATTT ATATTTATAT TTATATTATG TATCCCTTAT CTTTAT----
HM-1: 3-4   GTATATATTT ATATTTATAT TTATATTATG TATCCCTTAT CTTTAT----
HM-1: 9-4   GTATATATTT ATATTTATAT TTATATTATG TATCCCTTAT CTTTAT----

              301                                               350
  Tannich   --ATTATGTA TATATTTTAT ATTATGTATA TATTTTATAT TATGTGTAGT
Eichinger   ATATTATGTA TATATTT-AT ATTATGTATA TATTT-ATAT TATGTGTAGT
   Michel   --ATTATGTA TATATTTTAT ATTATGTATA TATTTTATAT TATGTGTAGT
HM-1: 3-4   --ATTATGTA TATATTTTAT ATTATGTATA TATTTTATAT TATGTGTAGT
HM-1: 9-4   --ATTATGTA TATATTT-AT ATTATGTATA TATTT-ATAT TATGTGTAGT
```

```
            351                                                       400
  Tannich   ATATATTTAT TTTATGTATA TGTGTAGTGT ACTATATGTG TAGTGTACTT
Eichinger   ATATATT-AT T--ATGTATA TGTGTAGTGT ACTATATGTG TAGTGTACTT
   Michel   ATATATTTAT TTTATGTATA TGTGTAGTGT ACTATATGTG TAGTGTACTG
HM-1: 3-4   ATATATTTAT TTTATGTATA TGTGTAGTGT ACTATATGTG TAGTGTACTT
HM-1: 9-4   ATATATT--- -----GTATA TGTGTAGTGT ACTATATGTG TAGTGTACTT

            401                                                       450
  Tannich   ATGTGTACTA TATATATATA TTCTTGCTAA AAAATTCCAT ---CGCCGGG
Eichinger   ATGTGTACTA TATATATATA TTCTTGCTAA AAAATTCCAT ---CGCCGGG
   Michel   ATGTGTACTA TATATATATA TTCTTGCTAA AAAATTCCAT CTACGCCGGG
    Lohia   .......... .......... .......... .......... .........G
HM-1: 3-4   ATGTGTACTA TATATATATA TTCTTGCTAA AAAATTCCAT ---CGCCGGG
HM-1: 9-4   ATGTGTACTA TATATATATA TTCTTGCTAA AAAATTCCAT ---CGCCGGG

            451                                                       500
  Tannich   AATCGAACCC ------GGGG CACAAACGTG AAAGGCTTGT ATGTTACCAA
Eichinger   AATCGAACCC -------GGG CACAAACGTG AAAGGCTTGT ATGTTACCAA
   Michel   AATCGAACC- TAGTCAGGGG CACAAACGTG AAGGGCTTGT ATGTTACCAA
    Lohia   -ATCGA-CCC -------GGG CACAA-CGTG AA-GGCTTGT ATGTTACCAA
HM-1: 3-4   AATCGAACCC ------GGGG CACAAACGTG AAAGGCTTGT ATGTTACCAA
HM-1: 9-4   AATCGAACCC ------GGGG CACAAACGTG AAAGGCTTGT ATGTTACCAA

            501                                                       550
  Tannich   C-TACACCAC AATGGATTTA CCATGACTAT AT--GCCCC- TTAGACTTTA
Eichinger   C-TACACCAC AATGGATTTA CCATGACTAT AT--GCCCC- TTAGACTTTA
   Michel   C-TACACCAC AATGGATTTA CCATGACTAT ATTAGCCCCC TTAGACTTTA
    Lohia   CCTACACCAC AATGGATTTA CCATGACTAT AT--GCCCC- -TAGACT--A
HM-1: 3-4   C-TACACCAC AATGGATTTA CCATGACTAT AT--GCCCC- TTAGACTTTA
HM-1: 9-4   C-TACACCAC AATGGATTTA CCATGACTAT AT--GCCCC- TTAGACTTTA

            551                                                       600
  Tannich   TTCTTTATAT TCCTTATTAT ACTTATTATA CTTATTATAC TTATTATACT
Eichinger   TTCTTTATAT T-CTTATTAT AC-TATTATA C-TATTATAC -TATTATAC-
   Michel   TTCTTTATAT TCCTTATTAT ACTTATTATA CTTATTATAC TTATTATACT
    Lohia   -TCT--ATAT TCCTTATTAT ACTTATTATA CTTATTATAC TTATTATACT
HM-1: 3-4   TTCTTTATAT TCCTTATTAT ACTTATTATA CTTATTATAC TTATTATACT
HM-1: 9-4   TTCTTTATAT TCCTTATTAT ACTTATTATA CTTATTATAC TTATTATACT

            601                                                       650
  Tannich   TATTATACTT ATTATACTTA TTATA----- ---------- ----------
Eichinger   TATTATAC-T ATTATAC-TA TTATAC-TAT TATAC-TATT ATAC-TATTA
   Michel   TATTATACTT ATTATACTTA TTATACTTAT TATACTTATT ATACTTATTA
    Lohia   TATTATACTT ATTATACTTA TTATACTTAT TATACTTATT ATACTTATTA
HM-1: 3-4   TATTATACTT ATTATA---- ---------- ---------- ----------
HM-1: 9-4   TATTATACTT ATTATACTTA TTATACTTAT TATA------ ----------

            651                                                       700
  Tannich   ---------- -CTTTACTTC TCTTTTACCA CGACTTATGT ATATTTATAT
Eichinger   TAC-TATTAT ACTT-ACTTC TCTTT-ACCA CGACTTATGT ATATTTATAT
   Michel   TACTTATTAT ACTTTACTTC TCTTTTACCA CGA-TTATGT ATATTTATAT
    Lohia   TACTTATTAT ACTTTACTTC TCTTTTACCA CGACTTATGT ATATTTATAT
HM-1: 3-4   ---------- -CTTTACTTC TCTTTTACCA CGACTTATGT ATATTTATAT
HM-1: 9-4   ---------- -CTTTACTTC TCTTTTACCA CGACTTATGT ATATTTATAT
```

```
                701                                                     750
  Tannich   ATGTTTCTAT TCCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT
Eichinger   ATGTTTCTAT TCCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT
   Michel   ATGTTTCTAT TGCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT
    Lohia   ATGTTTCTAT TCCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT
HM-1: 3-4   ATGTTTCTAT TCCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT
HM-1: 9-4   ATGTTTCTAT TCCTTTATAT TATTATTATC TTTATTATTA TCTTTATTAT

                751                                                     800
  Tannich   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA
Eichinger   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA
   Michel   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA
    Lohia   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA
HM-1: 3-4   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA
HM-1: 9-4   TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC TTTATTATTA

                801                                                     850
  Tannich   TCTTTATTAT TAT------- ---------- ---------- ---------C
Eichinger   TCTTTATTAT TAT------- ---------- ---------- ---------C
   Michel   TCTTTATTAT TATCTTTATT ATTAT----- ---------- ---------C
    Lohia   TCTTTATTAT TATCTTTATT ATTAT----- ---------- ---------C
HM-1: 3-4   TCTTTATTAT TATCTTTATT ATTATCTTTA TTATTATCTT TATTATTATC
HM-1: 9-4   TCTTTATTAT TAT------- ---------- ---------- ---------C

                851                                                     900
  Tannich   TTTATTATCT TTATTATCTT TATTATCTTT ATTATTTATG CTTATGTACT
Eichinger   TTTATTATCT TTATTATCTT TATTAT---- -----TTATG CTTATGTACT
   Michel   TTTATTATCT TTATTATCTT TATTAT---- -----TTATG CTTATGTACT
    Lohia   TTTATTATCT TTATTATCTT TATTAT---- -----TTATG CTTATGTACT
HM-1: 3-4   TTTATTATCT TTATTATCTT TATTAT---- -----TTATG CTTATGTACT
HM-1: 9-4   TTTATTATCT TTATTATCTT TATTATCTTT ATTATTTATG CTTATGTACT

                901                                                     950
  Tannich   TGACTATATA CTTATTTCTA CTTATTTATA CTCTTTTTCC TATACCTATA
Eichinger   TGACTATATA CTTATTTCTA CTTATTTATA CTCTTTT-CC TATACCTATA
   Michel   TGACTATATA CTTATTTCTA CTTATTTATA CTCTTTTTCC TATACCTATA
    Lohia   TGACTATATT CTTATTTCTA CTTATTTATA CTCTTTTTCC TATACCTATA
HM-1: 3-4   TGACTATATA CTTATTTCTA CTTATTTATA CTCTTTTTCC TATACCTATA
HM-1: 9-4   TGACTATATA CTTATTTCTA CTTATTTATA CTCTTTTTCC TATACCTATA

                951                                                    1000
  Tannich   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATCCG ACCAACCGGA
Eichinger   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATCCG ACCAACCGGA
   Michel   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATCCG ACCAACCGGA
    Lohia   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATC.. ..........
HM-1: 3-4   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATCCG ACCAACCGGA
HM-1: 9-4   CCTATCCCTA TACCTATCTA TTAGTATAAA AACTGATCCG ACCAACCGGA

          →1001  primer 4    ←
  Tannich   TTCGAACCAG TGACCTAAG
Eichinger   TTCGAACCAG TGACCTAAG
   Michel   TTCGAACCAG TGACCTAAG
HM-1: 3-4   TTCGAACCAG TGACCTAAG
HM-1: 9-4   TTCGAACCAG TGACCTAAG
```

Note: Multiple nucleotide sequence alignment is shown between; Willhoeft and **Tannich** (accession # AF181969), Rosales-Encina and **Eichinger** (accession #

AF265348; our designation: locus 9-4), **Michel** *et al.* (accession # M77091; our designation: locus 3-4), **Lohia** *et al* (accession # M55340) and **HM-1:IMSS Locus 3-4 & 9-4** clones (Chapter 3, section 3.3.5). The Eichinger sequence seen here is the modified version shown in Appendix VIII. Dots indicate start- and end-points of readable sequences. Primer R3, R9 & R4 sequences are highlighted. Dashes have been used to maximise alignment. Repeat regions are boxed. Bold highlighted regions show single nucleotide substitutions. The $tRNA^{Tyr}$ gene (Tannich & Eichinger: thick underline) $tRNA^{Leu}$ gene (Michel: double underline) and $tRNA^{Glu}$ gene (Tannich, Eichinger & Michel: underlined) are shown.

APPENDIX X

E. histolytica ISOLATES - GLOBAL SURVEY

No	Isolates	Clinical Diagnosis †	Microscopy †	Zymodeme	Serology ✳
1	HM-1:IMSS Clone 9#	Amoebic dysentery	NA	II	NA
2	200:NIH#	Amoebic dysentery	NA	II	NA
3	H-303:NIH#	Bilateral Amoebic empyema & Amoebic dysentery	NA	II	NA
4	*IULA:0593:2✠*	Amoebic dysentery	*E.h*T, blood & mucus	II	NA
5	IULA:1092:1✠	Amoebic dysentery	*E.h*T, blood & mucus	II	NA
6	HK9#	Amoebic dysentery	NA	II	NA
7	Rahman#	ACP	NA	II	NA
8	SD157#	Amoebic colitis	NA	II	NA
9	DKB#	Amoebic dysentery	NA	II	NA
10	HB301:NIH#	Amoebic dysentery	NA	II	NA
11	8691	Amoebic colitis	*E.h*T	II	NA
12	4530	Amoebic colitis	*E.h*T	XIV	NA
13	1320300	Amoebic colitis	*E.h*T	II	NA
14	2701	Amoebic colitis	*E.h*T	II	NA
15	48286	Amoebic colitis	*E.h*T	II	NA
16	1363050	Amoebic colitis	*E.h*T	II	NA
17	1375975	Amoebic colitis	*E.h*T	II	NA
18	44298	Amoebic colitis	*E.h*T	XIV	NA
19	58996	Amoebic colitis	*E.h*T	II	NA
20	1353600	Amoebic colitis	*E.h*T	II	NA
21	J1	Colitis	*E.h/ E.d* blood & mucus+	II	+ve
22	J2	Colitis	*E.h/ E.d*	XIX	+ve
23	J3	Colitis	*E.h/ E.d*	II	-ve
24	J4	NA	*E.h/ E.d*	XIX	+ve
25	J5	Asymptomatic	*E.h/ E.d*C	II	+ve
26	J6	Asymptomatic	*E.h/ E.d*	II	+ve
27	*J7*	Colitis & Liver abscess	*E.h/ E.d* blood & mucus+	II	+ve
28	J8	Colitis & Liver abscess	*E.h/ E.d*	XIV	+ve
29	J9	NA	*E.h/ E.d* blood+	II	+ve
30	J10	Colitis	*E.h/ E.d*T	II	+ve
31	J11	Liver abscess	*E.h/ E.d*	XIV	+ve
32	J12	Asymptomatic	*E.h/ E.d*C	II	+ve
33	J13	Asymptomatic	*E.h/ E.d*C	XIX	+ve
34	J14	Asymptomatic	*E.h/ E.d*C	XIX	+ve
35	J15	Colitis	*E.h/ E.d* blood+	XIX	+ve
36	J16	Colitis	*E.h/ E.d*C blood+	XIX	+ve
37	J17	Liver abscess	*E.h/ E.d*T	II	+ve
38	J18	NA	*E.h/ E.d*C	II	+ve
39	J19	NA	*E.h/ E.d*C	II	+ve
40	J20	NA	*E.h/ E.d*C	II	+ve
41	J21	NA	*E.h/ E.d*C	II	+ve
42	J22	NA	*E.h/ E.d*C	II	+ve

43	YS-27	Liver abscess	NA	II	NA
44	887C	Diarrhoea	*E.h/ E.d*C	ND	ND
45	20-2	AC	*E.h/ E.d*C	ND	+ve
46	26-6	AC	*E.h/ E.d*C, *Ascaris*, *T. trichiura*	ND	-ve
47	41-1	AC	*E. coli*	ND	-ve
48	46-5	AC	*E.h/ E.d*C, *Giardia*, *T. trichiura*	ND	+ve
49	66-1	AC	*E.h/ E.d*C	ND	+ve
50	72-1	AC	*T. trichiura*	ND	-ve
51	82-4	AC	*E.h/ E.d*C, *Ascaris*, *T. trichiura*	ND	+ve
52	84-3	AC	*Giardia*, *Trichuris*	ND	+ve
53	85-3	AC	*E.h/ E.d*C, *Ascaris*	ND	+ve
54	148-7	AC	*E.h/ E.d*C, *Ascaris*, *T. trichiura*, Hookworm	ND	+ve
55	25591C	Liver abscess	*E.h/ E.d*C	ND	+ve
56	25591F				
57	26825C	Fever/ Diarrhoea	*E.h/ E.d*C	ND	+ve
58	26825F				
59	27749C	NA	*E.h/ E.d*C	ND	+ve
60	27749F				
61	28577C	Bloody diarrhoea	*E.h/ E.d*C, *Ex. nana*	ND	+ve
62	28577F				
63	28639C	NA	*E.h/ E.d*C	ND	ND
64	28639F				
65	29146C	NA	*E.h/ E.d*C	ND	ND
66	29146F				
67	29978C	Asymptomatic	*Ex. nana*	ND	ND
68	29978F				
69	30325C	Diarrhoea	*E.h/ E.d*C	ND	+ve
70	30325F				
71	32083C	Liver abscess	*E. coli*	ND	+ve
72	32083F				
73	32257C	Diarrhoea	*E.h/ E.d*C	ND	ND
74	32257F				
75	1203	Asymptomatic	*D. fragilis*, *E.h/ E.d*C, *E. coli*, *B. hominis*	IIα-	+ve
76	1205	Asymptomatic	*E.h/ E.d*C, *E. coli*	IIα-	Weak +ve
77	822	Amoebic dysentery	*E.h*C/T	Hexokinase (Fast)	ND
78	2024	Amoebic dysentery	*E.h*C/T, *Salmonella* B (copro culture)	Hexokinase (Fast)	ND
79	37.0C	Index Case	*T. trichiura*+	XIX	Strong +ve
80	39.384C	ACP	*E.h/ E.d*C++, *E. coli*+++	XI	Weak +ve
81	2596	ACP	*E.h/ E.d*C++, *E. coli*+	XI	Strong +ve
82	CSP	Dysenteric colitis	*A. lumbricoides*, *T. trichiura*, *Taenia* sp & Hookworm	II	+ve
83	462	Asymptomatic	No Parasites Seen	II	-ve
84	RPS	Asymptomatic	*E. coli*	II	-ve

† Clinical Diagnosis & Microscopy: as provided by supplier of the sample except (#) where information was obtained from The American Type Culture Collection (ATCC) (www.atcc.org) and (✠) where the information was obtained from the

relevant citation (Urdaneta *et al*, 1995). "+, ++, +++" reflect the relative abundance of parasites seen by the microscopist.

Abbreviations; ACP: Asymptomatic cyst passer; AC: Asymptomatic carrier; Index case: Convalescent amoebic liver abscess patient, cured as at recruitment date, now asymptomatic. *E.h/ E.d* C/T: *Entamoeba histolytica/ Entamoeba dispar* Cyst/ Trophozoite; *Ex. nana*: *Endolimax nana*; *E. coli*: *Entamoeba coli*; *D. fragilis*: *Dientamoeba fragilis*; *B. hominis*: *Blastocystis hominis*; *T. trichiura*: *Trichuris trichiura*; *A. lumbricoides*: *Ascaris lumbricoides*; NA: Not available; ND: Not done.

✳ **Serology:**

Isolate No 21 - 42: serological analysis was made by one or more of three means (personal communication) and the results summarised by us (above) are derived from them. GDP (Gel Double Diffusion Precipitation Test), results reported as precipitation +ve or -ve after amidoblack staining. ELISA (Enzyme Linked Immunosorbent Assay), results reported as negative control value (NCV) (OD 405 of sample/ OD 405 of negative control) where >2.0 is -ve and ≤2.0 is +ve and; IFA (Immunofluorescence Antibody Test), results reported as +ve when fluorescence was observed on fixed amoebae in x 64 or more diluted serum.

Isolate No 45 - 54: serology results were obtained using an ELISA based on a recombinant *E. histolytica* surface antigen (Lotter *et al*, 1992). In addition samples were characterised using the *E. histolytica*-specific lectin coproantigen ELISA (TechLab Inc., Blacksburg VA, USA) stool antigen detection kit and PCR (Blessmann *et al*, 2002).

Isolate No 55 - 76: serological analysis was done with their home made ELISA using crude lysate of strain HM-1:IMSS where a titre of ≥1:40 was considered +ve.

Isolate No 79 - 81: Serological assessments were made using the amoebic agarose gel diffusion test (AGDT). Results were reported as strong +ve when precipitin arcs appeared within ±20 hours or weak +ve when the precipitin arcs appeared at or after 40 hours (personal communication).

Isolate No 82 - 84: sera were tested by ELISA (Silva *et al*, 1997) and the minimum sera dilution used to consider the test +ve was 1:100 (personal communication).

APPENDIX XI

SOUTH AFRICAN *E. dispar* ISOLATES - CLINICAL, PARASITOLOGICAL & SEROEPIDEMIOLOGICAL FEATURES

No	Isolates	Clinical Diagnosis †	Microscopy †	Zymodeme	Serology
1	16.156N	ACP	No Parasites Seen	I	-ve
2	19.190N	ACP	*T. trichiura*++, *A. lumbricoides* +++	XVII	-ve
3	22.211L	ACP	ND	III	-ve
4	22.212M	ACP	*Ex. nana*+, *T. trichiura*+	I	-ve
5	22.217L	ACP	*A. lumbricoides*+	I	-ve
6	27.270N	ACP	*E.h/E.d*C++, *E. coli*+, *T. trichiura*+	I	-ve
7	29.284M	ACP	No Parasites Seen	III	-ve
8	29.284N	ACP	*E.h/E.d*C++, *E. coli*+, *T. trichiura*+	XVII	-ve
9	36.352N	ACP	ND	I	-ve
10	36.352L	ACP	*E.h/E.d*C++, *E. coli*++, *C. mesnili*++, *T. trichiura*+, *A. lumbricoides* +++	I	-ve
11	40.395M	ACP	*E.h/E.d*C++, *E. coli*++, *T. trichiura*	XVII	-ve
12	40.397K	ACP	*G. lamblia*+++, *Ex. nana*, *S. mansoni*	I	-ve
13	40.397L	ACP	*E.h/E.d*C+, *Ex. nana*++, *E. coli*+, *T. trichiura*+	III	-ve
14	40.397M	ACP	*E.h/E.d*C+, *Ex. nana*+, *E. coli*+,	I	-ve
15	40.399M	ACP	*A. lumbricoides* ++, *Ex. nana*++	V	-ve
16	41.410K	ACP	*E.h/E.d*C++, *T. trichiura*+, *A. lumbricoides* ++	I	-ve
17	41.410M	ACP	*E. coli*+, *T. trichiura*+, *A. lumbricoides* ++	I	-ve
18	44.440K	ACP	*Ex. nana*++, *E. hartmanni*++, *E. coli*++, *T. trichiura*+,	I	-ve
19	49.484L	ACP	*E. coli*+	XVII	-ve
20	49.486L	ACP	*E.h/E.d*C+++	I	-ve
21	50.503I	ACP	No Parasites Seen	I	-ve
22	52.521K	ACP	*E.h/E.d*C++, *T. trichiura*+ *A. lumbricoides* ++	XVII	-ve
23	53.539J	ACP	No Parasites Seen	I	-ve
24	57.579K	ACP	No Parasites Seen	XVII	-ve
25	58.590L	ACP	*E.h/E.d*C++, *E. coli*++, *T. trichiura*+	III	-ve
26	58.588L	ACP	*E.h/E.d*C+	I	-ve
27	59.596K	ACP	*E. coli*+	I	-ve
28	59.595I	ACP	*E. coli*++, *T. trichiura*+	III	-ve
29	59.595K	ACP	No Parasites Seen	I	-ve
30	60.0J	IC	*E.h/E.d*C+++, *E. coli*++, *Ex. nana*+++	I	Weak +ve
31	60.0Ji	IC	*E.h/E.d*C+++, *E. coli*++, *Ex. nana*+++	I	Weak +ve
32	60.607J	ACP	*E.h/E.d*C+, *E. coli*+, *Ex. nana*+	I	-ve
33	60.607Ji	ACP	*E.h/E.d*C+, *E. coli*+, *Ex. nana*+	I	-ve
34	61.612JB	ACP	*E.h/E.d*C+++, *C. mesnili*++	XVIII	Strong +ve

35	61.612KA	ACP	*E.h/E.d*C++	XVII	Strong +ve
36	62.628K	ACP	*E.h/E.d*C++, *I. buetschlii*+	I	Weak +ve
37	62.629I	ACP	*A. lumbricoides* +	I	-ve
38	62.629K	ACP	*E.h/E.d*C+, *Ex. nana*+, *T. trichiura*+	I	-ve
39	64.641H	ACP	*E.h/E.d*C+++	I	-ve
40	64.648K	ACP	*E.h/E.d*C++, *T. trichiura*+, *Strongyloides*	III	-ve
41	69.692I	ACP	*E. coli*+	I	-ve
42	69.692Ii	ACP	*E. coli*+	I	-ve
43	69.698I	ACP	*E. hartmanni*+++, *E. coli*+,	III	-ve
44	69.698Ii	ACP	*E. hartmanni*+++, *E. coli*+,	III	-ve
45	69.699L	ACP	No Parasites Seen	I	-ve
46	71.719H	ACP	No Parasites Seen	III	-ve
47	72.721J	ACP	*E.h/E.d*C++, *E. coli*+, *Ex. nana*+, *T. trichiura*+,	I	-ve
48	78.782H	ACP	*E.h/E.d*C+, *E. coli*++, *E. hartmanni*	I	-ve
49	78.784H	ACP	*A. lumbricoides* ++, *T. trichiura*+	XVII	-ve
50	78.789H	ACP	*A. lumbricoides* +, *T. trichiura*+	III	-ve
51	81.814G	ACP	*C. mesnili*+++, *G. lamblia*++, *T. trichiura*+	I	-ve
52	81.815G	ACP	*G. lamblia*+++, *T. trichiura*+	I	-ve
53	83.840G	ACP	*T. trichiura*++	I	-ve
54	84.843I	ACP	*E.h/E.d*C+, *E. hartmanni*++, *T. trichiura*+, *E. coli*+++, *A. lumbricoides* +	I	-ve
55	84.843K	ACP	*E.h/E.d*C++, *E. hartmanni*++, *T. trichiura*+,	I	-ve
56	85.860H	ACP	*E. hartmanni*+, *E. coli*+++, *T. trichiura*+,	XVI	Weak +ve
57	86.861G	ACP	*E. coli*+++, *Taenia* spp	I	-ve
58	86.862G	ACP	No Parasites Seen	III	-ve
59	86.863G	ACP	*A. lumbricoides* +	I	-ve
60	86.863Gi	ACP	*A. lumbricoides* +	I	-ve
61	86.870J	ACP	No Parasites Seen	XVII	-ve
62	87.874H	ACP	*E. hartmanni*+, *I. buetschlii*+	I	-ve
63	87.874I	ACP	*E.h/E.d*C+, *Ex. nana*+, *E. coli*+,	I	-ve
64	87.878H	ACP	*E.h/E.d*C+++, *Ex. nana*++,	XVII	-ve
65	87.878I	ACP	*E.h/E.d*C+, *Ex. nana*++, *T. trichiura*+, *I. buetschlii*+	I	-ve
66	87.880G	ACP	*E.h/E.d*C++, *E. coli*+, *E. hartmanni*++	III	-ve
67	87.880Gi	ACP	*E.h/E.d*C++, *E. coli*+, *E. hartmanni*++	III	-ve
68	87.880H	ACP	*E.h/E.d*C+, *E. coli*+, *E. hartmanni*++	XVII	-ve
69	87.880I	ACP	*E.h/E.d*C+	I	-ve
70	88.881G	ACP	*T. trichiura*+	I	-ve
71	88.881H	ACP	*E.h/E.d*C+, *Ex. nana*++, *E. coli*+, *T. trichiura*+	I	-ve
72	89.891I	ACP	*E.h/E.d*C++, *T. trichiura*+	I	-ve
73	89.900K	ACP	*E. coli*++, Hookworm+, *A. lumbricoides* ++	XVII	-ve
74	90.902K	ACP	*E.h/E.d*C++, *E. coli*+, *I. buetschlii*+	I	-ve
75	90.904G	ACP	*E.h/E.d*C++, *T. trichiura*+	VIII	-ve
76	90.905G	ACP	*E.h/E.d*C++	XVII	-ve
77	90.905I	ACP	*E.h/E.d*C+, *E. hartmanni*++	I	-ve
78	90.907G	ACP	*E.h/E.d*C+	I	-ve
79	90.908GB	ACP	*E. coli*+	XIII	-ve

80	90.909HA	ACP	*A. lumbricoides* +, *T. trichiura*+	III	-ve
81	90.909IB	ACP	*E. hartmanni*+++, *I. buetschlii*+++	I	-ve
82	91.911I	ACP	*E. hartmanni*++	III	-ve
83	91.915I	ACP	*E.h/E.d*C++, *Ex. nana*++	III	-ve
84	91.919I	ACP	*E.h/E.d*C++	XVII	-ve
85	92.926G	ACP	*E.h/E.d*C++, *E. coli*+, *Ex. nana*++	III	-ve
86	92 926Gi	ACP	*E.h/E.d*C++, *E. coli*+, *Ex. nana*++	III	-ve
87	92.926J	ACP	*E. coli*+	XVII	-ve
88	93.931G	ACP	*E.h/E.d*C++	I	-ve
89	93.934G	ACP	*E.h/E.d*C++, *Ex. nana*+++	III	-ve
90	93.934H	ACP	*E.h/E.d*C++, *E. hartmanni*+++	I	-ve
91	93.934Hi	ACP	*E.h/E.d*C/T++,	XVII	-ve
92	93.939G	ACP	No Parasites Seen	I	-ve
93	94.942I	ACP	*E.h/E.d*C++, *E. coli*+, *T. trichiura*+	I	-ve
94	94.943G	ACP	*E. coli*+	V	-ve
95	94.943I	ACP	*E.h/E.d*C+++, *E. coli*+,	I	-ve
96	95.958I	ACP	*E.h/E.d*C++, *T. trichiura*+	I	-ve
97	97.0F	IC	No Parasites Seen	I	Weak +ve
98	97.0Fi	IC	No Parasites Seen	I	Weak +ve
99	97.0G	IC	*E.h/E.d*C++, *Ex. nana*+++, *E. coli*++	I	Weak +ve
100	97.0Gi	IC	*E.h/E.d*C++, *Ex. nana*+++, *E. coli*++	I	Weak +ve
101	99.993G	ACP	*E. coli*+, Ex. nana+, *T. trichiura*+	I	-ve
102	99.996F	ACP	*E.h/E.d*C+, *A. lumbricoides*+	I	-ve
103	99.996G	ACP	*A. lumbricoides*+++	XVII	-ve
104	11581	ACP	ND	I	ND
105	11590	ACP	ND	I	ND
106	11593	ACP	ND	I	ND
107	11597	ACP	ND	I	ND
108	11598	ACP	ND	I	ND
109	11628	ACP	ND	III	ND
110	11663	ACP	ND	I	ND
111	11691	ACP	ND	I	ND

† Clinical Diagnosis & Microscopy: as provided by supplier of the sample i.e. T.F.H.G Jackson and S. Reddy, Medical Research Council Durban, South Africa. "+, ++, +++" reflect the relative abundance of parasites seen by the microscopist.

Serology: was done using the amoebic agarose gel diffusion test (AGDT). Results were reported as strong +ve when precipitin arcs appeared within ±20 hours or weak +ve when the precipitin arcs appeared at or after 40 hours (personal communication).

Abbreviations; ACP: Asymptomatic cyst passer; IC: Index case (convalescent amoebic liver abscess patient, cured as at recruitment date, now asymptomatic). *E.h/ E.d*C/T: *Entamoeba histolytica/ Entamoeba dispar* Cyst/ Trophozoite; *Ex. nana*: *Endolimax nana*; *E. coli*: *Entamoeba coli*; *E. hartmanni*: *Entamoeba hartmanni*; *I. buetschlii*: *Iodamoeba buetschlii*; *C. mesnili*: *Chilomastix mesnili*; *G. lamblia*: *Giardia lamblia*; *T. trichiura*: *Trichuris trichiura*; *A. lumbricoides*: *Ascaris lumbricoides*; ND: Not done.

APPENDIX XII

SOUTH AFRICAN *E. histolytica* ISOLATES - CLINICAL, PARASITOLOGICAL & SEROEPIDEMIOLOGICAL FEATURES

No	Isolate #	Clinical Diagnosis †	Microscopy †	Zymodeme	Serology
1	37.0C	IC	*T. trichiura*+	XIX	Strong +ve
2	39.0L	IC	*E.h/E.d*C+++, *T. trichiura*+	II	Strong +ve
3	39.384C	ACP	*E.h/E.d*C++, *E. coli*+++	XI	Weak +ve
4	39.384KB	ACP	No Parasites Seen	XI	Weak +ve
5	2596	ACP	*E.h/E.d*C++, *E. coli*+	XI	Strong +ve
6	56.0L	IC	No Parasites seen	II	Strong +ve
7	63.0I	IC	*E.h/E.d*T+++, *E.h/E.d*C+, *E. coli*++	II	Strong +ve
8	63.0Ii	IC	*E.h/E.d*T+++, *E.h/E.d*C+, *E. coli*++	II	Strong +ve
9	63.0IA	IC	*E.h/E.d*T+++, *E.h/E.d*C+, *E. coli*++	II	Strong +ve
10	63.0JA	IC	*E.h/E.d*C+++	II	Strong +ve
11	63.0JB	IC	*E.h/E.d*C+++	II	Strong +ve
12	63.0K	IC	*E.h/E.d*C+	II	Strong +ve
13	63.632K	ACP	No Parasites Seen	II	-ve
14	63.635K	ACP	No Parasites Seen	II	-ve
15	63.638L	ACP	*E.h/E.d*C+, *I. buetschlii*+, *T. trichiura*+	II	-ve
16	73.733L	ACP	*E.h/E.d*C++, *E. coli*++, *Ex. nana*++, *T. trichiura*+	II	-ve
17	73.737I	ACP	*E.h/E.d*C+, *E. coli*+	II	Strong +ve
18	73.737IA	ACP	*E.h/E.d*C++, *E. coli*+, *T. trichiura*+	II	Strong +ve
19	73.737IB	ACP	No Parasites Seen	XI	Strong +ve
20	73.737L	ACP	*E. coli*+, *T. trichiura*+	II	Strong +ve
21	73.738HA	ACP	No Parasites Seen	XI	Strong +ve
22	73.738HAi	ACP	No Parasites Seen	XI	Strong +ve
23	73.738I	ACP	No Parasites Seen	II	Strong +ve
24	73.738IB	ACP	*A. lumbricoides*+	II	Strong +ve
25	73.738L	ACP	*A. lumbricoides*+	II	Strong +ve
26	73.740I	ACP	*E.h/E.d*C+, *T. trichiura*+	XI	-ve
27	73.740Ii	ACP	*E.h/E.d*C+, *T. trichiura*+	XI	-ve
28	85.857H	ACP	*Ex. nana*++	II	Strong +ve
29	85.858H	ACP	*E. coli*+, *T. trichiura*+	XI	Strong +ve
30	85.859G	ACP	No Parasites Seen	II	Strong +ve
31	85.859H	ACP	*E.h/E.d*C++, *E. coli*++	II	Strong +ve
32	87.0H	IC	No Parasites Seen	II	Strong +ve
33	89.892I	ACP	*E.h/E.d*C+, *T. trichiura*+, *A. lumbricoides*+	II	-ve
34	90.901HA	ACP	*A. lumbricoides*+	II	Weak +ve
35	90.901HAi	ACP	*A. lumbricoides*+	II	Weak +ve
36	90.901HAii	ACP	*A. lumbricoides*+	II	Weak +ve
37	90.901K	ACP	*E.h/E.d*C+++	II	Weak +ve
38	90.901KA	ACP	ND	XI	Weak +ve
39	90.903G	ACP	No Parasites Seen	II	Strong +ve

40	90.903I	ACP	*E.h/E.d*C+	XI	Strong +ve
41	90.903K	ACP	*E.h/E.d*C+++, *E. coli*++, *I. buetschlii*++, *E. hartmanni*+++	II	Strong +ve
42	90.906G	ACP	*E.h/E.d*C++, *T. trichiura*+	XI	Strong +ve
43	91.913G	ACP	*E.h/E.d*C+	II	Strong +ve
44	91.913I	ACP	No Parasites Seen	XI	Strong +ve
45	100.0F	IC	No Parasites Seen	II	Strong +ve
46	100.0Fi	IC	No Parasites Seen	II	Strong +ve
47	100.1006G	ACP	*E.h/E.d*C+, *E. coli*+, *T. trichiura*+, *S. mansoni*, *A. lumbricoides*+	XIV	Strong +ve

† **Clinical Diagnosis & Microscopy:** as provided by supplier of the sample i.e. T.F.H.G Jackson and S. Reddy, Medical Research Council Durban, South Africa. "+, ++, +++" reflect the relative abundance of parasite (C/T) seen by the microscopist.

Serology: was done using the amoebic agarose gel diffusion test (AGDT). Results were reported as strong +ve when precipitin arcs appeared within ±20 hours or weak +ve when the precipitin arcs appeared at or after 40 hours (personal communication).

Abbreviations; ACP: Asymptomatic cyst passer; IC: Index case (convalescent amoebic liver abscess patient, cured as at recruitment date, now asymptomatic). *E.h/ E.d*C/T: *Entamoeba histolytica/ Entamoeba dispar* Cyst/ Trophozoite; *Ex. nana*: *Endolimax nana*; *E. coli*: *Entamoeba coli*; *E. hartmanni*: *Entamoeba hartmanni*; *I. buetschlii*: *Iodamoeba buetschlii*; *S. mansoni*: *Schistosoma mansoni*; *T. trichiura*: *Trichuris trichiura*; *A. lumbricoides*: *Ascaris lumbricoides*; ND: Not done.

APPENDIX XIII

SOUTH AFRICAN *E. dispar* ISOLATES - FAMILY GROUPS & DWELLING TYPE

Family Unit	Isolate #	Relationship	Dwelling
16	16.156N	None	IS
19	19.190N	None	Farm
22	22.211L	NCA	IS
	22.212M	NCA	IS
	22.217L	NCA	IS
27	27.270N	None	Farm
29	29.284M	None	Suburb
	29.284N	None	Suburb
36	36.352N	None	NA
	36.352L	None	NA
40	40.395M	NCA	IS
	40.397K	NCA	IS
	40.397L	NCA	IS
	40.397M	NCA	IS
	40.399M	NCA	IS
41	41.410K	None	Farm
	41.410M	None	Farm
44	44.440K	None	IS
49	49.484L	NCA	IS
	49.486L	NCA	IS
50	50.503I	None	IS
52	52.521K	None	IS
53	53.539J	None	IS
57	57.579K	NCA	Suburb
58	58.590L	NCA	IS
	58.588L	NCA	IS
59	59.596K	NCA	Farm
	59.595I	NCA	Farm
	59.595K	NCA	Farm
60	60.0J	IC	IS
	60.0Ji	IC	IS
	60.607J	NCA	IS
	60.607Ji	NCA	IS
61	61.612JB	None	IS
	61.612KA	None	IS
62	62.628K	NCA	Suburb
	62.629I	NCA	Suburb
	62.629K	NCA	Suburb
64	64.641H	Mother	Farm
	64.648K	Daughter	Farm
69	69.692I	NCA	IS
	69.692Ii	NCA	IS
	69.698I	NCA	IS
	69.698Ii	NCA	IS
	69.699L	NCA	IS
71	71.719H	None	Farm

72	72.721J	None	Suburb
78	78.782H	NCA	IS
	78.784H	NCA	IS
	78.789H	NCA	IS
81	81.814G	Sister	IS
	81.815G	Sister	IS
83	83.840G	None	Suburb
84	84.843I	NCA	IS
	84.843K	NCA	IS
85	85.860H	NCA	Farm
86	86.861G	NCA	IS
	86.862G	Daughter	IS
	86.863G	Mother	IS
	86.863Gi	Mother	IS
	86.870J	NCA	IS
87	87.874H	Daughter	Farm
	87.874I	Daughter	Farm
	87.878H	Son	Farm
	87.878I	Son	Farm
	87.880G	NCA	Farm
	87.880Gi	NCA	Farm
	87.880H	NCA	Farm
	87.880I	NCA	Farm
88	88.881G	None	Suburb
	88.881H	None	Suburb
89	89.891I	Mother	Suburb
	89.900K	NCA	IS
90	90.902K	NCA	Farm
	90.904G	NCA	Farm
	90.905G	NCA	Farm
	90.905I	NCA	Farm
	90.907G	NCA	Farm
	90.908GB	Husband	Farm
	90.909HA	NCA	Farm
	90.909IB	NCA	Farm
91	91.911I	NCA	Farm
	91.915I	NCA	Farm
	91.919I	NCA	Farm
92	92.926G	None	IS
	92.926Gi	None	IS
	92.926J	None	IS
93	93.931G	NCA	IS
	93.934G	NCA	IS
	93.934H	NCA	IS
	93.934Hi	NCA	IS
	93.939G	NCA	IS
94	94.942I	NCA	IS
	94.943G	NCA	IS
	94.943I	NCA	IS
95	95.958I	None	Suburb
97	97.0F	IC	IS
	97.0Fi	IC	IS
	97.0G	IC	IS
	97.0Gi	IC	IS

99	99.993G	NCA	IS
	99.996F	NCA	IS
	99.996G	NCA	IS
School Children from Langebaan	11581	Same School	Suburb
	11590	Same School	Suburb
	11593	Same School	Suburb
	11597	Same School	Suburb
	11598	Same School	Suburb
	11628	Same School	Suburb
	11663	Same School	Suburb
	11691	Same School	Suburb

Abbreviations; IC: Index Case (cured of amoebic liver abscess (ALA), now asymptomatic); NCA: Neighbour or Close Associate; IS: Informal settlement.

APPENDIX XIV

SOUTH AFRICAN *E. histolytica* ISOLATES - FAMILY GROUPS AND DWELLING TYPE

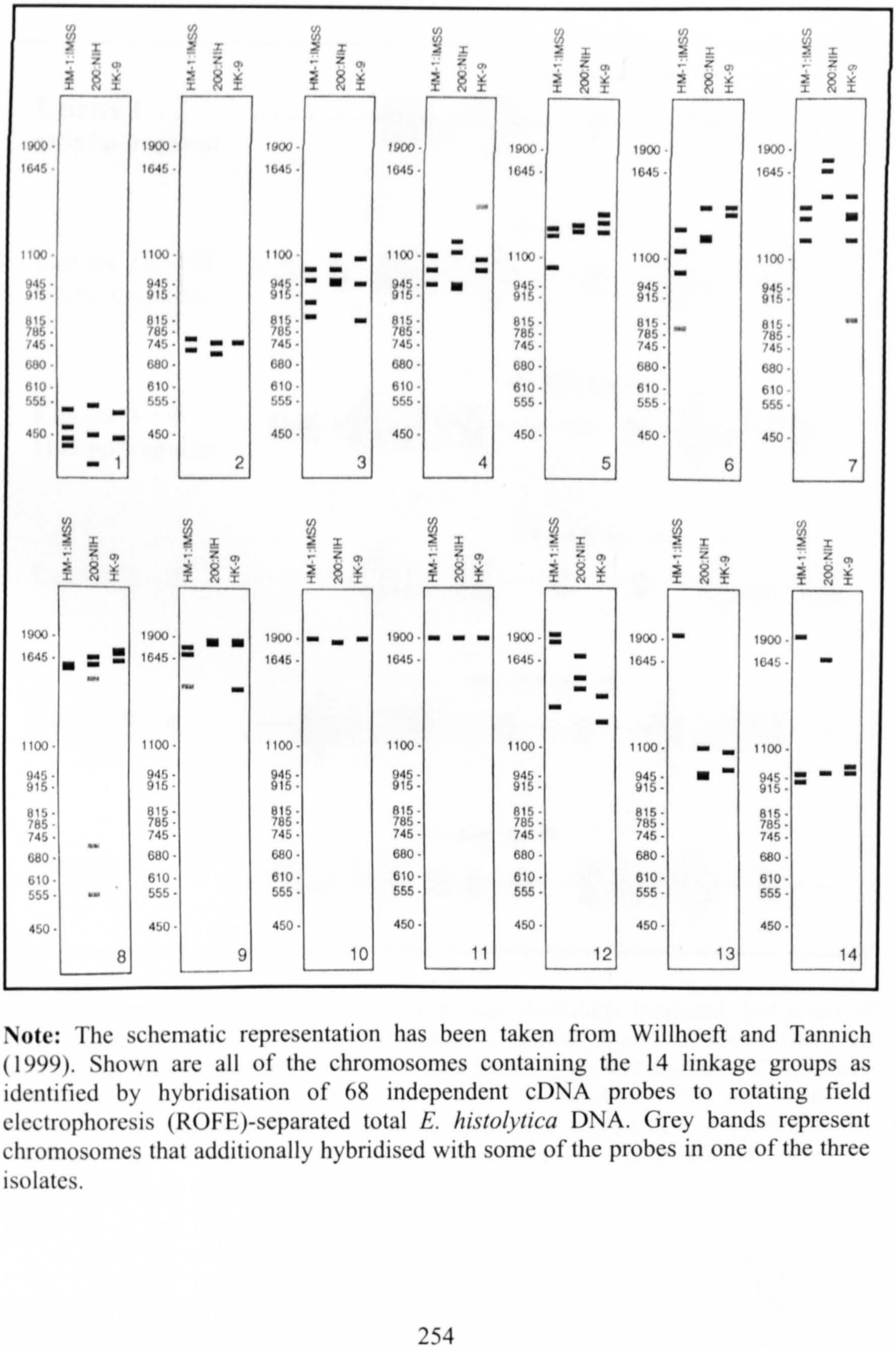
Family Unit	Isolate #	Relationship	Dwelling
37	37.0C	IC	Suburb
39	39.0L	IC	Farm
	39.384C	Sister	Farm
	39.384KB	Sister	Farm
2596	2596	NA	IS
56	56.0L	IC	Suburb
63	63.0I	IC	Suburb
	63.0Ii	IC	Suburb
	63.0IA	IC	Suburb
	63.0JA	IC	Suburb
	63.0JB	IC	Suburb
	63.0K	IC	Suburb
	63.632K	NCA	Suburb
	63.635K	NCA	Suburb
	63.638L	NCA	Suburb
73	73.733L	NCA	IS
	73.737I	NCA	IS
	73.737IA	NCA	IS
	73.737IB	NCA	IS
	73.737L	NCA	IS
	73.738HA	NCA	IS
	73.738HAi	NCA	IS
	73.738I	NCA	IS
	73.738IB	NCA	IS
	73.738L	NCA	IS
	73.740I	NCA	IS
	73.740Ii	NCA	IS
85	85.857H	NCA	Farm
	85.858H	NCA	Farm
	85.859G	NCA	Farm
	85.859H	NCA	Farm
87	87.0H	IC	Farm
89	89.892I	Son	IS
90	90.901HA	NCA	Farm
	90.901HAi	NCA	Farm
	90.901HAii	NCA	Farm
	90.901K	NCA	Farm
	90.901KA	NCA	Farm
	90.903G	Wife	Farm
	90.903I	Wife	Farm
	90.903K	Wife	Farm
	90.906G	Son	Farm
91	91.913G	NCA	Farm
	91.913I	NCA	Farm

100	100.0F	IC	IS
	100.0Fi	IC	IC
	100.1006G	NCA	IS

Abbreviations; IC: Index Case (cured of amoebic liver abscess (ALA), now asymptomatic); NCA: Neighbour or Close Associate; IS: Informal settlement.

APPENDIX XV

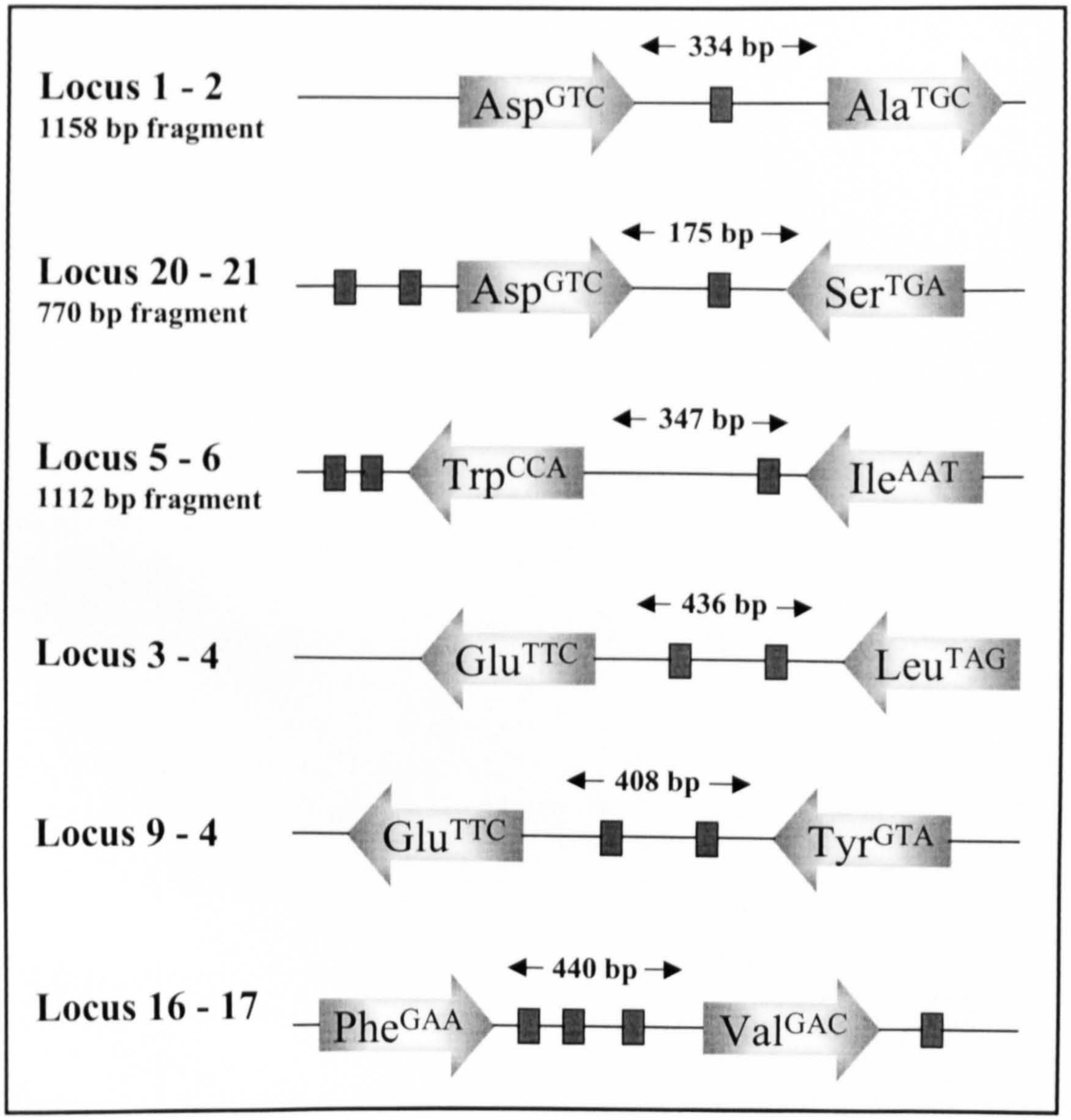
E. histolytica ELECTROPHORETIC KARYOTYPE



Note: The schematic representation has been taken from Willhoeft and Tannich (1999). Shown are all of the chromosomes containing the 14 linkage groups as identified by hybridisation of 68 independent cDNA probes to rotating field electrophoresis (ROFE)-separated total *E. histolytica* DNA. Grey bands represent chromosomes that additionally hybridised with some of the probes in one of the three isolates.

APPENDIX XVI

SCHEMATIC REPRESENTATION OF tRNA GENES & ASSOCIATED REPEATS



Note: The encoded tRNAs are shown with their anticodon indicated. Locations of major blocks of tandem repeats in the inter-tRNA spacer regions are shown (only the major repeat blocks are represented and the different types and numbers of individual repeat units are not shown). At loci 1-2 and 20-21 (section 7.2.4) and locus 5-6 (section 7.2.3) the information is based on the complete sequences as suggested by the TIGR database searches. Lengths of the inter- tRNA spacer regions are indicated.

SUPPLEMENT

PHYLOGENETIC ANALYSIS

Results of surveys carried out using the multiple polymorphic loci described in Chapter 3 revealed that *E. histolytica* is genetically highly variable (Chapters 4 & 6). All the isolates studied in Chapter 6 come from the same geographical area i.e. South Africa, and were asymptomatic at the time of sample collection (Appendix XII). On the other hand the isolates studied in the global survey (Chapter 4) were from geographically diverse populations and for the most part their hosts have been diagnosed with one or other form of invasive clinical disease (Appendix X). Despite the relatively large numbers of samples used and due in part to the limitations of the sampled populations (discussed in Chapter 8) no obvious correlation could be established between *E. histolytica* genotypes and clinical disease. However, it remains to be established whether any correlation exists between the observed polymorphism and the geographic distribution of the isolates. We have attempted to test this hypothesis using phylogenetic analysis and the results are presented here.

A total of eighty-four *E. histolytica* samples were included in the global survey (section 4.2; Appendix X). Twenty of these (Table 6G) represented duplicate samples with one sample containing DNA extracted directly from stool (suffix-F) while the other had culture-derived DNA (suffix-C) of the same infection. In all cases the PCR product sizes obtained from the same infection were identical (Table 8; Fig 18). Hence, for the phylogenetic analysis, data of only one of each pair (suffix-C) was included. Two of the Vietnamese isolates i.e. samples 82-4 and 84-3 (Table 6F), which failed to amplify at a majority of the loci (Table 8) were also removed from the present analysis. Similarly, five of the forty-one South African DNA samples (section 6.3.1; Table 12) were available in duplicate or triplicate (those identified by suffixes i and ii) and these too were removed from the data set. Thus the total number of isolates used for this analysis is seventy-two from the global survey and thirty-six from the South African study.

The presence or absence of a given band was scored using a discrete two-character state i.e. 0 (absent) and 1 (present) (data not shown). As the half loci 3-8 & 7-4, 9-11 & 10-4 and 16-19 & 18-17 present two halves of the complete loci 3-4, 9-4 and 16-17 respectively, only the data resulting from half loci were used in the phylogenetic analysis. Furthermore, since products at both loci 3-4 and 9-4 and their respective half loci appear for the most part to be very similar (Fig. 9 & Fig. 10) and as it has been suggested that these loci may in fact be members of the same family of repeated DNA (section 3.4) only locus 3-8 and 7-4 data were used in this analysis. Locus 1-2 and 5-6 being discrete and independent (section 3.3.2) were both included. In the final analysis therefore, data from six loci i.e. 1-2, 5-6, 3-8, 7-4, 16-19 and 18-17 were used for this study.

Genetic distances were estimated using the SYN-TAX-pc DOS version 5.0 package (Podani, 1993) which produces distance matrices based on the Jaccard index [1-(a/(a+b+c))]. Phylogenetic trees were then built from the distance matrices using the Neighbour-Joining method (Neighbour programme of PHYLIP package version 3.573C) (Felsenstein, 1993). Trees, both Phylograms and Cladograms, were displayed using TreeView version 1.6.2 (Page, 1996). Separate as well as pooled distance matrices and trees were produced for the South African and the global survey data sets.

Fig. 44 depicts the phylogram resulting from analysis of the thirty-six South African samples included in this analysis. Three samples (cluster 4) failed to amplify at one or more of the six loci included in this study. Data matrices both with and without these three samples were used to generate trees and it was found that the branching order of the other isolates did not alter (data not shown). Hence all three samples were included in the final analysis and the results presented herein. From these results it appears that for the most part the major clusters seen here (Fig. 44; cluster 1, 3, 5-7) are the same as the family groups identified by eye (section 6.3.1). Why some samples which are identical in the binary data (data not shown and Table 12) are not the same in the tree (Fig. 44; "*"cluster 3, 5 & 7) is unclear. However, with the exception of cluster 2 where two unrelated isolates are branching reasonably closely, there is no evidence of any significant relationships among the remaining samples as the internode distances where clusters are joined are very short.

As with the South African samples, results with the seventy-two global isolates (Fig. 45) also revealed clustering of Japanese isolates from two mass infections (cluster 1 & 3) which is in accord with the results previously seen by eye (Table 8; Fig. 16). Likewise, our previous results showing shared patterns between two samples which represent members of the same family (Table 8; Fig. 18) are upheld in the present analysis (Fig. 45; cluster 6). Interestingly, a number of Japanese isolates cluster together with the Korean isolate YS-27 (Fig. 45; cluster 2) even though there is no known epidemiological link among them. Similarly, all the Vietnamese isolates (Fig. 45; cluster 4) as well as a majority of the Bangladeshi isolates (Fig. 45; cluster 8) group together. These results may not be entirely surprising given that both these sample groups come from relatively restricted geographical localities. However, exceptions are also seen where one or more samples are scattered among others outside the geographic clusters e.g. Japanese isolates J6 & J8 and Bangladeshi isolates 8691, 1320300 & 1353600 (marked with asterisks in Fig. 45).

Surprisingly all ten samples provided by the University of Leiden form a close cluster together with one of the two Venezuelan isolates (IULA:1092:1; Fig. 45; cluster 5). Although all ten samples were obtained from the same source (Table 6G) they originate from different parts of the world and between them represent three different continents. Hence clustering on a scale similar to that seen for some of the other geographical clusters was not expected. This raises doubts as to the degree of credence that can be given to the geographical clusters mentioned above.

It must be pointed out that as originally seen and reported (Table 8; section 4.3.2) isolates 200:NIH, HK-9 and H-303:NIH are identical (Fig. 45; cluster 7). This result is highly suspicious in view of the high degree of diversity seen here and the fact that no other isolates from different countries were found to be identical. It is therefore possible that these three isolates have been misidentified and are in fact the same.

Trees based on distance matrices produced with the isolates from both the South African and global surveys were also generated (data not shown). All the major South African family clusters (Fig. 44; cluster 1, 3, 5-7) remain intact and even some of the deeper clusters do so. The single samples and many family clusters are interspersed among samples from other parts of the world, but no close relationships were

identified. Importantly, the South African samples did not form a geographical cluster on their own.

Taken together these results suggest that there is more similarity among samples from restricted geographical areas compared to random samples. However the link is not particularly strong and as such we can not predict, with any confidence, the origin of a given sample based on its polymorphic pattern alone. We have not been able to do a bootstrap or similar analysis to test the robustness of the clusters identified. However, it may be mentioned that when analysis was performed with different combinations of samples the main clusters remained intact.

REFERENCES

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.572c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences. **12**: 357-358.

Podani. 1993. Syn-Tax-Pc Computer programs for Multivariate Data Analysis in Ecology and Systematics. Budapest, Sciencia Publishing.

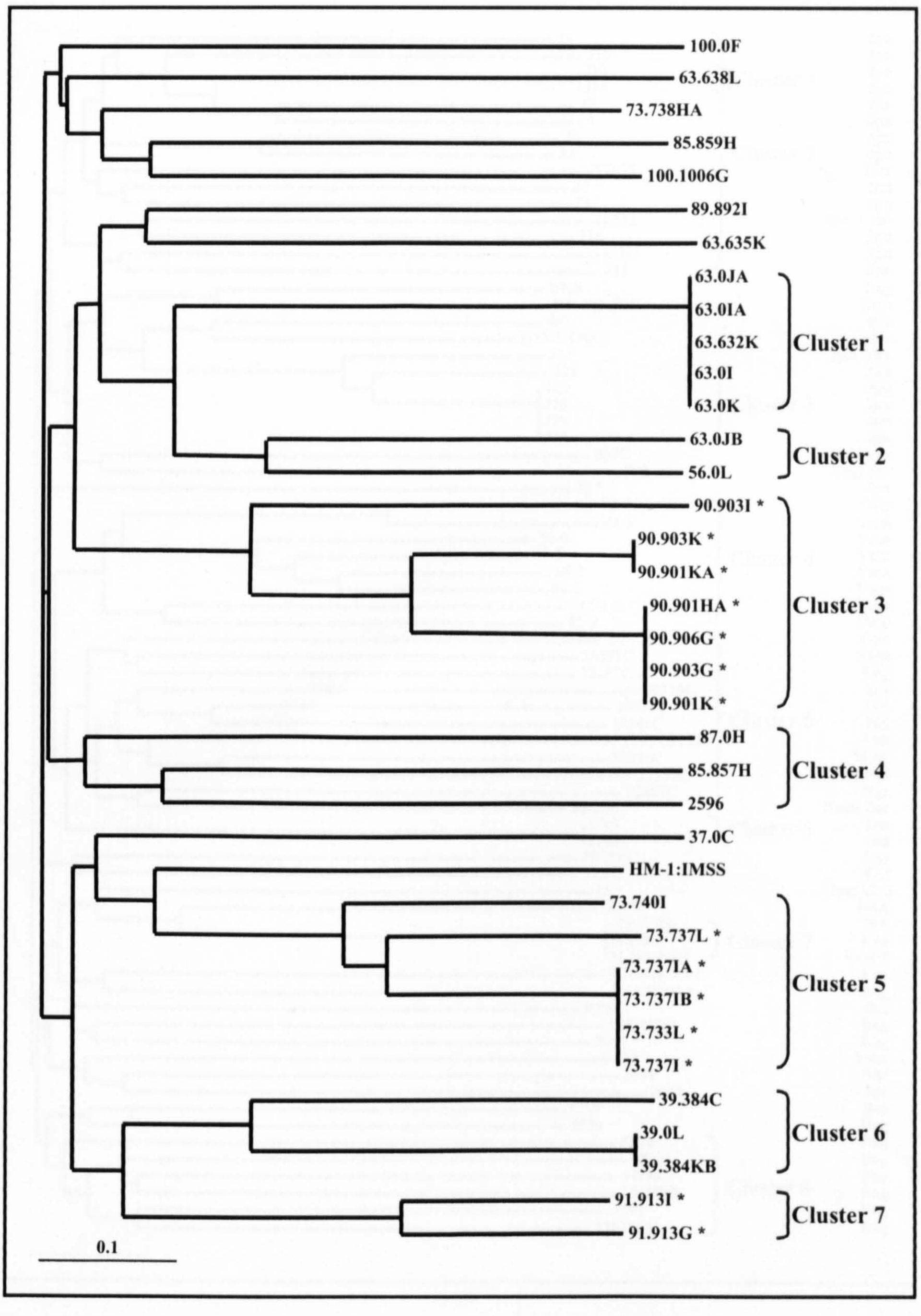


Fig. 44. Phylogenetic Analysis of South African *E. histolytica* Isolates. Neighbour-joining (phylogram) unrooted tree (PHYLIP) using Jaccard distances. Clusters mentioned in the text are highlighted. Specific samples mentioned in the text are marked with asterisks.

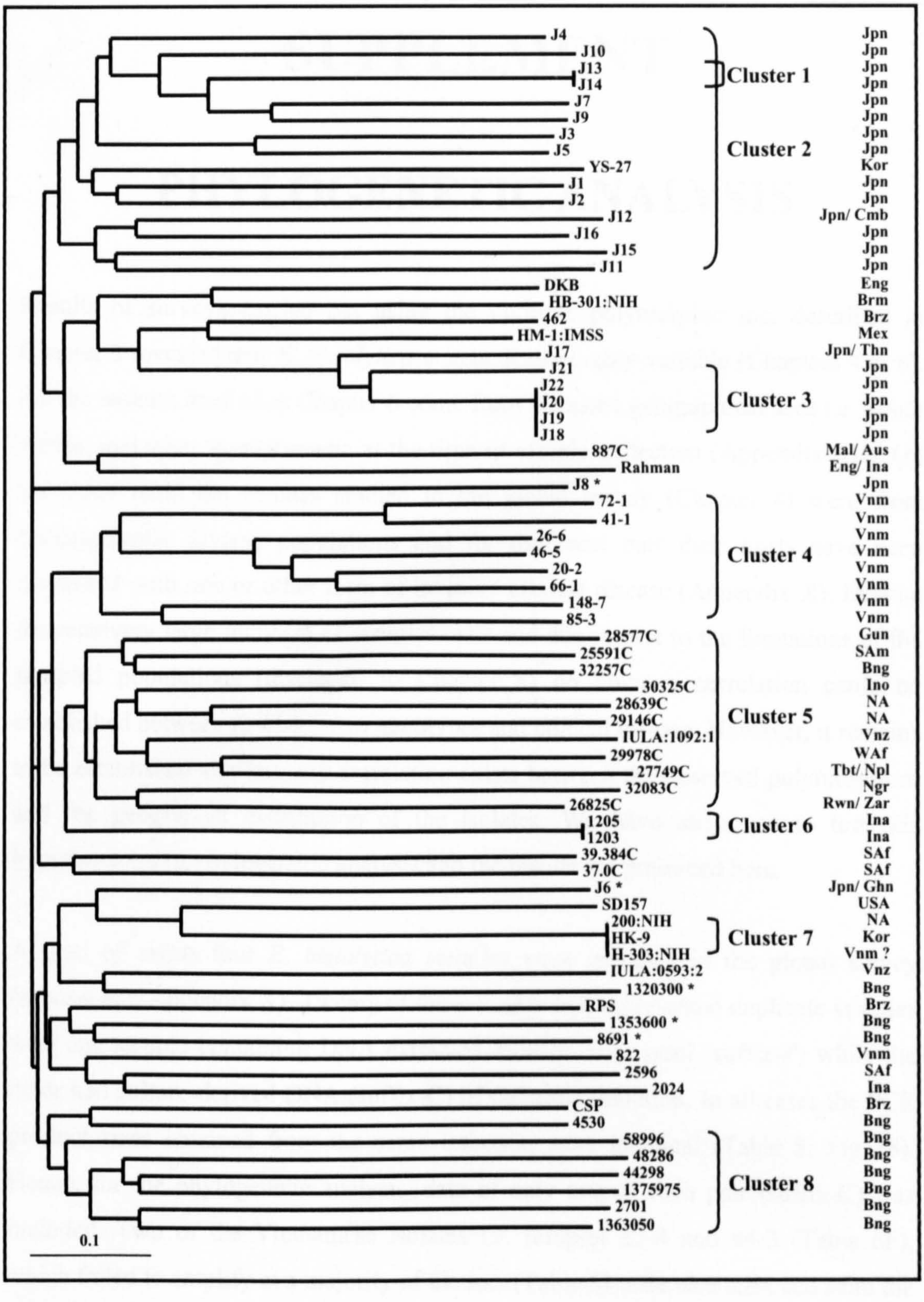


Fig. 45. Phylogenetic Analysis of Global *E. histolytica* Isolates. Neighbour-joining (phylogram) unrooted tree (PHYLIP) using Jaccard distances. Clusters mentioned in the text are highlighted. Specific samples mentioned in the text are marked with asterisks. The geographical origin of each isolate is given as a three-letter code. Abbreviations: **Aus** (Australia); **Bng** (Bangladesh); **Brm** (Burma); **Brz** (Brazil); **Cmb** (Cambodia); **Eng** (England); **Ghn** (Ghana); **Gun** (Guinea); **Ina** (India); **Ino** (Indonesia); **Jpn** (Japan); **Kor** (Korea); **Mal** (Malaysia); **Mex** (Mexico); **NA** (Not Available): **Ngr** (Nigeria); **Npl** (Nepal); **Rwn** (Rwanda); **SAf** (South Africa); **SAm** (South America); **Tbt** (Tibet); **Thn** (Thailand); **USA** (United States of America): **Vnm** (Vietnam); **Vnz** (Venezuela); **WAf** (West Africa); **Zar** (Zaire).

JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 2001, p. 897–905 Vol. 39, No. 3
0095-1137/01/$04.00+0 DOI: 10.1128/JCM.39.3.897–905.2001

Isolation and Characterization of Polymorphic DNA from *Entamoeba histolytica*

MEHREEN ZAKI AND C. GRAHAM CLARK*

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

Received 7 July 2000/Returned for modification 13 November 2000/Accepted 20 December 2000

An important gap in our understanding of the epidemiology of amebiasis is what determines the outcome of *Entamoeba histolytica* infections. To investigate the possible existence of invasive and noninvasive strains as one factor, the ability to differentiate individual isolates of *E. histolytica* is necessary. Two new loci containing internal repeats, locus 1-2 and locus 5-6, have been isolated. Each contains a single repeat block with two types of related direct repeats arranged in tandem. Southern blot analysis suggests that both loci are multicopy and may themselves be arranged in tandem arrays. Three other previously reported, internally repetitive loci containing at least two repeat blocks each with one or more related repeat units were also investigated. PCR was used to study polymorphism at each of these loci, which was detected to various degrees in each case. Variation was seen in the total number of bands obtained per isolate and their sizes. Nucleotide sequence comparison of loci 1-2 and 5-6 in five axenic isolates revealed differences in the number of repeat units, which correlated with the observed PCR product size variation, and in repeat sequence. Use of multiple loci collectively allowed differentiation of a majority of the 13 isolates studied, and we believe that these loci have the potential to be used as polymorphic molecular markers for investigating the epidemiology of *E. histolytica* and the potential existence of genetically distinct invasive and noninvasive strains.

The acceptance of *Entamoeba histolytica* and *Entamoeba dispar* as distinct species (2, 11) has had a major impact on our views of amebiasis, in particular its clinical management and epidemiology. It is likely that at least 90% of the infections previously ascribed to *E. histolytica* are actually *E. dispar*, while only the remaining 10% are infected with *E. histolytica* in its new sense. However, it also appears that many *E. histolytica* infections never progress to become symptomatic and are spontaneously lost. This observation raises some important questions. Are the organisms that produce invasive, symptomatic disease genetically distinct from those that give rise to asymptomatic infections? Or do all *E. histolytica* isolates have the potential to become invasive? Do certain invasive isolates show tropism for specific organs, with some preferentially ending up in the intestinal wall while others reach extraintestinal sites? To address the possibility of a relationship between parasite variation and infection outcome the ability to differentiate isolates of *E. histolytica* is necessary.

Our present knowledge of intraspecies variation in *E. histolytica* is limited. Isoenzyme analysis provided the first markers (25), but it now appears that isoenzyme patterns are not fixed (5) and therefore that many 'zymodeme' assignments are unreliable (16). A limited number of DNA markers have been shown to exhibit intraspecies diversity. Variation has been observed in the number of rRNA transcription units present on the extrachromosomal ribosomal DNA circles; only one rRNA gene copy has been seen in some strains, while the majority have two (27). Variation has also been detected in the non-coding families of short tandem repeats found both upstream and downstream of the rRNA genes (20, 21, 26). However, variability in the occurrence and instability in the length of some of these sequences limits their use for isolate identification (4). PCR amplification of the Strain-Specific Gene (6) or *Tr* (27), which is present upstream of one rRNA transcription unit and contains tandemly repeated internal elements, has revealed considerable variation in the number of repeats among strains of *E. histolytica* (8). However, the complete absence of this locus in certain strains (27) makes it a poor candidate for intraspecies typing.

At present, the most polymorphic gene of *E. histolytica* is that encoding the serine-rich *E. histolytica* protein (SREHP or K2) a surface antigen with tandem 8- and 12-amino-acid repeats (17, 28). Repeat number, sequence, and restriction enzyme site polymorphisms have been reported among different *E. histolytica* isolates (8, 14). However, more than one-third of the isolates tested gave the same restriction fragment pattern (8). The chitinase gene also encodes tandem repeats of a degenerate 7-amino-acid sequence (10), and a report on the use of chitinase repeat polymorphisms to distinguish isolates of *E. histolytica* has been published recently (14). However, there still exists a need for additional reliable polymorphic *E. histolytica* DNA markers.

The use of microsatellite locus analysis has gained considerable popularity as a tool for detecting intra- and interspecies variations in a number of organisms, including protozoan parasites such as *Trypanosoma* (22), *Leishmania* (24), and *Plasmodium* (1) spp. Using a method designed to isolate microsatellite loci, we have obtained two new polymorphic DNAs containing tandemly repeated sequences from *E. histolytica*. We present here the preliminary characterization of the two loci and the interstrain variations they display. In addition, three other loci showing the presence of tandemly repeated

* Corresponding author. Mailing address: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom. Phone: 44-207-927-2351. Fax: 44-207-636-8739. E-mail: graham.clark@lshtm.ac.uk.

TABLE 1. Origin of *E. histolytica* isolates

Isolate	Origin	Culture type	Clinical information
HM-1:IMSS clone 9	Mexico	Axenic	Isolated from rectal ulcer; patient with dysentery
200:NIH	(uncertain)	Axenic	Patient with amebic dysentery
H-303:NIH	VietNam (?)	Axenic	Patient with amebic empyema and dysentery
IULA:0593:2	Venezuela	Axenic	Patient with amebic colitis
IULA:1092:1	Venezuela	Axenic	Patient with amebic colitis
8691	Bangladesh	Xenic	Patient with amebic colitis
4530	Bangladesh	Xenic	Patient with amebic colitis
1320300	Bangladesh	Xenic	Patient with amebic colitis
48286	Bangladesh	Xenic	Patient with amebic colitis
2596	South Africa	Xenic	Patient asymptomatic; serology positive by antigen gel diffusion test
26.253C	South Africa	Xenic	Patient asymptomatic; serology positive by antigen gel diffusion test
37.0C	South Africa	Xenic	Convalescent amebic liver abscess patient; serology positive by antigen gel diffusion test
39.384C	South Africa	Xenic	Patient asymptomatic; serology positive by antigen gel diffusion test

sequences have been studied for their potential as polymorphic markers for use in investigating the molecular epidemiology of *E. histolytica*.

MATERIALS AND METHODS

***E. histolytica* isolates.** Except for HM-1:IMSS clone 9, the axenic isolates were provided by John Ackers (London School of Hygiene and Tropical Medicine) (Table 1). All axenic isolates were cultured in the casein-free medium YI-S (12) with 15% heat-inactivated adult bovine serum (Sigma-Aldrich).

Xenic isolates were obtained from two sources (Table 1). Four samples were from Rashidul Haque of the International Centre for Diarrhoeal Disease Research, Bangladesh, via Aura Aguirre (London School of Hygiene and Tropical Medicine), while four others were provided by Terry Jackson of the Medical Research Council of South Africa, Durban. The South African isolates were from a patient who had recovered from amebic liver abscess (39.0C) or close family contacts of such patients who were asymptomatic at the time of isolation. All xenic strains were originally isolated in Robinson's medium (23); there is no evidence that culture conditions or media have any effect on the markers studied.

Isolation of nucleic acids. DNA was isolated as previously described (7, 9), dissolved in 10 mM Tris-Cl (pH 8.5) and passed over a Microspin S-200 HR column (Amersham Pharmacia Biotech, Inc). RNA was removed by the addition of RNase A (Promega) to 0.05 μg ml^{-1}.

Isolation of repeated DNA containing sequences. A nonradioactive method designed for rapid isolation of microsatellite sequences (13, 22) was adapted. Genomic DNA of isolate HM-1:IMSS (ca. 500 ng) was digested for 2 h with a restriction enzyme, either *Alu*I or *Rsa*I (10 U/20-μl reaction) (Gibco-BRL), followed by incubation at 65°C for 15 min to inactivate the enzyme and passage through a S-200 column to remove the salts.

5′-Phosphorylated 24-mer (5′-pAGTCCGGATCCAAGCAAGAGCACA-3′) and nonphosphorylated 20-mer (5′-CTCTTGCTTGGATCCGGACT-3′) oligonucleotides with overlapping complementary sequences containing a *Bam*HI site were used to generate an adapter. Then, 2.5 pmol of adapter was ligated to approximately 250 ng of digested DNA with T4 DNA ligase at 14°C.

Ligated fragments (equivalent to ca. 50 ng of DNA) were annealed to 20 pmol of a biotinylated microsatellite oligonucleotide [GATGATCCGACGCAT$(CA)_{12}$, GATGATCCGACGCAT$(CT)_{12}$, $(CAA)_{12}$, $(CTT)_{12}$, $(CAT)_{12}$, $(CTA)_{12}$, or $(TAA)_{12}$] by denaturing at 95°C for 10 min and annealing at 60°C for 1 min; the hybrids were then bound to 100 μg of streptavidin-coated magnetic beads (Dynabeads KilobaseBinder kit; Dynal). Following incubation at room temperature for 3 h the Dynabead-DNA complexes were washed twice (10 mM Tris-Cl, pH 7.5; 1 mM EDTA; 2.0 M NaCl) and resuspended in 50 μl of TE buffer (10 mM Tris-Cl, pH 8.0; 1 mM EDTA; pH 8.0). The captured product was used as a template for PCR amplification using the 20-mer adapter oligonucleotide under standard conditions. Amplified products were analyzed on a 1.8% agarose gel (Gibco-BRL) using amplified adapter-ligated but unselected digested DNA as a control.

After electrophoresis, PCR products appearing to be enriched by the selection process were gel purified and cloned into the vector pGEM-T Easy (Promega). Recombinant plasmids were sequenced using an ABI PRISM 377 (Perkin-Elmer) and Thermo-Sequenase II dye terminator cycle sequencing kit (Amersham Pharmacia Biotech).

TABLE 2. Oligonucleotide primers

Primer	Primer sequence (5′ to 3′)
R1	CTG GTT AGT ATC TTC GCC TGT
R2	CTT ACA CCC CCA TTA ACA AT
R3	GCT ATG GTC GGT ATC GAT ATC
R4	CCT TAG GTC ACT GGT TCG AA
R5	CTA AAG CCC CCT TCT TCT AT
R5A	CTA AAG CCC CCT TCT TCT ATA ATT
R6	GTG CTA ATA ACG CCA GGG TC
R6A	CTC AGT CGG TAG AGC ATG GT
R7	CTT TAC TTC TCT TTT ACC ACG
R8	CGT GGT AAA AGA GAA GTA AAG
R9	CTA CAT CTA CAG TCC TCC GCT
R10	CTT ACT TCT CTT TAC CAC GAC
R11	GTC GTG GTA AAG AGA AGT AAG
R16	AAG CTT CCT TAG CTC AGC TG
R17	TAA AAG GGG AAA GAA TAG GAA
R18	GGT TTC ATG GTG TAG TTG GT
R19	ACC AAC TAC ACC ATG AAA CC

PCR product size polymorphisms and nucleotide sequence comparison. Primers were designed from repeat flanking region sequences of all the loci. The genomic DNA of *E. histolytica* was amplified using the primers listed in Table 2 and 30 cycles of 1 min at 94°C, 1 min at the primer-dependent annealing temperature, and 2 min at 72°C, with a final extension of 5 min at 72°C. Amplified products were analyzed using 2.4% NuSieve 3:1 agarose gels (FMC) in 1× Tris-boric acid-EDTA buffer (TBE).

PCR products from all five axenic isolates of *E. histolytica* were cloned pGEM-T Easy vector (Promega) and sequenced as described above.

Southern blot analysis. Genomic DNA of isolate HM-1:IMSS clone 9 was digested overnight with 10 U each of restriction enzymes *Alu*I, *Dde*I, *Dra*I, *Eco*RI, and *Rsa*I (Gibco-BRL or MBI Fermentas), and fragments were separated by electrophoresis using 0.8% agarose gels in 1× TBE buffer and transferred to BiodyneA membranes (Gibco-BRL) using standard methods. [α-^{32}P]dCTP-labeled double-stranded DNA probes were prepared by using the Rediprime II random prime labeling system (Amersham Pharmacia Biotech). Filters were hybridized overnight at 65°C in a solution of 1 M NaCl–1% sodium dodecyl sulfate–10% dextran sulfate and then washed to a final stringency of 0.2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate at 65°C before autoradiography at −70°C.

The nucleotide sequence data reported here have been submitted to the GenBank database under accession numbers AF276055 to AF276065.

RESULTS

Isolation of repeated DNA containing sequences from the *E. histolytica* genome. To try and obtain DNA fragments containing microsatellites, we employed a nonradioactive method

based on affinity capture of single-stranded restriction fragments annealed to biotinylated microsatellite oligonucleotides, with attachment to streptavidin-coated magnetic beads (Dynal), followed by adapter-mediated PCR (13, 22). A total of twenty two PCR fragments ranging in size from 250 to 700 bp were gel purified from the total amplification products of *Alu*I or *Rsa*I restriction fragments annealed to one of seven biotinylated oligonucleotides. These fragments were chosen on the basis of their apparent enrichment compared to control amplification products and were cloned, sequenced, and examined for the presence of microsatellites.

No products contained sequences corresponding to the microsatellite oligonucleotides used in their capture. Furthermore, the majority of the sequences did not reveal any tandemly repeated DNAs. However, two clones (clone 1 and clone 4) derived from an approximately 480-bp fragment, obtained from *Alu*I restriction fragments annealed to the $(CTT)_{12}$ oligonucleotide, showed the presence of internal tandem repeats. The repeats seen in clones 1 and 4 were distinct. Two other clones (clone 1′ and clone 5) derived from *Alu*I restriction fragments annealed to $(TAA)_{12}$ contained the same type of repeats as clone 4. Clones 1′ and 5 contained fragments of approximately 480 and 450 bp, respectively. Further analysis was carried out on clone 1, which represents locus 5-6, and clone 4, which represents locus 1-2.

Characterization of loci 1-2 and 5-6. The complete sequence of the locus 1-2 clone (Fig. 1A), not including the adapter sequence, is 402 bp long and contains a single repeat block with two related direct repeats arranged in tandem (Fig. 1B). In addition to the major repeat block, tandem duplications of 8 to 12 bp are present in the flanking regions. The complete sequence of the locus 5-6 clone (Fig. 2A), not including the adapter sequence, is 424 bp long and contains a single repeat block (Fig. 2B). As in locus 1-2, other tandem duplications in the regions flanking the repeat block are also evident. BLAST search results revealed no identity of either locus 1-2 or locus 5-6 to any previously reported *E. histolytica* sequences.

PCR product size polymorphisms at loci 1-2 and 5-6. Primers were designed in the regions flanking the repetitive blocks for both locus 1-2 and locus 5-6, and the PCR amplification products were analyzed on 2.4% NuSieve agarose gels to look for fragment size polymorphism among the 13 *E. histolytica* isolates (Fig. 3).

Amplification of locus 1-2 gave the expected product of ca. 400 bp in isolate HM-1:IMSS clone 9 (Fig. 3A). All of the *E. histolytica* isolates gave a single major product. The four South African isolates gave the most variable patterns.

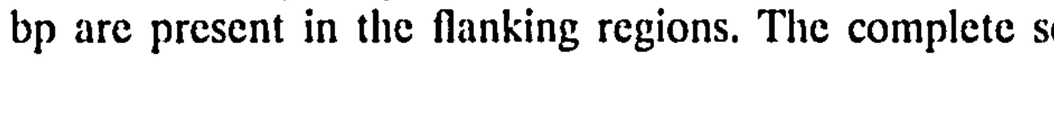
Amplification of locus 5-6 gave the expected product of ca. 420 bp in isolate HM-1:IMSS clone 9 (Fig. 3B), but two additional bands of ca. 480 and 520 bp were also seen. This locus is highly polymorphic. Variation is seen in the total number of bands per isolate and their sizes even within the same geographic area.

Nucleotide sequence analysis and characterization of the observed size polymorphism. In order to study the underlying nature of the observed size polymorphisms, the amplification

A

```
        10         20         30         40         50         60
CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCGTTGAAGAGTT
        70         80         90        100        110        120
CTCTTTTTATACTTTTATATGTTTATATGTTTATATCCTTATTTATTATTCTTTTATATT
       130        140        150        160        170        180
CTTATCACTTCCTACTACTCTTATTTATTATCCTTATTATATCTATTCTTACTCCCTATC
       190        200        210        220        230        240
TTTATTATCTTTATTATCTTTATTATCTTTATTATCTTTATTATCTTTATTACCTTTATT
       250        260        270        280        290        300
ACCTTTATTACCTTTATTACCTTTATTATCTTTATTACCTTTATTATATCTATTCTCACT
       310        320        330        340        350        360
TCCTATACGTACTCTTTTTACTACTCTTTTTACTACTCTTCTTACTATACCTCTTACTAC
       370        380        390        400
TCCTACTTTCACCTCCCTCTTTATTGTTAATGGGGGTGTAAG
```

B



■ CTTTATTAT

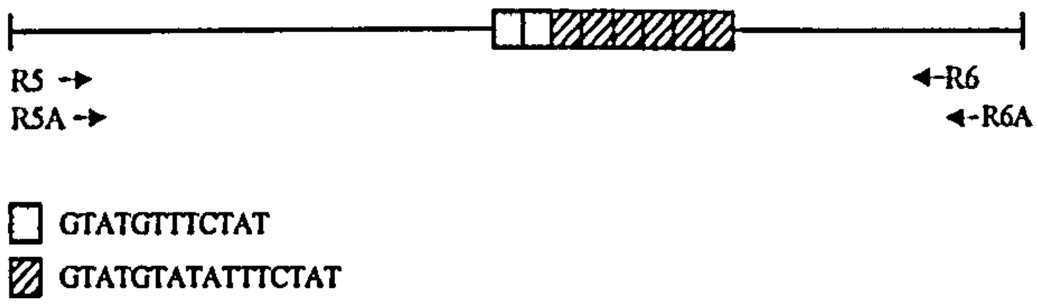
▨ CTTTATTAC

FIG. 1. Locus 1-2. (A) Nucleotide sequence. The main block of internal tandem repeats is in boldface. Underlined regions indicate one of the two types of repeat units. (B) Schematic representation. The two types of internal tandem repeats and their arrangement with respect to each other are shown. Tandem duplications in the flanking regions are not shown. The positions of the amplification primers are indicated.

```
        10         20         30         40         50         60
CTAAAGCCCCCTTCTTCTATAATTTATATATTATTCTCTTTGAGACTTATTTCTACTTTA
        70         80         90        100        110        120
TTTCTTATACTTATTATATCTACTTTAGATACTTTACTTCACTTTATATATTCTACTTTA
       130        140        150        160        170        180
TATTCCTGACTTATATCTTTTATGTTTATATGTTTATATGATTTTATGATTTTATGTTTA
       190        200        210        220        230        240
TATTCTTCCTATTTATACCACTATGTATATATATATGTATGTTTCTATGTATGTTTCTAT
       250        260        270        280        290        300
GTATGTATATTTCTATGTATGTATATTTCTATGTATGTATATTTCTATGTATGTATATTT
       310        320        330        340        350        360
CTATGTATGTATATTTCTATGTATGTATATTTCTATGTACGTCTTTAACTTTAAAAACAA
       370        380        390        400        410        420
TGCTCTCAGCAGGTTTCGAACCCTGCGACCCTGGCGTTATTAGCACCATGCTCTACCGAC

TGAG
```

B



□ GTATGTTTCTAT

▨ GTATGTATATTTCTAT

FIG. 2. Locus 5-6. (A) Nucleotide sequence. The main block of internal tandem repeats is in boldface. Underlined regions indicate one of the two types of repeat units. (B) Schematic representation. The two types of internal tandem repeats and their arrangement with respect to each other is shown. Tandem duplications in the flanking regions are not shown. The positions of the amplification primers are indicated. Two primer pairs were designed for locus 5-6 (Table 2). Amplification products generated by primers R5 and R6 were cloned, sequenced, and aligned for intrastrain nucleotide sequence comparisons (Fig. 4B), while the primer pair R5A-R6A was used for studying interstrain PCR product size polymorphisms (Fig. 3B).

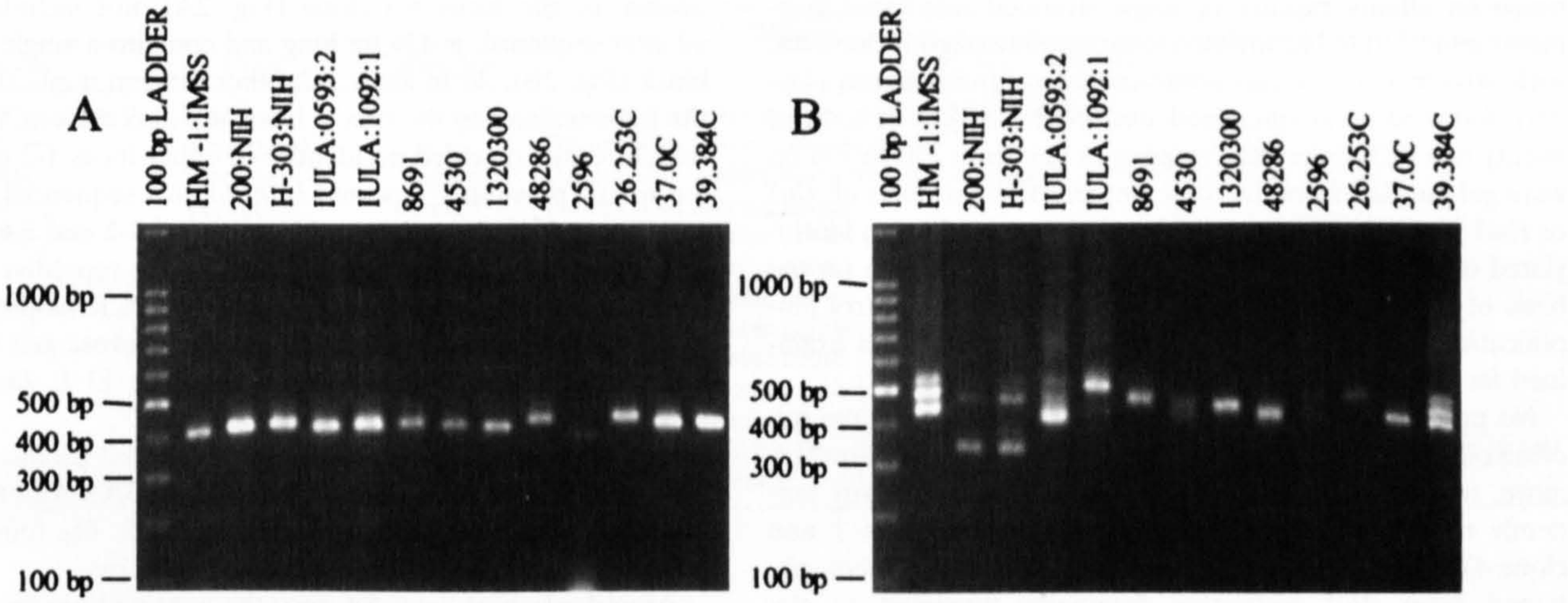


FIG. 3. Polymorphic DNA analysis of *Entamoeba histolytica* isolates. (A) Locus 1-2. Amplification products were generated using primers R1 and R2 at an annealing temperature of 53°C. (B) Locus 5-6. Amplification products were generated using primers R5A and R6A at an annealing temperature of 56°C. Isolate origins: HM-1:IMSS (Mexico); 200:NIH (uncertain); H-303:NIH (VietNam); IULA:1092:1 and IULA:0593:2 (Venezuela); 8691, 4530, 1320300, and 48286 (Bangladesh); 2596, 26.253, 37.0C, and 39.384C (South Africa).

products of all five axenic isolates at locus 1-2 and locus 5-6 were cloned and sequenced. This analysis revealed differences in the number and sequence of the repeat units, as well as sequence variation in the regions flanking the repeat blocks (Fig. 4).

There was very little variation in the total number of repeat units among the five samples at locus 1-2 (Fig. 4A). This is consistent with the slight differences in PCR product size observed in Fig. 3A. However, considerable variation existed between the isolates in the relative numbers of repeat units of type 1 versus type 2. In contrast to locus 1-2, there was considerable variation in the total number of units of one repeat type among the five strains at locus 5-6 (Fig. 4B). This high degree of variation is also reflected in the PCR product size comparison (Fig. 3B).

The locus 5-6 amplification products from isolate H-303:NIH revealed two distinct fragments of ca. 320 and 450 bp, respectively. Cloning of the PCR products from this locus also resulted in two inserts which differed in size by ca. 100 bp [designated H-303:NIH-(3) and H-303:NIH-(4); Fig. 4B].

Shared single-base alterations within the repeat blocks were seen at two positions in isolates 200:NIH and H-303:NIH at locus 1-2 (Fig. 4A). Another two single base changes were seen only within the repeat block of isolate IULA:1092:1 at locus 1-2, and this isolate was also missing a single copy of a 12-bp duplication in the 3′-flanking region. A single base change in the 5′-flanking region was present in isolate IULA:0593:2.

At locus 5-6 a single base change was seen within the repeat block (Fig. 4B) for isolate IULA:1092:1. Additionally, both isolate 200:NIH and H-303:NIH-(4) appeared to be missing the initial 10 bp of the first 12-bp repeat unit (GTATGTTTC TAT). A difference was also evident in the flanking regions of the repeat units in that isolate 200:NIH was missing a single copy of an 8-bp tandem duplication at the 5′ end. While it is possible that single nucleotide differences are PCR amplification artifacts, it is highly unlikely that shared nucleotide differences among isolates and repeats or repeat number variations could have this origin.

Southern blot analysis. Southern hybridization analysis was performed using HM-1:IMSS clone 9 genomic DNA digested with five enzymes (Fig. 5) and either locus 1-2 or locus 5-6 as specific probes. The specificity of the probes was ensured by

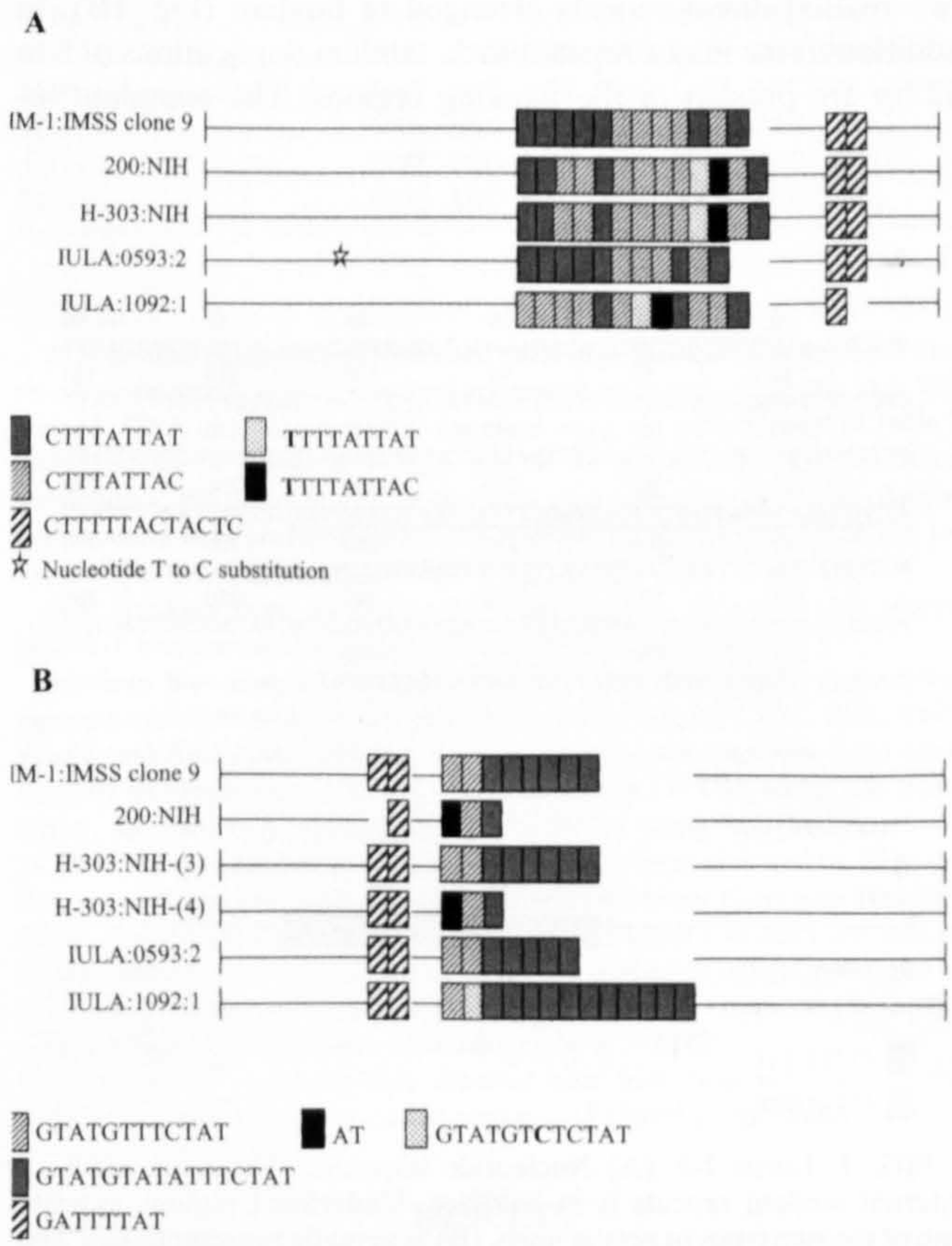


FIG. 4. Schematic representation of locus structure in five axenic isolates of *E. histolytica*. Variations in number, sequence, and arrangement of repeat units are shown. Gaps have been introduced to optimize alignment. (A) Locus 1-2. (B) Locus 5-6.

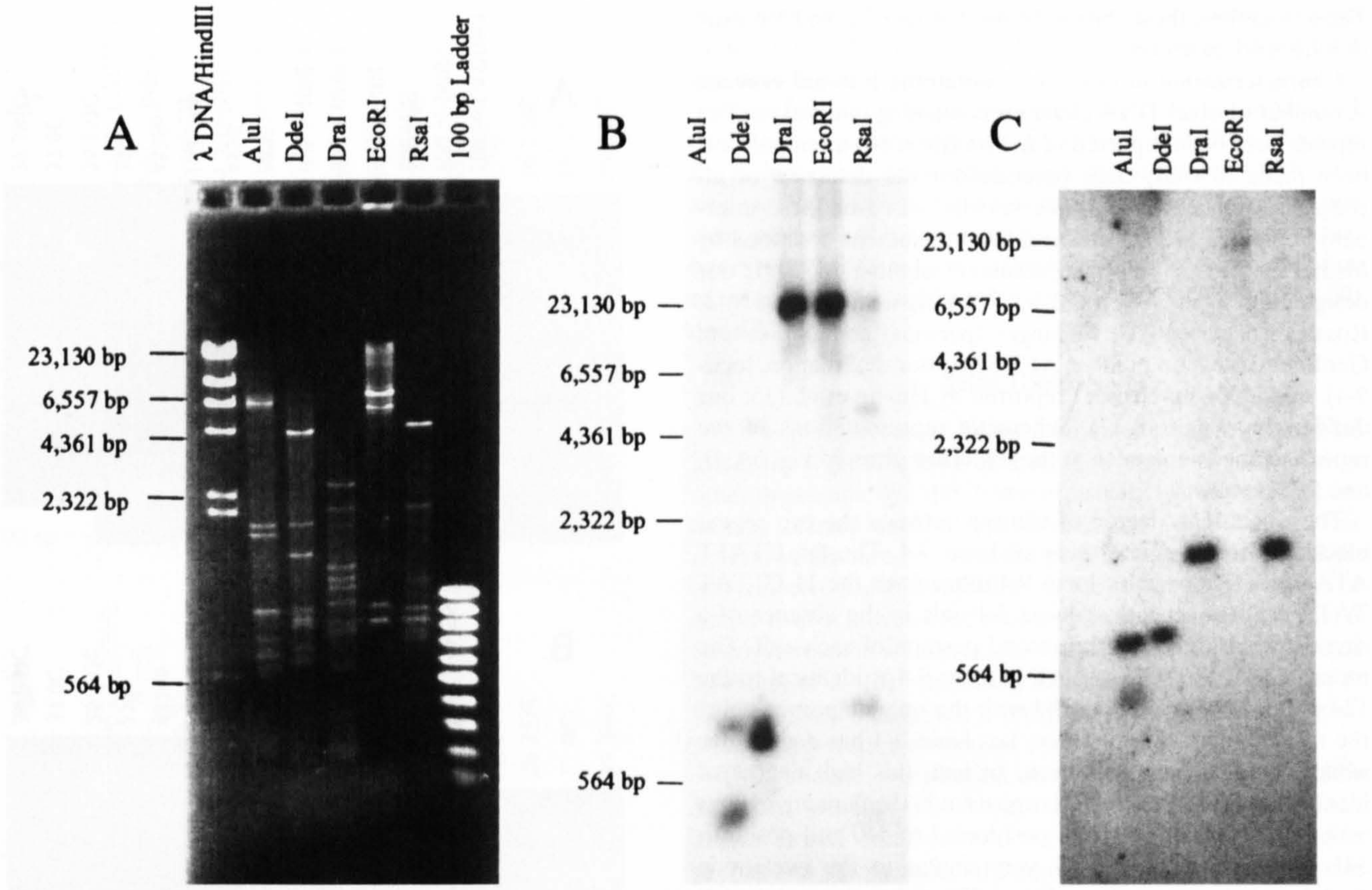


FIG. 5. Southern blot analysis. (A) Genomic DNA of *E. histolytica* isolate HM-1:IMSS digested with restriction enzymes and stained with ethidium bromide. (B) Blot of gel in panel A hybridized with a locus 1-2 specific ^{32}P-labeled probe. (C) Blot of gel in panel A hybridized with a locus 5-6 specific ^{32}P-labeled probe. Some of the faint bands seen in Fig. 5B may result from slight cross-hybridization to fragments of the abundant extrachromosomal circular DNA seen in the ethidium bromide-stained gel (Fig. 5A).

using PCR products produced from plasmid DNAs (clone 4 and clone 1, respectively).

With the locus 1-2 specific probe, the band of ca. 400 bp in the *Alu*I lane (Fig. 5B) was expected since the clone was obtained from an *Alu*I restriction fragment of about the same size. It was a surprise to find that the probe gave intense hybridization signals at ≥23 kb with the *Dra*I-digested DNA since this enzyme cuts frequently in *E. histolytica* DNA and usually produces much smaller fragments, as seen in Fig. 5A. It is notable that *Alu*I, *Dde*I, and *Rsa*I give major fragments of about the same size and that *Dra*I and *Eco*RI both give very large fragments.

Similarly, a band of ca. 400 bp was expected in the *Alu*I lane with the locus 5-6 specific probe (Fig. 5C), since the clone was obtained from an *Alu*I restriction fragment of about this size. Once again hybridization with this probe gave major fragments of about the same size with *Alu*I and *Dde*I and with *Dra*I and *Rsa*I. *Eco*RI again produces a very large fragment of ≥23 kb.

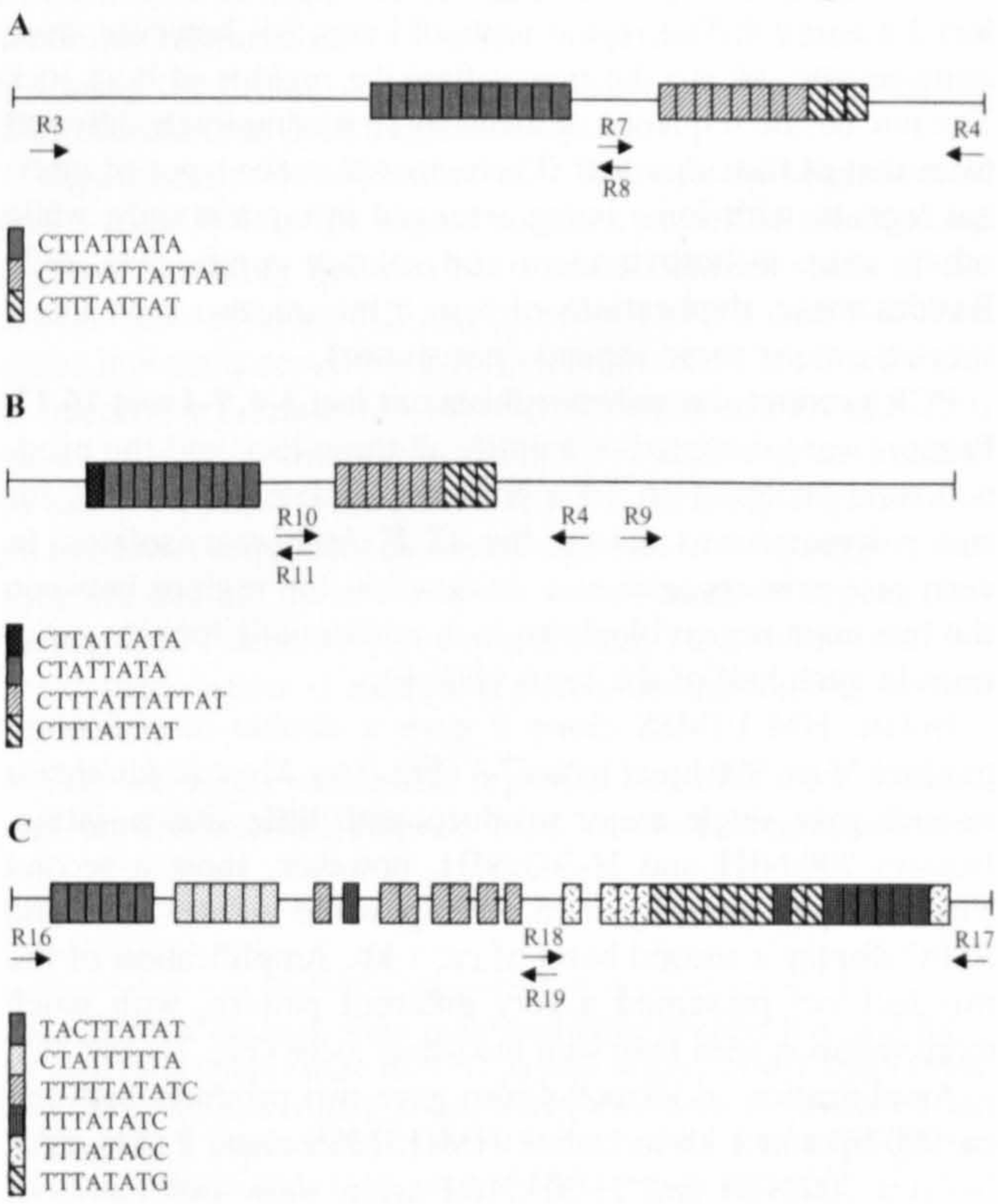


FIG. 6. Schematic representation of repeat arrangements at three other loci. (A) Locus 3-4. (B) Locus 9-4. (C) Locus 16-17. Only the major blocks of internal tandem repeats are shown. The positions of the primers for whole- and half-locus amplification are shown for all three loci.

Taken together, these data indicate that loci 1-2 and 5-6 exist in long tandem arrays.

Characterization of other loci containing internal repeats. A number of other DNA elements containing internal tandem repeats have been reported in *E. histolytica*. No attempts have been made to study their potential for the detection of intraspecies polymorphisms. We selected three of these internally repetitive loci for study: a 978-bp element described by Michel et al. (19; GenBank accession number M77091; our designation, locus 3-4), a 931-bp DNA element isolated by J. Rosales-Encina and D. Eichinger (personal communication; GenBank accession number AF265348; our designation, locus 9-4), and a 964-bp element reported by Huang et al. (15; our designation, locus 16-17). Schematic representations of the repeat arrangements seen at these loci are given in Fig. 6A, B, and C, respectively.

There is a high degree of identity between the two repeat blocks of locus 3-4 and those of locus 9-4. The ten CTATT ATA tandem repeats of locus 9-4 differ from the 11 CTTAT TATA tandem repeats of locus 3-4 only in the absence of a single nucleotide (T) at the second position of each unit. The repeat unit CTTTATTATTAT in locus 9-4 is identical to the 12-bp repeat units of locus 3-4 with the only difference being the total number of units seen, i.e., locus 3-4 has eight units, while locus 9-4 has only seven. In fact, this high degree of identity between the two loci is apparent in the flanking regions as well. The sequences from positions 1 to 540 and positions 541 to 931 of locus 9-4 are very similar to the nucleotide stretches spanning positions 401 to 977 and positions 1 to 400 in locus 3-4, respectively (data not shown).

On comparing the sequences of loci 1-2 and 5-6 with those of loci 3-4 and 9-4 we find that the repeat unit CTTTATTAT, which occurs a total of seven times in locus 1-2, is identical to the three 9-bp units present in the second repeat blocks of both loci 3-4 and 9-4. The repeat units of locus 5-6, however, were quite unique, as are the repeat flanking regions of both loci. The nucleotide sequence of locus 16-17 is completely different from that of the other loci. There are six major types of internal repeats, with some being arranged in tandem only, while others exist as both tandem and solitary copies (Fig. 6C). Besides these, duplications of 5 to 8 bp are also seen interspersed among these repeats (not shown).

PCR product size polymorphisms at loci 3-4, 9-4 and 16-17. Primers were designed to amplify all three loci, and the products were analyzed on 2.4% NuSieve agarose gels to look for size polymorphisms among the 13 *E. histolytica* isolates. In each case primers were also designed in the regions between the two main repeat blocks to look additionally for size variations in each half of the locus (Fig. 6).

Isolate HM-1:IMSS clone 9 gave a double amplification product of ca. 900 bp at locus 3-4 (Fig. 7A). Most *E. histolytica* isolates gave single major products with little size variation. Isolates 200:NIH and H-303:NIH, however, show a second band of equal intensity at ca. 850 bp, while isolates 8691 and 37.0C display a second band of ca. 1 kb. Amplification of the two half-loci presented a very different pattern, with much more variation seen than with the whole locus (Fig. 7B and 7C).

Amplification of locus 9-4 also gave two products between ca. 900 bp and 1 kb in isolate HM-1:IMSS clone 9 (Fig. 8A). Isolates 200:NIH and H-303:NIH again show two bands of

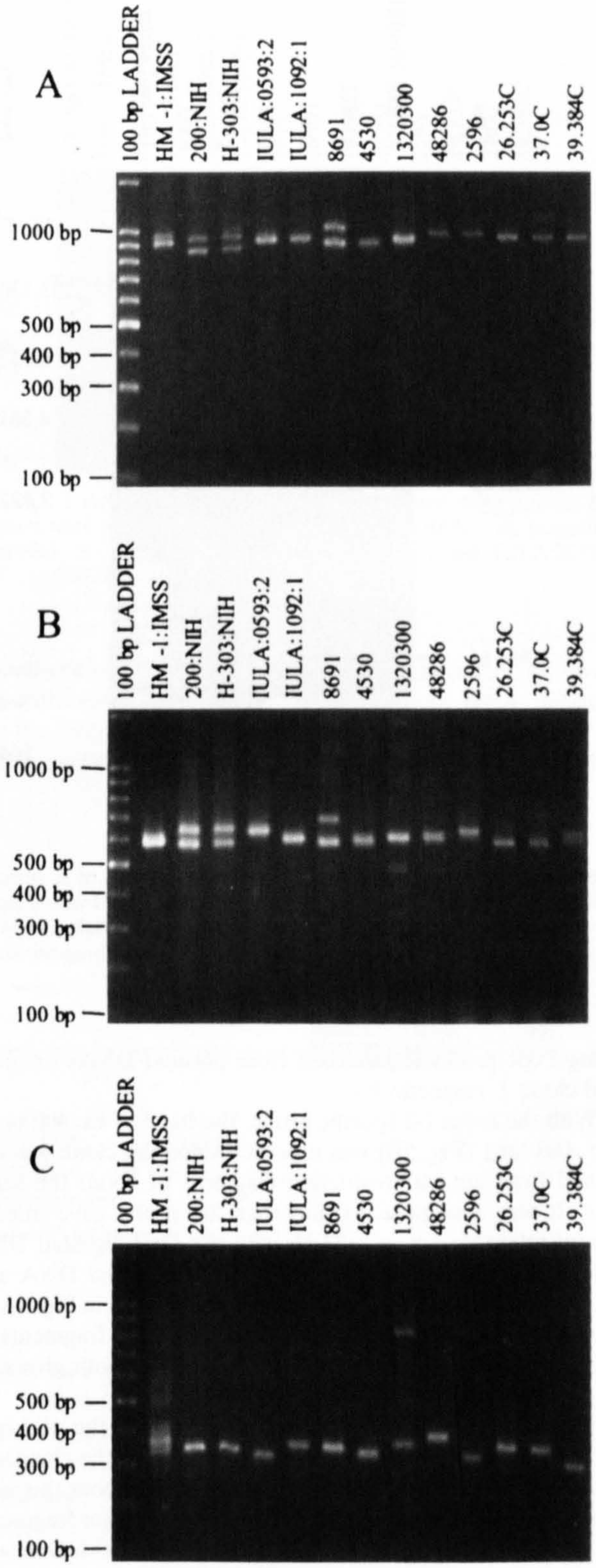


FIG. 7. Polymorphic DNA analysis of *E. histolytica* isolates. (A) Locus 3-4. Amplification products were generated using primers R3 and R4 at an annealing temperature of 55°C. (B) Half-locus 3-8. Amplification products were generated using primers R3 and R8 at an annealing temperature of 50°C. (C) Half-locus 7-4. Amplification products were generated using primers R7 and R4 at an annealing temperature of 50°C. Isolate origins: HM-1:IMSS (Mexico); 200:NIH (uncertain); H-303:NIH (VietNam); IULA:1092:1 and IULA:0593:2 (Venezuela); 8691, 4530, 1320300, and 48286 (Bangladesh); 2596, 26.253, 37.0C, and 39.384C (South Africa).

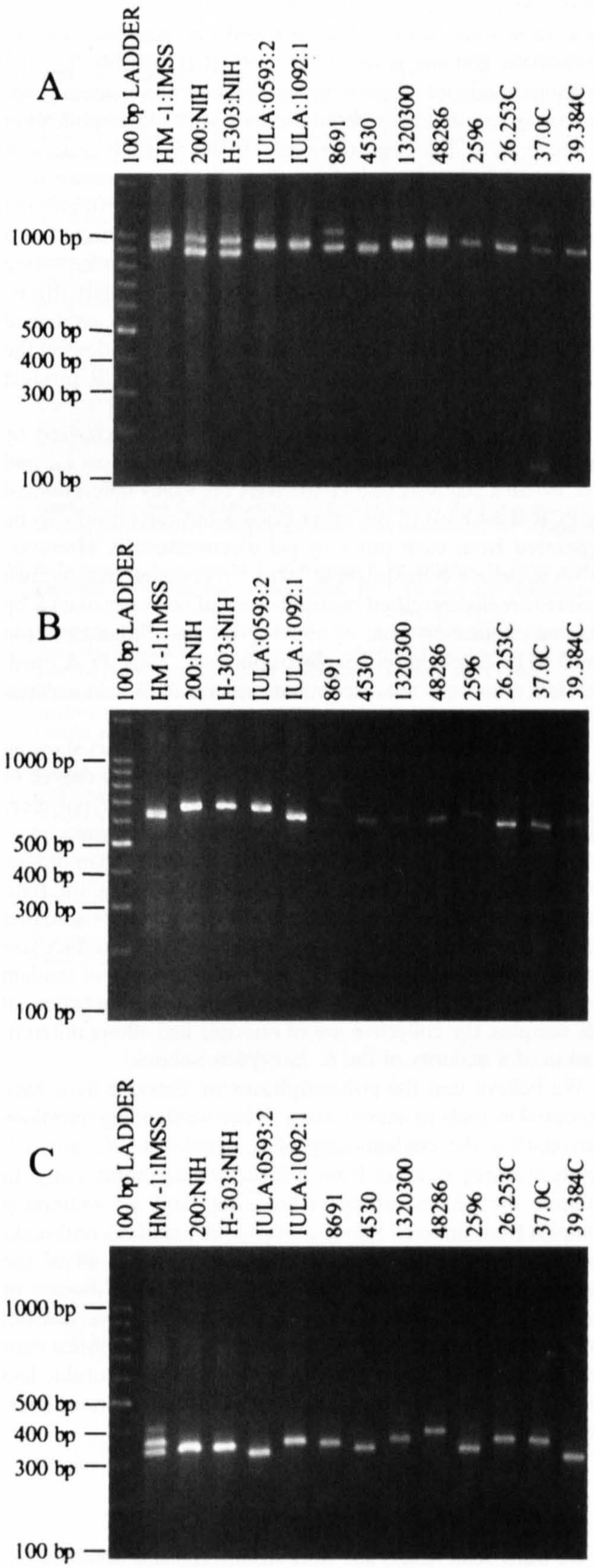


FIG. 8. Polymorphic DNA analysis of *Entamoeba histolytica* isolates. (A) Locus 9-4. Amplification products were generated using primers R9 and R4 at an annealing temperature of 55°C. (B) Half-locus 9-11. Amplification products were generated using primers R9 and R11 at an annealing temperature of 50°C. (C) Half-locus 10-4. Amplification products were generated using primers R10 and R4 at an annealing temperature of 50°C. Isolate origins: HM-1:IMSS (Mexico); 200:NIH (uncertain); H-303:NIH (VietNam); IULA:1092:1 and IULA:0593:2 (Venezuela); 8691, 4530, 1320300, and 48286 (Bangladesh); 2596, 26.253, 37.0C, and 39.384C (South Africa).

equal intensity. As before, amplification of both half-loci (Fig. 8B and C) produced a greater variety of banding patterns than was seen at the whole locus.

Amplification of locus 16-17 gave the expected product of ca. 900 bp in isolate HM-1:IMSS clone 9 (Fig. 9A). Many of the *E. histolytica* isolates gave single major products with little size variation among them, although isolate 4530 gave two clear products of equal intensity and certain others gave two bands very close in size. Amplification of the two half-loci again produced a highly polymorphic array of bands (Fig. 9B and C).

DISCUSSION

We were unsuccessful in our attempt to clone microsatellite loci using a modification of a method that was successful in other organisms. We did, however, isolate two novel loci containing internal tandem repeats (1-2 and 5-6). Whether this reflects an absence or a reduced population of the classical di- and trinucleotide microsatellites in the *E. histolytica* genome or simply their relative abundance is at present too early to say. A $(GA)_{27}$ stretch in an expressed sequence tag has been reported by Azam et al. (3), suggesting that microsatellites do exist in this organism. Three other, previously reported internally repetitive loci were also studied.

The genomic organization of the loci was investigated by Southern blotting. The large fragments generated by *Dra*I and/or *Eco*RI detected with the locus 1-2 and 5-6 specific probes and the presence of major fragments of the same size with two or more restriction enzymes suggests that both loci are arranged in long tandem arrays. The fact that some enzymes generate very large fragments also was reported by Michel et al. (19) for locus 3-4, and these authors also suggested that it was tandemly arrayed. Similarly, Southern hybridization with locus 16-17 indicated that this element was tandemly repeated (15). The successful amplification of locus 9-4 using the primer orientation indicated in Fig. 6B suggests that this element is in tandem arrays also.

Significant levels of identity exist between the sequences of loci 3-4 and 9-4, in the repeat domains as well as the flanking regions. Similarity has also been noted between some of the repeat units of locus 1-2 and those of loci 3-4 and 9-4. Two other internally repetitive DNA elements were reported independently by Lohia et al. (18) and Willhoeft and Tannich (30) which also bear remarkable similarity in their repeat domains to loci 3-4 and 9-4. However, complete sequence alignment of all five DNA sequences shows enough variation to suggest that they are distinct members of the same family of repeat containing DNA elements (data not shown). That loci 3-4 and 9-4 are indeed distinct is clearly seen from comparing Fig. 7B and Fig. 8B.

Size variations within the repeated domains were studied, and all loci studied showed PCR product length polymor-

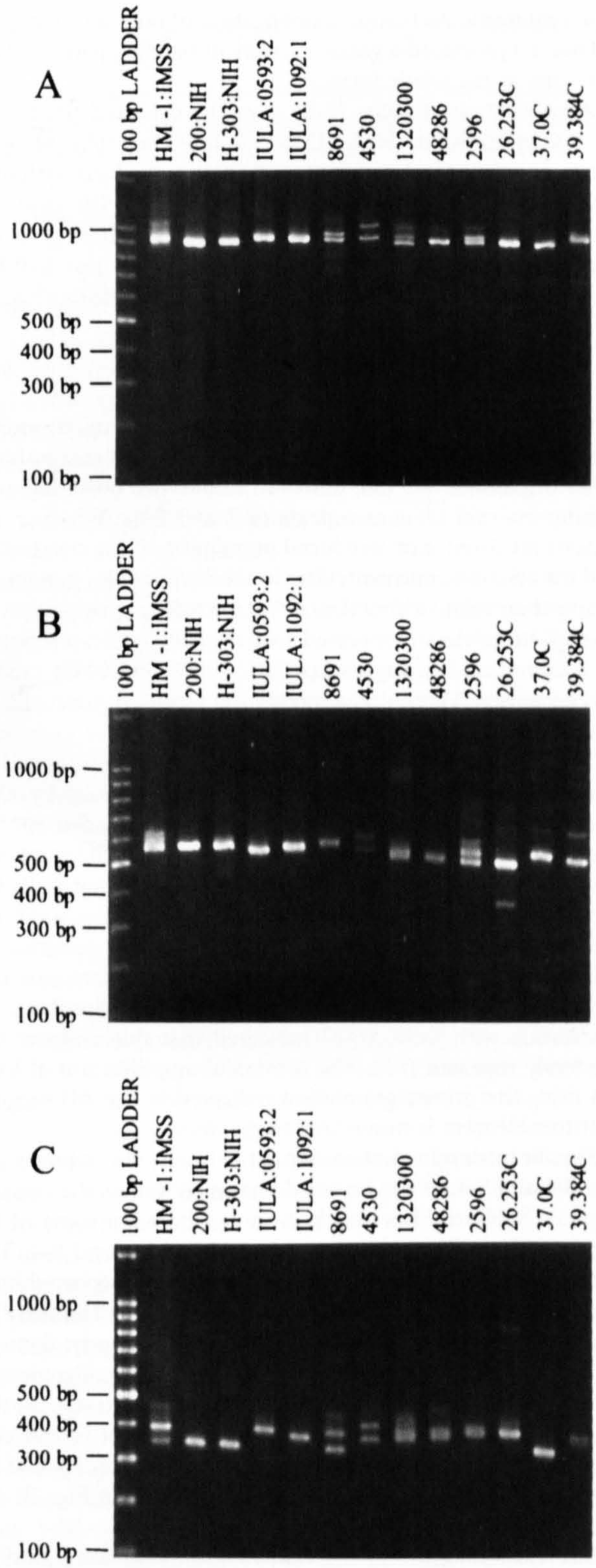


FIG. 9. Polymorphic DNA analysis of *E. histolytica* isolates. (A) Locus 16-17. Amplification products were generated using primers R16 and R17 at an annealing temperature of 55°C. (B) Half-locus 16-19. Amplification products were generated using primers R16 and R19 at an annealing temperature of 54°C. (C) Half-locus 18-17. Amplification products were generated using primers R18 and R17 at an annealing temperature of 54°C. Isolate origins: HM-1:IMSS (Mexico); 200:NIH (uncertain); H-303:NIH (VietNam); IULA:1092:1 and IULA:0593:2 (Venezuela); 8691, 4530, 1320300, and 48286 (Bangladesh); 2596, 26.253, 37.0C, and 39.384C (South Africa).

phism. At most loci, amplification results in two or three bands in at least some isolates. Present evidence suggests that the *Entamoeba* genome is tetraploid (29). It is possible that the multiple bands we observe reflect polymorphism among homologous loci on allelic chromosomes. Multiple amplification products have also been reported for the SREHP gene in a number of *E. histolytica* isolates (8), a finding consistent with the isolation of cDNAs containing variable numbers of internal tandem repeats (17). SREHP gives the pattern expected of a single copy gene when analyzed by Southern blotting indicating that the length differences are allelic variations. Alternatively, the presence of multiple bands in this study could be explained by the existence of these repeat loci at multiple locations in the *Entamoeba* genome, each with a characteristic PCR product size.

The observed size variation was further characterized by nucleotide sequence comparison of five isolates at loci 1-2 and 5-6. Isolates 200:NIH and H-303:NIH are easily differentiated by PCR from most of the other isolates but cannot reliably be separated from each other by gel electrophoresis. However, DNA sequence comparison at locus 5-6 revealed that the two isolates are distinguished by the absence of one copy of an 8-bp tandem duplication from strain 200:NIH; this difference is too small to be detected on gel electrophoresis. Thus, PCR product sizes alone may not discriminate among all distinct isolates, at least under these conditions.

From our results it appears that a number of loci showing size polymorphism are present in *E. histolytica*. The degree of diversity seen varies, with some loci showing more polymorphism and thus having a greater potential for detecting interstrain polymorphism among *E. histolytica* isolates than others. Despite being from geographically restricted regions, the Bangladeshi and South African samples could be differentiated with ease. For the most part variations in PCR product size appear to be a result of differences in the numbers of tandem repeat units. While no single locus discriminates between all the samples, the collective use of multiple loci allows differentiation of a majority of the *E. histolytica* isolates.

We believe that the polymorphisms we describe here have potential as tools to answer many of the outstanding questions surrounding the epidemiology of *E. histolytica*. We are currently studying samples from a broader geographic range to validate the general utility of these loci and are examining samples from infected family groups and amebiasis outbreaks for shared polymorphisms. In the present work, all of the isolates came from individuals who had invasive disease or were likely infected by someone who had invasive disease. Whether invasive and noninvasive strains of *E. histolytica* exist remains to be established, but hopefully the polymorphic loci described here will be useful in answering this important question.

ACKNOWLEDGMENTS

We thank John Ackers and Aura Aguirre (London School of Hygiene and Tropical Medicine), Terry Jackson (The Medical Research Council of South Africa, Durban), and Rashidul Haque (The International Center for Diarrhoeal Diseases Research, Bangladesh) for providing the DNA samples used in this study and John Ackers for reading the manuscript.

REFERENCES

1. **Anderson, T. J. C., X. Z. Su, M. Bockarie, M. Lagog, and K. P. Day.** 1999. Twelve microsatellite markers for characterisation of *Plasmodium falciparum* from finger-prick blood samples. Parasitology **119:**113–125.
2. **Anonymous.** 1997. WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. Epidemiol. Bull. PAHO **18:**13–14.
3. **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.
4. **Bhattacharya, S., A. Bhattacharya, and L. S. Diamond.** 1992. *Entamoeba histolytica* extrachromosomal circular ribosomal DNA: analysis of clonal variation in a hypervariable region. Exp. Parasitol. **74:**200–204.
5. **Blanc, D. S., and P. G. Sargeaunt.** 1991. *Entamoeba histolytica* zymodemes: exhibition of γ and δ bands only of glucose phosphate isomerase and phosphoglucomutase may be influenced by starch content in the medium. Exp. Parasitol. **72:**87–90.
6. **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.
7. **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, Kans.
8. **Clark, C. G., and L. S. Diamond.** 1993. *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. **77:**450–455.
9. **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.
10. **de la Vega, H., C. A. Specht, C. E. Semino, P. W. Robbins, D. Eichinger, D. Caplivski, S. Ghosh, and J. Samuelson.** 1997. Cloning and expression of chitinases of *Entamoebae*. Mol. Biochem. Parasitol. **85:**139–147.
11. **Diamond, L. S., and C. G. Clark.** 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (Emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. J. Eukaryot. Microbiol. **40:**340–344.
12. **Diamond, L. S., C. G. Clark, and C. C. Cunnick.** 1995. YI-S, a casein-free medium for axenic cultivation of *Entamoeba histolytica*, related *Entamoeba, Giardia intestinalis* and *Trichomonas vaginalis*. J. Eukaryot. Microbiol. **42:**277–278.
13. **Fischer, D., and K. Bachmann.** 1998. Microsatellite enrichment in organisms with large genomes (*Allium cepa* L.). BioTechniques **24:**796–802.
14. **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.
15. **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**(Suppl.)**:**S1–S4.
16. **Jackson, T. F. H. G., and S. Suparsad.** 1997. Zymodeme stability of *Entamoeba histolytica* and *E. dispar*. Arch. Med. Res. **28**(Suppl.)**:**S304–S305.
17. **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.
18. **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.
19. **Michel, B., A. Alagón, P. Lizardi, and M. Zurita.** 1992. Characterization of a repetitive DNA element from *Entamoeba histolytica*. Mol. Biochem. Parasitol. **51:**165–168.
20. **Mittal, V., A. Bhattacharya, and S. Bhattacharya.** 1992. Organization of repeated sequences in the region downstream to rRNA genes in the rDNA episome of *Entamoeba histolytica*. Arch. Med. Res. **23:**17–18.
21. **Mittal, V., D. Sehgal, A. Bhattacharya, and S. Bhattacharya.** 1992. A second short repeat sequence detected downstream of rRNA genes in the *Entamoeba histolytica* rDNA episome. Mol. Biochem. Parasitol. **54:**97–100.
22. **Oliveira, R. P., N. E. Broude, A. M. Macedo, C. R. Cantor, C. L. Smith, and S. D. J. Pena.** 1998. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proc. Natl. Acad. Sci. USA **95:**3776–3780.
23. **Robinson, G. L.** 1968. The laboratory diagnosis of human parasitic amoebae. Trans. R. Soc. Trop. Med. Hyg. **62:**285–294.
24. **Russell, R., M. P. Iribar, B. Lambson, S. Brewster, J. M. Blackwell, C. Dye, and J. W. Ajioka.** 1999. Intra and inter-specific microsatellite variation in the *Leishmania* subgenus *Viannia*. Mol. Biochem. Parasitol. **103:**71–77.
25. **Sargeaunt, P. G.** 1987. The reliability of *Entamoeba histolytica* zymodemes in clinical diagnosis. Parasitol. Today **3:**40–43.
26. **Sehgal, D., A. Bhattacharya, and S. Bhattacharya.** 1993. Analysis of a polymorphic locus present upstream of rDNA transcription units in the extrachromosomal circle of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **62:**129–130.
27. **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.
28. **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.
29. **Willhoeft, U., and E. Tannich.** 1999. The electrophoretic karyotype of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **99:**41–53.
30. **Willhoeft, U., and E. Tannich.** 2000. Fluorescence microscopy and fluorescence in situ hybridization of *Entamoeba histolytica* nuclei to analyse mitosis and the localization of repetitive DNA. Mol. Biochem. Parasitol. **105:**291–296.

JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 2002, p. 1271–1276 Vol. 40, No. 4
0095-1137/02/$04.00+0 DOI: 10.1128/JCM.40.4.1271–1276.2002

Simultaneous Differentiation and Typing of *Entamoeba histolytica* and *Entamoeba dispar*

Mehreen Zaki, Parool Meelu,† Wei Sun,‡ and C. Graham Clark*

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom

Received 17 October 2001/Returned for modification 13 December 2001/Accepted 20 January 2002

Sequences corresponding to some of the polymorphic loci previously reported from *Entamoeba histolytica* have been detected in *Entamoeba dispar*. Comparison of nucleotide sequences of two loci between *E. dispar* strain SAW760 and *E. histolytica* strain HM-1:IMSS revealed significant differences in both repeat and flanking regions. The tandem repeat units varied not only in sequence but also in number and arrangement between the two species at both the loci. Using the sequences obtained, primer pairs aimed at amplifying species-specific products were designed and tested on a variety of *E. histolytica* and *E. dispar* samples. Amplification results were in complete agreement with the original species classification in all cases, and the PCR products displayed discernible size and pattern variations among the isolates.

The realization that *Entamoeba histolytica* and *Entamoeba dispar* are two distinct but morphologically identical species (8) has had a major impact on all aspects of amebiasis research, most notably epidemiology (2, 3). The 10% prevalence estimate made by Walsh (21), though still widely quoted, was based on data collected before the formal redescription of the species and so does not distinguish between infections with *E. histolytica*, which can cause invasive intestinal and extraintestinal disease, and those with *E. dispar*, which do not. It is now known that even in areas where invasive amebiasis is common *E. dispar* is by far the more prevalent species (11).

Tools that allowed accurate differentiation of the two species were clearly needed, and in the past decade differentiation based on DNA amplification has been a research focus of many groups. Species-specific primers that amplify regions of several different genes have been used (1, 7, 14, 17, 18, 19). Using trophozoites in culture, comparisons showed that PCR is more sensitive and specific than the enzyme-linked immunosorbent assay-based stool antigen detection kits, which employ monoclonal antibodies for the detection and differentiation of *E. histolytica* and *E. dispar* (16). A significant percentage of individuals in areas of high endemicity could be simultaneously infected with both *E. histolytica* and *E. dispar* (16). However, the major drawback of using culture is that mixed infections are overlooked and were only detected after PCR was used on DNA extracted directly from stool (12). Field studies that compared PCR and antigen detection methods directly on stool samples suggest that these methods perform equally well (13).

But is *E. dispar* really nonpathogenic, and should it on this basis be completely dismissed as a subject for further investigation? It has been shown to be capable of producing variable focal intestinal lesions in animals (4, 9, 20) and of destroying epithelial cell monolayers in vitro (10). There is also some evidence that pathological changes may occur in some humans (15), though invasive lesions and symptomatic infections have to date not been reported. Whether these characteristics are variable among strains is unknown.

None of the above species differentiation methods can detect intraspecies variation, however. We have recently shown that a number of loci displaying PCR fragment size polymorphism exist in *E. histolytica* (22). All the loci contain tandemly arranged, complex internal repeat units ranging in length from 8 to 16 bp. Variations in the total number of PCR amplification products obtained per isolate and their sizes were seen. Nucleotide sequence analysis revealed that for the most part the observed size variation was a direct consequence of differences in the numbers of internal short tandem repeat units. The patterns appear to be stable over time in culture and in the same patient (unpublished data).

Using these same polymorphic locus-specific primers we have now observed intraspecies polymorphism among *E. dispar* isolates as well. The present study describes the interstrain variations seen in *E. dispar* and the design and testing of species-specific primers for two of the loci in both *E. histolytica* and *E. dispar*.

MATERIALS AND METHODS

***Entamoeba* samples.** Table 1 summarizes the *E. histolytica* and *E. dispar* samples used in this study. All the South African isolates were from individuals clinically classified as asymptomatic and were serology negative by an agarose gel diffusion test, except for isolate 62.628K, which was listed as weakly positive (T. F. H. G. Jackson and S. Reddy, personal communication).

Monoxenic *E. dispar* strain SAW760 and axenic *E. histolytica* strain HM-1:IMSS clone 9 were maintained in LYI-S-2 medium (7a) supplemented with 10% heat-inactivated adult bovine serum (Sigma-Aldrich) plus *Crithidia fasciculata* or 15% heat-inactivated adult bovine serum, respectively, at 36°C.

E. histolytica strains HM-1:IMSS, DKB, HB-301:NIH, IULA:0593:2, 887C, J2, and J3 are all from patients who presented with intestinal disease, while the other *E. histolytica* samples come from asymptomatic individuals.

* Corresponding author. Mailing address: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom. Phone: 44-207-927 2351. Fax: 44-207-636 8739. E-mail: graham.clark@lshtm.ac.uk.

† Present address: Department of Pediatric Hematology/Oncology, Medical School Hannover, D-30625 Hannover, Germany.

‡ Present address: Kirin Pharmaceuticals (China) Co., Ltd., Beijing, China.

TABLE 1. Origin of *Entamoeba* samples

Isolate	Strain origin	DNA origin	*Entamoeba* sp.	Source[a]	Clinical diagnosis[b]
16.156N	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
22.211L	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
22.212M	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
29.284N	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
36.352L	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
41.410K	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
49.484L	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
49.485L	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
49.486L	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
50.503I	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
59.595K	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
62.628K	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
88.881H	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
94.943I	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
99.996G	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
11691	South Africa	Lysate (xenic)	*E. dispar*	A	Asymptomatic cyst passer
A1	Colombia	DNA (xenic)	*E. dispar*	B	Asymptomatic
SAW 760	England	Culture (monoxenic)	*E. dispar*	C	Asymptomatic cyst passer#
HM-1:IMSS	Mexico	Culture (axenic)	*E. histolytica*	C	Amebic dysentery#
DKB	England	Culture (axenic)	*E. histolytica*	D	Amebic dysentery#
Rahman	England/India	Culture (axenic)	*E. histolytica*	D	Asymptomatic cyst passer#
HB-301:NIH	Burma	Culture (axenic)	*E. histolytica*	D	Amebic dysentery#
IULA:0593:2	Venezuela	Culture (axenic)	*E. histolytica*	E	Amebic dysentery
887C	Malaysia/Australia	DNA (xenic)	*E. histolytica*	F	Diarrhea
RPS	Brazil	DNA (axenic)	*E. histolytica*	G	Asymptomatic
462	Brazil	DNA (axenic)	*E. histolytica*	G	Asymptomatic
J2	Japan	Lysate (axenic)	*E. histolytica*	H	Colitis
J3	Japan	Lysate (axenic)	*E. histolytica*	H	Colitis
J5	Japan	Lysate (axenic)	*E. histolytica*	H	Asymptomatic
66-1	Vietnam	DNA (fecal)	*E. histolytica*	I	Asymptomatic carrier

[a] Source code definitions are as follows: A, T. F. H. G. Jackson and S. Reddy, Medical Research Council, Durban, South Africa, Robinson's medium; B, A. Aguirre, London School of Hygiene and Tropical Medicine, Robinson's medium, purified DNA provided; C, this laboratory, LYI-S-2 medium; D, D. Nolder, London School of Hygiene and Tropical Medicine, TYI-S-33 medium; E, J. P. Ackers and A. Shire, London School of Hygiene and Tropical Medicine, YI-S medium; F, J. Williams, London School of Hygiene and Tropical Medicine, Robinson's medium, purified DNA provided; G, M. A. Gomes, Universidade Federal de Minas Gerais, Brazil, TYI-S-33 medium, purified DNA provided; H, S. Kobayashi and T. Takeuchi, Keio University, Japan, TYI-S-33 medium; I, E. Tannich, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany, purified DNA provided, extracted with the Qiagen stool DNA extraction kit.

[b] Clinical diagnosis was provided by the supplier of the sample except for those marked with pound signs, where the information was obtained from the catalogue of the American Type Culture Collection (www.atcc.org).

Isolation of nucleic acids. DNA was isolated from cultures or lysates as previously described (5, 6), dissolved in 10 mM Tris-Cl, pH 8.5, and passed over a Microspin S-200 HR column (Amersham Pharmacia Biotech Inc.). RNA was removed by the addition of RNase A (Promega) to 0.05 μg ml^{-1}.

PCR amplification of repeated DNA-containing loci from *E. dispar*. Genomic DNA was subjected to PCR using primers designed to amplify the repeated DNA-containing loci (Table 2). Amplification consisted of 30 cycles of 1 min at 94°C, 1.5 min at the primer-dependent annealing temperature, and 2 min at 72°C with a final extension of 5 min at 72°C. Amplified products were analyzed by electrophoresis using 1.8% agarose gels (Gibco BRL) in 1× Tris-boric acid-EDTA buffer.

Isolation of repeated DNA-containing loci from *E. dispar* and nucleotide sequence analysis. PCR products obtained from *E. dispar* strain SAW760 with primers for loci 1-2 and 5-6 (22) were cloned into pGEM-T Easy (Promega) and sequenced (MWG Biotech). The resulting sequences were aligned by eye with those of *E. histolytica* from the respective loci. Species-specific primers were designed by using the aligned sequences of loci 1-2 and 5-6 for both *E. histolytica* and *E. dispar* (Table 3).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this paper have been submitted to the GenBank database with accession no. AY058216 and AY058217.

RESULTS AND DISCUSSION

PCR amplification of repeated DNA-containing loci from *E. dispar*. Our initial aim was to see if the intraspecies PCR fragment size polymorphism described for *E. histolytica* isolates (22) could be observed among *E. dispar* samples and also whether the primers used were species specific or not. To this end, previously described primers for *E. histolytica* (Table 2) were used to attempt amplification of the corresponding loci in *E. dispar* strain SAW760. Amplification products were obtained at all loci and half loci and were compared to those of *E. histolytica* strain HM-1:IMSS (Fig. 1). Although the patterns

TABLE 2. Polymorphic-locus-specific oligonucleotide primers

Primer	Sequence (5′→3′)
R1	CTG GTT AGT ATC TTC GCC TGT
R2	CTT ACA CCC CCA TTA ACA AT
R3	GCT ATG GTC GGT ATC GAT ATC
R4	CCT TAG GTC ACT GGT TCG AA
R5A	CTA AAG CCC CCT TCT TCT ATA ATT
R6A	CTC AGT CGG TAG AGC ATG GT
R7	CTT TAC TTC TCT TTT ACC ACG
R8	CGT GGT AAA AGA GAA GTA AAG
R9	CTA CAT CTA CAG TCC TCC GCT
R10	CTT ACT TCT CTT TAC CAC GAC
R11	GTC GTG GTA AAG AGA AGT AAG
R16	AAG CTT CCT TAG CTC AGC TG
R17	TAA AAG GGG GAA GAA TAG GAA
R18	GGT TTC ATG GTG TAG TTG GT
R19	ACC AAC TAC ACC ATG AAA CC

TABLE 3. Species-specific oligonucleotide primers

Primer	Sequence (5′→3′)
Hsp1	GAG TTC TCT TTT TAT ACT TTT ATA TGT T
Hsp2	ATT AAC AAT AAA GAG GGA GGT
Hsp5	CTA TAA TTT ATA TAT TAT TCT CTT TGA GA
Hsp6	CAT TGT TTT TAA AGT TAA AGA CG
Dsp1	TTG AAG AGT TCA CTT TTT ATA CTA TA
Dsp2	TAA CAA TAA AGG GGA GGG
Dsp5	CTA TAC TAT ATT CTT TTT ATG TAC TTC CC
Dsp6	CTG AGA GCA TTG TTT TTA AAG AA

for both strains appear in general to be similar, at loci 1-2, 3-4, 9-4, and 16-19 the SAW760 products are ca. 50 to 100 bp larger.

Following successful amplification using SAW760 we tested DNA from xenic isolates of *E. dispar*. A total of 111 South African isolates, characterized as *E. dispar* on the basis of zymodeme analysis, were available. Initially a randomly selected group of some 10 to 20 DNA samples were tested at all loci. Very few samples gave amplification products for loci 5-6, 3-8, 7-4, 9-11, 10-4, 16-17, and 18-17. At locus 5-6, 9 of the initial set of 20 DNA samples tested gave amplification products but the ca. 50 samples tested thereafter all failed to amplify. In all, amplification at locus 5-6 was attempted for 77 of the 111 samples before the analyses were discontinued. It was, however, notable that there was a high degree of PCR product size variation among the nine positive samples. At least four different patterns could be distinguished (data not shown). Amplifications at the remaining loci were either negative or gave nonspecific banding patterns.

Based on these results, PCR amplification was carried out on all 111 DNA samples at loci 1-2, 3-4, and 9-4 and at half locus 16-19. In the final analysis, DNA from 11 of the 111 samples failed to amplify at any of these four loci. Each of the remaining 100 samples produced amplification products with one or more of the four primer pairs, with 85 to 95% of the samples being positive for each of the four loci. The products displayed size polymorphism at all four loci. Results obtained for 12 representative samples are shown in Fig. 2. The fragment sizes displayed by most of the isolates are similar to those seen with SAW760 at the same locus.

Isolation of repeated DNA-containing loci from *E. dispar*. The relative ease with which some of the primer pairs gave amplification products suggests that these oligonucleotides are derived from sequences which are conserved between the sister species *E. histolytica* and *E. dispar*. On the other hand, the failure of other primer pairs suggests that these lie in regions of sequence divergence.

This raises a number of questions. What is the degree to which the sequences differ between species? Do the differences exist only in the repeat-flanking regions? Are the differences in PCR product size and pattern between HM-1:IMSS and SAW760 at some loci simply a reflection of different numbers of tandem repeat units or does variation extend to the repeat sequences themselves?

To address these questions, the amplification products of strain SAW760 at loci 1-2 and 5-6 were cloned. One representative clone from each locus was sequenced and compared to corresponding sequences from strain HM-1:IMSS (22).

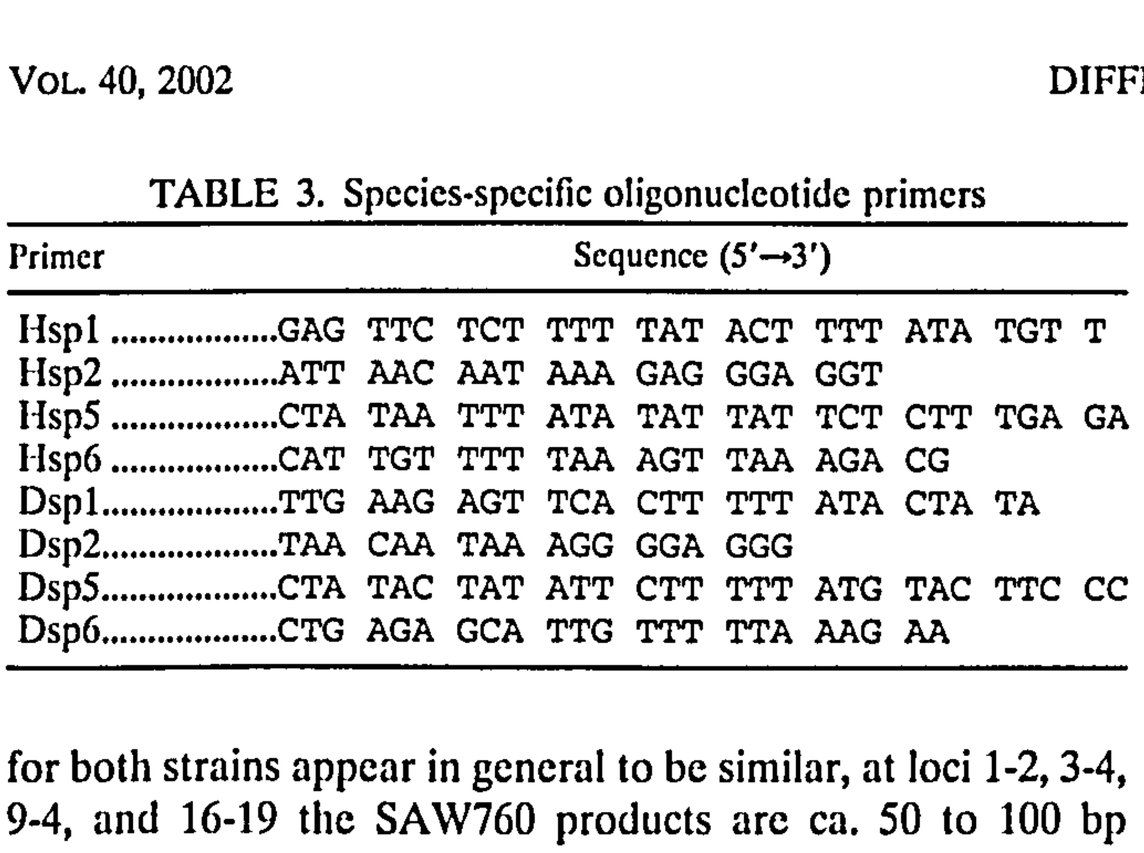


FIG. 1. Electrophoretic comparison of PCR amplification products of *E. histolytica* strain HM-1:IMSS (H) and *E. dispar* strain SAW760 (S) at 11 loci. The 100-bp DNA ladder is the size marker (M). Annealing temperatures used were as follows: locus 1-2, 53°C; locus 5-6, 56°C; locus 3-4, 55°C; locus 3-8, 50°C; locus 7-4, 50°C; locus 9-4, 55°C; locus 9-11, 50°C; locus 10-4, 50°C; locus 16-17, 55°C; locus 16-19, 54°C; locus 18-17, 54°C.

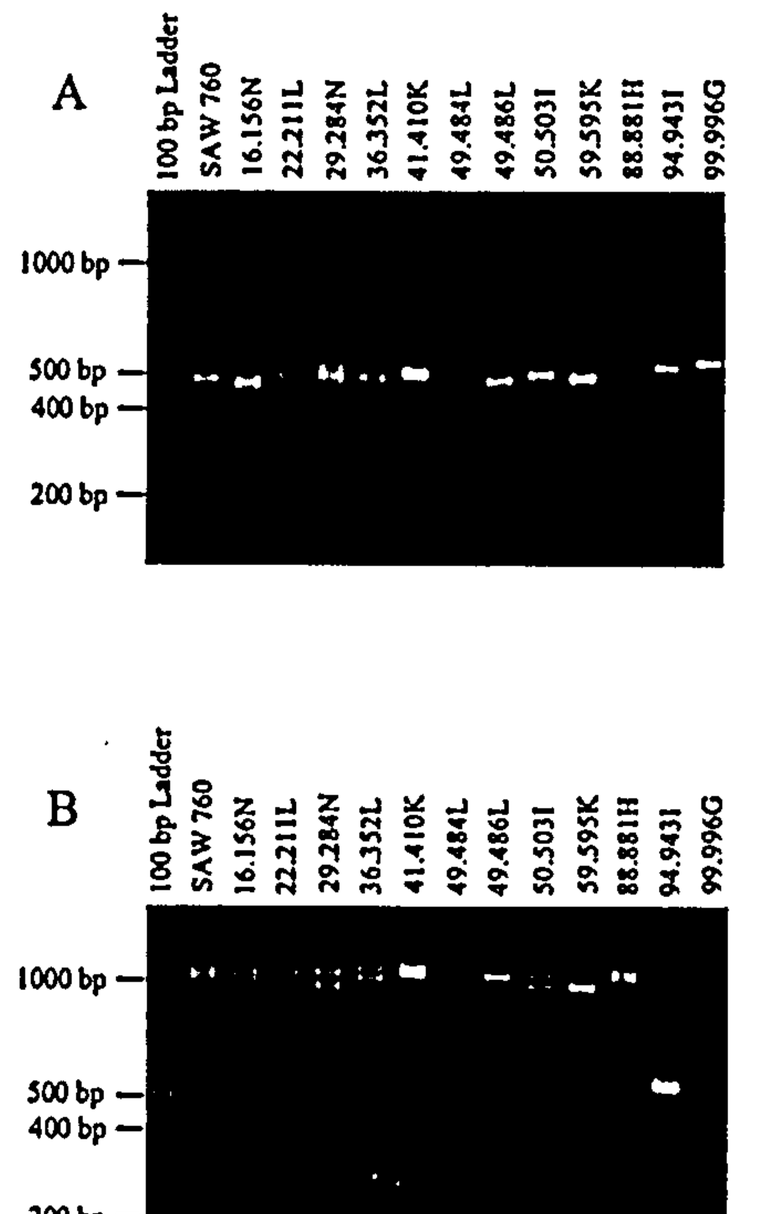
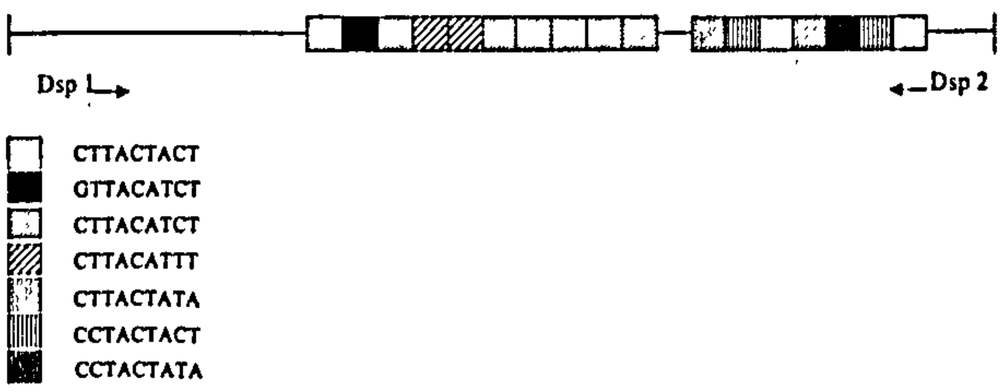


FIG. 2. Polymorphic DNA analysis of *E. dispar* isolates. (A) Locus 1-2. Amplification products were generated with primers R1 and R2 at an annealing temperature of 53°C. (B) Locus 9-4. Amplification products were generated with primers R9 and R4 at an annealing temperature of 55°C. (C) Locus 16-19. Amplification products were generated with primers R16 and R19 at an annealing temperature of 54°C.

A

```
        10         20         30         40         50         60
CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCGTTGAAGAGTT
        70         80         90        100        110        120
CACTTTTTATACTATATATATGTATATATTCTTATTTATTAGTCACTTATACTATTAACT
       130        140        150        160        170        180
TTATCTATTCCTCTTATATATTATACTTATTATACTTATTATATATATTATATCTATTCT
       190        200        210        220        230        240
TACTCTATACCTTTATTATATTATTTATATTATATATATTATATATATTTATTATCCTTA
       250        260        270        280        290        300
TTATTTTTATTATTCCACTTCTTATCTATTATATCTATTCTCACTTCTTATACTTACTAC
       310        320        330        340        350        360
TGTTACATCTCTTACATCTCTTACATTTCTTACATTTCTTACATCTCTTACATCTCTTAC
       370        380        390        400        410        420
ATCTCTTACATCTCTTACATCTCTTATTCCTCTTACTATACCTACTACTCTTACTACTCT
       430        440        450        460        470        480
TACTATACCTACTATACCTACTACTCTTACTACTCCTACTTTCCCCCTCCCCTTTATTGT
       490
TAATGGGGGTGTAAG
```

B



C

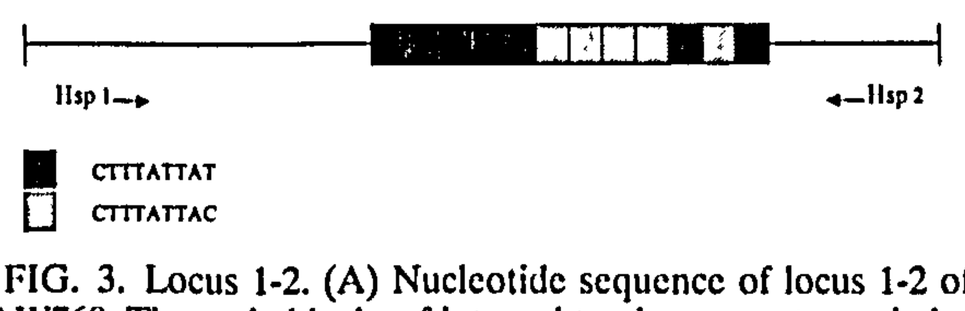


FIG. 3. Locus 1-2. (A) Nucleotide sequence of locus 1-2 of strain SAW760. The main blocks of internal tandem repeats are in boldface. Underlined and/or highlighted regions indicate the seven different types of repeat units. (B and C) Schematic representation of locus 1-2 in strain SAW760 (B) and strain HM-1:IMSS (C). The different types of internal tandem repeats and their arrangements with respect to each other are shown. Tandem duplications in the flanking regions are not shown. Positions of species-specific amplification primers are indicated.

Nucleotide sequence analysis at loci 1-2 and 5-6. The complete sequence of the locus 1-2 clone from strain SAW760 is 495 bp (Fig. 3A). There are two main repeat blocks which, between them, display seven distinct but highly related direct repeats arranged in tandem (Fig. 3B). Only one of the seven repeat types is represented in both repeat blocks. In addition, several tandem duplications of 5 to 9 bp are seen in the 5′ flanking regions of the major repeat blocks (not shown).

The complete sequence of the locus 5-6 clone from strain SAW760 is 510 bp (Fig. 4A) and consists of six major types of tandemly arranged repeats organized into five main blocks (Fig. 4B). Block 1 consists of sets of three and two repeats of the same sequence separated by 41 bp, in which a single copy of the same repeat is found. In contrast to the interspersed arrangement of repeats seen at locus 1-2, four of the five blocks seen in locus 5-6 consist of one repeat unit type exclusively. Here too, duplications and solitary copies of 5- to 9-bp sequences are interspersed among the major repeat blocks.

No open reading frames are found in either sequence.

Nucleotide sequence comparison at loci 1-2 and 5-6. When the sequences of loci 1-2 and 5-6 from strain SAW760 are compared to those from strain HM-1:IMSS (Fig. 3C and 4C), significant differences in the number, sequence, and arrangement of the repeats between the two are revealed. There were also differences in the 5′ and 3′ repeat-flanking regions between the two species at both loci. The highest degree of sequence identity was between the first 60 bp at the 5′ end of locus 1-2 (100%) (Fig. 5A) and between the last 75 bp at the 3′ end of locus 5-6 (98.6%) (Fig. 5B).

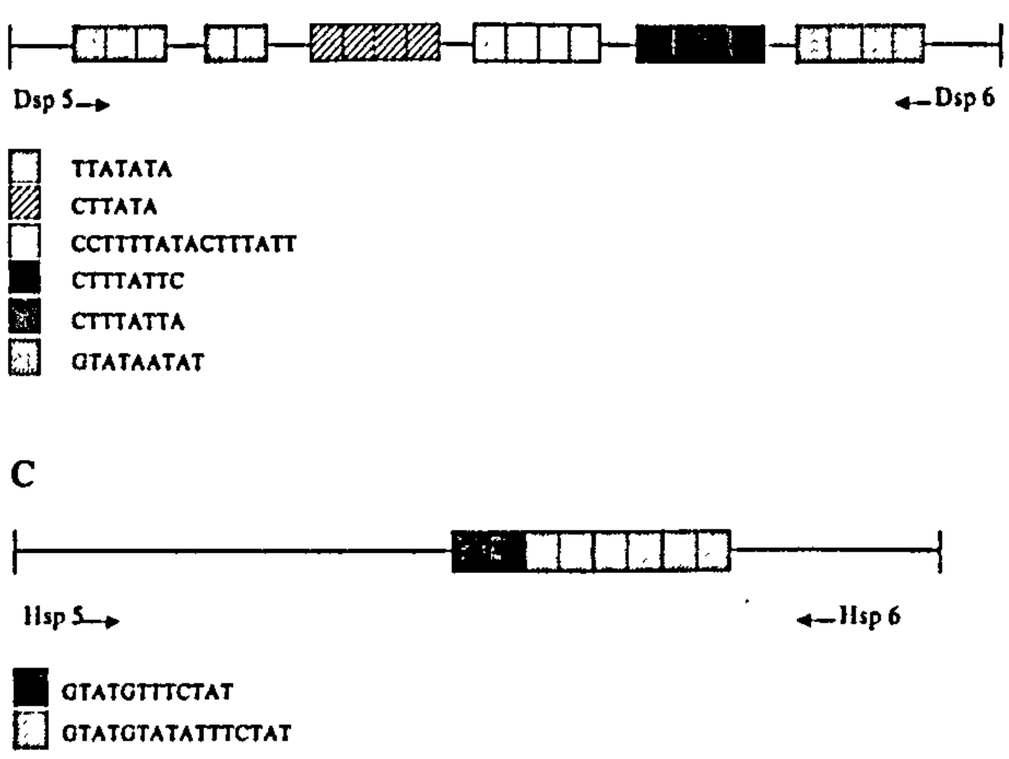
Significant differences between species in the repeat numbers and sequences and in the repeat-flanking regions of both loci were therefore found. The significantly greater number of repeat units accounts for the larger products for *E. dispar*.

Design and testing of species-specific primers. With the aim of amplifying species-specific products, four sets of primers in the repeat-flanking regions of both loci were designed (Table 3; Fig. 5). Amplification of strain HM-1:IMSS with *E. histolytica*-specific primers Hsp1 and Hsp2 gave the expected product of ca. 340 bp (Fig. 6A). No amplification was seen with strain SAW760. Similarly, strain SAW760 showed the expected product of ca. 430-bp with *E. dispar*-specific primers Dsp1 and

A

```
        10         20         30         40         50         60
CTAAAGCCCCCTTCTTCTATAATTAATCTATACTATATTCTTTTTATGTACTTCCCTTAT
        70         80         90        100        110        120
ATATTATATATTATATATTATTATATTTATTATATATATTATATGTATATATTCTACTTT
       130        140        150        160        170        180
ATATATTATATATTATATCCTTTATATTCTATATTCTATTTTATATTCATGACTTATACT
       190        200        210        220        230        240
TATACTTATACTTATATCCTTTATTATTATATTCCTTTATTACTATATTCCTTTTATACT
       250        260        270        280        290        300
TTATTCCTTTTATACTTTATTCCTTTTATACTTTATTCCTTTTATACTTTATTCCTTTAT
       310        320        330        340        350        360
TACTATATTACTATATTCTTTTTATACTTTATTCCTTTATTACTTTATTACTTTATTCCT
       370        380        390        400        410        420
TTTATACAAATACTCTCATGTATGTATATGTATAATATGTATAATATGTATAAATATGTAT
       430        440        450        460        470        480
AATATGTATGTATTTCTTTAAAAACAATGCTCTCAGCAGGTTTCGAACCTGCGACCCTGG
       490        500        510
CGTTATTAGCACCATGCTCTACCGACTGAG
```

B



C

FIG. 4. Locus 5-6. (A) Nucleotide sequence of locus 5-6 of strain SAW760. The main blocks of internal tandem repeats are in boldface. Underlined and/or highlighted regions indicate the six different types of repeat units. (B and C) Schematic representation of locus 5-6 in strain SAW760 (B) and strain HM-1:IMSS (C). The different types of internal tandem repeats and their arrangements with respect to each other are shown. Tandem duplications in the flanking regions are not shown. Positions of species-specific amplification primers are indicated.

A

```
Ed: 1    CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCG  50
Eh: 1    CTGGTTAGTATCTTCGCCTGTCACGTGAAAGACCGGGGTTCGAATCCCCG  50

Ed: 51   TTGAAGAGTTCACTTTTTATACTAT-ATATATGTATAT------------  87
Eh: 51   TTGAAGAGTTCTCTTTTTATACTTTTATATGTTTATAT------------  88

Ed: 446  CTTACTACTCCTACTTTCCCCCTCCCCTTTATTGTTAATGGGGGTGTAAG 495
Eh: 353  CTTACTACTCCTACTTTCACCTCCCTCTTTATTGTTAATGGGGGTGTAAG 402
```

B

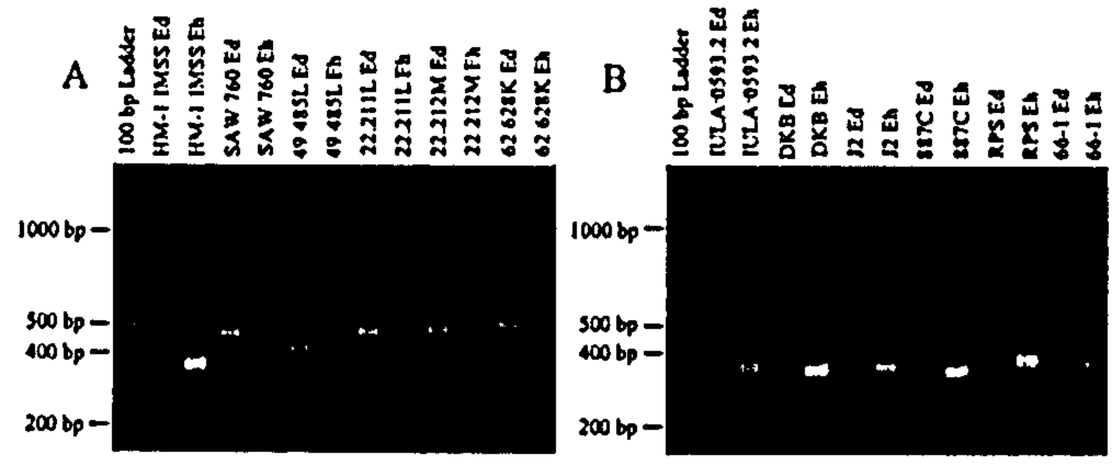
```
Ed: 1    CTAAAGCCCCCTTCTTCTATAATTAATCTATACTATATTCTTTTTATGTA  50
Eh: 1    CTAAAGCCCCCTTCTTCTATAATTTATATATTATTCTCTTTGAGACTTAT  50

Ed: 51   CTTCCCTTATATA-----------------------------------  63
Eh: 51   TTCTACTTTATTT-----------------------------------  63

Ed: 417  GTATAA---TATGTATGTATTT--CTTTAAAAACAATGCTCTCAGCAGGT 461
Eh: 325  GTATATTTCTATGTACGTCTTTAACTTTAAAAACAATGCTCTCAGCAGGT 374

Ed: 462  TTCGAACC-TGCGACCCTGGCGTTATTAGCACCATGCTCTACCGACTGAG 510
Eh: 375  TTCGAACCCTGCGACCCTGGCGTTATTAGCACCATGCTCTACCGACTGAG 424
```

FIG. 5. Alignment of nucleotide sequences from *E. histolytica* strain HM-1:IMSS (Eh) and *E. dispar* strain SAW760 (Ed). Only the repeat-flanking region sequences are shown. Nucleotides belonging to a repeat unit are in boldface. Single base differences are highlighted. Dashes indicate gaps introduced to optimize alignment. (A) Alignment at locus 1-2. Underlined regions show locations of the two pairs of 5′ and 3′ species-specific primer sequences (Dsp1 and Hsp1 and Dsp2 and Hsp2) which were used for PCR amplification. (B) Alignment at locus 5-6. Underlined regions show locations of the two pairs of 5′ and 3′ species-specific primer sequences (Dsp5 and Hsp5 and Dsp6 and Hsp6) which were used for PCR amplification.



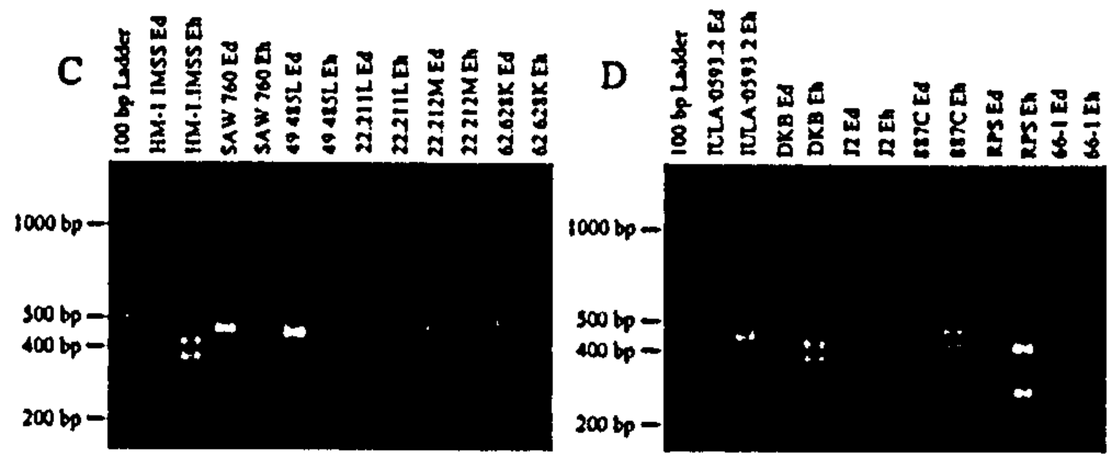


FIG. 6. Species-specific DNA analysis of *E. dispar* and *E. histolytica* isolates. (A and B) Locus 1-2. Amplification products were generated with primers Dsp1 and -2 (Ed) and primers Hsp1 and -2 (Eh) at an annealing temperature of 50°C. (C and D) Locus 5-6. Amplification products were generated using primers Dsp5 and -6 (Ed) at an annealing temperature of 52°C and primers Hsp5 and -6 (Eh) at an annealing temperature of 48°C.

Dsp2 and strain HM-1:IMSS failed to amplify with these primers. At locus 5-6, too, both *E. histolytica*-specific primers Hsp5 and Hsp6 and *E. dispar*-specific primers Dsp5 and Dsp6 gave the expected products of ca. 350 bp with strain HM-1:IMSS and 430 bp with strain SAW760, respectively, and in a species-specific manner (Fig. 6C).

Species specificity of these primers was then tested with 17 additional isolates, 11 of which had previously been characterized as *E. histolytica*. PCR amplification results for 10 of these isolates (Table 1) are shown for both locus 1-2 (Fig. 6A and B) and locus 5-6 (Fig. 6C and D). The presence or absence of PCR amplification products conformed to the original species classification.

The *E. histolytica* isolates tested came from a wide geographical range, and those of *E. dispar* came from three continents. All of them gave the expected PCR products with the appropriate species-specific locus 1-2 and 5-6 primers, and there was no amplification with the other species-specific primers. In addition, we tested the species specificity of all four primer pairs on *E. moshkovskii* strain Laredo. No amplification products were detected (data not shown). The source of the DNA, whether xenic or axenic culture or stool, was irrelevant. However, we have noted that the method by which the DNA is extracted from stool has a significant impact on amplification success.

One of the main questions in amebiasis research, which has not yet been resolved, is the basis for the wide spectrum of clinical manifestations observed among individuals infected with *E. histolytica* and/or *E. dispar*. The presence of both types of parasite and/or different strains of either parasite in the same patient could be one of the reasons for the differences in signs and symptoms in infected individuals. The availability of species-specific markers that simultaneously detect intraspecies polymorphisms provides us with the tools to address the role of parasite variation in the outcome of disease and to investigate patterns of transmission of both parasites.

ACKNOWLEDGMENT

We thank Aura Aguirre (London School of Hygiene and Tropical Medicine) for critical reading of the manuscript.

REFERENCES

1. Aguirre, A., D. C. Warhurst, F. Guhl, and I. A. Frame. 1995. Polymerase chain reaction-solution hybridization enzyme-linked immunoassay (PCR-SHELA) for the differential diagnosis of pathogenic and non-pathogenic *Entamoeba histolytica*. Trans. R. Soc. Trop. Med. Hyg. 89:187–188.
2. Anonymous. 1997. *Entamoeba* taxonomy. Bull. W. H. O. 75:291–292.
3. Anonymous. 1997. WHO/PAHO/UNESCO report. A consultation with experts on amoebiasis. Epidemiol. Bull. PAHO 18:13–14.
4. 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.
5. 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, Kans.
6. 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.
7. Clark, C. G., and L. S. Diamond. 1991. Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct. Mol. Biochem. Parasitol. 49:297–302.

7a. Clark, C. G., and L. S. Diamond. Methods for the cultivation of lumenal parasitic protists of clinical importance. Clin. Microbiol. Rev., in press.

8. Diamond, L. S., and C. G. Clark. 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (emended Walker, 1911) separating it from *Entamoeba dispar* Brumpt, 1925. J. Eukaryot. Microbiol. 40:340–344.
9. Espinosa-Cantellano, M., G. Castañon Gutierrez, and A. Martínez-Palomo.

1997. *In vivo* pathogenesis of *Entamoeba dispar*. Arch. Med. Res. 28(Suppl.): S204–S206.
10. **Espinosa-Cantellano, M., A. González-Robles, B. Chávez, G. Castañon, C. Argüello, A. Lázaro-Haller, and A. Martínez-Palomo.** 1998. *Entamoeba dispar*: ultrastructure, surface properties, and cytopathic effect. J. Eukaryot. Microbiol. 45:265–272.
11. **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.
12. **Haque, R., I. K. M. Ali, S. Akther, and W. A. Petri, Jr.** 1998. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. J. Clin. Microbiol. 36:449–452.
13. **Huston, C. D., R. Haque, and W. A. Petri, Jr.** 22 March 1999, posting date. Molecular-based diagnosis of *Entamoeba histolytica* infection. Exp. Rev. Mol. Med. [Online] http://www-ermm.cbcu.cam.ac.uk/99000599h.htm.
14. **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.
15. **McMillan, A., H. M. Gilmour, G. McNeillage, and G. R. Scott.** 1984. Amoebiasis in homosexual men. Gut 25:356–360.
16. **Mirelman, D., Y. Nuchamowitz, and T. Stolarsky.** 1997. Comparison of use of enzyme-linked immunosorbent assay-based kits and PCR amplification of rRNA genes for simultaneous detection of *Entamoeba histolytica* and *E. dispar*. J. Clin. Microbiol. 35:2405–2407.
17. **Núñez, Y. O., M. A. Fernández, D. Torres-Núñez, J. A. Silva, I. Montano, J. L. Maestre, and L. Fonte.** 2001. Multiplex polymerase chain reaction amplification and differentiation of *Entamoeba histolytica* and *Entamoeba dispar* DNA from stool samples. Am. J. Trop. Med. Hyg. 64:293–297.
18. **Tachibana, H., S. Kobayashi, M. Takekoshi, and S. Ihara.** 1991. Distinguishing pathogenic isolates of *Entamoeba histolytica* by polymerase chain reaction. J. Infect. Dis. 164:825–826.
19. **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.
20. **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.
21. **Walsh, J. A.** 1986. Problems in recognition and diagnosis of amoebiasis: estimation of the global magnitude of morbidity and mortality. Rev. Infect. Dis. 8:228–238.
22. **Zaki, M., and C. G. Clark.** 2001. Isolation and characterization of polymorphic DNA from *Entamoeba histolytica*. J. Clin. Microbiol. 39:897–905.