Theor Appl Genet (2001) 102:683–688

D.H. Xu · J. Abe · A. Kanazawa · J.Y. Gai
Y. Shimamoto

Identification of sequence variations by PCR-RFLP and its application to the evaluation of cpDNA diversity in wild and cultivated soybeans

Received: 14 April 2000 / Accepted: 11 July 2000

Abstract The diversity and maternal lineage in wild and cultivated soybeans have previously been assayed using restriction fragment length polymorphism (RFLP) and sequencing analyses of chloroplast DNA (cpDNA). Here we describe a method based on PCR-RFLP for the identification of nucleotides at four mutation sites in non-coding regions of cpDNA. Of the four sites, two were located in restriction enzyme sites and two were not. For the latter two sites, new primers were designed to artificially create restriction sites that spanned them. The PCR-RFLP method enabled us to identify nucleotides at each of the four mutation sites easily and reliably. Fifty-seven wild and sixty-seven cultivated soybeans of different origins and different cpDNA types (types I, II, and III) were assayed. All of the samples tested could be classified into four haplotypes. All of the type-I and -II accessions had the same nucleotides at each of the four mutation sites, while all of the type-III accessions, except for 3 wild ones, had nucleotides that were different from those of types I and II. A sequencing analysis revealed that the 3 wild accessions possessed other single-base variations in the non-coding regions of *trn*H-*psb*A and *trn*T-*trn*L. The results of this study suggest that the type-I and type-II chloroplast genomes form a group that is distinct from the type-III chloroplast genome.

Keywords Soybean · cpDNA · Sequence variation · PCR-RFLP

Communicated by R. Hagemann

D.H. Xu · J. Abe (✉) · A. Kanazawa · Y. Shimamoto
Laboratory of Plant Genetics and Evolution,
Graduate School of Agriculture, Hokkaido University,
Sapporo 060-8589, Japan
Fax: +81-11-7064933
e-mail: jabe@res.agr.hokudai.ac.jp

J.Y. Gai
Soybean Research Institute, Nanjing Agricultural University,
Nanjing, 210095, China

Introduction

Chloroplast DNA (cpDNA) polymorphism in wild (*Glycine soja*) and cultivated (*Glycine max*) soybeans has been used in several studies to evaluate cytoplasmic diversity and maternal lineage (Shoemaker et al. 1986; Close et al. 1989; Lee et al. 1992a, b; Hirata et al. 1996; Shimamoto et al. 1992, 1998, 2000; Abe et al. 1999; Xu et al. 2000). Three haplotypes (types I, II, and III) have been determined by means of restriction fragment length polymorphism (RFLP) analysis, and their geographical distributions have been extensively evaluated in both wild and cultivated soybeans (Shimamoto et al. 1992, 1998, 2000; Hirata et al. 1996; Abe et al. 1999). The type-I haplotype is the most predominant among cultivated soybeans and the type-III haplotype is the most predominant among wild soybeans, while the type-II haplotype is present at similar frequencies in both species (Shimamoto et al. 1998, 2000; Abe et al. 1999). The mutational events characterizing the RFLP profiles used for the classification of the three cpDNA types are two single-base substitutions. One is located in an *Eco*RI site in the non-coding region *rps*11-*rpl*36, which discriminates types I and II from type III, and the other is located in a *Cla*I site in the 3′ part of the coding region of *rps*3, which discriminates type I from types II and III (Kanazawa et al. 1998). Based on this distribution, the type-II chloroplast genome has been considered to be an intermediate in the evolutionary transition from type III to type I, or vice versa (Close et al. 1989; Kanazawa et al. 1998). A subsequent sequence analysis of non-coding regions of cpDNA (Xu et al. 2000) revealed that type III was different from types I and II in at least five mutations out of a total of 3,849 bases examined in nine non-coding regions of cpDNA and that no mutational differences could be found between types I and II. Based on the results obtained from the sequencing analysis and a comparison with the sequences of some species of the subgenus *Glycine* used as out-group, Xu et al. (2000) assumed that the type-I and type-II chloroplast genomes had not been established monophyletically from type III, but rather that these three types had diverged from a common ancestor. In order that

the evolutionary process of different cpDNA types be clarified, an extensive survey of wild and cultivated soybeans of different origins and different cpDNA types is needed.

Sequencing analysis is not practical for studies such as population genetics and germplasm evaluation, in which a large number of samples often need to be surveyed. Several methods, such as allele-specific oligonucleotide hybridization (Myers et al. 1985), single-strand conformation polymorphism (SSCP) (Orita et al. 1989), dideoxy-fingerprinting (ddF) (Sarkar et al. 1992), and directed termination polymerase chain reaction (DT-PCR) (Chen and Hebert 1999), have been developed for the detection of single-base variations. However, these all have some disadvantages. In particular, their high cost and the large amount of effort required are often nearly the same as for sequencing analyses. A single-base variation in a particular region can be detected by amplifying the region using PCR and then digesting it with the appropriate restriction enzyme – if the variation is located in a restriction site. This approach has been successfully applied to identify plant varieties and species in the genus *Gossypium* (Martsinkovskaya et al. 1996) and in the genus *Beta* (Shen et al. 1998). In order to detect a single-base variation that is not located in a restriction site, we artificially created an enzyme recognition site that spanned the variation site by PCR with a primer pair in which one of the primers has a mismatched base, following the protocol of Haliassos et al. (1989). In this paper, we report a simple and reliable method for the identification of bases at four informative mutation sites of the soybean chloroplast genome and discuss the genetic diversity and evolutionary process of different cpDNA haplotypes.

Materials and methods

Plant materials

Fifty-seven wild and 67 cultivated soybean accessions with different cpDNA types (types I, II, and III) were used in this study. The wild soybeans were from Japan (20), China (28), South Korea (6), and far-eastern Russia (3); the cultivated soybeans were from Japan (18), China (38), South Korea (4), North Korea (4), and far-eastern Russia (3). The chloroplast DNA types for both wild and cultivated soybeans were determined from RFLP profiles following digestion with *Eco*RI and *Cla*I, electrophoresis, and hybridization with a 10.9-kb sugarbeet cpDNA fragment (Shimamoto et al. 1992).

PCR-RFLP analysis

Based on the results of the cpDNA sequence analysis, four single-base variations in three non-coding regions of cpDNA were selected for developing the PCR-RFLP method. The variations examined are located (1) 6 bp downstream of *psb*A for the #1 mutation in the non-coding region *trn*H-*psb*A, (2) and (3) 197 bp and 519 bp upstream of *trn*T for the #1 and #3 mutations, respectively, in the non-coding region *trn*T-*trn*L, and (4) 156 bp downstream of *psb*B for the #1 mutation in the non-coding region *psb*B-*psb*H (Xu et al. 2000). Two of these mutation sites, *trn*H-*psb*A #1 and *trn*T-*trn*L #1, are located in the restriction sites of *Mfl*I and *Ssp*I, respectively. Thus, the PCR products for these two regions were directly subjected to enzyme digestion and subsequent electrophoresis. The primers used were the same as those described by Xu et al. (2000): 5′TGATCCACTTGGCTACATCCGCC3′ (forward) and 5′GCTAACCTTGGTATGGAAGT3′ (reverse) for the non-coding region *trn*H-*psb*A; 5′GGATTCGAACCGATGACCAT3′ (forward) and 5′TTAAGTCCGTAGCGTCTACC3′ (reverse) for the non-coding region *trn*T-*trn*L. For the other two mutations, *trn*T-*trn*L #3 and *psb*B-*psb*H #1, a primer pair was designed to artificially create a restriction site by PCR in the region flanking the mutation site, following the method shown in Fig. 1.

Total DNA was extracted from young leaves according the method of Doyle and Doyle (1990). The PCR mixture contained

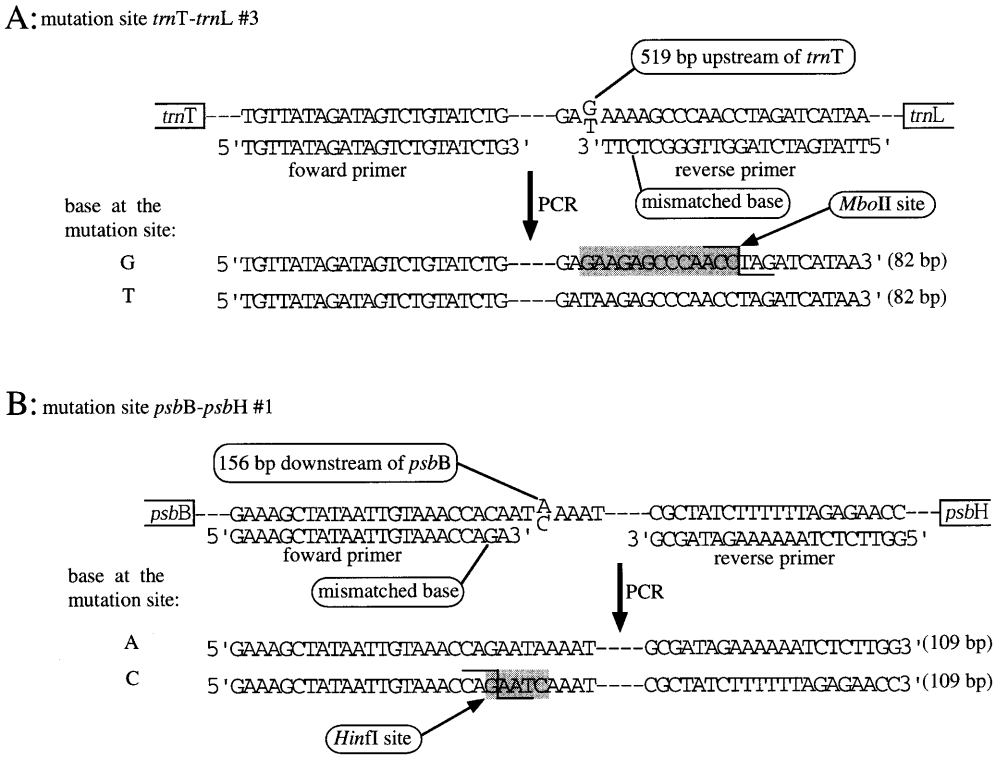

Fig. 1 Strategy for the identification of bases by PCR-RFLP at the mutation sites *trn*T-*trn*L #3 (**A**) and *psb*B-*psb*H #1 (**B**) that are not located in restriction sites. Restriction sites were artificially introduced by the use of a primer with a mismatched base

30 ng of total DNA, 0.25 μ*M* of 5′ and 3′ end primers, 100 μ*M* of nucleotides, 1 U *Tag* polymerase (TaKaRa, Japan) and 1 × PCR buffer (50 m*M* KCl, 10 m*M* TRIS-HCl, pH 8.3, 1.5 m*M* $MgCl_2$) in a total volume of 50 μl. Cycling consisted of three steps: (1) denaturation at 94°C for 5 min and annealing at 55°–60°C (depending on the primers) for 5 min, (2) 30 cycles at 94°C for 1 min, 55°–60°C for 2 min, and 72°C for 2 min, and (3) a final extension at 72°C for 10 min. The PCR was performed on the GeneAmp PCR System 9700 (Perkin-Elmer). The amplified products were then digested with the appropriate restriction enzymes, and the digests were separated on a 3% NuSieve 3:1 agarose gel in 1 × TBE buffer. After the electrophoresis, the gel was stained with ethidium bromide, and the DNA fragments were visualized under ultraviolet light.

Sequencing analysis

In order to confirm the results obtained by the PCR-RFLP method, we sequenced the three non-coding regions, *trn*H-*psb*A, *trn*T-*trn*L, and *psb*B-*psb*H, in which the four mutation sites are located, for 9 accessions with different cpDNA types. The sequencing analysis was performed as described by Xu et al. (2000).

Results and discussion

Identification of sequence variations with PCR-RFLP

Our previous sequencing analysis revealed that the base at the mutation site *trn*H-*psb*A #1 was either C or A and that the base at the mutation site *trn*T-*trn*L #1 was also either C or A in both the subgenus *Soja* and subgenus *Glycine* (Xu et al. 2000). The amplified product with a base C at the mutation site *trn*H-*psb*A #1 included a segment of AGATCC, which is an *Mfl*I restriction site (RGATCY; R: A or C; Y: C or T). Similarly, the amplified product with a base A at the mutation site *trn*T-*trn*L #1 included a segment of AATATT, which is an *Ssp*I restriction site. The amplified products of the non-coding region *trn*H-*psb*A were divided upon digestion with *Mfl*I into three fragments (250 bp, 58 bp, and 45 bp) if the nucleotide at the mutation site was C and into two fragments (295 bp and 58 bp) if the nucleotide was A, because a second *Mfl*I restriction site existed in the PCR fragment (Fig. 2A). Similarly, the amplified products of the non-coding region *trn*T-*trn*L were divided upon digestion with *Ssp*I into four fragments (549 bp, 168 bp, 73 bp and 61 bp) if the nucleotide at the mutation site was A and into three fragments (549 bp, 168 bp and 134 bp) if the nucleotide was C, because an additional two *Ssp*I restriction sites existed in the non-coding region (Fig. 2B).

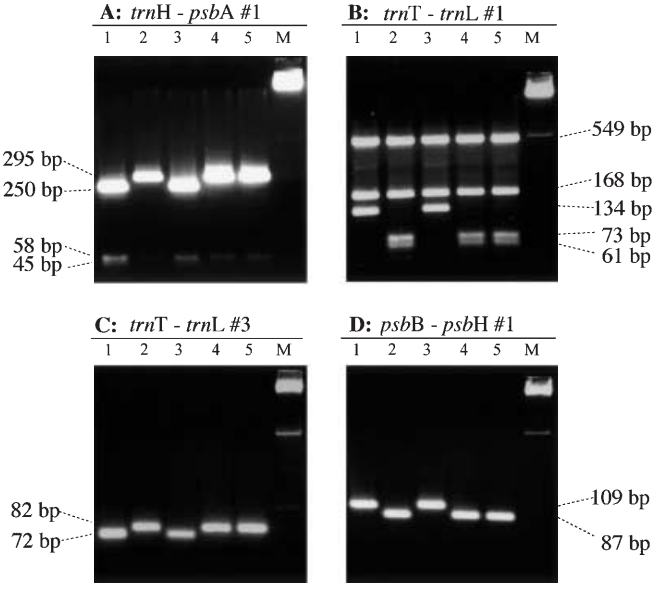

Fig. 2 Amplified products digested by restriction enzymes *Mfl*I (**A**), *Ssp*I (**B**), *Mbo*II (**C**), and *Hin*fI (**D**). These products identify the bases at the four mutation sites, *trn*H-*psb*A #1 (**A**), *trn*T-*trn*L #1 (**B**) *trn*T-*trn*L #3 (**C**), and *psb*B-*psb*H #1 (**D**), of soybean cpDNA. *Lane 1* Wasekin (type I), *lane 2* Tanbakuro (type III), *lane 3* Aogozen (type II), *lane 4* Chuteppou (type III), *lane 5* Kumadaizu (type III), *M* DNA marker (λ/*Hin*dIII)

For the other two mutations (*trn*T-*trn*L #3 and *psb*B-*psb*H #1), which were not located in restriction sites, new primers were designed to artificially create a restriction site in the region flanking the mutation site using a pair of primers, one of which contains a nucleotide that is mismatched with its target DNA (Fig. 1). The introduction of a mismatched base C in the reverse primer for the mutation site *trn*T-*trn*L #3 would result in the production of an *Mbo*II restriction site (GAAGANNNNNNNN) in the amplified product, if the nucleotide at the mutation site was G (Fig. 1A). Similarly, the introduction of a mismatched base G in the forward primer for the mutation site *psb*B-*psb*H #1 would result in the production of a *Hin*fI restriction site (GANTC) in the amplified product, if the nucleotide at the mutation site was C (Fig. 1B). Thus, the presence of nucleotide G at the *trn*T-*trn*L #3 mutation site could be recognized by the digestion of the PCR products with *Mbo*II and subsequent electrophoresis, which showed fragments differing in length by 10 bp (Fig. 2C). Similarly, the presence of nucleotide C at the *psb*B-*psb*H #1 mutation site could be recognized by the digestion of the PCR proucts with *Hin*fI and subsequent electrophoresis, which showed fragments differing in length by 22 bp (Fig. 2D). When the PCR products were not cleaved, we assigned the nucleotide T and A to mutation sites *trn*T-*trn*L #3 and *psb*B-*psb*H #1, respectively, since our previous sequencing analysis demonstrated that the base at mutation site *trn*T-*trn*L #3 was either G or T and that the base at mutation site *psb*B-*psb*H #1 was either A or C in both the *Soja* and *Glycine* subgenera (Xu et al. 2000). Therefore, the nucleotides at all of the four mutation sites were successfully and easily recognized by the PCR-RFLP method.

Newton et al. (1989) described a different method for the detection of point mutations by PCR in which one of the primers has a mismatched base. In this method, called the amplified refractory mutation system (ARMS), the primer is designed with the mutation site of interest at the 3′ end and the mismatched base generally placed the the third base position from the 3′ end of the primer. An amplified product was obtained with the mis-

Table 1 Nucleotides at four mutation sites in wild and cultivated soybeans revealed by PCR-RFLP

Species	cpDNA type	Mutation site							
		*trn*H-*psb*A #1		*trn*T-*trn*L #1		*trn*T-*trn*L #3		*psb*B-*psb*H #1	
		C	A	A	C	G	T	C	A
Glycine max	Type I (40)[a]	40	0	0	40	40	0	0	40
	Type II (10)	10	0	0	10	10	0	0	10
	Type III (17)	0	17	17	0	0	17	17	0
Glycine soja	Type I (1)	1	0	0	1	1	0	0	1
	Type II (11)	11	0	0	11	11	0	0	11
	Type III (45)	1	44	45	0	2	43	45	0

[a] Number in parenthesis indicates the number of accessions tested

matched primer for the target sequence but not for the non-target sequence, because a 1-base difference between the template DNA and the primer (due to the introduced mismatched base) is not enough to prevent the amplification, whereas a 2-base difference (due to the introduced mismatched base and mutated mutation base) does prevent amplification. Using the ARMS method, Shen et al. (1998) developed several taxa-specific primers for the internal transcribed spacer (ITS1 locus) of nuclear ribosomal DNA and successfully identified species and subspecies in the genus *Beta*. However, the PCR is often influenced by small fluctuations in experimental conditions, and consequently, the amplification itself may be unsuccessful. Moreover, we found that, in some cases, even a 2-base difference between the template DNA and the primer is not enough to prevent amplification. Compared with the ARMS method, the method used in the present study is more reliable; it can be applied to detect almost all of the single-base variations, if the sequence flanking the variation has the appropriate G+C content, a prerequisite of primer design in PCR. Using this method, Inazu et al. (1993) have developed a rapid method for the diagnosis of a human disease, cholesteryl ester transfer protein (CETP) deficiency. As more DNA sequence data become available, this single-base detection method will become an even more powerful tool in genotype identification, population genetics and ecological studies.

Diversity in wild and cultivated soybeans at four mutation sites: implications for the evolution of soybean chloroplast genome types.

Fifty-seven wild and 67 cultivated soybeans were assayed for nucleotides at the four mutation sites with the PCR-RFLP method described above. Table 1 shows the nucleotides at each of the four mutation sites in wild and cultivated soybeans with different cpDNA types (types I, II and III). Four combinations of nucleotides (haplotypes) were observed for all of the samples tested. Across the species, all of the type-I and type-II accessions tested possessed the same nucleotides, C, C, G and A, at the mutation sites, *trn*H-*psb*A #1, *trn*T-*trn*L #1 and #3, and *psb*B-*psb*H #1, respectively, whereas all of the type-III accessions, except for 3 wild ones, had the nucleotides A, A, T and C at these four sites. This result indicates that type-I and type-II chloroplast genomes have a closer relationship to each other than to the type-III chloroplast genome and are remote from the latter. Each of the 3 wild accessions, 2 from neighboring provinces of China, Henan (N24161) and Hubei (N24312), and 1 from Kochi prefecture, Japan (B07096), possessed the same base as types I and II at one of the four mutation sites. The 2 Chinese accessions had the nucleotide G at the *trn*T-*trn*L #3 mutation site, and the 1 Japanese accession had the nucleotide C at the *trn*H-*psb*A #1 mutation site. Thus, these 3 accessions formed another two haplotypes: A, A, G, C and C, A, T, C.

The results obtained by the PCR-RFLP analysis were confirmed by a sequencing analysis. The alignment of the sequences in the non-coding rgion *trn*H-*psb*A are presented as an example in Fig. 3. Of the 9 accessions sequenced, 4 that showed the digested phenotype (250 bp, 58 bp and 45 bp; see Fig. 2A) had the base C at the mutation site *trn*H-*psb*A #1 as expected by PCR-RFLP analysis. On the other hand, the remaining 5 that had the undigested phenotype (295 bp and 58 bp; see Fig. 2A) showed the base A. The same is true for the other three muation sites: the nucleotides that were expected from the PCR-RFLP for each of the 9 accessions were in fact observed in the sequencing analysis. The sequencing analysis also revealed that the 2 Chinese accessions (N24161 and N24312) and the 1 Japanese accession (B07096) each possessed other single-base variations in the non-coding regions *trn*T-*trn*L and *trn*H-*psb*A: 107 bp downstream of *psb*A in *trn*H-*psb*A for the 2 Chinese accessions and 130 bp downstream of *psb*A in *trn*H-*psb*A and 319 bp upstream of *trn*T in *trn*T-*trn*L for the Japanese accession. No sequence variation other than these mutations was detected among the 9 accessions sequenced.

The present study demonstrated that there are only two major combinations of nucleotides at the four mutation sites in the wild and cultivated soybean chloroplast genomes and that the other possible combinations were absent or rarely observed. As mentioned above, the 3 wild type-III accessions that had the same base as the type-I and -II accessions at one of the four mutation sites examined had unique sequence variations in other places in their cpDNAs. The two cpDNA haplotypes having nucleotide combinations of A, A, G, C and C, A, T, C are

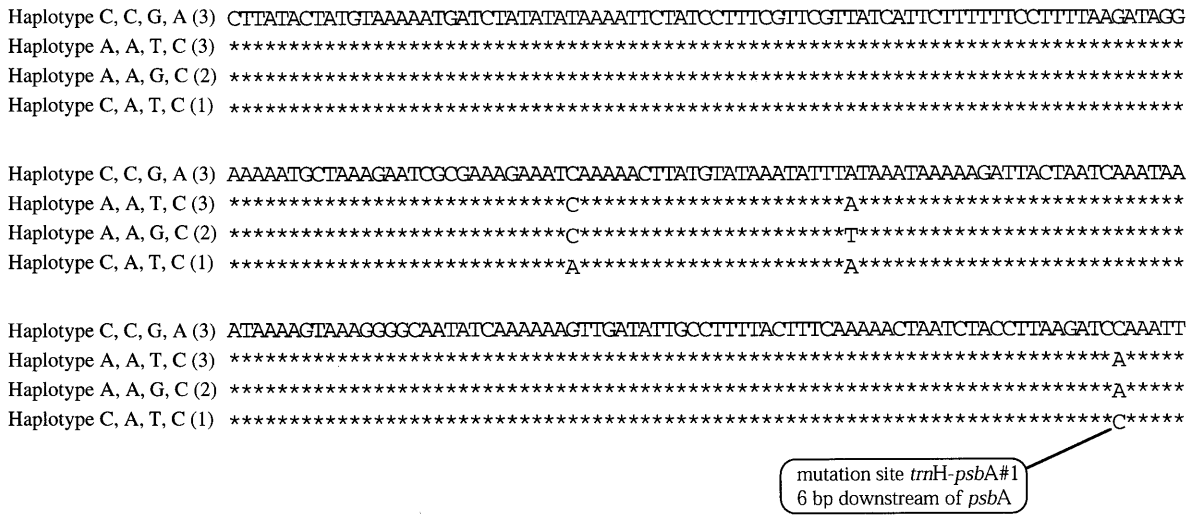

Fig. 3 Aligned sequences of the non-coding region *trn*H-*psb*A in four cpDNA haplotypes. *Asterisks* indicate the same base as in haplotype C, C, G, A. *Numbers* in *parenthesis* indicate the number of accessions sequenced. The three accessions of haplotype C, C, G, A are: Heihuangdou (type I, *G. max*), Aogozen (type II, *G. max*), and N23336 (type II, *G. soja*). The three accessions of haplotype A, A, T, C are: Hengshanhongdou (type III, *G. max*), B02138 (type III, *G. soja*), and B05018 (type III, *G. soja*). The two accessions of haplotype A, A, G, C are: N24312 (type III, *G. soja*) and N24161 (type III, *G. soja*). The one accession of haplotype C, A, T, C is: B07096 (type III, *G. soja*)

thus not likely to be evolutionary intermediates between the two major haplotypes (C, C, G, A and A, A, T, C), although they might be derivatives of possible intermediates. None of the wild and cultivated soybeans that we examined had a chloroplast genome type that suggests a transition from type III to type II and I. Close et al. (1989) implied that type I, a predominant type in *G. max*, had been derived from type III, a predominant type in *G. soja*, via type II based on the assumption that *G. soja* is the wild progenitor of *G. max*, although they also mentioned that there might be a missing ancestral type that is intermediate between type II and the type III. However, the findings obtained in the present study suggest that the type-I and -II chloroplast genomes are not directly derived from the extant type III. The predominant occurrence of haplotypes having the nucleotide combination C, C, G, A and A, A, T, C might have resulted from selection or random fixation of these haplotypes during the early stages of evolution of the subgenus *Soja*.

On the other hand, the type-I and -II chloroplast genomes showed a close relationship to each other than to those of type III. As has been revealed by an extensive survey of both wild and cultivated soybeans from various regions of East Asia (Shimamoto et al. 1992, 1998; Abe et al. 1999), the type-I chloroplst genome is the most predominant type in *G. max* but is rarely observed in *G. soja*. Wild accessions with the type-I chloroplast genome have been found in only four geographically separated sites in Japan (Abe et al. 1999). Abe et al. (1999) assumed that these may have been derived from hybrids between wild and cultivated soybeans or are relics of the direct progenitor of cultivated soybean with the type-I chloroplast genome. Further studies are needed to determine whether cultivars with the type-I chloroplast genome have been established from wild soybeans with the type-I chloroplast genome or from cultivars with the type-II chloroplast genome. In any case, the base substitution at the *Cla*I site in the coding region of *rps*3, which discriminates the type-I chloroplast genome from the type-II chloroplast genome, may have occurred during the early stages of soybean domestication.

Acknowledgements This research was supported by the Japan Society for the Promotion of Science (JSPS) (Grant No. PI98120) and the Heiwa Nakajima Foundation. The experiments of this study comply with the current laws of Japan.

References

Abe J, Hasegawa A, Fukushi H, Mikami T, Ohara M, Shimamoto Y (1999) Introgression between wild and cultiated soybeans of Japan revealed by RFLP analysis for chloroplast DNAs. Econ Bot 53:285–291

Chen J, Hebert PDN (1999) Directed termination of the polymerase chain reaction: kinetics and application in mutation detection. Genome 42:72–79

Close PS, Shoemaker RC, Keim P (1989) Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. Theor Appl Genet 77:768–776

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Haliassos A, Chomel JC, Tesson L, Baudis M, Kruh J, Kaplan JC, Kitzis A (1989) Modification of enzymatically amplified DNA for the detection of point mutations. Nucleic Acids Res 17:3606

Hirata T, Abe J, Shimamoto Y (1996) RFLPs of chloroplast and motochondrial genomes in summer and autumn maturing cultivar groups of soybean in Kyushu district of Japan. Soybean Genet Newsl 23:107–111

Inazu A, Koizumi J, Haraki T, Yagi K, Wakasugi T, Takegoshi T, Mabuchi H, Taketa R (1993) Rapid detection and prevalence of cholestryl ester transfer protein deficiency caused by an intron 14 splicing defect in hyperalphalipoproteinmia. Hum Genet 91:13–16

Kanazawa A, Tozuka A, Shimamoto Y (1998) Sequence variation of chloroplast DNA that involves *Eco*RI and *Cla*I restriction site polymorphisms in soybean. Genes Genet Syst 73:111–119

Lee DJ, Caha CA, Specht JE, Graef GL (1992a) Analysis of cytoplasmic diversity in an outcrossing population of soybean. Crop Sci 34:46–50

Lee DJ, Caha CA, Specht JE, Graef GL (1992b) Chloroplast DNA evidence for non-random selection of females in an outcrossed population of soybeans [*Glycine max* (L.)]. Theor Appl Genet 85:261–268

Martsinkovskaya AI, Moukhamedov RS, Abdukarimov AA (1996) Potential use of PCR-amplified ribosomal intergenic sequences for differentiation of varieites and species of *Gossypium* cotton. Plant Mol Biol Rep 14:44–49

Myers RM, Lumelsky N, Lerman LS, Maniatis T (1985) Detection of single base substitution in total genomic DNA. Nature 313:495–498

Newton CR, Graham A, Heptinstall LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res 17:2503–2516

Orita M, Suzuki TS, Hayashi K (1989) Rapid and sensitive detection of point mutation and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874–879

Sarkar G, Yoon H, Sommer SS (1992) Dideoxy fingerprinting (ddF): a rapid and efficient screen for the presence of mutations. Genomics 13:441–443

Shen Y, Newbury HJ, Ford-Lloyd BV (1998) Identification of taxa in the genus *Beta* using ITS1 sequence information. Plant Mol Biol Rep 16:147–155

Shimamoto Y, Hasegawa A, Abe J, Ohara M, Mikami T (1992) *Glycine soja* germplasm in Japan: isozyme and chloroplast DNA variation. Soybean Genet Newsl 19:73–77

Shimamoto Y, Fukushi H, Abe J, Kanazawa A, Gai J, Gao Z, Xu D (1998) RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. Genet Resour Crop Evol 45:433–439

Shimamoto Y, Abe J, Gao Z, Gai J, Thseng F (2000) Characterizing the cytoplasmic diversity and phyletic relationship of Chinese landraces of soybean, *Glycine max*, based on RFLPs of chloroplast and mitochondrial DNA. Gnet Resour Crop Evol (in press)

Shoemaker RC, Hatfield PM, Palmer RG, Atherly AG (1986) Chloroplast DNA variation in the genus *Glycine*, subgenus *Soja*. J Hered 77:26–30

Xu DH, Abe J, Sakai M, Kanazawa A, Shimamoto Y (2000) Sequence variation of non-coding regions of chloroplast DNA of soybean and related wild species and its implications for the evolution of different chloroplast haplotypes. Theor Appl Genet 101:724–732