Enolase from *Trypanosoma brucei*, from the Amitochondriate Protist *Mastigamoeba balamuthi*, and from the Chloroplast and Cytosol of *Euglena gracilis*: Pieces in the Evolutionary Puzzle of the Eukaryotic Glycolytic Pathway

Véronique Hannaert, Henner Brinkmann,† Ulrich Nowitzki,‡ Jennifer A. Lee,§ Marie-Astrid Albert,* Christoph W. Sensen,‖ Terry Gaasterland,§ Miklós Müller,§ Paul Michels,* and William Martin‡*

*Research Unit for Tropical Diseases, Christian de Duve Institute of Cellular Pathology, Department of Biochemistry, Université catholique de Louvain, Brussels, Belgium; †Institute of Evolutionary Biology, Department of Biology, Universität Konstanz, Konstanz, Germany; ‡Institut für Botanik, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany; §The Rockefeller University, New York; and ‖Institute of Marine Biosciences, National Research Council, Halifax, Nova Scotia, Canada

Genomic or cDNA clones for the glycolytic enzyme enolase were isolated from the amitochondriate pelobiont *Mastigamoeba balamuthi,* from the kinetoplastid *Trypanosoma brucei,* and from the euglenid *Euglena gracilis.* Clones for the cytosolic enzyme were found in all three organisms, whereas *Euglena* was found to also express mRNA for a second isoenzyme that possesses a putative N-terminal plastid-targeting peptide and is probably targeted to the chloroplast. Database searching revealed that *Arabidopsis* also possesses a second enolase gene that encodes an N-terminal extension and is likely targeted to the chloroplast. A phylogeny of enolase amino acid sequences from 6 archaebacteria, 24 eubacteria, and 32 eukaryotes showed that the *Mastigamoeba* enolase tended to branch with its homologs from *Trypanosoma* and from the amitochondriate protist *Entamoeba histolytica.* The compartment-specific isoenzymes in *Euglena* arose through a gene duplication independent of that which gave rise to the compartment-specific isoenzymes in *Arabidopsis,* as evidenced by the finding that the *Euglena* enolases are more similar to the homolog from the eubacterium *Treponema pallidum* than they are to homologs from any other organism sampled. In marked contrast to all other glycolytic enzymes studied to date, enolases from all eukaryotes surveyed here (except *Euglena*) are not markedly more similar to eubacterial than to archaebacterial homologs. An intriguing indel shared by enolase from eukaryotes, from the archaebacterium *Methanococcus jannaschii,* and from the eubacterium *Campylobacter jejuni* maps to the surface of the three-dimensional structure of the enzyme and appears to have occurred at the same position in parallel in independent lineages.

Introduction

The Embden-Meyerhof-Parnas pathway of glycolysis is the backbone of energy metabolism (ATP synthesis) in eukaryotes. Enolase (2-phospho-D-glycerate hydrolase; EC 4.2.1.11) catalyzes the Mg^{2+}-dependent dehydration of 2-phosphoglycerate to phosphoenolpyruvate, which is converted by pyruvate kinase into pyruvate with the concomitant generation of ATP in the subsequent, final step of glycolysis. Enolase also catalyzes the reverse reaction during gluconeogenesis, but in addition appears to have been recruited for other functions in at least two lineages: it is a major lens protein in some vertebrates (Wistow et al. 1988), and it plays a role in the acquisition of thermal tolerance in yeast (Iida and Yahara 1985). The enzyme from a large variety of organisms, including archaebacteria, eubacteria, and eukaryotes has been studied and is highly conserved. Although octameric enzymes have been described for *Bacillus subtilis* (Brown et al. 1998) and *Thermotoga maritima* (Schurig et al. 1995), in all eukaryotes and many prokaryotes, enolase is biologically active as a dimer, with subunits having an M_r of approximately 45,000. Catalytically active dimers may be generated with products of different enolase genes, for example, in *Saccharomyces cerevisiae* (McAlister and Holland 1982). Crystal structures for the enzyme are known from lobster and *S. cerevisiae* at 2.2 and 2.0 Å resolution, respectively (Duquerroy, Camus, and Janin 1995; Zhang et al. 1997).

In the vast majority of eukaryotes studied to date, glycolysis is a cytosolic pathway. This is always true in eukaryotes that lack organellar compartments involved in ATP synthesis, such as *Giardia lamblia* (Müller 1998) or *Mastigamoeba balamuthi,* an amoeboflagellate protist that possesses a strikingly barren cytoplasm and lacks morphologically recognizable mitochondria (Chavéz, Balamuth, and Gong 1986; Brugerolle 1993). However, among other protists, there are some notable exceptions to be found with regard to the compartmentation of glycolysis. In *Chlamydomonas reinhardtii,* for example, all glycolytic enzyme activites studied to date were found to be localized in the chloroplast, rather than in the cytosol (Schnarrenberger et al. 1990). A better-known and well-studied exception is the kinetoplastids (trypanosomes and related organisms), in which most of the gly-

Abbreviations: FBA, fructose-1,6-bisphosphate aldolase; FBP, fructose-1,6-bisphosphatase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GPI, glucose-6-phosphate isomerase; ORF, open reading frame; PGK, phosphoglycerate kinase; PCR, polymerase chain reaction; PFK, phosphofructokinase; PGM, phosphoglycerate mutase; PP_i-PFK, PP_i-linked phosphofructokinase; PVP, polyvinylpyrrolidone; SDS, sodium dodecyl sulphate; TPI, triosephosphate isomerase.

Key words: glycolysis, metabolism, amitochondriate protist, phylogeny, evolution, organelles, endosymbiosis.

Address for correspondence and reprints: William Martin, Institut für Botanik, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1, D-40225 Düsseldorf, Germany. E-mail: w.martin@tu-bs.de.

Mol. Biol. Evol. 17(7):989–1000. 2000
© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

colytic pathway is compartmentalized in specialized microbodies, glycosomes (Opperdoes and Borst 1977; Hannaert and Michels 1994; Clayton and Michels 1996). In these organisms, only the last three enzymes of the pathway leading to pyruvate production, including enolase, are found in the cytosol. In relatives of the kinetoplastids, the euglenids, which possess plastids but not glycosomes, some of the enzymes that are common to glycolysis and the Calvin cycle occur as distinct chloroplast-cytosol isoenzyme pairs (Kitaoka et al. 1989; Henze et al. 1995; Plaumann et al. 1997), as is also found in many higher plants (Martin and Schnarrenberger 1997). Several key glycolytic enzymes from trypanosomes and amitochondriate protists differ in their regulatory properties from those of other eukaryotes, e.g., yeast or vertebrates, differences that appear to correlate with the lifestyle of these organisms (Mertens, van Schaftingen, and Müller 1992; Mertens 1993; Hannaert and Michels 1994; Michels and Hannaert 1994; Clayton and Michels 1996; Bakker et al. 1997; Park, Schofield, and Edwards 1997).

The study of the evolution of pathways requires sequences for all of the enzymes involved. For the kinetoplastid *Trypanosoma brucei,* all glycolytic enzymes except enolase have been cloned. For amitochondriate protists and *Euglena,* whose glycolytic pathways differ from kinetoplastids not only with respect to compartmentation, but also with respect to regulation of the enzymes involved (Kitaoka et al. 1989; Bakker et al. 1997), fewer sequences for glycolytic enzymes are known. Here we report the molecular analysis of cytosolic enolase from the amitochondriate protist *M. balamuthi* and *T. brucei* and two isoenzymes of enolase from *Euglena gracilis,* one of which possesses a putative N-terminal plastid-targeting peptide. A protein phylogeny of enolase sequences from eubacterial, archaebacterial, and eukaryotic sources is presented that reveals insights into the complex evolutionary history and subcellular compartmentation of this glycolytic enzyme.

Material and Methods

Cloning and Molecular Analysis of *T. brucei* Enolase

A search in the EST subset of the GenBank nucleic acid database revealed a 369-bp *T. brucei rhodesiense* sequence (accession number AA023828; cDNA clone number T3246) coding for an amino acid sequence of which 122 residues showed 55% identity with the C-terminal part of human enolase. Based on this sequence, two oligonucleotides containing a *Bam*H1 site (italicized) were designed: 5′-CGC*GGATCC*CCAATTGTCAGTATTGAAGACCC-3′ and 5′-CGC*GGATCC*CCTCTGTCTCACCACTGCGG-3′. PCR amplification was performed using genomic DNA from *T. brucei* stock 427 as template with initial denaturation for 5 min at 95°C, followed by 30 cycles of 30 s at 95°C, 1 min at 60°C, and 1 min at 72°C and, finally, 10 min of incubation at 72°C. A major amplified product of 300 bp was cloned into pTZ19R (Amersham Pharmacia Biotech, Sweden), sequenced, and used to screen a genomic library of *T. brucei* in *Escherichia coli* in λGEM11 (Promega) (Michels et al. 1991). Hybridization was performed at stringent conditions as described previously (Kohl et al. 1996). Six positive recombinant phages were purified and rescreened. A 2.5-kb *Pst*I fragment from one of the hybridizing phages was subcloned into pTZ19R and sequenced on both strands.

Cloning and Molecular Analysis of *E. gracilis* Enolase Isoenzymes

Axenic *E. gracilis* cultures (SAG 1224-5/25) were grown as described (Henze et al. 1995) under a 14:10 h light : dark regime. Isolation of mRNA through two rounds of oligo-dT cellulose chromatography and synthesis of cDNA for amplification was performed as described (Henze et al. 1995) but without the *Not*I adapters. Degenerate primers 1F (5′-GAYWSIMGIGGIAAYCCIACIGTIGAR-3′) and 3R (5′RTCYTCIGTYTCICCISWICKRTG-3′) were designed against the conserved amino acid sequence motifs ''DSRGNPTVE'' and ''HRSGETED'' from the N-terminal and C-terminal regions, respectively, of an alignment of enolase sequences extracted from GenBank. Amplification was performed for 35 cycles of 2 min at 94°C, 1 min at 50°C, and 2 min at 72°C in 25 μl containing 10 μM of each primer, 50 μM dNTP, 1 mM $MgCl_2$, 10 ng cDNA, and 0.5 U *Taq* polymerase (Perkin-Elmer) using the supplier's buffer. The 1.1-kb fragment was subcloned and used to screen 300,000 recombinants of a *Euglena* cDNA library (Henze et al. 1995). From 150 hybridizing positives, 15 were subcloned and sequenced. One corresponded to pEgEno29. The other 14 represented an mRNA encoding a different enolase sequence, the largest clone of which was designated pEgEno02.

For Northern blots, inserts of pEgEno02 and pEgEno29 were purified by electroelution. Five micrograms of polyA+ mRNA purified through two rounds of oligo-dT cellulose affinity chromatography was loaded per lane on the same formaldehyde gel, electrophoresed as described (Sambrook, Fritsch, and Maniatis 1989), and transferred to Hybond N (Amersham) membranes according to the manufacturer's protocol. Lanes were stained, photographed, cut, and hybridized separately at 70 °C in 3 × SSPE, 0.1% SDS, 0.02% (w/v) PVP, 0.02% (w/v) Ficoll 400, 50 μg/ml polyadenylic acid, and 10 ng/ml of probe labeled to 5×10^7 cpm/μg. Washing was at 70 °C in 2 × SSPE, 0.1% SDS.

Spinach Enolase

Using cDNA from 10-day-old spinach seedlings, the primers 1F and 3R, described above for *Euglena,* were used under the same conditions to amplify a 1.1-kb fragment from spinach. The amplification product was cloned as above and used to screen a spinach cDNA library prepared as described (Nowitzki et al. 1998). Ten independent positive clones were identified, subcloned, and shown by terminal sequencing to represent the same mRNA. The longest of these, pSoEno8.3, was sequenced.

Organism	Localization	Protein	Length	Sequence
Euglena nuclear-encoded proteins	chloroplast	FbaI	(142)	<<TFVAPAAQSSNTFATSS**VAASIGMVMGAGAVLLA**RMNQKPVAMNAWTGSVYGVL|>>
		Lhcp	(142)	<<EKSTMGS**LLMIAAAGVAAAVAFVW**KSVPRQQDSVINVPLLPVSVATMATSGKKS|>>
		RbcS	(135)	<<PLMAASVGAESESRRWVAS**AILFPLSGLFAAVAL**KMAMMKPKVAAVLPFTSEKD|>>
		Hmbs	(139)	<<STVKGQERTIT**ILGVCSFLSASLFYIW**KQFGMKARTTKPADLQEVSGGRIWSLA|>>
		Ps30	(118)	<<EDPE**AVLAGAGRAMGAALIGAAVAG**SANAASLTYDELQSLSYLEVKSSGIAGTC|>>
		GapA	(147)	<<**PVAAAAAFFAY**KKGQSDAQTYEVADPAGVQTYGTELVFAMNATIGNKYPGNSNT|>>
Euglena	cytosolic	Enolase02		MPVIKTVKAR>>
(probable)	plastid	Enolase29	(~130)	<<**FFSILPIMA**NNTADIQYLRQHNIPELVNDLIQQLVTTKPAQPVAFLRDVLSKRK|AVIKVVKAR>>
Arabidiopsis	cytosolic	Enolase 1		MATITVVKAR>>
(probable)	plastid	Enolase 2		MSVQEYLDKHMLSRKIEDAVNAAVRAKTSDPVLFIANHLKKAVS|SVITKVKAR>>
T. brucei	cytosol	Enolase		MTIQKVH>>

FIG. 1.—N-terminal regions of enolase sequences. Plastid-targeting peptide-coding regions of nuclear-encoded chloroplast proteins from *Euglena gracilis* are shown (data summarized from Plaumann et al. [1997], references therein, and this paper). Numbers in parentheses indicate the lengths of the plastid-targeting peptides in amino acids. The second of two hydrophobic stretches in plastid-targeting peptides of *Euglena* are indicated in boldface type. "<<" and ">>" indicate that only part of the sequence is shown, except in the case of *Euglena* 29, for which a full-size clone was not obtained. "|" indicates the inferred processing site of the plastid-targeting peptides. Abbreviations of nuclear-encoded chloroplast proteins are as follows: FbaI, class I fructose-1,6-bisphosphate aldolase; Lhcp, light harvesting complex protein; RbcS, small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase; Hmbs, hydroxymethylbilane synthase; Ps30, photosystem protein; GapA, $NADP^+$-dependent GAPDH.

Enolase from *M. balamuthi*

Mastigamoeba balamuthi identical to the American Type Culture Collection isolate ATCC 30984, was obtained from Dr. Frederick Schuster (Brooklyn College of the City University of New York, Brooklyn, N.Y.). cDNA clone PHR60, containing the complete ORF for enolase, was identified in our ongoing *M. balamuthi* cDNA sequencing project, to be described elsewhere. The insert in this clone was sequenced on both strands by primer walking.

Sequence Handling and Phylogenetic Analysis

From sequences extracted from GenBank, an alignment was produced with CLUSTAL W (Thompson, Higgins, and Gibson 1994) that was manually refined using the program ED of the MUST package (Philippe 1993). Regions of uncertain alignment were omitted from the analysis, leaving 400 amino acid positions for analysis. Protein phylogeny was inferred using protein maximum-likelihood (ML) as implemented in MOLPHY, version 2.3 (Adachi and Hasegawa 1996), using local rearrangement starting with the neighbor- joining (NJ) tree and using the best tree obtained from 1,000 PROTML topologies generated using the quicksearch option of MOLPHY using the JTT-F model. Distance analyses were performed with the MUST package using the Kimura correction and the NJ method (Saitou and Nei 1987) using 1,000 bootstrap replicates. Parsimony analysis was performed using PAUP* with 500 bootstrap replicates and 10 times random addition (Swofford 1999). The significance of different ML topologies was tested using the Kishino-Hasegawa test implemented in the MOLPHY package. Accession numbers for all sequences and the alignment are available via anonymous ftp from 134.169.70.80/ftp/pub/incoming/enolase.

Molecular Modeling

The structure of yeast enolase (Wedekind, Reed, and Rayment 1994) was downloaded from the Protein Data Bank (http://www.rcsb.org/pdb/) and visualized using RASMOL (http://www.umass.edu/microbio/rasmol/) on a Macintosh computer.

Results

Enolase from *T. brucei*

Based on an EST sequence of a *T. brucei rhodesiense* cDNA fragment available in public databases, oligonucleotides were designed to amplify a homologous hybridization probe from genomic DNA of *T. brucei.* The amplified fragment was identical in sequence to that of the partial cDNA of *T. b. rhodesiense* present in the EST database and was used to screen a λGEM11 *T. brucei* genomic DNA library. Among hybidizing clones, restriction mapping revealed a 2.5-kb *Pst*I fragment that contained an open reading frame (ORF) for enolase encoding a polypeptide of 428 amino acids (excluding the initiator methionine) with a molecular mass of 46,461, a pI of 6.2, and a net charge of −5.

Enolase from the Amitochondriate Protist *M. balamuthi*

Of the approximately 200 clones sequenced from a random cDNA library of *M. balamuthi,* six clones encoded a typical enolase. Of these, one clone was completely sequenced and found to code for a putative enolase of 438 amino acid residues with a molecular mass of 47,596, a pI of 6.7, and a net charge of −3.

Cytosolic and Plastid-Targeting Peptide-Bearing Isoforms of Enolase in *Euglena*

Grown in rich medium in the light under aerobic conditions, *Euglena* expresses two different genes for enolase. The sequence of pEgEno02 is 1,523 bp long, including a 21-base polyA stretch, and encodes an ORF of 431 amino acids with a molecular mass of 46,501, a pI of 5.79, and a net charge of −7. pEgEno29 is 1,622 bp long with a 10-base polyA tail and encodes an ORF of 485 amino acids that lacks a start codon near the N-terminus of pEgEno02. Instead, the reading frame extends beyond the start codon of pEgEno02 by at least 53 amino acids. This region extends into, but terminates in, a highly hydrophobic region (fig. 1). The plastid targeting-peptides of nuclear-encoded chloroplast proteins in *Euglena* are roughly 120–150 amino acids long and possess two highly hydrophobic regions (Plaumann et

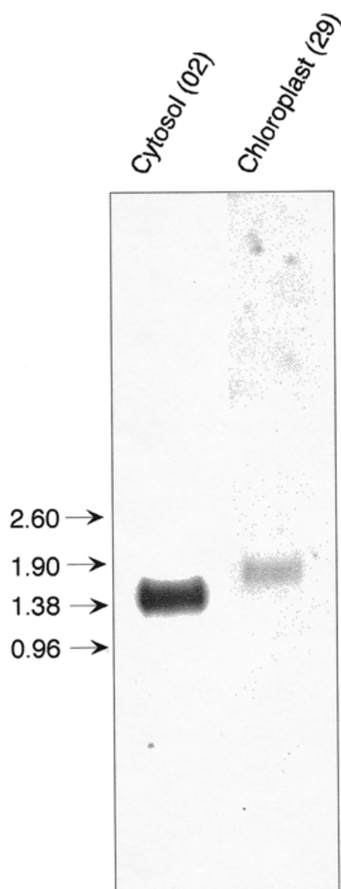

FIG. 2.—Northern blot of *Euglena gracilis* mRNA probed with cDNAs for cytosolic (02) and plastid (29) enolases. Sizes of molecular mass standards are indicated. Exposure was 2 days at −70°C.

al. 1997) for transport across the three membranes that surround the plastids of *Euglena* (Schwartzbach, Osafune, and Löffelhardt 1998). The position of the proximal hydrophobic region in the protein encoded by pEgEno29 relative to the N-terminus of other enolases is very similar to the position noted in the plastid-targeting peptide of *Euglena* chloroplast glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Henze et al. 1995), suggesting that this cDNA might encode a chloroplast isoform. Assuming the processing site indicated in figure 1, the molecular mass of the mature protein would be 46,751, and its pI would be 7.12, with a net charge of −1.

The Northern blot revealed that the polyadenylated mRNA for pEgEno29 was 1.9 kb long, ∼400 bases longer than that of pEgEno02 (fig. 2). This difference corresponds to a putative plastid-targeting peptide-coding region encompassing roughly 130 amino acids, in agreement with the lengths of other nuclear-encoded chloroplast proteins from *Euglena.* Although the chloroplast localization was not shown directly, the similarity in pattern to chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Henze et al. 1995) and chloroplast and cytosolic fructose-1,6-bisphosphate aldolase (FBA) (Plaumann et al. 1997) suggests that pEgEno29 and pEgEno02 encode chloroplast and cytosolic isoenzymes of enolase, respectively, from *Euglena.* This is consistent with the findings (1) that mitochondrial transit peptides in *Euglena* are roughly 30–50 amino acids long (Mukai et al. 1989; Cui et al. 1994), much shorter than its chloroplast plastid-targeting peptides; (2) that among eukaryotes, glycolytic enzymes are almost never imported into mitochondria, the exception being diatoms (Liaud et al. 2000); and (3) that other glycolytic/gluconeogenetic enzymes are known to exist in *Euglena* plastids, including triosephosphate isomerase (TPI) (Mo, Harris, and Gracy 1973), fructose-1,6-bisphosphatase (FBP) (Latzko and Gibbs 1969), GAPDH (Henze et al. 1995), and FBA (Plaumann et al. 1997).

Alignment and Phylogenetic Analysis

The alignment of enolase sequences revealed two length polymorphisms with notable phylogenetic distribution, as shown in figure 3 (see *Discussion*). One thousand trees were generated with the quicksearch option of MOLPHY and evaluated using the JTT-F model. Using the Kishino-Hasegawa test implemented in MOLPHY, 585 of the trees were not significantly worse than the best tree found, which had a log likelihood (lnL) of −26,714.8. Using that topology for local rearrangement with the JTT-F model, the topology in figure 4 was found, the lnL of which (−26,671.97) was slightly, but insignificantly, better than the lnL value for the starting topology. Using the NJ tree of Kimura distances as the starting topology, an ML tree was found that also was not significantly different at the 5% level from the topology in figure 4.

Discussion

Indels in Enolase Sequences and Evidence for Plastid Isoforms in *Euglena* and *Arabidopsis*

Enolase is one of the most conservatively evolving glycolytic enzymes (Fothergill-Gilmore and Michels 1993). With the availability of many new enolase sequences, including those from sequenced genomes, it was of interest to address the gene phylogeny of this enzyme. We extracted enolase homologs from the databases and produced an alignment encompassing 62 sequences and 400 sites. Overall, enolase sequences are highly conserved, showing roughly 40% amino acid identity across eukaryotes, eubacteria, and archaebacteria. Two notable indels found in enolase amino acid sequences are shown in figure 3.

The first of these, at positions 161–165 of our alignment, underscores the specificity of a five–amino acid insertion previously noted to be shared in cytosolic enolase from higher plants and enolase from the apicomplexans *Plasmodium falciparum* (Read et al. 1994) and *Toxoplasma gondii* (Dzierzinski et al. 1999). This indel suggests that apicomplexan parasites, which possess plastids (McFadden et al. 1996; McFadden and Roos 1999), probably obtained their nuclear enolase gene from the photosynthetic symbiont (Read et al. 1994; Dzierzinski et al. 1999), a view that receives additional support from the current analysis, due to the highly restricted phylogenetic distribution of the indel.

```
                           11111111111111111111111111111 ... 2222222222233333333333333333333333333333333
                           45555555555666666666677777777 ... 8999999999900000000001111111111222222222233
                           90123456789012345678901234567 ... 9012345678901234567890123456789012345678901

Rattus                     NLML-ELDGTEN-----KSKFGANAILGV ... EALELVKEAIDKAGYT------------EKMVIGMDVAASEFY
Homo                       NLML-ELDGTEN-----KSKFGANAILGV ... EALELVKEAIDKAGYT------------EKIVIGMDVAASEFY
Xenopus                    KLMI-EMDGTEN-----KSKFGANALLGV ... EALELLKTAINKAGYP------------DKIVIGMDVAASEFY
Gallus                     KLML-EMDGTEN-----KSKFGANAILGV ... EALELLKTAIGKAGYS------------DKVVIGMDVAASEFY
Loligo                     NFLL-ALDGTEN-----KNKLGANAILGV ... EGLELLKTAIANAGYT------------AEIEIGMDVAASEFC
Caenorhabditis*            DFMM-ALDGSEN-----KGNLGANAILGV ... EGLDLLNTAIDKAGYT------------GKISIGMDVAASEFF
Penaeus                    DFMC-KLDGTEN-----KSRLGANAILGV ... DALTLIQESIEKAGYT------------GKIEIGMDVAASEFY
Drosophila                 NFMI-KLDGTEN-----KSKFGANAILGV ... EALNLISDAIAKAGYT------------GKIEIGMDVAASEFY
Schistosoma japonica       QFMI-ELDGKEN-----KGNLGANAILGV ... KGLQLLEEAIKIAGYT------------GKVEIGMDCAASEYY
Schistosoma mansonii       KYMI-DLDGTEN-----KEKLGANAILGV ... KGLQLLEEAIKIAGYT------------GKVEIGMDCAASEFH
Fasciola hepatica          KFML-DLDGTPN-----KEKLGANAILGV ... EGLELLRTAIDKAGYT------------GKVKIAMDCAASEFY
Neocallimastix             DFLL-KLDGTPN-----KSKLGANAILGV ... EGLELLNEAIAKAGYT------------GKVKIGMDVASSEFY
Saccharomyces*             DFLI-SLDGTAN-----KSKLGANAILGV ... EALDLIVDAIKAAGHD------------GKVKIGLDCASSEFF
Candida                    EFLL-SLDGTPN-----KSKLGANAILGV ... EALDLIMDAIDKAGYK------------GKVGIAMDVASSEFY
Aspergillus                EFLK-KLDGSAN-----KSNLGANAILGV ... EALDLITEAIEQAGYT------------GKMKIAMDVASSEFY
Cladosporium               AFLN-KLDGTTN-----KTKIGANAILGV ... EALDLITDAIEEAGYT------------GQIKIAMDVASSEFY
Schizosaccharomyces        EFLL-KLDGTEN-----KSKLGANAILGV ... EALDLIVEAINKAGYE------------GKIKIGLDVTSSEFY
Pneumocystis               KFLI-DLDGTSN-----KSRLGANAILGV ... EALDLIVTAIEEAGYT------------GKIKIALDVAASEFY
Arabidopsis 1 cytosol      NFMVHELDGTQNEWGWCKQKLGANAILAV ... EGLELLKTAIEKAGYT------------GKVVIGMDVAASEFY
Spinach cytosol            NFMVQELDGTTNEWGWCKQKLGANAILAV ... EGLELLKTAIEKAGYT------------GKVVIGMDVAAAEFY
Ricinus                    NFMVQELDGTVNEWGWCKQKLGANAILAV ... EGLELLKTAIAKAGYT------------GKVVIGMDVAASEFY
Mesembryanthemum           NFMVQQLDGTVNEWGWCKQKLGANAILAV ... EGLELLKTAIEKAGYT------------GKVVIGMDVAASEFY
Zea                        NFMVQQLDGTSNEWGWCKQKLGANAILAV ... EGLELLKAAIEKAGYT------------GKVVIGMDVAASEFF
Lycopersicon               NFMVHQLDGTQNEWGWCKEKLGANAILAV ... EGLELLKTAIEKAGYT------------GKVVIGMDVAASEFY
Plasmodium                 NLMVEELDGSKNEWGWSKSKLGANAILAI ... EALDLLVTAIKSAGYE------------GKVKIAMDVAASEFY
Toxoplasma                 MLMVEQLDGTKNEWGYSKSKLGANAILGV ... EALNLLVEAIKAAGYE------------GKIKIAFDAAASEFY
Chlamydomonas              QKMK-DLDGTDN-----KGKLGANAILAV ... EGLNLVNEAIEKAGYT------------GKVKIGMDVASSEFY
Arabidopsis 2 plastid      QAMI-DLDKTEK-----KSELGANAILAV ... EGLELVKEAINRTGYN------------DKIKIAIDIAATNFC
Entamoeba                  EMMI-KLDGTNN-----KGKLGANAILGC ... EALDLLVEAIAKAGYT------------GKIEIAMDCAASEFY
Trypanosoma brucei         TLML-RLDGTPN-----KGKLGANAILGC ... EPLPILMEAIEEAGHR------------GKFAICMDCAASETY
Mastigamoeba balamuthi     RLML-QIDGTEN-----KTHLGANAILGC ... EGLELIVEAIKQAGYT------------GKIEIGMDVAASSFW
Methanococcus*             TIMI-ELDGTPN-----KSRLGANAILAV ... EALDLLTESVKKAGYE------------DEVVFALDAAASEFY
Pyrococcus furiosus        ALLL-ELDGTEN-----KSNVGANAILAV ... EPLDVLIKAIEEAGYKPG----------DEIALALDVASSELF
Pyrococcus horikoshii*     KLLI-ELDGTEN-----KSNLGANAILAV ... EPLDALIKAIEEAGYKPG----------DEIALAIDAASSEFF
Haloarcula                 TALH-AADGTDD-----FSGIGANSAVAI ... EAFEIMDEAVETVADDFG----------FAISFGLDVARAELY
Archaeoglobus*             EALR-ELDGTDN-----FSRIGGNFAITA ... QAFEILSEAIQRVEDELG----------VKVRMGIDVAATELW
Methanobacterium*          LVLK-EIDGTEN-----LSSLGGNTVVAV ... DALEIQATACEEVTDELG----------VEVRPSLDLAASEFW
Mycobacterium*             QALV-DLDGTPD-----KSRLGGNAILGV ... AALDLISRAIESAGLRPG----------ADVALALDAAATEFF
Synechocystis*             QAMI-DRDGTDN-----KKELGANAILGV ... QALDILIEAIEQAGYKPG----------SQIALAMDIAASEFF
Deinococcus                KALM-DVDGTSN-----KGNMGGNAILAV ... EALEVLLEAIQQAGYEPG----------KDICIALDPAVTELY
Euglena cytosol            NLMR-ELDGTPN-----KGKIGANGILGV ... EALDFVIAAIKAAGFEPG----------KQIAIALDCASSELF
Euglena plastid            ELMK-QIDGTPN-----KGKIGANGILGV ... EALEFVMKAIKAAGYTPG----------TQIGVALDCASSELF
Treponema*                 HAMR-NLDGTDN-----KSKLGANAMLGV ... QALEYIMKAIAKAGLAPR----------KDVCIALDCASSELF
Borrelia*                  RKML-ELDGTPT-----KEKLGANAILAV ... EACEVIIEAIKKAGYEPG----------KDIAIALDPATSELY
Chlorobium 1               AKLI-ELDGTPN-----KSKLGANAILGV ... QAIELVIEAIGMAGYKAGAPTDRGGLGDGHVMIALDPASSEFY
Haemophilus*               QIMI-DLDGTEN-----KSNFGANAILAV ... DALACIKEAVEKAGYVLG----------KDVTLAMDCASSEFY
Escherichia*               KIMI-DLDGTEN-----KSKFGANAILAV ... EALAVIAEAVKAAGYELG----------KDITLAMDCAASEFY
Nitrosomonas               KTLI-ELDGSEN-----KSRLGANAILAV ... EALALIVQAIDEAGYQPG----------SEVAIGVDCASSEFF
Pseudomonas                HAMI-ELDGTEN-----KAKLGANAILAV ... DALAAIAEAVEKAGYKLG----------DDVTLALDCASSEFF
Caulobacter                NLMI-ELDGTPN-----KARLGANAILGV ... AALDFIVKAGEKAGYKAG----------DDFVLGLDVASTEFF
Zymomonas                  QTMI-KLDGTPN-----KGKFGANAILGV ... SALDFIVDSISKAGYKPG----------EDVFIALDAASSEFY
Porphyromonas              RKLI-ELDGTKT-----KSNLGANAMLGV ... DAIESILKAVEAAGYVPG----------KDITIAMDCASSEFF
Bordetella                 RTLI-ELDGTDS-----KERLGANAMLAA ... AAIQLILKAITEAGYEPG----------TQIALGLDCASSEFY
Clostridium                QTML-EIDGTEN-----KSKLGANAMLGV ... EAIQIILEAVTKAGYEPG----------KDMFIAMDPASTEFY
Bacillus*                  QLLI-ELDGTEN-----KGKLGANAILGV ... EALQTIVEAIEKAGFKPG----------EEVKLAMDAASSEFY
Thermotoga                 KTLI-ELDGTEN-----KSKLGAHAILAV ... EAIKVLIEAIEKAGYKPG----------EEVFIALDCAASSFY
Aquifex*                   QILI-ELDGTEN-----KSKLGANAILGT ... EALDILMQAIEKAGYKPG----------EDILLALDVASSEFY
Campylobacter              DTLR-ELDGTNN-----YSNLGANATLGV ... EPIDLLMTCIKKAGYE------------NRVKIALDVASTEFF
Helicobacter 1*            ERLR-ALDGTPN-----YANLGANAVLGV ... EPLEIISQAIEKAGYKLG----------EEIALALDVASSELV
Helicobacter 2*            ERLR-ALDGTPN-----YANLGANAVLGV ... EPLEVISQAIEKAGYKLG----------EEIALALDVASSELV
Mycoplasma pneumoniae*     ALMI-QLDGTPN-----KAKLGANAILAV ... EALDLMVEAIKAAGYQPG----------SDIAIALDVAASEFY
Mycoplasma genitalium*     QAMI-KLDNTPN-----KAKLGANAILSV ... DALDIMVEAIKLAGYKPW----------DDIAIAIDVAASEFY
Chlorobium 2               YFLI-ELDETNN-----KSKLGANAILGV ... QAVEMILEGITKAGYKPS----------VDVSICLDPASSEMW
```

FIG. 3.—Conspicuous indels in enolase sequences. Position numbers in the alignment are indicated. Gaps are shown as dashes. Enolase sequences from completely sequenced genomes are indicated by asterisks.

Interestingly, two enolase sequences from plants, that from *Chlamydomonas* and a new enolase sequence from the *Arabidopsis* genome project (*Arabidopsis* 2 in fig. 3), do not share the insertion at positions 161–165. The *Arabidopsis* 2 protein also carries a transit peptide, suggesting that it represents a plastid isoform. Many higher plant tissues are known to possess distinct and separable chloroplast-cytosol isoenzymes of enolase, for example, developing *Ricinus* seeds (Miernyk and Dennis 1992) and developing barley leaves (Hoppe et al. 1993), but the plastid enzyme is usually not highly expressed in fully green tissues. A previous study indicated the lack of a plastid enolase in *Arabidopsis* (van der Straeten et al. 1991), but it seems likely that the plastid enzyme is expressed only in early stages of *Arabidopsis* development, as is the case in barley (Hoppe et al. 1993). Like the probable plastid isoform from *Arabidopsis,* the *Chlamydomonas* enolase also lacks the five–amino acid insertion characteristic of plant (and apicomplexan) cytosolic enolase. Furthermore, a study of subcellular distribution of several glycolytic enzymes in *Chlamydomonas* has shown that only a single isoform for each of the glycolytic enzymes glucose-6-phosphate isomerase (GPI), FBA, phosphoglycerate kinase (PGK), and TPI exists and that each is localized in the chloroplast (Schnarrenberger et al. 1990). Although enolase

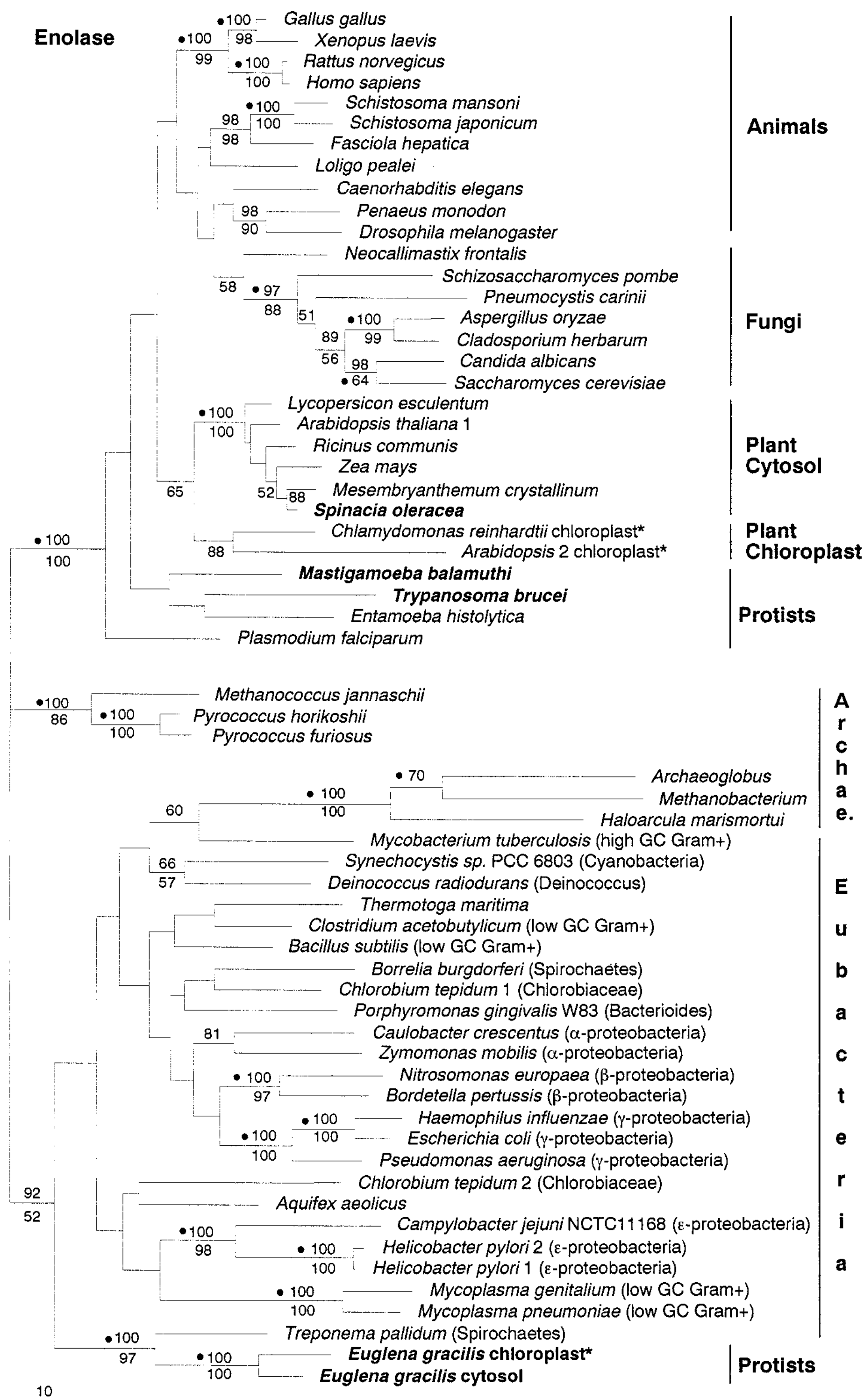

FIG. 4.—Phylogeny of enolase protein sequences constructed by maximum likelihood (ML) (Adachi and Hasegawa 1996). Numbers at nodes indicate bootstrap proportions, upper values indicate Kimura distances with neighbor-joining (1,000 replicates), and lower numbers indicate parsimony (see *Materials and Methods*). Solid dots indicate branches that were found in all 585 of the PROTML trees identified from quicksearch of MOLPHY that were not significantly worse than the best tree found at the 5% level. The scale bar at lower left indicates 10 substitutions. Sequences presented in this paper are given in boldface type. An asterisk indicates that the probable chloroplast enolase isoforms from *Chlamydomonas, Arabidopsis,* and *Euglena* have not been directly localized to the plastid (see text). Eubacterial groups as recognized by rRNA systematics are indicated.

was not assayed in that study, the lack of cytosolic isoenzymes for those glycolytic activities indicates that glycolysis in light-grown *Chlamydomonas* is primarily localized in the plastid, suggesting that the *Chlamydomonas* enolase is also a plastid isoform, consistent with its shared lack of the insertion at positions 161–165. Since the *Chlamydomonas* enolase sequence in the database lacks about 50 residues of the mature subunit, it is possible that the complete clone may possess a transit peptide.

Another notable indel occurs at positions 305–306 of the alignment. All enolase sequences from eukaryotic sources lack the motif ‘‘PG’’ (or its variants) at this position, with the notable exception of the *Euglena* isoenzymes. All prokaryotic sequences surveyed possess the PG motif, with the exception of the archaebacterium *Methanococcus* and the eubacterium *Campylobacter*. However, the positions of *Methanococcus, Campylobacter,* and eukaryotic enolases in the gene phylogeny (fig. 4) suggest that this indel does not represent a unique event in enolase evolution, but rather that this deletion has occurred several times independently. That independent indel events can occur at this position is furthermore seen by the unique event in the *Chlorobium* 1 sequence (fig. 3).

Conflict Between Indel Distribution and Protein Phylogeny in *Plasmodium* Enolase

The conspicuous indel in figure 3 that indicates common ancestry of the *Plasmodium* and plant cytosolic enolases is not consistent with the position of *Plasmodium* in the phylogeny of figure 4. Instead of branching with plant cytosolic enolase, the *Plasmodium* sequence tends to assume a basal position among eukaryotic homologs. Also in distance and parsimony analyses, the *Plasmodium* sequence did not branch with the plant homologs, but rather assumed a basal position, whereby it branched above the branch bearing *Trypanosoma, Mastigamoeba,* and *Entamoeba* in parsimony analysis and branched with *Trypanosoma* and *Entamoeba* when *Mastigamoeba* was excluded in PROTML.

The nature of the indel at positions 161–165 (fig. 3) suggests that *Plasmodium* enolase does indeed share a common ancestry with the plant cytosolic sequences and that its position in figure 4 may be a reconstruction artifact. Overall, the position of the *Plasmodium* sequence was rather unstable, and it seems likely that its basal position in figure 4 is due to a long-branch attraction artifact and is hence a misplacement. Using the Kishino-Hasegawa test, PROTML trees in which *Plasmodium* was forced with the plant sequences were not worse than the tree in figure 4 at the 5% level. In parsimony analyses of smaller data sets, consisting of 45 sequences through exclusion of many eubacterial sequences, the *Plasmodium* sequence shared a common branch with its homologs from higher plants, but with a bootstrap proportion of only 46%. In an analysis of the enolase sequence from the apicomplexan *T. gondii,* Dzierszinski et al. (1999) also found that with smaller data sets, the apicomplexan enolase sequences branched with plant homologs in parsimony analyses.

Among the 585 trees found with the quicksearch option of MOLPHY that were not significantly worse at the 5% level than the best tree found, only 23 grouped *Plasmodium* with the plant sequences. The position of *Plasmodium* enolase is thus difficult to resolve in phylogenetic inference from sequence data, with the conflict between the specific indel (fig. 3) and the position of *Plasmodium* enolase in the phylogeny suggesting that it is simply misplaced in figure 4. In light of the plastid that has been found in apicomplexans (McFadden et al. 1996; McFadden and Roos 1999), a link between apicomplexan and plant enolase, as evidenced by the indel at positions 161–165, is not surprising. Indeed, in analyses of a different glycolytic enzyme, GPI, sequences from apicomplexans did branch specifically with their plant homologs in PROTML analyses (Nowitzki et al. 1998; Dzierszinski et al. 1999).

Enolases from Amitochondriate Protists Branch with the *Trypanosoma* Homolog

The detection of enolase transcripts indicates that the core metabolism of *M. balamuthi* involves an Embden-Meyerhof-Parnas–type glycolysis. This conclusion is supported by the identification of cDNA clones from the EST project encoding PP_i-linked phosphofructokinase (PP_i-PFK), class II fructose1,6-bisphosphate aldolase, GAPDH, and phosphoglycerate mutase (PGM) (unpublished data). The organism contains no mitochondrion- or hydrogenosome-like structures (Chavéz, Balamuth, and Gong 1986; Brugerolle 1993) and grows in a cysteine-rich anaerobic medium, and lactate is one of its major end products (unpublished data). These characteristics permit an assignment of *M. balamuthi* to type I amitochondriate protists characterized by lack of compartmentalized energy metabolism (Martin and Müller 1998; Müller 1998).

Mastigamoeba balamuthi is a member of the amoeboflagellate group of pelobionts, which also contains the giant amitochondriate amoeba *Pelomyxa palustris.* Because of its simple cytological makeup, this group of eukaryotes is often regarded as ‘‘primitive’’ and is placed with other mitochondrion-lacking eukaryotes in some taxonomic schemes (Cavalier-Smith 1991, 1993). The evolutionary position of these organisms, however, remains unresolved (Embley and Hirt 1998). While some data indicate an ancestral (basal) position of this group among eukaryotes (Stiller, Duffield, and Hall 1998), others suggest a derived, nonbasal position (Hinkle at al. 1994). The *Mastigamoeba* sequence branches with *Trypanosoma* and *Entamoeba* enolase also in distance and parsimony analyses. The limited taxonomic sample provides no further insight into the evolution of enolase in these protists.

Gene Duplications for the Origin of Chloroplast-Cytosol Enolase Isoenzymes

As discussed above, available molecular and biochemical data suggest that the putative transit peptide-

bearing enolases from *Arabidopsis* and *Chlamydomonas* likely represent plastid isoforms of the enzyme. The position of these isoforms in figure 4 suggests that the genes for the chloroplast-specific isoforms were not acquired by the nucleus from the cyanobacterial ancestor of plastids, but, rather, arose through gene duplication involving the cytosolic isoform and acquisition of a transit peptide, as is the case for many compartment-specific enzymes that were common to hosts and symbionts during the origin of organelles (Martin and Schnarrenberger 1997). This duplication appears to have taken place relatively early in plant evolution, probably prior to the origin of the insertion at positions 161–165 that is specific to the plant cytosolic isoforms (under the reasonable premise that the *Plasmodium* sequence is misplaced in the phylogenetic analyses).

Enolase from *Euglena* Is an Exception

Many genes studied to date underscore a common ancestry of trypanosomes and euglenids. These include nuclear-encoded ribosomal RNA (Sogin 1994), tubulin (Levasseur, Meng, and Bouck 1994), Hsp60 (Yasuhira and Simpson 1997), and the ER-specific protein calreticulin (Navazio et al. 1998), in addition to mitochondrial-encoded cytochrome oxidase subunit I (Tessier et al. 1997; Yasuhira and Simpson 1997). Enolase is thus an exception in that it clearly does not reflect the common ancestry of the kinetoplastid (*Trypanosoma*) and euglenid lineages (fig. 4).

In contrast to trypanosomes, euglenids possess plastids. These are surrounded by three membranes, suggesting that *Euglena* may have acquired its plastids by secondary symbiosis through engulfment of a eukaryotic green alga (Gibbs 1978), a view that is supported by chloroplast genome phylogeny (Martin et al. 1998). As in the case of higher plants and *Chlamydomonas,* the cytosolic and plastid-targeting peptide-bearing forms of *Euglena* enolase also must have arisen through gene duplication. Yet, the source of the eukaryotic nuclear gene which gave rise to that duplication is different from that of all other eukaryotic enolase genes. Enolase genes in *Euglena* are much more similar to the homolog from the spirochaete *Treponema* than they are to enolase from any other source currently sampled. This is a mildly surprising, but by no means unprecedented, finding. The nuclear gene for cytosolic GAPDH in *Euglena* also branches quite robustly with its homolog from the *Treponema* genome (Figge et al. 1999). Further affinities between spirochaete and euglenozoan (kinetoplastid and euglenid) glycolytic enzymes are found for PFK, for which the *Trypanosoma* homolog branches with *Borrelia* and, curiously, as with enolase, *Entamoeba* homologs (Michels et al. 1997; Mertens et al. 1998).

At face value, the surprising affinity of nuclear-encoded *Euglena* enolases to the *Treponema* homolog could reflect an outright lateral (i.e., nonendosymbiotic) gene transfer from spirochaetes to the *Euglena* lineage. If so, then future sampling of eukaryotic groups should reveal that the acquisition is specific to the euglenid lineage. If, on the other hand, other, more distantly related, eukaryotic lineages are also found to possess enolase genes that branch with the *Treponema* homolog, then the acquisition could be traced more deeply into eukaryotic phylogeny, in which case differential loss could be invoked to explain the distribution of enolase genes among eukaryotes. In analogy, a similarly ''odd'' aldolase gene was recently found in *Giardia* that is closely related to α-proteobacterial homologs, whereas other eukaryotic aldolase genes described to date are not (Henze et al. 1998). As in the case of *Euglena* enolase, it is possible that other eukaryotes will be found to possess the same ''odd'' FBA gene as *Giardia* does.

Thus, further sampling of protist lineages is needed before a strong case for or against horizontal enolase gene transfer from prokaryotes to *Euglena* in a context that did not involve the origin of organelles can be argued. Several previous claims of outright prokaryote-eukaryote horizontal gene transfer (i.e., not involving the origins of chloroplasts or mitochondria) argued on the basis of small species samples have turned out, with time, to be just as easily or more easily explained by gene transfer in the context of the origins of organelles (Martin and Schnarrenberger 1997; Nowitzki et al. 1998), warranting caution when only few protist sequences are available for comparison. The finding that prokaryotes tend to exchange genes via horizontal transfer (Lawrence and Ochman 1998; Vorholt et al. 1999) lends an additional degree of complexity to the problem of trying to identify the source of eukaryotic nuclear genes (Brown and Doolittle 1997; Martin and Schnarrenberger 1997; Doolittle 1999). The non-rRNA-like phylogeny of some eubacterial enolases in figure 4—for example, a lack of common branching for the two *Chlorobium* enolases, the sequences from two spirochaetes, or the four sequences from low-GC Gram positives—would not exclude the possibility that some transfer of enolase genes among eubacteria has occurred.

Eukaryotic Enolase Is an Exception Among Glycolytic Enzymes

Previous phylogenetic studies of glycolytic/gluconeogenetic enzymes have revealed that most of the glycolytic pathway in the eukaryotic cytosol consists of enzymes that are more similar to eubacterial homologs than they are to archaebacterial homologs. This is true for GAPDH (Markos, Miretsky, and Müller 1993; Henze et al. 1995; Wiemer et al. 1995), PGK (Brinkmann and Martin 1996; Adjé, Opperdoes, and Michels 1998), FBA (Plaumann et al. 1997; Henze et al. 1998), FBP (Martin and Schnarrenberger 1997), TPI (Keeling and Doolittle 1997), PFK (Siebers, Klenk, and Hensel 1998; Mertens et al. 1998), and GPI (Nowitzki et al. 1998).

In marked contrast to those enzymes, eukaryotic enolase sequences are not more similar to eubacterial than to archaebacterial homologs, with the exception of *Euglena.* Rather, they tend to branch near the homologs from *Pyrococcus horihoshii, Pyrococcus furiosus,* and *Methanococcus jannaschii.* However, the position of the branch bearing these three archaebacterial sequences basal to eukaryotic enolase in figure 4 is not strongly

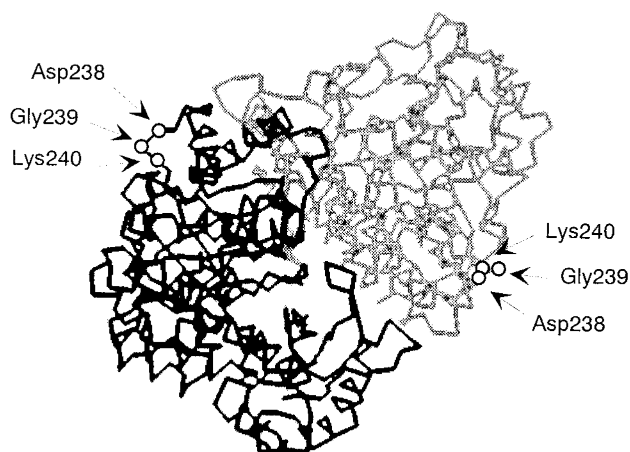

FIG. 5.—The region corresponding to positions 304–318 from the alignment in figure 3 mapped onto the crystal structure of the enolase dimer from *Saccharomyces cerevisiae.* The positions of the residues in the yeast sequence spanning this region (D . . . GK) are indicated as open circles with the numbering from the crystal structure (Wedekind, Reed, and Rayment 1994). The structure was generated with RASMOL from the Protein Data Base file and rotated such that the localization of the indel position at the surface of the protein is visible. The backbones of the individual enolase monomers are indicated in black and gray shading. A projection of the indel at positions 161–165 in figure 3 onto the three-dimensional structure is given in Dzierszinski et al. (1999).

supported, and the deeper branches of the enolase protein phylogeny are quite difficult to resolve. For example, among the 585 trees found using the quicksearch option of MOLPHY that were not significantly worse than the best tree at the 5% level, 52 placed the six archaebacterial sequences on a single common branch and 82 placed the branch bearing the *Euglena* and *Treponema* sequences basal to the eukaryotic homologs.

In this respect, it is notable that *Methanococcus* enolase, like the eukaryotic sequences, lacks the insertion at positions 305–306 in figure 3, which at face value might tend to favor an archaebacterial ancestry of the eukaryotic enzyme. However, this polymorphism is not phylogenetically unique among prokaryotes, since it is also found in the *Campylobacter* sequence (fig. 3), suggesting that it may have occurred in parallel in independent lineages. To examine this possibility in further detail, we mapped the corresponding position to the three-dimensional structure of enolase determined from yeast (Wedekind, Reed, and Rayment 1994). The residues of the yeast enzyme that correspond to the position of the indel in question are Asp^{238}, Gly^{239}, and Lys^{240} in the numbering of Wedekind, Reed, and Rayment (1994), which are numbered 304, 317, and 318 in figure 3. As shown in figure 5, these positions map to a small loop at the surface of the yeast enolase dimer, such that insertions and deletions at this position might be easily tolerated by the proteins without altering enzyme function. The additional 10 residues that are found at this position in the *Chlorobium* 1 sequence (fig. 3) may thus form a small external loop. That the plant- and apicomplexan-specific insertion at positions 161–165 (fig. 3) also maps to the surface of the enzyme, where it likely forms a small external loop, was recently shown by comparative modeling with the yeast enolase structure (Dzierszinski et al. 1999).

Thus, with the available sequence data and the inference models used, it is presently difficult to determine with confidence on the basis of protein phylogeny whether eukaryotic enolase sequences are of eubacterial or archaebacterial ancestry. Yet, independent data from gene cluster analysis may provide additional clues: enolase is the only glycolytic enzyme that occurs in the ancient gene cluster encompassing roughly 50 genes primarily for ribosomal proteins (but also for proteins of transcription) that was recently identified by Wächtershäuser (1998). Gene order within that cluster is markedly well conserved across 18 sequenced prokaryotic genomes. The enolase gene occurs within the cluster, but not among eubacterial genomes, rather only among archaebacterial genomes, where it is situated between the genes for RNA polymerase subunit K and ribosomal protein S2 (Wächtershäuser 1998). The cytosolic ribosomes (Yang et al. 1999) and transcriptional apparatus (Langer et al. 1995) of eukaryotes are much more similar to archaebacterial than to eubacterial counterparts. Such a pattern of similarity is not well pronounced for enolase, but its conserved linkage to such proteins in archaebacterial genomes provide a connection between eukaryotic and archaebacterial enolases that is independent of phylogenetic inference. Enolase in *Euglena* is unusual in that it has apparently been replaced by a eubacterial intruder, but cytosolic ribosomes in *Euglena* are also unusual in that their large subunit is composed of 14 discrete small RNA fragments (Smallman, Schnare, and Gray 1996).

Conclusions

Enolase is an exception among eukaryotic glycolytic enzymes in that it does not show markedly more similarity to eubacterial homologs than it does to archaebacterial homologs (except the enolases from *Euglena*). This is in contrast to other enzymes of energy metabolism in eukaryotes, which appear to be acquisitions from eubacteria, probably from the antecedent of mitochondria, even in eukaryotes that now lack mitochondria (Martin and Müller 1998). However, enolase is not the only exception. For example, transketolase in some eukaryotes is very similar to methanogen homologs, whereas most eukaryotic transketolases are more similar to eubacterial homologs (Martin and Schnarrenberger 1997). Furthermore, an important enzyme of the ATP-generating pathway downstream of pyruvate in *Giardia,* acetyl-CoA synthase, is also more similar to archaebacterial than to eubacterial homologs (Sánchez et al. 1999). Current hypotheses for the origin of eukaryotes postulate that the host that acquired the mitochondrion was a descendant or a member of the archaebacteria (Doolittle 1998). With only one exception (Martin and Müller 1998), all models for the origin of mitochondria assume that the host was a heterotrophic fermenting organism that obtained its ATP through glycolysis, a view that is very difficult to reconcile with the finding that eukaryotes, including amitochondriate

forms, generally possess eubacterial glycolytic enzymes. Current data indicate that the origin of mitochondria occurred very early in eukaryotic evolution, if not at its root (Embley and Hirt 1998; Gray, Burger, and Lang 1999), and that the origin of the glycolytic pathway that is the basis of the heterotrophic lifestyle in eukaryotes may trace to the origin of mitochondria (Martin and Müller 1998). In agreement with that view, some protists possess nuclear-encoded glycolytic enzymes that are imported into mitochondria (Liaud et al. 2000). The evolution of the glycolytic pathway entails aspects of gene phylogeny, protein compartmentation, and enzyme structure and function (Fothergill-Gilmore and Michels 1993). Further study of glycolytic enzymes, particularly from protists, should lead to a more complete picture of the complex evolutionary history of this pathway in eukaryotes.

Supplementary Material

Enolase sequences reported in this paper have been deposited in GenBank under the accession numbers AF152348 (*T. brucei*), AJ272112 (*Euglena* chloroplast), AJ272111 (*Euglena* cytosol), AJ271719 (spinach cytosol), and AF205070 (*Mastigamoeba*).

Acknowledgments

A culture of *M. balamuthi* was kindly provided by Dr. Frederick Schuster (Brooklyn College of the City University of New York, Brooklyn, N.Y.). M.M., P.M., and W.M. thank the National Science Foundation (DEB-9615659), the Interuniversity Poles of Attraction Programme–Belgium State, Prime Minister's Office–Federal Office for Scientific, Technical and Cultural Affairs, and the Deutsche Forschungsgemeinschaft, respectively, for financial support. We thank Christiane Lindemann for excellent technical assistance. The contributions of the authors in this paper were as follows: V.H., *T. brucei* cloning and sequencing, writing; H.B., phylogenetic analysis; M.-A.A., *T. brucei* cloning and sequencing; U.N., *Euglena* and spinach cloning and sequencing; J.A.L., *Mastigamoeba* cloning; C.W.S., *Mastigamoeba* sequencing; T.G., *Mastigamoeba* sequence analysis; M.M., P.M., and W.M., interpretation and writing.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Comput. Sci. Monogr. **28**:1--150.

ADJÉ, C., F. R. OPPERDOES, and P. A. M. MICHELS. 1998. Molecular analysis of phosphoglycerate kinase in *Trypanoplasma borreli* and the evolution of this enzyme in Kinetoplastida. Gene **217**:91–99.

BAKKER, B. M., P. A. M. MICHELS, F. R. OPPERDOES, and H. V. WESTERHOFF. 1997. Glycolysis in bloodstream form *Trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. J. Biol. Chem. **272**:3207–3215.

BRINKMANN, H., and W. MARTIN. 1996. Higher plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement. Plant Mol. Biol. **30**:65–75.

BROWN, C. K., P. L. KUHLMAN, S. MATTINGLY, K. SLATES, P. J. CALIE, and W. W. FARRAR. 1998. A model of the quaternary structure of enolases, based on structural and evolutionary analysis of the octameric enolase from *Bacillus subtilis.* Protein Chem. **17**:855–866.

BROWN, J. R., and W. F. DOOLITTLE. 1997. Archaea and the prokaryote-to-eukaryote transition. Microbiol. Mol. Biol. Rev. **61**:456–502.

BRUGEROLLE, G. 1993. Evolution and diversity of amitochondrial zooflagellates. J. Eukaryot. Microbiol. **40**:616–618.

CAVALIER-SMITH, T. 1991. Archamoebae: the ancestral eukaryotes? BioSystems **25**:25–38.

———. 1993. Kingdom protozoa and its 18 phyla. Microbiol. Rev. **57**:953–994.

CHAVÉZ, L. A., W. BALAMUTH, and T. GONG. 1986. A light and electron microscopical study of a new, polymorphic free-living amoeba, *Phreatamoeba balamuthi* n.g., n.sp. J. Protozool. **33**:397–404.

CLAYTON, C. E., and P. MICHELS. 1996. Metabolic compartmentation in African trypanosomes. Parasitol. Today **12**: 465–471.

CUI, J.-Y., K. MUKAI, K. SAEKI, and H. MATSUBARA. 1994. Molecular cloning and nucleotide sequences of cDNAs encoding subunits I, II, and IX of *Euglena gracilis* mitochondrial complex III. J. Biochem. (Tokyo) **115**:98–107.

DOOLITTLE, W. F. 1998. A paradigm gets shifty. Nature **392**: 15–16.

———. 1999. Phylogenetic classification and the universal tree. Science **284**:2124–2128.

DUQUERROY, S., C. CAMUS, and J. JANIN. 1995. X-ray structure and catalytic mechanism of lobster enolase. Biochemistry **34**:12513–12523.

DZIERSZINSKI, F., O. POPESCU, C. TOURSEL, C. SLOMIANNY, B. YAHIAOUI, and S. TOMAVO. 1999. The protozoan parasite *Toxoplasma gondii* expresses two functional plant-like glycolytic enzymes. J. Biol. Chem. **274**:24888–24895.

EMBLEY, T. M., and R. P. HIRT. 1998. Early branching eukaryotes? Curr. Opin. Genet. Dev. **8**:624–629.

FIGGE, R. M., M. SCHUBERT, H. BRINKMANN, and R. CERFF. 1999. Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence of intra- and inter-kingdom gene transfer. Mol. Biol. Evol. **16**:429–440.

FOTHERGILL-GILMORE, L. A., and P. A. M. MICHELS. 1993. Evolution of glycolysis. Prog. Biophys. Mol. Biol. **59**:105–235.

GIBBS, S. P. 1978. The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can. J. Bot. **56**:2883–2889.

GRAY, M. W., G. BURGER, and B. F. LANG. 1999. Mitochondrial evolution. Science **283**:1476–1481.

HANNAERT, V., and P. MICHELS. 1994. Structure, function and biogenesis of glycosomes in Kinetoplastida. J. Bioenerg. Biomembr. **20**:205–212.

HENZE, K., A. BADR, M. WETTERN, R. CERFF, and W. MARTIN. 1995. A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA **92**:9122–9126.

HENZE, K., H. G. MORRISON, M. L. SOGIN, and M. MÜLLER. 1998. Sequence and phylogenetic position of a class II aldolase gene in the amitochondriate protist, *Giardia lamblia.* Gene **222**:163–168.

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.

HOPPE, P., A. HEINTZE, A. RIEDEL, C. CREUZER, and G. SCHULTZ. 1993. The plastidic 3-phosphoglycerate → acetyl-CoA pathway in barley leaves and its involvement in the

synthesis of amino acids, plastidic isoprenoids and fatty acids during chloroplast development. Planta **190**:253–262.

IIDA, H., and I. YAHARA. 1985. Yeast heat shock protein of M_r 48,000 is an isoprotein of enolase. Nature **315**:688–690.

KEELING, P. J., and W. F. DOOLITTLE. 1997. Evidence that eukaryotic triosphosphate isomerase is of alpha-proteobacterial origin. Proc. Natl. Acad. Sci. USA **94**:1270–1275.

KITAOKA, S., Y. NAKANO, K. MIYATAKE, and A. YOKOTA. 1989. Enzymes and their functional location. Pp. 2–135 *in* D. E. BUETOW, ed. The biology of *Euglena,* Vol. 6. Subcellular biochemistry and molecular biology. Academic Press, San Diego.

KOHL, L., T. DRMOTA, C.-D. DO-THI, M. CALLENS, J. VAN BEEUMEN, F. R. OPPERDOES, and P. A. M. MICHELS. 1996. Cloning and characterization of the NAD-linked glycerol-3-phosphate dehydrogenases of *Trypanosoma brucei* and *Leishmania mexicana mexicana* and expression of the trypanosome enzyme in *Escherichia coli.* Mol. Biochem. Parasitol. **76**:159–173.

LANGER, D., J. HAIN, P. THURIAUX, and W. ZILLIG. 1995. Transcription in Archaea: similarity to that in Eukarya. Proc. Natl. Acad. Sci. USA **92**:5768–5772.

LATZKO, E., and M. GIBBS. 1969. Enzyme activities of the carbon reduction cycle in some photosynthetic organisms. Plant Physiol. **44**:295–300.

LAWRENCE, J. G., and H. OCHMAN. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95**:9413–9417.

LEVASSEUR, P. J., Q. MENG, and B. BOUCK. 1994. Tubulin genes in the algal protist *Euglena gracilis.* J. Eukaryot. Microbiol. **41**:468–477.

LIAUD, M.-F., C. LICHTLÉ, K. APT, W. MARTIN, and R. CERFF. 2000. Glycolytic enzymes in diatom mitochondria. Mol. Biol. Evol. **17**:213–223.

MCALISTER, L., and M. J. HOLLAND. 1982. Targeted deletion of a yeast enolase structural gene. J. Biol. Chem. **257**:7181–7188.

MCFADDEN, G. I., M. E. REITH, J. MUNHOLLAND, and N. LANG-UNNASCH. 1996. Plastid in human parasites. Nature **381**:482.

MCFADDEN, G. I., and D. ROOS. 1999. Apicomplexan plastids as drug targets. Trends Microbiol. **7**:328–332.

MARKOS, A., A. MIRETSKY, and M. MÜLLER. 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote *Trichomonas vaginalis.* J. Mol. Evol. **37**:631–643.

MARTIN, W., and M. MÜLLER. 1998. The hydrogen hypothesis of the first eukaryote. Nature **392**:37–41.

MARTIN, W., and C. SCHNARRENBERGER. 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr. Genet. **32**:1–18.

MARTIN, W., B. STOEBE, V. GOREMYKIN, S. HANSMANN, M. HASEGAWA, and K. V. KOWALLIK. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393**: 162–165.

MERTENS, E. 1993. ATP versus pyrophosphate: glycolysis revisited in parasitic protists. Parasitol. Today **9**:122–126.

MERTENS, E., U. S. LADROR, J. A. LEE, A. MIRETSKY, A. MORRIS, C. ROZARIO, R. J. KEMP, and M. MÜLLER. 1998. The pyrophosphate-dependent phosphofructokinase of the protist, *Trichomonas vaginalis,* and the evolutionary relationships of protist phosphofructokinases. J. Mol. Evol. **47**: 739–750.

MERTENS, E., E. VAN SCHAFTINGEN, and M. MÜLLER. 1992. Pyruvate kinase from *Trichomonas vaginalis,* an allosteric enzyme stimulated by ribose 5-phosphate and glycerate 3-phosphate. Mol. Biochem. Parasitol. **54**:13–20.

MICHELS, P. A. M., N. CHEVALIER, F. R. OPPERDOES, M. H. RIDER, and D. J. RIGDEN. 1997. The glycosomal ATP-dependent phosphofructokinase of *Trypanosoma brucei* must have evolved from an ancestral pyrophosphate-dependent enzyme. Eur. J. Biochem. **250**:698–704.

MICHELS, P. A. M., and V. HANNAERT. 1994. The evolution of kinetoplastid glycosomes. J. Bioenerg. Biomembr. **26**:213–219.

MICHELS, P. A. M., M. MARCHAND, L. KOHL, S. ALLERT, R. K. WIERENGA, and F. R. OPPERDOES. 1991. The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distinct evolutionary relationship. Eur. J. Biochem. **200**:19–27.

MIERNYK, J. A., and D. T. DENNIS. 1992. A developmental analysis of the enolase isozymes from *Ricinus communis.* Plant Physiol. **99**:748–750.

MO, Y., B. G. HARRIS, and R. W. GRACY. 1973. Triosephosphate isomerase and aldolases from light and dark-grown *Euglena gracilis.* Arch. Biochem. Biophys. **157**:580–587.

MUKAI, K., M. YOSHIDA, H. TOYOSAKI, Y. YAO, S. WAKABAYASHI, and H. MATSUBARA. 1989. An atypical heme-binding structure of cytochrome c1 of *Euglena gracilis* mitochondrial complex III. Eur. J. Biochem. **178**:649–656.

MÜLLER, M. 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists—a special case in eukaryotic evolution? Pp. 109–131 *in* G. H. COOMBS, K. VICKERMAN, M. A. SLEIGH, and A. WARREN, eds. Evolutionary relationships among protozoa. Kluwer, Dordrecht, the Netherlands.

NAVAZIO, L., C. NARDI, B. BALDAN, P. DAINESE, A. C. FITCHETTE, W. MARTIN, and P. MARIANI. 1998. Functional conservation of calreticulin from *Euglena gracilis.* J. Eukaryot. Microbiol. **45**:307–313.

NOWITZKI, U., A. FLECHNER, J. KELLERMANN, M. HASEGAWA, C. SCHNARRENBERGER, and W. MARTIN. 1998. Eubacterial origin of eukaryotic nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. Gene **214**:205–213.

OPPERDOES, F. R., and P. BORST. 1977. Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome. FEBS Lett. **80**:360–364.

PARK, J. H., P. J. SCHOFIELD, and M. R. EDWARDS. 1997. Pyruvate kinase is present in *Giardia intestinalis.* Exp. Parasitol. **87**:153–156.

PHILIPPE, H. 1993. MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res. **21**: 5264–5272.

PLAUMANN, M., B. PELZER-REITH, W. MARTIN, and C. SCHNARRENBERGER. 1997. Cloning of fructose-1,6-bisphosphate aldolases from *Euglena gracilis*: multiple recruitment of class I aldolase to chloroplasts and eubacterial origin of eukaryotic class II aldolase genes. Curr. Genet. **31**:430–438.

READ, M., K. E. HICKS, P. F. SIMS, and J. E. HYDE. 1994. Molecular characterisation of the enolase gene from the human malaria parasite *Plasmodium falciparum.* Evidence for ancestry within a photosynthetic lineage. Eur. J. Biochem. **220**:513–520.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMBROOK, J., E. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

SÁNCHEZ, L. B., H. G. MORRISON, M. L. SOGIN, and M. MÜLLER. 1999. Cloning and sequencing of an acetyl-CoA synthetase (ADP-forming) gene from the amitochondriate protist, *Giardia lamblia.* Gene **233**:225–231.

SCHNARRENBERGER, C., S. JACOBSHAGEN, B. MÜLLER, and I. KRÜGER. 1990. Evolution of isozymes of sugar metabolism in green algae. Pp. 743–746 *in* S. OGITA and J. S. SCANDALIOS, eds. Isozymes: structure, function, and use in biology and medicine. Wiley-Liss, New York.

SCHURIG, H., K. RUTKAT, R. RACHEL, and R. JAENICKE. 1995. Octameric enolase from the hyperthermophilic bacterium *Thermotoga maritima*: purification, characterization, and image processing. Protein Sci. **4**:228–236.

SCHWARTZBACH, S. D., T. OSAFUNE, and W. LÖFFELHARDT. 1998. Protein import into cyanelles and complex chloroplasts. Plant Mol. Biol. **38**:247–263.

SIEBERS, B., H. P. KLENK, and R. HENSEL. 1998. PP_i-dependent phosphofructokinase from *Thermoproteus tenax,* an archaeal descendant of an ancient line in phosphofructokinase evolution. J. Bacteriol. **180**:2137–2143.

SMALLMAN, D. S., M. N. SCHNARE, and M. W. GRAY. 1996. RNA: RNA interactions in the large subunit ribosomal RNA of *Euglena gracilis.* Biochim. Biophys. Acta. **1305**:1–6.

SOGIN, M. 1994. The origin of eukaryotes and evolution into major kingdoms. Pp. 181–192 *in* S. BENGSTON, ed. Early life on earth. Columbia University Press, New York.

STILLER, J. W., E. C. DUFFIELD, and B. D. HALL. 1998. Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. Proc. Natl. Acad. Sci. USA **95**:11769–11774.

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TESSIER, L. H., H. VAN DER SPECK, J. M. GUALBERTO, and J. M. GRIENENBERGER. 1997. The cox1 gene from *Euglena gracilis*: a protist mitochondrial gene without introns and genetic code modifications. Curr. Genet. **31**:208–213.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

VAN DER STRAETEN, D., R. A. RODRIGUES-POUSADA, H. M. GOODMAN, and M. VAN MONTAGU. 1991. Plant enolase: gene structure, expression, and evolution. Plant Cell **3**:719–735.

VORHOLT, J. A., L. CHISTOSERDOVA, S. M. STOLYAR, R. K. THAUER, and M. E. LIDSTROM. 1999. Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. J. Bacteriol. **181**:5750–5757.

WÄCHTERSHÄUSER, G. 1998. Towards a reconstruction of ancestral genomes by gene cluster alignment. Syst. Appl. Microbiol. **21**:473–477.

WEDEKIND, J. E., G. H. REED, and I. RAYMENT. 1994. Chelation of serine 39 to Mg^{2+} latches a gate at the active site of enolase: structure of the bis(Mg^{2+}) complex of yeast enolase and the intermediate, phosphonoacetohydroxamate, at 2.1 Å resolution. Biochemistry **33**:9333–9342.

WIEMER, E. A. C., V. HANNAERT, P. R. L. A. VAN DEN IJSSEL, J. VAN ROY, F. R. OPPERDOES, and P. A. M. MICHELS. 1995. Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanoplasma borelli.* J. Mol. Evol. **40**:443–454.

WISTOW, G. J., T. LIETMAN, L. A. WILLIAMS, S. O. STAPEL, W. W. DE JONG, J. HORWITZ, and J. PIATIGORSKY. 1988. τ-Crystalin/α-enolase: one gene encodes both an enzyme and a lens structural protein. J. Cell Biol. **107**:2729–2736.

YANG, D., I. KUSSER, A. K. KOPKE, B. F. KOOP, and A. T. MATHESON. 1999. The structure and evolution of the ribosomal proteins encoded in the *spc* operon of the archaeon (Crenarchaeota) *Sulfolobus acidocaldarius.* Mol. Phylogenet. Evol. **12**:177–185.

YASUHIRA, S., and L. SIMPSON. 1997. Phylogenetic affinity of mitochondria of *Euglena gracilis* and kinetoplastids using cytochrome oxidase I and hsp60. J. Mol. Evol. **44**:341–347.

ZHANG, E., J. M. BREWER, W. MINOR, L. A. CARREIRA, and L. LEBIODA. 1997. Mechanism of enolase: the crystal structure of asymmetric dimer enolase-2-phospho-D-glycerate/enolase-phosphoenolpyruv ate at 2.0 Å resolution. Biochemistry **36**:12526–12534.

GEOFFREY MCFADDEN, reviewing editor

Accepted February 14, 2000