

Fructose-1,6-bisphosphate aldolases in amitochondriate protists constitute a single protein subfamily with eubacterial relationships

Lidya B. Sánchez^{a,*}, David S. Horner^{b,1}, Dorothy V. Moore^a, Katrin Henze^{a,2},
T. Martin Embley^b, Miklós Müller^a

^aThe Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

^bDepartment of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Received 25 March 2002; received in revised form 14 June 2002; accepted 19 June 2002

Received by W. Martin

Abstract

Sequences of putative fructose-1,6-bisphosphate aldolases (FBA) in five amitochondriate unicellular eukaryotes, the diplomonads *Giardia intestinalis* (published earlier) and *Spironucleus barkhanus*, the pelobiont *Mastigamoeba balamuthi*, the entamoebid *Entamoeba histolytica*, and the parabasalid *Trichomonas vaginalis* all belong to Class II of FBAs and are highly similar to each other (>48% amino acid identity). The five protist sequences, however, do not form a monophyletic group. Diplomonad FBAs share a most recent common ancestor, while FBAs of the three other protist species are part of a lineage that also includes sequences from a few eubacteria (*Clostridium difficile*, *Treponema pallidum*, *Chlorobium tepidum*). Both clades are part of the Type B of Class II aldolases, a complex that contains at least three additional lineages (subgroups) of enzymes. Type B enzymes are distant from Type A Class II aldolases, which consists of a number of bacterial and fungal enzymes and also contains the cytosolic FBA of *Euglena gracilis*. Class II aldolases are not homologous to Class I enzymes, to which animal and plant enzymes belong. The results indicate that amitochondriate protists acquired their FBAs from separate and different sources, involving lateral gene transfer from eubacteria, than did all other eukaryotes studied so far and underscore the complex composition of the glycolytic machinery in unicellular eukaryotes. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Diplomonad; Embden–Meyerhof–Parnas pathway; Glycolysis; Pelobiont; Parabasalid; Protein evolution

1. Introduction

The major mechanism of glucose catabolism in all eukaryotic organisms is the Embden–Meyerhof–Parnas (EMP) glycolytic pathway (Fothergill–Gilmore and Michels, 1993), in contrast to the diversity of pathways seen in prokaryotes, suggesting that the EMP pathway was present in the last common ancestor of all eukaryotes. While the pathway itself is highly conserved, its individual enzymes can differ from one major eukaryotic group to another. In some cases isofunctional enzymes belong to functionally different classes in different lineages (Fothergill–Gilmore

and Michels, 1993). A study of the taxonomic distribution of such enzymes has great, but as yet unfulfilled, promise in attempts to decipher the history of eukaryotic core metabolism. In our comparative studies of organisms with some of the most divergent types of eukaryotic core metabolism, the amitochondriate protists (Müller, 1998), we have put special emphasis on such enzymes with the aim of unraveling the history of their peculiar metabolic machinery.

The fourth step of glycolysis, the aldol cleavage of fructose-1,6-bisphosphate to dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, catalysed by fructose-1,6-bisphosphate aldolase (FBA), is particularly promising for such a study (Rutter, 1964). The two main classes of FBAs and closely related enzymes differ in the mechanism of their action. Class I enzymes form a Schiff base with the substrate and Class II enzymes act with the participation of a coordinated divalent metal ion (Rutter, 1964; Zgiby et al., 2000). These enzymes show practically no sequence similarities and are assumed to represent two independent protein families (Marsh and Leberer, 1992; Rutter, 1964). Within Class II aldolases, two Types (A and B) are clearly deli-

Abbreviations: EST, expressed sequence tag; FBA, fructose-1,6-bisphosphate aldolase; ML, maximum likelihood; ORF, open reading frame; PCR, polymerase chain reaction

* Corresponding author. Tel.: +1-212-327-8144; fax: +1-212-327-7974.

E-mail address: sanchel@mail.rockefeller.edu (L.B. Sánchez).

¹ Present address: Dipartimento di Fisiologia e Biochimica Generali, Università di Milano, via Celoria 26, 20133 Milan, Italy.

² Present address: Institute of Botany III, Heinrich-Heine-University of Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany.

neated by their sequence properties (Henze et al., 1998; Nickol et al., 2000; Plaumann et al., 1997). Type B enzymes can be further divided into easily distinguished subgroups (Henze et al., 1998).

FBA and related enzymes also participate in anabolic reactions, gluconeogenesis and photosynthesis. In addition, Class II aldolases include bacterial enzymes acting on tagatose-1,6-bisphosphate, which are such close homologs of FBAs that the substrate specificity of these enzymes cannot be established from sequence data alone (Zgiby et al., 2000). Accordingly, we include in our analysis established and putative bacterial tagatose-1,6-bisphosphate aldolases.

The different classes and types of FBAs and related enzymes show a complex and peculiar distribution among major evolutionary lineages. Members of both classes are present in diverse eubacteria. Among eukaryotes Class I enzymes are characteristic of animals, plants, and some protists, while Class II enzymes are found in fungi (Marsh and Lebherz, 1992; Rutter, 1964). Diverse algae contain either Class I or Class II FBAs (Anita, 1967; Ikawa et al., 1972; Nickol et al., 2000). A eukaryotic organism that contains both Class I and Class II FBA is the euglenozoon, *Euglena gracilis* (Henze et al., 1995; Plaumann et al., 1997; Rutter, 1964). These findings show that in different organisms the same step in glycolysis (and gluconeogenesis) can be performed by enzymes of quite different origins. Interestingly, despite a lack of similarity and mechanistic differences, expression of the Class I enzyme from *Drosophila melanogaster* restored the wild-type phenotype in *Saccharomyces cerevisiae* mutants in which the endogenous Class II enzyme has been deleted, showing an extensive functional equivalency of the two classes (Boles and Zimmermann, 1993). This finding suggests the absence of a strong barrier to possible evolutionary replacement of FBAs by lateral gene transfer.

We previously sequenced the gene coding for FBA in the diplomonad *Giardia intestinalis* (Henze et al., 1998) and found that it belonged to Class II, Type B, quite distinct from Type A enzymes found in fungi and *E. gracilis*. This finding indicated that *G. intestinalis* acquired the corresponding gene from a different source than other eukaryotes. We report here that FBA sequences from four additional amitochondriate eukaryotes from different taxonomic groups are also of Class II, Type B. The results show that the evolutionary histories of FBAs of these organisms are separate from those of other eukaryotic FBAs studied so far.

2. Materials and methods

2.1. Organisms

The organisms studied were the free-living pelobiont *Mastigamoeba balamuthi* (ATCC 30984) (synonym *Phreatamoeba balamuthi*), the entamoebid *Entamoeba histolytica* (strain HM-1:IMSS, ATCC 30459), the parabasalid *Tricho-*

monas vaginalis (strain NIH-C1, ATCC 30001), and the diplomonad *Spironucleus barkhanus*, a parasite of fish (strain NOR-1A, ATCC 50380).

2.2. Isolation of clones and sequencing

Entamoeba histolytica FBA was sequenced starting from partial FBA sequences identified in several entries (ENTGQ91TR, ENTGS72TR and ENTCL85TR) of the ongoing *E. histolytica* genome project at the Institute for Genomic Research (<http://www.TIGR.org>), with the nucleotide sequence of *G. intestinalis* aldolase gene as query. A pair of oligonucleotide primers was synthesized (forward primer: 5'-TTC ACT TAC CCT TGA CCA TGG AGC-3'; reverse primer: 5'-AAG TGC TTG TCT AGC TGG TCC AAG-3', corresponding to amino acid residues 96–103 and 302–309 in the final sequence).

With *E. histolytica* genomic DNA as template, a fragment was amplified in standard PCR. The product was gel purified, random primer labeled and used to screen *E. histolytica* genomic and cDNA libraries in Lambda gt11. The several clones sequenced lacked the first four amino terminal residues. However, their amino-terminal sequences overlapped with contig zheng_2850 of the TIGR database. The missing amino acids were inferred from this contig.

A partial cDNA sequence of *T. vaginalis* FBA was recognized in our expressed sequence tag (EST) project on this organism (Hirt, R.P., Embley, T.M., Murphy, C., Ragan, M., Horner, D.S., Dyal, P.). In order to establish the full open reading frame, PCR primers (forward primer 5'-GTC ACT TGG TCT CGT CAA CTC-3' and reverse primer 5'-ACG TAC TCT TGT GGG ATG GAG-3', corresponding to amino acid residues 7–13 and 232–238 in the final sequence) were synthesized. With gDNA of *T. vaginalis* as template, a single fragment was amplified by standard PCR, inserted into pBluescript KS(+) and its identity verified by sequencing. The PCR product was randomly labeled and used to screen a *T. vaginalis* gDNA library in Lambda-ZapII. Of the 10 positive clones identified one was completely sequenced by primer walking.

The *S. barkhanus* and *M. balamuthi* FBA sequences are from random cDNA clones isolated in our ongoing EST projects (Horner, D.S., Dyal, P., Hirt, R.P., Embley, T.M., Murphy, C., Ragan, M and Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Gaasterland, T., Müller, M., respectively) and were sequenced on both strands by primer walking. The sequence of FBA from a second diplomonad, *Giardia intestinalis*, was established earlier (Henze et al., 1998).

2.3. Comparison of sequences and phylogenetic analysis

The sequences were aligned with putative homologs retrieved from the non-redundant GenBank database with a BLAST search using the *M. balamuthi* sequence as query. Matches were retained with an expected probability value below e^{-20} and duplicates were eliminated. Preliminary sequences of *Chlorobium tepidum* and *Clostridium difficile*

FBA were obtained from The Institute for Genomic Research website at <http://www.tigr.org> and The Sanger Center website at <http://www.sanger.ac.uk>, respectively. The sequences were formatted with programs (BLASTALI and SEQ) developed by Mr. J. Lewin. Multiple alignment was done with CLUSTAL_X.

Phylogenetic analysis was performed on a 52-taxon dataset of 214 unambiguously aligned amino acid residues. This dataset was too large for the calculation of exact bootstrap values, which was performed on a subset of the same alignment containing 26 sequences that were the most closely related to the five protist sequences. The identical topology of these taxa on trees derived from either the 52-taxon or the 26-taxon subsets validated this approach.

Bayesian searches of treespace were performed with the program MrBayes (Huelsenbeck and Rannala, 1997) using the JTT-f amino acid substitution matrix with one invariable and four variable Γ rate categories. Two hundred thousand Monte Carlo Markov Chain generations were constructed with trees sampled every 100 generations. For compilation of the Bayesian consensus topologies a ‘burn-in’ of 201 trees was used. For the maximum likelihood (ML) topologies shown, branch lengths were calculated under the same model in TREE-PUZZLE version 5.0 (Strimmer and von Haeseler, 1996). To test the robustness of alternative tree topologies within ML, constraint trees were generated using the ‘enforcecons’ option in the program MrBayes (Huelsenbeck and Rannala, 1997). Relationships outside of each constraint were allowed to optimize under likelihood in each analysis. Shimodaira–Hasegawa tests (Shimodaira and Hasegawa, 1999) were implemented using custom software within the Python phylogenetics package P4 (written by P. Foster, Natural History Museum, London). ML bootstrap analyses were performed using the custom software MrBoot (D.S.H. and P. Foster), which automates MrBayes analyses of resampled datasets generated by PAUP* (Swofford, 1998). For bootstrap replicate analyses, 20,001 search generations were sampled every 50 generations and a ‘burn-in’ of 100 trees was used.

2.4. Database accession numbers

The nucleotide sequences established have been deposited with GenBank under accession numbers AY05996 (*Entamoeba histolytica*), BE636694 (*Mastigamoeba balamuthi*), AF461037 (*Spironucleus barkhanus*), and AF390544 (*Trichomonas vaginalis*).

3. Results and discussion

3.1. Characterization of sequences

The open reading frames (ORF) established for the four amitochondriate protist species studied were similar in length and essentially colinear to each other and also to that of *G. intestinalis*, studied earlier (Henze et al., 1998)

(Fig. 1). The G + C content of the ORFs (*G. intestinalis*: 54.7%; *S. barkhanus*: 43.4%; *E. histolytica*: 35.2%; *M. balamuthi*: 64.6%; *T. vaginalis*: 53.3%) was typical of each species. Preliminary BLAST searches of the NCBI database with these sequences gave high similarity scores to Class II aldolases, and within this class to enzymes designated earlier as Type B (Henze et al., 1998; Nickol et al., 2000; Plaumann et al., 1997) (Fig. 1). In the *S. barkhanus* cDNA sequence six stop codons (TAA or TAG) were seen. Since this sequence was derived from a cDNA clone, we assumed that these do not interrupt the ORF but specify an amino acid. The use of stop codons to encode glutamine has been reported before for *S. barkhanus* and other diplomonads, but not for *G. intestinalis* (Keeling and Doolittle, 1996; Rozario et al., 1996). These positions were not included in the phylogenetic analysis.

The derived amino acid sequences of the five protist FBAs were highly similar (47–72% identity in 270 shared positions) (Fig. 1). Identity values and short indels divided the protist sequences into two groups. The first contained the two diplomonad (*G. intestinalis* and *S. barkhanus*) sequences, which were 70% identical to each other (330 shared positions). The second comprised FBA sequences of the three other species showing 65–67% identities to each other (325 shared positions). The identities between members of the two groups ranged from 46 to 53% (307 shared positions). The diplomonads lacked indel B and had 14–16-amino-acid long carboxyl-terminal extensions (Fig. 1).

The active site and the residues interacting with the substrates and the catalytically important metal have been established for the FBA of *Escherichia coli* (Blom et al., 1996; Cooper et al., 1996) and tagatose-1,6-bisphosphate aldolase (Zgiby et al., 2000). All these residues are conserved in the five protist enzymes studied (Fig. 1).

We interpret the sequences to correspond to FBAs that participate in glycolysis of the amitochondriate protists studied. This tentative assignment is based on several considerations. First, these organisms are all amitochondriate and glycolysis and its extensions play an important role in them (Müller, 1998). Second, Class II FBA activities have been detected in extracts of *T. vaginalis* (Baernstein, 1955) and *E. histolytica* (Kalra et al., 1969; Susskind et al., 1982) and in *E. coli* in which the *fba* genes of these two species were expressed (unpublished observations). Third, no additional FBA sequences have been recognized by BLAST searches of the databases of the *G. intestinalis* and *E. histolytica* genome projects on the NCBI server.

3.2. Sequence comparisons and phylogenetic analysis

The final alignment comprised 52 verified and putative FBAs and related hexose bisphosphate aldolase sequences. Based on sequence similarity and shared indels, the sequences were easily assigned to several groups recognized earlier (Henze et al., 1998; Nickol et al., 2000; Plaumann et al., 1997). A subset of this alignment shown on Fig. 1 consists

TYPE B

Diplomonads	G. i.	MP	CTLR	RM	LG	ARK	HKY	GV	GAF	N	V	N	N	LO	IG	HK	AV	V	Q	L	SP	V	L	C	GR	AL	K	YS	D	-----	A	-----	M	I	Y	K	K	E	C	E	A	L	E	71																																													
	S. b.	G	K	S	F	A	L	L	N	S	---	L	L	A	E	A	K	K	G	K	Y	G	V	G	A	N	V	N	N	M	E	D	I	O	A	I	L	A	A	Q	O	T	K	S	P	V	I	L	C	S	R	G	A	L	K	Y	A	D	72																														
	M. b.	M	A	A	H	K	V	S	Y	K	E	L	G	L	N	T	R	E	M	F	A	R	A	V	D	G	G	F	A	P	A	F	N	N	N	M	E	D	I	O	A	I	L	S	A	C	V	E	C	R	S	P	V	I	L	G	V	S	G	A	R	N	Y	A	N	81																							
	Other protists	E. h.	T	V	N	Y	K	E	L	G	L	C	N	H	K	E	M	F	E	H	A	I	K	G	G	F	A	P	G	E	F	N	N	L	E	D	I	O	A	I	O	A	C	T	E	A	K	S	P	V	I	L	G	V	S	G	A	R	E	Y	A	N	77																										
	T. v.	M	A	V	S	K	S	L	G	L	V	N	S	K	D	I	F	A	K	A	V	N	G	Y	A	I	P	G	Y	N	F	S	N	L	E	D	I	O	A	I	T	A	S	V	K	T	E	S	P	V	I	L	G	V	S	A	G	A	R	K	Y	A	N	78																									
	T. p.	M	T	S	Y	K	A	L	G	L	V	N	T	K	D	L	F	A	K	A	V	K	G	Y	A	I	P	A	N	F	N	N	M	E	D	I	O	A	I	O	A	C	V	E	T	R	S	P	V	I	L	G	V	S	G	A	R	K	Y	A	N	77																											
	C. t.	Y	K	E	L	G	L	V	N	S	R	E	L	F	A	K	A	I	S	G	Y	A	I	P	A	N	F	N	N	M	E	D	I	O	A	I	M	A	C	V	E	T	A	S	P	V	I	L	G	V	S	G	A	R	S	Y	A	N	74																														
	C. d.	M	A	L	I	T	T	K	E	M	F	K	A	Y	E	G	G	F	A	T	G	A	F	N	I	S	D	L	E	O	L	G	V	L	K	A	A	K	S	K	N	S	Y	M	I	G	A	S	M	A	V	K	Y	A	G	-----	P	H	T	V	E	M	V	K	A	A	S	D	71																				
	S. sp.	M	A	L	V	P	M	R	L	L	D	H	A	E	A	N	G	Y	G	I	P	A	F	N	V	N	N	M	E	D	I	O	A	I	T	A	S	V	K	T	E	S	P	V	I	L	G	A	S	A	G	A	R	S	Y	A	G	-----	E	N	F	R	H	E	V	L	G	A	V	E	71																		
	ii	R. e.	M	A	L	I	S	L	R	O	L	L	D	H	A	E	F	G	Y	G	V	P	A	F	N	V	N	N	M	E	D	I	O	A	I	M	A	E	A	E	T	O	S	P	V	I	L	G	A	S	A	G	A	R	K	Y	A	G	-----	E	A	Y	R	H	V	L	A	A	E	71																			
T. a.	M	L	V	T	G	L	E	L	L	K	A	R	E	E	Y	G	V	G	A	F	N	V	N	N	M	E	D	I	O	A	I	L	O	A	V	L	E	A	A	E	E	O	R	S	P	V	I	L	A	L	S	E	A	M	K	Y	G	-----	R	A	L	T	L	M	A	V	E	L	A	K	E	70																	
i	T. m.	M	T	M	P	Y	K	N	T	K	E	L	L	E	K	A	S	K	E	R	Y	A	I	G	A	F	N	N	M	E	D	I	O	A	I	L	E	A	E	E	E	K	A	P	V	I	V	A	T	S	E	G	A	I	K	Y	I	G	K	G	D	I	E	T	G	-----	A	K	L	A	V	E	M	V	R	T	Y	A	E	81									
E. c. T	M	S	I	S	T	K	Y	L	L	O	A	A	N	G	Y	A	V	P	A	F	N	I	H	A	T	I	O	A	L	E	V	C	S	E	M	R	S	P	V	I	L	A	G	T	P	G	T	F	K	H	I	A	-----	L	E	E	L	Y	L	C	S	A	Y	S	T	70																							
B. s.	M	P	L	Y	S	M	T	E	M	L	N	T	A	K	E	K	Y	A	V	G	O	N	N	N	L	F	T	O	A	L	O	A	E	E	E	K	S	P	V	I	L	G	V	S	A	G	R	Y	M	G	-----	F	K	T	V	A	V	K	A	M	72																												
E. c. F	M	S	K	I	F	D	F	V	K	P	G	V	I	T	G	D	D	V	K	V	F	Q	V	A	K	E	N	N	F	A	P	V	N	C	V	G	T	S	I	N	A	Y	L	E	T	A	A	K	V	K	A	P	V	I	V	F	P	N	G	A	S	F	I	A	G	K	V	K	S	D	V	P	O	G	A	I	L	G	A	I	S	G	A	H	H	O	M	A	99
S. c.	M	G	V	E	Q	I	L	R	K	T	G	V	I	G	V	E	D	V	H	N	L	F	Y	A	K	H	K	F	A	I	P	A	I	N	V	T	S	S	T	A	V	A	L	E	A	R	D	S	K	P	T	L	L	E	T	N	G	A	T	F	A	G	K	I	S	N	E	G	O	N	-	A	S	I	K	G	A	I	A	A	H	Y	I	R	S	I	A	P	99

Diplomonads	G. i.	K	H	P	D	-----	B	-----	I	P	C	H	L	D	H	G	D	L	E	S	V	K	M	A	-----	C	-----	D	-----	S	S	V	M	I	D	A	S	H	P	-----	E	-----	F	D	E	N	V	R	I	-----	F	-----	V	S	V	E	A	E	L	G	T	-----	139														
	S. b.	L	N	K	D	-----	I	P	V	C	H	L	D	H	G	D	T	F	E	S	C	K	N	A	-----	S	L	G	-----	F	S	S	V	M	I	D	A	S	H	S	-----	F	E	E	N	W	K	I	-----	K	O	V	V	Y	A	H	A	H	G	-----	V	S	V	E	A	E	L	G	O	-----	140						
	M. b.	Y	A	K	E	I	S	K	D	K	G	I	P	A	L	N	L	D	H	C	D	S	Y	E	L	V	S	C	D	S	G	-----	F	S	N	V	M	I	D	G	S	H	L	-----	Y	D	D	N	V	A	E	T	K	K	V	Y	A	H	A	H	O	-----	V	I	T	V	E	G	E	L	G	V	-----	156			
	E. h.	Y	S	K	E	I	D	P	E	H	K	V	F	P	S	L	T	D	H	C	A	T	F	D	I	C	K	E	C	D	N	G	-----	F	S	N	V	M	I	D	G	S	A	L	P	-----	Y	E	E	N	W	K	I	-----	K	O	V	V	Y	A	H	A	H	O	-----	V	I	T	V	E	G	E	L	G	V	-----	152
	T. v.	F	A	K	E	I	P	E	H	K	L	P	V	L	H	L	D	H	G	D	S	F	E	L	C	K	S	T	D	L	C	-----	F	S	S	V	M	I	D	G	S	H	P	-----	Y	D	E	N	V	A	L	-----	K	K	V	Y	A	H	A	H	O	-----	S	R	P	D	Y	V	T	V	E	G	L	V	-----	155	
	T. p.	Y	A	H	E	L	G	V	D	-----	I	P	V	L	H	L	D	H	G	D	S	L	E	L	C	I	D	C	E	S	G	-----	F	S	S	V	M	I	D	G	S	A	L	P	-----	Y	D	E	N	V	A	L	-----	S	R	K	V	Y	A	H	A	H	O	-----	V	I	T	V	E	G	L	V	-----	151			
	C. t.	Y	A	A	E	L	G	R	E	-----	I	P	V	L	H	L	D	H	G	D	S	F	E	L	C	K	D	C	I	E	T	G	-----	F	S	S	V	M	I	D	G	S	H	L	S	-----	Y	E	D	N	V	A	L	-----	T	R	K	V	Y	A	H	A	H	O	-----	V	I	T	V	E	G	L	V	-----	146		
	C. d.	E	I	G	-----	V	O	V	A	L	H	L	D	H	G	P	N	M	O	A	I	K	T	C	D	A	G	-----	F	S	S	V	M	I	D	G	S	H	F	D	-----	F	E	E	N	V	R	I	-----	K	E	A	V	Y	A	H	S	K	G	-----	V	V	V	E	A	E	L	G	V	-----	138						
	S. sp.	T	Y	P	H	-----	I	P	I	A	M	H	O	H	C	N	S	P	A	T	C	Y	S	A	R	N	G	-----	F	S	S	V	M	I	D	G	S	L	E	A	D	A	K	T	P	A	S	F	E	Y	N	V	N	Y	A	E	V	K	V	A	H	S	V	G	-----	A	S	V	E	G	E	L	G	L	-----	146	
	R. e.	T	H	P	D	-----	I	P	V	L	H	L	D	H	G	S	S	P	A	V	C	O	A	S	I	R	S	C	-----	F	S	S	V	M	I	D	G	S	L	E	D	M	K	T	P	S	D	Y	D	N	N	Y	I	R	R	V	C	M	A	H	A	V	G	-----	V	S	V	E	G	E	L	G	L	-----	146		
T. a.	A	R	-----	V	P	V	A	M	H	L	D	H	G	S	S	Y	E	S	V	L	R	A	L	R	A	G	-----	F	S	S	V	M	I	D	K	S	H	E	D	-----	F	E	T	N	V	R	E	-----	T	R	R	V	E	A	A	H	A	V	G	-----	V	I	T	V	E	A	E	L	G	L	-----	136					
T. m.	K	L	S	-----	V	P	V	A	M	H	L	D	H	G	R	D	F	K	V	I	M	A	A	K	A	G	-----	Y	S	S	V	M	I	D	A	S	H	P	-----	F	E	E	N	L	R																																

of the five amitochondriate protist sequences and selected bacterial sequences that represent all major groups recognized. The alignment reveals the existence of 12 insertions/deletions, which taken together clearly differentiate these groups. Fig. 2A presents a phylogenetic reconstruction based upon all of the sequences available. This reconstruction shows two major clusters that comprise Type A and Type B enzymes, respectively. Both clusters contain eubacterial and eukaryotic sequences. Type A and Type B sequences differ also in the presence or absence of six of the 12 indels seen in the dataset (indels A, C, D, H, J, L) (Fig. 1).

The sequences from the five amitochondriate protist species are all within a small subcluster, which also contains a number of bacterial sequences, in the Type B cluster (Fig. 2A). A 21 amino acid insertion (indel K) also supports the separation of this subcluster from the rest of the Type B cluster (Fig. 1). The five amitochondriate sequences are thus very different from other eukaryotic Class II enzymes. One more eukaryotic sequence in the Type B cluster but separate from the protist sequences is an unidentified ORF from *Arabidopsis thaliana*. The data also confirm that FBAs of fungi and the Class II enzyme of *E. gracilis* are part of the Type A cluster (Henze et al., 1998; Nickol et al., 2000; Plaumann et al., 1997).

The 26 sequences in the subcluster containing the amitochondriate FBA sequences were used to calculate maximum likelihood bootstrap values for each node (Fig. 2B). This subcluster consists of four clades separated by internal nodes that are strongly supported by bootstrapping (Fig. 2B). Notably, the five protist FBAs were not recovered as a monophyletic group in this analysis.

The two diplomonads (*Giardia* and *Spiroucleus*) were recovered together with strong bootstrap support, suggesting a single common origin for their FBA. The sequences from *Mastigamoeba* and *Entamoeba* were also recovered together (separate from the diplomonads) with strong bootstrap support, consistent with a common origin for their FBA. These two sequences formed a clade with the sequence from *Trichomonas vaginalis* and the sequences from three taxonomically diverse eubacteria, *Clostridium difficile*, *Chlorobium tepidum* and *Treponema pallidum*. With the exception of *C. difficile*, this subgroup shared an additional insertion (B) that is seven amino acids long in the protists but only four long in the two bacteria (Fig. 1). This indel event possibly took place after the separation of *C. difficile* but it

cannot be excluded that it happened earlier and *C. difficile* lost the insertion. There was no support for the FBA sequence from *T. vaginalis* forming a monophyletic group with the *Mastigamoeba* and *Entamoeba* FBA. The *Trichomonas* FBA clustered with the sequence from *T. pallidum*. The *T. pallidum* and *T. vaginalis* FBA also share a short indel (F) that is not found in any of the other sequences.

A third subgroup, designated earlier as subgroup i (Henze et al., 1998) unites eubacteria, which have different properties but often group together in phylogenetic reconstructions: *Deinococcus radiodurans*, *Thermus aquaticus*, *Thermotoga maritima*, *Aquifex aeolicus* and *Helicobacter pylori*. The last subgroup (subgroup ii (Henze et al., 1998)), separated from the others in the subcluster by two indels (insertions E and G), consists of proteobacterial and cyanobacterial sequences, together with the *C. paradoxa* chloroplast sequence. Members of this subgroup are assumed to function as anabolic enzymes (Nickol et al., 2000; Plaumann et al., 1997). The relationships of the remaining sequences of the Type B cluster are in good agreement with earlier results (Henze et al., 1998; Plaumann et al., 1997) (Fig. 2A).

Our data strongly indicate that the amitochondriate protists studied acquired their FBA from sources distinct from the sources of FBA in other eukaryotes. There is no evidence from our analyses that the amitochondriate protists acquired their FBA in a single, shared event, i.e. they do not support monophyly of the enzymes from all five species. To analyse the strength of the signal against monophyly, we compared the likelihood of trees in which monophyly was constrained with the likelihood of the best tree. Results of the Shimodaira–Hasegawa test showed that a monophyly of *E. histolytica*, *M. balamuthi* and *T. vaginalis* could not be rejected ($P = 0.79$) but that monophyly of the five amitochondriate enzymes could be rejected ($P \leq 0.01$).

According to classical and molecular taxonomic data, the five amitochondriate protists that harbor the unusual FBA discussed here belong to several separate eukaryotic lineages. *Giardia* and *Spiroucleus* species are diplomonads as confirmed by morphological and molecular data (Cavaliere-Smith and Chao, 1996; Vickerman, 1990). In the past, the genera *Mastigamoeba* and *Entamoeba* have been assigned to separate major lineages, Pelobionta and Entamoebida, respectively. Based on biological considerations they were placed recently, together with the mitochondriate

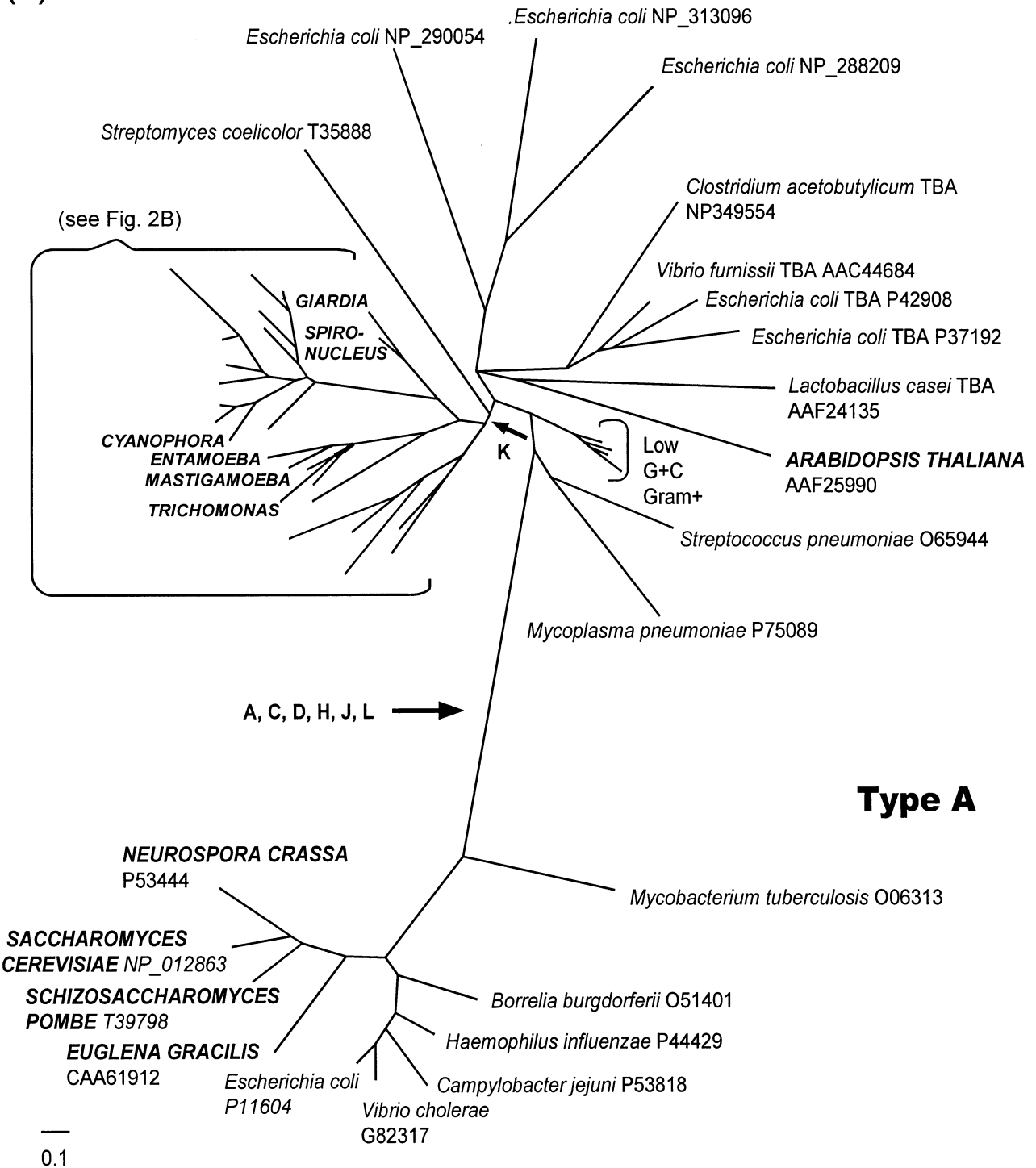
Fig. 1. Alignment of putative fructose-1,6-bisphosphate aldolase sequences from amitochondriate protists with selected Class II aldolase homologs. Residues identical in over 60% of the sequences are on black background, conservative substitutions in grey background. Conserved residues of known functional significance are marked by arrows (↓). Solid circles (●) over the alignment denote the positions where stop codons are assumed to code for Gln in the *Spiroucleus barkhanus* sequence. Letters A–M denote indels of taxonomic value. The phylogenetic position of these indels are marked by arrows in Fig. 2A,B. Sequences included are: G. i., *Giardia intestinalis* (AAD05239); S. b., *Spiroucleus barkhanus* (AF461037); M. b., *Mastigamoeba balamuthi* (BE636694); E. h., *Entamoeba histolytica* (AY05996); T. v., *Trichomonas vaginalis* (AF390544); T. p., *Treponema pallidum* (O83668); C. t., *Chlorobium tepidum* (preliminary sequence obtained from the websites of The DOE Joint Genome Institute at <http://www.jgi.doe.gov>); C. d., *Clostridium difficile* (preliminary sequence obtained from the websites of The Sanger Centre at <http://www.sanger.ac.uk>); S. sp., *Synechocystis* sp. (Q55664); R. e., *Ralstonia eutropha* (Q59101); T. a., *Thermus aquaticus* (AAF22441); T. m., *Thermotoga maritima* (G72397); E. c. T., *Escherichia coli* tagatose-1,6-bisphosphate aldolase (P42908); B. s., *Bacillus subtilis* (P13243); E. c. F., *Escherichia coli* FBA (P11604); S. c., *Saccharomyces cerevisiae* (NP_012863).

slime molds, into a monophyletic group, the Conosa (Cavalier-Smith, 1998). This assignment is now receiving support from molecular data (Arisue et al., 2002; Baptiste et al., 2002; Horner and Embley, 2001). Interestingly, so far no Class II FBA gene has been detected in the slime mold *Dictyostelium discoideum*, which has a gene for a typical Class I FBA, as indicated by a BLAST search of the *Dictyos-*

telium genome project database (<http://dicty.sdsc.edu/>). It is possible that transition from a typical mitochondrial metabolism to a fermentative one occurred in a common ancestor of *Mastigamoeba* and *Entamoeba* after its separation from the slime mold lineages. The acquisition of the Class II FBA and the loss of a Class I enzyme might have occurred about the time of this event. The last species included in this study, *T.*

(A)

Type B



vaginalis, is a parabasalid. This group is distinct from the others in its morphology and subcellular organization of metabolism (Müller, 1998). However, sequence information from an increasing number of genes indicates that diplomonads and parabasalids shared a most recent ancestor

to the exclusion of other groups (Embley and Hirt, 1998; Horner and Embley, 2001; Roger et al., 1999). Even if this is indeed the case, the present analysis suggests that they received their FBA from separate sources.

This study shows that a common feature of amitochondri-

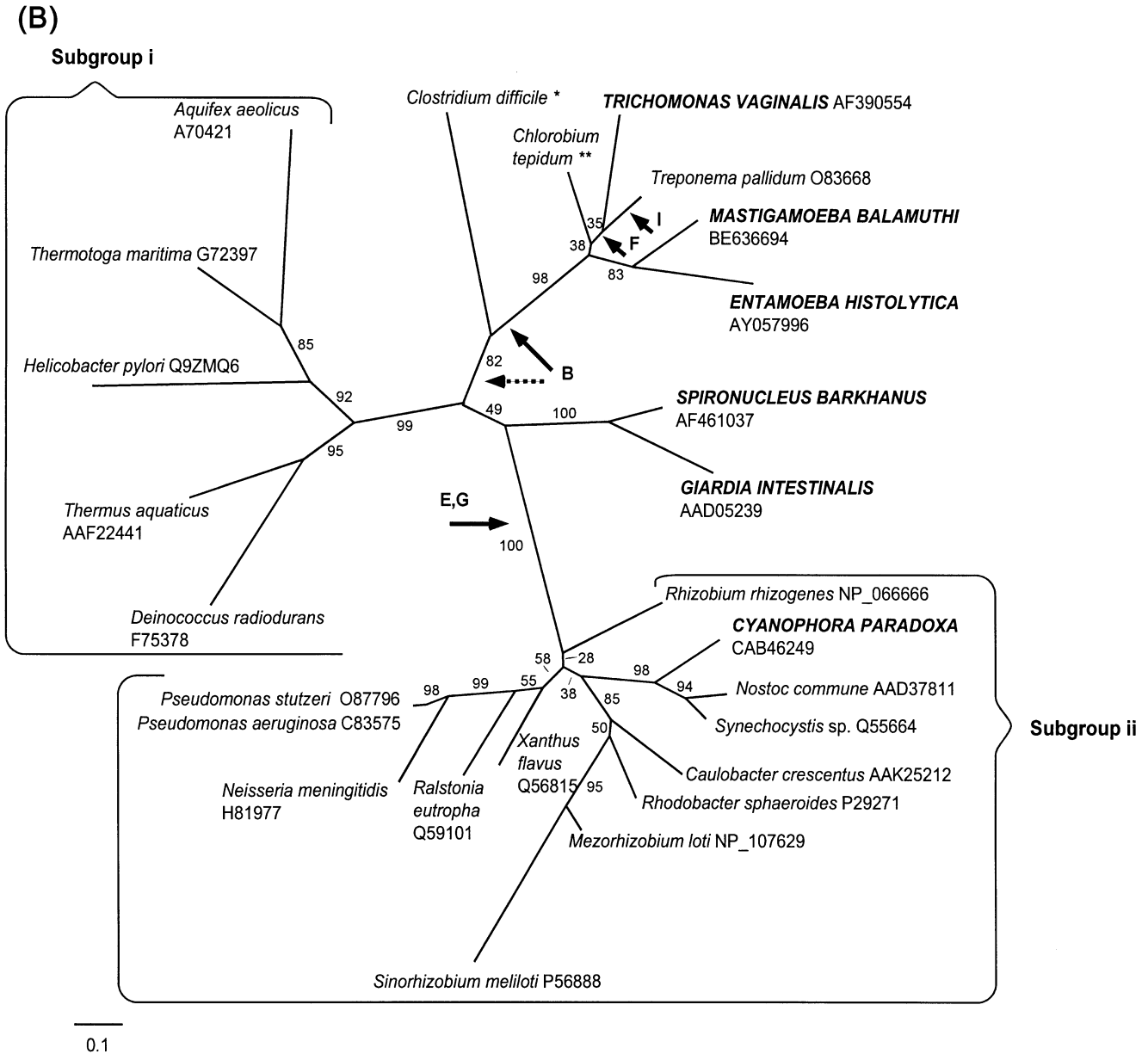


Fig. 2. Protein maximum-likelihood phylogenetic tree for 52 Class II fructose biphosphate aldolase and related sequences (214 unambiguously aligned amino acid characters). Eukaryotes are designated with boldface capitals. The branch of low G + C Gram + bacteria comprises *Bacillus halodurans* BAB07505, *B. stearothermophilus* P94453, *B. subtilis* P13243 and *Staphylococcus aureus* BAB58287. Analyses were performed using the program MrBayes (Huelsenbeck and Rannala, 1997), see Section 2. Arrows indicate branches supported by insertion/deletion events (labeled as in Fig. 1) observed in the sequence alignment. Note the long internal branch separating Type A and Type B enzymes. Taxa are labeled with species name and accession number, apart from the clade containing the protist sequences which is the subject of a more detailed analysis in (B). (B) Protein maximum-likelihood phylogenetic tree for 26 related Class II fructose biphosphate aldolase sequences (214 unambiguously aligned amino acid characters). Analyses were performed using the program MrBayes (Huelsenbeck and Rannala, 1997); numbers on branches are maximum likelihood bootstrap support values, see Section 2. Arrows indicate branches supported by insertion/deletion events (labeled as in Fig. 1) observed in the sequence alignment. Taxa are labeled by organism and accession number. Preliminary sequence data were obtained from websites of *The Sanger Centre at <http://www.sanger.ac.uk> and **The Institute for Genomic Research at <http://www.tigr.org>. Eukaryote taxa are labeled in capital letters. Note the non-monophyly of aldolase sequences from amitochondriate protists.

ate protists, which depend primarily on extended glycolysis for their ATP generation (Müller, 1998), is the presence of an FBA that is not homologous to FBAs in animals, plants and many protists, and is only distantly related to the enzyme found in fungi. Moreover, while the FBAs present in different amitochondriate protist lineages belong to the same unique subgroup, clearly separated from all other eukaryotic FBAs, they do not form a monophyletic group. Taken at face value the data indicate several separate acquisitions of genes for FBA from eubacteria, by the ancestors of amitochondriate eukaryotes. The data also do not exclude the possibility that some of the eubacteria gained their FBA from an ancestral eukaryote (i.e. the direction of transfer was from eukaryote to prokaryote). A broader taxonomic sampling of protists and eubacteria might help to resolve this conundrum.

The presence of closely related Class II Type B FBAs in five amitochondriate organisms belonging to three separate taxonomic groups is probably not a chance coincidence but might point to a selective advantage of this enzyme for amitochondriates. This assumption will have to be tested in detailed biochemical studies.

Acknowledgements

The initial culture of *Mastigamoeba balamuthi* and advice on its cultivation was kindly provided by Frederic Schuster (Brooklyn, NY). The collaboration of Christoph W. Sensen and Paul Gordon (Calgary, AL, Canada), Rama Singh (Halifax, NS, Canada) and Jennifer A. Lee and Terry Gaasterland (New York, NY) in the *M. balamuthi* EST project is gratefully acknowledged. Peter Foster (London, United Kingdom) generously provided access to the Python phylogenetics package P4 and gave invaluable help in its use. The *Entamoeba histolytica* cDNA and gDNA libraries were generous gifts of John Samuelson (Boston, MA) and Egbert Tannich (Hamburg, Germany). Cloning of the *Entamoeba histolytica* FBA was based on GSS data deposited in GenBank by the Institute of Genomic Research, Inc. (Bethesda, MD), where the *E. histolytica* genome project is performed with support from NIH-NIAID. Preliminary sequences of *Chlorobium tepidum* and *Clostridium difficile* FBA were obtained from the Institute for Genomic Research website at <http://www.tigr.org>, where sequencing of this organism was accomplished with support from DOE, and from The Sanger Centre website at <http://www.sanger.ac.uk>. This research was supported by grants to M.M. from the National Institutes of Health (AI 11942) and the National Science Foundation (MCB 9615659). K.H. held a fellowship from the European Molecular Biology Organization.

References

Anita, N.J., 1967. Comparative aldolase activity in marine planktonic algae and their evolutionary significance. *J. Phycol.* 3, 81–85.

- Arisue, N., Hashimoto, T., Lee, J.A., Moore, D.V., Gordon, P., Sensen, C.W., Gaasterland, T., Hasegawa, M., Müller, M., 2002. The phylogenetic position of the pelobiont *Mastigamoeba balamuthi* based on sequences of rDNA and translation elongation factors EF-1 α and EF-2. *J. Eukaryot. Microbiol.* 49, 1–11.
- Baernstein, H.D., 1955. Aldolase in *Trichomonas vaginalis*. *Exp. Parasitol.* 4, 323–334.
- Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Duruffé, L., Gaasterland, T., Lopez, P., Müller, M., Philippe, H., 2002. The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc. Natl. Acad. Sci. USA* 99, 1414–1419.
- Blom, N.S., Tétrault, S., Coulombe, R., Sygush, J., 1996. Novel active site in *Escherichia coli* fructose 1,6-bisphosphate aldolase. *Nat. Struct. Biol.* 3, 856–862.
- Boles, E., Zimmermann, F.K., 1993. *Sacchromyces cerevisiae* phosphoglucose isomerase and fructose bisphosphate aldolase can be replaced functionally by the corresponding enzymes of *Escherichia coli* and *Drosophila melanogaster*. *Curr. Genet.* 23, 187–191.
- Cavalier-Smith, T., 1998. A revised six-kingdom system of life. *Biol. Rev.* 73, 203–266.
- Cavalier-Smith, T., Chao, E.E., 1996. Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. *J. Mol. Evol.* 43, 551–562.
- Cooper, S.J., Leonard, G.A., McSweeney, S.M., Thompson, A.W., Naismith, J.H., Qamar, S., Plater, A., Berry, A., Hunter, W.N., 1996. The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold. *Structure* 4, 1303–1315.
- Embley, T.M., Hirt, R.P., 1998. Early branching eukaryotes? *Curr. Opin. Gen. Dev.* 8, 624–629.
- Fothergill-Gilmore, L.A., Michels, P.A.M., 1993. Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* 59, 105–235.
- Henze, K., Badr, A., Wettern, M., Cerff, R., Martin, W., 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., Morrison, H.G., Sogin, M.L., Müller, M., 1998. Sequence and phylogenetic position of a class II aldolase gene in the amitochondriate protist, *Giardia lamblia*. *Gene* 222, 163–168.
- Horner, D.S., Embley, T.M., 2001. Chaperonin 60 phylogeny provides further evidence for secondary loss of mitochondria among putative early-branching eukaryotes. *Mol. Biol. Evol.* 18, 1970–1975.
- Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276, 227–232.
- Ikawa, T., Asami, S., Nisizawa, K., 1972. Comparative studies on fructose diphosphate aldolases mainly in marine algae. *Proc. Int. Seaweed Symp.* 7, 526–531.
- Kalra, I.S., Dutta, G.P., Mohan Rao, V.K., 1969. *Entamoeba histolytica*: effects of metal ions, metal binders, therapeutics, antibiotics on aldolase activity. *Exp. Parasitol.* 24, 26–31.
- Keeling, P.J., Doolittle, W.F., 1996. A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.* 15, 2285–2290.
- Marsh, K., Leberer, H.G., 1992. Fructose-bisphosphate aldolases; an evolutionary history. *Trends Biochem. Sci.* 17, 110–118.
- Müller, M., 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists – a special case in eukaryotic evolution? In: Coombs, G.H., Vickerman, K., Sleigh, M.A., Warren, A. (Eds.). *Evolutionary Relationships Among Protozoa*, Kluwer, Dordrecht, pp. 109–131.
- Nickol, A.A., Müller, N.E., Bausenwein, U., Mayer, T.L., Schenk, H.E.A., 2000. *Cyanophora paradoxa*: nucleotide sequence and phylogeny of the nucleus encoded muroplast fructose-1,6-bisphosphate aldolase. *Z. Naturforsch.* 55c, 991–1003.
- Plaumann, M., Pelzer-Reith, B., Martin, W.F., Schnarrenberger, C., 1997. Multiple recruitment of class-I aldolase to chloroplasts and eubacterial

- origin of eukaryotic class-II aldolases revealed by cDNAs from *Euglena gracilis*. *Curr. Gen.* 31, 430–438.
- Roger, A.J., Morrison, H.G., Sogin, M.L., 1999. Primary structure and phylogenetic relationships of a malate dehydrogenase gene from *Giardia lamblia*. *J. Mol. Evol.* 48, 750–755.
- Rozario, C., Morin, L., Roger, A.J., Smith, M.J., Müller, M., 1996. Primary structure and phylogenetic relationships of glyceraldehyde-3-phosphate dehydrogenase genes of free-living and parasitic diplomonad flagellates. *J. Eukaryot. Microbiol.* 43, 330–340.
- Rutter, W.J., 1964. Evolution of aldolase. *Fed. Proc.* 23, 1248–1257.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969.
- Susskind, B.M., Warren, L.G., Reeves, R.E., 1982. A pathway for the interconversion of hexose and pentose in the parasitic amoeba *Entamoeba histolytica*. *Biochem. J.* 204, 191–196.
- Swofford, D.L., 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4, Sinauer Associates, Sunderland, MA.
- Vickerman, K., 1990. Phylum Zoomastigina class Diplomonadida. In: Margulis, L., Corliss, J.O., Melkonian, M., Chapman, M. (Eds.). *Handbook of Protozoa*, Jones and Bartlett, Boston, MA, pp. 200–210.
- Zgiby, S.M., Thomson, G.J., Qamar, S., Berry, A., 2000. Exploring substrate binding and discrimination in fructose 1,6-bisphosphate and tagatose 1,6-bisphosphate aldolases. *Eur. J. Biochem.* 267, 1858–1868.