Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 9122–9126, September 1995
Evolution

A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution

(endosymbiotic gene transfer/eukaryotic evolution/introns/glyceraldehyde-3-phosphate dehydrogenase/kinetoplastids)

KATRIN HENZE*, ABDELFATTAH BADR†, MICHAEL WETTERN‡, RÜDIGER CERFF*, AND WILLIAM MARTIN*§

*Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Federal Republic of Germany; ‡Institut für Botanik, Technische Universität Braunschweig, Mendelssohnstrasse 1, D-38023 Braunschweig, Federal Republic of Germany; and †Botany Department, Faculty of Science, Tanta University, Tanta, Egypt

Communicated by Russell F. Doolittle, University of California, San Diego, CA, June 26, 1995 (received for review April 5, 1995)

ABSTRACT **Genes for glycolytic and Calvin-cycle glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of higher eukaryotes derive from ancient gene duplications which occurred in eubacterial genomes; both were transferred to the nucleus during the course of endosymbiosis. We have cloned cDNAs encoding chloroplast and cytosolic GAPDH from the early-branching photosynthetic protist *Euglena gracilis* and have determined the structure of its nuclear gene for cytosolic GAPDH. The gene contains four introns which possess unusual secondary structures, do not obey the GT–AG rule, and are flanked by 2- to 3-bp direct repeats. A gene phylogeny for these sequences in the context of eubacterial homologues indicates that euglenozoa, like higher eukaryotes, have obtained their GAPDH genes from eubacteria via endosymbiotic (organelle-to-nucleus) gene transfer. The data further suggest that the early-branching protists *Giardia lamblia* and *Entamoeba histolytica*—which lack mitochondria—and portions of the trypanosome lineage have acquired GAPDH genes from eubacterial donors which did not ultimately give rise to contemporary membrane-bound organelles. Evidence that "cryptic" (possibly ephemeral) endosymbioses during evolution may have entailed successful gene transfer is preserved in protist nuclear gene sequences.**

Some genes for proteins essential to chloroplasts and mitochondria were encoded in the genomes of free-living antecedents of these organelles and were transferred to the nucleus during evolution. This process, endosymbiotic gene transfer, is a special case of interkingdom horizontal gene transfer and took place in a biologically meaningful context. The contemporary protein products of these genes are synthesized on cytosolic ribosomes and reimported into the organelle of their genetic origin. Although intracellular gene transfer is an ongoing process, the evidence suggests that most genes were transferred during the early phases of endosymbiosis (1–4). It is conceivable that DNA transferred from organelles to the nucleus may have carried not only coding sequences, but also the forerunners of spliceosomal introns now widespread in eukaryotic nuclei (5–7). Little is known about nuclear genes from protists which branched early in eukaryotic evolution.

Euglena gracilis is well suited for the study of endosymbiosis and organellar gene transfer. (*i*) *Euglena*'s plastids are surrounded by three membranes instead of two and possess chlorophylls *a* and *b*, findings which led to the suggestion (8) that *Euglena*'s plastids may have arisen through engulfment of a eukaryotic alga (secondary endosymbiosis), a notion supported by molecular sequence analyses (9, 10). (*ii*) Whereas some algae of secondary symbiotic origin possess a vestigial nucleus (nucleomorph) of the eukaryotic symbiont (11, 12), *Euglena* does not. Evolutionary degeneration of the symbiont may have entailed reduction and fusion of different genetic apparatuses which now comprise the three-membrane-bounded chloroplast genome. (*iii*) Cytological (13, 14) and DNA sequence (15) data indicate that the host cell of *Euglena*'s secondary symbiosis shared a common ancestor with kinetoplastids, a group of nonphotosynthetic protists encompassing trypanosomes and their relatives (16). Thus, in *Euglena*, nuclear genes of endosymbiotic origin may have been transferred twice: once to the algal nucleus and once more to the kinetoplastid nucleus.

Biochemical studies had indicated that *Euglena*, like other photosynthetic eukaryotes, possesses two distinct glyceraldehyde-3-phosphate dehydrogenase (GAPDH) enzymes, the Calvin-cycle GAPDH of chloroplasts (GapA, EC 1.2.1.13), and the glycolytic GAPDH of cytosol (GapC, EC 1.2.1.12) (17); neither is encoded in *Euglena*'s chloroplast DNA (cpDNA) (18). Here we show that the nuclear *GapC* and *GapA* genes of *Euglena*, as in other eukaryotes, are descendants of ancient gene duplications which occurred in eubacterial genomes (4, 19). We argue that during evolution, the kinetoplastid lineage as well as the amitochondriate protists *Giardia* and *Entamoeba* have independently obtained GAPDH genes from eubacteria through cryptic endosymbiosis—i.e., in an endosymbiotic context resulting in abortive organelle genesis. We also report the structure of *Euglena*'s expressed nuclear *GapC* gene.¶

MATERIALS AND METHODS

Isolation of Recombinant Clones. *Euglena* cultures (SAG 1224-5/25) were grown as described (20) under a 14-hr light/10-hr dark regime aerated with 1.5% CO_2. Nucleic acid isolation and cDNA cloning were performed as described (21). The cDNA library was screened by plaque hybridization (4) with an end-labeled oligonucleotide, 5′-TGGTAYGAYAAN-GART-3′. About 10^6 recombinants of an *Mbo* I genomic library in λEMBL4 (22) were screened by plaque hybridization with the random-labeled *Not* I insert of pEGC20 (encoding GapC; see *Results*). Five clones containing a 4.2-kb *Hin*dIII fragment identified by Southern hybridization of genomic DNA (data not shown) were purified. The hybridizing *Hin*dIII fragment of λEGCg110 was subcloned into pBluescript vectors (Stratagene) and sequenced. Other molecular methods were as described (22).

Phylogenetic Analysis. The amino acid alignment (available upon request) from which nucleotide sequences (368 codons per sequence) were aligned was produced with the LINEUP program of the WISGEN package (23). A matrix of divergence

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.

Abbreviations: GAPDH, glyceraldehyde-3-phosphate dehydrogenase; cpDNA, chloroplast DNA.

§To whom reprint requests should be addressed.

¶The sequences from *Euglena gracilis* reported in this paper have been deposited in the GenBank data base [accession nos. L21903 (*GapC* cDNA), L21904 (*GapA* cDNA), and L39772 (*GapC* gene)].

at nonsynonymous sites (24) was used to construct a neighbor-joining tree (25). The topology was tested by bootstrap neighbor-joining analysis using the Dayhoff matrix (PHYLIP 3.5) between protein sequences.

RESULTS AND DISCUSSION

GAPDH: A Eubacterial Gene Family. Previous work indicated that genes for eukaryotic GAPDH enzymes are descendants of an ancient gene family which existed in the common ancestor of extant eubacteria (4, 19, 26). This view is supported by the GAPDH gene phylogeny in Fig. 1. To date, GAPDH gene families of at least three members have been characterized in two eubacteria, *Anabaena variabilis* (4) and *E. coli* (28, 29) (the third *E. coli* sequence is not complete; GenBank accession no. L09067). For *gap1, gap2,* and *gap3* of *E. coli*, orthologous genes have been characterized in other free-living eubacteria, as shown in the schematic topologies (subtrees) tI, tIII, and tIV in Fig. 1. For *Anabaena gap2*, orthologues have been characterized only in cyanobacteria (unpublished data) and in the nuclei of photosynthetic eukaryotes, to which they were transferred from the antecedents of modern chloroplasts (tII). Broad-scale surveys of GAPDH gene diversity in eubacteria have not been reported, but sufficient isolated eubacterial *gap* gene sequences exist in the data base to reveal that the common eubacterial ancestor had four or more *gap* genes. Members of the ancestral gene family may have been lost independently in different eubacterial lineages or have not been characterized to date, as indicated by open branches in tI–tIV. Comparisons of sequences across different subtrees reveal an average of about 50% amino acid identity. Within subtrees, average amino acid identity is about 60% or greater, except within subtree tIII (45–50%), which contains the rapidly evolving *Anabaena gap3* and *E. coli gap2* sequences and receives only very weak support from bootstrap analysis, making the identification of this subtree tentative.

Subtree tI contains by far the greatest number of sequenced eubacterial genes. *E. coli gap1* and the *Serratia* sequence were chosen to represent the roughly 60 sequences reported from enterics (ref. 30 and GenBank release 84). The partial sequence of a *gap1* gene from the β-purple bacterium *Pseudomonas solanacearum* was recently reported (GenBank accession no. L19269); phylogenetic analysis of the C-terminal 75 aa which have been sequenced for that gene suggests that it is orthologous to *E. coli* and *Anabaena gap1*, branching robustly within tI (data not shown), supporting this interpretation of

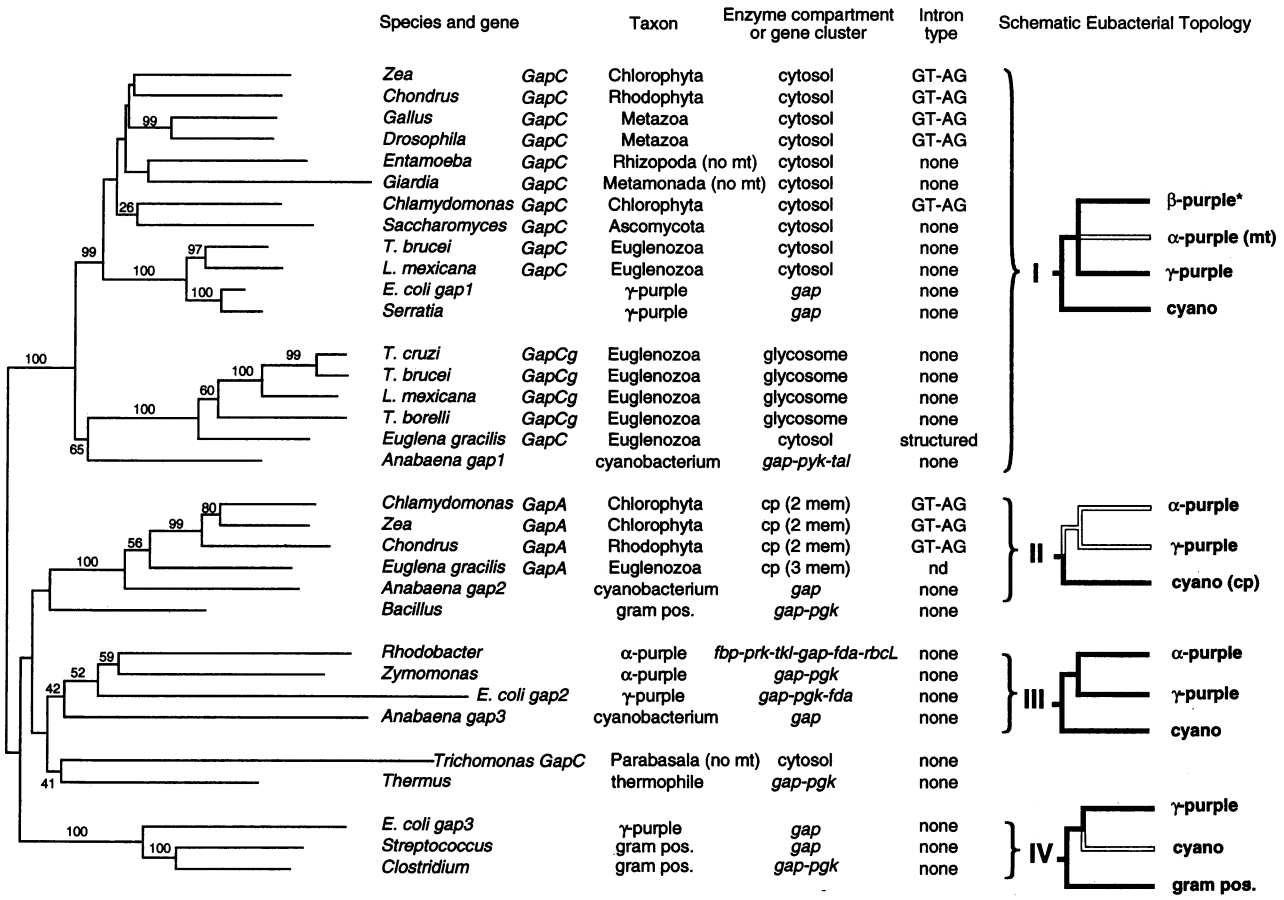

FIG. 1. Phylogenetic distribution of GAPDH gene types, enzyme compartmentalization, operon structures, and introns. The gene phylogeny was constructed as described in *Material and Methods.* Bootstrap numbers indicate the number of times that the branch was detected out of 100 neighbor-joining replicates using the Dayhoff matrix distances (PHYLIP 3.5) between aligned protein sequences. Branches lacking numbers were found in <40 replicates. Eukaryotic taxon designations refer to phyla recognized by Corliss (27); no mt, amitochondriate protists. Gene designations in eubacterial operons are *gap* (GAPDH), *pyk* (pyruvate kinase), *tal* (transaldolase), *fbp* (fructose-1,6-bisphosphatase), *prk* (phosphoribulokinase), *tkl* (transketolase), *fda* (fructose-1,6-bisphosphate aldolase), and *pgk* (phosphoglycerate kinase). Intron type refers to introns characterized in GAPDH genes; GT–AG, spliceosomal intron(s) present. Schematic eubacterial topologies tI–tIV are intended to support and clarify the interpretation of eukaryotic GAPDH genes as descendants of a eubacterial gene family; branches bearing characterized genes are shown solid; those carrying genes which have yet to be characterized (or have been lost during evolution) for the taxa indicated are not filled in. *, The β-purple branch in tI refers to the partial sequence reported for *Pseudomonas solanacearum* (not shown in tree; see text). *T., Trypanosoma* (*cruzi* or *brucei*) or *Trypanoplasma* (*borelli*); *L., Leishmania*; *E., Escherichia*; GapCg, glycosomal GapC; purple, purple (proteo-) bacteria; gram pos., Gram-positive eubacteria; mem, membranes; nd, not determined; mt, mitochondria; cp, chloroplasts. Sequences were retrieved from GenBank (release 84).

the subtree. The data do not indicate whether the *Thermus* and *Trichomonas* sequences in Fig. 1 are orthologs of tI–tIV or whether they represent branches of further duplicate *gap* gene subtrees.

From the *Euglena gracilis* cDNA library of 60,000 recombinants, we identified 41 positive clones which hybridized to the GAPDH-specific oligonucleotide probe. Nineteen clones containing >1-kb inserts were terminally sequenced, revealing that 10 were identical in 5′ and/or 3′ untranslated regions to pEGC20 (accession no. L21903), which encodes *Euglena*'s cytosolic GapC enzyme (see below). The remaining 9 positive clones for which terminal sequences were determined were identical to pEGA23, which encodes the chloroplast enzyme, GapA (accession no. L21904). pEGC20 and pEGA23 thus appear to represent *Euglena*'s major transcripts for GapC and GapA, respectively.

***GapA* of *Euglena* Does Not Encode a Polyprotein.** Two previously studied nuclear gene-encoded chloroplast enzymes of *Euglena*, ribulose-bisphosphate carboxylase small subunit (rbcS) and light-harvesting chloroplast protein I (LHCI), are translated as multimeric precursor polyproteins whose concatenate subunits are proteolytically processed to yield mature enzyme monomers subsequent to transport across the three chloroplast membranes (31, 32). *GapA* of *Euglena* encodes a chloroplast protein but is not organized in the same unusual manner. pEGA23 encodes a single 333-aa mature subunit preceded by a 147-aa N-terminal extension. Northern blots of *Euglena* poly(A)$^+$ RNA probed with pEGA23 clearly revealed a single band at 1.8 kb (data not shown), excluding the possibility that multimeric cDNAs might have recombined out during cloning in *E. coli*.

The N-terminal extension of GapA is considerably longer than typical transit peptides of higher plants (33). It shows no similarities to other chloroplast GAPDH transit peptides but is similar in length to the ≈140-aa transit peptides of three previously characterized nuclear gene-encoded chloroplast proteins of *Euglena* (34) (Fig. 2). These contain two highly hydrophobic domains (underlined in Fig. 2) separated by an ≈60-aa hydrophilic stretch, as well as topogenic signals for targeting to the endoplasmic reticulum during precursor import across *Euglena*'s three outer chloroplast membranes (32, 34).

Euglena rbcS, rbcL, psbA, and *tufA* sequences are more similar to chlorophyte than to rhodophyte or eubacterial homologues, suggesting a secondary endosymbiotic origin from chlorophyte antecedents for its plastids (9, 10). We expected *Euglena*'s *GapA* sequence to reveal the same pattern of homology. The gene phylogeny shows that although *GapA* of *Euglena* is clearly a homologue of cyanobacterial *gap2*, it assumes an outgroup position relative to both chlorophyte and rhodophyte *GapA* homologues (Fig. 1). This result is incongruent with phylogenies inferred from genes of cpDNA, which are more reliable markers for plastid origins than nuclear genes are (9, 10). Furthermore, bootstrap support for the outgroup position of *Anabaena gap2* in this subtree is low, and sequences of *gap2* homologues from purple bacteria, which in our prediction exist but have yet to be found, are lacking for internal reference. Thus, the position of *Euglena*'s *GapA* sequence in Fig. 1 indicates a eubacterial origin for the gene but neither supports the view of a secondary endosymbiotic origin for *Euglena*'s plastids nor clearly indicates from what type of eubacterium it was donated.

Cryptic Endosymbiosis in Kinetoplastids. In the gene phylogeny in Fig. 1, cytosolic GapC of *Euglena* is orthologous to glycosomal GAPDH from kinetoplastids and has thus been recompartmentalized during euglenozoan evolution. Yet some kinetoplastids—e.g., *T. brucei* and *L. mexicana*—also possess a second, distinct cytosolic GAPDH enzyme in addition to the glycosomal form (35, 36). The overall topology of Fig. 1 makes it exceedingly unlikely that enterobacterial *gap1* and kinetoplastid cytosolic *GapC* genes reflect prokaryote–eukaryote divergence. One or the other of these genes appears to reside in the wrong genome, suggesting that gene transfer has occurred between the antecedents of these organisms (37), but in which direction? We argue that *E. coli gap1* is native to the eubacterial genome, suggesting that kinetoplastids have acquired the gene for their cytosolic GAPDH from bacteria via an additional gene transfer event subsequent to the acquisition of the glycosomal homologue. Since contemporary kinetoplastids and euglenids are known to carry endosymbiotic bacteria (13, 16), it seems quite plausible that the common ancestor of *T. brucei* and *L. mexicana* acquired its gene for cytosolic GAPDH from a γ-purple bacterial donor in a symbiotic context which did not result in a membrane-bound organelle. More distantly related euglenozoans, such as *Trypanoplasma borelli* and *Euglena*, separated from the *Trypanosoma/Leishmania* lineage prior to the symbiotic event (Fig. 1 and ref. 38) and thus apparently never obtained the gene. The gene for cytosolic GAPDH in kinetoplastids provides evidence for an evolutionarily recent symbiotic event which entailed gene transfer but did not proceed to organellogenic symbiosis.

Amitochondriate Protists with Eubacterial GAPDH Genes. In light of this evidence, the positions of GAPDH genes from the amitochondriate protists *Giardia lamblia* and *Entamoeba histolytica* deserve particular attention. Their GAPDH genes assume an (albeit poorly supported) crown position within the eukaryotic phylogeny, which contrasts with results obtained with other phylogenetic markers (15, 39). Have *Giardia* and *Entamoeba* obtained eubacterial *gap* genes even though they possess no obvious mitochondrion which would betray a symbiotic event? Recently, *Entamoeba* was shown to possess nuclear genes for two typically mitochondrial proteins: chaperonin cpn60 and pyridine nucleotide transhydrogenase (40). The strongly mitochondrion-like phylogeny of those genes supports the view that *Entamoeba* is not primitively amitochondriate. Although its position in Fig. 1 is not resolved, *Entamoeba*'s GAPDH gene could derive from the same endosymbiotic donor as its cnp60 and pyridine nucleotide transhydrogenase sequences. Similar observations were made for *Giardia*'s bacterial-like 78-kDa glucose-regulated protein (GRP78)/70-kDa heat shock protein (HSP70) homologue (41). Those who would argue that these protists are not primitively amitochondriate might contend that their possession of eubacterial GAPDH genes circumstantially supports that view.

The position in the gene phylogeny of GAPDH from a different amitochondriate protist, *Trichomonas vaginalis*, is highly intriguing. This organism lacks mitochondria yet possesses a DNA-free, double-membrane-bound, energy-producing organelle: the hydrogenosome (42). *Trichomonas*

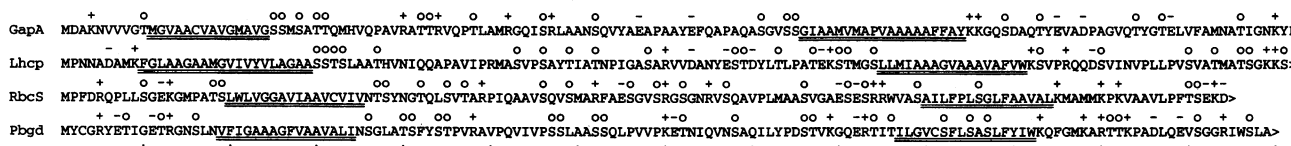

FIG. 2. Comparison of transit-peptide regions for nuclear gene-encoded chloroplast proteins of *E. gracilis*. Sequences were taken from this paper and from references in ref. 34. Positively charged, negatively charged, and hydroxylated amino acids are indicated with +, −, and ○, respectively. Hydrophobic domains are doubly underlined.

GAPDH has eubacterial features at both the biochemical and the sequence level (43), yet like *Thermus*, its gene cannot be assigned to any of the schematic topologies tI–tIV in Fig. 1. Though it appears that *Trichomonas* has also obtained its GAPDH gene from eubacteria (43), the data do not indicate clearly from what type of bacterium it might have obtained the gene, whether the gene donor was implicated in hydrogenosome evolution, or whether the gene represents a descendant of an otherwise uncharacterized member of the ancestral *gap* gene family. However, its unique position in the phylogeny would not exclude the possibility that *Trichomonas GapC* represents the only truly endogenous GAPDH of eukaryotes described to date.

Clearly, the schematic topologies in Fig. 1 represent a working hypothesis. Barring extensive differential loss, it would predict *gap* genes to be found which ultimately trace the overall course of eubacterial evolution in at least four independent subtrees. But alternative interpretations can also account for various aspects of the data. It is possible that widespread (conjugational or other) transfer of GAPDH genes between eubacteria has occurred in evolution, in which case tI–tIV may be erroneous, but an explanation of why such promiscuity is not observed for other genes is wanting. Also, subtree tI might consist of two separate eubacterial phylogenies in the upper and lower portions, as suggested by the unstable position of *Anabaena gap1*. As a further consideration, eukaryotic origins have not been fully resolved (44) and conceivably could have entailed fusion of archaebacterial and eubacterial genomes, the latter (or both) of which may have possessed several *gap* genes. In that case, any sequence of prefusion duplication and subsequent cryptic endosymbiotic transfer events involving GAPDH genes found in contemporary eukaryotic chromosomes would become extremely difficult to reliably reconstruct.

***Euglena GapC* Contains Four Unusual Introns.** The sequence of the 4.2-kb genomic *Hin*dIII fragment revealed that *GapC* contains no nucleotide substitutions relative to pEGC20. Three introns of 46, 251, and 378 bp occur in the coding region; intron 4, in the 3′ untranslated region, is 441 bp long. All four may assume an unusual secondary structure (Fig. 3). Less elaborate secondary structures were also noted by Tessier *et al.* (45) for introns in the *rbcS* gene from *Euglena*. The acceptor and donor ends of all four introns can assume a stem–loop structure, and introns 2–4 possess further internal structures very atypical of pre-mRNA introns in higher eukaryotes. We found no similarity between these secondary structures and those characteristic of group I (46) or group II (47) introns, nor could we identify a "bulging A" motif as in group III introns (48).

All four *GapC* introns are flanked by 2- to 3-bp repeats, so that their position could not be determined to the base. With the notable exception of intron 1, none of the introns can be placed within their short duplication so as to conform to the GT–AG rule (Fig. 4). The two previously characterized nuclear genes from *Euglena*, *rbcS* (45) and *lhcp2* (49), also contain introns which lack GT–AG consensus borders. The *Euglena GapC* introns might possess G at position +5 (Fig. 4), a residue also conserved in group III introns (48), contingent upon precise intron localization.

On the basis of Fig. 4, none of *Euglena*'s *GapC* introns are precisely conserved in position with introns from any known GAPDH gene. With a previous numbering scheme for GAPDH intron positions (19), intron 1 is 9 bp 3′ of position 15, intron 2 is 3 bp 3′ of position 32, and intron 3 is 4 bp 3′ of position 38. Intron 2 could be considered identical to position 32 if shifted to the extreme left within the 3-bp border duplication. These three positions are found in *GapC* but not in *GapA* or *GapB* and thus favor neither the view that conservation of intron positions in GAPDH genes may be due

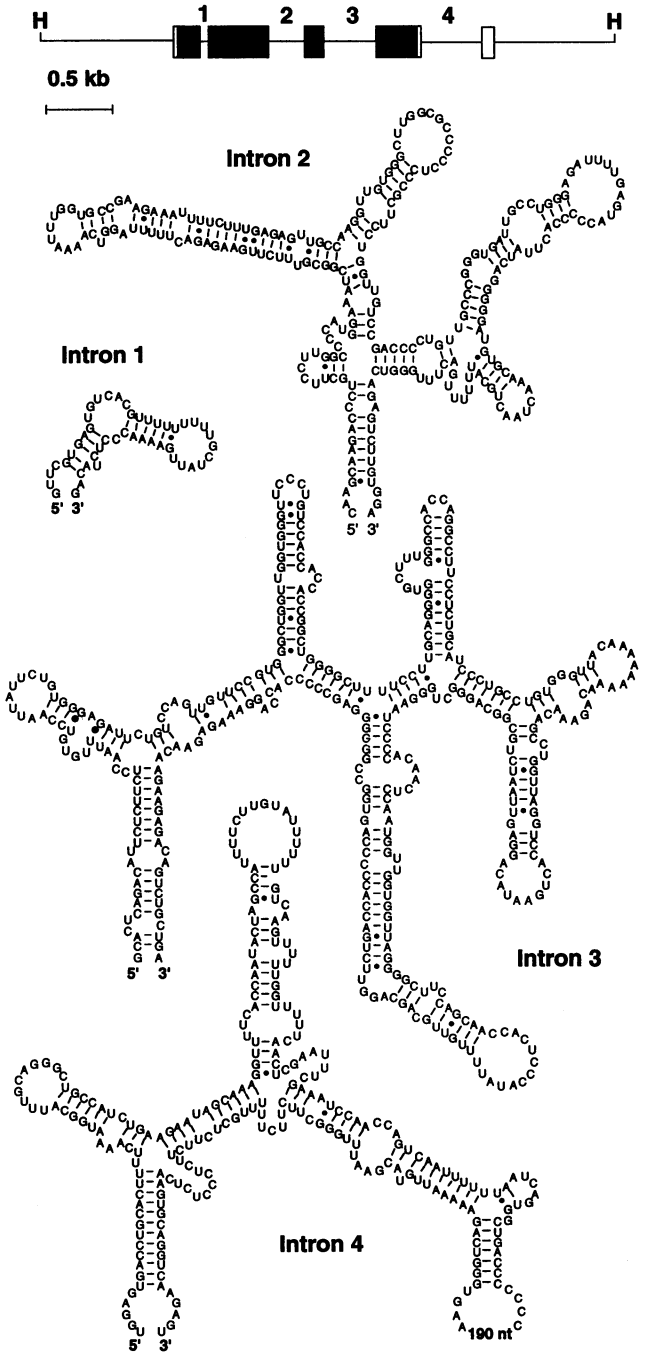

FIG. 3. Structure of the *Euglena GapC* gene on the sequenced 4.2-kb *Hin*dIII fragment and predicted intron RNA secondary structures. Exons are shown as boxes, coding regions are shaded black. The secondary structures of individual introns were generated with the MFOLD and SQUIGGLES programs of the Genetics Computer Group (Madison, WI) package. The 190-nt region not shown in intron 4 consists of a single 95-bp imperfect hairpin.

to common ancestry from eubacterial genomes (19) nor the view that it is due to independent insertion (50).

The lack of consensus borders and the presence of unusual secondary structures within *Euglena* nuclear introns raise the question of their origins, which in turn relates to the origin of spliceosomal introns in general. If one accepts the commonalities of group II and spliceosomal introns as evidence of common ancestry, three scenarios ensue for the origin of spliceosomal introns (7): (*i*) group II introns invaded the nucleus via endosymbiotic gene transfer from organelles and subsequently gave rise to spliceosomal introns, (*ii*) group II

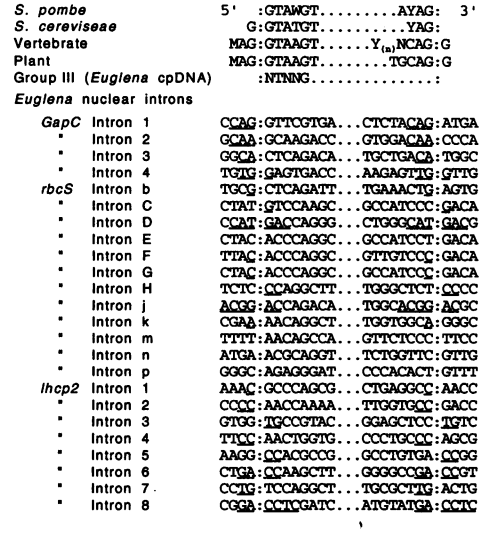

FIG. 4. Intron border consensus sequences from eukaryotic nuclei and *Euglena* chloroplast group III introns in comparison to *Euglena* nuclear intron borders. *Euglena* nuclear introns flanked by short direct repeats (underlined) cannot be unambiguously placed between exons; in such cases, the placement was arbitrarily chosen here to maximize the occurrence of G (or R) at position +5 of the intron, which is the only recognizable feature of these introns shared with other types listed. *Euglena* nuclear introns other than *GapC* were taken from the published sequences (45, 49). References to intron border consensus sequences given are found in refs. 45–49. *S. pombe, Schizosaccharomyces pombe; S. cerevisiae, Saccharomyces cerevisiae.*

introns were present in the common ancestor of eubacteria and eukaryotes but gave rise to spliceosomal introns only in the latter, or (*iii*) group II and spliceosomal introns coexisted in the common ancestor of eubacteria and eukaryotes, but spliceosomal introns became abundant only in the latter. Our findings suggest that the evolution of nuclear introns proceeded independently in *Euglena* and higher eukaryotes, but do not distinguish among Roger *et al.*'s scenarios (7). At the structural level, the *Euglena* introns are sufficiently distinct to suggest that they may represent a novel intron class, and their splicing mechanism deserves further attention.

Endosymbiosis and gene transfer have contributed to the nuclear complement of eukaryotic genes. Some symbiotic events have left very clear evidence of their occurrence in the form of organellar genomes in plastids and mitochondria. Yet other, "cryptic" symbioses appear to have entailed gene transfer but resulted either in organelles which no longer possess vestigial genomes or in no organelle at all. Nuclear genes in protists provide evidence for the existence of cryptic symbiotic events.

We thank M. Müller, H. Brinkmann, P. Michels, and C. Schnarrenberger for critical comments on the manuscript and C. Köhler for excellent technical assistance. A.B. was the recipient of an Alexander von Humboldt fellowship. This work was supported by Grants Ma1426/1-2 and Ma1426/3-1 from the Deutsche Forschungsgemeinschaft to W.M.

1. Weeden, N. F. (1981) *J. Mol. Evol.* **17,** 133–139.
2. Baldauf, S. & Palmer, J. (1990) *Nature (London)* **344,** 262–265.
3. Gannt, J. S., Baldauf, S., Calie, P. J., Weeden, N. F. & Palmer, J. D. (1991) *EMBO J.* **10,** 3073–3078.
4. Martin, W., Brinkmann, H., Savona, C. & Cerff, R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8692–8696.
5. Roger, A. J. & Doolittle, W. F. (1993) *Nature (London)* **364,** 289–290.
6. Ferat, J.-L. & Michel, F. (1993) *Nature (London)* **364,** 358–361.
7. Roger, A. J., Keeling, P. J. & Doolittle, W. F. (1994) *Molecular Evolution of Physiological Process* (Rockefeller Univ. Press, New York).
8. Gibbs, S. P. (1978) *Can. J. Bot.* **56,** 2883–2889.
9. Martin, W., Somerville, C. C. & Loiseaux-de Goër, S. (1992) *J. Mol. Evol.* **35,** 385–403.
10. Morden, C. W., Delwiche, C. F., Kuhsel, M. & Palmer, J. D. (1992) *BioSystems* **28,** 75–90.
11. Maier, U.-G. (1992) *BioSystems* **28,** 69–74.
12. McFadden, G., Gilson, P. R., Hoffmann, C. J. B., Adcock, G. J. & Maier, U.-G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 3690–3694.
13. Walne, P. L. & Kivic, P. A. (1989) in *Handbook of Protoctista*, eds. Margulis, L., Corliss, J. O., Melkonian, M. & Chapman, D. J. (Jones & Bartlett, Boston), pp. 270–287.
14. Surek, B. & Melkonian, M. (1986) *Protoplasma* **133,** 39–49.
15. Sogin, M., Gunderson, J., Elwood, H., Alonso, R. M. & Peattie, D. (1989) *Science* **243,** 75–77.
16. Vickerman, K. (1990) in *Handbook of Protoctista*, eds. Margulis, L., Corliss, J. O., Melkonian, M. & Chapman, D. J. (Jones & Bartlett, Boston), pp. 215–238.
17. Theiss-Seuberling, H.-B. (1984) *Plant Cell Physiol.* **25,** 601–609.
18. Hallick, R. B., Hong, L., Drager, R. G., Favreau, M., Monfort, A., Orsat, B., Spielmann, A. & Stutz, E. (1993) *Nucleic Acids Res.* **21,** 3537–3544.
19. Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.-X., Martin, W. & Cerff, R. (1994) *Nature (London)* **367,** 387–389.
20. Cook, J. R. (1968) in *The Biology of Euglena II*, ed. Buetow, D. E. (Academic, New York), pp. 243–314.
21. Henze, K., Schnarrenberger, C., Kellermann, J. & Martin, W. (1994) *Plant Mol. Biol.* **26,** 1961–1973.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
23. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
24. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
25. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
26. Meyer-Gauen, G., Cerff, R., Schnarrenberger, C. & Martin, W. (1994) *Plant Mol. Biol.* **26,** 1155–1166.
27. Corliss, J. (1994) *Acta Protozool.* **33,** 1–51.
28. Branlant, G. & Branlant, C. (1985) *Eur. J. Biochem.* **150,** 61–66.
29. Alefounder, P. R. & Perham, R. N. (1989) *Mol. Microbiol.* **3,** 723–732.
30. Nelson, K., Whittam, T. S. & Selander, R. K. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6667–6671.
31. Chan, R., Keller, M., Canaday, S., Weil, J. & Imbault, P. (1990) *EMBO J.* **9,** 333–338.
32. Houlné, G. & Schantz, R. (1988) *Mol. Gen. Genet.* **213,** 479–486.
33. von Heijne, G., Hirai, T., Klösgen, R.-B., Stepphuhn, J., Bruce, B., Keegstra, K. & Herrmann, R. (1991) *Plant Mol. Biol. Reporter* **9,** 104–126.
34. Kishore, R., Muchhal, U. & Schwartzbach, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 11845–11849.
35. Michels, P., Marchand, M., Kohl, L., Allert, S., Vellieux, F. M. D., Wierenga, R. & Opperdoes, F. (1991) *Eur. J. Biochem.* **198,** 421–428.
36. Hannaert, V., Blaauw, M., Kohl, L., Allert, S., Opperdoes, F. R. & Michels, P. A. M. (1992) *Mol. Biochem. Parasitol.* **55,** 115–126.
37. Doolittle, R., Feng, D., Anderson, K. & Alberro, M. (1990) *J. Mol. Evol.* **31,** 383–388.
38. Wiemer, E. A. C., Hannaert, V., van den IJssel, P. R. L. A., Van Roy, J., Opperdoes, F. & Michels, P. A. M. (1995) *J. Mol. Evol.* **40,** 443–454.
39. Hashimoto, T., Nakamura, Y., Nakamura, F., Shirakura, T., Adachi, J., Goto, N., Okamoto, K. & Hasegawa, M. (1994) *Mol. Biol. Evol.* **11,** 65–71.
40. Clark, C. G. & Roger, A. J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 6518–6521.
41. Gupta, R. S., Aitken, K., Falah, M. & Singh, B. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2895–2899.
42. Müller, M. (1988) *Annu. Rev. Microbiol.* **42,** 465–488.
43. Markos, A., Miretsky, A. & Müller, M. (1993) *J. Mol. Evol.* **37,** 631–643.
44. Doolittle, R. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 2421–2423.
45. Tessier, L. H., Paulus, F., Keller, M., Vial, C. & Imbault, P. (1995) *J. Mol. Biol.* **245,** 22–33.
46. Cech, T. R. (1990) *Annu. Rev. Genet.* **59,** 543–568.
47. Michel, F., Umesono, K. & Ozeki, H. (1989) *Gene* **82,** 5–30.
48. Copertino, D. W., Shigeoka, S. & Hallick, R. B. (1992) *EMBO J.* **11,** 5041–5050.
49. Muchhal, U. S. & Schwartzbach, S. D. (1992) *Plant Mol. Biol.* **18,** 287–299.
50. Stoltzfus, A., Spencer, D. F., Zuker, M., Logsdon, J. M. & Doolittle, W. F. (1994) *Science* **265,** 202–207.