Theor Appl Genet (2002) 105:645–653
DOI 10.1007/s00122-002-0972-7

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

Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean

Received: 1 November 2001 / Accepted: 19 December 2001 / Published online: 19 June 2002

Abstract Soybean [*Glycine max* (L.) Merr.] is one of the major crops in the world and was domesticated from a wild progenitor, *Glycine soja* Sieb. & Zucc., in East Asia. In order to address the questions concerning the evolution and maternal lineage of soybean, we surveyed the variation in chloroplast DNA simple sequence repeats (cpSSR) of 326 wild and cultivated soybean accessions that were collected from various Asian countries. Twenty-three variants were detected at six cpSSRs in the accessions tested. All of the variants were found in wild soybean, whereas only 14 variants existed in the cultigen. Combining the variants at the six cpSSRs gave 52 haplotypes in the former and eight haplotypes in the latter. Both analyses indicated a considerably higher genetic diversity in the wild soybean. Around 75% of the cultivated accessions tested possessed a common haplotype (no. 49), which was detected in only seven wild accessions, six from southern Japan and one from southern China. The predominant haplotype in the cultigen may therefore have originated from a rare haplotype of the wild soybean that is presently distributed in the southern areas of Japan and China. The remaining seven haplotypes in the cultigen were distributed regionally, and except for three rare haplotypes, largely overlapped with the distributions of wild accessions with the same respective haplotypes. Our results strongly suggest that the cultivated soybeans with different cpDNA haplotypes originated independently in different regions from different wild gene pools and/or hybrid swarms between cultivated and wild forms.

Communicated by R. Hagemann

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

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

Keywords Soybean · Chloroplast DNA · SSR · Diversity · Evolution

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the major crops in the world. It is widely accepted that the cultivated soybean was domesticated from the wild soybean, *Glycine soja* Sieb. & Zucc., in East Asia. Chloroplast DNA (cpDNA) has been used to address questions concerning the origins, maternal lineage, and degree of introgressions in soybean (Shoemaker et al. 1986; Close et al. 1989; Shimamoto et al. 1992, 1998, 2000; Hirata et al. 1996; Abe et al. 1999; Xu et al. 2000, 2001; Shimamoto 2001). Three cpDNA types (types I, II, and III) have been classified by means of restriction fragment length polymorphis (RFLP) analysis and their geographical distributions have been revealed for both wild and cultivated soybeans (Shimamoto et al. 1992, 1998, 2000; Hirata et al. 1996; Abe et al. 1999; Shimamoto 2001). The type-I chloroplast genome is predominant in the cultivated soybean, while the type-III chloroplast genome is observed at a high frequency in the wild soybean, although there is an overlapping between both species. The type-I chloroplast genome has been found only in wild plants collected from four geographically separated sites of Japan out of around 650 sites in Japan, China, South Korea, and far-eastern Russia (Shimamoto et al. 1998; Abe et al. 1999; our unpublished data). Abe et al. (1999) assumed that wild plants with the type-I chloroplast genome may be either derivatives of hybridization between cultivated and wild soybeans or relics of a direct progenitor of cultivated soybean with the type-I chloroplast genome.

Kanazawa et al. (1998) verified that 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 type I and type II from type III, and the other is located

Table 1 Six cpSSR markers analyzed in this study

cpSSR[a]	Location	Primer (5′–3′)	*Tm* (°C)	Motif of repeat[b]
gmcp1	*trn*T/*trn*L intergenic region	TCGATTCTATGCCCCTACTT (forward) AGACTCCCAAGTTTTCAGTCG (reverse)	58	$(T)_{12}$
gmcp2	*acc*D/*psa*I intergenic region	TCTAGTGATTATAGACTATCTTT (forward) GGAAAGTTTATCATAAAATGGG (reverse)	58	$(A)_{13}$
gmcp3	ORF203/ORF103 intergenic region	GCTTCAGAATTGTCCTATTTA (forward) ATCAAATAACGCCTCATCTAT (reverse)	52	$(A)_{12}CG(T)_{11}$
gmcp4	*atp*B/*rbc*L intergenic region	TATCACTGTCAAGATTAAGAG (forward) CTTTTATATGTATGGCGCAAC (reverse)	56	$(A)_{11}$
RD19	*rps*19	CTAAATATTACAAAATGGAATTCT (forward) ACCAATTCAAAAAATCGAATA (reverse)	55	$(A)_{14}$
SOYCP	*trn*M	CATAGATAGGTACCATCCTTTTT (forward) CGCCGTATGAAAGCAATAC (reverse)	55	$(T)_{13}(G)_{10}$

[a] Powell et al. (1995b) was referred for RD19 and SOYCP
[b] Motif of repeat in soybean cultivar Harosoy

in a *Cla*I site in the 3′ part of the coding region of *rps*3, which discriminates type I from type II and type III. Xu et al. (2000) identified a total of 3,849 bases for nine non-coding regions of cpDNA for 19 wild and cultivated soybean accessions with different cpDNA types to determine the mechanisms of molecular change among the three cpDNA types. They found no further base substitution between the type-I and type-II chloroplast genomes, although at least five mutational differences existed between the former two types and the type-III chloroplast genome. These results were further confirmed by a polymerase chain reaction (PCR)-RFLP method devised to identify the bases at four informative mutation sites and was applied to a total of 124 wild and cultivated soybean accessions from various regions of Asia (Xu et al. 2001). Based on these results, Xu et al. (2001) assumed that the type-I and type-II chloroplast genomes are genetically close to each other but remote from the type-III chloroplast genome. However, the variability of the cpDNA sequence of wild and cultivated soybeans is not high enough to clearly identify the origins of cultivated soybeans with different cpDNA types.

Simple sequence repeats in chloroplast genomes (cpSSRs) have been used for the study of genetic diversity of the cpDNA genome in plant species such as pine (Powell et al. 1995a), potato (Bryan et al. 1999), barley (Provan et al. 1999), rice (Provan et al. 1997; Ishii and McCouch 2000), and soybean (Powell et al. 1995b, 1996). Owing to the haploid nature and high copy number of the chloroplast genome, the cpSSRs can be easily analyzed by PCR. The higher level of intraspecific variation in cpSSRs compared with base substitutions and insertions/deletions revealed by RFLP and sequencing analyses can reveal the phylogenetic relationships between cultivated species and their wild progenitors in higher resolution.

We report here variation in six cpSSRs in soybean and its wild progenitor. The cpSSR analysis revealed that the chloroplast genomes of both wild and cultivated soybeans are highly variable. Our data strongly suggest that cultivated soybeans with different chloroplast genome types originated independently in different regions from different wild gene pools and/or hybrid swarms between cultivated and wild forms.

Materials and methods

Plant materials

One hundred and forty-three wild accessions and 183 cultivated accessions were used in this study. The wild plants were collected from the whole distribution area of the species: China (73), Japan (48), South Korea (14), and far-eastern Russia (8). The cultivated accessions tested were mostly landraces or pure-line selections and were selected to cover various Asian countries: China (92), Japan (41), South Korea (13), North Korea (6), far-east Russia (3), Vietnam (3), Thailand (3), Indonesia (4), Myanmar (5), India (3), Bhutan (2), Nepal (3), Pakistan (2), and Kyrgyz (3). The cultivar Harosoy was also included as a standard.

Classification of chloroplast genome types

Chloroplast genome types (types I, II, and III) were determined for all of the accessions using either RFLP analysis or PCR-RFLP analysis. The RFLP analysis followed the method of Shimamoto et al. (1998). Mutational events characterizing the RFLP profiles used for the classification of the three cpDNA types are single-base substitutions in an *Eco*RI site in the non-coding region *rps*11–*rpl*36 and a *Cla*I site in the 3′ part of the coding region of *rps*3 (Kanazawa et al. 1998). Two primer pairs were used to amplify the cpDNA regions that span the two mutation sites: 5′-GTATGGATATATCCATTTCGTG-3′ and 5′-TGAATAACTTACCCATGAATC-3′ for the mutation in the *Eco*RI site and 5′-ACTGAACAGGCGGGTACA-3′ and 5′-ATCCGAAGCGATGCGTTG-3′ for the mutation in the *Cla*I site. The cpDNA type was determined based on the digestion patterns of the PCR products with *Eco*RI and *Cla*I.

Analysis of cpSSRs

Six cpSSRs, including RD19 and SOYCP that were developed by Powell et al. (1995b), were used in the present study (Table 1). Primer pairs flanking the mononucleotide repeats region were de-

signed on the basis of our unpublished sequencing data and soybean cpDNA sequence data from GenBank (Reverdatto et al. 1995). For all of the cpSSRs, forward primers were labeled with either 6-FAM (blue), HEX (green), or NED (yellow) fluorescent dyes.

DNA was extracted as described by Doyle and Doyle (1990). The PCR reaction mixture contained 30 ng of total genome DNA, 0.25 μ*M* of 5′ and 3′ end primers, 200 μ*M* of each dNTPs, 0.5 U *Taq* polymerase (TaKaRa, Japan), and 1× PCR buffer (10 m*M* Tris–HCl, pH 8.3, 50 m*M* of KCl, 1.5 m*M* $MgCl_2$) in a total volume of 20 μl. PCR reactions were performed with a GeneAmp PCR System 9700 (Perkin Elmer/Applied Biosystems) using the following program: (1) denaturation at 94 °C for 2 min, (2) 30 cycles at 94 °C for 1 min; 52°–58 °C (depending on the primers) for 1 min; 72 °C for 1 min; (3) a final extension at 72 °C for 10 min. Following amplification, 1.5 μl 6-FAM-labeled, 4.0 μl HEX-labeled, and 2.0 μl NED-labeled PCR products were combined and brought to a total volume of 20 μl by adding distilled water. An aliquot (1.0 μl) of the mixture of PCR products was combined with loading buffer (1.0 μl) containing a ROX (red)-labeled internal size standard (GeneScan-500) and was then loaded and separated by an ABI 377 sequencer (Perkin Elmer/Applied Biosystems). GeneScan (version 3.1) software was used to visualize the variants and to estimate their sizes.

The genetic diversity index (*H*) based on variant or haplotype frequencies was calculated for each cpSSR using Nei's unbiased statistics (Nei 1987):

$$H = 1 - \sum P_i^2$$

where P_i is the frequency of the *i*th variant or haplotype.

Sequencing of cpSSR variants

In order to verify whether the observed length variations were due to changes in the mononucleotide repeats or due to other mutation events, all of the variants detected in this study were sequenced. Six new primer pairs were designed to amplify cpDNA regions that span each of the six cpSSR fragments. The amplified products were directly sequenced using the dRhodamine Terminator sequencing kit (Perkin-Elmer/Applied Biosystems) on an ABI 377 Sequencer following the manufacturer's instructions.

Results

Polymorphism at cpSSRs in wild and cultivated soybeans

All of the six cpSSRs showed polymorphism in both wild and cultivated soybeans (Table 2, Fig. 1). Variants usually consisted of several bands. This is a common feature of SSR because of the slippage of *Taq* polymerase during the amplification process. However, the variants could be unambiguously identified by assigning the fragment size of the strongest band to that of the variant. The number of variants found at each of the six SSRs ranged from three to six with an average of 3.8 (Table 2). The sequences of the cpSSRs revealed that the differ-

Table 2 Variation at six cpDNA SSRs in wild and cultivated soybeans, and three cpDNA types (*H* genetic diversity index)

cpSSR	Variant[a]) (bp)	Cultivated soybean (*n*=183)	Wild soybean (*n*=143)	cpDNA type		
				Type I (*n*=105)	Type II (*n*=65)	Type III (*n*=156)
gmcp1	121		0.014			0.013
	122	0.246	0.755			0.981
	123	0.754	0.231	1.000	1.000	0.006
	H	0.371	0.376	0.000	0.000	0.038
gmcp2	142		0.014			0.013
	143	0.246	0.664		0.015	0.609
	144	0.754	0.322	1.000	0.985	0.378
	H	0.371	0.455	0.000	0.030	0.486
gmcp3	107		0.014			0.013
	108		0.063			0.058
	109		0.147			0.135
	110	0.175	0.301		0.062	0.455
	111	0.825	0.448	1.000	0.892	0.333
	112		0.028		0.046	0.006
	H	0.289	0.683	0.000	0.198	0.660
gmcp4	134		0.196		0.292	0.058
	135	0.852	0.559	0.981	0.708	0.558
	136	0.148	0.245	0.019		0.385
	H	0.252	0.589	0.037	0.414	0.538
RD19	149	0.049	0.077			0.128
	150	0.951	0.888	0.981	1.000	0.853
	151		0.035	0.019		0.019
	H	0.094	0.204	0.037	0.000	0.256
SOYCP	91	0.044	0.056			0.103
	92	0.082	0.315	0.010	0.092	0.340
	93	0.869	0.455	0.990	0.908	0.391
	94	0.005	0.147			0.141
	95		0.028			0.026
	H	0.236	0.669	0.019	0.168	0.701
Mean *H*		0.269	0.496	0.016	0.135	0.446
Number of haplotypes		8	52	4	9	41
H at a level of haplotype		0.421	0.964	0.068	0.608	0.921

[a] Numbers that are underlined indicate the size of each cpSSRs in soybean cultivar Harosoy

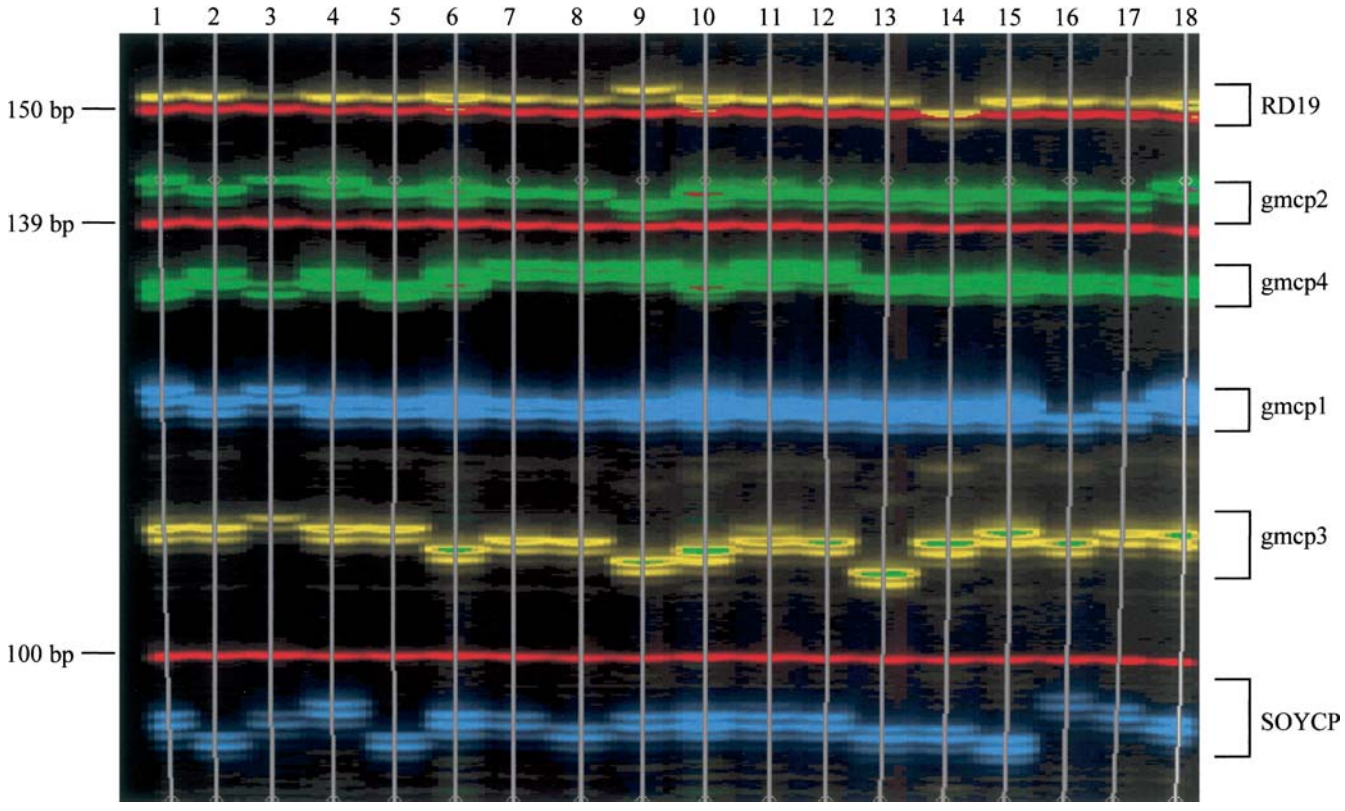

Fig. 1 Multiplex analysis of six cpSSRs in 17 wild soybean accessions on the ABI 377 sequencer. *Lane 18* is the standard cultivar Harosoy. *Red bands* are GeneScan-500 (ROX) size standards

```
B03046 (G. soja)        GCTTCAGAATTGTCCTATTTATTTTGTGTACCTTTTGAAATATTTTG----AAAAAAAAACG-
ZYD2727 (G. soja)       **************************************************----**********T
B05020 (G. soja)        **************************************************---A**********T
Yixianheidou (G. max)   **************************************************--AA**********T
Harosoy (G. max)        **************************************************-AAA**********T
ZYD4641 (G. soja)       **************************************************AAAA**********T

B03046 (G. soja)        TTTTTTTTTTCAAAATATTTCCATTTTTATAGATGAGGCGTTATTTGAT (107 bp)
ZYD2727 (G. soja)       ************************************************** (108 bp)
B05020 (G. soja)        ************************************************** (109 bp)
Yixianheidou (G. max)   ************************************************** (110 bp)
Harosoy (G. max)        ************************************************** (111 bp)
ZYD4641 (G. soja)       ************************************************** (112 bp)
```

Fig. 2 Aligned sequences of six variants of cpSSR gmcp3 in wild and cultivated soybeans. Asterisks indicate the same base as in B03046 and dashes indicate gaps. Primers used to amplify this region are indicated by underlines

ences in length among the PCR products were due to the differences in the number of mononucleotide repeats. No other length mutation such as an insertion or deletion was detected in the accessions sequenced. An example of the sequence alignment of the six variants at the gmcp3 is presented in Fig. 2.

Of the 23 variants detected, all were present in the wild soybean, while only 14 variants were observed in the cultigen. The *H* values ranged from 0.204 (RD19) to 0.683 (gmcp3) with an average of 0.496 in the wild soybean, and from 0.094 (RD19) to 0.371 (gmcp1 and gmcp2) with an average of 0.269 in the cultivated soybean (Table 2). Combining the variants for six cpSSRs generated 52 haplotypes (Table 3). All of the haplotypes were observed in the 143 wild accessions studied. The most frequent haplotype in the wild soybean was haplotype no. 25, although its frequency of occurrence was only 7.0%. Eleven other haplotypes (nos. 12, 13, 20, 21, 29, 30, 34, 40, 46, 47, and 49) were also observed at significant frequencies of 3.5–6.3%. These 12 haplotypes accounted for 58.7% of the wild accessions tested. Twenty-six (18.2%) of the wild accessions had a unique haplotype. On the other hand, eight haplotypes were observed in the cultivated soybean. No haplotype was unique to the cultigen. The *H* values at the level of haplotype in the wild and cultivated soybeans were 0.964 and 0.421, respectively, indicating a considerably high diversity at cpSSRs in the wild soybean.

The chloroplast genome typing with the RFLP or PCR-RFLP methods gave results similar to those of our previous studies (Shimamoto et al. 1998, 2000; Abe et al. 1999). The type-I chloroplast genome was predominant in the cultivated soybean (54.5%) and the type-III genome was predominant in the wild soybean (77.1%), while the type-II genome had similar frequencies in both the former (20.3%) and the latter (18.8%). In the wild soybean, the type-III chloroplast genome consisted of 40 cpSSR haplotypes, the type-II genome consisted of nine haplotypes, and the type-I genome consisted of three haplotypes. In the cultivated soybean, the type-III chloroplast genome consisted of six cpSSR haplotypes, the type-II genome consisted of one haplotype, and the type-I genome consisted of two haplotypes.

Table 3 Haplotypes constructed by six cpSSRs and their distribution in wild and cultivated soybeans, and three cpDNA types

Haplotype no.	cpSSR						Number of accessions		cpDNA type	Haplotype no.	cpSSR						Number of accessions		cpDNA type
	gmcp1	gmcp2	gmcp3	gmcp4	RD19	SOYCP	*G. max*	*G. soja*			gmcp1	gmcp2	gmcp3	gmcp4	RD19	SOYCP	*G. max*	*G. soja*	
1	121	143	110	134	150	93	–	1	III	27	122	143	111	134	150	91	–	1	III
2	121	143	110	135	150	95	–	1	III	28	122	143	111	135	149	92	–	1	III
3	122	142	108	136	151	93	–	1	III	29	122	143	111	135	150	91	8	6	III
4	122	142	110	135	150	94	–	1	III	30	122	143	111	135	150	92	–	8	III
5	122	143	107	135	150	92	–	1	III	31	122	143	111	135	150	93	–	1	III
6	122	143	107	136	150	92	–	1	III	32	122	143	111	135	150	94	–	2	III
7	122	143	108	134	150	95	–	1	III	33	122	143	111	135	151	93	–	2	III
8	122	143	108	135	150	93	–	2	III	34	122	143	111	136	150	92	5	5	III
9	122	143	108	136	150	92	–	2	III	35	122	143	111	136	150	93	–	2	III
10	122	143	108	136	150	94	–	1	III	36	122	143	111	136	150	94	–	1	III
11	122	143	109	134	150	93	–	3	III	37	122	144	108	136	150	92	–	1	III
12	122	143	109	135	150	92	–	6	III	38	122	144	108	136	150	93	–	1	III
13	122	143	109	135	150	93	–	7	III	39	122	144	111	135	150	92	–	1	III
14	122	143	109	135	150	94	–	1	III	40	122	144	111	135	150	94	–	9	III
15	122	143	109	136	150	92	–	1	III	41	122	144	112	135	150	92	–	1	III
16	122	143	109	136	150	93	–	2	III	42	123	143	111	134	150	93	–	1	II
17	122	143	110	134	150	92	–	1	III	43	123	144	109	134	150	93	–	1	II
18	122	143	110	134	150	93	–	1	III	44	123	144	110	134	150	93	–	3	II
19	122	143	110	135	149	91	–	1	III	45	123	144	110	135	150	93	–	1	II
20	122	143	110	135	149	92	9	9	III	46	123	144	111	134	150	92	–	5	II
21	122	143	110	135	150	93	1	5	III	47	123	144	111	134	150	93	–	7	II
22	122	143	110	135	150	94	–	2	III	48	123	144	111	135	150	92	1	1	I, II
23	122	143	110	135	150	95	–	2	III	49	123	144	111	135	150	93	137	7	I, II
24	122	143	110	136	150	92	–	1	III	50	123	144	111	135	151	93	–	2	I
25	122	143	110	136	150	93	21	10	III	51	123	144	111	136	150	93	–	2	I
26	122	143	110	136	150	94	1	4	III	52	123	144	112	134	150	93	–	3	II

Table 4 Distribution of cpSSR haplotypes of cultivated soybean in different regions of Asia

Region	cpDNA type/cpSSR haplotype							
	Type III						Type I and Type II	
	no.20	no.21	no.25	no.26	no.29	no.34	no.48	no.49
Northern Japan (21[a])						4		17
Southern Japan (20)	6							14
Korean peninsula (19)	2		4			1		12
Northeast China and far-eastern Russia (19)			7					12
Huang-Huai-Hai River Valley of China (21)			5					16
Changjiang River Valley of China (32)			1		6			25
Southern China (23)			2	1	1		1	18
Southeast Asia[b] (15)					1			14
South and Central Asia[c] (13)	1	1	2					9
Total (183)	9	1	21	1	8	5	1	137

[a] Number in parentheses indicate the number of accessions tested
[b] Vietnam, Indonesia, Thailand, and Myanmar
[c] India, Bhutan, Nepal, Pakistan, and Kirgyz

Geographical distribution of cpSSR haplotypes in the cultivated soybean

The distributions of the eight cpSSR haplotypes detected in the cultivated soybean are given in Table 4. All of the cultivated accessions with the type-I and type-II chloroplast genomes, except for one accession from Yunnan province of southern China (Mengzidaqingdou), have haplotype no. 49, the most predominant haplotype in the cultivated soybean (74.9%). Mengzidaqingdou, which had the type-I chloroplast genome, had haplotype no. 48, which differed from haplotype no. 49 in a one-step mutation at SOYCP (Table 3). On the other hand, the remaining six haplotypes, which all had the type-III chloroplast genome, were distributed in various regions of Asia. Haplotypes nos. 20 and 34 were mainly observed in Japan and the Korean peninsula: haplotype no. 20 was mainly distributed in southern Japan, whereas haplotype no. 34 was mainly distributed in northern Japan. Haplotype no. 25 was observed in relatively high frequencies in the Korean peninsula, northeast China and far-eastern Russia, and the Huang-Huai-Hai Valley area of China, and haplotype no. 29 was mainly distributed in the Changjiang River Valley area of China. Of the two remaining haplotypes, no. 21 was unique to an Indian accession (PI374189) and no. 26 was unique to a Chinese accession from Yunnan province of southern China (Yuxihuangdou).

Geographical distribution of cpSSR haplotypes in the wild soybean

The geographical distributions of the cpSSR haplotypes in the wild soybean are presented in Fig. 3 for the acces-

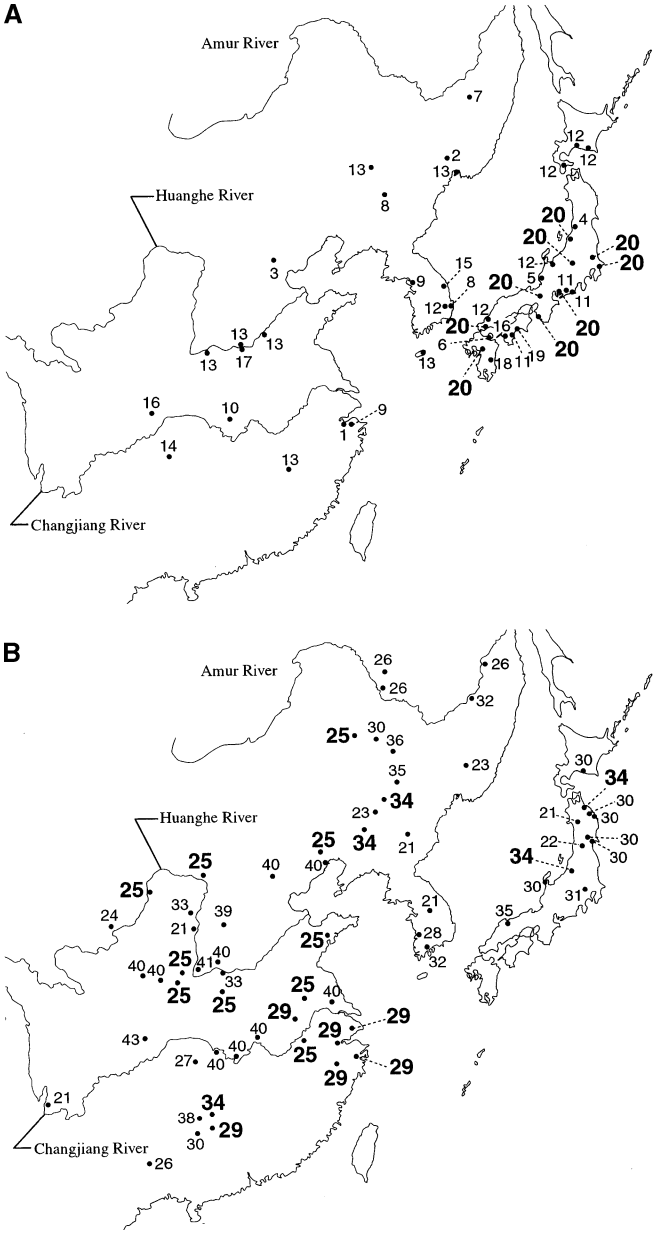

Fig. 3 Geographical distributions of 40 haplotypes in the wild soybeans with the type-III chloroplast genome. Bold figures indicate region-specific haplotypes that were observed in the cultivated soybean. A: Haplotypes no. 1 to no. 20. B: Haplotypes no. 21 to no. 40

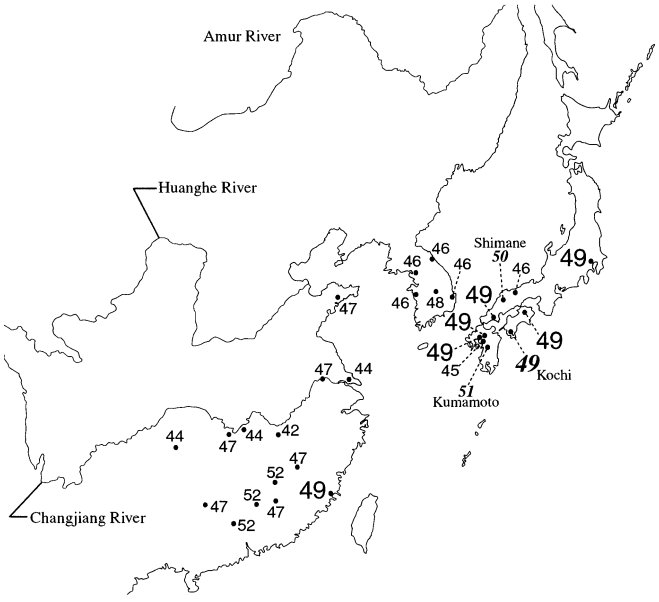

Fig. 4 Geographical distributions of 12 haplotypes in the wild soybeans with the type-I and type-II chloroplast genomes. Bold figures indicate the haplotype that was predominantly observed in the cultivated soybean. Italic figures indicate the accessions with the type-I chloroplast genome

sions with the type-III chloroplast genome and in Fig. 4 for those with the type-I and type-II chloroplast genomes. In addition to the 26 unique haplotypes, 9 of the 12 haplotypes with frequencies of more than 3% (nos. 12, 13, 20, 25, 29, 40, 46, 47, and 49) tended to be distributed regionally in various regions of the distribution area of the species. The most frequent haplotype (no. 25) was mainly distributed in the Huang-Huai-Hai Valley area of China. Another two major haplotypes, nos. 20 and 40, were observed in central to southern Japan and the Huang-Huai-Hai Valley and Changjiang Valley areas of China, respectively. On the contrary, haplotypes nos. 21, 30 and 34 were observed in relatively separated regions. Each of the four region-specific haplotypes of the cultivated soybean (nos. 20, 25, 29, and 34) was observed in the wild accessions that were collected from the areas nearly sympatric to the distribution of the cultivated accessions having the same haplotype: nine wild accessions of central to southern Japan had haplotype no. 20, seven wild accessions of the Huang-Huai-Hai Valley area of China had haplotype no. 25, six wild accessions of the Changjiang River Valley area of China had haplotype no. 29, and two wild accessions of northern Japan had haplotype no. 34 (Fig. 3).

The predominant haplotype among the cultivated accessions (no. 49) was detected only in six wild accessions from southern Japan and one accession from the Fujian province of southern China (Table 3, Fig. 4). In particular, the combination of haplotype no. 49 and the type-I chloroplast genome, which predominated in the cultigen, was observed only in one wild accession collected in Kochi prefecture of southern Japan. The other wild accessions with the type-I chloroplast genome that were collected from the Shimane and Kumamoto prefectures of southern Japan possessed haplotypes nos. 50 and nos. 51, respectively.

Discussion

Diversity at cpSSRs in wild and cultivated soybeans

The present study revealed that a rich variability is present at the six SSRs of the soybean chloroplast genome. All of the SSRs tested were polymorphic, with three to six variants in a total of 326 accessions analyzed, and had an average genetic diversity of 0.496 in the wild soybean and 0.269 in the cultigen. Furthermore, combining the variants at the six SSRs generated 52 haplotypes (H=0.964) in the wild soybean and eight haplotypes (H=0.421) in the cultigen. The high variability in the chloroplast SSRs is in sharp contrast to the limited variability for base substitutions and insertions/deletions in wild and cultivated soybeans that have been detected with RFLP and sequencing analyses (Close et al. 1989; Shimamoto et al. 1992, 1998; Kanazawa et al. 1998; Abe et al. 1999; Xu et al. 2000). In 3,849 bases identified in nine intergenic spacer regions of the chloroplast genome, Xu et al. (2000) found only 11 base substitutions and a 5-base deletion in 19 wild and cultivated soybean accessions tested. The high variability in the chloroplast SSRs may therefore enable us to more clearly resolve phylogenetic relationships not only among the different chloroplast genome types but also between wild and cultivated soybeans. Sequences of the whole chloroplast genome remain undetermined in soybean. As more data on cpDNA sequence are accumulated, more cpSSR markers may become available for studies on the evolutionary mechanism in soybean and its wild progenitor.

Geographical distribution of cpSSR haplotypes in the wild soybean

Fifty-two haplotypes were observed in the 143 wild soybean accessions tested. Most of the haplotypes exhibited regional distributions, suggesting extensive geographical differentiations at the level of chloroplast SSR haplotypes in the wild soybean. Non-random distributions of organelle genome types in the wild soybean have also been reported for two cpSSRs (RD19 and SOYCP; Powell et al. 1995b, 1996), and chloroplast and mitochondrial RFLPs (Tozuka et al. 1997; Shimamoto et al. 1998). The regional distributions of organelle genome types may be due to the fact that the wild soybean possesses no mechanism of long-distance seed dispersal, although the natural populations within regions are often subject to dispersal by human intervention and natural disturbances such as flooding (Abe 2000).

Origins of cultivated soybeans with different cpSSR haplotypes

Eight cpSSR haplotypes were found in the 183 cultivated accessions collected from various countries of Asia. All 52 of the haplotypes were present in the wild soybean, indicating a considerably high genetic diversity in the wild soybean relative to the cultigen. Therefore, the present study suggests that a severe 'cytoplasmic bottleneck' occurred during the domestication of soybean, as has been reported in barley (Neale et al. 1986; Provan et al. 1999), cowpea (Vaillancourt and Weeden, 1992), lentil (Ladizinsky 1999), onion (Friesen et al. 1999), and rice (Provan et al. 1997)

Haplotype no. 49 predominated in all of the regions, accounting for around 75% of the cultivated accessions. This haplotype was observed only in the accessions having the type-I and type-II chloroplast genomes. Our previous RFLP analysis of cpDNAs revealed that the cultivated soybean accessions of the type-I chloroplast genome may have been established from wild plants having the type-I chloroplast genome or those having the type-II genome via a single-base substitution at the *Cla*I site in the coding region of *rps*3 which discriminated type I from type II. In the wild soybean, the type-II chloroplast genome, which is an intermediate between type I and type III (Kanazawa et al. 1998), is presently distributed in a relatively wide area covering the lower Changjiang River region and southern China, the Korean peninsula, and central to southern Japan, while the type-I genome is restrictedly distributed in Japan (Shimamoto et al. 1998; Abe et al. 1999; our unpublished data). Our studies thus suggest that the cultivated soybean with the type-I chloroplast genome originated somewhere in the distribution area of the wild soybeans having the type-I or type-II chloroplast genomes.

The present cpSSR analysis enabled us to further subdivide the accessions of the type-II chloroplast genome into nine haplotype groupings, and those of the type-I chloroplast genome into four haplotype groupings. Of particular interest is the finding that all of the 138 cultivated accessions with the type-I and type-II chloroplast genomes, except for one accession from southern China, possessed haplotype no. 49. This suggests that the homeland of the cultivated soybeans with the type-I and type-II chloroplast genomes is limited to the distribution area of wild plants having haplotype no. 49, which is presently found in southern Japan and southern China. Furthermore, the combination of haplotype no. 49 and the type-I chloroplast genome, the most predominant chloroplast genome type in the cultivated soybean, was observed only in one wild accession (from southern Japan) of the 143 wild accessions tested. The chloroplast genome type characteristic of the present cultivated soybean population may therefore have originated in a rare type of the wild soybean.

In contrast to the wide distribution of haplotype no. 49, each of the four haplotypes with low but significant frequencies (nos. 20, 25, 29, and 34) was observed in different regions of Asia. All of the four haplotypes were different in at least three cpSSRs from haplotype no. 49 and were restricted to the accessions having the type-III chloroplast genome. Since the type-III chloroplast genome differs by at least five base substitutions from the type-I and type-II genomes (Xu et al. 2000, 2001), it is unlikely that the four haplotypes were derived from the predominant haplotype (no. 49) through the accumulation of mutations in the cultivated soybean, although size homoplasy was found between the different chloroplast genome types (Table 3).

Interestingly, for each of the four region-specific haplotypes found in the cultivated soybean, the same respective haplotype was also observed in the wild soybean accessions collected in almost the same area where the cultivated soybeans of the respective haplotypes predominated. Of the four haplotypes, only haplotype no. 34, which was unique to the cultivated accessions of northern Japan, was distributed separately in northern Japan, northeast China, and southern China. However, only the wild accessions from northern Japan possessed the same mitochondrial genome type found in the cultivated soybeans with haplotype no. 34 (our unpublished data). The present study thus suggests that the cultivated accessions with region-specific cpSSR haplotypes had been established in different regions from wild plants having the respective haplotypes. This possibility is supported by the findings obtained from the RFLP analyses for cpDNAs and mtDNAs, which revealed an overlapping of the distributions of organelle genome types between the wild and cultivated soybeans (Shimamoto et al. 1998; 2000; Shimamoto 2001). The overlapping distribution of the cpSSR haplotypes, and cpDNA and mtDNA RFLPs between the wild and cultivated soybeans provides two hypotheses that are not necessarily exclusive: (1) the cultivated soybeans had been domesticated independently from different wild gene pools and (2) the introgression of organelle genomes from the wild to cultivated soybeans had repeatedly occurred where the monophyletically-originated cultivated soybean had been disseminated. A comparative study may be needed for the genetic bases of domestication syndromes that discriminate between wild and cultivated forms, such as seed hardness, seed size, seed shattering, and growth habits, to determine the relative significance of the two events on the evolutionary process of soybean.

In conclusion, the results obtained in the present study demonstrate that the diversity of cpSSRs in the cultivated soybean resulted from multiple establishments of cultivated forms from different wild gene pools or hybrid swarms. The repeated gene flows from wild to cultivated gene pools that occurred as indicated by the present study further suggest that soybean possesses a relatively high genetic diversity in its nuclear genome, although the genetic poverty has often been stressed in improved cultivar populations (Keim et al. 1989; Kisha et al. 1998). It may be necessary to assess the diversity of the nuclear genome in the cultivated soybean, in particular to focus on landraces and local varieties that have been established in Asia.

Acknowledgements This research was supported by the Japan Society for the Promotion of Science (JSPS) (Grant no. PI98120 and no. 12460142) and the Heiwa Nakajima Foundation.

References

Abe J (2000) The genetic structure of natural populations of wild soybean revealed by isozymes and RFLPs of mitochondrial DNAs: possible influence of seed dispersal, cross-pollination and demography. In: Oono K (ed) 7th MAFF Int Workshop Genet Resourc. AFFRC and NIAR, Tsukuba, Japan, pp 143–158

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

Bryan GJ, McNicoll J, Ramssay G, Meyer RC (1999) Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor Appl Genet 99:859–867

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

Friesen N, Pollner S, Bachmann K, Blattner FR (1999) RAPDs and noncoding chloroplast DNA reveal a single origin of the cultivated *Allium fistulosum* from *A. altaicum* (*Alliaceae*). Am J Bot 86:554–562

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

Ishii T, McCouch SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. Theor Appl Genet 100:1257–1266

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

Keim P, Schoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Kisha TJ, Diers BW, Hoyt JM, Sneller CH (1998) Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci 38:1669–1680

Ladizinsky G (1999) Identification of the lentil's wild genetic stock. Genet Resour Crop Evol 46:115–118

Neale DB, Saghai-Maroof MA, Allard RW, Zhang Q, Jorgensen RA (1986) Chloroplast DNA diversity in populations of wild and cultivated barley. Genetics 120:1105–1110

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995a) Polymorphic simple sequence repeat regions in chloroplast genomes: application to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–7763

Powell W, Morgante M, Andre C, McNicol JW, Machray GC, Doyle JJ, Tingey SV, Rafalski JA (1995b) Hypervariable microstatellites provide a general source of polymorphic DNA markers for the chloroplast genome. Curr Biol 5:1023–1029

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV, Rafalski AJ (1996) Genepool variation in genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144:793–803

Provan J, Corbett G, McNicol JW, Powell W (1997) Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40:104–110

Provan J, Russell JR, Booth A, Powell W (1999) Polymorphic chloroplast simple sequence repeats primers for systematic and population studies in the genus *Hordeum*. Mol Ecol 8:505–511

Reverdatto SV, Beilinson V, Nielsen NC (1995) The *rps*16, *acc*D, *psa*I, ORF 203, ORF 151, ORF 103, ORF 229, and *pet*A gene cluster in the chloroplast genome of soybean (GenBank U26948). Plant Physiol 109:338

Shimamoto Y (2001) Polymorphism and phylogeny of soybean based on chloroplast and mitochondrial DNA analysis. Jpn Agric Res Quart 35:79–84

Shimamato 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 FS (2000) Characterizing the cytoplasmic diversity and phyletic relationship of Chinese landraces of soybean, *Glycine max*, based on RFLPs of chloroplast and mitochondrial DNA. Genet Resour Crop Evol 47:611–617

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

Tozuka A, Fukushi H, Hirata T, Ohara M, Kanazawa A, Mikami T, Abe J, Shimamoto Y (1998) Composite and clinal distribution of *Glycine soja* in Japan revealed by RFLP analysis of mitochondrial DNA. Theor Appl Genet 96:170–176

Vaillancourt RE, Weeden NF (1992) Chloroplast DNA polymorphism suggests Nigerian center of domestication for the cowpea, *Vigna unguiculata* (Leguminosa). Am J Bot 79:1194–1199

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

Xu DH, Abe J, Kanazawa A, Gai JY, Shimamoto Y (2001) Identification of sequence variations by PCR-RFLP and its application to the evaluation of cpDNA diversity in wild and cultivated soybeans. Theor Appl Genet 102:683–688