

Inheritance and Organization of Glycinin Genes in Soybean

Tae-Ju Cho,¹ Corinne S. Davies,² and Niels C. Nielsen³

U.S. Department of Agriculture/Agricultural Research Service and the Department of Agronomy, Purdue University, West Lafayette, Indiana 47907

Five genes (Gy_1 through Gy_5) encode most of the subunits that are assembled into glycinin, a predominant seed storage protein found in soybeans. Restriction fragment length polymorphisms are described that identify four of these five genes (Gy_1/Gy_2 , Gy_3 , and Gy_5). The fifth gene (Gy_4) is characterized by two alleles, one of which (gy_4) causes absence of the subunit. Genetic segregation studies indicate that the five genes are located at four genetic loci within the genome. Gy_1 and Gy_2 are in a direct tandem repeat at one locus, whereas there is a single glycinin gene at each of the other three loci. All four loci segregate independently from one another, and they also segregate independently from the genetic markers for tawny/grey pubescence (T/t), purple/white flower color (W_1/w_1), light/dark hilum pigmentation (I/i'), black/brown seed coat (R/r), and brown/tan pod color ($I_1I_1L_2L_2/I_1I_1l_2l_2$). The latter genetic markers are located on linkage groups 1 (t), 8 (w.), 7 (i), and 2 (r) in the soybean genome, respectively.

INTRODUCTION

Glycinin is a prevalent seed storage protein from soybeans. It is isolated from seeds as a hexamer composed of six nonidentical but homologous subunits. Thus far at least five such subunits have been purified and characterized biochemically (Moreira et al., 1979, 1981; Staswick et al., 1981). The complete sequences for genes that encode each of these five subunits have been determined and are summarized in the preceding paper (Nielsen et al., 1989). The gene symbols Gy_1 through Gy_5 are used to denote the five genes (Davies et al., 1987). Despite the considerable primary structural data generated about the subunits and the genes that encode them, relatively few formal genetic data have been accumulated that describe the inheritance of the glycinin genes and their organization within the genome.

Availability of detailed restriction maps and appropriate probes for each of five glycinin genes (Nielsen et al., 1989) suggested that it should be possible to identify restriction fragment length polymorphisms (RFLPs) characteristic of each gene, and that these could be used as genetic markers to determine linkage relationships among glycinin genes. This paper describes several such genetic markers and their use for genetic analysis of the glycinin gene family. Evidence is presented that the five glycinin genes

are distributed among four independently segregating genetic loci.

RESULTS

Description of Genetic Markers

Two soybean varieties, Raiden and Forrest, were used to perform these studies. The two varieties carried allelic alternatives for a number of pigment genes that could be scored visually. They also contained allelic alternatives for each of the genetic loci involved in production of glycinin subunits. Hence, examination of progeny from crosses between these two parents permitted simultaneous evaluation of a number of genetic markers for linkage relationships.

The genotypes for the lines of Forrest and Raiden used were $TTw_1w_1i'i'RRl_1l_1l_2l_2$ and $ttW_1W_1llrrl_1l_1L_2L_2$, respectively. The gene symbols denote tawny/grey pubescence (T/t), purple/white flower color (W_1/w_1), light/dark hilum pigmentation (I/i'), black/brown seed coat (R/r), and brown/tan pod color ($I_1I_1L_2L_2/I_1I_1l_2l_2$) (Bernard and Weiss, 1973; Palmer and Kilen, 1987). Flower and pubescence colors were scored on F_2 plants, whereas hilum and pod colors were scored on either pods or F_3 seeds from selfed F_2 plants.

Restriction fragment length polymorphisms (RFLPs) were used as genetic markers for three of the four glycinin

¹ Current address: Department of Biochemistry, Chung-buk National University, Cheon-ju, Korea.

² Current address: Department of Biology, McGill University, 1205 Docteur Penfield Ave., Montreal, Quebec, Canada H3A 1B1.

³ To whom correspondence should be addressed.

loci that have been identified. Figure 1 shows restriction endonuclease maps of the original clones from which the probes were obtained that were used in the linkage experiments. Genes Gy_1 and Gy_2 are tandemly linked at a single genetic locus, with Gy_1 located upstream from Gy_2 (Nielsen et al., 1989). Because of this tight linkage, a single RFLP that overlapped the 3'-end of Gy_1 was used as the marker for both genes (referred to as Gy_{1+2} for purposes of this discussion). The nick-translated probe used to recognize the RFLP originated as the 1.1-kb EcoRI/PstI fragment from Gy_1 cDNA clone pG134. This fragment extended downstream from the EcoRI site in Gy_1 to the 3'-end of the coding region (see Figure 1A; identified in red in Figure 2). When DNA gel blot transfers of genomic DNA digested with EcoRI/HindIII were probed at 55°C, a 2.0-kb fragment hybridized in the case of Raiden (Figure

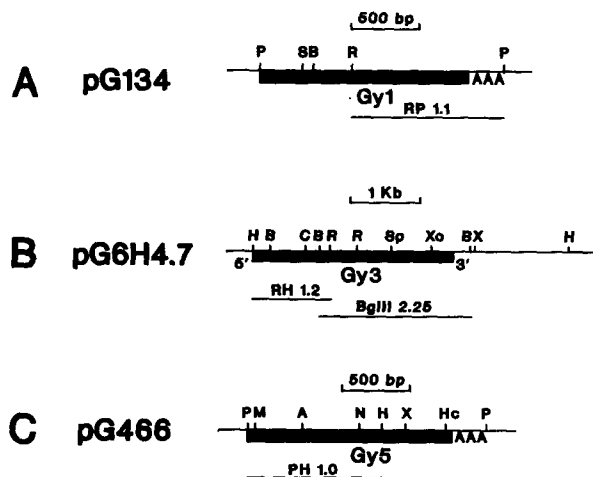


Figure 1. Restriction Maps of the Clones Used in Segregation Analysis Using DNA RFLPs as Genetic Markers.

(A) Clone pG134 contains a near full-length cDNA insert (1.8 kb) for the G1 glycinin subunit. It was prepared and isolated as described by Scallan et al. (1985). The cDNA insert was cloned into PstI site of the pUC8 plasmid. The 1.1-kb EcoRI/PstI fragment (RP 1.1) from pG134 was used as a probe, as indicated below the restriction map.

(B) Clone pG6H4.7 was derived from λ DA28-6 containing the Gy_3 glycinin gene (Nielsen et al., 1989). The 4.7-kb HindIII fragment from the λ DA28-6 was inserted into HindIII site of the pUC8 plasmid. The region in the clone pG6H4.7 that contains the gene is indicated by the black box. The 1.2-kb EcoRI/HindIII fragment (RH 1.2) and 2.25-kb BglII fragment (BglII 2.25) that were used as probes are indicated below the restriction map.

(C) Clone pG466 contains a 1.7-kb cDNA insert for the G5 glycinin subunit. It was isolated from the cDNA library described by Scallan et al. (1985). The cDNA insert was cloned into PstI site of pUC8 plasmid. The 1.0-kb PstI/HindIII fragment (PH 1.0) that was used as probe is shown below the restriction map. A, Avall; B, BglII; C, ClaI; H, HindIII; Hc, HincII; M, MluI; N, NcoI; P, PstI; S, SmaI; Sp, SphI; X, XhoI; Xo, XorII.

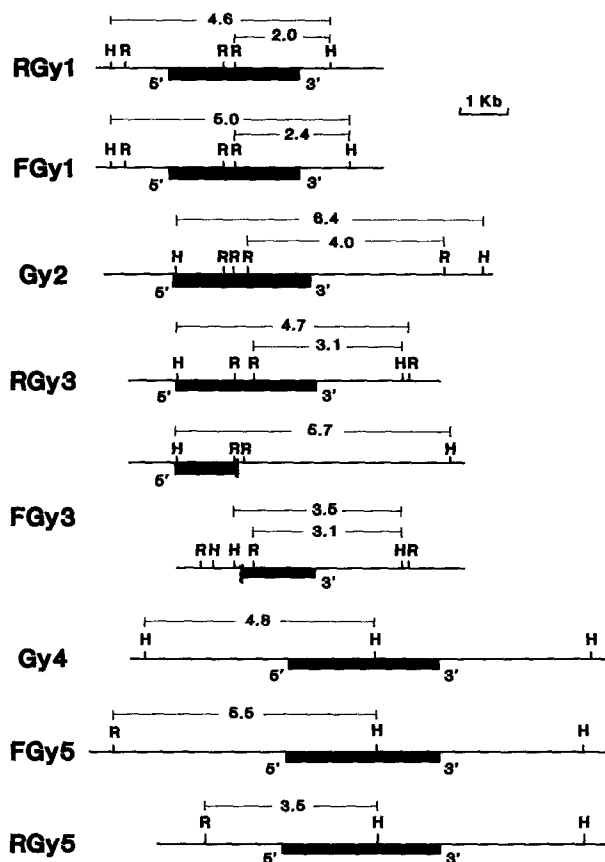


Figure 2. Restriction Maps of Group-I Glycinin Genes and Group-II Glycinin Genes in Forrest and Raiden.

Only those restriction sites showing length polymorphisms are shown. Fragments identified in red hybridized strongly to the probes, and represented the genes being monitored. Fragments identified in blue indicate fragments from homologous genes where possible cross hybridization might occur (see text). DNA RFLPs for Gy_2 and Gy_4 have not been found. F, Forrest; R, Raiden; H, HindIII; R, EcoRI.

3). In the case of the alternate allele in Forrest, a piece 2.4 kb in length was involved. The size difference of the two fragments was due to a small deletion/insertion immediately downstream from Gy_1 (T.-J. Cho, unpublished observation).

Equivalent fragments in the other two Group-I genes were also capable of hybridizing to the probe, and potentially could confuse the interpretation of results. As indicated in Figure 2, the corresponding portion of Gy_2 was located on a 4.0-kb EcoRI/EcoRI fragment (identified in blue), and the one from Gy_3 was located on a 3.1-kb EcoRI/HindIII fragment (identified in blue). Although these fragments were generally absent when genomic transfers were probed at high stringency (55°C), they occasionally were observed as considerably less intense bands on

autoradiographs. A decrease in hybridization temperature to 42°C caused an increase in the prominence of the 4.0-kb and 3.1-kb fragments of *Gy*₂ and *Gy*₃ compared with those from *Gy*₁ (not shown). However, even at the lower stringency, the size differences between the fragments permitted an easy identification of the bands.

The second glycinin locus contained the allele pair denoted *Gy*₃ and *gy*₃. Plants homozygous for *gy*₃ lacked the subunit because of a chromosomal rearrangement that separated the 5' and 3' ends of the gene (Cho et al., 1989). The rearrangement resulted in an RFLP that could be exploited for our linkage determinations. To detect the RFLP, a mixed probe was used that contained both the 1.2-kb EcoRI/HindIII and 2.25-kb BglII/BglII fragments from pG6H4.7 (Figure 1B). The former fragment encoded the 5' end of the gene, and the latter the 3' end. When DNA gel blot transfers of genomic DNA from Forrest that had been cut with HindIII were probed, two fragments hybridized, as shown in Figure 3. One was the 5.7-kb

fragment from the 5' end of *gy*₃, and the other was the 3.5-kb piece from the 3' end of this gene (identified in red, Figure 2). When genomic DNA from Raiden was probed, however, only a 4.7-kb HindIII fragment that contained an intact *Gy*₃ was observed (identified in red, Figure 2; see also Figure 3).

Analogous regions from the other Group-I genes also hybridized to the mixed probe used to analyze the second glycinin locus. The *Gy*₂ genes from both Raiden and Forrest were contained in 6.4-kb HindIII fragments (identified in blue, Figure 2). These fragments did not hybridize effectively to the mixed probe at 55°C, and were observed in autoradiographs only occasionally. In those few cases where fragments from *Gy*₂ were observed, they were easily identified because they hybridized to the mixed probe weakly and were larger than the fragments from both *Gy*₃ and *gy*₃. However, the equivalent sequences from *Gy*₁ were more troublesome because *Gy*₁ was more homologous to *Gy*₃ than it was to *Gy*₂ (Nielsen et al., 1989). As a result, the *Gy*₁ sequences cross-hybridized extensively to the mixed probe and were nearly always present in the autoradiographs. The analogous regions of *Gy*₁ were contained in 4.6-kb and 5.0-kb fragments from the Raiden and Forrest genomes, respectively (identified in blue, Figure 2). When the temperature was maintained at 55°C, the hybridization intensity of the 4.7-kb fragment with the *Gy*₃ probe was considerably stronger than it was for the fragments which contained *Gy*₁. Because of this difference in hybridization intensity, it was seldom difficult to distinguish the *Gy*₃/*Gy*₃ and *Gy*₃/*gy*₃ genotypes in the autoradiographs. This phenotype distinction became far less evident, however, when the hybridization temperature was reduced.

The third locus for a glycinin subunit contained *Gy*₅. The RFLP used for this locus is shown in Figure 2. It was identified using as a probe the 1.0-kb PstI/HindIII fragment from pG466 that included coding regions for the 5' end of the gene (Figure 1C). When hybridization was done with genomic DNA doubly digested with EcoRI and HindIII, a 3.5-kb fragment from Raiden hybridized to the probe. By contrast, a 5.5-kb fragment was present in Forrest DNA (Figure 3). In both cases, a 4.8-kb fragment was also observed. The latter fragment originated from the 5' end of *Gy*₄ (identified in red, Figure 2), and its presence in the autoradiographs could not be eliminated by adjustment of hybridization temperature.

The fourth genetic locus contained the alternate alleles *Gy*₄ and *gy*₄, the latter being recessive to the former (Kitamura et al., 1984). The two alleles have been characterized extensively, and differ from one another by only two nucleotides (Scallon et al., 1987). The critical difference that results in loss of expression is an ATG to ATA mutation in the initiator Met codon. Despite a careful search, an RFLP was not located that permitted distinction between the two alleles. However, because the G4 glycinin subunit encoded by this gene was absent in Raiden (*gy*₄/

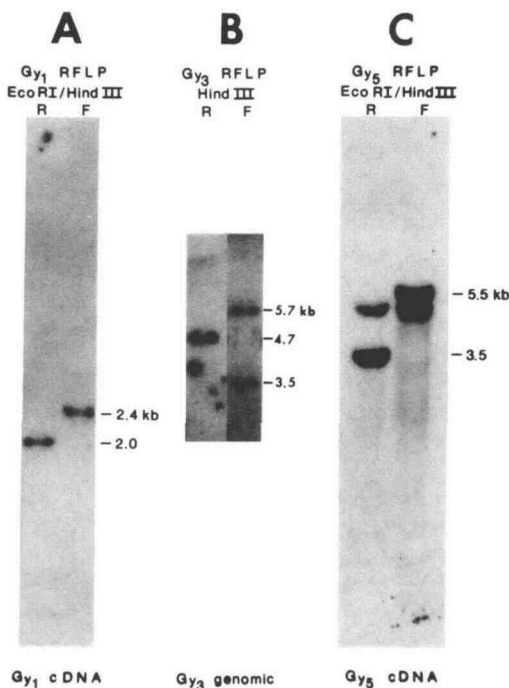


Figure 3. DNA Restriction Fragment Length Polymorphisms of *Gy*₁₊₂ (A), *Gy*₃ (B), and *Gy*₅ (C).

Leaf DNAs were digested with EcoRI and HindIII [(A) and (C)] or HindIII alone (B) and analyzed by DNA gel blot hybridization. The probes used were 1.1-kb EcoRI/PstI fragments from *Gy*₁ cDNA clone pG134 (A), 1.2-kb EcoRI/HindIII fragments and 2.25-kb BglII fragments from *Gy*₃ genomic clone pG6H4.7 (B), and 1.0-kb PstI/HindIII fragments from the *Gy*₅ cDNA clone pG466 (C). The sizes of the fragments that hybridize are indicated (kb). F, Forrest; R, Raiden.

gy₄) (Staswick et al., 1983; Kitamura et al., 1984) and present in Forrest (*Gy₄/-*), these two alleles were distinguished from one another by SDS-polyacrylamide electrophoresis (Fontes et al., 1984).

Segregation Analysis

Reciprocal crosses were performed between Raiden and Forrest, and then the *F*₁ progeny were checked to ensure that hybridization occurred. Flower color proved useful in this regard because the former parent had purple flower colors and the latter white ones. When Forrest was the female parent, the *F*₁s had purple flowers and could easily be distinguished from progeny of self-pollinations, which were white. However, when Raiden was the female parent, it was more difficult to distinguish between selfed plants and true *F*₁s because both had purple flowers. As a consequence, all potential *F*₁ progeny were subjected to RFLP analysis using the *Gy₁* insert in pG134 as probe. Since the RFLPs were inherited codominantly, the *F*₁ genotypes had fragments from both parents. Of the 13 potential *F*₁ seeds tested, 11 contained both the 2.0-kb and 2.4-kb fragments characteristic of *Gy₁* in Raiden and Forrest, respectively (data not shown). Approximately 250 *F*₂ seeds were subsequently obtained from six of these 11 plants and analyzed further.

Leaf DNA was purified from individual *F*₂ plants and subjected to RFLP analysis. Since the genetic markers that distinguished among *Gy₁₊₂* and *Gy₅* both required EcoRI/HindIII double digestions, segregation of alleles for these two genes was determined at the same time. To achieve this, a mixed probe was used that consisted of the appropriate inserts from pG134 and pG466, as indicated in Figure 1. A representative genomic DNA gel blot from one such experiment is reproduced in Figure 4. The heterozygous genotype and both types of homozygous ones could be identified and were easily scored. This permitted complete classification and reduced the number of progeny that needed to be scored to obtain statistically reliable data. The allelic pairs for both *Gy₁₊₂* and *Gy₅* segregated with the 1:2:1 ratios among the approximately 160 progeny scored. Hence *Gy₁₊₂* and *Gy₅* each segregated at a single genetic locus. The χ^2 values for goodness-of-fit tests to the 1:2:1 ratios were 1.93 ($0.25 < P < 0.5$) for *Gy₁₊₂* and 4.23 ($0.1 < P < 0.25$) for *Gy₅*.

Segregation for *Gy₃* versus *gy₃* was determined using leaf DNA digested with HindIII. DNA gel blot transfers prepared from the digests were probed with the insert from pG6H4.7 as described above (Figure 1). A representative autoradiograph from one such experiment is shown in Figure 5. The data indicated that *Gy₃* and *gy₃* segregated at a single genetic locus. The χ^2 value obtained for goodness-of-fit to a 1:2:1 ratio was 0.36 ($0.75 < P < 0.9$).

Segregation for *Gy₄* versus *gy₄* was determined by SDS-polyacrylamide electrophoresis. A picture of a representative gel is shown in Figure 6. The recessive *gy₄* allele

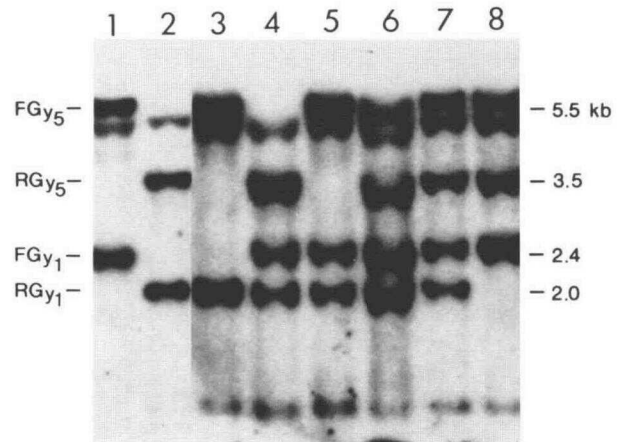


Figure 4. Determination of the Genotypes of the *Gy₁* and *Gy₅* Glycinin Genes.

DNAs from individual *F*₂ plants from Raiden × Forrest crosses were hybridized to nick-translated inserts from pG134 and pG466. Leaf DNAs were digested with EcoRI and HindIII. Lanes 1 and 2 show the parental lines (Forrest and Raiden), and the rest are for *F*₂ plants segregating for the two parental genotypes.

conditioned the simultaneous loss of the *A₅*, *A₄*, and *B₃* peptides from the seed (Staswick et al., 1983; Kitamura et al., 1984). To obtain maximum separation of the three peptides, urea was included in the gels (Fontes et al., 1984). *A₅* was the easiest of the three peptides to score since it was well separated from the other bands in the gel. Since the genotypes *Gy₄/gy₄* and *Gy₄/Gy₄* could not be distinguished from one another phenotypically the segregation data was tested for conformance to a 3:1 ratio. The χ^2 from this goodness-of-fit test was 0.545 ($0.25 < P < 0.50$). This result was consistent with previously reported data that showed these alleles resided at a single genetic locus (Kitamura et al., 1984).

Since the presence or absence of products from *Gy₄* was monitored rather than the gene itself, other genetic factors could have influenced the phenotype that was observed. To evaluate this possibility, data involving *Gy₄* were examined for maternal effects. When the segregation patterns of *Gy₄* and *gy₄* in *F*₂ progeny from reciprocal crosses were compared, differences were not evident in homogeneity tests ($\chi^2 = 0.84$, $0.25 < P < 0.50$). Hence, this concern seemed unfounded.

Linkage Relationships

χ^2 goodness-of-fit tests were performed to evaluate a null hypothesis for independent segregation between each of the four genetic loci involved in the production of glycinin subunits. The results of these tests are given in Table 1. In each case, the low χ^2 values observed were consistent

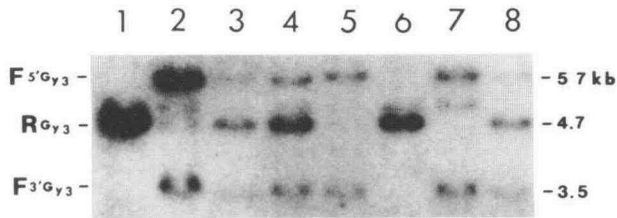


Figure 5. Determination of the Genotypes of the Gy_3 Glycinin Genes.

DNAs from individual F_2 plants from Raidein \times Forrest crosses were digested with HindIII and were hybridized to nick-translated 1.2-kb EcoRI/HindIII fragments and 2.25-kb BglII/BglII fragments from the clone pG6H4.7. Lanes 1 and 6 represent the homozygous Raidein genotype, lanes 2, 5, and 7 are the homozygous Forrest genotypes, and the rest are heterozygous genotypes.

with the null hypothesis. When recombination frequencies for linkage among loci were evaluated using the maximum likelihood method (Allard, 1956), it also indicated that segregation of the glycinin gene pairs followed the expected independent segregation ratios of either 1:2:1:2:4:2:1:2:1 or 3:6:3:1:2:1 in the F_2 generation.

The segregation data were also examined for linkage between the glycinin genes and the pigment markers T/t , W_1/w_1 , I/i , R/r , and L_2/l_2 . As anticipated from earlier reports (Bernard and Weiss, 1973, Palmer and Kilen, 1987), the allelic frequencies for each of the five pigment markers were consistent with a 3:1 ratio and they each segregated independently. Moreover, as indicated by the lack of significant χ^2 values (Table 2), the pigment markers were not linked to any of the glycinin genes. It should be pointed out, however, that Gy_{1+2} versus l_2 had a high χ^2 value (11.07) and a recombinant frequency estimated to be 0.43 ± 0.06 . When the χ^2 value was recalculated at $P = 0.43$, there was still a large deviation ($\chi^2 = 9.28$) from the expected segregation ratio. Consequently, the apparent loose linkage relationship between Gy_{1+2} and l_2 evident in our data was considered an erroneous random occurrence, and was probably due either to the small sample size or to an error in determining the L_2/l_2 phenotypes.

DISCUSSION

In this study the inheritance and organization of genes for five glycinin subunits were determined. Each of these genes corresponded to one of the five subunits that had previously been purified from the seed and characterized biochemically (Nielsen, 1984, 1985; Nielsen et al., 1989). The five genes were distributed in four loci, with Gy_1 and Gy_2 being in a direct tandem repeat at one of the loci (Nielsen et al., 1989). RFLPs were identified that served as codominantly inherited genetic markers for three of the

four loci. In the case of the fourth locus, the absence of a gene product from a recessive null allele served as the genetic marker. The results of a segregation analysis of progeny from a cross between the varieties Raidein and Forrest indicated that all four loci segregated independently. This result agrees with the molecular analysis of glycinin gene organization presented in the accompanying paper (Nielsen et al., 1989).

It is of interest to compare the organization of the glycinin genes with those of seed storage proteins in other cereals and legumes. In wheat, barley, and rye, it has been reported that homologous prolamine genes are clustered at several complex loci. Each of these loci encodes a polymorphic, but homologous, series of proteins (Shewry et al., 1984; Kreis et al., 1985; Payne, 1987). In maize, both the clustering of zein genes and coexistence of genes encoding different classes of polypeptides have been reported (Soave and Salamini, 1984; Heidecker and Messing, 1986). In legumes, the 11S subunit genes of pea (Matta and Gatehouse, 1982; Domoney et al., 1986) and the 7S proteins of *Phaseolus vulgaris* (Brown et al., 1981) are also clustered at several genomic loci. In the case of pea and common bean, however, the number of genes at each locus appears reduced compared with the situations reported in the cereals. In pea, for example, about 10 legumin genes are distributed among two or three loci (Casey et al., 1986). The situation in soybean is similar to

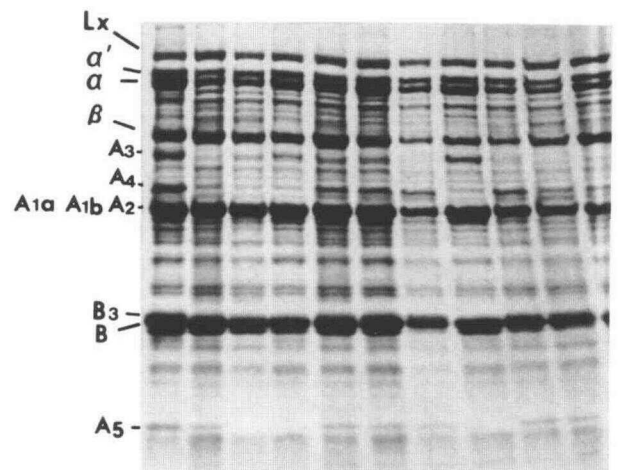


Figure 6. Determination of Gy_4/gy_4 Genotypes of F_2 Seeds from the Raidein \times Forrest.

The seeds were analyzed by SDS-polyacrylamide gel electrophoresis. Sample preparation and electrophoretic conditions are as described in "Methods." The gene products from Gy_4 are identified as A_5 , A_4 , and B_3 . Shown in this figure are F_2 seeds from a Raidein \times Forrest cross that are segregating for the gy_4 null allele. Lanes 3, 4, and 8 (from the left) lack the G4 glycinin subunit, whereas the others contain the subunit. Lx, lipoxigenase, α' , α , and β , subunits of β -conglycinin; A, acidic polypeptides of glycinin; B, basic polypeptides of glycinin.

Table 1. Test for Linkage between Gy_{1+2} , Gy_3 , Gy_4 , and Gy_5 Loci from the Cross Raiden \times Forrest

Loci	F ₂ Progeny in Each Genotypic Class ^a									n	χ^2 ^b	P ^c	r.f. ^d
	AABB	AABb	AAbb	AaBB	AaBb	Aabb	aaBB	aaBb	aabb				
Gy_{1+2}, Gy_3	7	18	10	28	32	16	10	23	13	157	4.880	>0.75	0.50 \pm 0.04
Gy_{1+2}, Gy_5	11	19	5	17	43	11	12	24	8	150	0.986	>0.99	0.48 \pm 0.04
Gy_3, Gy_5	7	23	9	22	39	10	9	24	6	149	3.280	>0.90	0.54 \pm 0.04
Gy_{1+2}, Gy_4		30 ^e	7		50	26		35	14	162	2.845	>0.50	0.46 \pm 0.05
Gy_3, Gy_4		28	15		63	18		30	12	166	2.307	>0.75	0.53 \pm 0.05
Gy_5, Gy_4		29	12		64	25		20	8	158	0.020	>0.99	0.51 \pm 0.05

^a The genotypes were classified by considering the alleles from the cultivar Forrest as dominant. A, first in gene pair; B, second in gene pair.

^b χ^2 goodness-of-fit tests to either 1:2:1:2:4:2:1:2:1 or 3:1:6:2:3:1 were performed to evaluate a null hypothesis for independent segregation of a pair of genetic loci.

^c Probability (P) for a larger value of χ^2 .

^d Recombination frequency (r.f.) calculated by the method of maximum likelihood (Allard, 1956).

^e Complete classification of genotypes containing Gy_4/gy_4 was not possible by scoring presence versus absence of the G4-subunit, so the sum of recombinants corresponding to each phenotype is given.

that observed for other legumes, although more extreme. Three of the four genetic loci in soybeans apparently have but a single functional glycinin gene, whereas the remaining one has two.

The family of glycinin genes in soybean has been divided into two groups based on properties of the subunits they encode (Nielsen, 1984). The genes in Group-I, which includes Gy_1 , Gy_2 , and Gy_3 , have nucleotide sequences that are more than 80% homologous to one another (Nielsen et al., 1989). The nucleotide sequences for members of Group-II, which includes Gy_4 and Gy_5 , are likewise more than 80% identical with one another, but are less than 60% homologous with those in Group-I. An equivalent situation exists for the homologous legumin gene families found in *Pisum* and *Vicia faba*. The Group-I glycinin genes correspond to the Lg-1 genes in pea and the A-legumins in field bean (*V. faba*), whereas the Lg-2 genes and B-legumins correspond to the Group-II glycinins (Casey et al., 1986).

A comparison of the distribution of legumin genes in the pea genome with that of the glycinin genes in soybeans is informative. The Lg-1 locus in pea contains five genes, four of which are in a direct tandem repeat, whereas the Lg-2 locus contains at least three genes (Casey et al., 1986). Hence, in pea, one locus is involved in the production of Group-I subunits and the other one produces the subunits that comprise Group-II. In soybean there are four loci: two each that encode Group-I and Group-II subunits, respectively. Although it may be fortuitous that there are twice as many loci involved in the production of 11S subunits in soybeans compared with pea, this arrangement may also be related to differences between these two species in genome organization. It has been speculated that soybeans are stable tetraploids with a diploidized genome, whereas *Pisum sativum* is considered a true diploid. The notion that soybeans are stable tetraploids is

supported by other findings. These include (1) duplicate factor inheritance at a number of loci in the soybean genome (Bernard and Weiss, 1973), (2) the absence of a detrimental effect of trisomics (Palmer, 1976), and (3) the fact that *Glycine* is the only genus in the tribe *Phaseoleae* where the chromosome number is either 40 or 80, and not 20 (Lackey, 1980). Evidence at the DNA level also supports the presumption that soybean is a stable tetraploid. Heteroduplex analysis has shown that the two Group-I glycinin loci share extensive homology in both the 5'-flanking and 3'-flanking regions, similar to what might be expected from homoeologous chromosomes (Nielsen et al., 1989).

There is only limited information in the literature that pertains to the comparative location of seed protein genes in genomes of different plant species. One of the best studied is the high molecular weight prolamine gene clusters found in wheat, barley, and rye. In wheat, they are found in three complex loci (Glu-A1, Glu-B1, and Glu-D1) located on the long arms of the homoeologous chromosomes 1A, 1B, and 1D (Payne, 1987). In barley, the high molecular weight prolamines are encoded by genes at the Hor 3 locus on the long arm of chromosome 5. Chromosome 5 of barley is homologous to chromosome 1 in wheat, and the Hor 3 and Glu-I loci are located at equivalent positions on the long arms of their respective chromosomes (Shewry et al., 1983). A similar situation exists for chromosome 1R in rye (Lawrence and Shepherd, 1981; Singh and Shepherd, 1984). Hence, in these three members of the tribe *Triticeae*, at least, loci with comparable functions occupy equivalent physical positions in their respective genomes.

Similar information about the location of the storage protein genes in genomes of members in tribe *Phaseoleae* is lacking. The Lg-1 locus in pea is located about 11 map units away from the wrinkled (*r*) locus on chromosome 7 (Davies, 1980; Matta and Gatehouse, 1982; Domoney et

Table 2. χ^2 Values^a for Test of Linkage of Glycinin Genes to Phenotypic Marker Genes in the F₂ Generation of Raiden × Forrest

Marker Gene	<i>t</i>	<i>w</i> ₁	<i>i</i>	<i>r</i>	<i>l</i> ₂
<i>Gy</i> ₁₊₂	1.836 (<i>P</i> > 0.075) (<i>n</i> = 158)	0.790 (<i>P</i> > 0.97) (<i>n</i> = 161)	2.464 (<i>P</i> > 0.75) (<i>n</i> = 98)	0.010 (<i>P</i> > 0.99) (<i>n</i> = 98)	11.07 (<i>P</i> > 0.05) (<i>n</i> = 119)
<i>Gy</i> ₃	1.725 (<i>P</i> > 0.75) (<i>n</i> = 166)	4.456 (<i>P</i> > 0.25) (<i>n</i> = 169)	1.856 (<i>P</i> > 0.75) (<i>n</i> = 103)	2.070 (<i>P</i> > 0.75) (<i>n</i> = 103)	0.987 (<i>P</i> > 0.95) (<i>n</i> = 128)
<i>Gy</i> ₄	0.834 ^b (<i>P</i> > 0.75) (<i>n</i> = 190)	0.006 ^b (<i>P</i> > 0.99) (<i>n</i> = 197)	0.472 ^b (<i>P</i> > 0.90) (<i>n</i> = 122)	0.868 ^b (<i>P</i> > 0.75) (<i>n</i> = 122)	0.539 ^b (<i>P</i> > 0.75) (<i>n</i> = 147)
<i>Gy</i> ₅	0.779 (<i>P</i> > 0.97) (<i>n</i> = 154)	2.706 (<i>P</i> > 0.50) (<i>n</i> = 158)	2.849 (<i>P</i> > 0.50) (<i>n</i> = 98)	0.322 (<i>P</i> > 0.99) (<i>n</i> = 98)	2.181 (<i>P</i> > 0.75) (<i>n</i> = 117)

^a χ^2 goodness-of-fit tests to either a 3:6:3:1:2:1 or 9:3:3:1 ratios were performed to evaluate a null hypothesis for independent segregation of two genetic loci.

^b Yates' correction term was applied when χ^2 values were calculated.

al., 1986), whereas Lg-2 occupies a locus approximately 7 map units from a-locus on chromosome 1 (Domoney et al., 1986). The pea genome contains a third legumin locus (Lg-3), which is also linked to the a-locus. It produces a precursor that is larger than those generated from genes in the other two loci, and is encoded in clone pCD32 (R. Casey, personal communication). When DNA gel blot transfers of restricted Forrest genomic DNA were probed at low stringency (at 42°C in 5 × SSC, no formamide) with the insert from pCD32, a strong reproducible signal was not observed (T.-J. Cho, unpublished observations). Hence, we were unable to obtain evidence for a locus in soybean that was similar to the Lg-3 locus in pea.

In addition to segregating independently from one another, the four glycinin loci also segregated independently from several linkage group markers. The loci responsible for *T/t*, *W₁/w₁*, *I/i*, *R/r*, and *L₂/l₂* are associated with linkage groups 1, 8, 7, 2, and an unknown one, respectively. Hence, either the glycinin loci are linked to one of the other linkage groups presently recognized in soybean, or else they were sufficiently far removed from the markers chosen so that they segregated independently. Additional work will be required to establish the location of these loci in the soybean genome.

METHODS

Plant Material

Crosses between the varieties Forrest and Raiden were made during December, 1985. The F₁ seeds that resulted were planted in the greenhouse the following fall and F₂ seeds were obtained by self-pollination. Just prior to planting the F₂ seeds, a small chip opposite the embryo was removed for analysis of protein (see below). Leaf DNA was purified from individual F₂ plants and the

plants were scored for both flower color and grey versus tawny pubescence. F₃ seeds were obtained from selfed F₂ plants and scored for hilum pigmentation. The pods containing the F₃ seeds were scored for pod color.

DNA Methodology

Leaf DNA was isolated as described by Dellaporta et al. (1983), and purified by CsCl density gradient centrifugation. The DNA was digested with restriction endonucleases using the specifications of the manufacturer. For genomic hybridization, 3 μg to 5 μg of restricted DNA was fractionated in 0.6% agarose gels and blot-transferred to nylon membranes (GeneScreen Plus) as described by New England Nuclear, Inc. (Catalog NEF-976). The membranes were prehybridized at 55°C in 50% formamide, 5 × SSC (0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.0), 5 × Denhardt's solution (0.1% bovine serum albumin, 0.1% Ficoll, and 0.1% polyvinylpyrrolidone), 1% SDS, 50 mM sodium phosphate, pH 6.5, 500 μg/ml denatured salmon sperm DNA, 1 mM EDTA, and 1 μg/ml poly(dG)-poly(dC). Hybridization was performed using nick-translated inserts from the following cDNA or genomic clones: pG134 (*Gy*₁ cDNA), pG466 (*Gy*₅ cDNA), and pG6H4.7 (*Gy*₃ genomic DNA). The origin of probes used in this study and physical maps of pG134, pG466, and pG6H4.7 have been described in Figure 1. Plasmid pG6H4.7 was a subclone of genomic clone λDA28-6 (Nielsen et al., 1989). Conditions were chosen for washing the membranes such that there would be minimal cross-hybridization between homologous gene sequences. Unless stated to the contrary, the filters were washed either two or three times for 1 hr at room temperature in 2 × SSC and 0.1% SDS, and then at 65°C for 2 hr in 0.1 × SSC and 0.1% SDS. The filters were autoradiographed with an intensifier screen either 2 days or 3 days.

Seed Protein Analysis

Individual F₂ seeds were chipped with a razor blade to remove approximately 5 mg of the cotyledon on the side of the seed

opposite the embryo. The chips were cut into small pieces and added to 0.5 ml of a sample buffer consisting of 0.05 M Tris-HCl, pH 8, 0.2% SDS, and 5 M urea. After 30 min at room temperature, the extracts were analyzed by SDS-polyacrylamide gel electrophoresis as described by Kitamura et al. (1984). The system contained a 5% stacking gel (2.5 cm) and a 7% to 13% running gel (7.5 cm).

ACKNOWLEDGMENTS

We thank Laura Coppoc for technical assistance. The experiments reported in this paper were the result of cooperative research between the U.S. Department of Agriculture/Agricultural Research Service and the Indiana Agricultural Experimental Station. Research was supported in part by grants from the American Soybean Association and Central Soya, Inc. This is Journal Paper No. 11,934 from the Indiana Agricultural Experiment Station.

Received January 23, 1989.

REFERENCES

- Allard, R.W. (1956). Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia* **24**, 235–278.
- Bernard, R.L., and Weiss, M.G. (1973). Qualitative genetics. In *Soybeans: Improvement, Production and Uses*, Agronomy Monograph 16, B.E. Caldwell, ed. (Madison, WI: American Society of Agronomy, Inc.), pp. 117–154.
- Brown, J.W.S., Bliss, F.A., and Hall, T.C. (1981). Linkage relationships between genes controlling seed proteins in french bean. *Theor. Appl. Genet.* **60**, 251–259.
- Casey, R., Domoney, C., and Ellis, T.H.N. (1986). Legume storage proteins and their genes. *Oxford Surv. Plant Mol. Cell Biol.* **3**, 1–95.
- Cho, T.-J., Davies, C.S., Fischer, R.L., Tumer, N.E., Goldberg, R.B., and Nielsen, N.C. (1989b). Molecular characterization of an aberrant allele for the *Gy₂* glycinin gene: A chromosomal rearrangement. *Plant Cell* **1**, 339–350.
- Davies, D.R. (1980). The *r_s* locus and legumin synthesis in *Pisum sativum*. *Biochem. Genet.* **18**, 1207–1219.
- Davies, C.S., Cho, T.-J., and Nielsen, N.C. (1987). Tests for linkage of seed protein genes *Lx₁*, *Lx₃*, and *Cgy₁* to markers on known linkage groups. *Soybean Genet. Newsletter* **14**, 248–252.
- Dellaporta, D.L., Wood, J., and Hicks, J.B. (1983). A plant DNA mini-preparation: Version II. *Plant Mol. Biol. Rep.* **1**, 19–21.
- Domoney, C., Ellis, T.H.N., and Davies, D.R. (1986). Organization and mapping of legumin genes in *Pisum*. *Mol. Gen. Genet.* **202**, 280–285.
- Fontes, E.P.B., Moreira, M.A., Davies, C.S., and Nielsen, N.C. (1984). Urea-elicited changes in relative electrophoretic mobility of certain glycinin and β -conglycinin subunits. *Plant Physiol.* **76**, 840–842.
- Heidecker, G., and Messing, J. (1986). Structural analysis of plant genes. *Annu. Rev. Plant Physiol.* **37**, 439–466.
- Kitamura, K., Davies, C.S., and Nielsen, N.C. (1984). Inheritance of alleles for *Cgy₁* and *Gy₄* storage protein genes in soybean. *Theor. Appl. Genet.* **68**, 253–257.
- Kreis, M., Shewry, P.R., Forde, B.G., Forde, J., and Mifflin, B.J. (1985). Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley and rye. *Oxford Surv. Plant Mol. Cell Biol.* **2**, 253–317.
- Lackey, J.A. (1980). Chromosome numbers in the *Phaseoleae* (*Fabaceae:Faboideae*) and their relation to taxonomy. *Am. J. Bot.* **67**, 595–602.
- Lawrence, G.J., and Shepherd, K.W. (1981). Chromosomal location of genes controlling seed proteins in species related to wheat. *Theor. Appl. Genet.* **59**, 25–31.
- Matta, N.K., and Gatehouse, J.A. (1982). Inheritance and mapping of storage proteins genes in *Pisum sativum* L. *Heredity* **48**, 383–392.
- Moreira, M.A., Hermodson, M.A., Larkins, B.A., and Nielsen, N.C. (1979). Partial characterization of the acidic and basic polypeptides of glycinin. *J. Biol. Chem.* **254**, 9921–9926.
- Moreira, M.A., Hermodson, M.A., Larkins, B.A., and Nielsen, N.C. (1981). Comparison of the primary structure of the acidic polypeptides of glycinin. *J. Biol. Chem.* **210**, 633–642.
- Nielsen, N.C. (1984). *The chemistry of legume storage proteins*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **304**, 287–296.
- Nielsen, N.C. (1985). The structure and complexity of the 11S polypeptides in soybeans. *J. Am. Oil Chem. Soc.* **62**, 1680–1686.
- Nielsen, N.C., Dickinson, C.D., Cho, T.-J., Thanh, V.H., Scallion, B.J., Sims, T.L., Fischer, R.L., and Goldberg, R.B. (1989). Characterization of the glycinin gene family. *Plant Cell* **1**, 313–328.
- Palmer, R.G. (1976). Chromosome transmission and morphology of three primary trisomics in soybeans (*Glycine max*). *Can. J. Genet. Cytol.* **18**, 131–140.
- Palmer, R.G., and Kilen, T.C. (1987). Qualitative genetics and cytogenetics. In *Soybeans: Improvement, Production, and Uses*, Agronomy Monograph 16, 2nd ed, J.R. Wilcox, ed. (Madison, WI: American Society of Agronomy, Inc.), pp. 135–209.
- Payne, P.I. (1987). Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annu. Rev. Plant Physiol.* **38**, 141–153.
- Scallion, B.J., Thanh, V.H., Floener, L.A., and Nielsen, N.C. (1985). Identification and characterization of DNA clones encoding group-II glycinin subunits. *Theor. Appl. Genet.* **70**, 510–519.
- Scallion, B.J., Dickinson, C.D., and Nielsen, N.C. (1987). Characterization of a null-allele for the *Gy₄* glycinin gene from soybean. *Mol. Gen. Genet.* **208**, 107–113.
- Shewry, P.R., Finch, R., Parmar, S., Franklin, J., and Mifflin, B.J. (1983). Chromosomal location of Hor 3, a new locus governing storage proteins in barley. *Heredity* **50**, 179–189.
- Shewry, P.R., Mifflin, B.J., and Kasarda, D.D. (1984). The structural and evolutionary relationships of the prolamins storage proteins of barley, rye and wheat. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **304**, 297–308.

- Singh, N.K., and Shepherd, K.W.** (1984). Mapping of the genes controlling the high-molecular-weight subunits of rye on the long arm of chromosome 1R. *Genet. Res.* **44**, 117–123.
- Soave, C., and Salamini, F.** (1984). Organization and regulation of zein genes in maize endosperm. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **304**, 341–347.
- Staswick, P.E., Hermodson, M.A., and Nielsen, N.C.** (1981). Identification of the acidic and basic subunit complexes of glycinin. *J. Biol. Chem.* **256**, 8752–8755.
- Staswick, P.E., Broue, P., and Nielsen, N.C.** (1983). Characterization of a soybean cultivar lacking certain glycinin subunits. *Arch. Biochem. Biophys.* **223**, 1–8.