Phylogeny of the Genera *Entamoeba* and *Endolimax* as Deduced from Small-Subunit Ribosomal RNA Sequences

Jeffrey D. Silberman, C. Graham Clark,† Louis S. Diamond,‡ and Mitchell L. Sogin**

*Marine Biological Laboratory, Woods Hole, Massachusetts; †Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, England; and ‡Laboratory of Parasitic Diseases, National Institutes of Health, Bethesda, Maryland

We sequenced small-subunit ribosomal RNA genes (16S-like rDNAs) of 10 species belonging to the genera *Entamoeba* and *Endolimax.* This study was undertaken to (1) resolve the relationships among the major lineages of *Entamoeba* previously identified by riboprinting; (2) examine the validity of grouping the genera *Entamoeba* and *Endolimax* in the same family, the Entamoebidae; and (3) examine how different models of nucleotide evolution influence the position of *Entamoeba* in eukaryotic phylogenetic reconstructions. The results obtained with distance, parsimony, and maximum-likelihood analyses support monophyly of the genus *Entamoeba* and are largely in accord with riboprinting results. Species of *Entamoeba* producing cysts with the same number of nuclei form monophyletic groups. The most basal *Entamoeba* species are those that produce cysts with eight nuclei, while the group producing four-nucleated cysts is most derived. Most phylogenetic reconstructions support monophyly of the Entamoebidae. In maximum-likelihood and parsimony analyses, *Endolimax* is a sister taxon to *Entamoeba,* while in some distance analyses, it represents a separate lineage. The secondary loss of mitochondria and other organelles from these genera is confirmed by their relatively late divergence in eukaryotic 16S-like rDNA phylogenies. Finally, we show that the positions of some (fast-evolving) eukaryotic lineages are uncertain in trees constructed with models that make corrections for among-site rate variation.

Introduction

The systematics of amoeboid eukaryotes is in a state of flux. Over the past several years, it has become clear that amoebae represent a highly diverse and polyphyletic grouping. Other than their form of locomotion, no phenotypic characters unite the group (Hinkle et al. 1994; Silberman, Clark, and Sogin 1996). The notion that primitive eukaryotic morphologies resembled amoeboid cells (Meza 1992; Bakker-Grunwald and Wöstmann 1993) is not supported by molecular data. The earliest-branching eukaryotic lineages consistently identified by ribosomal RNA gene (rDNA) sequence analyses—the Diplomonadida and Microspora (Cavalier-Smith 1993; Leipe et al. 1993)—do not include amoebae.

Based on the absence of mitochondria and other eukaryotic organelles (Martínez-Palomo 1986), Bakker-Grunwald and Wöstmann (1993) argued that *Entamoeba* was an appropriate model for the earliest eukaryote. Molecular phylogenies inferred from comparisons of alpha, beta, and gamma tubulins (Hasegawa et al. 1993; Edlind et al. 1996; Keeling, Poulsen, and McFadden 1998) support a basal placement of *Entamoeba.* In contrast, phylogenies based on analysis of small-subunit rDNA (Sogin and Silberman 1998) and elongation factors (Shirakura et al. 1994; Baldauf, Palmer, and Doolittle 1996) show *Entamoeba* to be of relatively late divergence. If *Entamoeba* does not represent an early-diverging lineage, its lack of mitochondria must represent secondary loss, rather than a primitive state. Support for this hypothesis came from studies that revealed the presence of genes in *Entamoeba histolytica* which encode proteins normally associated with mitochondria (Clark and Roger 1995). The recent detection of a mitochondrion-derived subcellular compartment in *E. histolytica* (Mai et al. 1999; Tovar, Fischer, and Clark 1999) confirms the hypothesized secondary loss of mitochondria in the *Entamoeba* lineage. However, additional data are needed to define the correct position of *Entamoeba* in broad-scale eukaryotic phylogenies.

The family Entamoebidae was originally erected by Chatton (1925) to encompass the parasitic intestinal amoebae of the genera *Entamoeba, Endamoeba, Endolimax, Dientamoeba,* and *Iodamoeba.* While *Endamoeba* is exclusively a parasite of invertebrates, the other genera generally inhabit vertebrate hosts. According to morphological (Camp, Mattern, and Honigberg 1974) and molecular evidence (Silberman, Clark, and Sogin 1996), *Dientamoeba fragilis* is now recognized as an aberrant trichomonad. However, most classification schemes retain the other genera in the Entamoebidae. Except for *Entamoeba,* most members of the Entamoebidae are not well studied. Even the life cycles of *Dientamoeba* and *Iodamoeba* are still incompletely known. This paucity of information is due to the relatively minor clinical importance of these taxa and difficulties associated with their laboratory cultivation.

Morphologically, there is little resemblance between *Entamoeba, Endolimax,* and *Iodamoeba,* which raises concern about their placement in the same family. They are alike only with regard to the stages in their life cycles and similarities in their modes of locomotion, with all of them moving by means of pseudopodia. All known *Endolimax* and *Iodamoeba* have two stages in their life cycles, a trophozoite stage and a cyst stage, as do the majority of *Entamoeba.* Mature cysts of *Endolimax* contain four nuclei and those of *Iodamoeba* contain

Key words: Entamoebidae, *Entamoeba, Endolimax,* 16S-like rDNA, phylogeny, evolution.

Address for correspondence and reprints: Mitchell L. Sogin, Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, 7 MBL Street, Woods Hole, Massachusetts 02543. E-mail: sogin@mbl.edu.

Mol. Biol. Evol. 16(12):1740–1751. 1999

Table 1
Physical Characteristics of the 16S-like rDNAs of *Entamoeba* Species, *Endolimax nana* and Other Amoeboid Taxa

Organism	No. Nuclei/ Cyst	G + C Content				rDNA Length (bp)	Normal Host
		All Positions	Aligned Positions				
			Fig. 1	Fig. 2	Fig. 3		
Entamoeba invadens IP-1	4	0.39	0.41	0.42	ø	1,968	Reptile
Entamoeba histolytica HM-1:IMSS	4	0.38	0.40	ø	ø	1,946	Primate
Entamoeba dispar	4	0.38	0.40	ø	ø	1,949	Primate
***Entamoeba ranarum* NIH:1092:1**	4	0.39	0.41	0.41	ø	1,934	Amphibian
***Entamoeba moshkovskii* Laredo**	4	0.41	0.40	0.41	0.42	1,944	Free-living
***Entamoeba insolita* NIH:1192:1**	4	0.40	0.40	0.41	ø	1,942	Tortoise
***Entamoeba hartmanni* 162-2005**	4	0.38	0.39	0.40	ø	1,960	Primate
***Entamoeba terrapinae* M**	4	0.37	0.39	0.40	ø	1,943	Turtle
***Entamoeba* sp. NIH:1091:1**	4	0.33	0.37	0.38	ø	2,004	Iguana
***Entamoeba coli* HU-1:CDC**	8	0.41	0.41	0.42	0.42	2,105	Primate
***E. coli* IH:96/135**	8	0.42	0.42	ø	ø	2,105	Primate
***Entamoeba polecki* NIH:1293:1**	1	0.33	0.37	ø	ø	1,858	Pig
***Entamoeba chattoni* NIH:0191:1**	1	0.34	0.37	0.38	0.39	1,864	Primate
Entamoeba gingivalis HU-304:NIH	—	0.34	0.37	0.38	0.39	1,918	Primate
***Endolimax nana* NIH:0591:1**	4	0.47	ø	0.46	0.46	2,589	Primate
Mastigamoeba balamuthi	1	0.47	ø	0.48	0.48	2,741	Free-living
Hartmannella vermiformis	1	0.49	ø	0.47	0.48	1,840	Free-living
Dictyostelium discoideum	—	0.42	ø	0.44	0.44	1,872	Free-living

Note.—The 16S-like rDNAs sequenced in this study are indicated in bold. ø indicates that the sequence was not part of a specific data set, and a dash indicates that an organism does not form cysts.

only one, but these nuclei differ dramatically in structure from their counterparts in *Entamoeba.* In fact, nuclear differences are so distinct, they form the most reliable means of separating the genera on morphological grounds.

Within the genus *Entamoeba,* species have classically been divided into groups based on the number of nuclei present in the mature cyst (Levine 1973). In a recent study, the validity of these groupings was upheld by riboprinting (Clark and Diamond 1997). However, the relationships among lineages producing cysts with one, four, and eight nuclei were unresolved, as were the deeper branches within the four-nucleated-cyst-producing group.

To further examine the relationships among the lineages of *Entamoeba,* between *Entamoeba* and *Endolimax,* and between these genera and other eukaryotes, we isolated, cloned, and sequenced the 16S-like rDNA of 10 species belonging to the genera *Entamoeba* and *Endolimax,* including members of all of the major lineages identified in the former genus. The results of our phylogenetic analyses give a much clearer picture of the evolution of these parasitic amoebae and offer likely explanations for the incongruity among the published rDNA and protein phylogenies.

Materials and Methods

Organisms

The origin, isolation, and cultivation of all strains of *Entamoeba* and of *Endolimax nana* used in this study (table 1) have been published previously (Clark and Diamond 1997). A second culture of *E. nana* was isolated (designation NIH:0192:1) from a siamang (*Symphalangus syndactylus*) at the National Zoological Park, Washington, D.C. Since the riboprint patterns of both isolates were identical, only one gene was sequenced. We determined 16S-like rDNA sequences for 10 species of Entamoebidae (table 1; GenBank accession numbers AF149906–AF149916). The 16S-like rDNA sequence of *Entamoeba invadens* was obtained from unpublished work by U. Edman, N. Agabian and M. Sogin (MBL, GenBank accession number AF149905), and all other sequences included in the analyses were obtained from GenBank.

Molecular Biology

DNA was isolated and 16S-like rDNA was amplified by the polymerase chain reaction as described (Clark and Diamond 1997). 16S-like rDNAs were cloned and sequenced as described (Silberman, Clark, and Sogin 1996). Briefly, the PCR fragments were electrophoresed through 1% low-melting-point agarose, and gel slices containing the appropriate DNA bands were excised. Ligations of the 16S-like rDNA into the T/A vector, pGem-T (Promega BioTech, Madison, Wis.), were performed either ''in gel'' or with DNA purified from the agarose using Gene Clean II (Bio 101, Vista, Calif.). For each species, between 7 and 12 independent clones were obtained.

Sequence data were generated on LI-COR 4000L automated sequencers using a cycle sequencing protocol (Epicentre, Madison Wis.) with IR-labeled primers. The 16S-like rDNA clones for each organism were pooled for sequencing to assess possible microheterogeneity among the multicopy rRNA genes and to overcome possible *Taq* polymerase errors in the initial amplification. No microheterogeneity or length variations were detected in the 16S-like rRNA genes of the species studied.

Phylogenetic Analyses

Sequence data were evaluated using distance, parsimony, and likelihood methods under a variety of evolutionary models for DNA substitution. DNA sequences

were aligned according to conserved primary and secondary structures. Only homologous positions were used in phylogenetic analyses; i.e., those positions for which the alignment was equivocal were excluded. All phylogenetic analyses were performed with the program PAUP*, version 4.0b1 (Swofford 1998).

The most appropriate model for calculating phylogenetic relationships was derived from an iterative process of tree and parameter estimations that were evaluated by likelihood methods (Sullivan, Markert, and Kilpatrick 1997). Using neighbor-joining trees, we calculated maximum-likelihood estimations of base frequency and proportion of invariant sites (PINVAR) and the alpha value for a discrete gamma shape parameter (α value) (four or eight rate categories of among-site rate variation). This process was reiterated using initial parameter settings to infer better trees and refine the parameter estimates. After six iterations, the parameters were incorporated into maximum-likelihood inferences for further refinement. Log-likelihood ratio tests (Yang, Goldman, and Friday 1995) were performed against other models of molecular evolution in order to determine whether a simpler (and hence less computationally intensive) model could estimate the data as well. A general time reversible (GTR) model of nucleotide substitution with estimation of the nucleotide rate substitution matrix, the PINVAR, and among-site rate heterogeneity described the phylogenetic reconstructions significantly better than all other models tested (Jukes-Cantor, F81, K_2P, HKY85, and GTR with equal rates of nucleotide substitution). This result was insensitive to the method used to generate the starting tree (parsimony, minimum evolution, or maximum likelihood). The refined estimates of base frequency, PINVAR, gamma shape parameter, and rate matrix were utilized in subsequent maximum-likelihood and minimum-evolution phylogenetic inferences under a GTR model of sequence evolution. Unless otherwise stated, all trees generated by maximum-likelihood and minimum-evolution methods were calculated under a GTR + gamma + PINVAR model of evolution. In the analyzed sequences, most insertions/deletions were of a single nucleotide; in rare instances, they involved two adjacent positions. Therefore, gaps were treated as a fifth character in parsimony analyses (virtually identical results were obtained in phylogenetic analyses that treated gaps as missing data). To assess the support for topological elements, 100 or more bootstrap replicates (Felsenstein 1985) were performed utilizing the maximum-likelihood and distance models described above plus parsimony. Because relative branching order can be influenced by the sequence composition of outgroup (and ingroup) taxa (Leipe et al. 1993), bootstrap analyses were performed with various combinations of outgroup and ingroup species.

We explored the potential effects of base composition differences among taxa in minimum-evolution estimates that employed LogDet/paralinear transformation to calculate evolutionary distances. Unfortunately, LogDet transformation of dissimilarity values precludes implementation of the gamma distribution, but effects of among-site rate heterogeneity were reduced by incorporating estimates of the PINVAR from a GTR model with equal rates among sites.

Results

Physical Characteristics of Entamoebidae Small-Subunit Ribosomal RNA Genes

Table 1 lists the lengths and base compositions of Entamoebidae 16S-like rDNAs. The rRNA genes from each of the three morphological *Entamoeba* subgroups have distinct lengths. Species producing cysts with one, four, and eight nuclei have 16S-like rDNA lengths of ca. 1,860, 1,960 and 2,100 bases, respectively. *Endolimax nana* has an exceptionally long 16S-like rRNA gene that contains 2,589 bases. The sequence length variations are primarily restricted to expansion regions that are also hypervariable in other eukaryotes. There is no evidence of introns in any of the Entamoebidae rRNA-coding regions.

The 16S-like rDNAs of *Entamoeba* species are relatively rich in A+T residues. Their G+C content ranges from 33% in *Entamoeba* sp. NIH:1091:1 and *Entamoeba polecki* to 42% in *Entamoebaa coli* IH:96/135. In contrast, the 47% G+C content of *E. nana* is comparable with the majority of eukaryotic 16S-like rDNAs. The prevalence of A+T is not restricted to hypervariable regions or expansion segments, but occurs throughout the gene. Because the A+T content varies over a small range within the genus *Entamoeba,* it should have little effect in phylogenetic reconstructions restricted to this genus. Inclusion of taxa outside the genus *Entamoeba* skews base composition among taxa. As discussed below, this variation in base composition may be misleading in phylogenetic inferences.

Phylogenetic Analyses

Phylogenetic analyses utilizing three different data sets were performed to answer the following questions: (1) What are the relative phylogenetic relationships among species of the genus *Entamoeba*? (2) What is the most basal *Entamoeba* taxon? (3) What is the phylogenetic relationship of *E. nana* to the genus *Entamoeba*? As a corollary to the last question, we explored the phylogenetic positions of *Endolimax* and *Entamoeba* in a data set that included a broad sampling of eukaryotic lineages.

Phylogenetic Relationships Within the Genus *Entamoeba*

To assess the phylogenetic relationships within the genus *Entamoeba,* we analyzed 16S-like rDNA sequences from 13 different species (14 taxa overall; fig. 1). The network in figure 1 is unrooted to maximize the number of positions available for phylogenetic analyses. We included 1,427 unambiguously aligned characters and utilized an eight-category discrete gamma among-site rate variation model in our phylogenetic reconstructions. To reduce the computational burden imposed by calculating phylogenetic trees under this complicated model, analyses of subsequent data sets utilized a four-category discrete gamma among-site rate variation model.

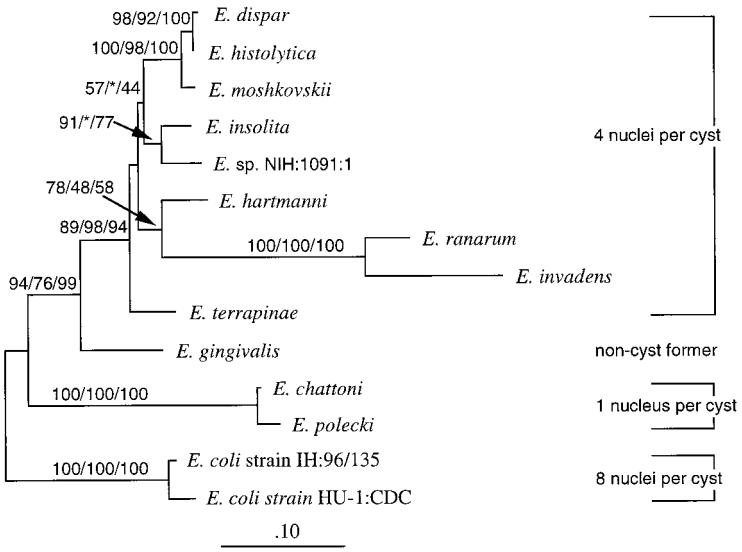

FIG. 1.—An unrooted phylogenetic reconstruction based on 16S-like rDNA exploring the relationships among *Entamoeba* species. A maximum-likelihood tree derived under a GTR model employing estimates of the proportion of invariant sites (PINVAR) and rate heterogeneity among sites (α value) is shown. Bootstrap numbers from 100 replicates of maximum likelihood, minimum evolution, and parsimony, respectively, are shown above the nodes. Nodes with significant bootstrap support are shown, and asterisks indicate bootstrap values less than 50. The scale bar represents the evolutionary distance for the number of changes per site. PINVAR = 0.478, α = 0.966.

Phylogenetic analyses of relationships among *Entamoeba* species delineated four clades that correlate with discrete morphological characteristics (fig. 1). *Entamoeba chattoni* and *E. polecki* form cysts containing a single nucleus and are sister taxa, a relationship that has 100% bootstrap support in all analyses. The two *E. coli* isolates (eight nuclei per cyst) are sister taxa in 100% of the bootstrap resamplings. The other well-supported group includes all species with four nuclei per cyst, and it is separated from the one- and eight-nuclei-per-cyst clusters by the non-cyst-forming species *Entamoeba gingivalis.*

Within the group of four-nucleated-cyst-producing species, *Entamoeba moshkovskii* joins *Entamoeba histolytica* and *Entamoeba dispar* in a well-supported clade. *Entamoeba ranarum* and *E. invadens* are sister taxa that form a clade with *Entamoeba hartmanni. Entamoeba* sp. NIH:1091:1 and *Entamoeba insolita* cluster together with high bootstrap support from maximum-likelihood (91%) and parsimony analyses (77%), but come together in less than 50% of the replicates in distance analyses. The position of *Entamoeba terrapinae* relative to these subclades is unresolved.

Rooting the *Entamoeba* Lineage

To determine the position of the root within the genus *Entamoeba,* four organisms in well-defined clades (*E. histolytica, E. dispar, E. polecki,* and one *E. coli*) were replaced by 16S-like rDNA sequences from the amoeboid species *E. nana, Mastigamoeba* (synonymous to *Phreatamoeba,* Simpson et al. 1997) *balamuthi, Hartmannella vermiformis,* and *Dictyostelium discoideum.* The number of characters that could be unambiguously aligned dropped to 1,325. The genus *Entamoeba* formed a robust monophyletic assemblage with high bootstrap support (fig. 2). The basal position of *E. coli* (eight nuclei per cyst) is followed by *E. chattoni* (a single nucleus per cyst). The non-cyst-forming species *E. gingivalis* diverged next, separating the monophyletic four-nuclei-per-cyst clade from the other *Entamoeba* groups.

Lower bootstrap values on nodes separating the *Entamoeba* lineages in the analyses depicted in figure 2 relative to values in figure 1 are most likely due to instabilities in the positions of lineages with moderate to long branches. In minimum-evolution analyses that did not employ corrections for among-site rate heterogeneity, the rapidly evolving *E. ranarum* plus *E. invadens* lineage migrated toward the base of the *Entamoeba* lineage and branched between *E. chattoni*/*E. polecki* and *E. gingivalis* (not shown). When corrected for among-site rate variation, the branching pattern reverted to that seen when only *Entamoeba* species were included in the analyses (fig. 1).

Although the A+T composition of the 16S-rDNAs is relatively homogeneous among *Entamoeba* species, we examined the possibility that the lower support for branching patterns in figure 2 reflected variant base compositions of the *Entamoeba* relative to the outgroup taxa. For this purpose, we inferred phylogenetic trees under a LogDet/paralinear model. When the most likely estimate for PINVAR was incorporated into the distance transformations, the topology recovered was that seen in figure 2. When an underestimate of PINVAR was used, the long *E. ranarum* plus *E. invadens* branch became unstable and moved the clade below *E. gingivalis* (not shown). Thus, base compositional bias does not seem to be the primary cause of branch instability. This phenomenon seems to be a classic case of long-branch attraction that is rectified by correction for among-site rate heterogeneity (Huelsenbeck 1997; Sullivan, Markert, and Kilpatrick 1997; Philippe and Laurent 1998).

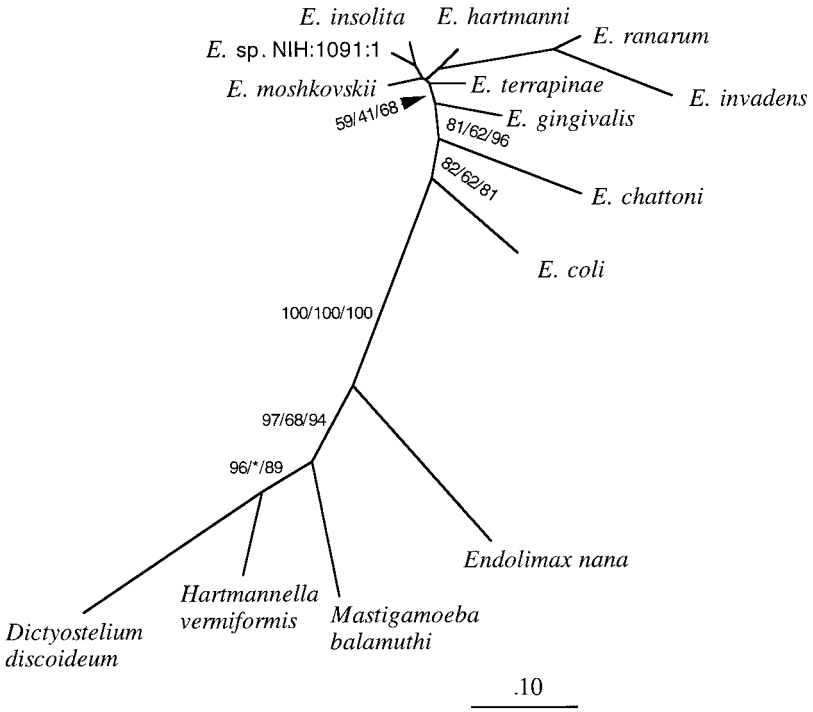

FIG. 2.—A rooted tree based on 16S-like rDNA defining the basal *Entamoeba* species. A maximum-likelihood tree derived under a GTR model employing estimates of the proportion of invariant sites (PINVAR) and rate heterogeneity among sites (α value) is shown. Bootstrap numbers from 100 replicates for maximum likelihood, minimum evolution, and parsimony, respectively, are shown for nodes leading to the *Entamoeba* clade and the nodes that separate *Entamoeba* lineages with differing numbers of nuclei per cyst. Asterisks indicate bootstrap values less than 50. Not shown are the bootstrap values for the *Entamoeba ranarum*/*Entamoeba invadens* node (100/100/100), the *Entamoeba hartmanni*/*E. ranarum*/*E. invadens* clade (86/*/48), and the *Entamoeba insolita*/*Entamoeba* sp. NIH:1091:1 node (86/41/76). The scale bar represents the evolutionary distance for the number of changes per site. PINVAR = 0.366, α = 0.924.

Phylogenetic Relationship of *E. nana* to the genus *Entamoeba*

Figure 2 shows that *E. nana* is the closest relative to the *Entamoeba* assemblage. For this collection of taxa, bootstrap values strongly support the affiliation of the genera *Endolimax* and *Entamoeba*, but there is a very long internal segment separating these taxa. We further explored the apparent sister taxon relationship between *Endolimax* and *Entamoeba* using a data set that included one species from each of the four divergent *Entamoeba* clades, *E. nana*, and representatives from 19 other divergent eukaryotic lineages. The number of characters that could be utilized in phylogenetic reconstructions dropped to 1,264. Analyses of this data set also place *E. nana* as the sister taxon to the genus *Entamoeba* (fig. 3). As in the previous analyses (fig. 2), strong bootstrap support in likelihood and parsimony analyses unites *Endolimax* with *Entamoeba*, although a large evolutionary distance separates these taxa. Conversely, distance analyses do not strongly support the specific relationship between the two genera. The placement and relative support for the *Endolimax* branch in distance analyses is dependent on the model of evolution used in phylogenetic reconstruction. Maximum-likelihood distance models suggest a sister taxon relationship of *Endolimax* and *Entamoeba* but with low bootstrap values of 42 and 55 with and without gamma + PINVAR corrections, respectively. However, minimum evolution under a GTR model weakly favored the alternative branching of *Endolimax* with *Mastigamoeba* (bootstrap values of 61 and 55, with and without gamma + PINVAR, respectively), as did LogDet/paralinear transformation of distances (bootstrap values of 61 and 75, with and without PINVAR, respectively). Although support in distance analyses is tenuous, collectively these analyses indicate that the genera *Endolimax* and *Entamoeba* shared a recent common ancestor, thus upholding the validity of the family Entamoebidae.

Phylogenetic Relationships of the Entamoebidae Within the Eukaryotic Subtree

An initial global maximum-likelihood tree was constructed that included the taxa depicted in figure 3 plus three eubacteria and three archaebacteria to root the eukaryotic subtree. As with previous phylogenetic studies utilizing 16S-like rDNA (Cavalier-Smith 1993; Leipe et al. 1993), the amitochondriate diplomonad and trichomonad lineages were among the earliest-diverging eukaryotes, followed by a polytomy consisting of the Euglenozoa, the Heterolobosea, and plasmodial and dictyostelid slime molds (data not shown). This analysis was limited to 1,104 aligned positions. In an effort to increase the resolution among the eukaryotic lineages, the prokaryotic outgroups were eliminated from subsequent analyses, thereby allowing 1,264 aligned positions to be used in phylogenetic reconstructions.

The trees shown in figure 3 confirm other ribosomal gene sequence analyses in placing *Entamoeba* (along

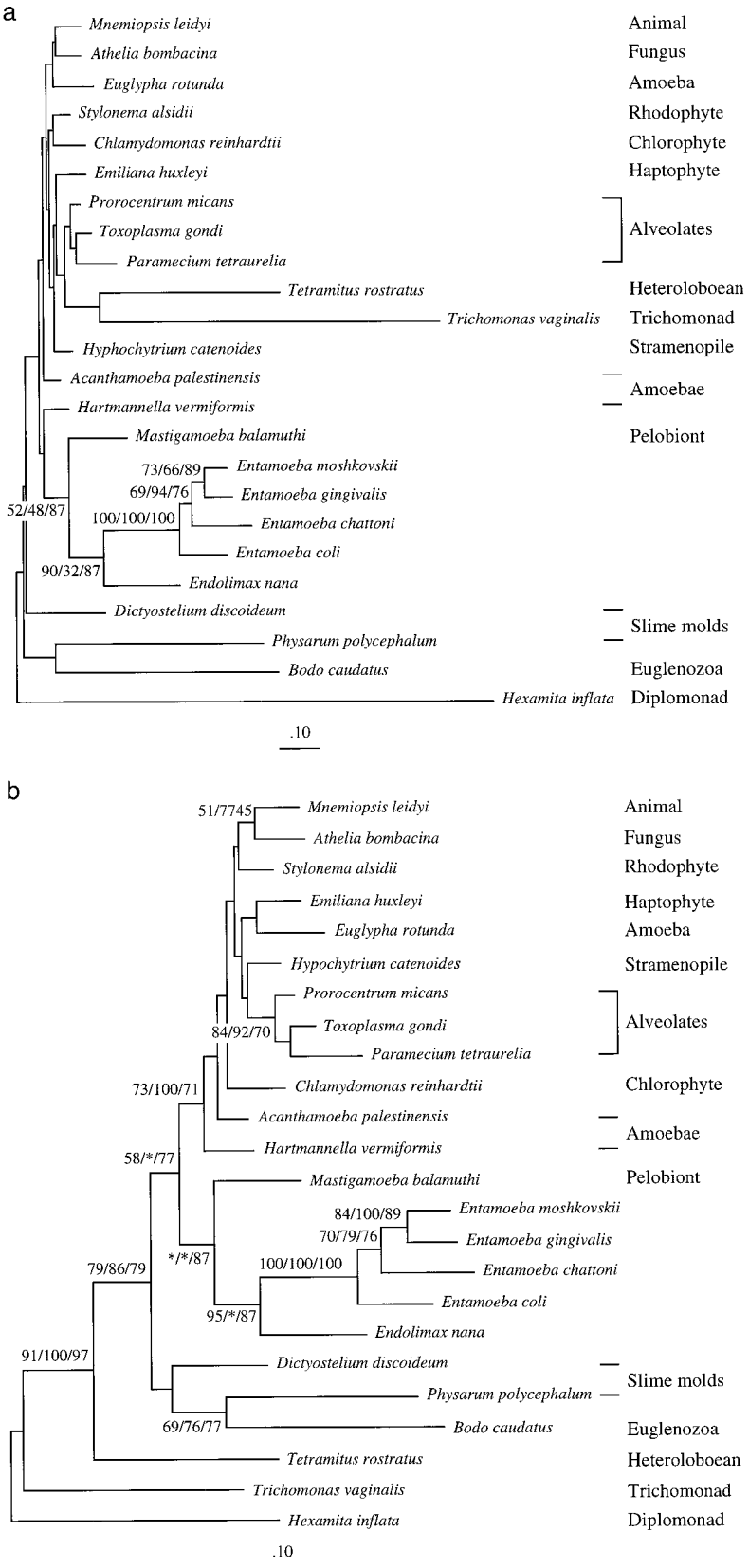

FIG. 3.—Phylogenetic association of *Endolimax* and *Entamoeba* within the eukaryotic subtree. *a*, The best tree (unrooted) under maximum likelihood with the GTR + gamma + PINVAR model of nucleotide evolution. Bootstrap values are shown over resolved nodes for maximum likelihood + GTR + gamma + PINVAR/distance + GTR + gamma + PINVAR/parsimony. PINVAR = 0.154, $\alpha = 0.532$. *b*, The best tree (unrooted) under maximum likelihood with the GTR, no gamma or PINVAR, model of nucleotide evolution. Bootstrap values are shown over resolved nodes for maximum likelihood + GTR, no gamma or PINVAR/distance + GTR, no gamma or PINVAR/parsimony. Asterisks indicate bootstrap values less than 50. Scale bar represents the evolutionary distance for the number of changes per site.

with *Endolimax*) among the mitochondriate lineages (Sogin 1991; Cavalier-Smith 1993; Sogin and Silberman 1998). The best trees under maximum likelihood, minimum evolution (both using a GTR + gamma + PINVAR model), and parsimony position the amitochondriate pelobiont *M. balamuthi* immediately before the *Entamoeba*/*Endolimax* group. The base of the *Mastigamoeba*/*Entamoeba*/*Endolimax* branch emerges close to the mitochondriate amoeba *H. vermiformis* and various other aerobic lineages. The bootstrap support for a pelobiont/*Endolimax*/*Entamoeba* group is weak in maximum-likelihood and minimum-evolution phylogenetic reconstructions but is moderate in parsimony analyses. However, the consistency of methods in recovering this group is indicative of a specific relationship between the pelobionts and *Entamoeba,* as previously suggested (Cavalier-Smith 1996/1997). There is strong support for a coherent alveolate group, but bootstrap support for the link between animals and fungi is low. The tenuous nature of the link between animals and fungi often appears in ribosomal RNA phylogenies that include divergent taxa and an attendant reduction in the number of unambiguously aligned positions. Most other nodes are not well resolved.

Models of Nucleotide Evolution Affect Branching Patterns

The branching positions of fast-evolving taxa that have previously been considered early-diverging eukaryotic lineages (Leipe et al. 1993; Cavalier-Smith and Chao 1996/1997) differed among trees constructed with and without estimates of among-site rate heterogeneity (see fig. 3).

Our phylogenetic reconstructions utilizing models that do not account for among-site rate heterogeneity corroborated previous 16S-like rDNA analyses (fig. 3*b*). A tree with a ''crown radiation'' (composed of taxa thought to have evolved relatively recently, e.g., animals, fungi, plants, alveolates, etc.; Knoll 1992) and ''early diverging taxa'' (represented by the diplomonads, trichomonads, heteroloboseans, Euglenozoa, and slime molds) was obtained when *Hexamita* was assigned a basal position. However, under the best model of phylogenetic reconstruction (maximum likelihood, GTR + gamma + PINVAR model), the relative branching positions of the trichomonads and the heteroloboseans were very different (fig. 3*a*). When a gamma distribution was used to correct for among-site rate variation in phylogenetic inferences, the trichomonads and heteroloboseans formed a sister clade to the alveolates, which are members of the crown radiation.

Kishino-Hasegawa ratio tests (Kishino and Hasegawa 1989) were conducted to assess the significance of differences in likelihood scores between trees with the heteroloboseans and trichomonads branching in the ''crown'' (fig. 3*a*) or ''deep'' (fig. 3*b*) under models of evolution with and without corrections for among-site rate variation (table 2). When comparing alternative tree topologies under a maximum-likelihood + GTR model of sequence evolution without accounting for among-site rate heterogeneity and invariable sites, the tree topology with the heteroloboseans and trichomonads diverging ''deep'' is significantly better than the alternative ($P < 0.017$). In contrast, in comparisons of the likelihoods of these trees under the maximum-likelihood + GTR + gamma + PINVAR model, the topology with the heteroloboseans and trichomonads diverging among the ''crown'' lineages is favored, but the likelihood of the alternative tree topology was not significantly different ($P > 0.45$). Moreover, resolution of the branching order among the typically ''deep'' taxa (diplomonads, trichomonads, heteroloboseans, Euglenozoa, and slime molds) disappears in phylogenetic reconstructions utilizing various maximum-likelihood and minimum-evolution methods (GTR, HKY85, K_2P, and Jukes-Cantor) that employ corrections for among-site rate heterogeneity and invariable sites. For instance, under a Jukes-Cantor + gamma + PINVAR model of evolution, the trichomonads, diplomonads, and heteroloboseans diverged near the alveolates. In addition, the moderate bootstrap support for a ''crown'' radiation and ''deep'' clades (e.g., heteroloboseans + trichomonads + diplomonads, *Physarum* + Euglenozoa) is not present in trees constructed with a gamma correction. These results indicate that the deep branching position of rapidly evolving lineages might be an artifact of long-branch attractions. Such misplacements can occur with methods that fail to account for accelerated evolutionary rates or among-site rate heterogeneity (or for suites of sites that vary at different rates in different lineages, i.e., a covarion model of rate variation; Lockhart et al. 1992, 1996, 1998).

Table 2
Kishino-Hasegawa Tests of Alternative Tree Topologies Under Models With and Without Correction for Among-Site Rate Heterogeneity and Proportion of Invariant Sites

Model of Evolution	−lnL		Kashino-Hasegawa Significance Test
	Tree in Figure 3*a*	Tree in Figure 3*b*	
ML + GTR + Γ + I	14,212.8*	14,225.6	$P > 0.45$
ML + GTR	15,514.6	15,436.5*	$P < 0.017$

Note.—The asterisk denotes the favored tree topology for each model of nucleotide substitution.

The effect of among-site rate correction on long-branch attraction was assessed by Kishino-Hasegawa tests of alternative tree topologies (fig. 3*a* vs. *b*), holding all model parameters constant except for the gamma shape parameter (fig. 4). The likelihood of each tree was estimated under a Jukes-Cantor model of nucleotide substitution (equal base frequencies, transition/transversion ratio = 1, PINVAR = 0) while varying the α value from 0.05 (extreme among-site rate heterogeneity) to infinity (equal rates among sites). Using either tree topology under such a model of sequence evolution (fig. 3*a* or *b*), the best estimate for the α value was approximately 0.4. Between α values of 0.05 and 0.50, the tree with the trichomonads and heteroloboseans branching among the ''crown'' lineages was the best. At α values above 0.55, the tree with the trichomonads and heteroloboseans branching ''deep'' was the best. There was no

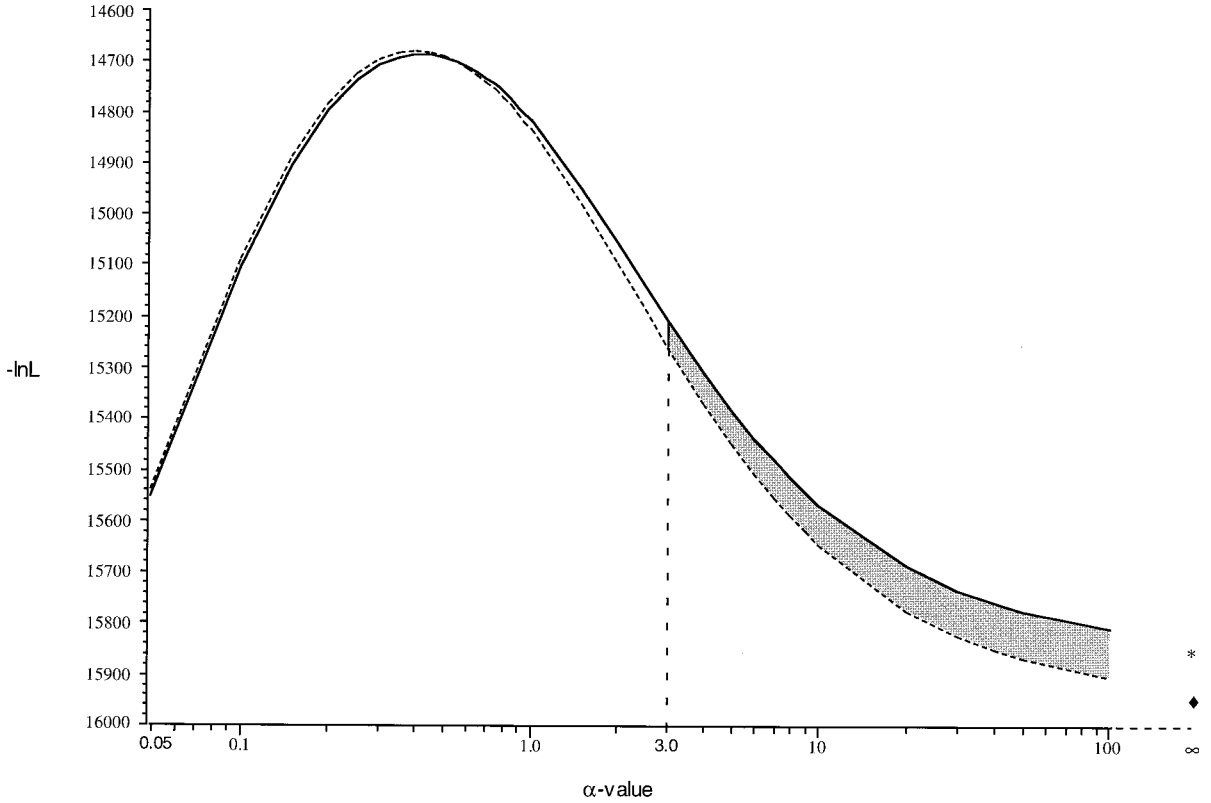

FIG. 4.—Plot of −lnL of trees in figure 3 with changing gamma shape parameter values. Note that the values on the y-axis are inverted (i.e., ascending −lnL values range from 16,000 to 14,600). The likelihood values of the trees in figure 3 were calculated under a Jukes-Cantor model of evolution (equal base frequency, transitions/transversions = 1, PINVAR = 0) at gamma shape parameters (α values) ranging from 0.05 to infinity. Likelihood scores were estimated for each tree at α values in 0.05 increments up to 2.0 and in 1.0 increments up to 10, then at α values of 20, 30, 40, 50, 100, 200, and, finally, ∞ (equal rates). The dashed curve represents the −lnL values of the tree depicted in figure 3*a*, and the solid curve represents the −lnL values of the tree depicted in figure 3*b*. Kishino-Hasegawa tests were conducted for each α value. The best likelihood estimate for the gamma shape parameter (α) is approximately 0.4 using either tree as the starting tree. At α values below 0.55, the tree shown in figure 3*a* is the most likely tree, but the difference is not significant. At α values lower than 3.0 (left of the vertical dashed line) there is no significant difference in the −lnL values of the tree topologies. At α values above 3.0 (shaded region between curves), the tree topology of figure 3*b* becomes significantly better than that of figure 3*a*. ♦ = −lnL tree of figure 3*a* at equal rates of evolution among sites; * = −lnL tree of figure 3*b* at equal rates of evolution among sites.

significant difference in the likelihoods of the tree topologies at α values lower than 3.0. At α values above 3.0, the tree topology with the trichomonads and heteroloboseans branching ''deep'' was significantly better ($P < 0.05$).

The likelihood values of the alternative tree topologies seen in figure 3 were also assessed under a complex model of evolution (GTR) at various α values while optimizing all other model parameters (estimated base frequency, substitution rate matrix, and PINVAR). Because the exact model of evolution for each tree at each α value was slightly different, Kishino-Hasegawa tests could not be performed. However, the shape of the resulting plot of −lnL versus α value mirrored that depicted in figure 4 (data not shown). Thus, we can generalize the effect of correcting for among-site rate variation in resolving long branches under simple and complex models of sequence evolution.

The instability of the heterolobosean and trichomonad lineages did not influence the relative branching positions of *Endolimax* or *Entamoeba*. If corrections for among-site rate heterogeneity were used, all analyses recovered the sister taxon relationship of *Endolimax* and *Entamoeba* and the internal branching order of the *Entamoeba* taxa (with *E. coli* always being the most basal species).

Discussion

Relationships Among *Entamoeba* Species

The number of nuclei in mature cysts is a defining morphological feature in *Entamoeba* taxonomy. *Entamoeba* species may have one, four, or eight nuclei per cyst, while others do not encyst. Phylogenetic inferences based on comparisons of full-length 16S-like rDNA sequences (figs. 1 and 2) identify similar groupings.

Within the *Entamoeba* group that produces four nuclei per cyst, branching patterns in the rRNA trees are similar to those observed in riboprinting, a method that indirectly samples variation in 16S-like rRNA coding regions through comparisons of restriction endonuclease digestion patterns (Clark and Diamond 1997). The rRNA trees recognize the clustering of *E. insolita* with *Entamoeba* sp. NIH:1091:1 and the grouping of *E. hartmanni* with *E. ranarum* and *E. invadens*. Neither of these relationships was detected in riboprinting analyses.

While riboprinting is a good method for assessing relationships among closely related taxa, analyses using full-length rRNA sequences and complex models for phylogenetic inference provide increased resolution, especially in the deeper regions of the tree.

Whenever eukaryotic outgroups are used to root the *Entamoeba* clade (figs. 2 and 3) the taxa with eight nuclei per cyst occupy a basal position, followed by *Entamoeba* species with a single nucleus per cyst. Finally, a non-cyst-forming species diverges prior to a clade of species with four nuclei per cyst. This distribution of groups with different numbers of nuclei per cyst makes it difficult to infer the state of the last common ancestor of all *Entamoeba* species. We can infer that since each internal node in the *Entamoeba* clade represents a cyst-bearing ancestral state, *E. gingivalis* must have secondarily lost its ability to encyst.

Two *Entamoeba* species are pathogens, but the others in this study are either commensal or free-living. *Entamoeba invadens* causes lethal infections in some reptiles, and *E. histolytica* is the causative agent of human amoebic dysentery. Both of these pathogenic species are members of separate well-supported clades. Since commensal forms diverge before pathogens in each of these clades, pathogenicity arose at least twice within the genus *Entamoeba.* Our trees yield no evidence of coevolution among *Entamoeba* species and their hosts. Species from reptilian and mammalian sources are interspersed in our phylogenies, which suggests that host switching has been a common event in this lineage.

All maximum-likelihood models and parsimony analyses recover tree topologies that are consistent with the monophyly of taxa based on numbers of nuclei per cyst. In contrast, under certain models of evolution, distance analyses do not support these clades. This is likely a consequence of the accelerated rates of evolutionary change in some lineages, as evidenced by the long branches leading to the *E. ranarum*/*E. invadens, E. chattoni*/*E. polecki,* and *E. coli* groups (figs. 1 and 2). For the rooted data set of figure 2, the position of the non-cyst-former *E. gingivalis* interchanges with *E. ranarum* plus *E. invadens* (not shown) in minimum-evolution analyses that do not correct for among-site rate heterogeneity. However, inferences based on gamma-corrected minimum-evolution estimates are identical to the likelihood tree in figure 2. Under a GTR model with a gamma correction, the tree topology shown in figure 2 is strongly favored by the Kishino-Hasegawa test ($P < 0.015$) over topologies obtained by minimal evolution without correction for among-site variation. Thus, accounting for among-site rate heterogeneity diminishes long-branch attraction in this data set. If the number of nuclei per cyst is a synapomorphy within the genus *Entamoeba,* then distance analyses under inappropriate evolutionary models yield incorrect tree topologies.

The Relationship Between *Entamoeba* and *Endolimax*

Our phylogenetic analyses generally support a common evolutionary history for *Endolimax* and *Entamoeba.* This agrees with prior analyses based on morphology and life history. Yet, the inclusion of *Endolimax* with *Entamoeba* in the family Entamoebidae is more difficult to assess. Except for very closely related bacterial species (Stackebrandt and Goebel 1994), there are no standards for converting genetic distances into taxonomic levels. *Endolimax* never branches within the *Entamoeba* clade. It emerges as a very deep lineage on the branch leading to the genus *Entamoeba.* In most analyses, *Endolimax* shares an immediate common ancestor with the genus *Entamoeba,* but they are separated by a large evolutionary distance (figs. 2 and 3). In distance-based phylogenetic inferences, *Endolimax* tends to cluster with the pelobiont *Mastigamoeba* at a basal position in the *Entamoeba* clade. This relationship between *Endolimax* and *Mastigamoeba* is not supported by phylogenetic reconstructions using parsimony and maximum likelihood (regardless of correction for among-site rate heterogeneity). The dependency of the branching pattern on the method of phylogenetic reconstruction limits our interpretation of how *Endolimax* might be related to *Entamoeba.* Additional molecular and ultrastructure data may clarify this relationship.

The pattern of cell and nuclear division following excystment (metacystic development) shows some parallels among *Endolimax* and *Entamoeba* species. Dobell (1928, 1938, 1943) studied the metacystic development of *E. nana, E. coli,* and *E. histolytica*. In all cases, a single, multinucleate amoeba emerges from the cyst and undergoes multiple rounds of cell division to produce uninucleate amoebae. However, there are significant differences in the details. *Endolimax nana* produces cysts with four nuclei, but no nuclear division takes place prior to cell division. Therefore, excystation produces four amoebae. Each of the four nuclei in the emerging *E. histolytica* undergoes one round of division, thus forming eight amoebae from a single cyst. In contrast, the eight nuclei of *E. coli* do not divide prior to cell division. There are no published reports of metacystic development in any species of *Entamoeba* that normally produces mature cysts with a single nucleus.

The molecular analyses presented in this study confirm the monophyly of the genus *Entamoeba* (figs. 2 and 3) and thus corroborate the use of the characteristic nuclear structure that is shared by all species examined as a diagnostic derived feature for the genus. As seen in stained preparations under the light microscope, the nucleus has a central ''karyosome,'' and the nuclear membrane is lined with peripheral chromatin. While the relative location and size of the karyosome and the thickness and distribution of the peripheral chromatin vary among species, the general appearance of the nucleus is distinctive enough to allow unambiguous identification at the generic level, clearly distinguishing *Entamoeba* from *Endolimax.*

In addition to nuclear structure, another morphological characteristic uniting *Entamoeba* species to the exclusion of *E. nana* is the presence of ''chromatoid material'' in *Entamoeba* cysts. These are semicrystalline arrays of ribosomes that are most prominent during cyst formation, but smaller arrays also appear in electron micrographs of the cytoplasm of trophozoites. Because

there are no published electron microscopic studies of *E. nana*, we do not know if similar structures occur. The phylogenetic proximity of *Endolimax* to the *Entamoeba* clade provides hints about features which may be discovered through more detailed investigations. This is the first DNA-based study of *E. nana*, and it will be of great interest to see whether some of the unusual features of *Entamoeba* species, such as their circular ribosomal RNA genes (Bhattacharya, Som, and Bhattacharya 1998), also occur in this species.

Relationships of *Entamoeba* and *Endolimax* to Other Eukaryotes

It is intriguing to consider a specific relationship among the pelobionts and *Endolimax*/*Entamoeba.* Like the Entamoebidae, pelobionts are amitochondriate amoebae. However, both *Endolimax* and *Entamoeba* are commensals or parasites that have exclusively amoeboid trophozoites, whereas the pelobionts are generally free-living and flagellated. These taxa group together in all analyses, albeit with weak bootstrap support (figs. 2 and 3). They share features such as an anaerobic life history and the ability to encyst and are remarkably similar in their apparently simplified intracellular organizations (Martínez-Palomo 1986; Simpson et al. 1997). However, many of these characteristics are also shared with other unrelated protist groups. The paucity of recognizable organelles in these taxa makes it difficult to identify any shared derived morphological features. Because the process of excystment has not been studied in pelobionts, comparison with either *E. nana* or the *Entamoeba* species is not yet possible. If a shared common ancestry were to be established, then the parasitic genera *Endolimax* and *Entamoeba* may be derived from an anaerobic, free-living, flagellated ancestor. Moreover, if mitochondrial remnant organelles (similar to those described for *Entamoeba*) were to be discovered in *Endolimax* and the pelobionts, it would point to an early degeneracy of mitochondria in this evolutionary assemblage, rather than a parallel adaptation to an anaerobic life style. As is the case for all exclusively amoeboid eukaryotes examined to date, *Endolimax* and *Entamoeba* must have secondarily lost structural components of their microtubule-based cytoskeletons, including the flagella. The absence of mitochondria, flagella, and other complex cytoskeletal features does not necessarily correspond to a primitive state.

The placement of the *Entamoeba*/*Endolimax* group within the eukaryotes is less clear when complex models are used to reconstruct evolutionary history. In all models that account for among-site rate variation, the positions of long-branched taxa normally found at the base of rooted eukaryotic trees become less resolved. For example, in maximum-likelihood analysis employing a gamma correction, the best tree places the heterolobosean and trichomonad lineages among taxa usually considered to be more recently evolved (fig. 3*a*). However, under this model of phylogenetic reconstruction, there was no significant difference in tree score for the placement of these taxa either adjacent to the alveolates (fig. 3*a*) or adjacent to the diplomonads, Euglenozoa, and slime molds (fig. 3*b*). Conversely, under a maximum-likelihood model without a gamma correction, the tree topology placing the heterolobosean and trichomonads among the latter groups (diplomonads, Euglenozoa, and slime molds) was significantly better than that placing them adjacent to the alveolates (table 2). Taken together, these data raise the possibility that the traditional placement of these normally ''deep'' branching taxa might be artifacts of long-branch attraction. These issues regarding phylogenetic methodologies do not influence results concerning the relationship of *Endolimax* to *Entamoeba*, but instead reflect the difficulties in confidently placing taxa that have skewed base compositions, accelerated rates of molecular change, and/or exaggerated among-site rate heterogeneity.

As discussed by Philippe and Adoutte (1998), long-branched lineages are typically located deep in phylogenetic trees and the constituent taxa of these deep nodes differ with the molecule under analyses. A synthesis of these data has led to the hypothesis that extant eukaryotes may have arisen over a short period, i.e., an evolutionary ''big bang.'' This view of eukaryote evolution is in sharp contrast to the fossil record, which clearly demonstrates the existence of eukaryotes in 1.7–2.2-billion-year-old rocks, the relatively late diversification of basal animal lineages at the precambrian, and the divergence of early chlorophytes and plants around 850 MYA. Unfortunately, the eukaryotes seen in the ancient rocks cannot be assigned to recognizable, extant protist lineages, which makes it difficult to directly link the early fossil record with current molecular hypotheses.

Among distantly related groups of organisms, it is apparent that most, if not all, genes suffer from rate or compositional anomalies in various lineages. For example, among-site rate heterogeneity and/or mutational saturation have been convincingly demonstrated in large- and small-subunit rRNA genes (Kumar and Rzhetsky 1996; Tourasse and Gouy 1998), elongation factors (Roger et al. 1999), tubulins and actin (Philippe and Adoutte 1998), and the largest subunit of RNA polymerase (Hirt et al. 1999). These genes are known to violate simple models of molecular evolution and can lead to erroneous conclusions about evolutionary history. Although the favored model used in this study (GTR + gamma + PINVAR) is among the most comprehensive models of nucleotide substitution currently available that is computationally tractable, it does not fully describe actual modes of molecular evolution. This model highlights, but may not resolve, certain artifacts that plague molecular phylogenies. Models that fully account for realistic modes of molecular evolution are still lacking, and reevaluation of molecular data sets should be concurrent with the development of more refined models. Additionally, the resolving power of each molecular marker should be critically evaluated for its appropriateness in addressing specific phylogenetic questions. It is becoming apparent that single-gene phylogenies may not be sufficient for resolving basal eukaryotic lineages. Phylogenetic distributions of well-understood phenotypic traits, including morphology,

biochemistry, developmental processes, and life history, provide a powerful means by which to evaluate the utility of any particular molecular phylogeny.

In summary, the 16S-like rDNA phylogenies presented here firmly establish the monophyly of the genus *Entamoeba*. The number of nuclei per cyst is a valid taxonomic character that delimits monophyletic *Entamoeba* lineages. The proposed Entamoebidae assemblage is largely supported by our molecular phylogenies. There is weak evidence that the Entamoebidae and pelobionts are sister lineages. It is clear that both *Endolimax* and *Entamoeba* are derived from mitochondrion-bearing ancestors, but our phylogenies highlight the difficulties in establishing their immediate common ancestor and their relationship to other eukaryotic groups.

Acknowledgments

We wish to thank Andrew Roger, Andrew McArthur, and Virginia Edgcomb for critical comments on earlier versions of the manuscript and A. Roger and A. McArthur for valuable suggestions on phylogenetic methodologies. This work was supported in part by NIH grant GM32964 to M.L.S.

LITERATURE CITED

BAKKER-GRUNWALD, T., and C. WÖSTMANN. 1993. *Entamoeba histolytica* as a model for the primitive eukaryotic cell. Parasitol. Today **9**:27–31.

BALDAUF, S. L., J. D. PALMER, and W. F. DOOLITTLE. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc. Natl. Acad. Sci. USA **93**:7749–7754.

BHATTACHARYA, S., I. SOM, and A. BHATTACHARYA. 1998. The ribosomal DNA plasmids of *Entamoeba*. Parasitol. Today **14**:181–185.

CAMP, R. R., C. F. MATTERN, and B. M. HONIGBERG. 1974. Study of *Dientamoeba fragilis* Jepps & Dobell. I. Electronmicroscopic observations of the binucleate stages. II. Taxonomic position and revision of the genus. J. Protozool. **21**: 69–82.

CAVALIER-SMITH, T. 1993. Kingdom protozoa and its 18 phyla. Microbiol. Rev. **57**:953–994.

———. 1996/1997. Amoeboflagellates and mitochondrial cristae in eukaryote evolution: megasystematics of the new protozoan subkingdoms Eozoa and Neozoa. Arch. Protistenkd. **147**:237–258.

CAVALIER-SMITH, T., and E. E. CHAO. 1996/1997. Sarcomonad ribosomal RNA sequences, rhizopod phylogeny, and the origin of euglyphid amoebae. Arch. Protistenkd. **147**:227–236.

CHATTON, E. P. L. 1925. *Pansporella perplexa*, amoebien à spores protégées, parasite des daphnies. Réflexions sur la biologie et la phylogénie des protozoaires. Ann. Sci. Nat. Zool. 10e **8**:5–85.

CLARK, C. G., and L. S. DIAMOND. 1997. Intraspecific variation and phylogenetic relationships in the genus *Entamoeba* as revealed by riboprinting. J. Eukaryot. Microbiol. **44**:142–154.

CLARK, C. G., and A. J. ROGER. 1995. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc. Natl. Acad. Sci. USA **92**:6518–6521.

DOBELL, C. 1928. Researches on the intestinal protozoa of monkeys and man. I. General introduction II. Description of the whole life-history of *Entamoeba histolytica* in cultures. Parasitology **20**:357–412.

———. 1938. Researches on the intestinal protozoa of monkeys and man. IX. The life-history of *Entamoeba coli*, with special reference to metacystic development. Parasitology **30**:195–238.

———. 1943. Researches on the intestinal protozoa of monkeys and man. XI. The cytology and life-history of *Endolimax nana*. Parasitology **35**:134–158.

EDLIND, T. D., J. LI, G. S. VISVESVARA, M. H. VODKIN, G. L. MCLAUGHLIN, and S. K. KATIYAR. 1996. Phylogenetic analysis of beta-tubulin sequences from amitochondrial protozoa. Mol. Phylogenet. Evol. **5**:359–367.

FELSENSTEIN, J. 1985. Confidence intervals on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

HASEGAWA, M., T. HASHIMOTO, J. ADACHI, N. IWABE, and T. MIYATA. 1993. Early branchings in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. J. Mol. Evol. **36**:380–388.

HINKLE, G., D. D. LEIPE, T. A. NERAD, and M. L. SOGIN. 1994. The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi*. Nucleic Acids Res. **22**:465–469.

HIRT, R. P., J. M. LOGSDON JR., B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLEY. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc. Natl. Acad. Sci. USA **96**:580–585.

HUELSENBECK, J. P. 1997. Is the Felsenstein Zone a fly trap? Syst. Biol. **46**:69–74.

KEELING, P. J., N. POULSEN, and G. I. MCFADDEN. 1998. Phylogenetic diversity of parabasalian symbionts from termites, including the phylogenetic position of *Pseudotrypanosoma* and *Trichonympha*. J. Eukaryot. Microbiol. **45**:643–650.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

KNOLL, A. H. 1992. The early evolution of eukaryotes: a geological perspective. Science **256**:622–627.

KUMAR, S., and A. RZHETSKY. 1996. Evolutionary relationships of eukaryotic kingdoms. J. Mol. Evol. **42**:183–193.

LEIPE, D. D., J. H. GUNDERSON, T. A. NERAD, and M. L. SOGIN. 1993. Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol. Biochem. Parasitol. **59**:41–48.

LEVINE, N. D. 1973. Protozoan parasites of domestic animals and man. 2nd edition. Burgess Publishing Company, Minneapolis, Minn.

LOCKHART, P. J., C. J. HOWE, D. A. BRYANT, T. J. BEANLAND, and A. W. LARKUM. 1992. Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. **34**:153–162.

LOCKHART, P. J., A. W. LARKUM, M. STEEL, P. J. WADDELL, and D. PENNY. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA **93**:1930–1934.

LOCKHART, P. J., M. A. STEEL, A. C. BARBROOK, D. H. HUSON, M. A. CHARLESTON, and C. J. HOWE. 1998. A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol. Biol. Evol. **15**:1183–1188.

MAI, Z., S. GHOSH, M. FRISARDI, B. ROSENTHAL, R. ROGERS, and J. SAMUELSON. 1999. Hsp60 is targeted to a cryptic mitochondrion-derived organelle (''Crypton'') in the microaerophilic protozoan parasite *Entamoeba histolytica*. Mol. Cell. Biol. **19**:2198–2205.

MARTÍNEZ-PALOMO, A. 1986. Biology of *Entamoeba histolytica*. Pp. 11–43 *in* A. MARTÍNEZ-PALOMO, ed. Amebiasis. Elsevier, Amsterdam.

MEZA, I. 1992. *Entamoeba histolytica*: phylogenetic considerations. A review. Arch. Med. Res. **23**:1–5.

PHILIPPE, H., and A. ADOUTTE. 1998. The molecular phylogeny of Eukaryota: solid facts and uncertainties. Pp. 25–56 *in* G. H. COOMBS, K. VICKERMAN, M. A. SLEIGH, and A. WARREN, eds. Evolutionary relationships among protozoa. Kluwer Academic Publishers, Dordrecht, the Netherlands.

PHILIPPE, H., and J. LAURENT. 1998. How good are deep phylogenetic trees? Curr. Opin. Genet. Dev. **8**:616–623.

ROGER, A. J., O. SANDBLOM, W. F. DOOLITTLE, and H. PHILIPPE. 1999. An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. Mol. Biol. Evol. **16**: 218–233.

SHIRAKURA, T., T. HASHIMOTO, Y. NAKAMURA, T. KAMAISHI, Y. CAO, J. ADACHI, M. HASEGAWA, A. YAMAMOTO, and N. GOTO. 1994. Phylogenetic place of a mitochondria-lacking protozoan, *Entamoeba histolytica*, inferred from amino acid sequences of elongation factor II. Jpn. J. Genet. **69**:119–135.

SILBERMAN, J. D., C. G. CLARK, and M. L. SOGIN. 1996. *Dientamoeba fragilis* shares a recent common evolutionary history with the trichomonads. Mol. Biochem. Parasitol. **76**: 311–314.

SIMPSON, A. G. B., C. BERNARD, T. FENCHEL, and D. J. PATTERSON. 1997. The organisation of *Mastigamoeba schizophrenia* n. sp.: more evidence of ultrastructural idiosyncrasy and simplicity in pelobiont protists. Eur. J. Protistol. **33**:87–98.

SOGIN, M. L. 1991. Early evolution and the origin of eukaryotes. Curr. Opin. Genet. Dev. **1**:457–463.

SOGIN, M. L., and J. D. SILBERMAN. 1998. Evolution of the protists and protistan parasites from the perspective of molecular systematics. Int. J. Parasitol. **28**:11–20.

STACKEBRANDT, E., and B. M. GOEBEL. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44**:846–849.

SULLIVAN, J., J. A. MARKERT, and W. C. KILPATRICK. 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. Syst. Biol. **46**:426–440.

SWOFFORD, D. L. 1998. PAUP*. Version 4.0b1. Sinauer, Sunderland, Mass.

TOURASSE, N. J., and M. GOUY. 1998. Evolutionary relationships among protist phyla constructed from LSU rRNAs accounting for unequal rates of substitution among sites. Pp. 57–75 *in* G. H. COOMBS, K. VICKERMAN, M. A. SLEIGH, and A. WARREN, eds. Evolutionary relationships among protozoa. Kluwer Academic Publishers, Dordrecht, the Netherlands.

TOVAR, J., A. FISCHER, and C. G. CLARK. 1999. The mitosome, a novel organelle related to mitochondria in the amitochondriate parasite *Entamoeba histolytica*. Mol. Microbiol. **32**: 1013–1021.

YANG, Z., N. GOLDMAN, and A. FRIDAY. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst. Biol. **44**:384–399.

GEOFFREY MCFADDEN, reviewing editor

Accepted August 18, 1999