ELSEVIER

Gene 214 (1998) 205–213

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

Eubacterial origin of nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase from spinach: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis[1]

Ulrich Nowitzki [a], Anke Flechner [b], Josef Kellermann [c], Masami Hasegawa [d], Claus Schnarrenberger [b], William Martin [a,]*

[a] *Institut für Genetik, Technische Universität Braunschweig, Spielmannstr. 7, D-38023 Braunschweig, Germany*
[b] *Institut für Pflanzenphysiologie und Mikrobiologie, Freie Universität Berlin, Königin-Luise-Str. 12-16a, D-14195 Berlin, Germany*
[c] *Max-Planck-Institut für Biochemie, Am Klopferspitz, D-82152 Martinsried, Germany*
[d] *The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*

Received 15 January 1998; accepted 24 March 1998

Abstract

Higher plants possess two distinct nuclear-encoded glucose-6-phosphate isomerase (GPI) isoenzymes, a cytosolic enzmye of the Embden–Meyerhof pathway and a chloroplast enzyme essential to storage and mobilization of carbohydrate fixed by the Calvin cycle. We have purified spinach chloroplast GPI to homogeneity, determined amino acid sequences from the active enzyme, and cloned cDNAs for chloroplast and cytosolic GPI isoenzymes from spinach. Sequence comparisons reveal three distantly related families of GPI genes that are non-uniformly distributed among contemporary eubacteria and archaebacteria, suggesting that ancient gene diversity existed for this glycolytic enzyme. Spinach chloroplast GPI is much more similar to its homologue from the cyanobacterium *Synechocystis* PCC6803 than it is to the enzyme from any other source, providing strong evidence that the gene for chloroplast GPI was acquired by the nucleus via endosymbiotic gene transfer from the cyanobacterial antecedants of chloroplasts. Eukaryotic nuclear genes for cytosolic GPI are more similar to eubacterial than to archaebacterial homologues, suggesting that these too were acquired by eukaryotes from eubacteria, probably during the course of the endosymbiotic origin of mitochondria. Chloroplast and cytosolic GPI provide evidence for a eubacterial origin of yet another component of the eukaryotic glycolytic pathway. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Evolution; Endosymbiosis; Glycolysis; Isoenzymes; Metabolism

1. Introduction

Both the host cell of plastid symbiosis and the cyanobacterial antecendant of plastids should have possessed more or less complete sets of genes for the enzymes of core carbohydrate metabolism (glycolysis, gluconeogenesis, and oxidative pentose phosphate pathway). During the course of endosymbiosis, plastids relinquished the majority of their genes, yet retained many of the biochemical pathways germane to their prokaryotic heritage. For enzymes of pathways that (i) were common to both symbiont and host at the onset of plastid symbiosis; and (ii) were functionally equivalent, endosymbiotic gene transfer to nuclear chromosomes resulted in at least two functionally redundant, nuclear-encoded copies, or more than two if either symbiotic partner possessed more than one copy of the given gene (Martin and Schnarrenberger, 1997).

For such enzymes of core carbohydrate metabolism, this functional redundancy was eliminated during evolution: one of the two genes (either that of the symbiont or that of the host) underwent loss. The encoded compartment-specific function was replaced through gene duplication of the persisting copy, whereby its product underwent evolutionary rerouting to the compartment requiring the activity, and duplication necessarily preceded loss of the eliminated copy. This general scenario has been observed for all chloroplast–cytosol isoenzymes

* Corresponding author. Tel: ++49 531 3915785; Fax: ++49 531 3915765; E-mail: w.martin@tu-bs.de

[1] The sequences reported in this paper have been deposited with GenBank under the accession numbers AJ000265 (spinach chloroplast GPI) and AJ000266 (spinach cytosolic GPI).

PII: S0378-1119(98)00229-7

of carbohydrate metabolism studied to date: glyceraldehyde-3-phosphate dehydrogenase (GAPDH), fructose-1,6-bisphosphatase (FBP), phosphoglycerate kinase (PGK), fructose-1,6-bisphosphate aldolase (FBA), and triosephosphate isomerase (TPI) (see Martin and Schnarrenberger, 1997, for an overview). The timing of the individual gene duplication and gene loss events were apparently not co-ordinated in any recognizable manner, since functional redundancy of host and plastid copies was eliminated for each enzyme independently and at very different times during the evolution of photosynthetic eukaryotes (Martin and Schnarrenberger, 1997).

Similar functional redundancy for enzymes of carbohydrate metabolism should also have existed during the origins of mitochondria. As in the scenario for plastids described above, both the mitochondrial symbiont and its host should have possessed more or less complete pathways of central carbohydrate metabolism, in which case endosymbiotic gene transfer resulted in functionally redundant copies in the host's chromosomes. A substantial body of molecular phylogenetic evidence from components of the genetic apparatus indicates that the host was a descendant of the archaebacteria (Iwabe et al., 1989; Edgell and Doolittle, 1997; Reeve et al., 1997). It is therefore all the more curious that the enzymes of core carbohydrate metabolism preserved in the cytosol of contemporary eukaryotes are of eubacterial ancestry (Martin and Schnarrenberger, 1997), also in some protists that have secondarily lost their mitochondria (Markos et al., 1993; Henze et al., 1995; Keeling and Doolittle, 1997). The simplest interpretation of these findings is that during the endosymbiotic origins of organelles, the elimination of functional redundancy in carbohydrate metabolism resulted in the preferential loss of the host's archaebacterial homologues (Martin and Schnarrenberger, 1997; Martin and Müller, 1998).

In the light of these considerations, we wished to address the evolutionary history of higher plant chloroplast and cytosolic glucose-6-phosphate isomerases (EC 5.3.1.9, GPI). GPI catalyses the reversible interconversion of glucose-6-phosphate and fructose-6-phosphate, and is thus an integral component of glycolysis and gluconeogenesis. These are cytosolic pathways in most eukaryotes, yet there are exceptions: in kinetoplastids (trypanosomes and relatives) glycolysis and GPI are compartmentalized in specialized microbodies, glycosomes (Marchand et al., 1989) and in higher plants, separable GPI isoenzymes exist in the chloroplast and cytosol (Schnarrenberger and Oeser, 1974) that are encoded by distinct nuclear genes. The chloroplast enzyme is involved in the synthesis and mobilization of assimilate in photosynthetic plastids (Schnarrenberger and Oeser, 1974), and in starch accumulation in nonphotosynthetic plastids (Plaxton, 1996). Clones for the cytosolic enzyme have been isolated from many eukaryotes, including plants (Thomas et al., 1993). Previous studies of GPI gene evolution have revealed that eukaryotic GPI genes share surprisingly high sequence identity to eubacterial homologues, in particular to the genes from *Escherichia coli* and related γ-proteobacteria. The currently accepted interpretation of that finding is that the common ancestor of *E. coli* (and its close relatives) have acquired their GPI genes via horizontal gene transfer from a eukaryotic source (Smith and Doolittle, 1992; Smith et al., 1992). Yet previous studies did not address the origin of the eukaryotic genes themselves that were postulated to have been donated to eubacteria and, as pointed out by Hattori et al. (1995), may have underestimated GPI gene diversity within prokaryotes. Here we report the purification, protein sequencing and cloning of chloroplast GPI from spinach leaves. We establish the identity of chloroplast GPI as a protein with only 30% amino acid identity to a sequence previously reported (Tait et al., 1988) as higher plant chloroplast GPI. The evolutionary history of eukaryotic GPI genes is investigated and reinterpreted in the context of prokaryotic gene diversity and endosymbiotic gene transfer.

2. Materials and methods

2.1. Protein purification

All steps were performed at 4°C unless otherwise indicated. About 1500 g of mature, deribbed spinach leaves were homogenized in buffer A (50 mM Tris–HCl (pH 8.5), 20 mM 2-mercaptoethanol, 100 g/l Polyklar AT) using a Waring blender, filtered through cheesecloth, and centrifuged for 40 min at $20000 \times g$. The 30–60% ammonium sulphate fraction of the supernatant was collected by centrifugation, dialysed against buffer B (10 mM Tris–HCl (pH 8.5), 20 mM 2-mercaptoethanol) to <3 μS/cm, and loaded onto a 3×13 cm DEAE Fractogel 650 S (Merck) column. The column was washed with 180 ml buffer B, proteins were eluted in a 180 ml gradient of 0–400 mM KCl in buffer B, fractions of 2.5 ml were collected. Two peaks of GPI activity were quantitatively separated: cytosolic GPI eluted at 50 mM KCl and cpGPI eluted at 190 mM KCl as previously described (Schnarrenberger and Oeser, 1974). Fractions containing cpGPI were pooled, dialysed against buffer C (10 mM potassium phosphate (pH 8.0), 20 mM 2-mercaptoethanol), and purified at 4°C on a 3×8.5 cm Biogel HTP (BioRad, Hercules, CA, USA) column. The flow through was applied at 20°C to a 1.6×8.5 cm Source 30Q (Pharmacia, Uppsala, Sweden) column. The column was washed with 35 ml buffer B, proteins were eluted in a 100 ml gradient of 0–400 mM KCl in buffer B. Fractions with GPI activity were dialysed against buffer D (10 mM Tris–HCl (pH 8.0), 1 M ammonium sulphate, 10 mM 2-mercaptoethanol) and loaded onto a 1.6 × 10 cm Octylsepharose 4 FF column (Pharmacia) equilibrated with buffer D. The column was washed with 40 ml buffer D, proteins were eluted at 20°C in a 50 ml gradient of 700–0 mM ammonium sulphate in buffer E (10 mM Tris–HCl (pH 8.0), 20 mM 2-mercaptoethanol). Fractions (0.2 ml each) with GPI activity were dialysed against buffer B and bound at 20°C to a MonoQ HR 5/5 (Pharmacia) column equilibrated in buffer B. The column was washed with 5 ml buffer B, proteins were eluted in a 12 ml gradient of 0–600 mM KCl in buffer B. Fractions with GPI activity were dialysed against buffer D and bound at 20°C to a Octylsepharose 4 FF HiTrap column (Pharmacia) equilibrated in buffer D. The column was washed with 5 ml buffer D, proteins were eluted in a 5 ml gradient of 500–0 mM ammonium sulphate in buffer E. Fractions with GPI activity were concentrated by ultrafiltration (Amicon, Beverley, MA, USA) to 60 μl, applied to a preparative 5.5 cm, 6% native polyacrylamide gel (Mini-Prepcell, BioRad), and electrophoresed at 300 V. Fractions of 250 μl were collected at 100 μl/min and assayed for GPI activity. This preparation was submitted to protein sequencing as described (Henze et al., 1994) both directly and after endopeptidase LysC digestion.

2.2. *Hybridization probe and cloning*

Messenger RNA and cDNA from 7-day-old, light grown spinach seedlings were prepared as described (Henze et al., 1994). PCR was performed for 35 cycles of 1 min 93°C, 1 min 45°C and 1 min 72°C in 25 μl of 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 0.5 mM $MgCl_2$, 0.05 mM of each dNTP, 0.02 U/μl Taq-Polymerase (Perkin Elmer, Foster City, CA, USA), 2 ng/μl spinach cDNA (lacking *Eco*/*Not* adaptors) and 0.8 μM each of the primers 5′-AARGAYATGGTNGTNYTNCCNTAYAA-3′ and 5′-TTRTTNCCRTCNARRTCRAAYTCYTT-3′ designed against the sequenced peptides (K)DMVVLPYK and (K)EFDLDGNK, respectively, obtained from purified chloroplast GPI. The resulting 109 bp amplification product was subcloned blunt into pBluecript SK+ (Stratagene), verified by sequencing, and used as a hybridization probe to screen 10^5 recombinant cDNA clones in Lambda ZAP II (Stratagene) as described (Henze et al., 1994). Nine independent positives were isolated and shown by sequencing to represent the same transcript. The sequence of one of the full-size clones (pcpGPI3) was determined using nested deletions.

A hybridization probe for cytosolic GPI was obtained by PCR as above using the primers 5′ TTYTGGGAYTGGGTIGGIG 3′ and 5′ TCIACICCCCAYTGRTCRAA 3′ designed against conserved regions of cytosolic GPI from various eukaryotes. The 741 bp amplification product was subcloned and used as a hybridization probe against 10^5 recombinant cDNA clones as described above. Five independent positives were isolated and shown by sequencing to represent the same transcript. The sequence of one of the full-size clones (pcyGPI14) was determined using nested deletions. Standard molecular techniques were performed as described (Sambrook et al., 1989).

2.3. *Other methods*

Sequences were extracted from GenBank, general data handling and sequence alignment (available upon request) was performed with the Wisconsin package (Genetics Computer Group, 1994). Phylogenetic analysis was performed with the maximum likelihood method (Adachi and Hasegawa, 1996a). The JTT-F substitution matrix (Adachi and Hasegawa, 1996b) was used in addition to a substitution matrix determined from the frequencies of 9958 amino acid positions in 45 protein-coding genes from plastids (J. Adachi and M. Hasegawa, unpublished). Both matrices gave very similar results, but the plastid subsitution matrix gave higher likelihoods (by 140 log-likelihood units for the 650 sites), suggesting that it better reflects the evolution of the genes under study here. Therefore, the results from the plastid matrix are reported. Enzyme activity was measured photometrically at 25°C in 1 ml of 25 mM Tris–HCl (pH 8.0), 7.5 mM $MgCl_2$, 1 U/ml glucose-6-phosphate dehydrogenase, 2 mM fructose-6-phosphate, and 170 μM $NADP^+$. One unit is the amount of enzyme that catalyses the fructose-6-phosphate-dependent oxidation of 1 μmol of NADPH in 1 min in the presence of glucose-6-phosphate dehydrogenase. Protein concentration was determined as described (Henze et al., 1994) using BSA as a standard.

3. Results

3.1. *Purification and cloning of spinach chloroplast GPI*

Chloroplast and cytosolic GPI from spinach were quantitatively separated by ion exchange chromatography. The fraction eluting at 190 mM KCl was previously shown to contain the chloroplast enzyme by virtue of its co-chromatography with the GPI activity found in isolated chloroplasts (Schnarrenberger and Oeser, 1974) and was purified further (Table 1). The final preparation of spinach chloroplast GPI contained 50 μg of electrophoretically homogeneous, 2700-fold purifed enzyme (Fig. 1) with a specific activity of 380 units per mg.

The N-terminal sequence of the purified protein from spinach chloroplasts and the sequence of three internal proteolytic fragments were determined (underlined in Fig. 2). The mature chloroplast subunit is preceded by

Table 1
Purification of chloroplast GPI from spinach

Purification step	Total activity (U)	Total protein (mg)	Specific activity (U/mg)	Purification (-fold)
Crude extract	754[a]	5253	0.14	—
DEAE Fractogel	274	450	0.60	4
Hydroxyapatite	259	216	1.21	8
Source 30Q	124	19.5	6.36	45
Octylsepharose	38	4.0	9.50	68
Mono Q	27	0.8	33.75	241
Octylsepharose	23	0.5	46.0	328
Native PAGE	19	0.05	380	2714

[a]Sum of activity of chloroplast and cytosolic GPI.

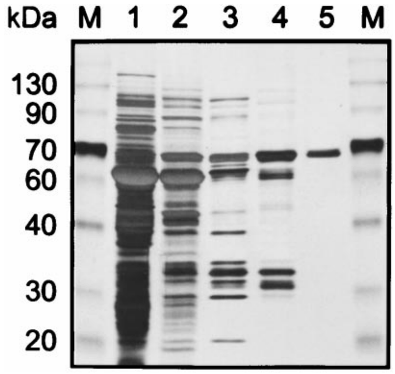

Fig. 1. SDS–PAGE of chloroplast GPI from spinach leaves from different steps of purification. Lane 1: crude extract (20 μg); lane 2, DEAE–Fractogel eluate; lane 3: source 30 Q eluate; lane 4: octylsepharose eluate; lane 5: preparative electrophoresis eluate (1 μg); M: molecular mass standard (sizes indicated).

a 60 amino acid transit peptide, the calculated M_r of the deduced amino acid sequence is 61 400 kDa, which is in reasonably good agreement with the M_r of 66 estimated from SDS–PAGE (Fig. 1). All four peptide sequences determined from the active, purified protein were found in the deduced amino acid sequence of pcpPGI3 (Fig. 2), establishing beyond doubt that pcpGPI3 contains the sequence of the authentic spinach chloroplast glucose-6-phosphate isomerase enzyme. This is notable, because an unconfirmed sequence for ‘chloroplast GPI’ from the angiosperm *Clarkia* was previously reported (Tait et al., 1988; accession number P11243), that differs markedly from the spinach chloroplast GPI sequence described here. Nuclear-encoded spinach chloroplast GPI shares 62% amino acid sequence identity with a sequence identified by similarity search as GPI in the genome of the cyanobacterium *Synechocystis* PCC6803 (Kaneko et al., 1996), but shares only about 30% amino acid identity with homologues from other sources, including the putative *Clarkia* chloroplast enzyme.

3.2. Eukaryotic GPI genes: branches on a eubacterial tree

Retrieval and alignment of GPI amino acid sequences revealed that three very distinct families of GPI proteins exist that share only about 30% sequence identity in between-family comparisons. For convenience, we arbitrarily designate these families as I, II, and III. ‘Family I’ contains genes encoding GPI enzymes described to date from the eukaryotic cytosol and from many eubacteria. GPI enzymes from several low-GC gram positive eubacteria and their homologue from the *Methanococcus jannaschii* genome (Bult et al., 1996) are assigned to family III. Family II currently consists only of the spinach chloroplast enzyme and its cyanobacterial homologue (Fig. 3).

Since these three families of GPI genes are highly divergent, several regions in the alignment are questionable (see also Fig. 2). To take this factor into account, phylogenetic analyses were performed in parallel both with the complete amino acid sequences (650 positions) and under exclusion of sites that contain gaps in one or more sequences in the alignment (336 positions). The PROTML analyses are summarized in Fig. 3. Fig. 3A shows the result with the complete alignment for the same species, Fig. 3B shows the ML result that is obtained when gapped sites are excluded. In both data sets, GPI enzymes belonging to families I, II, and III are robustly distinguished. Notably, eubacterial sequences occur in all three families. The branching orders for the GPI proteins in families II and III are the same for both data sets and possess no particularly surprising features.

However, within the family I GPI subtree, complex interleaving of eubacterial and eukaryotic GPI protein sequences is observed. For the complete data including highly variable sites (many of which are difficult to align), the deepest dichotomy within subtree I separates cytosolic GPI of plastid-bearing eukaryotes—to which *Plasmodium* belongs (McFadden et al., 1996)—from the

```
Zymomonas                                                             MARIANKAAIDAAWKQVSACSEKTLKQLFEEDSNRLSGLVVETAKLRFD
Spinach cytosol                                                     MAPSTLICDTDSWQNLKTHVAEIKKTHLRDLMSDADRCKSMMVEFDGLLLD
Methanococcus                                                                                                      M
Synechocystis                                                                                  MNNQQLWQRYQDWLYYHGGLDFYLD
Spinach chloroplast MASSLSNLYSSTSLKPKISHLPTITKNPISGPKSLSFKPISSVARDTPADLSTSSSSSTNNLPSLQKKKADGSLEKDPRALWARYVEWLYQHKDLGLYLD

Zymomonas           FSKNHLDSQKLTAFKKLLEACDFDARRKALFAGEKINITEDRAVEHMAERGQGAPASVARAKEYHA-------RMRTLIEAIDAGAFGE-----VKHLLH
Spinach cytosol     YSRQNATHDTMSKLFQLAEASHLKDKINQMFNGEHINSTENRSVLHVALRASRDAVINGDGKNVVPDVWQVLDKIRDFSEKIRSGSWVGVTGKPLTNVVA
Methanococcus       LS-YDYENALKVGEISLEDI------NKVDFANAYSNLMEKLD-----------NGVVGFRDVIYDENLDKYKSLNG-----------------YENVVV
Synechocystis       VSRMGFSDALVEDLQPKFAKA-FQDMVA-LEKGAIANPDEQRMVGHYWLR----NPALAPNDGIRAEITEPLRQIKAFVADVHQGNIKPPTAPKFTDLLA
Spinach chloroplast VSRIGFSDEFVKEMEPRFEKA-FKHMEG-IEKGAIANPDEGRMVGHYWLR----NSSLAPTTFLKNQIDVTLDRVWQFANDVISGKIKAPTGERFTHILS

Zymomonas           IGIGGSALGPKLLIDALTRESGRY------DVAVVSNVDGQALEEVFKKFNPH----KTLIAVASKTFTTAETMLNAESAMEWMKKHGVEDPQGRMIALT
Spinach cytosol     VGIGGSFLGPLFVHTALQTESEAAECAKGRQLRFLANVDPIDVAKNISGLNPE----TTLVVVVSKTFTTAETMLNARTLREWISS--ALGPAAVAKHMV
Methanococcus       IGMGGSILGTMAIYYAISPFNNNA--------YFIDNSDPEKTL-SILKKVDL---NESIIYIISKSGNTLETLVNYYLIKKRIEK--LNSFKGKLVFIT
Synechocystis       IGIGGSALGPQFVAQALAPNFPPL------AIHFIDNSDPDGIDRVLNCLKAQDKLKSTLVVTTSKSGGTPEPRNGLAETKAVFEAQ-GLHFADYAVAVT
Spinach chloroplast VGIGGSALGPQFVAEALAPDNPPL------KIRFIDNTDPAGIDHQIAQLGPE--LATTLVMVISKSGGTPETRNGLLEVQKAFRDA-GLVFAKQGVAIT
                     ♦  ♦♦  ♦                                                       ♦♦

Zymomonas           A---NPAKASEMGIDD-TRILPFAESIGGRYSLWSSIG-FPAALALGWEGFQQLLEGGAAMDRHFLEAAPEKNAPILAAFADQYYSAVRGAQTHGIFAYD
Spinach cytosol     AVSTNLTLVEKFGIDP-KNAFAFWDWVGGRYSVCSAVGVLPLSLQYGFPIVEKFLKGASSIDQHFHSAPLEKNLPVLLGLLSLWNVSFLGHPARAILPYC
Methanococcus       N---GGKLKREAEKNNY-DIFSIPENVPGRFSVFTAVGLAPLYSLGV--DISKILEGAREMDKICQNEDILKNPALLNGVIHYLY-DKRGKDISVIMSYV
Synechocystis       M--PGSKLSQQAQTEQWLQAFPMQDWVGGRTSELSAVGLLPAALQGI--DIQAMLDGAKTMDEATRVRELRQNPAALLALAWYYAGDGQGKKDMVILPYK
Spinach chloroplast Q--ENSLLDNTARIEGWIDRFPMFDWVGGRTSEMSAVGLLPAALQGI--DIKEMLAGAALMDEATKIPVLRSNPAALLAMSWYWASDGVGSKDMVVLPYK
                                                ♦♦       ♦                  ♦

Zymomonas           ERLQLLPFYLQQLEMESNGKRVDLDGNLIDHPSAFITWGGVGTDAQHAVFQLLHQGTRLVPIEFIAAIKA-----DDTLNPVHHKTLLTNAFAQGAALMS
Spinach cytosol     QALEKFAPHIQQVSMESNGKGVSIDGVVLPFEAGEIDFGEPGTNGQHSFYQLIHQG-RVIPCDFIGIAKSQQPVYLKGEVVSNHDELMSNFFAQPDALAY
Methanococcus       ESLKYFGDWYKQLIGESLGKN----------KHGITPLLSIGAKDQHSLLQLYMDGKKDKIITFMVAKKYRLDEEIEFEDIN----------DEKISCRY
Synechocystis       DRLLLFSRYLQQLVMESLGKERDLDGNVV--HQGIAVYGNKGSTDQHAYVQQLRDGVPNFFATFIEVLHDR-QGPSLELEPG----------VTSGDYLS
Spinach chloroplast DSLLLFSRYLQQLVMESLGKEFDLDGNKV--NQGLTVYGNKGSTDQHAYIQQLRDGVHNFFATFIEVLRDRPPGHDWELEPG----------VTCGDYLF
                               ♦   ♦♦ ♦♦

Zymomonas           GRDNKD-----------PARSYPGDRPSTTILMEELRPAQLGALIAFYEHRTFTNGVLLGINSFDQFGVELGKEMAHAIADHPENS--------DFDPST
Spinach cytosol     GKTQEELQKENISPHLVPHKTFTGNRPSLSLLLPSLTAYNVGQLLAIYEHRVAVEGFVWGINSFDQWGVELGKSLANQVRKQLHASRTNGEAVKGFNFST
Methanococcus       SDIIRSQQK------ATEIALTNNGVPNVRITLDEINEMAMGALLYMYEMQVGFMGELYNINAYNQPAVEEEKKICWRLIKQ*
Synechocystis       GFLQGTRQA-----------LFENQRDSITVTIPEVDATSVGALIALYERAVSFYGSLVNVNAYHQPGVEAGKKAAASILELQKAILSTLQNESG-----
Spinach chloroplast GMLQGTRSA-----------LYANNRESISVTVQEVTPRSVGAMVALYERAVGLYASLVNINAYHQPGVEAGKKAAAEVLALQKRVLAVLNEASCKDPVE
                                                             ♦                       ♦  ♦♦  ♦

Zymomonas           KALIAAALK*
Spinach cytosol     TTVMAKYLQETSDVPAELPTKLP*
Methanococcus
Synechocystis       PIALEALATKVQAPEQVETVYKIVRHLAANDRGVTLQGDRQFPQRLQIQWRS*
Spinach chloroplast PLTIEEVADHCHCPDDIEMIYKIIAHMAANDRVILAEGDCGSPRSIKAFLGECNVDELYA*
```

Fig. 2. Comparison of spinach chloroplast and cytosolic glucose-6-phosphate isomerases deduced from the sequence of the cDNA clone. Double-underlined regions indicate peptide sequences from spinach chloroplast GPI determined by microsequencing of the active, purified enzyme. The transit peptide of chloroplast GPI is indicated in italics, the processing site is inferred from the N-terminal sequence of the purified enzyme. Residues that are conserved in all GPI sequences analysed in this study (see Fig. 3) are indicated by '♦'. Gaps are indicated by dashes. The *Zymomonas, Methanococcus*, and *Synechocystis* sequences are shown to underscore the low degree of sequence similarity found across the three families genes distinguished here (see text).

enzyme of non-plastid-bearing eukaryotes and several eubacteria (Fig. 3A). The latter two groups of sequences are phylogenetically interleaved—due to the curious position of the *E. coli* and *Haemophilus* sequences (see below)—but they reveal two general patterns: (i) sequence diversity within eubacterial GPI sequences surveyed is greater than that observed across the eubacteria–eukaryote boundary; and (ii) the eukaryotic GPI sequences occur on branches of a eubacterial gene tree. Such patterns are generally characteristic for eukaryotic genes of eubacterial origin (Henze et al., 1995; Martin and Schnarrenberger, 1997; Doolittle, 1997).

The topology of GPI subtree I is quite different in the ML analysis of the more conservative core of 336 positions (excluding gapped sites, Fig. 3B). Again, sequence diversity within proteobacterial GPI sequences surveyed is greater than that observed across the eubacteria–eukaryote boundary, with all eukaryotic GPI sequences appearing on branches of a eubacterial gene tree. The branching of the *E. coli* and *Haemophilus* sequences with vertebrate homologues is unchanged, but greater interleaving is observed, since the *Mycobacterium* homologue shifts far up in the topology, branching with its homologues from plastid-bearing eukaryotes, which also move up in the topology relative to Fig. 3A.

3.3. The E. coli – Haemophilus *branch: horizontal transfer?*

The common branch of the *E. coli* and *Haemophilus* sequences with vertebrate GPI was of particular interest, since the unusually high similarity of *E. coli* PGI to eukaryotic homologues has been interpreted as evidence for horizontal transfer from eukaryotes to prokaryotes (Smith and Doolittle, 1992; Smith et al., 1992). However, the instability of the GPI topology with respect to the types of sites analysed raised questions concerning the strength of this position within eukaryotic sequences. Furthermore, the finding that eukaryotic GPI sequences in general tend to occur on branches of a eubacterial GPI gene tree suggests a eubacterial origin of eukaryotic GPI genes. Thus, to examine the *E. coli* (and *Haemophilus*) position further, alternative topologies were investigated.

A complete ML analysis of all topologies was not possible (26 genes, approx. 10^{30} trees). Therefore, we

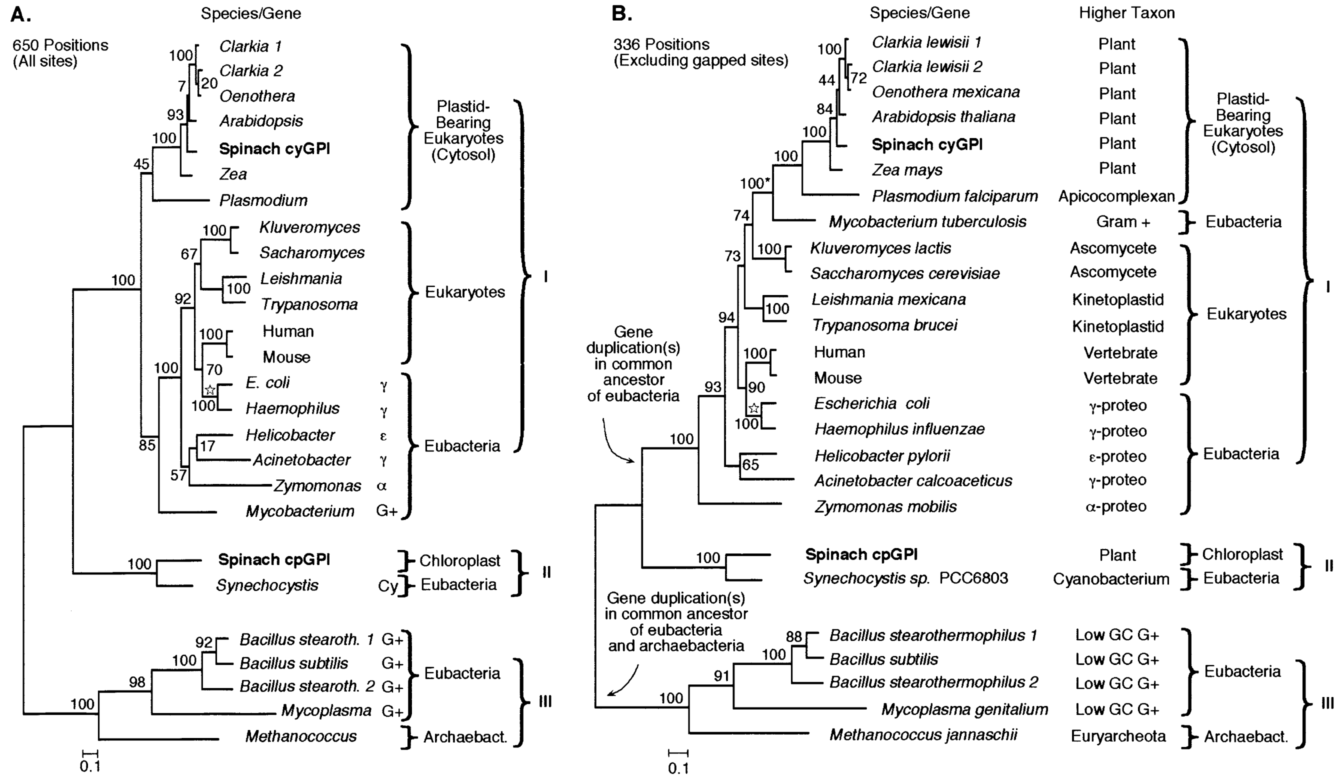

Fig. 3. Unrooted GPI gene phylogenies constructed with PROTML (Adachi and Hasegawa, 1996a,b). Local bootstrap values estimated by the RELL method (Adachi and Hasegawa, 1996b) are indicated next to branches. Scale bar at lower left indicates 0.1 substitutions per site. I, II, and III refer to the three familes of GPI genes distinguished here (see text). **(A)** Phylogeny inferred from the complete data set (including sites that possess a gap in one or more sequences). Abbreviated designations of higher taxa given in **(B)** are indicated next to species names. **(B)** Phylogeny inferred from the 336 sites of the alignment that possess no gaps in any sequence. Gram +, Gram positive eubacterium; proteo, proteobacterium; archaebact., archaebacterium. Gene duplications interpreted from the topology are indicated (see text). The sequence previously described as higher plant chloroplast GPI (32) (SWISSPROT P11243) is not included in the figure, since the source of the sequence is unclear (see text). It branches with *E. coli* and *Haemophilus* (indicated by a small asterisk). Accession numbers to sequences shown are *Acinetobacter calcoaceticus*, S58164; *Arabibopsis thaliana*, P34795; *Bacillus stearothermophilus*(1), P13375; *Bacillus stearothermophilus*(2), P13376; *Bacillus subtilis*, Z93936; *Clarkia lewisii*(1), P34796; *Clarkia lewisii*(2), P29333; *Escherichia coli*, P11537; *Haemophilus influenzae*, P44312; *Helicobacter pylorii*, HP1166 (obtained directly from http://www.tigr.org), human, P06744, *Kluyveromyces lactis*, P12341; *Leishmania mexicana*, P42861; *Zea mays*, P49105; *Methanococcus jannaschii*, D64500; mouse, P06745; *Mycobacterium tuberculosis*, 1524216; *Mycoplasma genitalium*, P47357; *Oenothera mexicana*, P54243; *Plasmodium falciparum*, P18240; *Synechocystis* PCC 6803, P52983; *Trypanosoma brucei*, P13377; *Saccharomyces cerevisiae*, P12709; *Zymomonas mobilis*, P28718.

employed local constraints, defining the following nine branches on the basis of results from Fig. 3: (1) plastid-bearing eukaryotes; (2) *Mycobacterium*; (3) *Kluveromyces, Saccharomyces*; (4) *Leishmania, Trypanosoma*; (5) human, mouse; (6) *E. coli, Haemophilus*; (7) *Helicobacter, Acinetobacter*; (8) *Zymonomas*; (9) GPI genes in families II and III. This left 135,135 possible trees to examine—still too large a number for an exhaustive ML search. The 135,135 trees were therefore examined by the approximate likelihood criterion, and the best 2000 candidates were selected. These 2000 topologies were tested by ML for both the 650-site data and the 336-site subset. The common branch ((*E. coli*, *Haemophilus*), (human, mouse)) received an average of 65% bootstrap proportion (BP) support across the 2000 trees for the complete data, and 80% BP for the 336-site data.

Importantly, a number of trees among the 2000 tested were found in both data sets, for which (i) the log-likelihood (lnL) was not significantly lower by one standard error than the lnL of the ML trees in Fig. 3; and (ii) the ((*E. coli*, *Haemophilus*), (human, mouse)) branch was disrupted. These trees are listed in Table 2. Thus, the ((*E. coli*, *Haemophilus*), (human, mouse)) branch occurs with reasonably high BP in the ML trees for both data sets, but it does not occur in many trees that are not significantly worse than the ML tree. Therefore, the GPI data do not provide any direct statistical support for the ((*E. coli*, *Haemophilus*), (human, mouse)) branch. As a consequence, arguments for lateral transfer of GPI genes from eukaryotes to prokaryotes on the basis of the weakly interleaving branching prokaryotic and eukaryotic GPI sequences in subtree I are not contradicted, but also are not supported at any level of significance by the data. The same instability is observed for the position of the

Table 2
Alternative topologies found with ML in the family I GPI tree that do not support common branching of (human/mouse) and (*E. coli* /*Haemophilus*) GPI genes

L[a]	Family I GPI topology[b]	ΔlnL ± SE[c]
650	((((1,M),7),(((3,4),H),E)),8)	11.1±16.1
	((1,M),((((3,4),H),E),(7,8)))	11.7±12.1
	((1,M),(((((3,4),H),E),8),7))	14.1±12.9
	((((1,M),7),((3,(4,H)),E)),8)	18.0±18.9
	(((1,M),((((3,4),H),E),7)),8)	11.0±15.4
	((((1,M),7),(((3,H),4),E)),8)	16.9±19.1
	((1,M),(((((3,4),H),E),7),8))	14.7±12.7
	(((1,M),(((3,(4,H)),E),7)),8)	17.9±18.4
	(((1,M),((((3,H),4),E),7)),8)	16.7±18.6
	((((1,M),(((3,4),H),E)),7),8)	9.1±15.5
	((((1,M),(((3,4),H),E)),8),7)	15.4±15.1
	(((1,M),(((3,4),H),E)),(7,8))	16.4±14.7
	(1,(M,(((((3,4),H),E),7),8)))	9.0±10.5
	((((1,M),((3,(4,H)),E)),7),8)	15.6±18.5
	(1,(M,(((((3,H),4),E),7),8)))	15.1±14.3
	(1,(M,(((((3,4),H),8),E),7)))	13.6±13.1
	(1,(M,(((((3,4),H),E),8),7)))	8.2±10.8
	(((((1,M),7),((3,4),H)),E),8)	17.9±18.8
	(1,(M,((((3,4),H),(E,8)),7)))	13.9±13.1
	(1,(M,((((3,4),H),E),(7,8))))	6.0±9.7
	(1,(M,((((3,(4,H)),E),8),7)))	15.9±14.1
	(1,(M,(((((3,H),4),E),8),7)))	14.4±14.5
	(1,(M,((((3,H),4),E),(7,8))))	12.3±13.6
	(1,(M,(((3,(4,H)),E),(7,8))))	13.5±13.3
336	((((1,M),7),(((3,4),H),E)),8)	18.1±15.7
	(((1,M),((((3,4),H),E),7)),8)	18.0±15.7
	((((1,M),(((3,4),H),E)),7),8)	14.6±14.9
	((((1,M),(((3,H),4),E)),7),8)	21.2±16.4
	(1,(M,(((((3,4),H),E),7),8)))	21.9±19.8
	((((1,M),((3,(4,H)),E)),7),8)	18.0±16.6
	(1,(M,(((((3,4),H),E),8),7)))	21.7±20.2
	(1,(M,((((3,4),H),E),(7,8))))	18.9±19.3
	(1,(M,(((3,(4,H)),E),(7,8))))	23.1±20.3
	(1,(M,((((3,H),4),E),(7,8))))	25.9±20.2

[a] Number of sites analysed, all sites (650) or excluding sites that possess gaps (336).

[b] Numbers designating defined groups are given in the text. M designates *Mycobacterium* GPI, E indicates the branch bearing *Escherichia* and *Haemophilus* GPI, H indicates the branch bearing human and mouse GPI. The remaining GPI sequences in Fig. 3 were used as the outgroups.

[c] Difference in log likelihood to the ML tree and one standard error of that difference.

Mycobacterium sequence in the GPI topology (Table 2). Thus, the branching pattern within the GPI subtree I is simply very unstable and currently unresolved.

4. Discussion

We have purified chloroplast glucose-6-phosphate isomerase from spinach chloroplasts to homogeneity, microsequenced the active protein and cloned the full-size cDNA that encodes that enzyme. Identity between all four sequenced peptides and the protein sequence deduced from the cDNA establish that pcpGPI13 encodes that authentic chloroplast enzyme. It shows 62% amino acid identity to GPI from the cyanobacterium *Synechocystis* PCC6803, but only 32% identity with a sequence previously described as higher plant chloroplast GPI (Tait et al., 1988; SwissProt accession number P11243, see legend to Fig. 3). The surprisingly high sequence identity (88%) between P11243 and *E. coli* GPI (Froman et al., 1989) had previously cast doubt upon the authenticity of P11243 (Thomas et al., 1992).

Our findings firmly establish the identity of higher plant chloroplast GPI as a nuclear-encoded descendant of cyanobacterial GPI. Notably, this phylogenetic result was predicted many years ago on the basis of immunological studies (Weeden et al., 1982). Furthermore, our findings suggest that the putative clone for *Clarkia* chloroplast GPI (P11243), which was cloned through complementation of *E. coli* mutants (Tait et al., 1988), represents a prokaryotic gene from an unknown eubacterial source, rather than the higher plant nuclear gene for the chloroplast enzyme. As a consequence, we suggest that arguments previously forwarded for eukaryote-to-prokaryote horizontal gene transfer on the basis of sequence similarity between P11243 and prokaryotic homologues (Smith and Doolittle, 1992; Smith et al., 1992) were forwarded soundly and in good conscience, but apparently on the basis of invalid data. More recently, horizontal transfer of GPI genes from eukaryotes to prokaryotes was argued independently of P11243, on the basis of the position of the *E. coli–Haemophilus* branch within the tree of eukaryotic sequences designated here as subtree I (Katz, 1996). The ML analysis (Fig. 3 and Table 2) indicate that the data resolve neither the position of that branch, nor the position of the *Mycobacterium* sequence with any statistical significance. Thus, we conclude that GPI genes do not harbor direct evidence for horiziontal gene transfer from eukaryotes to prokaryotes.

On the basis of previous analyses of the evolution of chloroplast–cytosol isoenzymes viewed in the context of endosymbiosis (Martin and Schnarrenberger, 1997), we interpret the overall picture of GPI gene evolution obtained here as indicating four general points.

(1) The gene for nuclear encoded chloroplast GPI was obtained by plants (represented by spinach) via endosymbiotic gene transfer from the cyanobacterial antecedants of chloroplasts.

(2) Nuclear genes for cytosolic (and glycosomal) GPI of non-photosynthetic eukaryotes were obtained via endosymbiotic gene transfer, but from the antecedants of mitochondria (Martin and Müller, 1998).

(3) Ancient (prokaryotic and eukaryotic) gene diversity of GPI genes can account as easily for much of the confusion in GPI gene evolution as horizontal transfer can (Hattori et al., 1995).

(4) GPI gene phylogeny within subtree I is extremely unstable and difficult to resolve, which is most easily explained as a manifestation of the inability of individual genes to accurately reflect ancient evolutionary processes due to the limitations of phylogenetic information contained within any individual gene (Martin et al., 1998).

GPI is a prominent example, but it is not the only gene suspected of involvement in transkingdom horizontal gene transfers outside the context of endosymbiosis. Other suspected cases included GAPDH, FBA, and glutamine synthase (GS) (Smith et al., 1992). Later studies with larger samples of prokaryotic and eukaryotic genes showed that the topologies which suggested the possibility of outright horizontal transfer for these three genes could easily be explained on the basis of other premises. In the case of GAPDH, the unexpected similarity of the prokaryotic and eukaryotic homologues is easily attributable to the (unexpected) eubacterial (endosymbiotic) origin of the eukaryotic nuclear genes for these cytosolic enzymes (Henze et al., 1995; Martin and Schnarrenberger, 1997). Much the same applies for FBA (Plaumann et al., 1997). In the case of GS, the suspected lateral transfer was readily explained by ancient paralogy revealed through more extensive sampling of eubacterial gene diversity (Kumada et al., 1993). All of these factors can easily mimic horizontal transfer (Martin and Schnarrenberger, 1997; Doolittle, 1997).

Previous studies of GPI gene evolution among prokaryotes and eukaryotes have been conducted on the basis of very sparse prokaryotic lineage samples. The phylogenies in Fig. 3 still embrace a very sparse sample, but the inclusion of a cyanobacterial, a plastid, and an archaebacterial protein reveal the existence of prokaryotic GPI gene diversity which far exceeds that previously recognized. This diversity is reflected in the skew distribution of three highly divergent (approx. 30% identity) families of GPI proteins across eubacterial lineages. Notwithstanding the uncertainties attached to the gene phylogeny itself, this could be interpreted—in principle—as evidence reflecting either of two phenomena. It could be that the common ancestor of archaebacteria and eubacteria sampled possessed all three families of GPI genes, that these were subject to differential loss in various lineages, and that eukaryotes inherited only a fraction of that ancient diversity during the course of endosymbiosis. Alternatively, it is possible that some degree of transfer of GPI genes between eubacteria has taken place in evolution, mimicking ancient diversity. Horizontal gene transfer between eubacteria is very common and is very well documented (Matic et al., 1995; Vulic et al., 1997). Obviously, a combination of these two factors (ancient gene diversity and lateral transfer between eubacteria) may be causal to the 'non-rRNA-like' picture of bacterial evolution provided by GPI gene phylogeny, and the problem of distingiushing between these two factors from the standpoint of the current distribution of genes across prokaryotic genomes is not trivial.

How severe is the general phylogenetic problem of ancient gene diversity in prokaryotes and its loss over time to the present? In their analysis of the *E. coli* genome, Blattner et al. (1997) provided some benchmark figures that help to outline the magnitude of the problem. Using the arbitrary threshold of 30% identity, only 111 protein-coding genes are common to the *E. coli*, *Haemophilus*, *Mycoplasma* and *Synechocystis* genomes (Blattner et al., 1997). However, using the same threshold, only 16 protein-coding genes (Blattner et al., 1997) are shared by those four and *Methanococcus* and yeast. Blattner et al. concluded that such findings are indicative of numerous gene losses over the course of genome evolution, an interpretation with which we generally agree. By inference, ancestral prokaryotic genomes would have possessed a much larger number of genes per genome than contemporary prokaryotes do, in order to have been able to endure such losses. Notwithstanding lateral gene transfer between eubacteria (see above), that general view is consistent with our interpretation that ancient prokaryotic diversity of GPI genes, differential gene loss in independent prokaryotic lineages (Blattner et al., 1997), sampling of eubacterial gene diversity through endosymbiosis (Martin and Schnarrenberger, 1997), elimination of functional redundancy following endosymbiotic gene transfer (Martin and Müller, 1998), gene transfer between eubacteria (Matic et al., 1995) and the phylogenetic limitations of individual genes (Martin et al., 1998) are the factors that result in phylogenetic interleaving of eubacterial and eukaryotic GPI genes, not horizontal gene transfer from eukaryotes to prokaryotes.

Acknowledgement

We thank Christiane Lindemann and Renate Grübnau for excellent technical assistance and Rainer Figge for critical reading of the manuscript. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged (Ma 1425/3-2 and Schn 28/21-2).

References

Adachi, J., Hasegawa, M., 1996a. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. 42, 459–468.

Adachi, J., Hasegawa, M., 1996b. MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. Computer Science Monographs, No. 28. Institute of Statistical Mathematics, Tokyo.

Blattner, F.R. et al., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1474.

Bult, C.J. et al., 1996. Complete genome sequence of the methanogenic Archeon *Methanococcus jannaschii*. Science 273, 1058–1073.

Doolittle, W.F., 1997. Fun with genealogy. Proc. Natl. Acad. Sci. USA 94, 12751–12753.

Edgell, D.R., Doolittle, W.F., 1997. Archaea and the origin(s) of DNA replication proteins. Cell 89, 995–998.

Froman, B.E., Tait, R.C., Gottlieb, L.D., 1989. Isolation and characterization of the phosphoglucose isomerase gene from *Escherichia coli*. Mol. Gen. Genet. 217, 126–131.

Hattori, J., Baum, B.R., Miki, B.L., 1995. Ancient diversity of the glucose-6-phosphate isomerase genes. Biochem. Syst. Ecol. 23, 33–38.

Henze, K., Schnarrenberger, C., Kellermann, J., Martin, W., 1994. Chloroplast and cytosolic triosephosphate isomerase from spinach: Purification, microsequencing and cDNA sequence of the chloroplast enzyme. Plant Mol. Biol. 26, 1961–1973.

Henze, K., Badr, A., Wettern, M., Cerff, R., 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.

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S., Miyata, T., 1989. Evolutionary relationship of archaebacteria, eubacteria and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86, 9355–9359.

Kaneko, T. et al., Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. 1996. DNA Res. 3, 109–136.

Katz, L., 1996. Transkingdom transfer of the phosphoglucose isomerase gene. J. Mol. Evol. 43, 453–459.

Keeling, P.J., Doolittle, W.F., 1997. Eukaryotic triosephosphate isomerase originated in the mitochondrion. Proc. Natl. Acad. Sci. USA 94, 1270–1275.

Kumada, Y., Benson, D.R., Hillemann, D., Hosted, T.J., Rochefort, D.A., Thompson, C.J., Wohlleben, W., Tateno, Y., 1993. Evolution of the glutamin synthase gene, one of the oldest existing and functioning genes. Proc. Natl. Acad. Sci. USA 90, 3009–3013.

Marchand, M., Kooystra, U., Wierenga, R.K., Lambeir, A.M., Van Beeumen, J., Opperdoes, F.R., Michels, P.A., 1989. Glucosephosphate isomerase from *Trypanosoma brucei*. Cloning and characterization of the gene and analysis of the enzyme. Eur. J. Biochem. 184, 455–464.

Markos, A., Miretsky, A., Müller, M., 1993. A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote *Trichomonas vaginalis*. J. Mol. Evol. 37, 631–643.

Martin, W., Müller, M., 1998. The hydrogen hypothesis for the first eukaryote. Nature 392, 37–41.

Martin, W., Schnarrenberger, C., 1997. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redendancy in ancient pathways through symbiosis. Curr. Genet. 32, 1–18.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., Kowallik, K.V., 1998. Gene transfer to the nucleus and the evolution of chloroplasts.. Nature. 393, 162–165.

Matic, I., Rayssiguier, C., Radman, M., 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. Cell 80, 507–515.

McFadden, G.I., Reith, M.E., Munholland, J., Lang-Unnasch, N., 1996. Plastid in human parasites. Nature 381, 482

Plaumann, M., Pelzer-Reith, B., Martin, W., Schnarrenberger, C., 1997. Cloning of fructose-1,6-bisphosphate aldolases from *Euglena gracilis*: multiple recruitment of class I aldolase to chloroplasts and eubacterial origin of eukaryotic class II aldolase genes. Curr. Genet. 31, 430–438.

Plaxton, W.C., 1996. The organization and regulation of plant glycolysis. Ann. Rev. Plant Physiol. Plant Mol. Biol. 47, 185–214.

Reeve, J.N., Sandman, K., Daniels, C.J., 1997. Archaeal histones, nucleosomes and transcription inititation. Cell 89, 999–1002.

Sambrook, J., Fritsch, E., Maniatis, T., 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, New York.

Schnarrenberger, C., Oeser, A., 1974. Two isoenzymes of glucosephosphate isomerase from spinach leaves and their intracellular compartmentation. Eur. J. Biochem. 45, 77–82.

Smith, M.W., Doolittle, R.F., 1992. Anomalous phylogeny involving the enzyme glucose-6-phosphate isomerase. J. Mol. Evol. 34, 544–545.

Smith, M.W., Feng, D.-F., Doolittle, R.F., 1992. Evolution by acquisition: the case for horizontal gene transfers. Trends Bioch. Sci. 17, 489–493.

Tait, R.C., Froman, B.E., Laudencia-Chingcuanco, D.L., Gottlieb, L.D., 1988. Plant phosphoglucose isomerase genes lack introns and are expressed in *Escherichia coli*. Plant Mol. Biol. 11, 381–388.

Thomas, B.R., Laudencia-Chingcuanco, D., Gottlieb, L.D., 1992. Molecular analysis of the plant gene encoding cytosolic phosphoglucose isomerase. Plant Mol. Biol. 19, 745–757.

Thomas, B.R., Ford, V.S., Pichersky, E., Gottlieb, L.D., 1993. Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in *Clarkia* and comparison to the single gene in *Arabidopsis*. Genetics 135, 895–905.

Vulic, M., Dionisio, F., Taddei, F., Radman, M., 1997. Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. Proc. Natl. Acad. Sci. USA 94, 9763–9767.

Weeden, N.F., Higgins, R.C., Gottlieb, L.D., 1982. Immunological similarity between a cyanobacterial enzyme and a nuclear DNA-encoded plastid-specific isoenzyme from spinach. Proc. Natl. Acad. Sci. USA 79, 5953–5955.