

## Effect of Six Decades of Selective Breeding on Soybean Protein Composition and Quality: A Biochemical and Molecular Analysis

AHMED A. MAHMOUD,<sup>†</sup> SAVITHIRY S. NATARAJAN,<sup>‡</sup> JOHN O. BENNETT,<sup>§</sup>  
THOMAS P. MAWHINNEY,<sup>||</sup> WILLIAM J. WIEBOLD,<sup>†</sup> AND HARI B. KRISHNAN<sup>\*,†,§</sup>

Plant Science Division and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211, Soybean Genomics & Improvement Laboratory, USDA-ARS, Beltsville, Maryland 20705, and Plant Genetics Research Unit, Agricultural Research Service, USDA, Columbia, Missouri 65211

To evaluate the extent of the genetic change and its effects on the seed protein composition of soybean cultivars released during the past 60 years, representative ancestral cultivars and those derived from selective breeding were grown in a side-by-side comparison. Total seed protein content, determined by combustion analysis of nitrogen, revealed a decline in the protein content after decades of selection and breeding. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis comparison of protein profiles of the soybean cultivars indicated that relative expression of most of the seed storage proteins had not varied substantially from the ancestral lines to the present commercial cultivars. There was noticeably less  $\beta$ -subunit of  $\beta$ -conglycinin, a protein devoid of sulfur amino acids, in the modern cultivars represented by Mustang, Pioneer 93B09, and Asgrow 3602. Comparison of the amino acid profiles of soybean seed, a benchmark of the protein's nutritional quality, revealed that the ancestral progenitor, *G. soja*, was significantly higher in cysteine, glutamic acid, histidine, and arginine than either the ancestral or the modern cultivars. Selective breeding over the past 60 years minimally affected the overall amino acid composition. The degree of divergence in the DNA sequence of the genes encoding glycinin and  $\beta$ -conglycinin in the ancestral and modern cultivars was investigated using Southern hybridization and the polymerase chain reaction. Even though some restriction fragment polymorphisms could be detected, overall, the banding patterns were remarkably similar among the ancestral cultivars and those derived from them, suggesting a high degree of conservation of seed-storage protein genes. The results of our study suggest that selection and breeding for yield during the past 60 years had no major influence on the protein composition, ostensibly because of limited genetic diversity among the parental lines.

**KEYWORDS:**  $\beta$ -Conglycinin; genetic diversity; glycinin; protein composition; protein quality

### INTRODUCTION

A plethora of writings, both factual and legendary, dealing with the history of the soybean [*Glycine max* (L.) Merr.] have been recently reviewed (1). Genetic and biochemical evidence indicates that cultivated soybean was derived from *Glycine soja*, which, in turn, is thought to have evolved from perennial vining plants (2). Archeological evidence points to domestication and cultivation of soybeans some 3000 years ago in northeast China, with the crop spreading to Japan and Southeast Asia along trading routes (3, 4). In 1765, the seafarer Samuel Bowen

brought soybeans to the United States from China and used them in the manufacture of a fermented product (5). Primarily used as a forage crop in the early part of the 20th century, soybean production for grain began in earnest in the United States and, by the 1940s, was supplying previously imported edible oil.

In the past, breeding programs have primarily focused on increasing the yield of the crop grown under regional climatic conditions and cultural practices (6, 7). Selective breeding using elite high-yielding parents has resulted in the production of hundreds of improved cultivars (8–10). Yield has steadily increased over this time period, with accelerated improvement seen in the last 25 years (11). However, the effect of selective breeding on protein composition and quality throughout this time period has not been thoroughly evaluated. Protein composition and quality are cultivar-dependent and can be significantly affected by environmental conditions (10, 12–15). To minimize the environmental effects, it is essential to examine the effect

\* Corresponding author. Address: Dr. Hari B. Krishnan, Plant Genetics Research Unit, USDA-ARS, 108W Curtis Hall, University of Missouri, Columbia, MO 65211. Phone: 573-882-8151. Fax: 573-884-7850. E-mail: KrishnanH@missouri.edu.

<sup>†</sup> Plant Science Division, University of Missouri.

<sup>‡</sup> Soybean Genomics & Improvement Laboratory, USDA-ARS.

<sup>§</sup> Plant Genetics Research Unit, Agricultural Research Service, USDA.

<sup>||</sup> Department of Biochemistry, University of Missouri.

**Table 1.** Ancestral Soybean Cultivars and Modern Cultivars Utilized in This Study

cultivar/ accession	maturity group	pedigree	origin	year of release
<i>Glycine soja</i>	perennial		China	
A. K.	III	ancestral genotype	China	1917
Mandarin	I	ancestral genotype	China	1920
Mukden	II	ancestral genotype	China	1921
Lincoln	III	ancestral genotype	China	1943
Illini	III	ancestral genotype	Illinois	1921
Adams	II	Illini × Dunfield	Iowa	1949
Clark	IV	Lincoln × Richland	Illinois	1953
Harosoy 63	II	Harosoy × Blackhawk	Illinois	1963
Wayne	III	L64-4091 [(Lincoln × Richland) × (Lincoln × CNS)] × Clark	Illinois	1964
Custer	IV	{[(Peking × Scott) × (i-I Rhg4 line from Peking Scott)] × Scott × Blackhawk} × (Peking × Scott)	Missouri	1967
Williams	III	Wayne × L57-0034 (Clark × Adams)	Illinois	1971
Franklin	IV	L12 (Clark 63 isolate) × Custer	Illinois	1977
Pyramid	IV	Franklin × J75-5{Forrest × [D68 (Dyer × Bragg) × PI88788]}	Southern Illinois	1985
Fayette	III	Williams × PI88788	Illinois	1987
Mustang	IV	Fayette × Pyramid	Missouri	1996
Pioneer 93B09	III			2002
Asgrow 3602	III			2004

of selective breeding on protein composition of cultivars released during the past 60 years grown at one location. Herein, we report the results of a study conducted at one location showing the effect of six decades of selective breeding on soybean protein composition and quality.

## MATERIALS AND METHODS

**Plant Materials and Growing Conditions.** A collection of ancestral cultivars and cultivars subsequently developed from them, for the Midwest United States, was grown at the University of Missouri Bradford Research and Extension Center near Columbia, MO, during the 2004 and 2005 seasons. Because the plots harbored *Bradyrhizobium* sp., amendment of the soil with commercial inoculants was not performed. Cultural practices were typical of those utilized for soybean production the Midwest United States. Cultivar name, pedigree, year of release, and maturity group of the genotypes are presented in **Table 1**.

**Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS–PAGE).** Seeds collected from the 2004 and 2005 harvests were ground, and each of four replicates of 15 mg of seed powder was extracted with 1 mL of SDS sample buffer [60 mM Tris-HCl, (pH 6.8), 2% SDS (w/v), 10% glycerol (v/v), 0.03 mM bromophenol blue, and 5% 2-mercaptoethanol (v/v)]. The slurry was heated in a boiling water bath for 5 min and subjected to centrifugation (15 800g, 15 min). Ten-microliter aliquots of the supernatant were loaded onto a 13.5% resolving gel (w/v) and fractionated by SDS–PAGE (16) using a Hoefer SE 260 minigel apparatus (Amersham Biosciences, Piscataway, NJ). Gels were stained overnight with Coomassie Blue G-250. After being destained for 15 min in 50% methanol/10% glacial acetic acid (v/v), the gels were transferred to 10% glacial acid (v/v) prior to visualization.

**Amino Acid Analysis.** Four independent aliquots of soybean seed powder from each cultivar were hydrolyzed in 6 N HCl at 155 °C for 16 h under a nitrogen atmosphere. Prior to hydrolysis, the sulfur-containing amino acids, methionine and cysteine, were quantified from duplicate samples oxidized with performic acid. Amino acids were separated using a cation-exchange resin column in conjunction with a Beckman 6300 Amino Acid Analyzer equipped with a postcolumn ninhydrin reaction detection system (Beckman Instruments, Fullerton, CA).

**Southern Hybridization.** Total genomic DNA was isolated from leaf tissues of the genotypes listed in **Table 1** using the CTAB method (17). Eight micrograms of genomic DNA was digested with the restriction enzyme *EcoR* I overnight at 37 °C. After size fractionation by electrophoresis on a 0.8% agarose gel, the DNA was transferred to

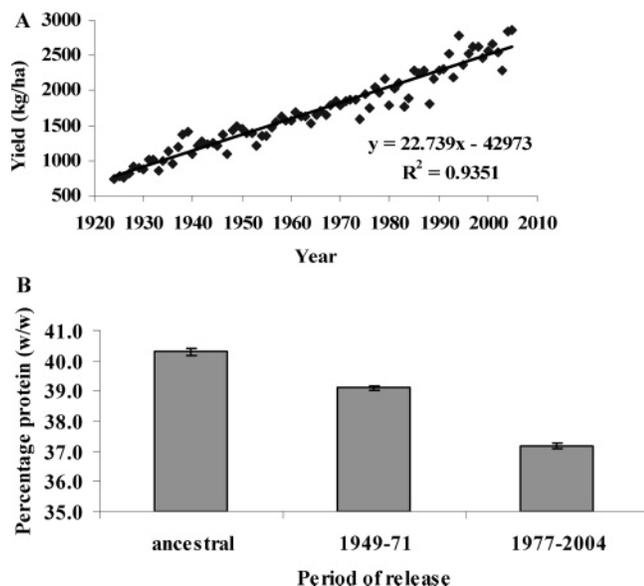
nylon membranes by capillary transfer using 0.4 M NaOH. The  $\beta$ -subunit of  $\beta$ -conglycinin EST clone and partial sequences of PCR-amplified glycinin *Gy2* and *Gy4* were selected as probes for Southern hybridization. The primers used for PCR amplification were as follows: *Gy2* forward 5' TTCGCCCTGAATTCTTGAAAGAAGCG 3', reverse 5' CTCAAAGTTATCGCTCTGGGATTT 3'; *Gy4* forward 5' TTACCTTGCTGGGAACCCAG 3', reverse 5' GGACAACATAT-TGGGCACTGA 3'. Utilizing total RNA from developing seeds as a template, the coding regions were amplified by reverse-transcriptase PCR (RT-PCR). PCR products with the predicted molecular size were purified and cloned into the pGEM-T easy vector (Promega Corp., Madison, WI) and confirmed by sequencing. Probe DNA was labeled with <sup>32</sup>P-dCTP using the Ladderman labeling kit (Takara Bio Inc., Shiga, Japan). Prehybridization of at least 10 h and overnight hybridization were carried out at 65 °C in 6X SSPE buffer (1X SSPE is 0.1 M NaCl, 0.01 M NaH<sub>2</sub>PO<sub>4</sub>, and 0.001 M EDTA), 5X Denhardt's solution, 0.5% SDS, and 50  $\mu$ g/mL sheared and denatured salmon sperm DNA (ssDNA). After hybridization, membranes were washed three times in 2X SSPE and 0.5% SDS at room temperature for 10 min and one or two times in 0.1X SSPE and 0.1% SDS at 65 °C for 30 min and then exposed to X-ray films at –80 °C.

**PCR Amplification of the Hypervariable Regions (HVRs) from Group-2 Glycinin Genes.** The primers for PCR amplification of *Gy4* gene described above were designed to amplify exon 3, which contains the hypervariable region from group-2 glycinin *Gy4* and *Gy5* genes (18). PCR amplification was carried out using Ready To Go PCR Beads (Amersham Pharmacia Biotech, Inc., Piscataway, NJ), following the supplier's instructions. The PCR amplification program consisted of one cycle of 95 °C (4 min), followed by 35 cycles of 94 °C (1 min), 60 °C (1 min), and 72 °C (1 min) and a final 5 min at 72 °C. After the amplification process, the products were separated on a 2% agarose gel.

**Statistical Analysis.** Analysis of variance was used to ascertain whether the amino acid contents of the seed protein and total seed protein of the wild progenitor, cultivated ancestral varieties, and modern varieties of soybean exhibited differences. Data were analyzed using SAS (8.2 version) with PROC General Linear Model (SAS Institute, Inc., Cary, NC).

## RESULTS

**Decline in Total Protein Is Seen in Modern Cultivars.** The existing compilation of data clearly shows that North American soybean yields have steadily increased over the past several decades (USDA NASS <http://151.121.3.33:8080/QuickStats/>) (**Figure 1A**). This has been attributed to both improved genetics through selective plant breeding and advanced agronomic practices (10, 11, 19, 20). However, the observed yield improvement has not been accompanied by increases in protein content, apparently because of an inverse relationship existing between these two characteristics (21–25). Genetics and environmental conditions influence the protein content and composition of soybean (26–29). In this study, we sought to minimize the disparity in protein accumulation and composition as related to environmental conditions by growing plants at one location. Seed protein content from representative ancestral cultivars and those derived from selective breeding during the past 60 years (**Table 1**) were determined by combustion analysis of nitrogen (**Figure 1B**). Protein content appears to have diminished over the past 60 years. This difference in protein content becomes evident when the cultivars are divided into three groups based on their period of release in the United States. The first group, consisting of A. K., Mandarin, Mukden, Lincoln, and Illini, some of which form the basis of the North American gene pool, contained 38–43% protein. The second group, Adams Clark, Harosoy 63, Wayne, and Williams, which were derived from crosses of the ancestral cultivars, had 36–41% protein. More recent releases, namely, Franklin, Fayette,



**Figure 1.** Yields and protein contents of historical soybeans. (A) Soybean yields from 1924 to 2005 reported by the United States Department of Agriculture National Statistics Service Quick Stats. (B) Average seed protein content of ancestral cultivars, some early releases (1949–1971), and more recent cultivars (1977–2004).

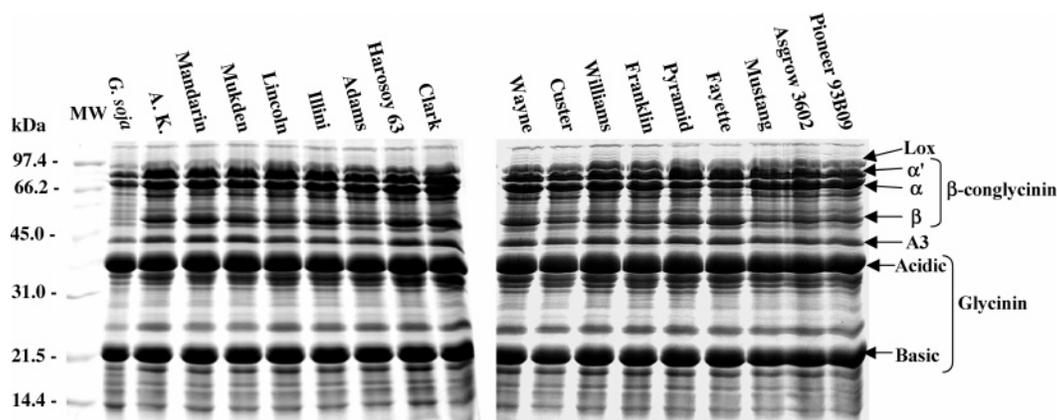
Mustang, Asgrow 3602, and Pioneer 93B09, representing the third group, ranged in protein content between 36 and 39% (Figure 1B). Conversely, *G. soja* the progenitor of modern soybeans contained 44% protein, which is higher than the contents of all cultivars examined in this study.

**Comparison of Protein Profiles of Ancestral and Modern Cultivars.** Because protein content was observed to diminish over the past 60 years, we wanted to determine whether protein composition had also been affected. Results from the fractionation of seed proteins by SDS–PAGE indicate that the relative expression of each seed storage protein did not vary substantially from that of the ancestral lines to that of the present commercial cultivars (Figure 2). Although chemical analysis shows *G. soja* to have the highest protein content (Figure 1B), this is not reflected in the SDS–PAGE gel (Figure 2). This discrepancy might be due to the presence of phenolic compounds in *G. soja* seeds that are known to interfere with effective extraction of the seed protein (30). The  $\beta$ -subunit of  $\beta$ -conglycinin showed little or no accumulation in the progenitor *G. soja* when compared to the ancestral and modern cultivars (Figure 2). The

low abundance of the  $\beta$ -subunit of  $\beta$ -conglycinin in *G. soja* is not surprising given that earlier work showed that some *G. soja* accessions lack this subunit (31). In the modern cultivars represented by Mustang, Pioneer 93B09, and Asgrow 3602, there is noticeably less  $\beta$ -subunit of  $\beta$ -conglycinin (Figure 2).

**Comparison of Amino Acid Compositions of Ancestral and Modern Cultivars.** Soybeans are a good protein source for monogastric nutrition. However, increasing the concentration of certain amino acids such as methionine, cysteine, threonine, tryptophan, isoleucine, valine, and arginine would obviate the use of expensive synthetics necessary to build complete rations for growing swine and poultry (32, 33). To determine whether changes in the relative content of amino acids have occurred during the past 60 years, seeds of the wild progenitor, *G. soja*; representative cultivars of ancestral introductions; and the modern cultivars were analyzed. Statistical differences did exist in amino acid content among the seeds from the tested cultivars (Table 2). Pairwise comparisons revealed that the ancestral progenitor, *G. soja*, was significantly higher in the amino acids cysteine, glutamic acid, histidine, and arginine than either the ancestral or modern cultivars. *G. soja* was significantly lower in the amino acids alanine, phenylalanine, and leucine and contained approximately 60% the amount of tryptophan as the domesticated cultivars (Table 2). With the exception of *G. soja*, significantly higher concentrations of methionine were found in the modern cultivars Asgrow 3602 and Pioneer 93B09 as well as Franklin. Phenylalanine and leucine declined, whereas alanine and glycine showed modest increases in the cultivars representative of successive years of selection and breeding. The sulfur amino acids showed opposing trends: methionine appears to be increasing, whereas cysteine is declining (Table 2).

**Genotypic Variation in the Genes Encoding Seed Storage Proteins at the DNA Level.** The degree of divergence in the DNA sequence of the genes encoding glycinin and  $\beta$ -conglycinin in the selected collection of U.S. soybean cultivars was investigated using Southern hybridization and PCR. In Southern hybridization, the coding regions of glycinin and  $\beta$ -conglycinin genes were used as probes to hybridize *Eco*R I-digested genomic DNA. Five major genes encode the different subunits of the hexameric glycinin proteins, which are divided into two groups based on DNA sequence similarities (18). Group-1 glycinins include three genes, *Gy1*, *Gy2*, and *Gy3*, whereas group-2 includes two genes, *Gy4* and *Gy5*. In addition, two other glycinin-related genes, a functional glycinin gene, *Gy7*, and a

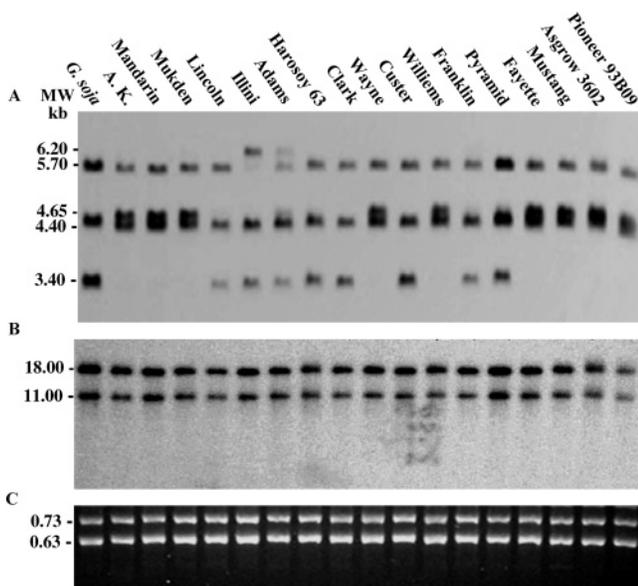


**Figure 2.** Protein profiles of ancestral cultivars and representative modern cultivars. Total seed proteins extracted from the wild progenitor, *G. soja*; ancestral; and modern cultivars released during the past six decades were fractionated by SDS–PAGE on a 13.5% gel and stained with Coomassie Blue. The most abundant seed storage proteins are identified.

**Table 2.** Relative Percentages of Individual Amino Acids as Related to Total Amino Acid Content

cultivar	Ala	Arg	Asp	Cys	Glu	Gly	His	Iso	Leu	Lys	Met	Phe	Pro	Ser	Thr	Try	Tyr	Val
<i>G. soja</i>	4.14	8.27	11.67	1.98	18.29	4.43	2.85	4.51	7.45	6.71	1.50	4.79	5.23	4.61	3.66	0.80	3.56	4.90
A. K.	4.34	7.38	11.43	1.74	17.83	4.26	2.72	4.84	7.91	6.74	1.42	5.19	5.13	4.64	3.84	1.40	3.72	5.09
Mandarin	4.22	7.83	11.66	1.67	18.11	4.23	2.72	4.80	7.90	6.61	1.37	5.26	5.21	4.50	3.70	1.12	3.65	5.10
Mukden	4.25	7.67	11.49	1.66	18.08	4.27	2.71	4.85	7.90	6.65	1.40	5.23	5.18	4.39	3.71	1.25	3.67	5.20
Lincoln	4.34	7.27	11.46	1.69	17.58	4.28	2.73	4.84	8.02	6.74	1.42	5.19	5.10	4.62	3.87	1.43	3.82	5.15
Illini	4.30	7.39	11.52	1.68	17.87	4.24	2.69	4.80	7.94	6.68	1.41	5.21	5.12	4.69	3.85	1.46	3.69	5.05
Adams	4.35	7.29	11.47	1.77	17.77	4.27	2.70	4.84	7.87	6.75	1.44	5.16	5.17	4.62	3.87	1.46	3.72	5.12
Clark	4.27	7.30	11.59	1.69	17.95	4.22	2.72	4.82	8.00	6.71	1.42	5.15	5.21	4.67	3.90	1.45	3.75	5.10
Harosoy 63	4.32	7.56	11.45	1.69	17.70	4.27	2.70	4.62	7.93	6.59	1.40	5.19	5.15	4.87	3.89	1.37	3.68	4.87
Wayne	4.31	7.55	11.51	1.57	17.95	4.29	2.71	4.73	7.94	6.64	1.39	5.16	5.18	4.74	3.88	1.40	3.66	5.06
Custer	4.36	7.34	11.37	1.69	17.58	4.34	2.70	4.73	7.88	6.69	1.49	5.15	5.22	4.84	3.98	1.46	3.72	5.03
Williams	4.29	7.44	11.46	1.56	17.89	4.27	2.70	4.67	7.90	6.64	1.45	5.12	5.23	4.87	3.96	1.45	3.74	4.97
Franklin	4.35	7.40	11.51	1.75	17.43	4.29	2.72	4.63	7.89	6.74	1.50	5.08	5.17	4.92	4.01	1.51	3.76	4.93
Pyramid	4.36	7.33	11.44	1.67	17.41	4.34	2.78	4.65	7.95	6.71	1.49	5.15	5.20	4.94	4.03	1.47	3.72	4.97
Fayette	4.35	7.28	11.46	1.62	17.43	4.37	2.76	4.70	7.95	6.60	1.44	5.22	5.22	4.91	4.02	1.39	3.79	5.09
Mustang	4.35	7.47	11.58	1.58	17.84	4.38	2.71	4.72	7.90	6.67	1.48	4.98	5.23	4.76	3.98	1.40	3.50	5.04
Asgrow 3602	4.40	7.37	11.52	1.64	17.95	4.44	2.65	4.77	7.79	6.71	1.52	4.96	5.15	4.66	3.88	1.58	3.50	5.12
Pioneer 93B09	4.31	7.72	11.61	1.65	17.94	4.37	2.69	4.66	7.73	6.66	1.51	4.94	5.19	4.69	3.84	1.46	3.49	5.02
LSD <sup>a</sup>	0.05	0.14	0.09	0.05	0.17	0.05	0.03	0.15	0.05	0.05	0.02	0.11	0.12	0.25	0.10	0.19	0.12	0.15

<sup>a</sup> Comparisons within columns are by LSD at the  $P \leq 0.05$  level.



**Figure 3.** Southern blot analysis of soybean glycinin. Genomic DNA from the leaf tissue of *G. soja* and ancestral and modern cultivars was digested by *EcoR* I, fractionated on a 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with <sup>32</sup>P-labeled (A) *Gy2* and (B) *Gy4* glycinin probes and then subjected to autoradiography. (C) PCR amplification of the hypervariable regions (HVR) from the *Gy4* and *Gy5* genes. The molecular weight markers in kilobases are shown on the left of the figure.

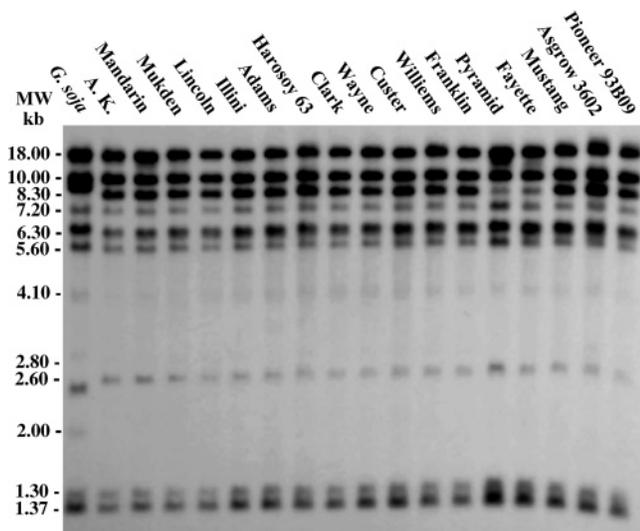
pseudogene, *Gy6*, have recently been identified and shown to have limited sequence homology to group-1 glycinin genes (34).

The Southern hybridization profile of the group-1 glycinin probe, presented in **Figure 3A**, showed hybridization to three to five different DNA fragments with varying banding patterns as well as hybridization intensities among the tested genotypes. The group-1 glycinin probe used in this study corresponds to the carboxyl terminus (including exons 3 and 4), which lack restriction sites for *EcoR* I in the previously characterized group-1 glycinin genes isolated from soybean cultivar “Dare” (18). Thus, the probe should hybridize to three fragments, each corresponding to one of the group-1 glycinin genes. The Southern hybridization profile presented in **Figure 3A** shows that the probe hybridized to three main fragments in most of

the genotypes with polymorphism in the size of these fragments, indicating sequence variation among the genotypes for group-1 glycinin genes. The genotypes could be grouped on the basis of their hybridization patterns into three groups; group 1 includes A. K., Mandarin, Mukden, Wayne, Williams, Fayette, Mustang, Asgrow 3602, and Pioneer 93B09; group 2 comprises *G. soja*, Lincoln, Harosoy 63, Clark, Custer, Franklin, and Pyramid; and group 3 contains Illini and Adams (**Figure 3A**). Of the three main hybridized fragments in both group-1 and group-2 genotypes, two fragments (5.70 and 4.40 kb) were in common, except for *G. soja*, which showed a slightly higher-molecular-weight fragment (4.50 kb) instead of the 4.40-kb fragment. The third fragment was polymorphic between the two groups (4.65 kb in group 1 compared to 3.40 kb in group 2). Group-3 genotypes, Illini and Adams, were more distinct in their hybridization pattern compared to the other genotypes in both the number and intensity of hybridized fragments. The less intense hybridization to the 5.70-kb fragment, which was strongly hybridized in all other genotypes, and the presence of a novel 6.20-kb fragment with different hybridization intensities in Illini and Adams (**Figure 3A**) suggest the presence of unique alterations in the DNA sequence and/or the genomic organization of one or more of the glycinin genes in these two varieties. It should be noted that Illini is one of the two parents of the cross from which Adams was selected (**Table 1**).

Unlike the group-1 glycinin genes, the Southern hybridization profile of the group-2 genes, *Gy4* and *Gy5*, did not reveal any DNA sequence variation with the probe/restriction enzyme combination used in our study (**Figure 3B**). The probe corresponding to exon 3, which shows 90% sequence homology between *Gy4* and *Gy5* (18), was used to hybridize *EcoR* I digest of genomic DNA. The probe hybridized strongly and equally to two high-molecular-weight fragments 11 and 18 kb in length (**Figure 3B**). These results suggest that group-2 glycinin genes are highly conserved among the investigated genotypes.

Sequence analysis of the glycinin genes has shown that one of the main regions of sequence divergence differentiating glycinin genes from one another and legumin-like genes in other species lies within exon 3. This region of the genome has been termed the hypervariable region (HVR) (18). The source of sequence divergence in this region seems to have evolved by the duplication in tandem of an aspartate-/glutamate-rich motif particularly in group-2 glycinin genes. To examine the pos-



**Figure 4.** Southern blot analysis of soybean  $\beta$ -conglycinin. Genomic DNA from leaf tissue of *G. soja* and the ancestral and modern cultivars was digested by *EcoR* I, fractionated on a 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with a  $^{32}\text{P}$ -labeled  $\beta$ -subunit of the  $\beta$ -conglycinin gene and then subjected to autoradiography. The molecular weight markers in kilobases are shown on the left of the figure.

sibility of sequence variation in the HVR among the collection of genotypes in the present study, primers were designed to amplify exon 3 of group-2 glycinin *Gy4* and *Gy5* genes using the PCR. The amplified 730- and 630-bp fragments correspond to *Gy4* and *Gy5* genes, respectively (Figure 3C). No detectable differences in the size of the two fragments were seen, indicating conservation of the sequence length of the HVR among all genotypes, including *G. soja*. Cloning and sequencing of the two fragments from some of the genotypes, including *G. soja*, revealed almost identical DNA sequences among the genotypes except for alterations of one or two base pairs (data not shown).

The  $\beta$ -conglycinin storage proteins are the second major class of proteins present in soybean seeds. They are encoded by a multigene family that includes at least 15 genes (37). Analyses of the DNA sequences of the  $\beta$ -conglycinin genes have revealed that they are highly homologous and can be divided into two main groups based on the presence or absence of specific DNA fragments (35). To investigate the level of conservation in the  $\beta$ -conglycinin gene family across the historic collection of soybean cultivars, the coding region of the  $\beta$ -subunit of  $\beta$ -conglycinin was used to probe the nylon membrane containing the *EcoR* I-digested genomic DNA (Figure 4). The probe hybridized strongly to nine fragments and weakly to three additional fragments. The hybridization profiles were almost identical among all genotypes with one exception. *G. soja* showed hybridization to a total of 12 fragments, but only eight were in common with other soybean cultivars. Three of the remaining four fragments were different in their molecular size, and a fourth 2-kb fragment was unique to *G. soja* (Figure 4).

## DISCUSSION

An earlier study utilizing data derived from the Soybean Uniform Tests showed yearly fluctuation in protein content of United States soybean crop (36). Among cultivars acclimated to the Midwest, maturity groups I and II showed a trend toward lowered protein content, whereas protein in maturity groups III and IV remained relatively unchanged (10). Similarly, an analysis of soybean samples from 30 states shows that protein

content has remained relatively constant since 1986, ostensibly because few ancestral lines account for a preponderance of the genetic base of United States cultivars (8, 37, 38–40). In contrast to previous studies, in which the protein data were compiled using seeds grown at different locations, our studies show diminished protein content in cultivars released successively over the past 60 years. The observed effect could also reflect the fact that, during the past several decades, breeding efforts have emphasized increased oil content. Our results indicate, however, that soybean protein composition and quality have not changed remarkably during this time period.

Globulins consisting of the multi-subunit 7S  $\beta$ -conglycinins and 11S glycinins cumulatively account for 70–80% of the total soybean seed protein (18, 37). Because of their combined abundance and differences in amino acid content, the relative accumulation of these two groups of proteins is a major determinant of soybean protein quality. Comparatively, the glycinins are a better source of the sulfur amino acids than the  $\beta$ -conglycinins, which exhibit a paucity of methionine and cysteine (18, 37, 41, 42). Relative accumulation of the seed storage proteins is influenced by the nutritional status of the plant (13, 43–45). A high nitrogen/sulfur ratio of nutrient availability favors accumulation of the  $\beta$ -subunit of  $\beta$ -conglycinin, which is devoid of the sulfur amino acids. In this study, we found that the accumulation of the  $\beta$ -subunit of  $\beta$ -conglycinin was variable among the cultivars. However, the modern cultivars Asgrow 3602 and Pioneer 93B09 accumulated less of the  $\beta$ -subunit of  $\beta$ -conglycinin. Because all of the cultivars were grown at one location and similar results were obtained from two successive growing seasons, the difference in accumulation of the  $\beta$ -subunit is probably not mediated by the nutritional status of the plant. This difference, possibly a result of selective breeding, is a positive step in improving protein quality, as the  $\beta$ -subunit is devoid of the sulfur amino acids (46).

The genetic diversity of Asian and North American soybean cultivars has been examined using molecular markers such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), and simple sequence repeats (SSR) (8, 40, 47–51). These studies have shown that molecular markers can be utilized in identifying the genetic relationships among cultivars. In addition, these molecular markers can be exploited by the breeders to select genetically diverse parental lines for expanding the diversity in soybean germplasms of North American cultivars. In this present study, we found little or no variation in the DNA sequences in the genes encoding the two major storage proteins of soybean among the different cultivars. The lack of variation might be attributable to the limited genetic diversity of the U.S. soybean cultivars and selection pressure to maintain sequence integrity of essential seed storage proteins. It has been established that a limited number of ancestral lines contributed to the gene pool of modern U.S. soybean cultivars (38, 52). Using pedigree analysis, Gizlice et al. (38) found that 80% of the genes of 258 public soybean cultivars released between 1947 and 1988 were derived from 13 ancestral lines. As few as five ancestral lines, A. K., Mandarin, Mukden, Lincoln, and Illini, which were included in this study, contributed more than 40% to the genetic base of all U.S. soybean cultivars and more than 56% to the Northern U.S. cultivars. Results from this study showed no variation in the hybridization patterns among these five ancestral lines for  $\beta$ -conglycinin and group-2 glycinin genes. Even in the case of group-1 glycinin genes, which showed a limited degree of polymorphism, the modern cultivars showed hybridization patterns that were identical to those of at least one of the five

ancestral lines. These results suggest that all of the ancestral lines involved in the development of the modern cultivars have little or no variation in the DNA sequences of the seed storage protein genes.

Although progress in North American soybean development has been significant, the advances have been derived from a narrow genetic base, making effective improvement of both agronomic characteristics and marketable components in the future a concern (10, 19, 20, 25, 38–40, 52). Introgression of diverse soybean germplasm from distinct geographical locations will help to broaden the genetic base of soybeans (8, 53). Exploitation of this diversity holds promise for the continued improvement of North American soybean cultivars.

#### ACKNOWLEDGMENT

We thank Dr. Randall Nelson for the soybean seeds used in this study.

#### LITERATURE CITED

- Hymowitz, T.; Shurtleff, W. R. Debunking soybean myths and legends in the historical and popular literature. *Crop Sci.* **2005**, *45*, 473–476.
- Hermann, F. J. *A Revision of the Genus Glycine and Its Immediate Allies*; Technical Bulletin 1268; United States Department of Agriculture: Washington, DC, 1962.
- Ho, P.-T. The loess and the origin of Chinese agriculture. *Am. Hist. Rev.* **1969**, *75*, 1–36.
- Hymowitz, T. On the domestication of the soybean. *Econ. Bot.* **1970**, *24*, 408–421.
- Hymowitz, T.; Harlan, J. R. The introduction of the soybean to North America by Samuel Bowen in 1765. *Econ. Bot.* **1983**, *37*, 371–379.
- Ustun, A.; Allen, F. L.; English, B. C. Genetic progress in soybean of the U.S. Midsouth. *Crop Sci.* **2001**, *41*, 993–998.
- Durham, D. The United Soybean Board's Better Bean Initiative: Building United States soybean competitiveness from the inside out. *AgBioForum* **2003**, *6*, 23–26.
- Brown-Guedira, G. L.; Thompson, J. A.; Nelson, R. L.; Warburton, M. L. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR Markers. *Crop Sci.* **2000**, *40*, 815–823.
- Thompson, J. A.; Nelson, R. L. Utilization of diverse germplasm for soybean yield improvement. *Crop Sci.* **1998**, *38*, 1362–1368.
- Wilcox, J. R. Sixty years of improvement in publicly developed elite soybean lines. *Crop Sci.* **2001**, *41*, 1711–1716.
- Specht, J. E.; Hume, D. J.; Kumudini, S. V. Soybean yield potential—A genetic and physiological perspective. *Crop Sci.* **1999**, *39*, 1560–1570.
- Wolf, R. B.; Canvins, J. F.; Kleiman, R.; Black, L. T. Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. *J. Am. Oil Chem. Soc.* **1982**, *59*, 230–232.
- Paek, N. C.; Imsande, J.; Shoemaker, R. C.; Shibles, R. Nutritional control of soybean seed storage protein. *Crop Sci.* **1997**, *37*, 498–503.
- Maestri, D. M.; Labuckas, D. O.; Meriles, J. M.; Lamarque, A. L.; Zygadlo, J. A.; Guzman, C. A. Seed composition of soybean cultivars evaluated in different environmental regions. *J. Sci. Food Agric.* **1998**, *77*, 494–498.
- Sexton, P. J.; Paek, N. C.; Shibles, R. Soybean sulfur and nitrogen balance under varying levels of available sulfur. *Crop Sci.* **1998**, *38*, 975–982.
- Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage. T4. *Nature* **1970**, *227*, 680–685.
- Saghai Maroof, M. A.; Soliman, K. M.; Jorgensen, R. A.; Allard, R. W. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 8014–8019.
- Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, V. H.; Scallon, B. J.; Fischer, R. L.; Sims, T. L.; Drews, G. N.; Goldberg, R. B. Characterization of the glycinin gene family. *Plant Cell* **1989**, *1*, 313–328.
- Leudders, V. D. Genetic improvement in yield of soybeans. *Crop Sci.* **1977**, *17*, 971–972.
- Wilcox, J. R.; Schapaugh, W. T., Jr.; Bernard, R. L.; Cooper, R. L.; Fehr, W. R.; Niehaus, M. H. Genetic improvement of soybeans in the Midwest. *Crop Sci.* **1979**, *19*, 803–805.
- Johnson, H. W.; Robinson, H. F.; Comstock, R. E. Estimates of genotypic and phenotypic correlations in soybeans and their implications in selection. *Agron. J.* **1955**, *47*, 477–483.
- Hanson, W. D.; Leffel, R. C.; Howell, R. W. Genetic analysis of energy production in the soybean. *Crop Sci.* **1961**, *1*, 121–126.
- Hartwig, E. E.; Hinson, K. Association between chemical composition of seed and seed yield of soybeans. *Crop Sci.* **1972**, *12*, 829–830.
- Shannon, J. G.; Wilcox, J. R.; Probst, A. H. Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. *Crop Sci.* **1972**, *12*, 824–828.
- Nelson, R. L.; Guzman, P. S.; Neece, D. J. Qtl associated with seed protein concentration from three sources of high protein. In *Abstracts of Contributed Papers and Posters. VII World Soybean Research Conference*; Moscardi, F., Hoffmann-Campo, O. F., Saraiva, P. R., Galerani, F. C., Krzyzanowski, F. C., Carrao-Panizzi, M. C., Eds.; Foz Do Iguassu: Pr Embrapa, Brazil, 2004; p 56.
- Murphy, P. A.; Resurreccion, A. Varietal and environmental differences in soybean glycinin and  $\beta$ -conglycinin. *J. Food Agric. Chem.* **1984**, *32*, 911–915.
- Hurburgh, C. R., Jr. Long-term soybean composition patterns and their effect on processing. *J. Am. Oil Chem. Soc.* **1994**, *71*, 1425–1427.
- Piper, E. L.; Boote, K. J. Temperature and cultivar effects on soybean seed oil and protein concentrations. *J. Am. Oil Chem. Soc.* **1999**, *76*, 1233–1241.
- Fehr, W. R.; Hoeck, J. A.; Johnson, S. L.; Murphy, P. A.; Nott, J. D.; Padilla, G. I.; Welke, G. A. Genotype and environment influence on protein components of soybean. *Crop Sci.* **2003**, *43*, 511–514.
- Bates, L. S. Evaluation of an ultra filtration system to minimize phenolic interference during protein extraction and fractionation. *Sep. Sci. Technol.* **1979**, *14*, 653–658.
- Fukuda, T.; Maruyama, N.; Kanazawa, A.; Abe, J.; Shimamoto, Y.; Hiemori, M.; et al. Molecular analysis and physicochemical properties of electrophoretic variants of wild soybean *Glycine soja* storage proteins. *J. Agric. Food Chem.* **2005**, *53*, 3658–3665.
- Imsande, J. Selection of soybean mutants with increased concentrations of methionine. *Crop Sci.* **2001**, *41*, 510–515.
- Kerley, M.; Allee, G. L. Modifications in soybean seed composition to enhance animal feed use and value: Moving from a dietary ingredient to a functional dietary component. *AgBioForum* **2003**, *6*, 14–17.
- Beilinson, V.; Chen, Z.; Shoemaker, R.; Fischer, R.; Goldberg, R.; Nielsen, N. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.* **2002**, *104*, 1132–1140.
- Harada, J. J.; Barker, S. J.; Goldberg, R. B. Soybean  $\beta$ -conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* **1989**, *1*, 415–425.
- Yaklich, R. W.; Vinyard, B.; Camp, M.; Douglass, S. Analysis of seed protein and oil from soybean northern and southern region uniform tests. *Crop Sci.* **2002**, *42*, 1504–1515.
- Brumm, T. J.; Hurburgh, C. R., Jr.; Naev, S. *Quality of the 2005 Soybean Crop from the United States*; Prepared for the American Soybean Association Quality Mission to Asia, 2005. ([www.soy-growers.com/international/quality/US-05-Quality.pdf](http://www.soy-growers.com/international/quality/US-05-Quality.pdf)).

- (38) Gizlice, Z.; Carter, T. E.; Burton, J. W. Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* **1994**, *34*, 1143–1151.
- (39) Gizlice, Z.; Carter, T. E., Jr.; Gerig, T. M.; Burton, J. W. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci.* **1996**, *36*, 753–765.
- (40) Kisha, T. J.; Diers, B. W.; Hoyt, J. M.; Sneller, C. H. Genetic diversity among soybean plant introductions and North American germplasm. *Crop Sci.* **1998**, *38*, 1669–1680.
- (41) Thanh, V. H.; Sibasaki, K. Major proteins of soybean seeds. Reversible and irreversible dissociation of  $\beta$ -conglycinin. *J. Agric. Food Chem.* **1979**, *27*, 805–809.
- (42) Utsumi, S.; Inaba, H.; Mori, T. Heterogeneity of soybean glycinin. *Photochemistry* **1981**, *20*, 585–589.
- (43) Ohtake, N.; Suzuki, M.; Takahashi, Y.; Fujiwara, T.; Chino, M.; Ikarashi, T.; Ohyama, T. Differential expression of  $\beta$ -conglycinin genes in nodulated and nonnodulated isolines of soybean. *Physiol. Plantarum* **1996**, *96*, 101–110.
- (44) Sexton, P. J.; Paek, N. C.; Shibles, R. M. Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. *Field Crops Res.* **1998**, *59*, 1–8.
- (45) Sexton, P. J.; Paek, N. C.; Naeve, S. L.; Shibles, R. M. Sulfur metabolism and protein quality of soybean. *J. Crop Prod.* **2002**, *5*, 285–308.
- (46) Thanh, V. H.; Shibasaki, K. Beta-conglycinin from soybean proteins. Isolation and immunological and physicochemical properties of the monomeric forms. *Biochim. Biophys. Acta* **1977**, *490*, 370–374.
- (47) Skorupska, H. T.; Shoemaker, R. C.; Warner, A.; Shipe, E. R.; Bridges, W. C. Restriction fragment length polymorphism in soybean germplasm of the southern USA. *Crop Sci.* **1993**, *33*, 1169–1176.
- (48) Lorenzen, L. L.; Boutin, S.; Young, N.; Specht, J. E.; Shoemaker, R. C. Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. *Crop Sci.* **1995**, *35*, 1326–1336.
- (49) Thompson, J. A.; Nelson, R. L. Core set of primers to evaluate genetic diversity in soybean. *Crop Sci.* **1998**, *38*, 1356–1362.
- (50) Thompson, J. A.; Nelson, R. L.; Vodkin, L. O. Identification of diverse soybean germplasm using RAPD markers. *Crop Sci.* **1998**, *38*, 1348–1355.
- (51) Li, Z.; Nelson, R. L. Genetic diversity among soybean accessions from three countries measured by RAPDS. *Crop Sci.* **2001**, *41*, 1137–1147.
- (52) Sneller, C. H. Pedigree analysis of Elite soybean lines. *Crop Sci.* **1994**, *34*, 1515–1522.
- (53) Singh, R. J.; Hymowitz, T. Soybean genetic resources and crop improvement. *Genome* **1999**, *42*, 605–616

---

Received for review February 9, 2006. Revised manuscript received March 27, 2006. Accepted March 31, 2006. Names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of product, and the use of the name by the USDA implies no approval of the product to the exclusion of others that may also be suitable.

JF060391M