Plant Molecular Biology **30**: 65–75, 1996.

Higher-plant chloroplast and cytosolic 3-phosphoglycerate kinases: a case of endosymbiotic gene replacement

Henner Brinkmann [1] and William Martin [2,*]
[1] *Institut für Botanik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany;* [2] *Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany (* author for correspondence)*

Received 10 July 1995; accepted in revised form 25 September 1995

Key words: Archaebacteria, endosymbiosis, molecular evolution, isoenzymes, carbohydrate metabolism

Abstract

Previous studies indicated that plant nuclear genes for chloroplast and cytosolic isoenzymes of 3-phosphoglycerate kinase (PGK) arose through recombination between a preexisting gene of the eukaryotic host nucleus for the cytosolic enzyme and an endosymbiont-derived gene for the chloroplast enzyme. We readdressed the evolution of eukaryotic *pgk* genes through isolation and characterisation of a *pgk* gene from the extreme halophilic, photosynthetic archaebacterium *Haloarcula vallismortis* and analysis of PGK sequences from the three urkingdoms. A very high calculated net negative charge of 63 for PGK from *H. vallismortis* was found which is suggested to result from selection for enzyme solubility in this extremely halophilic cytosol. We refute the recombination hypothesis proposed for the origin of plant PGK isoenzymes. The data indicate that the ancestral gene from which contemporary homologues for the Calvin cycle/glycolytic isoenzymes in higher plants derive was acquired by the nucleus from (endosymbiotic) eubacteria. Gene duplication subsequent to separation of *Chlamydomonas* and land plant lineages gave rise to the contemporary genes for chloroplast and cytosolic PGK isoenzymes in higher plants, and resulted in replacement of the preexisting gene for PGK of the eukaryotic cytosol. Evidence suggesting a eubacterial origin of plant genes for PGK via endosymbiotic gene replacement indicates that plant nuclear genomes are more highly chimaeric, i.e. contain more genes of eubacterial origin, than is generally assumed.

Abbreviations: PGK, 3-phosphoglycerate kinase; FBA, fructose-1,6-bisphosphate aldolase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; TPI, triosephosphate isomerase.

Introduction

The evolutionary origin of chloroplast/cytosol enzymes in plants traditionally has been viewed in light of a working hypothesis known as the gene transfer corollary to endosymbiotic theory [49], which predicts that the genes for cytosolic isoenzymes were endogeneous to the nuclear genome

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number L47295.

at the time of chloroplast acquisition and that, in the process of chloroplast genome reduction, genes for chloroplast isoenzymes were transferred to the nucleus where they acquired a transit peptide for reimport into the organelle of their genetic origin. Calvin cycle and glycolysis possess a number of parallel enzymatic reactions in cytosol and stroma which are catalysed by distinct isoenzymes specific to each compartment, whereas homologues of the Calvin cycle and the oxidative pentose phosphate pathway appear only to occur in chloroplasts [42]. Molecular sequences for these genes can help to shed light on the evolutionary mechanisms which led to the biochemical partitioning now observed in contemporary photosynthetic eukaryotes. Recent molecular studies of chloroplast/cytosol isoenzymes of Calvin cycle and glycolytic pathways revealed, perhaps surprisingly, that the gene transfer corollary has not been substantiated for any isoenzyme pair studied in plants to date. It was gene duplication, not endosymbiotic gene transfer, which provided the origin of the chloroplast/cytosol isoenzymes for fructose-1,6-bisphosphate aldolase (FBA, duplication in early eukaryotic evolution) [36], triosephosphate isomerase (TPI, duplication in chlorophyte evolution) [21, 41], and glyceraldehyde-3-phosphate dehydrogenase (GAPDH, duplication in eubacterial evolution) [32, 20].

In the initial report of molecular sequences for chloroplast and cytosolic 3-phosphoglycerate kinase from plants (PGK) [30] it was noted that the sequences are more similar to one another than one would expect under the gene transfer corollary. It was suggested that recombination had occurred between the respective isoenzyme genes in an attempt to explain the lack of expected similarity between plant and other eukaryotic cytosolic PGK. But, due to the lack at that time of sufficient reference sequences from prokaryotic and eukaryotic sources, an explicit hypothesis was not put forth for the origin of either gene in the context of either eukaryotic or eubacterial evolution. In a later study of PGK gene evolution, Vohra *et al.* [48] interpreted the phylogenetic results obtained as statistical support for the recombination model.

Sequences from archaebacteria are neccessary for comparison to eubacterial and eukaryotic homologues in order to distinguish between different possibilities of PGK gene origin. Two PGK sequences from methanogenic archaebacteria exist in the database [13], but methanogens might not be representative of all archaebacteria concerning PGK evolution. Methanogens use PGK in hexose biosynthesis rather than energy metabolism [8, 9], whereas halophilic archaebacteria possess rhodopsin-based photosynthetic membranes and have been reported to possess ribulose-1,5-bisphosphate carboxylase activity [2, 38], a key Calvin cycle enzyme which methanogens lack altogether. It is thus conceivable that aspects of sugar phosphate metabolism in halophiles are more similar to that in photosynthetic eubacteria and eukaryotes, making them important links for understanding the evolution of genes for enzymes of sugar phosphate metabolism in general. Here we describe the cloning and structure of the PGK gene from the photosynthetic extreme halophile *Haloarcula vallismortis* and reevaluate PGK gene evolution as it relates to the origin of genes for plant chloroplast and cytosol PGK.

Material and methods

Culture conditions and nucleic acid isolation

Haloarcula vallismortis type strain 3756 (ATCC 29715) was obtained from the Deutsche Sammlung für Mikroorganismen (Braunschweig) and grown under shaking at 43 °C in the medium described [37]. Cells were grown to an OD_{600} of 3.0, harvested by centrifugation and lysed by suspension in 50 mM Tris-HCl, 50 mM EDTA pH 8.0. The lysate was purified by phenol/chloroform extraction, nucleic acids were recovered by ethanol precipitation, DNA was purified by CsCl centrifugation in a Ti70 rotor.

Isolation of a pgk *gene from* Haloarcula vallismortis

A library was constructed in λEMBL4 via *Sau* 3AI partial digestion and purification of 18–22 kb fragments in saccharose gradients, ligation and packaging as described [40]. Recombinants were plated on *Escherichia coli* K803 and screened by plaque hybridisation at 32 °C in hybridization buffer (6 × SSPE, 0.1% SDS, 0.02% PVP, 0.02% Ficoll-400) with 10 ng/ml (5×10^6 cpm/ml) of an end-labelled 16-fold degenerate 16-mer constructed against the conserved amino acid motif WYDNEY found in peptides sequenced from the *H. vallismortis* GAPDH enzyme [37]. Positives were purified from which the hybridizing and overlapping 3.0 kb *Bss* HII, 3.3 kb *Pst* I and 2.4 kb *Sac* II fragments were subcloned into pBluescript vectors (Stratagene). The region encompassing the *gap* and *pgk* genes was sequenced on both strands via Exo III deletion series and synthetic oligonucleotide primers using the dideoxy method. Other molecular methods were performed as described [40].

Sequence analysis

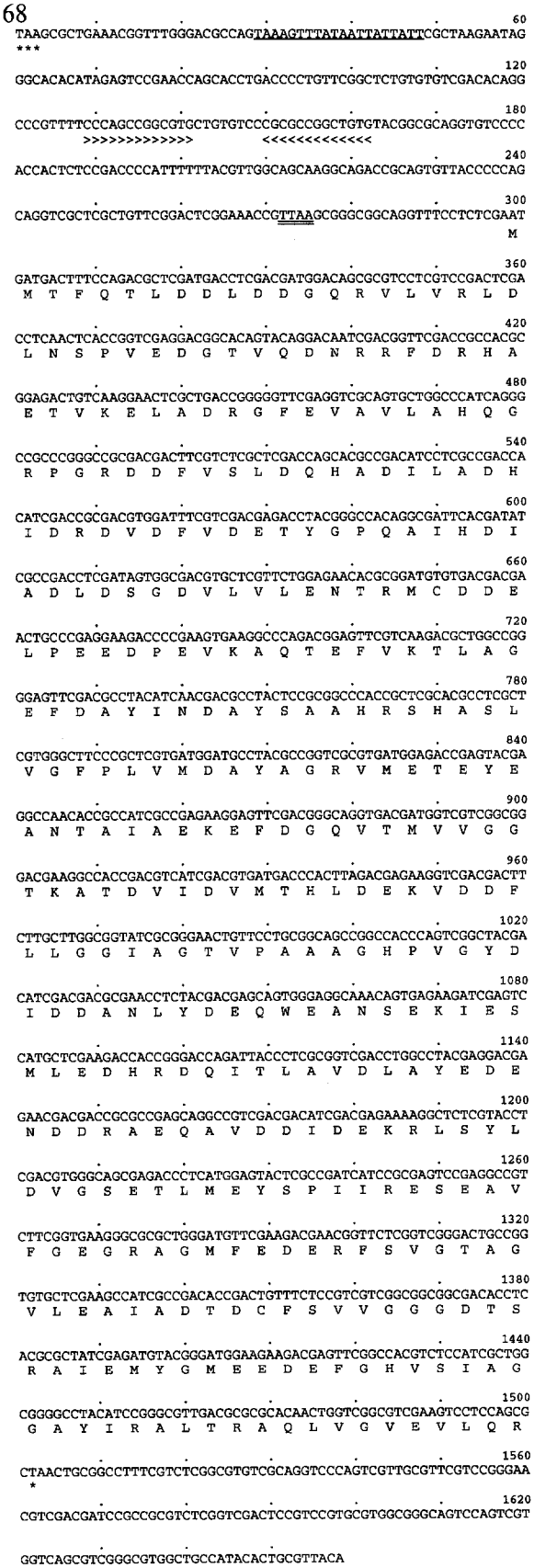
Standard sequence analysis was performed with the GCG Package [10]. Sources of sequences used for analysis are given in the legend to Fig. 3. Sequences were aligned with clustal V [22], the alignment was refined by eye with lineup. After exclusion of positions not occupied by an amino acid in all sequences, 308 positions remained in the final alignment which was used for phylogenetic inference. Distance between sequences was measured as numbers of amino acid substitutions per site corrected for multiple substitutions by assuming a gamma distribution for the variability of substitution rate across positions [24]. For this, a neighbour-joining tree [39] was constructed using the proportion of amino acid differences between PGK sequences, from which the gamma parameter (a) was estimated. The value of a thus determined (2.28) was used to estimate numbers of substitutions per site between sequences using the gamma correction. The resulting distance matrix yielded the final neighbour joining tree (calculated with a program kindly provided by M. Nei). The reliability of branches was estimated by bootstrapping (100 replicates) using the same gamma parameter and by bootstrap parsimony analysis using protpars of phylip [15].

Results and discussion

H. Vallismortis *possesses a* gap-pgk *gene cluster*

Ten phages which positively hydridized to the synthetic oligonucleotide probe were purified and shown by restriction mapping to overlap in the region encompassing the *gap* gene. Several overlapping fragments spanning the region were subcloned and sequenced. Due to the very high GC content of *H. vallismortis* DNA, a large number of sequencing runs on both strands in addition to the use of deaza analogues were necessary before all bases had been determined unambiguously. The *gap* gene sequence (GenBank accession number L47295) was found to contain both of the oligopeptide sequences reported by Prüss *et al.* [37] with no discrepancies. The deduced amino acid sequence of the *Haloarcula gap* gene product shares roughly 50% identical residues with GAPDH proteins from eubacteria and eukaryotes, but like eubacterial GAPDH, it shares only ca. 15% identical residues with sequences from methanogenic archaebacteria [19] suggesting that, in analogy to class I and class II aldolases [36], type I and type II glutamine syntheases [27], or form I and form II Rubisco [7, 33, 47], archaebacteria may possess 'class I' and 'class II' GAPDH (manuscript in preparation).

Downstream of the *gap* gene we found a 1203 bp open reading frame, the predicted product of which (Fig. 1) shows ~37% identity to PGK from methanogenic archaebacteria and <30% identity to PGK from eukaryotes and eubacteria. Between the stop codon of the *gap* gene and the start codon of the *pgk* gene there is a 296 bp stretch which contains several terminator-like hairpin structures and palindromic sequences, the



most prominent of which is indicated in the figure. Upstream of the palindrome is an AT-rich region which stands out due to the high GC content of the region sequenced (60% overall, 85% at third codon positions). A TATA-like motif is present 25 bp upstream of the *pgk* start codon.

Some gene clusters, such as the tryptophan operon, are similarly organized in archaebacteria and eubacteria [28, 29]. A *gap-pgk* gene cluster is found in both archaebacterial and eubacterial genomes (Fig. 3), which suggests that this could represent a gene organisation which might have been present in the progenote. The gene cluster *gap-pgk-tpi* has been found in gram positive eubacteria and *Thermotoga maritima*, although in *Thermotoga* PGK and TPI are fused to a single bifunctional 70 kDa protein [44]. In two γ-purple bacteria, the *fba* gene for fructose-1,6-bisphosphate aldolase (Class II) [1] instead of *tpi* is found downstream of *pgk* (Fig. 3). Although in *E. coli* the *gap-pgk* association is preserved, in the *Haemophilus* genome the *pgk-fba* cluster (accession number U32734) is separated from *gap* and *tpi* by over 500 and 100 kb, respectively, indicating that 'reshuffling' of these genes occurs at least occasionally in bacterial evolution. In accordance with that view is the finding that *pgk* in the thermoacidophilic archaebacterium *Sulfolobus* (not mentioned in [3]) is located next to a *gap* gene, but in *pgk-gap* order. Furthermore, the *Sulfolobus gap* gene encodes a Class II GAPDH enzyme (methanogen-like, see above), suggesting convergent *pgk* gene cluster organization in this thermoacidophile.

Fig. 1. Sequence of the *Haloarcula vallismortis pgk* gene in the *gap-pgk* cluster. An AT-rich region (underlined) and a conspicuous palindrome (>>> <<<) are indicated. The putative TATA box is double underlined, the stop codon of the *gap* gene is indicated with three asterisks, that of the *pgk* gene with one asterisk.

High net charge of PGK from an extreme halophile

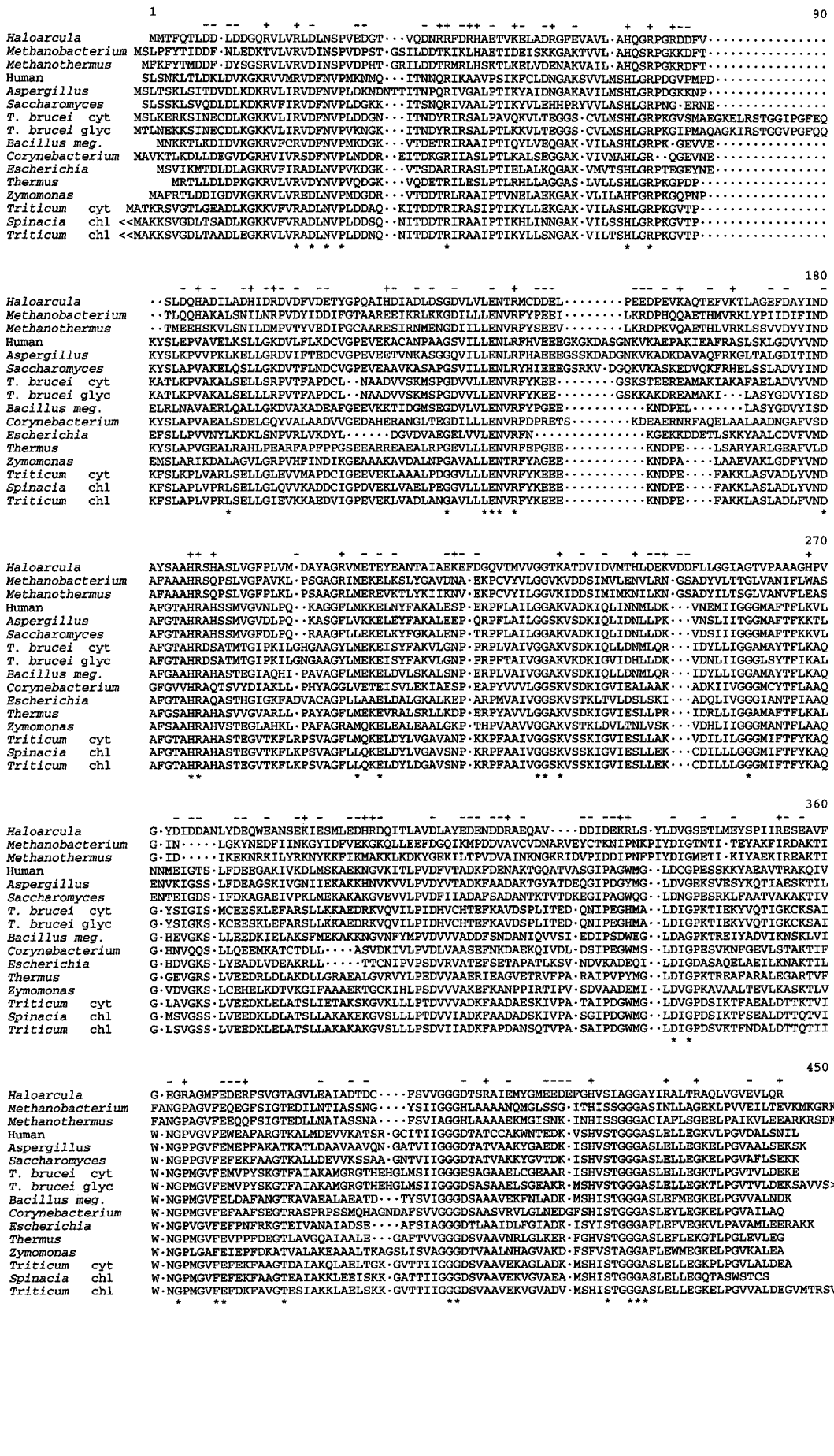
H. vallismortis grows well at 43 °C in 4–5 M NaCl, in our hands a doubling time of about 4 h

was observed. Enzymes of sugar phosphate metabolism from extreme halophilic archaebacteria require salt concentrations of 2 to 3 M in order to achieve full enzyme activity [2, 6, 11, 37], a concentration which would be sufficient to precipitate many proteins from non-halophilic organisms. Furthermore, intracellular salt concentrations in extreme halophiles are typically at least as high as those of the environment (ca. 3.5 M) [9], it is therefore perhaps not surprising that the deduced amino acid sequence of *H. vallismortis* PGK has 93 aspartate plus glutamate residues (D + E) in addition to 41 histidine, lysine plus arginine (H + K + R) residues for total of 134 out of 401 (33%) highly charged residues and an overall predicted charge of −63 from the simple primary sequence. By comparison, *E. coli* PGK has an overall charge of −11 (55 D + E and 48 H + K + R), human PGK a charge of +3 (50 D + E and 58 H + K + R), and PGK from the non-halophilic archaebacterium *Methanothermus fervidus* a charge of +5 (55 D + E and 68 H + K + R). The highly negative charge of *H. vallismortis* PGK is, however, comparable to that previously reported for various cytoplasmic enzymes from extreme halophiles: superoxide dismutase, −32 [25]; NADP-specific glutamate dehydrogenase, −40 [4]; dnaK, −102 [17]; enolase, −53 [26]. This charge is thus not the product of chance, but rather a common characteristic of cytoplasmic proteins from halophiles which likely serves to keep them in solution at high to saturating cellular salt concentrations. An alignment of phosphoglycerate kinase sequences from eukaryotic, eubacterial and archaebacterial genomes (Fig. 2) indicates that the distribution of charge in the *H. vallismortis* sequence is rather uniform, revealing no obvious concentration of either positively or negatively charged residues in any portion of the protein primary structure.

Plant chloroplast/cytosol PGK isoenzymes: a gene duplication

Based upon an expanded version of the alignment in Fig. 2, phylogenetic analysis of PGK sequences including chloroplast and cytosolic isoenzymes from plants provide an intruiging picture of gene evolution (Fig. 3). Both chloroplast and cytosolic PGK sequences from plants assume a very distinct position in the gene tree relative to *pgk* genes of non-photosynthetic eukaryotes. The cytosolic enzyme does not branch with its homologues from higher eukaryotes and very importantly, genes for both plant isoenzymes share a common branch which is located well below *pgk* genes from the kinetoplastid protists *Trypanosoma* and *Crithidia* with a quite high bootstrap percentage (94) (Fig. 3). This finding quite clearly indicates that neither genes for the cytosolic nor the chloroplast enzyme in plants surveyed (only chlorophytes) are orthologues of the cytosolic enzyme in non-photosynthetic eukaryotes, since orthologues would be expected to branch well above kinetoplastid genes, close to animals and fungi. This unexpected pattern of similarity was perceived by Longstaff *et al.* [30], who put forth a 'recombination' hypothesis (as one of several alternatives) to explain the lack of expected similarity between wheat cytosolic PGK and its eukaryotic homologues. Under that model, the gene for the plant cytosolic enzyme is orthologous to its counterpart in animals and fungi whereas the gene for the chloroplast enzyme derives from the cyanobacteria-like antecedents of plastids, and subsequent recombination was suggested to have occurred between them resulting in genes with intermediate similarity to eubacterial and eukaryotic homologues.

The recombination hypothesis was thus envoked under Weeden's [49] gene transfer corollary as a subordinate scenario to explain aspects of the data and was accepted in later studies of PGK evolution [14, 48]. However, recombination does not influence divergence *between recombining genes* and therefore cannot account for the finding that the plant isoenzymes are more similar (82% identity) to each other than to any eukaryotic *or* eubacterial homologue. The by far most straightforward interpretation of the data is that genes for the chloroplast and cytosolic isoenzymes of phosphoglycerate kinase in higher plants arose through gene duplication in eukaryotic ge-


Haloarcula
Methanobacterium
Methanothermus
Human
Aspergillus
Saccharomyces
T. brucei cyt
T. brucei glyc
Bacillus meg.
Corynebacterium
Escherichia
Thermus
Zymomonas
Triticum cyt
Spinacia chl
Triticum chl

nomes. That view is strongly supported by the position of the *Chlamydomonas* sequence, which branches robustly with other plant sequences but below the node corresponding to the duplication event which gave rise to the chloroplast and cytosolic isoenzymes of wheat (and presumably of other higher plants). The position of the spinach chloroplast PGK sequence suggests that the duplication occurred before the separation of monocots and dicots.

Gene duplications which resulted in enzymes of different eukaryotic cell compartments is a recurring theme of PGK evolution, as suggested by the portion of the tree bearing the kinetoplastid sequences (Fig. 3). It is evident that genes for recompartmentalized PGK enzymes destined for the glycosome, a specific glycolytic microbody present in some but not all kinetoplastids [18], have arisen several times independently in kinetoplastid evolution. The three PGK genes of *T. brucei* and *T. congolense* are organized in tandem [34, 35], the first in the array (*pgkA*) carries a large insertion of roughly 80 amino acids not found in other PGK enzymes. The resulting product is of higher molecular weight (ca. 56 kDa) and is located in glycosomes [35], although it is not the major glycosomal form in *T. brucei* [34]. The topogenic signals involved in glycosomal precursor import are quite different from those for chloroplasts [34, 46], but for both plants and protists, the genes for differentially compartmentalized PGK isoenzymes arose through duplication (Fig. 3).

Comparison of PGK sequences from the three urkingdoms

H. vallismortis PGK is more similar to PGK from two methanogenic archaebacteria (Fig. 3) than it is to eubacterial or eukaryotic sequences, suggesting that it represents an orthologue of other archaebacterial PGK genes (the partial *Sulfolobus* PGK sequence also branches with these homologues, data not shown). The elevated rate of amino acid substitution for the *H. vallismortis* sequences implied by the length of its terminal branch very likely relates to the roughly 40 additional acidic amino acids that this sequence contains relative to its homologues from nonhalophilic organisms. An important aspect of Fig. 3 is that eubacterial and eukaryotic PGK genes are much more similar to each other (ca. 50% identity) than either is to archaebacterial sequences (only ca. 30% amino acid identity). This is in marked contrast to the majority of protein coding genes which have been studied in the three urkingdoms; they depict a different urkingdom topology, showing greater similarity between archaebacterial and eukaryotic genes relative to eubacterial homologues [5, 12, 16, 23, 51, 52]. This discrepancy in PGK gene evolution was a finding upon which, at least in part, Zillig's hypothesis for the origin of the eukaryotic genome via the fusion of archaebacterial and eubacterial genomes was originally based [51, 52] (see [12] for a review). Due to the high degree of divergence between archaebacterial and remaining PGK sequences, the position of the archaebacterial branch may be a long-branch artefact, as suggested by low bootstrap values in basal portion of the tree.

Eubacterial origin of chlorophyte PGK genes

Note that the portion of the PGK tree for nonphotosynthetic eukaryotes is consistent with other molecular phylogenies regarding the branching order of trypanosomes, *Plasmodium*, animals and fungi. Note that if one disregards for a moment all eukaryotic sequences in the figure (i.e. by sim-

Fig. 2. Alignment of PGK sequences. Sources of sequences are as given in the legend to Fig. 3. Strictly conserved residues in the alignment are indicated with an asterisk. Positively and negatively charged amino acids in the *Haloarcula vallismortis* sequence are indicated above the alignment with '+' and '−' respectively. Gaps are indicated as dots. cyt, cytosolic; chl, chloroplast; glyc, glycosome. '≫' indicates that the C-terminus of the *T. brucei* glycosomal sequence (YASAGTGTLSNRWSSL) is not shown. '≪' indicates the presence of a transit peptide. Position numbering is arbitrary.

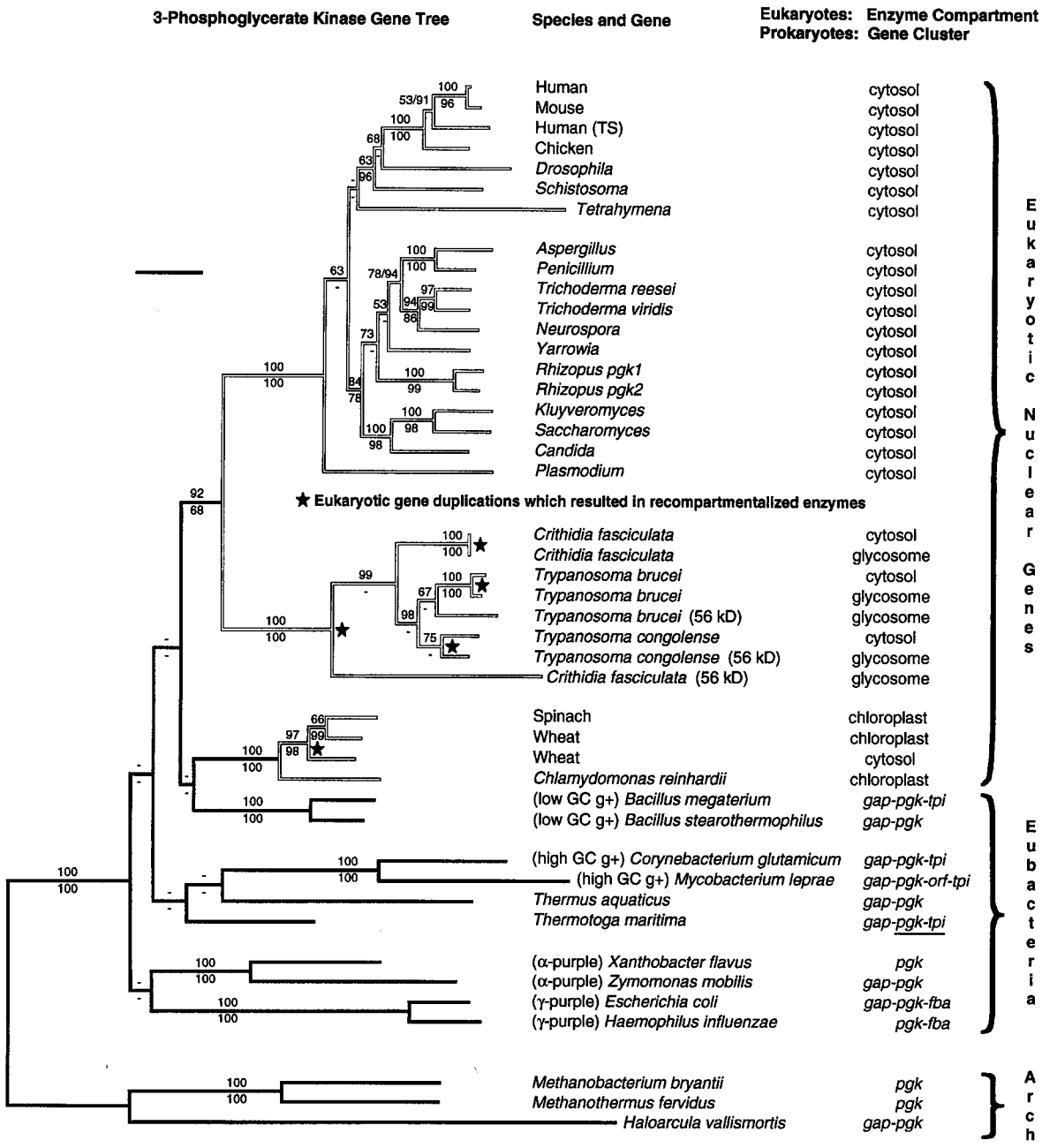


Fig. 3. Neighbour joining tree for PGK protein sequences constructed from pairwise estimates of numbers of amino acid substitutions obtained assuming a gamma distribution for the variabiliy of substitution rate across positions. Genes found in eukaryotic genomes are designated with open branches, those found in eubacterial and archaebacterial genomes are designated with solid branches. Gene duplication events which resulted in subcellular recompartmentalization of enzymes are indicated with stars. Abbreviations are: TS, testis-specific; g+, gram-positive bacteria; GC, guanosine plus cytosine content; purple, purple bacteria; arch, archaebacteria. The scale bar at the upper left indicates 0.1 substitutions per site. PGK and TPI are fused in a single cotranslated protein in *Thermotoga* [44], as indicated by underlining. Numbers above nodes indicate the bootstrap proportion for 100 replicates using the same tree construction method, numbers below nodes (or in parenthesis) indicate the bootstrap proportion using protein parsimony of phylip. Bootstrap values lower than 50/100 are indicated as '–'. Sources of sequences are: *Aspergillus nidulans* P11977, *Bacillus megaterium* P24269, *Bacillus stearothermophilus* P18912, *Candida maltosa* P41757, *Chlamydomonas*

ply covering them with a piece of paper), the eubacterial portion of the PGK tree is *also* consistent in many aspects with traditional molecular systematic views of eubacterial evolution [50]: (1) common branching of purple bacterial sequences [45], (2) common branching of low-G + C gram-positives [31], (3) common branching of high G + C-gram-positives [31], (4) separate branches for low and high G + C-gram-positives respectively [31], and (5) 'basal' positions within eubacteria for the thermophiles *Thermus* and *Thermotoga*. The 'problem' with the PGK tree is simply that PGK genes located in eukaryotic genomes are borne on eubacterial branches, and the plant sequences are distinct from remaining eukaryotic homologues. Two scenarios, each of which takes into account current theories on the origin of the eukaryotic nucleus can account for this finding.

As to the first of these, we assume that the fusion hypothesis [12] is correct. It can account for the PGK gene tree topology observed, if (and only if) one makes the corollary assumptions that (1) the position of the archaebacterial 'root' is a treeing artefact, its true position lying on the 92/100 bootstrap branch bearing non-photosynthetic eukaryotic PGK sequences (which cannot be excluded), (2) eukaryotes lost the archaebacterial copy and retained only the eubacterial *pgk* gene subsequent to fusion, (3) the position of the plant sequences is explained as the result of subsequent acquisition via endosymbiotic gene transfer from eubacteria of a *pgk* gene which then underwent duplication to yield the plant isoenzymes, *and* (4) the preexisting gene for the plant homologue of cytosolic PGK from animals and fungi was simply replaced (or lost) subsequent to duplication.

But what if the fusion hypothesis is not correct? As the second scenario, we assume that archaebacteria and eukaryotes are truly sister groups as suggested by several papers [12]. In that case, one can account for the PGK topology observed, if (and only if) the corollary assumptions are made that (1) *all* PGK sequences found in eukaryotic genomes are of eubacterial (endosymbiotic) origin, (2) the ancestral plant gene (which duplicated to yield the plant isoenzymes) is of eubacterial origin via endosymbiotic gene transfer *from chloroplasts*, and (3) the preexisting gene for the plant homologue of cytosolic PGK from animals and fungi was simply replaced (or lost) subsequent to duplication.

In either case, the ancestral gene which gave rise via duplication to the plant isoenzymes are of eubacterial origin, and any preexisting PGK enzyme of the primitive chlorophyte cytosol must have been replaced (lost) subsequent to the duplication event. That would easily account for the finding that *Chlamydomonas reinhardtii* does not possess a cytosolic isoenzyme for PGK [43], rather only a chloroplast enzyme. Non-green photosynthetic protists (or *Euglena*) which branched early in eukaryotic evolution might have retained different type(s) of PGK genes. Plant PGK sequences are slightly more similar (with low bootstrap support) to PGK from low-G + C-gram-positive eubacteria (*Bacillus*) than to homologues from other eubacteria, the same is found for another nuclear-encoded chloroplast enzyme of eubacterial origin [20, 32]. No PGK sequence data from cyanobacteria are currently available. Under our working hypothesis of chloroplast origin for plant PGK, we would predict cyanobac-

---

reinhardtii U14912, *Corynebacterium glutamicum* Q01655, *Crithidia fasciculata* pgkA glycosome (56PGK) P25055, *Crithidia fasciculata* pgkB cytosol P08966, *Crithidia fasciculata* pgkC glycosome P08967, *Drosophila melanogaster* Q01604, *Escherichia coli* P11665, chicken L37101, *Haemophilus influenzae* U32734, human P00558, human testis-specific X05246, *Kluveromyces lactis* P14828, *Methanobacterium bryantii* P20972, *Methanothermus fervidus* P20971, mouse P09411, *Mycobacterium leprae* U00013, *Neurospora crassa* P38667, *Penicillium citrinum* P33161, *Plasmodium falciparum* P27362, *Rhizopus niveus* pgk1 P29405, *Rhizopus niveus* pgk2 P29406, *Saccharomyces cerevisiae* P00560, *Schistosoma mansoni* L36833, *Spinacia oleracea* chloroplast P29409, *Tetrahymena thermophila* X63528, *Thermotoga maritima* X75437, *Thermus aquaticus* P09403, *Trichoderma reesei* P14228, *Trichoderma viride* P24590, *Triticum aestivum* chloroplast P12782, *Triticum aestivum* cytosol P12783, *Trypanosoma brucei* pgkA glycosome (56PGK) P08891, *Trypanosoma brucei* pgkB cytosol P07377, *Trypanosoma brucei* pgkC glycosome P07378, *Trypanosoma congolense* pgkA glycosome (p56) P41762, *Trypanosoma congolense* pgkC cytosol P41760, *Xanthobacter flavus* U08462, *Yarrowia lipolytica* P29407, *Zymomonas mobilis* P09404.

terial *pgk* genes to be found which branch above *Bacillus* PGK but below the nuclear gene for *Chlamydomonas* chloroplast PGK in the *pgk* gene phylogeny.

Acknowledgements

We thank C. Schnarrenberger for valuable discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft.

References

1. Alefounder PR, Perham SA: Identification, molecular cloning and sequence analysis of a gene cluster encoding the 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 (1989).
2. Altekar W, Rajagopalan R: Ribulose bisphosphate carboxylase activity in halophilic archaebacteria. Arch Microbiol 153: 169–174 (1990).
3. Arcari P, Dello-Russo A, Ianniciello G, Gallo M, Bocchini V: Nucleotide sequence and molecular evolution of the gene coding for glyceraldehyde-3-phosphate dehydrogenase in the thermoacidophilic archaebacterium *Sulfolobus solfataricus*. Biochem Genet 31: 241–251 (1993).
4. Benachenhou N, Baldacci G: The gene for a halophilic glutamate dehydrogenase: sequence, transcription analysis and phylogenetic implications. Mol Gen Genet 230: 345–352 (1991).
5. Brown JR, Doolittle WF: Root of the universal tree of life based on ancient aminoacyl-tRNA synthase gene duplications. Proc Natl Acad Sci USA 92: 2441–2445 (1995).
6. Cendrin F, Chroboczek J, Zaccia G, Eisenberg H, Mevarech M: Cloning, sequencing and expression in *Escherichia coli* of the gene coding for malate dehydrogenase from the extremely halophilic archaebacterium *Haloarcula marismortui*. Biochemistry 23: 4308–4313 (1993).
7. Chen J-H, Gibson JL, McCue LA, Tabita FR: Identification, expression, and deduced primary structure of transketolase and other enzymes encoded within the form II carbon dioxide fixation operon of *Rhodobacter sphaeroides*. J Biol Chem 266: 20447–20452 (1991).
8. Danson MJ: Archaebacteria: the comparative enzymology of their central matabolic pathways. Adv Microbial Physiol 29: 164–231 (1988).
9. Danson MJ: Central metabolism of the archaebacteria: an overview. Can J Microbiol 35: 58–64 (1989).
10. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the vax. Nucl Acids Res 12: 387–395 (1984).
11. Dhar NM, Altekar W: Distribution of class I and class II fructose bisphosphate aldolases in halophilic archaebacteria. FEMS Microbiol Lett 35: 177–181 (1986).
12. Doolittle RF: Of Archae and Eo: What's in a name? Proc Natl Acad Sci USA 92: 2421–2423 (1995).
13. Fabry S, Heppner P, Dietmaier W, Hensel R: Cloning and sequencing the gene encoding 3-phosphoglycerate kinase from mesophilic *Methanobacterium bryantii* and thermophilic *Methanothermus fervidus*. Gene 91: 19–25 (1990).
14. Fothergill-Gilmore LA, Michels PAM: Evolution of glycolysis. Progr Biophys Mol Biol 59: 105–238 (1993).
15. Felsenstein J: PHYLIP: Phylogeny Inference Package (version 3.2). Cladistics 5: 164–166 (1989).
16. Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF, Poole RJ, Date T, Oshima T, Konishi J, Denda K, Yoshida M: Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc Natl Acad Sci USA 86: 6661–6665 (1989).
17. Gupta RS, Singh B: Cloning of the HSP70 gene from *Halobacterium marismortui*: relatedness of archaebacterial HSP70 to is eubacterial homologs and a model for the evolution of the HSP70 gene. J Bact 174: 4594–4605 (1992).
18. Hannaert V, Michels P: Structure, function and biogenesis of glycosomes in Kinetoplastida. J Bioenerg Biomembr 20: 205–212 (1994).
19. Hensel R, Zwickl P, Fabry S, Lang J, Palm P: Sequence comparison of glyceraldehyde-3-phosphate dehydrogenase from the three urkingdoms: evolutionary implication. Can J Microbiol 35: 81–85 (1989).
20. Henze K, Badr A, Wettern M, Cerff R, Martin W: A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA 92: 9122–9126 (1995).
21. Henze K, Schnarrenberger C, Kellermann J, Martin W: Chloroplast and cytosolic triosephosphate isomerase from spinach: Purification, microsequencing and cDNA sequence of the chloroplast enzyme. Plant Mol Biol 26: 1961–1973 (1994).
22. Higgins DG, Sharp PM: CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244 (1988).
23. Iwabe N, Kuma K-I, Hasegawa M, Osawa S, Miyata T: Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci USA 86: 9355–9359 (1989).
24. Jin L, Nei M: Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7: 82–102 (1990).
25. Joshi PB, Dennis PP: Characterization of paralogous and orthologous members of the superoxide dismutase gene family from genera of the halophilic archaebacteria. J Bact 175: 1561–1571 (1993).
26. Krömer WJ, Arndt E: Halobacterial S9 operon: three ribosomal protein genes are cotranscribed with genes en-

coding a tRNA-Leu, the enolase, and a putative membrane protein in the archaebacterium *Haloarcula* (*Halobacterium*) *marismortui*. J Biol Chem 266: 24573–24579 (1991).

27. Kumada Y, Benson DR, Hillemann D, Hosted TJ, Rochefort DA, Thompson CJ, Wohlleben W, Tateno Y: Evolution of the glutamin synthase gene, one of the oldest existing and functioning genes. Proc Natl Acad Sci USA 90: 3009–3013 (1993).
28. Lam WL, Cohen A, Tsouluhas D, Doolittle WF: Genes for tryptophan biosynthesis in the halophile archaebacterium *Haloferax volcanii*: the *trpDFEG* cluster. Proc Natl Acad Sci USA 87: 6614–6618 (1990).
29. Lam WL, Logan SM, Doolittle WF: Genes for tryptophan biosynthesis in the halophile archaebacterium *Haloferax volcanii*: the *trpDFEG* cluster. J Bact 174: 1694–1697 (1992).
30. Longstaff M, Raines CA, McMorrow EM, Bradbeer JW, Dyer T: Wheat phosphoglycerate kinase: Evidence for recombination between the genes for the chloroplastic and cytosolic enzymes. Nucl Acids Res 17: 6569–6580 (1989).
31. Ludwig W, Kirchhof G, Klugbauer N, Weizenegger N, Betzl D, Ehrmann M, Hertel C, Jilg S, Tatzel R, Zitelsberger H, Liebl S, Hochberger M, Shah J, Lane D, Wallnöfer P, Scheifer KH: Complete 23S ribosomal RNA sequences of Gram positive bacteria with a low G + C content. System Appl Microbiol 15: 487–501 (1992).
32. Martin W, Brinkmann H, Savona C, Cerff R: 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 (1993).
33. Morse D, Salois P, Markovic P Hastings JW: A nuclear-encoded form II RuBisCO in dinoflagellates. Science 268: 1622–1624 (1995).
34. Osinga KA, Swinkels BW, Gibson WC, Borst P, Veeneman GH, van Boom JH, Michels PAM, Opperdoes F: Topogenesis of microbody enzymes: a sequence comparison of the genes for the glycosomal (microbody) and cytosolic phosphoglycerate kinases of *Trypanosoma brucei*. EMBO J 4: 3811–3817 (1985).
35. Parker HL, Hill T, Alexander K, Murphy NB, Fish WR, Parsons M: Three genes and two isoenzymes: Gene conversion and the compartmentalization and expression of the phosphoglycerate kinases of *Trypanosoma (Nannomonas) congolense*. Mol Biochem Parasitol 69: 269–279 (1995).
36. Pelzer-Reith B, Penger A, Schnarrenberger C: Plant aldolase: cDNA and deduced amino-acid sequence of the chloroplast and cytosol enzyme from spinach. Plant Mol Biol 21: 331–340 (1993).
37. Prüss B, Meyer HE, Holldorf AW: Characterization of the glyceraldehyde-3-phosphate dehydrogenase from the extremely halophilic archaebacterium *Haloarcula vallismortis*. Arch Microbiol 160: 5–11 (1993).
38. Rawal H, Kelkar SM, Altekar W: Ribulose-1,5-bisphosphate dependent CO_2 fixation in the halophilic archaebacterium *Halobacterium mediterranei*. Biochem Biophys Res Comm 156: 451–456 (1988).
39. Saitou N, Nei M: The neighbor-joining method: a new method for the reconstruction of phylogenetic trees. Mol Biol Evol 4: 406–425 (1987).
40. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989).
41. Schmidt M, Dvendsen I, Feierabend J: 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 (1995).
42. Schnarrenberger C, Flechner A, Martin W: Enzymatic evidence indicating a complete oxidative pentose phosphate in the chloroplasts and an incomplete pathway in the cytosol of spinach leaves. Plant Physiol 108: 609–614 (1995).
43. Schnarrenberger C, Jacobshagen S, Müller B, Krüger I: Evolution of isozymes of sugar phosphate metabolism in green algae. In: Ogita S, Scandalios JS (eds) Isozymes: Structure, Function, and Use in Biology and Medicine, pp. 743–7460. Wiley-Liss, New York (1990).
44. Schurig H, Beaucamp N, Ostendorp R, Jaenicke R, Adler E, Knowles JR: Phosphoglycerate kinase and triosephosphate isomerase from the hyperthermophilic bacterium *Thermotoga maritima* form a covalent bifunctional enzyme complex. EMBO J 14: 442–451 (1995).
45. Stackebrandt E, Murray RGE, Trüper HG: *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the 'purple bacteria and their relatives'. Int J Syst Bact 38: 321–325 (1988).
46. Swinkels BW, Evers R, Borst P: The topogenic signal of the glycosomal (microbody) phosphoglycerate kinase of *Crithida fasiculata* resides in a carboxy-terminal extension. EMBO J 7: 1159–1165 (1988).
47. Tabita RF: Carbon dioxide fixation and its regulation in cyanobacteria. In: Fay P, van Baalen C (eds) The Cyanobacteria, pp. 95–118 Elsevier, Amsterdam (1987).
48. Vohra GB, Golding GB, Tsao N, Pearlman RE: A phylogenetic analysis based on the gene encoding phosphoglycerate kinase. J Mol Evol 34: 383–395 (1992).
49. Weeden NF: Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J Mol Evol 17: 133–139 (1981).
50. Woese CR: Bacterial evolution. Microbiol Rev 51: 221–271 (1987).
51. Zillig W, Klenk H-P, Palm P, Leffers H, Pühler G, Gropp F, Garrett RA: Did eukaryotes originate by a fusion event? Endocyt C Res 6: 1–25 (1989).
52. Zillig W: Comparative biochemistry of Archaea and Bacteria. Curr Opin Genet Devel 1: 544–550 (1991).