A novel plant nuclear gene encoding chloroplast ribosomal protein S9 has a transit peptide related to that of rice chloroplast ribosomal protein L12

Shin-ichi Arimura[a], Shin Takusagawa[a], Shoji Hatano[a,b], Mikio Nakazono[a], Atsushi Hirai[a], Nobuhiro Tsutsumi[a,*]

[a] *Laboratory of Radiation Genetics, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*

[b] *Department of Information Science and Technology, National Agriculture Research Center, Kan-nondai 3-1-1, Tsukuba, Ibaraki 305-8666, Japan*

Received 10 February 1999; received in revised form 30 March 1999

Abstract **We have cloned a novel nuclear gene for a ribosomal protein of rice and *Arabidopsis* that is like the bacterial ribosomal protein S9. To determine the subcellular localization of the gene product, we fused the N-terminal region and green fluorescent protein and expressed it transiently in rice seedlings. Localized fluorescence was detectable only in chloroplasts, indicating that this nuclear gene encodes chloroplast ribosomal protein S9. The N-terminal region of rice ribosomal protein S9 was found to have a high sequence similarity to the transit peptide region of the rice chloroplast ribosomal protein L12, suggesting that these transit peptides have a common lineage.**

Key words: Chloroplast ribosomal protein; Transit peptide; Green fluorescent protein; Gene transfer; *Oryza sativa*

1. Introduction

Plant cells contain three distinct types of ribosomes that are found respectively in the cytosol, mitochondria and chloroplasts. Many structural similarities exist among the components of chloroplast and eubacterial ribosomes including strong sequence similarities in the ribosomal proteins [1] and 16S and 23S rRNAs [2,3], most likely reflecting their endosymbiotic origin. The chloroplasts of higher plants have about 55–60 ribosomal proteins. Of these, 21 are encoded by the chloroplast genome and the rest are assumed to be encoded by the nucleus [4,5]. This distribution between the nuclear and chloroplast genomes results from the translocation of genes from the endosymbiont to the nucleus [6]. Nuclear-encoded chloroplast proteins are synthesized as higher molecular weight precursors in the cytosol by 80S ribosomes and are translocated into chloroplasts and proteolytically processed to their mature sizes [7]. The processed N-terminal regions, termed the chloroplast transit peptides, are both necessary and sufficient to mediate precise transport to the chloroplast stroma. Many nuclear genes encoding chloroplast proteins are thought to come from the chloroplast genome, but where do their transit peptides come from? Neither the origin of chloroplast transit peptides nor the mechanism by which they are acquired is known. The difficulty in answering these questions is partly due to the fact that the amino acid sequences of chloroplast transit peptides have little homology with one another, though they have several common features [8]. Here, we describe for the first time the cDNA sequence of nuclear-encoded ribosomal protein S9 (RPS9), a constituent member of the small subunit of the chloroplast ribosome, in higher plants. In addition, we show that the transit peptides of this and another chloroplast protein have a common lineage.

2. Materials and methods

2.1. Construction and screening of a rice cDNA library

A rice cDNA library was constructed from 10 day old seedlings and screened as previously described [9]. The probe used to isolate rice *rps9* was prepared as follows. Reverse-transcribed (RT)-PCR was performed to amplify the rice cDNA fragments that encode S9 from poly(A)$^+$ RNA, isolated from 10 day old rice seedlings. Primers used for RT-PCR were designed based on the bacterial *rps9*-like sequences in rice that were found in the course of random cDNA sequencing. The amplified fragments were cloned to pBluescript SK$^-$ (Stratagene) and sequenced to confirm that they were genuine parts of this gene. The fragments amplified by RT-PCR were labelled by digoxigenin as described in the instruction manual of the DIG-System (Roche) and used as probes.

2.2. DNA sequencing and analysis

DNA sequences were determined by the dideoxynucleotide chain termination method with an automated DNA sequencer (model 373, Perkin Elmer ABD). The sequences were determined on both strands. The nucleotide sequences and the deduced amino acid sequences were aligned with the CLUSTAL W algorithm [10].

2.3. Oligonucleotides

The following oligonucleotides were used as primers for amplification by RT-PCR: a, 5′-GGTTTGCTGACAAGAGACAC-3′; b, 5′-CAGATAATTGCCTTTTGTGACC-3′ and for OsRPS9TP-GFP construction: c, 5′-ATGTCGACGATGGCGCTCTCCCTCACC-3′; d, 5′-TACCATGGTCCGCTGCGCCGCGAAG-3′. The underlined sequences show recognition sites of the restriction enzymes *Sal*I and *Nco*I.

2.4. Construction and visualization of GFP fusion proteins

The targeting sequence corresponding to the first 98 amino acids of rice *rps9* was PCR-amplified from cloned genes by using oligonucleotides that contained either a *Nco*I site or a *Sal*I site. The amplified fragments were digested with *Nco*I or *Sal*I and cloned in-frame into the *Nco*I and *Sal*I sites of the GFP expression vector S65TGFP (CaMV35Spro::S65TGFP::NOSpolyA) (a gift from Dr Y. Niwa, Shizuoka University) [11]. The sequence of the resulting plasmid, termed OsRPS9TP-GFP, was checked by sequencing. 10 μg of OsRPS9TP-GFP plasmid was precipitated onto 1.0 μm spherical gold beads (Bio-Rad). Sheaths from 10 day old rice green seedlings were cut into 3 cm long pieces and laid onto Petri plates containing solid media (0.5% GelRite agarose, pH 5.5, in water). 10 sheaths were bombarded at a time with three replicates using a PDS-1000 (Bio-Rad) according to the manufacturer's instructions. After bombard-

*Corresponding author. Fax: (81) (3) 5841 5183.
E-mail: atsutsu@hongo.ecc.u-tokyo.ac.jp

0014-5793/99/$20.00
PII: S0014-5793(99)00491-3

ment, the sheaths were placed on a benchtop for 24 h. Transformed rice sheath epidermal cells were examined with a Bio-Rad confocal laser scanning microscope (Micro-Radiance MR/AG-2: Bio-Rad). The samples were illuminated with an argon ion laser using the 488 nm band for green fluorescent protein (GFP) or a green HeNe laser using the 543 nm band for chlorophyll autofluorescence. The final image was usually obtained by averaging three scans.

3. Results

3.1. A novel gene for a bacterial-like RPS9 functions in rice chloroplasts

Genes for RPS9 have been identified in many bacterial genomes [12–15]. Similar genes have also been identified in the yeast nuclear genome (for mitochondrial RPS9 (MRPS9): [16]) and in the chloroplast genomes of lower unicellular plants [17–19]. In *Escherichia coli*, the RPS9 is cross-linked to the L5 and L18 proteins of the large subunit. It was concluded that the S9 protein is located at the interface between the small and large subunits [20]. The chloroplast genome of rice [21] contains no gene encoding chloroplast RPS9. To address the possibility that the *rps9* gene was transferred from the chloroplast genome to the nucleus and the possibility that it is transcribed in the nucleus, we sequenced the *rps9* gene as follows. cDNA clones were screened from a rice cDNA library as described in Section 2. As a result, eight clones were isolated and a sequence analysis revealed that the longest clone contained an open reading frame (ORF) of 223 amino acid residues (Fig. 1, DDBJ/EMBL/GenBank accession number AB022675). It is generally known that amino acid sequences of chloroplast ribosomal proteins are more similar to bacterial homologues than to mitochondrial homologues. The deduced amino acid sequence of the rice cDNA showed 48 and 44% identities with cyanobacterium RPS9 [14] and *E. coli* RPS9 [12], respectively (Fig. 2). Rice RPS9 showed a higher similarity to cyanobacterial RPS9 than to *E. coli* RPS9, suggesting that this gene comes from the chloroplast genome.

Long N-terminal extensions were observed in the rice nuclear gene for RPS9, suggesting that the extended region contains the targeting signal for chloroplasts or mitochondria. In fact, the first 48 amino acids of rice RPS9 are enriched for serine, threonine and arginine residues and contain no acidic residues. This biased amino acid composition is a common hallmark of both mitochondrial and chloroplast targeting sequences [8]. We analyzed the N-terminal region of rice RPS9 with the computer program PSORT [22]. PSORT predicted that this region, which we presumed to be a chloroplast transit peptide, is a mitochondrial matrix targeting sequence. To determine the subcellular localization of the rice *rps9* product in living cells, a recombinant plasmid (termed OsRPS9TP-GFP) was constructed in which the first 98 codons of the rice RPS9

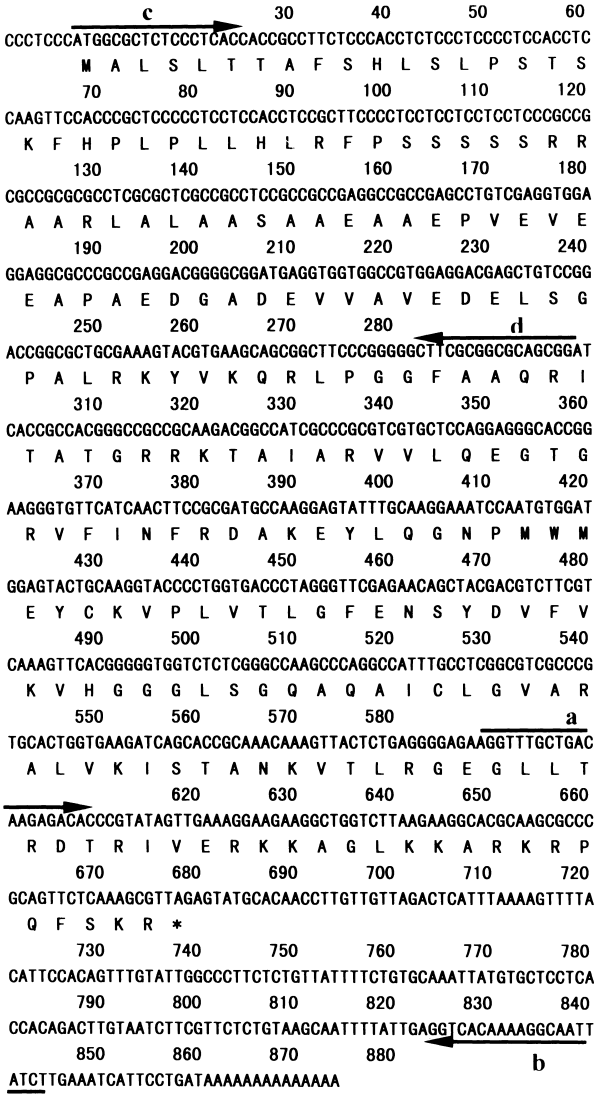

Fig. 1. Nucleotide and deduced amino acid sequences of cDNA for chloroplast RPS9 of rice (*Oryza sativa* L cv. Nipponbare). The stop codon is denoted by an asterisk. Arrows labelled a and b indicate the positions of the primers used for amplification by RT-PCR and arrows labelled c and d indicate the positions of primers used for GFP construction.

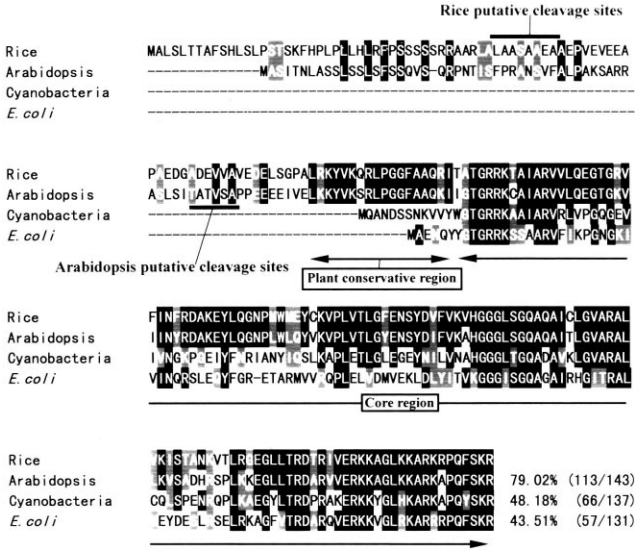

Fig. 2. Alignment and sequence comparison of RPS9 proteins or their deduced precursors from rice (*O. sativa*: this report), *Arabidopsis* (*Arabidopsis thaliana*: this report), cyanobacterium (*Synechocystis* sp. strain PCC6803 [14]) and *E. coli* [12]. Identical and similar amino acid residues are shaded by black and half tone gray, respectively. Hyphens represent gaps to maximize the alignment of the sequences. The putative cleavage sites for the chloroplast transit peptides of rice and *Arabidopsis* are indicated by bold lines [23]. Percentages at the bottom of the figure indicate the amino acid identities to rice RPS9.

cDNA, encoding the putative targeting peptide region, were fused to the encoding sequence of GFP [11]. This OsRPS9TP-GFP recombinant plasmid was introduced into rice sheath epidermal cells by particle bombardment. As shown in Fig. 3, transient expression of GFP could be only detected in a bombarded cell (the central area of the photograph) surrounded by normal cells. Confocal laser scanning microscopy revealed that GFP fluorescence was localized exclusively in oval structures of 3–8 μm in diameter (Fig. 3b), that perfectly matched the red autofluorescence of chlorophyll (Fig. 3a and c), demonstrating that OsRPS9TP-GFP fusion protein is efficiently targeted to chloroplasts in this system. Both GFP fusion protein with a mitochondrial presequence and GFP by itself showed subcellular localization patterns that are quite different from the localization pattern of OsRPS9TP-GFP (data not shown). Thus, we concluded that rice RPS9 encoded by *rps9* is a chloroplast ribosomal protein.

3.2. An Arabidopsis orthologue was found in an expressed sequence tag (EST) clone

To compare *rps9* with orthologues of other higher plant species, a homology search was performed on the EST database. One EST clone (GenBank accession number T13861) was found to have a high homology with a portion of rice *rps9*. This EST clone was from the *Arabidopsis* poly(A)$^+$ RNA and was expected to encode the full length of the encoding sequence of *Arabidopsis rps9*. We obtained this clone from the *Arabidopsis* Biological Resource Center and sequenced it completely (DDBJ/EMBL/GenBank accession number AB022676). The amino acid sequence deduced from the *Arabidopsis* cDNA showed 79, 54 and 47% identities with the deduced sequences of rice RPS9 (this study), cyanobacterium RPS9 and *E. coli* RPS9, respectively (Fig. 2). The sequences of many bacterial RPS9s have been reported [12–16]. These sequences show a high similarity in the C-terminal 124 amino acids (Fig. 2 and data not shown). In the adjacent upstream region, there is no similarity between the RPS9s of plants (rice and *Arabidopsis*) and those of bacteria, but we found a region that was highly conserved in rice and *Arabidopsis* (78.95% identity, 15/19 amino acids). The *Arabidopsis* nuclear gene for RPS9 was found to have a long N-terminal extension. This sequence, based on the properties of mitochondrial and chloroplast targeting peptides [8], is more similar to a chloroplast transit peptide than a mitochondrial transit peptide. Though the rice and *Arabidopsis* RPS9s showed a high similarity in the core region (79%), the N-terminal extensions of rice and *Arabidopsis* RPS9 (except for the plant-specific conserved region) did not show a significant sequence similarity and also differed in their lengths.

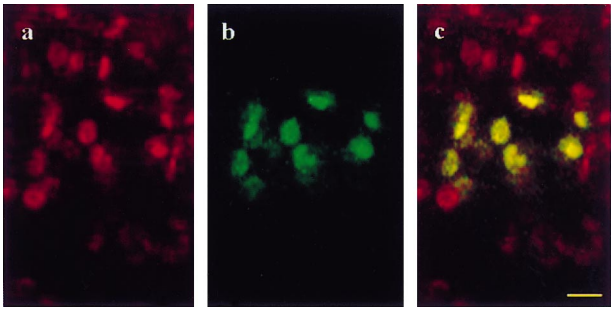

Fig. 3. Transient expression of the GFP fusion protein in rice sheath epidermal cells. A recombinant plasmid was transformed into rice by particle bombardment (PDS-1000: Bio-Rad). These images were obtained at 24 h after bombardment. Separate images of a cell expressing OsRPS9TP-GFP fusion protein were obtained with the red channel (image 'a': chlorophyll autofluorescence) and the green channel (image 'b': GFP fluorescence). Image 'c' is a superposition of 'a' and 'b'. The scale bar shows 10 μm.

Fig. 4. Nucleotide (a) and amino acid (b) sequence alignment of the N-terminal portion of the nuclear-encoded chloroplast RPS9 from rice (termed OsRPS9: this study) and RPL12-2 from rice (termed OsRPL12-2, [24]). The conserved nucleotides and amino acid residues are indicated by asterisks. Bold lines indicate the putative cleavage sites for the chloroplast transit peptides of the two proteins [8].

3.3. The chloroplast transit peptide of rice rps9 shows a high similarity to that of rice rpl12

The nucleotide sequence of the N-terminal extension of rice *rps9*, which showed little similarity to that of *Arabidopsis*, was used to search for homologues in the sequence databases. High similarity was found with the transit peptide regions of the nuclear-encoded chloroplast RPL12 from rice (*rpl12-1* and *rpl12-2*: [24]) and rye (*rpl12-1* and *rpl12-2*: [25]). The *rpl12-1* and *rpl12-2* genes of rice (and rye) are thought to be paralogous genes encoding L12 protein. Among these four nucleotide sequences, the rice *rpl12-2* transit peptide region showed the highest homology, 80.54% (120/149 nucleotides: Fig. 4a) to the rice *rps9* N-terminal region. A slightly lower homology was found between the N-terminal region of rice *rpl12-1* (sequence not shown) and that of rice *rpl12-2* (72.60%). The amino acid sequence of the N-terminal region of RPS9 was also the closest to that of the transit peptide region of rice RPL12-2 (42% identity and 72% similarity per 50 amino acids: Fig. 4b). These high similarities between the nucleotide and amino acid sequences suggest that the transit peptides of rice *rps9* and *rpl12* have a common evolutionary origin.

4. Discussion

A number of amino acid sequences for the N-terminal extensions of nuclear-encoded chloroplast proteins have been analyzed so far. A notable result is that no significant sequence homology has been detected among the sequences. However, the primary structures of chloroplast transit peptides do exhibit several common features. They are rich in serine or threonine residues, lack acidic amino acid residues and have three distinct structural domains [8]. The process of

organelle to nucleus gene transfer is thought to be ongoing [26,27]. How the chloroplast and mitochondrial copies that were integrated into the nuclear genome recruited a DNA sequence encoding a transit peptide remains an important question. Interestingly, Kadowaki et al. found that the N-terminal region of a gene that was recently transferred from the mitochondrion to the nucleus contained a mitochondrial transit peptide and that this region showed a high similarity to that of another pre-existing nuclear-encoded mitochondrial protein [28]. The sharing of the N-terminal region together with its 5′ flanking untranslated nucleotide sequence in different proteins strongly suggests an involvement of duplication/recombination for targeting signal acquisition after gene migration. On the other hand, for chloroplast transit peptides, there have been no reports indicating a common lineage between different proteins. In this study, our most notable findings are that highly conserved chloroplast transit peptides are present in different chloroplast proteins, as was found to be the case for the transit peptide of the mitochondrial protein. Our results suggest a possible mechanism for the acquisition of a transit peptide by a gene introduced from the chloroplast to the nucleus in plants. That is, the transit peptide is acquired from another protein that is already targeted to the chloroplast, possibly by a duplication/recombination event. Such a common use of one transit peptide for different proteins may accelerate the functional reconstruction of a gene that has migrated from one organelle to another.

Acknowledgements: We acknowledge Dr Y. Niwa for the generous gift of the enhanced GFP expression vector and the *Arabidopsis* Biological Resource Center (The Ohio State University) for providing the *Arabidopsis* EST clone used in this study. This work was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan and the Bio-oriented Technology Research Advancement Institution.

References

[1] Subramanian, A.R., Steinmetz, A. and Bogoard, L. (1983) Nucleic Acids Res. 11, 5277–5286.
[2] Schwartz, Z. and Kossel, H. (1980) Nature 283, 739–742.
[3] Dwards, K. and Kossel, H. (1981) Nucleic Acids Res. 9, 2853–2869.
[4] Subramanian, A.R. (1993) Trends Biol. Sci. 18, 177–181.
[5] Mache, R., Stutz, E. and Subramanian, A.R. (1991) NATO ASI series, series H, Cell Biol., 55.
[6] Smooker, P.M., Kruft, V. and Subramanian, A.R. (1990) J. Biol. Chem. 265, 16699–16703.
[7] Gantt, J.S. and Key, J.L. (1986) Mol. Gen. Genet. 202, 186–193.
[8] Von Heijne, G., Steppuhn, J. and Herrmann, R.G. (1989) Eur. J. Biochem. 180, 535–546.
[9] Tsutsumi, N., Takusagawa, S., Suzuki, H. and Hirai, A. (1996) Plant Sci. 121, 167–174.
[10] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[11] Chiu, W.L., Niwa, Y., Zeng, W., Hirano, T., Kobayashi, H. and Sheen, J. (1996) Curr. Biol. 6, 325–330.
[12] Isono, S., Thamm, S., Kitakawa, M. and Isono, K. (1985) Mol. Gen. Genet. 198, 279–282.
[13] Theisan, M. and Potter, A.A. (1992) J. Bacteriol. 174, 17–23.
[14] Kaneko, T. and Tabata, S. (1997) Plant Cell Physiol. 38, 1171–1176.
[15] Kimura, M. and Chow, C.K. (1984) Eur. J. Biochem. 139, 225–234.
[16] Koetter, P. and Entian, K.D. (1995) Curr. Genet. 28, 26–31.
[17] Reith, M.E. and Munholland, J. (1995) Plant Mol. Biol. Rep. 13, 333–335.
[18] Kowallik, K.V., Stoebe, B., Schaffran, I., Kroth-Pancic, P. and Freier, U. (1995) Plant Mol. Biol. Rep. 13, 336–342.
[19] Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A. and Stutz, E. (1993) Nucleic Acids Res. 21, 3537–3544.
[20] Wittmann, H.G. (1983) Annu. Rev. Biochem. 52, 35–65.
[21] Hiratsuka, J., Shimada, H., Whitter, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., 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.
[22] Claros, M.G. and Vincens, P. (1996) Eur. J. Biochem. 241, 779–786.
[23] Gavel, Y. and Von Heijne, G. (1990) FEBS Lett. 261, 455–458.
[24] Kusaka, M., Kurashige, M., Hirai, A. and Tsutsumi, N. (1998) Theor. Appl. Genet. 97, 110–115.
[25] Schmidt, M., Pichl, L., Lepper, M. and Feierabend, J. (1993) Biochim. Biophys. Acta 1172, 349–352.
[26] Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. and Kowallik, K.V. (1998) Nature 393, 162–165.
[27] Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. and Schuster, W. (1993) FEBS Lett. 325, 140–145.
[28] Kadowaki, K., Kubo, N., Ozawa, K. and Hirai, A. (1996) EMBO J. 15, 6652–6661.