Curr Genet (1997) 32: 1–18

REVIEW

William Martin · Claus Schnarrenberger

The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis

Received: 8 January / 20 May 1997

Abstract The evolutionary histories of the 12 enzymes that catalyze the reactions of the Calvin cycle in higher-plant chloroplasts are summarized. They are shown to be encoded by a mixture of nuclear genes of cyanobacterial and proteobacterial origin. Moreover, where cytosolic isoforms of these enzymes are found they are almost invariably encoded by genes of clearly endosymbiont origin. We infer that endosymbiosis resulted in functional redundancy that was eliminated through differential gene loss, with intruding eubacterial genes repeatedly replacing pre-existing nuclear counterparts to which they were either functionally or structurally homologous. Our findings fail to support the 'product-specificity corollary', which predicts re-targeting of nuclear-encoded gene products to the organelle from whose genome they originated. Rather it would appear that the enzymes of central carbohydrate metabolism have evolved novel targeting possibilities regardless of their origins. Our findings suggest a new hypothesis to explain organelle genome persistence, based on the testable idea that some organelle-encoded gene products might be toxic when present in the cytosol or other inappropriate cellular compartments.

Key words Chloroplast · Mitochondria · Endosymbiosis · Endosymbiotic gene transfer · Calvin cycle · Glycolysis · Evolution · Amitochondriate · Metabolism · Compartmentation · Hydrogenosome · Eukaryote · Origin

W. Martin (✉)
Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany

C. Schnarrenberger
Institut für Pflanzenphysiologie und Mikrobiologie, Freie Universität Berlin, Königin-Luise-Strasse 12–16a, 14195 Berlin, Germany

Communicated by H. T. Jacobs

Introduction

The Calvin cycle (reductive pentose phosphate, photosynthetic carbon reduction, Calvin-Benson-Bassham, Benson-Calvin, or CBB cycle) is the major route through which carbon enters metabolism in many eubacteria. It is still the only pathway of CO_2 fixation known in eukaryotes (Calvin 1956). Several reviews covering various aspects of the Calvin cycle in higher plants have appeared in recent years, including general biochemistry (Macdonald and Buchanan 1990), regulation of enzyme activity by light through the ferredoxin/thioredoxin system (Woodrow and Berry 1988; Scheibe 1990; Buchanan 1991; Wolosiuk et al. 1993), diurnal regulation (Geiger and Servaites 1994), and C4 photosynthesis (Nelson and Langdale 1992). Rubisco itself has been the subject of several recent reviews (Portis 1992; Spreitzer 1993; Gutteridge and Gatenby 1995).

In higher plants and photosynthetic protists, the enzymes of the cycle are localized in the stroma of plastids, where they catalyze light-driven CO_2 fixation just as in eubacteria. The reactions that constitute the Calvin cycle in spinach are summarized in Fig. 1. CO_2 fixed in chloroplasts as triose phosphate is exported in many plants, primarily as glyceraldehyde-3-phosphate (GA3P), to the cytosol in exchange for phosphate via the phosphate translocator (Stitt 1990; Flügge and Heldt 1991), the major protein of the chloroplast envelope (Flügge et al. 1991). In the cytosol, GA3P can *inter alia* enter the glycolytic pathway for ATP synthesis in mitochondria via oxidative phosphorylation, or enter the gluconeogenetic pathway for the synthesis of transport forms such as sucrose to meet the energy and storage needs of cells in other organs and tissues. Of the 11 enzymes involved in the cycle in spinach plastids, five have isoforms occurring in the glycolytic and gluconeogenetic pathways of the cytosol. The scheme of compartmentation in Fig. 1 is only valid for spinach leaves (Schnarrenberger et al. 1995; Martin et al. 1996a). It does not summarize data concerning the complete and partial glycolytic pathways in plastids from other sources and tissues (reviewed by Plaxton 1996). It can also not be gener-

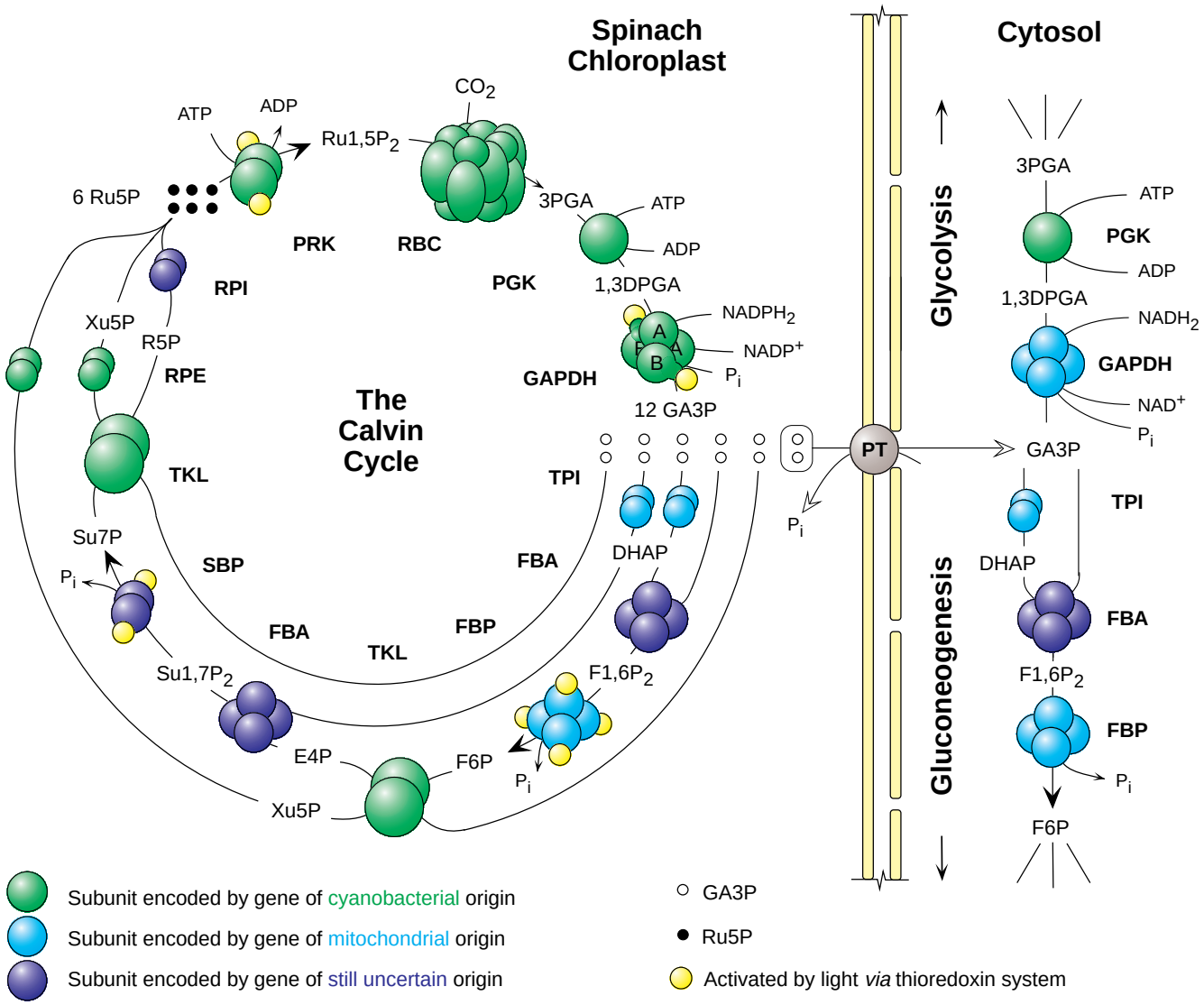

Figure 1 Enzymes of the Calvin cycle in spinach chloroplasts and their cytosolic homologues. Enzymes regulated through the thioredoxin system are indicated. Suggested evolutionary origins for the nuclear genes are *color-coded.* Enzyme abbreviations are: *FBA* fructose-1,6-bisphosphate aldolase; *FBP* fructose-1,6-bisphosphatase; *GAPDH* glyceraldehyde-3-phosphate dehydrogenase; *PGK* 3-phosphoglycerate kinase; *PRI* ribose-5-phosphate isomerase; *PRK* phosphoribulokinase; *RBC* ribulose-1,5-bisphosphate carboxylase/oxygenase; *RPE* ribulose-5-phosphate 3-epimerase; *SBP* sedoheptulose-1,7-bisphosphatase; *TKL* transketolase; *TPI* triosephosphate isomerase; *PT* phosphate translocator. Substrate/product abbreviations are: *Ru1,5P*$_2$ ribulose-1,5-bisphosphate; *3PGA* 3-phosphoglycerate; *1,3DPGA* 1,3-diphosphoglycerate; *GA3P* glyceraldehyde-3-phosphate; *DHAP* dihydroxyacetone phosphate; *F1,6P*$_2$ fructose-1,6-bisphosphate; F6P fructose-6-phosphate; *E4P* erythrose-4-phosphate; *Xu5P* xylulose-5-phosphate; *Su1,7P*$_2$ sedoheptulose-1,7-bisphosphate; *Su7P* sedoheptulose-7-phosphate; *R5P* ribose-5-phosphate; *Ru5P* ribulose-5-phosphate. *Open arrowheads* indicate transport rather than conversion. *Solid arrowheads* indicate physiologically irreversible reactions

alized to other plants, because compartmentation can vary dramatically across species. *Chlamydomonas*, for example, does not possess a glycolytic pathway in the cytosol; rather it occurs only in plastids (Schnarrenberger et al. 1990, 1994).

The pathway has been studied extensively in several photosynthetic eubacteria including *Rhodobacter sphaeroides* (Gibson et al. 1990, 1991; Qian and Tabita 1996), *Alcaligenes eutrophus* (Kusian et al. 1992, 1995; Schäferjohann et al. 1995; Bommer et al. 1996) and *Xanthobacter flavus* (Meijer et al. 1996). These organisms encode the enzymes of the pathway in *cbb* operons that vary in structure across species, and appear to be under the common regulation of the *cbbR* gene product, a member of the LysR family of transcriptional activators (Windhövel and Bowien 1991; Gibson and Tabita 1993). In the completely sequenced genome of the cyanobacterium *Synechocystis* PCC6803 (Kaneko et al. 1996) this operon organization is not conserved at all, the genes for Calvin enzymes being dispersed throughout the chromosome. Whether or not

cyanobacterial *cbb* genes constitute a regulon is unknown (Tabita 1994; Gibson and Tabita 1996), though *Synechocystis* notably does encode two homologues of *cbbR*. The genetics of the Calvin cycle and its regulation in eubacteria have been reviewed by Tabita (1994) and Gibson and Tabita (1996).

A complete Calvin cycle has not been described for any member of the archaebacteria, where the reverse citric acid cycle, the acetyl-CoA pathway, and the 3-hydroxypropionate pathway are used for carbon fixation (Beh et al. 1993; Schönheit and Schäfer 1995; Ishii et al. 1996). There have been reports of key Calvin-cycle enzymes in halophilic archaebacteria (Altekar and Rajagopalan 1990; Rajagopalan and Altekar 1994), and the recently sequenced *Methanococcus jannaschii* genome (Bult et al. 1996) contains a homologue of the large subunit of Rubisco that is thought to be specific for the pathway. These are interesting findings, but there is currently insufficient evidence to conclude whether some archaebacteria may have a complete, a partial, or a modified Calvin cycle.

In several previous contributions we have reported on the evolutionary history of individual higher-plant Calvin-cycle enzymes in the context of the endosymbiotic theory. The evolution of the pathway as a whole has not been previously addressed. The purpose of the present review is to summarize as briefly as possible the principles that have become evident from an investigation of the evolution of eukaryotic Calvin-cycle enzymes in the context of endosymbiotic theory.

Endosymbiotic gene transfer and the product-specificity corollary

The Calvin cycle is a chloroplast pathway, but genes for only one of its enzymes, Rubisco, can be found in plastid DNA, the remainder being encoded in the nucleus. A general working hypothesis exists for the evolution of nuclear-encoded genes for chloroplast enzymes. It is known as the product-specificity corollary to the endosymbiotic theory. It was explicitly and elegantly formulated by Weeden (1981). Due to its robustness, it has undergone only minor modification over the years to account for newer findings (Martin et al. 1990, 1993a). It postulates that the genes for chloroplast-specific enzymes were transferred from the genome of the cyanobacterial progenitor of plastids to the nucleus during the course of evolution. Once there, they are thought to have come under the regulatory hierarchy of the nucleus and to have acquired a transit peptide, with the help of which the encoded proteins have been re-imported into the organelle of their genetic origin (product specificity) ever since. The product-specificity corollary thus predicted that nuclear-encoded chloroplast proteins should reflect a cyanobacterial origin. It was formulated, however, on the basis of evidence from chloroplast-cytosol isoenzymes, and it carries one further important and often neglected prediction, namely that the cytosolic isoenzymes should reflect the history of the host cell under the endosymbiotic theory. Numerous cases of nuclear-encoded chloroplast proteins have been reported that confirm the first prediction, but the second prediction remained untestable for many years. This is because any test of the product-specificity hypothesis requires not only (1) an internal reference system, i.e. a cytosolic isoenzyme against which to contrast the evolutionary history of the chloroplast enzyme, but also (2) a null hypothesis for the expected evolutionary origin of both proteins. For chloroplast proteins, the expected origin is obvious: cyanobacteria. For cytosolic proteins, the expected origin is not at all obvious (Martin 1996, 1997), a topic that is considered in the section that follows.

What are the origins of the nucleus and cytosol under the endosymbiotic theory?

There is still no generally accepted answer to this question, and views on the evolutionary history of the eukaryotic nucleocytoplasm are in a state of flux. The prevalent view of the 1980s, that three independent urkingdoms of life, archaebacteria, eubacteria and eukaryotes, descended directly from the last common ancestor, is gradually giving way to the view that the core genetic apparatus (proteins of transcription and translation) in archaebacteria and eukaryotes share a more recent common ancestor that either does with eubacteria (Doolittle and Brown 1994; Brown and Doolittle 1995; Doolittle 1996). This notion can be traced to molecular phylogenetic studies of archaebacterial RNA polymerases, translational elongation factors and other proteins, carried out during the late 1980s (Gogarten et al. 1989; Iwabe et al. 1989; Pühler et al. 1989; Zillig et al. 1989; Zillig 1991), and is supported by more recent findings (Keeling and Doolittle 1995; Langer et al. 1995; Baldauf et al. 1996). If this notion is true, then it may be asked whether "primitive" eukaryotes that lack mitochondria are properly regarded as direct descendants of archaebacterial ancestors that never harbored organellar symbionts (Cavalier-Smith 1993). If so, they can be assumed to represent a "pristine" eukaryotic nuclear genome as the standard for phylogenetic comparison. This simple proposition is probably untrue, however, because cases are rapidly becoming known in which (by the measure of rRNA phylogeny) 'early branching' amitochondriate eukaryotes have been found to possess nuclear genes that are very likely of mitochondrial origin (Clark and Roger 1995; Henze et al. 1995; Keeling and Doolittle 1997). More importantly, data are accumulating which indicate a common evolutionary origin of mitochondria and hydrogenosomes (Bui et al. 1996; Germot et al. 1996; Horner et al. 1996; Roger et al. 1996), the double-membrane bound, DNA-lacking, 'powerhouse' of many amitochondriate eukaryotes[1] (Müller 1973, 1988, 1993, 1996).

[1] Hydrogenosomes (discovered by Müller) are the only double-membrane-bound organelles known in eukaryotes that lack DNA and are thus highly relevant to the endosymbiotic theory. They are found in

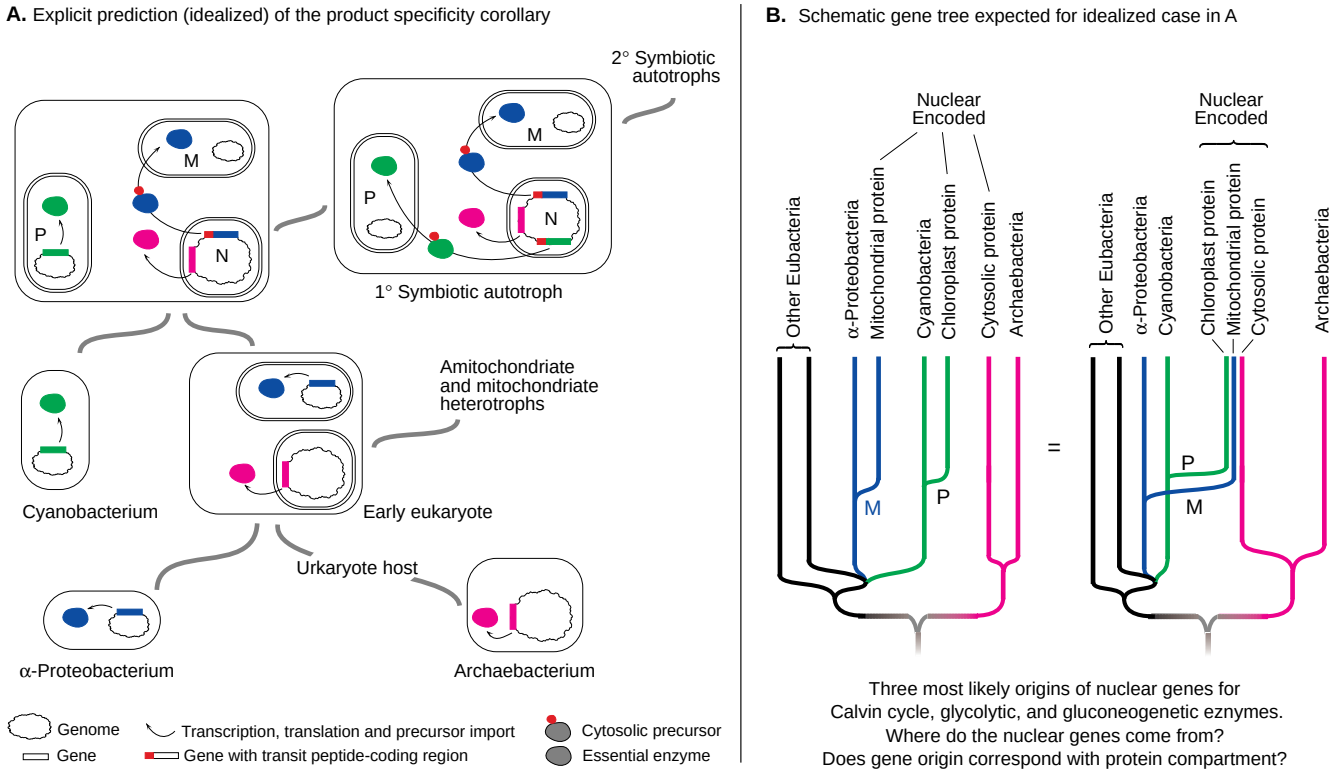

Figure 2 A, B An explicit formulation of the product-specificity corollary under the working hypothesis that the nuclear lineage of eukaryotes is of archaebacterial ancestry. **A** an idealized case of the product-specificity corollary for an enzyme common to all participants in the (primary) endosymbiotic series leading to photosynthetic eukaryotes. *N*, nucleus; *P*, plastids; *M*, mitochondria. The figure is intended to underscore redundancy of gene products as a result of endosymbiosis under the working hypothesis (see text). No attempt is made to explain the origin of membranes. It is not suggested that this scenario of evolutionary compartmentation is correct. The figure merely states in explicit terms the assumption that is implicitly accepted in many formulations of the symbiont hypothesis. Weeden's product-specificity corollary has been very robust for explaining the evolution of most chloroplast proteins. But its predictions fail for cytosolic proteins (see text) **B** schematic gene phylogeny expected for the idealized case in **A**

Such findings lend strength to the salient argument (Zillig 1989; Zillig et al. 1991) that the eukaryotic nucleus may have been a hybrid of archaebacterial and eubacterial genes from the very beginning. However, in contrast to Zillig's original view that a 'fusion' event between a eubacterium and an archaebacterium gave rise to a nucleus-containing eukaryote ancestor *prior* to mitochondrial origins (favored by Gupta and Golding 1996), current data would no longer require as a corollary the assumption that the presumed archaebacterium which associated with the mitochondrial/hydrogenosomal symbiont possessed eubacterial genes prior to that event. This view is still quite speculative and almost certainly too simple to be correct, but despite its simplicity it can account for a variety of otherwise puzzling findings, including the phylogenetic distribution of hydrogenosomes among eukaryotic crown groups (Müller 1996), much data from comparative genomics (Doolittle 1996), and complex phylogenies of nuclear genes for enzymes of carbohydrate metabolism (Martin 1996).

Thus, for the sake of structuring an unwieldy problem, the oversimplified view is embraced in this paper (as a working hypothesis) that eukaryotic nuclear genes may be *either* archaebacterial *or* eubacterial. If eubacterial, then they may be *either* mitochondrial (α-proteobacterial) *or* plastidial (cyanobacterial) in origin. This simple working hypothesis for the origin of nuclear genes and endosymbiotic gene transfer with re-import of the encoded protein to the compartment that donated the gene is summarized in Fig. 2 *for a protein that was common to all prokaryotes that participated in the evolution of the eukaryotic cell*, as is the case for several enzymes common to the Calvin cycle, glycolysis, and gluconeogenesis. It is obviously much too simple to be anywhere close to correct, but it *is* an explicit hypothesis for the evolutionary origin of nuclear genes for mitochondrial, chloroplast *and cytosolic* proteins, something that we have not had in the past. It merely takes a widely – and often implicity – accepted assumption under endosymbiotic theory (contemporary compart-

amitochondriate eukaryotes, both in early branching protists and in secondarily amitochondriate fungi and ciliates. As a rule they import pyruvate from cytosolic glycolysis and from it produce ATP, acetate, CO_2 and H_2; but variations on that theme are known. They occur only in amitochondriate protists with Type-II extended glycolysis, whereas amitochondriate protists that possess no hydrogenosomes use Type-I extended glycolysis. Both Type-I and Type-II extended glycolysis differ markedly from the classical glycolytic pathway of mitochondriate eukaryotes (Müller 1973, 1988, 1993, 1996)

mentation of proteins in the cell should reflect their ancestral compartmentation *sensu* Weeden) and extends it to the cytosol, taking into account findings that implicate archaebacteria as the source of several nuclear and cytoplasmic proteins of the genetic apparatus. Figure 2 is not an answer to a question, rather it is the simplest explicit formulation of a question that was impossible to ask in the absence of a good theory for the origin of the nucleocytoplasmic component of eukaryotes: *what do we expect the evolution of cytosolic enzymes to reflect under endosymbiotic theory?*

The foregoing considerations contain many unknowns, and the hypothesis in Fig. 2 (which makes no attempt to account for the history of eukaryotic membrane systems) is certainly debatable. But without such an explicit structure, the gene phylogenies that have been observed for Calvin-cycle enzymes and their cytosolic homologues would be altogether uninterpretable. If the working hypothesis for nuclear origins in Fig. 2 is completely incorrect, many of the arguments in this paper fail and our findings must then be re-interpreted on the basis of other premisses. However, as we will see, Fig. 2 can account fully for the origins of the nuclear genes considered here but, importantly, the product-specificity corollary cannot account for the compartmentation of the encoded gene products. The product-specificity corollary will be shown to fail in a sufficient number of specific examples of functional redundancy through symbiosis to suggest that it should be abandoned as the *a priori* expectation for the evolution of the compartmentation of enzymes that were common to symbionts and hosts.

So why draw a picture of a 16-year-old hypothesis only to dismantle it? Simply because many aspects of Weeden's elegant hypothesis were so thoroughly argued and are still so compelling that it is widely *assumed* to be correct (e.g. Palmer, 1997). It is deeply engrained in current views of eukaryotic evolution and endosymbiosis, though implicitly so (Martin 1997). To provide a current example, the sole basis for the arguments derived from molecular data that support a common evolutionary origin of hydrogenosomes and mitochondria (cited above) exists in Weeden's hypothesis. The hsp10, hsp60 and hsp70 *data* suggest that amitochondriate protists possessed *mitochondria*. The logical link between *mitochondria* and *hydrogenosomes* on the basis of that data is an unproven premise: the product-specificity corollary. It *does* hold in many cases, but equally it *does not* hold in many others. It has been taken for granted for many years that it has somehow been proven; but that is not the case. We do not doubt for a minute that hydrogenosomes and mitochondria share a commom ancestor, biochemical data support that view even in the absence of molecular data (Müller, 1996). But in an age of symbiotic theory, where the phylogeny of organelles that do not possess DNA is being inferred from the phylogeny of the proteins that they contain, it is essential to inspect the model more critically, lest it lead us catastrophically astray. Furthermore, compartmentalised proteins in protists of secondary symbiotic origin add a dimension to the problem which Weeden could not have been aware of (Maier 1992; Henze et al. 1995; van de Peer et al. 1996, McFadden et al. 1997). For these reasons, the model has been explicitly re-formulated in Fig. 2 so that its merits may be examined.

The Calvin cycle: a pathway of pieces (was it ever together?)

The evolution of the Calvin cycle as a whole (the pathway), consisting of parts (enzymes), can be expressed in general terms as follows. The history of the whole is the sum of the histories of the parts but, in different prokaryotes and eukaryotes, neither are the histories of the parts identical, nor are the same parts employed in all cells. What is conserved in the Calvin cycle across the tree of life is the set of substrate conversions given in Fig. 1. The enzymes that catalyze these conversions in two organisms may be orthologues, paralogues, or structurally unrelated. Such proteins, functionally equivalent but structurally unrelated (or nearly so), are known as class-I/class-II enzymes. Several such examples will be encountered.

Because of their very different histories, the only meaningful way to approach nuclear Calvin-cycle genes from an evolutionary standpoint is to address each enzyme independently and consider the known diversity of its counterparts from eubacteria and, where applicable, archaebacteria. The following sections cannot possibly more than scratch the surface of the vast literature for each enzyme. The sections are intended to provide the reader access to a few recent papers on the evolution each enzyme, and to formulate, where possible, a current hypothesis for the evolutionary origin of the eukaryotic nuclear genes in the context of endosymbiosis, noting the compartmentation and evolution of functionally redundant homologues. Taken together, this will enable the reader to recognize the complexities inherent in the problem.

Ribulose-1,5-bisphosphate carboxylase/oxygenase

Rubisco (EC 4.1.1.39) catalyzes the initial CO_2 fixation step of the Calvin cycle (for details of catalysis, see Gutteridge and Gatenby 1995). Rubisco is the only enzyme of the Calvin cycle which is still encoded – at least in part – in chloroplast DNA (cpDNA). Two structurally and evolutionarily distinct types of Rubisco are known. Class-I (or form-I) Rubisco has a native M_r of about 560 kDa and consists of eight large subunits (LSU, M_r approximately 55 kDa each, *rbc*L gene) and eight small subunits (SSU, M_r approximately 15 kDa each) to comprise the L_8S_8 holoenzyme. Class-II Rubisco consists only of large subunits of M_r approximately 55 kDa each ($L_{2–6}$ holoenzyme) (Tabita 1994; Gibson and Tabita, 1996). This subunit possesses several indels relative to the large subunit of class-I Rubisco and shares about 30% amino-acid identity with the latter, indicating that they share a common ancestor.

The class-I L_8S_8 enzyme is typical of photosynthetic eukaryotes, cyanobacteria, and numerous other eubacteria.

Within class-I Rubisco enzymes, a distinct dichotomy at the amino-acid-sequence level into 'red' type (R-type) and 'green' type (G-type) class-I Rubisco is found. R-type Rubisco derives its name from the fact that members of this class-I gene family are typically found in the cpDNA of red algae (rhodophytes) and in protists that derive their plastids from rhodophytes via secondary endosymbiosis. cpDNA-encoded R-type Rubisco operons also encode the small subunit (RbcS). G-type Rubisco derives its name from the fact that this gene family of the class-I enzyme is found in cyanobacteria and chlorophytes.

The deep dichotomy between R-type and G-type Rubisco is the result of an ancient gene duplication that involved both the large and small subunits in the common ancestor of contemporary eubacteria, followed by differential gene losses during the course of bacterial evolution and endosymbiosis (Martin et al. 1992, 1993b). The extreme discordance between bacterial phylogenies based on rRNA and *rbc*L gene evolution *across and within* the R-type and G-type families suggests that some lateral transfer of Rubisco genes between eubacteria might have occurred in evolution (Assali et al. 1991; Morden et al. 1992; Palmer 1995; Delwiche and Palmer 1996). The Rubisco sequences recently reported from the marine cyanobacterium *Synechococcus* sp. strain WH7803 (Watson and Tabita 1996) raise the question of whether Rubisco is truly subject to rampant lateral transfer, or whether we are just beginning to see the tip of an iceberg of the functional diversity of this enzyme among autotrophic photosynthesizers, the underlying biochemical and genetic basis of which is not yet fully understood. Conservatively interpreted, the extent of Rubisco diversity observed in eubacteria suggests that the latter may be the case (Watson and Tabita 1996).

Members of the R-type class-I Rubisco family share approximately 70–80% amino-acid identity between large subunits, whereas identity between R-type and G-type RbcL sequences is of the order of 50–60%. No organism is currently known which possesses both R-type and G-type class-I Rubisco enzymes. With the exception of *Cyanophora paradoxa*, where *rbcS* is cpDNA-encoded (Starnes et al. 1985; Stirewalt et al. 1995), the small subunit of G-type Rubisco is nuclear encoded in all eukaryotes studied to-date. The cyanobacterial and *Cyanophora* reference sequences make it evident from molecular phylogenies (Martin et al. 1992; Delwich and Palmer 1996) that the *rbc*S gene has been straightforwardly transferred to the nucleus during the course of chlorophyte evolution. However, when the early gene duplication in eubacteria, subsequent differential gene loss, and possible lateral transfer of class-I Rubisco genes between eubacteria are taken into account, a relatively complex picture of *rbc*S gene evolution on its way to the nucleus for function in the higher-plant Calvin cycle emerges, as schematically summarized in Fig. 3.

Prior to fixation in the nucleus, transferred genes must acquire a transit peptide in order to replace the function of the organellar copy. The transit-peptide-coding region of nuclear RbcS genes is separated in many species from the region encoding the mature subunit by an intron, suggesting that exon shuffling may have played a role in the acquisition of the transit-peptide (Wolter et al. 1988). A particularly clear example of exon shuffling in the acquisition of transit peptides was recently reported (Long et al. 1996), but it is still not known whether exon shuffling is the major mechanism of transit-peptide acquisition in plants.

Curiously, the *M. jannaschii* genome (Bult et al. 1996) encodes a homologue for a class-I RbcL subunit, but does not encode an identifiable *rbc*S gene, nor does it contain a recognizeable homologue of either class-I or class-II phosphoribulokinase (see below) to generate the ribulose-1,5-bisphosphate substrate. The *Methanococcus rbc*L sequence is an outgroup to both R-type and G-type class-I homologs, as depicted schematically in Fig. 3; but the function of the encoded protein is unknown.

As a last note on class-I Rubisco, the nuclear *rbc*S gene of *Euglena gracilis* encodes a polyprotein consisting of eight concatenated, nearly identical, *rbc*S subunits that are imported into the chloroplast with the aid of a single transit peptide. Upon import, the polyprotein is proteolytically cleaved to release the mature subunits for assembly of the enzyme (Chan et al. 1990; Tessier et al. 1995). Polyprotein precursor import has only been observed in eukaryotes of secondary symbiotic origin.

Class-II Rubisco, with two very notable and well-documented exceptions, is found only in proteobacteria. The exceptions are nuclear-encoded class-II Rubisco in the dinoflagellates *Gonyaulax* (Morse et al. 1995) and *Symbiodinium* (Rowan et al. 1996) – photosynthetic protists with complex plastids. This is in contrast to all previously studied photosynthetic eukaryotes, where the large subunit of Rubisco is cpDNA-encoded. Like RbcS of *Euglena*, the nuclear class-II RbcL of *Symbiodinium* is encoded as a polyprotein, but consists of three, instead of eight, concatenated subunits. Several eubacteria are known which possess both class-I and class-II Rubisco, most notably *Rhodobacter*, where their operons are differentially expressed in response to oxygen levels (Gibson and Tabita 1993), but also *Thiobacillus* and *Hydrogenovibrio* (see Watson and Tabita 1996). Complex scenarios of lateral transfer of Rubisco genes between plastids, their antecedants or other organelles, either prior to or subsequent to symbiosis, have been suggested to explain the dinoflagellate class-II Rubisco data (Palmer 1995; Delwiche and Palmer 1996). Such interpretations are very speculative. The most straightforward statement is that put forward by both Morse et al. (1995) and Rowan et al. (1996), namely that Rubisco gene diversity in eubacteria is insufficiently known to discriminate between the possibilities of cyanobacterial, mitochondrial, or other origins for these nuclear class-II genes. In subsequent sections, we will encounter examples of nuclear genes for Calvin-cycle enzymes of higher plant chloroplasts whose descent from mitochondria is much more clearly implied.

To summarize, both class-II Rubisco and two members of the eubacterial class-I Rubisco family, R-type and G-type, occur in eukaryotes. Genes for class-II Rubisco, when they have been found in eukaryotes, occur only in the nucleus, and their origin is unclear. Genes for G-type class-I

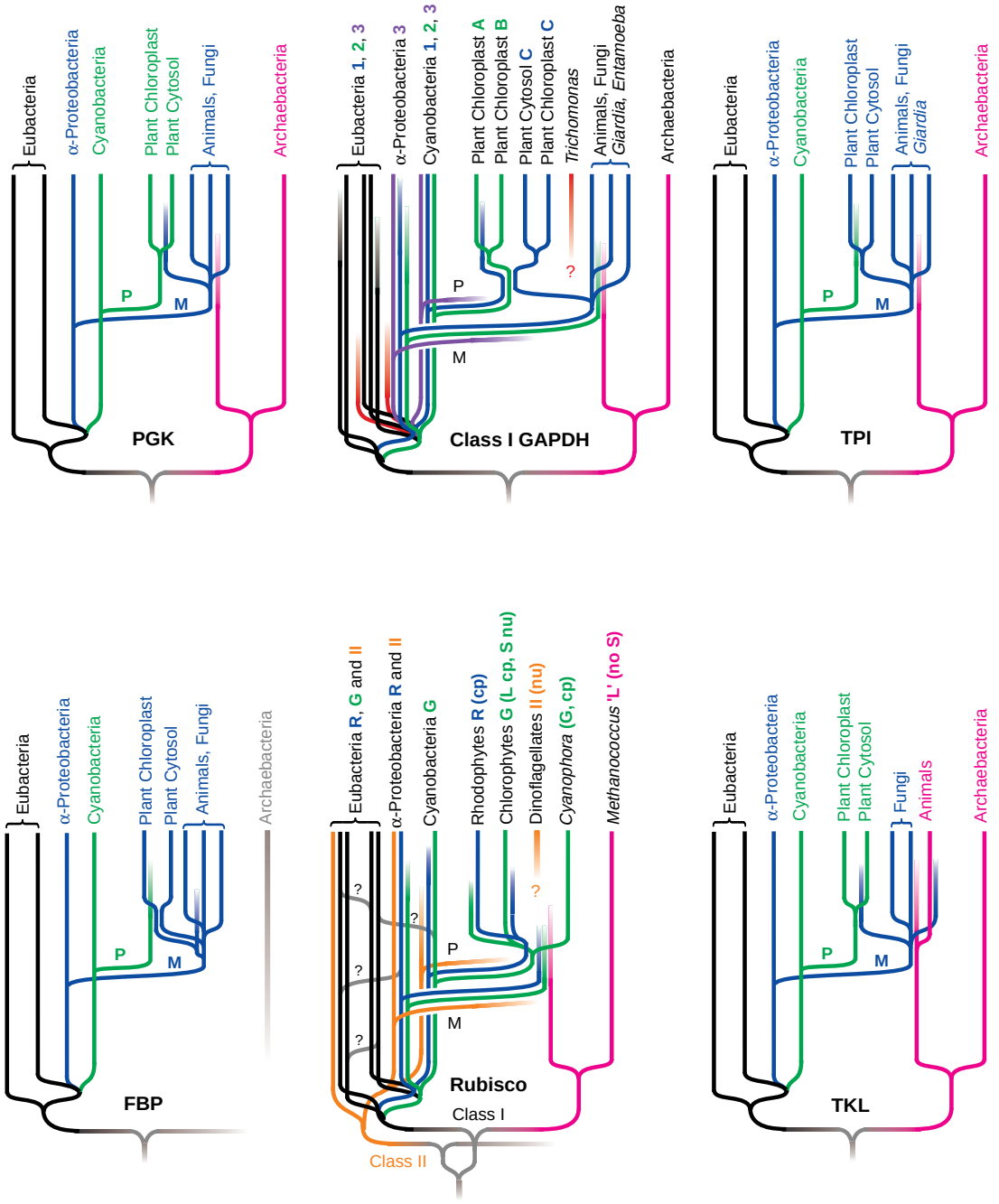

Figure 3 Schematic summary of the origins of genes for selected Calvin-cycle enzymes and their cytosolic homologues (see text). Branches of relevant gene lineages are color-coded in such a way as to depict gene phylogeny with priority over symbiont phylogeny, the crux of eubacterial gene sampling through endosymbiosis. *Colors* across different schematic trees do not consistently reflect a certain lineage. Genes which did not persist to the present (or have not been characterized to-date) are indicated by *tapered branches.* Cases in which a gene origin is unaccountable are indicated with *discontinuous lines. P*, plastids; *M*, mitochondria. See text

Rubisco are usually divided, part in the nucleus (*rbc*S) and part in the plastid (*rbc*L). Genes for R-type Rubisco (both *rbc*L and *rbc*S) have only been found in plastid DNAs (Kowallik et al. 1995; Reith and Munholland 1995), where, intruigingly, they are accompanied by plastid-encoded *cbbR*, apparently an orthologue of the regulator of proteobacterial Calvin-cycle gene expression, plus plastid-encoded *cbbX*, a protein of still unknown function encoded in proteobacterial *cbb*-operons (Gibson and Tabita 1997).

Only a very small sample of eubacterial Rubisco gene diversity is preserved within any one eukaryotic cell; but across the breadth of eukaryotic lineages nearly the full spectrum of eubacterial Rubisco diversity can be found. Since sound evidence exists for a single primary symbiotic origin of both chloroplasts (Melkonian 1996) and mitochondria (Gray 1993), this suggests that symbionts possessed Rubisco diversity equal to or greater than that found in any contemporary eubacterium, that endosymbiosis had a bottleneck effect on that diversity, and that the bottleneck

occurred independently in different eukaryotic lineages. In this view, endosymbiosis was a straightforward sampling process of eubacterial Rubisco genes. This principle can be observed in the history of several other Calvin-cycle enzymes.

Phosphoglycerate kinase

PGK (EC 2.7.2.3) is the only Calvin-cycle enzyme that is active as a monomer, its M_r being approximately 44 kDa. Numerous sequences for PGK from various sources, including several archaebacteria, are known (Fothergill-Gilmore and Michels 1993). No obvious ancient eubacterial gene duplications in the history of the enzyme have been identified, nor have class-I and class-II forms of PGK been described.

Higher plants possess chloroplast and cytosolic isoenzymes for PGK (Longstaff et al. 1988) whereas *Chlamydomonas* possesses solely a chloroplast form of the enzyme (Schnarrenberger et al. 1990). The nuclear genes for chloroplast and cytosolic PGK from higher plants arose through gene duplication during plant evolution, the copy encoding the chloroplast-targeted product having acquired a transit peptide. The ancestral gene from which contemporary homologues for the Calvin-cycle/glycolytic isoenzymes in higher plants derive was suggested to have been acquired by the nucleus from the cyanobacterial antecedants of plastids (Brinkmann and Martin 1996). Since that paper appeared, the PGK sequence from *Synechocystis* has become available (Kaneko et al. 1996). It branches very robustly with the plant nuclear genes to the exclusion of homologues from other sources, confirming the cyanobacterial origin suggested by Brinkmann and Martin (1996).

Interestingly, *the (presumed) pre-existing nuclear gene for cytosolic PGK that was replaced* in the plant lineage itself seems to have been an aquisition by eukaryotes from eubacteria (under the hypothesis of Fig. 2, from the mitochondrial ancestor), since cytosolic PGK of non-photosynthetic eukaryotes shares approximately 50% identity with eubacterial, and only approximately 30% identity with archaebacterial, homologues. The chloroplast/cytosol PGK gene duplication early in plant evolution almost certainly resulted in replacement of the pre-existing nuclear gene of mitochondrial origin, coding for cytosolic PGK. This scenario of endosymbiotic gene replacement is schematically depicted in Fig. 3.

Apicocomplexan parasites of humans (*Toxoplasma, Plasmodium* and relatives) possess a highly degenerate 35-kb chloroplast genome (Williamson et al. 1994) and were recently shown to contain plastids (McFadden et al. 1996). However, in contrast to plants, *Plasmodium* does not possess a *pgk* gene of plastid origin, rather the mitochondrial copy has persisted. The example of eukaryotic, and specifically higher-plant, PGK, reveals that for enzymes common to endosymbionts and host, each endosymbiotic event has a surprisingly high likelihood of successful endosymbiotic gene transfer. And in the ensuing competition for survival between endogenous and intruding genes, pre-existing endogenous nuclear copies do not appear to enjoy a significant or systematic advantage.

Glyceraldehyde-3-phosphate dehydrogenase

Class-I and class-II GAPDH (phosphorylating, EC 1.2.1.12) are known that share only 15–20% sequence identity. Both class-I and class-II native enzymes are tetramers with an M_r of about 150 kDa, consisting of approximately 37-kDa subunits. The class-I enzyme is typical of eukaryotes and eubacteria (Cerff 1995) and is found in one halophilic archaebacterium (Brinkmann and Martin 1996). The class-II enzyme has to-date only been found in archaebacteria (Zwickl et al. 1990).

Higher-plant Calvin-cycle GAPDH is the only heterotetrameric GAPDH known. The A and B subunits of chloroplast GAPDH ($NADP^+$-dependent, EC 1.2.1.13, GapA and GapB, respectively) arose via gene duplication during chlorophyte evolution (Meyer-Gauen et al, 1994), their amino-acid sequences sharing roughly 80% amino-acid identity (Brinkmann et al. 1989; Shih et al. 1992). GapB differs from GapA (and all other GAPDH subunits) by the presence of a highly negatively charged C-terminal extension (CTE) of roughly 30 amino acids that may be involved in light regulation (Baalmann et al. 1996).

The evolutionary origin of nuclear-encoded Calvin-cycle GAPDH has been uncontroversial. The first sequences for this plant enzyme were interpreted as evidence for its eubacterial origin via endosymbiotic gene transfer from chloroplasts (Martin and Cerff 1986), and these conclusions were later substantiated by subsequent analyses of GAPDH genes from cyanobacteria (Martin et al. 1993a) and *Euglena* (Henze et al. 1995). The sequence of the nuclear-encoded chloroplast enzyme in plants is significantly more similar to its homologues from cyanobacteria than it is to the enzyme from any other eubacterial or eukaryotic source, indicating that the gene product is re-imported into the organelle that donated the gene (a corroborating example of the product-specificity corollary).

The origin of the gene for the cytosolic (glycolytic) enzyme of higher plants and non-photosynthetic eukaryotes has been a matter of much confusion and debate. The crux of the problem concerning the evolution of cytosolic GAPDH was the discovery of a GAPDH gene in *Escherichia coli* that was surprisingly more similar to eukaryotic GAPDH than it was to other eubacterial sequences known at the time (Branlant and Branlant 1985). That finding was initially interpreted either as functional convergence for mesophily of the protein (Branlant and Branlant 1985) or as horizontal gene transfer from eukaryotes to *E. coli* (Martin and Cerff 1986). The latter eukaryote-to-prokaryote horizontal-transfer story was popularized as a strong case of inter-kingdom promiscuity (Doolittle et al. 1990; M. W. Smith et al. 1992) *in the direction opposite to that of endosymbiotic gene transfer* and is still often cited as such, but in all likelihood erroneously so. GAPDH genes studied from a broader eubacterial sample revealed that ancient duplications had produced a GAPDH gene family in the

common ancestor of cyanobacteria and proteobacteria that had been subject to extensive differential loss during evolution. Furthermore, the pre-existing nuclear (presumably archaebacterially related) gene for cytosolic GAPDH of eukaryotes appears to have been functionally replaced by a copy originating through endosymbiotic gene transfer from the antecedants of mitochondria (Martin et al. 1993a; Henze et al. 1995). The suggestion that eukaryotic cytosolic GAPDH might descend from mitochondrial DNA was first forwarded by T. L. Smith (1989). The fact that some amitochondriate protists (*Giardia lamblia* and *Entamoeba histolytica*) possess GAPDH genes of apparent mitochondrial origin was taken as evidence that these organisms had once possessed mitochondria that were secondarily lost (Henze et al. 1995). In *Trichomonas vaginalis*, a GAPDH was found that has no known orthologue among eubacterial GAPDH families (Markos et al. 1993), but which suggested the occurrence of a eubacterial symbiont in the history of that hydrogenosome-containing amitochondriate protist. The view that ancient eubacterial diversity, rather than phylogeny, is reflected in eukaryotic GAPDH genes is not universally accepted (Roger et al. 1996). However, the finding of a class-I GAPDH gene, only distantly related to those of eubacteria, in the archaebacterium *Haloarcula vallismortis* (Brinkmann and Martin 1996), supports the ancient diversity - endosymbiotic sampling view. The *Methanococcus* genome does not help to resolve this issue, because it does not encode a class-I GAPDH gene.

Considerable GAPDH gene diversity can be found in individual eubacteria. For example, *E. coli* and *Anabaena variabilis* each possess three different GAPDH genes with approximately 50% amino-acid identity within, and approximately 60% identity between genomes (Martin et al. 1993a; Henze et al. 1995). Notably, one of the *E. coli* genes, *gapB* (*E. coli gap2* in Henze et al. 1995), has evolved strong substrate specificity for erythrose-4-phosphate, rather than glyceraldehyde-3-phosphate (Zhao et al. 1995). Curiously, *E. coli gapB* appears to be the orthologue of Calvin-cycle GAPDH in *Rhodobacter* and *Alcaligenes*, whereas it is certainly not the orthologue of cyanobacterial Calvin-cycle GAPDH (Henze et al. 1995).

A novel plastid-localized NAD^+-dependent GAPDH was recently found in pine seedlings (Meyer-Gauen et al. 1994) and ferns (unpublished data), which seems to have arisen through duplication of the nuclear gene (of mitochondrial origin in the scenario favored here) for cytosolic GAPDH, with novel re-routing to the chloroplast during evolution. Similar such re-routing is also observed in the GAPDH from protists with complex plastids (Liaud et al. 1997).

Thus GAPDH, like Rubisco, occurs today in eubacteria as a family of genes that can be traced to a series of ancient duplications in eubacteria, and the full breadth of ancient eubacterial GAPDH diversity is not preserved in any single eubacterial genome. Endosymbiosis introduced only a small sample of that ancient gene diversity into the nucleus, but once established there, functional diversification and specialization through re-routing to various cell compartments began anew. Different samples of eubacterial GAPDH gene diversity have been preserved in different eukaryotic lineages.

Triosephosphate isomerase

TPI (EC 5.3.1.1) is a homodimer of approximately 27-kDa subunits in eubacteria, the eukaryotic cytosol, and higher-plant chloroplasts (Kurzok and Feierabend 1984; Fothergill-Gilmore and Michels 1993; Henze et al. 1994; Schmidt et al. 1995). In hyperthermophilic archaebacteria TPI is a homotetramer of 25-kDa subunits (Kohlhoff et al. 1996). Like PGK, class-I/class-II forms of the enzyme have not been described; nor have ancient eubacterial gene duplications/families been proposed.

The nuclear gene for higher-plant chloroplast TPI seems to have arisen during the course of eukaryotic evolution via simple duplication and transit-peptide acquisition of the pre-existing nuclear gene encoding cytosolic TPI in plants and non-photosynthetic eukaryotes (Henze et al. 1994; Schmidt et al. 1995). Like most cytosolic enzymes of glycolysis, TPI has traditionally been viewed as a *bona fide*, endogenous eukaryotic gene. But TPI sequences from several eubacteria, including α-proteobacteria, were recently obtained in a specific effort to identify the origin of eukaryotic TPI genes (Keeling and Doolittle 1997). That study revealed that nuclear-encoded cytosolic TPI of eukaryotes (including the amitochondriate *G. lamblia*) is more closely related to its α-proteobacterial homologues that it is to TPI from any other source. Futhermore, whereas TPI from eukaryotic and eubacterial sources share roughly 40–50% amino-acid identity, their identity with the archaebacterial homologues is only of the order of 20% (Kohlhoff et al. 1996). Taken together, this is strong evidence in favor of a mitochondrial origin for TPI of the eukaryotic cytosol, directly analogous to the case argued for cytosolic GAPDH (Martin et al. 1993; Henze et al. 1995) and PGK (Brinkmann and Martin, 1996). The scenario is more readily visible in the case of TPI due to an apparent paucity of eubacterial gene diversity preserved in contemporary genomes.

Fructose-1,6-bisphosphate aldolase

FBA (EC 4.12.1.13) represents the most prominent case of class-I and class-II enzyme occurrence in the Calvin cycle. Class-I FBAs are homotetramers whereas class-II FBAs are homodimers. The subunit size of both classes of FBA enzymes is approximately 40 kDa, but class-I and class-II FBA monomers share *no detectable sequence similarity*. This, in addition to their different catalytic mechanisms (Mo et al. 1973; Lebherz and Rutter 1969), clearly suggests that class-I and class-II FBA enzymes arose independently during evolution. Class-I and class-II FBA enzymes furthermore enjoy a complex and intriguing distribution across the kingdoms of life (for a review see Schnarrenberger et al. 1990). Most eubacteria, including all cya-

nobacteria studied to-date, typically possess class-II FBA (Rutter 1964; Antia 1967), although a few clearly documented instances of class-I FBA occurrence in eubacteria are known (Witke and Götz 1993). Halophilic archaebacteria can possess either class-I or class-II FBA (Dhar and Altekar 1986). Among higher eukaryotes, fungi typically possess class-II FBA whereas metazoa and higher plants possess class-I FBA enzymes (Schnarrenberger et al. 1990).

The distinct nuclear-encoded chloroplast (Calvin-cycle) and cytosolic FBA isoenzmyes of higher plants and charophytes are both of the class-I type (Jacobshagen and Schnarrenberger 1988, 1990). *Chlamydomonas reinhardtii* possesses only a single class-I FBA localized in chloroplasts (Schnarrenberger et al. 1994). In earlier-branching protists, the distribution of class-I and class-II FBA is more complex and has no recognizeable pattern (Rutter 1964; Antia 1967; Ikawa et al. 1972; Gross et al. 1994). Unique among the eukaryotes studied to-date, *E. gracilis* possesses both a class-I and class-II FBA: a class-I FBA in chloroplasts, as in higher plants, and a class-II FBA in the cytosol (Pelzer-Reith et al. 1994), as in fungi. This contrasts with *Euglena*'s close relatives, the kinetoplastids, where *Trypanosoma brucei* has only one class-I FBA, a glycosomal enzyme (Marchand et al. 1988). To round off the picture, *C. paradoxa* uses class-II FBA in the Calvin cycle of its cyanelles, as do cyanobacteria, but also utilizes a distinct class-II FBA in the cytosol (Gross et al. 1994).

Only one eubacterial class-I aldolase of known function has been sequenced to-date, that from *Staphylococcus carnosus* (Witke and Götz 1993). No FBA sequences are available from archaebacteria, and the *Methanococcus* genome does not contain an identifiable gene for FBA, although methanogens studied to-date possess FBA activity (Yu et al. 1994; Schönheit and Schäfer 1995), suggesting that a third independent origin of FBA-catalysis is waiting to be discovered, notwithstanding the report of experimentally selected antibodies that catalyze a very efficient aldolase reaction with the authentic class-I mechanism (Wagner et al. 1995). The *S. carnosus* class-I enzyme shares only about 25–30% identical residues with eukaryotic class-I FBA and is monomeric instead of tetrameric. Since most eukaryotic FBA enzymes are of the class-I type, the lack of reference sequences makes it impossible to cell from where eukaryotic nuclear class-I FBA genes came. Class-I FBA of the higher-plant Calvin cycle shares a common ancestor with cytosolic class-I FBA of other eukaryotes, indicating that the gene duplication that gave rise to the chloroplast/cytosol isoenzymes took place early in eukaryotic evolution (Schnarrenberger et al. 1994).

For class-II FBA, the situation is precisely the converse; numerous eubacterial sequences are known, but only three from eukaryotes: yeast, *Schizosaccharomyces pombe*, and the cytosolic enzyme from *Euglena*. Phylogenetic analyses (Plaumann et al. 1997) reveal that these nuclear genes are clearly of proteobacterial affinity, and the most straightforward explanation for that finding (in light of the results for GAPDH, PGK and TPI) is that the eukaryotic class-II FBA genes are of mitochondrial origin. Eubacterial class-II FBA genes reveal patterns of ancient gene duplications and differential loss that are at least as complex as those observed for Rubisco and GAPDH (Plaumann et al. 1997). These problems are compounded by the fact that duplicated *fba* genes in *E. coli* have undergone functional specialization towards substrate specificity for tagatose-1,6-bisphosphate. For example, in *E. coli*, GatY of the *gat*-cluster for galactitol utilization (Noblemann and Lengler 1995) and AgaY of the *aga*-cluster for N-acetylgalactosamine metabolism (Reizer et al. 1996) clearly belong to the same eubacterial class-II FBA gene family as the Calvin-cycle FBA enzymes encoded in the *cbb*-operons of proteobacteria (Gibson and Tabita 1996), whereas the functional class-II FBA enzyme of *E. coli* (found in an operon with *gapB* and PGK: Alefounder and Perham 1989) belongs to a quite different family of eubacterial FBA genes (Plaumann et al. 1997).

The fructose-1,6-bisphosphate aldolase (FBA) and sedoheptulose-1,7-bisphosphate aldolase reactions of the Calvin cycle are catalyzed by a single chloroplast FBA enzyme with dual substrate specificity (Brooks and Criddle 1966). This is true both for chloroplast class-I FBA in spinach and for chloroplast class-II FBA in *Cyanophora paradoxa* (Flechner and Schnarrenberger, unpublished).

The phylogenetic distribution of class-I and class-II FBA across the tree of life shows little, if any, recognizeable pattern. Proteins that perform the aldolase reaction arose independently at least twice, possibly three times, in evolution. The nuclear gene for the Calvin cycle enzyme of higher-plant chloroplasts (class-I) has a *functional*, but not a *structural*, homologue in the Calvin cycle both of contemporary cyanobacteria and proteobacteria, where the class-II enzyme is used. This is the clearest example where we can directly observe that the Calvin cycle evolves as a unit of function, whereby some of its structural components are freely interchangeable across species. We will encounter this principle again in the example of phosphoribulokinase. Functional exchange of class-I and class-II FBA in plastids has also occurred during eukaryotic evolution, because the early branching photosynthesizer *Cyanophora paradoxa* employs the class-II enzyme. An independent exchange occurred in the Calvin cycle of *Euglena*, where a class-I FBA has apparently been replaced by a class-I FBA of different origin. Great eubacterial diversity and seemingly unrestricted interchangeability of Calvin-cycle FBA enzymes provide an intruiging dimension of functional redundancy through symbiosis in the history of these proteins.

Fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase

The FBP (EC 3.1.3.11) and SBP (EC 3.1.3.37) steps of the Calvin cycle in eubacteria are catalyzed by a single enzyme with dual substrate specificity for both sugar bisophosphates (Yoo and Bowien 1995; Gerbling et al. 1986; Gibson and Tabita 1988). Higher-plant chloroplast FBP and SBP are encoded by distinct, but related, nuclear genes

(Raines et al. 1988, 1992) that have high specificity for their respective substrates (Cadet and Meunier 1988). The chloroplast and cytosolic isoenzymes for FBP in higher plants possess dramatically different regulatory properties (light for chloroplast FBP, allosteric regulation for cytosolic FBP: Stitt 1990) that can be attributed to functional specialization during eukaryotic evolution. This is because chloroplast and cytosolic FBP arose through a gene duplication that occurred during the course of eukaryotic evolution, apparently prior to crown group radiation (Martin et al. 1996a). Although several archaebacteria (including methanogens) clearly possess FBP enzyme activity (Yu et al. 1994; Schönheit and Schäfer 1995), no FBP sequences have been reported from archaebacteria, and the *Methanococcus* genome (Bult et al. 1996) does not encode a homologue of FBP, suggesting the existence of an entirely distinct gene encoding this enzyme in archaebacteria, i.e. the existence of class-I/class-II enzymes.

Despite the fact that archaebacterial homologues are not available for comparison, a mitochondrial origin for eukaryotic FBP genes can be proposed. This is because the FBP gene tree is characterized by two findings typical of other cases of eubacterial origin: (1) sequence divergence within eubacterial sequences (of the order of 40% identity) is greater than between eukaryotic and some eubacterial homologues (of the order of 50% identity), and (2) eukaryotic FBP sequences appearing on a branch that emerges from a 'trunk' of more highly diverse eubacterial sequences (Martin et al. 1996a). This pattern is strikingly similar to that observed for eukaryotic PGK, GAPDH and TPI (Martin 1996). Furthermore, the genes for higher-plant chloroplast and cytosolic FBP appear to descend from the same eubacterial source as their homologues from the non-photosynthetic eukaryotes thus far surveyed, and none of the eukaryotic nuclear genes reveal a significant similarity to FBP sequences from three cyanobacteria available for comparison. These findings suggest that the gene for the Calvin-cycle FBP of higher-plant chloroplasts was, in one further example, donated to the nucleus by a eubacterial symbiont distinct from the antecedant of the organelle into which the enzyme is now imported (Martin et al. 1996a). As sketched in Fig. 3, it appears that a nuclear gene of mitochondrial origin for cytosolic FBP underwent duplication to create the product that functions in chloroplasts.

Transketolase

TKL (EC 2.2.1.1) is essential to both the Calvin cycle of higher-plant chloroplasts and to the oxidative pentose phosphate pathway (OPPP). In spinach, both the Calvin cycle and the regenerative segment of the OPPP are localized in plastids and only one TKL enzyme is detectable which functions in both pathways (Schnarrenberger et al. 1995). In the dehydratable angiosperm *Craterostigma* two *TKL* genes are expressed, the products of which lack a transit peptide and, like their homologues of the OPPP in the cytosol of non-photosynthetic eukaryotes, are evidently cytosolic enzymes (Bernacchia et al. 1995). The evolution of TKL genes was considered in Flechner et al. (1996). Several critical TKL sequences have been reported since then, most notably those from *Synechocystis* PCC6803 (Kaneko et al. 1996) and *Methanococcus* (Bult et al. 1996). The *Methanococcus* TKL homologue is split across two genes that correspond to the N- and C-terminal halves, respectively, of the single-chain eukaryotic and eubacterial subunit. A phylogenetic analysis of TKL genes, including the cyanobacterial and archaebacterial sequences, reveals that TKL enzymes from the mammalian cytosol (long-branch species in Flechner et al. 1996) are much more similar to the archaebacterial sequence (average 60% sequence identity) than they are to homologues from other sources, including the yeast and higher-plant cytosol (average 50% identity) (Martin 1997). This similarity is underscored by the presence, in eubacterial homologues, of an approximately 40 amino-acid insertion found in the region corresponding to the break in the N-terminal TKLα and C-terminal TKLβ chains of the *Methanococcus* enzyme that is lacking in the mammalian homologues (Martin 1997). Also, *Saccharomyces* and plant (both chloroplast and cytosol) TKL sequences share much more similarity to eubacterial homologues (65–70% identity) than they do with the either the archaebacterial or the mammalian cytosolic enzymes. The higher-plant chloroplast and cytosolic TKL isoenzymes derive from a gene-duplication event which appears to have occurred specifically in the plant lineage (Flechner et al. 1996), and the plant sequences branch very robustly with the *Synechocystis* homologue. The TKL topology is generally complex, gene-duplications being evident among eubacterial sequences (Flechner et al. 1996) and there is no clear evidence for a specific affinity between fungal and proteobacterial *TKL* genes.

Despite the paucity of sequences from a large eubacterial and eukaryotic sample, and some other complications, the general pattern of similarity can be explained in a reasonably straightforward manner, although numerous details remained to be clarified. Recalling the general hypothesis for the origin of nuclear genes (Fig. 2), the overall topology of robustly supported branches in the TKL tree suggests that genes of mitochondrial, cyanobacterial, *and in a rare case also archaebacterial-type* (i.e. eukaryote ancestral) genes for an enzyme of central carbon metabolism, have persisted to the present in nuclear genomes as depicted in Fig. 3. Although the overall scenario for TKL is still incomplete, the case for a cyanobacterial origin of the plant genes and the origin of the plant chloroplast/cytosol isoenzymes via recent duplication is quite solid. The nature of the archaebacterial affinity of the mammalian enzymes will become clearer with larger eukaryotic samples. TKL will be an interesting enzyme for further study.

Ribulose-5-phosphate 3-epimerase

RPE (EC 5.1.3.1) is a homodimer of approximately 23-kDa subunits in animals (Karmali et al. 1983), eubacteria (Kusian et al. 1992) and plants (Nowitzki et al. 1995;

Teige et al. 1995). Very little is known about this Calvin-cycle enzyme. In spinach leaves, RPE is found only in the chloroplast, with no cytosolic activity detectable (Schnarrenberger et al. 1995). No class-I/class-II enzymes been described for RPE, but three very distantly related *rpe* genes occur in *E. coli*, indicating the existence of a relatively ancient gene family (Nowitzki et al. 1995). Including the homologues identifiable in the *Synechocystis* and *Methanococcus* genomes, still fewer than ten RPE sequences have been deposited in GenBank at the time of this review. The only statement that can currently be made is that the nuclear genes for chloroplast RPE from plants are robustly more similar to the *Synechocystis* homologue than to proteobacterial or other sequences, indicating that the plant nuclear gene was acquired from the chloroplast ancestor. The phylogenetic positions of the archaebacterial, yeast, and nematode sequences currently permit no conclusions.

Ribose-5-phosphate isomerase

RPI (EC 5.3.1.6) is a homodimer of approximately 23-kDa subunits in most source studies (Hove-Jensen and Maigaard 1993; Martin et al. 1996b). RPI and TPI are the only two Calvin-cycle enzymes that have not been identified in the *cbb* operons of photosynthetic proteobacteria (Gibson and Tabita 1996). As in the case of TKL and RPE, spinach leaves possess only one detectable RPI enzyme, localized in chloroplasts (Schnarrenberger et al. 1995). *E. coli* has two genes for functional RPI enzymes, *rpiA* and *rpiB*, the encoded products of which share no sequence similarity. RpiA is similar to the enzyme from eukaryotes, whereas RpiB of *E. coli* is a homodimer of 16-kDa subunits and shares 40% identity with galactose-6-phosphate isomerase of *Staphylococcus aureus*, (Sørensen and Hove-Jensen 1996). The functional interchangeability of the RpiA and RpiB gene products tentatively suggests that they represent *bona fide* class-I (RpiA) and class-II (RpiB) RPI enzymes. Even fewer sequences are currently available for RPI than for RPE. The spinach chloroplast (class-I) RPI enzyme has identifiable homologues encoded in the *Synechocystis* and *Methanococcus* genomes, but too little data for this Calvin-cycle enzyme are currently available to address the evolutionary origin of the plant nuclear gene.

Phosphoribulokinase

PRK (EC 2.7.1.19) regenerates the primary CO_2 acceptor for Rubisco, completing the cycle. Class-I and class-II PRK enzymes are known (Tabita 1994; Brandes et al. 1996), although previously they have not formally been designated as such. The best studied class of PRK is that found in photosynthetic proteobacteria, an octamer of approximately 30-kDa subunits. The enzyme from cyanobacteria (a tetramer) and higher plants (a dimer) consists of approximately 40-kDa subunits (Tabita 1988, 1994). The proteobacterial enzymes were first studied in detail (Gibson et al. 1990) and can be designated as class-I, with PRK from cyanobacteria and higher plants designated as class-II. Class-I and class-II PRK share only about 20% amino-acid identity. The question is open as to whether they might share a common ancestor as in the case of class-I/II Rubisco, or whether they arose through functional convergence as in the case of class-I/II FBA. PRK activity has not been detected in non-photosynthetic organisms. It is therefore curious that the *E. coli* genome possesses a gene with great similarity to class-II PRK. PRK has been cloned from a few plant sources (Milanez and Mural 1988; Raines et al. 1989; Roesler and Ogren 1990; Horsnell and Raines 1991), and a gene phylogeny based on a few higher plants and *Chlamydomonas* suggests that the nuclear-encoded PRK genes are descendants of the class-II cyanobacterial enzyme, and constitute a classic example of gene transfer with re-targeting to the organelle of origin. However, like FBA, the occurrence of class-I and class-II PRK in the Calvin cycle from different sources demonstrates that the pathway does not evolve as a unit of structure, but rather as a unit of function with interchangeable parts.

Principles: what we see, what we don't see, and many, many uncertainties

The non-specialist should once again be reminded that several premisses, upon which many of the arguments in this paper hinge, were explicitly stated in the introductory passages and that the actual gene phylogenies upon which many of our interpretations are based are more complex than the corresponding diagrams in Fig. 3 suggest. It is also important to note that lateral gene transfer between eubacteria is not uncommon (Lawrence and Ochman 1997) but, at the same time, appears to be much more common between closely related species than between distantly related ones (Matic et al. 1995). If lateral transfer of Calvin-cycle genes between distantly related eubacteria occurred with appreciable frequency prior or subsequent to symbiosis, as sketched for Rubisco in Fig. 3, then the problems introduced by ancient gene duplications and differential loss will be compounded by a factor of ancient eubacterial allele sampling. The potential result of such lateral transfer, as it applies to interpreting the history of eubacterial genes in eukaryotic genomes, is simply that many different eubacterial lineages would become erroneously implicated as gene donors to eukaryotes. If gene transfer between eubacteria and archaebacteria outside the context of eukaryotic origins was common in evolution – comparative genomics will soon provide hints as to whether this has occurred – then we will have great difficulties ultimately reconstructing the origins of eukaryotic genes. Worse yet, if many different prokaryotes contributed genes to the eukaryotic nucleus, without having left organelles behind as a trace of their existence (cryptic symbiosis: Henze et al. 1995), the nucleus can be expected to assume the contours of a genetic garbage bin.

Table 1 Summary of findings and interpretations from evolutionary analyses of Calvin-cycle genes. Cp, chloroplast; Mt, mitochondria; Eub, eubacteria; Euk, eukaryotes; n.a. not applicable; ?, unknown

Enzyme	Origin of chloropl. isoenzyme	Origin of cytosol. isoenzyme	Chloropl.-cytosol. isoenzyme gene dupl.	Eubact. gene families	Class-I and class-II enzymes
Rubisco	Cp	n.a.	n.a.	Yes	Yes
GAPDH	Cp	Mt	Eub	Yes	Yes
PGK	Cp	Cp	Plants	?	?
TPI	Mt	Mt	Plants	?	?
FBA[a]	Eub?	Eub?	Early Euk.	Yes	Yes
FBP	Mt	Mt	Early Euk.	?	Yes?[d]
SBP[b]	?	n.a.	n.a.	?	?
TKL[c]	Cp	Cp	Plants	Yes	?
RPE	Cp	n.a.	n.a.	Yes	?
RPI	?	n.a.	n.a.	?	Yes[e]
PRK	Cp	n.a.	n.a.	?	Yes

[a] Refers to class-I FBA (see text)
[b] SBP and FBP share a common ancestor, due to lack of outgroups the timing of the duplication is difficult to estimate
[c] The cytosolic TKL of animals may be an archaebacterial enzyme
[d] See text
[e] Class-I/II RPI has only been found in *E. coli*, see text

At present, however, with (1) the help of endosymbiotic theory, (2) an explicitly structured working hypothesis, and (3) familiarity with the uncertainties inherent to molecular phylogenetics, the evolutionary histories of eukaryotic enzymes of carbohydrate metabolism need not be dismissed as confusing and inconsistent. Rather, they can tell us many things about endosymbiosis and eukaryotic evolution, but appear to be doing so in an unfamiliar dialect.

As summarized and interpreted in Table 1 and Fig. 3, the history of the enzymes of the Calvin cycle demonstrate a substantial contribution of endosymbiotic gene transfer to eukaryotic genomes. The enzymes of the Calvin cycle, glycolysis, and gluconeogenesis furthermore provide an overall picture of eukaryotic evolution that is quite different from that depicted by components of the genetic apparatus (rRNA, translation factors, polymerases and the like). The evolutionary patterns of all of the above can be accounted for by the simplified scenario of eukaryotic evolution given in Fig. 2, except that (1) the product-specificity corollary for the compartmentation of enzymes must be abandoned (Martin 1997), and (2) the assumption of widespread paralogy and differential gene loss in eubacteria and eukaryotes must be invoked in order to explain the data.

For enzymes of central carbohydrate metabolism that were common to symbionts and hosts, we see that the products of genes transferred to the nucleus from the antecedants of chloroplasts and mitochondria were evidently quite free to explore various possibilities of cellular compartmentation during evolution. It seems that many different novel compartmentation variants, which apparently were not detrimental to the cell, have become fixed in different eukaryotic lineages. From this we may conclude that the contemporary compartmentation of the pathways studied is very likely the end result of numerous evolutionary experiments in individual eukaryotic lineages. Deleterious 'compartmentation variants' were unsuccessful and hence eliminated; advantageous or non-deleterious ones survived. This view predicts that in early branching photosynthetic protists, novel scenarios of individual enzyme compartmentation (regardless of gene origin), and novel patterns from the sampling of prokaryotic gene diversity for enzymes of the pathway will be found.

Among the 12 enzymes considered here only one has been identified so far that reflects an archaebacterial ancestry of cytosolic sugar phosphate metabolism in eukaryotes, namely transketolase. We know of no cases of archaebacterially related enzymes that have assumed functions in organelles, but Fig. 2, now freed of the product-specificity corollary, clearly predicts that such examples might eventually be discovered. It appears that not only the Calvin cycle and much of the oxidative pentose phosphate pathway of higher-plant plastids, but also major portions of the glycolytic and gluconeogenetic pathways of the cytosol, are catalyzed by eubacterial enzymes currently encoded in nuclear chromosomes. Extrapolating from these findings, it seems that a general picture of eubacterial (endosymbiotic) origin for central carbohydrate metabolism in the eukaryotic cytosol may eventually emerge. Yet it also seems likely that, in diverse eukaryotes, cytosolic pathways, like the Calvin cycle, may turn out to be units of function (substrate conversion) with interchangeable parts (enzymes), that constitute different samples of eubacterial gene diversity.

Obviously, endosymbiotic gene transfer has resulted in a significant increase of nuclear genetic diversity over that which existed prior to symbosis. However, for several enzymes of the Calvin-cycle pathway, the diversity observed among eubacterial genes is much greater than that observed across eubacteria and eukaryotes when individual organisms are compared. Thus, in individual eukaryotic lineages, a bottleneck in eubacterial gene diversity for Calvin-cycle enzymes seems to have occurred, with differential

gene loss. In different eukaryotic lineages this bottleneck preserved different eubacterial genes. Therefore much of the ancient eubacterial diversity of Calvin-cycle enzymes is only observable when the breadth of eukaryotic lineages is examined. Under our interpretation, this result is produced by the differential elimination of the functional redundancy that followed endosymbiosis.

It also seems that, in the antecedants of organelles, diversity in the pathway may have been considerably greater than that found in individual contemporary eubacteria. If these interpretations are correct, they suggest that reduction of diversity through selective elimination among functionally redundant Calvin-cycle enzymes has occurred both in eubacterial and in eukaryotic genomes subsequent to the time of symbiosis. Expressed in another way, these data tend to re-kindle an old idea that stream-lining (Doolittle 1978) of prokaryotic genomes with respect to functionally redundant genes had not gone to completion before the symbiotic origins of mitochondria and chloroplasts.

The present findings also allow us to address, in a completely new light, an important question in endosymbiotic theory that has been posed time and time again without an answer (Palmer 1997): *why were organelle genomes preserved at all?* Traditional conjecture on the issue has yet to bring forth a satisfactory explanation. It has focused on the notion that one or several organelle-encoded proteins are refractory to import, locking the organellar genetic apparatus in place. However, the broad diversity of proteins that can successfully be imported by organelles indicates that this is unlikely to be the cause of organelle genome persistence (Clausmeyer et al. 1993). Our findings have shown that the products of transferred genes are not directed back to their donor organelle with a 'homing device'. *Rather they have to find the donor organelle by themselves,* probably by trial and error. Could it be, as R.-B. Klösgen has suggested (personal communication), that persistently organelle-encoded proteins, when transferred and synthesized as soluble cytosolic precursors, wander about the eukaryotic cell and interfere with processes in such a way as to be toxic or otherwise deleterious? Many seemingly harmless foreign proteins when expressed in *E. coli* or yeast *are* toxic to those cells, and endsymbiotic gene transfer can be viewed as nothing more than Nature's ongoing heterologous expression experiment. If the errors are lethal, then both the gene product and the gene are trapped in the organelle, not because they can't get back in, but rather because the cytosol won't tolerate them outside. This hypothesis is testable.

One question emerges from these studies that could not be formulated previously: *why have the eubacterial enzymes of glycolysis surveyed here, instead of their archaebacterial homologues, been preserved in the eukaryotic cytosol?* Since we can only extrapolate to the past from known facts in the present, would all that follow would be speculation, and is therefore invited to adopt this not as the last sentence of this paper, but rather as the first sentence of the next.

Acknowledgements We thank Miklos Müller, Paul Michels, Ralf-Bernd Klösgen, Reinhardt Hensel, Ford Doolittle, Roger Milkman, Martin Embley and Reinhold Herrmann for numerous valuable discussions. Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

References

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

Altekar W, Rajagopalan R (1990) Ribulose bisphosphate carboxylase activity in halophilic archaebacteria. Arch Microbiol 153: 169–174

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

Assali N-E, Martin W, Loiseaux-de Goër S (1991) Evolution of the Rubisco operon from prokaryotes to algae: structure and analysis of the *rbc*S gene of the brown alga *Pylaiella littoralis*. Plant Mol Biol 17:853–863

Baalmann E, Scheibe R, Cerff R, Martin W (1996) Functional studies of chloroplast glyceraldehyde-3-phosphate dehydrogenase subunits A and B expressed in *Escherichia coli*: formation of highly active A_4 and B_4 homotetramers and evidence that aggregation of the B_4 complex is mediated by the B subunit carboxy terminus. Plant Mol Biol 32:505–514

Baldauf S, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation-factor phylogeny. Proc Natl Acad Sci USA 93:7749–7754

Beh M, Strauss G, Huber R, Stetter KO, Fuchs G (1993) Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. Arch Microbiol 160: 306–311

Bernacchia G, Schwall G, Lottspeich F, Salamini F, Bartels D (1995) The transketolase gene family of the resurrection plant *Craterostigma plantagineum*: differential expression during the re-hydration phase. EMBO J 14:610–618

Bommer D, Schäferjohann J, Bowien B (1996) Identification of *cbbBc* as an additional distal gene of the chromosomal *cbb* CO_2 fixation operon from *Ralstonia eutropha*. Arch Microbiol 166: 245–251

Brandes HK, Hartmann FC, Lu T-Y, Larimer FW (1996) Efficient expression of the gene for spinach phosphoribulokinase in *Picia pastoris* and utilization of the recombinant enzyme to explore the role of regulatory cysteinyl residue by site-directed mutagenesis. J Biol Chem 271:6490–6496

Branlant G, Branlant C (1985) Nucleotide sequence of the *Escherichia coli gap* gene. Eur J Biochem 150:61–66

Brinkmann H, Cerff R, Salomon M, Soll J (1989) Cloning and sequence analysis of cDNAs encoding the cytosolic precursors of subunits GapA and GapB of chloroplast glyceraldehyde-3-phosphate dehydrogenase from pea and spinach. Plant Mol Biol 13:81–94

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

Brooks K, Criddle RS (1966) Enzymes of the carbon cycle of photosynthesis. I. Isolation and properties of spinach chloroplast aldolase. Arch Biochem Biophys 117:650–659

Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthase gene duplications. Proc Natl Acad Sci USA 92:2441–2445.

Buchanan BB (1991) Regulation of CO_2 assimilation in oxygenic photosynthesis: the ferredoxin/thioredoxin system. Arch Biochem Biophys 288:1–9

Bui ETN, Bradley PJ, Johnson PJ (1996) A common evolutionary origin for mitochondria and hydrogenosomes. Proc Natl Acad Sci USA 93:9651–9656

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, (35 other co-authors), Venter JC (1996) Complete genome sequence of the methanogenic Archeon, *Methanococcus jannaschii*. Science 273:1058–1073

Cadet F, Meunier J-C (1988) pH and kinetic studies of chloroplast sedoheptulose-1,7-bisphosphatase from spinach (*Spinacia oleracea*). Biochem J 253:249–254

Calvin M (1956) The photosynthetic carbon cycle. J Chem Soc 1956:1895–1915

Cavalier-Smith T (1993) Kingdom Protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cerff R (1995) The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system. In: Go M, Schimmel P (eds) Tracing biological evolution in protein and gene structures. Elsevier, Amsterdam, pp 205–227

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

Clark CG, Roger AJ (1995) Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. Proc Natl Acad Sci USA 92:6518–6521

Clausmeyer S, Klösgen R-B, Herrmann RG (1993) Protein import into chloroplasts: the hydrophilic lumenal proteins exhibit unexpected import and sorting specificities in spite of structurally conserved transit peptides. J Biol Chem 268:13869–13876

Delwiche CF, Palmer JD (1996) Rampant horizontal transfer and duplication of Rubisco genes in eubacteria and plastids. Mol Biol Evol 13:873–882

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

Doolittle R, Feng D, Anderson K, Alberro M (1990) A naturally occurring horizontal gene transfer from a eukaryote to a prokaryote. J Mol Evol 31:383–388

Doolittle WF (1978) Genes in pieces: were they ever together? Nature 272:581–582.

Doolittle WF (1996) Some aspects of the biology of cells and their possible evolutionary significance. In: Roberts D, Sharp P, Alserson G, Collins M (Eds) Evolution of microbial life. Soc Gen Microbiol Symp 54, Cambridge University Press, Cambridge, UK, pp 1–21.

Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci USA 91:6721–6728

Flechner A, Dreßen U, Westhoff P, Henze K, Schnarrenberger C, Martin W (1996) Molecular characterization of transketolase (EC 2.2.1.1) active in the Calvin cycle of spinach chloroplasts. Plant Mol Biol 32:475–484

Flügge U-I, Heldt HW (1991) Metabolite translocators of the chloroplast envelope. Annu Rev Plant Physiol Plant Mol Biol 24:129–144

Flügge U-I, Weber A, Fischer K, Lottspeich F, Eckerskorn C, Waegemann K, Soll J (1991) The major chloroplast envelope polypeptide is the phosphate translocator and not the protein import receptor. Nature 353:364–367

Fothergill-Gilmore LA, Michels PAM (1993) Evolution of glycolysis. Progr Biophys Mol Biol 59:105–238

Geiger DR, Servaites JC (1994) Diurnal regulation of photosynthetic carbon metabolism in C3 plants. Annu Rev Plant Physiol Plant Mol Biol 45:235–256

Gerbling KP, Steup M, Latzko E (1986) Fructose 1,6-bisphosphatase form B from *Synechococcus leopoliensis* hydrolyzes both fructose and sedoheptulose bisphosphate. Plant Physiol 80:716–620

Germot A, Philippe H, Le Guyader H (1996) Presence of a mitochondrial-type 70-kDa heat-shock protein in *Trichomonas vaginalis* suggests a very early mitochondrial endosymbiosis in eukaryotes. Proc Natl Acad Sci USA 93:14614–14617

Gibson JL, Tabita FR (1988) Localization and mapping of CO_2 fixation genes within two gene clusters in *Rhodobacter sphaeroides*. J Bacteriol 170:2153–2158

Gibson JL, Tabita FR (1993) Nucleotide sequence and functional analysis of CbbR, a positive regulator of the Calvin-cycle operons of *Rhodobacter sphaeroides*. J Bacteriol 175:5778–5784

Gibson JL, Tabita FR (1996) The molecular regulation of the reductive pentose phosphate pathway in proteobacteria and cyanobacteria. Arch Microbiol 166:141–150

Gibson JL, Tabita FR (1997) Analysis of the *cbbXYZ* operon in *Rhodobacter sphaeroides*. J Bacteriol 179 663–669.

Gibson JL, Chen J-H, Tower PA and Tabita FR (1990) The form II fructose-1,6-bisphosphatase and phosphoribulokinase genes form part of a large operon in *Rhodobacter sphaeroides*: primary structure and insertional mutagenesis analysis. Biochemistry 29:8085–8093

Gibson JL, Falcone DL, Tabita FR (1991) Nucleotide sequence, transcriptional analysis and expression of genes encoded within the form-I CO_2-fixation operon of *Rhodobacter sphaeroides*. J Biol Chem 266:14646–14653

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 (1989) Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes Proc Natl Acad Sci USA 86:6661–6665

Gray MW (1993) Origin and evolution of organelle genomes. Curr Opin Genet Dev 3:884–890

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

Gupta RS, Golding GB (1996) The origin of the eukaryotic cell. Trends Biochem Sci 21:166–171

Gutteridge S, Gatenby AA (1995) Rubisco synthesis, assembly, mechanism and regulation. Plant Cell 7:809–819

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

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

Horner DS, Hirt RP, Kilvington S, Lloyd D, Embley TM (1996) Molecular data suggest an early acquisition of the mitochondrion endosymbiont. Proc R Soc Lond B 263:1053–1059

Horsnell PR, Raines CA (1991) Nucleotide sequence of a cDNA clone encoding chloroplast phosphoribulokinase from *Arabidopsis thaliana*. Plant Mol Biol 17:183–184.

Hove-Jensen B, Maigaard M (1993) *Escherichia coli rpiA* gene encoding ribose phosphate isomerase A. J Bacteriol 175:5628–5635

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

Ishii M, Miake T, Satoh T, Sugiyama H, Kodama T, Igarashi Y (1996) Autotrophic carbon dioxide fixation in *Acidianus brierleyi*. Arch Microbiol 166:368–371

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

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

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

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, (18 other co-authors), Tabata S (1996) 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. DNA Res 3:109–136

Karmali A, Drake AF, Spencer N (1983) Purification, properties and assay of D-ribulose-5-phosphate 3-epimerase from human erythrocytes. Biochem J 211:617–623

Keeling PJ, Doolittle WF (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. Proc Natl Acad Sci USA 92:5761–5764

Keeling PJ, Doolittle, WF (1997) Eukaryotic triosephosphate isomerase originated in the mitochondrion. Proc Natl Acad Sci USA 94:1270–1275..

Kohlhoff M, Dahm A, Hensel R (1996) Tetrameric triosephosphate isomerase from hyperthermophilic Archaea. FEBS Lett 383: 245–250

Kowallik K, Stöbe B, Schaffran I, Kroth-Pancic P, Freier U (1995) The chloroplast genome of a chlorophyll *a*+*c*-containing alga, *Odontella sinensis*. Plant Mol Biol Rep 13:336–342

Kurzok HG, Feierabend J (1984) Comparison of a cytosolic and a chloroplast triosephosphate isomerase isoenzyme from rye leaves. I. Purification and catalytic properties. Biochim Biophys Acta 788:214–221

Kusian B, Yoo JG, Bednarski R, Bowien B (1992) The Calvin-cycle enzyme pentose-5-phosphate 3-epimerase is encoded within the *cfx* operons of the chemoautotroph *Alcaligenes eutrophus*. J Bacteriol 174:7337–7344

Kusian B, Bednarski R, Husemann M, Bowien B (1995) Characterization of the duplicate ribulose-1,5-bisphosphate carboxylase genes and *cbb* promoters of *Alcaligenes eutrophus*. J Bacteriol 177:4442–4450

Langer D, Hain J, Thuriaux P, Zillig W (1995) Transcription in Archaea: similarity to that in Eukarya. Proc Natl Acad Sci USA 92:5768–5772

Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 44:383–397

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

Liaud MF, Brandt U, Scherzinger M, Cerff R (1997) Evolutionary origin of cryptomonad algae: two novel chloroplast/cytosol specific GAPDH genes as potential markers of ancestral symbiont and host -cell components. J Mol Evol 44:S38-S43

Long M, Souza SJ de, Rosenberg C, Gilbert W (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. Proc Natl Acad Sci USA 93:7727–7731

Longstaff M, Raines CA, McMorrow EM, Bradbeer JW, Dyer T (1989) Wheat phosphoglycerate kinase: evidence for recombination between the genes for the chloroplastic and cytosolic enzymes. Nucleic Acids Res 17:6569–6580

Macdonald FD, Buchanan BB (1990) The reductive pentose phosphate pathway and its regulation. In: Dennis DT, Turpin DH (eds) Plant physiology, biochemistry and molecular biology. Longman, Singapore, pp 249–262

Maier U-G (1992) The four genomes of the cryptomonad alga *Pyrenomonas salina*. BioSystems 28:69–74

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

Markos A, Miretsky A and 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 (1996) Is something wrong with the tree of life? BioEssays 18:523–527

Martin W (1997) Endosymbiosis and the origins of chloroplast-cytosol isoenzymes: revising the product-specificity corollary. In: Syvanen M, Kado C (eds) Horizontal gene transfer. Chapman Hall, London (in press)

Martin W, Cerff R (1986) Prokaryotic features of a nucleus-encoded enzyme: cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). Eur J Biochem 159:323–331

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

Martin W, Lagrange T, Li Y-F, Bisanz-Seyer C, Mache R (1990) Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. Curr Genet 18:553–556

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

Martin W, Jouannic S, Loiseaux-de Goer S (1993b) Molecular phylogeny of the *atp*B and *atp*E genes of the brown alga *Pylaiella littoralis*. Eur J Phycol 28:111–113

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

Martin W, Henze K, Kellermann J, Flechner A, Schnarrenberger C (1996b) Micro-sequencing and cDNA cloning of the Calvin-cycle/OPPP enzyme ribose-5-phosphate isomerase (EC 5.3.1.6) from spinach chloroplasts. Plant Mol Biol 30:795–805

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 GI, Reith ME, Munholland J, Lang-Unnasch N (1996) Plastids in human parasites. Nature 381:482

McFadden GI, Gilson PR, Douglas SE, Cavalier-Smith T, Hofmann CJ, Maier UG (1997) Bonsai genomics: sequencing the smallest eukaryotic genomes. Trends Genet 13:46–49

Meijer WG, van den Bergh ERE, Smith LM (1996) Induction of the *gap-pgk* operon encoding glyceraldehyde-3-phosphate dehydrogenase and phosphoglycerate kinase of *Xanthobacter flavus* requires the LysR-type transcriptional activator CbbR. J Bacteriol 178:881–887

Melkonian M (1996) Systematics and evolution of the algae: endocytobiosis and the evolution of the major algal lineages. Prog Bot 57:281–311

Meyer-Gauen G, Schnarrenberger C, Cerff R, Martin W (1994) Molecular characterization of a novel, nuclear-encoded, NAD^+-dependent glyceraldehyde-3-phosphate dehydrogenase in plastids of the gymnosperm *Pinus sylvestris* L. Plant Mol Biol 26:1155–1166

Milanez S, Mural RJ (1988) Cloning and sequencing of cDNA encoding the mature form of phosphoribulokinase from spinach. Gene 66:55–63

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

Morden CW, Delwiche C, Kuhsel M, Palmer JD (1992) Gene phylogenies and the endosymbiotic origin of plastids. BioSystems 28:75–90

Morse D, Salois P, Markovic P, Hastings JW (1995) A nuclear-encoded form-II RuBisCO in dinoflagellates. Science 268:1622–1624

Müller M (1973) Biochemical cytology of trichomonad flagellates. I. Subcellular localization of hydrolases, dehydrogenases and catalase. J Cell Biol 57:453–474

Müller M (1988) Energy metabolism of protozoa without mitochondria. Annu Rev Microbiol 42:465–488

Müller M (1993) The hydrogenosome. J Gen Microbiol 139:2879–2889

Müller M (1996) Energy metabolism of amitochondriate protists, an evolutionary puzzle. In: Schlegel M, Hausmann K (eds) Christian Gottfried Ehrenberg-Festschrift anläßlich der 14. Wissenschaftlichen Jahrestagung der Deutschen Gesellschaft für Protozoologie, 9.–11. Marz 1995 in Delitzsch (Sachsen). Leipziger Universitätsverlag, Leipzig, pp 63–76

Nelson T, Langdale JA (1992) Developmental genetics of C4 photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 43:25–47

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

Nowitzki U, Westhoff P, Henze K, Schnarrenberger C, Martin W (1995) Cloning of the amphibolic Calvin-cycle/OPPP enzyme D-ribulose-5-phosphate 3-epimerase (EC 5.1.3.1) from spinach chloroplasts: functional and evolutionary aspects. Plant Mol Biol 29:1279–1291

Palmer J (1995) Rubisco rules fall, gene transfer triumphs. Bioessays 17:1005–1008
Palmer J (1997) Organelle genomes: going, going, gone! Science 275:790–791
Peer Y van de, Rensing S, Maier U-G, De Wachter R (1996) Substitution-rate calibration of small subunit RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. Proc Natl Acad Sci USA 93:7744–7748
Pelzer-Reith B, Wiegand S, Schnarrenberger C (1994) Plastid class-I and cytosol class-II aldolase of *Euglena gracilis*. Plant Physiol 106:1137–1144
Plaumann M, Pelzer-Reith B, Martin W, Schnarrenberger C (1977) 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 WC (1996) The organization and regulation of plant glycolysis. Annu Rev Plant Physiol Plant Mol Biol 47:185–214
Portis AR (1992) Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase activity. Annu Rev Plant Physiol Plant Mol Biol 43:415–437
Pühler G, Leffers H, Gropp F, Palm P, Klenk H-P, Lottspeich F, Garrett R, Zillig W (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci USA 86:4569–4573
Qian Y, Tabita FR (1996) A global signal transduction system regulates aerobic and anaerobic CO_2 fixation in *Rhodobacter sphaeriodes*. J Bacteriol 178:12–18
Raines CA, Lloyd JC, Longstaff M, Bradley D, Dyer T (1988) Chloroplast fructose-1,6-bisphosphatase: the product of a mosaic gene. Nucleic Acids Res 16:7931–7942
Raines CA, Lloyd JC, Willingham NM, Potts S, Dyer TA (1992) cDNA and gene sequences of wheat chloroplast sedoheptulose-1,7-bisphosphatase reveal homology with fructose-1,6-bisphosphatases. Eur J Biochem 205:1053–1059.
Raines CA, Longstaff M, Lloyd JC and Dyer TA (1989) Complete nucleotide sequence of wheat phosphoribulokinase: developmental and light-dependent expression of the mRNA. Mol Gen Genet 220:43–48
Rajagopalan R, Altekar W (1994) Characterisation and purification of ribulose-bisphosphate carboxylase from heterotrophically grown halophilic archaebacterium, *Haloferax mediterranei*. Eur J Biochem 221:863–869
Reith M, Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol Biol Reptr 13:333–335
Reizer J, Ramseier TM, Reizer A, Charbit A, Saier MH (1996) Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative *N*-acetylgalactosamine metabolic pathway in *Escherichia coli*. Microbiology 142:231–250
Roesler KR, Ogren WL (1990) *Chlamydomonas reinhardtii* phosphoribulokinase: sequence, purification and kinetics. Plant Physiol 93:188–193
Roger AJ, Clark CG, Doolittle WF (1996) A possible mitochondrial gene in the early branching amitochondriate protist *Trichomonas vaginalis*. Proc Natl Acad Sci USA 93: 14618–14622
Roger AJ, Smith MW, Doolittle RF, Doolittle WF (1996) Evidence for the Heterolobosea from phylogenetic analysis of genes encoding glyceraldehyde-3-phosphate dehydrogenase. J Eukaryot Microbiol 43:475–485
Rowan R, Whitney SM, Fowler A, Yellowlees D (1996) Rubisco in marine symbiotic dinoflagellates: form II enzymes in eukaryotic oxygenic phototrophs encoded by a nuclear multigene family. Plant Cell 8:539–553
Rutter WJ (1964) Evolution of aldolase. Fed Proc 23:1248–1257
Schäferjohann J, Yoo J-G, Bowien B (1995) Analysis of the genes forming the distal parts of the two *cbb* CO_2-fixation operons from *Alcaligenes eutrophus*. Arch Microbiol 163:291–299
Scheibe R (1990) Light/dark modulation: regulation of chloroplast metabolism in a new light. Bot Acta 103:327–334
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-phosphate metabolism in green algae. In: Ogita S, Scandalios JS (eds) Isozymes: structure, function, and use in biology and medicine. Wiley-Liss, New York, pp 743–746
Schnarrenberger C, Pelzer-Reith B, Yatsuki H, Freund S, Jacobshagen S, Hori K (1994) Expression and sequence of the only detectable aldolase in *Chlamydomonas reinhardtii*. Arch Biochem Biophys 313:173–178
Schnarrenberger C, Flechner A, Martin W (1995) 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
Schönheit P, Schäfer T (1995) Metabolism of hyperthermophiles. World J Microbiol Biotechnol 11:26–57
Shih MC, Heinrich P, Goodman HM (1992) Cloning and chromosomal mapping of nuclear genes encoding chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenase from *Arabidopsis thaliana*. Gene 119:317–319
Smith MW, Feng D-F, Doolittle RF (1992) Evolution by acquisition: the case for horizontal gene transfers. Trends Bioch Sci 17:489–493
Smith TL (1989) Disparate evolution of yeasts and filamentous fungi indicated by analysis of glyceraldehyde-3-phosphate dehydrogenase genes. Proc Natl Acad Sci USA 86:7063–7066
Sørensen KI, Hove-Jensen B (1996) Ribose catabolism in *Escherichia coli*: characterization of the *rpiB* gene encoding ribose phosphate isomerase B and the *rpiR* gene, which is involved in regulation of *rpiB* expression. J Bacteriol 178:1003–1011
Spreitzer RJ (1993) Genetic dissection of Rubisco structure and function. Annu Rev Plant Physiol Plant Mol Biol 44:411–34
Starnes SM, Lambert DH, Maxwell ES, Stevens SE, Portis RD, Shively JM (1985) Co-transcription of the large and small subunit genes of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Cyanophora paradoxa*. FEMS Microbiol Lett 28: 165–196
Stirewalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ, Bryant DB (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol Biol Rep 13:327–332
Stitt M (1990) The flux of carbon between the chloroplast and the cytosol. In: Dennis DT, Turpin DH (eds) Plant physiology, biochemistry and molecular biology. Longman, Singapore, pp 309–326
Tabita FR (1988) Molecular and cellular regulation of autotrophic carbon-dioxide regulation in microorganisms. Microbiol Rev 52:155–189
Tabita FR (1994) The biochemistry and molecular regulation of carbon dioxide metabolism in cyanobacteria. In: Bryant DA (ed) The molecular biology of cyanobacteria. Kluwer, Dordrecht, pp 437–467
Teige M, Kopriva S, Bauwe H, Süß K-H (1995) Chloroplast ribulose-5-phosphate 3-epimerase from potato: cloning, cDNA sequence, and tissue-specific enzyme accumulation. FEBS Lett 377: 349–352
Tessier LH, Paulus F, Keller M, Vial C, Imbault P (1995) Structure and expression of *Euglena gracilis* nuclear *rbc*S genes encoding the small subunit of the ribulose 1,5-bisphosphate carboxylase/oxygenase: a novel splicing process for unusal intervening sequences. J Mol Biol 245:22–33
Wagner J, Lerner RA, Barbas III CF (1995) Efficient aldolase catalytic antibodies that use the enamine mechanism of natural enzymes. Science 270:1797–1800
Watson GM, Tabita FR (1996) Regulation, unique gene organization, and unusual primary structure of carbon fixation genes from a marine phycoerythrin-containing cyanobacterium. Plant Mol Biol 32:1103–1115
Weeden NF (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J Mol Evol 17:133–139

Williamson DH, Gardner MJ, Preise P, Moore DJ, Rangachari K, Wilson RJ (1994) The evolutionary origin of the 35-kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry. Mol Gen Genet 243:249–252

Windhövel U, Bowien B (1991) Identification of *cfxR*, an activator gene of autotrophic carbon dioxide fixation in *Alcaligenes eutrophus*. Mol Microbiol 5:2695–2706

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

Wolosiuk RA, Ballicora MA, Hagelin K (1993) The reductive pentose phosphate cycle for photosynthetic CO_2 assimilation: enzyme modulation. FASEB J 7:622–637

Wolter F, Fritz C, Willmitzer L, Schell J, Schreier P (1988) *rbc*S genes in *Solanum tuberosum*: conservation of transit peptide and exon shuffling during evolution. Proc Natl Acad Sci USA 85: 846–850

Woodrow IE, Berry JA (1988) Enzymatic regulation of photosynthetic CO_2 fixation in C3 plants. Annu Rev Plant Physiol Plant Mol Biol 39:533–594

Yoo JG, Bowien B (1995) Analysis of the *cbbF* genes from *Alcaligenes eutrophus* that encode fructose-1,6-/sedoheptulose-1,7-bisphosphatase. Curr Microbiol 31:55–61

Yu JP, Ladapo J, Whitman WB (1994) Pathway of glycogen metabolism in *Methanococcus maripaludis*. J Bacteriol 176: 325–332

Zhao G, Pease AJ, Bharani N, Winkler ME (1995) Biochemical characterization of gapB-encoded erythrose 4-phosphate dehydrogenase of *Escherichia coli* K-12 and its possible role in pyridoxal 5′-phosphate biosynthesis. J Bacteriol 177:2804–2812

Zillig W (1991) Comparative biochemistry of Archaea and Bacteria. Curr Opin Genet Dev 1:544–551

Zillig W, Klenk H-P, Palm P, Leffers H, Pühler G, Gropp F, Garrett RA (1989) Did eukaryotes originate by a fusion event? Endocyt Cell Res 6:1–25

Zwickl P, Fabry S, Bogedian C, Haas A, Hensel R (1990) Glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaebacterium *Pyrococcus woesei*: characterisation of the enzyme, cloning and sequencing of the gene, and expression in *Escherichia coli*. J Bacteriol 172:4329–4338