Mol Breeding (2013) 32:291–298
DOI 10.1007/s11032-013-9870-8

Identification of a plant introduction soybean line with genetic lesions affecting two distinct glycinin subunits and evaluation of impacts on protein content and composition

Won-Seok Kim · Jason D. Gillman · Hari B. Krishnan

Received: 14 February 2013 / Accepted: 16 April 2013 / Published online: 30 July 2013

Abstract Unlike other oilseeds, soybean (*Glycine max* [L.] Merr) is also valuable due to its direct conversion into human food. One notable example is the cheese-like product tofu. The quality of tofu is improved when protein subunits derived from two glycinin genes, *Gy1* and *Gy4*, are reduced or absent. Here we report the discovery that one exotic soybean plant introduction line, PI 605781 B, has not only a previously described loss-of-expression mutation affecting one glycinin gene (*gy4*), but also bears an extremely rare, potentially unique, frameshift mutation in the *Glycinin1* gene (*gy1-a*). We analyzed glycinin gene expression via qRT-PCR with mRNA from developing seeds, which revealed that the novel allele dramatically reduced *Gy1* mRNA accumulation. Similarly, both $A_4A_5B_3$ and $A_{1a}B_{1a}$ protein subunits were absent or at undetectable levels, as determined by two-dimensional protein fractionation. Despite the reduction in glycinin content, overall seed protein levels were unaffected. The novel *gy1-a* allele was found to be unique to PI 605871B in a sampling of 247 diverse germplasm lines drawn from a variety of geographic origins.

Keywords Glycinin · Soybean · Tofu · Seed proteins

Introduction

Soybean (*Glycine max* [L.] Merr) seeds feature exceptionally high levels of both protein (approx. 40 %) and oil (approx. 20 %) and are one of the largest sources of edible oil for human consumption (as well as a feedstock for industrial applications). During oil purification, protein-rich soybean meal is produced, which forms the primary source of value for the soybean crop; it is estimated that 75 % of protein meal for animal feed worldwide is derived from soybean (Boland et al. 2013).

A portion of soybean seeds is also used directly for human food, an example of which is the cheese-like food tofu. Tofu quality is directly correlated with the protein composition of the soybean seeds used (Poysa et al. 2006). Soybean seed proteins predominantly correspond to two protein classes: glycinins (11S globulin) and β-conglycinins (7S globulin), which

Won-Seok Kim and Jason D. Gillman contributed equally to this work.

Mention of a trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendors that may also be suitable.

Electronic supplementary material The online version of this article (doi:10.1007/s11032-013-9870-8) contains supplementary material, which is available to authorized users.

W.-S. Kim · J. D. Gillman · H. B. Krishnan (✉)
United States Department of Agriculture-Agricultural Research Service, University of Missouri-Columbia, 205 Curtis Hall, Columbia, MO 65211, USA
e-mail: Hari.Krishnan@usda.ars.gov

collectively comprise approximately 70 % of the protein in a typical soybean seed (Derbyshire et al. 1976). Glycinin subunits form a 12-subunit heteromeric protein consisting of equal numbers of acidic and basic units (Kitamura et al. 1976).

In the sequenced genotype 'Williams 82', a total of eight glycinin genes have been identified (Li and Zhang 2011), although only five of these genes (*Gy1–5*) are expressed at significant levels (Beilinson et al. 2002) [Electronic Supplemental Material (ESM) Table 1]. Elite high-yielding soybean lines grown in the USA are the result of an extreme "bottleneck event", with >80 % of the genetic diversity present in high-yielding cultivars deriving from <12 lines (Gizlice et al. 1994, 1996). As a result, studies on the ancestors of high-yielding North American commercial lines have shown that all glycinin genes are intact (*Gy1–5*; Mahmoud et al. 2006; Kim et al. 2008).

In contrast to the relatively small number of glycinin-encoding genes, as many as 15 β-conglycinin genes are present in the 'Williams 82' genome (Li and Zhang 2011). However, despite this high number of potential β-conglycinin encoding genes, the majority appear to be pseudogenes. Only three subunit encoding genes (α, α′, and β; Li and Zhang 2011) apparently correspond to the most highly seed-expressed protein subunits (Thanh and Shibasaki 1978), and null alleles for each have been combined into lines with a virtual absence of β-conglycinin (Takahashi et al. 1994).

In recent decades, several researchers have identified (or developed through radiation induced mutagenesis) soybean lines with loss of function mutations affecting all five of the glycinin subunits (Scallon et al. 1987; Cho et al. 1989a; Yagasaki et al. 1996). Despite the high levels of these proteins in seeds, all five glycinins and all three major β-conglycinins are dispensable; null lines which lack glycinin and beta-conglycinin are competent during all stages of reproduction and growth (Takahashi et al. 2003). The most successful strategy, in terms of elimination of total number of glycinin subunits, has been a combination of radiation-induced deletion (Yagasaki et al. 1996; Jenkinson and Fehr 2010) along with traditional breeding methods to incorporate a naturally occurring mutation affecting *Gy4* (Yagasaki et al. 1996; Jegadeesan et al. 2012).

In this study, we report on the identification of an exotic soybean line from the US. Department of Agriculture–Agriculture Research Service (USDA–ARS) germplasm collection which has mutations eliminating the protein expression of two glycinin encoding genes, *Gy1* and *Gy4*. We also report the development of effective molecular assays to detect these alleles and demonstrate that inheritance of both alleles virtually re-creates the protein phenotype of the exotic germplasm lines. The discovery of these two distinct mutations that reduce the level of the two subunits and of molecular markers to track these alleles will assist in the development of new "improved" tofu varieties in terms of value and quality.

Materials and methods

Plant materials

Soybeans of line PI 605871 B and 'Williams 82' were obtained from the USDA-ARS GRIN system. Seeds from 'Patriot' were obtained from Dr. David Sleper of the University of Missouri.

Protein extraction from soybean seeds

Electrophoresis chemicals were obtained from GE Healthcare [Piscataway, NJ; ammonium persulfate, acrylamide, bis-acrylamide, CHAPS, dithiothreitol, immobilized pH gradient (IPG) strips, sodium dodecyl sulfate (SDS), TEMED, thiourea], from Bio-Rad [Hercules, CA; ampholytes, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) molecular weight markers, urea], or from Sigma-Aldrich (St. Louis, MO; glycerol, 2-mercaptoethanol, Tris). Seeds were individually ground to a fine powder using a mortar and pestle and proteins were extracted from a 10-mg sample by incubation with 1 ml of SDS sample buffer [5 % 2-mercaptoethanol, 2 % SDS (w/v), 10 % glycerol (v/v) and 60 mM Tris–HCl, pH 6.8] and boiling for 5 min. Samples were then clarified by centrifugation at 15,800*g* for 5 min, and the supernatant was transferred to a clean Eppendorf tube.

One-dimensional SDS-PAGE

A 10-µl aliquot of the protein extract supernatant was loaded onto a SDS-PAGE gel apparatus (model Hoefer SE 260 minigel; Hoefer Inc., Holliston, MA, USA) and proteins were resolved according to manufacturer's recommendations.

Two-dimensional gel electrophoresis

A 100-mg aliquot of finely ground seed powder was prepared and extracted in 2 ml of extraction solution [10 % TCA (w/v), 0.07 % (v/v) 2-mercaptoethanol in acetone]. Seed proteins were resuspended in lysis buffer [1 % CHAPS (w/v), 1 % DTT (w/v), 1 % ampholytes (w/v) pH 3–10, 9 M urea] and sonicated. Samples were clarified by centrifugation at 20,800*g* for 20 min at 4 °C, and an aliquot was used in isoelectric focusing with IPG strips (pH 3–10, length 13 cm) in an IPGphor system (GE Healthcare, Little Chalfont, UK). Each protein sample (100 µg) was loaded onto strips by active rehydration at 50 V in 250 µl of buffer [2 % CHAPS (w/v), 0.5 % ampholytes, 1 % dithiothreitol (DTT), 0.00 2 % bromphenol blue, 8 M urea]. Two-dimensional (2-D) separation was performed with a Hoefer SE 600 Ruby electrophoresis unit (GE Healthcare) according to the manufacturer's recommendations.

SDS-PAGE image analysis

Images of Coomassie stained gels were obtained with a UMAX PowerLook 2100XL scanner with Adobe Photoshop V7.0 (Adobe Systems, San Jose, CA, USA), and images were scanned at 100-µm resolution (600 dpi). Images were analyzed with Phoretix 2D Advanced software (v6.01; Nonliner Dynamics Ltd., NewCastle upon Tyne, UK).

Reverse transcription-PCR and PCR sequencing analysis of glycinin genes

A 1-µg sample of DNase I-treated total RNA was used as the template for reverse transcription (RT)-PCR with gene-specific primers [Electronic Supplemental Material (ESM) Table 1]. The One-Step RT-PCR kit was used to amplify transcripts according to the procedures recommended by the manufacturer (Qiagen, Valencia, CA, USA). For *Gy1* and *Gy4* sequence analysis, leaf genomic DNA from PI 605871 B was isolated using the DNeasy kit (Qiagen) according to manufacturer's recommendations, and approximately 25 ng DNA was used in the PCR reactions in place of cDNA. The amplified PCR products were electrophoresed in a 1.0 % agarose gel, and resolved PCR products were isolated and purified using Ultrafree-DA columns (Millipore Corp., Bedford, MA, USA). The purified PCR products were individually cloned into pGEM-T easy vector (Promega, Madison, WI, USA) and then Sanger sequenced, using M13 primers, on a 3730xl 96-capillary DNA Analyzer with the Applied Biosystems BigDye Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) at the University of Missouri DNACore Facility.

qRT-PCR of glycinin genes during seed development

RNA was isolated from mid-maturation green soybean seeds (size 8–10 mm) from lines PI 605781 B and 'Williams 82' and then treated with DNase and purified as previously described (Gillman et al. 2013). A 1-µg sample of total RNA was used to generate cDNA using the SMARTscribe RT (Clonetech, Mountain View, CA, USA) with random hexamers, and 1/20th of a 20-µl RT reaction was used in gene-specific quantitative PCR using the Quatitect SYBR Green PCR kit (Qiagen). The primers used are listed in ESM Table 1. For each genotype, RNA from three individual biological replicates (each consisting of two seeds) were used for quantitation, using the ΔΔCt method (Livak and Schmittgen 2001), with the *cons7* gene as a reference gene (Libault et al. 2008).

Total protein analysis of seeds via LECO

Fifteen seeds per line, grown in the same field location in Columbia, Missouri during the summer of 2011, were freeze-dried, ground and used for total nitrogen/protein analysis as previously described (Bilyeu et al. 2008). Three separate biological replicates were evaluated per line, and the results were compared using Student's *t* test method (Samuels and Witmer 1999).

Development of molecular markers for *Gy1* and *Gy4* mutations

Molecular marker assays to track the *gy1-a* allele (NCBI accession KC460320) and *Gy4* polymorphism (Scallon et al. 1987) were designed using Lightcycler Probe Design Software, ver. 1 (Roche Applied Sciences, Indianapolis, IN, USA) to be complimentary to the 'Williams 82' sequences (Schmutz et al. 2010). The sequences for the *Gy1* SimpleProbe and *Gy4* Simple Probe were "Fluorescein-SPC-GTCTTTACCCTTGC

ACTGTGG-Phosphate" and "Fluorescein-SPC-GAGA GTGAAGGGCTTCCCCATAT-Phosphate", respectively, with SPC referring to a proprietary SimpleProbe sequence (Fluorosentric, Park City, UT, USA). Assymetric PCR with manually designed gene-specific primers (5:1 ratio of forward to reverse; ESM Table 1) was used to generate the template for binding with the SimpleProbes. SimpleProbes were used at a 0.2 μM final concentration. Amplification reactions and melting curve analysis were as previously described (Bilyeu et al. 2011). Examples of typical "melting peak" analysis, using a Lightcycler 480 II real-time PCR instrument (Roche Applied Sciences) are shown in ESM Fig. 1.

Results

Identification of a *Gy4* null line with additional reduction in *Gy1*-encoded subunits

In our previous work (Kim et al. 2008), we identified 38 plant introduction accessions with genetic reductions in the $A_4A_5B_3$ glycinin peptides, all of which are encoded by the *Gy4* locus. In these accessions, we detected a single base change which alters the start codon (ATG > ATA) such that it cannot be translated. In subsequent analysis, we identified one plant introduction line, PI 605871 B, which also featured a further reduction in either A_{1a}, A_{1b}, A_2 or A_3 subunits as detected by SDS-PAGE (Fig. 1). In order to determine which of these highly similar glycinin genes were reduced, we performed 2-D fractionation of proteins (isoelectric focusing followed by SDS-PAGE) which revealed that the A_{1a} (*Gy1*) and the $A_4A_5B_3$ (*Gy4*) proteins are either absent or at undetectable levels in seeds of PI 605871B (Fig. 2).

Examination of *Gy1* gene transcripts by RT-PCR and DNA sequencing

In order to assess if the alterations in glycinin subunits could potentially be due to reductions in glycinin mRNA abundance, we utilized quantitative (q)RT-PCR to quantify expression of a subset of glycinin subunit encoding genes (*Gy1*, *Gy2*, *Gy4*, *Gy5*), using a constitutively expressed gene (*cons7*) (Libault et al. 2008) as a control (Fig. 3). *Gy3* and *Gy7* were expressed at an extremely low level, which was too

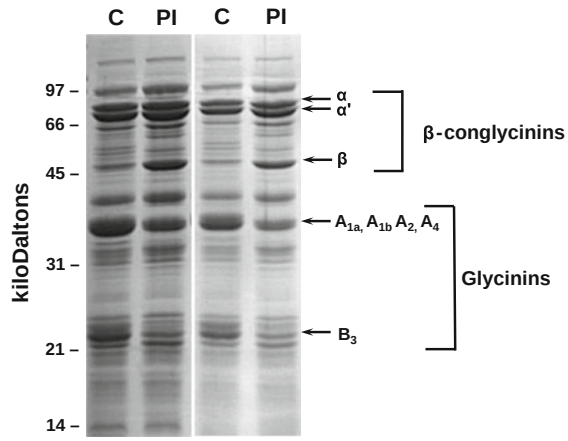

Fig. 1 One-dimensional SDS-PAGE analysis of soybean proteins seeds from line PI 605781 B in comparison to a public cultivar 'Patriot'. *Lanes*: *C* Proteins from 'Patriot' seeds, *PI* protein from PI 605781 B seeds. The *first two lanes* contain 40 μg of seed protein, and the *second two lanes* contain 20 μg seed protein, separated on a 12.5 % SDS-PAGE gel. Approximate protein masses were inferred from comparison to a standard protein ladder

low to be reliably quantified (Cp > 30). These findings are consistent with a recent RNA-seq analysis which revealed very low expression for *Gy3* and *Gy7* as compared to *Gy1*, *Gy2*, *Gy4* and *Gy5* in developing seeds (Severin et al. 2010). *Gy2* and *Gy5* did not show significantly differential expression in mid-maturation seeds between 'Williams 82' and PI 605871 B. In contrast, transcripts from the *Gy1* locus were at a significantly lower expression level ($P = 0.0099$, approx. 1,000 fold lower; Fig. 3) in seeds from line PI 605871 B as compared to 'Williams 82', as were transcripts from the *Gy4* locus ($P = 0.01$, approx. 10-fold lower; Fig. 3).

We hypothesized that a polymorphism affecting either the promoter or coding sequence of the *Gy1* gene could be acting to reduce transcript abundance and protein accumulation. To assess this possibility, we performed PCR with gene-specific primers (ESM Table 1) and cloned full-length transcripts into pGEM-T Easy Vector (Promega). Individual clones were then Sanger sequenced for *Gy1*, *Gy2*, *Gy3* and *Gy5*. Upon sequencing, *Gy2*, *Gy3* and *Gy5* were found to contain no sequence differences, as compared to the reference 'Williams 82' genome. In contrast, the *Gy1* locus from PI 605871 B was found to bear a single base deletion within exon 3 (G292Δ relative to start

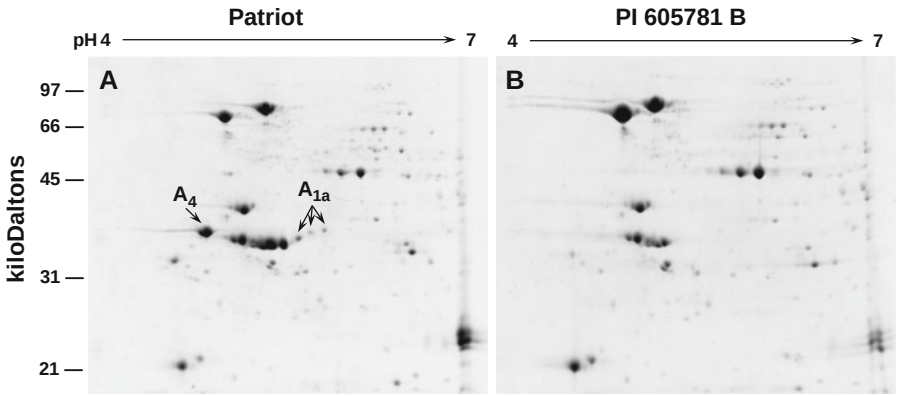

Fig. 2 Two-dimensional proteomic analysis of soybean seeds from line PI 605781 B in comparison to a public cultivar 'Patriot'. Seed proteins were separated by isoelectric focusing on pH 4–7 IPG strips followed by separation by SDS-PAGE on 16 % gels. The gels were stained with colloidal Coomassie blue G-250. Approximate protein masses were inferred from comparison to a standard protein ladder

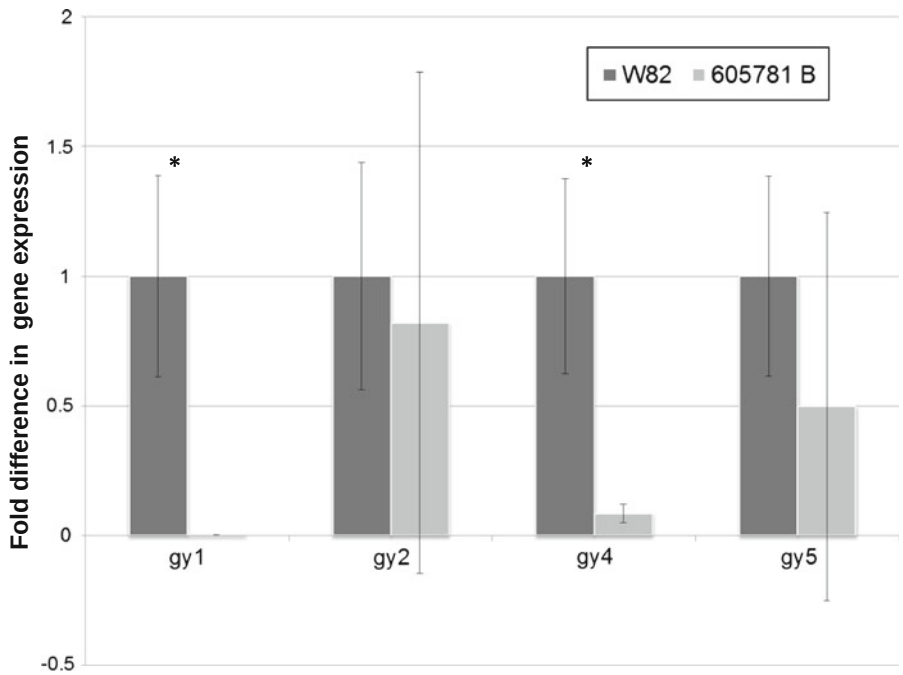

Fig. 3 Quantitative reverse-transcription (qRT)-PCR analysis of glycinin encoding genes in mRNA from developing (8–10 mm) seeds. Gene expression was normalized to W82 samples for each gene, respectively, using the ΔΔCt method and *cons7* (see "Methods") as a reference gene. Each histogram represents three independent biological replicates, and one standard deviation is indicated in *parenthesis*. *Asterisk* indicates significant differences at $P \leq 0.01$

codon; Fig. 4), which generates an immediate frame-shift mutation (Fig. 4). In this work, we will refer to this allele as *gy1-a* (NCBI accession KC460320). The reduction in *gy1-a* transcript abundance is most likely due to the phenomenon termed "nonsense-mediated mRNA decay" (Peltz et al. 1993), which has been shown to be a conserved mRNA surveillance pathway in plants (recently reviewed in Chang et al. 2007).

Development of molecular markers for *gy1-a* and *gy4* alleles

In order to confirm that the *gy1-a* single base deletion was causative for the protein alterations, we developed a segregating population by crossing PI 605781 B to a public soybean cultivar 'Patriot', developed by David Sleper and Kerry Clark of the University of Missouri.

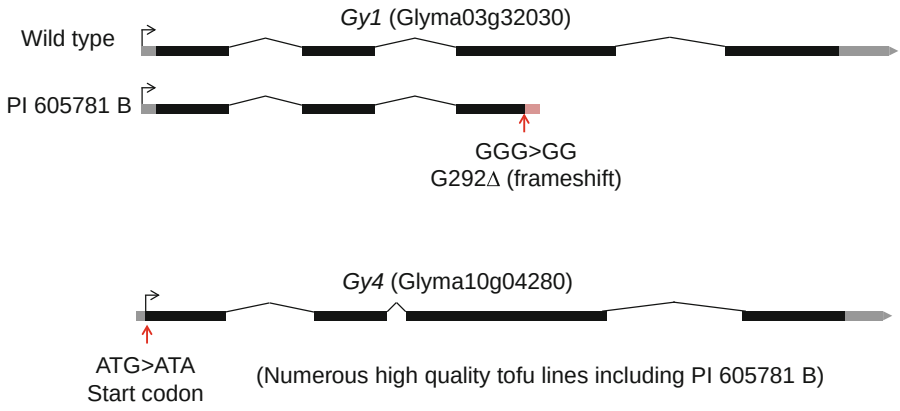

Fig. 4 Genetic lesions affecting Glycinin 1 (*Gy1*) and Glycinin 4 (*Gy4*) loci in plant introduction line PI 605781 B. Cartoon depiction of polymorphisms present in PI 605871 B within the *gy1-a* (Glyma03g32030) and *gy4* (Glyma10g04280) genes. *Dark bars* Exons, *lines* spliced introns, *gray bars* untranslated regions

We developed locus-specific markers for *Gy4* using SimpleProbe technology and evaluated F_2 plants for genotype and $F_{2:3}$ seeds by SDS-PAGE analysis for presence of A_4B_3 A_5 subunits. As previously shown for a number of lines, reductions in the A_4B_3 A_5 subunits were perfectly correlated with the presence of homozygosity for *Gy4* alleles from PI 605781 B (ESM Fig. 2).

The co-migration of A_{1a}, A_{1b} and A3 proteins on 1-D SDS-PAGE gels results in significant difficulty separating A_{1a} from A_{1b}, A_2 and A_3 on SDS-PAGE gels (Fig. 5). Performing sufficient 2-D gels for breeding purposes and/or genetic analysis for hundreds of samples is logistically and financially prohibitive. This may in part explain why a single *Gy1* mutation has not previously been reported in the literature. The results of our 1-D SDS-PAGE analysis revealed a reduction in the protein band corresponding to A_{1a} in seeds from PI 605871 B (Fig. 1), and this reduction was present in the re-creation of the *gy1-a/gy4* genotype (Fig. 5). We anticipate complete correlation between this genetic marker and the protein phenotype, as has been shown previously for genetic deletions encompassing both *Gy1* and *Gy2* (Cho et al. 1989b; Beilinson et al. 2002; Jenkinson and Fehr 2010). The re-creation of the *gy1-a*, *gy4* double homozygote results in an overall seed protein profile very similar to that of PI 605871 B (Fig. 5).

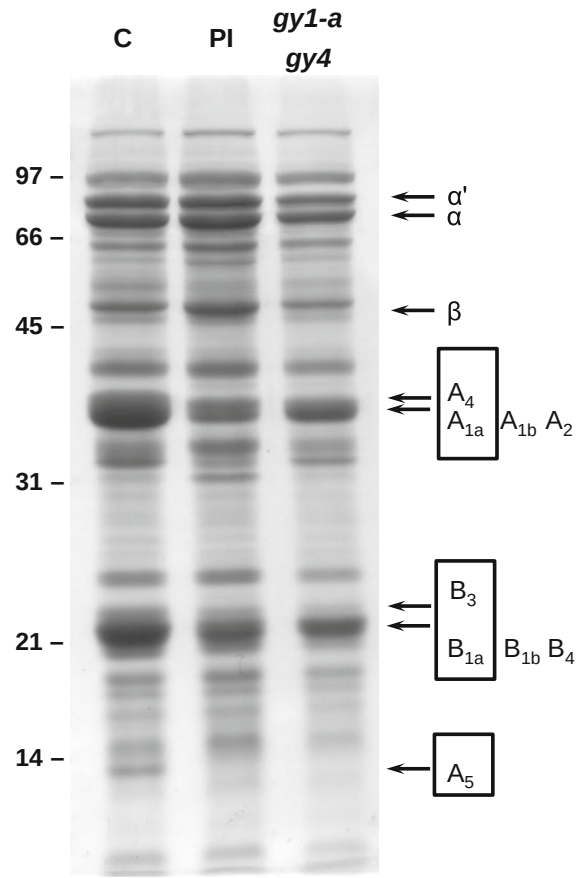

Fig. 5 SDS-PAGE separation of proteins from Patriot (*C*), PI 605871 B (*PI*) and a *gy1-a*, *gy4* double homozygote $F_{2:3}$ line. A 40-µg sample of seed protein was separated on 12.5 % SDS-PAGE gels. *Arrows* Typical migration pattern of β-conglycinin and glycinin subunits, *box* protein subunits encoded by the *Gy1* or *Gy4* loci

Examination of a diverse set of genotypes for presence of the novel *Gy1* allele

The development of single nucleotide polymorphism genotyping markers allowed us to assess the prevalence of the novel *Gy1* allele in a diverse sampling of 247 germplasm lines drawn from the USDA-ARS GRIN collection (ESM Table 2). We examined four representations of soybean genetic diversity: (1) all of the ancestral lines which make up >95 % of the

genetic diversity in North American high-yielding elite cultivars; (2) germplasm lines previously shown to bear the *Gy4* mutation, including PI 605781 B (Kim et al. 2008); (3) a set of samples drawn from a geographically diverse *Glycine max* and *G. soja* lines (Gillman et al. 2013); (4) The 31 wild and domesticated genomes which were recently resequenced (Lam et al. 2010). Remarkably, the novel *gy1-a* allele was present only in PI 605781 B, suggesting that it is an extremely rare, possibly unique, allele. A full list of the 247 lines examined is presented in ESM Table 2.

Discussion

Despite their strong similarity, distinct glycinin subunits have different functional characteristics for gelling and, consequently, for tofu quality. The presence of the $A_{1b}B_{1b}$, A_2B_2 and A_3B_4 protein subunits are correlated with higher quality tofu, whereas the absence of $A_{1a}B_{1a}$ and $A_4A_5B_3$ subunits are correlated with improved quality (Poysa et al. 2006). Mutations affecting *Gy3*/$A_{1b}B_{1b}$ have been reported (Cho et al. 1989a), but the presence of the *Gy3* derived A_3 subunit is positively correlated with tofu quality (Poysa et al. 2006), and such mutations are unlikely to result in superior tofu varieties. Another spontaneous mutation which prevents translation of the *Gy4* locus (Scallon et al. 1987) is relatively widespread in Japanese cultivars with superior tofu quality (Kim et al. 2008).

Two glycinin genes in tight genetic linkage have opposite impacts on tofu quality: $A_{1a}B_{1a}$ subunits (derived from the *Gy1* locus) negatively impact tofu quality when present, whereas the presence of A_2B_2 subunits (encoded by the *Gy2* locus) have been shown to have a positive association with tofu quality/firmness (Poysa et al. 2006). These two gene copies most likely arose due to an ancestral tandem duplication event. No germplasm source of a specific reduction in the *Gy1* locus, aside from complete deletion of both the *Gy1* and *Gy2* locus (Yagasaki et al. 1996), has to our knowledge been reported in the literature.

We have not evaluated the impacts of the presence of the novel *gy1-a* mutation on seed yield or tofu quality, as has been done for the *gy4* mutation. Because PI 605781 B is an agronomically poor-quality plant introduction line which has not been optimized for either seed yield nor other aspects of tofu quality (high protein, clear hilum, sugar content, etc.), such studies are best done after additional breeding. We do not anticipate that the alleles per se will have an effect on either seed yield; studies with irradiation lines featuring loss of all glycinin subunits have revealed no significant effect of loss of glycinin subunits on overall seed yield (Jenkinson and Fehr 2010).

Although we detected differences in specific glycinin subunit composition via protein gels, there was no significant difference ($P = 0.217$) in total protein content between seeds from PI 605871 B and 'Patriot' (39.0 ± 0.4 % and 39.3 ± 0.1 %, respectively). This finding is consistent with other studies, where the combination of multiple glycinin mutations (Jenkinson and Fehr 2010) or the combination of both glycinin and β-conglycinin mutations in a single line (Takahashi et al. 2003) did not alter the total amount of protein/amino acids present, only how they were partitioned. Similar results were obtained using RNAi gene silencing methods to reduce β-conglycinin (Kinney et al. 2001) or to silence both β-conglycinin and glycinin genes using a single gene silencing construct (Schmidt et al. 2011). These studies demonstrated a remarkable plasticity in the soybean seed proteome, with the mechanisms that control the total accumulation of amino acids operating independently of how those amino acids are fractionated into particular seed storage proteins.

Acknowledgments The authors would like to acknowledge the superb technical assistance of Sungchan Jang, Jeremy Mullis and Nathan Oehrle. This research was supported by the USDA-Agricultural Research Service.

References

Beilinson V, Chen Z, Shoemaker RC, Fischer RL, Goldberg RB, Nielsen NC (2002) Genomic organization of glycinin genes in soybean. Theor Appl Genet 104(6–7):1132–1140

Bilyeu KD, Zeng P, Coello P, Zhang ZJ, Krishnan HB, Bailey A, Beuselinck PR, Polacco JC (2008) Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. Plant Physiol 146(2): 468–477

Bilyeu K, Gillman JD, LeRoy AR (2011) Novel FAD3 mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Sci 51(1):1–5

Boland MJ, Rae AN, Vereijken JM, Meuwissen MPM, Fischer ARH, van Boekel MAJS, Rutherfurd SM, Gruppen H,

Moughan PJ, Hendriks WH (2013) The future supply of animal-derived protein for human consumption. Trends Food Sci Technol 29(1):62–73

Chang Y-F, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 76:51–74

Cho TJ, Davies CS, Fischer RL, Turner NE, Goldberg RB, Nielsen NC (1989a) Molecular characterization of an aberrant allele for the *Gy3* glycinin gene: a chromosomal rearrangement. Plant Cell 1(3):339–350

Cho TJ, Davies CS, Nielsen NC (1989b) Inheritance and organization of glycinin genes in soybean. Plant Cell 1(3): 329–337

Derbyshire E, Wright D, Boulter D (1976) Legumin and vicilin, storage proteins of legume seed. Phytochemistry 15(1): 3–24

Gillman JD, Baxter I, Bilyeu K (2013) Phosphorus partitioning of soybean lines containing different mutant alleles of two soybean seed-specific adenosine triphosphate-binding cassette phytic acid transporter paralogs. Plant Genome 6(1):1–10

Gizlice Z, Carter TE Jr, Burton JW (1994) Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci 34(5):1143–1151

Gizlice Z, Carter TE Jr, Gerig TM, Burton JW (1996) Genetic diversity patterns in North American public Soybean cultivars based on coefficient of parentage. Crop Sci 36(3): 753–765

Jegadeesan S, Yu K, Woodrow L, Wang Y, Shi C, Poysa V (2012) Molecular analysis of glycinin genes in soybean mutants for development of gene-specific markers. Theor Appl Genet 124(2):365–372

Jenkinson JE, Fehr WR (2010) Agronomic and seed characteristics of soybean lines with alleles for modified glycinin concentration. Crop Sci 50(5):1896–1903

Kim W-S, Ho HJ, Nelson RL, Krishnan HB (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(23):11320–11326

Kinney AJ, Jung R, Herman EM (2001) Cosuppression of the α subunits of β-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. Plant Cell 13(5):1165–1178

Kitamura K, Takagi T, Shibasaki K (1976) Subunit structure of soybean 11S globulin. Agric Biol Chem Tokyo 40(9): 1837–1844

Krishnan HB, Natarajan SS, Mahmoud AA, Nelson RL (2007) Identification of glycinin and beta-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. J Agric Food Chem 55(5): 1839–1845

Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, Li M-W, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Wang J, Sun SS-M, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42(12):1053–1059

Li C, Zhang YM (2011) Molecular evolution of glycinin and beta-conglycinin gene families in soybean (*Glycine max* L. Merr.). Heredity 106(4):633–641

Libault M, Thibivilliers S, Bilgin DD, Radwan O, Benitez M, Clough SJ, Stacey G (2008) Identification of four soybean reference genes for gene expression normalization. Plant Gen 1(1):44–54

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25(4):402–408

Mahmoud AA, Natarajan SS, Bennett JO, Mawhinney TP, Wiebold WJ, Krishnan HB (2006) Effect of six decades of selective breeding on soybean protein composition and quality: a biochemical and molecular analysis. J Agric Food Chem 54(11):3916–3922

Peltz SW, Brown AH, Jacobson A (1993) mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. Genes Dev 7(9):1737–1754

Pieretti MM, Chung D, Pacenza R, Slotkin T, Sicherer SH (2009) Audit of manufactured products: use of allergen advisory labels and identification of labeling ambiguities. J Allergy Clin Immun 124(2):337–341

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

Samuels ML, Witmer JA (1999) Statistics for the life sciences, 2nd edn. Prentice-Hall, London

Scallon BJ, Dickinson CD, Nielsen NC (1987) Characterization of a null-allele for the *Gy4* glycinin gene from soybean. Mol Gen Genet 208(1):107–113

Schmidt MA, Barbazuk WB, Sandford M, May G, Song Z, Zhou W, Nikolau BJ, Herman EM (2011) Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. Plant Physiol 156(1):330–345

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Severin A, Woody J, Bolon Y-T, Joseph B, Diers B, Farmer A, Muehlbauer G, Nelson R, Grant D, Specht J, Graham M, Cannon S, May G, Vance C, Shoemaker R (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol 10(1):160

Takahashi K, Banba H, Kikuchi A, Ito M, Nakamura S (1994) An induced mutant line lacking the α-subunit of β-conglycinin in soybean (*Glycine max* (L.) Merrill). Jpn J Breed 44(1):65–66

Takahashi M, Uematsu Y, Kashiwaba K, Yagasaki K, Hajika M, Matsunaga R, Komatsu K, Ishimoto M (2003) Accumulation of high levels of free amino acids in soybean seeds through integration of mutations conferring seed protein deficiency. Planta 217(4):577–586

Thanh V, Shibasaki K (1978) Major proteins of soybean seeds: subunit structure of β-conglycinin. J Agric Food Chem 26(3):692–695

Yagasaki K, Kalzuma N, Kitamura K (1996) Inheritance of glycinin subunits and characterization of glycinin molecules lacking the subunits in soybean (*Glycine max* (L.) Merr.). Jpn J Breed 46(1):11–15