Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 6518–6521, July 1995
Evolution

Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*

(Archezoa/chaperonin/phylogeny)

C. GRAHAM CLARK*† AND ANDREW J. ROGER‡

*Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892-0425; and ‡Department of Biochemistry, Dalhousie University, Halifax, NS Canada B3H 4H7

Communicated by William Trager, The Rockefeller University, New York, NY, April 17, 1995 (received for review March 13, 1995)

ABSTRACT **Archezoan protists are thought to represent lineages that diverged from other eukaryotes before acquisition of the mitochondrion and other organelles. The parasite *Entamoeba histolytica* was originally included in this group. Ribosomal RNA based phylogenies, however, place *E. histolytica* on a comparatively recent branch of the eukaryotic tree, implying that its ancestors had these structures. In this study, direct evidence for secondary loss of mitochondrial function was obtained by isolating two *E. histolytica* genes encoding proteins that in other eukaryotes are localized in the mitochondrion: the enzyme pyridine nucleotide transhydrogenase and the chaperonin cpn60. Phylogenetic analysis of the *E. histolytica* homolog of cpn60 confirmed that it is specifically related to the mitochondrial lineage. The data suggest that a mitochondrial relic may persist in this organism. Similar studies are needed in archezoan protists to ascertain which, if any, eukaryotic lineages primitively lack mitochondria.**

Entamoeba histolytica has been considered by some authors to be an early branching eukaryotic lineage, as it appears to lack such typical organelles as mitochondria, peroxisomes, rough endoplasmic reticulum, and Golgi dictyosomes (1) and has an unusual glycolytic metabolism (2). Many of these features are shared with other amitochondriate protists known as Archezoa (3). However, small-subunit ribosomal RNA-based phylogenetic trees (3, 4) place *E. histolytica* on a branch that arises more recently than several lineages with typical eukaryotic organelles and metabolism. This is in contrast to what is found for the Archezoa, implying that the unusual features of *E. histolytica* are the result of secondary loss and are not due to primitive absence. The ribosomal tree placement has proven controversial, and several authors have challenged its accuracy (5–7). We have sought to resolve this debate by searching for genes of mitochondrial origin in *E. histolytica*.§

The mitochondrion, an organelle consisting of over 300 proteins involved in several metabolic pathways (8), appears to have its origins in an endosymbiosis between a eukaryotic cell and an α-proteobacterium (9, 10). Almost all mitochondrial proteins are now encoded in the nuclear genome. We reasoned that if ancestors of *E. histolytica* had indeed lost this organelle, some mitochondrion-specific enzymes might have remained useful to the ameba despite the absence of the subcellular compartment. Those genes would have been retained, and their detection would provide clear support for secondary loss.

MATERIALS AND METHODS

DNA. Axenic amebae were grown and DNA was isolated from *E. histolytica* isolate HM-1:IMSS clone 6, as described (11, 12).

Cloning of the *E. histolytica* Pyridine Nucleotide Transhydrogenase (PNT) Gene. Degenerate primers were designed on the basis of alignment of the *Escherichia coli* and bovine mitochondrial PNT gene sequences (13, 14) taking into account the *E. histolytica* codon bias. The primers (THB5.1, ATTGGAGGAGCWGATATGCCAGTTGT, and THB3.1, CCWGCAACTGGATGAATTCCRAATCT) were designed to prime at amino acids 209–217 and 341–349 of the *E. coli* PNT β sequence. PCR amplification used standard conditions (AmpliTaq kit; Perkin–Elmer/Cetus) with 30 cycles of 94°C for 1 min, 55°C for 1.5 min, and 72°C for 2 min each. A single product of the expected size was obtained, cloned, and used as a probe to screen a genomic library of *E. histolytica* HM-1:IMSS DNA. A single hybridization-positive plaque was obtained that had an insert of approximately 4.4 kb, which was sequenced on both strands by primer walking at the Johns Hopkins University DNA Analysis Center (GenBank accession no. L39933).

Cloning of the *E. histolytica* Chaperonin (cpn) 60 Gene. Degenerate primers were designed on the basis of an alignment of various cpn60 homologs taking into account the *E. histolytica* codon bias. Several primer combinations gave PCR products of the expected size, and the largest PCR product (amplification conditions as above except annealing at 45°C; primers HSP5.4, CCAAAARTTACWAAAGATGGAGTTACWGTT, and HSP3.4, TCCTCCTCCTGGAACAATTCC) of about 1.0 kb (covering amino acids 47–416 of the *E. coli* protein or 67% of the total length) was cloned and sequenced on both strands by primer walking at the Johns Hopkins University DNA Analysis Center (GenBank accession no. L39934). A complete genomic clone has not yet been isolated.

Phylogenetic Analyses. The partial sequence of the *E. histolytica* cpn60 gene was aligned with cpn60 homologs and analyses were performed on 521 alignment positions. Gaps and missing regions of sequence were scored as missing data. Parsimony, distance, and maximum likelihood phylogenetic analyses were performed by using the PAUP 3.1 (15), PHYLIP 3.5c (16), and MOLPHY 2.2 (17) phylogeny analysis software packages, respectively.

Maximum Parsimony (PAUP). The most parsimonious trees were obtained by use of heuristic searches employing 20 replicates of random stepwise addition of taxa and branch swapping. Branch lengths for the parsimony tree were calculated with the optimization criterion based on accelerated transformation (ACCTRAN; ref. 15).

Distance (PHYLIP). Distance matrices were obtained by using PROTDIST, employing the Dayhoff PAM 001 matrix to make an amino acid transition probability matrix, and neighbor-joining

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: BP, bootstrap proportion; cpn, chaperonin; PNT, pyridine nucleotide transhydrogenase.

†To whom reprint requests should be addressed.

§The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L39933 and L39934).

trees were constructed by using the NEIGHBOR program (16). Bootstrap proportions (BPs) were determined by analysis of 500 resamplings of the data set for the maximum parsimony analysis and 400 resamplings for the distance analysis.

Maximum likelihood (MOLPHY). To employ the maximum likelihood method, we removed all regions with missing data from the alignment (except for single amino acid deletions which were scored as a 21st amino acid) leaving 362 positions. Maximum likelihood analysis was performed by using the PROTML program (version 2.2) (17) and a semi-constrained tree, employing the Jones, Taylor, and Thornton amino acid substitution matrix (18) adjusted for the amino acid frequencies in the data set (the JTT-F option). BPs were estimated by PROTML, which employs the RELL method, with 10,000 replications (19).

RESULTS

The enzyme PNT is exclusively located in the inner mitochondrial membrane in eukaryotes and in the cytoplasmic membrane of eubacteria, where it transfers reducing equivalents between NADH and NADPH (20). PNT activity has been detected in *E. histolytica* (21), making it a strong candidate for an *Entamoeba* enzyme of endosymbiotic origin.

The presence of this enzymatic activity in *E. histolytica* need not imply enzyme homology. Therefore, we attempted to detect an *E. histolytica* homolog of the bacterial/mitochondrial PNT gene by using PCR amplification with degenerate primers directed against highly conserved regions of the PNT β region. Sequencing of the single PCR product and subsequent analysis of a genomic clone of the gene proved that *E. histolytica* did encode a homologous protein. The deduced *E. histolytica* protein sequence (1083 amino acids; data not shown) shows several regions of high sequence similarity to the four other PNT homologs. Although the high degree of sequence identity is strong evidence for a eubacterial (and most likely mitochondrial) origin, the PNT gene is not well suited to phylogenetic analysis, as all nodes on parsimony and distance trees are poorly supported by BPs (data not shown). After this work was completed similar results were published by Yu and Samuelson (22). No strong conclusions regarding the phylogenetic origin of the *E. histolytica* PNT could be drawn. We therefore sought an *E. histolytica* gene of mitochondrial origin that was more amenable to phylogenetic reconstruction.

The amino terminus of the *E. histolytica* PNT shows an extension when compared to those of other organisms (Fig. 1), which is consistent with the presence of a transit peptide. The length and amino acid composition of the extension are similar to the transit peptides of trichomonad hydrogenosomal proteins (28, 29). The presence of the extension, coupled with the observation that the *E. histolytica* PNT activity is associated with the membrane fraction of the cell (21), suggested that a relic of the original mitochondrion might still be present.

Refolding of proteins after transit across the mitochondrial membrane involves the mitochondrial 60-kDa chaperonin, cpn60. This protein and its eubacterial homologs have been used extensively for phylogenetic analyses (9, 30) so we looked for an *E. histolytica* homolog of cpn60. Once again, PCR amplification with degenerate primers produced products of the expected size, and sequencing confirmed the homology of this gene to mitochondrial and eubacterial chaperonins. Although clearly a divergent sequence, as indicated by the length of its branch, phylogenetic analysis of the *E. histolytica* sequence shows that it clusters with mitochondrial cpn60s to the exclusion of the eubacterial cpn60 homologs by using distance, parsimony, and protein maximum likelihood analyses (Fig. 2 *A–D*). BPs give a high degree of support to this topology, as would be expected if the *E. histolytica* cpn60 were a mitochondrially derived gene. There are minor differences between the distance and parsimony trees, but none of the variable nodes are given significant support by bootstrap analysis.

Phylogenetic analyses of cpn60 sequences suggest that mitochondrial sequences form a clade with two α-proteobacteria, *Rickettsia tsutsugamushi* and *Ehrlichia chaffeensis*, to the exclusion of all other taxa (9). This result is consistent with analyses of ribosomal RNA, which indicate a similar affinity (10). Our analysis of a cpn60 data set which included *Ehr. chaffeensis* indicated that, in some of the distance analyses, the *E. histolytica* sequence showed a tendency to cluster with the cpn60 of *Ehr. chaffeensis* rather than those of mitochondria.

Since *Ehrlichia* and *Rickettsia* form a specific clade (9, 10), if the relationship of the *E. histolytica* and *Ehr. chaffeensis* sequences were real, we reasoned that deletion of *Ehr. chaffeensis* cpn60 from the data set should result in an *Entamoeba/Rickettsia* clade. On the contrary, however, removal of *Ehr. chaffeensis* causes the bootstrap support for the alternative topology, where *E. histolytica* cpn60 forms a clade with the mitochondrial sequences, to increase substantially in both distance and parsimony analyses and the BP for a specific *Entamoeba/Rickettsia* relationship becomes negligible (not shown). We suggest, therefore, that the affinity of the *Ehr. chaffeensis* and *E. histolytica* sequences, both of which are quite divergent, is the result of artifactual clustering of long branches. This has long been known to occur in parsimony analysis (31), and recent simulation studies (32, 33) have shown that both parsimony and distance methods perform poorly under conditions of substantial branch length inequality. For this reason, we chose to remove the *Ehr. chaffeensis* sequence from the final parsimony and distance analyses.

Unlike distance and parsimony methods, maximum likelihood analysis has been shown to be relatively insensitive to the effects

A

```
Consensus        TATTTAAA..........(N)19-27.......ATTCA(N)0-12ATG
Entamoeba PNT    TATTTAAA..........(N)24.........ATTCT(N)6..ATG
                 TATTTAAAGAGAAAAAAAAAATAGATAATAACATTCTAAATAAATG
```

B

```
                                                          |50
Entamoeba histolytica  MSTSSSIEEEVFNYMKIINNFVSVGNIIVSLCFILALRGLSTQISAKMGNIYGIIGMTVAFIAAVVDKIG
Bovine mitochondrion                ...SGY-IEQIMYLGSG--CVG--A-----GT-RL--AL-M--VAGGLA-TLGGLKP
Eimeria tenella                        MPSLLGAVYLFSAI----C------PQT--R---L-LV--VA-VVVTFTEAGF
Escherichia coli                      MSGGL-TAAY-VAAIL--FS-A---KHETSRQ--NF--A--AI-L--TIFGPDT
Rhodospirillum rubrum                 MTHSLTMAAY-VAGVL---------NPE--RN--RM-MV--AI-ILTTLLSPSV
```

FIG. 1. Sequence of the 5′ end of the *E. histolytica* PNT. (*A*) Upstream of the first in-frame methionine (italicized ATG) of *E. histolytica* PNT is a very close match, in both sequence and relative spacing, to the *E. histolytica* putative promoter element consensus (23). (*B*) Deduced amino terminal sequence of *E. histolytica* PNT aligned with homologs from bovine mitochondrion, *Eimeria tenella*, *E. coli*, and *Rhodospirillum rubrum*. The standard single letter amino acid code is used. Dashes indicate amino acid identity. The *E. coli* PNT operon encodes two proteins: the α and β subunits (13). The *Rhodospirillum* PNT α homolog consists of two proteins (AA and AB) (24). The *E. histolytica* PNT is a β-α fusion protein as is the *Eimeria* PNT (25), which is found in refractile bodies whose relationship to mitochondria is unclear (26). Bovine mitochondrial PNT is an α-β fusion protein (14), and therefore no amino terminus is shown. The bovine mitochondrial PNT α subunit carries a 43-amino acid signal peptide (27).

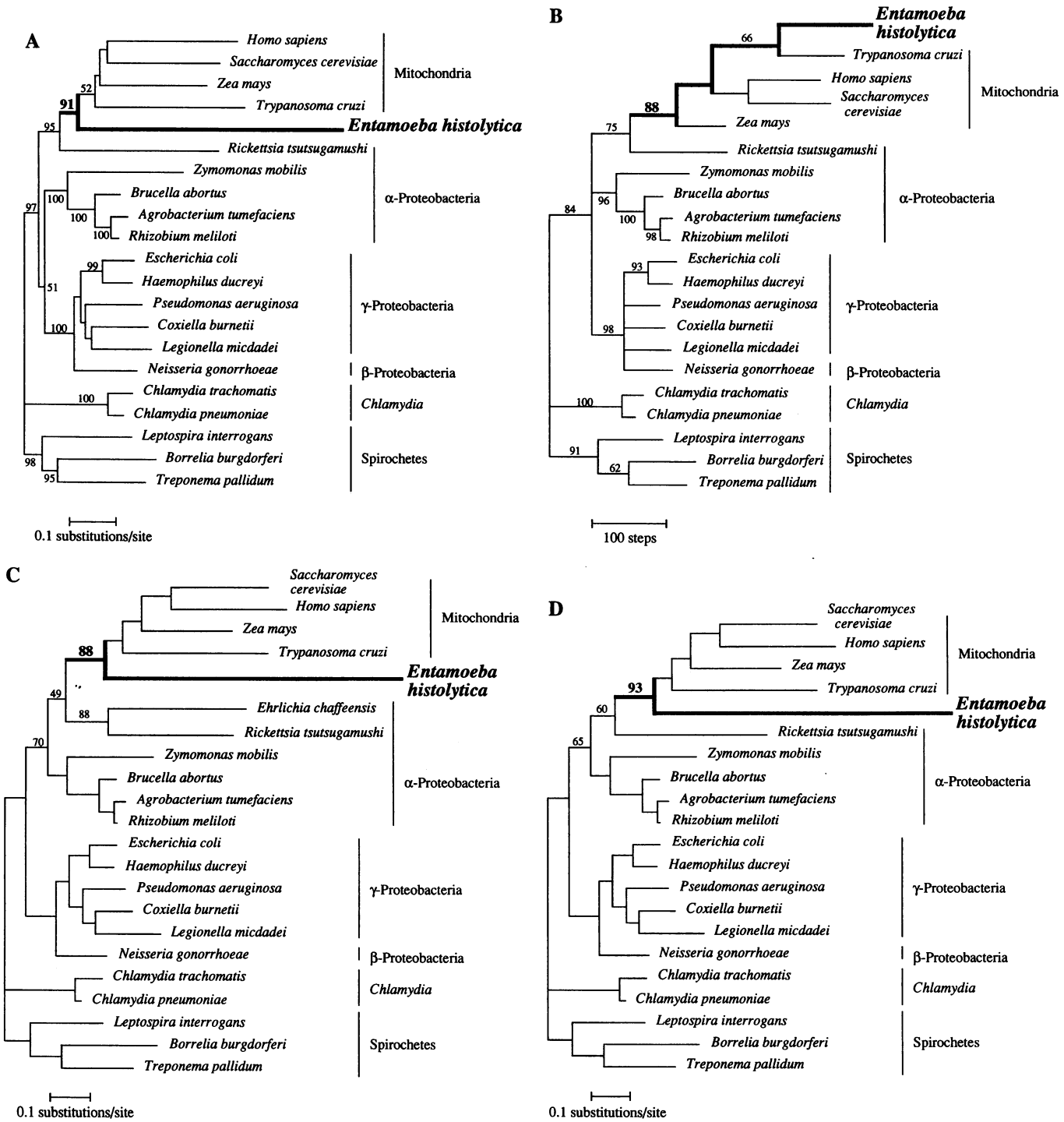

FIG. 2. Cpn60 phylogeny. Sequences were selected from organisms known to branch in the region of the eubacterial/mitochondrial cpn60 clade [α-, β-, and γ-proteobacteria with *Chlamydia* and spirochetes as outgroups (9, 30)]. (*A*) Protein distance matrix-based tree. (*B*) Maximum parsimony tree. This tree has a length of 2180 steps and is a strict consensus of three equally parsimonious trees. All nodes with BPs of more than 50% are indicated on both the distance and parsimony trees. (*C* and *D*) Maximum likelihood trees. The data set with (*C*) and without (*D*) the *Ehrlichia* sequence was analyzed as stated in *Materials and Methods.* Robust cpn60 groupings (α- and β-proteobacteria except for *Rickettsia* and *Ehrlichia*, γ-proteobacteria, chlamydiae, spirochetes, and the human, yeast, and maize mitochondrial sequences) were constrained. Estimated BPs for the relevant nodes are indicated. The ln likelihood values are −8990.17 and −8451.66 for *C* and *D*, respectively.

of evolutionary rate inequality among lineages (33). The maximum likelihood trees are shown in Fig. 2 *C* and *D* and estimated BPs are shown for the relevant nodes. Maximum likelihood analysis provides strong evidence for the *Entamoeba*/mitochondria clade with significant bootstrap support regardless of whether the *Ehr. chaffeensis* sequence is included (BP = 87.8%) or not (BP = 93.4%). Moreover, maximum likelihood, unlike the other two methods, provides reasonable support for the α-proteobacteria being a paraphyletic group that includes the mitochondria, consistent with published phylogenies based on both cpn60 (9, 30) and small subunit ribosomal RNA sequences (10).

DISCUSSION

We have demonstrated the presence in *E. histolytica* of two genes encoding proteins that in other eukaryotes are located

in the mitochondrion. Phylogenetic analysis of the *E. histolytica* cpn60 sequence strongly supports a mitochondrial origin for that gene. To reconcile these data with primitive absence of mitochondria from the *Entamoeba* lineage, the ribosomal RNA-based tree placement would have to be artifactual and lateral transfer of genes from a bacterium related to the mitochondrial lineage to the *Entamoeba* nucleus would have to have taken place. In contrast, secondary loss of mitochondrial functions explains the presence of the PNT and cpn60 genes in *E. histolytica* and is consistent with the relatively late branch point suggested by the small-subunit ribosomal RNA phylogenetic tree (3). The *E. histolytica* iron-containing superoxide dismutase, previously described as a separate lateral transfer from bacterium to eukaryote (34), may represent a third *E. histolytica* gene acquired during mitochondrial endosymbiosis, as it also has been shown to cluster with proteobacterial sequences.

Secondary loss is also consistent with the presence in *E. histolytica* of ubiquinone (35, 36), which is most commonly mitochondrially located in eukaryotes. Although found in significantly lower amounts than in mitochondrion-bearing eukaryotes, ubiquinone appears to form part of the *E. histolytica* electron transport chain. The presence of mitochondria in the cytoplasm of *Entamoeba* species was reported in the preelectron microscopy era by using vital stains and cytochemistry (37–40). The possibility that these structures, reported to be ovoid bodies of approximately 1 μm in length, represent mitochondrially derived organelles with a reduced array of functions deserves further attention.

We have shown that the *Entamoeba* lineage may have secondarily lost many mitochondrial functions by identifying genes that encode mitochondrially located proteins in other eukaryotes. This approach could reveal a similar story in other organisms. Three groups of early diverging eukaryotic microorganisms are widely thought to be primitively amitochondriate: the trichomonads and the archezoan diplomonads and microsporidia (3, 41). A protein that is immunologically related to cpn60 has been reported in the diplomonad genus *Giardia* (42, 43), but, unfortunately, sequence information for this protein is not yet available for phylogenetic analysis. This raises the possibility that, like *E. histolytica*, diplomonads and perhaps other amitochondriate eukaryotes are derived from mitochondrion-bearing ancestors. Our results also indicate that, rather than being illustrative of a "living metabolic fossil" (44), many of the biochemical characteristics *E. histolytica* shares with amitochondriate eukaryotes (2) are likely to be the result of secondary convergence.

We especially thank L. S. Diamond, in whose laboratory the experimental work was undertaken, for encouraging the pursuit of what initially seemed like a longshot project. We thank O. Feeley for access to a Sun Microsystems Workstation; R. Ashworth for DNA sequence analysis; J. Samuelson for the *E. histolytica* genomic DNA library; and M. Müller, J. E. Logsdon, T. E. Nash, and M. R. Mowatt for comments on the work. A.J.R. was supported by a 1967 Science and Engineering Scholarship from the Canadian Natural Sciences and Engineering Research Council, and a grant (MT4467) to W. F. Doolittle from the Medical Research Council of Canada.

1. Martínez-Palomo, A. (1986) in *Amebiasis*, ed. Martínez-Palomo, A. (Elsevier, Amsterdam), pp. 12–43.
2. Müller, M. (1988) *Annu. Rev. Microbiol.* **42,** 465–488.
3. Cavalier-Smith, T. (1993) *Microbiol. Rev.* **57,** 953–994.
4. Sogin, M. L. (1991) *Curr. Opin. Genet. Dev.* **1,** 457–463.
5. Meza, I. (1992) *Arch. Med. Res.* **23,** 1–5.
6. Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N. & Miyata, T. (1993) *J. Mol. Evol.* **36,** 380–388.
7. Bakker-Grunwald, T. & Wöstmann, C. (1993) *Parasitol. Today* **9,** 27–31.
8. Tyler, D. (1992) *The Mitochondrion in Health and Disease* (VCH, New York).
9. Viale, A. M. & Arakaki, A. K. (1994) *FEBS Lett.* **341,** 146–151.
10. Olsen, G. J., Woese, C. R. & Overbeek, R. (1994) *J. Bacteriol.* **176,** 1–6.
11. Diamond, L. S., Harlow, D. R. & Cunnick, C. C. (1978) *Trans. R. Soc. Trop. Med. Hyg.* **72,** 431–432.
12. Clark, C. G. (1992) in *Protocols in Protozoology*, eds. Lee, J. J. & Soldo, A. T. (Allen, Lawrence, KS), Vol. 1, pp. D-3.1–D-3.2.
13. Clarke, D. M., Loo, T. W., Gillam, S. & Bragg, P. D. (1986) *Eur. J. Biochem.* **158,** 647–653.
14. Yamaguchi, M., Hatefi, Y., Trach, K. & Hoch, J. A. (1988) *J. Biol. Chem.* **263,** 2761–2767.
15. Swofford, D. L. (1993) PAUP; Phylogenetic Analysis Using Parsimony (Illinois Nat. Hist. Surv., Champaign, IL), Version 3.1.
16. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Univ. of Washington, Seattle), Version 3.5c.
17. Adachi, J. & Hasegawa, M. (1992) *Computer Science Monographs* (Inst. of Statist. Math., Tokyo), No. 27.
18. Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Comput. Appl. Biosci.* **8,** 275–282.
19. Kishino, H., Miyata, T. & Hasegawa, M. (1990) *J. Mol. Evol.* **30,** 151–160.
20. Jackson, J. B. (1991) *J. Bioenerg. Biomemb.* **23,** 715–741.
21. Harlow, D. R., Weinbach, E. C. & Diamond, L. S. (1976) *Comp. Biochem. Physiol. B* **53,** 141–144.
22. Yu, Y. & Samuelson, J. (1994) *Mol. Biochem. Parasitol.* **68,** 323–328.
23. Bruchhaus, I., Leippe, M., Lioutas, C. & Tannich, E. (1993) *DNA Cell Biol.* **12,** 925–933.
24. Williams, R., Cotton, N. P. J., Thomas, C. M. & Jackson, J. B. (1994) *Microbiology (Reading, U.K.)* **140,** 1595–1604.
25. Kramer, R. A., Tomchak, L. A., McAndrew, S. J., Becker, K., Hug, D., Pasamontes, L. & Hümbelin, M. (1993) *Mol. Biochem. Parasitol.* **60,** 327–332.
26. Vermeulen, A. N., Kok, J. J., van den Boogaart, P., Dijkema, R. & Claessens, J. A. J. (1993) *FEMS Microbiol. Lett.* **110,** 223–230.
27. Yamaguchi, M., Hatefi, Y., Trach, K. & Hoch, J. A. (1988) *Biochem. Biophys. Res. Commun.* **157,** 24–29.
28. Länge, S., Rozario, C. & Müller, M. (1994) *Mol. Biochem. Parasitol.* **66,** 297–308.
29. Lahti, C. J., Bradley, P. J. & Johnson, P. J. (1994) *Mol. Biochem. Parasitol.* **66,** 309–318.
30. Viale, A. M., Arakaki, A. K., Soncini, F. C. & Ferreyra, R. G. (1994) *Int. J. Syst. Bact.* **44,** 527–533.
31. Felsenstein, J. (1978) *Syst. Zool.* **27,** 401–410.
32. Hillis, D. M., Huelsenbeck, J. P. & Cunningham, C. W. (1994) *Science* **264,** 671–677.
33. Hasegawa, M. & Fujiwara, M. (1993) *Mol. Phylogenet. Evol.* **2,** 1–5.
34. Smith, M. W., Feng, D.-F. & Doolittle, R. F. (1992) *Trends Biochem. Sci.* **17,** 489–493.
35. Weinbach, E. C., Takeuchi, T., Claggett, C. E., Inohue, F., Kon, H. & Diamond, L. S. (1980) *Arch. Invest. Méd.* **11,** Suppl. 1, 75–81.
36. Ellis, J. E., Setchell, K. D. R. & Kaneshiro, E. S. (1994) *Mol. Biochem. Parasitol.* **65,** 213–224.
37. Causey, D. (1925) *Univ. Calif. Publ. Zool.* **28,** 1–18.
38. Ray, H. N. & Sen Gupta, P. C. (1954) *J. Indian Med. Assoc.* **23,** 529–533.
39. Ray, H. N., Mukherjee, A. K., Sen Gupta, P. C. & Das Gupta, B. (1955) *Bull. Calcutta Sch. Trop. Med.* **3,** 63–64.
40. Ray, H. N., Guha, A., Mukherjea, A. K. & Sen Gupta, P. C. (1955) *Bull. Calcutta Sch. Trop. Med.* **3,** 171.
41. Patterson, D. J. & Sogin, M. L. (1992) in *The Origin and Evolution of Prokaryotic and Eukaryotic Cells*, eds. Hartman, H. & Matsuno, K. (World Scientific Publishers, Singapore), pp. 13–46.
42. Reiner, D. S., Shinnick, T. M., Ardeshir, F. & Gillin, F. D. (1992) *Infect. Immun.* **60,** 5312–5315.
43. Soltys, B. J. & Gupta, R. S. (1994) *J. Parasitol.* **80,** 580–590.
44. Reeves, R. E. (1984) *Adv. Parasitol.* **23,** 105–142.