

M. Kusaka · M. Kurashige · A. Hirai · N. Tsutsumi

Characterization of two rice genes for nuclear-encoded chloroplast ribosomal protein L12 and phylogenetic analysis of the acquisition of transit peptides and gene duplication

Received: 4 October 1997 / Accepted: 5 January 1998

Abstract We have identified two genes coding for chloroplast ribosomal protein L12 encoded in the nuclear genome of rice (*Oryza sativa*). These genes were designated *rpl12-1* and *rpl12-2* (*rpl12*, ribosomal protein L12). Northern analysis with specific probes revealed that both genes are transcribed. The expression of each gene seems to have a different regulatory machinery. It is also suggested that the expression of *rpl12-1* is controlled in an organ-specific manner. The deduced amino-acid sequences of the mature peptide parts are more conserved than those of the transit peptide parts in both monocotyledonous and dicotyledonous plants. A phylogenetic tree was constructed using the nucleotide sequences of the transit peptide region of the *rpl12*s of reported plant. The tree includes estimates of when the transit peptides were acquired, and when the genes were duplicated, in the course of evolution. According to our hypothesis, the nuclear-translocated chloroplast ribosomal protein L12 gene obtained its transit peptide after the divergence of monocots and dicots, then gene duplications occurred independently in monocots and dicots, and subsequently rice and rye branched apart.

Key words *rpl12* · *Oryza sativa* · Expressoin analysis · Phylogenetic tree

Introduction

In higher plants, approximately two-thirds of the 55–60 chloroplast ribosomal proteins are encoded in the nucleus, while the genes for the remaining third are located in the chloroplast genome (Subramanian 1993). A similar pattern appears to occur in rice. The complete nucleotide sequence of the chloroplast genome of rice (Hiratsuka et al. 1989) reveals that it contains 21 ribosomal protein genes, with the remaining 30–40 chloroplast ribosomal proteins presumably encoded in the nuclear genome. These proteins are synthesized as high-molecular-weight precursors from poly(A)⁺ RNA in the cytosol by 80S ribosomes and are translocated into chloroplasts and proteolytically processed to their mature size (Gantt and Key 1986).

We have analyzed these nuclear-encoded chloroplast ribosomal protein genes in rice to clarify their structure and to determine how they were transferred from the chloroplast to the nucleus, how the nuclear and chloroplast genes are expressed cooperatively, and how the proteins are translocated into the chloroplast.

In bacteria, L12 is the only multicopy component in the ribosome and forms a distinct structural entity (the stalk) in the large subunit (Wittmann 1983). Together with two other ribosomal proteins (L10 and L11) and a domain of the 23S rRNA, L12 forms an important functional center in the ribosome that is responsible for binding various translational factors and catalyzing GTP hydrolysis (Liljas et al. 1986; Möller and Maassen 1986; Egebjerg et al. 1990). In higher plants, nucleotide sequences of the L12 gene (*rpl12*) have been reported from spinach (Giese and Subramanian 1989), tobacco (Elhag et al. 1992), its relative, *Nicotiana sylvestris* (Li et al. 1992), rye (Schmidt et al. 1993), and *Arabidopsis thaliana* (Weglohner and Subramanian 1994). All of these genes are encoded in the nuclear genome, and have a chloroplast transit peptide in the 5' region, and a highly conserved mature peptide in the 3' region.

Communicated by K. Oono

M. Kusaka · M. Kurashige · A. Hirai · N. Tsutsumi (✉)
Laboratory of Radiation Genetics, Graduate School of Agricultural
Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku,
Tokyo 113-8657, Japan
Fax: + 81 3 3815 9335
E-mail: atsutsu@hongo.ecc.u-tokyo.ac.jp

Although all other reported chloroplast ribosomal proteins are encoded by single-copy genes, some *rpl12* genes are reported to be multi-copy (rye, at least two copies; *A. thaliana*, three copies).

Here we describe the cloning and characterization of two *rpl12* genes (*rpl12-1* and *rpl12-2*) encoded in the nuclear genome of rice. Northern analysis revealed that both genes are transcribed, and that the expression of *rpl12-1* is controlled in an organ-specific manner. A phylogenetic tree was constructed with the nucleotide sequences of the transit peptides of several L12 genes, and the evolutionary relationships are discussed.

Materials and methods

RT-PCR

Total rice RNA was isolated from 10-day-old green leaves of rice (*Oryza sativa* L. cv Nipponbare), by the guanidinium thiocyanate/CsCl method. The poly(A)⁺ RNA fraction was separated from the total RNA by double purification over OligotexTM dT30 (Takara Shuzo, Kyoto, Japan). The cDNA for RT-PCR was synthesized from poly(A)⁺ RNA by reverse transcriptase RAV-2 (Takara Shuzo). PCR was carried out using a DNA amplification system (Perkin Elmer ABD, Foster City, Calif.) with reverse-transcribed rice cDNA. The denaturation, annealing and extension times were 1 min at 94°C, 3 min at 50°C, and 2 min at 72°C, respectively. The cycle was repeated 30 times. Primers used for RT-PCR were based on L12 nucleotide sequences that were highly conserved among plant species and in the *Escherichia coli* homologue. The amplified fragments were cloned to pBluescript SK (Stratagene), and sequenced to confirm that genuine parts of the rice ribosomal protein gene sequence were amplified. The fragments were labelled by digoxigenin as described in the instruction manual of the DIG-System (Boehringer Mannheim, Germany), and used as probes.

Construction and screening of genomic DNA library

Total rice DNA was isolated from 10-day-old etiolated leaves of rice (*O. sativa*), by the modified CTAB method, and partly digested by *Sau3A*I. The genomic DNA library was constructed with the Lambda FIX vector (Stratagene), following the supplier's instructions. Screening of the positive clones was performed as described in the instruction manual of the Lambda FIX vector kit (Stratagene), except that probes were labelled and colorimetrically detected using the DIG-System (Boehringer Mannheim). Plaques that gave positive signals were further purified by two additional rounds of screening, and lambda phage DNAs of the positive clones were isolated.

DNA sequencing and analysis

DNA sequences were determined by the dideoxynucleotide chain-termination method with an automated DNA sequencer (model 373; Perkin Elmer ABD). DNA sequencing data and deduced amino-acid sequences were aligned with the CLUSTAL W algorithm (Thompson et al. 1994).

After aligning the nucleotide sequences of the transit peptide region, the percent divergence figures were calculated between all pairs of the sequence. These divergence figures were then used by the Neighbor-Joining method (Saitou and Nei 1987) to give

the phylogenetic tree. The distances are calculated without excluding positions with gaps and without correcting for multiple substitutions.

3'-RACE

The cDNA for 3'-RACE (rapid amplification of cDNA ends) was synthesized from poly(A)⁺ RNA by reverse transcriptase Superscript II (Gibco BRL). PCRs were carried out twice using a DNA amplification system (Perkin Elmer ABD, Foster City, Calif.) with reverse-transcribed rice cDNA. The denaturation, annealing and extension times were 1 min at 94°C, 2.5 min at 42°C, and 1.5 min at 72°C, respectively. The cycle was repeated 39 times. (c) and (f) primers were used in the 1st PCR, and (d) and (f) or (e) and (f) primers were used in the 2nd PCR (see primers and text). The primers used in this study are indicated in Fig. 2.

Primers

The following oligonucleotides were used as primers for amplification by RT-PCR: (a) 5'-GAAGACGGAGTTTCGACGT-3', (b) 5'-CTCGAGCTGCTTCTTGGC-3'. The following oligonucleotides were used as primers for 3'-RACE: (c) 5'-CAAGAAGCAGCTCGAGGAGG-3'; (d) 5'-GACCTCCCTGACGAACACAC-3'; (e) 5'-CTCACCTAATCTTGTGTAC-3'; (f) 5'-CTCTAGAAGCTTTTTTTTTTTTTTTTTTTT-3'.

Results

Identification of rice *rpl12* genes

A partial sequence of the rice *rpl12* was amplified by RT-PCR using primers designed from highly conserved regions of spinach and rye [primers (a) and (b), see Materials and methods]. The RT-PCR fragment, which had the desired length, was cloned into plasmid vectors and its identity was confirmed by sequencing. Then, the fragment was labelled with digoxigenin and used as a probe. A rice genomic library was screened and ten positive clones were isolated. They were digested with several restriction enzymes, and two distinct restriction maps were obtained (Fig. 1). These results indicated that there were at least two *rpl12* genes (designated as *rpl12-1* and *rpl12-2*) encoding the chloroplast ribosomal protein L12 in the rice nuclear genome.

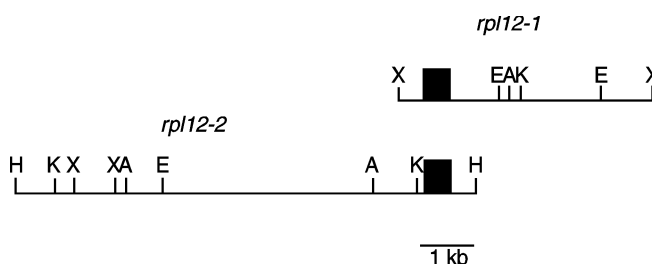


Fig. 1 Restriction-enzyme map of nuclear-encoded chloroplast ribosomal protein genes *rpl12-1* and *rpl12-2* of rice. Filled boxes indicate open reading frames. Restriction recognition sites are indicated for *Apal*(A), *EcoRI*(E), *HindIII*(H), *KpnI*(K) and *XbaI*(X)

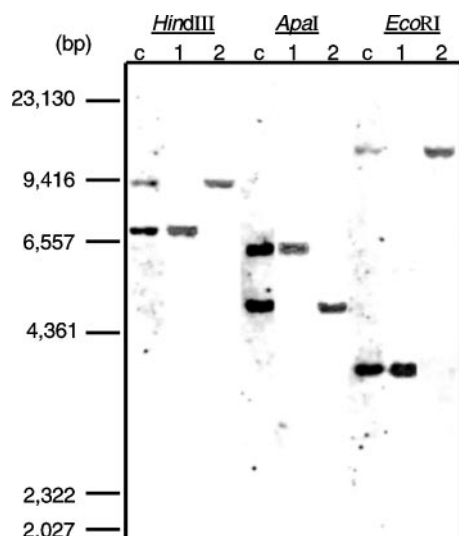


Fig. 3 Southern blots of electrophoretic separations on 0.6% agarose gels of fragments obtained after digestion of total DNA from rice green leaves with *Hind*III, *Apa*I, *Eco*RI. The genomic DNA fragments were hybridized with HRP-labelled probe. The probes were a conserved region probe (c), an *rpl12-1*-specific probe (1) and an *rpl12-2*-specific probe (2)

and we determined their poly(A) sites (Fig. 2). The products of 3'-RACE were used as *rpl12-1*- or *rpl12-2*-specific probes in the following experiments. Each probe produced the same signal intensity when hybridized to equal amounts of cloned DNAs (data not shown).

Figure 4a shows the accumulation of transcripts of *rpl12* in 10-day-old rice green leaves. This indicates that both *rpl12-1* and *rpl12-2* were transcribed, but that the amounts transcribed were different, suggesting that the expression of *rpl12-1* and *rpl12-2* is regulated differently.

Figure 4b shows the accumulation of transcripts of *rpl12-1* in leaf and root of 5-day-old rice under light and dark conditions. Using the *rpl12-2*-specific probe, no signal was detected in 5-day-old rice. This figure shows that *rpl12-1* was expressed in leaves regardless of the presence of light, but not in roots. These data suggest that *rpl12-1* was expressed in an organ-specific manner.

Comparison of L12 amino-acid sequences from other plants

The amino-acid sequences of rice *rpl12-1* and *rpl12-2* were deduced from their nucleotide sequences (Fig. 2). A comparison with amino-acid sequences from rye and other plants suggests that the putative cleavage sites for the transit peptide of the L12 precursors from rice were located between residues 46 and 47 for *rpl12-1*, and between residues 50 and 51 for *rpl12-2*. The sequence

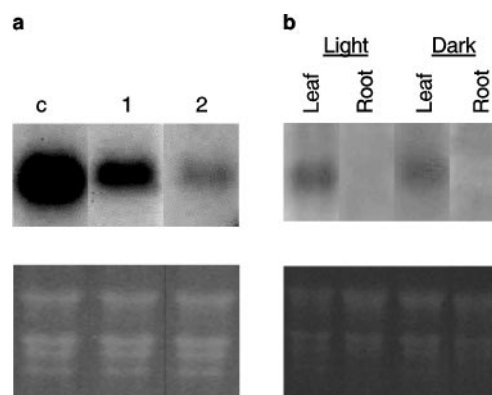


Fig. 4a, b Northern-blot analysis of the nuclear-encoded chloroplast ribosomal protein L12 genes. Lower pictures show Ethidium bromide-stained RNA. **a** Equal amounts of total RNA from 10-day-old rice leaves were separated on a 1% denaturing agarose gel and hybridized with digoxigenin-labelled *rpl12* probes. Probes (c, 1, 2) are described in the legend of Fig. 3. **b** Equal amounts of total RNA from 5-day-old rice leaves raised in light (lane 1) or dark (lane 3) and roots raised in light (lane 2) or dark (lane 4) were separated on a 1% denaturing agarose gel and hybridized with a digoxigenin-labelled *rpl12-1*-specific probe

similarities of the putative transit peptides are lower than those of the mature peptides (Fig. 5), but the amino-acid compositions of the transit peptides have characteristic features of chloroplast transit peptides, such as high contents of threonine, serine and proline and an absence of acidic residues, as well as absences of tyrosine and tryptophane (Schmidt and Miskind 1986; Keegstra et al. 1989; von Heijne et al. 1989).

In the mature protein region, sequence similarities are very high, especially in the conserved surface of the crystal structure of the *E. coli* L12 carboxyl-terminal fragment (Fig. 5). N-terminal extension and C-terminal extension, which are frequently found in the nuclear-encoded chloroplast ribosomal proteins (Phua et al. 1989; Mache 1990; Johnson and Subramanian 1991; Tsutsumi et al. 1996), are not shown in rice *rpl12*. These features suggest that the characteristic structures of the L12 gene are highly conserved in various plants, because L12 is a part of GTPase center protein which has one of the most important functions in the chloroplast ribosome.

Phylogenetic analysis of transit peptides

As mentioned above, the homologies of the transit peptide sequences of L12 proteins are lower than those of the mature protein sequences in plants. In general, it has been thought that genes acquire transit peptides after their transfer from the chloroplast to the nucleus (Subramanian 1993). In order to reveal inter- and intra-specific relationships between transit sequences of rice

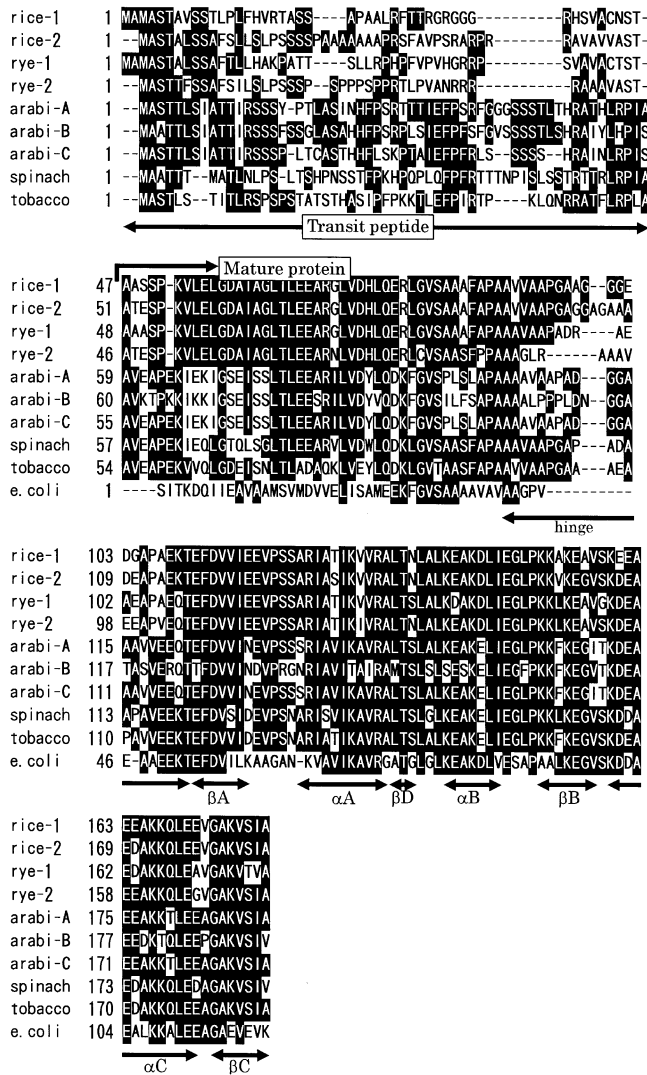


Fig. 5 Alignment and sequence comparison of L12 proteins or their precursors from rice (this report), rye (Schmidt et al. 1993), *Arabidopsis* (Weglohner and Subramanian 1994), spinach (Giese and Subramanian 1989), tobacco (Elhag et al. 1992) and *E. coli* (Post et al. 1979). Identical amino-acid residues are shaded. Dashes mark gaps that optimize sequence alignment. The bent arrow indicates the putative first amino acid of the mature protein. Parts of the hinge region, α -helices, and β -strands in the crystal structure of the *E. coli* L12 carboxyl-terminal fragment (Liljas et al. 1986) and transit peptide are indicated under the sequence

rpl12 and those from other plants, we constructed a distance-based, neighbor-joining tree with the nucleotide sequences of the transit peptide region of *rpl12* genes (Fig. 6). The transit peptides of L24s were forced to be incorporated in this phylogenetic tree, in order to compare their divergence from the transit peptides of L12.

The transit peptides were found to form three distinct clusters: a cluster of monocot L12s, a cluster of dicot L12s, and a cluster of L24s. The distance between the L12s of monocots and dicots are as large as the distances between the L24s and either of the L12

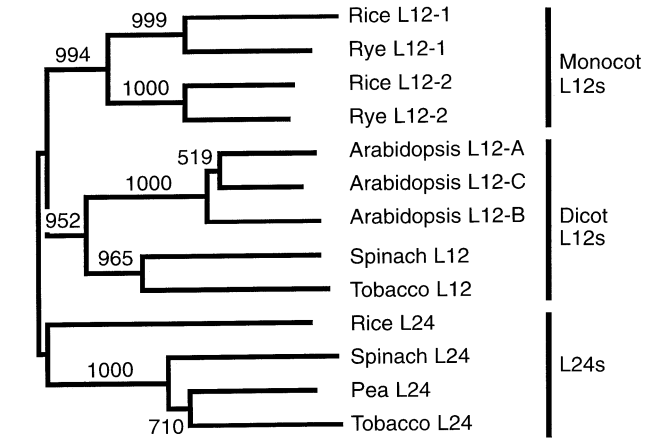


Fig. 6 Phylogenetic tree descended from the bootstrap analysis of nucleotide sequences of the transit peptides of chloroplast L12 genes. This tree was constructed by the neighbor-joining method (Saitou and Nei 1987). Horizontal branches are drawn proportional to genetic distances (the ratio of base differences). Numbers represent the bootstrap value per 1000 trials. The L24 cluster was used as an outgroup to bend the branch

clusters. The transit peptides of the monocotyledonous L12 proteins are highly divergent from those of the dicotyledonous L12 proteins.

On the other hand, the tree shows a one-to-one correspondence between rice *rpl12-1* and rye *rpl12-1*, and between rice *rpl12-2* and rye *rpl12-2*. This indicates that rice *rpl12-1* is phylogenetically closer to rye *rpl12-1* than to rice *rpl12-2*, and that rice *rpl12-2* is closer to rye *rpl12-2* than to rice *rpl12-1*.

Discussion

The preceding results demonstrated the existence of two nuclear-encoded chloroplast ribosomal protein genes (*rpl12-1* and *rpl12-2*) in rice. In rice and rye, the deduced amino-acid sequence of *rpl12-1* has two neighboring potential initiation codons. A similar configuration of start codons in spinach has been suggested to enhance the efficiency of translation (Giese and Subramanian 1990).

As in the rye L12 genes, the frequencies of G or C in the third codon position of rice *rpl12-1* (96.1%) and *rpl12-2* (96.2%) are high relative to the frequencies in L12 sequences from dicots. In addition, the numbers of codons used in *rpl12-1* (33) and *rpl12-2* (32) are small. It has been suggested that these codon biases occur in many nuclear genes of monocots, particularly among nuclear-encoded chloroplast protein genes (Brinkmann et al. 1987; Campbell and Gowri 1990).

Whereas the mature peptide regions of the L12 genes had high homologies among several plants, the transit peptide regions had low homologies. Assuming that

each of the nuclear-encoded chloroplast ribosomal proteins acquired a transit sequence with a different origin, the following conclusions can be drawn.

Because the genetic distances between the L12s of monocots and dicots are as large as those between the L12s and L24s, and L12 genes are thought to have secured their respective transit peptides after the divergence between monocots and dicots.

rpl12 is a multicopy gene in rice, rye and *Arabidopsis*, but corresponding relationships of their transit peptides exist only between rice and rye: rice *rpl12-1* is more similar to rye *rpl12-1* than to rice *rpl12-2*, and rice *rpl12-2* is more similar to rye *rpl12-2*. The distances among the L12s of rice and rye are close enough to regard them as having a common origin. These features suggest that the gene duplication event occurred after acquisition of the transit peptides independently in monocots and *Arabidopsis*, and before rice and rye branched apart.

Based on these results, we can speculate on the order of evolutionary events as follows: after the divergence of monocots and dicots, (1) nuclear-translocated chloroplast ribosomal protein L12 genes in monocots and dicots independently obtained transit peptides, (2) gene duplications occurred in monocots and dicots independently, and (3) rice and rye diverged.

rpl12s are the only known multicopy genes among the reported nuclear-encoded chloroplast ribosomal protein genes. We have suggested that the gene-duplication events of *rpl12* occurred in monocots and dicots independently. If this is true, a possible explanation is that a relatively large amount of L12 protein is needed, as it is the only multicopy protein (four copies) in the chloroplast ribosome.

Acknowledgements The authors express their appreciation to Dr. M. Nakazono of The University of Tokyo for helpful discussions. This research was supported in part by grants-in-aid from the Ministry of Education, Science, Sports and Culture, and the Ministry of Agriculture, Forestry and Fisheries, of Japan.

References

- Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R (1987) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. *J Mol Evol* 26: 320–328
- Campbell WH, Gowri G (1990) Codon usage in higher plants, green algae, and cyanobacteria. *Plant Physiol* 92: 1–11
- Egebjerg J, Douthwaite SR, Liljas A, Garrett RA (1990) Characterization of the binding sites of protein L11 and the L10. (L12)₄ pentameric complex in the GTPase domain of 23S ribosomal RNA from *Escherichia coli*. *J Mol Biol* 213: 275–288
- Elhag GA, Thomas FJ, McCreery TP, Bourque DP (1992) Nuclear-encoded chloroplast ribosomal protein L12 of *Nicotiana tabacum*: characterization of mature protein and isolation and sequence analysis of cDNA clones encoding its cytoplasmic precursor. *Nucleic Acids Res* 20: 689–697
- Gantt JS, Key JL (1986) Isolation of nuclear-encoded plastid ribosomal protein cDNAs. *Mol Gen Genet* 202: 186–193
- Giese K, Subramanian AR (1989) Chloroplast ribosomal protein L12 is encoded in the nucleus: construction and identification of its cDNA clones and nucleotide sequence including the transit peptide. *Biochemistry* 28: 3525–3529
- Giese K, Subramanian AR (1990) Enhanced translational utilization of chloroplast ribosomal protein mRNAs from two AUG codons shown by site-directed mutation. *Biochemistry* 29: 10562–10566
- Heijne G von, Steppuhn J, Herrmann RG (1989) Domain structure of mitochondrial- and chloroplast-targeting peptides. *Eur J Biochem* 180: 535–545
- Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice *Oryza sativa* chloroplast genome: Intermolecular recombination between distinct transfer RNA genes accounts for the major plastic DNA inversion during the evolution of the cereals. *Mol Genet* 217: 185–194
- Johnson CH, Subramanian AR (1991) Chloroplast ribosomal protein L15, like L1, L13, and L21, is significantly larger than its *E. coli* homologue. *FEBS Lett* 282: 268–272
- Keegstra K, Olsen LJ, Theg SM (1989) Chloroplastic precursors and their transport across the envelope membranes. *Annu Rev Plant Physiol Plant Mol Biol* 40: 471–501
- Li, Y, Itadani H, Sugita M, Sugiura M (1992) cDNA cloning and sequencing of tobacco chloroplast ribosomal protein L12. *FEBS Lett* 300: 199–202
- Liljas A, Kirsebom LA, Leijonmarck M (1986) Structural studies of the factor-binding domain. In: Hardesty B, Kramer G (eds) Structure, function, and genetics of ribosomes. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo, pp 379–390
- Mache R (1990) Chloroplast ribosomal proteins and their genes. *Plant Sci* 72: 1–12
- Möller W, Maassen JA (1986) On the structure, function, and dynamics of L7/L12 from *Escherichia coli* ribosomes. In: Hardesty B, Kramer G (eds) Structure, function, and genetics of ribosomes. Springer-Verlag, New York Berlin Heidelberg London Paris Tokyo, pp 309–325
- Phua SH, Srinivasa BR, Subramanian AR (1989) Chloroplast ribosomal protein L13 is encoded in the nucleus and is considerably larger than its bacterial homologue. *J Biol Chem* 264: 1968–1971
- Post LE, Strycharz GD, Nomura M, Lewis H, Dennis PP (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit beta in *Escherichia coli*. *Proc Natl Acad Sci USA* 76: 1697–1701
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425
- Schmidt GW, Mishkind ML (1986) The transport of proteins into chloroplasts. *Annu Rev Biochem* 55: 879–912
- Schmidt M, Pichl L, Leeper M, Feierabend J (1993) Identification of the nuclear-encoded chloroplast ribosomal protein L12 of the monocotyledonous plant *Secale cereale* and sequencing of two different cDNAs with strong bias. *Biochim Biophys Acta* 1172: 349–352
- Subramanian AR (1993) Molecular genetics of chloroplast ribosomal proteins. *Trends Biochem Sci* 18: 177–181
- Thompson, JD, Higgins, DG, Gibson, TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680
- Tsutsumi N, Takusagawa S, Suzuki H, Hirai A (1996) Molecular cloning and nucleotide sequencing of nuclear genes coding for the chloroplast ribosomal proteins L13, L24, L28 of rice (*Oryza sativa* L.). *Plant Sci* 121: 167–174
- Weglohner W, Subramanian AR (1994) Multicopy GTPase center protein L12 of the *Arabidopsis* chloroplast ribosome is encoded by a clustered nuclear gene family with the expressed members closely linked to tRNA^{Pro} genes. *J Biol Chem* 269: 7330–7336
- Wittmann HG (1983) Architecture of prokaryotic ribosomes. *Annu Rev Biochem* 52: 35–65