Theor Appl Genet (1989) 77:768–776

Distribution of restriction site polymorphism within the chloroplast genome of the genus *Glycine*, subgenus *Soja*

P.S. Close [1]*, R.C. Shoemaker [2] and P. Keim [3]

[1] Department of Genetics, Iowa State University, Ames, IA 50011, USA
[2] USDA-ARS, Departments of Agronomy and Genetics, Iowa State University, Ames, IA 50011, USA
[3] Department of Biology, Northern Arizona University, Flagstaff, AZ 86001, USA

Received August 15, 1988; Accepted February 15, 1989
Communicated by A.L. Kahler

Summary. Restriction fragment length polymorphisms (RFLPs) have been used to detect intragenic sequence diversity in *Glycine* subgenus *soja* chloroplast DNA. The distribution of these RFLPs allow *Glycine max* and *G. soja* accessions to be grouped according to cytoplasmic genetic relatedness. DNA clones from mung bean chloroplast DNA were used to locate the RFLPs to specific regions of the chloroplast genome. In the course of the experiments, several previously unobserved RFLPs were also identified. At least six molecular changes were detected, including both restriction site loss or gain and insertion/deletion events. Three of the fragment polymorphisms detected are due to changes in the juncture region between one inverted repeat region and the large single-copy region. Probes detecting polymorphisms in three representative soybean genotypes were used to screen additional cultivars and Plant Introductions. The distribution of RFLP patterns in these accessions were consistent with the patterns of previously described cytoplasmic groupings, with the exception of one accession, which formed a new plastome group.

Key words: Chloroplast DNA – RFLP – Soybean

Introduction

Plant chloroplasts contain 20–200 copies of closed circular DNA molecules, which range from 120–180 kbp in size. The chloroplast genome is known to contain all the chloroplast rRNA genes, tRNA genes, and probably all the genes for proteins synthesized within the chloroplast (Shinozaki et al. 1986; Dyer 1985). Most plant chloroplast DNAs (ctDNAs), including those of genus *Glycine*, are organized into four components: a large (78–100 kbp) single-copy region and a small (12–30 kbp) single-copy region, separated by two repeated sequences (20–24 kbp) arranged in an inverted orientation. This repeat is characteristic of all land plant chloroplast genomes with the exception of one group of leumes whose ctDNAs lack this repeated element (Palmer and Thompson 1981; Palmer et al. 1983b).

As a whole, the ctDNA sequence is remarkably conserved within genera and across species. Sequence divergence between related species is low. Rearrangements are rare except among some of the species lacking the inverted repeat, such as pea (*Pisum sativum*), broad bean (*Vicia faba*) (Palmer and Thompson 1982), or subclover (*Trifolium subterraneum*) (Palmer et al. 1987c); or in geranium (*Pelargonium hortorum*) (Palmer et al. 1987b), containing a much enlarged inverted repeat. Changes resulting in variation within genera and species result primarily from nucleotide substitutions, which occur at a low rate, or from small insertions or deletions of one to ten nucleotides (Curtis and Clegg 1984; Gordon et al. 1982; Palmer and Zamir 1982; Palmer et al. 1983a; Palmer 1987). Insertion/deletion events involving up to 1200 bp have also been reported (Palmer 1987), but mutations of much greater length have been reported only for *Nicotiana acuminata* (Shen et al. 1982) and *Linum* species (Coates and Cullis 1987). The changes that occur in the inverted repeats occur symmetrically in both segments (Palmer 1985).

Changes in the patterns of DNA fragments produced by digestion with site-specific restriction endonucleases reveal diversity in nucleotide sequence. Restriction fragment length polymorphism (RFLP) analysis of interspecific diversity of ctDNA is often used to delineate phylogenetic relationships between closely related genera

* To whom correspondence should be addressed

and species. However, the slow rate of change in ctDNA often limits the usefulness of ctDNA analyses for studying intraspecific variation. Although variation within a species is considered rare (Palmer 1987), Shoemaker et al. (1986) detected enough ctDNA sequence diversity within the genus *Glycine* subgenus *Soja* to delineate five groupings based upon digestions with only three restriction endonucleases (*Ava*I, *Cla*I, and *Eco*RI).

The genus *Glycine* subgenus *Soja* includes the cultivated soybean *Glycine max* (L.) Merr. and its annual wild counterpart *G. soja* Sieb. and Zucc. The subgenus *Soja* also includes a form designated *G. gracilis* (Skvortzov 1927), a semicultivated or weedy form intermediate in morphology between *G. max* and *G. soja*, and considered to be an introgression product between *G. soja* and *G. max* (Hymowitz 1970; Broich and Palmer 1981). Hermann (1962) considered *G. gracilis* to be a varient of *G. max*. Broich and Palmer (1980) demonstrated that distinct differences exist between the typical *G. max* phenotype and the *'gracilis'* form and suggested *Glycine max* forma *gracilis* (L.) Merr. *G. max, G. soja,* and *G. max.* forma *gracilis* are generally interfertile, although chromosome interchanges are known (Palmer et al. 1987a), and make up the primary gene pool of the soybean (Hymowitz and Singh 1987).

The objectives of this research were to (1) evaluate selected soybean cultivars and Plant Introductions for new ctDNA RFLPs in an attempt to further subdivide the soybean germplasm into additional cytoplasmic groupings, and (2) to physically locate these and previously identified RFLPs to specific regions of the chloroplast genome. This type of research would provide information on the availability of cytoplasmic diversity within the soybean collection and might also provide insight into the nature of ctDNA evolution within the genus *Glycine* subgenus *Soja*.

Materials and methods

Total plant DNA was isolated according to the methods of Keim et al. (1988) from three soybean genotypes ('Peking', 'Illini', and PI 79593) representing the various patterns of polymorphisms previously observed (Shoemaker et al. 1986), as well as from 53 additional accessions of *G. max* or *G. soja*. DNA was digested with Type II restriction endonucleases (New England Biolabs and BRL) *Ava*I, *Cla*I, *Eco*RI, *Dra*I, *Hind*III, *Sty*I, or *Eco*RV. Fragments were separated electophoretically on gels of 0.8% agarose in TBE (TRIS Borate EDTA) buffer at 35 mA for 16–18 h. DNA was transferred to Biotrace nylon membrane under high alkaline conditions (Reed and Mann 1985).

DNAs bound to membranes were hybridized to radioactively labelled probes prepared from a library of 13 mung bean ctDNA clones graciously provided by Dr. Jeff Palmer (University of Michigan). This library spans the length of the mung bean chloroplast genome, which is essentially colinear with that of soybean, with the exception of a 78-kbp inversion involving most of the large single-copy (LSC) region (Palmer 1987; Palmer et al. 1988). A *Pvu*II restriction site mapping at the edge of the large single-copy region, just inside of the inverted repeat (IR) region (Palmer et al. 1983b), was used to divide clone 4 into two fragments (4a and 4b). This site was also used to isolate a fragment from clone 1 specific for the inverted repeat (designated 1a). These three fragments were isolated from low-melting-point agarose (BRL no. 557UA) minigels. All probes were labelled with [^{32}P]dCTP by random hexamer primer extension (Feinberg and Vogelstein 1983, 1984). Hybridizations were carried out according to the protocol of Maniatis et al. (1982) with modifications according to Apuya et al. (1988). Autoradiograms were produced by exposing the hybridized nylon membranes to Kodak X-Omat film for 16–24 h.

Clones which detected polymorphic fragments in Peking, Illini or PI 79593 DNAs digested with *Cla*I or *Eco*RI were then used to screen the additional 53 cultivars and Plant Introductions for *Eco*RI and *Cla*I polymorphisms.

Double digests of Peking, Illini, and PI 79593 DNAs using *Cla*I + *Xho*I were probed with clones, 1a, 4a and 4b to more accurately place the polymorphic *Cla*I restriction site(s) with respect to the large single-copy and inverted repeat regions.

Results

RFLP identification and localization

In digests with *Ava*I, *Cla*I, and *Eco*RI, polymorphisms detected included those previously described (Shoemaker et al. 1986) as well as additional polymorphic fragments. However, one of the two *Ava*I polymorphic fragments previously reported (a 1.7-kbp fragment present in accessions represented by Illini and Peking but not in

Table 1. Soybean chloroplast genome polymorphic fragments produced by restriction endonuclease digestion

Mutation	RE/Probe	Pattern: fragment size (in kbp)
1	*Eco*RI/MB-10	A: 4.7*
		B: 6.4**
	*Eco*RI/MB-11	A: 4.7*+1.6
		B: 6.4**
2	*Eco*RI/MB-12	A: 2.1
		B: 1.1+1.1
3	*Eco*RI/MB-12	A: 0.52
		B: 0.55
4	*Eco*RI/MB-13	A: 4.8
		B: no 4.8 fragment
5	*Eco*RI/MB-4b	A: 4.8
		B: 2.5
6	*Cla*I/MB-4b	A: 2.4+1.1
		B: 3.5 or 3.6
7	MB-4b	A: *Cla*I 3.5; *Eco*RV 0.89; *Eco*RI 1.2; *Ava*I 4.1; *Sty*I 3.4
		B: *Cla*I 3.6; *Eco*RV 0.98; *Eco*RI 1.3; *Ava*I 4.2; *Sty*I 3.5

*,**: like symbols designate polymorphic fragments detected with more than one probe

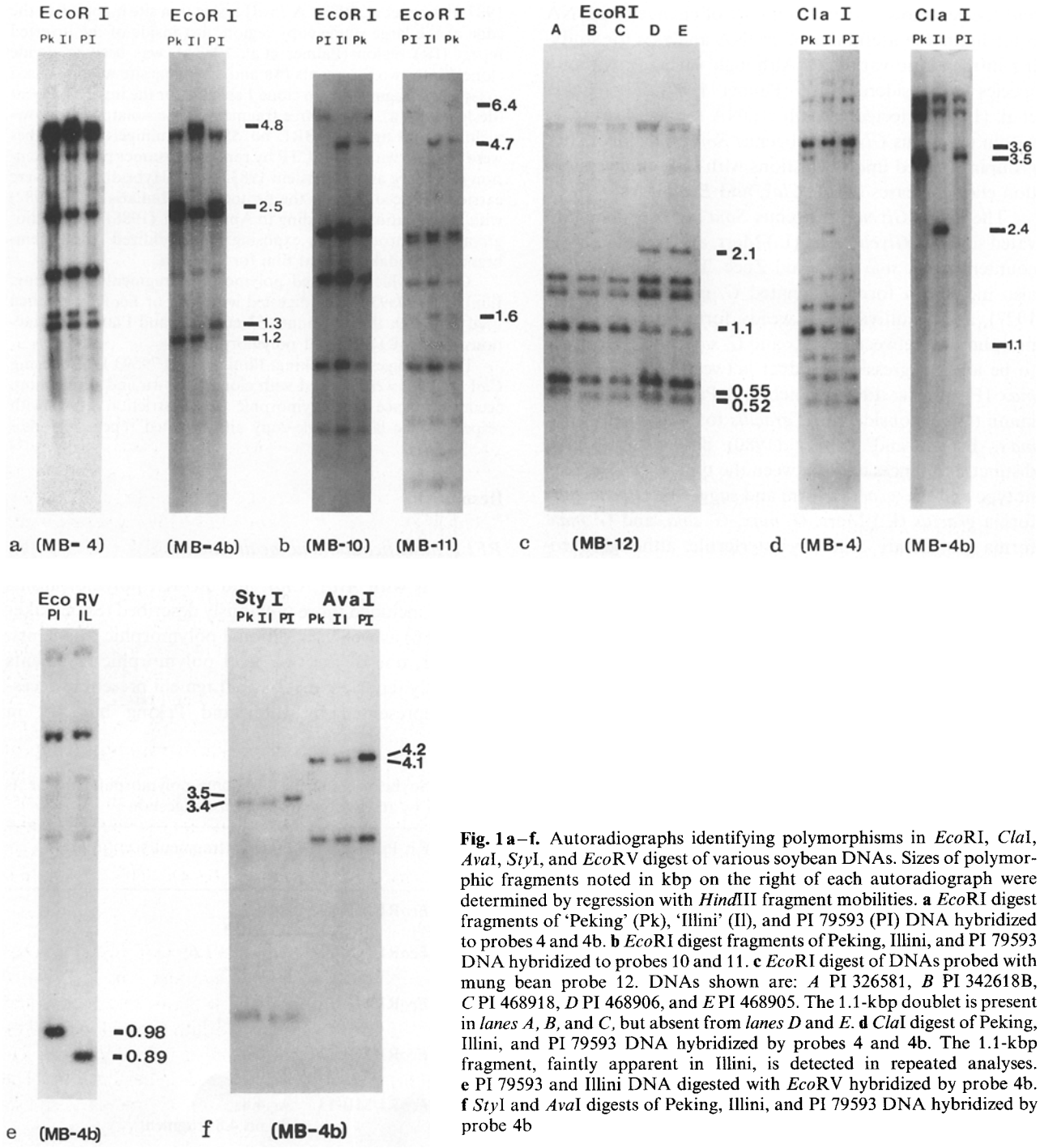

Fig. 1 a–f. Autoradiographs identifying polymorphisms in *Eco*RI, *Cla*I, *Ava*I, *Sty*I, and *Eco*RV digest of various soybean DNAs. Sizes of polymorphic fragments noted in kbp on the right of each autoradiograph were determined by regression with *Hind*III fragment mobilities. **a** *Eco*RI digest fragments of 'Peking' (Pk), 'Illini' (Il), and PI 79593 (PI) DNA hybridized to probes 4 and 4b. **b** *Eco*RI digest fragments of Peking, Illini, and PI 79593 DNA hybridized to probes 10 and 11. **c** *Eco*RI digest of DNAs probed with mung bean probe 12. DNAs shown are: *A* PI 326581, *B* PI 342618B, *C* PI 468918, *D* PI 468906, and *E* PI 468905. The 1.1-kbp doublet is present in *lanes A, B,* and *C,* but absent from *lanes D* and *E.* **d** *Cla*I digest of Peking, Illini, and PI 79593 DNA hybridized by probes 4 and 4b. The 1.1-kbp fragment, faintly apparent in Illini, is detected in repeated analyses. **e** PI 79593 and Illini DNA digested with *Eco*RV hybridized by probe 4b. **f** *Sty*I and *Ava*I digests of Peking, Illini, and PI 79593 DNA hybridized by probe 4b

PI 79593) was not identified. *Eco*RI site polymorphisms were observed by using probes 4, 4b, 10, 11, 12 (Fig. 1a–c) and 13 (not shown). Polymorphism of the *Cla*I restriction pattern was detected by using probes 4 and 4b (Fig. 1d) and probes 4a, 1, and 1a (not shown). 4b hybridization to *Eco*RV-digested DNAs revealed polymorphic fragments (Fig. 1e). The data identify polymorphism in *Sty*I restriction fragments hybridizing to clone 4b as well (Fig. 1f). Mung bean probes 2, 3, 5, 6, 7, 8, and 9 identified no RFLPs using the restriction enzymes listed above. Polymorphism of *Dra*I or *Hind*III restriction patterns was not detected using any of the mung bean clones. The RFLPs detected with the mung bean probes are summarized in Table 1. These are categorized as mutations 1 through 7 and are described below.

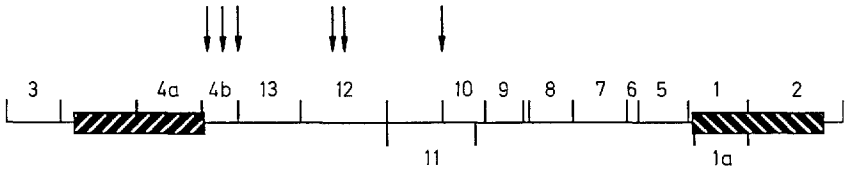

Fig. 2. *Arrows* indicate the distribution of RFLPs along the chloroplast DNA of soybean. [Arrangement of regions of homology to mung bean clones along the soybean ctDNA adapted from Palmer et al. (1988), and Palmer et al. (1983b)]. Soybean ctDNA is *circular*, but is presented in a *linear diagram* for ease of illustration

Mutation 1. Clone 10 detected an EcoRI fragment of 4.7 kbp present in Peking and Illini, but missing in PI 79593 (Fig. 1b). PI 79593 contained an additional *Eco*RI fragment of 6.4 kbp. The 6.4 kbp fragment was one of those previously identified (Shoemaker et al. 1986). The 4.7-kbp fragment had not been detected prior to these studies. These *Eco*RI fragments are also detected by clone 11, which is not surprising since clone 10 is largely contained within clone 11 (Fig. 2). Clone 11, however, revealed an additional *Eco*RI fragment of 1.6 kbp in Peking and Illini, which was not found in PI 79593 (Fig. 1b). This 1.6-kbp fragment had not been previously identified.

Mutation 2. A 2.1-kbp fragment was identified with probe 12 in *Eco*RI digests of Illini but not in digests of Peking or PI 79593. Peking and PI 79593 appear to contain a "doublet" band at 1.1 kbp, whereas Illini contains only a single band of that size hybridizing to mung bean clone 12 (Fig. 1c).

Mutation 3. The germplasm survey facilitated the identification of a new polymorphism with probe 12 (Fig. 1c). A single *G. soja* accession, PI 326581, showed a polymorphic fragment of 0.55 kbp replacing a fragment of 0.52 kbp found in all other genus *Glycine* accessions surveyed.

Mutation 4. Clone 13 hybridized to a 4.8-kbp fragment in Illini *Eco*RI digests that was not detected in digests of Peking or PI 79593 (not shown).

Mutation 5. Mung bean clone 4 overlaped the juncture of the inverted repeat and large single-copy region of the mung bean genome and included sequences from both regions. As a consequence of an inversion event in mung bean relative to soybean (Palmer 1987; Palmer et al. 1988), probe 4 hybridized to sequences in soybean ctDNA adjacent to those sequences hybridizing to probe 13. Probe 4 hybridized, albeit faintly, to an *Eco*RI fragment of 4.8 kbp in Illini, which is not present in Peking or PI 79593 (Fig. 1a), which instead showed hybridization to a 2.5-kbp fragment. (This 4.8-kbp fragment may be the same 4.8-kbp fragment detected in Illini with probe 13.) The 4.8-kbp and 2.5-kbp polymorphic fragments were previously described by Shoemaker et al. (1986). Cleavage of clone 4 into primarily inverted repeat (4a) and single-copy (4b) homologous segments, and subsequent hybridizations using these probes, revealed additional information concerning the location of these *Eco*RI polymorphic sites. Clone 4a detected no polymorphisms of *Eco*RI restriction sites, but strong hybridization of clone 4b to the 4.8-kbp fragment in Illini and the 2.5-kbp fragment in Peking and PI 79593 was observed. Clone 1 showed no polymorphism in *Eco*RI digests. Mung bean clone 1 could be cleaved into fragment 1a, containing only inverted repeat homologous sequences. This fragment, by definition, should be identical to 4a, and indeed, all hybridization data obtained with probe 1a are identical to those obtained with 4a. The other cleavage product, between the *Pst*I end point of the clone (located in the LSC region just outside the inverted repeat) and the *Pvu*II restriction site (in the inverted repeat just inside of the juncture with the LSC region) (Palmer et al. 1983b), is much too small to detect on an agarose gel, possibly being only a few nucleotides pairs in length. An additional polymorphism of the *Eco*RI restriction pattern was detected with probe 4b. This will be described under Mutation 7.

*Mutation 6. Cla*I restriction site polymorphisms was detected with probes 1, 1a, 4, 4a, and 4b. In Illini, a 2.4-kbp fragment was detected with all these probes and a 1.1-kbp fragment was detected with probe 4b (Fig. 1c). The 1.1-kbp fragment showed only a faint signal with 4b, but could be detected in repeated hybridizations. The 2.4-kbp and 1.1-kbp fragments were not detected in Peking digests; instead, a fragment of 3.5 kbp was identified with probes 1, 1a, 4a, and 4b. None of these three fragments were detected in hybridizations to PI 79593 *Cla*I digests; however, a fragment of 3.6 kbp was identified with all these probes. With the exception of the 1.1-kbp fragment, this RFLP pattern had been described previously (Shoemaker et al. 1986). Double digestion with *Cla*I and *Xho*II resulted in a uniform size reduction of the three polymorphic fragments (3.6, 3.5, or 2.4 kbp) of about 0.1 kbp, placing the nonvariable end point of these *Cla*I polymorphic fragments approximately 0.1 kbp in-

Table 2. Plastome groupings within the genus *Glycine*, subgenus *Soja* based upon RFLP patterns. (See Table 1 for a description of mutations 1–7). *G. max*, *G. max* forma *gracilis*, and *G. soja*, denoted by (m), (g) and (s) respectively; [T/T] indicate the presence of a nuclear genomic translocation in *G. soja* relative to *G. max* (RG Palmer et al. 1987); [N/N] indicate a normal arrangement of the genome relative to *G. max*; [n.t.] indicate that the accession has not been tested for the translocation

Accession	Maturity	Accession	Maturity
Group I: Mutations 1-A, 2-A, 3-A, 4-A, 5-A, 6-A, 7-A			
Illini (m)[a,b]	III	Harper (m)	III
Dunfield (m)[a,b]	III	Hodgson 78 (m)	I
Mukden (m)[a,b]	I	Wilson (m)	IV
Roanoke (m)[a,b]	VII	BSR101 (m)	I
Tokyo (m)[a,b]	VII	Jilin 3 [PI 427099] (m)	I
Mandarin (m)[a,b]	I	Maple Arrow (m)	00
Minsoy [PI 27890] (m)[b]	0	Altona (m)	00
Harosoy (m)[b]	II	Seneca (m)	II
A.K. Harrow (m)[b]	III	Mandarin [Ott.] (m)	0
Hark (m)[b]	I	S100 (m)	V
Manchu (m)[b]	III	A80-244036 (m)	II
Richland (m)[b]	II	A81-356022 (m)	III
Polysoy (m)[b]	V	PI 54610 (m)	VI
Corsoy (m)	II	PI 91732 (m)	I
Williams 82 (m)	III	PI 437477B (m)	III
Pride B216 (m)	II	T248 [PI 83945-4] (m)	IV
M.W. Oil Seed. 3010 (m)	III	PI 468904 (s) [N/N]	00
TriValley Charger (m)	III	PI 468905 (s) [N/N]	0
Stine 3200 (m)	III	PI 468906 (s) [n.t.]	0
Group II: Mutations 1-A, 2-A, 3-A, 4-A, 5-A, 6-B, 7-A			
CNS (m)[b]	VII	Noir I [PI 290136] (m)	0
Medium Green (m)[b]	I	PI 398303 (m)	IV
Aoda [PI 81043] (m)	IV	PI 407900 (m)	IV
Biloxi (m)	VIII	PI 240664 (m)	X
Palmetto (m)	VIII	PI 408272B (m)	IV
Group III: Mutations 1-A, 2-B, 3-A, 4-B, 5-B, 6-B, 7-A			
Peking (m)[b]	IV	PI 468918 (s) [N/N]	III
PI 424004A (s)[b] [n.t.]	II	PI 468916 (s) [N/N]	III
PI 342618B (s) [T/T]	I		
Group IV: Mutations 1-B, 2-B, 3-A, 4-B, 5-B, 6-B, 7-A			
PI 153292 (g)[b]	III		
Group V: Mutations 1-B, 2-B, 3-A, 4-B, 5-B, 6-B, 7-B			
PI 79593 (g)[b]	II		
Group VI: Mutations 1-B, 2-B, 3-B, 4-B, 5-B, 6-B, 7-B			
PI 326581 (s) [N/N]	II		

[a] Cultivar identified as a maternal ancestor of commonly grown soybean cultivars
[b] Cultivar or Plant Introduction analyzed by Shoemaker et al. (1986)

side the large single-copy proximal end of the 5.5-kbp *Xho*I fragment (IR region) (Palmer et al. 1983b) and the polymorphic *Cla*I sites in the 4b homologous region (LSC region). The mutation resulting in fragments of 2.4 kbp + 1.1 kbp versus a single fragment of 3.5 kbp is considered to be mutation 6. The change resulting in a 3.6-kbp fragment instead of a 3.5 kbp fragment is described as Mutation 7 (see below).

Mutation 7. This mutation is identified as a size difference of 100 bp detected in digests with five different restriction endonucleases. Probes 4a, 4b, 1 and 1a all detect a size difference of approcimately 100 bp in *Cla*I restriction fragments resulting in a 3.6-kbp fragment in PI 79593 versus a 3.5-kbp fragment in Peking. A fragment of 1.2-kbp in Peking and Illini that hybridized strongly to clone 4b with at least twice the intensity seen

in PI 79593 was detected in *Eco*RI digested DNA. The reverse situation existed for a 1.3-kbp band in PI 79593, which had at least twice the intensity of a 1.3-kbp band in Peking and Illini. No differences in signal intensity of these bands were observed with clone 4a or 1a. Hybridization of *Eco*RV-digested DNAs to probes 1, 4, or 4b, detected a fragment of 0.89-kbp in Peking and Illini, but not in PI 79593, where a unique fragment of 0.98 kbp was identified (Fig. 1e). These fragments are not detected with probes 4a or 1a, placing the polymorphic *Eco*RV site(s) within the 4b homologous region. A mobility shift attributed to a size difference of approximately 100 bp was also detected by using AvaI-digested DNAs hybridized with clone 4b (Fig. 1f). A fragment of approximately 4.1 kbp was observed in both Illini and Peking but not PI 79593, while a fragment of approximately 4.2 kbp was seen only in PI 79593. This same size difference was detected in *Sty*I digested DNAs with clone 4b (Fig. 1f). A fragment of approcimately 3.4 kbp was detected in Illini and Peking, but not in PI 79593, where a fragment approximating 3.5 kbp was observed instead. All five of these polymorphisms reflect a size difference of 100 bp between PI 79593 and both Illini and Peking.

Germplasm survey

Fifty-three additional cultivars and Plant Introductions were screened for the *Eco*RI and *Cla*I polymorphisms described. This germplasm survey partly overlapped the previous study and included additional *G. max,* and *G. soja* cultivars and Plant Introductions. Analysis of the RFLP patterns of the DNAs common to both studies confirmed that we had identified the polymorphic fragments previously described as well as additional polymorphisms. The distribution of RFLP patterns allowed the assignment of these genotypes into six groups of cytoplasmic relatedness. The results are listed in Table 2.

Discussion

The chloroplast genome of land plants is characterized by an extremely slow rate of evolution compared with the nuclear genome. This high conservation of sequence, combined with the cytoplasmic inheritance of ctDNA, makes the molecular analysis of chloroplast DNA restriction enzyme fragment patterns a particularily useful tool for understanding evolutionary relationships between closely related species (Palmer 1987; Palmer 1985; Curtis and Clegg 1984). Rates of sequence divergence between individual or related species are relatively low (Curtis and Clegg 1984). Of the changes that have occurred, most of the divergence is due to nucleotide substitutions or small deletions and additions of 1–10 bp in noncoding regions (Palmer 1987). The distribution of variant sites is reported to be decidedly nonrandom with respect to the inverted repeat and single-copy regions (Curtis and Clegg 1984). The inverted repeat evolves in tandem and at a much slower rate (Palmer 1985). Several reports (Kung et al. 1982; Palmer et al. 1983b; Palmer 1985; Salts et al. 1984; Sytsma and Gottleib 1986) have implicated a higher rate of interspecific and intraspecific sequence divergence in the large single-copy area near the junctures with the inverted repeats than in other areas of the genome.

Interpretation of hybridization data from our study places the majority of polymorphic restriction sites within one portion of the large single-copy region. Figure 2 summarizes the distribution of restriction site polymorphisms located to date. At least six molecular changes have been detected. Both insertion/deletion events, as well as restriction site change, are evident from the data. PI 79593 carries a 0.1-kbp addition relative to other accessions studied (Mutation 7). This size difference is evident with several different restriction endonucleases (*Cla*I, *Eco*RI, *Eco*RV, *Ava*I and *Sty*I), as one would expect. A restriction site change has likely occurred, resulting in the nearly additive ($4.7+1.6\cong6.4$) polymorphic fragments detected with clone 11 (Mutation 1). The polymorphic fragments detected in *Cla*I digests probably result from a restriction site change as well (Mutation 6), yielding additive fragments ($2.4+1.1=3.5$). These fragments, when combined with the 100 bp size difference discussed previously, result in a 3.6-kbp fragment in PI 79593. *Eco*RI polymorphisms detected by using clone 12 probably reflect an insertion/deletion (Mutation 3) of approximately 30 nucleotides, seen in PI 326581, as well as a restriction site change (Mutation 2) resulting in a fragment of 2.1 kbp, seen in Illini, instead of two 1.1-kbp fragments. The events resulting in a 4.8-kbp *Eco*RI fragment detected with use of either clone 13 or 4b (seen in Illini e.g.) versus a 2.5-kbp fragment detected with clone 4b (Mutation 4/5) are not clear. If an insertion/deletion of 2.3 kbp occurred, it should be evident in DNA digests with other restriction endonucleases, but this was not the case. No additive polymorphic fragments have been identified to indicate a restriction site mutation, however; it is possible that other restriction fragments hybridizing to clone 13 or 4b obscured the fragment.

The mutations just described have been designated 1 through 7 (numbers 4 and 5 are quite possibly the same mutation). These are listed in Table 1. The mutations can be used to divide the germplasm surveyed into the six plastome groups listed in Table 2. Treating each group as a haplotype, it is possible to construct a network of plastome groups (Lansman et al. 1983) showing possible phylogenetic relationships. This network is diagrammed in Fig. 3. This network suggests a branching between plastome groups I, II, and III–VI, breaking the subgenus

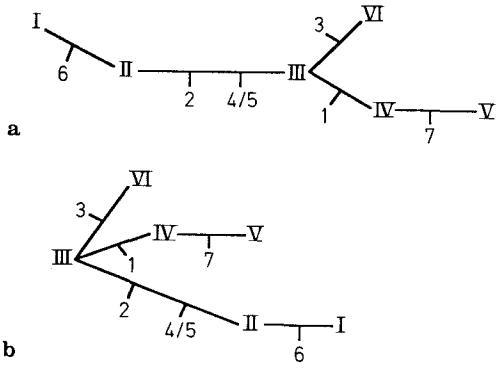

Fig. 3 a and b. Networks of plastome groups and ctDNA mutations. *Arabic numbers* identify each mutation (Table 1). *Roman numerals* denote plastome groups (haplotypes). **a** Unrooted network showing relatedness based upon mutation alone. **b** Network constructed using plastome group III (containing both *G. max* and *G. soja*) as the progenitor group. Group VI contains a *G. soja* accession. Groups IV and V contain *G. max* forma *gracilis*, and Group II contains *G. max*. Group I contains primarily modern *G. max* cultivars and their maternal ancestors plus three *G. soja* accessions of presumed *G. max* × *G. soja* hybrid origin

Soja into two major groups (*soja* and *max*), each with subgroups.

Group I, the largest of the cytoplasmic groupings, previously contained exclusively *G. max* lines, primarily modern cultivars and their maternal ancestors. All the *G. soja* accessions had been placed into group III, a diverse group containing primitive *G. max* lines and *G. max* forma *gracilis* as well. The patterns of RFLPs from our study confirmed the placing of the modern cultivars together into group I, but surprisingly, three *G. soja* Plant Introductions were found to share this same pattern of RFLPs. These are PI 468904, PI 468905, and PI 468906. These were characterized at the time of collection and inclusion in the USDA wild soybean germplasm collection as semiwild soybeans, seemingly of *G. max* × *G. soja* hybrid origin (Bernard 1986). The molecular analysis of these *G. soja* accessions demonstrates that they posses a chloroplast genome uniquely characteristic of *Glycine max*. Maternal inheritance of ctDNA in genus *Glycine*, subgenus *Soja* was demonstrated by Hatfield et al. (1985) and Corriveau and Coleman (1988). Our ctDNA analysis suggests that PI 468904, PI 468905, and PI 468906 are indeed of hybrid origin and, furthermore, that the maternal parent of the outcrossing was likely *Glycine max*.

The germplasm survey detected a new polymorphism in one of the *G. soja* accessions screened. PI 326581 demonstrated a pattern unique among all the cultivars and Plant Introductions screened. Previously, all the *G. soja* had been placed into group III. The identification of this mutation placed this *G. soja* accession alone into a new plastome group VI. This accession was introduced from the USSR and does not carry the nuclear chromosome translocation, relative to *G. max* (Palmer et al. 1987a), characteristic of many *G. soja* accessions from the Soviet Union.

Both *G. soja* and *G. max* have a chromosome number of $2n=40$ and are generally interfertile (Hadley and Hymowitz 1973). However, a chromosome interchange is reported in some *G. soja* accessions relative to the *G. max* chromosome complement (Palmer and Heer 1984; Palmer et al. 1987a). Evidence from morphological, cytological, and biochemical studies have supported the hypothesis that *G. soja* is the wild progenitor of the cultivated soybean *G. max* (Hymowitz and Newell 1980; Hymowitz and Singh 1978). *G. max* forma *gracilis* is found only where these two species hae come in contact (Hymowitz 1970) and it is thought that the *G. max* forma *gracilis* is the weedy product of hybridization between these two species (Broich and Palmer 1980). It has been recommended that *G. soja* and *G. max* be considered one species, representing a subspecies (Smartt 1984). Other reports of molecular studies undertaken to investigate the relationships among the taxa of the subgenus *Soja* include analyses of the nuclearly encoded 18S–25S ribosomal RNA multigene family and the 5S ribosomal RNA genes of *G. soja* and *G. max* (Doyle and Beachy 1985; Doyle 1988). Their data indicate little differentiation between the two taxa (the only variants being two *G. soja* accessions differing in 5S repeat size compared to 31 other *G. soja* and *G. max* accessions analyzed) and concluded their findings to be consistent with the existence of a single species.

Our analyses of ctDNA variation in subgenus *Soja* lead to three observations. (1) There is low mutational novelty (of ctDNA) within the subgenus, which indicates a recent common ancestry for all accessions. The maximum mutational difference, even between the most distant groups (Fig. 3), is three steps (possibly only two steps). (2) Since no outgroups were included (representatives of genus *Glycine* subgenus *Glycine*, e.g.), Fig. 3 cannot be polarized (i.e., the direction of mutation relative to ancestral type indicated). However, since *G. soja* is the presumed wild progenitor of the cultigen *G. max*, a rearrangement of the unrooted diagram in Fig. 3a is presented in Fig. 3b. (3) The occurrence of at least two independent mutations separating group III from group II provokes a question as to whether there is a missing ancestral type (i.e., a *G. soja*) bearing the intermediate haplotype) or have these mutations arisen in the history of cultivation?

It is interesting that legumes emerge as a particularly divergent family in terms of ctDNA. Although major structural rearrangements are rare in ctDNA, three prominent rearrangements mark the legume family. Palmer and Thompson (1982) describe a 50-kbp inversion unique to the legume subfamily Papilionoideae, which places the gene coding for the large subunit of Rubisco (*rbcL*) within a few kilobases of the *psbA* gene.

The chloroplast DNA genome of one branch of the legume family (including pea and broad bean) has lost one of the inverted repeat segments. Within some of these legume species, the loss of this repeat had been correlated with an increase in sequence divergence (Palmer 1985; Palmer 1987). An inversion of 78 kbp has been noted by Palmer et al. (1988) in a branch of the *Phaseoleae* tribe containing *Vigna* and *Phaseolus*, but not the genus *Glycine*. Apart from the major rearrangements described, intergeneric comparisons of legume ctDNAs also yield higher values for base sequence divergences than similar comparisons in other dicotyledonous plant families (Palmer et al. 1983b).

Molecular analysis of the ctDNA sequence diversity in soybean, both interspecific and intraspecific, has proved a means for evaluating relationships between cultivars and accessions. We are interested in characterizing the nature of the molecular changes involved. Palmer (1987) notes that the small additions or deletions responsible for much of ctDNA sequence divergence often occur close to or are flanked by short direct repeats. The occurrence of these structural motifs may be more frequent in the juncture regions designated as 'hotspots' for sequence changes. Detailed mapping studies are essential to determine the structural 'landscape' of the polymorphic sites. The use of mung bean probes is not adequate for such mapping in soybean. As a consequence of the 78-kbp inversion in mung bean relative to soybean, the colinearity between the ctDNA molecules becomes somewhat confused. Also, regions of deletion and additions reported to exist between the two genomes (Palmer et al. 1983b), specifically in the areas of interest, make the use of mung bean probes impractical for any further mapping. The development of homologous ctDNA libraries for each of these soybean genomes, fine-structure mapping with genome specific probes, and subsequent sequence analysis of the specific regions flanking the polymophic sites sould enable us to better characterize the juncture regions of the variant plastomes.

Acknowledgements. The authors acknowledge Dr. J. Doyle, Dr. J. Wendel, and Dr. R. Palmer for their critical reading of the manuscript and helpful suggestions. We also thank Dr. Doyle for his generous assistance providing us with ctDNA clones. This is Journal Paper No. J-13124 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa; Project 2763. Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the United States Department of Agriculture and does not imply its approval to the exclusion of other products that also may be suitable.

References

Apuya NR, Frazier BL, Keim P, Roth EJ, Lark KG (1988) Restriction fragment polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theor Appl Genet 75: 889–901

Bernard RL (1986) USDA Soybean germplasm collection: Wild soybeans (*Glycine soja* Sieb. & Zucc.) Department of Agronomy, University of Illinois, Urbana

Broich S, Palmer RG (1980) A cluster analysis of wild and domesticated soybean phenotypes. Euphytica 29:23–32

Broich S, Palmer RG (1981) Evolutionary studies of the soybean: the frequency and distribution of alleles among collections of *Glycine max* and *G. soja* of various origins. Euphytica 30:55–64

Coates D, Cullis CA (1987) Chloroplast DNA variability among *Linum* species. Am J Bot 74:260–268

Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. Am J Bot 75: 1443–1458

Curtis SE, Clegg MT (1984) Molecular evolution of chloroplast DNA sequences. Mol Biol Evol 1:291–301

Doyle JJ (1988) 5S ribosomal gene variation in the soybean and its progenitor. Theor Appl Genet 75:621–624

Doyle JJ, Beachy RN (1985) Ribosomal gene variation in soybean (*Glycine*) and its relatives. Theor Appl Genet 70: 369–376

Dyer TA (1985) The chloroplast genome and its products. Oxford Surv Plant Mol Biol 2:147–177

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6

Feinberg AP, Vogelstein B (1984) (Addendum) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266

Gordon KHJ, Crouse EJ, Bohnert HJ, Herrmann RG (1982) Physical mapping of differences in chloroplast DNA of the five wild-type plastomes in *Oenothera* subsection *Eudenothera*. Theor Appl Genet 61:373–384

Hadley HH, Hymowitz T (1973) Speciation and cytogenetics. In: Caldwell BE (ed) Soybeans: Improvement, production, and uses. American Society of Agronomy, Madison/WI pp 97–116

Hatfield PM, Shoemaker RC, Palmer RG (1985) Maternal inheritance of chloroplast DNA within the genus *Glycine*, subgenus *soja*. J Hered 76:373–374

Hermann FJ (1962) A revision of the genus *Glycine* and its immediate allies. US Dep Agric Tech Bull 1268:1–79

Hymowitz TA (1970) On the domestication of the soybean. Econ Bot 23:408–421

Hymowitz TA, Newell CA (1980) Taxonomy, speciation, domestication, dissemination, germplasm resources, and variation in the genus *Glycine*. In: Summerfield RJ, Bunting AG (eds) Advances in legume science. Royal Botanical Gardens, Kew pp 251–264

Hymowitz TA, Singh RJ (1987) Taxonomy and speciation. In: Wilcox JR (ed) Soybeans: Improvement, production, and uses. American Society of Agronomy, Madison/WI pp 23–48

Keim P, Olson TC, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. Soybean Genet Newslett 15: 150–152

Kung SD, Zhu YS, Shen GF (1982) *Nicotiana* chloroplast genome III. Chloroplast DNA evolution. Theor Appl Genet 61:73–79

Lansman RA, Avise JC, Aquadro CF, Shapira JF, Daniel SW (1983) Extensive genetic variation in mitochondrial DNA's among geographic populations of the deer mouse, *Peromyscus maniculatus*. Evolution 37:1–17

Maniatis TE, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor/NY

Palmer JD (1985) Comparative organization of chloroplast genomes. Annu Rev Genet 19:325–354

Palmer JD (1987) Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. Am Nat Suppl 130:S74–S100

Palmer RG and HE Heer (1984) Agronomic characteristics and genetics of a chromosome interchange in soybean. Euphytica 33:651–663

Palmer JD, Thompson W (1981) Rearrangements in the chloroplast genomes of mung bean and pea. Proc Natl Acad Sci USA 78:5533–5537

Palmer JD, Thompson W (1982) Chloroplast DNA rearrangements are more frequent when a large inverted sequence is lost. Cell 29:537–550

Palmer JD, Zamir D (1982) Chloroplast DNA evolution and phylogenetic relationships in *Lycopersicon*. Proc Natl Acad Sci USA 79:5006–5010

Palmer JD, Shields CR, Cohen DB, Orten TJ (1983a) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. Theor Appl Genet 65:181–189

Palmer JD, Singh GP, Pillay DTN (1983b) Structure and sequence evolution of three legume chloroplast DNAs. Mol Gen Gent 190:13–19

Palmer RG, Newhouse KE, Graybosch RA, Delannay X (1987a) Chromosome structure of the wild soybean. J Hered 78:243–247

Palmer JD, Nugent JM, Herbon LA (1987b) Unusual structure of geranium chloroplast DNA: A triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. Proc Natl Acad Sci USA 84:769–773

Palmer JD, Osorio B, Aldrich J, Thompson WF (1987c) Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. Curr Genet 11:275–286

Palmer JD, Osorio B, Thompson WF (1988) Evolutionary significance of inversions in legume chloroplast DNAs. Curr Genet 14:65–74

Reed KC, Mann DA (1985) Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res 13: 7207–7221

Salts Y, Herrman RG, Peleg N, Lavi U, Izhar S, Frankel R, Beckmann JS (1984) Physical mapping of plastid DNA variation among eleven *Nicotiana* species. Theor Appl Genet 69:1–14

Shen GF, Chen K, Wu M, Kung SD (1982) *Nicotiana* chloroplast genome IV. *N. accuminata* has larger inverted repeats and genome size. Mol Gen Genet 187:12–18

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and sequence. EMBO J 5:2043–2049

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

Skvorzov B (1927) The soybean – wild and cultivated in Eastern Asia. Proc Manchurian Res Soc Publ Ser A, Nat History Sect No. 22:1–8

Smartt J (1984) Gene pools in grain legumes. Econ Bot 38:24–35

Sytsma KJ, Gottleib LD (1986) Chloroplast DNA evolution and phylogenetic relationships in *Clarkia* sect. *Peripetasma* (Onograceae). Evolution 40:1248–1261