JARQ 35(2), 79 – 84 (2001) http://www.jircas.affrc.go.jp

Polymorphism and Phylogeny of Soybean Based on Chloroplast and Mitochondrial DNA Analysis

Yoshiya SHIMAMOTO*

Graduate School of Agriculture, Hokkaido University
(Sapporo, Hokkaido, 060–8589 Japan)

Abstract

Restriction fragment length polymorphisms (RFLPs) of chloroplast and mitochondrial DNAs of various soybean landraces and wild soybeans collected in China and Japan were analyzed to characterize their cytoplasmic genome and to consider the polymorphism and phylogeny of soybean from the maternal lineage. Three chloroplast and 6 mitochondrial types were distinguished and, in their combinations, 8 cytoplasmic genome types were detected in specimens of soybean landraces collected from China and Japan. The cytoplasmic genome diversity was higher in China, especially in the Yangtze River Valley. The main group of soybean landraces including current cytoplasmic types was composed of the 5 cytoplasmic types which were related phylogenetically to each other and were considered to occur mainly in the Yangtze River Valley of China, though one of them seemed to have differentiated in Japan. Three unique types of soybean cytoplasmic genome were markedly different in the genome constitution from each other, and considered to be derived separately from wild soybean with the same cytoplasmic genome and to have occurred independently in remote regions, i.e. the northern region and the Yangtze River Valley of China and the southern and northern regions of Japan, respectively. Based on the constitution of the cytoplasmic genome, it is suggested that domestication of soybean may have possibly occurred in various areas and shows multiple origins.

Discipline: Genetic resources
Additional key words: cytoplasmic genome, diversity, multiple origin, RFLPs, soybean landraces

Introduction

Soybean, *Glycine max* (L.) Merr., is the most important grain legume crop in the world from the standpoint of total production and international trade. World annual production of soybean amounts to about 160 Mt and ranks fourth after maize, wheat and rice. Soybean had been cultivated mainly in East Asia several decades ago. At present more than 80% of the world soybean production is located in the New World far from East Asia, its center of origin.

It had been commonly considered that soybean was domesticated in some areas of China, presumably the northern and central regions, about 4,000 years ago, from a wild soybean species, *G. soja* Sieb. & Zucc., as progenitor[10,11]. The wild soybean is densely distributed in China, Korea, Japan and the Far East region of Russia, without mountainous area. *G. max* and *G. soja* are included in the subgenus *Soja* in the genus *Glycine* and are growing occasionally side by side and can hybridize sexually to produce fertile offspring. People living in China and Japan may have utilized wild soybeans in prehistoric times, as well as used generally various traditional foods made from soybean and have so far conserved many landraces of soybean, respectively.

Polymorphism of the cytoplasmic genome in plants has been recognized based on the analysis of restriction fragment length polymorphisms (RFLPs) for organelle DNA and identified in soybean for the diversity in the subgenus *Soja* in terms of chloroplast DNA[2,9,20,22] and mitochondrial DNA[4–6,8,17,20.23,24]. This molecular information about the cytoplasmic genome has enabled to elucidate the phylogenetic relationships and the evolutionary process of soybeans[3,7].

We have carried out studies on the diversity of the cytoplasmic genome of landraces (*G. max*) and wild specimens (*G. soja*) of soybean worldwide. The objective of the present paper was to examine the center of diversity, differentiation, phylogeny and some evolution-

---

*Corresponding author: fax+81–11–706–4933, e-mail yshimamo@res.agr.hokudai.ac.jp

Received 22 March 2000, accepted 7 April 2000.

ary aspects of the cultivated soybean from the Chinese and Japanese landraces, in taking account of the information from the wild soybean based on polymorphisms of the cytoplasmic genome.

Diversity based on cytoplasmic genomes among soybean landraces

In the profile combination of 3 chloroplast(cp) and 6 mitochondrial (mt) RFLPs, 8 types, cpI+mtIIIb (cytoplasmic genome type with type I chloroplast and type III in the *cox2* region and type b in the *atp6* region of mitochondria), cpI+mtIVb, cpI+mtIVc, cpII+mtIVc, cpII+mtIVb, cpIII+mtIe, cpIII+mtIVa and cpIII+mtVIIIc, designated according to the terminology of Shimamoto et al.[20], were detected in a cytoplasmic genome survey of landraces of soybean collected from China[21] and Japan (partly cited from Hirata et al.[8]), as shown in Table 1.

Based on the collection records of landraces, the specimens were grouped geographically into 4 regions of China and 3 regions of Japan, respectively. Table 2 shows the frequency distribution of the cytoplasmic types of soybean for the 7 regions. Seven and 6 of the 8 types were found in China and Japan, respectively. Four of the 8 types have been detected in Korean soybean landraces (unpublished data). The cytoplasmic genome type with cpI+mtIVb was fairly predominant over East Asia and is found in most of the cultivars improved.

The Chinese germplasm of soybean landraces has conserved most of the types except for cpIII+mtVIIIc and a greater diversity (H value of Table 2) of cytoplasmic genomes, compared with the Japanese germplasm where 6 types without cpI+mtIIIb and cpIII+mtIVa are conserved. This is particularly true in the Yangtze River Valley, which could be a center of diversity of the cytoplasmic genome of soybean[21]. Also Shimamoto et al.[20] observed that the wild soybeans of the Yangtze River Valley had conserved a higher diversity than those in other regions of China and Japan as shown in the H value listed in Table 2.

Phylogenetic relationship among the cytoplasmic genome types

These 8 cytoplasmic types of soybean landraces were divided into 3 groups with different chloroplast genome types, cpI, cpII and cpIII, which were composed of 3, 2 and 3 types of mitochondrial genome, respectively (Table 1). Kanazawa et al.[15], based on RFLPs of 3 restriction enzymes with 6 mitochondrial gene probes, showed that 2 mitochondrial types, the predominant type mtIVb followed by mtIVc, with both cpI and cpII, exceptionally, and IIIb with cpI, were closely related to each other and differed at several sites of the mitochondrial genome from other 3 ones, mtIe, mtIVa and mtVIIIc carrying cpIII. Therefore, mtIIIb, mtIVb and mtIVc which formed the main current cultivar group and accounted for 86% of the Chinese and Japanese landraces, as shown in Table 2, may differ in the phylogenetic process from the 3 unique types, mtIe, mtIVa and mtVIIIc carrying cpIII.

Evolutionary process of differentiation among the cytoplasmic genome types with cpI or cpII

In most of the Chinese specimens of wild soybean, monomorphism has been observed in the chloroplast genome type among the plants which carry the same mitochondrial genome types[20]. The differences among the mitochondrial genome types are mainly attributed to recombination events[12,16] or selective amplification (or

Table 1. Polymorphism of cytoplasmic genome types combined with chloroplast and mitochondrial genome profiles in soybean landraces

Cytoplasmic genome type	Probe	Chloroplast		Mitochondria		
		H2 clone		*cox2*		*atp6*
	enzyme	*Eco*RI	*Cla*I	*Hin*dIII	*Bam*HI	*Bam*HI
cpI+mtIIIb		4.8	1.1, 2.4	1.2	8.5	2.9, 5.0
cpI+mtIVb		4.8	1.1, 2.4	3.5	8.1	2.9, 5.0
cpI+mtIVc		4.8	1.1, 2.4	3.5	8.1	5.0
cpII+mtIVb		4.8	3.5	3.5	8.1	2.9, 5.0
cpII+mtIVc		4.8	3.5	3.5	8.1	5.0
cpIII+mtIe		2.5	3.5	1.6	5.8	5.0, 12.0
cpIII+mtIVa		2.5	3.5	3.5	8.1	2.4, 5.0
cpIII+mtVIIIc		2.5	3.5	8.5, 10.0	11.0, 15.0	5.0

Note: Fragment sizes (kb) of hybridization signals are shown.

Table 2. Frequency distribution(%) and genetic diversity (H value) of cytoplasmic genome types combined with chloroplast and mitochondrial genome types detected in soybean landraces and their corresponding cytoplasmic types in wild soybean

Nation	No. of samples	Cytoplasmic genome type									Other types	H[c] value
Region		cp[a]	I	I	I	II	II	III	III	III		
		mt[b]	IIIb	IVb	IVc	IVb	IVc	Ie	IVa	VIIIc		
China[d]												
landrace	338		1.2	53.8	3.0	3.9	25.4	5.3	7.4	–	–	.635
wild soybean	753		–	–	–	22.3	1.2	3.3	55.2	–	18.0	.638
Northeastern region												
landrace	36		–	55.6	–	5.5	25.0	–	13.9	–	–	.606
wild soybean	213		–	–	–	.9	–	–	88.7	–	10.4	.205
Huang River Valley												
landrace	112		1.8	58.9	2.7	3.6	19.6	–	13.4	–	–	.594
wild soybean	188		–	–	–	10.1	–	–	73.4	–	16.5	.431
Yangtze River Valley												
landrace	144		1.4	45.5	3.5	3.5	31.2	12.5	1.4	–	–	.668
wild soybean	224		–	–	–	29.9	1.8	11.2	29.0	–	28.1	.799
Southern region												
landrace	46		–	63.0	4.4	4.4	21.7	–	6.5	–	–	.548
wild soybean	128		–	–	–	62.5	3.9	–	18.8	–	14.8	.555
Japan[e]												
landrace	180		–	68.3	10.6	1.1	6.1	7.2	–	6.7	–	.509
wild soybean	704		.4	1.3	.9	3.4	–	.3	19.9	*	73.8	.791
Northern region												
landrace	63		–	58.7	12.7	1.6	4.8	20.6	–	1.6	–	.594
wild soybean	449		–	.4	–	2.7	–	.2	26.5	–	71.0	.732
Central region												
landrace	45		–	75.6	6.7	2.2	11.1	–	–	4.4	–	.409
wild soybean	88		–	1.1	–	1.1	–	1.1	21.6	–	75.1	.674
Southern region												
landrace	72		–	72.2	11.1	–	4.2	–	–	12.5	–	.449
wild soybean	167		1.8	3.6	3.6	6.6	–	–	1.2	–	83.2	.562

a) : Chloroplast genome type,
b) : Mitochondrial genome type,
c) : Calculated by $1 - \Sigma pi^2$, where pi is the proportion of ith cytoplasmic genome type,
d) : Data of landraces[21] and wild soybean[20],
e) : Data of landraces[8], in addition to unpublished data and data of wild soybean from mitochondrial type[24], were combined with the chloroplast type[1],
– : Not detected,
* : A few plants were found (personal information).

loss) of subgenomic mitochondrial DNA molecules[15]. Meanwhile, differences among chloroplast genome types are considered to be due to single base substitutions[14]. If point mutations at the same position of chloroplast DNA are less likely to occur repeatedly than structural changes based on recombination in mitochondrial DNA during the evolutionary process of domestication of soybean, as described by Kanazawa et al.[13], the differences between mtIVb and mtIVc, probably due to a single event, may have occurred after mutation (single base substitution) in the chloroplast genome, leading to a distinction between cpI and cpII, as shown by the thick solid line in Fig.1. If cpII with mtIVb changed to cpI with mtIVb, mtIVc and mtIIIb with cpI could have been derived from cpI+mtIVb through a single event in the *atp6* and *cox2* regions of mitochondrial DNA, respectively. The change to cpI as a source of landrace chloroplast genome might have occurred in the landraces with cpII+mtIVb.

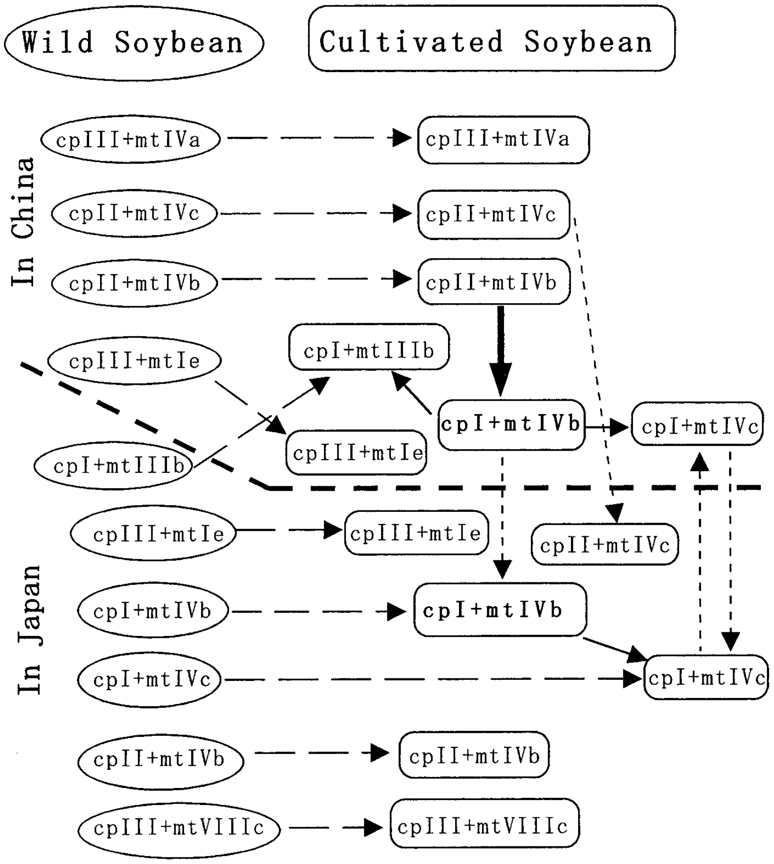

Fig. 1. Putative pathways and areas of soybean domestication based on cytoplasmic genome analysis of soybean landraces in relation to wild soybeans

Broken lines indicate the domestication to cultivar from the same cytoplasmic genome. A thick-solid line indicates a mutation on the chloroplast genome from cpII to cpI. Thin-solid line indicates a change of mitochondrial genome to the genome type (arrow point) through single event in the *atp6* or *cox2* region of mitochondrial DNA. Each thin-dotted line indicates the same genome type (arrow point) through possible dissemination between China and Japan.

The cpI+mtIVb type is the cytoplasmic genome type of most of the current cultivars, and its wild form has been found sporadically in Japan but not in China, as shown in Table 2. The wild cpII+mtIVb form which was predominant in China could have changed to the cultivated cpI+mtIVb with the change of cpII to cpI through cpII+mtIVb during domestication in the southern part of the Yangtze River Valley where the wild cpII+mtIVb form predominantly grew, as shown in Table 2.

The landrace with cpI+mtIVc which was commonly observed everywhere in Japan, was scarcely found in China (Table 2) and has not been detected so far in Korea (unpublished data). Hirata et al.[8] showed that all the specimens identified as cpI+mtIVc were the summer type which corresponded to the early maturing group of Japanese landraces distributed in the southern area of Japan and had been designated as 'NATSU-DAIZU' in Kyushu[18]. This cytoplasmic genome may have been derived from the common cultivar type cpI+mtIVb through single event in the *atp6* region in Kyushu, Japan, or/and in the southern part from the Huang River Valley of China where the landrace with this type was distributed.

The landrace with cpI+mtIIIb could have been derived from cpI+mtIVb through a single event in the *cox2* region. This cytoplasmic genome type as well as mtIVb and mtIVc carrying cpI was occasionally observed in wild soybean collected in Japan (Table 2), which might be used as direct progenitors of the cultivars, though these wild forms could be the progenies from the cultivars crossed with wild soybean[1].

The soybean landrace with cpII+mtIVc was observed all over East Asia and more frequently in China than in Japan. Therefore it is considered that this type was differentiated in China, possibly in the southern regions of China, because the putative ancestral type, wild cpII+mtIVc, had been distributed at least in the southern part from the Yangtze River Valley. The landrace with cpII+mtIVc also may have been derived from the cpII+mtIVb type through a single event in the *atp6* region.

Origin of 3 cytoplasmic genome types with cpIII in landraces

The 3 types, mtIe, mtIVa and mtVIIIc carrying cpIII, differed from each other at several sites in their mitochondrial genome[15]. Therefore, several steps may be required for a change in the 3 mitochondrial genomes of soybean carrying cpIII from one type to another. So far, each cytoplasmic genome with cpIII appears to be a unique group of soybean landraces, which may have evolved on its own and have a specific progenitor or origin.

Both landraces and wild soybeans with cpIII+mtIe were observed only in the Yangtze River Valley of China and the northern region of Japan (Table 2). The soybean landraces with cpIII+mtIe may have been directly derived from the wild soybean with cpIII+mtIe in the Yangtze River Valley of China and the northern part of Japan, as shown by the 2 broken lines in Fig.1.

The landrace with cpIII+mtIVa was observed more frequently in the northeastern part and the Huang River Valley than in the Yangtze River Valley and the southern part of China, and was detected frequently in Korea (unpublished data) and not in Japan (Table 2). Also the wild soybean with cpIII+mtIVa was found predominantly in the northeastern part and the Huang River Valley of China (Table 2) and in South Korea[19]. This cytoplasmic genome of the landrace may have been derived directly from a wild soybean with the same cytoplasmic genome as that distributed in the northeastern part or the Huang River Valley of China and possibly in Korea.

The soybean type with cpIII+mtVIIIc was found only in the autumn type corresponding to the late maturing group of Japanese landraces which was designated as 'AKI-DAIZU' in Kyushu, Japan[18]. A few wild soybeans with cpIII+mtVIIIc have been found in Japan (personal information) but not, so far, in natural habitats of China (Table 2) and South Korea[19]. Many specimens with this type were detected in fodder or manure soybean accessions which were improved in Japan and showed a growth habit similar to that of wild soybean. Soybean with this cytoplasmic type may have originated in Kyushu or the southern region of Japan and may have been derived from a wild soybean with cpIII+mtVIIIc.

Conclusion

In conclusion, a hypothetical illustration of the domestication of soybeans based on their cytoplasmic genomes is shown in Fig. 1. Several events of domestication shown by the broken lines in Fig. 1 may have led to the 8 cytoplasmic types of soybean landraces. The main groups of soybean landraces, including cpI+mtIIIb, cpI+mtIVb, cpI+mtIVc, cpII+mtIVc and cpII+mtIVb, may have evolved through a few events from the wild form of cpII+mtIVb. However, further studies should be carried out to elucidate the evolutionary process. The unique groups, cpIII+mtIe, cpIII+mtIVa and cpIII+mtVIIIc, have a different origin from the main group and each other, and became domesticated in different areas from those with the same cytoplasmic types of the wild form. Based on the analysis of the cytoplasmic genomes of soybean landraces and wild soybean accessions, soybean domestication may have been achieved in several regions of East Asia and thus has multiple origins.

References

1) Abe, J. et al. (1999): Introgression between wild and cultivated soybeans of Japan revealed by RFLP analysis for chloroplast DNAs. *Econ. Bot.*, **53**, 285–291.
2) Close, P. S., Shoemaker, R.C. & Keim, P. (1989): Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*. *Theor. Appl. Genet.*, **77**, 768–776.
3) Corriveau, J. L. & Coleman, A. W. (1988): Cytological evidence for maternal inheritance of plastid DNA in the genus *Glycine* subgenera *Glycine* and *Soja*. *Soybean Genet. Newsl.*, **15**, 89–92.
4) Grabau, E. A. et al. (1989): Restriction fragment length polymorphism in subclass of the 'Mandarin' soybean cytoplasm. *Crop Sci.*, **29**, 1554–1559.
5) Grabau, E. A. et al. (1992): Classification of soybean cultivars based on mitochondrial DNA restriction fragment polymorphisms. *Crop Sci.*, **32**, 271–274.
6) Hanlon, R. & Grabau, E. A. (1995): Cytoplasmic diversity in old domestic varieties of soybean using two mitochondrial markers. *Crop Sci.*, **35**, 1148–1151.
7) Hatfield, P. M., Shoemaker, R. C. & Palmer, R. G. (1985): Maternal inheritance of chloroplast DNA within the genus *Glycine* subgenus *Soja*. *J. Hered.*, **76**, 373–374.
8) 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.
9) Hirata, T., Abe, J. & Shimamoto, Y. (1999): Genetic structure of the Japanese soybean population. *Genet. Resour. Crop Evol.*, **46**, 441–453.
10) Hymowitz, T. (1970): On the domestication of the soybean. *Econ. Bot.*, **23**, 408–421.

11) Hymowitz, T. & Newell, C. A. (1981): Taxonomy of the genus *Glycine*, domestication and uses of soybeans. *Econ. Bot*., **35**, 272–288.

12) Kanazawa, A. et al. (1998): Small interspersed sequences that serve as recombination sites at the *cox2* and *atp6* loci in the mitochondrial genome of soybean are widely distributed in higher plants. *Curr. Genet*., **33**, 188–198.

13) Kanazawa, A. et al. (1998): Evolutionary relationships between changes in the mitochondrial and chloroplast genomes in the genus *Glycine* subgenus *Soja*. *Soybean Genet. Newsl*., **25**, 93–94.

14) 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.

15) Kanazawa, A. et al. (1998): Phylogenetic relationships of the mitochondrial genomes in the genus *Glycine* subgenus *Soja*. *Genes Genet. Syst*., **73**, 255–261.

16) Kato, S. et al. (1998): Evolutionary changes in the structures of the *cox2* and *atp6* loci in the mitochondrial genome of soybean involving recombination across small interspersed sequences. *Curr. Genet*., **34**, 303–312.

17) Moeykens, C. A., Mackenzie, S. A. & Shoemaker, R. C. (1995): Mitochondrial genome diversity in soybean ; repeats and rearrangements. *Plant Mol. Biol*, **29**, 245–254.

18) Nagata, T. (1960): Studies on the differentiation of soybeans in Japan and the world. *Mem. Hyogo Univ. Agric., Agron. Ser*., **4**, 63–102.

19) Shimamoto, Y. et al. (1997): Polymorphisms and differentiations of cytoplasmic genome in wild soybean growing in Korea. *In* Proceeding of the 8th SABRAO General Congress and the Annual Meeting of the Korean Breeding Society, 411–412.

20) Shimamoto, Y. et al. (1998): RFLPs of chloroplast and mitochondrial DNA in wild soybean, *Glycine soja*, growing in China. *Genet. Resour. Crop Evol*., **45**, 255–261.

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

22) Shoemaker, R. C. et al. (1986): Chloroplast DNA variation in the genus *Glycine* subgenus *Soja*. *J. Hered*., **77**, 26–30.

23) Sisson, V. A., Brim, C. A. & Levings III, C. S. (1978): Characterization of cytoplasmic diversity in soybean by restriction endonuclease analysis. *Crop Sci*., **18**, 991–996.

24) Tozuka, A. et al. (1998): Composite and clinal distribution of *Glycine soja* in Japan revealed by RFLP analysis of mitochondrial DNA. *Theor. Appl. Genet*., **96**, 170–176.