RESEARCH

Accumulation of Leginsulin, a Hormone-Like Bioactive Peptide, is Drastically Higher in Asian than in North American Soybean Accessions

Won-Seok Kim, Sungchan Jang, and Hari B. Krishnan*

ABSTRACT

Leginsulin, a peptide made up of 37 amino acids, is homologous to pea (*Pisum sativum* L.) albumin (PA1b) and belongs to the cysteine-knot family. Even though the physiological function and three-dimensional structure of leginsulin have been explored, little is known about its expression, accumulation, and distribution among soybean [*Glycine max* (L.) Merr.] accessions. An antibody generated against leginsulin was used to screen a diverse array of soybean accessions from the USDA Soybean Germplasm Collection to identify soybean accessions that are enriched in leginsulin. Analysis of 50% isopropanol-soluble proteins from 485 soybean lines by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis enabled the identification of 75 accessions that significantly accumulated leginsulin. Remarkably, all these accessions with the exception of two had their origin from Japan, Korea, or China. Leginsulin was barely detected in most of the commercial North American soybean cultivars used in this study. Western blot analysis revealed low accumulation of leginsulin in the embryonic axis but not in the cotyledons of North American soybean cultivar Williams 82 while in the Chinese PI 458249 leginsulin was abundantly present in both types of tissue. Examination of the soybean cultivar Williams 82 genome sequence revealed the presence of two homologous leginsulin genes (Gm13 g26330 [leginsulin 1] and Gm13 g26340 [leginsulin 2]) on chromosome 13. We have cloned the two leginsulin genes from PI 458249 and found them to be highly similar to that of Williams 82. Northern blot analysis indicated that leginsulin messenger RNA (mRNA) was abundant in Williams 82 embryonic axis but not in the cotyledon. In contrast, leginsulin mRNA was abundantly present in PI 458249, both in the embryonic axis and the cotyledons.

W.-S. Kim and H.B. Krishnan, USDA-ARS, Plant Genetics Research Unit, 205 Curtis Hall, Columbia, MO 65211; S. Jang and H.B. Krishnan, Plant Science Division, Univ. of Missouri, 1-41 Agriculture Bldg., Columbia, MO 65211. Received 25 Aug. 2011. *Corresponding author (Hari.Krishnan@ars.usda.gov).

Abbreviations: MALDI-TOF, matrix-assisted laser desorption time-of-flight; mRNA, messenger RNA; PA1b, pea albumin; PCR, polymerase chain reaction; RT-PCR, reverse transcriptase polymerase chain reaction; SDS, sodium dedecyl sulfate; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; SEF 4, soybean embryo factor 4; TBST, Tris-buffered saline containing 0.05% Tween-20.

The major seed storage proteins of soybean [*Glycine max* (L.) Merr.] are the globulins (Nielsen, 1996; Krishnan, 2000). They are grouped into 7S and 11S based on their sedimentation coefficients. The 11S proteins are termed as glycinin and the 7S as β-conglycinin (Nielsen, 1996; Krishnan, 2000). Glycinins are encoded by multigene families that are classified into three groups based on their sequence homology (Nielsen et al., 1989; Beilinson et al., 2002; Li and Zhang, 2011). Group I glycinin includes Gy1, Gy2, and Gy3, group II comprises Gy4 and Gy5, and group III contain Gy6, Gy7, and Gy8. Gy6 and Gy8 are pseudogenes while Gy7 is expressed at very low levels in soybean seeds (Fischer and Goldberg, 1982; Beilinson et al., 2002; Li and Zhang, 2011). β-conglycinin is also encoded by multigene family that is clustered in several genomic regions (Tierney et al., 1987; Harada et al., 1989; Sebastiani et al., 1990). β-conglycinin is a glycoprotein made up of α′, α, and β subunits (Thanh and Shibasaki, 1976). These two groups of proteins account for most of the seed storage proteins of soybean (Nielsen, 1996; Krishnan, 2000).

Published in Crop Sci. 52:262–271 (2012).
doi: 10.2135/cropsci2011.08.0454
Published online 17 Nov. 2011.

Soybean also accumulates another class of 7S protein that is soluble only in high ionic strength salt solutions (Yamauchi et al., 1984). This basic protein is synthesized as a precursor protein of 43 kDa and posttranslationally cleaved into α- (27 kDa) and β- (16 kDa) subunits that are held together by a disulfide bond (Yamazaki et al., 2003; Shutov et al., 2010). Interestingly, the 43 kDa shows protein kinase activity (Komatsu et al., 1994) and contains a consensus adenosine triphosphate-binding motif implicated in protein phosphorylation (Watanabe and Hirano, 1994). In addition to the protein kinase activity, the 7S basic globulin is able to bind insulin and insulin-like growth factors indicating that this protein may function as hormone receptor protein (Watanabe and Hirano, 1994). This observation led to the search for the hormone-like peptides in soybean that can bind with the 43 kDa protein. After considerable effort, a 4 kDa peptide was isolated from germinating soybean radicles (Watanabe et al., 1994). This hormone-like peptide was named leginsulin and was shown to stimulate the protein kinase activity of the 43 kDa protein suggesting a role in signal transduction pathway (Watanabe et al., 1994).

Leginsulin, a 37 amino acid peptide, is synthesized as a precursor polypeptide that is posttranslationally cleaved to yield 4 and 6 kDa peptides (Watanabe et al., 1994; Hirano, 2006). Immunolocalization studies have shown that the 4 kDa leginsulin is localized around the plasma membrane and cell walls while the 6 kDa peptide is localized in the protein storage vacuoles (Nishizawa et al., 1994). Leginsulin belongs to the cysteine-knot family (Jouvensal et al., 2003) and is homologous to pea (*Pisum sativum* L.) albumin (PA1b), a peptide toxin lethal to certain insects (Rahioui et al., 2007). Leginsulin and PA1b share about 65% amino acid sequence identity and include six cysteine residues suggesting a role of these residues in the structure and function of these two peptides (Jouvensal et al., 2003; Rahioui et al., 2007). In addition to the insecticidal activities it has been suggested that members of cysteine-knot peptide family may also influence glucose metabolism in mice (*Mus musculus* L.) (Dun et al., 2008). Even though the effect of leginsulin on glucose metabolism has not been reported, the structural similarity of this peptide with several of the 37 amino acid peptides in pea suggest leginsulin could also regulate blood glucose concentration in mammals (Dun et al., 2008). In spite of its potential role in signal transduction, defense response, and glucose metabolism, very little is known about leginsulin expression, accumulation, and distribution among soybean accessions. In this study we demonstrate that the accumulation of leginsulin is prevalent among Asian accessions and that the paucity of leginsulin accumulation in North American soybean accessions is related to low level expression of leginsulin genes in the developing cotyledons.

MATERIALS AND METHODS

Plant Material

A total of 485 soybean accessions exhibiting maximum genetic diversity and geographical diversity were obtained from the USDA Soybean Germplasm Collection, Urbana, IL. A brief description of the relevant characteristics of these soybean accessions have been described earlier (Kim et al., 2008). Soybean cultivar Williams 82 and PI 458249 were grown in a greenhouse (16-h daylength and 30/22°C day/night temperatures) in 25-cm plastic pots containing commercial Pro-Mix (Premier Horticulture, Riviere-du-Loup, QB, Canada). Periodically the plants were fertilized with Miracle-Gro according to the manufacturer's (Scotts Miracle-Gro Company, Marysville, OH) recommendations.

Isolation of Seed Proteins

Mature seeds from 485 soybean lines were individually ground to a fine powder in liquid N using a mortar and pestle. Total seed proteins were obtained from 10 mg of seed powder by extracting with 1 mL of sodium dedecyl sulfate (SDS)-sample buffer [60 mM Tris-HCl, pH 6.8, 2% SDS (w/v), 10% glycerol (v/v), and 5% 2-mercaptoethanol] followed by boiling for 5 min. The slurry was clarified by centrifugation (15800 × *g* for 5 min) and the resulting supernatant was transferred to a clean Eppendorf tube. Isopropanol-soluble proteins were isolated by extracting 100 mg seed powder with 1 mL of 50% isopropanol in a 2-mL Eppendorf tube. The tubes were left in a 30°C shaker for 1 h followed by centrifugation (15800 × *g* for 5 min). Two volumes of acetone were added to the resulting supernatant and this solution was incubated overnight at −20°C. Precipitated proteins were recovered by centrifugation as before, air dried, and suspended in 200 μL of SDS-sample buffer. Aliquots (10 μL) of protein samples were resolved by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis Analysis

Soybean total seed proteins were resolved by SDS-PAGE (Laemmli, 1970) using a Hoeffer SE 260 minigel apparatus (GE Healthcare, Piscataway, NJ) according to the manufacturer's recommendations. Electrophoretic separation was performed with 20 mA per gel under constant current. Resolved proteins were visualized by staining the gel overnight with Coomassie Blue R-250.

Generation of Leginsulin Antibody

An 18 amino acid peptide of the leginsulin protein (NGACSPFEVPPCRSRDCR) was obtained commercially. The purified peptide was conjugated to the keyhole limpet hemocyanin protein and used to immunize two rabbits (*Oryctolagus cuniculus*). Antibodies were purified on an affinity column and the purified serum was stored in small aliquots at −80°C.

Immunoblot Analysis

Soybean seed proteins resolved by SDS-PAGE were electrophoretically transferred to nitrocellulose membranes (Protran 85, Schleicher & Schuell Inc., Keene, NH). Following the transfer the membranes were blocked with 5% milk in Tris-buffered saline (pH 7.3) for 1 h and incubated with leginsulin peptide antiserum or pre-immune serum at 1:10,000 dilutions overnight at room temperature with gentle rocking. The membrane was washed four times with Tris-buffered saline containing 0.05% Tween-20 (TBST) for 10 min each wash and incubated for 2 h

in affinity purified goat anti-rabbit IgG-horseradish peroxidase conjugate (Bio-Rad Laboratories, Hercules, CA) at a 1:3000 dilution. Following this step, the membrane was washed four times with TBST as described above. Immunoreactive polypeptides were detected with an enhanced chemiluminescent substrate (Super Signal West Pico Kit; Pierce Biotechnology, Rockford, IL) according to manufacturer's protocol.

Northern Analysis

Total RNA from developing soybean seeds (stage 4, weighing between 300 and 350 mg; Phartiyal et al., 2006) was extracted following the Trizol Reagent protocol (Invitrogen, Carlsbad, CA). Ten micrograms of total RNA was electrophoresed in a 1.5% (w/v) agarose–formaldehyde gel, transferred by capillary blotting to a Hybond-N+ charged nylon membrane (GE Healthcare, Piscataway, NJ), and fixed to the membrane by ultraviolet cross-linking. A 431 bp coding region of leginsulin was [^{32}P]-labeled and hybridized to the membrane. Prehybridization (10 h at 65°C) and hybridization (30 h at 55°C) were performed in 7% SDS (w/v), 0.191 M Na_2HPO_4, 0.058 M NaH_2PO_4, 1% (w/v) bovine serum albumin, and 100 μg mL^{-1} salmon (*Salmo salar* L.) sperm DNA. After hybridization, the membrane was washed three times for 10 min at room temperature in 2x saline–sodium citrate and 0.5% SDS and then exposed to X-ray films at −80°C.

Semiquantitative Reverse Transcriptase Polymerase Chain Reaction Analysis

Total RNA (1 μg) was used as template for reverse transcriptase polymerase chain reactions (RT-PCRs). Before RT-PCR, the RNA was treated with deoxyribonuclease (DNase) I (Invitrogen, Carlsbad, CA) to remove any contaminating DNA. The reverse transcriptase reaction was performed in a volume of 50 μL using the OneStep RT-PCR kit (Qiagen, Valencia, CA) with primers designed to amplify a 431 bp coding region of leginsulin (forward: 5′-AATGGCTTATGCTAGGCTTGCTCC-3′; reverse: 5′-TAGTAGAAGATGCATCAAGCATGCCAC-3′). The polymerase chain reaction (PCR) products were gel purified and an aliquot was digested with *Spe*I and separated on a 0.7% agarose gel. The resolved PCR products were visualized using the GeneWizzard bioimaging system (LabRepco, Horsham, PA).

Polymerase Chain Reaction Amplification of Leginsulin from Soybean Genomic DNA

Total genomic DNA from soybean cultivar Williams 82 and PI 458249 was isolated from freshly harvested young leaf using the cetyltrimethylammonium bromide (CTAB) method (Saghai Maroof et al., 1984). Leginsulin 1 gene was amplified from genomic DNA using the following primers sets: forward primer 5′-AGCAAGTTAAGTGTTTATTAAACATTTTCC-3′ and reverse primer 5′-TAGTAGAAGATGCATCAAGCATGCCCAC-3′. Similarly, leginsulin 2 was amplified using the following primer sets: forward primer 5′-ATTTACGATCCCATGC-ACTCCAAAGGCAAAATC-3′ and reverse primer 5′-TAGTAGAAGATGCATCAAGCATGC-CCAC-3′. Polymerase chain reactions were performed 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 deoxyribonucleotide triphosphates (dNTPs), 1 μM of each primer, 100 ng of genomic DNA, and two units of ExTaq DNA polymerase (Pan Vera Corp., Madison, WI). The amplified PCR products were resolved on a 1.0% agarose gel. The desired DNA fragments were excised from the gel and purified with the help of 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) and the DNA sequences were elucidated at the University of Missouri DNA Core Facility (Columbia, MO). The sequence of leginsulin 1 and leginsulin 2 genes of PI 458249 have been deposited in the GenBank database (Benson et al., 2008) under accession numbers JF718836 and JF718837, respectively.

RESULTS

Leginsulin is Enriched in 50% Isopropanol Extracted Protein Fraction

We have previously shown that 50% isopropanol is effective in extracting proteins that are rich in cysteine (Krishnan, 2004). Since leginsulin is a cysteine-rich peptide we examined if isopropanol can be used to preferentially extract soybean leginsulin also. For this purpose, a North American soybean cultivar (Williams 82) and a Chinese soybean line (PI 458249) were used in our initial analysis. Sodium dodecyl sulfate polyacrylamide gel electrophoresis analysis of soybean seed proteins extracted with 50% isopropanol resulted in the resolution of various sized proteins. Prominent among them were 52, 32, 34, 21, 18, 16, 14, and 12 kDa proteins (Fig. 1A). We have previously identified the 52 kDa protein as β-amylase (Kim and Krishnan 2010) and the 21 and 14 kDa proteins as Kunitz trypsin inhibitor and Bowman Birk protease inhibitor, respectively (Krishnan, 2004). Immunoblot blot analysis utilizing antibodies raised against the 4 kDa peptide demonstrated a marked difference in the accumulation of this protein between the Asian soybean accession and the North American soybean cultivar (Fig. 1B and 1C). When equal amount of protein was loaded on the gel the accumulation of the 4 kDa protein was only detected in soybean accession PI 458249 (Fig. 1B). However, when twice the amount of protein from soybean accession PI 458249 was loaded a faint immunoreactive protein was detected (Fig. 1C) indicating that soybean accession PI 458249 accumulated the 4 kDa protein in greater amounts than soybean cultivar Williams 82. Two lines of evidence confirmed that this 4 kDa protein is leginsulin. The first line of evidence was obtained by immunoblot analysis using antibodies raised against the leginsulin peptide. The antibody specifically reacted against the 4 kDa peptide from soybean PI 458249 (Fig. 1B). Longer exposure of the immunoblot also detected the 4 kDa peptide in Williams 82 (data not shown) indicating very low levels of leginsulin in this North American cultivar. Second, we excised this protein band from the gel from soybean PI 458249, subjected it to in-gel digestion with trypsin, and analyzed with matrix-assisted laser desorption time-of-flight

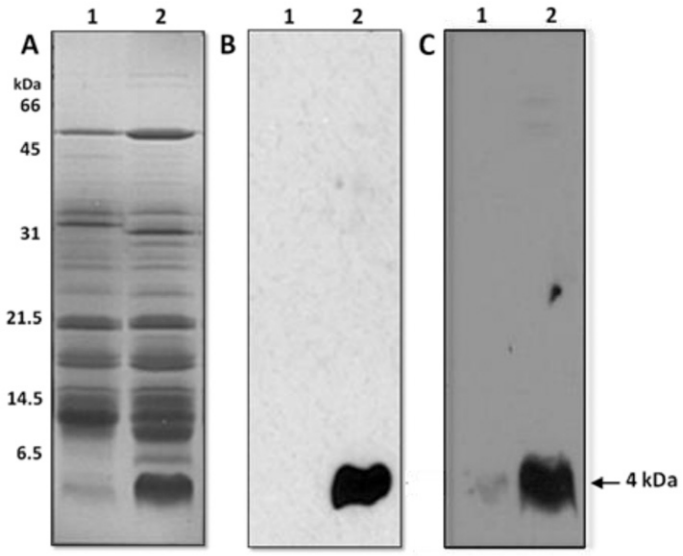

Figure 1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of 50% isopropanol soluble proteins from soybean seeds. Panel A: Isopropanol soluble proteins were extracted from dry seeds of soybean cultivar Williams 82 (Lane 1) and PI 458249 (Lane 2) and analyzed on a 15% SDS-PAGE. Resolved proteins were visualized by staining the gel with Coomassie Blue. Panel B: Proteins shown in Panel A were transferred to nitrocellulose (Williams 82, Lane 1; PI 458249, Lane 2) and probed with antibodies raised against the leginsulin protein. Panel C is the same as panel B except that twice the amount protein from Williams 82 and half the amount of protein from PI 458249 was loaded on the gel. Immunoreactive proteins were detected by chemiluminescence using anti-rabbit IgG horseradish peroxidase conjugate. The position of the protein molecular weight markers in kilodaltons are shown on the left of the figure.

(MALDI-TOF) mass spectrometry. Using Mascot (Matrix Science, 2010), the empirically determined mass-to-charge ratios of peptides were compared with peptides of known proteins listed in the National Center for Biotechnology Information (Bethesda, MD) nonredundant database. Mascot search results showed two peptides from the 4 kDa peptide having significant sequence homology to soybean leginsulin (Table 1).

Leginsulin is Abundantly Present in Asian Soybean Accessions

Since our initial analysis indicated differences in the accumulation of leginsulin in soybean accessions, an extensive analysis using a diverse array of soybean accessions from the USDA Soybean Germplasm Collection was conducted. We isolated 50% isopropanol-soluble proteins from 485 soybean accessions to identify soybean accessions that are enriched in leginsulin. Western blot analysis using leginsulin-specific antibodies led to the identification of 75 accessions that significantly accumulated leginsulin (Table 2). Remarkably, all these accessions with the exception of two (PI 548488 and PI 438471) had their origin from Japan, Korea, or China (Table 2). Leginsulin was barely detected in most of the North American soybean accessions examined in this study.

Leginsulin Accumulates Both in Embryonic Axis and Cotyledons of PI 458249

Even though leginsulin is abundantly present in several of the Asian soybean accessions it is not known if it accumulates in all parts of the seed. To examine its tissue distribution we isolated 50% isopropanol-soluble proteins from embryonic axis and cotyledons of soybean cultivar Williams 82 and PI 458249 (Fig. 2A) and performed immunoblot analysis. The leginsulin antibodies strongly reacted against a 4 kDa peptide from both the cotyledon and embryonic axis of PI 458249 (Fig. 2B). In the case of Williams 82 a positive reaction with the 4 kDa protein from the embryonic axis was detected. However, no reaction was detected against any proteins from the cotyledons of Williams 82 even after prolonged exposure of the X-ray film.

Differential Accumulation of Leginsulin in Soybean Cultivar Williams 82 and PI 458249

To examine the molecular basis of the differential accumulation of leginsulin between Williams 82 and PI 458249 we investigated the steady state messenger RNA (mRNA) levels in the developing seeds by northern analysis. Total RNA isolated from soybean seeds at developmental stage 4 (weighing 300 to 350 mg) were fractionated in 1.5% formaldehyde gel and hybridized with a 0.4 kb leginsulin probe. Northern blot analysis detected leginsulin mRNA in Williams 82 embryonic axis but not in the cotyledon (Fig. 3A). In contrast, leginsulin mRNA was abundantly present both in the embryonic axis and the cotyledons of PI 458249 (Fig. 3A). The results from the northern blot analysis reveals differential expression of leginsulin mRNA in the cotyledon and the embryonic axis of soybean cultivar Williams 82.

Two Leginsulin Genes are Present in Soybean Genome

Examination of the soybean cultivar Williams 82 genome sequence (Joint Genome Institute, 2011) revealed the presence

Table 1. Identification of the 4 kDa peptide from soybean PI 458249 as leginsulin by mass spectrometry†.

Mass spectrometry identified peptides	Mass	Percent coverage	Mass spectrometry/mass spectrometry score	Mass spectrometry fit: data base, accession, and identification
ADCNGACSPFEVPPCR	1879.87 Da	49	104	NCBInr‡, gi|159162564, chain A, leginsulin
PIGLFVGFCIHPTG	1774.84 Da	43	NA	NCBInr, gi|159162564, chain A, leginsulin

†The larger peptide was identified using matrix-assisted laser desorption time-of-flight (MALDI-TOF) mass spectrometry (MS) and MS/MS and the smaller peptide with MALDI-TOF MS alone.

‡NCBI, 2011.

Table 2. Leginsulin accumulating soybean Plant Introduction lines and their relevant characteristics†.

Plant Introduction no.	Accession name	Country	Maturity group	Seed coat color	Protein (%)	Oil (%)
548438	Arksoy	North Korea	VI	Yellow	46.4	17.2
548379	Mandarin (Ottawa)	China	0	Yellow	43.3	20.3
548402	Peking	China	IV	Black	38.2	18.5
548484	Ralsoy	North Korea	VI	Yellow	46.3	17.3
548488	S-100	United States	V	Yellow	46.7	17.6
163453	Quail Haven	China	VII	Black	44.7	12.0
366120		Japan	IV	Black	54.9	9.7
378694		Japan	VI	Black	49.5	10.7
407127		Japan	VII	Black	46.8	10.2
407197		South Korea	V	Black	52.9	12.0
487431	K113	Japan	IX	Black	52.1	9.2
507580	Tsuru Mame	Japan	V	Black	52.4	10.8
507597	NIAR 030005	Japan	V	Black	49.8	11.2
507613	NIAR 040021	Japan	VI	Black	45.5	11.4
507645	NIAR 070004	Japan	VI	Black	46.3	10.8
507646	NIAR 080001	Japan	VII	Black	44.8	10.5
507652	NIAR 080005	Japan	VII	Black	48.3	9.3
507654	NIAR 090002	Japan	VII	Black	51.7	10.1
507656	NIAR 090011	Japan	VII	Black	48.8	10.6
80837	Mejiro	Japan	IV	Yellow	42.4	19.0
89138	Zontanorukon	South Korea	II	Yellow	38.6	20.2
90245	Neihen	South Korea	IV	Yellow	40.9	19.0
96783	932	South Korea	IV	Yellow	40.0	18.5
159764		South Korea	00	Green	52.6	15.7
379559B	Komagi dadacha	Japan	I	Brown	45.9	17.7
398290	KAERI-GNT 172-10	South Korea	IV	Black	40.6	17.7
407788A	ORD 8113	South Korea	IV	Yellow	50.7	15.1
407849	KAS 510-1	South Korea	III	Yellow	43.8	18.1
423706	Oh won No. 1	South Korea	I	Yellow	45.7	19.0
424472B	KAS 581-20	South Korea	IV	Reddish brown	40.9	18.7
438471	Fiskeby III	Sweden	00	Yellow	40.3	20.3
464940	Wen Feng No. 1538	China	III	Yellow	41.4	18.6
464941	Wen Feng No. 1893	China	II	Yellow	41.0	18.1
468408C	Qi Huang No. 1	China	IV	yellow	39.8	19.2
508296D	Kumjong Kong/Maeju Kong	South Korea	IV	Yellow	51.3	15.3
592935	Fu dou 1	China	III	Yellow	41.7	17.9
592950	Yu dou 11	China	III	Yellow	40.2	18.6
593976	Gokuwase chishima	Japan	000	Black	55.6	14.4
593979	Wasekosode	Japan	000	Green	55.8	14.7
594252A	Orihime	Japan	II	Yellow	42.1	16.3
594398B	87-32	China	IV	Yellow	40.4	17.9
594403	85-135-1	China	IV	Yellow	41.0	18.0
603147	GL 1738/82	North Korea	000	Green	56.0	15.2
603149	GL 2216/84	North Korea	000	Green	57.9	14.8
603155	GL 2623/96	North Korea	IV	Green	51.5	14.7
603372	ZDD00679	China	II	Black	44.6	18.1
603424C	ZDD007871	China	I	Yellow	40.9	18.5
603429A	ZDD007968	China	0	Reddish brown	51.3	17.1
90406	7401	China	VI	Yellow	45.7	17.9
97094	1243	South Korea	VII	Yellow	45.0	17.7
200477	Hondo Daizu	Japan	VII	Yellow	54.2	14.8
398473	KAERI-GNT 220-3	South Korea	VI	Black	40.9	18.3
398969	KLS 629-1	South Korea	V	Yellow	41.0	18.7
398983	KLS 711-1	South Korea	VI	Yellow	51.5	11.7
407781C	ORD 8105	South Korea	VI	Yellow	52.1	13.1

(cont'd)

Table 2. Continued.

Plant Introduction no.	Accession name	Country	Maturity group	Seed coat color	Protein (%)	Oil (%)
407801		South Korea	VI	Yellow	48.8	12.9
407808-1	ORD 8146	South Korea	V	Green	51.5	13.5
407825	ORD 8195	South Korea	V	Green	50.5	15.7
407876	KAERI 511-10	South Korea	V	Green	52.0	12.8
407892B	KAERI 523-4	South Korea	V	Yellow	51.1	14.1
408056	KAERI 576-4	South Korea	V	Yellow	53.1	12.1
417120	Kyushu 25	Japan	VIII	Yellow	50.6	13.4
423986	Akasaya	Japan	VI	Yellow	42.4	20.4
424376	KAERI-GNT 353-15	South Korea	VI	Yellow	41.7	18.9
424433	KAERI-GNT 544-15	South Korea	VI	Yellow	53.7	10.9
504288	S	Japan	V	Black	50.4	8.2
506676	Furuzato Daizu	Japan	VII	Yellow	42.5	18.5
506792	Iwa A 2	Japan	VI	Green	40.3	20.9
507327	Tamahomare	Japan	VI	Yellow	40.8	20.8
567335B	Lai da dou	China	V	Yellow	50.2	17.8
567342	Ma hei dou	China	V	Greenish brown	51.9	14.3
567378	Ba yue zha	China	VI	Black	42.5	19.9
593987	SY 9514004	South Korea	VII	Black	44.6	20.3
594217C	Misao	Japan	VII	Yellow	50.5	15.8
341256	CMS	Vietnam	IX			

†Data compiled from USDA-ARS Germplasm Resources Information Network (USDA-ARS, 2010).

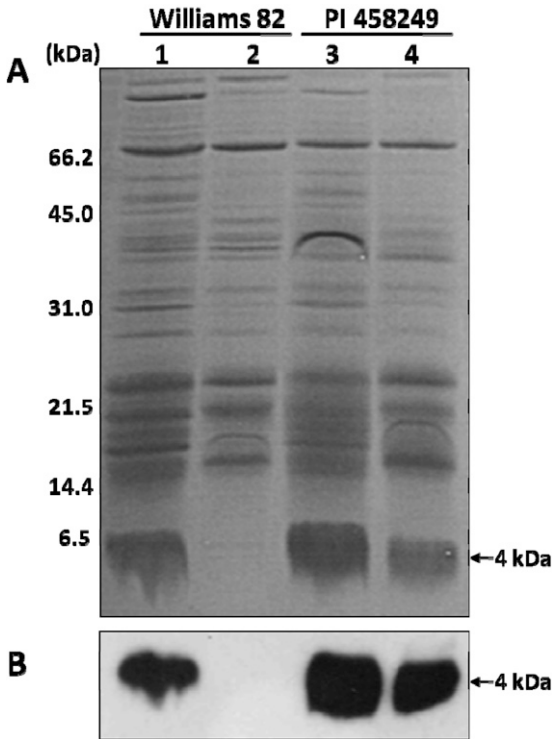

Figure 2. Differential accumulation of leginsulin in soybean seeds. Panel A: 50% isopropanol soluble proteins were extracted from embryonic axis (Lanes 1 and 3) and cotyledons (Lanes 2 and 4) of soybean cultivar Williams 82 and PI 458249, respectively. Proteins were resolved on a 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by staining the gel with Coomassie Blue. Panel B: Proteins shown in Panel A were transferred to a nitrocellulose membrane and probed with antibodies raised against the leginsulin protein. Immunoreactive proteins were detected by chemiluminescence using anti-rabbit IgG horseradish peroxidase conjugate. The position of the protein molecular weight markers in kilodaltons are shown on the left of the figure.

of two homologous leginsulin genes (Gm13 g26330 [leginsulin 1] and Gm13 g26340 [leginsulin2]) on chromosome 13. The genetic organization of these two closely related genes is shown in Fig. 4A. These two leginsulin genes are 5.8 kb apart from each other. Leginsulin 1 and 2 shows 92.5% sequence similarity. Leginsulin 1 contains a single intron of 489 bp while the leginsulin 2 is interrupted by a 371 bp intron (Fig. 4A). Leginsulin 1 and 2 comprise 146 and 147 amino acids and encode proteins of 16026 and 16048 Da, respectively. Both the leginsulins are synthesized as precursor polypeptides in which residues 48 through 84 from the amino terminus represents the leginsulin. A comparison of the amino acid sequence of leginsulin 1 and 2, which share 88.9% sequence identity, shows differences in four positions (Fig. 4B). Since soybean cultivar Williams 82 and PI 458249 reveal striking differences in the accumulation of leginsulin we wanted to examine if there are structural differences in the leginsulin genes between these soybean lines. We cloned the two leginsulin genes from PI 458249 and found them to be highly similar to that of Williams 82. As in the case of Williams 82, the leginsulin 1 and leginsulin 2 from PI 458249 were synthesized as precursor protein with molecular weight of 16 kDa and show 98.6% sequence identity. Unlike Williams 82, the amino acid sequence of the mature leginsulin 1 and leginsulin 2 were identical (Fig. 4B).

Different Leginsulin Genes are Expressed in Developing Seeds of Williams 82 and PI 458249

A comparison of the nucleotide sequence reveals that leginsulin 1 and leginsulin 2 of Williams 82, in spite of their extensive sequence homology, can be distinguished from

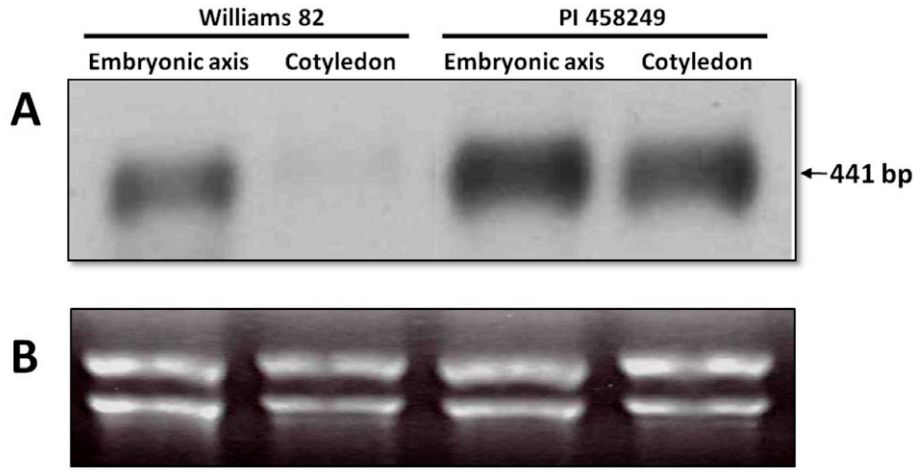

Figure 3. Northern blot analysis of leginsulin messenger RNA (mRNA) in soybean seeds. Total RNA isolated separately from embryonic axis and cotyledon of soybean cultivar Williams 82 and PI 458249 was resolved on a formaldehyde gel and probed with ^{32}P-labeled leginsulin coding sequences. The lower panel shows ethidium bromide stained gel, which served as a loading control.

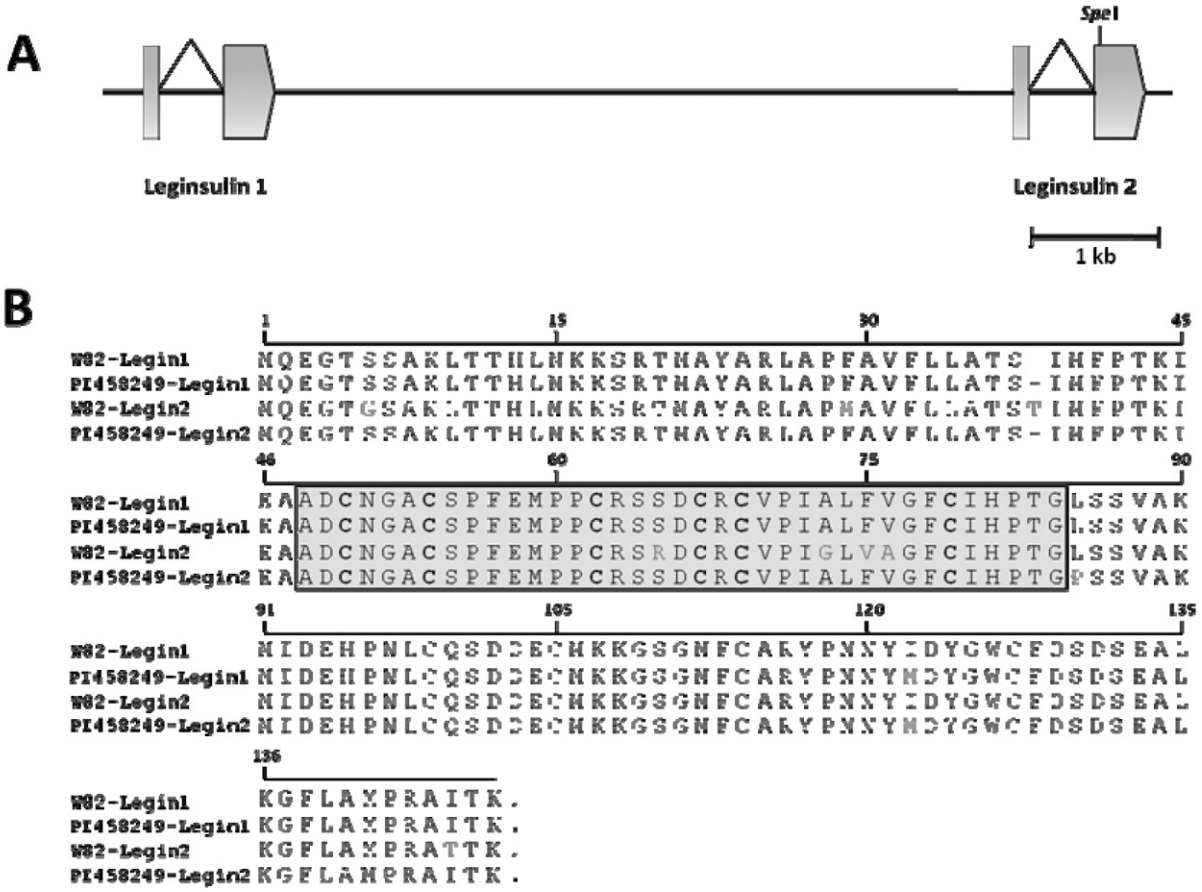

Figure 4. Genetic organization of soybean leginsulin genes. Panel A: Soybean cultivar Williams 82 genome sequence (Joint Genome Institute, 2011) revealed the presence of two homologous leginsulin genes [Gm13 g26330 (leginsulin 1) and Gm13 g26340 (leginsulin 2)] on chromosome 13 that are 5.9 kb apart. Both genes contain a single intron (shown as a triangle). Panel B: Amino acid sequence alignment of leginsulin 1 and 2 from soybean cultivar Williams 82 and PI 458249. The 37 amino acids of leginsulin peptide are boxed and the six conserved cysteine residues in the leginsulin peptide are shown in bold font.

each other. Leginsulin 2 contains a unique *Spe*I site (Fig. 4A) that can be conveniently exploited to distinguish these two closely related genes. In contrast, this unique *Spe*I site is absent in leginsulin 2 from PI 458249. The expression levels of leginsulin 1 and leginsulin 2 in developing soybean seeds were examined by RT-PCR utilizing total RNA isolated from the cotyledons and embryonic axis of Williams and PI 458249. Reverse transcriptase polymerase chain reaction analysis (23 cycles) also confirmed very low leginsulin transcript levels in the Williams 82 cotyledons. In order

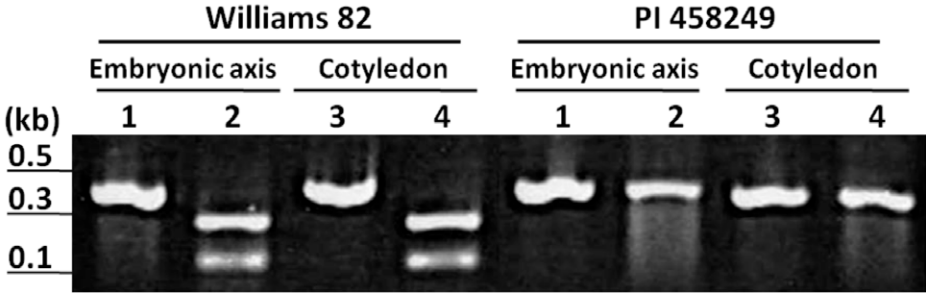

Figure 5. Differential expression of leginsulin 1 and 2 in developing soybean seeds. Total RNA was isolated separately from cotyledons and embryonic axis of developing soybean seeds and used in reverse transcriptase polymerase chain reaction (RT-PCR) assays. Leginsulin specific primers were used to amplify a 431 bp fragment. An aliquot of the RT-PCR product was digested with the restriction enzyme *Spe*I (lanes 2 and 4) and separated along with the undigested (lanes 1 and 3) fragment on a 1.5% agarose gel. The sizes of the polymerase chain reaction (PCR) products in kilobase pairs are shown to the left of the figure.

obtain enough transcripts from the cotyledonary tissue we extended the RT-PCR to 35 cycles, which resulted in the amplification of a 400 bp fragment from both the cotyledons and embryonic axis of Williams and PI 458249. To distinguish between leginsulin 1 and leginsulin 2 we purified the RT-PCR products and digested them with *Spe*I restriction enzyme. Analysis of the restriction enzyme products by agarose gel electrophoresis (Fig. 5) revealed that the RT-PCR product from Williams 82 resulted in the appearance of 160 and 270 bp fragments indicating that it is a leginsulin 2 product. In contrast, the RT-PCR product from PI 458249 remained intact since both leginsulin 1 and leginsulin 2 lack the *Spe*I site (Fig. 5).

DISCUSSION

The identification of a 43 kDa protein capable of binding animal insulin (Komatsu et al., 1994) with demonstrated protein kinase activity promoted an intensive search for a ligand in soybean. This led to the discovery of a leginsulin (4 kDa peptide) from soybean. Extensive research has been conducted to understand the interaction of leginsulin with the 43 kDa basic globulin (Hanada et al., 2003; Hanada and Hirano, 2004). Leginsulin expressed in *Escherichia coli* have been used to demonstrate that the C-terminal hydrophobic region of the peptide is essential for its binding to the 43 kDa protein (Hanada et al., 2003). The three-dimensional structure of leginsulin elucidated by nuclear magnetic resonance spectroscopy revealed the spatial arrangement of the hydrophobic residues at the solvent-exposed surface of the two stranded β sheet is responsible for the binding activity and stimulatory effect on protein phosphorylation (Yamazaki et al., 2003; Hirano, 2006). The physiological function of the leginsulin was also explored by transiently expressing this peptide in cultured carrot (*Daucus carota* L.) and bird's-foot trefoil (*Lotus corniculatus* L.) cells (Yamazaki et al., 2003). The transgenic callus was found to grow rapidly in comparison to the wild-type callus. An important role in signal transduction pathway is proposed for this hormone-like peptide since it was shown that the expression of leginsulin resulted in accelerated cell growth and differentiation (Yamazaki et al., 2003).

Leginsulin-like peptides are widespread in legumes (Kagawa et al., 1987; Louis et al., 2004). The amino acid sequence of these peptides is highly conserved. They all contain six cysteine residues that are involved in three disulfide bonds and are essential for ligand function (Watanabe et al., 1994; Yamazaki et al., 2003). Pea albumin (PA1b), a protein highly homologous to leginsulin, has been shown to possess insecticidal properties (Louis et al., 2004, 2007). By employing a combination of molecular, biochemical and specific insect bioassays, the presence of A1b peptides have been identified in several members of subfamilies Papilionoidease and Mimosideae but not in Caesalpinioideae (Louis et al., 2007). In spite of the fact that leginsulin is widespread in legumes its restricted distribution among North American soybean accessions is intriguing. Additionally, further research is required to ascertain if soybean leginsulin has any insecticidal properties.

Based on the available literature, two physiological functions can be assigned to leginsulin. One relates to signal transduction pathway and the other in defense against insect pests. We speculate that leginsulin may have only a limited role in regulating soybean growth and development since they are barely present in most of the commercially grown North American soybean cultivars. Since leginsulin-like peptides are widely distributed in legumes and have demonstrated insecticidal activities, it is tempting to suggest that the primary function of leginsulin is to protect the seeds from insects. Clearly, further in-depth studies are required to clarify the physiological function of leginsulin in soybean seeds. In this study we have demonstrated that the accumulation of leginsulin is significantly higher in most of the Asian soybean accessions. Since these soybean accessions are not commercially grown and are prone to insect attacks, one may expect that these accessions accumulate higher concentrations of leginsulin to benefit seed survival. This would lead to an evolutionary

pressure to maintain a higher concentration of this peptide in these soybean accessions. During the course of domestication and artificial selection for higher yield, protein, and oil, some of the desirable traits found in the wild soybeans may have been lost. The low level accumulation of leginsulin in commercial soybean cultivars may be a result of this continuing strong artificial selection. This possibility is strengthened by the observation that there is higher genome diversity in wild soybeans compared to cultivated soybeans (Lam et al., 2010).

The low levels of leginsulin accumulation in Williams 82 seeds appears not due to any major deletion or insertion in the coding region of these two closely related genes. The paucity of leginsulin accumulation in Williams 82 could be related to the low expression of leginsulin genes in this North American cultivar. A comparison of the 1 kb upstream sequences from the translation start site of leginsulin1 and 2 indicate the presence of several conserved promoter motifs in both these genes. Interestingly, significant differences in the number and location of MYB and soybean embryo factor 4 (SEF 4) motifs were detected in the promoter of leginsulin 1 and leginsulin 2. It has been previously shown that the SEF 4 motif serves as the binding site for the embryo factor that activates the differential and stage-specific expression of soybean seed storage protein genes (Itoh et al., 1994). Further investigation is required to examine the role, if any, of MYB and SEF 4 motifs in regulating the expression of leginsulin genes in the embryonic axis and the cotyledon of Williams 82.

Acknowledgments

The authors thank Nathan Oehrle (USDA-ARS, University of Missouri) for critical reading of this manuscript. Funding for this research was provided by ARS project 3622-21000-000-00D. Product names are necessary to report factually on available data; however, the University of Missouri and the USDA neither guarantees nor warrants the standard of product, and the use of the name by the University of Missouri and the USDA implies no approval of the product to the exclusion of others that may be suitable.

References

Beilinson, V., Z. Chen, R. Shoemaker, R. Fischer, R. Goldberg, and N. Nielsen. 2002. Genomic organization of glycinin genes in soybean. Theor. Appl. Genet. 104:1132–1140. doi:10.1007/s00122-002-0884-6

Benson, D.A., I. Karsch-Mizachi, D.J. Lipman, J. Ostell, and D.L. Wheeler. 2008. GenBank. Nucleic Acids Res. 36:D25–D30.

Dun, X., F. Li, J. Wang, and Z. Chen. 2008. The effect of pea albumin 1F on glucose metabolism in mice. Peptides 29:891–897. doi:10.1016/j.peptides.2008.01.013

Fischer, R.L., and R.B. Goldberg. 1982. Structure and flanking regions of soybean seed protein genes. Cell 29:651–660. doi:10.1016/0092-8674(82)90181-7

Hanada, K., and H. Hirano. 2004. Interaction of a 43-kDa receptor-like protein with a 4-kDa hormone-like peptide in soybean. Biochemistry 43:12105–12112. doi:10.1021/bi049587a

Hanada, K., Y. Nishiuchi, and H. Hirano. 2003. Amino acid residues on the surface of soybean 4-kDa peptide involved in the interaction with its binding protein. Eur. J. Biochem. 270:2583–2592. doi:10.1046/j.1432-1033.2003.03627.x

Harada, J.J., S.J. Baker, and R.B. Goldberg. 1989. Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and post-transcriptional processes. Plant Cell 1:415–425. doi:10.1105/tpc.1.4.415

Hirano, H. 2006. 4-kDa peptide. p. 1–4. *In* A.J. Kastin (ed.) Handbook of biologically active peptides. Academic Press, Burlington, MA.

Itoh, Y., Y. Kitamura, and C. Fukazawa. 1994. The glycinin box: A soybean embryo factor binding motif within the quantitative regulatory region of the 11S seed storage globulin promoter. Mol. Gen. Genet. 243:353–357. doi:10.1007/BF00301071

Joint Genome Institute. 2011. Phytozome v7.0: Glycine max. Available at http://www.phytozome.net/soybean (verified 8 Nov. 2011). University of California Regents, Oakland, CA.

Jouvensal, L., L. Quillien, E. Ferrasson, Y. Rahbe, J. Gueguen, and F. Vovelle. 2003. PA1b, an insecticidal protein extracted from pea seeds (*Pisum sativum*) (1)H-2D-NMR study and molecular modeling. Biochemistry 42:11915–11923. doi:10.1021/bi034803l

Kagawa, H., F. Yamauchi, and H. Hirano. 1987. Soybean basic 7S globulin represents a protein widely distributed in legume species. FEBS Lett. 226:145–149. doi:10.1016/0014-5793(87)80568-9

Kim, W.S., H.J. Ho, R.L. Nelson, and H.B. Krishnan. 2008. Identification of several *gy4* nulls from the USDA soybean germplasm collection provides new genetic resources for the development of high-quality tofu cultivars. J. Agric. Food Chem. 56:11320–11326. doi:10.1021/jf801831w

Kim, W.S., and H.B. Krishnan. 2010. The lack of beta-amylase activity in soybean cultivar Altona sp_1 is associated with a 1.2 kb deletion in the 5′ region of beta-amylase I gene. Crop Sci. 50:1942–1949. doi:10.2135/cropsci2009.11.0660

Komatsu, S., O. Koshio, and H. Hirano. 1994. Protein kinase activity and insulin-binding activity in plant basic 7S globulin. Biosci. Biotechnol. Biochem. 58:1705–1706. doi:10.1271/bbb.58.1705

Krishnan, H.B. 2000. Biochemistry and molecular biology of soybean seed storage proteins. J. New Seeds 2:1–25. doi:10.1300/J153v02n03_01

Krishnan, H.B. 2004. A simple and rapid method to isolate low molecular weight proteinase inhibitors from soybean. Korean J. Crop Sci. 49:342–348.

Lam, H., X. Xu, X. Liu, W. Chen, G. Yang, F. Wong, M. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J. Wang, G. Shao, J. Wang, S.S. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. 42:1053–1059. doi:10.1038/ng.715

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685. doi:10.1038/227680a0

Li, C., and Y.M. Zhang. 2011. Molecular evolution of glycinin and β-conglycinin gene families in soybean (*Glycine max* L. Merr.). Heredity 106:633–641. doi:10.1038/hdy.2010.97

Louis, S., B. Delobel, F. Gressent, G. Duport, O. Diol, I. Rahioui, H. Charles, and Y. Rahbe. 2007. Broad screening of the

legume family for variability in seed insecticidal activities and for the occurrence of the A1b-like knottin peptide entomotoxins. Phytochemistry 68:521–535. doi:10.1016/j.phytochem.2006.11.032

Louis, S., B. Delobel, F. Gressent, I. Rahioui, L. Quillien, A. Vallier, and Y. Rahbe. 2004. Molecular and biological screening for insect-toxic seed albumins from four legume species. Plant Sci. 167:705–714. doi:10.1016/j.plantsci.2004.04.018

Matrix Science. 2010. Mascot. Available at http://www.matrixscience.com/ (verified 8 Nov. 2011). Matrix Science, Inc., Boston, MA

National Center for Biotechnology Information (NCBI). 2011. NCBI nr database. Available at http://www.ncbi.nlm.nih.gov/. NCBI, U.S. National Library of Medicine, Bethesda, MD.

Nielsen, N.C. 1996. Soybean seed composition. p. 127–163. *In* D.P.S. Verma and R.C. Shoemaker (ed.) Soybean: Genetics, molecular biology and biotechnology. CABI, Wallingford, UK.

Nielsen, N.C., C.D. Dickinson, T.J. Cho, B.H. Thanh, B.J. Scallon, R.L. Fischer, T.L. Sims, G.N. Drews, and R.B. Goldberg. 1989. Characterization of the glycinin gene family. Plant Cell 1:313–328. doi:10.1105/tpc.1.3.313

Nishizawa, N.K., S. Mori, Y. Watanabe, and H. Hirano. 1994. Ultrastructural localization of the basic 7S globulin in soybean (*Glycine max*) cotyledons. Plant Cell Physiol. 35:1079–1085.

Phartiyal, P., W.-S. Kim, R.E. Cahoon, J.M. Jez, and H.B. Krishnan. 2006. Soybean ATP sulfurylase, a homodimeric enzyme involved in sulfur assimilation, is abundantly expressed in roots and induced by cold treatment. Arch. Biochem. Biophys. 450:20–29. doi:10.1016/j.abb.2006.03.033

Rahioui, I., C. Laugier, S. Balmand, P. Da Silva, Y. Rahbe, and F. Gressent. 2007. Toxicity, binding and internalization of the pea-A1b entomotoxin in Sf9 cells. Biochimie 89:1539–1543. doi:10.1016/j.biochi.2007.07.021

Saghai Maroof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA. 81:8014–8019.

Sebastiani, F.L., L.B. Farrell, M.A. Schuler, and R.N. Beachy. 1990. Complete sequence of a cDNA of the a subunit of soybean β-conglycinin. Plant Mol. Biol. 15:197–201.

Shutov, A.D., K. Prak, T. Fukuda, S.V. Rudakov, A.S. Rudakova, M.R. Tandang-Silvas, K. Fujiwara, B. Mikami, S. Utsumi, and N. Maruyama. 2010. Soybean basic 7S globulin: Subunit heterogeneity and molecular evolution. Biosci. Biotechnol. Biochem. 74:1631–1634. doi:10.1271/bbb.100234

Thanh, V.H., and K. Shibasaki. 1976. Major proteins of soybean seeds. A straight forward fractionation and their characterization. J. Agric. Food Chem. 24:1117–1121. doi:10.1021/jf60208a030

Tierney, M.L., E.A. Bray, R.D. Allen, Y. Ma, R.F. Drong, J. Slightom, and R.N. Beachy. 1987. Isolation and characterization of a genomic clone encoding the β-subunit of β-conglycinin. Planta 172:356–363. doi:10.1007/BF00398664

USDA-ARS. 2010. Germplasm resources information network (GRIN). National plant germplasm system. Available at http://www.ars-grin.gov/npgs/acc/acc_queries.html (verified 8 Nov. 2011). National Germplasm Resources Laboratory, Beltsville, MD.

Watanabe, Y., S.F. Barbashov, S. Komatsu, A.M. Hemmings, M. Miyagi, S. Tsunasawa, and H. Hirano. 1994. A peptide that stimulates phosphorylation of the plant insulin-binding protein. Isolation, primary structure and cDNA cloning. Eur. J. Biochem. 224:167–172. doi:10.1111/j.1432-1033.1994.tb20008.x

Watanabe, Y., and H. Hirano. 1994. Nucleotide sequence of the basic 7S globulin gene from soybean. Plant Physiol. 105:1019–1020. doi:10.1104/pp.105.3.1019

Yamauchi, F., K. Sato, and T. Yamagishi. 1984. Isolation and partial characterization of a salt-extractable globulin from soybean seeds. Agric. Biol. Chem. 48:645–650. doi:10.1271/bbb1961.48.645

Yamazaki, M., M. Takaoka, E. Katoh, K. Hanada, M. Sakita, K. Sakata, Y. Nishiuchi, and H. Hirano. 2003. A possible physiological function and the tertiary structure of a 4-kDa peptide in legumes. Eur. J. Biochem. 270:1269–1276. doi:10.1046/j.1432-1033.2003.03489.x