Theor Appl Genet (1994) 89:297–304

B. W. Diers · V. Beilinson · N. C. Nielsen
R. C. Shoemaker

Genetic mapping of the *Gy4* and *Gy5* glycinin genes in soybean and the analysis of a variant of *Gy4*

Received: 20 June 1993 / Accepted: 10 November 1993

Abstract The predominant storage protein of soybean [*Glycine max* (L.) Merr.] seed is a globulin called glycinin. Thus far five genes encoding glycinin subunits have been described, and these are denoted by the gene symbols *Gy1* to *Gy5*. The objectives of this study were to map two of these genes, *Gy4* and *Gy5*, and to conduct a genetic analysis of a subunit size-variant from an allele of *Gy4*. For this purpose a population was formed with an interspecific cross between PI 468916 (*G. soja*) and A81-356022 (*G. max*). The two size forms of G4, the subunit from *Gy4*, segregated codominantly in the mapping population, and were due to a short insertion in the hypervariable region of the mutant protein. The biochemical and molecular characteristics of the two subunits indicate that they are produced from alternate alleles of the same gene. The gene symbols Gy^a and Gy^b have been assigned to the normal and variant genes, respectively. When genomic DNA from the two parents was probed with a *Gy4* cDNA, RFLPs were identified for both *Gy4* and *Gy5*. Using these genetic markers, the *Gy4* and *Gy5* glycinin genes were mapped in linkage group "O" and "F" on the public soybean genomic map.

Joint contribution of North Central Region, USDA-ARS and Journal Paper No. J-14736 of the Iowa Agric. and Home Economics Exp. Stn., Ames, IA 50011; Project 2763. This work was supported, in part, with grants from the Iowa State Biotechnology Program (No. 480-46-09) and the Iowa Soybean Promotion Board to RCS, and the American Soybean Association to NCN

Communicated by A. L. Kahler

B. W. Diers
Department of Crops and Soil Science, Michigan State University, East Lansing, MI 48824, USA

V. Beilinson · N. C. Nielsen
USDA-ARS, Departments of Biochemistry and Agronomy, Purdue University, West Lafayette, IN 47907, USA

R. C. Shoemaker (✉)
USDA-ARS Field Crop Research and Departments of Agronomy and Zoology/Genetics, Iowa State University, Ames, IA 50011, USA

Key words Storage protein · Genome · RFLP

Introduction

The predominant storage protein in soybean [*Glycine max* (L.) Merr.] seed is an 11 *s* globulin called glycinin. Because of the importance of these proteins, their properties and the genes that encode them have been studied extensively (for a review see Shotwell and Larkins 1989). To-data, five genes that encode glycinin subunits have been cloned and assigned the gene symbols *Gy1* to *Gy5*. The subunits encoded by these genes are denoted G1 to G5, respectively, to distinguish them from the genes (Nielsen et al. 1989). The five glycinin genes are classified into two groups based on their sequence homology and size. The group-1 genes include *Gy1*, *Gy2* and *Gy3*, while group 2 consists of *Gy4* and *Gy5*.

Several mutant glycinin genes have been described. The cultivar 'Raiden' lacks the G4 glycinin subunit encoded by *Gy4* (Kitamura et al. 1980; Staswick and Nielsen 1983). Molecular studies later showed that the absence of the G4 subunit in Raiden was the result of a mutation in the initiation codon for the *Gy4* gene (Scallon et al. 1987). The cultivar 'Forest' lacks the G3 subunit of glycinin. This null is caused by a chromosomal rearrangement that separates and inverts the two halves of the *Gy3* gene (Cho et al. 1989). Staswick et al. (1983) conducted a germplasm survey to identify storage protein variants. They screened 120 lines of *Glycine soja*, an annual that shares the subgenus *Soja* with *G. max*. The electrophoretic mobility of the storage-protein components were compared using SDS-PAGE, but no obvious variants were found. Staswick et al. (1983) widened their search to include four perennial species of the subgenus *Glycine*: *G. canescens, G. tomentella, G. tabacina and G. clandestina*. A unique electrophoretic pattern for storage proteins was observed for each species. There were differences both in the number of bands and in their mobilities.

Efforts have been made to map glycinin genes. *Gy1* and *Gy2* are arranged in a direct tandem linkage in the

genome (Nielsen et al. 1989). However, this group of two genes and the other three glycinin genes all segregate independently from one another (Cho et al. 1989). The morphological markers T/t, W1/w1, l/i, R/r and l_1L_2/l_1l_2 were also tested by these workers and found to be genetically unlinked to the glycinin genes. Related to these efforts is the development of a RFLP map for soybean. Keim et al. (1990) have reported linkage relationships between 150 RFLP markers. Diers et al. (1992) reported linkage relationships between 252 RFLP markers, and presently the map includes in excess of 450 loci (Shoemaker, unpublished data). The map is being constructed using an F2 population formed from an interspecific cross between the *G. max* experimental line, A81-356022, and the *G. soja* plant introduction, PI 468916. As this public RFLP map has been generated, an effort has been made to include markers of known biochemical functions. In this report, we describe a new allele of *Gy4* in the *G. soja* parent of the mapping population. In addition, the map location of the two group-2 glycinin genes, *Gy4* and *Gy5*, are identified.

Material and methods

Genetic material

The *Gy4* and *Gy5* genes were mapped using an F2 population derived from a cross between the *G. max* experimental line A81-356022 and the *G. soja* PI 468916. The F2 seed for this population were obtained from Dr. W. R. Fehr, Iowa State University. F2 plants were grown from this seed during the summer of 1987 near Ames, Iowa. Sixty F2 plants were harvested and threshed separately to form F2-derived lines in the F3 generation (F2:3). The F2:3 lines were grown in a replicated trial at three locations near Ames, Iowa, during 1988, with two replications per location (Keim et al. 1990). Each plot was a single row 1.5 m long with a seeding rate of 33 seeds per meter. F2:4 seed from each plot were harvested in bulk. The F2-derived families were scored for the protein variant using SDS-PAGE. Each family was scored twice, first using protein isolation from a bulk of four F2:3 seed harvested in 1987, and then from a bulk of four F2:4 seed harvested in 1988. DNA for RFLP analysis was isolated from leaf material harvested from the F2 plants.

Protein extraction

Protein was extracted by crushing the seed between two sheets of glassine paper with a pestle. The crushed seed were placed in a 1.5-ml microcentrifuge tube and soluble protein was extracted in 0.5–1.0 ml of extraction buffer. Protein for gels that did not contain urea was extracted in a buffer containing 50 mM Tris (pH 8.0), 20 g/l SDS, 10 m/l β-mercaptoethanol. The protein used for gels containing urea were extracted in the same buffer, but it contained 5 M urea. The samples were extracted for 15 min on ice with periodic mixing and then centrifuged for 10 min at 4 °C at 15,000 g. This supernatant was pipetted into fresh tubes and the protein was quantified with the Biorad protein assay kit (Bio-rad Laboratories, Richmond, Calif.). Bovine serum albumin served as the protein standard.

Protein separation

A SDS-PAGE system adapted from Laemmli (1970) was used for protein separation. The resolving gels were 11 cm long and the stacking gels were 2 cm long. The electrophoretic buffer contained 0.025 M Tris, 1 g/l SDS and 0.1875 M glycine. Protein staining and destaining procedures were as described by Honeycutt et al. (1989). The F2-derived families were scored as having an F2 genotype of homozygous *G. max*, heterozygous or homozygous *G. soja*. In the case of electrophoretic gels without urea, the resolving gels were composed of a gel of 80 to 160 g/l acrylamide with 0.44 M Tris (pH 8.8) and 1.2 g/l SDS. The stacking gels contained 55 g/l acrylamide with 0.125 M Tris (pH 6.8) and 1.0 g/l SDS. Polymerization of the resolving gels was catalyzed with 0.417 g/l ammonium persulfate and 0.167 m/l TEMED (N, N, N′, N′-tetramethylethylethylenediamine). Polymerization of the stacking gels was catalyzed with 0.625 g/l ammonium persulfate and 0.33 ml/l TEMED. For gels that contained urea, the resolving gels were composed of an acrylamide gradient of 100–185 g/l acrylamide that contained 0.58 g/l SDS, 0.44 M Tris and 5 M urea. In these experiments, the stacking gels contained 50 g/l acrylamide, 0.44 M Tris, 0.58 g/l SDS and 5M urea. The resolving and stacking gels with urea were polymerized with 0.31 g/l ammonium persulfate and 0.167 ml/l TEMED. Both kinds of gels were run approximately 11 h at a constnat 35 milliamps/gel.

RFLP methods

The *Gy4* and *Gy5* genes were mapped using the cDNA clone pG248 (Dickinson et al. 1987). The portion of the clone encoding *Gy4* was cut from the plasmid, gel-purified and radiolabeled using random hexamer primers. After separation by electrophoresis, *Hin*dIII fragments of F2 genomic DNA were blotted onto nylon membranes and probed with the labeled *Gy4* cDNA. To determine the relative location of the *Hin*dIII polymorphism associated with *Gy4*, the purified cDNA was cut at an internal *Hin*dIII site and each end was isolated by gel purification. The two ends were used separately as a hybridization probe using a protocol described previously (Apuya et al. 1988; Keim et al. 1989).

Mapping *Gy4* and *Gy5*

RFLPs associated with *Gy4* and *Gy5* were mapped in the F2 population using pG248 glycinin DNA as a probe. This DNA hybridized strongly with DNA encoding *Gy4*, weakly with DNA encoding *Gy5*, but not at all with DNA encoding *Gy1*, *Gy2*, and *Gy3* (Scallon et al. 1985). *Gy4* also was mapped by analyzing the F2 population for the G4 subunit size variant using SDS-PAGE without urea. Data from these experiments, together with an existing set of F2 data consisting of over 300 markers, were analyzed using the program 'Mapmaker' (Lander et al. 1987).

Germplasm screen

Twenty-four *G. soja* plant introductions were screened with SDS-PAGE gels to determine if the variant *Gy4* polypeptide found in PI 468916 could be identified in these *G. soja* accessions. Protein was extracted from four seeds of each accession. The SDS-PAGE system used was the same as described for the F2 population and parents.

PCR and sequencing

Two oligonucleotide primers (TAAATCTAGACTCCTCATGTTC TTTCCA and GTGTGAGGATGAGGCCATCTGCAGAAATAA) were used for PCR amplification of the *Gy4* gene from PI 468916. The sequence of these oligonucleotides was based on the known sequence of *Gy4* from the soybean variety 'Dare' (Scallon et al. 1987). Restriction endonuclease cleavage sites for *Xbal* and *Pst*I were added to the 5′ and 3′ primers, respectively, to facilitate cloning. After 35 cycles

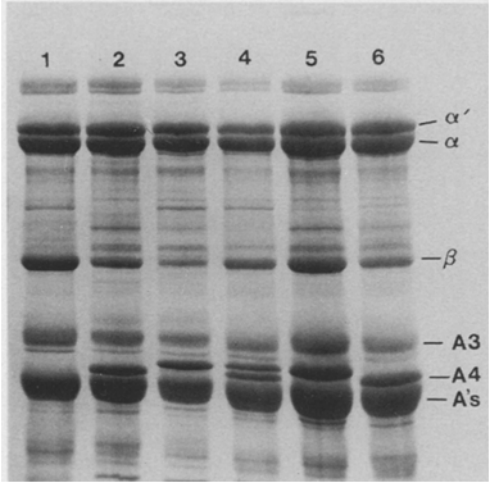
(94 °C 1 min 30 s; 53 °C 2 min; 72 °C 3 min) of PCR, a DNA fragment of about 2.7 kb was separated in an agarose gel and purified using a Geneclean II kit. The purified fragment was digested with *XbaI* and *PstI* restriction enzymes and ligated with a pBSK vector that also had been digested with these enzymes. After transformation into *E. coli* HMS (174) cells, four recombinant colonies were picked and characterized. For characterization, mini-preparations of plasmid DNA were analyzed by restriction analysis. A plasmid, pBSGM, that contained the complete *Gy4* gene sequence was obtained and sequenced.

Results and discussion

PI 468916 contains a variant seed protein

The parents of the population used for construction of a public RFLP map were screened for variation in storage protein subunit composition using SDS-PAGE. The compilation of soybean storage protein mutants described by Fontes et al. (1984) was used for comparison. Figure 1 shows that there is a mobility difference for one of the main protein bands in extracts of seeds from PI 468916 and A81-356022 (denoted A4, lanes 2 vs 3). The time of electrophoretic separation that led to the results shown in Fig. 1 were much longer than those used previously (Staswick et al. 1983), and this resulted in an increased resolution of the polypeptides in the region of the gel where the glycinin acidic polypeptides migrate.

Fig. 1 Analysis of the storage proteins of parents of the mapping population and F2-derived lines by SDS-PAGE. Conditions for electrophoretic conditions are described in Methods. The α', α and β are subunits of β-conglycinin. Bands designated *A3*, *A4* and *A's* are acidic polypeptides of glycinin (Fontes et al. 1984). *A3* and *A4* are derived from the G5 and G4 glycinin subunits, respectively, by post-translational modification (Nielsen et al. 1989). The band designated *A* contains a mixture of acidic chains from Group-1 subunits. The seed extracts used for each lane are : *1*, Raiden; *2*, A81-356022; *3*, PI 468916; and *4*, *5* and *6* are each from four seeds of F2-derived families that are heterozygous, homozygous *G. soja* and homozygous *G. max*, respectively

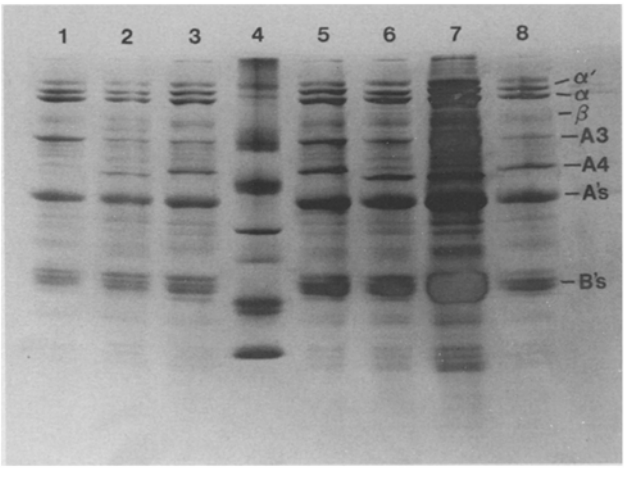


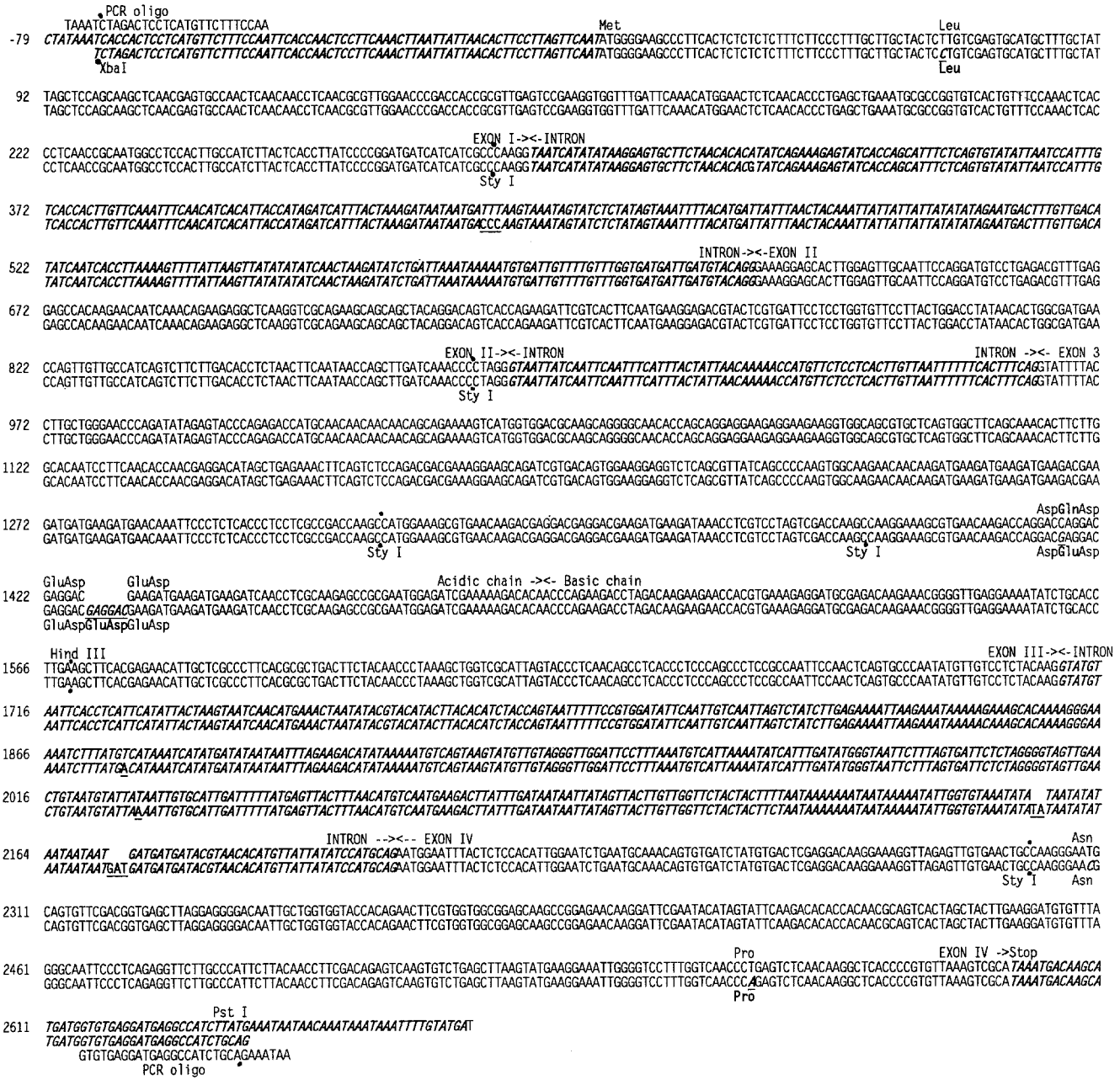
The size variant is derived from the G4 glycinin subunit

The variant seed protein in PI 468916 is derived from the glycinin *Gy4* gene. Three types of data support this conclusion. First, the variant band is absent in seed extracts from the variety Raiden (Figs. 1 and 2, lanes 1). Raiden carries a recessive null-allele, *gy4*, that causes the G4 subunit of glycinin to be absent from the seed, and A4 is derived from G4 by post-translational modification (Kitamura et al. 1984). Second, the electrophoretic mobility of the variant polypeptide is strongly influenced by the presence of urea in SDS gels. This can be seen by comparing the electrophoretic profiles reproduced in Figs. 1 and 2. By including urea in SDS gels, the electrophoretic mobility of acidic polypeptides A4 and A3 from the two Group-2 glycinin subunits, G4 and G5, respectively, are retarded as compared to those from the three Group-1 subunits (denoted as As in Figs. 1 and 2). This phenomena was described earlier by Fontes et al. (1984). Finally, as indicated in Fig. 3, DNA primers corresponding to the coding region at the 5′ and 3′ end ot *Gy4* from the variety Dare (Scallon et al. 1987) were constructed and used to amplify the corresponding region from PI 468916 DNA by PCR. By determining the nucleotide sequence of the amplified fragment, it was possibile to compare directly the genes in Dare and PI 468916.

Despite the high conservation of homology between the G4 subunits from Dare and PI 468916, a number of differences were observed in both the coding and noncoding regions of the genes. The most conspicuous difference was short insertion in the third exon beginning at position 1428. The insertion appears to be the

Fig. 2 Analysis of the storage proteins of parents for the mapping population and *G. soja* accessions with a variant A4 polypeptide by SDS-PAGE containing 5 M urea in the gel. Conditions for electrophoresis are given in the Materials and methods section. The protein bands are labeled as in Fig. 1 except that *B's* denote the basic polypeptides of glycinin. The extracts in each lane were: *1*, Raiden; *2*, A81-356022; *3*, PI 468916; *4*, molecular weight standards; *5*, PI 407288; *6*, PI 366122; 7, PI 339735; and *8*, PI 468918





result of a GAGGAC duplication, because a direct tandem array of three of these sequences is located in the mutant gene from PI 468916. The insertion occurs in the "hypervariable" region of the glycinin genes, a region identified earlier because of the extreme amount of variation that occurs in this region of all 11 *s* genes (Nielsen et al. 1989). The insertion is probably responsible for the increased apparent molecular weight associated with the mutant polypeptide in PI 468916. Additional differences were found between the two sequences, a number of which were point mutations. Two occurred in that part of the gene encoding the acidic polypeptide: a silent T-to-C point mutation in the codon for LEU at position 49, and a C-to-G transition at position 1 366 that caused GLU to be substituted for GLN. Two point mutations were also found in the basic chain, but both were silent. One affected the ASN codon at position 2 305 and the other was found in the PRO codon at position 2 561. Finally, several mutations were found in non-coding regions of the gene.

Fig. 3 Nucleotide sequence of *Gy4*[a] and *Gy4*[b]. The sequence of *Gy4*[a] (Dare) is shown on top and is taken from Scallon et al. (1987). The sequence of *Gy4*[b] originated from PI 468916 and was contained in clone pBSGM after PCR amplification. The primers used for PCR amplification are given at the beginning and end of the sequence. Cloning and sequencing details are given in the Materials and methods section. Noncoding regions are *underlined*, while the beginning and end of the coding sequence are in *bold print*. The mutations in the pBSGM sequence are indicated in *bold italic print*. *Dots* below or above the sequence at indicated restriction enzyme recognition sequences denotes points of cleavage

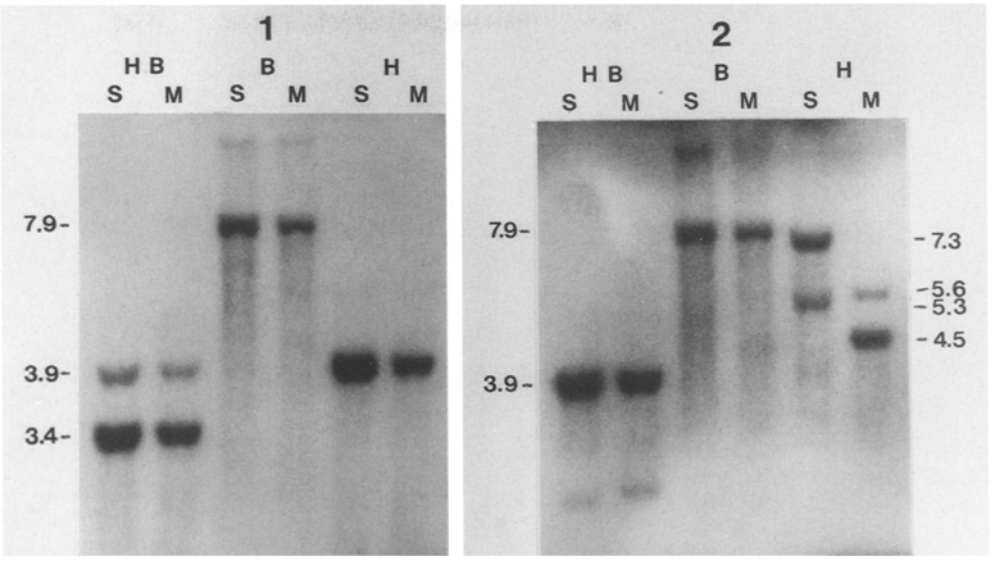


Fig. 4 Autoradiographs of the hybridization by the 3′ end (*1*) and 5′ end (*2*) of pG248 to *Bam*HI + *Hin*dIII (*HB*)-, *Bam*HI (*B*)- and *Hin*dIII (*H*)-digested genomic DNA from the *G. max* parent, A81-356022 (*M*) and the *G. soja* parent, PI 468916 (*S*). Both autoradiographs were obtained using the same membrane; the membrane was first hybridized with the fragment from the 5′ end from pG248, stripped, and then hybridized with the 3′ end of *Gy4*[a]. The sizes of the fragments are given in kilobases

There was a CCC for TTT substitution at positions 434–36 in intron-1, and two small insertions were identified in intron-3. A TA insertion occurred at positions 2 154–55, and a GAT insertion was found at position 2 173–75.

DNA purified from PI 468916 and A81-356022 was evaluated to determine if RFLP markers could be associated with glycinin *Gy4* and *Gy5* genes. For this purpose the G4 coding sequence in pG248 was purified and used to probe DNA from the two parents that had been cut with the restriction endonuclease *Hin*dIII. As illustrated in Fig. 4, restriction fragments in the digests consistently could be divided into those that hybridized strongly with the probe versus those that hybridized weakly. Those hybridizing strongly were considered to be from *Gy4*, while those hybridizing more weakly were from *Gy5*. This behavior was anticipated based upon similar results reported by Scallon et al. (1985).

The location of the polymorphism associated with *Gy4* that was detected upon digestion by *Hin*dIII is illustrated by the data in Fig. 5. For these experiments pG248 was cut at the unique internal *Hin*dIII site in the this gene (Fig. 3), and then the 5′ and 3′ fragments were used separately as hybridization probes. When *Hin*dIII-cut DNA of the parents was hybridized with the 5′-end of the clone a clear polymorphism was observed between a 4.5-kb fragment in the *G. max* and a 7.3-kb fragment in *G. soja*. The 5.6- and 5.3-kb fragments visible in these genomic blots were consistently less intense than the others and were therefore considered to be due to *Gy5*. No polymorphism was observed when the 3′ end was used to probe *Hin*dIII-cut parental DNA, and no polymorphism was observed when either end of pG248 was hybridized to *Bam*HI-cut DNA. This indicates the polymorphism that distinguishes the two parents is located 5′ to the internal *Hin*dIII site. Because the *Gy4* fragments that hybridize to the 5′-end of pG248 in *Hin*dIII/*Bam*HI double digests are of an approximately equal size, we concluded that the polymorphism lay

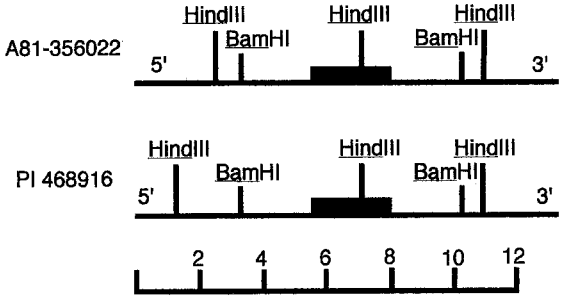


Fig. 5 Restriction maps of *Gy4* and surrounding regions for A81-356022 and PI 468916. The *darkened box* is the *Gy4* gene. The scale is given in kilobases

between the *Bam*HI and *Hin*dIII sites that are 5′ from *Gy4*.

The above interpretation is consistent with the DNA sequence data shown in Fig. 3. The insertion in the hypervariable region lies 5′ to the internal *Hin*dIII site, but the size of the insertion is small and unlikely to account for the 2.8-kb difference in the sizes of the RFLP fragments from the two parents. An insertion of this size would be difficult to detect as illustrated by the similar-sized *Bam*HI/*Hin*dIII fragments from each parent that hybridize to the 5′-pG248 probe (Fig. 6). Nonetheless, the hybridization data shown in Fig. 6 demonstrate that a size polymorphism does exist within *Gy4*. When *Hin*dIII/*Sty*1 double digests were prepared with genomic DNA derived from both the parents, an 0.2-kb fragment polymorphism was observed. Examination of the DNA sequence shown in Fig. 3 revealed that there is a *Sty*I site in the *G. soja* parent 182 bp 5′ from the unique *Hin*dIII, and this fragment contains the 6-bp insertion in the hypervariable region responsible for the polymorphism. Thus, the molecular and biochemical data clearly indicate that the mobility variant in PI 468916 is different from the one described earlier in the variety Dare; we propose that the allele in PI 468916 be denoted *Gy4*[b] to

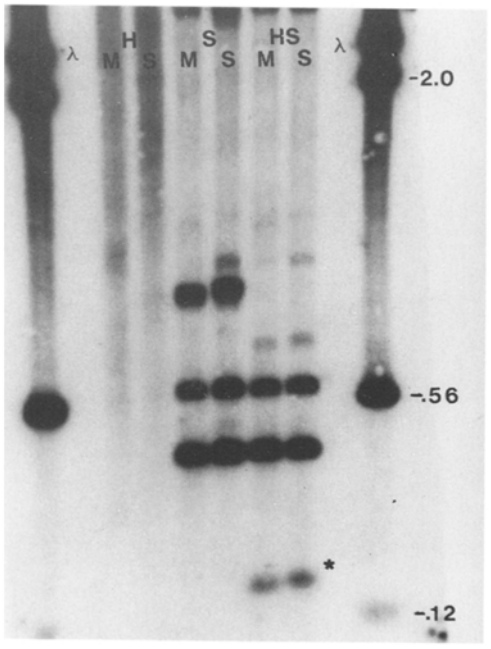


Fig. 6 Autoradiograph of the hybridization of the 5′ end of pG248 to *Hin*dIII-, *Sty*I-and *Hin*dIII + *Sty*I-digested genomic DNA from the *G. max* parent, A81-356022 (*M*) and the *G. soja* parent, PI 468916 (*S*). Also included is *Hin*dII-cut DNA (*l*). The sizes of the fragments are given in kilobases. The *asterisk* (*) denotes the *Hin*dIII + *Sty*I fragment that is polymorphic

distinguish it from the original allele, and that the original allele described by Scallon et al. (1987) now be denoted *Gy4*[a].

Gy4 and Gy5 are in linkage groups "O" and "F", respectively

Approximately 60 F2 plants from a population derived from an interspecific cross between A81-356022 and PI 468916 were examined for *Hin*dIII polymorphisms in *Gy4* and *Gy5* using pG248. This population was also monitored for the variant *G4* subunit using SDS-PAGE, and the genotype of each F2 plant was identical for both variant *G4* subunit and for the *Gy4* RFLP. Examples of the electrophoretic behavior of seed extracts from seeds of F2-derived families are given in Fig. 1 and show that the variant allele of *Gy4* in PI 468916 segregates codominantly with the normal allele A81-356022. The actual segregation data are summarized in Table 1, and show that the RFLPs associated with both *Gy4* and *Gy5* segregate with the 1:2:1 ratios expected for codominant genes.

The segregation data for *Gy4* and *Gy5*, together with data of over 300 other markers scored in this population, were analyzed using the program "Mapmaker" (Shoemaker et al., unpublished data) in order to place these two genes on the public soybean RFLP map. The LOD (logarithm of the odds ratio) that establish the linkage of *Gy4* and *Gy5* to several of these RFLP markers is given in Table 2, and a representation of their positions in the linkage map is given in Fig. 7. *Gy4* is

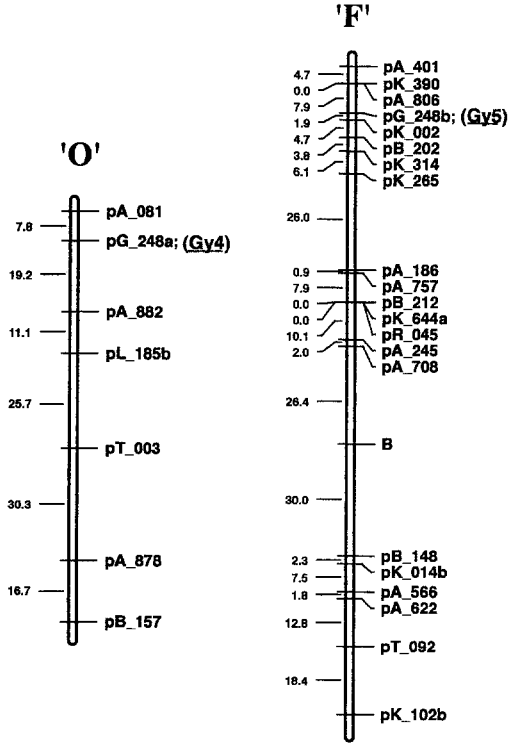
Table 1 Segregation of *Gy4* and *Gy5* alleles

Cross	Class	*Gy4*		*Gy5*	
		Observed	Expected	Observed	Expected
P1468916 (soja)	max/max	13	11.5	10	12
× A81-356022	max/soja	27	23	20	24
	soja/soja	16	11.5	18	12
	Total	46	46	48	48
		$\chi^2_{1:2:1} = 0.39$		$\chi^2_{1:2:1} = 4.0$	

Table 2 Linkage relationships between *Gy4*, *Gy5* and adjacent RFLP markers

Linkage relationship	Gene	LOD	Map separation
pG248a-pA-081	*Gy4*	14.87	7.85
pG248a-pA-882	*Gy4*	11.17	12.58
pG248b-pA-401	*Gy5*	10.81	10.66
pG248b-pK-390	*Gy5*	12.41	9.32
pG248b-pA-806	*Gy5*	11.89	9.44
pG248b-pK-002	*Gy5*	18.96	2.15
pG248b-pB-202	*Gy5*	12.64	8.13

Fig. 7 Linkage groups that include *Gy4* and *Gy5*. Linkage arrangement and distances were determined using 'Mapmaker' (Lander et al. 1987). The distances between markers are given in centiMorgans to the left of the linkage group. Markers labeled *pA* and *pK* were developed at Iowa State University and markers labeled *pT*, *pB*, *pG*, *pR* and *pL* were developed at the University of Utah and were kindly provided by Dr. K. G. Lark. The locus marked *B* is a seed coat bloom gene (Palmer and Kilen 1987)



located between pA-81 and pA-882 on linkage group "O", while *Gy5* is located on linkage group "F" between pA-806 and pK-2. This placement is consistent with the finding of Cho et al. (1989), who reported these two loci to segregate independently.

Glycinin subunit variants appear associated with centers of origin

Table 3 summarizes the *G. soja* accessions that were screened in this study, and shows that four additional size variants were identified. Two accessions, PI 407288 and PI 468918, had G4 subunits whose A4 peptide had an electrophoretic mobility identical to that of the PI 468916 parent (Fig. 8). The other two accessions, PI 336122 and PI 339735, had a predominant band with a greater electrophoretic mobility than A4 from the normal allele in A81-356022. The acidic polypeptides in these subunits were not resolved from the acidic subunits of the Group-1 subunits even though the electrophoresis was carried out for an extended period of time. In these cases the electrophoretic pattern that was produced resembled the one obtained with extracts from seeds with *Gy4* null-allele. However, a band with a mobility similar to that of the normal allele was evident when extracts of PI 336122 and PI 339735 were resolved in gels that contained urea (Fig. 2, lanes 6 and 7). It is therefore apparent that these accessions differed from the null-allele, and are referred to as "fast G4" in Table 3. Nonetheless, the genes encoding the rapidly-migrating, putative G4 subunit in PI 336122 and PI 339735 must be characterized more fully at either a genetic or molecular level before a gene symbol is assigned.

Table 3 *G. soja* plant introductions screened for storage protein *Gy4* variants

PI Number	Country of origin	Genotype
65549	China	$Gy4^a$
101404	China	$Gy4^a$
135624	China	$Gy4^a$
339735	South Korea	"fast G4"[a]
326581	Soviet Union	$Gy4^a$
326582	Soviet Union	$Gy4^a$
342618	Soviet Union	$Gy4^a$
366122	Japan	"fast G4"
407162	South Korea	$Gy4^a$
407184	South Korea	$Gy4^a$
407200	South Korea	$Gy4^a$
407275	South Korea	$Gy4^a$
407288	China	$Gy4^a$
407293	China	$Gy4^a$
407299	China	$Gy4^a$
423988	Soviet Union	$Gy4^a$
423991	Soviet Union	$Gy4^a$
424004	South Korea	$Gy4^a$
440913	China	$Gy4^a$
468904	China	$Gy4^a$
468904	China	$Gy4^a$
468906	China	$Gy4^a$
468916	China	$Gy4^b$
468918	China	$Gy4^b$

[a] Higher mobility than normal allele in SDS-PAGE but normal in SDS-PAGE and SDS/Urea-PAGE; not characterized genetically

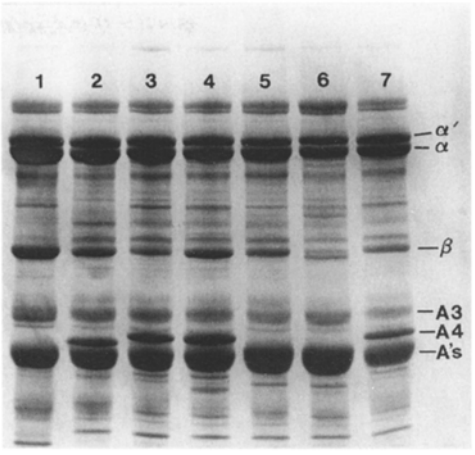


Fig. 8 Analysis of storage proteins of the parents for the mapping population and the variant *G. soja* accessions using nondenaturing SDS-PAGE. The electrophoretic conditions were given in the Materials and methods section and the proteins are labeled as in Fig. 1. The lanes were loaded with the following: *1*, Raiden; *2*, A81-356022; *3*, PI 468916; *4*, PI 407288; *5*, PI 366122; *6*, PI 339735; and 7, PI 468918

A consideration of the origin of the accessions with variant G4 subunits is informative. All three accessions with the $Gy4^b$ allele were collected in Northeastern China. PI 468916 and PI 468918 were collected near Shenyang in Liaoning Province, while PI 407288 came from near Kungchuling in Jilin Province. These two provinces lie adjacent to one another. The accessions PI 366122 and PI 339735 were collected in Japan and Korea, respectively; again indicating some regional localization. By using relationships such as these, it may be possible to study the evolution of storage protein genes and gene families.

References

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

Cho T, Davies CS, Fischer RL, Turner NE, Goldberg RB, Nielsen NC (1989) Molecular characterization of an aberrant allele for the *Gy3* glycinin gene: a chromosomal rearrangement. Plant Cell 1:339–350

Dickinson CD, Floener LA, Lilley GG, Nielsen NC (1987) Self-assembly of proglycinin and hybrid proglycinin synthesized in vitro from cDNA. Proc Nat Acad Sci USA 84:5525–5529

Diers BW, Keim P, Fehr WR, Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83:608–612.

Fontes BPB, Moreira MA, Davies CS, Nielsen NC (1984) Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β-conglycinin subunits. Plant Physiol 76:840–842

Honeycutt RJ, Burton JW, Palmer RG, Shoemaker RC (1989) Association of major seed components with a shriveled-seed trait in soybean. Crop Sci 29:804–809

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

Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. Genetics 126:735–742

Kitamura K, Toyokawa Y, Harada K (1980) Polymorphism of glycinin in soybean seeds. Phytochemistry 19:1841–1843

Kitamura K, Davies CS, Nielsen NC (1984) Inheritance of Alleles for *Cgy1* and *Gy4* storage protein genes in soybean. Theor Appl Genet 68:253–257

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Nielsen NC, Dickinson CD, Cho T, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB (1989) Characterization of the glycinin gene family in soybean. Plant Cell 1:313–328

Palmer RG, Kilen TC (1987) Qualitative genetics and cytogenetics. In: Wilcox JR (ed) Soybean: improvement, production and uses, 2nd edn. Agronomy 16:23–48.

Scallon B, Thanh VH, Floener LA, Nielsen NC (1985) Identification and characterization of DNA clones encoding group-II glycinin subunits. Theor Appl Genet 70:510–519

Scallon BJ, Dickinson CD, Nielsen NC (1987) Characterization of a null-allele for the *Gy4* glycinin gene from soybean. Mol Gen Genet 208:107–113.

Shotwell MA, Larkins BA (1989) The biochemistry and molecular biology of seed storage proteins. In: Marcus, A (ed) The biochemistry of plants: a comprehensive treatise, Vol 15. Academic Press New York, pp 296–345

Staswick PE, Nielsen NC (1963) Characterization of a soybean cultivar lacking certain glycinin subunits. Arch Biochem Biophys 223:1–8

Staswick PE, Boure P, Nielsen NC (1993) Glycinin composition of several perennial species related to soybean. Plant Physiol 72:1114–1118