Sequence variation of chloroplast DNA that involves EcoRI and ClaI restriction site polymorphisms in soybean

Akira Kanazawa*, Azumi Tozuka, and Yoshiya Shimamoto

Laboratory of Plant Genetics and Evolution, Faculty of Agriculture, Hokkaido University, Sapporo 060-8589, Japan

(Received 10 April 1998, accepted 22 May 1998)

Restriction fragment length polymorphisms (RFLPs) of EcoRI- and ClaI-digested chloroplast DNA (cpDNA) within the genus Glycine subgenus Soja were characterized. Two mutations were found to be responsible for the EcoRI and ClaI restriction site polymorphisms, and both were located in a region in which many ribosomal protein genes are clustered. This region is within the large single copy region of cpDNA and is located close to an inverted repeat. The locations of restriction sites of EcoRI and ClaI in the cpDNA region were analyzed by DNA gel-blot analyses and PCR amplification, which were followed by sequencing analyses. The EcoRI site polymorphism was found to have occurred in the intergenic spacer between rps11 and rpl36, while the ClaI site polymorphism was located within the 3' part of the coding region of rps3. The mutations that cause EcoRI and ClaI polymorphisms were both found to be single base substitutions. In addition to these polymorphisms, novel sequence variations in soybean cpDNA were detected near the sites of these mutations. Previously, it was shown that cultivated soybeans could be classified into three groups (I, II, and III) based on their cpDNA RFLPs. A comparison of the cpDNA sequences of soybeans in the present study was consistent with the notion that the cpDNA of group II soybeans is an intermediate between the cpDNAs of groups I and III.

INTRODUCTION

The genus Glycine consists of two subgenera, subgenus Soja and subgenus Glycine. The subgenus Soja includes two annual self-pollinated plants, the cultivated soybean [G. max (L.) Merr.] and its presumed wild progenitor G. soja Sieb. et Zucc. (Hymowitz, 1970), while the subgenus Glycine comprises more than a dozen wild perennial species (Doyle et al., 1990a). The plants of G. soja and the subgenus Glycine are distributed in East Asia and Australia, respectively.

Variations of chloroplast DNA (cpDNA), as well as variations of the nuclear ribosomal repeat region, has been widely used for plant systematics (Olmscheid and Palmer, 1994). In the subgenus Soja, variations of cpDNA have been detected (Shoemaker et al., 1986; Close et al., 1989; Lee et al., 1992, 1994; Shimamoto et al., 1992; Hirata et al., 1996; Powell et al., 1996), although they are not so many as those detected in the subgenus Glycine (Doyle et al., 1990a, 1990b; Doyle, 1991). Shoemaker et al. (1986) detected restriction fragment length polymorphisms (RFLPs) by gel electrophoresis of isolated cpDNA following restriction digestion. Close et al. (1989) detected RFLPs by DNA gel-blot analyses using mung bean cpDNA clones as probes. They reported seven mutations of cpDNA in the subgenus Soja. Some of these RFLP markers have been used to study the diversity of cytoplasm in an outcrossed population of soybean (Lee et al., 1992, 1994).

We have collected numerous cultivated soybean (G. max) and wild soybean (G. soja) from various regions of East Asia. In terms of DNA variation, we have been studying the diversity of cpDNA (Shimamoto et al., 1992; Hirata et al., 1996), as well as mitochondrial DNA (Tozuka et al., 1998; Kanazawa et al., 1998), of these plants. We have detected two hybridization profiles for each of EcoRI- and ClaI-digested DNAs when the digested DNAs were hybridized with a fragment of sugarbeet cpDNA designated as “H2”. This H2 clone is a 10.9-kb HindIII fragment that contains genes between 5'-rps12 and rps19 that are located adjacent to one of the inverted repeats (IR) in the large single copy (LSC) region of cpDNA (Kishima et al., 1986). The many soybeans examined could be classified into three groups, each with a different combination of EcoRI and ClaI RFLPs detected with the H2 probe. These three patterns
appeared to be identical to those detected by Close et al. (1989) and were designated groups I–III (Hirata et al., 1996). Although the RFLPs of cpDNA have been shown to provide valuable information for studying the evolution of the subgenus *Soja*, the precise locations of the mutations, as well as the mechanisms of the mutations, are unknown. In the present study we dissected the evolutionary changes of soybean cpDNA focusing mainly on the *Eco*RI and *Cla*I restriction site polymorphisms. Since the mutations were found to have occurred within the region of a ribosomal protein gene cluster in cpDNA, from which some genes were shown to have been lost during evolution (Spielmann et al., 1988; Gantt et al., 1991; Sugiura, 1992; Nakamura et al., 1997), we analyzed the structure of the region in terms of the location of genes in soybean.

**MATERIALS AND METHODS**

Soybean cultivars (*G. max*) 'Harosoy', 'Noir I', and 'Peking' were used as materials. A tobacco cultivar (*Nicotiana tabacum* L. var. Burley 21) was also used to construct probes for the hybridization analyses of soybeans. Total DNA was extracted from leaves of soybean and tobacco by the methods of Doyle and Doyle (1987), and Kanazawa and Tsutsumi (1992), respectively. Total DNA digested with restriction endonucleases was fractionated by electrophoresis on a 0.8% agarose gel. The DNA was transferred to nylon membranes and allowed to hybridize with the labelled probes. Hybridization was carried out using ECL direct nucleic acid labelling and detection systems (Amersham) as described previously (Tozuka et al., 1998). The final stringency of washing of membranes was 0.5 × SSC at 42°C. The membranes were used to expose X-ray films for 30 min.

PCR was performed using a DNA-amplification system (Perkin-Elmer). Each cycle of PCR consisted of denaturation for 1 min at 94°C, annealing for 2 min at 50°C, and extension for 2 min at 72°C. This cycle was repeated 30 times. The nucleotide sequences of the PCR products were analyzed either after being cloned into pBluescript II (Stratagene) or directly using a PRISM Cycle Sequencing Kit (Applied Biosystems). Primers for PCR were designed based on the sequence data of cpDNA from tobacco (Shinozaki et al., 1986), pea (Purton and Gray, 1989), *Oenothera* (Wolfson et al., 1991) and soybean (Spielmann et al., 1988). Primers were also made based on the sequence data obtained during the course of this study. The primers are listed in Table 1.

**Table 1. Nucleotide sequences of primers for PCR used in the present study**

<table>
<thead>
<tr>
<th>Name of primer</th>
<th>Sequence (5’ to 3’)</th>
<th>Position</th>
<th>Direction</th>
<th>Source of sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>petD-1’</td>
<td>CATGTAATGTAGGCTTAGCGG</td>
<td>petD exon 2</td>
<td>f</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>petD-2</td>
<td>GACCTAAGTAGTGGATTATCTC</td>
<td>petD exon 2</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>perL-3</td>
<td>TGGAGCAACATTACCTATTTG</td>
<td>petD exon 2</td>
<td>f</td>
<td>soybean (this study)</td>
</tr>
<tr>
<td>rpoA-1’</td>
<td>CCATTCCCATAAGAATGATAC</td>
<td>rpoA</td>
<td>f</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rpoA-2</td>
<td>CACTACATGGGAAGTGTGTTTG</td>
<td>rpoA</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rps11-1</td>
<td>GAGGCTCTACAGGCATTATG</td>
<td>rps11</td>
<td>f</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>S8-S11-1</td>
<td>ACTCTGCCTGTTGACATGCC</td>
<td>rps11</td>
<td>f</td>
<td>pea (Purton and Gray, 1989)</td>
</tr>
<tr>
<td>S8-S11-f</td>
<td>GTATGGATATATCCATTCTTG</td>
<td>rps11-1-rpl36 spacer</td>
<td>f</td>
<td>soybean (this study)</td>
</tr>
<tr>
<td>S8-S11-r</td>
<td>TGAATTAACCTACATGAAATC</td>
<td>infA</td>
<td>r</td>
<td>soybean (this study)</td>
</tr>
<tr>
<td>rpl14-2</td>
<td>CTGATCGCCTTCCCAACG</td>
<td>rpl14</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rps11-1</td>
<td>CAGATATAAACAACAAAATCTCCTCC</td>
<td>rps8</td>
<td>f</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>S8-S11-2</td>
<td>CTAGCACATAGAAGGATAAGGA</td>
<td>rps8</td>
<td>r</td>
<td><em>Oenothera</em> (Wolfson et al., 1991)</td>
</tr>
<tr>
<td>rpl16-1</td>
<td>AATGAAATGACAGTTGATTAGGCA</td>
<td>rpl16 exon2</td>
<td>f</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>cpl-5</td>
<td>CTCGATAGGATTATCCTTC</td>
<td>rpl16 exon2</td>
<td>f</td>
<td>soybean (this study)</td>
</tr>
<tr>
<td>rpl16-2</td>
<td>GAACCAAGTTTCCGTTAAAACAC</td>
<td>rpl16 exon2</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rps3-1</td>
<td>CTCCGATAGGATTCCGAACCT</td>
<td>rps3</td>
<td>r</td>
<td>soybean (this study)</td>
</tr>
<tr>
<td>rps3-2</td>
<td>GGTCTCACTGAGTGTTTACAC</td>
<td>rps3</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rpl22-1</td>
<td>CGAATCTACATCCTAATCCTAC</td>
<td>rpl22</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rpl22-2</td>
<td>GAAGTGATATGTTGAGTGAAC</td>
<td>rpl22</td>
<td>r</td>
<td>tobacco (Shinozaki et al., 1986)</td>
</tr>
<tr>
<td>rps19-1</td>
<td>TAAACCGGAGCATATTATTCATT</td>
<td>rps19</td>
<td>f</td>
<td>soybean (Spielmann et al., 1988)</td>
</tr>
<tr>
<td>rps19-2</td>
<td>AATCTTGTGGGAGGAAATCTT</td>
<td>rps19</td>
<td>r</td>
<td>soybean (Spielmann et al., 1988)</td>
</tr>
</tbody>
</table>

*f*, the direction from left to right on the gene maps in Figs. 2 and 3; *r*, the reverse.

The order of primers corresponds to the positions from left to right on the gene maps in Figs. 2 and 3.
RESULTS

The mutations that cause EcoRI and ClaI polymorphisms of soybean cpDNA have occurred within the region of a ribosomal protein gene cluster. We previously detected polymorphisms of cpDNA in different soybean cultivars when their DNAs were digested with the restriction endonucleases EcoRI or ClaI and hybridized with the H2 clone (for the cpDNA region in the H2 clone, see above) in DNA gel-blot analyses (Shimamoto et al., 1992; Hirata et al., 1996; see Fig. 1). Among all the soybeans examined, two hybridization patterns were detected with EcoRI-digested DNAs [patterns A (lanes 1 and 2 in Fig. 1A) and B (lane 3 in Fig. 1A)], and two hybridization patterns were detected with ClaI-digested DNAs [patterns a (lane 1 in Fig. 1B) and b (lanes 2 and 3 in Fig. 1B)]. It was found that all of the soybeans could be classified into three groups based on these patterns: group I had hybridization patterns of A and a, group II had hybridization patterns of A and b, and group III had hybridization patterns of B and b. In the present study, one cultivated soybean accession in each of these groups was used: ‘Harosoy’, was used to represent group I, ‘Noir I’ was used to represent group II, and ‘Peking’, was used to represent group III. To identify the location of the mutations that are responsible for the variations of EcoRI and ClaI fragments, at first, the H2 clone, which is 10.9 kb in length, was divided into four portions by a digestion with PstI and each of the four fragments was used as a probe for a DNA gel-blot analysis (data not shown). The resulting hybridization patterns indicated that the mutations occurred in a region close to the inverted repeat, in which many genes for ribosomal protein are located as a cluster. In order to make probes for this region, we amplified several portions of the region by PCR from tobacco total DNA. The primers were based on the sequence data of tobacco cpDNA (Shinozaki et al., 1986) except for the region of rps19, which was amplified from soybean using primers that were based on soybean sequence reported by Spielmann et al. (1988) (for primers, see Table 1). The amplified fragments include the genes located in this region in many higher plants: exon 2 of petD; rpoA; rps11, rpl36 and infA; rps8 and rpl14; exon 2 of rpl16; rps3; rpl22; rps19 (see Figs. 2 and 3). Then the PCR-amplified fragments were each used as a probe for a DNA gel-blot analysis of the groups I–III soybean cpDNAs. One or two strongly hybridized signals were detected with each probe except for the probe for rpl22 (see Figures 2 and 3). The rpl22 was previously shown to be absent from the chloroplast genome in legumes including soybean (Gantt et al., 1991). The rpl22 was found to be encoded in the nuclear genome as a single-copy gene in pea (Gantt et al., 1991). We could not detect a hybridization signal of the nucleus-encoded rpl22 of soybean by the hybridization condition and a short exposure (30 min) to X-ray films.

With regard to EcoRI-digested DNA, RFLPs were detected when rpoA, rps11-rpl36-infA, rps8-rpl14, and rpl16 were used as probes (Fig. 2). These probes hybridized with a 4.8-kb fragment in ‘Harosoy’ (group I) and ‘Noir I’ (group II), while they hybridized with a 1.9-kb and/or a 2.9-kb fragment in ‘Peking’ (group III). Among these four probes, the rps11-rpl36-infA probe hybridized with both the 1.9-kb and 2.9-kb fragments in ‘Peking’. This suggests that an EcoRI site that causes the RFLP between the groups I and II cpDNAs and the group III cpDNA is probably located within the region that includes rps11, rpl36, and infA in the group III cpDNA. The hybridization signals of petD differed from those of rpoA, and the hybridization signals of rps3 differed from the signals of both rpl16 and rps19. No difference was detected among groups I–III cpDNA with probes petD, rps3, and rpl19. Therefore, EcoRI sites, which are conserved among the cpDNAs of groups I–III, appear to be located at least 1) between petD and rpoA, 2) between exon 2 of rpl16 and rps3, and 3) between rps3 and rps19, if we assume that the order of genes in this region is conserved between tobacco and soybean.

With regard to ClaI-digested DNA, RFLPs were detected when rps3 and rps19 were used as probes, while no RFLP was detected by other probes (Fig. 3). Both rps3 and rps19 hybridized with a 2.5-kb fragment in ‘Harosoy’ (group I) and a 3.8-kb fragment in ‘Noir I’ (group II) and ‘Peking’ (group III). This suggests that the variation of the ClaI

![Fig. 1. RFLPs of EcoRI- and ClaI-digested cpDNA in soybean. Gel-blot analysis of total DNA isolated from ‘Harosoy’ (group I cpDNA; lane 1), ‘Noir I’ (group II cpDNA; lane 2), and ‘Peking’ (group III cpDNA; lane 3) probed with the H2 clone after digestion with EcoRI (panel A) or ClaI (panel B).](image-url)
site occurs around \textit{rps3} and \textit{rps19}.

The structure of the region was also analyzed by PCR amplification of soybean DNAs. Based on sequence data of other plants, primers were chosen to amplify regions that contained two or more adjacent genes (for primers, see Table 1). As a result, fragments of similar size were amplified from soybean and tobacco except that the products of \textit{rps11-rpl36-infA} from soybean was longer than that from tobacco, and the product of \textit{rps3-rps19} from soybean was shorter than that from tobacco (data not shown). Subsequent
analyses of nucleotide sequences showed that the former
difference was due to a longer sequence of the intergenic
spacer between rps11 and rpl36 in soybean cpDNA. The
latter difference is probably due to the lack of rpl22 in soy-
bean cpDNA. Amplification of the fragments of similar
size with tobacco DNA, as well as the observed patterns of
the signals in the DNA gel-blot analyses, indicated that
the order of genes in the region is also conserved in soy-
bean. Moreover, since no variation in the size of PCR
products was detected among the three groups of cpDNA,
the mutations that cause the RFLPs among these three
groups of cpDNA were expected to involve very short
sequences (e.g., point mutations) rather than large dele-
tions or insertions.

**Single base substitutions account for both EcoRI and
ClaI polymorphisms.** The amplified PCR fragments
were served for digestion with EcoRI or ClaI in order
to check the presence or absence of restriction sites of these enzymes (data not shown). To identify the precise loca-
tions of the EcoRI and ClaI restriction sites, the nucleotide
sequences of some of the PCR products were analyzed: the
petD-rpoA from ‘Harosoy’; the rps11-rpl36-infA-rps8 from
both ‘Harosoy’ and ‘Peking’; and the rpl16-rps3 from
‘Harosoy’. The regions that include mutations in the
EcoRI or ClaI restriction sites were further amplified by
PCR (Fig. 4A and B) and the nucleotide sequences were
analyzed with regard to all three groups of cpDNA.

The sequence analyses revealed that the EcoRI site in the
petD-rpoA region was located 69 bp downstream from
the termination codon of petD in the intergenic spacer
between petD and rpoA (for restriction map, see Fig. 5).
The two EcoRI sites in the rpl16-rps3 region were both
located within intron 1 of rpl16. Concerning the EcoRI
site in the rps3-rps19 region, a partial sequence of this
region was available (Spielmann et al., 1988). In this
sequence, we found an EcoRI site 57 bp downstream from
the termination codon of rps19 in the intergenic spacer
between rps3 and rps19. The position of this EcoRI site
was consistent with the digestion pattern of PCR products.

![Fig. 4](image)

**Fig. 4.** Presence or absence of EcoRI and ClaI sites in the PCR-amplified fragments. The PCR-amplified fragments that contained a region from the intergenic spacer between rps11 and rpl36 to infA (panel A, lanes 1–3), and a region from rpl16 intron 1 to rps3 (panel B, lanes 1–3) were digested with EcoRI (panel A, lanes 4–6) or ClaI (panel B, lanes 4–6). The PCR fragments amplified from ‘Harosoy’ (lanes 1 and 4), ‘Noir I’ (lanes 2 and 5), and ‘Peking’ (lanes 3 and 6) were analyzed in both experiments. The primers used for PCR were S8-S11-f and S8-S11-r (panel A) and cpI-1 and cpI-2 (panel B; for primers, see Table 1). M, phiX174 RF DNA digested with HaeIII was loaded for size markers.

![Fig. 5](image)

**Fig. 5.** Locations of the restriction sites of EcoRI and ClaI in the region close to IR within the LSC region of cpDNA in ‘Harosoy’ (group I cpDNA), ‘Noir I’ (group II cpDNA), and ‘Peking’ (group III cpDNA). Lines below the gene map indicate the regions in which nucleotide sequences were analyzed in the present study. E, EcoRI; C, ClaI.
in this study (data not shown). The EcoRI site in the rps11-rpl36-infa region, which showed polymorphism between 'Harosoy' and 'Noir I', and 'Peking', was located at 35 bp downstream from the termination codon of rpl36 in the intergenic spacer between rps11 and rpl36. The difference in the EcoRI site was found to be due to a single base

**Figure 6.** Nucleotide sequences of 'Harosoy', 'Noir I', and 'Peking' around the restriction sites that showed variation in soybean. Sequences around the EcoRI site in the intergenic spacer between rps11 and rpl36 (panel A), and those of the 3' part of the coding region of rps3 that contains Clai site and the intergenic spacer between rps3 and rpl16 (panel B), are shown. The coding regions are shown by boxes. Dots indicate the same nucleotide as in 'Harosoy' and dashes indicate gaps.
substitution from C to A, or the reverse, within the six bases of GAATTC, which is the target of EcoRI (Fig. 6A).

The ClaI sites were detected in intron 1 of rpl16 and in the coding region of rps3 in 'Harosoy', while only the former ClaI site was detected in 'Noir I' and 'Peking'. As in the EcoRI site, the difference in the latter ClaI site was also found to be due to a single base substitution from C to A, or the reverse, within the six bases of ATCGAT, which is the target of ClaI (Fig. 6B). The mutation in the coding region would not change the amino acid encoded at this location.

This analysis also revealed that the exact size of the smallest ClaI fragment in the region found in group I is 1,288 bp (data not shown), which has previously been reported to be 1.1 kb (Close et al. 1989). The locations of restriction sites of EcoRI and ClaI in the cpDNA region, based on the results of DNA gel-blot analyses, PCR amplification and nucleotide sequencing, are shown in Figure 5.

Detection of novel variations of the cpDNA of soybean. In addition to the variations in the EcoRI and ClaI restriction sites, we found two novel variations of the cpDNA of soybean in the regions around the EcoRI and ClaI sites (shown in Fig. 6A and B, respectively). In the intergenic spacer between rps3 and rpl16 (Fig. 6A), the 5-bp sequence of AAAAT was present 166 bp downstream from the stop codon of rpl36 in 'Peking', while it was present as a duplicated form of AAAATAAAAT in 'Harosoy' and 'Noir I'. In the intergenic spacer between rps3 and rpl16 (Fig. 6B), the nucleotide 75 bp upstream from the initiation codon of rpl16 was C in both 'Harosoy' and 'Noir I', while it was A in 'Peking'. These sequence variations were not likely to be due to amplification error, since more than two PCR products that were amplified independently showed the same sequence. If we count the duplication or deletion of the 5-bp sequence and a base substitution as single mutational events, then, based on the differences so far detected, 'Harosoy' (group I) and 'Noir I' (group II) differ by one mutation, 'Noir I' and 'Peking' (group III) differ by three mutations, and 'Harosoy' and 'Peking' differ by four mutations. Thus, 'Noir I' (group II) appeared to be intermediate between 'Harosoy' (group I) and 'Peking' (group III) with regard to these four mutational events.

DISCUSSION

In the present study we identified the locations of EcoRI and ClaI sites that showed polymorphisms among the accessions of cultivated soybean, as well as several other restriction sites of these enzymes that showed no polymorphism. The restriction sites that showed polymorphisms were found to be located in the region close to an inverted repeat within the LSC region of cpDNA, in which many ribosomal protein genes are present as a cluster. The mutations that cause EcoRI or ClaI polymorphisms were both found to be single base substitutions.

The region in which variations of EcoRI and ClaI sites are located is of interest in terms of evolutionary changes in the chloroplast genome. The genes for ribosomal protein in the region form an operon (rpl23 operon) in tobacco chloroplast, which is similar to S10, spc, and α operons in Escherichia coli (Sugiura, 1992). Rpl29 and rps17, which are located between rpl16 and rpl14, and rpl24 and rpl5, which are located between rpl14 and rps8, in both the genome of cyanobacteria (Kaneko et al., 1996), and the plastid genomes of algae (Reith and Munholland, 1995), were lost from the region in the plastid genomes of land plants such as liverwort (Ohyama et al., 1986), tobacco (Shinozaki et al., 1986), rice (Hiratsuka et al., 1989), and black pine (Wakasugi et al., 1994). Furthermore, in legumes, the rpl22 were lost from the region and transferred to the nucleus (Spielemann et al., 1988; Gantt et al., 1991). Thus, the structure of this region seems unstable. Transfer of a gene, cox2, from mitochondria to the nucleus was also shown to have occurred during the evolution of legumes (Nugent and Palmer, 1991; Covello and Gray, 1992). There may be other differences with regard to the location of genes in the chloroplast genome in soybean. With such a possibility in mind, we examined whether other genes might have been lost from the cpDNA region in soybean. Both the PCR results and the DNA gel-blot analysis indicated that no other sequence in this region of soybean cpDNA was lost, although the expression of the genes in this region remains to be examined.

Close et al. (1989) detected a difference among many cultivars in which the EcoRI fragments showed either a 4.8-kb signal or a 2.5-kb signal (in addition to several other hybridization signals) with the probe “MB-4b”. They designated this mutation as “mutation 6”. (In the present study, the size of the “2.9-kb” signal was estimated as 2.9 kb). They were unable to determine what type of mutation was present (shown to be a point mutation in the present study) because they used large fragments as probes. This resulted in hybridization signals that obscured the other polymorphic signal of the group III cpDNA. This is very similar to our case in which the long H2 clone was used as a probe: although we failed to find a polymorphic signal other than the 4.8-kb and 2.9-kb signals with the H2 probe (see Fig. 1), we detected the polymorphic signal of 1.9 kb in the group III cpDNA with probes that cover short regions of rpoA and rps11-rpl36-infA (see Fig. 2). We also found that a 1.9-kb EcoRI fragment, which is common to the three accessions, actually hybridized with the region from 5'-rps12 to petD (data not shown). This signal obscured the 1.9-kb polymorphic signal. Although we have not examined the identity with the 1.9-kb common signal, a weak hybridization signal of 1.9 kb was detected by the petD probe (see Fig. 2). A search of the EMBL database showed that petD shared slight homology with mitochondrial DNA in higher plants. Thus, the weak hybridization signal of soybean may be due to cross-hybridization of the petD probe.
to mitochondrial DNA. Further analysis may be required to characterize this signal.

The present results showed that “mutation 4” and “mutation 5” reported by Close et al. (1989) are the same. Mutation 4, which Close et al. (1989) detected with the probe “MB-13”, is shown by the presence or absence of a 4.8-kb EcoRI fragment. They stated that the 4.8-kb EcoRI fragment may be the same as the 4.8-kb fragment of mutation 5 detected with the probe MB-4b. The facts that the MB-4b probe contains a region between the inverted repeat and a region covered by MB-13 probe and that the MB-13 probe failed to hybridize with the 2.9-kb EcoRI fragment of the group III cpDNA (Close et al., 1989) were consistent with our restriction maps. The MB-13 probably contains rpoA and/or rps11, but it would not contain an extensive part of the region closer to the inverted repeat, which allows hybridization with the 4.8-kb EcoRI fragment in the groups I and II cpDNA but not with the 2.9-kb EcoRI fragment in the group III cpDNA, while MB-4b probably contains an extensive part of the region from rpl36 to rpl16, which allows hybridization with the 4.8-kb in the groups I and II, and with the 2.9-kb fragments in the group III. The expected signal of a 1.9-kb fragment by the MB-13 probe may be obscured by other signals as mentioned above. Thus, the observed differences in the hybridization patterns detected by using the MB-13 and MB-4b appeared to be due to the difference in the regions that were covered by these probes. Therefore, the mutation 4 and mutation 5 are the same and the point mutation that occurs between rps11 and rpl36, which we identified, presumably accounts for the mutation.

We have analyzed the EcoRI and ClaI site polymorphisms in more than 3,000 plants of wild soybean and cultivated soybean that are growing in East Asia (Abe et al., unpublished; Shimamoto et al., unpublished data). In wild soybean, groups III and II were predominant, while group I was very rarely detected. On the other hand, group I is predominant in cultivated soybean (unpublished data). Close et al. (1989) implied that the evolutionary order among these three groups is group III, group II, group I, since most of the plants of G. soja were placed into group III, while the plants of G. max were placed into groups I or II in their analysis. Close et al. (1989) also mentioned the possibility that there may be a missing ancestral type that is an intermediate between group II cpDNA and group III cpDNA, since more than one polymorphism that include mutation 4 (or 5) were detected between these two types. In the present study, we detected mutations other than EcoRI and ClaI site mutations. The number of differences so far analyzed among the three accessions was consistent with the notion that group II cpDNA is an intermediate group between group I cpDNA and group III cpDNA. Our ongoing analysis of the sequencing of the respective regions of the plants in the subgenus Glycine may help us polarize the mutations that generate groups I–III, as well as subdivide the three groups of cpDNA in the subgenus Soja. The results of the present study provide the bases for these analyses.

We would like to thank Jun Abe for many helpful discussions and comments with regard to soybean evolution.

REFERENCES


Sequence variation of cpDNA in soybean

and assignment of potential protein-coding regions. DNA Res. 3, 109–136.