JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2000, p. 3815–3821 Vol. 38, No. 10
0095-1137/00/$04.00+0

Molecular Epidemiology of *Entamoeba* spp.: Evidence of a Bottleneck (Demographic Sweep) and Transcontinental Spread of Diploid Parasites

SUDIP GHOSH,[1] MARTA FRISARDI,[1] LYNN RAMIREZ-AVILA,[1] STEVEN DESCOTEAUX,[1] KATHERINE STURM-RAMIREZ,[1] OSCAR ALBERTO NEWTON-SANCHEZ,[2] JOSE IGNACIO SANTOS-PRECIADO,[2]† CHAITI GANGULY,[3] ANURADHA LOHIA,[3] SHARON REED,[4] AND JOHN SAMUELSON[1]*

Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts[1]; Division of Infectious Disease, Hospital Infantil, Mexico City, D.F., Mexico[2]; Department of Biochemistry, Bose Institute, Calcutta, India[3]; and Department of Medicine and Pathology, UCSD Medical Center, San Diego, California[4]

Received 21 December 1999/Returned for modification 25 May 2000/Accepted 9 June 2000

Entamoeba histolytica **causes amebic colitis and liver abscess in developing countries such as Mexico and India.** ***Entamoeba dispar*** **is morphologically identical but is not associated with disease. Here we determined the ploidy of** ***E. histolytica*** **and developed PCR-based methods for distinguishing field isolates of** ***E. histolytica*** **or** ***E. dispar*****. Fluorescence in situ hybridization showed that** ***E. histolytica*** **trophozoites are diploid for five "single-copy" probes tested. Intergenic sequences between** ***superoxide dismutase*** **and** ***actin 3*** **genes of clinical isolates of** ***E. histolytica*** **from the New and Old Worlds were identical, as were those of** ***E. dispar*****. These results suggest a bottleneck or demographic sweep in entamoebae which infect humans. In contrast,** ***E. histolytica*** **and** ***E. dispar*** **genes encoding repeat antigens on the surface of trophozoites (Ser-rich protein) or encysting parasites (chitinase) were highly polymorphic.** ***chitinase*** **alleles suggested that the early axenized strains of** ***E. histolytica*****, HM-1 from Mexico City, Mexico, and NIH-200 from Calcutta, India, are still present and that similar** ***E. dispar*** **parasites can be identified in both the New and Old Worlds.** ***Ser-rich protein*** **alleles, which suggested the presence of the HM-1 strain in Mexico City, included some** ***E. histolytica*** **genes that predicted Ser-rich proteins with very few repeats. These results, which suggest diversifying selection at** ***chitinase*** **and** ***Ser-rich protein*** **loci, demonstrate the usefulness of these alleles for distinguishing clinical isolates of** ***E. histolytica*** **and** ***E. dispar*****.**

Entamoeba histolytica is a protozoan parasite that causes amebic colitis and liver abscess in developing countries such as Mexico, India, and Bangladesh (6, 17, 24, 32). The HM-1 strain of *E. histolytica*, which was isolated from a dysenteric patient in Mexico more than 30 years ago, still causes disease in experimental animals and has been used for nearly all immunological, biochemical, and molecular biological studies of amebae (14). *E. histolytica* is morphologically indistinguishable from *Entamoeba dispar*, which remains in the colonic lumen and so does not cause disease (8, 39, 44). Sequences of *E. histolytica* and *E. dispar* small-subunit rRNA genes differ by 1.7%, suggesting that the parasites diverged from each other tens of millions of years ago (9, 29, 40).

The haploid genome of *E. histolytica* contains 14 chromosomes and totals ~20 Mb (47). Homologous chromosomes vary in length and number (from one to four), suggesting that parasites may be polyploid (47). *E. histolytica* coding regions are AT rich, contain few introns, and are separated by relatively short intergenic regions (4, 22, 43). 5′ untranslated regions of amebic genes contain conserved TATA, CAAT, GAAC, and initiation sequences (5, 41). Two palindromic copies of each rRNA gene are present on 24-kb episomal plasmids, which are present in 100 to 200 copies per nucleus (40).

The relationships between different clinical isolates of *E. histolytica* are not known. No monoclonal antibodies have been able to distinguish isolates of *E. histolytica* (17). Isoenzyme groups or zymodemes have not proven useful for differentiating strains of *E. histolytica*, as the majority of strains fall into two main zymodeme groups (3, 39). In addition, sequences of the genes encoding hexokinase and phosphoglucomutase, which were used to distinguish parasites, failed to reveal any heterogeneity among *E. histolytica* and *E. dispar* isolates, respectively (30, 31). Although sequences of internal transcribed spacers (ITS) between rRNA genes discriminate bacterial isolates and strains of *Leishmania* spp., ITS sequences failed to distinguish isolates of *E. histolytica* or isolates of *E. dispar* from Mexico City, Mexico (11, 21, 28). In contrast, axenic strains of *E. histolytica* have been distinguished by restriction fragment length polymorphisms of PCR products of the genes encoding the Ser-rich *E. histolytica* protein, also known as the K2 protein (10, 20, 42). The Ser-rich protein, which is present on the surface of amebae, is composed of an N-terminal signal sequence and a hydrophobic C-terminal anchor, surrounding a series of tetrapeptide and octapeptide repeats of hydrophilic and acidic amino acids. The Ser-rich protein is an important amebic vaccine candidate, and antibodies to the Ser-rich protein correlate with infection (27, 46, 49). The amebic chitinase, which is expressed only by cysts, also contains a series of acidic and antigenic repeats between a putative N-terminal lectin domain and a C-terminal half catalytic domain (12).

* Corresponding author. Mailing address: Department of Immunology and Infectious Diseases, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115. Phone: (617) 432-4670. Fax: (617) 738-4914. E-mail: jsamuels@hsph.harvard.edu.

† Present address: The National Immunization Council and The National Child Health Program, CONAVA, Mexico City, D.F., Mexico.

While little is known about the molecular epidemiology of amebae, much has been learned from numerous excellent molecular epidemiological studies of *Plasmodium falciparum*, the cause of severe malaria. First, there was apparently a recent bottleneck in the evolution of *P. falciparum* infecting humans, such that parasites from all over the world are identical in sequences of genes encoding housekeeping proteins (34). This bottleneck may have occurred as recently as 7,000 years ago. Second, malaria genes encoding repeat antigens on the parasite surface (e.g., circumsporozoite protein, merozoite surface protein 2, or merozoite S antigen) are extraordinarily polymorphic (2, 16, 26). Third, most of the diversity of *circumsporozoite protein* genes is secondary to DNA slippage rather than meiotic recombination, while maintenance of variation is likely caused by immune selection (33).

To better understand the genetics and molecular epidemiology of *E. histolytica* and *E. dispar*, we asked three questions in these studies. First, what is the ploidy of *E. histolytica*? Second, can intergenic sequences between *superoxide dismutase* and *actin 3* (*sod-actin*) genes be used to distinguish isolates of *E. histolytica* and *E. dispar* from Mexico City, San Diego, Calif., and Calcutta, India? Third, can isolates of *E. histolytica* and *E. dispar* be distinguished by sequences of genes that encode chitinase and the Ser-rich protein?

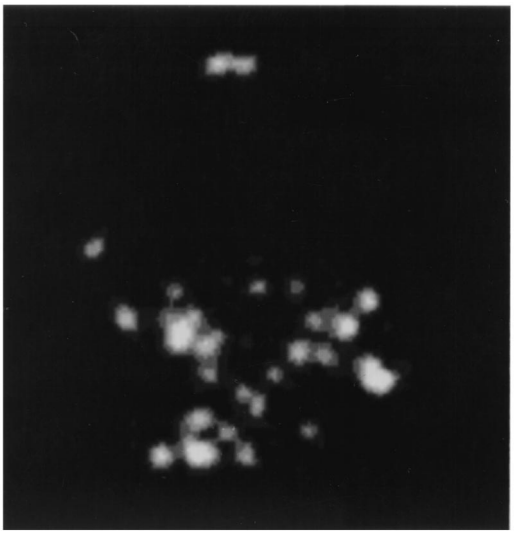

FIG. 1. Fluorescence micrograph of a colchicine-treated *E. histolytica* trophozoite stained with propidium iodide. Condensed chromosomes number 28, which is twice the number of chromosomes (14) identified on pulsed-field gels.

MATERIALS AND METHODS

Fluorescence in situ hybridization (FISH) of *E. histolytica* trophozoites. The HM-1 strain of *E. histolytica* was grown axenically at 37°C in TYI medium. For production of parasites with condensed chromosomes, amebae were incubated in 7 mg of colchicine per ml for 12 h, fixed in methanol-acetic acid (3:1), treated with 20 μg of RNase per ml for 30 min, and stained with 0.3 μg of propidium iodide per ml in Antifade (Oncor). For FISH, parasites were washed in phosphate-buffered saline, swelled in 70 mM potassium chloride, fixed in methanol-acetic acid, and dropped onto uncoated slides. DNA was denatured by dipping slides in 70% formamide at 70°C. *E. histolytica* genes encoding chitinase, pyruvate:ferredoxin oxidoreductase, nicotinamide nucleotide transhydrogenase, plasma membrane calcium-transporting ATPase, cysteine proteinase 5, and P-glycoprotein 6 genes were labeled by nick translation with biotin using an Oncor kit (12, 13, 15, 18, 37, 48). A negative control was pBluescript without an insert. Hybridizations were performed with 4 μg of each biotinylated probe per ml in 30% formamide and 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 37°C for 16 h (25). Slides were washed in the same buffer, 2× SSC, and phosphate buffer plus detergent (Oncor). Probes were detected with fluorescein isothiocyanate-avidin and amplified with antiavidin antibody, followed by a second incubation with fluorescein isothiocyanate-avidin. Parasite nuclei were counterstained with propidium iodide, and slides were examined with a Leitz Orthoplan epifluorescence microscope or a Leica confocal microscope.

Isolation of amebic DNA from clinical isolates of *E. histolytica* and *E. dispar*. Stools containing amebae as shown by light microscopy came from patients presenting to (i) the Hospital Infantil in Mexico City or a neighborhood clinic in Netzahualcoyotl, which is a barrio of 100,000 persons outside Mexico City; (ii) the Kothari Medical Center in Calcutta; or (iii) the University of California at San Diego. Other parasite DNA came from axenized *E. histolytica* (HM-1, HK-9, and NIH-200) and *E. dispar* (SAW 760) strains. For the most part, amebae from clinical isolates were cultured in Robinson's medium, concentrated by low-speed centrifugation, washed in phosphate-buffered saline, and lysed in 1% sodium dodecyl sulfate–50 mM EDTA–50 mM Tris, pH 8 (1, 35). Amebic DNA was then extracted with glass milk (Elutip; Schleicher and Schuell) (36). On some occasions, DNA was isolated directly from stool parasites, which were enriched by low-speed centrifugation and lysed in the same buffer (19). Identification of isolates as *E. histolytica* or *E. dispar* was made by restriction fragment length polymorphism analysis of PCR products using primers that spanned the ITS between rRNA genes (28). The *E. histolytica* rRNA ITS PCR product was cut with *Rsa*I, while *E. dispar* rRNA ITS PCR products were cut with *Eco*RV.

PCR amplification and sequencing of *sod-actin*, *chitinase*, and *Ser-rich protein* genes. The intergenic regions between amebic *sod-actin* genes of *E. histolytica* and *E. dispar* were amplified using PCR and the same set of primers (see Fig. 1) (4). A sense PCR primer (TTGGTGGAATGTAGTCAACTG) was located at the 3′ end of the coding region of the *superoxide dismutase* gene. An antisense primer (AAATCCGGCTTTACACATTCC) bound to a sequence located at the 5′ end of the coding region of the *actin 3* gene. The amebic *chitinase* gene repeats were amplified using PCR and the same antisense primer (TCTGTATTGTGC CCAATT) for both *E. histolytica* and *E. dispar* (12). An *E. histolytica chitinase*-specific sense primer was GGAACACCAGGTAAATGTATA. An *E. dispar chitinase*-specific sense primer was GGAACACCAGGTAAATGCCTT. The amebic *Ser-rich protein* gene repeats were amplified using PCR and primers specific for *E. histolytica* and *E. dispar*, respectively. An *E. histolytica Ser-rich protein*-specific sense primer was GCTAGTCCTGAAAAGCTTGAAGAAGC TG, while an *E. histolytica Ser-rich protein*-specific antisense primer was GGAC TTGATGCAGCATCAAGGT (20, 42). An *E. dispar Ser-rich protein*-specific sense primer was AGATACTAAGATTTCAGTC, while an *E. dispar*-specific *Ser-rich protein* antisense primer was CATAATGAAAGCAAAGAG (20).

The PCR products of cultured parasites were identified on agarose gels and sequenced without cloning, using *Taq* polymerase and cycle sequencing. For some *Ser-rich protein* gene analyses, DNA was extracted with glass milk from children's stools at the Hospital Infantil, and PCR was performed with the *E. histolytica Ser-rich protein* primers described above. A second PCR was performed with the *E. histolytica Ser-rich protein*-specific nested primers (sense, GTAGCTCAGCAAAACCAGAATC; antisense, TATCGTTATCTGAACTAC TTC) (42). The nested *E. histolytica Ser-rich protein* PCR product was cloned into the TA vector (Invitrogen) and sequenced by dideoxy methods. PCR products were named for the species (*E. histolytica* [Eh] or *E. dispar* [Ed]), the source (Hospital Infantil [HI], San Diego [SD], or Kothari [K]), and the isolate number.

Methods for alignments of *Ser-rich protein* and *chitinase* gene repeats. Common algorithms for aligning sequences are not adequate for sequences containing numerous degenerate repeats. Therefore, amebic *chitinase* and *Ser-rich protein* repeat sequences were coded by methods used to compare *P. falciparum circumsporozoite* gene sequences (33). Briefly, each repeat that predicted a unique amino acid sequence was given a number. Silent changes in nucleotide sequences of each repeat were identified, and the numbers (names) of each repeat were modified by a unique underbar. Sequences were then assembled from these repeats, using two assumptions. First, only identical or nearly identical repeats were aligned. Second, gaps (indicated by dashes) were added wherever needed.

RESULTS AND DISCUSSION

Condensed chromosomes of *E. histolytica* trophozoites number 28, suggesting that amebae are diploid. *E. histolytica* trophozoites form condensed chromosomes, which are rare in untreated amebae and frequent in amebae treated with 7 mg of colchicine per ml (Fig. 1). Each trophozoite averaged 22 ± 5, with a maximum of 28 and a minimum of 15. Assuming 14 chromosomes in haploid parasites (47) and assuming some inefficiency in counting closely opposed or small chromosomes, *E. histolytica* trophozoites appear to be diploid. Diploidy is consistent with the presence of two *E. histolytica Ser-rich protein* genes and two *E. dispar Ser-rich protein* genes in cDNA libraries of each parasite and the presence of two *Ser-rich protein* PCR products in numerous axenized strains of *E. histolytica* (10, 20). In contrast, the presence of as many as four

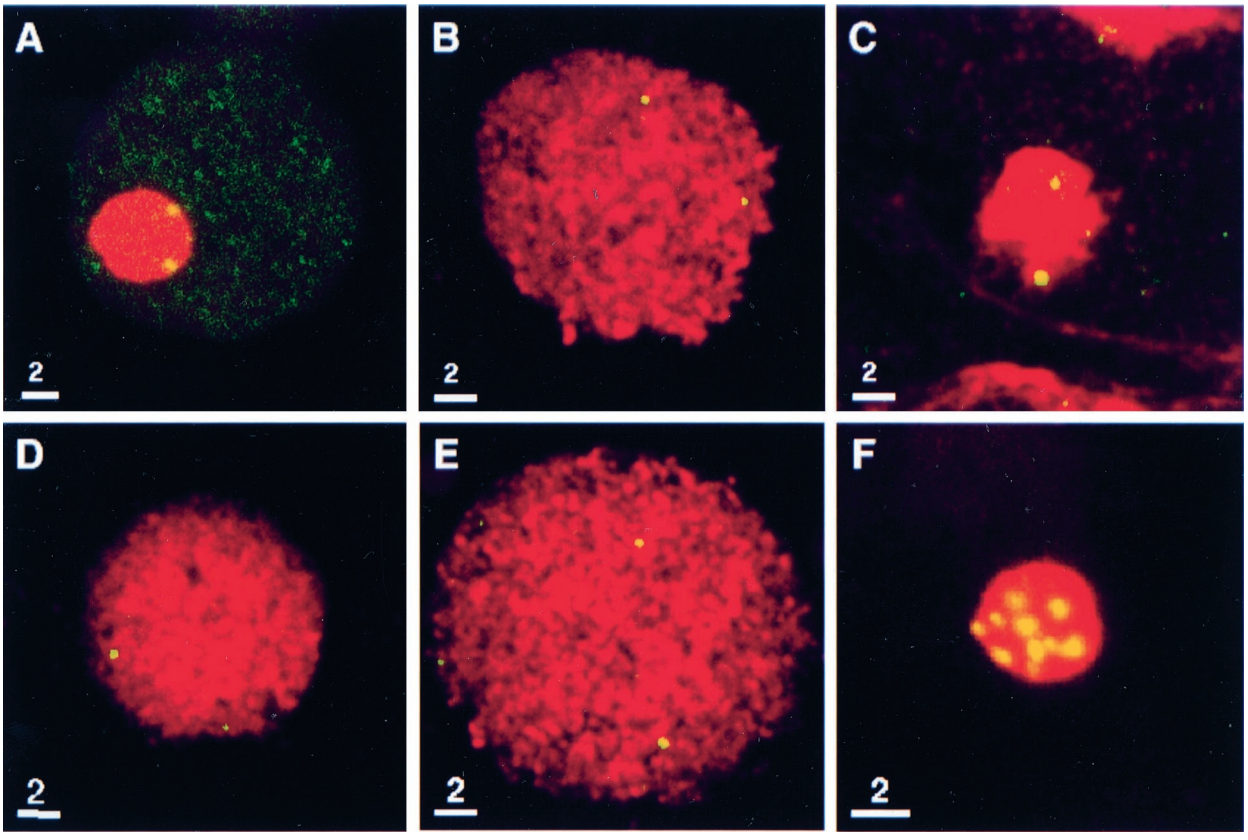

FIG. 2. Confocal micrographs of FISH of *E. histolytica* with single-copy genes encoding chitinase (A), pyruvate:ferredoxin oxidoreductase (B), nicotinamide nucleotide transhydrogenase (C), plasma membrane calcium-transporting ATPase (D), and cysteine proteinase 5 (E), which bind twice (yellow) to each nucleus. In contrast, FISH with *p-glycoprotein* genes (F), which are in multiple copies, shows numerous spots.

bands within a linkage group on Southern blots of pulsed-field gels of *E. histolytica* chromosomes suggests the possibility that amebae are tetraploid for some chromosomes (47).

***E. histolytica* is also diploid for five "single-copy" genes.** To determine the ploidy of amebae by an independent method, we performed FISH with an arbitrary set of amebic genes, which appear to be present in a single copy as shown by Southern blotting or by repetitive probing of cDNA or genomic DNA libraries. These *E. histolytica* genes encoded a diverse set of proteins involved in cyst wall destruction (chitinase), fermentation (pyruvate:ferredoxin oxidoreductase), exchange of electrons (nicotinamide nucleotide transhydrogenase), ion transport (plasma membrane calcium-transporting ATPase), or host tissue destruction (cysteine proteinase 5) (12, 15, 18, 37, 48). With all five *E. histolytica* genes, FISH showed a mixture of diploid and tetraploid parasites (Fig. 2A through E). Diploid parasites were presumably in G_1 prior to DNA synthesis, while tetraploid parasites were presumably in G_2 prior to mitosis. Negative controls with vector sequences alone showed no staining. FISH was also performed with *E. histolytica p-glycoprotein* genes, which are present in at least six copies and encode proteins associated with emetine resistance (13). As expected, *p-glycoprotein* gene probes bound numerous times to amebic trophozoites (Fig. 2F). These results suggest that the basic ploidy of *E. histolytica* is diploid, although these results do not rule out the possibility that some amebic genes will be present in more than two copies (if genes or portions of chromosomes are duplicated) or once (if a copy of the gene is deleted).

Evidence for bottlenecks or demographic sweeps in populations of *E. histolytica* and *E. dispar*. Previously, we found no differences among ITS among rRNA genes of 10 *E. histolytica* isolates (28). Here the intergenic sequences between *superoxide dismutase* and *actin 3* genes of *E. histolytica* were sequenced, because these sequences are single copy and are longer (380 nucleotides) than ITS between rRNA genes (165 nucleotides total) (4, 28). The *sod-actin* intergenic sequences of three isolates of *E. histolytica* from Mexico City, San Diego, and Calcutta were each the same as that of the HM-1 strain (Fig. 3), even though these *E. histolytica* isolates differed at *chitinase* or *Ser-rich protein* loci (see below). The *sod-actin* intergenic sequences of five isolates of *E. dispar*, which came from two continents and differed from each other at *chitinase* or *Ser-rich protein* loci, were also identical to each other. The *E. dispar sod-actin* intergenic sequences differed in 83 nucleotides (22%) from those of *E. histolytica*. Conserved in the *E. histolytica* and *E. dispar* sequences was a TATA-like box upstream from the *actin 3* coding region (5, 41). These results, which are consistent with the idea of a bottleneck or demographic sweep like that described previously for *P. falciparum* (34), indicate that human infections with each *Entamoeba* species derived from a single organism or from an identical group of organisms. Although these amebic bottlenecks were recent relative to the divergence of *E. histolytica* and *E. dispar*, how recent is not clear. With a great deal more sequence data available for *P. falciparum*, a bottleneck as recent as 7,000 years ago has been argued for (34).

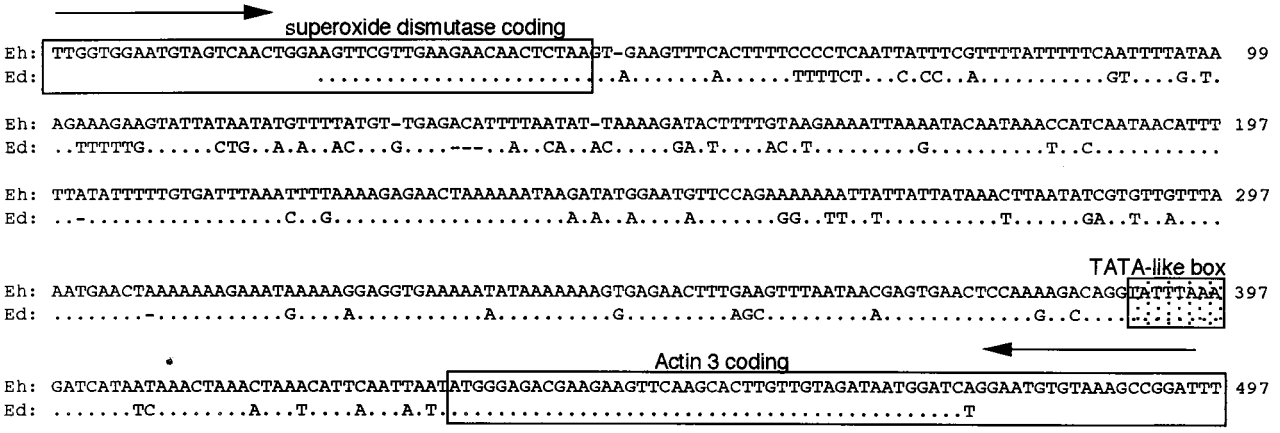

FIG. 3. Alignment of the intergenic sequences between *superoxide dismutase* and *actin 3* genes of *E. histolytica* and *E. dispar*. Periods indicate identity of *E. dispar* with *E. histolytica*, and dashes indicate gaps. Unshaded boxes indicate *superoxide dismutase* and *actin* gene coding regions, and the dotted box indicates the TATA-like sequence upstream of the start codon of *actin 3*, while arrows indicate locations of PCR primers. The *E. histolytica* sequence was identical to that reported previously with GenBank accession no. X70852 (40).

***chitinase* alleles tentatively identify *E. histolytica* HM-1 strain amebae in Mexico City and NIH-200 strain amebae in Calcutta.** Primers flanking the *E. histolytica chitinase* gene repeats produced a single PCR product from each clinical isolate (data not shown). This result suggests that the amebic *chitinase* genes are homozygous, as parasites are diploid at this locus (Fig. 2A). The *chitinase* PCR products were sequenced without cloning, and the sequences of the repeats were coded by methods used to compare *P. falciparum circumsporozoite protein* gene sequences (Fig. 4) (33). The *E. histolytica chitinase* gene repeats ranged from 84 to 252 nucleotides and so predicted chitinases with four heptapeptide repeats (28 amino acids) to 12 heptapeptide repeats (84 amino acids). The 168-nucleotide *chitinase* gene repeat of a San Diego isolate (Eh SD1) was identical to that of the HM-1 strain, while the 84-nucleotide *chitinase* gene repeat of a Calcutta isolate (Eh K1) was the same as those of NIH-200 amebae (12). These results, which suggest that the original axenized strains of *E. histolytica* may still be located in the New and Old Worlds, should be reassuring to bench scientists, who study these amebae. Because the amebae for the present studies came from individuals with amebic cysts in their stools, it is likely that *E. histolytica* chitinases with varying numbers of heptapeptide repeats are equally functional. This conclusion is consistent with the idea that heptapeptide repeats are spacers between lectin and catalytic domains of amebic chitinases (12).

The *E. histolytica chitinase* gene repeats were remarkable for the paucity of different heptapeptide sequences encoded (four) and their rigid and idiosyncratic codon usage (Fig. 4). For example, each heptapeptide repeat started with Glu, contained a Lys residue, and ended with two Ser residues. Heptapeptide EIKPDSS differed from EVKPDSS by a single, nonsilent point mutation. Glu was encoded by GAG in the first heptapeptide repeat of each *chitinase* repeat, while Glu was encoded by GAA in all other heptapeptide repeats. Ser in all heptapeptide repeats was encoded by TCT, which is used in only 23% of Ser residues in nonrepeat sequences of the *E. histolytica chitinase* gene (12). Similarly, Asp was encoded by GAC, which is infrequent in nonrepeat portions of the amebic *chitinase* gene and in other amebic coding sequences (43). In the absence of flanking sequences, it cannot be determined whether the *chitinase* gene diversity was generated by meiotic recombination or by DNA slippage, as shown previously for *P. falciparum circumsporozoite protein* genes (33).

***E. dispar chitinase* gene polymorphisms suggest transcontinental spread of parasites.** Primers flanking the *E. dispar chitinase* gene repeats produced a single PCR product from each clinical isolate (data not shown). Eleven different *E. dispar chitinase* gene alleles were found in 25 clinical isolates of *E. dispar* from Mexico City, San Diego, and Calcutta (Fig. 4). The average number of *E. dispar chitinase* gene repeats (16 ± 2) was greater than those of *E. histolytica* (8 ± 4; $P < 0.01$). This result suggests that forces which cause diversifying selection at the *chitinase* locus are at least as active against *E. dispar* parasites. Five different *chitinase* gene repeats, which predicted 12 to 18 heptapeptides, were identified among PCR products of 10 *E. dispar* isolates from Mexico City. Interestingly, one of these Mexican *E. dispar chitinase* alleles (Ed HI1) was also identified in an *E. dispar* isolate from San Diego, the *E. dispar* cDNA library made by Egbert Tannich (SAW 760), and the strain of *E. dispar* axenized by Graham Clark (SAW 1734) (7, 44). These results suggest that this *E. dispar* strain may have traveled between the New World and the Old World, as both SAW 760 and SAW 1734 were isolated from Ethiopian Jews in Israel (David Mirelman, personal communication). A second *E. dispar* Mexican *chitinase* gene allele (Ed HI5) was identified six times in PCR products of 10 *E. dispar* isolates from Calcutta (Ed K1), suggesting the transcontinental movement of this *E. dispar* strain as well. These studies, however, cannot determine the direction of travel of the parasites.

The major difference between the predicted chitinase repeats of the North American *E. dispar* isolates was in the number of EIKPDSS blocks, while the major difference between predicted chitinase repeats of the Asian *E. dispar* isolates was in the number of EVKDSS blocks (Fig. 4). A second DTKPDSS repeat present in North American *E. dispar* chitinase sequences was absent from Asian *E. dispar* chitinase sequences. These results suggest the possibility that strains may be tentatively identified as North American type or Asian type by the patterns of their *chitinase* gene repeats.

***Ser-rich protein* gene polymorphisms among clinical isolates of *E. histolytica* suggest the persistence of the early axenized strain in Mexico City.** Primers flanking the *E. histolytica* and *E. dispar Ser-rich protein* gene repeats frequently produced two PCR products on agarose gels from one clinical isolate (data not shown). This result suggests that the amebae are often heterozygous at the *Ser-rich protein* locus, as has been shown previously for axenic strains of *E. histolytica* and the SAW 760

A.

```
     E  K  L  P  D  S  S
1 = GAGAAGTTACCAGACTCTTCT

     D  T  K  P  D  S  S
2 = GATACTAAACCAGATTCTTCT

     E  S  K  H  E  S  S
3 = GAATCTAAACATGAATCTTCT
3 = GAGTCTAAACATGAATCTTCT

     E  I  K  P  D  S  S
4 = GAAATTAAACCAGATTCTTCT

     D  C  K  L  D  S  S
5 = GATTGTAAACTAGACTCTTCT

     D  C  K  P  D  S  S
6 = GACTGTAAACCAGATTCTTCT

     E  V  K  P  D  S  S
7 = GAAGTTAAACCAGATTCTTCT
7 = GAAGTCAAACCAGATTCCTCT

     E  K  S  P  D  S  S
8 = GAGAAGTCACCAGATTCTTCT
```

B.

```
HM-1     83437343----
Eh SD1   83437343----
Eh SD2   834373437343

NIH-200 8343--------
Eh K1   8343--------

SAW760  123244444454677------373
Ed HI1  123244444454677------373
Ed SD1  123244444-54677------373
Ed SD2  12324444--54677------373
Ed SD3  1232444---546777-----373
Ed HI2  1232444---54677------373
Ed HI3  123244----546777-----373
Ed SD4  1232444---5467-------373
Ed HI4  1232444---54-7-------373

Ed HI5  123-4-----5767777----373
Ed K1   123-4-----5767777----373
Ed K2   123-4-----57677777777373
Ed K3   123-4-----576777777--373
```

FIG. 4. Patterns of *E. histolytica* and *E. dispar chitinase* repeats. (A) Building blocks of *chitinase* repeats. Each 21-nucleotide sequence, which encodes a unique heptapeptide repeat, is given a number. The numbers are marked to indicate silent nucleotide changes, which are underlined. (B) Patterns of *chitinase* repeats. Shown are the *chitinase* PCR products from clinical isolates of *E. histolytica* (Eh) and *E. dispar* (Ed) in Mexico City (HI), San Diego (SD), and Calcutta (K). Each PCR product is coded using the numbers in panel A, while gaps are indicated by dashes.

strain used to make the *E. dispar* cDNA library (10, 20). Because only one *Ser-rich protein* PCR product was usually obtained from each clinical isolate (indicating that one *Ser-rich protein* PCR product was lost), we were unable to determine

A.

```
E  K  A  S  S  S  D  N  S
GAAAAAGCAAGTTCAAGTGATAACTCA 1

E  S  S  S  S  D  K  P
GAATCAAGCTCAAGTGATAAACCA 2

D  N  K  P
GATAATAAACCA 3

E  A  S  S  S  D  K  P
GAAGCAAGTTCAAGTGATAAACCA 4
GAAGCAAGCTCAAGTGATAAACCA 4
GAAGCAAGCTCAAGTGATAAGCCA 4
GAAGCAAGTTCAAGTGATAAGCCA 4

E  A  S  S  T  N  K  P
GAAGCAAGCTCAACTAATAAACCA 6

E  A  S  S  T  S  N  S
GAAGCAAGCTCAACTAGTAATTCA 7

E  A  S  S  N  D  K  S
GAAGCAAGTTCAAACGATAAGTCA 5

D  D  K  V
GATGATAAAGTA 8

E  E  S  S  S  D  K  P
GAAGAAAGTTCAAGTGATAAACCA 9
GAAGAAAGTTCAAGTGATAAGCCA 9

E  A  S  S  A  T  N  S
GAAGCAAGCTCAGCTACTAATTCA 0
```

B.

```
HM-1(1)   12344343434343666 67
HM-1(2)   123443434343-----67
Eh HI1    123443434343-----67
Eh HI2    123---------------7
Eh HI3    123-------------667

SAW760(1)  5844-34-4439344--343430
SAW760(2)  584443494439344-----343430
Ed HI1     584-43494439344--434343430

Ed K1      584443494439344-----343430
Ed K2      584--34-44393444434343430
Ed K3      584-434944393---434343430
```

FIG. 5. Patterns of *E. histolytica* and *E. dispar Ser-rich protein* repeats. (A) Building blocks of *Ser-rich protein* repeats. Each 24-nucleotide sequence, which encodes a unique octapeptide repeat, or 12-nucleotide sequence, which encodes a unique tetrapeptide repeat, is given a number. The numbers are marked to indicate silent nucleotide changes, which are underlined. (B) Patterns of *Ser-rich protein* repeats. Shown are *Ser-rich protein* PCR products from clinical isolates of *E. histolytica* (Eh) and *E. dispar* (Ed) in Mexico City (HI) and Calcutta (K). Each PCR product is coded using the numbers in panel A, while gaps are indicated by dashes.

the linkage between *chitinase* gene and *Ser-rich protein* gene polymorphisms. The *Ser-rich protein* PCR product from one Mexican *E. histolytica* isolate (Eh HI1) matched one of the cloned *Ser-rich protein* genes of HM-1 parasites (Fig. 5). Two other isolates of *E. histolytica* from stools of children in Mexico City showed *Ser-rich protein* alleles, which encoded proteins with few (four to six) tetrapeptide and octapeptide repeats.

Multiple clinical isolates of *E. dispar* contained varying patterns of *Ser-rich protein* repeats, consistent with diversifying selection at this locus in noninvasive parasites (Fig. 5). These *Ser-rich protein* repeats differed from each other according to rules which were not quite so rigid as those of *chitinase* repeats. As was the case with *chitinase* gene repeats, the average number of *Ser-rich protein* repeats of *E. dispar* strains (21 ± 2) was greater than that of *E. histolytica* strains (8 ± 5; $P < 0.01$). Because the function of the Ser-rich protein is not known, the significance of these differences in Ser-rich protein repeat number is not clear.

It is possible that diversifying selection of Ser-rich proteins is immunologically mediated, as has been argued previously for malaria surface antigens (2, 16, 26). The tetrapeptide and octapeptide repeats of the Ser-rich protein are targets of human B-cell activation and anti-amebic antibody production, and animals immunized with the Ser-rich protein are subsequently resistant to amebic challenge of the liver (27, 46, 49). Other evidence for human immunity to *E. histolytica* parasites is the correlation between noninvasive disease and antibodies to certain epitopes of the Gal/GalNAc lectin, which is another important amebic vaccine candidate (23).

Implications of these data for genetic and molecular epidemiological studies of amebae. The data here are too fragmentary for us to make any strong conclusions concerning populations of amebae which infect humans. This is particularly the case for *E. histolytica* clinical isolates. Still, a few implications of this data are worth noting. First, the *E. histolytica* genome appears to be diploid like those of most other eukaryotes. Whether amebae have a cryptic haploid stage and sex or reproduce clonally remains to be determined (45). Second, the HM-1 strain, which is the *E. histolytica* genome sequencing project strain, is likely identical in most aspects to other *E. histolytica* strains (*sod-actin* data). Although there is no guarantee that genes have not been lost from the HM-1 strain, it is reassuring that these parasites still cause lesions in animal models (23, 49). A possible explanation for the demographic sweep or bottleneck in human populations of *E. histolytica* is that amebae, like *P. falciparum*, predominantly infect humans and do not infect other mammalian hosts. Third, it appears that *Ser-rich protein* and *chitinase* alleles may be used to discriminate isolates of *E. histolytica* or *E. dispar*, at least in big cities, where these studies were performed (38). As parasites were homozygous at the *chitinase* locus, *chitinase* alleles may be easier to work with than *Ser-rich protein* alleles. Fourth, it appears that the HM-1 strain of *E. histolytica*, which was isolated more than 30 years ago from Mexico City, is still there and in San Diego (*chitinase* and *Ser-rich protein* data) (14). Previously, the pattern of the *Ser-rich protein* repeats was shown to remain constant during 30-plus years of culture (10). Fifth, *chitinase* gene alleles suggest that some *E. dispar* parasites have spread between the Old and New Worlds, although the direction of spread cannot be determined.

ACKNOWLEDGMENTS

This work was supported in part by grants from the NIH (AI-33492 and GM-31818 [J.S.] and AI-28035 and DK-35108 [S.R.]), the US-Mexico Foundation for Science (J.S.), the US-India Fund (J.S. and A.L.), CONACYT (J.I.S.P.) and CSIR (A.L.).

We thank K. N. Jalan of the Kothari Institute in Calcutta for cultures of *E. histolytica* and *E. dispar*. We thank Graham Clark of the London School of Tropical Medicine and Hygiene for frozen pellets of axenized amebae. We acknowledge the expert technical support of Jean Lai of the Harvard School of Public Health for confocal microscopy.

REFERENCES

1. **Acuna-Soto, R., J. Samuelson, P. de Gerolami, L. Zarate, G. Schoolnik, F. Millan-Velasco, and D. Wirth.** 1992. Application of polymerase chain reaction to the epidemiology of pathogenic and nonpathogenic *Entamoeba histolytica*. Am. J. Trop. Med. Hyg. **48:**58–70.
2. **Anders, R. F., D. J. McColl, and R. L. Coppel.** 1993. Molecular variation in *Plasmodium falciparum*: polymorphic antigens of asexual erythrocyte stages. Acta Trop. **53:**239–253.
3. **Blanc, D., and P. G. Sargeaunt.** 1991. *Entamoeba histolytica* zymodemes: exhibition of δ and γ bands only of glucose phosphate isomerase and phosphoglucomutase may be influenced by starch content in the medium. Exp. Parasitol. **72:**87–90.
4. **Bruchhaus, I., M. Leippe, C. Lioutas, and E. Tannich.** 1994. Unusual gene organization in the protozoan parasite *Entamoeba histolytica*. DNA Cell Biol. **12:**925–933.
5. **Bub, H., C. Lioutas, S. Dobinsky, R. Nickel, and E. Tannich.** 1995. Analysis of the 170-kDa lectin gene promoter of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **72:**1–10.
6. **Caballero-Salcedo, A., M. Viveros-Rogel, B. Salvatierra, R. Tapia-Conyer, J. Sepulveda-Amor, G. Gutierrez, and L. Ortiz-Ortiz.** 1994. Seroepidemiology of amebiasis in Mexico. Am. J. Trop. Med. Hyg. **50:**412–419.
7. **Clark, C. G.** 1995. Axenic cultivation of *Entamoeba dispar* Brumpt 1925, *Entamoeba insolita* Geiman and Wichterman 1937 and *Entamoeba ranarum* Grassi 1879. J. Eukaryot. Microbiol. **42:**590–593.
8. **Clark, C. G.** 1998. Amoebic disease. *Entamoeba dispar*, an organism reborn. Trans. R. Soc. Trop. Med. Hyg. **92:**361–364.
9. **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.
10. **Clark, C. G., and L. S. Diamond.** 1993. *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. **77:**450–455.
11. **Cupolillo, E., G. Grimaldi, Jr., H. Momen, and S. M. Beverley.** 1995. Intergenic region typing (IRT): a rapid molecular approach to the characterization and evolution of *Leishmania*. Mol. Biochem. Parasitol. **73:**145–155.
12. **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.
13. **Descoteaux, S., P. Ayala, J. Samuelson, and E. Orozco.** 1995. Increase in mRNA of multiple *Eh pgp* genes encoding P-glycoprotein homologues in emetine-resistant *Entamoeba histolytica* parasites. Gene **164:**179–184.
14. **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.
15. **Ghosh, S. K., B. Rosenthal, R. Rogers, and J. Samuelson.** 2000. Vacuolar localization of an *Entamoeba histolytica* homologue of the plasma membrane calcium-transporting ATPase (PMCA). Mol. Biochem. Parasitol. **108:**125–130.
16. **Gupta, S., K. Treenholme, R. M. Anderson, and K. P. Day.** 1994. Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. Science **263:**961–963.
17. **Haque, R., A. S. Faruque, P. Hahn, D. M. Lyerly, and W. A. Petri, Jr.** 1997. *Entamoeba histolytica* and *Entamoeba dispar* infection in children in Bangladesh. J. Infect. Dis. **175:**734–736.
18. **Jacobs, T., I. Bruchhaus, T. Dandekar, E. Tannich, and M. Leippe.** 1998. Isolation and molecular characterization of a surface-bound proteinase of *Entamoeba histolytica*. Mol. Microbiol. **27:**269–276.
19. **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.
20. **Kohler, 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.
21. **Kostman, J. R., M. B. Alden, M. Mair, T. D. Edlind, J. J. LiPuma, and T. L. Stull.** 1995. A universal approach to bacterial molecular epidemiology by polymerase chain reaction ribotyping. J. Infect. Dis. **171:**204–208.
22. **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.
23. **Lotter, H., T. Zhang, K. B. Seydel, S. L. Stanley, Jr., and E. Tannich.** 1997. Identification of an epitope on the *Entamoeba histolytica* 170-kD lectin conferring antibody-mediated protection against invasive amebiasis. J. Exp. Med. **185:**1793–1801.

24. **Martinez-Palomo, A., and M. Martinez-Baez.** 1983. Selective primary health care: strategies for control of disease in the developing world. X. Amebiasis. Rev. Infect. Dis. **5:**1093–1102.
25. **McNeil, J. A., C. V. Johnson, K. C. Carter, R. H. Singer, and J. B. Lawrence.** 1991. Localizing DNA and RNA within nuclei and chromosomes by fluorescence in situ hybridization. Genet. Anal. Tech. Appl. **8:**41–58.
26. **Miller, L. H., T. Roberts, M. Shahabuddin, and T. McCutchan.** 1993. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol. Biochem. Parasitol. **59:**1–14.
27. **Myung, K., D. Burch, T. F. H. G. Jackson, and S. L. Stanley, Jr.** 1992. Serodiagnosis of invasive amebiasis using a recombinant *Entamoeba histolytica* antigen-based ELISA. Arch. Med. Res. **23:**285–288.
28. **Newton-Sanchez, O., K. Sturm-Ramirez, J. L. Romero-Zamora, J. I. Santos-Preciado, and J. Samuelson.** 1997. High rate of occult infection with *Entamoeba histolytica* among non-dysenteric Mexican children. Arch. Med. Res. **28S:**311–313.
29. **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.
30. **Ortner, S., M. Binder, O. Scheiner, G. Wiedermann, and M. Duchene.** 1997. Molecular and biochemical characterization of phosphoglucomutases from *Entamoeba histolytica* and *Entamoeba dispar*. Mol. Biochem. Parasitol. **90:** 121–129.
31. **Ortner, S., C. G. Clark, M. Binder, O. Scheiner, G. Wiedermann, and M. Duchene.** 1997. Molecular biology of the hexokinase isoenzyme pattern that distinguishes pathogenic *Entamoeba histolytica* from nonpathogenic *Entamoeba dispar*. Mol. Biochem. Parasitol. **86:**85–94.
32. **Ravdin, J. I.** 1995. Amebiasis. State-of-the-art clinical article. Clin. Infect. Dis. **20:**1453–1466.
33. **Rich, S. M., R. R. Hudson, and F. J. Ayala.** 1997. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. Proc. Natl. Acad. Sci. USA **94:**13040–13045.
34. **Rich, S. M., M. C. Licht, R. R. Hudson, and F. J. Ayala.** 1998. Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **95:**4425–4430.
35. **Robinson, G.** 1968. The laboratory identification of human parasitic amoebae. Trans. R. Soc. Trop. Med. Hyg. **62:**285–294.
36. **Romero, J. L., S. Reed, E. Orozco, J. I. Santos, and J. Samuelson.** 1992. Use of the polymerase chain reaction and nonradioactive DNA probes to diagnose *Entamoeba histolytica* in clinical samples. Arch. Med. Res. **23:**277–279.
37. **Rosenthal, B., Z. Mai, D. Caplivski, S. Ghosh, H. de la Vega, T. Graf, and J. Samuelson.** 1997. Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. J. Bacteriol. **179:**3736–3745.
38. **Samuelson, J., D. Caplivski, K. Sturm-Ramirez, K. Kretsinger, S. Descoteaux, G. Hernandez, G. Vargas, R. Mammo, O. Newton, J. Santos, J. H. de la Vega, P. Robbins, C. Ganguly, and A. Lohia.** 1997. A proposal for a molecular biologic system for classifying isolates of *Entamoeba histolytica* and *Entamoeba dispar*. Arch. Med. Res. **28:**274–275.
39. **Sargeaunt, P. G., J. E. Williams, and J. D. Greene.** 1978. The differentiation of invasive and non-invasive *E. histolytica* by isoenzyme electrophoresis. Trans. R. Soc. Trop. Med. Hyg. **72:**519–521.
40. **Sehgal, D., V. Mittel, 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.
41. **Singh, U., J. B. Rogers, B. J. Mann, and W. A. Petri, Jr.** 1997. Transcription initiation is controlled by three core promoter elements in the *hgl5* gene of the protozoan parasite *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA **94:** 8812–8817.
42. **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.
43. **Tannich, E., and R. D. Horstmann.** 1992. Codon usage in pathogenic *Entamoeba histolytica*. J. Mol. Evol. **34:**272–273.
44. **Tannich, E., R. D. Horstmann, J. Knobloch, and H. H. Arnold.** 1989. Genomic DNA differences between pathogenic and nonpathogenic *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA **86:**5118–5122.
45. **Tibayrenc, M., F. Kjellberg, J. Arnaud, B. Oury, S. F. Breniere, M.-L. Darde, and F. J. Ayala.** 1991. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc. Natl. Acad. Sci. USA **88:**5129–5133.
46. **Wang, L., J. Calderon, and S. L. Stanley, Jr.** 1997. Short report: identification of B-cell epitopes in the serine-rich *Entamoeba histolytica* protein. Am. J. Trop. Med. Hyg. **57:**723–726.
47. **Willhoeft, U., and E. Tannich.** 1999. The electrophoretic karyotype of *Entamoeba histolytica*. Mol. Biochem. Parasitol. **99:**41–53.
48. **Yu, Y., and J. Samuelson.** 1994. Primary structure of *Entamoeba histolytica* nicotinamide nucleotide transhydrogenase. Mol. Biochem. Parasitol. **68:**323–328.
49. **Zhang, T., P. R. Cieslak, and S. L. Stanley, Jr.** 1994. Protection of gerbils from amebic liver abscess by immunization with a recombinant *Entamoeba histolytica* antigen. Infect. Immun. **62:**1166–1170.