JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Characterization of Storage Proteins in Wild (*Glycine soja*) and Cultivated (*Glycine max*) Soybean Seeds Using Proteomic Analysis

SAVITHIRY S. NATARAJAN,*,† CHENPING XU,‡ HANHONG BAE,§ THOMAS J. CAPERNA,|| AND WESLEY M. GARRETT⊥

Soybean Genomics and Improvement Laboratory, Sustainable Perennial Crops Laboratory, Growth Biology Laboratory, and Biotechnology and Germplasm Laboratory, Agricultural Research Service, U.S. Department of Agriculture, PSI, Beltsville, Maryland 20705, and Department of Natural Resource Sciences and Landscape Architecture, University of Maryland, College Park, Maryland 20742

A combined proteomic approach was applied for the separation, identification, and comparison of two major storage proteins, β-conglycinin and glycinin, in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) with three different immobilized pH gradient (IPG) strips was an effective method to separate a large number of abundant and less-abundant storage proteins. Most of the subunits of β-conglycinin were well-separated in the pH range 3.0−10.0, while acidic and basic glycinin polypeptides were well-separated in pH ranges 4.0−7.0 and 6.0−11.0, respectively. Although the overall distribution pattern of the protein spots was similar in both genotypes using pH 3.0−10.0, variations in number and intensity of protein spots were better resolved using a combination of pH 4.0−7.0 and pH 6.0−11.0. The total number of storage protein spots detected in wild and cultivated genotypes was approximately 44 and 34, respectively. This is the first study reporting the comparison of protein profiles of wild and cultivated genotypes of soybean seeds using proteomic tools.

KEYWORDS: Soybean; *G. soja*; *G. max*; 2D-PAGE; MALDI-TOF-MS; LC-MS/MS; β-conglycinin; glycinin

INTRODUCTION

In soybean seeds, two major storage proteins, β-conglycinin and glycinin, account for about 70−80% of the total proteins. These are largely responsible for the nutritional and physicochemical properties of soybeans (*1*, *2*). β-Conglycinin, a 7S globulin (vicilin family), is a trimeric glycoprotein consisting of three types of subunits, α, α′, and β, in which only the α-subunit of β-conglycinin has allergenic reactions (*2*, *3*). Glycinin, an 11S globulin and a hexamer (legumin family), consists of five subunits, G1, G2, G3, G4, and G5, in which G1 and G2 are allergens. Each subunit of glycinin consists of acidic (A) and basic (B) chains, which are interlinked by a single disulfide bond, except for the acidic chain A4, present in the G4 subunit (*4*).

In recent years, the application of proteomic tools such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF-MS), and liquid chromatography mass spectrometry (LC-MS/MS) has become popular, and these tools are powerful methodologies for accurately detecting and examining changes in protein composition. These tools have been extensively used to examine the composition of both natural and transgenic soybean storage protein profiles and to determine seed qualities of soybeans. However, limited studies are available for detecting both abundant and nonabundant soybean seed proteins at the subunit level since it remains a challenging issue (*5*, *6*). Recently, Moony and Thelen (*7*) studied the protein composition of the cultivated genotype *Glycine max* cv. Jefferson by 2D-PAGE analysis with the pH range between 3 and 10.0.

Although isozyme-based assays do not provide enough resolution of variations, a substantial amount of information has been reported on the genotypic variation of soybeans (*8*−*10*). However, knowledge of soybean genotypic variation alone does not provide adequate information concerning the alteration of protein due to environmental interactions. For a better understanding of the consequences of genetic manipulation, elucidation of the protein composition is necessary because of its direct relationship to phenotype (*11*). Proteomics involves high-resolution separation techniques such as 2D-PAGE combined with microanalytical processes such as mass spectrometry. The

* To whom correspondence should be addressed. Tel: 301-504-5258. Fax: 301-504-5728. E-mail: natarajs@ba.ars.usda.gov.
† Soybean Genomics and Improvement Laboratory, U.S. Department of Agriculture.
‡ University of Maryland.
§ Sustainable Perennial Crops Laboratory, U.S. Department of Agriculture.
|| Growth Biology Laboratory, U.S. Department of Agriculture.
⊥ Biotechnology and Germplasm Laboratory, U.S. Department of Agriculture.

10.1021/jf052954k CCC: $33.50
Published on Web 03/28/2006

objective of this investigation was to establish a combined proteomic approach for the thorough analysis of storage protein compositions of soybean seeds. We have studied the protein profiles of wild (*Glycine soja*) and cultivated (*G. max*) genotypes using three different immobilized pH gradient (IPG) strips (pH 3.0−10.0, 4.0−7.0, and 6.0−11.0) for the first dimension of isoelectric focusing (IEF) and second dimension of PAGE and identified storage proteins by combining both MALDI-TOF-MS and LC-MS/MS analysis.

MATERIALS AND METHODS

Plant Materials. Soybean seeds of *G. soja* PI 393551, which originated from Taiwan, and *G. max* PI 423954 cv. Shirome, which originated from Japan, were obtained from the U.S. Department of Agriculture soybean germplasm collection (Urbana, IL). Seeds were stored at −80 °C until used.

Chemicals. Chemicals for electrophoresis including acrylamide, bisacrylamide, sodium dodecyl sulfate (SDS), TEMED, ammonium persulfate, thiourea, dithiothreitol, CHAPS, and IPG strips were purchased from GE Healthcare (Piscataway, NJ). Urea and ampholytes (pH 3.0−10.0, 4.0−7.0, and 6.0−11.0) were purchased from Bio-Rad Laboratories (Hercules, CA). Tris-HCl (pH 8.8), 2-mercaptoethanol (2-ME), trichloroacetic acid (TCA), and glycerol were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO). α-Cyanohydroxycinnamic acid (CHCA) matrix was purchased from Bruker Daltonics (Billerica, MA). Water from a Millipore Milli-RO4 reverse osmosis system was used for making all solutions.

Extraction of Proteins from Seeds: Modified TCA/Acetone Precipitation/Urea Solubilization Extraction. This protocol was performed according to Natarajan et al. (*12*). For this method, soybean seeds were powdered in liquid nitrogen using a mortar and pestle. One hundred milligrams of the soybean seed powder was homogenized with 2 mL of a solution containing 10% (w/v) TCA in acetone with 0.07% (v/v) 2-ME. The total protein was precipitated for 1 h at −20 °C. The extract was centrifuged at 20800*g* for 20 min at 4 °C. The pellet was washed 2−3 times with acetone containing 0.07% (v/v) 2-ME. Then, the pellet was dried under vacuum for 30 min and the acetone dry powder was resuspended in 1 mL of lysis buffer [9 M urea, 1% CHAPS, 1% (w/v) one of the ampholytes (pH 3.0−10.0, 4.0−7.0, and 6.0−11.0)] and 1% dithiothreitol (DTT) followed by sonication for 30 min. The insoluble material was removed by centrifugation at 20800*g* for 20 min at 4 °C, and the supernatant was used in 2D-PAGE analysis. The protein concentration was determined according to Bradford method (*13*) using a commercial dye reagent from Bio-Rad. We have taken an aliquot of supernatant that contained 100 μg of protein for 2D-PAGE analysis.

2D-PAGE Analysis. The first dimension IEF was performed using 13 cm pH 3.0−10.0, 4.0−7.0, and 6.0−11.0 linear IPG strips in a IPGphor apparatus (GE Healthcare), according to the manufacturer's recommendations. A protein molecular weight standard was used in the second dimension in each gel. For the second dimension, the IPG strips were incubated with 50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue, and 1% DTT for 15 min and acetylated with iodoacetamideand subsequently placed onto 12% polyacrylamide gels prepared as described by Laemmli (*14*). The electrophoresis was performed using a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare) according to the manufacturer's recommendations. The 2D-PAGE gels were visualized by staining with Colloidal Coomassie Blue G-250 as described by Newsholme (*15*). The gels were stored in 20% ammonium sulfate solution and scanned using laser densitometry (GE Healthcare). Three separate seed samples extracted individually were taken for 2D-PAGE analysis.

In-Gel Digestion of Protein Spots. Protein spots were excised from the stained gel and washed first with distilled water to remove ammonium sulfate and then with 50% acetonitrile containing 25 mM ammonium bicarbonate to destain the gel plug. The gel plug was dehydrated with 100% acetonitrile, dried under vacuum, and then reswollen with 20 μL of 10 μg/mL trypsin (modified porcine trypsin, sequencing grade, Promega, Madison, WI) in 25 mM ammonium bicarbonate. Digestion was performed overnight at 37 °C. The resulting tryptic fragments were extracted with 100 μL of 50% acetonitrile and 5% trifluoroacetic acid with sonication. The extract was dried to completeness and dissolved in 5 μL of 50% acetonitrile and 0.1% trifluoroacetic acid.

Protein Identification: MALDI-TOF-MS Analysis. A Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems, Framingham, MA) operated in positive ion reflector mode was used to analyze tryptic peptides. Samples were cocrystallized with CHCA matrix, and spectra were acquired with 50 shots of a 337 nm nitrogen laser operating at 20 Hz. Spectra were calibrated using the trypsin autolysis peaks as internal standards at *m*/*z* 842.51 and 2211.10.

LC-MS/MS Analysis. A Thermo Finnigan LCQ Deca XP plus Ion Trap mass spectrometer was used to analyze proteins that were not positively identified by MALDI-TOF-MS. Peptides were separated on a reverse phase column using a 30 min gradient of 5−60% acetonitrile in water with 0.1% formic acid. The instrument was operated with a duty cycle that acquired MS/MS spectra on the three most abundant ions identified by a survey scan from 300 to 2000 Da. Dynamic exclusion was employed to prevent the continuous analysis of the same ions. Once two MS/MS spectra of any given ion had been acquired, the parent mass was placed on an exclusion list for a duration of 1.5 min. The raw data were processed by Sequest to generate DTA files for database searching. The *merge pl* script from Matrix Science (Boston, MA) was used to convert multiple Sequest DTA files into a single Mascot generic file suitable for searching in Mascot.

Data Interpretation Using Database. The trypsin-digested proteins generated reproducible peptide fragments whose molecular masses could be accurately determined by MALDI-TOF-MS. These experimentally determined mass-to-charge ratio (*m*/*z*) values were used to search theoretically generated peptide masses from known protein sequences (*16*, *17*). Protein identification was performed by searching the National Center for Biotechnology Information (NCBI) nonredundant database using the Mascot search engine (www.matrixscience.com), which uses a probability-based scoring system (*18*). The following parameters were used for database searches with MALDI-TOF-MS peptide mass fingerprinting data: monoisotopic mass, 25 ppm mass accuracy, trypsin as digesting enzyme with one missed cleavage allowed, and carbamidomethylation of cysteine, oxidation of methionine, and N-terminal pyroglutamic acid formation from glutamic acid or glutamine as allowable variable modifications. For database searches with MS/MS spectra, the parameters were used as follows: average mass, 1.0 Da peptide and MS/MS mass tolerance, peptide charges of +1, +2, or +3, trypsin as digesting enzyme with one missed cleavage allowed, carbamidomethylation of cysteine as a fixed modification, oxidation of methionine, and N-terminal pyroglutamic acid formation from glutamic acid or glutamine as allowable variable modifications. Taxonomy was limited to green plants for both MALDI-TOF-MS and MS/MS ion searches. To qualify the MALDI-TOF-MS data as a positive identification, the molecular weight search (MOWSE) score was equal to or exceeded the minimum significant score of 64. For LC-MS/MS, the positive identification required a minimum of two unique peptides with at least one peptide having a significant ion score.

RESULTS AND DISCUSSION

In this study, we applied 2D-PAGE with three different pH ranges and combined with MALDI-TOF-MS and LC-MS/MS to study and compare the protein composition of two major soybean storage proteins, β-conglycinin and glycinins, in wild (*G. soja*) and cultivated (*G. max*) soybean genotypes. These proteins are grouped into two types based on their sedimentation coefficients (7S and 11S). β-Conglycinin is encoded by two mRNA groups (*19*). The first mRNA group encodes α and α′ of β-conglycinin subunits, and the second mRNA group encodes the β-subunit of β-conglycinin (*3*, *19*). These three subunits of β-conglycinin are encoded by a gene family containing at least 15 members, which are divided into two major groups encoding 2.5 and 1.7 kb mRNAs and are clustered in several regions within the soybean genome (*20*). Glycinins have five nonallelic

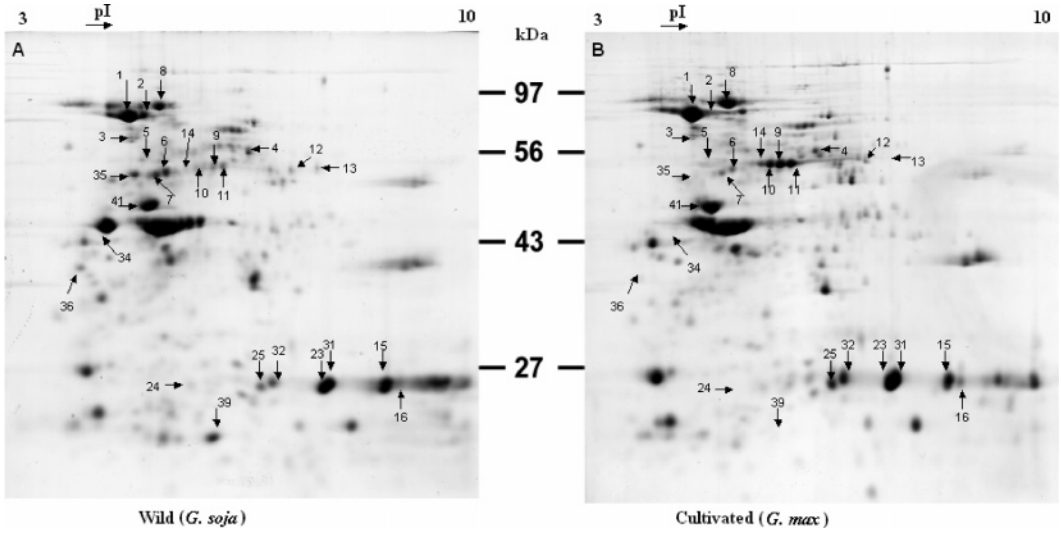

Figure 1. Proteomic comparison of the storage proteins of wild, *G. soja* PI 393551 (**A**), and cultivated, *G. max* PI 423954 (**B**), soybean seeds. The first dimension was run using a pH gradient from 3 to 10.0, and the second dimension was a 12% SDS-PAGE. Gels were stained with Colloidal Coomassie Blue stain G-250. Arrows indicate the spots that were abundant/nonabundant storage proteins, and numbered proteins are described in the text.

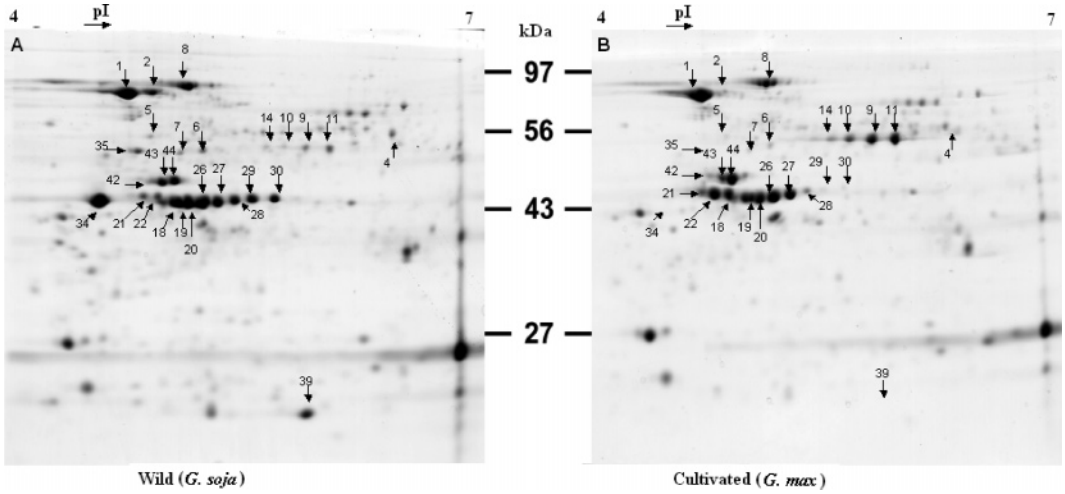

Figure 2. Proteomic comparison of the storage proteins of wild, *G. soja* PI 393551 (**A**), and cultivated, *G. max* PI 423954 (**B**), soybean seeds. The first dimension was run using a pH gradient from 4 to 7.0, and the second dimension was a 12% SDS-PAGE. Gels were stained with Colloidal Coomassie Blue stain G-250. Arrows indicate the spots that were abundant/nonabundant storage proteins, and numbered proteins are described in the text.

genes, Gy1, Gy2, Gy3, Gy4, and Gy5, which code for five glycinin protein precursor molecules, G1, G2, G3, G4, and G5, respectively (*21*). Each precursor subunit protein consists of two or three chains, which are cleaved posttranslationally.

Comparison of Storage Proteins. Our results showed that the overall distribution pattern of proteins is quite similar in wild and cultivated soybeans following separation with broad pH range (pH 3.0−10.0) IPG strips (**Figure 1A**,**B**). However, the number of protein spots (polypeptides) and spot intensity of the storage proteins, β-conglycinin and glycinin, varied between wild and cultivated soybean when we used narrow pH ranges. The total number of storage protein spots detected in wild and cultivated genotypes was approximately 44 and 34 in all three pH ranges, respectively (**Figure 1A**,**B**). Because the 2D-PAGE analysis with the pH 3.0−10.0 strip did not adequately separate some of the storage protein subunits, we used two narrow pH IPG strips, pH range 4.0−7.0, for acidic proteins (**Figure 2A**,**B**) and pH 6.0−11.0 range for basic proteins (**Figure 3A**,**B**). These gels showed distinct separation of most of the storage proteins of both wild and cultivated soybean seeds. A total of 47 different protein spots were excised from the 2D-PAGE of IEF pH ranges 3.0−10.0, 4.0−7.0, and 6.0−11.0 and were characterized using mass spectrometry. Thirty-seven protein spots were identified using MALDI-TOF-MS, and protein spots, which could not be identified, were subsequently analyzed by LC-MS/MS. Ten protein spots were identified, including less-abundant storage proteins, using LC-MS/MS. Recently, Mooney and Thelan (*3*) reported 17 storage protein spots of β-conglycinin and glycinin using 2D-PAGE with an IEF pH range of 3.0−10.0 in a cultivated genotype. All of the proteins identified by MALDI-TOF-MS, LC-MS/MS, and NCBI nonredundant database searches are listed in **Table 1**. The table consists of an assigned protein spot number, theoretical isoelectric point (pI), and molecular weight (M_r), protein identity, number of peptides matched, percent sequence coverage, MOWSE score, expect value, and NCBI database accession number of the best match.

Variation of β-Conglycinin in Wild and Cultivated Genotypes. Comparison of proteins from the wild type and the cultivated soybean genotypes by 2D-PAGE using pH 3.0−10.0 IPG strips (**Figure 1A**,**B**) showed three subunits of β-conglycinin, the α-, α'-, and β-subunits. The α-subunit showed five protein spots (#1, 3, 4, 6, and 7) in the cultivated genotype (**Figure 1B**). In the wild genotype, seven spots were resolved,

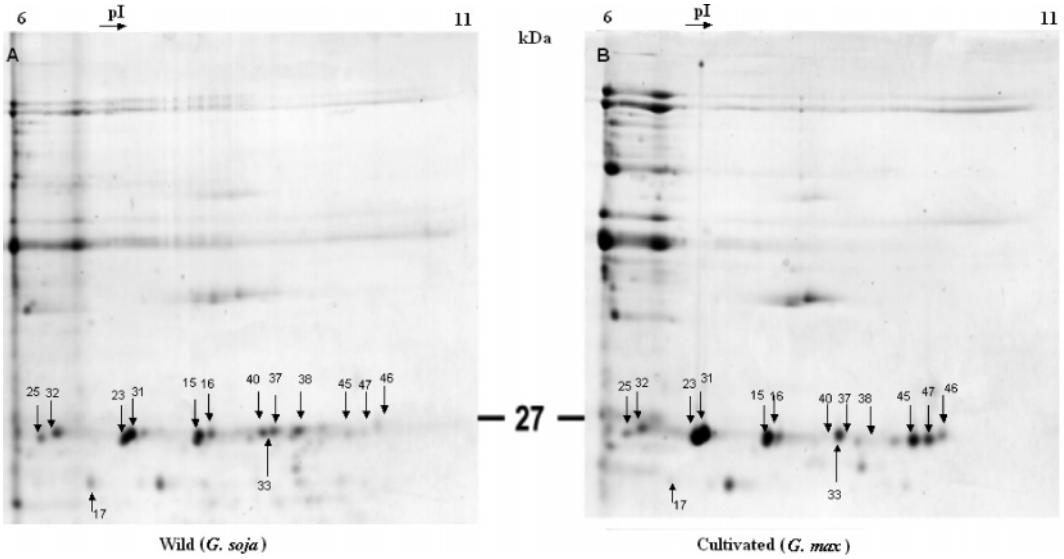

Figure 3. Proteomic comparison of the storage proteins of wild, *G. soja* PI 393551 (**A**), and cultivated, *G. max* PI 423954 (**B**), soybean seeds. The first dimension was run using a pH gradient from 6 to 11.0, and the second dimension was a 12% SDS-PAGE. Gels were stained with Colloidal Coomassie Blue stain G-250. Arrows indicate the spots that were abundant/nonabundant storage proteins, and numbered proteins are described in the text.

which include the above five spots and two additional spots, #2 and 5 (**Figure 1A**). The intensity of #6 and 7 spots were higher in the wild genotype as compared to the cultivated genotype. The α-subunit of β-conglycinin (spot #1) resolved at a molecular mass of 68 kDa (observed value) using pH 3.0−10.0 (**Figure 1**). This was similar to the previous report of 65 kDa for the α-subunit of β-conglycinin by Mooney and Thelan (*3*). The remaining spots demonstrated heterogeneity in molecular weight and isoelectric point (pI). These differences could be the result of posttranslational modifications, proteolysis, or the result of alternate splicing from the same gene or products of a multigene family, which have not been identified. The α′-subunit of β-conglycinin showed one spot (#8) in both wild and cultivated genotypes (**Table 1**), but the intensity of this spot was markedly lower in the wild (**Figure 1A**) as compared to the cultivated genotype (**Figure 1B**).

The β-subunit of β-conglycinin resolved into four protein spots, (#9 and 11−13) in the wild genotype (**Figure 1A**) and five protein spots (#9−12 and 14) in the cultivated genotype using pH 3.0−10.0 IEF gradients (**Figure 1B**). All separated polypeptides of the β-subunit of β-conglycinin showed a higher intensity of protein spots in the cultivated genotype as compared to the wild genotype (**Figure 1A**,**B**). In addition, all six spots ran at approximately 48 kDa, which is similar to the previous report of 45−49 kDa (*1*, *3*, *22*). In our investigation, β-conglycinin showed heterogeneity in its subunit composition, which is similar to other reports (*23*, *24*). Schuler et al. (*25*) reported that β-conglycinin subunits are products of a multigene family, and the variation in the distribution of protein spots in our study could also be due to posttranslational modifications (*3*). Similarly, Davies et al. (*24*) reported that β-conglycinin is known to undergo extensive co- and posttranslational modification, although the exact steps between translation and packaging into protein bodies remain unclear. Sengupta et al. (*22*) suggested that a heterogeneous banding pattern observed in one dimensional SDS-PAGE of β-conglycinin could be due to a sequence of glycosylation, deglycosylation, and proteolysis.

Variation of Glycinin in Wild and Cultivated Soybean Genotypes. Glycinin, the second major soybean storage protein, showed both acidic and basic polypeptides in pH 3.0−10.0 ranges of IEF; however, many did not separate well in this wide pH range and thus could not be identified (**Figure 1A**,**B**). Hence, we used narrow pH range IEF strips (4.0−7.0 and 6.0−11.0) to separate and identify the acidic and basic chains of glycinin subunits, G1, G2, G3, G4, and G5.

In all, 10 of a total of 11 glycinin chains were separated and identified. The A1a chain was not observed in our separation system, and full-length precursors were not observed for any of the glycinin molecules. The G1/A1aBx subunit showed three basic polypeptides (spots #15−17) in the pH 6.0−11.0 range that were similar in both wild and cultivated genotypes (**Figure 3A**,**B**), which were identified as the Bx polypeptide of G1. Similarly, using narrow pH range IEF strips (pH 4.0−7.0), we identified glycinin G2/A2B1a, which showed eight spots (#18−25) of acidic and basic polypeptides in the wild genotype and seven spots in cultivated genotype with the absence of spot #24 (**Figure 2A**,**B**). Among these spots, #18−22 were identified as the A2 chain and #23−25 were identified as the B1a chain of G2. We also observed considerable variations in the intensity of protein spots between the genotypes. The variation in the distribution of protein spots with different pI values observed from 2D-PAGE image is most likely due to posttranslational modifications. G3/A1ab1B glycinin acidic and basic polypeptides resolved into seven spots (#26−32) in the wild genotype, with the absence of polypeptide #33 (**Figure 3A**,**B**). Among these, spots #26−30 were identified as A1a and spots #31−33 were identified as B1b components of G3. However, the cultivated genotype showed an absence of two spots (#29 and 30), weak intensity of spot #28, and the presence of an additional spot, #33 (**Figure 2B** and **Table 1**). Two acidic polypeptides, which were identified as glycinin G4/A5A4B3 (spot #34, abundant, and spot #35, less abundant), were present in the wild (*G. soja*) genotype (**Figure 1A**), while these spots were completely absent in the cultivated (*G. max*) genotype (**Figure 1B**). Glycinin G4/A5A4B3 acidic and basic polypeptides separated into five spots (#36−40), were only present in the wild type seeds and showed strong intensity (**Figures 1A**,**B** and **3A**,**B**). Among these spots, #37, 38, and 40 were identified as the B3 components, spot #39 represented the A5 component, and spot #36 was identified as the A4 component of G4. Genetic studies have shown that the coordinate loss of A4, A5, and B3 is apparently due to a single recessive genetic characteristic (*26*). The absence of G4 subunits in cultivated genotypes observed in our study could be due to the absence of the gene(s) encoding

Table 1. Storage Proteins Identified by MALDI-TOF-MS and LC-MS/MS Analysis in Wild and Cultivated Soybean Genotypes

spot ID	theoretical pI/M_r	protein identity	peptides matched	sequence coverage (%)	MOWSE score	expected value	NCBI accession no.	wild (*G. soja*)	cultivated (*G. max*)	method/MS
1	4.92/63127	α-subunit of β-conglycinin	25	39	217	3.00E−17	gi|9967357	+	+	MALDI-TOF
2	4.92/63127	α-subunit of β-conglycinin	25	41	167	4.00E−12	gi|9967357	+	−	MALDI-TOF
3	4.92/63184	α-subunit of β-conglycinin	27	43	250	1.50E−20	gi|9967357	+	+	MALDI-TOF
4	4.92/63184	α-subunit of β-conglycinin	20	34	204	6.00E−16	gi|9967357	+	+	MALDI-TOF
5	5.32/72717	α-subunit of β-conglycinin	9	18	73	7.20E−03	gi|15425633	+	−	MALDI-TOF
6	5.32/72717	α-subunit of β-conglycinin	14	24	112	1.10E−03	gi|15425633	+	+	MALDI-TOF
7	5.32/72717	α-subunit of β-conglycinin	9	23	238		gi|51247837	+	+	LC-MS/MS
8	5.23/65160	α′-subunit of β-conglycinin	20	41	194	5.90E−15	gi|9967361	+	+	MALDI-TOF
9	5.67/48358	β-conglycinin β-subunit	26	47	216	5.10E−17	gi|63852207	+	+	MALDI-TOF
10	5.67/48358	β-conglycinin β-subunit	18	40	176	5.10E−13	gi|63852207	-	+	MALDI-TOF
11	5.67/48358	β-conglycinin β-subunit	27	49	256	5.10E−21	gi|63852207	+	+	MALDI-TOF
12	5.67/48358	β-conglycinin β-subunit	22	46	127	4.00E−08	gi|63852207	+	+	MALDI-TOF
13	5.67/48358	β-conglycinin β-subunit	11	28	85	6.40E−04	gi|63852207	+	−	MALDI-TOF
14	5.67/48358	β-conglycinin β-subunit	23	45	218	3.20E−17	gi|63852207	−	+	MALDI-TOF
15	5.89/56299	glycinin G1/A1aBx subunit	7	15	145		gi|18635	+	+	LC-MS/MS
16	5.89/55672	glycinin G1/A1aBx subunit	5	10	157		gi|18635	+	+	LC-MS/MS
17	6.15/56134	glycinin G1/A1aBx subunit	4	9	254		gi|72296	+	+	LC-MS/MS
18	5.46/54927	glycinin G2/A2B1 precursor	9	19	91	1.20E−04	gi|1212177	+	+	MALDI-TOF
19	5.46/54927	glycinin G2/A2B1 precursor	9	19	73	7.20E−03	gi|1212177	+	+	MALDI-TOF
20	5.46/54927	glycinin G2/A2B1 precursor	8	15	71	1.30E−02	gi|1212177	+	+	MALDI-TOF
21	5.46/54927	glycinin G2/A2B1 precursor	9	14	74	6.70E−03	gi|1212177	+	+	MALDI-TOF
22	5.46/54927	glycinin G2/A2B1 precursor	12	21	104	6.10E−06	gi|1212177	+	+	MALDI-TOF
23	5.78/54047	glycinin G2/A2B1 precursor	6	37	67	3.20E−02	gi|169967	+	+	MALDI-TOF
24	5.56/54903	glycinin G2/A2B1 precursor	3	9	175		gi|72295	+	−	LC-MS/MS
25	5.56/54903	glycinin G2/A2B1 precursor	9	14	313		gi|72295	+	+	LC-MS/MS
26	5.78/54047	glycinin subunit G3/A1ab1B	8	19	72	1.00E−02	gi|15988117	+	+	MALDI-TOF
27	5.78/54047	glycinin subunit G3/A1ab1B	8	18	70	1.40E−02	gi|15988117	+	+	MALDI-TOF
28	5.78/54047	glycinin subunit G3/A1ab1B	9	20	74	6.40E−03	gi|15988117	+	+	MALDI-TOF
29	5.78/54047	glycinin subunit G3/A1ab1B	10	20	75	6.00E−03	gi|15988117	+	−	MALDI-TOF
30	5.78/54047	glycinin subunit G3/A1ab1B	9	18	96	4.20E−05	gi|15988117	+	−	MALDI-TOF
31	5.78/54047	glycinin subunit G3/A1ab1B	8	18	72	9.40E−03	gi|15988117	+	+	MALDI-TOF
32	5.78/54047	glycinin subunit G3/A1ab1B	9	18	116	3.90E−07	gi|15988117	+	+	MALDI-TOF
33	5.73/54835	glycinin subunit G3/A1ab1B	2	4	148		gi|18639	−	+	LC-MS/MS
34	5.38/64097	glycinin G4/A5A4B3 precursor	13	16	76	4.20E−03	gi|99910	+	−	MALDI-TOF
35	5.38/64097	glycinin G4/A5A4B3 precursor	13	26	96	3.50E−05	gi|99910	+	−	MALDI-TOF
36	4.46/24349	glycinin G4/A5A4B3 precursor	10	38	110	1.50E−06	gi|6015515	+	−	MALDI-TOF
37	6.47/31065	glycinin G4/A5A4B3 precursor	10	45	142	9.50E−10	gi|81785	+	−	MALDI-TOF
38	6.47/31065	glycinin G4/A5A4B3 precursor	8	40	105	5.40E−06	gi|81785	+	−	MALDI-TOF
39	5.38/64136	glycinin G4/A5A4B3 precursor	6	8	160		gi|99910	+	−	LC-MS/MS
40	5.38/64136	glycinin G4/A5A4B3 precursor	14	14	408		gi|99910	+	−	LC-MS/MS
41	5.46/55850	glycinin G5/A3B4 subunit	13	26	96	3.50E−05	gi|33357661	+	+	MALDI-TOF
42	5.46/55850	glycinin G5/A3B4 subunit	11	18	76	4.30E−03	gi|33357661	+	+	MALDI-TOF
43	5.46/55850	glycinin G5/A3B4 subunit	12	22	110	1.50E−06	gi|33357661	+	+	MALDI-TOF
44	5.46/55850	glycinin G5/A3B4 subunit	10	21	86	3.70E−04	gi|33357661	+	+	MALDI-TOF
45	5.69/26938	glycinin G5/A3B4 subunit	7	30	65	5.00E−02	gi|541941	+	+	MALDI-TOF
46	5.69/26938	glycinin G5/A3B4 subunit	8	37	69	2.50E−02	gi|541941	+	+	MALDI-TOF
47	9.64/21482	glycinin G5/A3B4 subunit	3	12	109		gi|625538	+	+	LC-MS/MS

G4 subunits or could be present in amounts below our detection level. Nielson et al. (*21*) reported the absence of G4 in soybean cv. Raiden and suggested that heterogeneity of the G4 subunit may not have functional relevance. The glycinin G5/A3B4 polypeptide (spot #41) that was clustered in the pH 3.0−10.0 gels (**Figure 1A**,**B**) was clearly separated into three polypeptides (spots #42−44) in the narrow pH 4.0−7.0, in both wild and cultivated genotypes (**Figure 2A**,**B**), and they were identified as the A3 component of G5. These results indicate that the narrow pH range IPG strips are necessary to enhance the separation and resolution of the detectable protein spots of acidic polypeptides. Three basic polypeptides of G5/A3B4 (spots #45−47) showed strong intensity in the cultivated genotype and weak intensity in the wild genotype, and they were identified as the B4 component of glycinin G5/A3B4 (**Figure 3A**,**B**). Others have reported variation in glycinin mRNAs (*27*) and polypeptides (*6*). Our study showed considerable variations of both acidic and basic glycinin polypeptides between wild and cultivated genotypes.

In conclusion, analyses of subunit composition of the major soybean storage proteins, β-conglycinin and glycinin, demonstrate a higher degree of overall electrophoretic heterogeneity of the proteins in wild as compared to cultivated genotypes as demonstrated by the appearance of more distinct and diverse protein spots. Similar heterogeneity of storage proteins between wild and cultivated beans was reported in broad bean and pea legumin (*28*, *29*). In the present study, the wild genotype showed fewer protein spots for the β-subunit of β-conglycinin and the G3 and G5 polypeptides, while the cultivated genotype showed fewer protein spots of the α-subunit of β-conglycinin. The G4 polypeptides were completely absent in the cultivated genotype. Saio et al. (*30*) reported that the proportion of β-conglycinin and glycinin is important and is responsible for the differences

in the physical properties of tofu gel. Similar variations of storage proteins among genotypes have been reported in peas and soybeans (*6*, *31*−*33*). The large amount of variation in seed protein composition between wild and cultivated genotypes may be due to a different complement of genes in the wild genotype that control expression of β-conglycinin and glycinin protein composition as compared to the cultivated genotypes (*34*). Sebolt et al. (*35*) reported that the wild genotype has increased protein content that was associated with a specific quantitative trait locus allele. The larger quantities of glycinin subunits observed in wild soybeans contain greater methionine and cysteine contents and also have properties vital for gelatination in the preparation of soy food products (*36*). The observed variation between wild and cultivated genotypes might be due to the breeder's effort to select better genotypes for protein and oil production in the United States (*37*, *38*). Our study demonstrated that proteomic analysis in general could help to define specific changes in protein level and composition, which can occur in the generation of new soybean varieties. The comparative studies of the storage proteins of wild and cultivated soybeans would help us to understand the evolutionary relationship between them.

ACKNOWLEDGMENT

We thank Dr. N. C. Talbot and Dr. R. Sicher for their critical review of this manuscript.

LITERATURE CITED

(1) Maruyama, N.; Adachi, M.; Takahashi, K.; Yagasaki, K.; Kohna, M.; Takenaks, Y.; Okuda, E.; Nakagawa, S.; Mikami, B.; Utsumi, S. Crystal structures of recombinant and native soybean β-conglycinin β homotrimers. *Eur. J. Biochem*. **2001**, *268*, 3595−3607.

(2) Thanh, V. H.; Shibasaki, K. Major proteins of soybean seeds: A straightforward fractionation and their characterization. *J. Agric. Food Chem*. **1976**, *24*, 1117−1121.

(3) Koshiyama, I. Storage proteins of soybean. In *Seed Proteins Biochemistry, Genetics, Nutritive Value*; Gottschalk, W., Miller, H. P., Eds.; Martinus Nijhoff/Dr. W. Junk Publisher: The Hague, 1983; pp 427−450.

(4) Renkema, J. M. S.; Knabben, J. H. M.; Vliet, T. Gel formation by beta-conglycinin and glycinin and their mixtures. *Food Hydrocolloids* **2001**, *15*, 407−414.

(5) Herman, E. M.; Helm, R. M.; Jung, R.; Kinney, A. J. Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol.* **2003**, *132*, 36−43.

(6) Hurkman, W. J.; Tanaka, C. K. Polypeptide changes induced by salt stress, water deficit, and osmotic stress in barley roots: A comparison using two-dimensional gel electrophoresis. *Electrophoresis* **1988**, *9*, 781−787.

(7) Mooney, B. P.; Thelen, J. J. High-throughput peptide mass fingerprinting of soybean seed proteins: Automated workflow and utility of UniGene expressed sequence tag databases for protein identification. *Phytochemistry* **2004**, *65*, 1733−1744.

(8) Griffin, J. D.; Palmer, R. G. Variability of 13 isozymes loci in the USDA soybean germplasm collections. *Crop Sci*. **1995**, *35*, 897−904.

(9) Maughan, P. J.; Saghai-Maroof, M. A.; Buss, G. R.; Huestis, G. M. Amplified fragment length polymorphism (AFLP) in soybeans: Species diversity, inheritance, and near-isogenic line analysis. *Theor. Appl. Genet.* **1996**, *93*, 392−401.

(10) Li, Z.; Nelson, R. L. RAPD marker diversity among cultivated and wild soybean accessions from our Chinese provinces. *Crop Sci*. **2002**, *42*, 1737−1744.

(11) Skylas, D. J.; Mackintosh, J. A.; Cordwell, S. J.; Basseal, D. J.; Walsh, B. J.; Harry, J.; Blumenthal, C.; Copeland, L.; Wrigley, C. W.; Rathmell, W. Proteome approach to the characterization of protein composition in the developing and mature wheat-grain endosperm. *J. Cereal Sci.* **2000**, *32*, 169−188.

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

(13) Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding. *Anal. Biochem.* **1976**, *72*, 248−254.

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

(15) Newsholme, S. J.; Maleeft, B. F.; Steiner, S.; Anderson, N. L.; Schwart, L. W. Two-dimensional electrophoresis of liver proteins: Characterization of a drug-induced hepatomegaly in rats. *Anal. Biochem*. **2000**, *21*, 2122−2128.

(16) Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1993**, *33*, 1−19.

(17) Eng, J.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein databases. *J. Am. Soc. Mass Spectrum.* **1994**, *5*, 976−989.

(18) Perkins, D. N.; Pappinl, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551−3567.

(19) Schuler, M. A.; Ladin, B. F.; Fryer, G.; Pollaco, J. G.; Beachy, R. N. Structural sequences are conserved in the genes coding for the α', α, and β subunits of the soybean seed storage protein and other seed proteins. *Nucleic Acids Res.* **1982b**, *10*, 8245−8260.

(20) Harada, J. J.; Barker, S. J.; Goldberg, R. B. Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*. **1989**, *1*, 415−425.

(21) Nielsen, N. C.; Dickinson, C. D.; Cho, T. J.; Thanh, V. H.; Scallon, B. J.; Fischer, R. L.; Sims, T. L.; Goldberg, G. N. Characterization of the glycinin gene family in soybean. *Plant Cell*. **1989**, *1*, 313−328.

(22) Sengupta, C.; Deluca, V.; Bailey, D. S.; Verma, D. P. S. Posttranslational processing of 7S and 11S components of soybean storage proteins. *Plant Mol. Biol.* **1981**, *1*, 19−34.

(23) Coates, J. B.; Medeiros, J. S.; Thanh, V. H.; Nielsen, N. C. Characterization of the subunits of beta-conglycinin. *Arch. Biochem. Biophys*. **1985**, *243*, 184−194.

(24) Davies, C. S.; Coates, J. B.; Nielson, N. C. Inheritance and biochemical analysis of four electrophoretic variants of β-conglycinin from soybeans. *Theor. Appl. Genet*. **1985**, *71*, 351−358.

(25) Schuler, M. A.; Schmitt, E. S.; Beachy, R. N. Closely related families of genes code for the α and α' subunitsof the soybean 7S storage protein complex. *Nucleic Acids Res*. **1982**, *10*, 8225−8243.

(26) Staswick, P. E.; Nielsen, N. C. Characterization of soybean cultivars lacking certain glycinin subunits. *Arch. Biochem. Biophys*. **1983**, *223*, 1−8.

(27) Fukazawa, C.; Momma, H.; Hirano, H.; Harada, K.; Udaka, K. Glycinin A3B4 mRNA. *J. Biol. Chem*. **1985**, *260*, 6234−6239.

(28) Utsumi, S.; Inaba, H.; Mori, T. Formation of pseudo and hybrid 11S globulins from subunits of soybean and broad bean 11S globulins. *Agric. Biol. Chem*. **1980**, *44*, 1891−1896.

(29) Thomson, J. A.; Shroeder, H. E. Cotyledonary storage proteins in *Pisum sativum*. II. Hereditary variation in components of legumin and vicillin fractions. *Aust. J. Plant Physiol.* **1978**, *5*, 263−90.

(30) Saio, K.; Kamiya, M.; Watanable, T. Food processing characteristics of soybean 11S and 7S proteins. *Agric. Biol. Chem.* **1969**, *33*, 1301−1308.

(31) Gatehouse, J. A.; Croy, R. R. D.; Boulter, D. The synthesis and structure of pea storage proteins. *CRC Crit. Rev. Plant Sci.* **1984**, *1*, 287−314.

(32) Casey, R.; Domoney, C.; Smith, A. M. Biochemistry and molecular biology of seed products. In *Peas: Genetics, Molecular Biology and Biotechnology*; Casey, R., Davies, D. R., Eds.; CAB International: Wallingford, United Kingdom, 1993; pp 121−164.
(33) Yaklich, R.; Helm, R.; Herman, E. Analysis of the distribution of the major soybean allergen in a core collection of *Glycine max* accessions. *Crop Sci*. **1999**, *39*, 1444−1447.
(34) Kwanyuen, P.; Pantalone, V. R.; Burton, J. W.; Wilson, R. F. A new approach to genetic alteration of soybean protein composition and quality. *JAOCS* **1997**, *74*, 983−987.
(35) Sebolt, A. M.; Shoemaker, R. C.; Diers, B. W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentrations in soybean. *Crop. Sci*. **2000**, *40*, 1438−1444.
(36) Tezuka, M.; Taira, H.; Igarashi, Y.; Yagasaki, K.; Ono, T. Properties of tofu's and soymilks prepared from soybeans having different subunits of glycinin. *J. Agric. Food. Chem.* **2000**, *48*, 1111−1117.
(37) Renkema, J. M. S.; Knabben, J. H. M.; Vliet, T. Gel formation by beta-conglycinin and glycinin and their mixtures. *Food Hydrocolloids* **2001**, *15*, 407−414.
(38) Hughes, S. A.; Murphy, P. A. Varietal influence on the quantity of glycinin in soybeans. *J. Agric. Food. Chem.* **1983**, *31*, 376−379.

Received for review November 26, 2005. Revised manuscript received February 18, 2006. Accepted March 7, 2006. Funding for this research was provided by ARS project 1275-21000-156-00D. Mention of trade name, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture or imply its approval to the exclusion of other products or vendors that also may be suitable.

JF052954K