Eur. J. Biochem. **269**, 868–883 (2002) © FEBS 2002

Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants

A case study of endosymbiotic gene transfer

Claus Schnarrenberger[1] and William Martin[2]

[1]*Institut für Biologie, Freie Universität Berlin, Germany;* [2]*Institut für Botanik III, Universität Düsseldorf, Germany*

The citric acid or tricarboxylic acid cycle is a central element of higher-plant carbon metabolism which provides, among other things, electrons for oxidative phosphorylation in the inner mitochondrial membrane, intermediates for amino-acid biosynthesis, and oxaloacetate for gluconeogenesis from succinate derived from fatty acids via the glyoxylate cycle in glyoxysomes. The tricarboxylic acid cycle is a typical mitochondrial pathway and is widespread among α-proteobacteria, the group of eubacteria as defined under rRNA systematics from which mitochondria arose. Most of the enzymes of the tricarboxylic acid cycle are encoded in the nucleus in higher eukaryotes, and several have been previously shown to branch with their homologues from α-proteobacteria, indicating that the eukaryotic nuclear genes were acquired from the mitochondrial genome during the course of evolution. Here, we investigate the individual evolutionary histories of all of the enzymes of the tricarboxylic acid cycle and the glyoxylate cycle using protein maximum likelihood phylogenies, focusing on the evolutionary origin of the nuclear-encoded proteins in higher plants. The results indicate that about half of the proteins involved in this eukaryotic pathway are most similar to their α-proteobacterial homologues, whereas the remainder are most similar to eubacterial, but not specifically α-proteobacterial, homologues. A consideration of (a) the process of lateral gene transfer among free-living prokaryotes and (b) the mechanistics of endosymbiotic (symbiont-to-host) gene transfer reveals that it is unrealistic to expect all nuclear genes that were acquired from the α-proteobacterial ancestor of mitochondria to branch specifically with their homologues encoded in the genomes of contemporary α-proteobacteria. Rather, even if molecular phylogenetics were to work perfectly (which it does not), then some nuclear-encoded proteins that were acquired from the α-proteobacterial ancestor of mitochondria should, in phylogenetic trees, branch with homologues that are no longer found in most α-proteobacterial genomes, and some should reside on long branches that reveal affinity to eubacterial rather than archaebacterial homologues, but no particular affinity for any specific eubacterial donor.

Keywords: glyoxysomes; microbodies; mitochondria; pathway evolution, pyruvate dehydrogenase.

Metabolic pathways are units of biochemical function that encompass a number of substrate conversions leading from one chemical intermediate to another. The large amounts of accumulated sequence data from prokaryotic and eukaryotic sources provide novel opportunities to study the molecular evolution not only of individual enzymes, but also of individual pathways consisting of several enzymatic substrate conversions. This opens the door to a number of new and intriguing questions in molecular evolution, such as the following. Were pathways assembled originally during the early phases of biochemical evolution, and subsequently been passed down through inheritance ever since? Do pathways evolve as coherent entities consisting of the same group of enzyme-coding genes in different organisms? Do they evolve as coherent entities of enzymatic activities, the individual genes for which can easily be replaced? Do they evolve as coherent entities at all? During the endosymbiotic origins of chloroplasts and mitochondria, how many of the biochemical pathways now localized in these organelles were contributed by the symbionts and how many by the host?

One approach to studying pathway evolution is to use tools such as BLAST [1] to search among sequenced genomes for the presence and absence of sequences similar to individual genes. This has been carried out for the glycolytic pathway, for example [2]. However, the presence or absence of a gene bearing sequence similarity to a query sequence for a given enzyme makes no statement about the relatedness of the sequences so identified, hence such information does not reveal the evolution of a pathway at all because lateral gene transfer, particularly among prokaryotes, can, in principle, result in mosaic pathways consisting of genes acquired from many different sources [3–5].

In previous work, our approach to the study of pathway evolution has been based on conventional phylogenetic analysis for all of the enzymes of an individual pathway and comparison of trees obtained for the individual enzymes of the pathway, to search for general patterns of phylogenetic

Correspondence to C. Schnarrenberger, Institut für Biologie, Königin-Luise-Str. 12–16a, 14195 Berlin, Germany. Fax: + 030 8385 4313, Tel.: + 030 8385 3123, E-mail: schnarre@zedat.fu-berlin.de

Abbreviations: TCA, tricarboxylic acid; PDH, pyruvate dehydrogenase; OGDH, α-oxoglutarate dehydrogenase; OADH, α-oxoacid dehydrogenase; CS, citrate synthase; IRE-BP, iron-responsive element-binding protein; IPMI, isopropylmalate isomerase; ICDH, isocitrate dehydrogenase; STK, succinate thiokinase; SDH, succinate dehydrogenase; ICL, isocitrate lyase; MS, malate synthase.

(Received 27 July 2001, accepted 3 December 2001)

similarity or disconcordance among enzymes. This has been performed for the Calvin cycle (a pathway of CO_2 fixation that consists of 11 different enzymes [3,6]), the glycolytic/gluconeogenic pathway [3,6], and the two different pathways of isoprenoid biosynthesis [7]. Recently, the evolution of the biosynthetic pathway leading to vitamin B6 was studied in detail [8], as was the evolution of the chlorophyll-biosynthetic pathway [9]. In essence, these studies revealed a large degree of mosaicism within the pathways studied in both prokaryotes and eukaryotes. These findings indicate that pathways tend to evolve as coherent entities of enzymatic activity, the individual genes for which can, however, easily be replaced by intruding genes of equivalent function acquired through lateral transfer. Very similar conclusions were reached through the phylogenetic analysis of 63 individual genes belonging to many different functional categories among prokaryotes and eukaryotes [10] and through the distance analysis of normalized BLAST scores of several hundred genes common to six sequenced genomes [11].

In prokaryotes, there are several well-known mechanisms of lateral gene transfer, including plasmid-mediated conjugation, phage-mediated transduction, and natural competence [4,5,12,13]. In eukaryotes, by far the most prevalent form of lateral transfer documented to date is endosymbiotic gene transfer, i.e. the mostly unidirectional donation of genes from organelles to the nucleus during the process of organelle genome reduction in the wake of the endosymbiotic origins of organelles from free-living prokaryotes [3,6,14–20]. By studying the evolution of nuclear-encoded enzymes of pathways that are biochemically compartmentalized in chloroplasts and mitochondria and thought to have once been encoded in the respective organellar DNA, one can gain insights into the evolutionary dynamics of (a) pathway evolution, (b) organelle-to-nucleus gene transfer, and (c) the rerouting of nuclear-encoded proteins into novel evolutionary compartments.

In eukaryotes, the citric acid cycle (Krebs cycle, or tricarboxylic acid cycle) is an important pathway in that it is the primary source of electrons (usually stemming from pyruvate) donated to the respiratory membrane in mitochondria. It is not ubiquitous among eukaryotes, because not all eukaryotes possess mitochondria [21,22]. In anaerobic mitochondria, it occurs in a modified (shortened) form suited to fumarate respiration [23]. In *Euglena* it occurs in a modified form lacking α-oxoglutarate dehydrogenase (OGDH), a variant also found in the α-proteobacterium *Bradyrhizobium japonocum* [24]. The enzymatic framework of the tricarboxylic acid cycle was formulated by Krebs & Johnson [25] at a time when endosymbiotic theories for the origins of organelles were out of style (see [26]). Sixty-four years later, gene-for-gene phylogenetic analysis can provide insights into the origin of its individual enzymes.

However, the study of the enzymes of the tricarboxylic acid cycle necessarily also entails the study of the several enzymes involved in the glyoxylate cycle in plants, because three enzymatic steps common to both the tricarboxylic acid cycle and the glyoxylate cycle are catalyzed by differentially compartmentalized isoenzymes, analogous to the chloroplast cytosol isoenzymes involved in the Calvin cycle and glycolysis in plants. The glyoxylate cycle was discovered in bacteria by Kornberg & Krebs [27] as a means of converting C_2 units of acetate (a growth substrate) for synthesis of other cell constituents such as hexoses. The same cycle was subsequently found in germinating castor beans to convert acetyl-CoA from fat degradation into succinate and subsequently carbohydrates during conversion of fat into carbohydrate [28]. The enzymes of the glyoxylate cycle were later found to be associated in a novel organelle of plants, the glyoxysome [29]. The cycle apparently operates in all cells that have the capacity to convert acetate to carbohydrates, including eubacteria, plants, fungi, lower animals, and also mammals [30]. The glyoxylate cycle involves five enzyme activities that are all compartmentalized in the glyoxysomes of plants [31], the single exception being aconitase, which is localized in the cytosol [32,33]. Here we investigate the evolution of the enzymes of the pyruvate dehydrogenase (PDH) complex, the tricarboxylic acid cycle, and the glyoxylate cycle by examining the individual phylogenies of the 21 subunits comprising the 14 enzymes of these pathways as they occur in eukaryotes, specifically in higher plants.

MATERIALS AND METHODS

Amino-acid sequences for individual plant tricarboxylic acid cycle and glyoxylate cycle enzymes and their constituent subunits were extracted from the databases and compared with GenBank using BLAST [1]. We were frequently confronted with more than 400 hits per enzyme. To be able to make sense out of the data and in order to make the phylogenies tractable, we had to limit the number of proteins to be retrieved for analysis. In selecting sequences, we tried to include at least three sequences from plants, animals, and fungi, in addition to a representative sample of gene diversity and ancient gene families from eubacteria and archaebacteria. In some cases, homologues were not available from all sources. Furthermore, in the eukaryotes, particular care was taken to include sequences for the various compartment-specific isoenzymes (mitochondria, glyoxysomes, plastids and the cytosol where relevant). Importantly, very few homologues for these sequences from protists or algae were available in GenBank.

In the bacteria, we tried to include homologues from α-proteobacteria and cyanobacteria because they are thought to be the progenitors of mitochondria and plastids, respectively. However, the spectrum of α-proteobacteria and cyanobacteria available for comparison is limited. Homologues of these enzymes from achaebacteria were, in general, extremely scarce and were included where ever possible. Classes of enzymes were defined as proteins that show marginal (< 25%) amino-acid sequence identity.

Sequences were aligned using PILEUP from the Wisconsin package [34] and formatted using CLUSTALW [35]. Regions of alignment in which more than half of the positions possessed gaps were excluded from analysis. Trees were inferred with the MOLPHY package [36] using PROTML with the JTT-F martix and starting from the NJ tree of ML distances. We often encountered distantly related genes encoding related protein families for different enzyme activities. These were usually included in the analysis if they helped to elucidate a general evolution pattern within a gene family, but at the same time, without overloading the data.

RESULTS

Inferring the evolutionary history of a biochemical pathway on an enzyme-for-enzyme basis is more challenging than it might seem at first sight. In the case of the tricarboxylic acid cycle, many enzymes consist of multiple subunits. The only way we see to approach the problem is to analyze one enzyme at a time and, if applicable, one subunit at a time, describing the reaction catalyzed, some information about the enzyme, its subunits, and their evolutionary affinities. This is given in the following for the enzymes studied here.

Pyruvate dehydrogenase (PDH)

$$\text{Pyruvate} + \text{NAD}^+ + \text{CoASH} \rightarrow \text{acetyl-CoA} + \text{NADH} + \text{CO}_2$$

Pyruvate enters the tricarboxylic acid cycle through the action of PDH, a thiamine-dependent mitochondrial enzyme complex with several nonidentical subunits. Plants possess an additional PDH complex in plastids. The subunits of PDH are designated E1 (EC 1.2.4.1), E2 (EC 2.3.1.12) and E3 (EC 1.8.1.4), and E1 consists of two subunits, E1α and E1β. The reaction catalyzed by PDH (oxidative decarboxylation of an organic acid with a keto group at the α carbon) is mechanistically very similar to the reactions catalyzed by OGDH and by branched-chain α-oxoacid dehydrogenases (OADH). It is therefore not surprising that all three enzymes have an E1, E2, E3 subunit structure, and that some of the subunits of PDH, OGDH and OADH are related. The functional and evolutionary relationships between the subunits of these enzymes are somewhat complicated. In a nutshell, the E1α subunits of PDH and OADH are closely related to one another (≈ 30% identity) and more distantly related (≈ 20% identity) to the E1 subunit of OGDH, which has a single E1 subunit, rather than an E1α/E1β structure. The E1β subunits of PDH and OADH are also closely related to one another (≈ 30% identity) and more distantly related (≈ 20% identity) to the 'class II' E1β subunit of several eubacteria. The E2 subunits of PDH, OGDH and OADH (dihydrolipoamide acyl transferase; EC 2.3.1.12) share about 35% identity.

The tree of PDH E1α subunits (Fig. 1A) contains three branches in which eubacterial and eukaryotic sequences are interleaved. One branch relates mitochondrial E1α to α-proteobacterial homologues, a second connects E1α of chloroplast PDH to cyanobacterial homologues, and a third branch connects E1α of mitochondrial branched-chain OADHs to eubacterial homologues. No α-proteobacterial homologues of mitochondrial OADH E1α were found. The E1 subunit of mitochondrial OGDH (Fig. 1B) branches with α-proteobacterial homologues.

The tree of the E1β subunit of PDH and OADH (Fig. 1C) has the same overall shape as that found for the E1α subunit. Namely, chloroplast and mitochondrial PDH E1β branch with cyanobacterial and α-proteobacterial homologues, respectively, whereas the related OADH E1β does not. The E1β subunit occurs as a class II enzyme in some eubacteria (Fig. 1D) that is only distantly related to the class I enzyme (Fig. 1C). But both the class I and class II E1β (Fig. 1C,D) are related at the level of sequence similarity (≈ 20–30% identity) and tertiary structure [37,38] to other thiamine-dependent enzymes that perform biochemically similar reactions: transketolase, which catalyzes the transfer of two-carbon units in the Calvin cycle and oxidative pentose phosphate pathway, 1-deoxyxylulose-5-phosphate synthase, which transfers a C_2 unit from pyruvate to D-glyceraldehyde 3-phosphate in the first step of plant isoprenoid biosynthesis [7], and pyruvate–ferredoxin oxidoreductase, an oxygen-sensitive homodimeric enzyme that performs the oxidative decarboxylation of pyruvate in hydrogenosomes [21,22] and in *Euglena* mitochondria [39].

The E2 subunit of PDH contains the dihydrolipoamide transferase activity. The mitochondrial form of the E2 subunit for PDH is related to the E2 subunits of OADH and OGDH. All three E2 subunits in eukaryotes are encoded by an ancient and diverse eubacterial gene family which is largely preserved in eukaryotic chromosomes (Fig. 1E). Mitochondrial PDH E2 and OGDH E2 branch very close to α-proteobacterial homologues, whereas chloroplast PDH E2 branches with the cyanobacterial homologue. Mitochondrial OADH branches with eubacterial, but not specifically with, α-proteobacterial homologues (Fig. 1E).

The E3 subunit of PDH contains the dihydrolipoamide dehydrogenase activity. Mitochondrial PDH, OGDH and OADH all use the same E3 subunit [40]; it branches with α-proteobacterial homologues (Fig. 1F). The chloroplast PDH E3 subunit branches with its cyanobacterial homologue (Fig. 1F). The E3 subunit is related to eubacterial mercuric reductase and eukaryotic glutathione reductase.

In general, one can conclude that all four nuclear-encoded subunits of the mitochondrial PDH complex are acquisitions from the α-proteobacterial ancestor of mitochondria, whereas the four subunits of nuclear-encoded chloroplast PDH are acquisitions from the cyanobacterial ancestor of plastids. The E1α and E1β subunits of chloroplast PDH are even still encoded in the chloroplast genome of the red alga *Porphyra* [41], the genes having been transferred to the nucleus in higher plants (Fig. 1A,C).

Citrate synthase (CS)

$$\text{Oxalacetate} + \text{acetyl-CoA} \rightarrow \text{citrate} + \text{CoASH}$$

In eukaryotes, CS (EC 4.1.3.7) is usually found as isoenzymes in mitochondria and glyoxysomes, respectively [42,43]. They usually have a molecular mass of ≈ 90 kDa and are typically homodimers of 45-kDa subunits [44,45]. In the presence of Mg^{2+}, glyoxysomal CS of plants also forms tetramers [43]. However, there are also a number of bacteria for which the molecular mass of the enzyme has been reported to be ≈ 280 kDa or even more [46]. Many regulatory compounds [NADH, α-oxoglutarate, 5,5′-dithiobis-(2-nitrobenzoic acid), AMP, ATP, KCl, aggregation state] can influence the CS activity from various sources [46–48].

The tree of CS enzymes is shown in Fig. 2A. The mitochondrial enzymes of plants, animals, and fungi in addition to the fungal peroxisomal CS enzymes are separated from the remaining sequences by a very long branch. The peroxisomal enzyme of fungi arose through duplication of the gene for the mitochondrial enzyme during fungal evolution. By contrast, the glyoxysomal

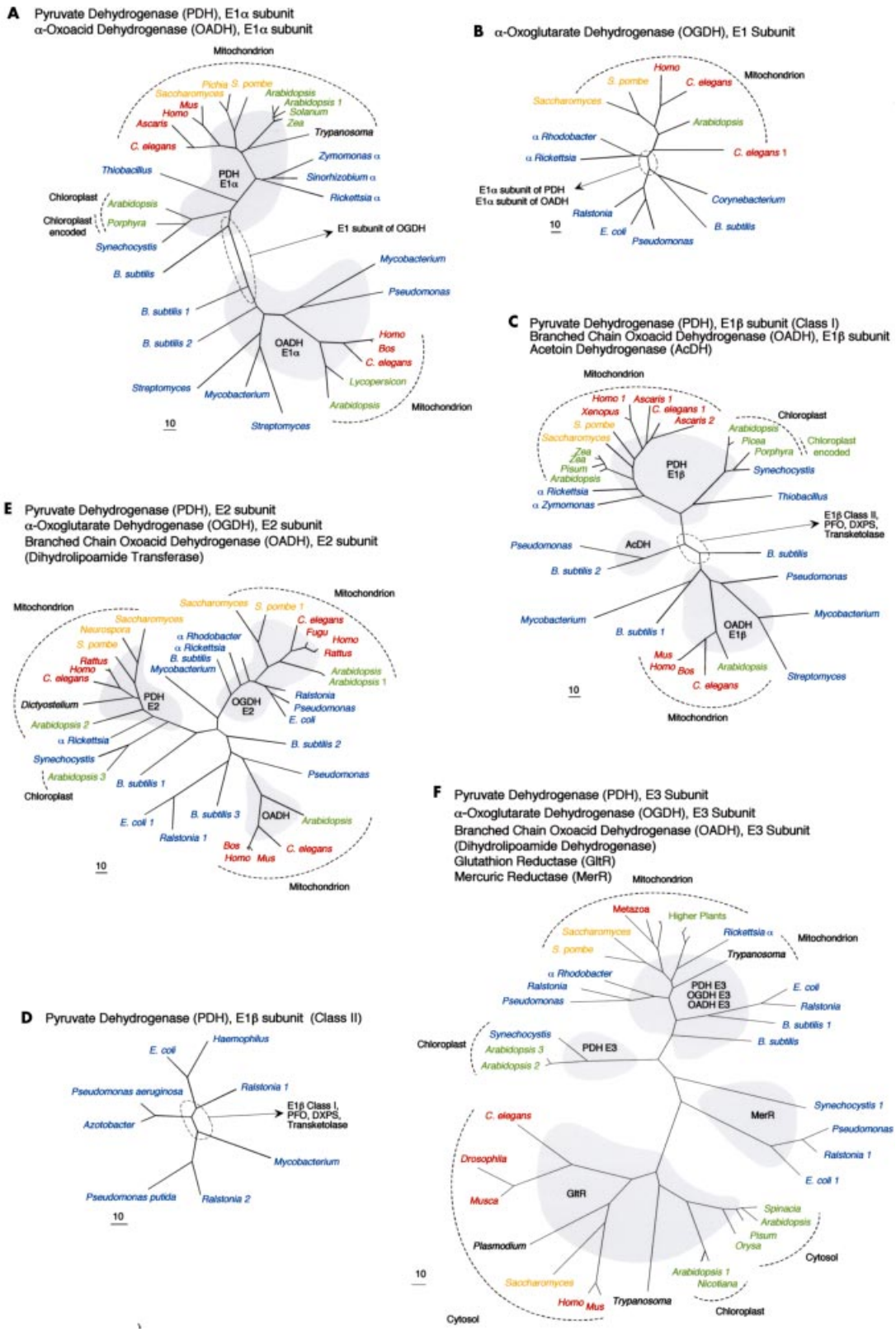

Fig. 1. Phylogenetic results. Protein maximum likelihood trees for PDH and OGDH subunits (see text). Color coding of species names is: metazoa, red; fungi, yellow; plants, green; protists, black; eubacteria, blue; archaebacteria, purple. Protein localization is indicated as is organelle-coding of individual genes (for example, α and β subunits of *Porphyra* PDH E1.

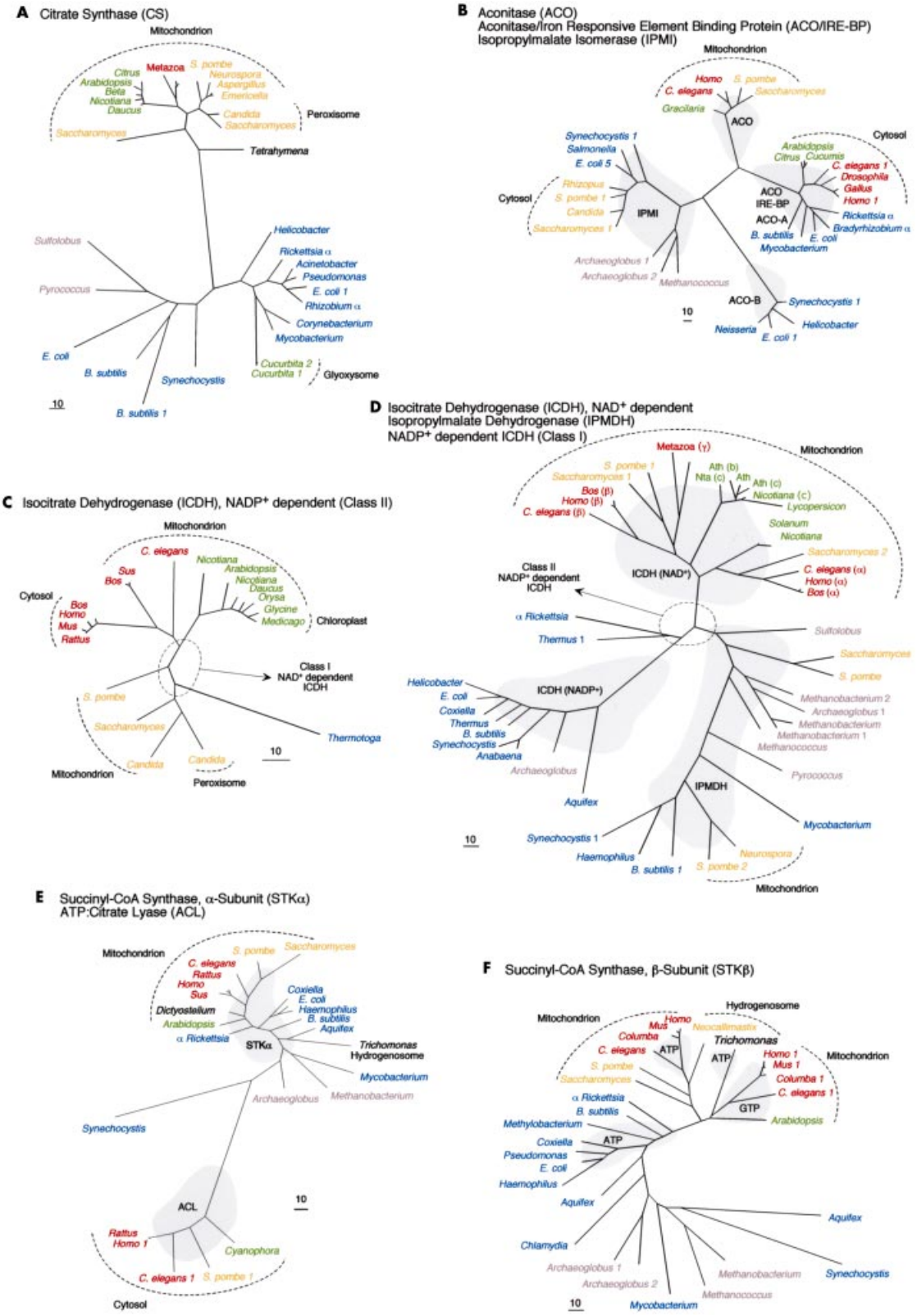

Fig. 2. Phylogenetic results. Protein maximum likelihood trees for CS, aconitase, ICDH ($NADP^+$), ICDH (NAD^+) and the α and β subunits of STK (see text). Color coding of species names is as in Fig. 1.

enzyme of plants branches within a cluster of eubacterial enzymes, suggesting that this gene was acquired from eubacteria; however, it branches with neither α-proteobacterial nor cyanobacterial homologues. Notwithstanding the fact that long branches are notoriously difficult to place correctly in a topology, the position of the long

branch bearing the eukaryotic genes for the mitochondrial (and fungal peroxisomal) enzymes is notable, because it places these enzymes within a tree of eubacterial genes. Thus, the eukaryotic enzymes seem to be more similar to eubacterial than to archaebacterial homologues (which exist for this enzyme), although a specific evolutionary affinity for a particular group of eubacterial enzymes is not evident.

Aconitase

$$\text{Citrate} \rightarrow \text{isocitrate}$$

Aconitase (EC 4.2.1.3) contains a 4Fe–4S cluster and is usually a monomer. There are two isoenzymes in eukaryotes: mitochondrial and cytosolic. Another activity of cytosolic aconitase, at least in animals, is that of an iron-responsive element-binding protein (IRE-BP), which binds to mRNA of ferritin and the transferrin receptor and thus participates in regulating iron metabolism in animals [49,50]. The latter activity is accomplished by a transition from the 4Fe–4S state of the protein (active form of aconitase) to a 3Fe–4S state (inactive as aconitase, but active as IRE-BP). Two forms of aconitase are known in eubacteria, aconitase A and aconitase B [51–53]. They are differently expressed [54]. Isopropylmalate isomerase (IPMI), which is involved in the biosynthetic pathway to leucine, is related to the aconitases.

The sequences of aconitase, IRE-BP and IPMI belong to a highly diverse gene family (Fig. 2B). The true aconitases, which include IRE-BP, are large enzymes (780–900 amino acids). The bacterial IPMI genes encode much smaller proteins (about 400 amino acids) than the fungal IMPI genes (about 760 amino acids). Cytosolic aconitase/IRE-BP from plants and animals is closely related to the eubacterial aconitase homologues termed here aconitase A. The sequences for eubacterial aconitase B proteins fall into a separate gene cluster and are only distantly related ($\approx$ 20% identity) with the eubacterial aconitase A enzymes, but share $\approx$ 30% identity with archaebacterial IPMI, indicating a nonrandom level of sequence similarity. Although we detected genes for three different aconitase isoenzymes in the *Arabidopsis* genome data, we did not detect one with a mitochondrion-specific targeting sequence. Although the eukaryotic cytosolic enzymes (aconitase and IRE-BP) do not branch specifically within eubacterial aconitase A sequences, they branch very close to them, and a case could be made for a eubacterial origin of the cytosolic enzyme, homologues of which were not found among archaebacteria. Database searching revealed no clear-cut prokaryotic homologue to the mitochondrial enzyme, the sequences of which have a very distinct position in the tree (Fig. 2B). IPMI from fungi is more closely related to eubacterial than to archaebacterial homologues, and appears to be a eubacterial acquisition.

Isocitrate dehydrogenase (ICDH)

$$\text{Isocitrate} + \text{NAD}^+ \rightarrow \alpha\text{-oxoglutarate} + \text{NADH}$$

$$\text{Isocitrate} + \text{NADP}^+ \rightarrow \alpha\text{-oxoglutarate} + \text{NADPH}$$

Two distinct types of ICDH (EC 1.1.1.41) exist which differ in their specificity for NAD^+ and $NADP^+$, respectively, and which share $\approx$ 30% sequence identity. Both enzymes are found in typical mitochondria, but the $NADP^+$-dependent enzyme can be localized in other eukaryotic compartments as well. The NAD^+-dependent enzyme is typically an octamer consisting of identical or related subunits [55,56]; however, dimeric forms have been characterized in archaebacteria [57]. Sequences of eukaryotic NAD-ICDH and NADP-ICDH share about 30% identity; the former shares about 40% identity with prokaryotic NADP-ICDH homologues and with isopropylmalate dehydrogenase, which is involved in leucine biosynthesis. Thus, in the case of aconitase/IPMI and NADP-ICDH/isopropylmalate dehydrogenase, consecutive and mechanistically related steps in the tricarboxylic acid cycle and leucine biosynthesis are catalyzed by related enzymes.

The evolutionary trees of class II NADP-ICDH (Fig. 2C) and NAD-ICDH plus class I NADP-ICDH (Fig. 2D) are somewhat complicated. The mitochondrial, peroxisomal, chloroplast and cytosolic forms of class II $NADP^+$-dependent ICDH in eukaryotes seem to have arisen from a single progenitor enzyme, with various processes of recompartmentalization of the enzyme having occurred during eukaryotic evolution. Direct homologues of this enzyme in prokaryotes are rare, one having been identified in the *Thermotoga* genome (Fig. 2C). Yet there is a clear but distant relationship with the NAD^+-dependent and class I $NADP^+$-dependent ICDH enzymes, which are found in eubacteria, archaebacteria and eukaryotes (Fig. 2D). The mitochondrial NAD-ICDH of eukaryotes has about as much similarity to an α-proteobacterial homologue as it does to the homologue from the archaebacterium *Sulfolobus* (Fig. 2D), so the evolutionary origin of this enzyme remains unresolved. The mitochondrial isopropylmalate dehydrogenase of fungi is clearly descended from eubacterial homologues (Fig. 2D).

α-Oxoglutarate dehydrogenase (OGDH)

$$\alpha\text{-Oxoglutarate} + \text{NAD}^+ + \text{CoASH} \rightarrow \text{succinyl-CoA} + \text{NADH} + \text{CO}_2$$

Like PDH and its relative OADH, OGDH consists of several nonidentical subunits. Subunit E1 (EC 1.2.4.2) is involved in substrate and cofactor (thiamine pyrophosphate) binding, subunit E2 (EC 2.3.1.61) is a dihydrolipoamide succinyl transferase, and subunit E3 (EC 1.8.1.4) is a dihydrolipoamide dehydrogenase. E1 and E2 are different proteins in OGDH, PDH, and OADH, but all three enzymes use one and the same E3 subunit. In eukaryotes, OGDH is thought to be located exclusively in the mitochondria.

The tree of OGDH E1 indicates that the eukaryotic sequences of animals, plants and fungi are most similar to homolgues in α-proteobacteria (Fig. 1B). As mentioned in the section on PDH above, the OGDH E1 subunit is related to the E1α subunit of PDH and OADH. The tree of eukaryotic OGDH E2 subunits also indicates a very close relationship to α-proteobacterial homologues (Fig. 1E). The OGDH E2 tree also indicates an early differentiation within eubacteria of PDH-specific, OADH-specific and

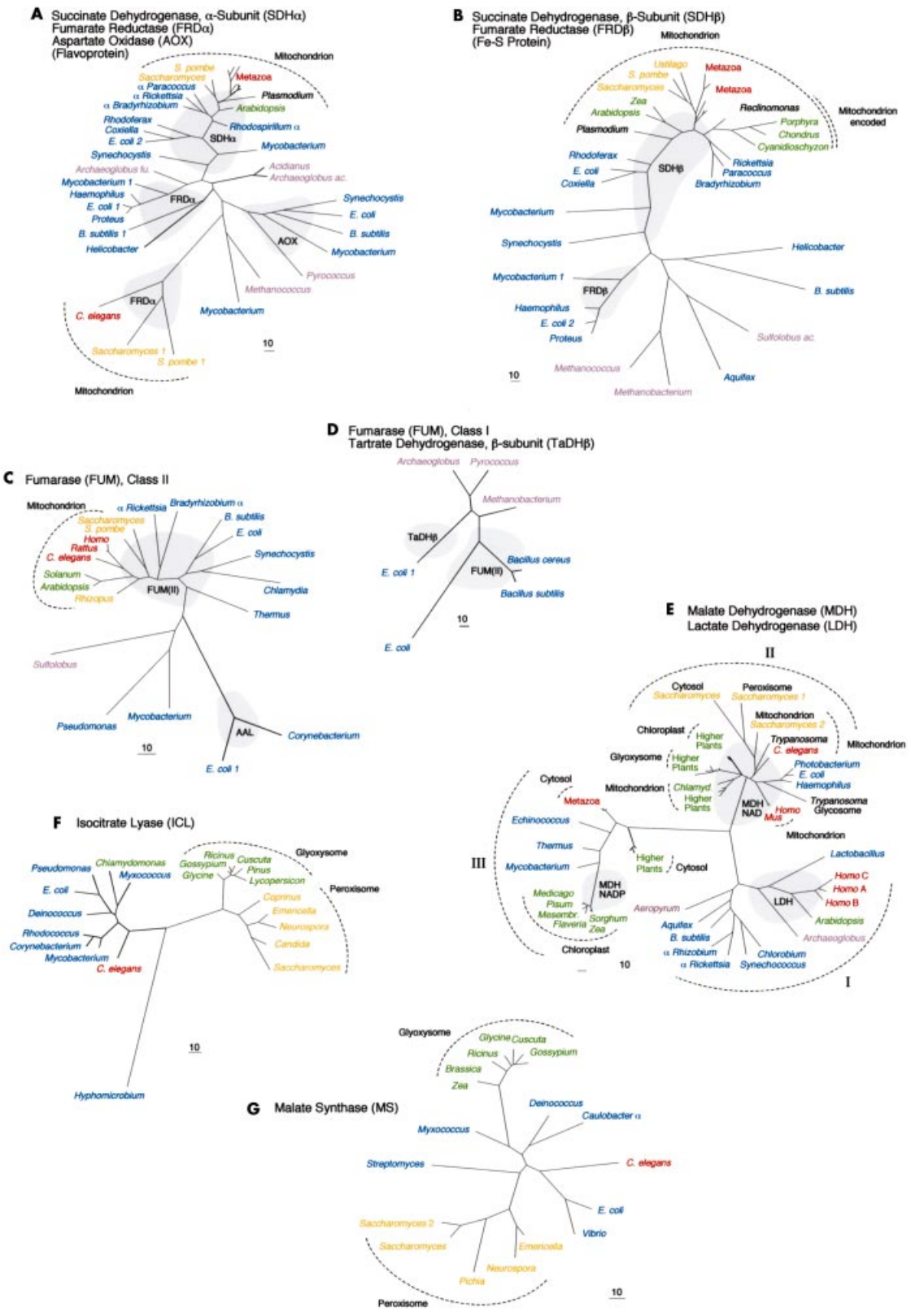

Fig. 3. Phylogenetic results. Protein maximum likelihood trees for the α and β subunits of SDH, class I and class II fumarase, MDH, ICL, and MS (see text). Color coding of species names is as in Fig. 1.

OGDH-specific subunits. Archaebacteria, which preferentially use the distantly related ferredoxin-dependent pyruvate–ferredoxin oxidoreductase and α-oxoacid–ferredoxin oxidoreductases instead of the corresponding NAD-dependent dehydrogenases, seem to lack clear homologues for E1, E2 and E3 subunits. The tree for OGDH E3 (Fig. 1F)

differs from the trees for E1 and E2 in that it contains branches encoding additional enzyme activities, glutathione reductase and mercuric reductase. Eukaryotic OGDH E3 is most similar to α-proteobacterial homologues. The eukaryotic glutathione reductases are roughly 30% identical with OGDH and are cytosolic enzymes, except in plants where an additional plastid isoenzyme exists. The cluster of glutathione reductases has split in early eukaryote evolution to produce plant and animal sequences. The two isoenzymes in the plant kingdom originated from a gene duplication in early plant evolution.

Succinate thiokinase (STK)

$$\text{Succinate} + \text{GTP(orATP)} + \text{CoASH} \rightarrow \text{succinyl-CoA} + PP_i + \text{GMP(orAMP)}$$

STK (EC 6.2.1.5) is also known as succinyl-CoA synthase; it consists of α and β subunits. It is usually an $\alpha_2\beta_2$ heterotetramer, but in some Gram-negative eubacteria it can have an $\alpha_4\beta_4$ structure. The β subunit carries the specificity for either ATP (EC 6.2.1.5) or GTP (EC 6.2.1.4). In eukaryotes, the enzyme is localized only in mitochondria or hydrogenosomes anaerobic forms of mitochondria that are found in some amitochondriate protists [21,22].

The sequences of STK α and β subunits have no siginificant sequence similarity to each other. Homologues are found in eukaryotes, eubacteria and archaebacteria for both STKα (Fig. 2E) and for STKβ (Fig. 2F). In the tree of the β subunits (Fig. 2F), a common ancestry for the GTP-specific and ATP-specific eukaryotic sequences is seen. In both trees (α and β), the eukaryotic STKs branch with α-proteobacterial homologues, with the single exception of the hydrogenosomal STKα, which, unlike STKβ, shows a slightly longer, and thus perhaps unreliably placed, branch. The STKα subunit is related to the C-terminus of eukaryotic cytosolic ATP-citrate lyases, which are homotetrameric proteins, and the STKβ subunit is related to the N-terminus of ATP-citrate lyases [113].

Succinate dehydrogenase (SDH)

$$\text{Succinate} + \text{FAD} \rightarrow \text{fumarate} + FADH_2$$

SDH (EC 1.3.5.1) is located in mitochondria and is attached to the inner membrane, where it is a component of complex II, which contains a cytochrome *b*, an anchor protein, and several additional subunits in the inner mitochondrial membrane. SDH consists of nonidentical subunits. The α subunit (SDHα) is a 70-kDa flavoprotein and possesses a [2Fe–2S] cluster. The β subunit is 30 kDa in size and has a [4Fe–4S] cluster. The electrons that are donated to the flavin cofactor of SDH are ultimately donated within complex II to quinones in the respiratory membrane. SDH is related to fumarate reductase. In some prokaryotes and eukaryotes, under anaerobic conditions, there is a preference for fumarate reductase to produce succinate, because of the presence of different kinds of quinones (with redox potentials better suited to fumarate reductase) in the respiratory membrane under anaerobic conditions [23]. Structures for fumarate reductase have been determined [58]. The SDH α subunit is also related to aspartate oxidase found in some prokaryotes.

The tree for the SDH α subunit (Fig. 3A) shows that the nuclear-encoded mitochondrial protein in eukaryotes is most similar to α-proteobacterial homologues. Proteins related to both the α and β subunits of SDH are also found in archaebacteria. The SDH β subunit in eukaryotes is also most closely related to the homologue from α-proteobacteria (Fig. 3B), indicating a mitochondrial origin for the eukaryotic gene. Very unusually for tricarboxylic acid cycle enzymes, the SDH β subunit it still encoded in the mitchondrial DNA, but only in a few protists [59]. Although their proteins branch slightly below the α-proteobacterial homologues in Fig. 3B, the genes for SDHβ from plants and *Plasmodium* were very probably also acquired from the mitochondrion.

Fumarase

$$\text{Fumarate} + H_2O \rightarrow \text{L-malate}$$

Fumarase (EC 4.2.1.2) catalyzes the reversible addition of a water molecule to the double bond of fumarate to produce L-malate. The enzyme occurs as class I and class II types which have no detectable sequence similarity. Class I fumarases have only been found in prokaryotes to date whereas class II fumarases, the more widespread of the two enzymes, are found in archaebacteria, eubacteria and eukaryotes. The class II fumarases are typically homotetramers of ≈ 50-kDa subunits [60,61]. In eukaryotes the enzyme is almost exclusively restricted to mitochondria. In some vertebrates, such as rat, there is an additional cytosolic enzyme, which is encoded by the same gene as the mitochondrial enzyme and which is produced by an alternative translation-initiation site [62].

The class II fumarases represent a group of highly conserved sequences; the mitochondrial enzyme in the eukaryotic tricarboxylic acid cycle is most closely related to α-proteobacterial homologues (Fig. 3C), indicating that the genes were acquired from the mitochondrial symbiont. More distantly related to the class II fumarases are genes in *Escherichia coli* and *Corynebacterium* encoding aspartate ammonia lyase activity. Class I fumarases and related sequences, including the β subunit of the heterotetrameric tartrate dehydrogenase from *E. coli*, are found in eubacteria and archaebacteria (Fig. 3D).

Malate dehydrogenase (MDH)

$$\text{Malate} + NAD^+ \rightarrow \text{oxalacetate} + \text{NADH} + H^+$$

$$\text{Malate} + NADP^+ \rightarrow \text{oxalacetate} + \text{NADPH} + H^+$$

MDH catalyzes the reversible oxidation of L-malate to oxalacetate. NAD^+-dependent (EC 1.1.1.37) and $NADP^+$-dependent (EC 1.1.1.82) forms of the enzyme exist. MDH is a homodimeric enzyme and it is well known for the many cell compartment-specific isoenzymes that have been characterized from various organisms [63,64]. There is a mitochondrial MDH that functions in the tricarboxylic acid cycle which is usually NAD^+-dependent. There are

two chloroplast enzymes in plants, one $NADP^+$-dependent and one NAD^+-dependent. Most eukaryotes that have been studied also have a cytosolic MDH isoform, and many microbodies contain MDH activity, for example yeast peroxisomes [65], plant peroxisomes [64] and *Trypanosoma* glycosomes [66]. Among other functions, these compartment-specific isoforms help to shuttle reducing equivalents in the form of malate/oxalacetate across membranes and into various cell compartments where they are needed. Whereas the $NADP^+$-dependent MDH from chloroplasts has long been known for its role in a mechanism for exporting reducing equivalents during photosynthesis [67], the NAD^+-dependent enzyme was only discovered recently [68] and is known to be induced during root nodule formation in legumes [69].

The gene tree of MDH (Fig. 3E) is very complex because of various cell compartment-specific isoenzymes and because the gene family is also related to genes of lactate dehydrogenase, which are tetrameric proteins located in the cytosol of eukaryotic cells. There are three main MDH clusters. The first (cluster I, Fig. 3E lower right) contains sequences of some eubacterial MDHs, including *Rhizobium leguminosarum* (α-proteobacteria) and *Synechocystis* (cyanobacteria), and the sequences for lactate dehydrogenases from archaebacteria, eubacteria, animals and plants. This seems to represent the oldest branch of the tree. We found no lactate dehydrogenase sequences for fungi in the databases.

MDH cluster II (Fig. 3E, top) contains eukaryotic NAD^+-dependent MDH of mitochondria, glyoxysomes and plastids of eukaryotes and *Saccharomyces cerevisiae* (the latter also including a cytosolic enzyme). Several homologues from γ-proteobacteria are interdispersed in this group. The three isoenzymes of *S. cerevisiae* and the two isoenzymes of *Trypanosoma brucei* are excellent examples of cell-compartment-specific isoenzymes that have evolved by gene duplication within one major phylum. Also, the close grouping of the mitochondrial, glyoxysomal and plastid MDHs of plants support this idea. The origin of the eukaryotic mitochondrial MDH is not clear, but that the closest homologues of the eukaryotic enzymes are found in proteobacteria, albeit γ-proteobacteria instead of α-proteobacteria, suggests a eubacterial origin. The glyoxysomal enzymes have evolved several times independently by gene duplication of apparently mitochondrial-specific forebears.

The most complex MDH cluster from the phylogenetic standpoint is designated here as cluster III (Fig. 3, left), which contains the cytosolic isoenzymes of animals and plants, the plastid $NADP^+$-specific isoenzymes of plants, and several interleaving eubacterial homologues. In contrast with fungi, the cytosolic MDHs of animals and plants fall into a cluster different from that of the mitochondrial and glyoxysomal enzymes. Also, the $NADP^+$-dependent enzymes of plants seem to descend from cytosolic NAD^+-dependent progenitors and not from the respective gene for the plastid NAD^+-specific isoenzyme, indicating that MDH gene evolution is, to a degree, independent from cofactor specificity. That a group of eubacterial sequences interrupts the sequences of the cytosolic MDHs and the $NADP^+$-dependent MDHs underscores the complexity of MDH gene evolution.

A problem with the MDH tree is sequence divergence between groups. Some MDH sequences show as little as 20% identity and, in some, individual comparisons appear not to be related at all. However, calculating the identity between closest neighboring sequences, all sequence members form a continuum of clearly related sequences, which includes some lactate dehydrogenase isoforms. A similar situation was also observed for the aconitases (see above). Rather than convergent gene evolution, it seems that the sequence divergence from a common ancestor and functional specialization of these enzymes underlies the overall spectrum of MDH (and lactate dehydrogenase) sequence diversity [70].

Isocitrate lyase (ICL)

$$\text{Isocitrate} \rightarrow \text{succinate} + \text{glyoxylate}$$

ICL (EC 4.1.3.1) catalyzes the cleavage of isocitrate into succinate and glyoxylate. The reactions catalyzed by ICL and malate synthase (MS) do not occur in the tricarboxylic acid cycle. They are usually catalyzed by separate enzymes in higher plants, fungi and animals, but they are encoded as a fusion protein with two functional domains in *Caenorhabditis elegans*. Both enzymes are located in microbodies. ICL is typically a homotetramer of ≈64-kDa subunits [71,72]. Using eukaryotic ICL sequences as a query, eubacterial but no archaebacterial sequences were detected, as indicated in the gene tree (Fig. 3F). The eukaryotic ICLs fall into two groups: (a) one that contains the eukaryotic sequences from *Caenorhabditis* and *Chlamydomonas* and is very similar to homologues in γ-proteobacterial genomes and (b) one that encodes the glyoxysomal enzymes of plants and fungi.

Malate synthase (MS)

$$\text{Glyoxylate} + H_2O + \text{acetyl-CoA} \rightarrow \text{malate} + \text{CoASH}$$

MS (EC 4.1.3.2) catalyzes the transfer of the acetyl moeity of acetyl-CoA to glyoxylate to yield L-malate. The glyoxysomal enzyme has been isolated as an octamer of identical ≈60-kDa subunits in maize [73] and other plants [74], as a homotetramer in the fungus *Candida* [75], and as a homodimer in eubacteria [76]. In *C. elegans*, MS is fused to the C-terminus of ICL, yielding a single bifunctional protein [77]. Relatively few sequences of MS are available from prokaryotes. None were found from archaebacteria, and MS activity is extremely rare in archaebacteria, but the activity is present in *Haloferax volcanii* [78].

The tree of MS sequences (Fig. 3G) indicates the distinctness of the plant, fungal and *C. elegans* enzymes, but the available sequence sample is too sparse to generate a solid case for the evolutionary history of the enzyme, other than the finding that the eukaryotic sequences emerge on different branches of a tree of eubacterial gene diversity, with no detectable homologues from archaebacteria.

DISCUSSION

For the 14 different enzymes involved in the higher-plant PDH complex, tricarboxylic acid cycle, and glyoxylate cycle, there are 21 different subunits involved, the sequence similarity patterns of which can be summarized in 19

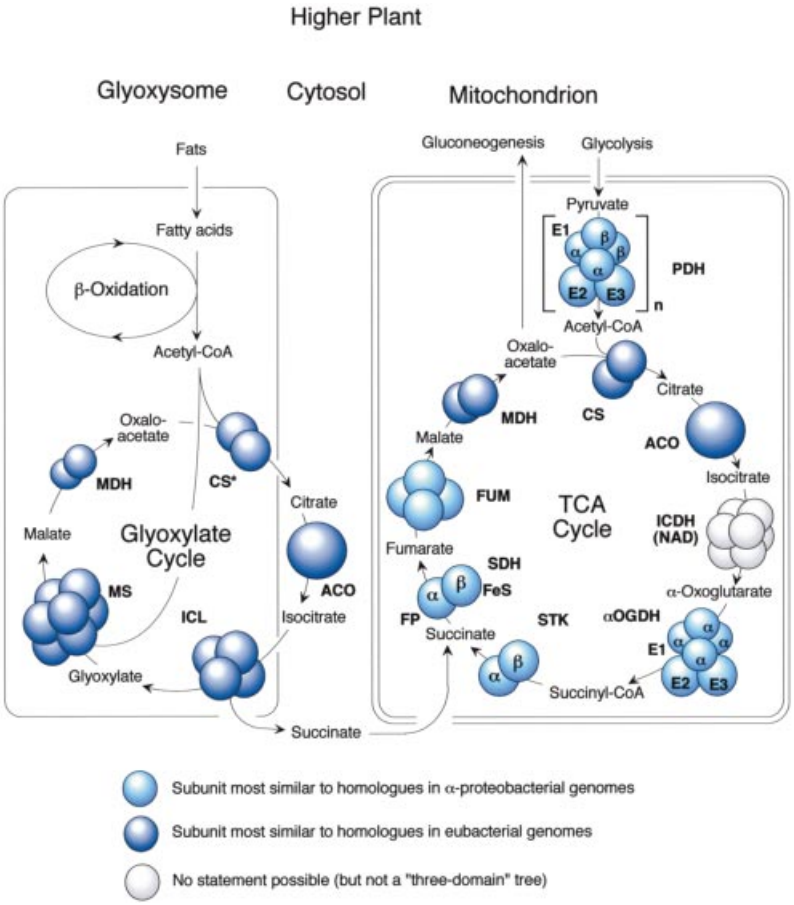

Fig. 4. Schematic summary of similarites of tricarboxylic acid cycle and glyoxylate cycle proteins. Subunit sizes are drawn roughly proportional to molecular mass subcellular compartmentalization. Color coding of subunit sequence simlarities as inferred from the phylogenies indicated. The multimeric nature of the PDH complex is indicated by brackets. FP, flavoprotein; FeS, iron-sulfur subunit. An asterisk next to the glyoxysomal CS indicates that its sequence is highly distinct from that of the mitochondrial enzyme. All of the enzymes in the figure are nuclear encoded in higher plants. Double and single membranes around mitochondria and glyoxysomes, respectively, are schematically indicated. Enzyme and subunit abbreviations are given in the text.

different trees. The trees that we have constructed and shown here do not explain exactly how these enzymes evolved, rather they describe general patterns of sequence similarity. In no case have we analyzed all the sequences available, and in no case have we performed exhaustive applications of the various methodological approaches that molecular phylogenetics has to offer (for example, substitution rate heterogeneity across alignments, significance tests, parametric bootstrapping, topology testing, and the like). Thus, it is possible to perform a more comprehensive analysis of the evolution of these enzymes than we have performed here. However, our aim was not to perform an exhaustive analysis but to obtain an overview of the patterns of similarity for the enzymes of these pathways in plants and the relationships of their differentially compartmentalized isoenzymes. Condensing the information from many individual trees into a single figure that would summarize these patterns of similarity at their most basic level for the plant enzymes, we obtain a simple schematic diagram that will fit on a printed page (Fig. 4). Despite its shortcomings, a few conclusions can be distilled from the present analysis, in particular the relatedness of several of the enzymes investigated to other enzyme families (Table 1).

Higher-plant tricarboxylic acid cycle and glyoxylate cycle: eubacterial enzymes

All of the plant enzymes surveyed here, except cytosolic aconitase (Fig. 2B) and mitochondrial NAD-ICDH (Fig. 2E), are clearly more similar to their eubacterial homologues than they are to their archaebacterial homologues. This is not only true for the plant enzymes, but for almost all of the eukaryotic enzymes studied. Only for about half of the enzymes surveyed were archaebacterial homologues even detected. This is important because many archaebacteria use the reductive tricarboxylic acid cycle, which contains most of the same activities as the tricarboxylic acid cycle, as a major pathway of central carbon metabolism [79]. In no case were the eukaryotic enzymes specifically more related to archaebacterial homologues than to eubacterial homologues.

This is a noteworthy finding because when thinking about the relatedness of eukaryotic archaebacterial and eubacte-

Table 1. Activities related to tricarboxylic acid cycle and glyoxylate cycle enzymes.

Enzyme	Related activity
Aconitase	IRE-BP, IPMI
NAD-ICDH	NADP-ICDH, isopropylmalate dehydrogenase
Fumarase	Aspartate ammonia lyase
NAD-MDH	NADP-MDH, lactate dehydrogenase
PDH, E1	OADH, acetoin dehydrogenase
ОГДН, E2	OADH, PDH
OGDH, E3	Glutathione reductase, mercuric reductase
STK	ATP-citrate lyase[a]
SDH, α subunit	Fumarate reductase, aspartate oxidase
SDH, β subunit	Fumarate reductase

[a] See [113].

rial genes (and proteins), most biologists still tend to envisage, by virtue of a prior knowledge default, the rRNA tree in its most classic form [80] depicting eukaryotes as being more closely related to archaebacteria than to eubacteria [81,82]. In this view, the a priori expectation of the relatedness of a given eukaryotic gene is that it should be more similar to its archaebacterial homologues than to its eubacterial homologues. This pattern was not found for any of the 21 proteins studied here, nor has it been reported for any of 40 other enzymes (and their subunits) (with three exceptions, see below) involved in central carbon metabolism in eukaryotes (glycolysis, gluconeogenesis, the Calvin cycle or the oxidative pentose phosphate cycle) that we have previously studied [3,83–85] (reviewed in [6]). In these analyses, we found no evidence to support the occasionally entertained notion [86,87] that microbodies, to which the glyoxysomes belong and which are surrounded by one membrane rather than two as in the case of chloroplasts and mitochondria, might be descendants of endosymbiotic bacteria.

Eubacterial genes for eukaryotic enzymes of energy metabolism: why?

Not only the cytosolic rRNA, but also most of the proteins involved in the gene-expression machinery in eukaryotes are more similar to their archaebacterial homologues than they are to their eubacterial homologues, including RNA polymerase [88], transcription factors [89], proteins involved with DNA replication [90], ribosomal proteins [91], and the like. In contrast, eukaryotic proteins involved in basic metabolic functions, in particular core carbohydrate metabolism and ATP synthesis, are more similar to eubacterial homologues (cited above). This general pattern is also supported at the level of genome-wide phylogenies for yeast in comparison with eubacerial and archaebacterial reference genomes [11,92]. The observation of genomic chimerism in eukaryotes has been a very surprising one for biologists. There are currently about four biological models that could, in principle, account for this finding.

One model includes the notion that, before the separation of eukaryotes, eubacteria and archaebacteria several billion years ago, there was widespread lateral gene transfer among all organisms, and one combination of such transfers gave rise to the eukaryotic lineage, which some time later obtained mitochondria (the 'genetic annealing' or 'transfer early' model [93]). Another model supposes that eukaryotes are an ancestrally phagocytosing lineage, and that, during the course of eating prokaryotes to survive, they ended up incorporating many genes from their food prokaryotes into their chromosomal genes, and that this process continued when eukaryotes later obtained their mitochondria (the 'you are what you eat' or 'transfer late' model [94]). A third model envisages the origin of eukaryotes as involving the cellular union of an archaebacterium and a eubacterium, in various formulations with the archaebacterium giving rise to the nucleus [92,95–98], yielding a nucleated cell with chimeric chromosomes that later acquired mitochondria (the 'fusion' or 'nucleosymbiosis' model). A fourth model posits that the host of the endosymbiont that became the mitochondrion was not a eukaryote, but rather an autotrophic archaebacterium that acquired roughly a genome's worth of eubacterial genes (and the heterotrophic lifestyle) from the once free-living ancestor of mitochondria; it addresses the common origin of mitochondria and hydrogenosomes (H_2-producing organelles of anaerobic ATP synthesis in eukaryotes that lack typical mitochondria; the 'hydrogen' model [83]).

Taken at face value, the first three models would predict a patchwork of eubacterial and archaebacterial genes in eukaryotic central carbon metabolism, whereas the hydrogen model specifically predicts a eubacterial origin for the enzymes of eukaryotic energy metabolism, of which central carbon metabolism is the backbone. Although the present data do not unambiguously discriminate between these models, it is a noteworthy finding that all of the roughly 40 enzymes involved in central carbon metabolism in eukaryotes that have been studied to date, now including those of the tricarboxylic acid cycle and the glyoxylate pathway in plants, are more similar to eubacterial homologues than they are to archaebacterial homologues. Known exceptions, in which the eukaryotic enzymes are more similar to archaebacterial homologues, are enolase (except *Euglena*) [99], the acetyl-CoA synthase of several mitochondrion-lacking eukaryotes [100,101], and transketolase of animals [8,102], all of which are more similar to their homologues from 'euryarchaeotes' (methanogens and relatives) than they are to homologues from 'crenarchaeotes' (the remaining archaebacteria). Such findings are directly accounted for by the hydrogen model, which posits that the host of mitochondrial symbiosis was a methanogen [83], but not by the other three. As discussed elsewhere [39,103], other traits also link eukaryotes to methanogens, for example histones [104]. Notwithstanding phylogenetic links between eukaryotes and methanogens, the finding that eukaryotes in general possess eubacterial genes for enzymes of carbohydrate and energy metabolism is a striking observation that is usually given insufficient attention in models designed to account for the origins of eukaryotes and their genes.

The eukaryotic tricarboxylic acid cycle: an inhertance from eubacteria, but from which?

The tricarboxylic acid cycle is a specifically mitochondrial pathway in eukaryotes and in some lineages, some of the genes for its enzymes are still encoded in mitochondrial DNA [59]. Furthermore, those tricarboxylic acid cycle genes that are encoded in mitochondria are most closely related to their homologues from α-proteobacteria (Fig. 3B), the lineage of prokaryotes from which mitochondria are thought to descend [105]. However, in most eukaryotes, all of the enzymes of the tricarboxylic acid cycle are encoded in the nucleus. (A very similar situation exists for the Calvin cycle in plastids, where almost of the genes of this typically eubacterial pathway are encoded in the nucleus [3]). This is not completely surprising, because it is known that mitochondrial genomes (and, analogously, plastid genomes) are very highly reduced compared with the genomes of their free-living eubacterial relatives, α-proteobacteria (and cyanobacteria in the case of plastids), and that many genes have been transferred from organelle genomes to the nucleus during the course of evolution [19,20,84].

Thus, one might expect all of the proteins of the tricarboxylic acid cycle to reflect an α-proteobacterial origin, even though they are encoded in the nucleus.

Previous phylogenetic studies focusing on yeast have revealed that several enzymes of the tricarboxylic acid cycle do indeed branch with their α-proteobacterial homologues [106], these cases are relatively easy to explain as above. But if one considers the evolution of all of the enzymes of the pathway (Fig. 4), it is clear that only about half of the enzymes of the tricarboxylic acid cycle, the major pathway of carbon metabolism in mitochondria of oxygen-respiring eukaryotes, can be traced specifically to an α-proteobacterial donor. These enzymes are shaded light blue in Fig. 4. The remaining enzymes are either equivocal (ICDH) or they are most similar to eubacterial, but not specifically α-proteobacterial, homologues (MDH, CS and aconitase in the tricarboxylic acid cycle, and all of the enzymes of the glyoxylate cycle.

There are two general patterns among the 'eubacterial but not specifically α-proteobacterial' proteins observed here and elsewhere [10,39] that deserve explanation. The first (pattern I) are those eukaryotic proteins that branch very close to eubacterial homologues, for example subtree II of MDH (Fig. 3E, top). The second (pattern II) are those eukaryotic proteins that branch within a broader cluster of eubacterial gene diversity, but are somewhat removed from the remaining eubacterial homologues and/or tend to reside on a long branch separating them from eubacterial homologues.

Pattern I. The 'pattern I' protein phylogenies, taken at face value and notwithstanding the vagaries of inferring the ancient past from trees, would tend to indicate that eukaryotes acquired these genes through independent lateral gene transfers from various eubacterial donors (OGDH E1, Fig. 1B; glyoxysomal CS, Fig. 2A; MDH, Fig. 3E; MS, Fig. 3G). However, the eukaryotes sampled here seem, in most cases, all to possess the same acquired gene. Thus, if these kinds of acquisition involved donor(s) that were not the ancestor of mitochondria, then the acquisitions must have occurred very early, and only very early, in eukaryotic evolution (for a discussion see [39,107,108]). However, this is not the only possibility, because it is also possible that the ancestor of mitochondria (or chloroplasts, in the case of plant-specific acquisitions) donated these 'pattern I' genes, even though they do not branch with their homologues found in α-proteobacterial genomes today. The reason for this is simple. Lateral gene transfer is known to occur today among prokaryotes, particularly eubacteria [5,12,13]. Therefore we can assume that it also occurred in the distant past.

Thus, if the free-living descendants of the α-proteobacterium that became mitochondria happened to exchange genes with other free-living eubacteria in the roughly 2 billion years [109,110] that have elapsed since the origin of mitochondria (which is not unlikely), then some (or many) of the genuinely (at that time) α-proteobacterial genes that were in fact donated to eukaryotes by the mitochondrial ancestor would no longer be encoded in α-proteobacterial genomes today [4,20]. As there is very strong evidence to indicate that horizontal transfer occurs today (pathogenicity islands are an excellent example), the principle of uniformitarianism would require us to assume that it existed in the past as well. Thus, if we embrace this assumption (which we should), then the a priori expectation for the phylogeny of eukaryotic genes that come from mitochondria would no longer be that they branch specifically with homologues found on the same contemporary eubacterial chromosomes as 16S rRNA genes, which possess the sequence characteristics necessary to be called α-proteobacterial (the current working definition of 'an α-proteobacterial gene').

Pattern II. The 'pattern II' protein phylogenies depict the eukaryotic proteins as being (a) somehow related to the eubacterial proteins, (b) not specifically related to any eubacterial homologue sampled (this of course can easily change as more sequences are included and as more become available), and (c) on long branches (cytosolic aconitase, Fig. 2B; glyoxysomal ICL, Fig. 3F; mitochondrial CS, Fig. 2A). As the simplest possibilities, this could reflect one of two things. First pattern II might reflect the genuine phylogenetic relationships of the respective proteins and their cellular lineages. However, looking at these trees, this somehow seems unlikely because of the overall failure of pattern II proteins to reflect interpretable evolutionary history. The second possibility, which is well worth considering, is that these patterns reflect sequence similarity that is due to factors other than processes of gene lineage sorting, i.e. that there have been major discontinuities in the evolutionary mode of these proteins during their transition from prokaryotic to eukaryotic chromosomes.

As a specific example of what is meant by the very general foregoing statement, we can consider the fate of a gene that is transferred from the genome of the ancestral mitochondrial symbiont to the genome of its host. Although the term 'endosymbiotic gene transfer' is well established to designate this process, the genes are not really transferred; they are copied, because a functional copy has to remain in the organelle until the nuclear copy obtains the proper expression and routing signals needed to produce a protein that is functional in the organelle, and hence can relieve the organelle copy from selection so that it can become lost to complete the transfer process [84]. However, when genes for symbiont-specific functions become incorporated into the chromosome of their host (by whatever means [19]), they are usually not incorporated in such a way as to immediately acquire the proper expression and targeting signals (current genome data indicates this to be true [19]), and the inevitable process of mutation sets it. At that point, there are basically four things that can happen [3,19,84]. (a) As mutations at otherwise conserved positions are accumulating, the gene acquires (by recombination) the proper expression signal (promoter) but no targeting sequence (transit peptide), and it thus ends up expressing a cytosolic protein (one that thus cannot compete in the organelle with the organelle-encoded protein). (b) As mutations at otherwise conserved positions are accumulating, the gene acquires (by recombination) the proper expression signal (promoter) and targeting sequence (transit peptide) to enable the protein to be imported into the organelle so that it can begin to compete with the organelle-encoded copy. (c) It eventually acquires expression signals and mutates or recombines in a manner so as to acquire a new function. (d) It never acquires the proper expression signals and becomes a pseudogene.

In all of the above cases, by virtue of lacking selection (release from functional constraint), the gene copy in the host's chromosomes will acquire mutations at positions that are otherwise conserved in the copy encoded and functioning in the organelle's (symbiont's) genome. In terms of molec-

ular phylogenetics, this will lead to an accelerated number of substitutions, hence a long branch in the trees, and furthermore it will lead to the mutation of conserved motifs otherwise common to the sequence family to which the gene belonged at the time of endosymbiosis. The dissolution of family-defining motifs through relaxed constraint at the time of relocation to the host's chromosomes more than a billion years ago will have a very concrete impact on the molecular phylogenetic inference of today's sequences; the expectation in such cases would be a long branch separating the eukaryotic sequences from their eubacterial homologues and a placement of that branch markedly removed from (below) its eubacterial progenitor cluster. In essence, this is what is observed in the pattern II phylogenies.

Endosymbiotic gene transfer as it occurred in the beginning

Today, nuclear-encoded mitochondrial proteins are imported into the organelle with the help of the protein translocation apparatus of the inner and outer mitochondrial membrane [111]. However, during the very earliest phases of mitochondrial origins, there must have been a time when the symbiont lived within the cellular confines of its host but had not yet evolved a molecular machinery to import proteins from the host cytosol. During that phase of evolution, symbiont genes that managed their way to the host's chromosomes would have been completely unable to encode products that could compete with the organelle-encoded copy, and thus they only could have been maintained as active genes if their products performed selectable functions in the cytosol. In this way, many pathways once germane to the symbiont could have been transferred to the cytosol of the host [3,83]. For the tricarboxylic acid cycle, a complete transfer of the pathway to the cytosol would not work, because some of its enzymes are intergral components of the inner mitochondrial membrane (for example SDH in complex II), hence inextricably linking the pathway to the organelle (for a more detailed discussion, see [112]). For the enzymes common to the tricarboxylic acid cycle and the glyoxylate cycle, gene-transfer events that did not immediately result in proper targeting of the protein to the mitochondrion may underly the origin of these highly diverse compartment-specific isoforms.

ACKNOWLEDGEMENTS

We thank Marianne Limpert for help in preparing the manuscript, Dr Christine Gietl (Munich) for discussions on plant MDH, and Dr Mikio Nishimura (Okasaki) for discussions on plant peroxisomal enzymes. This work was funded by the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
2. Dandekar, T., Schuster, S., Snel, B., Huynen, M. & Bork, B. (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem. J.* **343**, 115–124.
3. Martin, W. & Schnarrenberger, C. (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32**, 1–18.
4. Martin, W. (1999) Mosaic bacterial chromosomes: a challenge en route to a tree of genomes. *Bioessays* **21**, 99–104.
5. Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128.
6. Henze, K., Schnarrenberger, C. & Martin, W. (2001) Endosymbiotic gene transfer: a special case of horizontal gene transfer germane to endosymbiosis, the origins of organelles and the origins of eukaryotes. In *Horizontal Gene Transfer* (Syvanen, M. & Kado, C., eds), pp. 343–352. Academic Press, London.
7. Lange, B.M., Rujan, T., Martin, W. & Croteau, R. (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl Acad. Sci. USA* **97**, 13172–13177.
8. Mittenhuber, G. (2001) Phylogenetic analysis and comparative genomics of vitamin B6 (pyridoxine) and pyridoxal phosphate biosynthesis pathways. *J. Mol. Microbiol. Biotechnol.* **3**, 1–20.
9. Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M. & Bauer, C.E. (2000) Molecular evidence for the early evolution of photosynthesis. *Science* **289**, 1724–1730.
10. Brown, J.R. & Doolittle, W.F. (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502.
11. Rivera, M.C., Jain, R., Moore, J.E. & Lake, J.A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244.
12. Ochman, H., Lawrence, J.G. & Groisman, E.S. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature (London)* **405**, 299–304.
13. Eisen, J. (2000) Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Curr. Opin. Genet. Dev.* **10**, 606–611.
14. Weeden, N.F. (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J. Mol. Evol.* **17**, 133–139.
15. Martin, W. & Cerff, R. (1986) Prokaryotic features of a nucleus encoded enzyme: cDNA sequences for chloroplast and cytosolyic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). *Eur. J. Biochem.* **159**, 323–331.
16. Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. & Schuster, W. (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett.* **325**, 140–145.
17. Martin, W., Brinkmann, H., Savona, C. & Cerff, R. (1993) Evidence for a chimaeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. *Proc. Natl Acad. Sci. USA* **90**, 8692–8696.
18. Abdalla, F., Salamini, F. & Leister, D. (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant. Sci.* **5**, 141–142.
19. Henze, K. & Martin, W. (2001) How are mitochondrial genes transferred to the nucleus? *Trends Genet.* **17**, 383–387.
20. Rujan, T. & Martin, W. (2001) How many genes in *Arabidopsis* come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends Genet.* **17**, 113–120.
21. Müller, M. (1988) Energy metabolism of protozoa without mitochondria. *Annu. Rev. Microbiol.* **42**, 465–488.
22. Müller, M. (1993) The hydrogenosome. *J. Gen. Microbiol.* **139**, 2879–2889.
23. Tielens, A.G.M. & Van Hellemond, J. (1998) The electron transport chain in anaerobically functioning eukaryotes. *Biochim. Biophys. Acta* **1365**, 71–78.
24. Green, L.S., Li, Y., Emerich, D.W., Bergersen, F.J. & Day, D.A. (2000) Catabolism of α-ketoglutarate by a sucA mutant of *Bradyrhizobium japonicum*: evidence for an alternative tricarboxylic acid cycle. *J. Bacteriol.* **182**, 2844.

25. Krebs, H.A. & Johnson, W.A. (1937) The role of citric acid in intermediate metabolism in animal tissues. *Enzymologia* **4**, 148–156.
26. Mereschkowsky, C. (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Centralbl.* **25**, 593–604. English translation (1999) in *Eur. J. Phycol.* **34**, 287–295.
27. Kornberg, H.L. & Krebs, H.A. (1957) Synthesis of cell constituents from C2-units by a modified tricarboxylic acid cycle. *Nature (London)* **179**, 988–991.
28. Kornberg, H.L. & Beevers, H. (1957) The glyoxylate cycle as a stage in the conversion of fat to carbohydrates in castor beans. *Biochim. Biophys. Acta* **26**, 531–537.
29. Breidenbach, R.W. & Beevers, H. (1967) Association of the glyoxylate cycle enzymes in a novel subcellular particle from castor bean endosperm. *Biochem. Biophys. Res. Commun.* **27**, 462–469.
30. Popov, V.N., Igamberdiev, A.U., Schnarrenberger, C. & Volvenkin, S.V. (1996) Induction of the glyoxylate cycle enzymes in rat liver upon food starvation. *FEBS Lett.* **390**, 258–260.
31. Beevers, H. (1979) Microbodies in higher plants. *Annu. Rev. Plant Physiol.* **30**, 159–193.
32. DeBellis, L., Tsugeki, K., Alpi, A. & Nishimura, M. (1993) Purification and characterization of aconitase isoforms from etiolated pumpkin cotyledons. *Physiol. Plant.* **88**, 485–492.
33. DeBellis, L., Hayashi, M., Biagi, P.P., Hara-Nishimura, I., Alpi, A. & Nishimura, M. (1994) Immunological analysis of aconitase in pumpkin cotyledons: the absence of aconitase in glyoxysomes. *Physiol. Plant.* **90**, 757–762.
34. Genetics Computer Group (2000) *Wisconsin Package, Version 10.2*. Genetics Computer Group (GCG), Madison, WI.
35. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
36. Adachi, J. & Hasegawa, M. (1996) Programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs, no. 28. MOLPHY, Version 2.3*. Institute of Statistical Mathematics, Tokyo.
37. Lindqvist, Y., Schneider, G., Ermler, U. & Sundstrøm, M. (1992) Three-dimensional structure of transketolase, a thiamine diphosphate dependent enzyme. *EMBO J.* **11**, 2373–2379.
38. Muller, Y.A., Lindqvist, Y., Furey, W., Schulz, G.E., Jordan, F. & Schneider, G. (1993) A thiamin diphosphate binding fold revealed by comparison of the crystal structures of transketolase, pyruvate oxidase and pyruvate decarboxylase. *Structure* **1**, 95–103.
39. Rotte, C., Stejskal, F., Zhu, G., Keithly, J.S. & Martin, W. (2001) Pyruvate:$NADP^+$ oxidoreductase from the mitochondrion of *Euglena gracilis* and from the apicomplexan *Cryptosporidium parvum*: a fusion of pyruvate:ferredoxin oxidoreductase and NADPH-cytochrome P450 reductase. *Mol. Biol. Evol.* **18**, 710–720.
40. Millar, A.H., Hill, S.A. & Leaver, C.J. (1999) Plant mitochondrial 2-oxoglutarate dehydrogenase complex: purification and characterization in potato. *Biochem. J.* **343**, 327–334.
41. Reith, M. & Munholland, J. (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Reptr.* **13**, 333–335.
42. Sakamoto, R., Honma, K., Fujii, M. & Honda, K. (1970) Isolation and characterization of citrate synthase isozymes from castor bean endosperm. *Biochemistry* **45**, 405–411.
43. Zehler, H., Thomson, K.-St & Schnarrenberger, C. (1984) Citrate synthases from germinating castor bean seeds. I. Purification and properties. *Physiol. Plant.* **60**, 1–8.
44. Mitchell, C.G., Anderson, S.C. & el-Mansi, E.M. (1995) Purification and characterization of citrate synthase isoenzymes from *Pseudomonas aeruginosa*. *Biochem. J.* **309**, 507–511.
45. Russell, R.J., Ferguson, J.M., Hough, D.W., Danson, M.J. & Taylor, G.L. (1997) The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9Å resolution. *Biochemistry* **36**, 9983–9994.
46. Weitzman, P.D.J. & Dunmore, P. (1969) Citrate synthases: allosteric regulation and molecular size. *Biochim. Biophys. Acta* **171**, 198–200.
47. Flechtner, V.R. & Hanson, R.S. (1970) Regulation of the tricarboxylic acid cycle in bacteria. A comparison of citrate synthases from different bacteria. *Biochim. Biophys. Acta* **222**, 253–264.
48. Schnarrenberger, C., Fitting, K.-H., Tetour, M. & Zehler, H. (1980) Inactivation of the glyoxysomal citrate synthase from the castor bean endosperm by 5,5′-dithiobis (2-nitrobenzoic acid) (DTNB). *Protoplasma* **103**, 299–307.
49. Klausner, R.D., Rouault, T.A. & Harford, J.B. (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell* **72**, 19–28.
50. Gruer, M.J., Artymiuk, P.J. & Guest, J.R. (1997) The aconitase family: three structural variations on a common theme. *Trends Biochem. Sci.* **22**, 3–6.
51. Fouet, A., Jin, S.-F., Raffel, G. & Sonenshein, A.L. (1990) Multiple regulatory sites in the *Bacillus subtilis* citB promoter region. *J. Bacteriol.* **172**, 5408–5415.
52. Prodromou, C., Artymiuk, P.J. & Guest, J.R. (1992) The aconitase of *Escherichia coli*. Nucleotide sequence of the aconitase gene and amino acid sequence similarity with mitochondrial aconitases, the iron-responsive-element-binding protein and isopropylmalate isomerases. *Eur. J. Biochem.* **204**, 599–609.
53. Bradbury, A.J., Gruer, M.J., Rudd, K.E. & Guest, J.R. (1996) The second aconitase (AcnB) of *Escherichia coli*. *Microbiology* **142**, 289–400.
54. Gruer, M.J. & Gues, J.R. (1994) Two genetically-distinct and differentially-regulated aconitases (AcnA and AcnB) in *Escherichia coli*. *Microbiology* **140**, 2531–2541.
55. Lancien, M., Gadal, P. & Hodges, M. (1998) Molecular characterization of higher plant NAD-dependent isocitrate dehydrogenase: evidence for a heteromeric structure by the complementation of yeast mutants. *Plant J.* **16**, 325–333.
56. Panisko, E.A. & McAlister-Henn, L. (2001) Subunit interactions of yeast NAD^+-specific isocitrate dehydrogenase. *J. Biol. Chem.* **276**, 1204–1210.
57. Steen, I.H., Lien, T. & Birkeland, N.K. (1997) Biochemical and phylogenetic characterization of isocitrate dehydrogenase from a hyperthermophilic archaeon, *Archaeoglobus fulgidus*. *Arch. Microbiol.* **168**, 412–420.
58. Lancaster, C.R., Kröger, A., Auer, M. & Michel, H. (1999) Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature (London)* **402**, 377–385.
59. Burger, G., Lang, B.F., Reith, M. & Gray, M.W. (1996) Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proc. Natl Acad. Sci. USA* **93**, 2328–2332.
60. Weaver, T.M., Levitt, D.G., Donnelly, M.I., Stevens, P.P. & Banaszak, L.J. (1995) The multisubunit active site of fumarase C from *Escherichia coli*. *Nat. Struct. Biol.* **2**, 654–662.
61. Weaver, T.M., Lees, M., Zaitsev, V., Zaitseva, I., Duke, E., Lindley, P., McSweeny, S., Svensson, A., Keruchenko, J., Keruchenko, I., Gladilin, K. & Banaszak, L. (1998) Crystal structures of native and recombinant yeast fumarase. *J. Mol. Biol.* **280**, 431–242.
62. Suzuki, T., Yoshida, T. & Tuboi, S. (1992) Evidence that rat liver mitochondrial and cytosolic fumarases are synthesized from one species of mRNA by alternative translational initiation at two in-phase AUG codons. *Eur. J. Biochem.* **207**, 767–772.

63. Goward, C.R. & Nicholls, D.J. (1994) Malate dehydrogenase: a model for structure, evolution, and catalysis. *Protein Sci.* **3**, 1883–1888.
64. Gietl, C. (1992) Malate dehydrogenase isoenzymes: cellular locations and role in the flow of metabolites between the cytoplasm and cell organelles. *Biochim. Biophys. Acta* **1100**, 217–234.
65. Parish, R.W. (1975) The isolation and characterization of peroxisomes (microbodies) from baker's yeast, *Saccharomyces cerevisiae*. *Arch. Microbiol.* **105**, 187–192.
66. Anderson, S.A., Carter, V., Hagen, C.B. & Parsons, M. (1998) Molecular cloning of the glycosomal malate dehydrogenase of *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **96**, 185–189.
67. Ocheretina, O., Haferkamp, I., Tellioglu, H. & Scheibe, R. (2000) Light-modulated NADP-malate dehydrogenases from mossfern and green algae: insights into evolution of the enzyme's regulation. *Gene* **258**, 147–154.
68. Berkemeyer, M., Scheibe, R. & Ocheretina, O. (1998) A novel, non-redox-regulated NAD-dependent malate dehydrogenase from chloroplasts of *Arabidopsis thaliana*. *J. Biochem. (Tokyo)* **273**, 27927–27933.
69. Miller, S.S., Driscoll, B.T., Gregerson, R.G., Gantt, J.S. & Vance, C.P. (1998) Alfalfa malate dehydrogenase (MDH): molecular cloning and characterization of five different forms reveals a unique nodule-enhanced MDH. *Plant J.* **15**, 173–184.
70. Wu, G., Fiser, A., ter Kuile, B., Sali, A. & Müller, M. (1999) Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. *Proc. Natl Acad. Sci. USA* **96**, 6285–6290.
71. Jameel, S. & McFadden, B.A. (1985) *Caenorhabditis elegans*: purification of isocitrate lyase and the isolation and cell-free translation of poly $(A)^+$ RNA. *Exp. Parasitol.* **59**, 337–346.
72. Khan, A.S., Van Driessche, E., Kanarek, L. & Beeckmans, S. (1992) The purification and physicochemical characterization of maize (*Zea mays* L.) isocitrate lyase. *Arch. Biochem. Biophys.* **297**, 9–18.
73. Khan, A.S., Van Driessche, E., Kanarek, L. & Beeckmans, S. (1993) Purification of the glyoxylate cycle enzyme malate synthase from maize (*Zea mays* L.) and characterization of a proteolytic fragment. *Prot. Expr. Purif.* **4**, 519–528.
74. Guex, N., Henry, H., Flach, J., Richter, H. & Widmer, F. (1995) Glyoxysomal malate dehydrogenase and malate synthase from soybean cotyledons (*Glycine max* L.): enzyme association, antibody production and cDNA cloning. *Planta* **197**, 369–375.
75. Okada, H. & Tanaka, A. (1986) Purification of peroxisomal malate synthase from alkane-grown *Candida tropicalis* and some properties of the purified enzyme. *Arch. Microbiol.* **144**, 137–141.
76. Watanabe, S., Takada, Y. & Fukunaga, N. (2001) Purification and characterization of a cold–adapted isocitrate lyase and a malate synthase from *Colwellia maris*, a psychrophilic bacterium. *Biosci. Biotechnol. Biochem.* **65**, 1095–1103.
77. Liu, F., Thatcher, J.D., Barral, J.M. & Epstein, H.F. (1995) Bifunctional glyoxylate cycle protein of *Caenorhabditis elegans*: a developmentally regulated protein of intestine and muscle. *J. Dev. Biol.* **169**, 399–414.
78. Serrano, J.A., Camacho, M. & Bonete, M.J. (1998) Operation of glyoxylate cycle in halophilic archaea: presence of malate synthase and isocitrate lyase in *Haloferax volcanii*. *FEBS Lett.* **434**, 13–16.
79. Schäfer, S., Götz, M., Eisenreich, W., Bacher, A. & Fuchs, G. (1989) ^{13}C-NMR study of autotrophic CO_2 fixation in *Thermoproteus neutrophilus*. *Eur. J. Biochem.* **184**, 151–156.
80. Woese, C., Kandler, O. & Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579.
81. 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.
82. Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., 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.
83. Martin, W. & Müller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature (London)* **392**, 37–41.
84. Martin, W. & Herrmann, R.G. (1998) Gene transfer from organelles to the nucleus: how much, what happens and why? *Plant Physiol.* **118**, 9–17.
85. Krepinsky, K., Plaumann, M., Martin, W. & Schnarrenberger, C. (2001) Purification and cloning of chloroplast 6-phosphogluconate dehydrogenase from spinach: cyanobacterial genes for chloroplast and cytosolic isoenzymes encoded in eukaryotic chromosomes. *Eur. J. Biochem.* **268**, 2678–2686.
86. Cavalier-Smith, T. (1997) Cell and genome coevolution: facultative anaerobiosis, glycosomes and kinetoplastan RNA editing. *Trends Genet.* **13**, 6–9.
87. Latruffe, N. & Vamecq, J. (2000) Evolutionary aspects of peroxisomes as cell organelles, and of genes encoding peroxisomal proteins. *Biol. Cell* **92**, 389–395.
88. 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.
89. Todone, F., Weinzierl, R.O.J., Brick, P. & Onesti, S. (2000) Crystal structure of RPB5, a universal eukaryotic RNA polymerase subunit and transcription factor interaction target. *Proc. Natl Acad. Sci. USA* **97**, 6306–6310.
90. Tye, B.K. (2000) Insights into DNA replication from the third domain of life. *Proc. Natl Acad. Sci. USA* **97**, 2399–2401.
91. Hansmann, S. & Martin, W. (2000) Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. *Int. J. Syst. Evol. Microbiol.* **50**, 1655–1663.
92. Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. (2001) Origin of eukaryotic cell nuclei by symbiosis of archaea in bacteria is revealed by homology hit analysis. *Nat. Cell Biol.* **3**, 210–214.
93. Woese, C. (1998) The universal ancestor. *Proc. Natl Acad. Sci. USA* **95**, 6854–6859.
94. Doolittle, W.F. (1998) You are what you eat: a gene transfer ratchet could account for eubacterial genes in eukaryotic genomes. *Trends Genet.* **14**, 307–311.
95. Zillig, W., Klenk, H.-P., Palm, P., Leffers, H., Pühler, G., Gropp, F. & Garrett, R.A. (1989) Did eukaryotes originate by a fusion event? *Endocyt. C. Res.* **6**, 1–25.
96. Lake, J.A. & Rivera, M.C. (1994) Was the nucleus the first endosymbiont? *Proc. Natl Acad. Sci. USA* **91**, 2880–2881.
97. Gupta, R.S. (1998) Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol. Mol. Biol. Rev.* **62**, 1491.
98. Moreira, D. & Lopez-Garcia, P. (1998) Symbiosis between methanogenic archaea and δ-proteobacteria as the origin of eukaryotes: the syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530.
99. Hannaert, V., Brinkmann, H., Nowitzki, U., Lee, J.A., Albert, M.-A., Sensen, C., Gaasterland, T., Müller, M. & Martin, W. (2000) Enolase from *Trypanosoma brucei*, from the amitochondriate protist *Mastigamoeba balamuthi*, and from the chloroplast and cytosol of *Euglena gracilis*: pieces in the evolutionary puzzle of the eukaryotic glycolytic pathway. *Mol. Biol. Evol.* **17**, 989–1000.

100. Sánchez, L.B., Morrison, H.G., Sogin, M.L. & Müller, M. (1999) Cloning and sequencing of an acetyl-CoA synthetase (ADP-forming) gene from the amitochondriate protist, *Giardia lamblia*. *Gene* **233**, 225–231.
101. Sánchez, L.B., Galperin, M.Y. & Müller, M. (2000) Acetyl-CoA synthetase from the amitochondriate eukaryote *Giardia lamblia* belongs to the newly recognized superfamily of acyl-CoA synthetases (nucleoside diphosphate-forming). *J. Biol. Chem.* **275**, 5794–5803.
102. Martin, W. (1998) Endosymbiosis and the origins of chloroplast-cytosol isoenzymes: a revision of the gene transfer corollary. In *Horizontal Gene Transfer: Implications and Consequences* (Syvanen, M. & Kado, C., eds), pp. 363–379. Chapman & Hall, London.
103. Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. (2001) An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol. Chem.* **382**, 1521–1539.
104. Sandman, K., Periera, S.L. & Reeve, J.N. (1998) Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell. Mol. Life Sci.* **54**, 1350–1364.
105. Gray, M.W., Burger, G. & Lang, B.F. (1999) Mitochondrial evolution. *Science* **283**, 1476–1481.
106. Kurland, C.G. & Andersson, S.G.E. (2000) Origin and evolution of the mitochondrial proteome. *Mircobiol. Mol. Biol. Rev.* **64**, 786–820.
107. Horner, D.S., Hirt, R.P. & Embley, T.M. (1999) A single eubacterial origin of eukaryotic pyruvate: ferredoxin oxidoreductase genes: implications for the evolution of anaerobic eukaryotes. *Mol. Biol. Evol.* **16**, 1280–1291.
108. Embley, T.M. & Martin, W. (1998) A hydrogen-producing mitochondrion. *Nature (London)* **396**, 517–519.
109. Feng, D.-F., Cho, G. & Doolittle, R.F. (1997) Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl Acad. Sci. USA* **94**, 13028–13033.
110. Doolittle, W.F. (1997) Fun with genealogy. *Proc. Natl Acad. Sci. USA* **94**, 12751–12753.
111. Schatz, G. & Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* **271**, 1519–1526.
112. Allen, J.F. (1993) Control of gene expression by redox potential and the requirement for chloroplast and mitochondrial genomes. *J. Theor. Biol.* **165**, 609–631.
113. Ma, J., Jakowitsch, J., Maier, T.L., Bayer, M.G., Müller, N.E., Schenk, H.E.A. & Löffelhardt, W. (2001) ATP-citrate lyase of the glaucocystophyte alga *Cyanophora paradoxa* is a cytosolic enzyme: characterization of the large subunit at the cDNA and genome level. *Mol. Gen. Genomics* **266**, 231–238.