JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Identification of Several *gy4* Nulls from the USDA Soybean Germplasm Collection Provides New Genetic Resources for the Development of High-Quality Tofu Cultivars

WON-SEOK KIM,[†] HEO JAE HO,[†] RANDALL L. NELSON,[‡] AND HARI B. KRISHNAN*,[†,§]

Plant Science Division, University of Missouri, Columbia, Missouri 65211; Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, USDA Agricultural Research Service, Department of Crop Sciences, University of Illinois, Urbana, Illinois 61801; and Plant Genetics Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Columbia, Missouri 65211

Tofu, a cheese-like food made by curdling soy milk, is a major dietary staple of Asian countries. Consumption of tofu and other soy products is steadily increasing in North America due to its well-known health benefits. Soybean A_5, A_4, and B_3 peptide null lines 'Enrei' and 'Raiden' are commonly utilized in breeding programs to develop high-quality tofu cultivars. To expand the genetic diversity it is desirable to identify and utilize other A_5, A_4, and B_3 null genotypes in the development of improved tofu cultivars that are adapted to North American conditions. In this study were screened diverse soybean accessions from the USDA Soybean Germplasm Collection to identify *Gy4* mutants, the locus that controls A_5, A_4, and B_3 peptide production. Analysis of total seed proteins from 485 soybean lines by SDS-PAGE enabled the identification of 38 accessions that lacked the A_5, A_4, and B_3 peptides. These accessions showed marked differences in seed size and seed coat color and represented different maturity groups ranging from 0 to IX. To ascertain the molecular basis for the lack of A_5, A_4, and B_3 peptides in the newly identified *Gy4* mutants, the nucleotide sequence of a portion of the *Gy4* gene was determined from eight soybean accessions representing different maturity groups. These eight *Gy4* mutants revealed a single point mutation that changed the translation initiation codon ATG to ATA, resulting in the A_5, A_4, and B_3 null phenotype. The newly identified *Gy4* mutants from this study will enable plant breeders to expand the genetic diversity of North American food-quality soybeans and also aid in the development of hypoallergenic soybeans.

KEYWORDS: ***Glycine max*; glycinin; point mutation; SDS-PAGE; storage protein; tofu**

INTRODUCTION

Soybeans are an excellent source of high-quality protein. They serve as the important source of protein in China, Korea, Japan, and other rice-consuming Asian populations. Tofu, which is made from soy milk, is traditionally consumed in these countries because of its high protein and low saturated fat content (*1*). Although the consumption of tofu in Asian countries has been practiced for centuries, it has only recently gained popularity in North America and Europe. Health-conscious consumers have started using tofu as a meat alternative in their diet because it reduces the risk of cardiovascular disease, and the worldwide demand for tofu and other food-quality soybeans has increased. As a result, soybean cultivars with both desirable seed composition traits and superior agronomic characteristics are needed.

Tofu quality is profoundly influenced by soybean seed protein composition (*2, 3*). The two major seed storage proteins of soybean are glycinin (11S) and β-conglycinin (7S). Together, these two groups of proteins account for about 70% of soybean total seed proteins (*4, 5*). Both the glycinins and β-conglycinins are encoded by multigene families. Six functional genes encoding glycinin (*Gy1*, *Gy2*, *Gy3*, *Gy4*, *Gy5*, and *Gy7*) have been reported (*6, 7*). Each glycinin is synthesized as a precursor protein, which is post-translationally cleaved into acidic and basic polypeptides that are linked by a disulfide bond (*8*). In contrast to other glycinins, the glycinin produced by *Gy4* undergoes cleavage at two sites resulting in two acidic (A_4 and A_5) peptides and one basic (B_3) peptide (*9*) with theoretical

* Corresponding author [telephone (573) 882-8151; fax: (573) 884-7850; e-mail KrishnanH@missouri.edu].
[†] Plant Science Division, University of Missouri.
[‡] Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, USDA Agricultural Research Service.
[§] Plant Genetics Research Unit, Agricultural Research Service, U.S. Department of Agriculture.

10.1021/jf801831w CCC: $40.75
Published on Web 11/08/2008

molecular masses of 11 kDa (A_5), 29 kDa (A_4), and 21 kDa (B_3), respectively (*9*).

Superior tofu quality has been associated with the absence of A_4 glycinin peptide (*10*). Consequently, soybean breeders have utilized A_4 null lines such as 'Enrei' and 'Raiden' as donor parents in their breeding programs for the development of commercial tofu cultivars. Both cultivars contain a point mutation in the *Gy4* gene that alters the translation initiation codon from ATG to ATA (*11, 12*). Repeated use of few genotypes in the breeding programs has resulted in a narrow genetic base, which could hamper effective improvement of agronomic and seed traits in the future. Utilizing genetically diverse parents will broaden the genetic base of tofu cultivars. In this study, we have identified several diverse A_5, A_4, and B_3 nulls by screening soybean accessions from the USDA Soybean Germplasm Collection. The identification of these nulls provides an opportunity to develop food-quality soybeans with a broad genetic base.

MATERIALS AND METHODS

Plant Materials. Four sets of *Glycine* accessions were selected for this research. The first set was soybean germplasm selected for maximum genetic diversity in a small number of accessions. China, Japan, and Korea are the countries of most ancient soybean cultivation, and research with DNA markers has demonstrated that soybean germplasm from China is more diverse than germplasm from either Japan or Korea (*13*). Primitive germplasms for the same province in China are more likely to be closely related than accessions from different provinces (*14, 15*). On the basis of these assumptions 2 to 4 *G. max* accessions were selected from each of the 24 provinces of China represented in the USDA Soybean Germplasm Collection. Within each province accessions were selected to represent extremes in maturity group and morphological characteristics. A smaller number of accessions from Japan, North Korea, and South Korea based on diversity of origin, maturity group, and morphological diversity were also included. There were 96 accessions in this group. The second set of accessions consisted of 50 *Glycine soja* accessions selected to represent the areas of geographical diversity of China, Korea, Russia, and Japan where this species is native. In the third set, 176 accessions with high protein concentrations (generally >50%) were selected, and a corresponding set of 127 accessions with low protein concentrations with similar origins (*16*). Finally, 36 lines representing the major ancestral lines of current U.S. cultivars were included (*17*). All seeds were obtained from the USDA Soybean Germplasm Collection.

Reagents. Chemicals for electrophoresis (acrylamide, bis-acrylamide, SDS, TEMED, ammonium persulfate, thiourea, dithiothreitol, CHAPS, and IPG strips) were purchased from GE Healthcare (Piscataway, NJ). Urea, ampholytes, and SDS-PAGE molecular weight markers were purchased from Bio-Rad Inc. (Hercules, CA). Tris, 2-mercaptoethanol, and glycerol were purchased from Sigma (St. Louis, MO).

Extraction of Soybean Seed Proteins. Seeds from 485 soybean lines were individually ground to a fine powder using a mortar and pestle. A 10 mg aliquot from each of the soybean accessions was extracted with 1 mL of sodium dodecyl sulfate (SDS) sample buffer [60 mM Tris-HCl, pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), and 5% 2-mercaptoethanol] and then boiled for 5 min. Following extraction, the lysate was clarified by centrifugation (15800*g*, 5 min) and the resulting supernatant transferred to Eppendorf tubes. Aliquots (10 μL) of the supernatant were utilized for SDS-polyacrylamide gel electrophoresis (PAGE) analysis (*18*). Soybean lines tentatively identified as lacking the A_5B_3 peptides during the initial screen were subjected to two additional independent extractions and verified by SDS-PAGE.

SDS-PAGE. Soybean total seed proteins were resolved by SDS-PAGE using a Hoeffer SE 260 minigel apparatus according to the manufacturer's recommendations (GE Healthcare). Resolved proteins were visualized by Coomassie Brilliant Blue R-250 staining.

2-D Gel Electrophoresis. One hundred milligrams of replicated seed powder was homogenized with 2 mL of a solution containing 10% (w/v) TCA in acetone with 0.07% (v/v) 2-ME as described (*19*). Seed proteins were 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 was performed using 13 cm, IPG strips (pH 3−10) in the IPGphor System (GE Healthcare). A 100 μg sample of protein was loaded onto the strips by active rehydration (50 V) in 250 μL of buffer containing 8 M urea, 2% CHAPS (w/v), 0.5% ampholytes, 1% dithiothreitol (DTT), and 0.002% bromphenol blue, and 2-D gel electrophoresis was conducted as described earlier (*19*). 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 stained for 1 h in 34% methanol, 17% ammonium sulfate, 3% phosphoric acid, and 0.066% Coomassie Brilliant Blue G-250 (w/v). Electronic images of Coomassie stained gels images were acquired using a UMAX PowerLook 2100XL scanner through Adobe Photoshop v7.0. Images were scanned using 100 μm resolution at 600 dpi and analyzed using the Phoretix 2D Advanced v6.01 software (Nonlinear Dynamics Ltd., Newcastle upon Tyne, U.K.).

Molecular Techniques. Total genomic DNA was isolated from leaf tissues of selected *Gy4* null genotypes using the CTAB method (*20*). Utilizing genomic DNA as a template, a segment of the *Gy4* gene was amplified by Polymerase Chain Reaction (PCR). The primers (forward 5′-TTCACCAACTCCTTCAAACTTAATT-3′; reverse 5′-GGTGAG-TAAGATGGCAAGTGGAG-3′) were designed to amplify 307 bp of the N-terminal region of the *Gy4* gene including the 47 bp untranslated 5′ sequences (*11*). PCR reactions were carried out in a final volume of 50 μL containing 10 mM Tris-HCl, pH 8.0, 50 mM KCl, 1 mM $MgCl_2$, 200 μM each of dNTP, 1 μM of each primer, 100 ng of genomic DNA, and 2 units of ExTaq DNA polymerase (Pan Vera Corp., Madison, WI). The PCR products were isolated and purified from a 1.0% agarose gel using Ultrafree-DA columns (Millipore Corp., Bedford, MA). The gel-purified PCR products were individually cloned into pGEM-T Easy vector (Promega, Madison, WI). Plasmid DNA was prepared using WizardPlus SV Minipreps DNA Purification System (Promega). DNA sequences were determined at the University of Missouri DNA Core Facility using SP6 and T7 primers.

RESULTS

Identification of *Gy4* Mutants by SDS-PAGE Analysis. Total seed proteins were extracted from 485 soybean lines from the USDA Soybean Germplasm Collection and examined for their protein profile by SDS-PAGE. **Figure 1** represents a typical gel picture of total soybean seed protein profile. All soybean lines revealed significant accumulation of the α′ (72 kDa), α (70 kDa), and β (52 kDa) subunits of β-conglycinin. The relative concentration of each of these subunits among different soybean lines was variable, with the highest variation observed in the accumulation of the β-subunit of β-conglycinin (**Figure 1**). Electrophoretic separation of soybean seed proteins utilizing 12% SDS-PAGE gels enabled the clear separation the B_3 peptide from B_{1a}, B_{1b}, B_2, and B_4 glycinin peptides and the easy identification of A_5 glycinin peptide (**Figure 1**). A detailed analysis of seed protein profile from 485 soybean lines, for which the countries of origin are shown in **Table 1**, led to the identification of 38 soybean genotypes that failed to accumulate the A_5 and B_3 peptides. The 38 soybean lines lacking the A_5 and B_3 peptides ranged from maturity groups 0 to IX (**Table 2**) and came from China, Korea, Japan, North Korea, South Korea, Nepal, and Vietnam, where >90% of the accessions tested originated (**Table 3**).

Separation of soybean seed proteins by 12% SDS-PAGE gels enabled us to identify soybean lines that failed to accumulate the A_5 and B_3 peptides. However, under these conditions all of the glycinin acidic subunits migrated as a single abundant polypeptide (**Figure 1**). To overcome this problem, we resolved soybean seed proteins in 13.5% SDS-PAGE gels. Increasing

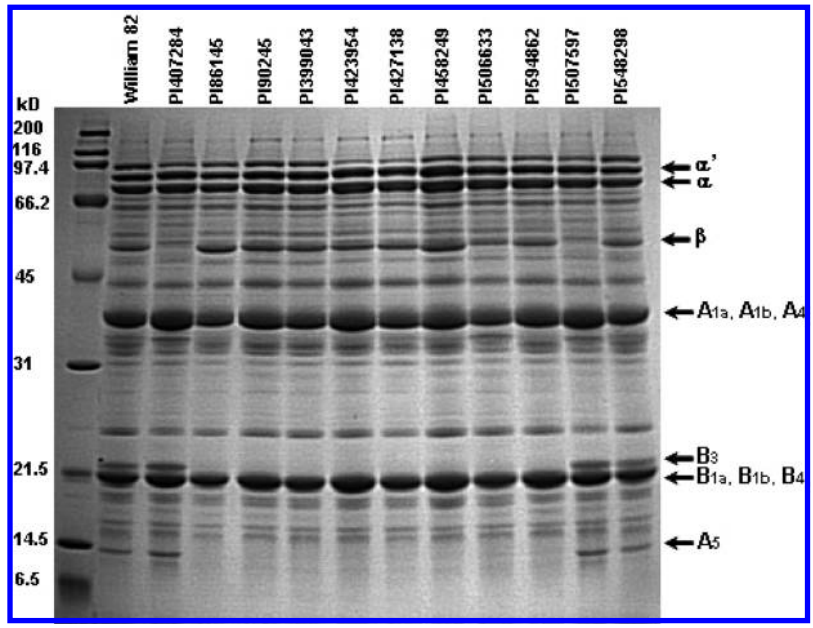

Figure 1. SDS-PAGE analysis of total seed proteins from select *Gy4* mutant lines. Seed proteins were extracted in SDS sample buffer and resolved via 12% SDS-PAGE. Arrows indicate the three subunits of β-conglycinin (α', α, and β) and the B_3 and A_5 glycinin peptides. Accession numbers of different soybean lines are indicated at the top of the figure, and the sizes of protein markers in kilodaltons are shown at the left of the figure.

Table 1. Origin of Soybean Accessions Used in This Study

country	no. of accessions
China	245
Germany	1
Indonesia	21
Japan	68
North Korea	20
South Korea	77
Nepal	8
Philippines	1
Russian Federation	10
Sweden	4
Taiwan	2
unknown	4
United States	3
Vietnam	21
total	485

Table 2. Frequency of *Gy4* Mutants Among Different Maturity Groups

maturity group screened	no. of accessions identified	no. of *Gy4*mutants
0	26	6
00	12	0
000	10	0
I	34	3
II	29	0
III	44	6
IV	63	7
V	62	5
VI	82	3
VII	51	2
VIII	40	5
IX	27	1
X	5	0
total	485	38

the acrylamide concentration of the resolving gels resulted in the slower migration of the A_4 glycinin peptide when compared with the other glycinin acidic subunits. By employing 13.5% SDS-PAGE gels we confirmed that all 38 soybean lines identified as missing the A_5 and B_3 peptides also failed to accumulate the A_4 glycinin peptide (data not shown). Further confirmation for the lack of the A_4 glycinin peptide in some of these newly identified *Gy4* mutants was obtained by 2-D gel analysis. **Figure 2** shows the 2-D protein profile of BARC-6, a high-protein soybean line and PI 427138, a *Gy4* mutant. The use of pH 3−10 linear IPG strips enabled the resolution of both the β-conglycinin and the glycinin subunits (**Figure 2**). Earlier it was shown that the A_4 peptide, which has a p*I* value of 4.8, migrates as a distinct spot separate from other glycinin acidic subunits (*21*). The A_4 peptide was clearly seen in protein extracts from BARC-6 but was missing from PI 427138, one of the 38 *Gy4* mutant lines identified in this study (**Figure 2**). The protein profiles of four other A_5 B_3 nulls were also examined, and all lack the A_4 peptide (data not shown).

Sequence Analysis Reveals the Same Point Mutation in the *Gy4* Gene in Diverse Soybean Accessions. This study has identified 38 soybean accessions that failed to accumulate the $A_5A_4B_3$ peptides. We amplified a portion of the *gy4* gene from eight soybean lines, from different maturity groups, that failed to accumulate the $A_5A_4B_3$ peptides. For comparison we also amplified the *gy4* gene fragment from the soybean cultivar Williams 82. A 307 bp DNA fragment was successfully amplified from all of these soybean lines (**Figure 3A**). DNA sequences of the PCR products of all eight *gy4* mutants were identical. Compared to the Williams 82 sequence, a single point mutation that changed deoxyguanosine (G) to deoxyadenine (A) at the initiation codon of the *gy4* gene was observed in all eight *Gy4* mutants (**Figure 3B**). Thus, the point mutation identified in these newly identified *Gy4* soybean lines was identical to the mutation in Raiden and Enrei (*11, 12*).

DISCUSSION

Glycinin, the most abundant soybean seed storage protein, is made up of five major subunits that are classified into groups I and II on the basis of their size and sequence homology (*6*). Group I comprises three subunits, $A_{1a}B_2$, $A_{1b}B_{1b}$, and A_2B_{1a}, whereas group II includes two subunits, A_3B_4 and $A_5A_4B_3$. SDS-PAGE analysis provides excellent resolution of most soybean seed proteins; however, it is not adequate to separate the

Table 3. *Gy4* Mutants and Their Relevant Characteristics[a]

PI no.	name	province	country	maturity group	seed coat color	100 seed wt (g)	protein (%)	oil (%)
PI603487A	Diao ya dou	Shandong	China	IV	green	22.5	52.9	16.0
PI594739B	Jin zhong shan da huang dou	Guangxi	China	VIII	yellow	9.3	52.5	14.5
PI471939		unknown	Nepal	IX	yellow	12.9	52.0	12.0
PI423949	Saikai 20	Kumamoto	Japan	I	yellow	13.3	51.4	14.4
PI594643	Ba yue huang No. 4	Guizhou	China	V	yellow	8.5	51.4	13.7
PI588005C	Da bai mao	Sichuan	China	V	yellow	12.4	50.8	14.7
PI594585	An hua chi huang dou	Hunan	China	VII	yellow	9.6	50.6	13.9
PI423966	Kumaji 2	Kumamoto	Japan	VIII	yellow	14.4	50.4	17.7
PI427141	Seuhae No. 20	Kyonggi	South Korea	I	yellow	13.6	49.5	14.1
PI424148	Shirome	Kyongsang Puk	South Korea	0	yellow	8.9	48.3	15.0
PI427138	Choseng No. 1	Kyonggi	South Korea	0	yellow	10.5	48.0	14.4
PI423954	Shirome	Kumamoto	Japan	0	yellow	11.8	47.3	15.6
PI399074		Chungchong Nam	South Korea	0	yellow	11.1	47.1	15.6
PI594615	Liu yue zao	Guizhou	China	IV	yellow	9.3	47.1	14.3
PI567395	Lai wa dou	Shaanxi	China	IV	grayish green	9.6	46.8	15.1
PI407765		Guangdong	China	V	brown	25.8	46.4	17.7
PI424242		Kyonggi	South Korea	0	yellow	10.1	46.3	16.4
PI399043		Cheju	South Korea	III	yellow	23.1	44.0	18.3
PI407849		Cholla Puk	South Korea	III	yellow	22.6	43.8	18.1
PI603521	Huang dou	Shaanxi	China	VIII	green	13.8	43.7	17.5
PI605781B		Cao bang	Vietnam	V	yellow	9.2	43.5	19.4
PI416886	Ginsui zairai	Kyushu	Japan	VIII	black	11.1	42.7	14.7
PI 88788		Liaoning	China	III	black	9.4	42.4	17.5
PI567683B	Zheng zhou niu yao qi	Henan	China	VI	yellow	11.8	42.4	18.5
PI458512	Qi huang No. 10	Shandong	China	III	yellow	14.5	41.9	17.4
PI506633	Chousen oiyarukon	Kyushu	Japan	VIII	yellow	9.4	41.8	17.6
PI 59845	Sohgetsu	Akita	Japan	VI	yellow	14.3	41.7	18.1
PI 86145	Monbetsu nagaba daizu	Hokkaido	Japan	III	brown	18.2	41.7	19.1
PI506885	Kinako mame	Tohoku	Japan	VI	green	5.5	41.1	16.3
PI458061A		Kangwon	South Korea	III	yellow	17.1	41.0	19.5
PI507573	Zairai misonimame	Kanto	Japan	IV	yellow	17.8	41.0	17.9
PI 90245	Neihen	Pyongan Nam	North Korea	IV	yellow	14.6	40.9	19.0
PI458249		Cholla Nam	South Korea	IV	green	22.3	40.8	17.5
PI594862	Cha huang dou	Yunnan	China	VII	reddish brown	11.9	40.7	17.0
PI603375	Qian guo jian ye he jia dou	Jilin	China	I	yellow	13.5	40.7	20.2
PI476879	Bach hoa thao moc chau	unknown	Vietnam	V	yellow	7.5	40.6	18.7
PI408295B		Kyongsang Nam	South Korea	IV	black	23.2	40.5	17.8
PI467323A	Jiu nong 13	Jilin	China	0	yellow	16.2	40.5	18.9

[a] Data compiled from USDA-ARS Germplasm Resources Information Network (http://www.ars-grin.gov/npgs/acc/acc_queries.html).

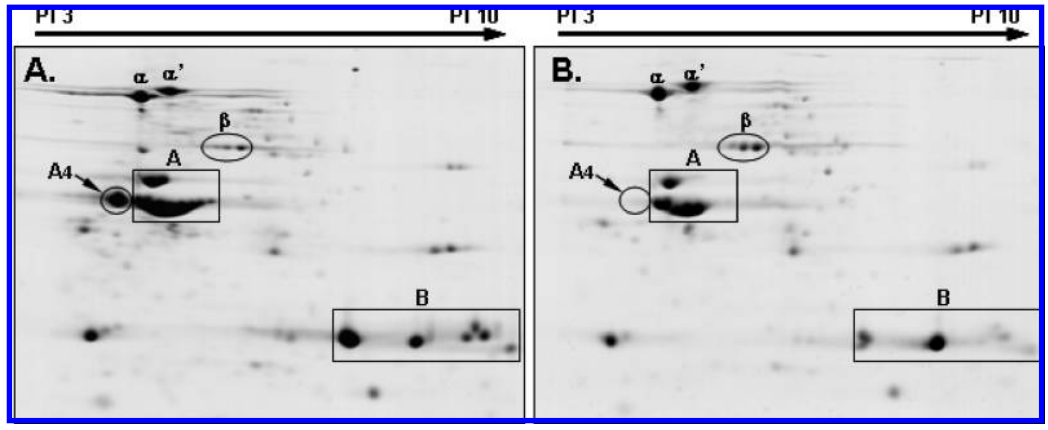

Figure 2. Comparison of the storage proteins of BARC-6 (**A**) and PI 427138 (**B**) soybean by 2D SDS-PAGE. Proteins were first separated by isoelectric focusing on IPG strips (13 cm, pH 3−10) and then by SDS-PAGE on a 13.5% gel. The protein spots corresponding to α′, α, and β subunits of β-conglycinin are indicated. The acidic (A) and basic (B) subunits of glycinin are enclosed in a rectangle, and the A_4 peptide is indicated with a circle.

different glycinin subunits. Fontes and co-workers (*22*) reported using gradient gels and incorporating 6 M urea resulted in easy identification of A_3 and A_4 glycinin peptides. Recently, this protocol was further modified by Poysa and co-workers and enabled reproducible identification of soybean genotypes deficient in glycinin subunits A_4 and A_5 (*3*). However, these modified protocols are cumbersome and not conducive to high-throughput screening. Because the presence or absence of A_5 glycinin peptide can be easily detected by regular SDS-PAGE without any modifications, we focused our attention on identification of soybean lines that lacked the A_5 peptide. Thus, this procedure enabled us to easily and rapidly identify 38 *Gy4* mutants from the USDA Soybean Germplasm Collection.

Previous studies have established that the accumulation of the β-subunit of β-conglycinin is influenced by the availability of nitrogen and sulfur. An overabundance of nitrogen favors

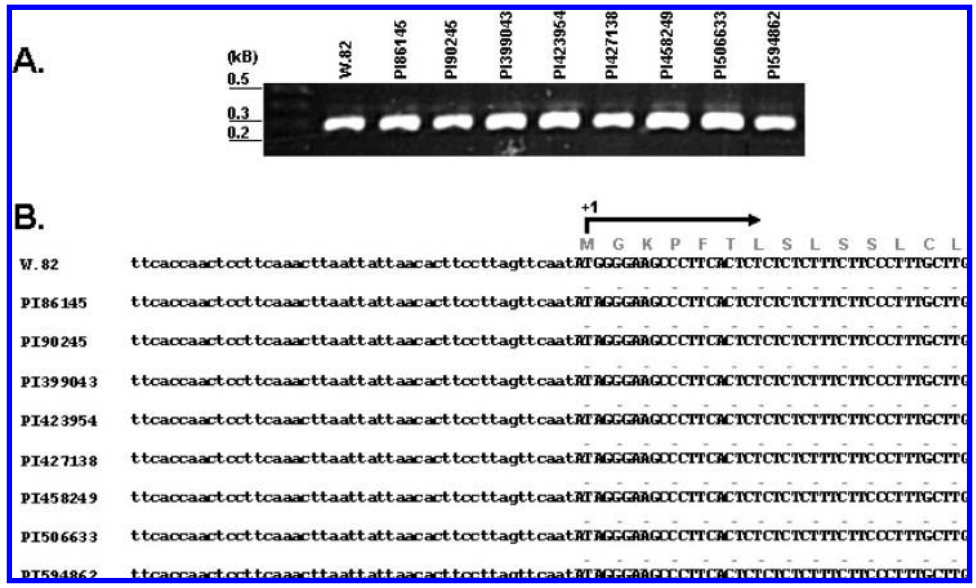

Figure 3. PCR amplification of a portion of *gy4* gene (**A**) and the nucleotide and deduced amino acid sequence comparison of a portion of the *gy4* gene from diverse soybean lines (**B**). (**A**) Genomic DNA isolated from Williams 82 and select *Gy4* mutant lines were used as a template to amplify a segment of the *gy4* gene by PCR. Accession numbers of different soybean lines are indicated at the top of the figure, and the sizes of DNA markers in kilo base pairs (kB) are shown at the left of the figure. (**B**) Nucleotide sequence of the PCR amplified DNA fragment from eight diverse soybean lines showing identical point mutation. The protein translation start position is marked +1. Note the initiation codon "ATG" in Williams 82 is mutated to "ATA" in all eight *Gy4* mutants.

the accumulation of the β-subunit of β-conglycinin, whereas an overabundance of sulfur leads to increased accumulation of the glycinin (*23*−*25*). Thus, the differences observed in the accumulation of the β-subunit of β-conglycinin may be related to the nitrogen status of the soil in which the different soybean lines were grown. Additionally, differences in the interaction between soybean lines inoculated with *Bradyrhizobium* spp. can also influence the accumulation of the β-subunit of β-conglycinin (*26, 27*). Successful interaction between *Bradyrhizobium* spp. and soybean during early growth leads to the formation of root nodules. Rhizobia within the root nodules convert atmospheric nitrogen into ammonia, which is then utilized by the soybean for the synthesis of proteins during seed development. Non-nodulating soybeans accumulate low amounts of the β-subunit of β-conglycinin in contrast to soybean cultivars that form nitrogen-fixing root nodules (*26, 27*). The variation in the accumulation of the β-subunit of β-conglycinin observed in our study could be attributed to differences in the ability of these soybean lines to form nitrogen-fixing nodules.

Soybean cultivar 'Raiden' does not accumulate A_4, A_5, and B_3 polypeptides (*28, 29*). Genetic characterization revealed the absence of these glycinin subunits results from a single recessive allele, *gy4* (*11, 28*). DNA sequence analysis of the *gy4* gene from Raiden revealed a point mutation in the initiation codon. As a consequence, transcripts of this gene do not accumulate, leading to the absence of the A_5, A_4, and B_3 peptides in the soybean cultivar 'Raiden' (*11*). Recently the molecular basis for the absence of A_4 peptide in glycinin of the soybean cultivar 'Enrei' was determined to be caused by the same point mutation that was first observed (*12*). In this study we have identified another 38 *Gy4* mutant soybean lines by SDS-PAGE. Partial sequence analysis of the *gy4* gene from eight of these newly identified mutant soybean lines established the presence of the same mutation as the two previously identified *Gy4* mutants. These mutant lines represent between 5 and 13% of the accessions analyzed from China, Japan, North Korea, South Korea, Nepal, and Vietnam. The only other countries with more than five accessions characterized were Indonesia (*21*) and Russia (*10*), and no *Gy4* mutants were found in lines from these countries. The frequency of *Gy4* mutants in maturity group 0 was higher than that in late maturity groups. No *Gy4* mutants were identified in maturity groups II and X. However, the lack of *Gy4* mutants in these maturity groups may be related to the limited number of accessions tested in this study. No wild soybean (*G. soja*) accession carried the *Gy4* mutant. Nine of 96 lines selected for maximum genetic diversity from China, Japan, and Korea had the *gy4* allele. This is similar to the percentage noted above for all seven countries with accessions carrying the mutant allele. Nearly 13% of the accessions selected for lower protein concentration and over 6% of the accessions selected for higher protein concentration carried the *gy4,* so there is no apparent association with total protein. PI 88788 from Liaoning, China, was the only ancestral line with the *gy4* allele. This line is the most common source of soybean cyst nematode (SCN) resistance currently deployed in the United States (http://web.extension.uiuc.edu/livingston/reports/i120/index.html). The cultivar 'Fayette' with the pedigree Williams (2) × PI 88788 is the intermediate source of this resistance for most current cultivars (*30*). 'Fayette' was tested and found not to have the *gy4* allele.

The remarkable conservation of the same point mutation among diverse soybean lines raises interesting questions about the origin and spread of this mutation. No wild soybean accessions were found with the *gy4* allele. This was a small sample of wild soybean accessions, but given the relative high percentage of soybean accessions with the mutant allele, if it does exist in *G. soja*, the gene frequency is much lower. Because the *gy4* allele was found in soybean lines of diverse origins ranging from Nepal to Vietnam to North Korea, it seems likely that this mutation may have occurred in multiple, independent events. If that is true, it seems unlikely that the same base in the translation initiation codon would be changed in all accessions simply by random chance. Another possibility is that some of these mutants originated from a common genetic stock. These materials could have been exchanged among farmers from North and South Korea, China, and Vietnam due to their

geographic proximity. A definitive answer about the origin and preservation of this mutation will depend on future research.

Japan and a few other Asian counties import significant amounts of food-quality soybeans from United States for the production tofu and other traditional food products. Most of the soybean cultivars that are grown in the Untied States are not particularly suited for tofu production. As a consequence, significant breeding efforts have gone into the development of food-grade soybean cultivars. These cultivars have specific characteristics that are targeted to the needs of end users. High protein and large seed size are the two main characteristics that are preferred for high-quality tofu production. It is well established that the yield and quality of tofu are greatly influenced by the primary seed proteins glycinin and β-conglycinin (*1, 31*). Gels made from glycinin are harder than gels from β-conglycinin (*2, 3*). Furthermore, the different subunits of glycinin and β-conglycinin contribute differently to gel formation (*32*−*34*). For example, it is known that tofu prepared from soybean cultivars lacking the A_5 peptide is harder (*35*). Studies using several spontaneous and induced mutants with altered seed protein components have shown their profound effect on tofu quality (*3, 32, 35*). On the basis of these studies, it is apparent that soybean cultivars that lack both the 7S α' and the glycinin $A_5A_4B_3$ peptides will yield high-quality tofu (*3*).

Currently, only a few *Gy4* mutants are utilized in the breeding programs for developing cultivars that are suited for tofu production. There are over 17000 *Glycine max* accessions in the USDA Soybean Germplasm Collection, and identification of *Gy4* mutants from this germplasm will provide an opportunity for expanding the genetic diversity of North American tofu cultivars. This study identified 38 *Gy4* mutants by screening a diverse sample of soybean germplasm. Most of the newly identified *Gy4* mutants contain relatively high concentrations of protein (**Table 1**), a desirable trait for tofu production.

The identification of several *Gy4* mutants from the USDA Soybean Germplasm Collection could also be exploited to develop hypoallergenic soybeans. A number of soybean proteins have been identified as food allergens (*36, 37*). Certain components of both the β-conglycinin and the glycinin are known to elicit food hypersensitivity reactions (*36, 37*). Additionally, several other IgE-binding soybean proteins including Gly m Bd 28K, Gly m Bd 30 K, and Gly m BD 60 K have been identified and characterized (*38*−*40*). In the United States, an allergic reaction to cow's milk is one of the main food allergies in infants (*41*). To avoid immunological reactions in infants, soy-based formulas are often used as a substitute for cow's milk. Recent studies, however, have demonstrated the cross-reactivity of A_5B_3 glycinin peptides with cow's milk specific antibodies (*42*). On the basis of this observation, it was concluded that the A_5B_3 glycinin peptides could be involved in allergic reactions in infants fed soy-based formulas (*42*). Thus, the utilization of *gy4* null soybean cultivars for soy milk production could minimize the problem of milk intolerance in infants.

ACKNOWLEDGMENT

We thank Sungchan Jang for technical assistance and Nathan Oehrle for critically reading the manuscript.

LITERATURE CITED

(1) Liu, K. *Soybeans: Chemistry, Technology, and Utilization*; Chapman and Hall: New York, 1997.

(2) Poysa, V.; Woodrow, L. Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. *Food Res. Int.* **2002**, *35*, 337–345.

(3) Poysa, V.; Woodrow, L.; Yu, Y. Effect of soy protein subunit composition on tofu quality. *Food Res. Int.* **2006**, *39*, 309–317.

(4) Nielsen, N. C. Soybean seed composition In *Soybean: Genetics, Molecular Biology and Biotechnology*; Verma, D. P. S., Shoemaker, R. C., Eds.; CAB:Tucson, AZ, 1996; pp127−163.

(5) Krishnan, H. B. Biochemistry and molecular biology of soybean seed storage proteins. *J. New Seeds* **2000**, *2*, 1–25.

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

(7) Beilinson, V.; Chen, Z.; Shoemaker, R. C.; Fischer, R. L.; Goldberg, R. B.; Nielsen, N. C. Genomic organization of glycinin genes in soybean. *Theor. Appl. Genet.* **2002**, *104*, 1132–1140.

(8) Staswick, P. E.; Hermodson, M. A.; Nielsen, N. C. Identification of the cystnes which link the acidic and basic components of the glycinin subunits. *J. Biol. Chem.* **1984**, *259*, 13431–13435.

(9) Momma, T.; Negoro, T.; Hirano, H.; Matsumoto, A.; Udaka, K.; Fukazawa, C. Glycinin $A_5A_4B_3$ mRNA: cDNA cloning and nucleotide sequencing of a splitting storage protein subunit of soybean. *Eur. J. Biochem.* **1985**, *149*, 491–496.

(10) Chang, K. C.; Hou, H. J. Science and technology of tofu making In *Handbook of Vegetable Preservation and Processing*; Hui, Y. H., Ghazala, S., Graham, D. M., Murrell, K. D., Nip, W., Eds.; Dekker: New York, 2003; pp 443−478.

(11) Scallon, B. J.; Dickinson, C. D.; Nielsen, N. C. Characterization of a null-allele for the Gy_4 glycinin gene from soybean. *Mol. Gen. Genet.* **1987**, *208*, 107–113.

(12) Yu, K.; Poysa, V.; Haffner, M.; Zhang, B.; Woodrow, L. Absence of the A4 peptide in the glycinin subunit of soybean cultivar Enrei is caused by a point mutation in the Gy4 gene. *Genet. Mol. Biol.* **2005**, *28*, 440–443.

(13) Li, Z.; Nelson, R. L. Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci.* **2001**, *41*, 1337–1347.

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

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

(16) GRIN (Germplasm Resources Information Network); available at www.ars-grin.gov/npgs/index.html (verified Aug 15, 2006).

(17) 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*, 1143–1147.

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

(19) 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.

(20) 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.

(21) Krishnan, H. B.; Natarajan, S. S.; Mahmoud, A. A.; Bennett, J. O.; Krishnan, A. H.; Prasad, B. N. Assessment of indigenous Nepalese soybean as a potential germplasm resource for improvement of protein in North American cultivars. *J. Agric. Food Chem.* **2006**, *54*, 5489–5497.

(22) Fontes, E.; Moreira, M.; Davies, C.; Nielsen, N. C. Urea elicted changes in relative electrophoretic mobility of certain glycinin and β-conglycinin subunits. *Plant Physiol.* **1984**, *76*, 840–842.

(23) Sexton, P. J.; Naeve, S. L.; Paek, N. C.; Shibles, R. Sulfur availability, cotyledon nitrogen:sulfur ratio, and relative abundance of seed storage proteins of soybean. *Crop Sci.* **1998**, *38*, 983–986.

(24) Gayler, K. R.; Sykes, G. E. Effects of nutritional stress on the storage proteins of soybeans. *Plant Physiol.* **1985**, *78*, 582–585.

(25) Paek, N. C.; Imsande, J.; Shoemaker, R. C.; Shibles, R. Nutritional control of soybean seed storage protein. *Crop Sci.* **1997**, *37*, 498–503.
(26) Ohtake, N.; Suzuki, M.; Takahashi, Y.; Fujiwara, T.; Chino, M.; Ikarashi, T.; Ohyama, T. Differential expression of β-conglycinin genes in nodulated and nonnodulated isolines of soybean. *Physiol. Plant.* **1996**, *96*, 101–110.
(27) Krishnan, H. B.; Jiang, G.; Krishnan, A. H.; Wiebold, W. J. Seed storage protein composition of nonnodulating soybean (*Glycine max* (L.) Merr.) and its influence on protein quality. *Plant Sci.* **2000**, *157*, 191–199.
(28) Kitamura, K.; Davies, C. S.; Nielsen, N. C. Inheritance of alleles for Cgy1 and Gy4 storage protein genes in soybean. *Theor. Appl. Genet.* **1984**, *68*, 253–257.
(29) Staswick, P. E.; Nielsen, N. C. Characterization of a soybean cultivar lacking certain glycinin subunits. *Arch. Biochem. Biophys.* **1983**, *223*, 1–8.
(30) Bernard, R. L.; Cremeens, C. R. Registration of 'Williams 82' soybean. *Crop Sci.* **1988**, *28*, 1027–1028.
(31) Murphy, P.; Chen, H.; Hauck, C.; Wilson, J. Soybean protein composition and tofu quality. *Food Technol.* **1997**, *51*, 86–88.
(32) Yagasaki, K.; Kousaka, F.; Kitamura, K. Potential improvement of soymilk gelation properties using soybean with modified protein subunit compositions. *Breed. Sci.* **2000**, *50*, 101–107.
(33) Mujoo, R.; Trinh, D. T.; Ng, P. Characterization of storage proteins in different soybean varieties and their relationship to tofu yield and texture. *Food Chem.* **2003**, *82*, 265–273.
(34) Mohamad Ramlan, B. M. S.; Maruyama, N.; Takahashi, K.; Yagasaki, K.; Higasa, T.; Matsumura, Y.; Utsumi, S. Gelling properties of soybean β-conglycinin having different subunit compositions. *Biosci., Biotechnol., Biochem.* **2004**, *68*, 1091–1096.
(35) Nishinari, K.; Kohyama, K.; Zhang, Y.; Kitamura, K.; Sugimoto, T.; Saio, K.; Kawamura, Y. Rheological study of the effect of A5 subunit on the gelation characteristics of soybean proteins. *Agric. Biol. Chem.* **1991**, *55*, 351–355.
(36) Ogawa, T.; Bando, N.; Tsuji, H.; Okajima, H.; Nishikawa, K.; Sasaoka, K. Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis. *J. Nutr. Sci. Vitaminol.* **1991**, *37*, 555–65.
(37) Burks, A. W., Jr.; Brooks, J. R.; Sampson, H. A. Allergenicity of major component proteins of soybean determined by enzyme-linked immunosorbent assay (ELISA) and immunoblotting in children with atopic dermatitis and positive soy challenges. *J. Allergy Clin. Immunol* **1988**, *81*, 1135–42.
(38) Tsuji, H.; Bando, N.; Hiemori, M.; Yamanishi, R.; Kimoto, M.; Nishikawa, K.; Ogawa, T. Purification of characterization of soybean allergen Gly m Bd 28K. *Biosci., Biotechnol., Biochem.* **1997**, *61*, 942–947.
(39) Ogawa, T.; Tsuji, H.; Bando, N.; Kitamura, K.; Zhu, Y.L.; Hirano, H.; Nishikawa, K. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kD oil-body-associated protein. *Biosci., Biotechnol., Biochem.* **1993**, *57*, 1030–1033.
(40) Xiang, P.; Haas, E. J.; Zeece, M. G.; Markwell, J.; Sarath, G. C-Terminal 23 kDa polypeptide of soybean Gly m Bd 28 K is a potential allergen. *Planta* **2004**, *220*, 56–63.
(41) Bock, S. A.; Lee, W. Y.; Remigio, L. K.; May, C. D. Studies of hypersensitivity reactions to foods in infants and children. *J. Allergy Clin. Immunol.* **1978**, *62*, 327–334.
(42) Rozenfeld, P.; Docena, G. H.; Anon, M. C.; Fossati, C. A. Detection and identification of a soy protein component that cross-reacts with caseins from cow's milk. *Clin. Exp. Immunol.* **2002**, *130*, 49–58.

Received for review June 14, 2008. Revised manuscript received August 22, 2008. Accepted August 29, 2008. 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 USDA implies no approval of the product to the exclusion of others that may be suitable.

JF801831W