JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. 2006, 54, 5489−5497 5489

Assessment of Indigenous Nepalese Soybean as a Potential Germplasm Resource for Improvement of Protein in North American Cultivars

HARI B. KRISHNAN,*,†,§ SAVITHIRY S. NATARAJAN,# AHMED A. MAHMOUD,§ JOHN O. BENNETT,§ AMMULU HARI KRISHNAN,§ AND BRAJ NANDAN PRASAD⊥,||

Plant Genetics Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Columbia, Missouri 65211; Soybean Genomics and Improvement Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland 20705; Plant Science Division, University of Missouri, Columbia, Missouri 65211; and Central Department of Botany, Biotechnology Unit, Tribhuvan University, Kirtipur, Kathmandu, Nepal

Soybeans contain approximately 40% protein and 20% oil and represents an important source of protein in animal rations and human diets. Attempts are being made to increase further the overall protein content of soybeans by utilization of exotic germplasms. In this study, soybean cultivars from Nepal have been characterized and their potential as a germplasm resource for improvement of the protein content and quality of North American cultivars assessed. Soybean cultivars 'Sathia', 'Seti', 'Kavre', and 'Soida Chiny', indigenous to various regions of Nepal, contained 42−45% protein, which is significantly higher in comparison to that of the North American cultivar 'Williams 82' (39%). Fractionation of seed protein by high-resolution two-dimensional gel electrophoresis revealed differences in the protein profiles of these cultivars. Various isoelectric forms of glycinin and β-conglycinin were identified by comparing the matrix-assisted laser desorption ionization time-of-flight mass fingerprinting data against the National Center for Biotechnology Information nonredundant database. Nepalese cultivar Sathia was distinct, lacking some isoelectric forms of acidic and basic glycinin subunits while expressing other unique forms. The contribution of these unique protein spots present in either Sathia or Williams 82 to the total protein content was quantified using scanning laser densitometry. Distinct restriction fragment length polymorphisms (RFLP) for group 1 glycinin genes were observed among the tested Nepalese genotypes, indicating sequence variation among the cultivars. Conversely, evaluation of RFLP for the genes encoding group 2 glycinins, β-conglycinin, and Bowman−Birk proteinase inhibitors indicated a high degree of conservation in these genes. Determination of amino acid composition, a reflection of protein quality, indicated that the arginine content of the Nepalese soybeans ranged from 7.7 to 8.1%, which was 5−10% higher than the 7.4% expressed in Williams 82. Additionally, Karve and Seti contained significantly more cysteine than Williams 82. Nepalese high-protein soybeans having a desirable amino acid composition hold potential to increase the protein quality and diversity of North American cultivars.

KEYWORDS: Amino acid composition; β-conglycinin; genetic diversity; glycinin; soybean; storage protein

INTRODUCTION

Soybeans (*Glycine max* L. Merrill) have become an integral part of the world economy, serving as a vital component of human diets, animal rations, and feedstock for industrial applications. The genetic base of North American cultivars is narrow, being derived from a small number of introductions primarily from the northern region of East China (*1*−*4*). Crossing cultivars of elite lines, each exhibiting desirable agronomic traits, has provided a steady increase in yield over the past several decades (*5*−*7*). Although selective breeding has benefited yield, it has produced soybeans with complex pedigrees of limited diversity, reducing the gene pool by ≈25% from that exhibited by ancestral cultivars (*3*, *8*, *9*). Consequently, concerns have arisen as to the continued ability to maintain productivity and enhance various attributes of seed components through selective breeding (*10*, *11*).

* Address correspondence to this author at the Plant Genetics Research Unit, USDA-ARS, 108W Curtis Hall, University of Missouri, Columbia, MO 65211 [telephone (573) 882-8151; fax (573) 884-7850; e-mail krishnanh@missouri.edu].
† Plant Genetics Research Unit, U.S. Department of Agriculture.
§ Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture.
# University of Missouri.
⊥ Tribhuvan University.
|| Current address: School of Biotechnology, Karunya Institute of Technology and Sciences, Coimbatore, 641114, India.

10.1021/jf0610559 CCC: $33.50 © 2006 American Chemical Society
Published on Web 07/04/2006

Asia holds the largest variety of soybean cultivars (*12*, *13*). The immense geographical area of the continent with a variety of growing conditions and selection based on various food type preferences has given rise to thousands of phenotypically diverse landraces (*14*). A study of over 20000 Chinese soybean accessions, excluding those bred in the past 50 years, has provided evidence of marked germplasm diversity (*12*). Similar evaluations of the genetic diversity among soybean germplasms in other countries have also been performed (*15*−*17*). Genetic diversity observed between germplasms found in the northern and southern growing regions of the United States has been ascribed to their different ancestral backgrounds (*1*, *9*, *18*−*20*). In Japan, the diversity in the soybean germplasm has been attributed to continual introduction of domestic landraces and plant introductions into the breeding program (*21*, *22*). Because significant genetic diversity exists between Chinese, Japanese, and Korean germplasms, each has the potential for improving U.S. cultivars (*11*, *17*).

At present 15000 accessions are available in the USDA soybean germplasm collection, most of which do not appear in pedigrees of U.S. cultivars (*23*). Introduction of exotic germplasm holds promise for expanding the genetic diversity of North American cultivars, facilitating the continued improvement in yield and enhancing the content of desired seed components (*24*). These accessions are being scrutinized to identify unique genes that hold potential for improving U.S. soybean yield and seed characteristics. Incorporation of these exotic germplasms, however, must be done in a calculated manner to maximize the desired effect of the combined genetic material and to minimize introduction of deleterious traits (*25*−*28*). Several plant introduction lines (PI) from China have been shown to contain high levels (>50%) of protein (*12*). These PIs are currently utilized in the breeding programs to incorporate the high protein trait in North American commercial cultivars. To expand the genetic diversity of North American cultivars, it would be desirable to utilize soybean germplasm from other regions of the world. Toward this goal, we have examined the potential of some soybean cultivars from Nepal as a germplasm resource for improvement of North American cultivars. As a first step in this direction, we have carried out proteonomic and molecular studies on the seed storage proteins of soybean cultivars from Nepal.

MATERIALS AND METHODS

Plant Materials. Soybean cultivars 'Sathia', 'Seti', 'Kavre', and 'Soida Chiny' indigenous to various regions of Nepal (all belonging to maturity group V) and the North American cultivar 'Williams 82' (maturity group III) were grown during the 2004 and 2005 seasons at the University of Missouri Bradford Research Center near Columbia, MO.

Amino Acid Analysis. Four replicates of dry seed powder from each cultivar representing both growing seasons were hydrolyzed under nitrogen at 155 °C for 16 h in 6 N HCl. The content of the sulfur amino acids was determined separately from samples, which had been pretreated with performic acid prior to standard acid hydrolysis. Tryptophan content was analyzed by enzymatic hydrolysis of seed protein followed by colorimetric determination. Amino acids were separated on a high-performance cation-exchange resin column installed in the Beckman 6300 Amino Acid Analyzer (Beckman Instruments, Fullerton, CA). Quantification was performed colormetrically using the postcolumn ninhydrin reaction detection system.

Modified Trichloroacetic Acid (TCA)/Acetone/Urea Extraction and Solubilization of Seed Protein for Two-Dimensional (2-D) Polyacrylamide Gel Electrophoresis (PAGE). One hundred milligrams of replicated samples from each cultivar and year of study was extracted with 2 mL of 10% TCA in acetone (w/v) containing 0.07% 2-mercaptoethanol (w/v) as described (*29*). Precipitated proteins were recovered by centrifugation and washed three times in acetone containing 0.07% 2-mercaptoethanol (w/v). The pellet was dried, resuspended in 1 mL of lysis buffer composed of 9 M urea, 1% CHAPS (w/v) 1% ampholytes (w/v), pH 3−10, and 1% DTT (w/v), and then sonicated. After centrifugation (20800*g*, 20 min, 4 °C), aliquots of the supernatant were analyzed by 2-D PAGE.

Isoelectric focusing (IEF) was performed using 13 cm IPG strips (pH 3−10, 4−7, or 6−11) in the IPGphor System (Amersham Biosciences, Piscataway, NJ). A 100 μg sample of protein was loaded onto the strips by active rehydration in 250 μL of buffer containing 8 M urea, 2% CHAPS (w/v), 0.5% ampholytes, and 0.002% bromphenol blue. Conditions for IEF were as described for 2-d PAGE. Electrophoretic separation according to mass was performed on a 13.5% SDS-PAGE gel (*30*). The gels were fixed overnight in a solution of 50% ethanol (v/v) and 3% phosphoric acid. After a distilled water wash, the gels were prestained for 1 h in 34% methanol, 17% ammonium sulfate, and 3% phosphoric acid and then stained in 0.066% Coomassie blue G250 (w/v). Until analysis, the gels were stored in 20% ammonium sulfate.

Image Analysis. Scanning laser densitometry in conjunction with analysis software (Personal Densitometer SI, ImageQuant software, Amersham Biosciences) was used to quantify the proteins resolved by 2-D PAGE. Software algorithms were used to calculate spot volume, pixel area, mean pixel intensity, standard deviation, and background values.

In-Gel Digestion of Protein Spots. After excision from the gel, protein spots selected for mass spectrographic analysis were washed in distilled water and then destained in a 50% solution of acetonitrile (v/v) containing 25 mM ammonium bicarbonate. After a 100% acetonitrile wash, the proteins contained in the gel spots were subjected to digestion using 20 μL (10 μg/mL) of modified porcine trypsin in 25 mM ammonium bicarbonate (Promega, Madison, WI). The fragmented proteins were sonicated in 50% acetonitrile/5% trifluoroacetic acid (TFA) (v/v), dried, and then reconstituted in 50% acetonitrile/0.1% TFA (v/v).

Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF-MS) Analysis of Trypsin-Digested Proteins. Peptides resulting from tryptic digestion of soybean seed proteins were analyzed by mass spectroscopy (Voyager DE-STR MALDI-TOF, Applied Biosystems, Framingham, MA). The peptides were cocrystallized with the matrix material α-cyano-4-hydroxycinnamic acid. A 337 nm nitrogen laser operating at 20 Hz was used in sample ionization. Trypsin autolysis peaks of charge mass ratios 842.51 and 2211.10 served as internal standards.

Liquid Chromatography−Mass spectrometry (LC-MS) Analysis. Peaks that were not identified by MALDI-TOF were analyzed using an ion trap mass spectrometer (Finnigan LCQ Deca XP Plus, Thermo Electron Corp., Waltham, MA). Prior to mass analysis, peptides were separated by liquid chromatography on a reverse phase column with a 5−60% gradient of acetonitrile/water containing 0.1% formic acid. When two MS/MS spectra of an ion were acquired, dynamic exclusion circumvented continuous analysis. Initial processing of the data was done using Sequest (University of Washington, Seattle, WA). The multiple Sequest DTA files were then concatenated into a single Mascot file using merge pl. script (Matrix Science, Boston, MA).

Data Interpretation. Using Mascot, the empirically determined mass-to-charge ratios of peptides were compared with those derived from peptides of known sequences listed in the National Center for Biotechnology Information nonredundant database. Mascot search parameters of MALDI-TOF and MS/MS generated data were set forth previously (*29*). At least two unique peptides with one or more having a significant ion score were required for positive identification by tandem mass spectroscopy. Positive identification by MALDI-TOF required that the protein meet or exceed the minimum significant score.

Southern Hybridization. Total genomic DNA was isolated from leaf tissue of the soybean cultivars using the cetyltrimethylammonium bromide (CTAB) method (*31*). Genomic DNA (8 μg) was digested overnight at 37 °C with *Eco*RI. Restricted DNA was electrophoretically separated on an 0.8% agarose gel and then transferred to a nylon membrane by capillary action using 0.04 M NaOH as the carrier. DNA

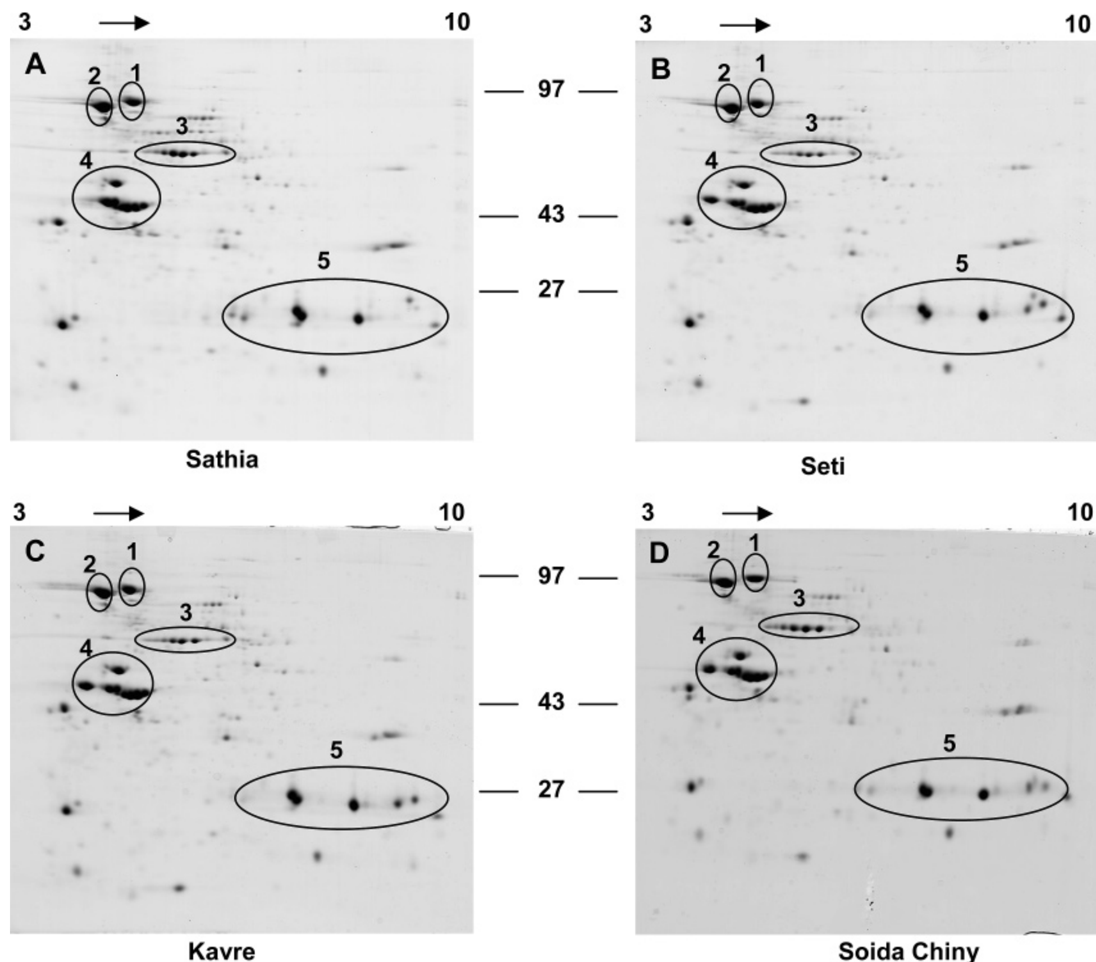

Figure 1. Comparison of Nepalese soybean seed protein profiles by two-dimensional gel electrophoresis: (**A**) Sathia; (**B**) Seti; (**C**) Kavre; (**D**) Soida Chiny. Proteins were first separated by isoelectric focusing on IPG strips (pH 3–10) and then by SDS-PAGE on a 13.5% gel. The gels were stained with Coomassie blue. Numerical designations identify the abundant seed proteins: 1, 2, and 3 mark the α′-, α-, and β-subunits, respectively, of β-conglycinin; 4 and 5 are the acidic and basic subunits of glycinin, respectively.

fragments corresponding to the coding regions of the glycinin *Gy2* and *Gy4*, β-conglycinin, BBI, soy CII, and KTI were used as probes for Southern hybridization. The probe DNA was labeled with ^{32}P-dCTP using the Ladderman kit (Takara Bio Inc., Otsu, Shiga, Japan). At minimum a 10 h prehybridization and an overnight hybridization were both conducted at 65 °C using 6× SSPE buffer (1× SSPE is 0.1 M NaCl, 0.01 M Na_2PO_4, and 0.001M EDTA), 5× Denhardt's solution, 0.5% SDS, and 50 μg/mL sheared and denatured salmon sperm. After hybridization, the membranes were washed three times in a solution of 2× SSPE and 0.5% SDS for 10 min at room temperature and then twice in a solution of 0.1× SSPE and 0.1% SDS at 65 °C for 30 min. Following washing the nylon membrane was exposed overnight at −80 °C to X-ray film.

Statistical Analysis. Data were analyzed using SAS (8.2 version) with PROC General Linear Model (SAS Institute, Inc., Cary, NC).

RESULTS

Protein Composition of Selected Nepalese Soybean Cultivars. The protein content of soybean cultivars that were obtained from Nepal along with the North American cultivar Williams 82 was determined by near-infrared reflectance spectroscopy. Analysis indicated that the Nepalese cultivars had higher protein contents than Williams 82. To minimize the effect of environmental conditions and agronomic practices on protein accumulation, the Nepalese cultivars and Williams 82 were planted side by side at the Bradford Research Center during the 2004 and 2005 growing seasons. The protein content of four replications of seed samples from these two growing seasons was ascertained by the combustion method for nitrogen determination. The four Nepalese cultivars, Soida Chiny, Sathia, Kavre, and Seti, and the North American cultivar Williams 82 contained 45.1 ± 0.1, 43.8 ± 0.3, 43.0 ± 0.4, 42.3 ± 0.1, and 39.9 ± 0.2% protein, respectively. To determine whether the increased protein content is accompanied by preferential accumulation of specific proteins, 2-D PAGE was performed. IEF on an immobilized pH gradient (3−10) gel followed by SDS-PAGE and Coomassie staining resolved the seed proteins into more than 100 distinct spots (**Figure 1**). Comparison of 2-PAGE patterns revealed subtle variations among the protein profiles of Nepalese cultivars and between them and Williams 82 (**Figure 2A−D**). Even though the number of protein spots representing the α′-, α-, and β-subunits of β-conglycinin were similar among the cultivars, the abundance of the β-subunit varied, with Seti showing the least accumulation. The acidic and basic subunits of glycinins were resolved into several different isoelectric forms (**Figure 1**). There were strong similarities in the number and distribution of these isoelectric forms among Seti, Kavre, and Soida Chiny. In contrast, Sathia was distinct and lacked some isoelectric forms representing both the acidic and basic subunits of glycinins (**Figure 1**). In addition, several unidentified protein spots were clustered in the gel above the β-subunit of β-conglycinin, being more abundant in Sathia than in the other Nepalese cultivars (**Figure 1**). Conversely, Sathia lacked a low molecular weight protein spot, which was present in the remaining Nepalese cultivars.

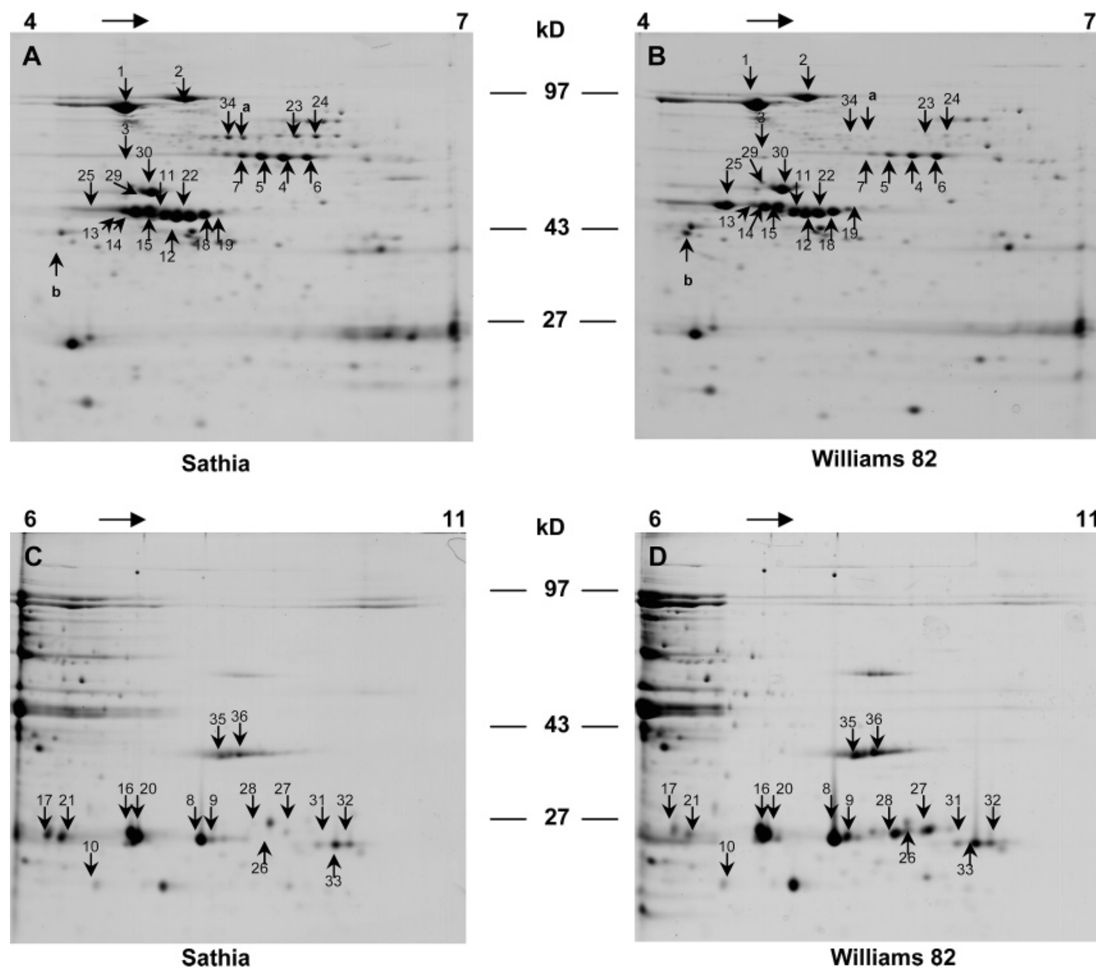

Figure 2. Comparison of seed protein profiles of Williams 82 and Sathia by 2-D PAGE: (**A**, **C**) Sathia; (**B**, **D**) Williams 82. Isoelectric focusing of proteins was first performed using either pH 4–7 (**A**, **B**) or pH 6–11 (**C**, **D**) narrow-range IPG strips followed by SDS-PAGE on a 13.5% gel. The gels were stained with Coomassie blue. Protein spots were excised digested with trypsin and the resulting peptides analyzed using MALDI-TOF-MS (**Table 1**).

Comparative Analysis and Identification of Williams 82 and Sathia Seed Protein Components by High-Resolution 2-D PAGE and MALDI-TOF-MS. To elucidate fully the differences between the Nepalese cultivar Sathia and the North American cultivar Williams 82, we performed comparative analyses of seed proteins resolved by 2-D PAGE. Our initial analysis was performed using immobilized pH gradient (3−10) (IPG) gel strips (**Figure 1**). IEF in this pH range gave limited resolution of the various isoelectric forms of the glycinins and β-conglycinin subunits. To improve separation of the protein spots IEF was performed with IPG strips ranging from pH 4 to 7 (**Figure 2A**,**B**) and from pH 6 to 11 (**Figure 2C**,**D**). Resolution of β-conglycinin subunits and acidic glycinin subunits was significantly enhanced when the proteins were subjected to IEF using the pH 4−7 IPG strip (**Figure 2A**,**B**). Similarly, the use of a pH 6−11 gradient enabled resolution of the acidic glycinin polypeptides (**Figure 2C**,**D**). Several of the protein spots identified by MALDI-TOF-MS are listed in **Table 1**. Thirty-six protein spots representing the different isoelectric forms of the glycinins and β-conglycinin were identified in this study. The Nepalese cultivar Sathia was either deficient in or contained low amounts of glycinin G4/A5A4B3 precursors (spots 25−28) and glycinin G1/A1aBx subunit (spot 9) (**Figure 2**). In contrast, Sathia accumulated higher amounts of glycinins G2/A2B1 precursor (spot 17) and glycinins subunit G3/A1ab1B (spot 21). Interestingly, Sathia also contained several protein spots that were identified as proglycinins (spots 23 and 24), which were not seen in Williams 82 (**Figure 2**). The contribution of these unique protein spots present in either Sathia or Williams 82 to the total protein content was quantified using scanning laser densitometry in conjunction with software algorithms. The relative concentration of the selected protein spots is shown in **Table 2**. This table clearly shows that the marked heterogeneity in the seed storage protein profiles of the Nepalese cultivar Sathia and the North American cultivar Williams 82.

Amino Acid Composition of Nepalese Soybean Cultivars. The seed storage protein content of the Nepalese soybean cultivars is significantly higher than that of the North American cultivar Williams 82. To ascertain whether the increased protein content is accompanied by enhanced protein quality, we determined the amino acid composition of these cultivars (**Table 3**). The arginine content of the Nepalese seed storage protein ranged from 7.7 to 8.1%, which was 5−10% higher ($P \leq 0.05$) than the 7.4% expressed in Williams 82. Seed storage proteins in Karve and Seti contained more cysteine than Williams 82. Cysteine accounted for 1.64% of the total amino acids in Seti and 1.66% in Karve, 6 and 7% more, respectively ($P \leq 0.05$), than found in Williams 82.

Restriction Fragment Length Polymorphism (RFLP) in the Genes Encoding 11S Storage Proteins. Southern hybridization was used to determine the levels of variation in the DNA sequences of the seed protein genes among the soybean accessions under investigation. Glycinin, an 11S globulin, accounts for >50% of the total soybean seed protein. Five major genes, encoding the different subunits of the hexameric glycinin proteins, are classified into two groups based on DNA sequence

Table 1. Proteins Identified in Sathia and Williams 82 by MALDI-TOF-MS and LC-MS/MS Analysis

spot ID	theor p*I*/M_r	protein identity	peptides matched	sequence coverage (%)	MOWSE score	expected value	NCBI accession no.	method
1	4.92/63127	α subunit of β conglycinin	25	39	217	3.00E−17	gi|9967357	MALDI-TOF-MS
2	5.23/65160	α′ subunit of β conglycinin	20	41	194	5.90E−15	gi|9967361	MALDI-TOF-MS
3	5.32/72717	α subunit of β conglycinin	9	18	73	7.20E−03	gi|15425633	MALDI-TOF-MS
4	5.67/48358	β conglycinin β subunit	26	47	216	5.10E−17	gi|63852207	MALDI-TOF-MS
5	5.67/48358	β conglycinin β subunit	18	40	176	5.10E−13	gi|63852207	MALDI-TOF-MS
6	5.67/48358	β conglycinin β subunit	27	49	256	5.10E−21	gi|63852207	MALDI-TOF-MS
7	5.67/48358	β conglycinin β subunit	11	28	85	6.40E−04	gi|63852207	MALDI-TOF-MS
8	5.89/56299	glycinin G1/A1aBx subunit	7	15	145		gi|18635	LC-MS/MS
9	5.89/55672	glycinin G1/A1aBx subunit	5	10	157		gi|18635	LC-MS/MS
10	6.15/56134	glycinin G1/A1aBx subunit	4	9	254		gi|72296	LC-MS/MS
11	5.46/54927	glycinin G2/A2B1 precursor	9	19	73	7.20E−03	gi|1212177	MALDI-TOF-MS
12	5.46/54927	glycinin G2/A2B1 precursor	8	15	71	1.30E−02	gi|1212177	MALDI-TOF-MS
13	5.46/54927	glycinin G2/A2B1 precursor	9	14	74	6.70E−03	gi|1212177	MALDI-TOF-MS
14	5.46/54927	glycinin G2/A2B1 precursor	12	21	104	6.10E−06	gi|1212177	MALDI-TOF-MS
15	5.46/54927	glycinin G2/A2B1 precursor	9	19	91	1.20E−04	gi|1212177	MALDI-TOF-MS
16	5.78/54047	glycinin G2/A2B1 precursor	6	37	67	3.20E−02	gi|169967	MALDI-TOF-MS
17	5.56/54903	glycinin G2/A2B1 precursor	9	14	313		gi|72295	LC-MS/MS
18	5.78/54047	glycinin subunit G3/A1ab1B	8	18	70	1.40E−02	gi|15988117	MALDI-TOF-MS
19	5.78/54047	glycinin subunit G3/A1ab1B	9	20	74	6.40E−03	gi|15988117	MALDI-TOF-MS
20	5.78/54047	glycinin subunit G3/A1ab1B	8	18	72	9.40E−03	gi|15988117	MALDI-TOF-MS
21	5.78/54047	glycinin subunit G3/A1ab1B	9	18	116	3.90E−07	gi|15988117	MALDI-TOF-MS
22	5.78/54047	Glycinin subunit G3/A1ab1B	8	19	72	1.00E−02	gi|15988117	MALDI-TOF-MS
23	5.78/54047	proglycinin A1ab1B	16	35	154	7E−10	gi|15988119	MALDI-TOF-MS
24	5.78/54047	proglycinin A1ab1B	13	30	132	1.40E−08	gi|15988119	MALDI-TOF-MS
25	5.38/64097	glycinin G4/A5A4B3 precursor	13	16	76	4.20E−03	gi|99910	MALDI-TOF-MS
26	6.47/31065	glycinin G4/A5A4B3 precursor	10	45	142	9.50E−10	gi|81785	MALDI-TOF-MS
27	6.47/31065	glycinin G4/A5A4B3 precursor	8	40	105	5.40E−06	gi|81785	MALDI-TOF-MS
28	5.38/64136	glycinin G4/A5A4B3 precursor	14	14	408		gi|99910	LC-MS/MS
29	5.46/55850	glycinin G5/A3B4 subunit	12	22	110	1.50E−06	gi|33357661	MALDI-TOF-MS
30	5.46/55850	glycinin G5/A3B4 subunit	10	21	86	3.70E−04	gi|33357661	MALDI-TOF-MS
31	5.69/26938	glycinin G5/A3B4 subunit	7	30	65	5.00E−02	gi|541941	MALDI-TOF-MS
32	5.69/26938	glycinin G5/A3B4 subunit	8	37	69	2.50E−02	gi|541941	MALDI-TOF-MS
33	9.64/21482	glycinin G5/A3B4 subunit	3	12	109		gi|625538	LC-MS/MS
34	4.46/24349	glycinin	10	38	84	0.00078	gi|6015515	MALDI-TOF-MS
35	5.46/54927	glycinin	13	32	82	0.0015	gi|l295800	MALDI-TOF-MS
36	8.68/47036	7S seed globulin precursor	18	17	321		gi|l401240	LC-MS/MS

Table 2. Relative Concentration of Selected Soybean Seed Proteins Resolved by Two-Dimensional Gel Electrophoresis

spot ID	protein ID	Sathia	Williams 82
25	glycinin G4 precursor	0	1065.3 ± 932.1
28	glycinin G4 precursor	0	3338.6 ± 280.6
36	7S seed globulin precursor	2486.7 ± 1107.5	5270.4 ± 454.0
26	glycinin A5A4B3 precursor	0	4968.2 ± 802.7
27	glycinin A5A4B3 precursor	0	3882.4 ± 836.4
b	unidentified	0	1651.6 ± 244.9
34	glycinin	1220.9 ± 33.7	0
a	unidentified	967.8 ± 25.8	0
23	proglycinin A1ab1b homotrimer	1037.0 ± 76.3	0
24	proglycinin A1ab1b homotrimer	1139.0 ± 19.5	511.0 ± 19.3
9	glycinin G1 subunit	1509.3 ± 122.9	4683.4 ± 904.4
35	glycinin	2063.9 ± 265.3	5387.4 ±204.2
6	β conglycinin β homotrimer	3674.2 ± 43.0	3953.6 ± 184.5

Table 3. Relative Percentage of Individual Amino Acids As Related to Total Amino Acid Content

amino acid	Soida Chiny	Sathia	Kavre	Seti	Williams 82	LSD[a]
alanine	4.24	4.33	4.23	4.27	4.29	0.04
arginine	8.07	7.76	8.02	7.72	7.44	0.22
aspartic acid	11.81	11.79	11.67	11.74	11.46	0.21
cysteine	1.46	1.52	1.65	1.64	1.56	0.07
glutamic acid	18.43	18.13	17.98	18.13	17.89	0.22
glycine	4.31	4.29	4.31	4.29	4.27	0.12
histidine	2.64	2.63	2.65	2.62	2.70	0.06
isoleucine	4.75	4.73	4.77	4.71	4.67	0.15
lysine	6.41	6.64	6.57	6.62	6.64	0.10
leucine	7.82	7.87	7.71	7.82	7.90	0.07
methionine	1.33	1.48	1.43	1.46	1.45	0.04
phenylalanine	5.19	5.14	5.19	5.04	5.12	0.18
proline	5.26	5.07	5.26	5.25	5.23	0.17
serine	4.68	4.78	4.54	4.64	4.87	0.21
threonine	3.72	3.89	3.75	3.81	3.96	0.08
tryptophan	1.26	1.32	1.27	1.40	1.45	0.14
tyrosine	3.34	3.36	3.67	3.44	3.74	0.27
valine	4.99	4.87	4.92	4.97	4.97	0.23

[a] Comparisons within rows are by LSD at the $P \leq 0.05$ level.

similarity (*32*). Group 1 glycinin genes include three genes, *Gy* 1, *Gy* 2, and *Gy* 3, whereas group 2 consists of two genes, *Gy* 4 and *Gy* 5. To test the level of sequence variation in group 1 glycinin genes, restriction digests of genomic DNA from the Nepalese varieties and Williams 82 were hybridized with a probe corresponding to exons 3 and 4 of the *Gy* 2 gene. On the basis of DNA sequence similarity among the glycinin genes in this region, the probe is expected to hybridize to the three group 1 genes. The Southern hybridization profile of the *Eco*RI digest (**Figure 3A**) showed strong hybridization to three fragments, which should correspond to the three group 1 glycinin genes. Different hybridization patterns were observed among the tested genotypes, indicating sequence variation among these genotypes for group 1 glycinin genes. The cultivar Soida Chiny appeared to be unique in its hybridization pattern in both *Eco*RI and *Hin*dIII restriction digests when compared with the other genotypes. Each of the four Nepalese cultivars was different in their respective hybridization patterns when compared to that of Williams 82 (**Figure 3A**). To determine the level of sequence variation in group 2 glycinin genes, the same genomic restriction

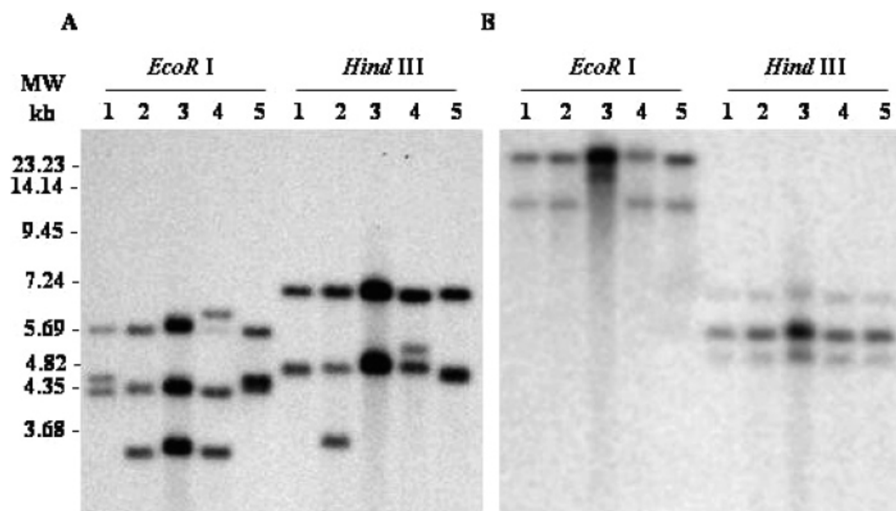

Figure 3. Southern blot analysis of soybean glycinin. Genomic DNA from leaf tissue of Williams 82 (lane 1), Kavre (lane 2), Sathia (lane 3), Soida Chiny (lane 4), and Seti (lane 5) was individually digested with either *Eco*RI or *Hin*dIII, fractionated on an 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with ^{32}P-labeled *Gy2* (**A**) and *Gy4* (**B**) glycinin probes and then subjected to autoradiography. Molecular weight markers in kb are shown on the left.

digests were hybridized to a probe corresponding to exon 3 of the *Gy* 4 gene. This probe is expected to hybridize to both genes of group 2 glycinin on the basis of their DNA sequence similarity. The hybridization profile of the *Eco*RI digest did show strong hybridization to two fragments, which correspond to the group 2 glycinin genes (**Figure 3B**). Unlike group 1 glycinin genes, group 2 genes showed very similar hybridization patterns among all of the cultivars except Sathia, which showed a difference in its hybridization pattern in the *Eco*RI digest (**Figure 3B**).

RFLP in the Genes Encoding 7S Storage Proteins. The β-conglycinin protein, a 7S globulin, is the second major seed storage protein in soybean. Three major subunits, α', α, and β, that make up the trimeric β-conglycinin are encoded by a multigene family of at least 15 genes. β-Conglycinin genes are highly homologous in their DNA sequences; some are found clustered in tandem arrays, whereas others are dispersed at different locations in the genome (*33*). A β-conglycinin-specific probe, when hybridized to the restriction digests of genomic DNA, revealed hybridization to 7−13 different fragments with varying intensities (**Figure 4**). The majority of the hybridized fragments were common among the tested genotypes, indicating a high level of sequence similarity among the β-conglycinin genes. However, a few fragments showed polymorphism among the genotypes, especially in the *Eco*RI and *Xba*I digests (**Figure 4**). Because the *Eco*RI and *Xba*I restriction enzymes are known to exhibit sensitivity to methylation, these polymorphic fragments possibly result from different methylation patterns among the tested genotypes. The hybridization pattern of Kavre was nearly identical to that of Williams 82 in the three restriction digests except for the 6.5- and 10.5-kb polymorphic fragments in the *Xba*I digest. The hybridization patterns of the varieties Soida Chiny and Seti differed from each other and showed the least similarity to that of Williams 82 (**Figure 4**).

RFLP in the Genes Encoding Bowman−Birk-Type Proteinase Inhibitors. The Bowman-Birk-type proteinase inhibitors (BBi) are common constituents of plant seeds and have been linked to a variety of functions, including regulation of proteinase activity during germination and protection against insects and pathogens (*34*, *35*). BBi have received increased interest recently as potential anticarcinogens (*36*). BBi are encoded by a multigene family, which produce at least seven

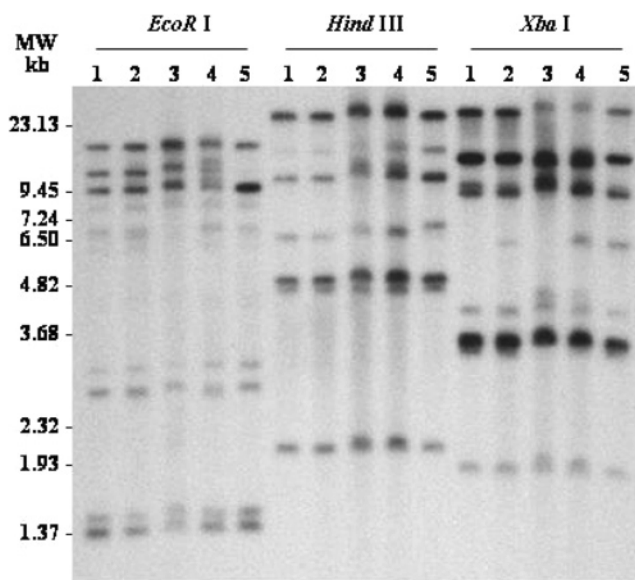

Figure 4. Southern blot analysis of soybean β-conglycinin. Genomic DNA from leaf tissue of Williams 82 (lane 1), Kavre (lane 2), Sathia (lane 3), Soida Chiny (lane 4), and Seti (lane 5) was individually digested with either *Eco*RI or *Hin*dIII or *Xba*I, fractionated on an 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with a ^{32}P-labeled β-subunit of the β-conglycinin gene and then subjected to autoradiography. Molecular weight markers in kb are shown on the left.

isoinhibitors that differ in protease inhibition specificities (*37*). Two well-characterized isoinhibitors in the BBi gene family include BBi A and BBi C-II (previously known as soy C-II). To examine the level of DNA sequence similarity of these BBi genes among the investigated genotypes, probes corresponding to the BBi A and BBi C-II genes were hybridized to restriction digests of genomic DNA. The Southern hybridization results presented in **Figure 5** showed very similar hybridization patterns among the tested genotypes, for both BBi A and BBi C-II. The only exception was in Soida Chiny, which showed relatively strong hybridization to one additional *Eco*RI fragment with the BBi C-II probe not seen in the other genotypes (**Figure 5B**). Thus, the results obtained with the probe/enzyme combinations

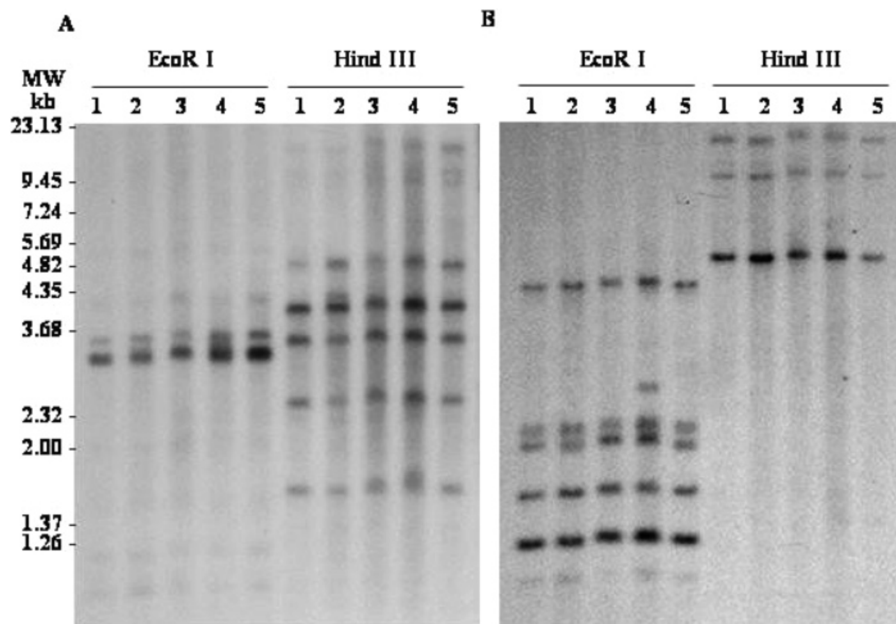

Figure 5. Southern blot analysis of soybean Bowman–Birk-type proteinase inhibitors. Genomic DNA from leaf tissue of Williams 82 (lane 1), Kavre (lane 2), Sathia (lane 3), Soida Chiny (lane 4), and Seti (lane 5) was individually digested with either *Eco*RI or *Hin*dIII, fractionated on an 0.8% agarose gel, and then transferred to a nylon membrane. The DNA fragments were hybridized with a ^{32}P-labeled BBi A (**A**) or BBi C-II (**B**) and then subjected to autoradiography. Molecular weight markers in kb are shown on the left.

used in the present study indicate very little or no sequence variation in the BBi genes among the tested genotypes.

DISCUSSION

Genetic diversity of the elite North American cultivars, derived primarily from a few plant introductions, is limited (*1*, *3*, *19*). Although continued yield increases through selective breeding over the past 60 years have been possible, there is concern that sustained improvement may be limited as the genetic base narrows (*7*, *38*−*40*). The narrow genetic base also affects our ability to improve other traits such as protein content, which depend on combinations of superior alleles through chromosome recombination or assortment (*41*). Protein content, a heritable trait, is ostensibly controlled by a minimal number of genes (*42*−*44*). Analysis of over 20000 accessions of soybean collected in China shows protein content ranging from 30 to 53% (*12*). Because the average protein content of soybeans produced in the United States is ≈35% (http://www.soygrowers.com/international/quality/US-05-Quality.pdf) there is optimism for increasing this component through the use of introduced germplasm.

Molecular genetic mapping and assessment of quantitative and qualitative traits of Asian soybeans provide evidence of substantial heterogeneity among the multitude of accessions. Whereas accessions from China and Japan exhibit distinct genetic differences, those found in southeast and south central Asia appear to be more similar to Chinese accessions (*12*, *14*, *16*, *17*, *45*, *46*). Significant diversity exists among Asian landraces, ostensibly due to selection for adaptability to the local environment and for qualities amenable for the preparation of native foods. In this study, we found RFLPs among group 1 glycinin genes of the Nepalese cultivars. RFLPs of these genes were distinct from those of the North American cultivar Williams 82. The genetic polymorphism is reflected in the differential accumulation of protein subunits and amino acid composition among the cultivars. Thus, our work reiterates the potential of diverse germplasm for improving the quality of North American soybean seed protein.

Seeds of the soybean cultivars, which were originally obtained from Nepal, had significantly higher protein content than Williams 82. The Nepalese cultivars are normally grown at altitudes ranging from 900 to 1500 m. The higher protein content of these cultivars appears to be controlled by genetic rather than environmental factors because they maintained higher protein content even when grown near Columbia, MO.

Two-dimensional electrophoretic analysis of seed protein of the Nepalese cultivar Sathia reveals marked heterogeneity of glycinin and β-conglycinin subunits. The cultivar Sathia was unique because it was devoid of the G4 glycinin A5A4B3 subunit. The absence of the G4 glycinin subunit in Sathia does not appear to be a trait that is shared by all of the Nepalese cultivars. Mutations resulting in the absence of specific proteins among soybean cultivars have been noted, and the molecular basis for their absence has been established (*32*, *47*−*51*). Previous studies have established that the absence of the A4 subunit results in the production of superior tofu (*52*). Soybean cultivars such as Raiden and Enrei, which lack the A4 subunit, are used in breeding programs to incorporate this trait. The absence of the A4 subunit of glycinin in Raiden has been attributed to point mutation in the initiator codon ATG of *Gy4* (*47*, *49*). Currently, we do not know the molecular basis for the absence of the A4 subunit in the Nepalese cultivar Sathia.

Enhancing the protein quality to meet the demands of modern livestock production and ostensibly to produce foods for direct human consumption is one facet of soybean seed improvement. Domestic swine and poultry consume >75% of the soybean protein meal produced in the United States. Although this meal is an excellent source of protein, quantities of specific amino acids in the prevalent corn−soybean meal rations are not optimal for growing swine and poultry. Lysine and methionine are the first limiting amino acids, whereas arginine, threonine, isoleucine, tryptophan, and valine are considered to be second-tier amino acids (*53*). Increased expression of several second-tier amino acids, which include arginine, could obviate the use of the expensive synthetic amino acids in ration formulation (*53*). The significantly higher concentrations of cysteine and arginine

in the Nepalese cultivars are desirable traits that could be incorporated into U.S. cultivars.

ACKNOWLEDGMENT

We thank Dr. Mel Oliver for critically reading the manuscript.

LITERATURE CITED

(1) Delannay, X.; Rodgers, D. M.; Palmer, R. G. Relative genetic contributions among ancestral lines to North America soybean cultivars. *Crop Sci.* **1983**, *23*, 944−949.

(2) Hymowitz, T.; Bernard, R. L. Origin of the soybean and germplasm introduction and development in North America. In *Use of Plant Introductions in Cultivar Development Part 1*; Shands, H. L., Ed.; Crop Science Society of America: Madison, WI, 1991; pp 149−167.

(3) Gizlice, Z.; Carter, T. E., Jr.; Burton, J. W. Genetic base of North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* **1994**, *34*, 1143−1151.

(4) Gizlice, Z.; Carter, T. E.; Gerig, T. M.; Burton, J. M. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci.* **1996**, *36*, 753−765.

(5) Thompson, J. A.; Nelson, R. L. Utilization of diverse germplasm for soybean yield improvement. *Crop Sci.* **1998**, *38*, 1362−1368.

(6) Specht, J. E.; Hume, D. J.; Kumudini, S. V. Soybean yield potential—a genetic and physiological perspective. *Crop Sci.* **1999**, *39*, 1560−1570.

(7) Wilcox, J. R. Sixty years of improvement in publicly developed elite soybean lines. *Crop Sci.* **2001**, *41*, 1711−1716.

(8) Gizlice, Z.; Carter, T. E., Jr.; Burton, J. W. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. *Crop Sci.* **1993**, *33*, 614−620.

(9) Kisha, T.; 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.

(10) Cornelious, B. K.; Sneller, C. H. Yield and molecular diversity of soybean lines derived from crosses of Northern and Southern elite parents. *Crop Sci.* **2002**, *42*, 642−647.

(11) Ude, G. N.; Kenworthy, W. J.; Costa, J. M.; Cregan, P. J.; Alvarnaz, J. Genetic diversity of soybean cultivars from China, Japan, North America, and North American ancestral lines determined by amplified fragment length polymorphism. *Crop Sci.* **2003**, *43*, 1858−1867.

(12) Dong, Y. S.; Zhao, L. M.; Liu, B.; Wang, Z. W. The genetic diversity of cultivated soybean grown in China. *Theor. Appl. Genet.* **2004**, *108*, 931−936.

(13) Chen, Y.; Nelson, R. L. Relationship between origin and genetic diversity in Chinese soybean germplasm. *Crop Sci.* **2005**, *45*, 1645−1652.

(14) Abe, J.; Xu, D. H.; Suzuki, Y.; Kanazawa, A.; Shimamoto, Y. Soybean germplasm pools in Asia revealed by SSR. *Theor. Appl. Genet.* **2003**, *106*, 445−453.

(15) Hymowitz, T.; Kaizuma, N. Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: hypotheses on paths of dissemination of soybeans from China. *Econ. Bot.* **1981**, *35*, 10−23.

(16) Hirata, T.; Abe, J.; Shimamoto, Y. Genetic structure of the Japanese soybean population. *Genet. Resour. Crop Evol.* **1999**, *46*, 441−453.

(17) Li, Z.; Nelson, R. J. Genetic diversity among soybean accessions from three countries measured by RAPDS. *Crop Sci.* **2001**, *41*, 1337−1347.

(18) Keim, P.; Beavis, W.; Schupp, J.; Freestone, R. Evaluation of soybean RFLP marker diversity in adapted germplasm. *Theor. App. Genet.* **1992**, *85*, 205−212.

(19) Sneller, C. H.; Miles, J. W.; Hoyt, J. M. Agronomic performance of soybean plant introductions and their genetic similarity to elite lines. *Crop Sci.* **1997**, *37*, 1595−1600.

(20) Nelson, R. L.; Cregan, P.; Boerma, H. R.; Carter, T. E.; Quigley, C. V.; Welsh, M. M.; et al. DNA marker diversity among modern North American Chinese and Japanese soybean cultivars. In *Agronomy Abstracts*; American Society of Agronomy: Madison, WI, 1998; p 164.

(21) Zhou, X.; Carter, T. E., Jr.; Cui, Z.; Miyazaki, S.; Burton, J. W. Genetic base of Japanese soybean cultivars released during 1950 to 1988. *Crop Sci.* **2000**, *40*, 1794−1802.

(22) Zhou, X.; Carter, T. E., Jr.; Cui, Z.; Miyazaki, S.; Burton, J. W. Genetic diversity patterns in Japanese soybean cultivars based on coefficient of parentage. *Crop Sci.* **2002**, *42*, 1331−1342.

(23) 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.

(24) Cui, Z.; Carter, T. E., Jr.; Burton, J. W.; Wells, R. Phenotypic diversity of modern Chinese and North American soybean cultivars. *Crop Sci.* **2001**, *41*, 1954−1967.

(25) Brim C. A.; Cockerham, C. C Inheritance of quantitative characters in soybean. *Crop Sci.* **1961**, *1*, 187−190.

(26) Reese, P. E., Jr.; Kenworthy, W. J.; Cregan, P. B.; Yocum, J. O. Comparison of selection systems for the identification of exotic soybean lines for use in germplasm development. *Crop Sci.* **1988**, *28*, 237−241.

(27) St. Martin, K. K.; Aslam, M. Performance of progeny of adapted and plant introduction soybean lines. *Crop Sci.* **1986**, *26*, 753−756.

(28) Cregan, P. B.; Jarvik, T.; Bush, A. L.; Shoemaker, R. C.; Lark, K. G.; Kahler, A. L.; et al. An integrated genetic linkage map of the soybean genome. *Crop Sci.* **1999**, *39*, 1464−1490.

(29) Natarajan, S.; Xu, C.; Caperna, T. J.; Garrett, W. A. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. *Anal. Biochem.* **2005**, *342*, 214−220.

(30) Laemmli, U. Cleavage of structureal proteins during the assembly of the head of the bacteriophage T4. *Nature* **1970**, *227*, 680−685.

(31) 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*, 8014−8018.

(32) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, B. H.; Scallon, B. J.; Fischer, R. L.; et al. Characterization of the glycinin gene family. *Plant Cell* **1989**, *1*, 313−328.

(33) Harada, K.; Toyokawa, Y.; Kitamura, K. Genetic analysis of the most acidic 11S globulin subunit and related characters in soybean seeds. *Jpn. J. Breed.* **1983**, *33*, 23−30.

(34) Mayer, A. M.; Shain, Y. Control of seed germination. *Annu. Rev. Plant Physiol.* **1974**, *25*, 167−193.

(35) Rahbe, Y.; Ferrasson, E.; Rabesona, H.; Quillien, L. Toxicity to the pea aphid *Acyrthosiphon pisum* of anti-chymotrypsin isoforms and fragments of the Bowman−Birk protease inhibitors from pea seeds. *Insect Biochem. Mol. Biol.* **2003**, *33*, 299−306.

(36) Kennedy, A. R. The Bowman−Birk inhibitor from soybeans as an anti-carcinogenic agent. *Am. J. Clin. Nutr.* **1998**, *68*, 1406S−1412S.

(37) Deshimaru, M.; Yoshimi, S.; Shioi, S.; Terada, S. Multigene family for Bowman−Birk type proteinase inhibitors of wild soja and soybean: The presence of two BBI-A genes and pseudogenes. *Biosci., Biotechnol., Biochem.* **2004**, *68*, 1279−1286.

(38) Fehr, W. R. Breeding methods for cultivar development. In *Soybeans: Improvement, Production, and Uses*, 2nd ed.; Wilcox, J. R., Ed.; American Society of Agronomy: Madison, WI, 1987; pp 249−294.

(39) Rasmussen, D. C.; Phillips, R. L. Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. *Crop Sci.* **1997**, *37*, 303−310.

(40) 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: Foz Do Iguassu, Pr, Brazil, 2004; p 56.

(41) Stefaniak, T. R.; Hyten, D. L.; Pantalone, V. R.; Klarer, A.; Pfeiffer, T. W. Soybean cultivars resulted from more recombination events than unselected lines in the same population. *Crop Sci.* **2006**, *46*, 43−51.
(42) Wilcox, J. R.; Cavins, J. F. Backcrossing high seed protein to a soybean cultivar. *Crop Sci.* **1995**, *35*, 1036−1041.
(43) Wilcox, J. R. Increasing seed protein content in soybean with eight cycles of recurrent selection. *Crop Sci.* **1998**, *38*, 1536−1540.
(44) Chung, J.; Babka, H. L.; Graef, G. L.; Staswick, P. E.; Lee, D. J.; Cregan, P. B.; Shoemaker, R. C.; Specht, J. E. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci.* **2003**, *43*, 1053−1067.
(45) Hymowitz, T.; Kaizuma, N. Dissemination of soybean (*Glycine max*): seed protein electrophoresis profiles among Japanese cultivar. *Econ. Bot.* **1979**, *33*, 311−319.
(46) Xu, D. H.; Gai, J. Y. Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. *Plant Breed.* **2003**, *122*, 503−503.
(47) Kitamura, K.; Davies, S. S.; Nielsen, N. C. Inheritance of alleles for Cgy_1 and Gy_4 storage protein genes in soybean. *Theor. Appl. Genet.* **1984**, *68*, 253−257.
(48) Ladin, B. F.; Doyle, J. J.; Beachy, R. N. Molecular characterization of a deletion mutation affecting the α'-subunit of β-conglycinin. *J. Mol. Appl. Genet.* **1984**, *2*, 372−389.
(49) Scallon, B. J.; Dickinson, C. D.; Nielsen, N. C. Characterization of a null-allele for the Gy 4 glycinin gene from soybean. *Mol. Genet. Genomics* **1986**, *208*, 107−113.
(50) Jofuku, K. D.; Schipper, R. D.; Goldberg, R. B. A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. *Plant Cell* **1989**, *1*, 427−435.
(51) Cho, T.-J.; Davies, C. S.; Fischer, R. L.; Turner, N. E.; Goldberg, R. B.; Nielsen, N. C. Molecular characterization of an aberrant allele for the Gy_3 glycinin gene: a chromosomal rearrangement. *Plant Cell* **1989**, *1*, 339−350.
(52) Murphy, P.; Chen, H.; Hauck, C.; Wilson, J. Soybean protein composition and tofu quality. *Food Technol.* **1997**, *51*, 86−88.
(53) 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.

Received for review April 14, 2006. Revised manuscript received May 30, 2006. Accepted June 2, 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 be suitable.

JF0610559