Nucleic Acids Research, Vol. 19, No. 23 6485–6490

Diversity of a ribonucleoprotein family in tobacco chloroplasts: two new chloroplast ribonucleoproteins and a phylogenetic tree of ten chloroplast RNA-binding domains

Lizhen Ye, Yuqing Li, Kaoru Fukami-Kobayashi[1], Mitiko Go[1], Tomokazu Konishi[2], Akira Watanabe[2] and Masahiro Sugiura
Center for Gene Research, [1]Department of Biology, [2]Research Institute for Biochemical Regulation, Nagoya University, Nagoya 464-01, Japan

Received August 23, 1991; Revised and Accepted November 8, 1991 EMBL accession nos X61113 and X61114

ABSTRACT

Two new ribonucleoproteins (RNPs) have been identified from a tobacco chloroplast lysate. These two proteins (cp29A and cp29B) are nuclear-encoded and have a less affinity to single-stranded DNA as compared with three other chloroplast RNPs (cp28, cp31 and cp33) previously isolated. DNA sequencing revealed that both contain two consensus sequence-type homologous RNA-binding domains (CS-RBDs) and a very acidic amino-terminal domain but shorter than that of cp28, cp31 and cp33. Comparison of cp29A and cp29B showed a 19 amino acid insertion in the region separating the two CS-RBDs in cp29B. This insertion results in three tandem repeats of a glycine-rich sequence of 10 amino acids, which is a novel feature in RNPs. The two proteins are encoded by different single nuclear genes and no alternatively spliced transcripts could be identified. We constructed a phylogenetic tree for the ten chloroplast CS-RBDs. These results suggest that there is a sizable RNP family in chloroplasts and the diversity was mainly generated through a series of gene duplications rather than through alternative pre-mRNA splicing. The gene for cp29B contains three introns. The first and second introns interrupt the first CS-RBD and the third intron does the second CS-RBD. The position of the first intron site is the same as that in the human hnRNP A1 protein gene.

INTRODUCTION

Chloroplasts contain their own genetic system and near 100 distinct chloroplast genes have so far been described (for review see 1, 2). Eighteen chloroplast genes from higher plants are known to contain introns and chloroplast introns can be classified into at least three groups (3–5). The intron boundary sequences of the largest group are conserved and resemble those of nuclear genes, suggesting that the splicing of many chloroplast pre-mRNAs occurs in a similar way to that for nuclear pre-mRNAs.

In the nucleus, pre-mRNAs are bound with specific proteins to form hnRNP particles. Some of the hnRNP proteins have been shown to be involved in pre-mRNA splicing (6,7). The hnRNP complex consists of six core proteins, A1, A2, B1, B2, C1 and C2 (for review see 8) and these core proteins contain one or two conserved domains of 80–100 amino acids named consensus sequence-type RNA-binding domain (CS-RBD) (9) or RNA recognition motif (10). CS-RBD includes a most conserved octamer sequence termed ribonucleoprotein consensus sequence (RNP-CS)(11,12) and a less well conserved hexamer, RNP2 (8). CS-RBD has been shown to be the minimum structure for RNA-binding activity (10,13–17). In animals, alternative pre-mRNA splicing is well known to generate the diversity of hnRNP proteins (18–22).

Three plant RNPs or RNA-binding proteins were isolated from tobacco chloroplasts (23). These three proteins were eluted from single-stranded (ss) DNA cellulose with 2 M NaCl and each contains two CS-RBDs. Since different hnRNP core proteins have different affinities to ssDNA (24), we tried to isolate RNP-like proteins with a less affinity to ssDNA. Two chloroplast proteins of around 29 kd were isolated and their genes were analyzed. Both proteins were found to be new RNPs containing two CS-RBDs and nuclear-encoded. This finding suggests the existence of a sizable RNP family in chloroplasts like in yeast and mammal nuclei. No alternatively spliced transcripts could be identified for the two new RNPs. We have constructed a phylogenetic tree for the ten chloroplast CS-RBDs. The tree supports that a diversity of chloroplast RNPs was mainly generated through the gene duplication from a common ancestor containing two CS-RBDs rather than through alternative splicing of single pre-mRNAs. The diversity of a RNP family is probably necessary for pre-mRNA processing and/or splicing in chloroplasts.

MATERIALS AND METHODS

Isolation of ssDNA-binding proteins

Chloroplast ssDNA-binding proteins (a 0.6 M NaCl fraction) were prepared and sequenced as described (23).

cDNA cloning

Leaf cDNA libraries of *N.tabacum* and *N.sylvestris* in λgt10 were constructed using a cDNA synthesis kit (Pharmacia). Oligonucleotide probes for cp29A and hybridization temperatures were:

probe A: CCRTCTTCIACRTCTTCDATTTGICCRAA, 51°C
probe B: CCRTCYTCIACRTCYTCIATYTGICCRAARTCISWIARIGTIAC, 58°C

(R= A and G, Y= C and T, S= G and C, W= A and T, D= A, G and T, I= Inosine). 108 Positive clones from a *N.tabacum* cDNA library (1×10^6 clones) screened by probe A were further verified by probe B. A *N.sylvestris* cDNA library was then screened using a subcloned 5'-fragment (341 bp) of a *N.tabacum* cDNA (λNT29AC, the longest insert of 998 bp) at a hybridization temperature of 65°C and a final washing at 65°C with 0.1×SSPE containing 0.1% SDS. About 100 positive clones were obtained from 1.3×10^6 clones.

Recombinant λDNA preparation and DNA sequence analysis were performed using a λDNA sequencing method (25). Two λgt10 primers 1233 and 1234 (23) and several other oligonucleotides were used as sequencing primers. Sequenase version 2.0 DNA sequencing kit (USB) was used.

In vitro import assay

A sequence encoding the precursor protein of cp29A was synthesized on λNS29AC template by PCR. Two PCR primers contained either *Pst*I or *Nco*I site were used. The amplified DNA was cloned into a *Nco*I/*Pst*I digested Bluescript version II SK^+ vector (Stratagene) and the insert sequence (pNS29AC) was verified by plasmid DNA sequencing.

Pea chloroplast preparation and import experiment were performed essentially as described (23,26). Ten μg of plasmid DNA was linearized with *Bam*HI, treated with proteinase K and extracted with phenol/chloroform. RNA synthesis was carried out according to the instruction manual of mCAP™ mRNA Capping kit (Stratagene). Translation was performed in 50 μl of a rabbit reticulocyte lysate system (Amersham) containing 70 μCi [^{35}S]methionine (specific activity > 1000 Ci/mmol) and 3 μl of the capped transcript (about 0.5 μg RNA) as described in the instruction manual.

Genomic cloning

The cp29A cDNA insert was excised from pNS29AC and used as probe to screen a *N.sylvestris* genomic library (in λdash) which was constructed previously (27). The insert in a clone (λNS29BG) was excised with *Xba*I and subcloned into Bluescript SK^+ vector (pNS29BG). Deletion, ssDNA preparation and sequencing was carried out according to the standard method (28) and the instruction manual of TaKaRa.

Reverse transcription and PCR cloning

A cDNA plasmid encoding the mature protein of cp29B was constructed using a cDNA Cycle™ kit (Invitrogen). Reverse transcription was carried out in a 20 μl reaction containing 20 μg of total *N. sylvestris* leaf RNA. Phenol/chloroform extraction and ethanol precipitation were then performed. The resultant single-stranded cDNA mixture was used as template for PCR using two specific primers for cp29B (containing either *Nco*I or *Xba*I site). The amplified fragment was cloned into a *Nco*I/*Xba*I-digested pTMV-3 vector (29) and the insert sequence (pNS29BC) was compared to the genomic sequence to determine the intron sites.

Computer analyses

DNA sequences were analyzed as described by Hiratsuka *et al.*(30). The evolutionary distance of each pair of aligned CS-RBD sequences was calculated according to Jukes and Cantor (31). Phylogenetic tree construction was carried out using the neighbor-joining (N-J) method (32). The region from the 15th to the 83rd residues of each CS-RBD sequence was used because the sequence of sorghum AAIPS1 is truncated. We have confirmed that the region we ignored does not essentially affect the results.

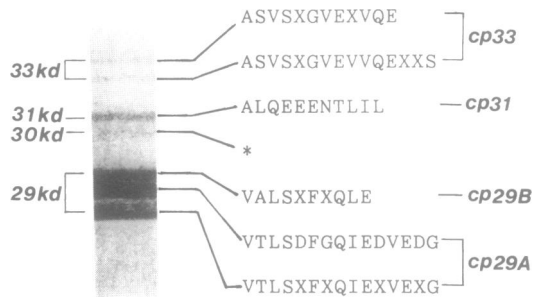

Figure 1. SDS-PAGE pattern of proteins eluted at 0.6 M NaCl from an ssDNA cellulose column. Determined N-terminal sequences are shown on right. An asterisk indicates a blocked N-terminus and X denotes unidentified residues.

RESULTS

Isolation of cp29A and cp29B proteins

When total soluble proteins from tobacco (*N. tabacum*) chloroplasts were applied to an ssDNA cellulose column, a group of 29−33 kd proteins were eluted between 0.3−0.6 M NaCl as shown in Figure 1. Based on the N-terminal amino acid sequences, two minor bands of 33 kd and a band of 31 kd correspond to cp33 and cp31, respectively (23). The N-terminus of a 30 kd protein is blocked. Among three major bands of around 29 kd, two of them have near identical N-terminal sequences and this is due probably to the amphidiploidy of *N.tabacum*. These two 29 kd proteins were designated as cp29A and the other as cp29B. We failed to find their N-terminal sequences in any frames translated from the complete sequence of tobacco chloroplast genome (33) and they must be nuclear-encoded.

cDNA sequence analysis of cp29A

cDNA clones for cp29A were isolated from a *N.sylvestris* leaf cDNA library (*N.sylvestris* is the female progenitor of *N.tabacum* and thus is a simpler system for the analysis of nuclear genes). The longest clone (λNS29AC, a 1155 bp insert) was sequenced directly by the λDNA sequencing method (25). As shown in Figure 2A, the determined N-terminal sequence of one of the cp29A bands matches the deduced sequence (underlined, D at the 7th position). We have also found that one of the cDNA clones for cp29A from *N.tabacum* contains G in the 7th position (data not shown). This polymorphism is likely to be caused by the alloploidy of *N.tabacum* (see Figure 1). The cDNA sequence revealed a reading frame of 273 amino acids. The first 58 amino acid sequence, rich in hydroxylated residues and has an overall positive charge, is likely to be a transit peptide (e.g. 34) and has been shown to be a genuine one by an *in vitro* import experiment (Figure 3). The mature protein is 215 amino acids long (mol. wt, 23,512) and contains two CS-RBDs, confirming that it is also an RNP.

Genomic sequence analysis of cp29B

A *N.sylvestris* genomic library was screened with the above cp29A cDNA as probe. Five positive clones were obtained from

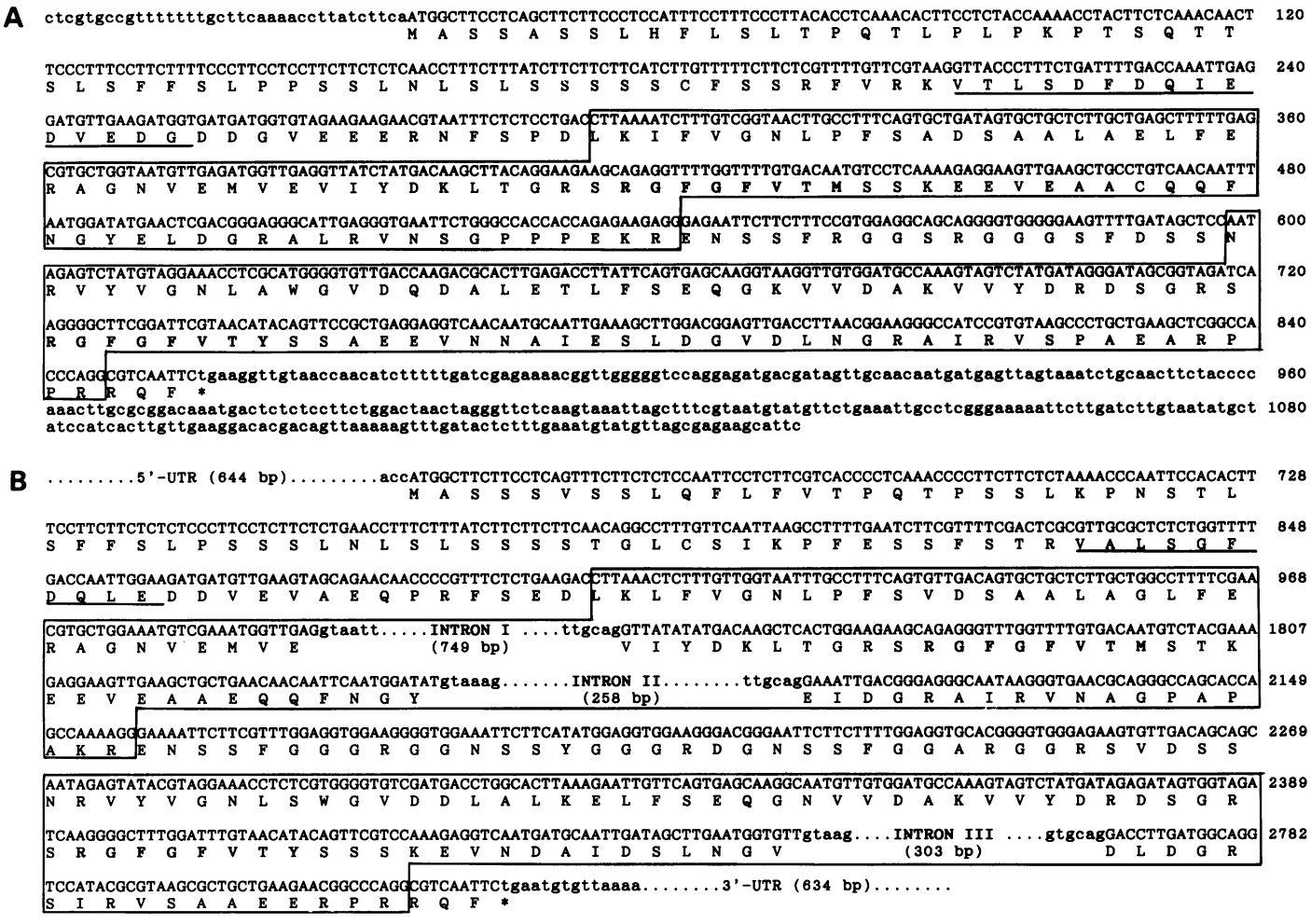

Figure 2. (A) cDNA and deduced amino acid sequences of cp29A, (B) Genomic DNA and deduced amino acid sequences of cp29B. Determined N-termini are underlined. CS-RBDs are boxed. Uppercase letters indicate coding regions and lowercase letters indicate untranslated regions (UTR) and introns. Asterisks denote stop codons.

1.4×10^6 clones after several rounds of screening. A clone (λNS29BG) containing a 4.3 kbp *Xba*I fragment was subcloned into a plasmid vector and sequenced (pNS29BG). The coding regions and intron sites were determined by reverse transcription and polymerase chain reaction (PCR) amplification using *N. sylvestris* leaf RNAs and specific primers. As shown in Figure 2B, the deduced amino acid sequence fits the determined N-terminal sequence of another protein, cp29B.

The deduced protein contains two CS-RBDs and a transit peptide-like sequence highly homologous to that of cp29A (see Figure 4), indicating that cp29B is also a chloroplast RNP. This raises an interesting possibility that the chloroplast has a sizable RNP family. The four exons constitute a reading frame of 291 amino acids. The mature protein is 229 amino acids long (mol. wt, 24,605), which is slightly larger than cp29A (see Figure 1).

The intron sites in the cp29B gene are identical as in cp31, namely CS-RBD I is split into three parts by the first two introns and CS-RBD II into two by the third intron (27). The position of the first cp29B intron site is again conserved as in the maize abscisic acid-induced glycine-rich protein (AAIP) (35) and the human hnRNP A1 (18), suggesting that a common origin of CS-RBDs in plant and animal kingdoms.

Figure 3. Import of cp29A into intact chloroplasts. **1**: [^{35}S] labeled products of *in vitro* translation; **2**: a processed cp29A mature protein found in protease-treated chloroplasts.

An extra l9 amino acids and a tandem repeat structure in cp29B

Comparison of the amino acid sequences of cp29A and cp29B revealed a l9 amino acid insertion so that a l0 amino acid sequence

```
cp29A  MASS-ASSLHFLSLTPQTLPLPKPTSQTTSLSFFSLPPSSLNLSLSSSS--SC----FSSRFVRKVTLSDFDQIED-VEDGDDGVEEERNFSPD  86
cp29B  MASSSVSSLQFLFVTPQT---PSSLKPNSTLSFFSLPSSSLNLSLSSSSTGLCSIKPFESSFSTRVALSGFDQLEDDVE-----VAEQPRFSED  86

       |<------------------- transit peptide ------------------->||<------ acidic domain ------>|

A      LKIFVGNLPFSADSAALAELFERAGNVEMVEVIYDKLTGRSRGFGFVTMSSKEEVEAACQQFNGYELDGRALRVNSGPPPEKR  170
B      LKLFVGNLPFSVDSAALAGLFERAGNVEMVEVIYDKLTGRSRGFGFVTMSTKEEVEAAEQQFNGYEIDGRAIRVNAGPAPAKR  170

       |<------------------------------------ CS-RBD I ------------------------------------>|

A      E--------------------NSSFRGGSRGGGSFDSS 187
B      ENSSFGGGRGGNSSYGGGRDGNSSF-GGARGGRSVDSS 206

A      NRVYVGNLAWGVDQDALETLFSEQGKVVDAKVVYDRDSGRSRGFGFVTYSSAEEVNNAIESLDGVDLNGRAIRVSPAEARPPRRQF 274
B      NRVYVGNLSWGVDDLALKELFSEQGNVVDAKVVYDRDSGRSRGFGFVTYSSSKEVNDAIDSLNGVDLDGRSIRVSAAEER-PRRQF 291

       |<------------------------------------ CS-RBD II ------------------------------------>|
```

Figure 4. Comparison of the amino acid sequences of cp29A and cp29B. Asterisks indicate identical residues and dots conservative substitutions. Dashes denote gaps introduced to optimize sequence alignment. RNP-CS motifs in CS-RBDs are boxed and tandem repeat units are underlined.

(consensus sequence; **NSSFGGGRGG**) is repeated three times in cp29B and once in cp29A within the spacer separating the two CS-RBDs (underlined in Figure 4). This repeat is rich in glycine and a novel feature distinct from other RNPs. It may contribute in part to the diversity of chloroplast RNP family through small peptide duplications. In addition we found a 4 amino acid deletion in the acidic amino-terminal domain in cp29B. This decreases a net negative charge from 39% in cp29A to 29% in cp29B.

To investigate whether there is any alternative pre-mRNA splicing, we carried out PCR amplification using a leaf cDNA pool as template and specific primer sets for each intron. No alternatively spliced transcripts could be found for cp29A and cp29B (data not shown). Each of cp29A and cp29B is coded for by a single-copy nuclear gene in *N. sylvestris* (Ye *et al.*, unpublished).

A phylogenetic tree of ten chloroplast CS-RBDs

We have so far identified five chloroplast RNPs, each containing two CS-RBDs. This suggests that the ancestor is a two-domain protein. Comparison of the ten chloroplast CS-RBDs shows high homology to each other within either CS-RBDs I or CS-RBDs II (see Figure 4 and Figure 5 in 23). For example, the CS-RBD I of cp29A shares 88% amino acid identity with that of cp29B while only 46% and 47% with the CS-RBD II of cp29A and cp29B, respectively. In contrast to these chloroplast RNPs, the maize AAIP (35) and sorghum AAIP-like proteins (36) contain only one CS-RBD. Their CS-RBDs show higher homologies to the CS-RBD II (especially the CS-RBD II of cp33) than to the CS-RBD I of the five chloroplast RNPs, suggesting that AAIP CS-RBDs were originated from CS-RBD II of a two-domain ancestor and CS-RBD I was lost during evolution. Another possibility is that AAIP CS-RBDs were derived from an one-domain protein containing an ancestral CS-RBD II.

We constructed a phylogenetic tree for the ten chloroplast and three AAIP CS-RBDs (Figure 5). The tree again indicates that these chloroplast CS-RBDs were likely originated from a common ancestor through gene duplications and that the CS-RBDs of AAIPs and the CS-RBD II of cp33, were of a common origin.

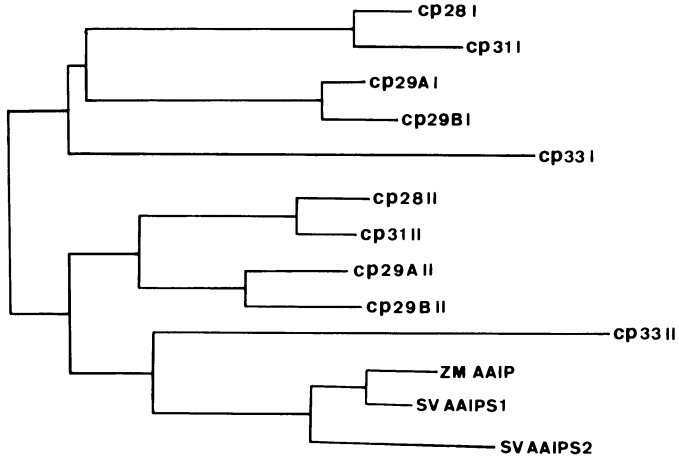

Figure 5. A phylogenetic tree of plant CS-RBDs from five tobacco chloroplast RNPs, one maize (ZM) AAIP (35) and two sorghum (SV) AAIP-like proteins S1 and S2 (36). cp28 I indicates cp28 CS-RBD I and so on. The lengths of horizontal lines represent the number of amino acid substitutions per site computed according to Jukes and Cantor (31).

DISCUSSION

We have identified two new chloroplast hnRNP-like proteins (cp29A and cp29B) which have a lower affinity to ssDNA than the three RNPs (cp28, cp31 and cp33) analyzed previously (23). This result implies the presence of a sizable RNP family in chloroplasts like in yeast and mammal nuclei. Additional lines of evidence for this are: (1) several other proteins eluted from an ssDNA column (between 0.3–2 M NaCl) remain to be analyzed because of their low amounts, (2) using a degenerate oligonucleotide probe corresponding to the consensus chloroplast RNP-CS motif we obtained many other cDNA clones which are distinct from those for the five identified RNPs. (3) several weak bands were observed in genomic Southern hybridization experiments using cp29A and cp29B cDNAs as probes even after washing at high stringent conditions.

Alternative splicing and alternative polyadenylation of pre-mRNAs play an important role in the diversity of hnRNP proteins (12,18–22,37). In contrast, we have not so far identified any alternatively spliced transcripts for chloroplast RNPs, suggesting that a different mechanism was utilized to produce the diversity. We propose that the diversity of chloroplast RNPs was mainly generated through a series of gene duplications rather than alternative pre-mRNA splicing. The phylogenetic tree shown in Figure 5 also supports this proposal.

cp29A and cp29B have a lower affinity to ssDNA as compared to cp28, cp31 and cp33 and contain a shorter acidic amino-terminal domain (24–28 residues in cp29A and B / 39–64 residues in cp28, 31 and 33) which results in a lower net negative charge. A negatively charged domain has been postulated to be involved in protein-protein interaction by neutralizing the positive charge of basic proteins when assembling spliceosomes (38) and also to influence RNA-binding (for review see 9). RNA-protein and protein-protein binding specificities have recently been shown to be determined by only a few amino acid residues in the auxiliary sequence of CS-RBD rather than by the most conserved RNP-CS motif itself (16,39). Therefore the RNA binding specificities of the five chloroplast RNPs are considered to be different as their sequences flanking CS-RBDs are different from each other. Actually, different binding affinities to ssDNA and ribonucleotide homopolymers have been shown for cp28, cp31 and cp33 (29).

Another unique feature is that the last residues of RNP-CS (consensus sequence, **RGFGFVTM**) are all methionines in the first CS-RBD of the five chloroplast RNPs. This is distinct from other RNPs so far reported, which generally contain an aromatic residue at the last position. This feature may be necessary for the common function of chloroplast RNPs.

The function of a RNP family in chloroplasts is interesting. Mammalian hnRNP particles contain 20 or more protein components. The hnRNP C proteins have been shown to be involved in pre-mRNA splicing (6,7). The hnRNP A1 and A2 proteins have been found to be differentially expressed in different developing-stages (40,41). The hnRNP A1 and B2, which are produced from a single gene by alternative pre-mRNA splicing but have different affinities to ssDNA, are expressed at different levels (20). Relevant to this, the tobacco cp28, cp31 and cp33 proteins show different light-modulated expression patterns (23). A group of yeast pre-RNA processing proteins have been shown to be involved in pre-mRNA splicing but function at different steps (for review see 42). Chloroplast pre-mRNAs are also thought to be processed and spliced by complex machineries, like hnRNP particles and spliceosomes, which probably require multiple protein components. Therefore, the existence of a sizable RNP family is probably necessary for proper and differential regulation of chloroplast gene expression at the post-transcriptional level. Recently, a spinach chloroplast RNP of 28 kd was reported to be required for chloroplast mRNA 3'end processing (43).

ACKNOWLEDGMENTS

We thank Dr.R.F.Whittier for a cDNA library from *N.tabacum* leaves, Mrs.C.Sugita for pea chloroplasts, Dr.N.Saitou and Mr. T.Imanishi for their program for computing and drawing an evolutionary tree, and Drs. M.Sugita and T.Wakasugi for valuable discussions. L.Ye and Y.Li were supported by Monbusho fellowships. This research was supported in part by a Grant-in-aid from the Ministry of Education, Science and Culture (Japan).

REFERENCES

1. Sugiura,M. (1989) *Annu. Rev. Cell Biol.*, **5**, 51–70.
2. Shimada,H. and Sugiura,M. (1991) *Nucleic Acids Res.*, **19**, 983–995.
3. Shinozaki,K., Deno,H., Sugita,M., Kuramitsu,S. and Sugiura,M. (1986) *Mol. Gen. Genet.*, **202**, 1–5.
4. Christopher,D.A. and Hallick,R.B. (1989) *Nucleic Acids Res.*, **17**, 7591–7608.
5. Sugiura,M. (1991) In Bogorad, L. and Vasil,I.K. (ed.) *Cell Culture and Somatic Cell Genetics of Plants*, Academic Press, San Diego, vol.7A, pp.125–137.
6. Choi,Y.D., Grabowski,P.J., Sharp,P.A. and Dreyfuss,G. (1986) *Science*, **231**, 1534–1539.
7. Sierakowska,H., Szer,W., Furdon,P.J. and Kole,R. (1986) *Nucleic Acids Res.*,**14**, 5241–5254.
8. Dreyfuss,G., Swanson,M.S. and Pinol-Roma,S. (1988) *Trends Biochem.Sci.*, **13**, 86–91.
9. Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.* **3**, 431–437.
10. Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell* **57**, 89–101.
11. Adam,S.A., Nakagawa,T., Swanson,M.S., Woodruff,T.K. and Dreyfuss,G. (1986) *Mol.Cell.Biol.*, **6**, 2932–2943.
12. Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol.Cell.Biol.*, **7**, 1731–1739.
13. Sachs,A.B., Davis,R.W. and Kornberg,R. (1987) *Mol.Cell.Biol.* **7**, 3268–3276.
14. Scherly,D., Boelens,W., Venrooij,W.J.van, Dathan,N.A., Hamm,J. and Mattaj,I.W. (1989) *EMBO J.* **8**, 4163–4170.
15. Surowy,C.S., van Santen,V.L., Scheib-Wixted,S.M. and Spritz,R.A. (1989) *Mol.Cell.Biol.*, **9**, 4179–4186.
16. Scherly,D., Boelens,W., Dathan,N.A., Venrooij,W.J.van and Mattaj,I.W. (1990) *Nature*, **345**, 502–506.
17. Nietfeld,W., Mentzel,H. and Pieler,T. (1990) *EMBO J.*, **9**, 3699–3705.
18. Biamonti,G., Buvoli,M., Bassi,M.T., Morandi,C., Cobianchi,F. and Riva,S. (1989) *J.Mol.Biol.*, **207**, 491–503.
19. Burd,C.G., Swanson,M.S., Gorlach,M. and Dreyfuss,G. (1989) *Proc.Natl.Acad.Sci.USA*, **86**, 9788–9792.
20. Buvoli,M., Cobianchi,F., Bestagno,M.G., Mangiarotti,A., Bassi,M.T., Biamonti,G. and Riva,S. (1990) *EMBO J.*, **9**, 1229–1235.
21. Kay,B.K., Sawhney,R.K. and Wilson,S.H. (1990) *Proc.Natl.Acad.Sci.USA*, **87**, 1367–1371.
22. Haynes,S.R., Johnson,D., Raychaudhuri,G. and Beyer,A.L. (1991) *Nucleic Acids Res.*, **19**, 25–31.
23. Li,Y. and Sugiura,M. (1990) *EMBO J.*, **9**, 3059–3066.
24. Pandolfo,P., Valentini,O., Biamonti,G., Rossi,P. and Riva,S. (1987) *Eur.J.Biochem.* **162**, 213–220.
25. Manfioletti,G. and Schneider,C. (1988) *Nucleic Acids Res.* **16**, 2873–2884.
26. Minami,E., Shinohara,K., Kuwabara,T. and Watanabe,A. (1986) *Arch. Biochem. Biophys.*, **244**, 517–527.
27. Li,Y., Ye,L., Sugita,M. and Sugiura,M. (1991) *Nucleic Acids Res.*, **19**, 2987–2991.
28. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
29. Li,Y. and Sugiura,M. (1991) *Nucleic Acids Res.*, **19**, 2893–2896.
30. Hiratsuka,J., Shimada,H., Whittier,R.F., Ishibashi,T., Sakamoto,M., Mori,M., Kondo,M., Honji,Y., Sun,C.R., Meng,B.Y., Li,Y.Q., Kanno,A., Nishizawa,Y., Hirai,A., Shinozaki,K. and Sugiura,M. (1989) *Mol.Gen.Genet.*, **217**, 185–194.
31. Jukes,T.H. and Cantor,C.R. (1969) In Munro,H.N. (ed) *Mammalian Protein Metabolism*. Academic Press, New York, pp21–132.
32. Saitou,N. and Nei,M. (1987) *Mol.Biol.Evol.*, **4**, 406–425.
33. Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwongse,J., Obokata,J., Shinozaki,K.Y., Ohto,C., Torazawa,K., Meng,B.Y., Sugita,M., Deno,H., Kamogashira,T., Yamada,K., Kusuda,J., Takaiwa,F., Kato,A., Tohdoh,N., Shimada,H. and Sugiura,M. (1986) *EMBO J.* **5**, 2043–2049.
34. Heijne,G.V., Steppuhn,J. and Herrmann,G. (1989) *Eur.J.Biochem.*, **180**, 535–545.

35. Gomez,J., Sanchez-Martinez,D., Stiefel,V., Rigau,J., Puigdomenech,P. and Pages,M. (1988) *Nature*, **334**, 262–264.
36. Cretin,C. and Puigdomenech,P. (1990) *Plant Mol.Biol.*, **15**, 783–785.
37. Buvoli,M., Biamonti,G., Tsoulfas,P., Bassi,M.T., Ghetti,A., Riva,S. and Morandi,C. (1988) *Nucleic Acids Res.*, **16**, 3751–3770.
38. Preugschat,F. and Wold,B. (1988) *Proc.Natl.Acad.Sci.USA* **85**, 9669–9673.
39. Scherly,D., Dathan,N.A., Boelens,W., Venrooij,W.J.van and Mattaj,I.W. (1990) *EMBO J.*, **9**, 3675–3681.
40. Celis,J.E., Bravo,R., Arenstorf,H.P. and LeStourgeon,W.M. (1986) *FEBS Lett.*, **194**, 101–109.
41. Dreyfuss,G. (1986) *Annu. Rev. Cell. Biol.*, **2**, 459–498.
42. Ruby,S.W. and Abelson,J. (1991) *Trends Genet.*, **7**, 79–85.
43. Schuster,G. and Gruissem,W. (1991) *EMBO J.*, **10**, 1493–1502.