Rapid publication

A Proposal for Identifying the Short ID Sequence Which Addresses the Plastid Subtype of Higher Plants

Ikuo Nakamura¹⁾, Nanako Kameya¹⁾, Yusuke Kato²⁾, Shin-ichi Yamanaka²⁾, Hiroshi Jomori³⁾ and Yoichiro Sato²⁾¹⁾ Iwate Biotechnology Research Center, 22-174-4 Narita, Kitakami, Iwate 024, Japan²⁾ Faculty of Agriculture, Shizuoka University, 836 Ooya, Shizuoka 422, Japan³⁾ Faculty of Agriculture, Iwate University, 3-18-6 Ueda, Morioka, Iwate 020, Japan

Summary

Recent complete sequence data of the cyanobacteria genome and the plastid genome of algae and land plants clearly show that plastids originated from the endosymbiotic integration of a photosynthetic prokaryote into eukaryotic host cells. One indication of this event is that the organization of ribosomal protein genes has been conserved from bacteria to plastids of land plants. We are interested in the *rpl16* and *rpl14* genes, which are interrupted by the *rpl29* and *rps17* genes in both the cyanobacteria genome and the plastid genome of red alga, but are adjacent in liverwort to higher plants. This suggests that a linker between *rpl16* and *rpl14* was made by deletion of the *rpl29* and *rps17* genes at the appearance of land plants, and therefore the *rpl16-rpl14* linkers of land plants have diverged from a common ancestral sequence. The *rpl16* and *rpl14* genes are highly conserved (more than 84 % homology) among rice, spinach and tobacco, whereas their linkers show low homology (28 %). The plastid *rpl16-rpl14* linker sequences of higher plants, such as barnyard grass, gentian and oak tree could be amplified and sequenced using a pair of common primers, even if their sequences are unknown. We propose short DNA sequence in the 3' noncoding region from the stop codon of the *rpl16* gene as plastid subtype ID sequence (PS-ID). PS-IDs of ca. 50 bp are thought to be long enough to address plastid subtypes of higher plants and the establishment of a PS-ID database could contribute to plant phylogeny.

Key Words: chloroplast, plastid subtype ID, ribosomal protein genes, higher plant, phylogeny.

Introduction

The complete DNA sequences of the plastid genome have already been determined in some land plants such as tobacco (Shinozaki *et al.* 1986), black pine (Wakasugi *et al.* 1994) and liverwort (Ohyama *et al.* 1986). Recently, the plastid genome of the three algae *Porphyra purpurea* (Reith and Munholland 1995), *Odontella sinensis* (Kowallik *et al.* 1995) and *Cyanophora paradoxa* (Stirewalt *et al.* 1995), and the nuclear genome of the cyanobacterium *Synechocystis* sp. (Kaneko *et al.* 1996) have been fully sequenced. Comparison of these DNA

sequences clearly shows that plastids originated from the endosymbiotic integration of a photosynthetic prokaryote into eukaryotic host cells (Palmer 1993).

Plastid DNA is considered suitable for analyzing the phylogenetic relationships among plant species, since it is more stable than mitochondrial and nuclear DNA, in contrast to the nuclear genome, the plastid genome shows simple inheritance in that it is usually transmitted only from maternal (or paternal) cells and the protein coding sequences change more slowly (Clegg 1993). Therefore, if a simple common method for identifying plastid subtypes of higher plants could be developed, it could provide a lot of information about the evolutionary lineages of the host plants and a different standpoint than those available from the data of the nuclear genome analyzed by various molecular techniques such as RFLP and DNA fingerprinting.

In this paper, we propose a simple and sensitive procedure for determining the PS-ID sequence which addresses the plastid subtype of higher plants. PS-ID is about 50 base long from the stop codon of *rpl16* to an upstream noncoding sequence of *rpl14* on the plastid, and can be amplified and sequenced using common primers.

Materials and Methods

Total genomic DNA was isolated by the procedure of Dellaporta *et al.* (1983) from leaf tissues (100 mg) of *japonica* (ten lines) and *indica* rice (two lines), and wild barnyard grass (two lines), gentian (two lines) and oak tree (one line). PCR amplification of the plastid DNA fragment containing *rpl16* and *rpl14* was performed using one pair of common primers (A primer: 5'-AAAGATC-TAGATTTTCGTAAACAACATAGAGGAAGAA-3' and B primer: 5'-ATCTGCAGCATTTAAAAGGGTCTGAGGTTGAATCAT-3'). The A and B primers were 36-mers designed based on the conserved sequences (84,051-84,016 and 83,509-83,544) of *rpl16* and *rpl14* in tobacco chloroplast genome (Shinozaki *et al.* 1986), respectively. The PCR reactions were performed in a volume of 25 μ l containing 1 $\times$ Taq DNA polymerase buffer (Takara Co.), 0.2 mM dNTPs, 1.0 μ M of primers, 1 U of Taq DNA polymerase and 10-50 ng of total genomic DNA as a template. Amplification conditions were 35 cycles of 1 min at 94 $^{\circ}$ C, 1 min at 52 $^{\circ}$ C, and 2 min at 72 $^{\circ}$ C using a DNA Engine (PT-200, MJ Research). Amplified DNA fragments were purified by a PCR

Received April 18, 1997. Accepted October 6, 1997.

purification kit (Promega Co.) and the nucleotide sequence of the amplified fragments were directly sequenced from both strands using the A or B primer with a PRISM dye terminator cycle sequence kit using a 373A sequencer (Applied Biosystems). The nucleotide sequences around the *rpl16-rpl14* linker of the plastid genome from various species of plants were cited from the DDBJ databank under the accession numbers listed in Table 1, and the sequences of cyanobacteria were also cited from the World Wide Web (cyanobase@kazusa). Homology analysis was performed using Genetyx software (SDC) or the blast e-mail server of the NCBI.

Results and Discussions

The sequence data of the *rbcL* gene, which encodes the large subunit of ribulose-bisphosphate carboxylase, has been determined for over 750 species. A comparative study of *rbcL* data makes it possible to determine precise relationships between the plastid subtypes of related plant species, but does not give insight into the relationships between distantly related species (Clegg 1993, Clegg *et al.* 1994). Certain problems exist if functional DNA sequences encoding proteins or RNAs are compared: since mutations preferentially occur at sites that do not affect their biological activities, two or more mutations at the same site could lead to a misunderstanding of the relationships between plastid subtypes of distantly related plant species (Clegg 1993). Thus, while it is better to compare non-functional sequences to study phylogenetic relationships, it is very difficult to find evidence for the homologous origin of non-functional sequences.

Comparative studies of noncoding sequences downstream of the *rbcL* gene have demonstrated accurate re-

lationships between related species, such as among wheat relatives (Ogihara *et al.* 1992) and in the grass family (Morton and Clegg 1993), and many structural changes due to inversion and insertion/deletion have been found in this region. In addition, the *rbcL* gene has been exported to mitochondrial DNAs in many higher plants and lost from plastids of the non-photosynthetic plant *Epifagus virginiana* (Wolfe *et al.* 1992). Thus, we are interested in two genes, *rpl16* and *rpl14*, which encode large subunits of ribosomal protein, since these two ribosomal protein genes are significantly conserved among cyanobacteria and red algae, in which they are interrupted by the *rpl29* and *rps17* genes. However, they are adjacent in all of the characterized plastids of land plants, from liverwort to higher plants (Fig. 1). This suggests that a linker between *rpl16* and *rpl14* was made by deletion of the *rpl29* and *rps17* genes at the appearance of land plants. Therefore, the *rpl16-rpl14* linker sequences of plastid DNA in all land plants may have been derived from the same ancestral sequence. The plastid sequences of the *rpl16-rpl14* region were compared among rice (Hiratsuka *et al.* 1989), spinach (Zhou *et al.* 1989) and tobacco (Shinozaki *et al.* 1986). The coding sequences of the *rpl16* and *rpl14* genes were highly conserved (85 % and 84 % homology), whereas the linker sequences showed only 28 % homology among the three plant species (Fig. 2). Therefore, a pair of common primers could be designed in the conserved region of *rpl16* and *rpl14* to amplify the linker sequence of plastids even if the sequence itself is unknown. In fact, DNA fragments (*ca.* 550 bp) containing *rpl16-rpl14* linker sequences of *indica* and *japonica* rice, the barnyard grass, gentian and oak tree could be amplified and sequenced using the common primers (Table 1).

Fifty-base-long segments of the 3'-noncoding se-

Fig. 1. Organization of ribosomal protein genes around *rpl16-rpl14* region in the known plastid genomes of land plants and alga, and nuclear genome of cyanobacteria. The *rpl14** of *E. virginiana* is a pseudogene.

Fig. 2. Structure of *rpl16* and *rpl14* and their linker in the plastids among tobacco, rice and spinach. The *rpl16* gene contains a large intron. As shown in boxes, the regions of sequences in *rpl16* and *rpl14* were conserved at 85 % and 84 %, respectively, whereas the linker has low homology of 28 % among three plant species. The PS-ID sequences could be amplified and sequenced from *rpl16*-*rpl14* linker region by using the common (A and B) primers shown in the Figure. Stippled boxes indicate exons of *rpl16* and the numbers indicate the positions from start codon of *rpl16* gene.

Table 1. List of plastid subtype ID (PS-ID) sequences in plants.

Plant Name	Species	PS-ID sequence	LL	Accession
Rice <i>japonica</i>	<i>Oryza sativa</i>	5' - TA ACCCCC AAAAATA GTAGTATTGAAAATAAAAAACC AGGTTCTTCTTT-3'	109	X15901
Rice <i>indica</i>	<i>Oryza sativa</i>	5' - TA ACCCCCCAAAAAAGTAGTATTGAAAATAAAAAACC AGGTTCTTCT-3'	112	this study
Barnyard grass	<i>Echinochloa frumentacea</i>	5' - TAG AACCCC AAAA GGAGTATTGAAGATAAAAAACCCCTGGTTTCTCTTT-3'	109	this study
Maize	<i>Zea mays</i>	5' - TAG AACCCC AAAA GGAATATTGAAGATAAAAAACCCCTGGTTTCTCTTT-3'	109	X86563
Common bean	<i>Phaseolus vulgaris</i>	5' - TA ATAAATGAGAACCAAGGAAATAGGCTTTTGAATGAAAAACAAAAAG-3'	119	X68048
Cowpea	<i>Vigna unguiculata</i>	5' - TA ATAAAGAGAACCAAGGAAATAGGCTTTTGAATGAAAAACAGTTGCG-3'	124	M80799
Spinach	<i>Spinach oleracea</i>	5' - TA GAATGTGCAACTAAAGGAAATGGATCTTTTGGATAAAAA CAAATAGA-3'	132	X03834
Oenothera	<i>Oenothera ammophila</i>	5' - TAG GAACTGAGAATCAAGGAAAGAG TCTTAGGGATAAAAAACAGTTGCG-3'	144	M60179
Tobacco	<i>Nicotiana tabacum</i>	5' - TA ATGTAATAGAAAAAATGCGGAAGGGCTTAGATATGAACGAAAC-3'	124	Z00044
Gentian	<i>Gentiana scabra</i>	5' - TA AAAGTATAAAAAGAACAATAACAAAAACGGTTTTTCATTATGAAAAA-3'	123	this study
Oak tree	<i>Quercus sp.</i>	5' - TA ACAAAGAAAAGAAATCTTTTGAATGAAAAAACAATAATATAGG-3'	129	this study
Mistletoe	<i>Epifagus virginiana</i>	5' - TA AAAAACGTATAACCAACAGAAATATTCTTGAACGAAACAAAARRGA-3'	99	M81884
Black pine	<i>Pinus thunbergii</i>	5' - TA ACCCAACCATCCATACAGAAATCAAGAGRRRRCTGGTGAAGAAG-3'	120	D17510
Liverwort	<i>Marchantia polymorpha</i>	5' - TA AAAAAATTACTAATTAGTAAATATATAAATTTTAAATATTAAATT-3'	97	X04465

The PS-ID sequences are 50 bp of 3' noncoding sequence from stop codon (bold letter) of *rpl16* gene of each plastid DNA. The DNA sequences were cited from each accession number of DDBJ, rice (X15901; Hiratsuka *et al.* 1989), maize (X86563; Maier *et al.* 1995), common bean (X68048; Carelse *et al.* 1992), cowpea (M80799; Arief *et al.* 1994), spinach (X03834; Zhou *et al.* 1989), oenothera (M60179; Wolfson *et al.* 1991), tobacco (Z00044; Shinozaki *et al.* 1986), black pine (D17510; Wakasugi *et al.* 1994), mistletoe (M81884; Wolfe *et al.* 1992), liverwort (Ohya *et al.* 1986), LL shows length of linker sequence between *rpl16* and *rpl14*. Underlines indicate mismatches between barnyard grass and corn, between common bean and cowpea and between spinach and oenothera. Italics between *japonica* and *indica* cultivars show slippages at C and A tracts in rice.

quences from the stop codon of *rpl16* were taken from the DNA database as candidates for the ID sequence to address plastid subtype (PS-ID), even though the *rpl16*-*rpl14* linker sequences varied from 97 to 144 bp among land plant species (Table 1). Three kinds of variation [substitution, short deletion (or insertion) and slippage] were found among the PS-IDs. Interestingly, the PS-ID

of barnyard grass, a companion weed in a rice field, is more closely related to that of the maize (Maier *et al.* 1995) than to that of rice. Although *japonica* and *indica* cultivars were considered to belong to the same species, two slippages in the A and C tracts were found between their PS-IDs. Three, four and fourteen independent point substitutions in the PS-IDs were found between

maize and barnyard grass, between *Phaseolus vulgaris* (Carelse *et al.* 1992) and *Vigna unguiculata*, and between *Spinach oleracea* and *Oenothera ammophila* (Wolfe *et al.* 1992), respectively (Table 1). The 50 bp PS-IDs proposed in this paper, representing on the *rpl16-rpl14* linker, were thought to be long enough to address plastid subtypes of higher plants, and the *rpl16-rpl14* linker might be a non-functional sequence that evolved neutrally and randomly. A database consisting of short DNA sequences would be very useful for managing and analyzing PS-IDs of a large number of plant species. Once a PS-ID database is established, it could contribute, through the identification of plastid subtypes and host plant species from small pieces of plant tissues, excavated remains and materials, to areas of plant science such as phylogeny, paleobiology, and plant breeding.

Acknowledgements

This research was supported in part by grants for the Promotion Project of Fundamental Biotechnology Development from the Iwate Prefectural Government.

Literatures cited

- Carelse, O., J. Kempf, J. L. Toussaint, C. J. Chetsanga and M. V. Mubumbila (1992) The common bean chloroplast *trnH* (GUG) gene and its eukaryotic putative promoter elements. *Nuc. Acid. Res.* 20: 5481.
- Clegg, M. T. (1993) Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA.* 90: 363-367.
- , B. S. Gaut, G. H. Learn and B. R. Morton (1994) Rate and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA.* 91: 6795-6801.
- Dellaporta, S. L., J. Wood and J. B. Hicks (1983) A plant DNA miniprep: Version II. *Plant Mol. Biol. Rept.* 1: 19-21.
- Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C. R. Sun, B. Y. Meng, Y. Q. Li, A. Kanno, Y. Nishizawa, A. Hirai, K. Shinozaki and M. Sugiura (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217: 185-194.
- Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda and S. Tabata (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* 3: 109-136.
- Kowallik, K. V., B. Stoebe, I. Schaffr, P. Kroth-Pancic and U. Freier (1995) The Chloroplast Genome of a chlorophyll a+c-containing Alga, *Odontella sinensis* *Plant Mol. Biol. Rept.* 13: 336-342.
- Maier, R. M., K. Neckermann, G. L. Igloi and H. Koessel (1995) Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J. Mol. Biol.* 251: 614-628.
- Morton, B. R. and M. T. Clegg (1993) A chloroplast DNA mutational hotspot and gene conversion in a noncoding region near *rbcL* in the grass family (*Poaceae*). *Curr. Genet.* 24: 357-365.
- Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umetsu, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi and H. Ozeki (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572-574.
- Ogihara, Y., T. Terachi and T. Sasakuma (1992) Structural analysis of length mutations in a hot-spot region of wheat chloroplast DNAs. *Curr. Genet.* 22: 251-258.
- Palmer, J. D. (1993) A genetic rainbow of plastids. *Nature* 364: 762-764.
- Reith, M. and J. Munholland (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Mol. Biol. Rept.* 13: 327-332.
- Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B. Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada and M. Sugiura (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5: 2043-2049.
- Stirewalt, V. L., C. B. Michalowski, W. Löffelhardt, H. J. Bohnert and D. A. Bryant (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Mol. Biol. Rept.* 13: 327-332.
- Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki and M. Sugiura (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc. Natl. Acad. Sci. U. S. A.* 91: 9794-9798.
- Wolfe, K. H., C. W. Morden and J. D. Palmer (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. U. S. A.* 89: 10648-10652.
- Wolfson, R., K. G. Higgins and B. B. Sears (1991) Evidence for replication slippage in the evolution of *Oenothera* chloroplast DNA. *Mol. Biol. Evol.* 8: 709-720.
- Zhou, D. X., F. Quigley, O. Massenet and R. Mache (1989) Co-transcription of the *S10*- and *spc*-like operons in spinach chloroplasts and identification of three of their gene products. *Mol. Gen. Genet.* 216: 439-445.