Updated information and services can be found at: http://jcm.asm.org/content/39/3/897

REFERENCES

This article cites 26 articles, 5 of which can be accessed free at: http://jcm.asm.org/content/39/3/897#ref-list-1

CONTENT ALERTS

Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: http://journals.asm.org/site/misc/reprints.xhtml
To subscribe to another ASM Journal go to: http://journals.asm.org/site/subscriptions/
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 variation has been seen in some strains, while the majority still exists a need for additional reliable polymorphic DNA markers.

Variation was seen in the number of rRNA transcription units present on the isolates tested gave the same restriction fragment pattern 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 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.

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-
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 (39.0°C) 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, 15) 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 5-200 column to remove the salts.

5'-Phosphorylated 24-mer (5'-pAGTCCGGATCCAAGCAAGGACA-3') and nonphosphorylated 20-mer (5'-CTGGTGGTGGATCCGACT-3') oligonucleotides with overlapping complementary sequences containing a BamHI 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 (GAGTCCGGTCCGGATCCAAGCAAGGACA-3') and 1 nmol of *Alu*I, *Dra*I, *Eco*I, and *Rsa*I (Gibco-BRL) 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 DNA as a control.

**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

---

**TABLE 1. Origin of *E. histolytica* isolates**

<table>
<thead>
<tr>
<th>Isolate Type</th>
<th>Origin</th>
<th>Culture type</th>
<th>Clinical information</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM-1:IMSS clone 9</td>
<td>Mexico</td>
<td>Axenic</td>
<td>Isolated from rectal ulcer; patient with dysentery</td>
</tr>
<tr>
<td>200:NIIH</td>
<td>(uncertain)</td>
<td>Axenic</td>
<td>Patient with amebic dysentery</td>
</tr>
<tr>
<td>H-303:NIH</td>
<td>VietNam (?)</td>
<td>Axenic</td>
<td>Patient with amebic dysentery</td>
</tr>
<tr>
<td>IULA:1190:2-3</td>
<td>Venezuela</td>
<td>Axenic</td>
<td>Patient with amebic colitis</td>
</tr>
<tr>
<td>8691</td>
<td>Bangladesh</td>
<td>Xenic</td>
<td>Patient with amebic colitis</td>
</tr>
<tr>
<td>4530</td>
<td>Bangladesh</td>
<td>Xenic</td>
<td>Patient with amebic colitis</td>
</tr>
<tr>
<td>1320300</td>
<td>Bangladesh</td>
<td>Xenic</td>
<td>Patient with amebic colitis</td>
</tr>
<tr>
<td>48286</td>
<td>Bangladesh</td>
<td>Xenic</td>
<td>Patient with amebic colitis</td>
</tr>
<tr>
<td>2596</td>
<td>South Africa</td>
<td>Xenic</td>
<td>Patient asymptomatic; serology positive by antigen gel diffusion test</td>
</tr>
<tr>
<td>26.25AC</td>
<td>South Africa</td>
<td>Xenic</td>
<td>Convalescent amebic liver abscess patient; serology positive by antigen gel diffusion test</td>
</tr>
<tr>
<td>37.0C</td>
<td>South Africa</td>
<td>Xenic</td>
<td>Patient asymptomatic; serology positive by antigen gel diffusion test</td>
</tr>
<tr>
<td>39.384C</td>
<td>South Africa</td>
<td>Xenic</td>
<td>Patient asymptomatic; serology positive by antigen gel diffusion test</td>
</tr>
</tbody>
</table>

**TABLE 2. Oligonucleotide primers**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Primer sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>CTG GAT TCA TTA CAG CCC TGG</td>
</tr>
<tr>
<td>R2</td>
<td>CTT ACA CCC CCA TTA ACA ATC</td>
</tr>
<tr>
<td>R3</td>
<td>GCT ATG GTG TCT GAT ATC</td>
</tr>
<tr>
<td>R4</td>
<td>CTT TAG TCA GTG TCT AA</td>
</tr>
<tr>
<td>R5</td>
<td>CTA AAG CCC CCT TCT TCT AT</td>
</tr>
<tr>
<td>R6</td>
<td>CTA AAG CCC CCT TCT TCT ATA ATT</td>
</tr>
<tr>
<td>R7</td>
<td>CTT AAC GAG GAA GTA AAG</td>
</tr>
<tr>
<td>R8</td>
<td>CTA AAG CCC CCT TCT TCT ATA ATT</td>
</tr>
<tr>
<td>R9</td>
<td>CTA CAT CTA AGC TCC GCC TGT</td>
</tr>
<tr>
<td>R10</td>
<td>CTT ACT CCT TCT AAC GAC CAC</td>
</tr>
<tr>
<td>R11</td>
<td>GTC GTT GTA AAG AGA AGT AAG</td>
</tr>
<tr>
<td>R12</td>
<td>AAG CCT GTC ACT GGT TCG AA</td>
</tr>
<tr>
<td>R13</td>
<td>ACC AAC TAC ATG GAA AAC</td>
</tr>
</tbody>
</table>

**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 24% NuSieve 3:1 agarose gels (EMD) in 1× Tris-borate-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, *Dra*I, *Eco*I, 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-32P]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× 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 and 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 AF275005 to AF275065.
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 AluI or RsaI 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 AluI 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 19 and clone 5) derived from AluI restriction fragments annealed to (TAA)$_{12}$ contained the same type of repeats as clone 4. Clones 19 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 product of clone 1 (representing locus 5-6) was sequenced and compared to the sequences of other clones representing the same locus. The results showed that the polymorphism is due to the number of tandem repeats in the repeat block. The tandem repeat blocks in clones 19 and 5 contained 8, 11, and 10 repeats, respectively, while clone 4 contained 12 repeats. The positions of the tandem repeats were determined by sequencing the PCR products generated by primers R5 and R6 (Table 2). 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. 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.

**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).
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:1092:1.

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 (GTATGTTCTAT). 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. 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 39.38C (South Africa).

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.
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 AluI lane (Fig. 5B) was expected since the clone was obtained from an AluI restriction fragment of about the same size. It was a surprise to find that the probe gave intense hybridization signals at \( \sim 23 \text{ kb} \) with the DraI-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 AluI, DdeI, and RsaI give major fragments of about the same size and that DraI and EcoRI both give very large fragments.

Similarly, a band of ca. 400 bp was expected in the AluI lane with the locus 5-6 specific probe (Fig. 5C), since the clone was obtained from an AluI restriction fragment of about this size. Once again hybridization with this probe gave major fragments of about the same size with AluI and DdeI and with DraI and RsaI. EcoRI again produces a very large fragment of \( \geq 23 \text{ kb} \).

**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 \(^{32}\text{P}\)-labeled probe. (C) Blot of gel in panel A hybridized with a locus 5-6 specific \(^{32}\text{P}\)-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).

**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 A265348; 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 CTATTATA tandem repeats of locus 9-4 differ from the 11 CTTATATA 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...
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 DraI and/or EcoRI 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-
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.

---

**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-1:1MSS (Mexico); 200:NIH (uncertain); H-303:NIH (VietNam); IULA:1092:1 and IULA:30593:2 (Venezuela); 8691, 4530, 1320300, and 48286 (Bangladesh); 2596, 26.253, 37.0C, and 39.384C (South Africa).
REFERENCES


