*J. Eukaryot. Microbiol.*, 55(3), 2008 pp. 170–177

DOI: 10.1111/j.1550-7408.2008.00324.x

# Triosephosphate Isomerase Genes in Two Trophic Modes of Euglenoids (Euglenophyceae) and Their Phylogenetic Analysis

**GUI-LING SUN,[a,b,1] WANG SHEN[a,b,1] and JIAN-FAN WEN[a]**

[a]*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China, and*

[b]*Graduate School of the Chinese Academy of Sciences, Beijing 100039, China*

ABSTRACT. Two different length cDNAs encoding triosephosphate isomerase (TIM) were identified in the two trophic modes of euglenoids, the phototrophic *Euglena gracilis* and *Euglena intermedia* and the saprotrophic *Astasia longa*. Sequence analyses and presequence prediction indicated that the shorter cDNA encodes a cytosolic TIM and the longer cDNA encodes a plastid TIM (pTIM). The typical presequence of the putative *A. longa* pTIM and the high sequence similarity between *A. longa* pTIM and *E. gracilis* pTIM imply that *A. longa* pTIM is targeted to plastids. Therefore, although the plastids of *A. longa* have lost the ability of photosynthesis, they might retain other TIM-related function(s), such as glycolysis and the synthesis of isopentenyl diphosphate or fatty acids. Including the TIM sequences obtained by us from chlorophytes and rhodophytes, our phylogenetic analyses indicated that euglenoid TIMs group neither with TIMs of kinetoplastids, which share the nearest common ancestor with euglenoids, nor are closely related to TIMs of chlorophytes, which are considered to be the donors of euglenoid plastids through secondary endosymbiosis. Instead, they group with TIMs of rhodophytes. In addition, our amino acid sequence alignment and structure modeling showed that TIMs of euglenoids and rhodophytes share a unique 2-aa insertion within their loop-4 areas. Therefore, either *tim* convergent evolution or lateral gene transfer (more probably) might have occurred between euglenoids and rhodophytes after the divergence of euglenoids with kinetoplastids.

**Key Words.** *Astasia longa*, chlorophytes, *Euglena gracilis*, kinetoplastids, phylogeny, plastids, rhodophytes.

TRIOSEPHOSPHATE isomerase (TIM) is a glycolytic enzyme catalyzing the interconversion between hydroxyacetone phosphate and D-glyceraldehyde 3-phosphate. It occurs ubiquitously in prokaryotes and in the cytosol of eukaryotic cells (cTIM) (Keeling and Doolittle 1997). Besides a cTIM, photoautotrophic plant cells also contain an extra TIM isozyme localized in their photosynthetic plastids (pTIM), where it participates in the photosynthetic carbon reduction cycle (Schmidt, Svendsen, and Feierabend 1995). However, in some photoautotrophic algae, such as the unicellular chlorophyte *Chlamydomonas reinhardtii* and the rhodophyte *Cyanidioschyzon merolae*, only a single pTIM isoform or gene was detected (Klein 1986; Matsuzaki et al. 2004; Schnarrenberger et al. 1990). It is likely that cTIM genes of eukaryotic cells were descended from a homolog in eubacteria rather than from that in archaebacteria (Keeling and Doolittle 1997), and cpTIM genes in photoautotrophic plants were derived from the pre-existing cTIM genes through duplication, rather than through gene transfer from a prokaryotic endosymbiont (Schmidt et al. 1995).

Evidence from cytology and molecular genetics indicated that euglenoids share a common ancestor with kinetoplastids, a group of non-photosynthetic protists including trypanosomes and their relatives (Triemer and Farmer 1991). Biochemical, morphological, and genetic data indicated phototrophic euglenoids were most likely derived from phagotrophic euglenoids through endosymbiosis with chlorophytes (Dawson and Walene 1994; Gibbs, 1978; Hallick et al. 1993; Leander, Esson, and Breglia 2007; Müllner et al. 2001). Whereas the saprotrophic (osmotrophic) euglenoid *Astasia longa*, whose plastids no longer perform photosynthesis, was considered to be probably derived from a phototrophic euglenoid through a secondary loss of the photosynthetic function (secondary saprotrophy) (Gockel and Hachtel 2000; Marin et al. 2003; Müllner et al. 2001; Siemeister and Hachtel 1989).

Biochemical studies indicated that like photoautotrophic plants, the phototrophic euglenoid *Euglena gracilis* possesses two active TIM isozymes, a glycolytic cTIM and a Calvin-cycle pTIM (Mo, Harris and Gracy 1973), neither of which is encoded by *E. gracilis*'s plastid DNA (Hallick et al. 1993). But, so far the genes encoding the two isozymes in *E. gracilis* have not been identified. Furthermore, it is not known whether the saprotrophic euglenoid *A. longa* also contains both cTIM and pTIM; and if it does, whether the pTIM remains functional as the photosynthetic pTIM because the plastids of *A. longa* have lost their photosynthetic ability (Hallick et al. 1993). Furthermore, the study of *tim* genes might provide new evidence for the evolution of euglenoids and their evolutionary relationships with other lineages. In the present work, the cDNA or DNA sequences of TIMs in both phototrophic and secondary saprotrophic euglenoids, as well as in the chlorophytes *C. reinhardtii* and *Volvox carteri* and the rhodophytes *Porphyra yezoensis* and *Galdieria sulphuraria* are identified by experiments and bioinformatics analyses. The phylogenetic relationships among TIMs in these organisms were also investigated.

## MATERIALS AND METHODS

**Materials.** Strains of *E. gracilis* (fachb-47), *Euglena intermedia* var., and *A. longa* (fachb-921) were given generously by Zhixin Shi (Institute of Hydrobiology, Chinese Academy of Sciences [CAS], Wuhan, Hubei), and were maintained in *Polytomella* medium (Starr and Zeikus 1993). A strain of *C. reinhardtii* (fachb-479) was purchased from the Freshwater Algae Culture Collection (Institute of Hydrobiology, CAS), and was cultured in SE medium (http://algae.ihb.ac.cn). All these algae were cultured at room temperature (20 °C–25 °C) under a 13 h/11 h light/dark cycle. *Porphyra yezoensis* was collected by Dr. Jinxian Liu (Dep. Biol. Sci. and Tech., Ocean University of China) from East China Sea and stored at − 20 °C.

**Obtaining the cDNA or DNA sequences of TIMs from euglenoids, chlorophytes, and rhodophytes.** Total RNA was extracted from *E. gracilis*, *E. intermedia*, *A. longa*, and *C. reinhardtii* with the UNIQ-10 total RNA minipreps classic kit (Sangon, Shanghai, China). 3′-Rapid Amplification of cDNA Ends (RACE) and 5′-RACE were carried out using a TaKaRa RNA PCR Kit (AMV) Ver.2.1 (TaKaRa, Dalian, China) and a 5′-full RACE core set (TaKaRa), respectively, following the manufacturer's instructions. For 3′-RACE, the cDNA products were amplified by using a sense primer P3 (designed on the basis of the conserved region of reported eukaryotic TIMs) for *E. gracilis*,

[1]Equal contribution.

Corresponding Author: J.-F. Wen, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China—Telephone number/FAX number: 86 0871 5198682; e-mail: wenjf@mail.kiz.ac.cn

*E. intermedia*, and *A. longa*, and a specific sense primer Sp4 (designed according to the 5′-terminal cDNA sequence downloaded from the web site of the *C. reinhardtii* expressed sequence tag (EST) project (Asamizu et al. 2000; http://www.kazusa.or.jp/en/plant/chlamy/EST/) and the antisense adaptor primer M13 Primer M4 provided by the kit. Amplification products of expected sizes were gel isolated and sequenced after cloning into the PMD18-T cloning vector (TaKaRa).

Sequences obtained by 3′-RACE were searched against GenBank nr database using BLAST to examine whether they are really *tim* gene sequences. Then, 5′-RACE was performed as follows: first-strand cDNAs were synthesized using AMV reverse transcriptase and primer Rt; polymerase chain reaction (PCR) was performed with primer pairs Ps1/Pn1 and Ps2/Pn2; a second PCR was carried out with primer pairs Ps3/Pn1 and Ps3/Pn2 to amplify the two *tim* genes, respectively. Primer sequences were P3, 5′-GGAAACTGGAAATGCAATGG-3′; Sp4, 5′-GCGTGCCGTGGGTTATTC-3′; Rt, 5′- CACCCAGTCAAATCC-3′; Ps1, 5′-GGAGTTGCATTCCAGAACTG-3′; Ps2; 5′-GGTGCAGAACTGTTGGATTG-3′; Ps3, 5′-TGGTGCCTTCACTGGCGA-3′; Pn1, 5′-GGGTGGTGCAACAACAA-3′; and Pn2, 5′-ACCTCAACATCAGCATC-3′.

To obtain *P. yezoensis tim* sequences, genomic DNA was extracted as described before (Wattier, Prodöhl, and Maggs 2000). A primer pair RedP3 (5′-AACTGGAAGTGCAACCTG-3′)/RedPr6 (5′-CCCACCAGGAAGCCGTC-3′), designed based on the two partial cDNA sequences obtained from the *P. yezoensis* EST project (Nikaido et al. 2000; http://www.kazusa.or.jp/en/plant/porphyra/EST/), was used to amplify the unknown internal region of the *tim* gene. PCR products were cloned and sequenced as described above.

To obtain *tim* sequences for *G. sulphuraria*, the TIM sequence of *C. merolae* was used as a query to search Galdieria Genome contigs (October 13, 2005) (Michigan State University Galdieria Database http://genomics.msu.edu/galdieria/sequence_data.html) by TBLASTN, and then introns were removed by searching the EST database and GenBank nr database and according to the GT–AG rule. For *V. carteri*, *tim* sequences were assembled from the trace archive database of *V. carteri* whole genome shotgun (WGS) in NCBI (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml) by BLAST with *C. reinhardtii tim* sequence as a query. The complete sequences were assembled with the SEQMAN and EDITSEQ programs from Lasergene package (DNAstar, Madison, WI).

**Presequence identification.** The two longer-deduced TIM protein sequences of *E. gracilis* and *A. longa*, which are considered to be pTIMs containing extra presequences for transport across the three-layered plastid membrane, were analyzed by the SignalP program (Bendtsen et al. 2004; http://www.cbs.dtu.dk/services/SignalP/) to predict their signal sequences. After excluding the signal sequences, ChloroP program (Emanuelsson, Nielsen, and von Heijne 1999; http://www.cbs.dtu.dk/services/ChloroP/) was used to predict the transit sequences. Finally, the dense alignment surface (DAS) program (Cserzo et al. 1997; http://www.sbc.su.se/~miklos/DAS/) and the Transmembrane Helix (TMH) prediction based on a Hidden Markov Model (TMHMM) program (Krogh et al. 2001; http://www.cbs.dtu.dk/services/TMHMM-2.0/) were applied to analyze the hydrophobic regions in both the predicted signal and transit sequences.

The deduced TIM protein sequences with N-terminal extensions from chlorophytes *C. reinhardtii* and *V. carteri* and rhodophytes *P. yezoensis* and *G. sulphuraria* and one protein sequence with an N-terminal extension of rhodophyte *C. merolae* retrieved from GenBank, were analyzed with ChloroP program to predict their transit sequences.

**Sequence alignment and phylogenetic analyses.** A BLASTP search was carried out with *E. gracilis* cTIM protein sequence as a query against GenBank. Considering that cTIM genes in eukaryotic cells descended from homologs in eubacteria rather than from those in archaebacteria (Keeling and Doolittle 1997), only eubacterial and eukaryotic TIM homologous sequences were used for the following phylogenetic analyses. For the sequences with high similarities from the same lineage (except euglenoids), only one sequence was selected as a representative. Finally, 40 sequences from 31 eukaryotes and four eubacteria were obtained. All these sequences were aligned with CLUSTALX version 1.8.1 (Thompson et al. 1997) with default settings, the N- and C-terminals were trimmed, and equivocal residues were removed to yield a common block of 234 amino acids. The alignment and a list of GenBank accession numbers for the sequences used are available upon request.

The best-fitting model Whelan and Goldman (WAG) (Whelan and Goldman 2001) with heterogeneous gamma rate distribution across sites and proportion of invariant sites was selected using program PROTTEST 1.2.6 (Abascal, Zardoya, and Posada 2005). One phylogenetic network and three phylogenetic trees using three methods, maximum likelihood (ML) distance, ML, and Markov chain Monte Carlo Bayesian, were reconstructed based on the selected model. A phylogenetic network was generated using Splitstree4 (Huson and Bryant 2006) with ProteinML distance and Neighbor-Net (Bryant and Moulton 2004) for clustering. The three phylogenetic trees were constructed as follows. The ML distance analysis was performed using PUZZLEBOOT (Holder and Roger, version 1.03, http://hades.biochem.dal.ca/Rogerlab/Software/software.html#puzzleboot) upon 1,000 replicates produced by Seqboot in PHYLIP3.6a3 (Felsenstein 1995) software and then the calculated distance matrices were used to reconstruct trees by the Neighbor Joining method. An extended majority rule consensus tree from the replicates was inferred by CONSENSE with the PHYLIP package. The ML method was implemented using PHYML with four rate categories and optimization of tree topology, branch lengths, and rate parameters (1,000 bootstraps) (Guindon and Gascuel 2003). The Metropolis-coupled Markov chain Monte Carlo Bayesian analysis was performed using the program MrBayes V3.0b4 (Ronquist and Huelsenbeck 2003), initiating from a random starting tree, and running four chains simultaneously for 2,500,000 generations with sampling trees every 100 generations. After discarding the first 6,250 trees (''burn-in''), posterior probabilities were calculated from the remaining trees.

Seven alternative topologies differing in the relative position of euglenoid cTIMs and pTIMs were generated with Treeview 1.6.6 (Page 1996). Approximately unbiased (AU) and weighted Shimodaira–Hasegawa (WSH) tests were performed with CONSEL (Shimodaira and Hasegawa 2001) using the likelihoods calculated with Tree-puzzle 5.2 (Strimmer and Von Haeseler 1996) with the same model and parameters as described above.

## RESULTS

**Identification and analyses of *tim* cDNA or DNA sequences from euglenoids, chlorophytes, and rhodophytes.** After assembling the fragments obtained by 3′-RACE and 5′-RACE, two cDNA sequences with different lengths from both *E. gracilis* and *A. longa*, and only one shorter cDNA sequence from *E. intermedia* were obtained. More attempts were made to get the other cDNA sequence from *E. intermedia* with different primers, but these failed. The presence of poly-A tail at the 3′-end, the spliced leader sequence (5′-TCATTCACTTTCTGAGTGTCTATTTTTTTTCG-3′, the signature of mature mRNA in *Euglena*) at the 5′-end (Tessier et al. 1991), and the contiguous open reading frame in each of these obtained sequences indicated that all these cDNA sequences were complete. The deduced protein sequences contain all the

conserved amino acid residues of known TIMs (e.g. Lys12, His96, and Glu166) and conserved motifs (e.g. the conserved sequence 169-WAIGTG-174, which is critical for the substrate binding to the active center). These indicate that all the obtained cDNA sequences encode TIM isozymes. Sequence comparison showed the TIM protein sequence deduced from the shorter cDNA sequence of *A. longa* shares 95% and 87% identities with those from *E. gracilis* and *E. intermedia*, respectively; and the deduced longer TIM protein sequence of *A. longa* shares 82% identity with its counterpart of *E. gracilis*.

One complete cDNA sequence of TIM in the chlorophyte *C. reinhardtii* was obtained by assembling three cDNA fragments, one obtained by our 3′-RACE reaction and two from the cDNA library (Asamizu et al. 2000; http://www.kazusa.or.jp/en/plant/chlamy/EST/), before the genomic data were available. The genomic data of *C. reinhardtii* available later indicated the *tim* gene we obtained was a single copy gene. This is consistent with Schnarrenberger et al. (1990) who detected only one TIM protein. One *tim* DNA sequence without intron of the chlorophyte *V. carteri* was retrieved from the WGS database.

For rhodophytes, the sequence obtained by our PCR from *P. yezoensis* overlapped in its termini with the two partial cDNA sequences retrieved from the EST project (Nikaido et al. 2000; http://www.kazusa.or.jp/en/plant/porphyra/EST/), so a full-length contig was assembled. Two *tim* gene sequences with high similarities to the reported *C. merolae* TIM were obtained from *G. sulphuraria* contig_970_Oct13_2005 through BLAST search. Sequence analysis indicated the shorter one does not possess any introns and is consistent with the cDNA sequence assembled from the downloaded EST data, while the longer one, whose corresponding ESTs were not found in the EST database, contains two shorter introns that can be identified by the canonical GT–AG rule.

The alignment of the above sequences with known cTIMs and pTIMs from representative species indicated that all the shorter-deduced TIM protein sequences do not possess any N-terminal extensions and thus they should be cTIM, while all the longer-deduced TIM protein sequences possess an N-terminal extension, so do the sequences from *C. reinhardtii*, *V. carteri*, and *P. yezoensis* (only one TIM sequence was obtained from each of them). All these sequences with N-terminal extensions were subject to presequence analyses.

**Identification of pTIM sequences in *Euglena gracilis*, *Astasia longa*, *Chlamydomonas reinhardtii*, *Volvox carteri*, *Porphyra yezoensis*, *Galdieria sulphuraria*, and *Cyanidioschyzon merolae*.** Euglenoid secondary plastids have three layers of membrane, and cytoplasmically synthesized plastid precursor proteins are transported in vesicles as integral membrane proteins from ER to Golgi apparatus and finally to plastids. Therefore, unlike presequences of proteins imported into plastids with two-layered membrane, the presequences of euglenoid plastid proteins contain an ER targeting signal peptide domain with a TMH region and a stromal targeting transit peptide domain. In the transit peptide domain, there is another TMH (the second TMH) in most plastid proteins (~86%), which acts as a stop-transfer sequence (Durnford and Gray 2006; Sulli et al. 1999).

Our sequence alignment showed that the longer sequences of both *E. gracilis* and *A. longa* had an extra fragment of ~100 aa at the N-terminals. Our presequence analyses indicated that these N-termimal extensions contain the signal peptide with one TMH and the transit peptide, the features of the nuclear-encoded, plastids-localized proteins of *Euglena*. The details are as follows: in their N-terminals, the beginning ~37 aa consist of a signal peptide followed by a transit peptide of ~64-aa long. In the signal peptide domain, one TMH of 7 aa was predicted according to the loose cutoff value of the DAS program, but the TMH reaches ~20 aa in the prediction of the TMHMM program. Considering that the TMHMM program has excellent discriminatory power for *Euglena* targeting sequences (Durnford, pers. commun.), a TMH should be present in the signal peptide domain. The comparison of the presequences of the two longer euglenoid TIMs with those of previously characterized nuclear-encoded plastids-localized proteins in *E. gracilis* demonstrated that they share similar lengths and structures (Fig. 1), indicating the two longer sequences with

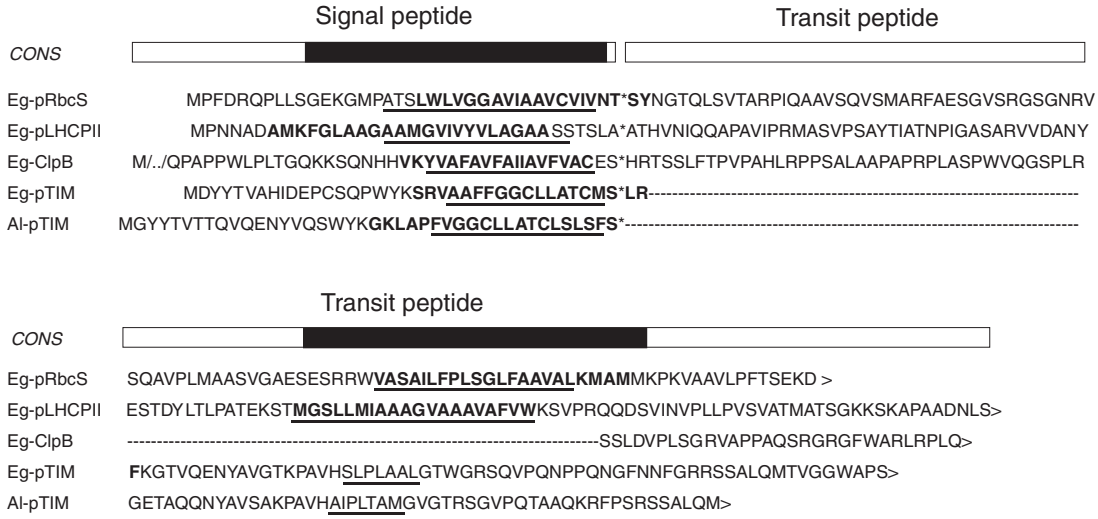


Fig. **1.** Presequence comparison between *Astasia longa* pTIM (Al-pTIM) and *Euglena gracilis* nuclear-encoded plastid proteins. All sequences were aligned at their predicted signal peptidase cleavage site (*). The transmembrane helixes predicted by the dense alignment surface and the transmembrane helix prediction based on a hidden Markov model programs within the signal peptide and transit peptide are underlined and bolded, respectively. ''>''predicted processing site of the transit-targeting peptides. ''-''blanks. ''/../''14 amino acids, which were left out for the convenience of illustration. *CONS* is a diagrammatic representation of the consensus presequence structure with hydrophilic regions in white, hydrophobic regions in gray. Besides our *Euglena gracilis* pTIM (Eg-pTIM), all other *E. gracilis* presequences are according to Sulli et al. (1999). Their abbreviations and accession numbers are as follows: Eg-pRubcS, small subunit of Rubsco (P16881); Eg-ClpB, heat shock protein HSP100; Eg-pLhcpII, light-harvesting complex protein II (S53596).

N-terminal extensions obtained from *E. gracilis* and *A. longa* should be pTIMs. As for the stop transfer sequence, a shorter TMH region appears to be present based on the loose cutoff value of the DAS program but was not predicted by the TMHMM program. Therefore, the pTIMs might lack the stop transfer sequence.

As for the ChloroP predicted cleavage sites of the transit peptides, they are inconsistent between *E. gracilis* and *A. longa* pTIM: in *E. gracilis* pTIM it is 2-aa distant from the start of the corresponding cTIM sequence, while in *A. longa* pTIM the distance is 10 aa.

ChloroP predicted that all three single sequences from *C. reinhardtii*, *V. carteri*, and *P. yezoensis*, and two longer sequences from *G. sulphuraria* and *C. merolae* contain transit sequences with high probabilities, and the lengths of these transit sequences were 26, 32, 36, 37, and 36 aa accordingly. The lengths of their putative mature proteins are similar to known cTIMs of other organisms. Therefore, the five sequences from the two chlorophytes (*C. reinhardtii* and *V. carteri*) and the three rhodophytes (*P. yezoensis, G. sulphuraria*, and *C. merolae*) should be pTIMs.

All our identified sequences have been deposited in GenBank (Accession nos. AY342150 to AY342151, AY742322 to AY742325, DQ459379) and the Third Party Annotation Section of the DDBJ/EMBL/GenBank databases (Acession no. TPA: BK005848).

**Sequence alignment and phylogenetic analyses of TIMs.** Besides the above sequences, our BLAST search retrieved one cTIM from the rhodophyte *Gracilaria gracilis*, one pTIM from the chlorophyte *Prototheca wickerhamii* (containing non-photosynthetic plastids), one cTIM and one pTIM from the chlorophyte *Chlorella pyrenoidosa*, and many TIMs from other eukaryotes (e.g. animals, plants, fungi, and protists) and eubacteria. Interestingly, the alignment of these sequences revealed that TIMs of euglenoids and rhodophytes exclusively shared a unique 2-aa insertion between two conserved motifs: the upstream one is ''GHSERR'' and the downstream one is ''E'' (Fig. 2). The insertion is present in the area that can be defined as loop 4 by our structure modeling with chicken TIM as a modeling template (using CPHmodels 2.0 Server, http://www.cbs.dtu.dk/services/CPHmodels/) (data not shown). Moreover, the two amino acids of the insertion are relatively conserved: PE presents in most euglenoids and all the unicellular rhodophytes (*G. sulphuraria* and *C. merolae*), PQ in *E. intermedia*, and AQ and LQ in all the multicellular rhodophytes *G. gracilis* and *P. yezoensis* respectively.

A network and three phylogenetic trees were produced by our phylogenetic analysis. The three phylogenetic trees showed nearly identical topologies. Although the phylogenetic network shows some conflicting phylogenetic signals (see supplementary materials Fig. S1), it presents congruence with the three trees in exhibiting the following six clades: eubacteria, kinetoplastids, fungi, animals, rhodophytes/euglenoids, and chlorophytes/green plants. Here, we choose the ML tree as the representative tree to illustrate the two new clades (Fig. 3). In the chlorophyte/green plant clade, pTIMs and cTIMs form two sister clades, indicating that the gene duplication previously proposed for plants by Schmidt et al. (1995) occurred most probably before the split of green plants and chlorophytes. Euglenoid TIMs are not closely related to TIMs of the euglenoids' closest relatives, the kinetoplastids, but group with TIMs of rhodophytes, though the bootstrap supports and the posterior probabilities are not high (Fig. 3). Within the clade, euglenoid cTIMs and pTIMs cluster, respectively, to form two branches with strong support.

In order to gain an insight into how well the data supported the phylogenetic clade of euglenoids and rhodophytes, we performed AU and WSH tests for comparing the likelihoods of the topology of Fig. 3 (designated as T0) and seven alternative topologies produced by constraining the relative positions of euglenoid TIMs. Firstly, euglenoid cTIMs and pTIMs were alternatively placed as sister branches, which are indicated by T1 and T2 in Fig. 3, respectively. The topologies T0 and T1 were not rejected by the AU test at a 5% level, and the topology of T2 was rejected by the AU test but was not rejected by the WSH test (see supplementary materials Table S1). Secondly, euglenoid TIMs with cTIMs and pTIMs clustered together were alternatively placed as a sister branch to the kinetoplastid TIMs, to the chlorophyte/green plant TIMs, to green plant cTIMs and pTIMs/chlorophyte pTIMs, to green plant pTIMs/chlorophyte pTIMs, to chlorophyte pTIMs, which are indicated by T3–T7, respectively. All these five topologies were significantly rejected by the AU and WSH tests at a 5% level.

## DISCUSSION

**Two *tim* genes in euglenoids, a pTIM gene in the non-photosynthetic euglenoid *Astasia longa* and only one *tim* in *Cyanidioschyzon merolae* and *Chlamydomonas reinhardtii*, and their implications.** Although Mo et al. (1973) have detected the activities of cTIM and pTIM in *E. gracilis* cytosol and plastids, respectively, no gene evidence of the two isoenzymes had ever been reported in euglenoids, especially in different trophic modes of euglenoids. For the first time we identified the cDNAs of the two TIM isozyme genes in two photosynthetic and one non-photosynthetic euglenoids. Among the three cTIMs, *A. longa* cTIM shares higher identities with *E. gracilis* cTIMs (95%) than with *E. intermedia* cTIM (87%), which indicates that *A. longa* has a closer relationship with *E. gracilis* than with *E. intermedia*. This supports the claim that *A. longa* descended from an *E. gracilis*-like ancestor (Siemeister and Hachtel 1989). Unexpectedly, besides a cTIM gene, a pTIM gene was also identified from the non-photosynthetic euglenoid *A. longa*. This pTIM contains a leader sequence highly similar to that of *E. gracilis* pTIM, and its mature protein sequence shares high identity (91.6%) with that of *E. gracilis* pTIM and also possesses the conserved active site amino acids and important sequence motifs of all other known TIMs. All these suggest it is located in plastids and functions as *E. gracilis* pTIM does. Because pTIM is involved in the Calvin cycle of photosynthesis in plastids, and because plastids in *A. longa* have lost the function of photosynthesis, then what does the expressed pTIM in *A. longa* imply?

Like *A. longa*, apicomplexan parasites (e.g. *Plasmodium* and *Toxoplasma*) and the green algae *P. wickerhamii* and *Helicosporidium* sp. also contain degenerate plastids. Although these plastids are no longer photosynthetic, they were also reported to contain a pTIM, and these organisms have apparently become dependent upon this organelle for other metabolic functions, including fatty acid and isoprenoid synthesis (Borza, Popescu, and Lee 2005; de Koning and Keeling 2004; Waller and McFadden 2005). Furthermore, Shashidhara and Smith (1991) reported that like plastids of *E. gracilis*, plastids of *A. longa* contained an active nuclear-encoded plastid protein—porphobilinogen deaminase, one of the early enzymes involved in porphyrin synthesis, suggesting that the plastids of *A. longa* retain the function of porphyrin synthesis. Therefore, the presence of pTIM in *A. longa* may also suggest that the plastids have function(s) in which TIM is involved. Because, besides the Calvin Cycle, pTIMs take part in glycolysis or/and fatty acid synthesis in plastids of higher plants (Tobin and Bowsher 2005), and because pTIMs in the degenerate plastids of Apicomplexa (apicoplasts) participate in isopentenyl diphosphate biosynthesis (Ralph et al. 2004), then the presence of pTIM probably suggests that plastids of *A. longa* may also have such functions as glycolysis, isopentenyl diphosphate or fatty acid synthesis, though further evidence is awaited.

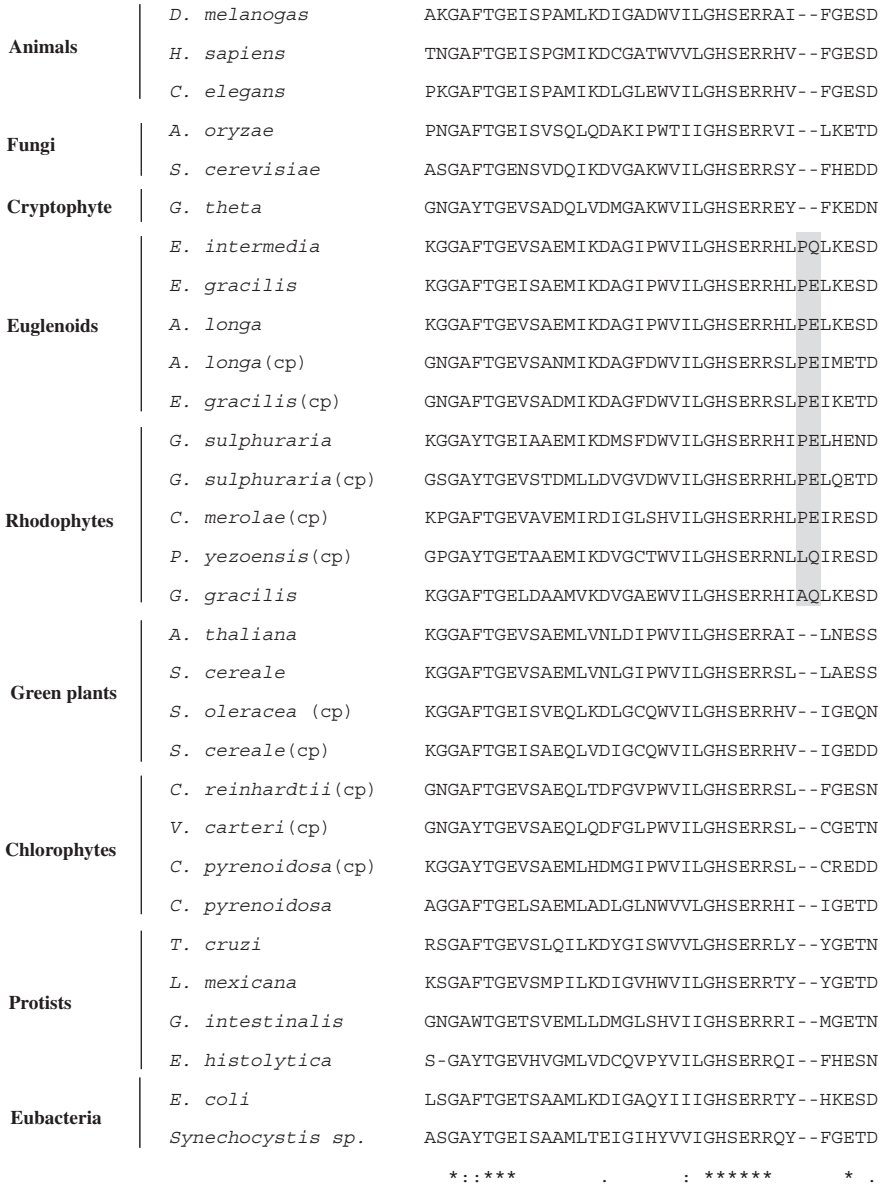


Fig. **2.** Partial alignment results of triosephosphate isomerase (TIM) loop 4 areas of different organisms. TIMs of euglenoids and rhodophytes share a unique 2-aa insertion (characters with shadow) in the loop 4 area and between two highly conserved motifs—the upstream ''GHSERR'' and downstream ''E.'' The highly conserved amino acids are marked with ''*'' beneath their characters.

In *C. reinhardtii*, it was reported that only one activity peak of TIM was isolated from cell fragments (Schnarrenberger et al. 1990). Our investigation also indicated that cells like *C. merolae* and *C. reinhardtii* possess only one *tim* gene—a pTIM gene. How do these cells cope with only one pTIM? The answer may lie in the unique compartmentation of some metabolic pathways in these cells: besides photosynthesis, glycolysis also mainly takes place in plastids in *C. reinhardtii* cells rather than in cytosol (Klein 1986; Schnarrenberger et al. 1990).

In addition, we found that the cleavage sites of the transit peptides predicted by ChloroP were inconsistent between *E. gracilis* and *A. longa* pTIMs. Considering the high identity of pTIMs from *E. gracilis* and *A. longa*, the cleavage sites predicted by ChloroP may be unreliable, and such unreliability of the program prediction was once mentioned by Durnford and Gray (2006). Because ChloroP is developed primarily for use with vascular plant sequences and the criteria for transit peptide cleavage prediction are based on a semi-conserved motif, the disability to predict precisely the cleavage sites might be due to the semi-conserved motif is not yet conserved in plastid proteins of euglenoids and therefore the program may be unsuitable for euglenoid sequences.

**The possible evolutionary relationship between TIMs in euglenoids and rhodophytes.** Previous TIM phylogenetic analysis showed four clades—eubacteria, fungi, animals, and green plants (Keeling and Doolittle 1997). Our phylogenetic trees also indicated the four clades, but showed two other intriguing clades after including chlorophyte, rhodophyte, and euglenoid sequence data. Because much evidence indicated that euglenoids and kinetoplastids had a common ancestor (Triemer and Farmer 1991) and that photosynthetic euglenoids have chlorophyte-derived plastids resulted from a secondary endosymbiosis, it might be expected that euglenoid TIMs would group with kinetoplastid TIMs (due to

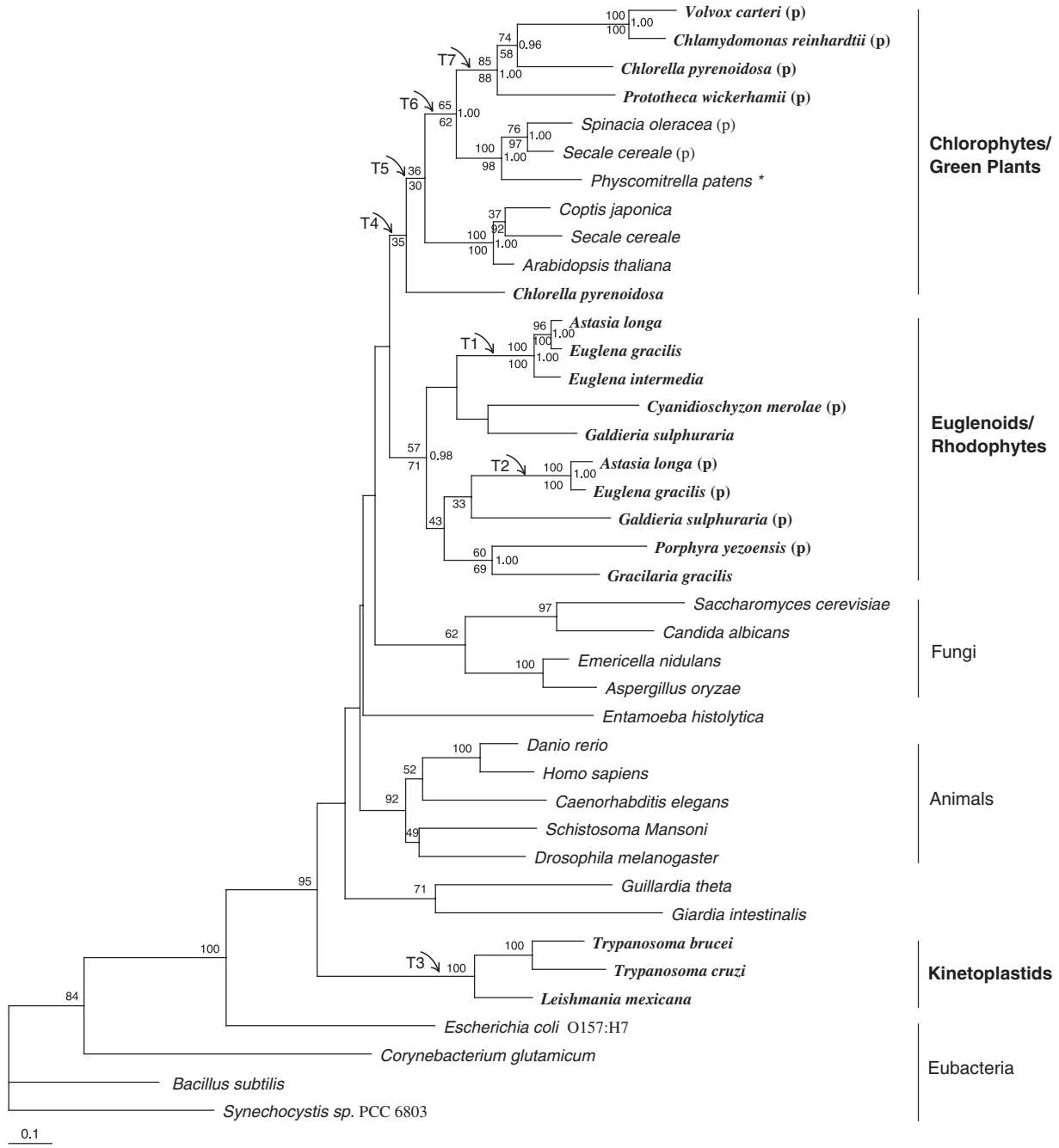


Fig. **3.** Triosephosphate isomerase (TIM) protein phylogeny based on an alignment of 234 amino acid positions. The tree was constructed by the maximum likelihood (ML) method with 1,000 bootstrap replicates under WAG+I+G model using PHYML, and was designated as T0. T1–T7 before arrows indicated topologies used to evaluate seven alternative positions of euglenoid TIMs by performing AU and WSH tests. T1 indicated euglenoid pTIMs clustered as a sister branch to euglenoid cTIMs; T2 indicated euglenoid cTIMs clustered as a sister branch to euglenoid pTIMs; T3–T7 indicated euglenoids were placed as a sister branch to kinetoplastids and different chlorophyte and plant groups. Numbers at nodes correspond to bootstrap values $>30\%$ obtained from ML analysis (top left), ML distance analysis (bottom left) and posterior probabilities from Bayesian analysis $>0.95$ (right) for chlorophyte/green Plant and euglenoid/rhodophyte clade. The scale bar indicates the estimated sequence divergence per unit branch length. Protein sequences containing presequence (located in plastid) were marked with ''(p)'' after their species names. ''*'' behind *Physcomitrella patens* indicated the sequence might be incomplete. WAG, Whelan and Goldman; WSH, weighted Shimodaira–Hasegawa; AU, approximately unbiased.

vertical gene transfer) or chlorophyte TIMs (due to endosymbiotic gene transfer) in phylogenetic analyses. However, our phylogenetic trees indicated that euglenoid cTIMs and pTIMs form two sub-groups, and then they group together with neither kinetoplastid nor chlorophyte but with rhodophyte TIMs. Although such an unusual euglenoid/rhodophyte branch is not very robust, probably due to the limited length of the TIM sequence, and the cTIMs and pTIMs of euglenoids did not cluster as a sister group, the constrained trees by pairwise combination of euglenoids with chlorophytes/green plants and kinetoplastids, respectively, were

all rejected by the AU and WSH tests, and the grouping of euglenoid pTIMs with cTIMs was not rejected in the tests. Therefore, the cluster of euglenid and rhodophyte TIMs at the moment is credible. Nevertheless, whether euglenoid pTIMs and cTIMs group together within the cluster cannot be determined with certainty by our data. One possible reason for this may be that the tree can resolve the cluster of euglenid/rhodophyte TIMs but cannot resolve the branching order within the cluster.

All these result suggest the following: (1) like *tim* genes in chlorophytes/green plants, neither pTIM nor cTIM genes in euglenoids derived from a prokaryotic endosymbiont directly; (2) either convergent evolution of *tim* or its lateral gene transfer (LGT) might have occurred between euglenoids and rhodophytes. Nevertheless, the unique 2-aa insertion shared by euglenoid and rhodophyte TIMs tends to support a common ancestry, and thus a *tim* LGT event between euglenoids and rhodophytes seems more probable. Although euglenoid cTIMs and pTIMs did not group together in the tree, suggesting two separate lateral transfer events of the cytosolic and plastid genes, the sister group of euglenoid cTIMs and pTIMs (T1) was not rejected by the statistical tests, suggesting the *tim* LGT event from rhodophytes to euglenoids most probably occurred only once. Considering that phototrophic euglenoids were descended from phagotrophs (Leander et al. 2007), the LGT probably occurred once by a ''euglenoid'' engulfing a ''rhodophyte'' as a food item, followed by gene duplication in euglenoids to generate a cytosolic and a plastid copy. The pre-existing cytosolic ''euglenoid'' TIM was presumably lost/replaced or else has not been found yet.

Actually, similar LGT events have already been found for *Euglena* cytosolic glyceraldehyde-3-phosphate dehydrogenase and enolase, both of which were considered to be from eubacteria (spirochaetes) (Hannaert et al. 2000). Another less parsimonious, but not impossible, explanation is that the unusual phylogeny resulted from gene duplication events followed by differential gene losses, but the number of losses needed to explain the complete phylogeny would be very large. Therefore, to confirm the evolutionary relationship between euglenoid and rhodophyte TIMs more clearly, it will be necessary to include more data, especially sequences from bodonids and phagotrophic euglenoids, in future phylogenetic analyses.

## ACKNOWLEDGMENTS

We thank Prof. Zhi-Xin Shi and Dr. Jin-Xian Liu for presenting us euglenoid strains and rhodophyte materials, respectively. Thanks also go to Dr. De-Dong Xin and Dr. Hai-Feng Tian for their help in the phylogenetic analyses and helpful discussion. Financial support for this study was provided by the National Nature Science Foundation of China (30570241; 30021004; 90408016; 30623007), 973 Program (2007CB815705) and the CAS (XiBu ZhiGuang Project 2003; The Knowledge Innovation Program KSCX2-YW-R-091).

## LITERATURE CITED

Abascal, F., Zardoya, R. & Posada, D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, **21**:2104–2105.

Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. & Tabata, S. 2000. Generation of Expressed Sequence Tags from low-$CO_2$ and high-$CO_2$ adapted cells of *Chlamydomonas reinhardtii*. *DNA Res.*, **7**:305–307.

Bendtsen, J. D., Nielsen, H., Von Heijne, G. & Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**:783–795.

Borza, T., Popescu, C. E. & Lee, R. W. 2005. Multiple metabolic roles for the nonphotosynthetic plastid of the green alga *Prototheca wickerhamii*. *Eukaryot. Cell*, **4**:253–261.

Bryant, D. & Moulton, V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.*, **21**: 255–265.

Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. 1997. Prediction of transmembrane alpha-helices in procariotic membrane proteins: the Dense Alignment Surface method. *Prot. Eng.*, **10**: 673–676.

Dawson, N. S. & Walene, P. L. 1994. Evolutionary trends in euglenoids. *Arch. Protistenkd.*, **144**:221–225.

de Koning, A. P. & Keeling, P. J. 2004. Nucleus-encoded genes for plastid-targeted proteins in *Helicosporidium*: functional diversity of a cryptic plastid in a parasitic alga. *Eukaryot. Cell*, **3**:1198–1205.

Durnford, D. G. & Gray, M. W. 2006. Analysis of *Euglena gracilis* plastid-targeted proteins reveals different classes of transit sequences. *Eukaryot. Cell*, **5**:2079–2091.

Emanuelsson, O., Nielsen, H. & von Heijne, G. 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.*, **8**:978–984.

Felsenstein, J. 1995. PHYLIP (Phylogeny Inference Package). Department of Genetics, University of Washington, Seattle.

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

Gockel, G. & Hachtel, W. 2000. Complete gene map of the plastid genome of the non-photosynthetic euglenoid flagellate *Astasia longa*. *Protist*, **151**:347–351.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.*, **52**: 696–704.

Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.*, **21**:3537–3544.

Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J. A., Albert, M. A., Sensen C, . W., Gaasterland, T., Muller, M., Michels, P. & Martin, W. 2000. Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.*, **17**:989–1000.

Huson, D. H. & Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.*, **23**:254–267.

Keeling, P. J. & Doolittle, W. F. 1997. Evidence that eukaryotic triosephosphate isomerase is of alpha-proteobacterial origin. *Proc. Natl. Acad. Sci. USA*, **94**:1270–1275.

Klein, U 1986. Compartmentation of glycolysis and of the oxidative pentose-phosphate pathway in *Chlamydomonas reinhardtii*. *Planta*, **167**: 81–86.

Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**:567–580.

Leander, B. S., Esson, H. J. & Breglia, S. A. 2007. Macroevolution of complex cytoskeletal systems in euglenids. *BioEssays*, **29**:987–1000.

Marin, B., Palm, A., Klingberg, M. & Melkonian, M. 2003. Phylogeny and taxonomic revision of plastid-containing euglenophytes based on SSU rDNA sequence comparisons and synapomorphic signatures in the SSU rRNA secondary structure. *Protist*, **154**:99–145.

Matsuzaki, M., Misumi, O., Shin-I, T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., Yoshida, Y., Nishimura, Y., Nakao, S., Kobayashi, T., Momoyama, Y., Higashiyama, T., Minoda, A., Sano, M., Nomoto, H., Oishi, K., Hayashi, H., Ohta, F., Nishizaka, S., Haga, S., Miura, S., Morishita, T., Kabeya, Y., Terasawa, K., Suzuki, Y., Ishii, Y., Asakawa, S., Takano, H., Ohta, N., Kuroiwa, H., Tanaka, K., Shimizu, N., Sugano, S., Sato, N., Nozaki, H., Ogasawara, N., Kohara, Y. & Kuroiwa, T. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature*, **428**:653–657.

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

Müllner, A. N., Angeler, D. G., Samuel, R., Linton, E. W. & Triemer, R. E. 2001. Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using nuclear 18S rDNA sequence. *Int. J. Syst. Evol. Microbiol.*, **51**:783–791.

Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. & Tabata, S. 2000. Generation of 10, 154 expressed sequence tags from a leafy

gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res.*, **7**:223–227.
Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**:357–358.
Ralph, S.A, van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S. & McFadden, G. I. 2004. Metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat. Rev. Microbiol.*, **2**:203–216.
Ronquist, F. R. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**:1752–1754.
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., Jacobshagen, S., Müller, B. & Krüger, I. 1990. Evolution of isozymes of sugar metabolism in green algae. *In*: Rattazzi, M. C., Scandalios, J. G. & Whitt, G. S. (ed.), Isozymes: Structure, Function, and Use in Biology and Medicine. Alan R. Liss Inc., New York. p. 743–764.
Shashidhara, L. S. & Smith, A. G. 1991. Expression and subcellular location of the tetrapyrrole synthesis enzyme porphobilinogen deaminase in light-grown *Euglena gracilis* and three non-chlorophyllous cell lines. *Proc. Natl. Acad. Sci. USA*, **88**:63–67.
Shimodaira, H. & Hasegawa, M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics*, **17**:1246–1247.
Siemeister, G. & Hachtel, W. 1989. A circular 73 kb DNA from the colourless flagellate *Astasia longa* that resembles the chloroplast DNA of *Euglena*: restriction and gene map. *Curr. Genet.*, **15**:435–441.
Starr, R. C. & Zeikus, J. A. 1993. UTEX: the culture collection of algae at the University of Texas at Austin, 1993 list of cultures. *J. Phycol.*, **29**:1–106.
Strimmer, K. & Von Haeseler, A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**:964–969.
Sulli, C., Fang, Z., Muchhal, U. & Schwartzbach, S. D. 1999. Topology of *Euglena* chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. *J. Biol. Chem.*, **274**:457–463.
Tessier, L. H., Keller, M., Chan, R. L., Fournier, R., Weil, J. H. & Imbault, P. 1991. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. *EMBO J.*, **10**:2621–2625.
Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**:4876–4882.
Tobin, A. K. & Bowsher, C. G. 2005. Nitrogen and carbon metabolism in plastids: Evolution, integration and coordination with reactions in the cytosol. *Adv. Bot. Res.*, **42**:113–165.
Triemer, R. E. & Farmer, M. A. 1991. An ultrastructural comparison of the mitotic apparatus, feeding apparatus, flagellar apparatus and cytoskeleton in euglenoids and kinetoplastids. *Protoplasma*, **164**:91–104.
Waller, R. F. & McFadden, G. I. 2005. The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr. Issues Mol. Biol.*, **7**:57–79.
Wattier, R. A., Prodöhl, P. A. & Maggs, C. A. 2000. DNA isolation protocol for red seaweed (Rhodophyta). *Plant Mol. Biol. Rep.*, **18**: 275–281.
Whelan, S. & Goldman, N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.*, **18**:691–699.

## SUPPLEMENTARY MATERIAL

The following material is available for this article online:

**Table S1.** Comparison of eight alternative positions of euglenoids (topologies shown in Fig. 3) with *P* values of AU and WSH tests.

**Fig. S1.** Phylogenetic network of triosephosphate isomerase protein. The phylogenetic network shows conflicting phylogenetic signals but presents congruence with the three phylogenetic trees in forming of the following six clades: Bacteria, Kinetoplastids, Fungi, Animals, Rhodophytes/Euglenoids and Chlorophytes/Green plants.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1550-7408.2008.00324.x (This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

*Received: 08/20/07, 011/07/07, 01/27/08, 02/17/08; accepted: 02/20/08*