Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8692–8696, September 1993
Evolution

Evidence for a chimeric nature of nuclear genomes: Eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes

(endosymbiosis/lateral gene transfer/paralogous genes/*Anabaena variabilis*/purple bacteria)

WILLIAM MARTIN*†, HENNER BRINKMANN*, CATHERINE SAVONNA‡, AND RÜDIGER CERFF*

*Institut für Genetik, Technische Universität Braunschweig, Postfach 3329, D-38023 Braunschweig, Germany; and ‡Laboratoire de Biologie Moleculaire Végétale, Université Joseph Fourier, B.P. 53X, F-38041 Grenoble Cedex, France

Communicated by Robert K. Selander, July 1, 1993

ABSTRACT **Higher plants possess two distinct, nuclear gene-encoded glyceraldehyde-3-phosphate dehydrogenase (GAPDH) proteins, a Calvin-cycle enzyme active within chloroplasts and a glycolytic enzyme active within the cytosol. The gene for the chloroplast enzyme was previously suggested to be of endosymbiotic origin. Since the ancestors of plastids were related to cyanobacteria, we have studied GAPDH genes in the cyanobacterium *Anabaena variabilis*. Our results confirm that the nuclear gene for higher plant chloroplast GAPDH indeed derives from the genome of a cyanobacterium-like endosymbiont. But two additional GAPDH genes were found in the *Anabaena* genome and, surprisingly, one of these sequences is very similar to nuclear genes encoding the GAPDH enzyme of glycolysis in plants, animals, and fungi. Evidence that the eukaryotic nuclear genes for glycolytic GAPDH, as well as the Calvin-cycle genes, are of eubacterial origin suggests that eukaryotic genomes are more highly chimeric than previously assumed.**

Plastids were once free-living prokaryotes and must have possessed all genes necessary for photoautotrophic growth at the time of endosymbiosis. Yet higher plant chloroplast DNA encodes at least an order of magnitude fewer genes than the genomes of free-living prokaryotes. The majority of higher plant genes involved in photosynthesis, a metabolic pathway surely possessed by the endosymbiont, are currently located in the nucleus. Complete sequences for three plastid genomes have revealed that no known Calvin-cycle enzyme other than the large subunit of ribulose-bisphosphate carboxylase/oxygenase is encoded by chloroplast DNA (for review see ref. 1). Under the gene-transfer corollary to the endosymbiotic theory, plant nuclear genes for those proteins essential to photoautotrophism in cyanobacteria (2) were ultimately derived from the endosymbiont's genome (3). They were transferred to the nucleus and their products were then reimported into the organelle of their origin with the help of a transit peptide (4, 5).

In higher plants, glycolytic and Calvin-cycle pathways possess a number of enzymatic reactions in common which are catalyzed by distinct enzymes unique to each (6). In our previous studies of plant nuclear genes encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) enzymes of the Calvin cycle (GAPA and GAPB; EC 1.2.1.13) and glycolysis (GAPC; EC 1.2.1.12), our working hypothesis predicted that *GapA* and *GapB* genes should reflect the evolution of the endosymbiont, whereas *GapC* genes should reflect that of the eukaryotic host nucleus (refs. 5, 7–10; see also ref. 11). To test this prediction, we have screened a gene library of *Anabaena variabilis* with a degenerate oligonucleotide designed against a highly conserved region of GAPDH amino acid sequences. We have found three GAPDH genes in *Anabaena*.§ Our comparative analyses revealed that this cyanobacterium possesses the expected homologue of Calvin-cycle GAPDH genes of plants and, surprisingly, a GAPDH gene closely related to *GapC* homologues from plants, animals, and fungi. Here we present evidence which strongly suggests that all eukaryotic GAPDH genes studied to date were derived by lateral (endosymbiotic) gene-transfer events early in eukaryotic evolution.

MATERIALS AND METHODS

Molecular Methods. DNA from light-grown (12) axenic cultures of *A. variabilis* (ATCC 29413) was prepared through cesium chloride gradients. *Hin*dIII fragments were cloned into λNM1149 (13) and screened by plaque hybridization at 32°C in hybridization buffer (900 mM NaCl/60 mM sodium phosphate, pH 7.4/6 mM EDTA/0.1% SDS/0.02% polyvinylpyrrolidone/0.02% Ficoll 400) with probe at 10 ng/ml (5×10^6 cpm/ml). The probe was an end-labeled 16-fold degenerate 16-mer oligodeoxynucleotide constructed against the conserved amino acid motif WYDNE(W/Y/F). Three classes of positively hybridizing clones could be differentiated on the basis of inserts which correspond to the three *Hin*dIII bands found in Southern blots of *A. variabilis* genomic DNA when either the oligonucleotide or the respective inserts were used as probes (data not shown). Hybridizing *Hin*dIII restriction fragments for each class were subcloned into pBluescript SK (Stratagene) vectors and sequenced on both strands. Cloning and Southern hybridization experiments were performed for two separate cultures of *A. variabilis* which were independently obtained and independently grown. In both Southern and cloning experiments, the three GAPDH genes were present in equimolar amounts, indicating that all three genes are endogenous to the *A. variabilis* genome.

Phylogenetic Data Analysis. Deduced amino acid sequences were aligned with the Genetics Computer Group package (14). Phylogenetic trees were constructed by the neighbor-joining method (15), which has been shown to be very efficient in recovery of the correct topology under a variety of sequence parameters (16). Distances were measured as nucleotide divergence at nonsynonymous sites (d_N; ref. 17) calculated on the basis of the average of 990 homologous nucleotide positions (average of 760 nonsynonymous sites). Reliability of the distance matrix tree was estimated through

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

†To whom reprint requests should be addressed.

§The sequences from *Anabaena variabilis* reported in this paper have been deposited in the GenBank data base [accession nos. L07497 (*gap1*), L07498 (*gap2*), and L07499 (*gap3*)].

comparison to bootstrap parsimony trees (DNABOOT of PHYLIP; ref. 18) performed on a data set consisting of first and second codon positions for the 27 sequences (734 sites each). This procedure approximates the removal of synonymous sites and improves the parsimony result when divergence between sequences is great (19).

Nomenclature of the GAPDH Genes from Plants and Eubacteria. Gene products (mRNAs, cDNAs, proteins) encoding or corresponding to subunits A and B of chloroplast GAPDH (Calvin cycle) and subunit C of cytosolic GAPDH (glycolysis) are specified as GAPA, GAPB, and GAPC; products from different members of the same gene family are numbered consecutively: GAPA1, GAPA2, GAPB1, GAPB2, etc. The corresponding genes are designated *GapA1*, *GapA2*, *GapB1*, *GapB2*, etc. GAPDH genes from eubacteria are numbered consecutively: *gap1*, *gap2*, *gap3*, etc. This nomenclature corresponds to that previously proposed (5) except that the plant genes are now designated by the same four letters as their corresponding products and that the last letter, designating the subunit type, has been capitalized (e.g., *GapA1* instead of *Gpa1*).

RESULTS AND DISCUSSION

***A. variabilis* Contains Three Divergent GAPDH Genes.** Each of the three *A. variabilis* GAPDH genes occurs as a single copy as determined by Southern blot experiments (data not shown) using specific probes. The gene encoding *Anabaena*'s homologue of the higher plant chloroplast enzyme has been designated *gap2*. The derived protein is more similar to its higher plant Calvin-cycle counterparts (roughly 65% amino acid identity) than it is to any other GAPDH enzyme including the glycolytic enzymes of the eukaryotic cytosol (roughly 48% amino acid identity; Table 1). These data confirm the endosymbiotic origin of nuclear *GapA* and *GapB* genes.

We were surprised to find a second type of GAPDH gene in *Anabaena*, *gap1*, which is more similar to eukaryotic GAPC (Table 1) than to any eubacterial enzyme (except *gap1* of *Escherichia coli*; see below). Like its eukaryotic homologues, *gap1* of *Anabaena* is an NAD^+-specific GAPDH as indicated by the presence of proline-188 in the polypeptide chain, a residue which confers NAD^+ specificity on GAPDH holoenzymes (21, 22). *gap1* of *Anabaena* is located within an operon which also encodes pyruvate kinase and transaldolase, as revealed by data base searches with two open reading frames immediately 5′ to the *gap1* gene. Pyruvate kinase and transaldolase are not involved in the Calvin cycle (23) yet are integral to the oxidative pentose phosphate cycle, the major pathway of carbohydrate breakdown in cyanobacteria (24, 25). *Anabaena*'s *gap1* is therefore probably involved in carbohydrate catabolism. The third GAPDH gene isolated from *A. variabilis* has been termed *gap3*.

Table 1. Average divergence of *gap1* and *gap2* of *A. variabilis* relative to other GAPDH sequences

	Eukaryotic glycolytic *GapC* (n = 10)	Plant Calvin-cycle *GapA/GapB* (n = 6)	Eubacterial GAPDH (n = 6)
Anabaena gap1			
d_N	0.376 (0.028)	0.511 (0.035)	0.540 (0.037)
p	0.401 (0.027)	0.512 (0.027)	0.539 (0.072)
Anabaena gap2			
d_N	0.532 (0.036)	0.322 (0.024)	0.506 (0.035)
p	0.530 (0.027)	0.369 (0.028)	0.510 (0.027)

n, Number of pairwise sequence comparisons from which averages were calculated; d_N, averages of mean numbers of nonsynonymous substitutions per nonsynonymous site (17); p, averages of proportions of amino acid differences per site (20). Numbers in parentheses give the averages of standard errors attached to individual values of d_N and p. Values of average divergence between *Anabaena gap2* and *GapA/GapB* (chloroplast GAPDH) and between *Anabaena gap1* and eukaryotic sequences for glycolytic GAPDH are underlined. Comparisons were between sequences of those species shown in Fig. 1. Eukaryotic sequences refer to plant, animal, and fungal glycolytic GAPDH.

Nuclear GAPDH Genes Are of Eubacterial Origin. The thrust of our interest focused upon *gap1* of *Anabaena*. This gene shows a similar degree of identity to eukaryotic *GapC* genes as *Anabaena gap2* does to higher plant *GapA* and *GapB* (Table 1). Indeed, this finding initially led us to believe that our *A. variabilis* cultures may have been contaminated with some eukaryotic organism which had escaped both microscopic and axenic test detection; for this reason, cloning and Southern hybridizations were performed from two independently grown axenic cultures (see *Materials and Methods*). Equimolarity of the three genes was found in both cultures. Additionally, we have isolated and identified by partial sequencing the specific homologues of *A. variabilis gap1* and *gap2* genes from *Synechocystis* PCC6803 (data not shown), so that the possibility of contamination for the origin of *A. variabilis gap1* clones can be excluded.

Prior to constructing a gene phylogeny, we examined base composition in GAPDH sequences, since fluctuations in G+C content can influence topology (26). The nucleotide composition at third codon positions, which contain about 95% of all synonymous sites (17, 27), varies drastically in GAPDH sequences from 97% G+C in maize *GapA* to 15% G+C in *Clostridium*. First and second codon positions contain about 90% of all nonsynonymous sites, and only about 2% of first and second codon positions are synonymous (17, 27). At first and second codon positions, G+C content is remarkably constant across the 27 GAPDH genes surveyed here, ranging only from 44% in *Clostridium* to 52% in yeast. Since we constructed dendrograms on the basis of either (*i*) divergence at nonsynonymous sites or (*ii*) bootstrap parsimony analysis at first plus second codon positions, G+C bias at third positions in these genes should have virtually no effect upon our topology.

Phylogenetic inference revealed a very complex picture of GAPDH gene evolution (Fig. 1). The most notable result is that the *gap1* gene of *Anabaena* is significantly (99/100 bootstrap parsimony replicates) more closely related to eukaryotic *GapC* than it is to any eubacterial sequence (except *gap1* of *E. coli*; see below). In striking analogy to the plastid origin for nuclear *GapA* genes in higher plants (Table 1), this finding suggests that *GapC* genes of eukaryotes, in addition to *GapA* and *GapB* of plants (5, 11), were laterally transferred from eubacteria to the nucleus early in eukaryotic evolution. This would not only account for the surprising similarity of *A. variabilis gap1* to eukaryotic *GapC* genes but would also easily explain two otherwise puzzling observations previously reported concerning GAPDH gene evolution: (*i*) the unusual relationship between archaebacterial GAPDH genes and those of eubacteria/eukaryotes and (*ii*) the surprisingly high similarity between *E. coli gap1* and eukaryotic enzymes (Fig. 1).

Anomalous Divergence of Archaebacterial GAPDH Genes. *GapC* genes of eukaryotes were previously assumed to have been present in the nucleus prior to any endosymbiotic events, yet GAPDH comparisons between the urkingdoms (34, 35) reveal an anomalous evolutionary behavior for this gene. GAPDH genes from archaebacteria are quite distinct from both their eubacterial and their eukaryotic homologues. These enzymes share only about 15% identical residues with eubacterial and eukaryotic sequences (34), whereas homologues of the latter two share roughly 45–65% identity (Table 1). This pattern of similarity for GAPDH is precisely the converse of that observed for other conservatively evolving

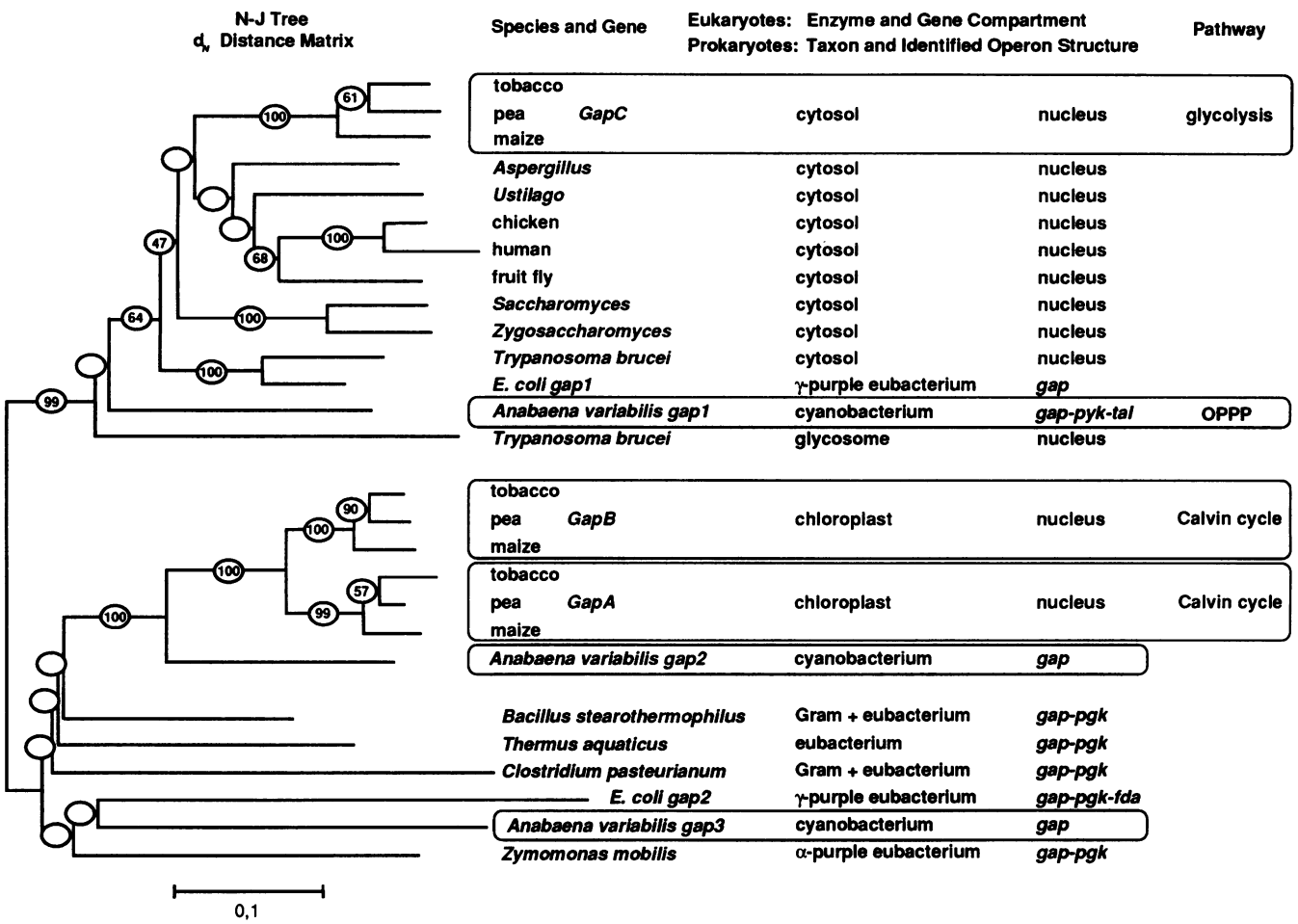

FIG. 1. Unrooted phylogenetic tree constructed by the neighbor-joining (N-J) method (15) from a matrix of values for nucleotide divergence at nonsynonymous sites, d_N (16). Branch lengths are drawn to scale. Higher plant and cyanobacterial sequences are boxed for clarity. Compartmental localization of the enzyme products (eukaryotes) and genes, as well as pathway involvement and operon structure (prokaryotes) as known, is indicated. Numbers in ovals indicate the number of times out of 100 bootstrap parsimony replicates that the corresponding branch was detected (see *Materials and Methods*); ovals without numbers indicate those branches which were detected in <18 replicates. Three branches were detected in >50/100 bootstrap replicates which were not detected in the neighbor-joining tree; these were (*i*) *Zymononas* with *E. coli gap2* (81/100), (*ii*) plant *GapC* sequences with *Saccharomyces* and *Zygosaccharomyces* (64/100), and (*iii*) *Ustilago* with *Aspergillus* (58/100). The close relationship between *Ustilago, Aspergillus*, and metazoan *GapC* sequences seen in the figure and in ref. 28 was detected in <18/100 bootstrap replicates. The *E. coli gap1* sequence is designated *gapA* in ref. 29; the other 16 enterobacterial sequences reported are borne on the *E. coli gap1* branch, diverged from one another within about the last 30 million years (29), and were omitted from analysis here for clarity. The leftmost branch of the unrooted tree was simply "bent" here to permit display of species names in a convenient manner. Abbreviations: OPPP, oxidative pentose phosphate pathway; *gap*, glyceraldehyde-3-phosphate dehydrogenase; *pgk*: phosphoglycerate kinase; *fda*: fructose-1,6-bisphosphate aldolase; *pyk*: pyruvate kinase; *tal*: transaldoase. Sources for sequences other than those given in refs. 28 and 30 are as follows: *Anabaena gap1, gap2*, and *gap3* (this paper); tobacco (11); *GapA* and *GapB* of pea (5); *GapA* and *GapC* of maize (9); *E. coli gap2* (31); *Clostridium* (32); *GapC* of pea (33). Operon structures are not necessarily complete and show only those reading frames adjacent to *gap* genes which have been identified. Unidentified reading frames and regions for which sequences are unavailable are not indicated. Nucleotide and amino acid alignments upon which the figure was based are available upon request.

genes (36–41), as schematically summarized in Fig. 2. A number of explanations for this finding have been put forth (28, 30, 34, 46–49), including the suggestion that archaebacterial GAPDH may not be at all homologous to other GAPDH genes but, rather, functionally converged from a different gene altogether (46, 48). Convergent molecular evolution for GAPDH genes could account for the anomalous relationship of the archaebacterial homologues, but not for the similarity of *gap1* from *Anabaena* and *E. coli* to eukaryotic *GapC*. A eubacterial origin for *GapC* could account for both findings.

The "Eukaryotic-Like" GAPDH Gene of *E. coli*. In 1985, a GAPDH gene was isolated from *E. coli* (50) which, like *Anabaena gap1*, was more similar to GAPDH of eukaryotes than to eubacterial counterparts (*E. coli gap1*; Fig. 1). This eukaryotic-like GAPDH gene of *E. coli* was initially quite puzzling and has been cited (8, 30, 46) as an example for lateral gene transfer from eukaryotes to prokaryotes. A second GAPDH gene was later found in *E. coli* (*gap2* in Fig. 1; ref. 31). Our results place the *E. coli gap1* gene in an entirely new light; they strongly suggest that *E. coli gap1* is endogenous to this eubacterium and was not acquired from eukaryotes. Clearly, the gene duplication which gave rise to the *gap1* and *gap2* genes of *A. variabilis* preceded the separation of the diverse eubacteria surveyed in Fig. 1. Therefore we would expect some eubacteria other than *Anabaena* to have retained more than one GAPDH gene. Indeed, the eukaryotic-like GAPDH genes found in *E. coli* and other enterobacteria (29) appear to represent such endogenous descendants of early eubacterial genomes; their similarity to nuclear *GapC* simply reflects the eubacterial origin of the latter.

Fig. 1 shows a gene phylogeny of duplicated eubacterial sequences, some of which occur today in eukaryotic nuclei, others of which occur in the genomes of free-living eubac-

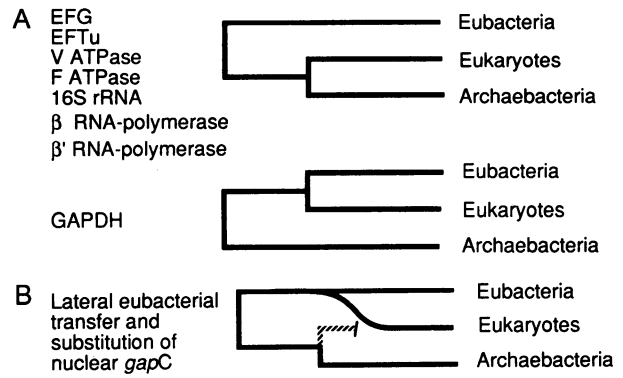

FIG. 2. (*A*) Comparative evolutionary patterns for the evolution of GAPDH and several other genes in eubacteria, eukaryotes, and archaebacteria. The upper cladogram depicts schematically the phyletic relationships observed between seven different genes from the urkingdoms; these include the two translational elongation factors EF-G/2 (36) and EF-Tu/1a (36, 40, 42), vacuolar (V) and F_0-type (F) ATPases (36, 38), 16S ribosomal RNA (39), and two subunits of the DNA-dependent RNA polymerase (36, 41, 43). The lower cladogram depicts schematically the same relationships observed for GAPDH. The anomalous evolutionary pattern for GAPDH was first noted through the cloning (35, 44, 45) and phylogenetic analysis (34) of the archaebacterial genes. (*B*) Schematic depiction of lateral endosymbiotic transfer which could account for various aspects of the data (see text).

teria. We interpret the closely related genes *GapC*, *Anabaena gap1*, and *E. coli gap1* on the one hand and *GapA/GapB* and *Anabaena gap2* on the other as orthologous members of the eubacterial GAPDH gene family. The other, more deeply rooted genes in the lower part of Fig. 1 are probably paralogous descendants of early eubacterial gene duplications and therefore do not reflect true bacterial phylogeny. Orthologous counterparts of *Anabaena gap1* thus either have been lost or have not been detected to date in bacteria other than enterobacteria and *Synechocystis* (see *Materials and Methods*). With the possible exception of the glycosomal GAPDH gene from trypanosomes (see below), nuclear *GapC* genes are more closely related to *E. coli gap1* than to *Anabaena gap1*, suggesting that these *GapC* genes may have derived from purple bacteria (*sensu* Woese; ref. 51) rather than cyanobacteria, possibly from an ancestor of present-day mitochondria (28).

The extraordinarily close relationship between cytosolic *GapC* of *Trypanosoma brucei* and *E. coli gap1* (see Fig. 1) merits special attention, since it suggests that *T. brucei* received this gene by an independent lateral transfer from an *E. coli*-like ancestor, relatively recently in evolution, perhaps in a symbiotic context. If this were the case, kinetoplastids which diverged from the *T. brucei* lineage prior to such a prokaryote-to-eukaryote lateral transfer should possess only the glycosomal *GapC* gene and lack the cytosolic form. Precisely that is observed. In a phylogenetic study of kinetoplastid GAPDH genes, Michels *et al.* (52) have shown (*i*) that *Trypanoplasma borelli* diverged early in evolution from the *T. brucei* lineage, (*ii*) that *Trypanoplasma borelli* possesses only the glycosomal GAPDH enzyme, and (*iii*) that kinetoplastids which are more closely related than *Trypanoplasma borelli* to *T. bruceii* possess both the glycosomal and the *E. coli*-like cytosolic *GapC* forms. Whether glycosomal *GapC* of trypanosmes was derived from a different transfer event than *GapC* of other eukaryotes cannot presently be determined. Glycosomal GAPC shares a common branch with *Anabaena gap1* in 45/100 replicates but also possesses several insertions and deletions that are found in no other published GAPDH sequences (30), which may affect its behavior in phylogenetic inference.

The Biological Context of *GapC* Origin: Endosymbiotic Gene Replacement? The genes for Calvin-cycle GAPDH (*GapA* and *GapB*) were transferred from the plastid genome to the nucleus, where they became established with a transit peptide and were subjected to proper regulation by the nuclear transcription machinery (5, 8–11). The transfer scenario summarized in Fig. 2 for *GapC* of glycolytic function would appear to be more or less identical to that for *GapA/GapB*, except that no transit peptide was necessary to yield a properly compartmentalized, active cytosolic enzyme. Endosymbiotic gene transfer of *GapA* led to loss of the organellar gene (53), but transfer of *GapC* appears to have resulted in loss of both the organellar and the endogenous nuclear genes; *GapC* would thus be the first documented case of eubacterium-to-eukaryote nuclear gene replacement. This working hypothesis for *GapC* evolution provides us with several testable predictions. Among these, we would expect that further analyses of GAPDH genes of purple bacteria should eventually reveal true orthologues of cyanobacterial *gap2*. Furthermore, we might expect that in some very primitive eukaryotes, glycolytic GAPDH genes will be found which are not derived from the *gap1* type but, rather, from some other member of the eubacterial GAPDH gene family. *Trichomonas vaginalis* appears to possess such a GAPDH gene (54).

The evidence for eubacterial origin of eukaryotic GAPC, an essential glycolytic enzyme, suggests that eukaryotic genomes are more highly chimeric than previously assumed. Whereas most organellar proteins are currently encoded in the nucleus, endosymbionts may have donated many genes to the nucleus without organellar reimport of the protein, thereby enriching the genetic and metabolic potential of the host. Those genes which were present in the DNA of both endosymbionts and hosts should be likely candidates for further endosymbiotic gene-replacement events.

We are grateful to W. Ford Doolittle (Halifax) and coworkers for their valuable suggestions on an earlier version of the manuscript, to the Gesellschaft für Biotechnologische Forschung (Braunschweig) for the use of their computer facilities, to Siegfried Scherer (Universtät Konstanz, Konstanz, F.R.G.) for the *Anabaena* inoculum, and to Jean Pelmont (Université Joseph Fourier, Grenoble) and Michael Wettern (Technische Universität, Braunschweig) for help in growing the *Anabaena* cultures. This work was funded by grants from the Centre National de la Recherche Scientifique (URA 1178), the Ministère de l'Education Nationale, and the Deutsche Forschungsgemeinschaft (doctoral grant for H.B.).

1. Shimada, H. & Sugiura, M. (1991) *Nucleic Acids Res.* **19**, 983–995.
2. Tabita, R. F. (1987) in *The Cyanobacteria*, eds. Fay, P. & van Baalen, C. (Elsevier, Amsterdam), pp. 95–118.
3. Weeden, N. F. (1981) *J. Mol. Evol.* **17**, 133–139.
4. Smeekens, S., Weisbeek, P. & Robinson, C. (1990) *Trends Biochem. Sci.* **15**, 73–76.
5. Liaud, M.-F., Zhang, D. X. & Cerff, R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8918–8922.
6. Gottlieb, L. D. (1982) *Science* **216**, 373–380.
7. Cerff, R. (1982) in *Methods in Chloroplast Molecular Biology*, eds. Edelman, M., Hallick, R. B. & Chua, N.-H. (Elsevier/North Holland, Amsterdam), pp. 683–694.
8. Martin, W. & Cerff, R. (1986) *Eur. J. Biochem.* **159**, 323–331.
9. Brinkmann, H., Martinez, P., Martin, W., Quigley, F. & Cerff, R. (1987) *J. Mol. Evol.* **26**, 320–328.
10. Quigley, F., Martin, W. F. & Cerff, R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2672–2676.
11. Shih, M.-C., Lazar, G. & Goodman, H. M. (1986) *Cell* **47**, 73–80.
12. Tandeau de Marsac, N., Borrias, W. E., Kuhlemeier, C. J., Castets, A. M., van Arkel, G. A. & van den Handel, C. (1982) *Gene* **20**, 111–119.
13. Murray, N. (1983) in *Lambda II*, ed. Hendrix, R. W. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 395–432.

14. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
15. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
16. Nei, M. (1991) in *Recent Advances in Phylogenetic Studies of DNA Sequences*, eds. Miyamoto, M. M. & Cracraft, J. L. (Oxford Univ. Press, New York), pp. 90–128.
17. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
18. Felsenstein, J. (1985) *Evolution* **39**, 783–791.
19. Martin, W., Somerville, C. C. & Loiseaux-de Goër, S. (1992) *J. Mol. Evol.* **35**, 385–403.
20. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York), pp. 39–46.
21. Biesecker, G., Harris, J. I., Thierry, J. C., Walker, J. E. & Wonacott, A. J. (1977) *Nature (London)* **266**, 328–333.
22. Corbier, C., Clermont, S., Billard, P., Skarzynski, T., Branlant, C., Wonacott, A. & Branlant, G. (1990) *Biochemistry* **29**, 7101–7106.
23. Miernyk, J. A. (1990) in *Plant Physiology, Biochemistry and Molecular Biology*, eds. Dennis, D. T. & Turpin, D. H. (Longman, Essex, U.K.), pp. 77–100.
24. Pelroy, R. A., Rippka, R. & Stanier, R. Y. (1972) *Arch. Microbiol.* **87**, 303–322.
25. Peschek, G. A. (1987) in *The Cyanobacteria*, eds. Fay, P. & van Baalen, C. (Elsevier, Amsterdam), pp. 119–162.
26. Weisburg, W. G., Giovannoni, S. J. & Woese, C. R. (1989) *Syst. Appl. Microbiol.* **11**, 128–134.
27. Li, W.-H., Wu, C.-I. & Luo, C.-C. (1985) *Mol. Biol. Evol.* **2**, 150–174.
28. Smith, T. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7063–7066.
29. Nelson, K., Whittam, T. S. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6667–6671.
30. Michels, P. A. M., Marchand, M., Kohl, L., Allert, S., Vellieux, F. M. D., Wierenga, R. & Opperdoes, F. (1991) *Eur. J. Biochem.* **198**, 421–428.
31. Alefounder, P. R. & Perham, R. N. (1989) *Mol. Microbiol.* **3**, 723–732.
32. Oster, T. (1989) *Ph.D. thesis* (Université de Nancy I, Nancy, France).
33. Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H. & Cerff, R. (1993) *Mol. Biol. Evol.* **10**, 140–162.
34. Hensel, R., Zwickl, P., Fabry, S., Lang, J. & Palm, P. (1989) *Can. J. Microbiol.* **35**, 81–85.
35. Fabry, S., Lang, J., Niermann, T., Vingron, M. & Hensel, R. (1989) *Eur. J. Biochem.* **179**, 405–413.
36. Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. & Miyata, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9355–9359.
37. Jess, W., Palm, P., Evers, R., Köck, J. & Cornellisson, A. W. C. A. (1990) *Curr. Genet.* **18**, 547–551.
38. 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) *Proc. Natl. Acad. Sci. USA* **86**, 6661–6665.
39. Woese, C., Kandler, O. & Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4576–4579.
40. Auer, J., Spicker, G., Mayerhofer, L., Pühler, G. & Böck, A. (1991) *Syst. Appl. Microbiol.* **14**, 14–22.
41. Iwabe, N., Kuma, K.-I., Kishino, H., Hasegawa, M. & Miyata, T. (1991) *J. Mol. Evol.* **32**, 70–78.
42. Creti, R., Citarella, F., Tiboni, O., Sanangelantoni, A., Palm, P. & Cammarano, P. (1991) *J. Mol. Evol.* **33**, 332–342.
43. Pühler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. & Zillig, W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.
44. Hensel, R., Laumann, S., Lang, J., Heumann, H. & Lottspeich, F. (1987) *Eur. J. Biochem.* **170**, 325–333.
45. Zwickl, P., Fabry, S., Bogedian, C., Haas, A. & Hensel, R. (1990) *J. Bacteriol.* **172**, 4329–4338.
46. Doolittle, R., Feng, D., Anderson, K. & Alberro, M. (1990) *J. Mol. Evol.* **31**, 383–388.
47. Zillig, W., Klenk, H.-P., Palm, P., Leffers, H., Pühler, G., Gropp, F. & Garret, R. A. (1989) *Endocytobiosis Cell Res.* **6**, 1–25.
48. Fothergill-Gilmore, L. A. & Michels, P. A. M. (1992) *Prog. Biophys. Mol. Biol.* **59**, 105–238.
49. Fabry, S. & Hensel, R. (1988) *Gene* **64**, 189–197.
50. Branlant, G. & Branlant, C. (1985) *Eur. J. Biochem.* **150**, 61–66.
51. Woese, C. R. (1987) *Microbiol. Rev.* **51**, 221–271.
52. Michels, P. A. M., Opperdoes, F. R., Hannaert, V., Wiemer, E. A. C., Allert, S. & Chevalier, N. (1992) *Belg. J. Bot.* **125**, 164–173.
53. Baldauf, S. L., Manhart, J. R. & Palmer, J. D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 5317–5321.
54. Markos, A., Miretsky, A. & Müller, M., *J. Mol. Evol.*, in press.