Genetic diversity in *Entamoeba histolytica*

C GRAHAM CLARK*, MEHREEN ZAKI and IBNE KARIM MD. ALI

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

**Corresponding author (Fax, 44-20-7636-8739; Email, graham.clark@lshtm.ac.uk)*

Genetic diversity within *Entamoeba histolytica* led to the re-description of the species 10 years ago. However, more recent investigation has revealed significant diversity within the re-defined species. Both protein-coding and non-coding sequences show variability, but the common feature in all cases is the presence of short tandem repeats of varying length and sequence. The ability to identify strains of *E. histolytica* may lead to insights into the population structure and epidemiology of the organism.

[Clark C G, Zaki M and Ali I K M 2002 Genetic diversity in *Entamoeba histolytica*; *J. Biosci*. (*Suppl. 3*) **27** 603–607]

1. Introduction

Intestinal amoebae of the genus *Entamoeba* offer very few morphological features on which identification can be based. Not surprisingly, therefore, with the advent of molecular tools there have been revisions made to the taxonomy of *Entamoeba* species. The most important changes are those that affect the pathogen *E. histolytica*.

The first involves the '*E. histolytica*-like' amoebae that are able to grow at a wide range of temperatures (the 'Laredo' strain and others). These have been identified as *E. moshkovskii*, which is typically a free-living species (Clark and Diamond 1991a). These organisms are occasionally isolated from humans, although how common such infections are is unclear.

A more important change, however, is the separation of *E. histolytica* from *E. dispar* (Diamond and Clark 1993). These two species are morphologically identical parasites but only the former causes invasive disease. The distinction initially was proposed based on isoenzyme analyses (Sargeaunt 1987) and subsequently this has been upheld through the discovery of many antigenic and genetic differences. The recognition of *E. dispar* as a distinct species has profound implications for the diagnosis, treatment and epidemiology of amoebiasis (Anonymous 1997). With respect to the latter, it means that *E. histolytica* prevalence data that were collected using stool microscopy alone are no longer of any use, as the species present cannot be identified retrospectively with any confidence. Clearly, new data on the prevalence of *E. histolytica* are needed, now that we have the tools necessary to identify this species specifically.

One of the long-standing mysteries of amoebiasis has been the disparity between the large number of people infected and the relatively small number who are symptomatic. Despite recognition of the non-invasive *E. dispar* as the more common species, it is apparent that many individuals infected with *E. histolytica*, as newly defined, never develop disease (Gathiram and Jackson 1987). One possible explanation for this is that genetic subgroups exist within *E. histolytica* that give rise to infections with different outcomes. In order to examine this possibility, it is necessary to have tools that detect variation within the species. The aim of this review is to examine where we stand at present with respect to achieving this goal.

Keywords. Isoenzymes; polymorphism; strain typing

Abbreviations used: PCR, Polymerase chain reaction; SREHP, serine-rich *E. histolytica* protein; STRs, short tandem repeats.

2. Isoenzymes

The first indication of variation within *E. histolytica* came from isoenzyme studies (Sargeaunt *et al* 1978). These identified not only the 'non-pathogenic' and 'pathogenic' subgroups that are now known as *E. dispar* and *E. histolytica*, respectively, but also the existence of variation within each group. Sargeaunt and colleagues defined 22 zymodemes divided between the two species (Sargeaunt 1987) and others have since reported more. These data are difficult to interpret, however, as the patterns do not appear to be reliable. Blanc and Sargeaunt (1991) reported that the most variable bands in the zymodeme patterns were influenced by the presence of starch in the medium. Jackson and Suparsad (1997) have reported that many zymodemes 'disappear' upon removal of the bacterial flora, suggesting that at least some of the bands are of bacterial rather than amoebal origin. If the zymodemes defined by stable bands alone are counted, only three remain for *E. histolytica* (II, XIV and XIX) and one for *E. dispar* (I).

In addition to low diversity, zymodeme analysis has other drawbacks. It relies on establishing the amoebae in culture – a large number of cells are needed for the enzyme analyses. This process is not always successful. In addition, cultivation may lead to selection – in the case of mixed infections one species or strain may outgrow the other, which is not desirable when studying diversity. Finally, cultivation is also labour intensive.

In combination, these potential and real problems have meant that isoenzyme analysis has been superseded by DNA-based methods as the approach of choice for detecting variation in *E. histolytica*. The use of the polymerase chain reaction (PCR) means that, in theory, no cultivation is necessary – extraction of DNA directly from stool specimens should provide appropriate and sufficient material for analysis. In practise, this is not straightforward, and most studies still rely on cultivated cells. However, with improvements in extraction methods, the analysis of DNA derived directly from stool is likely to become more widespread.

PCR targets for the detection of diversity fall into two categories – protein-encoding genes and non-coding DNA. In all cases, the basis of the detection system is variation in the number and/or sequence of short tandem repeats (STRs) in the DNA.

3. Protein

Two protein-encoding genes have been shown to exhibit polymorphism in the coding region – the "serine-rich *E. histolytica* protein" (SREHP; Stanley *et al* 1990) and chitinase (de la Vega *et al* 1997).

SREHP was first described as an antigen (Stanley *et al* 1990; Köhler and Tannich 1993) and is being investigated as a vaccine candidate. An early paper showed that independent cDNA sequences from the same strain encoded different numbers of amino acid STRs (Köhler and Tannich 1993; figure 1A). This observation was extended using PCR and PCR-RFLP to detect a large amount of variation in this gene in a geographically diverse collection of axenic *E. histolytica* isolates (Clark and Diamond 1993). Subsequently, DNA sequencing has been used to further define the basis of variation and to show that repeat number alone does not give the full picture of the diversity that exists in both *E. histolytica* and *E. dispar* (Ghosh *et al* 2000). A nested SREHP-PCR approach has recently been used to investigate *E. histolytica* diversity in a single human population using DNA extracted from stool specimens (Ayeh-Kumi *et al* 2001).

The gene encoding chitinase (figure 1B) has been used less frequently but repeat number variation is detected by PCR and sequencing in both *E. histolytica* and *E. dispar* (Ghosh *et al* 2000).

4. Ribosomal DNA

The ribosomal RNA genes (rDNA) of *E. histolytica* are carried on circular episomes, usually as two copies per circle organized in inverted repeats. In addition, these molecules carry several classes of repeated sequences (Bhattacharya *et al* 1998). Polymorphism has been observed on several levels. One class of repeats appears to undergo expansion and contraction that is observable in the same isolate over time *in vitro* (Bhattacharya *et al* 1992). Another region (SSG or *Tr*; figure 1C) shows variation in the number of STRs between strains but appears to be relatively stable (Clark and Diamond 1993). On a larger scale, in some isolates the episome has undergone recombinational loss of one rDNA coding region and the adjacent STR classes (Cázares *et al* 1994; Sehgal *et al* 1994; Ghosh *et al* 2001). This loss has never been observed in recent isolates from patients and is likely to have occurred during long term *in vitro* cultivation. The rDNA recombination event leads to the loss of the SSG region, making this polymorphic region unsuitable for typing certain isolates. The rDNA coding regions themselves appear to be relatively homogeneous in sequence within the species (Clark and Diamond 1991b; Newton-Sánchez *et al* 1997; Som *et al* 2000).

5. tRNA genes

A recent search for new polymorphic sequences in *E. histolytica* unearthed a number of loci suitable for strain typing, and revealed a unique gene organization at the

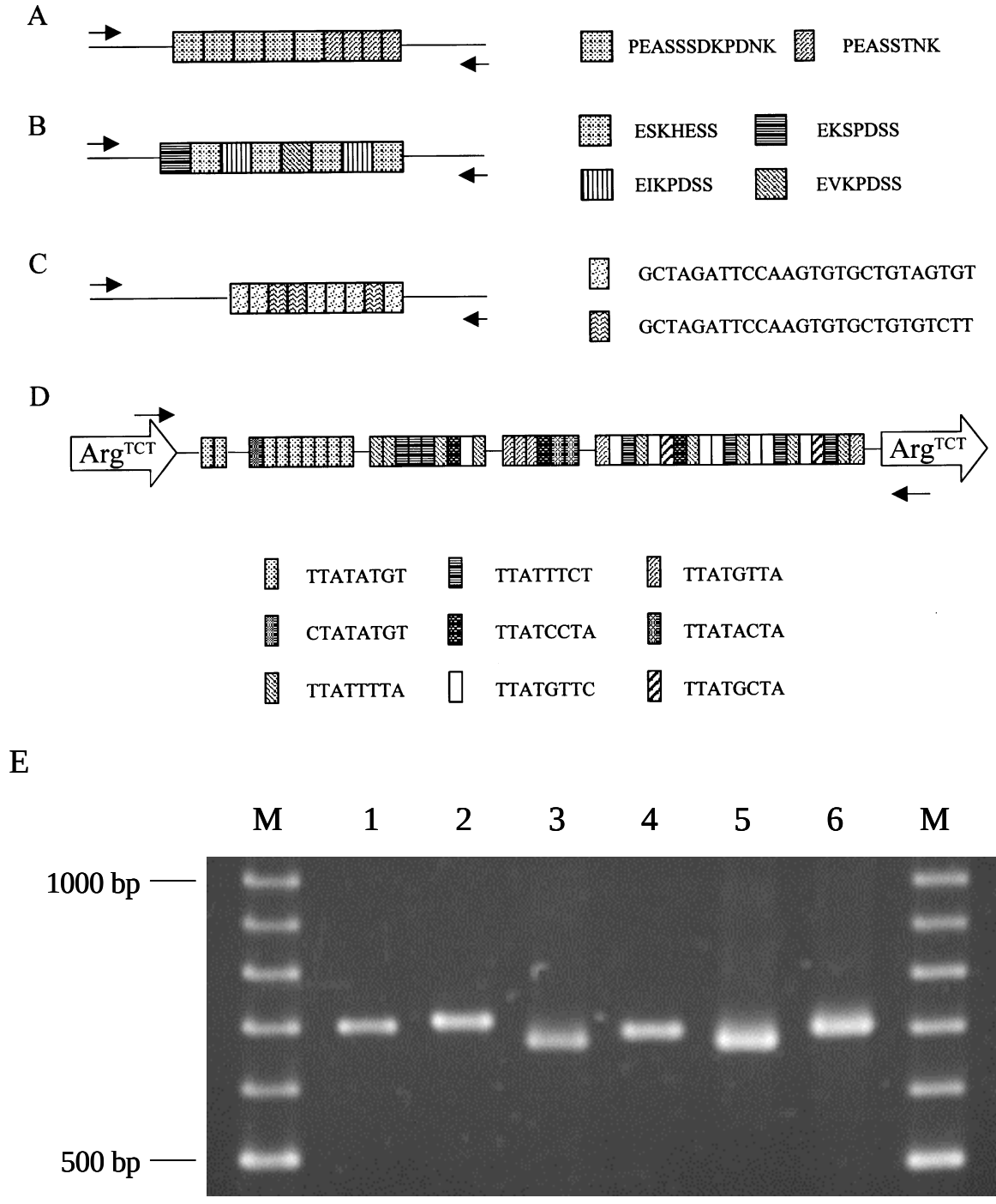

Figure 1. (**A**) Schematic representation of the SREHP locus. (**B**) Schematic representation of the chitinase locus. (**C**) Schematic representation of the SSG/*Tr* locus. In each case the sequence and number of the repeat elements in the gene from *E. histolytica* HM-1:IMSS is shown (the larger of the two allelic variants in the case of SREHP). The single letter amino acid code is used in **A** and **B**. The location of the primers used in PCR amplification is indicated by arrows. Not to scale. (**D**) Schematic representation of the Arg^{TCT} locus. The location of the primers used in PCR amplification is indicated by small arrows, the tRNA genes are indicated by large arrows, and the STRs by shaded boxes. Not to scale. (**E**) Polymorphism in the Arg^{TCT} locus. DNA was extracted from axenically grown isolates of *E. histolytica* as described in Clark and Diamond (1991b) and the Arg^{TCT} locus was amplified by PCR. Lanes, 1, HM-1:IMSS; 2, HK-9; 3, HB-301:NIH; 4, Rahman; 5, SD157; 6, DKB; M, 100 bp ladder size marker. The primers used were TCT5 (AGCATCAGCCTTCTAAGCTG) and TCT3 (CTTCCGACTGAGCTAACAAG). PCR amplification consisted of 30 cycles of 94°C, 1 min; 55°C, 1·5 min; 72°C, 2 min. PCR products were separated in a 1·3% agarose gel using Tris-Borate-EDTA electrophoresis buffer.

same time. These loci are all characterized by the presence of A + T-rich STRs that vary in both number and sequence between strains (Zaki and Clark 2001). The loci also share the characteristic of being closely linked to tRNA genes (figure 1D) and in a few cases to 5S RNA genes also (C G Clark, unpublished). In fact, almost all the tRNA genes and all of the 5S RNA genes of *E. histolytica* are found in this context. The tRNA genes are found in clusters of one to five distinct types, interspersed with STRs and these clusters are in turn repeated to form long arrays. The arrays make up a significant proportion of the genome – approximately 13% of the sequences being generated by the *E. histolytica* genome project contain tRNA genes.

While almost all tRNA genes are separated by STRs, not all of these intergenic regions exhibit detectable polymorphism (I K M Ali, unpublished). However, several loci that are highly polymorphic have already been identified (e.g. figure 1E) and a number remain to be tested. These tRNA loci are also found in *E. dispar* but sufficient sequence divergence exists to allow the design of specific primers that permit simultaneous species differentiation and strain typing (Zaki *et al* 2002).

6. Conclusions

Epidemiological studies to determine the role of parasite genetics in the outcome of infection by *E. histolytica* will require the examination of a large number of samples. For this reason, amoeba cultivation and DNA sequencing are likely to be too labour intensive and too expensive, respectively. DNA extraction directly from stool samples for PCR-based species identification is now a reality and we have recently shown that, depending on the extraction method, such samples can also be used to type strains at the tRNA-linked polymorphic loci (M Zaki, unpublished). Although PCR products of the same length need not necessarily have the same sequence, the use of multiple loci for strain typing makes it unlikely that two unrelated strains will produce the same fragment pattern. Likewise, the apparent stability of the loci makes it unlikely that two individuals infected with the same strain will yield distinct polymorphic patterns. We feel that the time is near for an investigation to be undertaken into the role of genetic diversity in the outcome of *E. histolytica* infection.

References

Anonymous 1997 *Entamoeba* taxonomy; *Bull. WHO* **75** 291–292

Ayeh-Kumi P F, Ali I M, Lockhart L A, Gilchrist C A, Petri W A Jr and Haque R 2001 *Entamoeba histolytica*: genetic diversity of clinical isolates from Bangladesh as demonstrated by polymorphisms in the serine-rich gene; *Exp. Parasitol.* **99** 80–88

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

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

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

Cázares F, Manning-Cela R and Meza I 1994 Heterogeneity of the ribosomal DNA episome in strains and species of *Entamoeba*; *Mol. Microbiol.* **12** 607–612

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

Clark C G and Diamond L S 1991b Ribosomal RNA genes of 'pathogenic' and 'nonpathogenic' *Entamoeba histolytica* are distinct; *Mol. Biochem. Parasitol.* **47** 297–302

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

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

Diamond L S and Clark C G 1993 A redescription of *Entamoeba histolytica* Schaudinn 1903 (Emended Walker 1911) separating it from *Entamoeba dispar* Brumpt 1925; *J. Euk. Microbiol.* **40** 340–344

Gathiram V and Jackson T F H G 1987 A longitudinal study of asymptomatic carriers of pathogenic zymodemes of *Entamoeba histolytica*; *S. Afr. Med. J.* **72** 669–672

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

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

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

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

Newton-Sánchez O L, Sturm-Ramírez K, Romero-Zamora J L, Santos-Preciado J I and Samuelson J 1997 High rate of occult infection with *Entamoeba histolytica* in non-dysenteric Mexican children; *Arch. Med. Res.* (*Suppl.*) **28** S311–S313

Sargeaunt P G 1987 The reliability of *Entamoeba histolytica* zymodemes in clinical diagnosis; *Parasitol. Today* **3** 40–43

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

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

Som I, Azam A, Bhattacharya A and Bhattacharya S 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

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

Zaki M and Clark C G 2001 Isolation and characterization of polymorphic DNA from *Entamoeba histolytica*; *J. Clin. Microbiol*. **39** 897–905

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