Curr Genet (1997) 31: 430–438

ORIGINAL PAPER

Marlies Plaumann · Birgit Pelzer-Reith
William F. Martin · Claus Schnarrenberger

Multiple recruitment of class-I aldolase to chloroplasts and eubacterial origin of eukaryotic class-II aldolases revealed by cDNAs from *Euglena gracilis*

Received: 20 December 1996 / 5 February 1997

Abstract The photosynthetic protist *Euglena gracilis* is one of few organisms known to possess both class-I and class-II fructose-1,6-bisphosphate aldolases (FBA). We have isolated cDNA clones encoding the precursor of chloroplast class-I FBA and cytosolic class-II FBA from *Euglena*. Chloroplast class-I FBA is encoded as a single subunit rather than as a polyprotein, its deduced transit peptide of 139 amino acids possesses structural motifs neccessary for precursor import across *Euglena*'s three outer chloroplast membranes. Evolutionary analyses reveal that the class-I FBA of *Euglena* was recruited to the chloroplast independently from the chloroplast class-I FBA of chlorophytes and may derive from the cytosolic homologue of the secondary chlorophytic endosymbiont. Two distinct subfamilies of class-II FBA genes are shown to exist in eubacteria, which can be traced to an ancient gene duplication which occurred in the common ancestor of contemporary gram-positive and proteobacterial lineages. Subsequent duplications involving eubacterial class-II FBA genes resulted in functional specialization of the encoded products for substrates other than fructose-1,6-bisphosphate. Class-II FBA genes of *Euglena* and ascomycetes are shown to be of eubacterial origin, having been acquired via endosymbiotic gene transfer, probably from the antecedants of mitochondria. The data provide evidence for the chimaeric nature of eukaryotic genomes.

Key words *Euglena gracilis* · Endosymbiosis · Endosymbiotic gene transfer · Molecular evolution

M. Plaumann · B. Pelzer-Reith · C. Schnarrenberger (✉)
Institut für Pflanzenphysiologie und Mikrobiologie,
Freie Universität Berlin, Königin-Luise Strasse 12–16a,
D-14195 Berlin, Germany

W. F. Martin
Institut für Genetik, Technische Universität Braunschweig,
Spielmannstrasse 7, D-38106 Braunschweig, Germany

Communicated by F. K. Zimmermann

Introduction

Fructose-1,6-bisphosphate aldolase (FBA) (EC 4.12.1.13) catalyzes the reversible aldol condensation of dihydroxyacetone phosphate and glyceraldehyde-3-phosphate in the Calvin cycle, glycolysis and gluconeogenesis, and is thus essential to primary metabolism in all cells. Two very distinct types of FBA enzymes occur in nature, which differ in their catalytic mechanism and can be distinguished by biochemical means. Class-I FBA enzymes form a Schiff-base with the substrate during catalysis via condensation of the ε-amino group of an active-center lysine residue with the carbonyl group of the substrate, and can be inhibited by borohydride reagents. Class-II FBA enzymes require divalent cations as co-factors which stabilize the carbanion intermediate formed during the reaction, and can be inhibited by EDTA (Rutter 1964; Lebherz and Rutter 1969). Class-I FBAs are homotetramers, whereas class-II FBAs are homodimers. The subunit size of both classes of FBA enzymes is 40 kDa but, importantly, class-I and class-II FBA monomers share no detectable sequence similarity. This suggests that class-I and class-II FBA enzymes arose independently during evolution.

The phylogenetic distribution of FBA enzymes is complex and intriguing (for a review see Schnarrenberger et al. 1992). Eubacteria, including cyanobacteria, typically possess class-II FBAs (Rutter 1964; Antia 1967), although a few clearly documented instances of class-I FBA occurrence in eubacteria are known (Witke and Götz 1993). Halophilic archaebacteria possess either class-I or class-II FBAs (Dhar and Altekar 1986); the distribution of the enzyme across thermophilic archaebacteria has not been studied. Among higher eukaryotes, fungi typically possess class-II FBAs whereas metazoa and higher plants possess class-I FBAs (Rutter 1964; Schnarrenberger et al. 1992). The distinct chloroplast and cytosolic FBA isoenzymes of higher plants are both of the class-I type (Anderson and Advani 1970; Krüger and Schnarrenberger 1983; Lebherz et al. 1984). The chlorophytes *Chara foetia* and *Klebsormidium flaccidum* also possess class-I FBAs in the chloroplast and the cytosol (Jacobshagen and Schnarrenberger 1988, 1990) whereas *Chlamydomonas reinhardtii* possesses only a chloroplast class-I FBA (Schnarrenberger et al. 1994).

In earlier-branching protists, the distribution of class-I and class-II FBAs is more complicated (Rutter 1964; Antia 1967; Ikawa et al. 1972). *Euglena gracilis* is unique among eukaryotes studied to-date in that it possesses both class-I *and* class-II FBAs. The Calvin cycle enzyme of *Euglena*'s chloroplasts is a class-I FBA, as in higher plants, whereas the glycolytic/gluconeogenetic enzyme of the cytosol is a class-II FBA (Mo et al. 1973; Pelzer-Reith et al. 1994), as in fungi. This contrasts sharply with the situation found in the kinetoplastid *Trypanosoma brucei,* which possesses only one class-I FBA active in a specialized glycolytic microbody, the glycosome (Marchand et al. 1988). *Trypanosoma* shares a common line of nucleo-cytoplasmic descent with *Euglena* (Surek and Melkonian 1986; Sogin et al. 1989; Walne and Kivic 1989) and *Euglena'*s plastids are thought to have arisen through engulfment of a eukaryotic, possibly chlorophytic, alga (Gibbs 1978; Lefort-Tran 1981). In order to determine whether the difference between *Euglena* and *Trypanosoma* with regard to FBA enzymes may have involved endosymbiotic gene transfer surrounding the origin of *Euglena*'s chloroplasts, we have investigated *Euglena*'s nuclear-encoded chloroplast class-I and cytosolic class-II fructose-1,6-bisphosphate aldolases.

Materials and methods

Cultivation of E. gracilis cells. E. gracilis strain Z (No. 1224-5/25) was obtained from the Sammlung von Algenkulturen of the University of Göttingen (FRG). Heterotrophic cells were raised from autotrophic 15-l cultures in the medium described by Böger and San Pietro (1967) supplied with 5% CO_2 in transparent polycarbonate vessels at 27°C. After transfer to darkness, the cells were additionally supplied with 5% sucrose.

Isolation and analysis of cDNA clones. The *E. gracilis* strain Z (1224-5/25) cDNA library previously described by Henze et al. (1995) was screened using a spinach cDNA clone for chloroplast aldolase (Pelzer-Reith et al. 1993) and a cDNA clone coding for the class-II aldolase (pGHS001) from *Saccharomyces cerevisiae* (Schwelberger et al. 1989) as hybridization probes. Filters were hybridized overnight in 3× SSPE, 0.2% polyvinylpyrrolidone, 0.2% Ficoll 400, 0.1% SDS, 50 μg/ml of denatured salmom-sperm DNA and the respective radioactively labelled hybridization probe. Hybridization and washing was performed at 53°C for class-I FBA and at 49°C for class-II FBA. Filters were washed three times for 10 min with 2× SSC and 0.1% SDS at the hybridization temperature. Positive plaques were purified, cDNA inserts of positively hybridizing phages were subcloned in pUC18 and sequenced on both strands using the dideoxy method on double-stranded DNA templates with T_7 polymerase (Pharmacia) according to the supplier's protocol. Oligonucleotides (17-mers) were synthesized for sequencing as needed.

Data handling. Sequence analyses were carried out with the GCG program (Devereux et al. 1984). Amino-acid sequences were aligned with CLUSTAL W (Thompson et al. 1994), the aligment was refined by eye with the LINEUP program of GCG. Pairwise distances between sequences were estimated using the Dayhoff matrix option of PROTDIST in PHYLIP (Felsenstein 1993); trees were constructed by the neighbor-joining method (Saitou and Nei 1987). The reliability of branches was estimated by bootstrapping.

DNA and RNA analysis. Genomic DNA was isolated by ethanol precipitation from the LiCl supernatant of the RNA preparation. For genomic Southern blots, 15 μg of DNA was digested with either *Bam*HI, *Hin*dIII, *Kpn*I, or *Pst*I, fractionated on 0.8% agarose gels, blotted onto nylon membranes overnight, and fixed by exposure to UV light. Total RNA was extracted and purified as described (Houlné and Schantz 1987). For Northern blots, 10 μg of total RNA were separated by electrophoresis on a 1.2% agarose-formaldehyde gel, blotted onto a nylon membrane, and fixed by exposure to UV light. Southern and Northern blots were probed with the corresponding *Euglena* cDNA inserts labelled by random-primed synthesis. Hybridization and washing conditions were the same as described for the screening procedures except that temperatures were increased to 65°C for all hybridization and washing steps.

Results and discussion

Several aspects of nuclear gene structure and organization in *Euglena* differ from that in higher eukaryotes. Some of *Euglena*'s nuclear-encoded chloroplast proteins are translated as multimeric polyproteins which are proteolytically processed upon import into the chloroplast, e.g. the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Chan et al. 1990) and the LHCII proteins (Houlne and Schantz 1988, 1993). Also, chloroplast transit peptides in *Euglena* contain an ER-specific signal sequence required for import of the cytosolic precursors across the outer of its three chloroplast membranes (Sharif et al. 1989; Kishore et al. 1993). To further investigate *Euglena*'s nuclear gene organization, we isolated and characterized cDNA clones for class-I and class-II fructose-1,6-bisphosphate aldolases.

Clones for chloroplast class-I FBA from *Euglena*

From 120 000 independent recombinants of the *Euglena* cDNA library, screened with a heterologous cDNA for spinach chloroplast class-I aldolase, we obtained 25 hybridizing positives. The five longest *Eco*RI inserts were subcloned and shown by terminal sequencing to represent one and the same transcript since they contained identical nucleotide sequences at their 3′ and the 5′ ends. The complete nucleotide sequence of the cDNA insert of plasmid pEgAlp1 is 1662 bp long (Fig. 1). Starting with the first in-frame methionine codon, pEgAlp encodes a precursor protein of 495 amino acids with a predicted molecular mass of 52.46 kDa. Comparison of the deduced protein with chloroplast aldolases from higher plants and *C. reinhardtii* suggests an N-terminus for the mature subunit at Leu^{139} (Fig. 1). This cleavage site yields a mature protein with a predicted M_r of 40.7 kDa, which is in good agreement with the molecular mass of the purified *Euglena* class-I enzyme previously determined as 40 kDa by SDS polyacrylamide-gel electrophoresis (Pelzer-Reith et al. 1994).

The putative transit peptid of 139 amino acids possesses a hydrophilic N-terminal region followed by a long stretch of non-polar amino acids (amino acids 23 to 44), as found in signal peptides of secreted proteins. Comparison of the transit-peptide region of class-I FBA to other transit peptides of nuclear-encoded chloroplast proteins in *Euglena* revealed that it possesses two highly hydrophobic domains separated by a roughly 60 amino-acid hydrophilic stretch rich in hydroxylated amino acids (Fig. 2). These bipartite transit peptides are much longer than those typical for higher-plant chloroplasts (von Heijne et al. 1989) and are known to contain topogenic signals for targeting to the endoplasmic reticulum during precursor import across *Euglena*'s three outer chloroplast membranes (Kishore et al. 1993).

The gene for chloroplast class-I FBA in *Euglena* is transcribed and translated as a precursor which encodes a single subunit, rather than as a polyprotein as in the case of nuclear-encoded genes for the small subunit of Rubisco (Chan et al. 1990) and LHCP (Houlné and Schantz 1993). The size of the mRNA in a Northern blot (Fig. 3) corresponds to the length of the cDNA. A Southern blot of genomic DNA probed with the complete cDNA insert encoding class-I FBA shows several bands with different intensities (Fig. 4). This suggests that chloroplast FBA in *Euglena* is organized as a small multigene family, or else that the gene(s) contain(s) several introns, as has previously been described for the *RbcS*, *Lhcp* and *GapC* genes (Muchhal and Schwartzbach 1992; Henze et al. 1995; Tessier et al. 1995).

Clones for cytosolic class-II FBA from *Euglena*

From the same library we found 18 clones that hybridize to the insert for class-II FBA from yeast. The longest of these, pEgAlc1, has an insert of 1203 bp and encodes an open reading frame of 1068 bp (but lacks the start codon) plus a 3′ non-coding region of about 135 bp (Fig. 5). A

```
                                                          60
ctttatccaaaaATGAGCGCAACTCAATTGCTGGGTGGTTACCAGACCATTTCTGAGCGC
            M  S  A  T  Q  L  L  G  G  Y  Q  T  I  S  E  R
                                                         120
ACTGCTCCCCAGAACAAGCTTGCTGTTGTTGGTGCCGTGGCTGTCATTGCTGTGGCCTCG
T  A  P  Q  N  K  L  A  V  V  G  A  V  A  V  I  A  V  A  S
                                                         180
GGCGTTGCAGCTGGCTATGCGCTCAATCTCTCTATGCCCCCTCTGTTGCCTCCGTGCGCA
G  V  A  A  G  Y  A  L  N  L  S  M  P  P  L  L  P  P  C  A
                                                         240
CCACTTGCGCACGGGCACCTATCGAGAACAACTTGGCACAGGCCAACTCCCGCTGCAGTT
P  L  A  H  G  H  L  S  R  T  T  W  H  R  P  T  P  A  A  V
                                                         300
GCCTTTGCCTCTGCCTCTGAAGGTGCTCAAACCTTTGTGGCTCCTGCCGCTCAGTCCTCC
A  F  A  S  A  S  E  G  A  Q  T  F  V  A  P  A  A  Q  S  S
                                                         360
AACACCTTTGCAACTTCCTCCGTGGCTGCCTCCATTGGTATGGTGATGGGAGCAGGCGCC
N  T  F  A  T  S  S  V  A  A  S  I  G  M  V  M  G  A  G  A
                                                         420
GTGCTGCTGGCCCGGATGAACCAGAAGCCAGTTGCCATGAACGCCTGGACGGGCTCTGTT
V  L  L  A  R  M  N  Q  K  P  V  A  M  N  A  W  T  G  S  V
                                                         480
TATGGTGTGCTCTCTCAGCTGTGGAGAAGCCATGGACCAAGTATGCTGATGAGCGTAAGA
Y  G  V  L  S  Q  L  W  R  S  H  G  P  S  M  L  M  S  V  R
         <||>                                            540
CAAGATCATCTGCACCCCTGGCAAGGTATTCTTGCTGCGGATGAGTCCAGACCGAACAAG
Q  D  H  L  H  P  W  Q  G  I  L  A  A  D  E  S  R  P  N  K
                                                         600
ACCTGCGGTGCGCGCCTGAAGTCAATTGGTGTGGAGAACACAGAGGAGAACGTGCACAGT
T  C  G  A  R  L  K  S  I  G  V  E  N  T  E  E  N  V  H  S
                                                         660
AGCCGCAGTTGTGTTCACCGCACCTGGCTTCAGCGAGGAAATATTCTTGGTGTGATCATG
S  R  S  C  V  H  R  T  W  L  Q  R  G  N  I  L  G  V  I  M
                                                         720
TACGAGGAGACTTTGTACCAGAAGGATAAGAATGGAAAGCCTTTTGTGCAAATCATCAAC
Y  E  E  T  L  Y  Q  K  D  K  N  G  K  P  F  V  Q  I  I  N
                                                         780
GAGGCTGGAGCTGTTGCTGGTGTCAAGGTGGACACTGGCATTGCACCGCTGCCAGGTGCC
E  A  G  A  V  A  G  V  K  V  D  T  G  I  A  P  L  P  G  A
                                                         840
GATGATGAGGGCTACACCATGGGTCTGGATGGCCTTCGTGAGAGGTGCCAGGAATACTAC
D  D  E  G  Y  T  M  G  L  D  G  L  R  E  R  C  Q  E  Y  Y
                                                         900
AGACAGGGTGCGCGCTTTGCTAAGTGGCGTGCAGTGCTCCGCATCGACTCCAAGGGACTG
R  Q  G  A  R  F  A  K  W  R  A  V  L  R  I  D  S  K  G  L
                                                         960
CCATCTGACCGCTCCATCCTGGCCAATGCTACTGGTTTGGCACAGTATGCTGCCATCTGC
P  S  D  R  S  I  L  A  N  A  T  G  L  A  Q  Y  A  A  I  C
                                                        1020
CAGGAGTGTGGCTTGGTTCCAATTGTTGAGCCTGAGATCCTGATGGATGGTGACCATGAC
Q  E  C  G  L  V  P  I  V  E  P  E  I  L  M  D  G  D  H  D
                                                        1080
ATTGAGACCGCAGCTGCTGCTGCCGAGCGTGTCCTTGTTGCTGTCTATGACGCTCTTGCA
I  E  T  A  A  A  A  A  E  R  V  L  V  A  V  Y  D  A  L  A
                                                        1140
ACACAGGGAGTTCTTTTGGAGGGGACTCTGTTGAAGCCCAACATGGTGACTCCTGGTGTG
T  Q  G  V  L  L  E  G  T  L  L  K  P  N  M  V  T  P  G  V
                                                        1200
GACTCTGGTATCAAGGCCACCCCAGAGACATTGCCTATTTTCACTTGCGCACTTCTTGCC
D  S  G  I  K  A  T  P  E  T  L  P  I  F  T  C  A  L  L  A
                                                        1260
ACTGTTCCAGCTGCTGTTGCAGGTATCAGTTTCTTGTCCGGAGGTAGTTCTGAGGAGGAT
T  V  P  A  A  V  A  G  I  S  F  L  S  G  G  S  S  E  E  D
                                                        1320
GCATCCTTGAACCTGAACGCCATCAATGCAATCCCATACGAGCGCAAGCCTTGGGCACTG
A  S  L  N  L  N  A  I  N  A  I  P  Y  E  R  K  P  W  A  L
                                                        1380
ACCTTCTCGTTTGGTCGTGCCCTCCAGGCCTCCACCCTCAAGACTTGGGGAGGCAAGGAC
T  F  S  F  G  R  A  L  Q  A  S  T  L  K  T  W  G  G  K  D
                                                        1440
GAGAACATCGCGCGGCTCAGAAGGTGTTCGCTGAGCGTGCCAAGCCAATGGCCTGGCAAG
E  N  I  A  R  L  R  R  C  S  L  S  V  P  S  Q  W  P  G  K
                                                        1500
TACCAGGGCTCTGGGCAGGCTGGTGAGTCCTTGTTTGTCAAGGGCTACAAGTACTAAttt
Y  Q  G  S  G  Q  A  G  E  S  L  F  V  K  G  Y  K  Y  *
                                                        1560
gaaacatcctgccttcctctttctcaccttatctattgttgggcgagaaatcccatgcta
                                                        1620
gtcagggatcacctagtgaaaacaggaccatcgacagaaggctactctctgttagaccgc
                                     1661
agcgaaagaccttgacccatggggacccctgttttgagcgg
```

Fig. 1 Nucleotide and deduced amino-acid sequence of the cDNA pEgAlp1 encoding the chloroplast class-I FBA precursor of *E. gracilis*. The putative transit peptide is indicated in *italics*, the putative processing site is indicated by "<||>". Non-coding regions are shown in *lower case*

Fig. 2 Comparison of transit peptide regions for nuclear-encoded chloroplast proteins of *E. gracilis*. Sequences were taken from this paper and from the data base. Positively charged, negatively charged, and hydroxylated amino acids are indicated by "+", "–", and (○), respectively. Hydrophobic domains are in *bold type and double underlined*. Precursor abbreviations and accession numbers to the sequences are: *FbaI* fructose-1,6-bisphosphate aldolase (X89768); *GapA* glyceraldehyde-3-phosphate dehydrogenase (L21904); *RbcS* small subunit of Rubisco (X79154); *Lhcp* light-harvesting complex protein II (U03392); *Hmbs* hydroxymethylbilane synthase (X15743); *PS30* extrinsic 30-kDa protein (OEC30) of photosystem II (D14702); *IF3* chloroplast initiation factor 3 (P36177)

```
       o o       o o-+o    +             o          o            + + o+oo ++ o       o o-   o       oo o  ooo   o             +   +       o o
FbaI  MSATQLLGGYQTISERTAPQNKLAVVGAVAVIAVASGVAAGYALNLSMPPLLPPCAPLAHGHLSRTTWHRPTPAAVAFASASEGAQTFVAPAAQSSNTFATSSVAASIGMVMGAGAVLLARMNQKPVAMNAWTGSVYGVL>
           -  +                    oooo   o+           +  o  o  o  o     o +  -   -oo-  o   o-+oo  o                 +o  +  -o         o  o  oo ++o
Lhcp  MPNNADAMKFGLAAGAAMGVIVYVLAGAASSTSLAATHVNIQQAPAVIPRMASVPSAYTIATNPIGASARVVDANYESTDYLTLPATEKSTMGSLLMIAAAGVAAAVAFVWKSVPRQQDSVINVPLLPVSVATMATSGKKS>
         -+    o -+    oo                 oo   o  o o +      o  o  +  -o  o+ o  + o        o   -o-o++   o      o        +    + +       oo-+-
RbcS  MPFDRQPLLSGEKGMPATSLWLVGGAVIAAVCVIVNTSYNGTQLSVTARPIQAAVSQVSMARFAESGVSRGSGNRVSQAVPLMAASVGAESESRRWVASAILFPLSGLFAAVALKMAMMKPKVAAVLPFTSEKD>
          + -o  -o+  o                             o   oo  oo  +        oo   oo      +-o     o      -oo +  -+o o     o  o o     +    + +oo+  -  - o  +  o
Hmbs  MYCGRYETIGETRGNSLNVFIGAAAGFVAAVALINSGLATSFYSTPVRAVPQVIVPSSLAASSQLPVVPKETNIQVNSAQILYPDSTVKGQERTITILGVCSFLSASLFYIWKQFGMKARTTKPADLQEVSGGRIWSLA>
       - +     o              oo o oo  +     + oo+   o   +   o+    o    -     -       o  oo                   ++  o-  o -  -     o  o-       o   +    o o
GapA  MDAKNVVVGTMGVAACVAVGMAVGSSMSATTQMHVQPAVRATTRVQPTLAMRGQISRLAANSQVYAEAPAAYEFQAPAQASGVSSGIAAMVMAPVAAAAAFFAYKKGQSDAQTYEVADPAGVQTYGTELVFAMNATIGNKYPGNSNT>
        +oo    -  o  o oo            o        o oo o   o    oo+         oo    o   ++  +  + +  o- oo  o + o       oo          o           o ++-    o
IF3   MVRSSCLQCDQPSQGSNSTFGCGGPLAAVCATGLLVLVLYSPSSQTANWSAQGISTKALYPAVPVPSTLLPGSAPAKHQLHVWRAHAMSEATTNNSFKQSLFGYNAISSIWLQLAGVAATFFAFGALMAAVTQRKEIAVFS>
                         o+        +   +   --   -  o o         o            +++ o o -  +-- -       +             o    o o --  o o  - +oo    o
Ps30                    MSRVNLAAPANHVVPRAAQEEVQGEYQTASNNWAVAAMASAGAAVGAAVLAMRRRATNTYEAIREDPEAVLAGAGRAMGAALIGAAVAGSANAASLTYDELQSLSYLEVKSSGIAGTC>
```

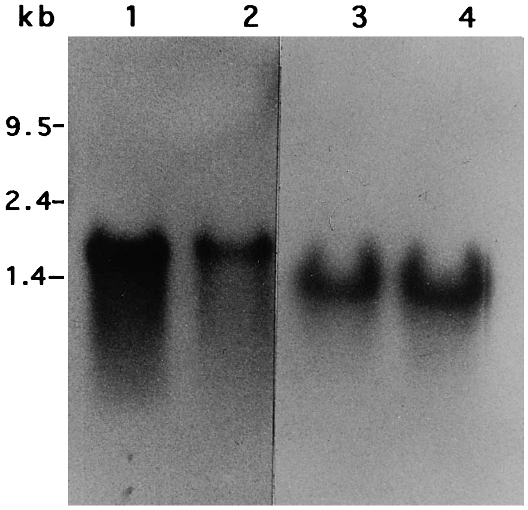

Fig. 3 Northern-blot analysis of RNA from *E. gracilis* probed with plastidic class-I (*lanes 1 and 2*) and cytosolic class-II (*lanes 3 and 4*) aldolase cDNA. *Lanes 1 and 3* contain 10 µg each of RNA isolated from autotrophically grown cells; *lanes 2 and 4* contain 10 µg of RNA from heterotrophically grown cells

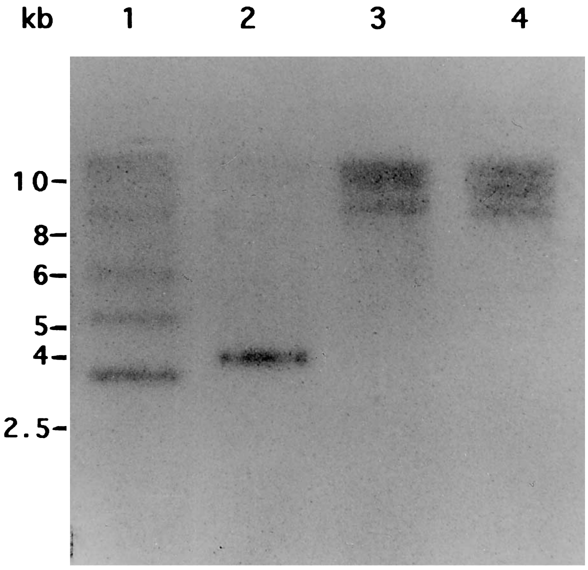

Fig. 4 Southern-blot analysis of the class-I aldolase gene in *E. gracilis* genomic DNA; 15 µg of nuclear DNA was digested with *Bam*HI (*lane 1*), *Pst*I (*lane 2*), *Kpn*I (*lane 3*), and *Hin*dIII (*lane 4*). The probe was the cDNA fragment coding for the class-I aldolase from *E. gracilis*. *Numbers* on the left indicate the size (kb) of DNA markers

```
                                                          60
CCTGATTTTCCCAAAGACCTGAAGGGTGTTTTGGACGGCAACCAAGTCCGAACTCTGTTC
P  D  F  P  K  D  L  K  G  V  L  D  G  N  Q  V  R  T  L  F
                                                         120
GACTTCGCGCAGAAGAAGGGCTTCGCAATCCCTGCTGTGAACTGCACATCGTCTTCCACC
D  F  A  Q  K  K  G  F  A  I  P  A  V  N  C  T  S  S  S  T
                                                         180
GTGAATGTGGTGCTGGAACGGGCCCGAGACACCCACAACCCGGTCATCATCCAGGTGTCC
V  N  V  V  L  E  R  A  R  D  T  H  N  P  V  I  I  Q  V  S
                                                         240
CAGGGTGGTGCTGCCTTCTATTGTGGCAAGGGAGTGAAGGATGAGAAGCTGATTGCCAGT
Q  G  G  A  A  F  Y  C  G  K  G  V  K  D  E  K  L  I  A  S
                                                         300
GTGGACGGCTCCGTGGCGTTGGCACACCACGTGCGGGCCGTTGCCCATACTATGGCGCCA
V  D  G  S  V  A  L  A  H  H  V  R  A  V  A  H  T  M  A  P
                                                         360
GTTGTCGTTCATTCCGACCATTGTGCCAAGAAGCTGCTTCCGTGGTTCGATGGCATGTTG
V  V  V  H  S  D  H  C  A  K  K  L  L  P  W  F  D  G  M  L
                                                         420
GATGCCGATGGAGAAATATTTTGCGAGCACGGCGTCCCCCTCTTCTCCAGCCACATGCTC
D  A  D  G  E  I  F  C  E  H  G  V  P  L  F  S  S  H  M  L
                                                         480
GATTTGAGTGAAGAAAATGACGAGGAGGACATTGGCACGTGCGTGAAGTATTTCACTCGA
D  L  S  E  E  N  D  E  E  D  I  G  T  C  V  K  Y  F  T  R
                                                         540
ATGGCCAAGTTGAACCTGTGGTTGGAGATGGAGATTGGTATGACTGGAGGTGTGGAAGAT
M  A  K  L  N  L  W  L  E  M  E  I  G  M  T  G  G  V  E  D
                                                         600
GGCGTGGACAACAGTGGGGTGGCCAATGACAAGCTGTACACGTCGTCGGAGCAGGTCTTT
G  V  D  N  S  G  V  A  N  D  K  L  Y  T  S  S  E  Q  V  F
                                                         660
GCTGTCCACAAAGCACTCGGCGCCAGCTCGCCAAACTTCTCCATTGCTGCTGCTTTCGGC
A  V  H  K  A  L  G  A  S  S  P  N  F  S  I  A  A  A  F  G
                                                         720
AATGTCCACGGTGTGTACAAGCCAGGCAATGTGAAGTTGCAGCCCAATCTGTTGAAGGAG
N  V  H  G  V  Y  K  P  G  N  V  K  L  Q  P  N  L  L  K  E
                                                         780
CACCAGGATTACGCCCGGAAGCAGTTGTCCTCCTCTGAGGACCATCCTCTCTACCTCTGG
H  Q  D  Y  A  R  K  Q  L  S  S  S  E  D  H  P  L  Y  L  W
                                                         840
TTCCATGGTCCTTCAGGCTCGACAGATGCAGAGATTCATGAAGCAGTGCGGAATGGGGTG
F  H  G  P  S  G  S  T  D  A  E  I  H  E  A  V  R  N  G  V
                                                         900
GTCAAGATGAACCTTGATACTGACATGCAGTGGGCATACTGGGATGGATTGCGCCAGTTT
V  K  M  N  L  D  T  D  M  Q  W  A  Y  W  D  G  L  R  Q  F
                                                         960
GAGGCAAAGAAACACGATTATTTGCAGGGACAGATTGGGAACCCCGAGGGCCCTGACAAG
E  A  K  K  H  D  Y  L  Q  G  Q  I  G  N  P  E  G  P  D  K
                                                        1020
CCGAACAAGAATTATTACGACCCACGAAAGTGGATCCGTGAGGCCGAGCTTGGCATGCTG
P  N  K  N  Y  Y  D  P  R  K  W  I  R  E  A  E  L  G  M  L
                                                        1080
GCCCGCGTCAAGGTGGCCTTCAAGGCGGTTGAATTGCCCGGTGGCCTAAAGGAGTTCATT
A  R  V  K  V  A  F  K  A  V  E  L  P  G  G  L  K  E  F  I
                                            1128
GGTATCCCCTGAaccacacgactcgtacctggaacgaaacgtttatgg
G  I  P  *
```

Fig. 5 Nucleotide sequence and deduced amino-acid sequence of cDNA pEgAlc1 encoding the cytosolic class-II FBA of *E. gracilis*. The start codon and perhaps one or two additional N-terminal residues are not contained in the cDNA as judged by comparison to the size of the isolated protein and by alignment to other class-II FBA enzymes (Pelzer-Reith et al. 1994; see also text)

Northern blot of *Euglena* RNA probed with pEgAlc1 (Fig. 3) reveals a single band of 1.3 kb, indicating that the cDNA clone lacks a total of about 100 bp of 5′ and/or 3′ UTR. The predicted molecular mass of the encoded product is 39.1 kDa, in good agreement with the molecular mass of 38 kDa previously determined for the cytosolic class-II FBA subunit in SDS-polyacrylamide gels (Pelzer-Reith et al. 1994), suggesting that pEgAlc1 lacks only very few of the N-terminal residues contained in *Euglena*'s class-II FBA subunit. A Southern blot probed with the insert of pEgAlc1 (Fig. 6) revealed only a few hybridizing bands, indicating a less-complex gene organization for class-II FBA than for class-I FBA in *Euglena*.

Origin of the gene for chloroplast class-I FBA in *Euglena*

In the most straightforward scenario for the origin of *Euglena*'s nuclear-encoded chloroplast class-I FBA gene, the class-I chloroplast FBA of the endosymbiont simply could have been transferred to the kinetoplastid nucleus in the course of the degeneration of the secondary symbiont's nucleus, as was suggested for *Euglena*'s *RbcS* (Martin et al. 1992) and *GapA* (Henze et al. 1995) genes. In that

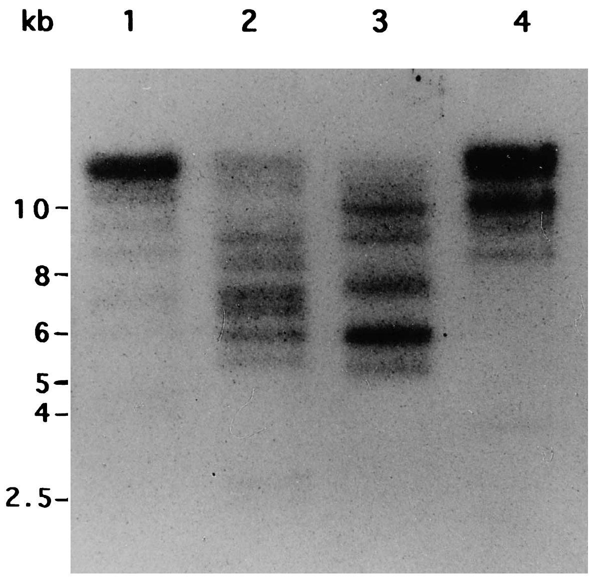

Fig. 6 Southern-blot analysis of the class-II aldolase gene of the *E. gracilis* genomic DNA; 15 µg of nuclear DNA was digested with the enzymes *Bam*HI (*lane 1*), *Pst*I (*lane 2*), *Kpn*I (*lane 3*), and *Hin*dIII (*lane 4*). The probe was the cDNA fragment of the class-II aldolase of *E. gracilis*. *Numbers* on the left indicate the size (kb) of DNA markers

case, the class-I FBA of *Euglena* should share a common branch with chloroplast homologues of chlorophytes. But the gene tree reveals that the chloroplast class-I FBA of *Euglena* shows no specific affinity to the chloroplast FBA enzymes of chlorophytes (Fig. 7). Rather, it branches with weak bootstrap support between the cytosolic FBA enzymes of higher plants and those of metazoa. Therefore, alternative evolutionary routes for its origin have to be considered.

A second reasonable possibility would be that *Euglena* chloroplast FBA was recruited via duplication of the preexisting gene for cytosolic FBA from the kinetoplastid host nucleus following secondary endosymbiosis. In that case one would expect the *Euglena* class-I enzyme to share a common branch with the homologue from *T. brucei* (Sogin et al. 1989; Levasseur et al. 1994; Henze et al. 1995). This is also not the case (Fig. 7), indicating that recruitment from the cytosolic homologue of the secondary host is unlikely.

As a third possibility, the very low bootstrap values separating the *Euglena* class-I aldolase from its cytosolic homologues of chlorophytes could be reconciled with the view that *Euglena*'s chloroplast class-I FBA might have been recruited from the endosymbiotic chlorophyte's gene for cytosolic aldolase. Although the present data do not lend strong support to this notion, it is in our view the alternative that can most easily account for the data, notwithstanding the possibility of as yet unrecognized gene duplications. The very weak affinity observed between *Euglena* and *Plasmodium* aldolase genes in Fig. 7 is not detected with other distance-estimation methods (data not shown) and is probably altogether insignificant.

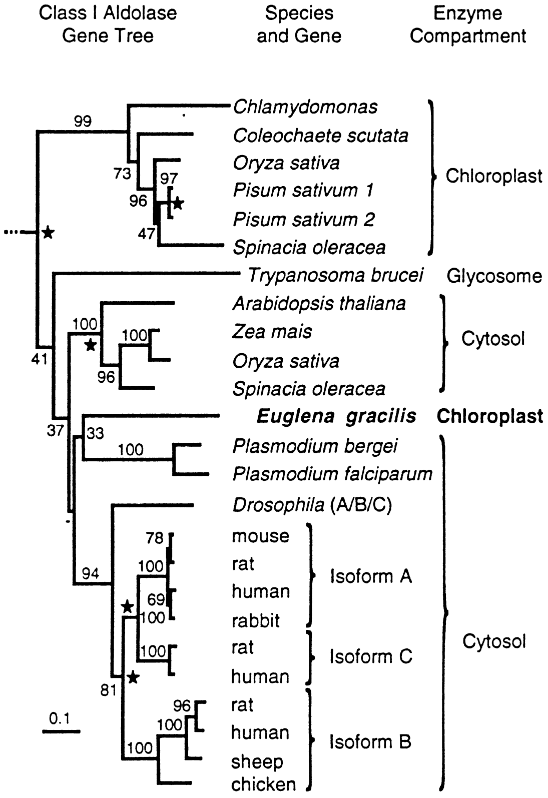

Fig. 7 A class-I fructose-1,6-bisphosphate aldolase gene tree. The tree was constructed by the neighbor-joining method for the matrix of numbers of amino-acid substitutions per site estimated with the Dayhoff matrix option of PROTDIST in PHYLIP. *Numbers at branches* indicate the bootstrap proportion for 100 replicates using the same distance-estimation method. *Stars at nodes* indicate the presence of possible gene duplications. The scale bar indicates 0.1 substitutions per site; the length of the branch bearing the *S. carnosus* outgroup is 1.4. Compartmentalization of the respective gene product is indicated. Sequences were extracted from GenBank, Swissprot and PIR data bases. Accession numbers are: *Chlamydomonas* chloroplast S48639, *Coleochaete* chloroplast (R. Kämmerer unpublished), rice chloroplast D13513, pea chloroplast 1 S29047, pea chloroplast 2 S29048, spinach chloroplast P16096, *T. brucei* P07752, *Arabidopsis* cytosol P22197, maize cytosol P08440, rice cytosol P17784, spinach cytosol P29356, *Euglena* chloroplast X89768, *Plasmodium berghei* A45610, *Plasmodium falciparum* P14223, *Drosophila* D10446, mouse A J05517, rat A P05065, human A P04075, rabbit A P00883, rat B P00884, human B P05062, chicken B P07341, sheep B S47540, rat C P09117, human C P09972

Independent origins of chloroplast class-I FBAs in *Euglena* and chlorophytes

Although class-I aldolases exist in both eubacteria and archaebacteria (Dhar and Altekar 1986; Lebherz and Rutter 1973; London and Kline 1973; Stribling and Perham 1973), only the eubacterial class-I aldolase from *Staphylococcus carnosus* has been sequenced to-date (Witke and Götz 1993). The *S. carnosus* class-I enzyme shares only about 25–30% identical residues with eukaryotic class-I FBAs. Assuming that divergence between *Staphylococcus* and eukaryotic FBA sequences does reflect eubacterial-eukaryotic divergence, genes for chloroplast and cytosolic aldo-

```
                                                                                                                                   100
Euglena cytosol                      <<PDFPKDLKGVLDGNQVRTLFDFAQK··KGFAIPAVNCTSSSTVNVVLERARDTHNPVIIQVSQGGAAFYCGKGVKDE·KLIASVDGSVALAHHVRAVAHTMA···P
S. pombe                            MGILDIV··PTGVIAGDNVLKLFTYARE··HGFAIPAINVTSSSTAIAALEAAREARSPIILQTSNGGAHFFAGKESSNE·GQKASIAGSIAAAHYIRSIAPFFG··VP
S. cereviseae                        GVEQILKRKTGVIVGEDVHNLFTYAKE··HKFAIPAINVTSSSTAVAALEAARDSKSPIILQTSNGGAAYFAGKGISNE·GQNASIKGAIAAAHYIRSIAPAYG··IP
Escherichia                         SKIFDFVK··PGVITGDDVQKVFQVAKE··NNFALPAVNCVGTDSINAVLETAAKVKAPVIVQFSNGGASFIAGKGVKSDVPQGAAILGAISGAHHVHQMAEHYG··VP
Haemophilus                        MAKLLDIVK··PGVVTGEDVQKVFAYAKE··HNFAIPAVNCVGSDSVNAVLETAARVKAPVIIQFSNGGAAFYAGKGIKPTSGTRPDVLGAIAGAKQVHTLAKEYG··VP
Campylobacter                       MGVLDIVK··AGVISGDELNKIYDYAKA··EGFAIPAVNVVGTDSINAVLEAAKKVNSPVIIQFSNGGAKFYAGKNC·····PNGEVLGAISGAKHVHLLAKAYG··VP
Corynebacterium                                PIATPEVYNEMLDRAKE··GGFAFPAINCTSSETINAALKGFAEAESDGIIQFSTGGAEFGSGLAV······KNKVKGAVALAAFAHEAAKSYG··IN
Rhodobacter 1                                    MALITLRQLLDHAAERLQGYGVPAFNINNMEQGLAILAAARACDAPVI··ASRGARSYAGDIM················LRHIVEALAEMYP·QIP
Rhodobacter 2                                    MALITLRQLLDHAAE··QGYGVPAFNINNMEQGLAIMEAARACDAPVIIQASRGARSYANDIM················LAKMIEALAAIYP·EIP
Alcaligenes 1                                    MALISLRQLLDHAGE··FGYGVPAFNVNNLEQIHAIMEAAAEETDSPVILQASAGARKYAGEAY················LRHMVLAAAETHP·DIP
Xanthobacter                                     MALVSMRQLLDHAAD··DSYGLPAFNVNNMEQVKAIMDAARATSSPVILQGSAGARKYAGEPF················LRHLIAAAVEAYP·EIP
Synechocystis                                    MALVPMRLLLDHAAE··NGYGIPAFNVNNMEQIISIMQAADETDSPVILQASRGARSYAGENF················LRHLVLGAVETYP·HIP
E. coli AgaY    (N-AcGal)                        MSIISTKYLLQDAQA··NGYAVPAFNIHNAETIQAILEVCSEMRSPVILAGTPGTFKHIALEE················IYALCSAYSTTY··NMP
E. coli GatY   (Galactitol)                    MKMYVVSTKQMLNNAQR··GGYAVPAFNIHNLETMQVVVETAANLHAPVIIAGTPGTFTHAGTEN················LLALVSAMAKQY··HHP
B. subtilis OrfY (spoOF locus)                   MPLVSMTEMLNTAKE··KGYAVGQFNLNNLEFTQAILQAAEEEKSPVILGVSEGAGRYMGGFKT···············VVAMVKALMEEYKVTVP
B. subtilis B65C region                          MAFVSMKELLEDAKR··EQYAIGQFNINGLQWTKAILQAAQKEQSPVIAAASDRLVDYLGGFKT···············IAAMVGALIEDMAITVP
Mycoplasma                                        MLVNFKLMLQKAKL··GKYAIPHININNYEWAKAVLTAANQANSPIIVSVSEGALKYMSGYSV···············VIPLVKGLIESLSVKVP
                                                                       *                      *

                                                                                                                         200
Euglena cytosol                    VVVHSDHCAKKLL·PWFDGMLDADGEIFCEHGVPLFSSHMLDLS·······EENDEEDIGTCVKYFTRMAKLNLWLEMEIGMTGGVEDGVDNSG········VANDKLYT
S. pombe                           VVMHSDHCAKKLL·PWMDGMFEADEAYFKIHGEPLFSSHMLDLS·······EEPKKENIAQVKEYCKRAVPMKIWIEMEIGITGGEEDGVDNSH········VSHTELYT
S. cereviseae                      VVLHSDHCAKKLL·PWFDGMLEADEAYFKEHGEPLFSSHMLDLS·······EETDEENISTCVKYFKRMAAMDQWLEMEIGITGGEEDGVNNEN········ADKEDLYT
Escherichia                        VILHTDHCAKKLL·PWIDGLLDAGEKHFAATGKPLFSSHMIDLS·······EESLQENIEICSKYLERMSKIGMTLEIELGCTGGEEDGVDNSH········MDASALYT
Haemophilus                        VILHTDHAAKKLL·PWIDGLLDAGEKHFAETGRPLFSSHMIDLS·······EESMEENMAICREYLARMDKMGMTLEIEIGITGGEEDGVDNSD········VDESRLYT
Campylobacter                      VILHTDHAARKLL·PWIDGLIEANAQYKKTHGQALFSSHMLDLS·······EESLEENLSTCEVYLQKLDALGVALEIELGCTGGEEDGVDNTG········IDNSKLYT
Corynebacterium                    VALHTDHCQKEVLDEYVRPLLAISQERVDRGELPLFQSHMWDGS·······AVPIDENLEIAQELLAKAKAANIILEVEIGVVGGEEDGVEAKA········GAN··LYT
Rhodobacter 1                      ICLHQDH··············GNNEATCLSAIRHGFTSVMMDGSLQADMKTVASYDYNVDITRRVTDAAHWVGASVEGELGVLGSLEKGEAEAEDGSGAEGKLDHSQMLT
Rhodobacter 2                      LCMHQDH··············GNNEATCMTAIRHGFTSVMMDGSLKADAKTPADYDYNVDITARVSHMAHWVGASVEGELGVLGSLETGESEAEDGHGAEGKLDHSQLLT
Alcaligenes 1                      IVLHQDH··············GSSPAVCQASIRSGFTSVMMDGSLREDMKTPSDYDYNVDVTRRVCEMAHAVGVSVEGELGCLGSLETGQAGEEDGVGAAGTLSHDMMLT
Xanthobacter                       VVMHQDH··············GASPAVCMGAIKSGFSSVMMDGSLKEDGKTPADYDYNVSVTAKVVELAHAVGVSVEGELGCLGSLETGKGEAEDGHGAEEALDHSKLLT
Synechocystis                      IAMHQDH··············GNSPATCYSAIRNGFTSVMMDGSLEADAKTPASFEYNVNVTAEVVKVAHSVGASVEGELGCLGSLETGQGEAEDGHGFEGKLDHSQLLT
E. coli AgaY                       LALHLDH··············HESLDDIRRKVHAGVRSAMIDGS·······HFPFAENVKLVKSVVDFCHSQDCSVEAELGRLGGVEDDMSVDAESA··········FLT
E. coli GatY                       LAIHLDH··············HTKFDDIAQNLRSGVRSVMIDAS·······HLPFAQNISRVKEVVDFCHRFDVSVEAELGQLGGQEDDVQVNEADA··········FYT
B. subtilis OrfY                   VAIHLDH··············GSSFESCAKAIHAGFTSVMIDAS·······HHPFEENVATTAKVVELAHFHGVSVEAELGTVGGQEDD··VIAEGV··········IYA
B. subtilis B65C                   VVLHLDH··············GSSAERCRQAIDAGFSSVMIDGS·······HQPIDENIAMTKEVTDYAAKHGVSVEAEVGTVGGMEDG··LVG·GV··········RYA
Mycoplasma                         VTLHLDH··············G·SYDACIQALQAGFSSVMFDGS·······HLPFEENFNKSKKLIEIAQKTNASVELEVGTIGGEEDG··VIGQG···········ELA
                                      * **                              * * * *             *                  *   *     *

                                                                                                               300
Euglena cytosol                    SSEQVFAVHKALGAS·SPNFSIAAAFGNVHGVYKPGNVKLQPNLLKEHQDYARKQLSSSEDH·PLYLWFHGPS·····················GSTDAEIHEAVRNGVV
S. pombe                           QPEDIWDVYRELSSV·TPYFSIAAAFGNVHGVYKPGNVKLQPALLGQHQAYVKEQLKTTNDK·PVFFVFHGGS·····················GSSVNEFRTGIKCGVV
S. cereviseae                      KPEQVYNVYKALHPI·SPNFSIAAAFGNCHGLY·AGDIALRPEILAEHQKYTREQVGCKEEK·PLFLVFHGGS·····················GSTVQEFHTGIDNGVV
Escherichia                        QPEDVDYAYTELSKI·SPRFTIAASFGNVHGVYKPGNVVLTPTILRDSQEYVSKKHNL·PHN·SLNFVFHGGS·····················GSTAQEIKDSVSYGVV
Haemophilus                        QPSDVLYVYDQLHPV·SPNFTVAAAFGNVHGVYKPGNVKLKPSILGESQEFVSKERNL·PAK·PINFVFHGGS·····················GSSREEIREAIGYGAI
Campylobacter                      QPEDVALAYERLGKI·SDKFSIAASFGNVHGVYKPGNVSLQPEILKNSQKFVKDKFALNSDK·PINFVFHGGS·····················GSELKDIKNAVSYGVI
Corynebacterium                    SPEDFEKTIDAIGTGEKGRYLLAATFGNVHGVYKPGNVKLRPEVLLEGQQVARKKLGLADDALPFDFVFHGGS·····················GSEKEKIEEALTYGVI
Rhodobacter 1                      DPDQAVEFVQATRVD·····ALAIAMGTSHGAYKFSRKPDGEILAMRVIEEIHARLP······ATHLVMHGSSSVAARLQDLINAHGADMPQTYGVPVEEIERGIRHGVR
Rhodobacter 2                      DPDQAVDFVKKTQVD·····ALAIACGTSHGAYKFSRKPDGEILAMSVIEAIHKKLP······DTHLVMHGSSSVPQELQDIINAFGGAMPQTFGVPVEEIVRGIKMGVR
Alcaligenes 1                      DPAQARDFVARTGVD·····ALAIAIGTSHGAYKFSRKPTGDILAIDRIREIHEQIP······DTHLVMHGSSSVPQEWLEIIRQYGGDIKETYGVPVEEILRGIKTGVR
Xanthobacter                       DPDEAAQFVKATQCD·····ALAIAIGTSHGAYKFTRKPTGDILAIDRIKAIHQRIP······TTHLVMHGSSSVPQELLEEIRTYGGDIKETYGVPVEEIQEGIRYGVR
Synechocystis                      DPEEAVEFVNKTQVD·····ALAVAIGTSHGAYKFTRKPTGEVLAISRIEEIHRLLP······NTHLVMHGSSSVPQEWIDMINEFGGAIPETYGVPVEEIQKGIKSGVR
E. coli AgaY                       DPQEAKRFVELTGVD·····SLAVAIGTAHGLY····SKT·PKIDFQRLAEI·REVV······DVPLVLHGAS·····················DVPDEFVRRTIELGVT
E. coli GatY                       NPAQAREFAEATGID·····SLAVAIGTAHGMY····ASA·PVLDFSRLENI·RQWV······NLPLVLHGAS·····················GLSTKDIQQTIKLGIC
B. subtilis OrfY                   DPKECQELVERTGID·····CLAPALGSVHGPY····KGE·PNLGFKEMEEI·GKST······GLPLVLHGGT·····················GIPTADIKKSISLGTA
B. subtilis B65C                   DITECERIVKETNID·····ALAAALGSVHGKY····QGE·PNLGFKEMEAI·SRMT······DIPLVLHGAS·····················GIPQDQIKKAITLGHA
Mycoplasma                         NVDECKQIASLKP·D·····ALAAGIGNIHGIY····PKNWKGLNFPLIETI·SKIT······NLPLVLHGGS·····················GILENDVKKAISLGIC
                                                         *   *  ** *                                  * **

                                                                                                     400                                 436
Euglena cytosol                    KMNLDTDMQWAYWDGLRQFEAKKHDYLQGQIGNPEGPDKPNKNYYDPRKWIREAELGMLARVKVAFKAVELPGGLKEFIGIP························
S. pombe                           KVNIDTDTQFAYVEGVRDYVLKYKDYLMTPVGNPEGADKPNKKKFDPRVWIHEGEKTMTKRVLTALEDFYTVNTL······························
S. cereviseae                      KVNLDTDCQYAYLTGIRDYVLNKKDYIMSPVGNPEGPEKPNKKFFDPRVWVREGEKTMGAKITKSLETFRTTNTL······························
Escherichia                        KMNIDTDTQWATWEGVLNYYKANEAYLQGQLGNPKGEDQPNKKYYDPRVWLRAGQTSMIARLEKAFQELNAIDVL······························
Haemophilus                        KMNIDTDTQWASWNGILNFYKANEAYLQGQLGNPEGPDAPNKKYYDPRVWLRKMEESMSKRLEQSFEDLNCVDVL······························
Campylobacter                      KMNIDTDTQWAFWDGVREYELKNRAYLQGQIGNPEGDDKPNKKYYDPRVWLRSGEESMIKRLEIAFEDLNCINKN······························
Corynebacterium                    KMNVDTDTHYAFTRPIVSHMFENYNGVL·····KIDGEVGNKKAYDPRSYMKKAEQSMSERIIESCQDLKSVGKTTSK···························
Rhodobacter 1                      KVNIDTDCRMALTGQFRKVAMESPKE··················FDARKFMIPAMKEMEALVRDRFERFGTAGNASKITVIPMDDMAKRYASGALDPAVATAKAA
Rhodobacter 2                      KVNIDTDCRMAMTGQFRRIAQQTPSE··················FDPRKFLKPAMDAMRDLCKQRLEAFGTAGQAGKIRIIPMDDMAKRYASGALAPKTA·····
Alcaligenes 1                      KVNIDTDIRLAMIGAIRKSLAEDRSE··················FDPRKALLAAKKGARSVVKLRFEAFGCAGQASKIKPIAMEQLAQWYR···············
Xanthobacter                       KVNIDTDIRLAMTAAIRRVGAKNKSE··················FDPRKFMAAAMEEAKKVCIARFEAFGSAGKAEKIRAIELDEMAKRYASGELAQVVH·····
Synechocystis                      KVNIDTDNRLAITAAFREAAAKDPKN··················FDPRHFLKPSIKYMKQVCADRYQQFWTAGNASKIKQLTLDDYAAKYAKGELTATSRTSVAV·
E. coli AgaY                       KVNVATELKIAFAGAVKAWFAENPQG··················NDPRYYMRVGMDAMKEVVRNKINVCGSANRISA····························
E. coli GatY                       KINVATELKNAFSQALKNYLTAHPEA··················TDPRDYLQSAKSAMRDVVSKVIADCGCEGRA······························
B. subtilis OrfY                   KINVNTENQISSAKAVRETLAAKPDE··················YDPRKYLGPAREAIKETVIGKMREFGSSNQA······························
B. subtilis B65C                   KININTECMVAWTDETRRMFQENSDL··················YEPRGYLTPGIEAVEETVRSKMREFGSAGKAAKQQVG························
Mycoplasma                         KLNINTECQLAFAHEIRKYIESNKDLDLNKKG············YDPRKLLKEPTQAIVDTCLEKIDLCGSRNKA·······························
                                   * *                                            *
```

Fig. 8 Alignment of class-II FBA proteins. Gaps are indicated as *dots*, strictly conserved residues in the alignment are indicated with an *asterisk*. Sequences were extracted from GenBank, Swissprot and PIR data bases. Accession numbers to sequences are *Campylobacter* S52413, *Corynebacterium* P19537, *Escherichia coli fba* P11604, *Euglena* X89769, *Haemophilus* P44429, *S. pombe* P36580, yeast P14540, *B. subtilis* OrfY P13243, *B. subtilis* B65C P42420, *E. coli* AgaY P42908, *E. coli* GatY P37192, *Mycoplasma* L43967, *Alcaligenes* (plasmid) U12423, *Rhodobacter* 1 P27995, *Rhodobacter* 2 P29271, *Synechocystis* D64000, *Xanthobacter* U29134. "≪" indicates that the start codon is not contained in the *Euglena* cDNA clone

lase isoenzymes of higher plants presumably arose through gene duplication very early in eukaryotic evolution. This is in contrast to TPI, where the higher-plant chloroplast/cytosol isoenzymes arose through gene duplication of the cytosolic enzyme relatively late in early plant evolution (Henze et al. 1994; Schmidt et al. 1995), and is also in contrast to GAPDH where the chloroplast/cytosol isoforms are related by duplication in eubacterial genomes (Martin et al. 1993). A number of later gene duplications (Fig. 7) have occurred in vertebrates (A, B, and C isoforms) and in higher plants (Kukita et al. 1988; Razdan et al. 1993).

Despite some unclarified problems, one important conclusion can be drawn. Chloroplast-localized forms of FBA arose twice during evolution: one in the lineage of chlorophytes and another in the lineage of *Euglena* because of their extensive separation in the gene tree (Fig. 7). And, notably, *neither* of these chloroplast enzymes is likely to be of cyanobacterial origin, because all cyanobacteria analyzed to-date were found to possess class-II FBAs (Rutter 1964; Antia 1967; Schnarrenberger et al. 1992).

Class-II FBA: a eubacterial gene family

Class-II aldolases have long been known to be distributed in *Euglena*, numerous fungi, numerous eubacteria and ar-

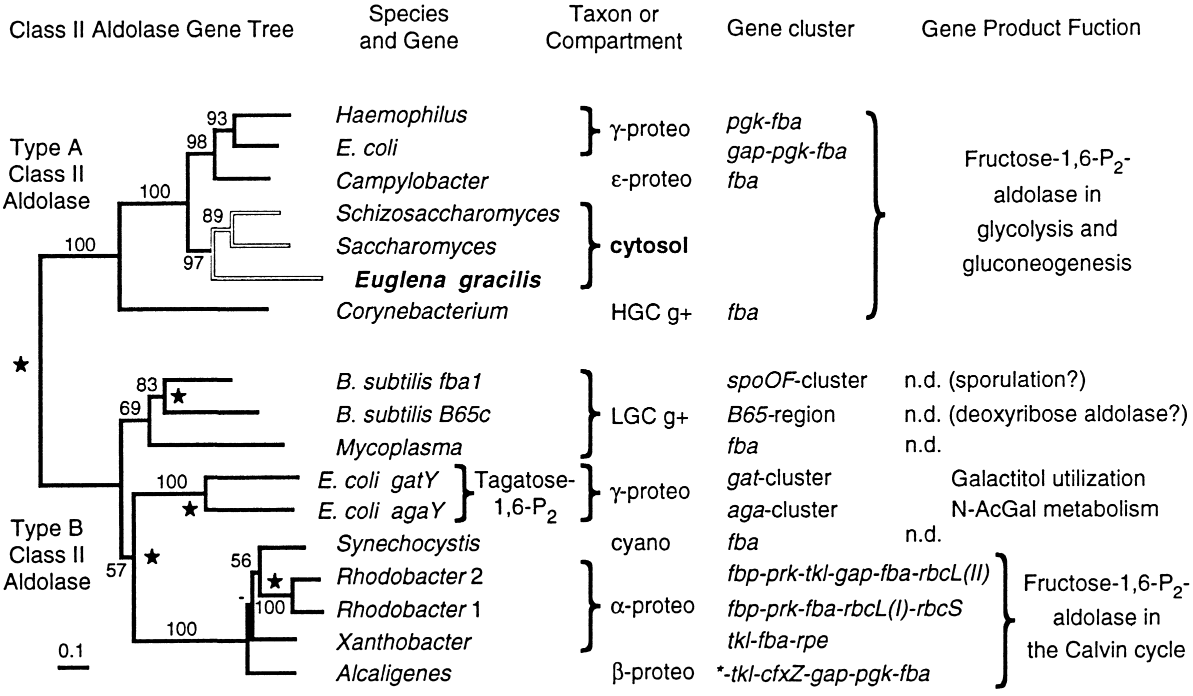

Fig. 9 A class-II fructose-1,6-bisphosphate aldolase gene tree. The tree was constructed by the neighbor-joining method for the matrix of numbers of amino-acid substitutions per site estimated with the Dayhoff matrix option of PROTDIST in PHYLIP. *Numbers at branches* indicate the bootstrap proportion for 100 replicates using the same distance estimation method. *Stars at nodes* indicate the presence of possible gene duplications. The scale bar indicates 0.1 substitutions per site. Gene-cluster information and the function of the encoded product where available is indicated. The accession numbers to sequences are given in the legend to Fig. 8. The plasmid and chromosomal copies of *fba* from *Alcaligenes* are nearly identical in sequence, but only the plasmid sequence is shown here. *N.d.* not determined. An *asterisk* in the *Alcaligenes* operon structure indicates the presence of six additional genes for enzymes of the Calvin cycle which are not shown here. The sequences encoded in eukaryotic genomes are indicated in *open branches*

chaebacteria (Dhar and Altekar 1986; Schnarrenberger et al. 1992). In order to examine the evolutionary history of class-II FBAs, we aligned sequences for eubacterial and eukaryotic class-II FBAs and related enzymes (Fig. 8) and constructed a tree of their gene evolution.

Several gene duplications for class-II FBAs occurred during eubacterial evolution (indicated with stars in the Fig. 9). One of them involved the common ancestor of gram-positive and proteobacteria and led to the separation of two families of class-II FBA enzymes, designated here as type "A" and type "B" class-II FBAs (Fig. 9). Amino-acid sequence identity between the type A and type B class-II FBA enymes is of the order of 25–30% in individual comparisons, whereas within type A and type B comparisons it is of the order of 40% (data not shown).

The type A class-II FBA enyzmes encompass those that have been characterized to-date from the three eukaryotes studied, *Corynebacterium*, and γ- and ε-proteobacteria. For the type A class-II FBA enzymes surveyed here, the function of the encoded product is known to be fructose-1,6-bisphosphate aldolase (Alefounder and Perham 1989; Schwelberger et al. 1989; Mutoh and Hayashi 1994; Cenatiempo and Fauchere 1995). By contrast, only those type B class-II FBA genes that are found in the Calvin-cycle operons of photautotrophic proteobacteria are known to encode fructose-1,6-bisphosphate aldolase (Tabita et al. 1993). Several of the other eubacterial type B class-II FBA-related proteins have other functions, e.g. tagatose-1,6-bisphosphate-accepting aldolases (Lengler 1977; Noblemann and Lengler 1995; Reizer et al. 1996) or deoxyribose aldolase activity (Yoshida et al. 1994). Because the specific fructose-1,6-bisphosphate aldolase function is found among both the type A and type B enzymes, it seems that fructose-1,6-bisphosphate aldolase activity was the original function of the type A and type B enzymes.

This gene phylogeny indicates that class-II FBA enzymes existed as a eubacterial gene family very early in evolution and underwent a complex series of recurrent duplication events, similar to those found in Rubisco (Martin et al. 1992) and GAPDH (Henze et al. 1995) gene evolution. Furthermore, differential loss has obviously also occurred in eubacterial class-II FBA gene evolution, since the complete *Haemophilus* genome does not encode a type B class-II FBA and the complete *Mycoplasma* genome does not encode type A.

Eukaryotic class-II FBA genes: endosymbiotic rather than "horizontal" transfer

For class-II FBA, it was contended that the possession of a gene for class-II aldolase in *S. cereviseae* represents a "likely" example of "horizontal transfer of a type II aldolase gene from some eubacterium to yeast" (Smith et al. 1992). The evolutionary tree for these enzymes reveals that no inter-kingdom horizontal transfer event of the type envisaged, i.e. from bacteria to yeast, has occurred. Rather, the topolgy of Fig. 9 very clearly indicates that the common ancestor of *Euglena* and the two ascomycetes simply possessed a gene for a class-II FBA.

The position of the eukaryotic class-II FBA genes as a branch in the eubacteria is highly reminiscent of the situation found for eukaryotic GAPDH (Henze et al. 1995), eukaryotic fructose-1,6-bisphosphatase genes (Martin et al. 1996) and plant PGK genes (Brinkmann and Martin 1996), and the gene for class-II FBA might have been transferred from eubacteria to eukaryotes by an endosymbiotic event. The position of the three eukaryotic sequences in the eubacterial tree furthermore suggests that these nuclear genes were transferred to the nucleus during the process of mitochondrial (rather than plastid) origins. Finally, the fact that *Euglena* possesses a cytosolic class-II FBA, whereas the "homolog" from kinetoplastids is a class-I FBA enzyme, suggests that their common ancestor possessed both class-I and class-II FBAs, and that differential loss of the genes for these enzymes has occurred during the evolution of these two eukaryotes.

Chloroplast and cytosolic fructose-1,6-bisphosphate aldolases of *Euglena* are the first pair of chloroplast-cytosol isoenzymes that share no evolutionary relationship at all. This is in sharp contrast to chloroplast-cytosol isoenzymes of sugar-phosphate metabolism of higher plants, which are known to be related by gene duplications (Schnarrenberger and Martin 1997). The class-II FBA enzymes of *Euglena* and the two ascomycetes studied here appear to have been subject to endosymbiotic gene transfer from the antecedants of mitochondria during the course of endosymbiosis. The contemporary chloroplast-localized class-I FBA enzymes of higher plants and *Euglena* arose independently in evolution, but the ultimate origin of *Euglena*'s nuclear gene for chloroplast class-I FBA could not be clarified. However it seems most likely to have been recruited via duplication of the gene for the cytosolic FBA of the chlorophyte endosymbiont. But still a third independent origin of chloroplast-localized FBA must have occurred during evolution: the cyanelles (plastids) of *Cyanophora paradoxa* possess a class-II FBA (Gross et al. 1994), as do cyanobacteria.

Acknowledgements We gratefully acknowledge financial support from the Deutsche Forschungsgemeinschaft. Sequences in this paper have been submitted to GenBank under the accession numbers X89768 (class-I aldolase) and X89769 (class-II aldolase).

References

Alefounder PR, Perham RN (1989) Identification, molecular cloning and sequence analysis of a gene cluster encoding class-II fructose-1,6-bisphosphate aldolase, 3-phosphoglycerate kinase and a putative second glyceraldehyde-3-phosphate dehydrogenase of *Escherichia coli*. Mol Microbiol 3:723–732

Anderson LE, Advani VR (1970) Chloroplast and cytoplasmic enzymes. Three distinct isoenzymes associated with the reductive pentose-phophate cycle. Plant Physiol 45:583–585

Antia NJ (1967) Comparative studies on aldolase activity in marine planktonic algae and their evolutionary significance. J Phycol 3:81–84

Böger P, San Pietro A (1967) Ferredoxin and cytochrome f in *Euglena gracilis*. Z Pflanzenphysiol 58:70–75

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

Cenatiempo Y, Fauchere JL (1995) Nucleotide sequence and characterization of *peb4A* encoding an antigenic protein in *Campylobacter jejuni*. Res Microbiol 146:467–476

Chan RL, Keller M, Canaday J, Weil J-H, Imbault P (1990) Eight small subunits of *Euglena* ribulose-1,5-bisphosphate carboxylase/oxgygenase are translated from a large mRNA as a polyprotein. EMBO J 9:333–338

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence-analysis programs for the vax. Nucleic Acids Res 12:387–395

Dhar NM, Altekar W (1986) Distribution of class-I and class-II fructose-bisphosphate aldolases in halophilic archaebacteria. FEMS Microbiol Lett 35:177–181

Felsenstein J (1993) Phylip manual, version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Gibbs SH (1978) The chloroplast of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56:2883–2889

Gross W, Bayer MG, Schnarrenberger C, Gebhart UB, Maier TL, Schenk HEA (1994) Two distinct aldolases of the class-II type in the cyanoplasts and in the cytosol of the alga *Cyanophora paradoxa*. Plant Physiol 105:1393–1398

Henze K, Schnarrenberger C, Kellermann J, Martin W (1994) Chloroplast and cytosolic triosephosphate isomerases from spinach: purification, microsequencing and cDNA cloning of the chloroplast enzyme. Plant Mol Biol 26:1961–1973

Henze K, Badr A, Cerff R, Wettern M, Martin W (1995) A nuclear gene of eubacterial origin in *Euglena* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA 92: 9122–9126

Houlné G, Schantz R (1987) Molecular analysis of the transcripts encoding the light-harvesting chlorophyll a/b protein in *Euglena gracilis*: unusual size of the mRNA. Curr Genet 12:611–616

Houlné G, Schantz R (1988) Characterization of cDNA sequences for LHCI apoprotein in *Euglena gracilis*: the mRNA encodes a large precursor containing several consecutive polypeptides. Mol Gen Genet 213:479–486

Houlné G, Schantz R (1993) Expression of polyproteins in *Euglena*. Crit Rev Plant Sci 12:1–17

Ikawa T, Asami S, Nisizawa K (1972) Comparative studies on fructose-diphosphate aldolases mainly in marine algae. Proc Int Seaweed Symp 7:526–531

Jacobshagen S, Schnarrenberger C (1988) Two class-I aldolases in the green alga *Chara foetida* (Charophyceae). Plant Physiol 87:78–82

Jacobshagen S, Schnarrenberger C (1990) Two class-I aldolases in *Klebsormidium flaccidum* (Charophyceae): an evolutionary link from chlorophytes to higher plants. J Phycol 26:312–317

Kishore R, Muchhal US, Schwartzbach ST (1993) The pre-sequence of *Euglena* LHCPII, a cytoplasmically synthesized chloroplast protein, contains a functional endoplasmic reticulum-targeting domain. Proc Natl Acad Sci USA 90:11845–11849

Krüger I, Schnarrenberger C (1983) Purification, subunit structure, and immunological comparison of fructose-bisphosphate aldolases from spinach and corn leaves. Eur J Biochem 136:101–106

Kukita A, Mukai T, Miyata T, Hori K (1988) The structure of brain-specific rat aldolase C mRNA and the evolution of aldolase isozyme genes. Eur J Biochem 171:471–478

Lebherz HG, Rutter WJ (1969) Distribution of fructose-diphosphate aldolase variants in biological systems. Biochemistry 8:109–121

Lebherz HG, Rutter WJ (1973) A class-I (Schiff-base) fructose-diphosphate aldolase of prokaryotic origin. J Biol Chem 24: 1650–1659

Lebherz HG, Leadbetter MM, Bradshaw RA (1984) Isolation and characterization of the cytosolic and chloroplast form of spinach leaf fructose-diphosphate aldolase. J Biol Chem 259:1011–1017

Lefort-Tran M (1981) The triple layered organization of the *Euglena* chloroplast envelope (significance and function). Ber Deutsch Bot Ges 94:463–476

Lengler J (1977) Analysis of mutations affecting the dissimilation of galactitol (dulcitol) in *Escherichia coli* K12. Mol Gen Genet 152:83–91

Levasseur PJ, Meng Q, Bouck B (1994) Tubulin genes in the algal protist *Euglena gracilis*. J Euk Microbiol 41:468–477

London J, Kline K (1973) Aldolase of lactic-acid bacteria: a case history in the use of an enzyme as an evolutionary marker. Bacteriol Rev 37:453–478

Marchand M, Poliszczak A, Gibson WC, Wierenga RK, Opperdoes FR, Michels PA (1988) Characterization of the genes for fructose-bisphosphate aldolase in *Trypanosoma brucei*. Mol Biochem Parasitol 29:65–75

Martin W, Somerville CC, Loiseaux-de Goër S (1992) Molecular phylogenies of plastid origins and algal evolution. J Mol Evol 35:385–403

Martin W, Brinkmann H, Savona C, Cerff R (1993) Evidence for a chimaeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 90:8692–8696

Martin W, Mustafa A-Z, Henze K, Schnarrenberger C (1996) Higher plant chloroplast and cytosolic fructose-1,6-bisphophosphatase isoenzymes: origins via duplication rather than prokaryote-eukaryote divergence. Plant Mol Biol 32:485–491

Mo Y, Harris BG, Gracy RW (1973) Triosephosphate isomerase and aldolases from light- and dark-grown *Euglena gracilis*. Arch Biochem Biophys 157:580–587

Muchhal US, Schwartzbach SD (1992) Characterisation of a *Euglena* gene encoding a polyprotein precursor to the light-harvesting chlorophyll *a/b*-binding protein of photosystem II. Plant Mol Biol 18:287–299

Mutoh N, Hayashi Y (1994) Molecular cloning and nucleotide sequencing of a *Schizosaccharomyces pombe* homologue of the class-II fructose-1,6-bisphosphate aldolase gene. Biochim Biophys Acta 1183:550–552

Noblemann B, Lengler J (1995) Sequence of the *gat* operon for galactitol utilization from a wild-type strain EC3132 of *Escherichia coli*. Biochim Biophys Acta 1262:69–72

Pelzer-Reith B, Penger A, Schnarrenberger C (1993) Plant aldolase: cDNA and deduced amino-acid sequences of the chloroplast and cytosol enzyme from spinach. Plant Mol Biol 21:331–340

Pelzer-Reith B, Wiegand S, Schnarrenberger C (1994) Plastid class-I and cytosol class-II aldolase of *Euglena gracilis*. Plant Physiol 106:1137–1144

Razdan KK, Heinrikson RL, Zurcher-Neely HA, Morris P, Anderson LE (1993) Chloroplast and cytoplasmic enzymes: isolation and sequencing of cDNAs coding for two distinct pea chloroplast aldolases. Arch Biochem Biophys 298:192–197

Reizer J, Ramseier TM, Reizer A, Charbit A, Saier MH (1996) Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative *N*-acetylgalactosamine metabolic pathway in *Escherichia coli*. Microbiology 142:231–250

Rutter WJ (1964) Evolution of aldolase. Fed Proc 23:1248–1257

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425

Schmidt M, Svendsen I, Feierabend J (1995) Analysis of the primary structure of the chloroplast isozyme of triosephosphate isomerase from rye-leaves by protein and cDNA sequencing indicates a eukaryotic origin of its gene. Biochim Biophys Acta 1261: 257–264

Schnarrenberger C, Martin W (1997) The Calvin cycle: a historical perspective. Photosynthica 33:331–345

Schnarrenberger C, Gross W, Pelzer-Reith B, Wiegand S (1992) The evolution of isoenzymes of sugar-phosphate metabolism in algae. In: Stabenau H, Tolbert NE (eds) Phylogenetic changes in peroxisomes in algae. Phylogeny of plant peroxisomes. University of Oldenburg, Oldenburg, pp 310–329

Schnarrenberger C, Pelzer-Reith B, Yatsuki H, Freund S, Jacobshagen S, Hori K (1994) Expression and sequence of the only detectable aldolase in *Chlamydomonas reinhardtii*. Arch Biochem Biophys 313:173–178

Schwelberger HG, Kohlwein SD, Paltauf F (1989) Molecular cloning, primary structure and disruption of the structural gene of aldolase from *Saccharomyces cerevisiae*. Eur J Biochem 180: 301–308

Sharif AL, Smith AG, Abell C (1989) Isolation and characterisation of a cDNA for a chlorophyll synthesis enzyme from *Euglena gracilis*. Eur J Biochem 184:353–359

Smith MW, Feng D-F, Doolittle RF (1992) Evolution by acquisition: the case for horizontal gene transfers. Trends Biochem Sci 17: 489–493

Sogin M, Gunderson J, Elwood H, Alonso R, Peattie D (1989) Phylogenetic meaning of the kingdom concept: an unusal ribosomal RNA from *Giardia lamblia*. Science 243:75–77

Stribling D, Perham RN (1973) Purification and characterization of two fructose-diphosphate aldolases from *Escherichia coli* (Crookes' strain). Biochem J 131:833–841

Surek B, Melkonian M (1986) A cryptic cytostome is present in *Euglena*. Protoplasma 133:39–49

Tabita FR, Gibson JL, Falcone DL, Wang X, Li L-A, Read BA, Terlesky KC, Paoli GC (1993) Current studies on the molecular biology and biochemistry of CO_2 fixation in phototrophic bacteria. In: Murrel JC, Kelley DP (eds) Microbial growth on C1 compounds. Intercept Scientific, Andover, UK, pp 469–479

Tessier LH, Paulus F, Keller M, Vial C, Imbault P (1995) Structure and expression of *Euglena gracilis* nuclear rbcS genes encoding the small subunit of the ribulose 1,5-bisphosphate carboxylase/oxygenase: a novel splicing process for unusal intervening sequences. J Mol Biol 245:22–33

Thompson JD, Higgins DG, Gibson TJ (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

Von Heijne G, Stepphuhn J, Herrmann RG (1989) Domain structure of mitochondrial- and chloroplast-targeting peptides. Eur J Biochem 180:535–545

Walne PL, Kivic PA (1989) Phylum Euglenida. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds) Handbook of Protoctista. Jones and Bartlett, Boston, pp 270–287

Witke C, Götz F (1993) Cloning, sequencing, and characterization of the gene encoding the class-I fructose-1,6-bisphosphate aldolase of *Staphylococcus carnosus*. J Bacteriol 175:7495–7499

Yoshida K, Sano H, Miwa Y, Ogasawara N, Fujita Y (1994) Cloning and nucleotide sequencing of a 15-kb region of the *Bacillus subtilis* genome containing the *iol* operon. Microbiology 140: 2289–2298