JOURNAL OF BACTERIOLOGY, Nov. 1997, p. 6798–6806 Vol. 179, No. 21
0021-9193/97/$04.00+0
Copyright © 1997, American Society for Microbiology

Origin and Evolution of Group I Introns in Cyanobacterial tRNA Genes

BRUNO PAQUIN,[1] SCOTT D. KATHE,[2] SANDRA A. NIERZWICKI-BAUER,[2] AND DAVID A. SHUB[1]*

Department of Biological Sciences and Center for Molecular Genetics, University at Albany-SUNY, Albany, New York 12222,[1] and Plant Research Group, Biology Department, Rensselaer Polytechnic Institute, Troy, New York 12181[2]

Received 24 April 1997/Accepted 21 August 1997

Many $tRNA^{Leu}_{UAA}$ genes from plastids contain a group I intron. An intron is also inserted in the same gene at the same position in cyanobacteria, the bacterial progenitors of plastids, suggesting an ancient bacterial origin for this intron. A group I intron has also been found in the $tRNA^{fMet}$ gene of some cyanobacteria but not in plastids, suggesting a more recent origin for this intron. In this study, we investigate the phylogenetic distributions of the two introns among cyanobacteria, from the earliest branching to the more derived species. The phylogenetic distribution of the $tRNA^{Leu}_{UAA}$ intron follows the clustering of rRNA sequences, being either absent or present in clades of closely related species, with only one exception in the *Pseudanabaena* group. Our data support the notion that the $tRNA^{Leu}_{UAA}$ intron was inherited by cyanobacteria and plastids through a common ancestor. Conversely, the $tRNA^{fMet}$ intron has a sporadic distribution, implying that many gains and losses occurred during cyanobacterial evolution. Interestingly, a phylogenetic tree inferred from intronic sequences clearly separates the different tRNA introns, suggesting that each family has its own evolutionary history.

Ever since their discovery, the origin of introns has been a subject of controversy. One view, the introns-late hypothesis, proposes that introns are recent invaders and that split genes arose by late insertion of introns into originally uninterrupted genes (28). In that scenario, horizontal transfer and transposition of introns are frequent events, accounting for the scattered phylogenetic distribution of introns. Although the debate has focused on spliceosomal introns, such a scenario could apply as well to other types of introns, some of which are known to be mobile (22). In contrast, the introns-early view implies that introns are very ancient, being present in the progenote (universal ancestor) (7). The demonstration that some members of group I and group II introns are capable of in vitro autocatalytic activity (19, 29, 39) lends further support to the presence of these introns at an early stage of evolution, maybe as early as the putative precellular RNA world (13). In such a scenario, the observed phylogenetic distribution of introns could be explained by multiple losses in different lineages during evolution (7) and by their mobility, which is assumed to be a derived feature (2). A major obstacle for the introns-early hypothesis was the apparent absence of introns in eubacteria, although this was tentatively rationalized by pressure to streamline the genome in rapidly dividing bacteria (7). Discovery of group I introns in bacteriophages of gram-positive and gram-negative bacteria did not help to resolve the issue, due to uncertainties concerning the origin of the bacteriophages themselves (see discussion in reference 35). The recent discovery of both group I and group II introns in divergent eubacteria (4, 11, 12, 20, 31, 44) was acclaimed as a breakthrough by introns-early proponents. In most cases, however, the relationship of the eubacterial introns to their eukaryotic counterparts is still unclear. Evidence for either of these two hypotheses is at best circumstantial, and the issue remains highly speculative.

All the characterized eubacterial group I introns have been found in the anticodon loop regions of tRNAs (4, 20, 31, 44). Several cyanobacteria were shown to possess a group I intron in either or both of their $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ genes (4, 20, 44). The $tRNA^{Leu}_{UAA}$ intron represents the most likely example of an ancient intron, because a similar intron is known to be present in the corresponding gene of many plastids at the same position as those recently discovered in some cyanobacteria. Since the plastids arose from a cyanobacterial endosymbiont (8), it has been suggested that this intron was inherited from a common ancestor. However, one could argue that an alternative explanation of these findings is invasion of a mobile intron in cyanobacteria. In fact, the striking similarity between the plastid and cyanobacterial introns (20, 44) is consistent with a recent divergence. In addition, although the $tRNA^{Leu}_{UAA}$ intron was originally found in five different strains (20, 44), these represent only a small subset of the cyanobacterial lineages, and more-recent data show that the intron is absent in two other strains representing one of the previously unexplored lineages (4).

The intron inserted in the $tRNA^{fMet}$ genes was found in only 3 of 11 cyanobacteria, two of which did not have the $tRNA^{Leu}_{UAA}$ intron (4). On the basis of the observation that one of the $tRNA^{fMet}$ introns contains an open reading frame (ORF) which presumably assists the intron in its own mobility, and the absence of correlation between relatedness of the intron sequences and relatedness of the rRNAs of the three strains in which they were found, Biniszkiewicz et al. (4) suggested that the $tRNA^{fMet}$ intron arose recently in cyanobacteria by horizontal transfer. To help resolve the origin and evolution of both introns, we have undertaken a systematic investigation of their phylogenetic distribution among cyanobacteria.

MATERIALS AND METHODS

Strains, growth conditions, and DNA preparation. Strains were obtained as living cultures either from John Waterbury (Marine Biological Laboratory, Woods Hole, Mass.), the American Type Culture Collection (Rockville, Md.), or the Pasteur Culture Collection (Institut Pasteur, Paris, France). Cultures were grown and DNA extractions were done as described previously (4, 44).

PCR amplification, cloning, and sequencing. PCR amplifications were carried out with 100 ng of genomic DNA by using *Taq* DNA polymerase (Boehringer Mannheim) and degenerate primers targeting either the $tRNA^{Leu}_{UAA}$ gene (TL25

* Corresponding author. Mailing address: Department of Biological Sciences and Center for Molecular Genetics, University at Albany-SUNY, Albany, NY 12222. Phone: (518) 442-4324. Fax: (518) 442-4767. E-mail: shub@cnsunix.albany.edu.

TABLE 1. Cyanobacterial strains used in this study

Species or strain	Leu[a]	fMet[b]	Reference[c]
Anabaena cylindrica	+	−	This study; U83251
Calothrix desertica PCC7102	+	−	This study; U83252
Chamaesiphon subglobosus	−	−	This study
Chlorogloeopsis fritschii	+	−	This study; U83257
Cylindrospermum strain PCC7417	+	+	This study; U83250, U83261
Dermocarpa strain PCC7437	−	+	4
Fischerella ambigua UTEX 1903	+	−	This study; U83258
Gloeobacter violaceus PCC7421	−	+	This study; U83262
Microcystis aeruginosa	−	−	This study
Myxosarcina strain PCC7312	−	−	This study
Nostoc strain PCC7120	+	−	44; this study
Nostoc strain PCC73102	+	−	This study; U83254
Oscillatoria strain PCC6304	+	+	This study; U83255, U83260
Oscillatoria strain PCC7105	−	+	This study; U83259
Phormidium ectocarpi PCC7375	+	−	This study; U83256
Plectonema boryanum ATCC 18200	−	−	This study
Prochlorothrix hollandica	+	−	38; this study
Pseudanabaena strain PCC7403	+	−	This study; U83253
Pseudanabaena strain PCC6903	−	−	This study
Scytonema strain PCC7110	+	ND[d]	20
Scytonema hofmanii UTEX 2349	(+)[e]	+	4
Spirulina strain PCC6313	−	−	This study
Synechococcus strain R2	+	−	20; this study
Synechocystis strain PCC6803	−	+	4
Synechocystis strain PCC6906	−	(+)[e]	This study

[a] Presence (+) or absence (−) of insertion of a group I intron in the $tRNA^{Leu}_{UAA}$ gene.
[b] Presence (+) or absence (−) of insertion of a group I intron in the $tRNA^{fMet}$ gene.
[c] The GenBank accession numbers are given for the intron sequences determined in this study.
[d] ND, not determined.
[e] (+), presence of the intron deduced solely from the size of a PCR amplification product.

and TL23) or the $tRNA^{fMet}$ gene (SM1 and SM2) (4). The amplification consisted of 30 three-step cycles of denaturation (30 s at 95°C), annealing (1 min at 37°C), and elongation (30 s at 72°C) followed by a final incubation at 72°C for 10 min. The PCR products were then trimmed by a 30-min incubation at 12°C with T4 DNA polymerase (New England Biolabs) and purified by passage through a column (QIAQUICK PCR purification kit; QIAGEN). The fragments were cloned in a *Sma*I-linearized pBluescript II plasmid (Stratagene). Chain termination sequencing reactions were performed with selected clones by using phage T7 DNA polymerase (Sequenase II; U.S. Biochemicals) according to the manufacturer's instructions. Sequence samples were separated by electrophoresis in a 5% polyacrylamide–8 M urea gel. To overcome any possibility of PCR-induced sequence error, two different clones from independent PCRs were analyzed. In case of discrepancies, a third independent PCR clone was sequenced and a consensus was deduced. We confirmed the source of the PCR products by amplifying a region of the gene coding for the small-subunit (SSU) rRNA. Two sets of primers were used, either rRN-5′ (5′-CAGCAGTGGGGAATTTTCCGCAA-3′) and rRN-3′ (5′-GTTGCGCTCGTTGCGGGACTTAA-3′) or rRN2-5′ [5′-CTCGCGTC(A/T)GATTAGCTAG(A/T)TGG-3′] and rRN2-3′ [5′-CCAGGCGGGA(C/T)ACTTAACGCG-3′].

Southern hybridization and genomic cloning. Two micrograms of genomic DNA from *Gloeobacter violaceus* PCC7421 was digested overnight with an excess of *Pst*I, and the restriction fragments were separated on a 0.8% agarose gel. The DNA was transferred by capillary blotting to a Hybond-N nylon membrane (Amersham) according to the manufacturer's instructions. The PCR-amplified $tRNA^{Leu}_{UAA}$ intron from *Synechococcus* sp. strain R2 (20) was internally labeled with [α-^{32}P]dATP by random priming, and the hybridization was carried out at 55°C by standard protocols (34). Construction of a subgenomic library in pBluescript II and colony hybridization screening were also performed according to standard protocols (34).

Phylogenetic analyses. For phylogenetic inference, we extracted the prealigned SSU rRNA sequences of cyanobacteria from the Ribosomal Database Project (RDP) (23). The intron sequences were aligned by comparing both their primary and their secondary structures. Only unambiguously aligned positions corresponding to the catalytic core were considered. Phylogenetic trees were computed by using programs supplied in the PHYLIP package, version 3.5c (10), with default parameters. First, a distance matrix was calculated from the Kimura two-parameter equation (transversions weighted two times transitions) (17), and the trees were built by a neighbor-joining algorithm (33). The robustness of the tree topologies was tested by bootstrap analysis (9). Alternatively, trees (not shown) were also inferred by using fastDNAml (27).

Nucleotide sequence accession numbers. The nucleotide sequences generated for this paper have been deposited in GenBank under accession no. U83250 to U83262. See Table 1 for a detailed list.

RESULTS

The aim of this study was to determine the phylogenetic distribution of group I introns inserted in both $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ genes among the cyanobacteria. We selected a variety of strains (Table 1) representative of each of the lineages within the cyanobacterial subtree from the earliest branching to the more derived species, based on the phylogenetic tree compiled in the RDP (23).

PCR amplification of the $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ genes from various cyanobacteria. Genes coding for both $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ were amplified by PCR of genomic DNA, using primers annealing to mature tRNA sequences. Amplification of an uninterrupted gene should give a product of ~70 to 85 bp, whereas an intron-containing tRNA gene is expected to produce a fragment of 300 to 400 bp. Products longer than 400 bp could indicate an ORF-containing intron (e.g., see reference 4). As depicted in Fig. 1, when amplification is done with the $tRNA^{Leu}$ primers, most of the templates gave a single major product consistent with either an intron-containing or an uninterrupted gene. Since the primers were designed to amplify a portion of the tRNA gene comprising the anticodon stem and loop and part of the extra arm (4), the specificity of the PCR products could be confirmed by sequencing (Fig. 2). In all cases where an intron-containing fragment was amplified, it was shown to include a $tRNA^{Leu}_{UAA}$ with a group I intron inserted between the U and A of the anticodon, exactly the same position where similar introns were localized previously (20, 44). The sizes of the different introns range from 216 (*Oscillatoria* strain PCC6304) to 308 (*Scytonema* strain PCC7110 [20]) nucleotides; none of them encodes an ORF.

For the strains showing only a tRNA-size band, the absence

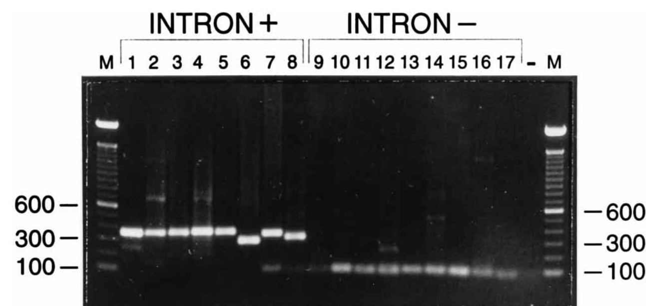

FIG. 1. PCR amplification products of cyanobacterial $tRNA^{Leu}_{UAA}$ genes. Migration in a 1% agarose gel of the amplification products of PCR with the $tRNA^{Leu}$ primers and various cyanobacterial DNAs is shown. Lane 1, *Anabaena cylindrica*; lane 2, *Chlorogloeopsis fritschii*; lane 3, *Cylindrospermum* strain PCC7417; lane 4, *F. ambigua* UTEX 1903; lane 5, *Nostoc* strain PCC73102; lane 6, *Oscillatoria* strain PCC6304; lane 7, *Phormidium ectocarpi* PCC7375; lane 8, *Pseudanabaena* strain PCC7403; lane 9, *C. subglobosus*; lane 10, *G. violaceus* PCC7421; lane 11, *Microcystis aeruginosa*; lane 12, *Myxosarcina* strain PCC7312; lane 13, *Oscillatoria* strain PCC7105; lane 14, *P. boryanum* ATCC 18200; lane 15, *Pseudanabaena* strain PCC6903; lane 16, *Spirulina* strain PCC6313; lane 17, *Synechocystis* strain PCC6906; lane −, no DNA (negative control); lane M, 100-bp DNA marker (Gibco/BRL) (sizes are indicated on the sides of the gel).

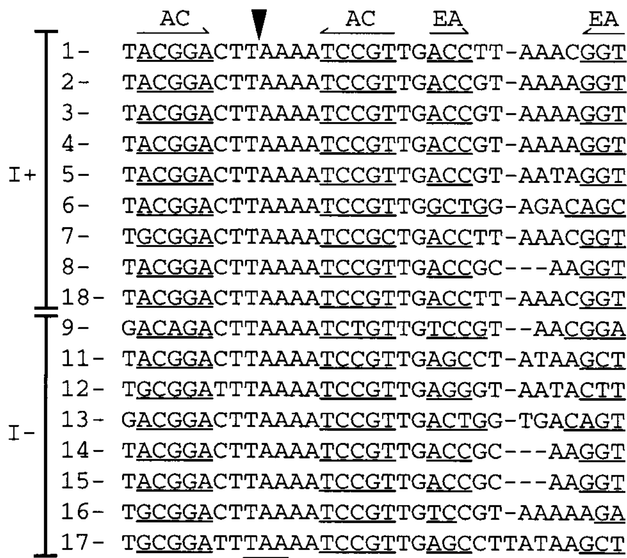

FIG. 2. Alignment of the $tRNA^{Leu}_{UAA}$ gene sequences amplified by PCR. Sequence numbers correspond to the lanes in Fig. 1, with the addition of no. 18 for *Calothrix* strain PCC7102. Base-pairing (underline) in the anticodon stem (AC) and extra arm (EA), the anticodon (double underline), and the intron insertion site (arrowhead) are indicated. I+ and I−, intron positive and negative, respectively.

of the intron was also confirmed by sequencing. In many cases the band was heterogeneous in composition, comprising different $tRNA^{Leu}$ isoacceptor gene sequences. However, with the exception of *Gloeobacter* strain PCC7421 (Fig. 1, lane 10), we were able to find an intronless form of the $tRNA^{Leu}_{UAA}$ gene (Fig. 2). With some of the DNAs, both intron-containing and tRNA-size bands were detectable, especially with *Phormidium ectocarpi* PCC7375 and *Pseudanabaena* strain PCC7403 (Fig. 1, lanes 7 and 8, respectively). Consistent with the heterogeneity of tRNA-size bands, the sequences of six clones (three from each species) showed that the lower band resulted from amplification of $tRNA^{Leu}_{NAG}$ and $tRNA^{Leu}_{CAA}$ genes, not from a second, intronless copy of the $tRNA^{Leu}_{UAA}$ gene. Fragments of intermediate or greater size were occasionally amplified (e.g., Fig. 1, lanes 1, 2, 12, and 13) but were not systematically analyzed. A few of them were sequenced, but none had characteristics expected for group I intron or tRNA sequences.

Amplification of the $tRNA^{fMet}$ genes yielded a single product for each template, either an intron-containing band or a tRNA-size band (data not shown; for examples, see reference 4). In addition to the three strains from which the $tRNA^{fMet}$ intron was already characterized (4), it was found in a few other strains: *Gloeobacter* strain PCC7421, *Oscillatoria* strain PCC7105, *Oscillatoria* strain PCC6304, and *Cylindrospermum* strain PCC7417. Two intronless bands were sequenced (*Myxosarcina* strain PCC7312 and *Synechococcus* strain R2), both displaying an uninterrupted 7-bp sequence consistent with the anticodon loop of a $tRNA^{fMet}$ gene. Except for the *Synechocystis* strains PCC6803 (4) and PCC6906, none of the $tRNA^{fMet}$ introns encodes an ORF.

To ensure that the PCR products were from the expected strains and not from contaminating DNA, we also amplified a portion of the SSU rRNA gene. We then aligned our sequences (172 to 551 bp) with the corresponding ones extracted from the database (23). No more than three mismatches were observed in any alignment (<1.6%), a level of discrepancy that could be caused by PCR-induced errors, except for the *Fischerella* pair, for which 19 mismatches had to be assumed in a 223-bp segment (8.5%). Our sequence of *Fischerella* was obtained from *Fischerella ambigua* UTEX 1903, whereas the database sequence is from *Fischerella* sp. strain PCC7414, probably explaining the discrepancies between the two sequences. Still, the *F. ambigua* rRNA sequence was closer to those of PCC7414 and other closely related strains (as defined by branch G in Fig. 5) than to any other sequence (not shown). Therefore, it is unlikely that the PCR products have been amplified from a contaminating DNA.

***Gloeobacter* strain PCC7421 has no $tRNA^{Leu}_{UAA}$ intron homolog in its genome.** All attempts to amplify the $tRNA^{Leu}_{UAA}$ gene of *G. violaceus* PCC7421 have been unsuccessful. We tried different combinations of $tRNA^{Leu}$ primers shortened by one or two bases at the 3′ end, and also those used by Kuhsel et al. (20), without success. A total of 15 cloned amplification products were analyzed, and all of them had sequences consistent with uninterrupted genes of the $tRNA^{Leu}_{NAG}$ family (12 CAG, 2 TAG, and 1 GAG). *G. violaceus* PCC7421 is a key strain for our study, since it has been identified as the earliest-branching cyanobacterium (8, 23, 41). Therefore, we performed a Southern hybridization, using a labeled intron-containing PCR product from *Synechococcus* strain R2 (also known as PCC7942 or *Anacystis nidulans* R2) (20) to probe its presence in *Gloeobacter*. A *Pst*I fragment of 3.5 kb hybridized to the probe (Fig. 3A). Except for a fragment of about 1.4 kb that was barely visible on the original autoradiogram, no other signal was observed. The 3.5-kb fragment was cloned, and a 1.8-kb *Bam*HI subclone containing the hybridizing region (not shown) was completely sequenced on both strands (Fig. 3B) (accession no. U83262). A typical group I intron within a tRNA gene was identified, but rather than revealing a $tRNA^{Leu}$ intron, the sequence was identical to the one that had been amplified by PCR using the $tRNA^{fMet}$ primers. The exon sequences can be folded into a structure possessing the hallmarks of $tRNA^{fMet}$ (30), including the absence of a Watson-Crick base pair at the top of the acceptor stem and the critical GC base pairs in both the acceptor and the anticodon stems (Fig. 4). The intron sequence is also typical of the $tRNA^{fMet}$ introns previously described, harboring an extended P1 stem (comprising up to 62 nucleotides) (reference 4 and this study). Downstream of the $tRNA^{fMet}$ gene, the clone ends in a large ORF of more than 395 codons (ORF395). BLASTP analysis (1) of the translated sequence revealed a high level of identity to hypothetical ORFs of unknown function in the cyanobacteria *Synechocystis* strain PCC6803 (276 of 384, or 72% [accession no. D90910]) and *Prochlorococcus marinus* (199 of 371, or 54% [32]).

Most of the remainder of the 3.5-kb *Pst*I fragment was sequenced on at least one strand, and neither another group I intron nor a $tRNA^{Leu}$ was identified. It is unlikely that while hybridizing to a similar intron from a different gene the probe would have missed the homologous intron. Therefore, we concluded that there is no $tRNA^{Leu}_{UAA}$ intron in *Gloeobacter* strain PCC7421. The presence of a single TTA codon in ORF395 suggests, however, that $tRNA^{Leu}_{UAA}$ is encoded in this genome, despite our numerous unsuccessful attempts to amplify its gene by PCR.

The $tRNA^{Leu}_{UAA}$ intron has a wide distribution among cyanobacteria. SSU rRNA sequences were extracted from the RDP (23). A phylogenetic tree was then constructed by using a neighbor-joining algorithm (17, 33), and its robustness was tested by bootstrap analysis (9). The topology of the tree (Fig. 5) is in general agreement with the tree inferred with fast DNAml (27) (tree not shown) and those published by other groups (23, 40–42). The main branches defining the different

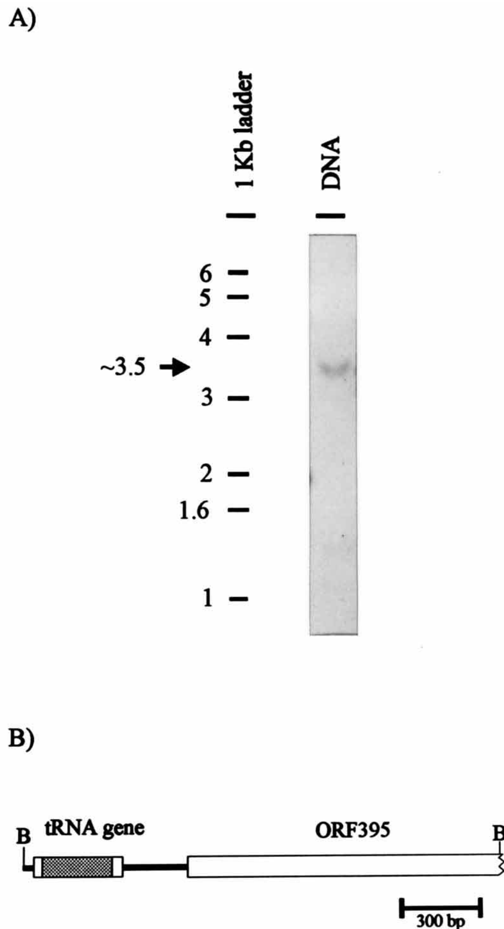

FIG. 3. Lack of $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ intron homolog in *Gloeobacter*. (A) Southern hybridization of *Gloeobacter* DNA digested with *Pst*I. Positions and approximate sizes (in kilobases) of the 1-kb marker fragments (Gibco/BRL) are indicated on the left. The 3.5-kb fragment hybridizing with the probe is shown (arrow). The probe consisted of the radiolabeled, intron-containing PCR product obtained by using the $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ primers with *Synechococcus* DNA. (B) Schematic representation of the sequence of a *Bam*HI (B) subclone containing the hybridization signal. The sequence has been deposited in GenBank under accession no. U83262. Genes or putative genes (open boxes), the location of a group I intron within the tRNA gene (shaded area), and the intergenic spacers (thick black lines) are indicated. The scale for 300 bp is shown.

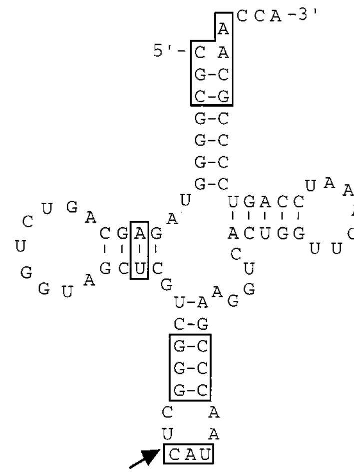

FIG. 4. Structure of the *G. violaceus* PCC7421 $\text{tRNA}^{\text{fMet}}$. The sequence of the $\text{tRNA}^{\text{fMet}}$ was determined from a genomic clone. Diagnostic features for a $\text{tRNA}^{\text{fMet}}$ (30) (boxed) and the intron insertion site (arrow) are shown.

lineages (branches B to E in Fig. 5) are consistently recovered, but their relative branching order may change, as suggested by some unsupported branches. The positions of *Chamaesiphon subglobosus* and *Plectonema boryanum* (within branch F) and those of the two *Oscillatoria* sequences (strains PCC6304 and PCC7105) are unresolved in this tree, and their relative placement is also highly variable in other published trees (14, 23, 40–42). In fact, the bootstrap values were fairly low throughout the tree (values in parentheses in Fig. 5), but for many of the branches, they were significantly higher when the *Oscillatoria* strain PCC6304 and *Oscillatoria* strain PCC7105 sequences were omitted, although the topology was preserved (both values are shown only when significantly different [Fig. 5]).

The phylogenetic distributions of both introns are outlined along with the rRNA tree. Not only is the $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ intron widely distributed, but it also follows the clustering of rRNA sequences, being either present or absent in a clade of closely related species. A significant exception is encountered in the *Pseudanabaena* group, where the corresponding intron was found in one strain (PCC7403) but was absent in another (PCC6903). The *Pseudanabaena* strain PCC7403 SSU rRNA sequence is not available for phylogenetic analysis, but it has been reported to form a monophyletic group with the strain PCC6903 sequence supported by very high bootstrap values (40).

Strains harboring an intron in their $\text{tRNA}^{\text{fMet}}$ genes are also indicated; although the intron was found in only a restricted number of strains, it covers a wide portion of the tree. Unlike the $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ intron, the $\text{tRNA}^{\text{fMet}}$ intron seems to be randomly dispersed throughout the cyanobacteria.

Highly conserved intron structures. The sequences of the newly discovered introns in this study were aligned with those of the other, previously published cyanobacterial tRNA group I introns. The degree of sequence similarity between the different introns is surprisingly high. Although it is particularly striking between homologous introns, it is also obvious between the two intron families. For example, while 61% (120 of 196) of the nucleotides included in the alignment are identical among the $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ intron sequences (Fig. 6) and 63% (113 of 179) are identical among the $\text{tRNA}^{\text{fMet}}$ sequences, 38% (76 of 200) are identical in both data sets. Most of these positions are also conserved in the $\text{tRNA}^{\text{Arg}}_{\text{CCU}}$ (69 of 76 nucleotides) and $\text{tRNA}^{\text{Ile}}_{\text{CAU}}$ (65 of 76 nucleotides) intron sequences. Considering the generally low level of conservation between homologous group I intron sequences in organelles (24), these numbers are astonishingly high. Moreover, within the cyanobacterial $\text{tRNA}^{\text{Leu}}_{\text{UAA}}$ sequences, 83% (162 of 196) of the positions are identical in at least 10 of 13 sequences (Fig. 6).

There is also a high degree of secondary structural similarity. The region of the intron core (defined in Fig. 6) can be practically superimposed from one intron to another. The most characteristic features for these tRNA group I introns are (i) an unusually short exon-intron interaction in P1 (3 bp), (ii) a 9-bp P2 stem with a GAAA tetraloop, and (iii) a T-shaped P9 region harboring a 5-bp P9b stem with a GAAA tetraloop (Fig. 6). The P9 region in the $\text{tRNA}^{\text{Ile}}_{\text{CAU}}$ intron from *Azoarcus* strain BH72, however, comprises only a single 5-bp stem with a GAAA tetraloop (31). In addition, an extended P1 stem-loop region distinguishes the $\text{tRNA}^{\text{fMet}}$ introns from the other eubacterial tRNA group I introns (4). In *Synechocystis*, this extension culminates in an ORF.

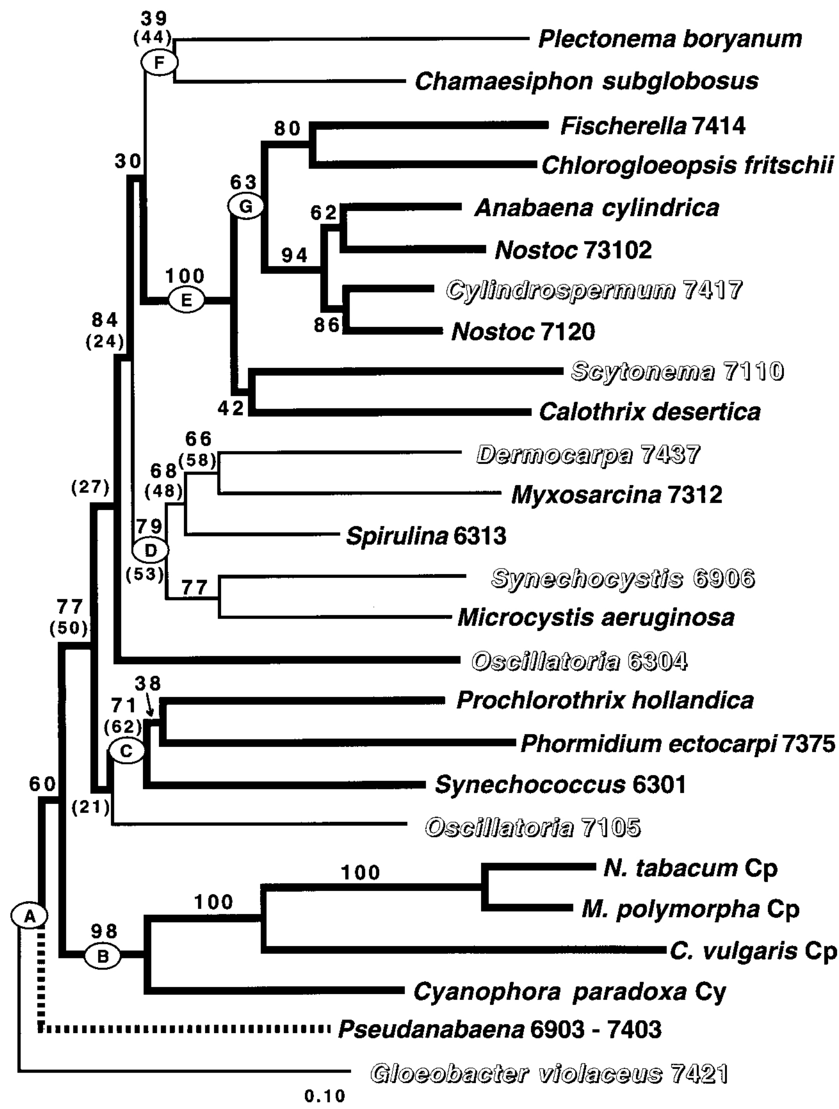

FIG. 5. SSU rRNA phylogenetic tree. The prealigned SSU rRNA sequences were extracted from the RDP (23), and the tree was obtained by using distance methods as described in Materials and Methods. Bootstrap values are given in percentages as calculated from 100 replicates by omitting the *Oscillatoria* strain PCC6304 and PCC7105 sequences. The values that differed by more than 5% when the *Oscillatoria* sequences were included, as well as those of the branches separating these two sequences, are shown in parentheses. The distribution of the $tRNA^{Leu}_{UAA}$ introns is shown as follows: thick branches, intron-positive strains; thin branches, intron-negative strains. The dotted line leading to the *Pseudanabaena* sequence indicates that one strain (PCC7403) has the intron while another (PCC6903) does not. The tree was drawn by assuming an origin of the intron before the endosymbiosis. The strains harboring an intron in their $tRNA^{fMet}$ gene (outlined characters) and branches discussed in the text (letters in ovals) are indicated. *Synechocystis* strain PCC6906 and *Fischerella* strain PCC7414, whose SSU rRNA sequences were used in this tree, are not exactly the same strains from which the intron sequences were obtained (*Synechocystis* strain PCC6803 and *F. ambigua* UTEX 1903). The presence of a highly similar $tRNA^{fMet}$ intron and the absence of a $tRNA^{Leu}_{UAA}$ intron in *Synechocystis* strain PCC6906 were confirmed. In this tree, we chose four representatives of plastids harboring a $tRNA^{Leu}_{UAA}$ intron, but, as noted in the text, this does not imply that the intron is universally present in this lineage. *N. tabacum*, *Nicotiana tabacum*; *M. polymorpha*, *Marchantia polymorpha*; *C. vulgaris*, *Chlorella vulgaris*; Cp, chloroplast; Cy, cyanelle.

Comparison of an intron phylogeny to the SSU rRNA tree. An extended alignment of the core sequences (28a) was produced by adding three $tRNA^{Leu}_{UAA}$ introns from chloroplasts, as well as the $tRNA^{Arg}$ intron from *Agrobacterium tumefaciens* and the $tRNA^{Ile}$ intron from *Azoarcus* (31). A phylogeny based on the intron core sequences was inferred by the same methods used for the SSU rRNA tree (Fig. 7). The branches separating the different intron families are well supported (bootstrap values between 74 and 100%). This separation was reproduced in trees inferred by different methods (fastDNAml and parsimony). The $tRNA^{fMet}$ intron subtree is relatively well supported in bootstrap analysis. However, the only branching order that remained intact in maximum-likelihood analysis was that of *Cylindrospermum*, *Dermocarpa*, and *Scytonema* spp. Even though the early position of the *Gloeobacter* $tRNA^{fMet}$ intron is supported by a high bootstrap value (91% [Fig. 7]), it was not reproduced in the maximum-likelihood tree. In the latter, its branch diverged just prior to the *Cylindrospermum* sequence. Within the $tRNA^{Leu}_{UAA}$ intron subtree, most of the branches are poorly supported, with bootstrap values below 50%. In addition, the topology of the $tRNA^{Leu}_{UAA}$ intron subtree varies extensively depending on the method used (maximum likelihood versus neighbor joining) and the data set (the complete set as in Fig. 7 versus the $tRNA^{Leu}_{UAA}$ introns alone). Given the high

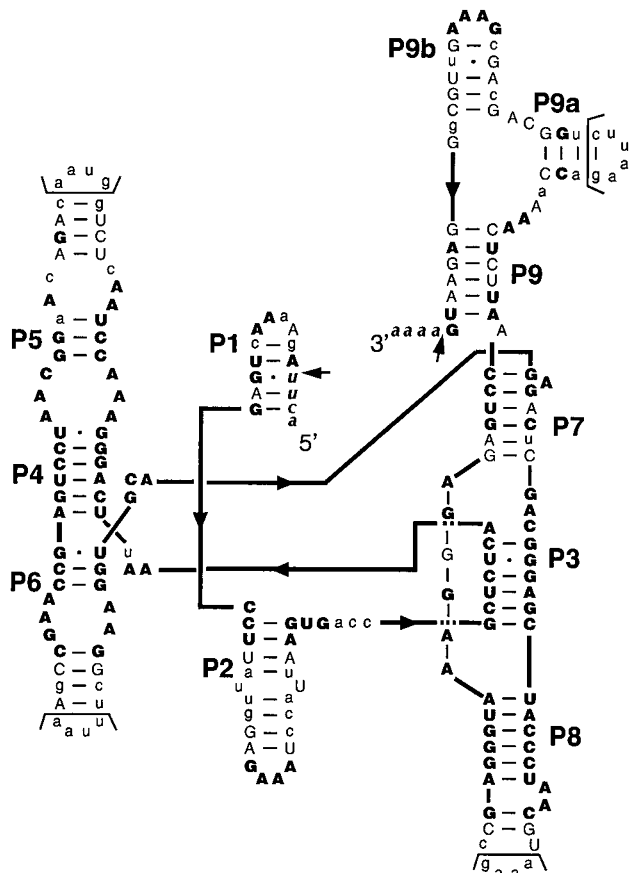

FIG. 6. Secondary structure of the *Oscillatoria* strain PCC6304 $tRNA^{Leu}_{UAA}$ intron. The secondary structure of the intron (accession no. U83255) is shown in the format described by Cech et al. (5). P1 to P9 refer to conserved helices of group I introns (24). The regions defined by brackets have been excluded from the alignment of the core sequences used in phylogenetic analysis. Putative splice sites (arrows), nucleotides identical throughout the cyanobacterial $tRNA^{Leu}_{UAA}$ intron data set (boldface), highly conserved positions (at least 10 of 13 identical) (uppercase), the otherwise less conserved nucleotides (lowercase), and the exon sequence (italics) are indicated.

degree of sequence conservation of the intron core, the lack of resolution is not surprising. Therefore, to avoid confusion, we chose not to show a particular topology for this subtree (Fig. 7).

Upon comparison of the two phylogenetic trees, we found that the $tRNA^{fMet}$ intron subtree and the SSU rRNA tree harbor a major discrepancy. In the latter, *Cylindrospermum* and *Scytonema* are closely related (within branch E), as are *Synechocystis* and *Dermocarpa* (branch D). This topology is not reproduced in the intron tree; instead, a close relationship between *Dermocarpa* and *Scytonema* (88%) is inferred, while *Synechocystis* branches more deeply. This discrepancy was already pointed out by Biniszkiewicz et al. (4), even though they had only these three sequences available. Within the $tRNA^{Leu}_{UAA}$ intron subtree, except for the branch grouping the plastid introns which is consistently recovered with high bootstrap values, the lack of resolution does not permit such a comparison.

DISCUSSION

Group I introns in $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ genes were previously identified in some cyanobacteria, but their respective relationships and phylogenetic distributions remained undetermined. In this study, we have addressed these questions by sampling an ensemble of strains that include representatives from every clade of the cyanobacterial subtree. Our results show that while each intron is present in a wide variety of cyanobacteria, their distribution patterns are very different.

Group I introns in cyanobacteria: migrant or nonmigrant? The $tRNA^{Leu}_{UAA}$ intron was found in a large number of cyanobacteria, and its distribution pattern correlates to the SSU rRNA grouping, with one exception in the *Pseudanabaena* branch. Another possible exception was the report of Kuhsel et al. (20), who, solely on the basis of PCR fragment size, proposed that an intron was inserted in the corresponding genes of both *Plectonema* strain L2 and *Gloeothece* strain PCC6501. Further analyses led J. D. Palmer (27a) to conclude that these fragments (as well as the one from the plastid of *Gracilaria lemaneiformis*) were not bona fide group I intron-containing $tRNA^{Leu}$ products from these organisms. These results are in agreement with ours, since no intron was found in *P. boryanum* ATCC 18200, and *Gloeothece* SSU rRNA is positioned within branch D (not shown, but see references 23, 40, and 42), which defines a group of strains encoding an intronless $tRNA^{Leu}_{UAA}$ gene.

Our data suggest that the $tRNA^{Leu}_{UAA}$ intron is relatively stable and that a cluster of closely related strains harboring the intron inherited it from a common ancestor. For example, the intron was most probably present in the ancestors of each of the branches C and E but absent in the ancestor of branch D (Fig. 5). The intron was also most likely present in the ancestor of branch B, the branch leading to the plastid sequences. It should be noted, however, that although $tRNA^{Leu}_{UAA}$ is widely distributed among plastids (see subgroup IC3 in reference 6a), its evolution in this lineage is punctuated by a few independent losses; for example, the intron is absent from the chloroplast genomes of some algae (6, 15, 18). From our data, it is not possible to predict with certainty the occurrence of the intron in the ancestor of branch A. Its presence in only one of two *Pseudanabaena* strains can be equally explained by horizontal or vertical transmission. In sharp contrast, the $tRNA^{fMet}$ intron is present in a more restricted number of cyanobacteria, and its distribution pattern is sporadic, with no correlation to the SSU rRNA clustering. In this regard, this intron appears less stable, and the presence of an intronic ORF in one strain (4) that has endonuclease activity (4a) suggests that this intron is capable of mobility (for a review of intron mobility, see reference 22).

Origin of the tRNA introns: early versus late. The phylogenetic distribution of the $tRNA^{Leu}_{UAA}$ intron implies that it is relatively ancient. In contrast, the high similarity of the intron sequences points to a recent propagation of a mobile intron. However, a late scenario requires a minimum of 13 transfers (one for each of the intron-containing strains), a number of events significantly greater than the five predicted by early scenarios (see below). Thus, assumption of an ancient origin for the $tRNA^{Leu}_{UAA}$ intron is the most parsimonious solution.

If the intron were present in branch A (before the endosymbiosis), its present distribution would require a total of four independent losses, one each for *Pseudanabaena* strain PCC6903, *Oscillatoria* strain PCC7105, and branches D and F. Alternatively, if the intron first originated in the plastid lineage (after the endosymbiosis), four gains (transfers) are postulated, one each to *Pseudanabaena* strain PCC7403, *Oscillatoria* strain PCC6304, and branches C and E. The number of unweighted events for these two scenarios is five (including the original insertion event). Although we cannot systematically rule out one of them, we favor the "before" scenario as the more likely, since gains and losses are not equally probable in

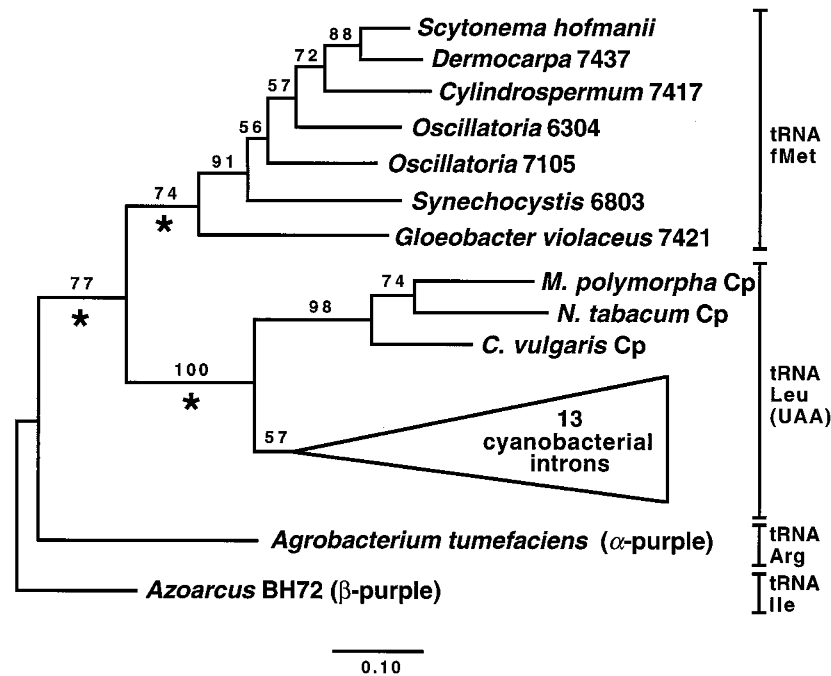

FIG. 7. Intron phylogenetic tree, inferred by using an alignment of the intron core sequences (defined in Fig. 6 and reference 28a) as described in Materials and Methods and the legend to Fig. 5. Branches separating the different intron families are marked (asterisks). Because the $tRNA^{Leu}_{UAA}$ intron subtree is unresolved and highly unstable, a particular topology is not shown (see text). References for all the cyanobacterial tRNA intron sequences are given in Table 1. The sources of the other sequences are as follows: $tRNA^{Arg}_{CCU}$ intron of *A. tumefaciens* A136 and $tRNA^{Ile}_{CAU}$ intron of *Azoarcus* strain BH72, reference 31; and $tRNA^{Leu}_{UAA}$ introns of the chloroplasts of *Chlorella vulgaris*, *M. polymorpha*, and *N. tabacum*, references 20, 25, and 45, respectively.

this system. A gain implies a horizontal transfer, and this particular intron is unlikely to be mobile due to the absence of a homing endonuclease. Although propagation by reverse splicing followed by cDNA recombination is not impossible (43), it is clearly less likely than intron losses, which do not involve any type of interspecific genetic exchange. Thus, the "after" scenario, which entails four transfers, appears less probable. Consequently, the $tRNA^{Leu}_{UAA}$ intron was likely present in the common ancestor of cyanobacteria and plastids before the endosymbiosis event. Inheritance of an ancient group I intron has also been proposed in other cases (3, 26).

The correlation between the intron distribution and the SSU rRNA tree (Fig. 5) is only as solid as the tree itself. Although most of the important branches are reasonably well supported, the two *Oscillatoria* branches (strains PCC7105 and PCC6304, supported at 21 and 27%, respectively) and branch F (39%) and its preceding branch (30%) are clearly not stable. However, regardless of the actual placement of the unsupported branches, our conclusion would remain the same because the number of events (losses or gains) is unlikely to change significantly. For example, according to a topology published by Wilmotte et al. (42), *Oscillatoria* strain PCC7105 would be placed in branch D, subtracting one loss in the before scenario. However, this topology is also poorly supported.

In contrast to the case for the $tRNA^{Leu}_{UAA}$ intron, many observations point to a recent origin of the $tRNA^{fMet}$ intron. These include its apparent absence in plastids, the lack of congruence between the intron and the SSU rRNA phylogenetic trees, its sporadic distribution among cyanobacteria, and the presence in one strain of an intron containing a motility-conferring ORF. Following these criteria, its distribution would be better explained by horizontal propagation. On the other hand, the inferred number of losses in an early scenario is not significantly higher than the number of gains in a late scenario (11 losses versus 7 gains). More important, the early branching position of the *Gloeobacter* $tRNA^{fMet}$ intron, a placement congruent to that of the corresponding SSU rRNA sequence (Fig. 5 and 7), could have suggested an early origin for this intron. However, for reasons already outlined in Results, this branching order is uncertain. Therefore, we believe that the $tRNA^{fMet}$ intron probably originated late in the evolution of cyanobacteria.

Presence of $tRNA^{Leu}_{UAA}$ and $tRNA^{fMet}$ introns in other eubacterial phyla. Using the same primers, we tried to amplify the $tRNA^{Leu}_{UAA}$ gene of other eubacteria. In every case except one, the PCR products resulted from amplification of one of the $tRNA^{Leu}_{NAG}$ genes. Only with *Chlorobium thiosulfatopilum*, a green sulfur bacterium, were we able to amplify the $tRNA^{Leu}_{UAA}$ gene, which has no intron (data not shown). Moreover, the intron is absent in all published sequences of homologous eubacterial genes (reference 36 and references therein), comprising data from diverse phyla such as β-purple (*Azoarcus*), γ-purple (*Escherichia coli* and *Haemophilus influenzae*), and high-GC (*Streptomyces coelicolor*) and low-GC (*Bacillus subtilis*, *Staphylococcus aureus*, and *Mycoplasma*) gram-positive bacteria. Preliminary evidence for a homologous intron in purple bacteria and *Thermotoga* spp. was reported (21), but no confirmation of these data has been published. Currently, the evidence does not allow us to conclude whether the origin of the $tRNA^{Leu}_{UAA}$ intron predated the cyanobacterial radiation.

An extensive search for the $tRNA^{fMet}$ intron in other eubacteria has already been conducted; the intron was not found in any of the tested species (3a, 4).

Does the $tRNA^{Leu}_{UAA}$ intron have a biological function? Considering their ancient origin, the level of identity among the $tRNA^{Leu}_{UAA}$ intron sequences is striking. As a result of a long

coexistence, it is possible that the $tRNA^{Leu}_{UAA}$ intron has evolved to provide the host genome with a positive selective value. This would also explain why the intron was retained in virtually all the descendants of an intron-containing ancestor (Fig. 5, clades E and C). In *Nostoc* strain PCC7120 (previously known as *Anabaena* strain PCC7120), it was shown that formation of the anticodon stem in the pre-tRNA is required for optimal splicing of the intron (46). This interaction between the tRNA and the intron, which is believed to compensate for the very short exon-intron pairings in P1 and the apparent absence of P10, may reflect an adaptation of the intron to its location. More striking is the necessity of an intron in some eukaryotic nuclear tRNA genes for a base modification to occur in the tRNA anticodon (16, 37). Even though the nuclear tRNA introns are not related to group I introns, such a role could account for the stability and the high degree of sequence conservation of the $tRNA^{Leu}_{UAA}$ introns in cyanobacteria. However, any selective value provided by the intron must be compensated by other factors in intron-negative strains.

Concluding remarks. The evolution of the two group I intron families in cyanobacteria can be explained by different evolutionary pathways, lateral transfer ($tRNA^{fMet}$ introns) and common ancestry ($tRNA^{Leu}_{UAA}$ introns). Furthermore, even though the intron phylogenetic tree (Fig. 7) shows that the two intron families have a monophyletic origin, the degree of structural conservation between them suggests that they share a more recent common ancestor to the exclusion of the other group I introns.

ACKNOWLEDGMENTS

We thank John Waterbury for providing us with some of the cultures, Annick Wilmotte for helpful discussions and sharing of unpublished sequences, Jeffrey D. Palmer for communication of unpublished observations, Robin Gutell for help with Fig. 6, and Caro-Beth Stewart and Randall Collura for helpful discussions and critical reading of the manuscript.

B.P. was supported in part by a fellowship from the Natural Sciences and Engineering Research Council (Canada). This work was supported by grant GM37746 (D.A.S.) from the National Institutes of Health.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Bell-Pedersen, D., S. Quirk, J. Clyman, and M. Belfort.** 1990. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. Nucleic Acids Res. **18:**3763–3770.
3. **Bhattacharya, D., B. Surek, M. Rüsing, S. Damberger, and M. Melkonian.** 1994. Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). Proc. Natl. Acad. Sci. USA **91:**9916–9920.

3a. **Biniszkiewicz, D., and D. A. Shub.** Unpublished results.

4. **Biniszkiewicz, D., E. Cesnaviciene, and D. A. Shub.** 1994. Self-splicing group I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO J. **13:**4629–4635.

4a. **Bonocora, R. P., and D. A. Shub.** Unpublished observations.

5. **Cech, T. R., S. H. Damberger, and R. R. Gutell.** 1994. Representation of the secondary structure of group I introns. Nat. Struct. Biol. **1:**273–280.
6. **Chen, Z. Y., and J. V. Moroney.** 1995. Identification of a *Chlamydomonas reinhardtii* chloroplast gene with homology to bacterial genes involved in cytochrome c biosynthesis. Plant Physiol. **108:**843–844.

6a. **Damberger, S. H., and R. A. Gutell.** 1994. A comparative database of group I intron structures. Nucleic Acids Res. **22:**3508–3510.

7. **Darnell, J. E., and W. F. Doolittle.** 1986. Speculations on the early course of evolution. Proc. Natl. Acad. Sci. USA **83:**1271–1275.
8. **Delwiche, C. F., M. Kuhsel, and J. D. Palmer.** 1995. Phylogenetic analysis of *tuf*A sequences indicates a cyanobacterial origin of all plastids. Mol. Phylogenet. Evol. **4:**110–128.
9. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
10. **Felsenstein, J. (Department of Genetics, University of Washington, Seattle).** 1993. PHYLIP (Phylogeny Inference Package), version 3.5c.
11. **Ferat, J.-L., and F. Michel.** 1993. Group II self-splicing introns in bacteria. Nature (London) **364:**358–361.
12. **Ferat, J.-L., M. Le Gouar, and F. Michel.** 1994. Multiple group II self-splicing introns in mobile DNA from *Escherichia coli*. C. R. Acad. Sci. Paris **317:**141–148.
13. **Gilbert, W.** 1986. The RNA world. Nature (London) **319:**618.
14. **Giovannoni, S. J., S. Turner, G. J. Olsen, S. Barns, D. J. Lane, and N. R. Pace.** 1988. Evolutionary relationships among cyanobacteria and green chloroplasts. J. Bacteriol. **170:**3584–3592.
15. **Hallick, R. B., L. Hong, R. G. Drager, M. R. Favreau, A. Monfort, B. Orsat, A. Spielmann, and E. Stutz.** 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. **21:**3537–3544.
16. **Johnson, P. F., and J. Abelson.** 1983. The yeast $tRNA^{Tyr}$ gene intron is essential for correct modification of its tRNA product. Nature **302:**681–687.
17. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
18. **Kowallik, K. V., B. Stoebe, I. Schaffran, P. Kroth-Pancic, and U. Freier.** 1995. The chloroplast genome of a chlorophyll a+c-containing alga, *Odontella sinensis*. Plant Mol. Biol. Rep. **13:**336–342.
19. **Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech.** 1982. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell **31:**147–157.
20. **Kuhsel, M. G., R. Strickland, and J. D. Palmer.** 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science **250:**1570–1573.
21. **Kuhsel, M., and J. D. Palmer.** 1991. Introns shared by bacteria and chloroplasts. J. Cell. Biochem. Suppl. **15D:**68.
22. **Lambowitz, A. M., and M. Belfort.** 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. **62:**587–622.
23. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The ribosomal database project (RDP). Nucleic Acids Res. **24:**82–85.
24. **Michel, F., and E. Westhoff.** 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. **216:**585–610.
25. **Oda, K., K. Yamato, E. Ohta, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, T. Kanegae, Y. Ogura, T. Kohchi, and K. Ohyama.** 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J. Mol. Biol. **223:**1–7.
26. **Ohta, E., K. Oda, K. Yamato, Y. Nakamura, M. Takemura, N. Nozato, K. Akashi, K. Ohyama, and F. Michel.** 1993. Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res. **21:**1297–1305.
27. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:**41–48.

27a. **Palmer, J. D.** Personal communication.

28. **Palmer, J. D., and J. M. Logsdon.** 1991. The recent origins of introns. Curr. Opin. Genet. Dev. **1:**470–477.

28a. **Paquin, B., and D. A. Shub.** 1997. Aligned intron sequences. http://barbara.bio.albany.edu/~carole/ali-Fig.html.

29. **Peebles, C. L., P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell, and H.-L. Cheng.** 1986. A self-splicing RNA excises an intron lariat. Cell **44:**213–223.
30. **RajBhandary, U. L.** 1994. Initiator transfer RNAs. J. Bacteriol. **176:**547–552.
31. **Reinhold-Hurek, B., and D. A. Shub.** 1992. Self-splicing introns in tRNA genes of widely divergent bacteria. Nature (London) **357:**173–176.
32. **Richter, S., W. R. Hess, M. Krause, and W. Messer.** 1997. Cloning of the *dnaA* gene of *Prochlorococcus marinus*. Mol. Gen. Genet., in press.
33. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Shub, D. A.** 1991. The antiquity of group I introns. Curr. Opin. Genet. Dev. **1:**478–484.
36. **Sprinzl, M., C. Steegborn, F. Hübel, and S. Steinberg.** 1996. Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res. **24:**68–72.
37. **Sweykowska-Kulinska, Z., B. Senger, G. Keith, F. Fasiolo, and H. Grosjean.** 1994. Intron-dependent formation of pseudouridines in the anticodon of *Saccharomyces cerevisiae* minor tRNA(Ile). EMBO J. **13:**4636–4644.
38. **Tsinomeras, N. F., M. S. Nalty, K. A. Sukkawala, and S. S. Golden.** 1995. Unpublished results (accession no. U29955).
39. **van der Veen, R., A. C. Arnberg, G. van der Horst, L. Bonen, H. F. Tabak, and L. A. Grivell.** 1986. Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. Cell **44:**225–234.
40. **Wilmotte, A.** 1994. Molecular evolution and taxonomy of the cyanobacteria,

p. 1–25. *In* D. A. Bryant (ed.), The molecular biology of cyanobacteria. Advances in Photosynthesis, vol. 1. Kluwer Academic Publishers, Dordrecht, The Netherlands.

41. **Wilmotte, A., S. Turner, Y. Van de Peer, and N. R. Pace.** 1992. Taxonomic study of marine oscillatoriacean strains (cyanobacteria) with narrow trichomes. II. Nucleotide sequence analysis of the 16S ribosomal RNA. J. Phycol. **28:**828–838.
42. **Wilmotte, A., J.-M. Neefs, and R. De Wachter.** 1994. Evolutionary affiliation of the marine nitrogen-fixing cyanobacterium *Trichodesmium* sp. strain NIBB 1067, derived by 16S ribosomal RNA sequence analysis. Microbiology **140:**2159–2164.
43. **Woodson, S. A., and T. R. Cech.** 1989. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell **57:**335–345.
44. **Xu, M.-Q., S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki-Bauer, and D. A. Shub.** 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. Science **250:**1566–1570.
45. **Yamada, K., K. Shinozaki, and M. Sugiura.** 1986. DNA sequences of tobacco chloroplast genes for tRNA-Ser (GGA), tRNA-Thr (UGU), tRNA-Leu (UAA), tRNA-Phe (GAA): the tRNA-Leu gene contains a 503 bp intron. Plant Mol. Biol. **6:**193–199.
46. **Zaug, A. J., M. M. McEvoy, and T. R. Cech.** 1993. Self-splicing of the group I intron from *Anabaena* pre-tRNA: requirement for the base-pairing of the exons in the anticodon stem. Biochemistry **32:**7946–7953.