Theor Appl Genet (2012) 124:365–372
DOI 10.1007/s00122-011-1711-8

ORIGINAL PAPER

Molecular analysis of glycinin genes in soybean mutants for development of gene-specific markers

Souframanien Jegadeesan · Kangfu Yu · Lorna Woodrow · Yi Wang · Chun Shi · Vaino Poysa

Received: 6 May 2011 / Accepted: 13 September 2011 / Published online: 29 September 2011

Abstract Soybean mutant lines that differ in 11S glycinin and 7S β-conglycinin seed storage protein subunit compositions were developed. These proteins have significant influence on tofu quality. The molecular mechanisms underlying the mutant lines are unknown. In this study, gene-specific markers for five of the glycinin genes (*Gy1* to *Gy5*) were developed using three 11S null lines, two A_4 null Japanese cultivars, Enrei and Raiden, and a control cultivar, Harovinton. Whereas gene-specific primers produced the appropriate products in the control cultivar for the *Gy1*, *Gy2*, *Gy3* and *Gy5* genes, they did not amplify in mutants missing the $A_{1a}B_2$, A_2B_{1a}, A_{1b} B_{1b}, and A_3B_4 subunits. However, ecotype targeting induced local lesions in genomes (EcoTILLING) and sequencing analysis revealed that the absence of the A_4 peptide in the mutants is due to the same point mutation as that in Enrei and Raiden. Selection efficiency of the gene-specific primer pairs was tested using a number of breeding lines segregating for the different subunits. Primer pairs specific to each of the *Gy1*, *Gy2*, *Gy3*, and *Gy5* genes can be used to detect the presence or absence of amplification in normal or mutant lines. The *Gy4* null allele can be selected for by temperature-switch PCR (TS-PCR) for identification of the A_4 (G4) null genotypes. In comparison to protein analysis by SDS-PAGE, gene-specific markers are easier, faster and more accurate for analysis, they do not have to use seed, and can be analyzed at any plant growth stage for marker-assisted selection.

Communicated by J. Ray.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1711-8) contains supplementary material, which is available to authorized users.

S. Jegadeesan · K. Yu (✉) · L. Woodrow · Y. Wang · C. Shi · V. Poysa
Greenhouse and Processing Crops Research Centre, Agriculture and Agri-Food Canada, 2585 County Road 20, Harrow, ON N0R 1G0, Canada
e-mail: kangfu.yu@agr.gc.ca

Present Address:
S. Jegadeesan
Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Maharashtra, Mumbai 400085, India

Introduction

Soybean seeds are distinguished from other legume grains by their especially high protein content, averaging 40% of total dry matter, and are therefore an important source of vegetable protein in both food and industry. The two major storage proteins, β-conglycinin and glycinin, corresponding to 7S and 11S globulins, account for about 70% of total soy seed protein (Koshiyama 1968; Derbyshire et al.1976). β-conglycinin is a glycol protein composed of three subunits, α (~67 kDa), α' (~71 kDa), and β (~50 kDa). In contrast, glycinin is a hexamer composed of six subunits and each subunit has an acidic (A) polypeptide linked by a disulfide bond to a specific basic (B) polypeptide (Staswick et al. 1984). The glycinin subunits were divided into Group I (G1, G2, G3 or $A_{1a}B_2$, A_2B_{1a}, A_{1b} B_{1b}) coded by *Gy1*, *Gy2*, *Gy3* genes; Group IIa (G4 or $A_5A_4B_3$) coded by *Gy4;* and Group IIb (G5 or A_3B_4) coded by *Gy5* (Nielsen et al. 1989; Yagasaki et al. 1996; Beilinson et al. 2002). Beilinson et al. (2002) identified and mapped two additional glycinin genes, namely a pseudogene, *gy6*, and a functional gene, *Gy7*, which encodes the sixth glycinin subunit, G7. For each subunit, there is more than 84% homology within a group and 45–49% between groups (Nielsen et al. 1989; Prak et al. 2005).

The quantity and quality of storage protein in soybean seed are the major biochemical components influencing the quality of tofu and other soy food products (Poysa and Woodrow 2002). Null soybean cultivars lacking the α′, and all the α, α′, and β-subunits of β-conglycinin, have been identified in Japanese soybean germplasm collections (Takahashi et al. 2003). Takahashi et al. (2003) have shown that null mutant soybeans lacking both glycinin and β-conglycinin storage components could grow and reproduce normally. This allows the possibility of genetically altering the entire subunit composition of glycinin and β-conglycinin of soybeans to improve the tofu quality. Poysa et al. (2006) developed a series of 28 new mutant soybean genotypes differing in seed storage glycinin and β-conglycinin subunit composition and tested their effects on tofu quality. Their results showed that Group IIb (A_3) glycinin played the major role in contributing to tofu firmness with any coagulant, while the Group IIa (A_4) subunit could have a negative effect on tofu quality. Soybeans with the Group I (A_1A_2) subunit resulted in tofu about one-third more firm than tofu prepared from soybeans without the Group I subunit. The individual components of Group I had contradictory effects on glucono-δ-lactone (GDL) coagulated tofu quality, with the A_1 subunit having a negative effect and A_2 having a major positive effect on tofu firmness. Lack of the α′ subunit of β-conglycinin increased gel hardness relative to the complete 7S protein. These mutant lines extend the range of genetic variability available to the breeder to improve soy protein functional properties for specific end uses.

Previous research with these mutant lines has focused on the effects of protein subunit composition on tofu quality (Poysa et al. 2006) and amino acid composition (Zarkadas et al. 2007). So far, the molecular mechanism underlying these mutant lines is largely unknown. Molecular changes that result in the loss of subunits have been reported for *Gy4* (Scallon et al. 1986; Yu et al. 2005) and *Gy3* (Cho et al. 1989) genes of glycinin and the α′-subunit gene of β-conglycinin (Ladin et al. 1984).

Breeder-friendly markers associated with soybean protein quality are in high demand. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of total seed storage protein is a standard method used by breeders to select for the presence or absence of the (A_4) peptide (Fontes et al. 1984). Although it is effective for A_4, it is not sufficient to separate G1, G2 and G3 subunits because of their similar molecular weights. In addition, breeders must wait until seeds are harvested for protein analysis. Alternatively, DNA markers can be tested at any stage of plant growth. Since glycinin gene sequences are available, gene-specific markers could be developed for *Gy1–Gy5* genes. Selection based on gene-specific markers would be more efficient than linked markers (Andersen and Lubberstedt 2003). Targeting Induced Local Lesions IN Genomes (TILLING) is a new technique that can identify single nucleotide polymorphisms (SNP) in a target gene by heteroduplex analysis. A variation of this technique ecotype targeting induced local lesions in genomes (EcoTILLING) presents a mean to determine the natural variation of genes in crops (McCallum et al. 2000; Comai et al. 2004; Banik et al. 2007). Temperature-Switch PCR (TS-PCR) is another new technique that can improve the speed and efficiency for SNP genotyping (Hayden et al. 2009). It is a biphasic three or four-primer PCR technique that permits amplification of the target locus in the first phase of thermal cycling before switching to the detection of the SNP alleles (Hayden et al. 2009). Combining EcoTILLING and TS-PCR could be a cost-effective approach for small scale SNP discovery and genotyping.

The objectives of this study were to (1) characterize the genes that underlie the absence of specific 11S protein subunits in three mutant lines, (2) develop gene-specific markers for *Gy1–Gy5* genes for use in marker-assisted selection (MAS), and (3) evaluate the specificity and efficiency of the *Gy1–Gy5* gene-specific markers in MAS with 19 breeding lines derived from the same source as the mutant lines.

Materials and methods

Materials

The soybean lines used in this study included the tofu-type cultivar, Harovinton, which has a complete complement of glycinin and β-conglycinin subunits; two previously characterized A_4 null lines, Enrei and Raiden; three mutant lines, SQ2-1 (A_1,A_2, A_3, A_4 null), SQ2-3 (A_1,A_2, A_3, A_4 null) and SQ3-1a (α′, A_1,A_2, A_3, A_4 null); and 19 F_5 breeding lines. The three mutant lines were developed by crossing and backcrossing Harovinton with a Japanese line lacking all glycinin subunits and the α′ subunit of β-conglycinin. The line lacking all 11S protein subunits was developed by crossing a γ-irradiation induced mutation line deficient for all Group I and IIa subunits with a Group IIb null wild soybean accession (Yagasaki et al. 1996; Kaizuma, personal communication). Seeds of Harovinton, Enrei, Raiden, the three mutant lines, and the 19 F_5 lines segregating for different glycinin subunits were grown in the greenhouse under normal conditions at the Greenhouse and the Processing Crops Research Centre (GPCRC), Harrow, Ontario in 2007.

SDS-PAGE

The protein compositions of the parent varieties (Harovinton, Enrei, and Raiden) and the three mutant lines (SQ2-1, SQ2-3 and SQ3-1a) were identified using sodium dodecyl

SDS-PAGE. SDS-PAGE was performed with 12.5% Tris–HCl Criterion™ precast gel with 4% (v/v) stacking gel in a BioRad mini-protean electrophoresis unit (BioRad Laboratories, Hercules, CA, USA). The electrophoresis buffer was 25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3. Samples were prepared by mixing 15 µg of ground soybean powder with 3 ml of sample buffer containing 10 mM Tris/HCl, 1 mM EDTA, pH 8.0, 2.5% SDS, 5.0% β-mercaptoethanol, 0.01% bromophenol blue in a plastic tube. Samples were mixed well and incubated in a boiling water bath for 5 min and cooled. 1 ml of the extract was transferred into a 1.5 ml microcentrifuge tube and centrifuged for 5 min. Aliquots of 10 µl of the prepared samples were loaded into each well. The electrophoretic separation was conducted at 200 V for 55 min. The gels were stained with Coomassie Blue solution for 60 min with constant shaking, and then destained twice with 50% (v/v) distilled deionized water, 40% (v/v) methanol and 10% (v/v) acetic acid solution for 1 h and stored in 10% acetic acid solution. The gels were scanned with a DeskJet 6300C scanner (Hewlett-Packard Canada Co., Mississauga, Canada).

Two-dimensional (2D) electrophoresis (2D-E)

Because SDS-PAGE could not effectively distinguish G1 ($A_{1a}B_2$), G2 (A_2B_{1a}) and G3 ($A_{1b}B_{1b}$) subunits, 2D-E was carried out to evaluate Harovinton, SQ2-3 and the 19 F_5 breeding lines (Supplementary Table S1). Proteins were extracted from finely ground soybean seeds with buffer containing 8 M urea, 4% CHAPS, 0.2% carrier ampholytes (125 each of Bio-Lyte 4/6 and Bio-Lyte 5/7, Bio-Rad), 0.0002% Bromophenol Blue and 50 mM dithiothreitol. Immobilized pH gradient strips containing the protein extract (IPG ReadyStrip 11 cm, pH 4–7 or 4.7–5.9, Bio-Rad) were re-hydrated overnight at room temperature in the sample extraction buffer. Isoelectric focusing was performed in Protean IEF cell (Bio-Rad) according to the manufacturer's instructions. After isoelectric focussing, IPG strips were equilibrated with 2% dithiothreitol followed by 2.5% iodoacetamide, each in 50 mM Tris/HCl buffer (pH 8.8) containing 6 M urea, 2% SDS and 20% glycerol. The second dimension, SDS-PAGE, was conducted using Criterion 12.5% polyacrylamide gels (Bio-Rad) and was followed by Coomassie Blue staining.

Gene-specific primer design

Gene-specific primers were designed based on genomic sequence information obtained from Genbank accessions X15121, X15122, X15123, X05651 and AB003680 and Soybase (http://soybase.org/) for *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5*, respectively. DNA sequences were aligned using Vector NTI Advance 11 (Invitrogen, Carlsbad, CA, USA). Overlapping primer pairs amplifying a fragment of 800–1,200 bp and covering the entire gene sequence were designed for each gene (Table 1). As the genes were highly homologous at the sequence level, gene-specific primers were selected based on the 3′ mismatches with other genes. Oligo design was carried out using GENE RUNNER version 3.05 (Hastings Software Inc., Las Vegas, NV, USA) program.

PCR amplification and EcoTILLING

The PCR amplifications were performed in a 25 µl reaction volume containing 50 ng template DNA using 0.5 units of *Taq* DNA polymerase (Sigma Aldrich Canada Ltd., Oakville, ON, Canada). Amplification conditions were 1 cycle at 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 52°C for 1 min, and 72°C for 2 min, and 1 final cycle at 72°C for 10 min. The amplified products were separated on a 1.5% agarose gel along with 1 Kb plus DNA ladder (Invitrogen, CA, USA) in 1X TAE buffer at 75 V. The gels were visualized under UV after staining with SYPRO orange dye bromide and were documented with a KODAK Electrophoresis documentation and analysis system (KODAK, Rochester, NY, USA).

EcoTILLING followed the procedures as described by Till et al. (2006). PCR amplification was carried out as described above. Heteroduplex PCR was performed by mixing 10 µl of PCR products of Harovinton (normal) and the null lines. After brief mixing, the PCR products were denatured at 99°C for 10 min and re-natured initially at 70°C for 20 s followed by 69 cycles where the temperature decreased by 0.3°C per cycle. Celery juice extract was extracted as described by Till et al. (2006). EcoTILLING on an agarose gel was performed using the method described by Chitra et al. (2007).

Reverse transcriptase PCR (RT-PCR)

Total RNA was isolated from immature seeds (0.5 g/seed fresh weight) using the ultra clean plant RNA isolation kit (MO BIO, Carlsbad, CA, USA) and the cDNA was synthesized using the Superscript II kit (Invitrogen, Carlsbad, CA, USA). The cDNA was then amplified by PCR with Sigma *Taq* DNA polymerase (Sigma Aldrich Canada Ltd., Oakville, ON, Canada) and primer pairs specific to *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5* (Table 2). The Soybean Actin 3 (*SAc3*) gene was used as a control. The primer pairs specific to *SAc3* used in this study were described earlier by Tsubokura et al. (2006). RT-PCR conditions for the Actin gene were the same as those described above for the glycinin genes.

Table 1 List of gene-specific primer sets covering each of the entire *Gy1*-*Gy5* glycinin genes

Gene	Primer names	Forward 5′–3′	Reverse 5′–3′	Product size (bp)
Gy1 gene	*Gy1-1*	AAGTGTACGTGGTTAACATTAG	CCCTCATTTATCAAACCC	801
	Gy1-2	CATCTTGTCCAAACTTCAC	CTGTGGCTTCTCATCCTC	1,097
	Gy1-3	TTAAAATATGCCAATTGA	GTACATCAATGCATGCAAG	1,127
	Gy1-4	GGAGTTATCTTAACACTTGC	TCCTGTCCTGCCAACAC	906
Gy2 gene	*Gy2-1*	ATGATATGTAAACCAAAGC	GTTGCTGAGATTCTTGCG	1,133
	Gy2-2	CTTCTTAACCGTCCAACC	GTGCTAAGATAACTTTGTC	1,335
	Gy2-3	GGTACGTACATCATACATGT	TATTTACTGCTACCAGCAC	636
	Gy2-4	GGAGAGCTGTGGCTTAGAGAG	CAGACGATGACATGGAGC	1,134
Gy3 gene	*Gy3-1*	GAAAGGAAAGAAAGAAGC	ACATTATCGATGTCCTTAACG	995
	Gy3-2	CTTGCGGGATTTATCAAC	GCTTCTGTGGCTGATACTG	864
	Gy3-3	CGTGAGCCACATATATAGC	TCTCATTCCTTAACCAATATCC	1,059
	Gy3-4	AAATATGCGCTGTATGTC	AACATGTCACGAAATTCTAC	1,053
Gy4 gene	*Gy4-1*	TTTACAGAGACAATTATGCT	GATATGTCAACAAAGTCATTC	1,108
	Gy4-2	AACATCACATTACCATAGAT	TCTTTTTCGATCTCCATTC	1,091
	Gy4-3	CGTGAACAAGACCAGGACCAG	TCATGCTTGTCATTTATGCGAC	1,216
Gy5 gene	*Gy5-1*	TCAATATGGGGAAGCCCTTCTTC	ATCACTGCTCATTATAATTACG	943
	Gy5-2	GGAGCAATTGGATTTGCATT	CCTCGTTGGTGTTGAAGGAT	864
	Gy5-3	GACGAAGAATATGAACAAACT	CATAACCTCTAAAAATCAGTTATC	638
	Gy5-4	ATCATAATTCATAGGTCATATACG	TTTCGTTAATAGACAGACATCTC	944

Table 2 Gene-specific primers used in the reverse transcriptase polymerase chain reaction (RT-PCR) gene expression experiments

Primer name	Forward 5′–3′	Reverse 5′–3′	Product size (bp)
SAc3[a]	CGACCTCGACATACTGGTGTTAT	TGCCATATAGATCCTTTCTGATA	778
*Gy1*cDNA	GTATAGAGTCAGAAGGAGGGC	CTTCTTCTTCTTCCTCTTCCTGGG	720
*Gy2*cDNA	TGAGTGCCAGATCCAAAAGCTG	TGGCTTCCTCATGGCTGGAGCTG	730
*Gy3*cDNA	GCAAAACGAGTGCCAGATCCAACG	TCTGGCTTCTCCTCTTCCTCGGG	765
*Gy4*cDNA	TATCCCCGGATGATCATCATCGCC	TTCCTTGGCTTGGTCGACTAG	727
*Gy5*cDNA	CGTTGAGTCCGAAGGTGGTC	TCGGGCCTGCTTGGTCGCTGT	855

[a] Soybean actin gene was used as a control

Temperature-switch PCR (TS-PCR)

TS-PCR assay was performed as described by Hayden et al. (2009) with little modification. The reaction volume was kept at 6 µl with 50 ng of DNA without drying the samples. Locus-specific primers were designed for *Gy4* gene SNP (TTCCTTAGTTCAATAT(G/A) GGGAAGCCCTTC ACT) reported by Yu et al. (2005). The locus-specific primers were positioned such that the forward primer on the same template DNA strand as the allele-specific primer was located at least 100 nucleotides away from the SNP. Forward primer *Gy4*LS1F (5′-GAGGGGACCATTCAGAGTGCC-3′) and the reverse primer *Gy4*LS1R (5′-GCAAGTGGAGGCCAT TGCGG-3′) were designed, respectively, 248 and 224 bp away from the SNP location using GENE RUNNER program (Table 3). The allele-specific forward primer, *Gy4*a1F (5′-GGCACTTCCTTAGTTCAA TATA-3′), was designed to be complementary to the A_4 null allele (**A**) present at the locus.

Sequencing

Genomic sequences of glycinin genes were amplified with their respective gene-specific primer pairs. PCR products were purified using QIAquick PCR Purification Kit (QIAGEN Inc. Mississauga, ON, Canada). PCR fragments were cloned in pGEM-T Easy Vector System-I (Promega, Madison, WI, USA). The specificity of each gene-specific primer pair was ascertained by sequencing (Roberts Research Institute at the University of Western Ontario, London, ON, Canada) the cloned PCR fragments and

Table 3 The gene-specific primers used for screening of A_{1a}, A_2, A_{1b}, A_3, and A_4 glycinin subunits

Subunit	Forward 5′–3′	Reverse 5′–3′	Band size (bp)
A_{1a} (*Gy1*)	TCTTCCTGGGGTCTTTGTTG	TCTTCCTGGGGTCTTTGTTG	766
A_2 (*Gy2*)	TTCTTAACCGTCCAACCTATCTT	GGGCACTGAGTTTGAGAAGC	1,211
A_{1b} (*Gy3*)	TTGACACCAACAGCTTCCAG	TACAGCGCATATTTGCACCA	1,037
A_3 (*Gy5*)	GGAGCAATTGGATTTGCATT	CCTCGTTGGTGTTGAAGGAT	864
A_4 (*Gy4*)[a]	GAGGGGACCATTCAGAGTGCC	GCAAGTGGAGGCCATTGCGG	514
	GGCACTTCCTTAGTTCAATATA		266

[a] TS-PCR primers, The locus-specific primer pairs amplify a 514 bp fragment, and the allele-specific primer and reverse locus-specific primer amplify a 266 bp fragment. The underlined bases in the allele-specific primers are the addition of two base pairs at the 5′ end and the SNP at the 3′ end, respectively

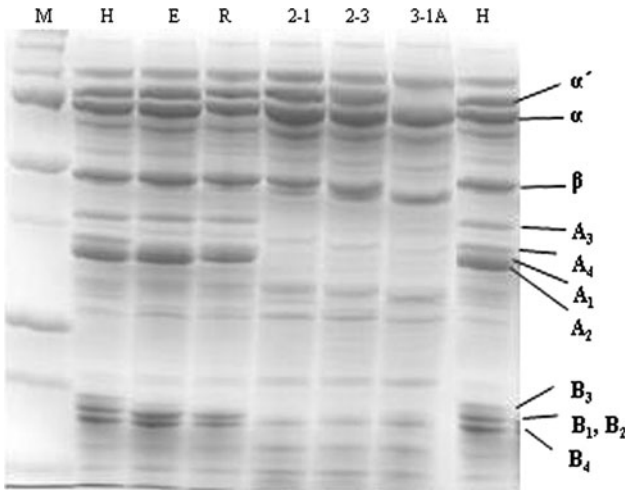

Fig. 1 Analysis of the storage proteins by SDS-PAGE. The α′, α and *β* are subunits of *β*-conglycinin. A_1, A_2, A_3 and A_4 and B_1, B_2, B_3, and B_4 are glycinin acidic and basic subunits, respectively. *Lanes M* Protein Ladder, *H* Harovinton, *E* Enrei, *R* Raiden, 2-1, 2-3, and 3-1A are mutant lines SQ2-1, SQ2-3, and SQ3-1a, respectively

aligning the sequences to each corresponding glycinin gene sequence in Vector NTI Advance 11 (Invitrogen, Carlsbad, CA, USA).

Results

Analyses of seed protein variants by SDS-PAGE and 2D-PAGE

The globulin fractions extracted from seeds were resolved by SDS-PAGE. The fractions obtained from the commercial variety, Harovinton, possessed all the α, α′, and *β* subunits of *β*-conglycinin as well as all acidic and basic polypeptide components of glycinin. The two A_4 null cultivars, Enrei and Raiden, lacked the A_4 glycinin subunit as expected and the three mutant lines, SQ2-1, SQ2-3 and SQ3-1a, each lacked the A_1, A_2, A_3, and A_4 glycinin subunits. SQ3-1a also lacked the α′,*β*-conglycinin subunit (Fig. 1). To achieve better separation among G1, G2, and G3 subunits, 2D-E analyses were used to characterize each of the glycinin subunits, $A_{1a}B_2$, $A_{1b}B_{1b}$, and A_2B_{1a}. Figure 2 illustrates the representative differences of Group I (A_{1a}, A_2, A_{1b},) and Group II (A_4, A_3) glycinin acidic polypeptides between the SQ2-3 mutant line and the control, Harovinton. The presence or absence of specific acidic subunits of the glycinin proteins are indicated by arrows (Fig. 2).

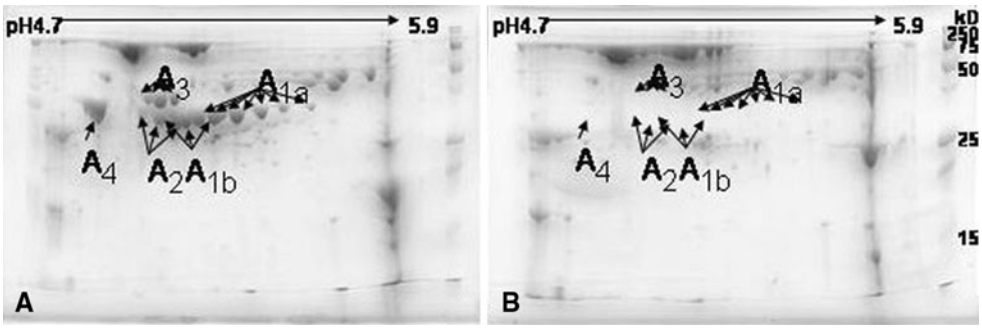

Fig. 2 Examples of a two-dimensional gel electrophoresis (2D-GE) analysis of the glycinin proteins of null lines. The first dimension was run on a pH 4–6 linear pH gradient. The second dimension was a 12% SDS-PAGE. Molecular masses are shown on the right. The specific glycinin protein spots are indicated by *arrows*. **a** 2D-GE of Harovinton, **b** 2D-GE of the SQ2-3 mutant line

Amplification of glycinin genes

Despite the high DNA sequence homology among Group I and Group II genes, gene-specific primer pairs were designed for *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5* genes. Four primer pairs covering the entire gene sequences of each of *Gy1*, *Gy2*, *Gy3*, and *Gy5* genes were designed (Table 1). For the *Gy4* gene, only three primer pairs were needed. All the primer pairs specific to each of *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5* showed amplification in the control cultivar, Harovinton. No amplification of any of the primer pairs specific to *Gy1* or *Gy2*, however, was observed in the mutant lines (Supplementary Fig. S1a and S1b). Three primer pairs for *Gy3* and two primer pairs for *Gy5* at their 5′ end also had no amplification in the mutant lines, which could be due to partial indels of the *Gy3* and *Gy5* genes or sequence variations on the primer binding sites in these lines (Supplementary Fig. S1c and S1e).

All three primer pairs specific to *Gy4* showed amplification in Harovinton and the three mutant lines (Supplementary Fig. S1d). EcoTILLING procedures revealed the presence of a SNP in the *Gy4* gene amplified by the *Gy4*-F1 and *Gy4*-R1 (Table 1) primer pair (Fig. 3). Cloning and sequencing of these fragments revealed that the three mutant lines (SQ2-1, SQ2-3 and SQ3-1a) have the same SNP involving one base pair substitution of deoxyguanosine (G) by deoxyadenine (A) at the initiation codon in the *Gy4* gene (data not shown). cDNA from Harovinton and the three mutants was used for RT-PCR amplification using cDNA specific primers from *Gy1*, *Gy2*, *Gy3*, *Gy4* and *Gy5* (Table 2). All cDNA specific primer pairs showed amplification in Harovinton, however, amplification was not observed in the mutant lines possibly due to the absence of gene expression or mRNA (data not shown). Sequencing of the RT-PCR amplicon from each gene in the control was also carried out to confirm the gene-specific amplification (data not shown).

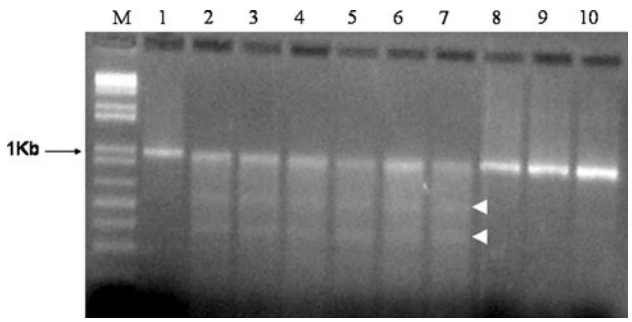

Fig. 3 EcoTILLING of the PCR amplicon from *Gy4* gene using *Gy4-1* primer pairs. The PCR products were subjected to heteroduplex formation and digested with CEL I (*lanes 2*, *3*, *4*) or celery juice extract (*lanes 5*, *6*, *7*). *Lane M* is molecular weight standard. *Lanes 1*, *8*, *9*, *10* are control digestions (homoduplex). *Arrows* indicate the presence of SNP in the mutant lines

PCR and 2D-E analysis using F_5 individuals

PCR and 2D-E analyses were conducted using 19 F_5 individuals obtained from crosses involving the three mutants, SQ2-1, SQ2-3 and SQ3-1a, and Harovinton (Supplementary Table S1). Genotypes at the *Gy1*, *Gy2*, *Gy3*, and *Gy5* loci in the F_5 individuals were determined by PCR using primers specific for each of the *Gy* genes (Table 3). The segregation patterns of A_1 to A_5 glycinin subunits as determined by 2D-E matched perfectly with the PCR results (Supplementary Table S1). *Gy1*, *Gy2*, *Gy3* and *Gy5* primer pairs acted like sequence characterized amplified region (SCAR) markers because they showed either presence or absence of a specific PCR fragment from its respective gene (Fig. 4).

The SNP found by Yu et al. (2005) in the *Gy4* gene was tested using TS-PCR. The SNP allele complementary to the null allele-specific primer produced the 'reference SNP allele' product (smaller PCR product), while those without complement produced the 'alternate SNP allele' product (larger PCR product) (Fig. 5). However, in the present study both bands were observed for the allele in the mutants, indicating the poor specificity of the allele-specific primer for the target SNP. TS-PCR clearly differentiated the normal and null allele for the *Gy4* gene. Thus, the gene-specific primer pairs described here can be used for genotyping

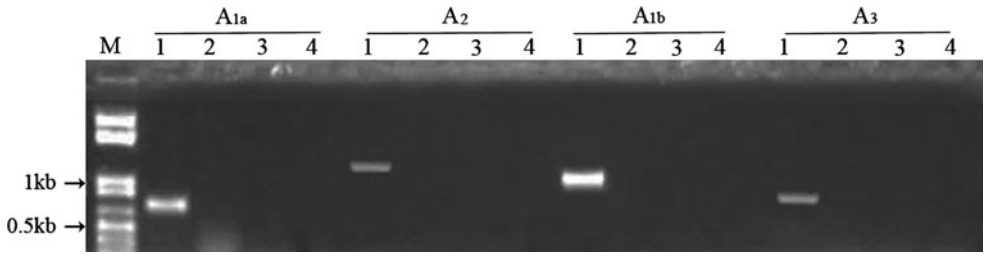

Fig. 4 PCR amplification of genes encoding G1 (A_{1a}), G2 (A_2), G3 (A_{1b}) and G5 (A_3) subunits in Harovinton and the three mutant lines with gene-specific primers that can be used for marker-assisted selection. *Lane M* is the molecular weight standard, *lanes 1*, *2*, *3*, and *4* are Harovinton, SQ2-1, SQ2-3 and SQ3-1a, respectively

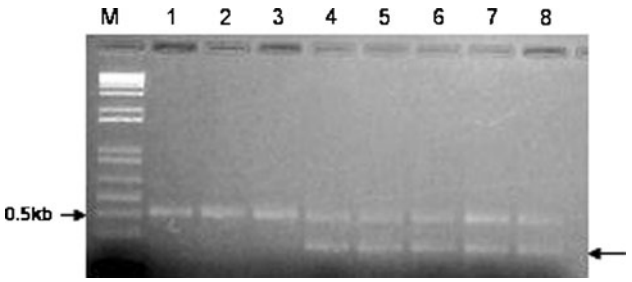

Fig. 5 Temperature-switch PCR assay performed using *Gy4* null allele and locus-specific primer pairs. *Lane M* is molecular weight standard. *Arrow* on the right indicates the band specific to A4 null allele (A). *M* molecular weight standard, *Lanes 1* Harovinton, *2* OX744, *3* Williams, *4* Enrei, *5* Raiden, *6* SQ2-1, *7* SQ2-3, *8* SQ3-1a

Gy1, *Gy2*, *Gy3*, *Gy4* and *Gy5* genes encoding the $A_{1a}B_2$; $A_{1b}B_{1b}$; A_2B_{1a}; $A_5A_4B_3$ and A_3B_4 subunits of glycinin.

Discussion

Molecular mechanisms underlying the null lines (Enrei, Raiden, SQ2-1, SQ2-3 and SQ3-1a) lacking 11S protein subunits were investigated in this study. A mutation in the initiation codon for the *Gy4* and chromosomal rearrangements for the *Gy3* genes were previously reported (Scallon et al. 1986; Cho et al. 1989; Diers et al. 1994). Lack of *Gy1*, *Gy2*, *Gy3* and *Gy5* gene-specific amplification in the mutant lines (S-Fig. 1a–c, e) indicated that gene indels or sequence variations could be responsible for the absence of the protein subunits. RT-PCR indicated that the lack of expression in the glycinin genes occurred at the transcriptional level (data not shown). According to genetic map in Soybase (http://soybase.org/) the two gylcinin genes, *Gy1* and *Gy2*, are linked in the same orientation about 1.5 kb apart in one domain, whereas the *Gy3* glycinin gene is located in another domain. All the primer pairs designed for *Gy1* and *Gy2* were able to amplify in the control cultivar, but failed to amplify in the three mutant lines, possibly due to gene indels or sequence variations in this domain. Of the four primer pairs used for amplifying *Gy3*, only one primer at the 3′ end showed amplification in the mutant lines, which could be due to partial indels or sequence variations in the gene. Similarly for *Gy5*, two of the four primer pairs at the 3′ end showed amplification in the mutant lines indicating that partial gene indels or sequence variations may be responsible for the null phenotype. For *Gy4* gene, a single base pair substitution is responsible for the loss of the protein subunit. This was the same change as reported earlier for Raiden and Enrei (Scallon et al. 1986; Yu et al. 2005).

EcoTILLING and TS-PCR were combined for SNP discovery and genotyping. Firstly, the EcoTILLING technique was employed to identify the point mutations underlying the glycinin protein subunits. Considering that high levels of variation might exist in diverse genotypes, pooling more than two samples was avoided (Chitra et al. 2007). Each mutant genotype was contrasted separately against a reference genotype, Harovinton, which has all the protein subunits. No amplification for Group I genes was observed in the mutants, possibly due to gene indels or sequence variations. All the primer pairs for Group II genes (*Gy4* and *Gy5*) showed amplification except two primer pairs at the 5′ end of the *Gy5* gene. A SNP was observed only for the *Gy4* gene. The SNP site was confirmed by sequencing. TS-PCR was employed to differentiate null and normal alleles for the *Gy4* gene. TS-PCR has been used to develop allele-specific PCR markers for co-dominant SNP genotyping in barley (Hayden et al. 2009). An allele-specific primer was designed specific to the A_4 null SNP (from G to A), and hence a band was observed for null alleles. The SNP allele complementary to the allele-specific primer produced the 'reference SNP allele' product (smaller band), while those without complement produced the 'alternate SNP allele' product (larger band) (Fig. 5). The poor specificity of the allele-specific primer resulted in the amplification of both bands in the mutants with target SNP. Hayden et al. (2009) reported that such SNP allele specificity problems can be solved by targeting the alternate SNP allele. The allele-specific primer for the alternate allele was tested, but it did not resolve the problem of specificity. Nevertheless, the TS-PCR primer pair designed and tested in this study can still be used for the identification of the *Gy4* null allele by the presence or absence of the 'reference allele' in a dominant fashion.

Currently, SDS-PAGE analysis of total seed storage protein is a standard method used by breeders to select superior lines when breeding to improve tofu quality through modification of the protein subunit profile (Poysa et al. 2006). It requires seeds for protein extraction and analysis, thus breeders have to wait until the seeds are harvested to select the lines in the following generation. Furthermore, SDS-PAGE cannot efficiently distinguish among A_{1a}, A_2 and A_{1b} subunits. Although 2D-E can be used for separating those subunits, it is tedious, expensive and time consuming. With the advent of DNA marker technology, DNA-based molecular markers have become more popular than protein markers. Previous studies have reported markers linked with glycinin genes. RFLP markers associated with *Gy4* and *Gy5* glycinin genes were mapped in linkage group O and F (Diers et al. 1994; Chen and Shoemaker 1998). Panthee et al. (2004) found three glycinin QTL associated SSR markers (Satt461, Satt292 and Satt156) that were distributed on linkage group D2, I and L, respectively, whereas two β-conglycinin QTL associated SSRs (Satt461 and Satt249) were distributed on LG D2 and J. However, gene-specific markers are derived from the genes involved in phenotypic trait variation, they can more accurately and

efficiently identify the alleles in a population, thereby giving them an advantage over linked markers (Andersen and Lubberstedt 2003). Respective gene-specific markers were developed for *Gy1–Gy5* genes (Table 3). These markers matched the 2D-E results perfectly for the lines analysed in this study and should permit efficient MAS for improved protein subunit profiles for tofu quality and other functional properties of soy.

In conclusion, the absence of the G_1–G_5 glycinin subunits in the three mutant lines studied probably resulted from gene indels or sequence variations in the Group I genes (*Gy1*, *Gy2*, *Gy3*) and the Group IIb gene (*Gy5*) and a point mutation in the *Gy4* gene. The five gene-specific markers reported in this study for *Gy1–Gy5* can be used to replace protein markers for MAS. Since the *Gy1*, *Gy2*, *Gy3* and *Gy5* markers are dominant-recessive PCR markers similar to SCAR markers, and the *Gy4* marker is a TS-PCR marker, they will be a time- and cost-effective method for MAS of soybean varieties with superior combinations of 11S glycinin subunits for improved tofu quality or other protein functionality. The successful detection of the SNP with EcoTILLING and TS-PCR demonstrated that combining both techniques might be a cost-effective approach for the identification and application of SNP markers in a breeding program of limited budget.

Acknowledgments The authors would like to thank Ontario Soybean Growers and Agriculture and Agri-Food Canada for their financial support. The technical assistance of Dale Anderson, Barb Harwood and Bob Armstrong is acknowledged as well.

References

Andersen JR, Lubberstedt T (2003) Functional markers in plants. Trends Plant Sci 8:554–560

Banik M, Liu S, Yu K, Poysa V, Park SJ (2007) Molecular TILLING and EcoTILLING: effective tools for mutant gene detection in plants. Genes Genomes Genomics 1(2):123–132

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

Chen Z, Shoemaker RC (1998) Four genes affecting seed traits in soybean map to linkage group F. J Hered 89:211–215

Chitra R, Naredo MEB, Wang H, Atienza G, Liu B, Qiu F, McNally KL, Leung H (2007) RAPID method for detecting SNPs on Agarose gels and its application in candidate gene mapping. Mol Breed 19:87–101

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

Comai L, Young K, Till BJ, Reynolds SH, Green EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Henikoff S (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. The Plant J 37:778–786

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

Diers BW, Beilinson V, Nielsen NC, Shoemaker RC (1994) Genetic mapping of the *Gy4* and *Gy5* glycinin genes in soybean and the analysis of a variant of *Gy4*. Theor Appl Genet 89:297–304

Fontes E, Moreira M, Davies C, Nielsen NC (1984) Urea elicited changes in relative electrophoretic mobility of certain glycinin subunits. Plant Physiol 76:840–842

Hayden MJ, Tabone T, Mather DE (2009) Development and assessment of simple PCR markers for SNP genotyping in barley. Theor Appl Genet 119:939–951

Koshiyama I (1968) Chemical and physical properties of a 7S protein in soybean globulins. Cereal Chem 45:393–404

Ladin BF, Doyle JJ, Beachy RN (1984) Molecular characterization of deletion mutation affecting the α' subunits of β-conglycinin. J Mol Appl Genet 2:372–389

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genome (TILLING) for plant functional genomics. Plant Physiol 123:439–442

Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB (1989) Characterization of the glycinin gene family in soybean. Plant Cell 1:313–328

Panthee DR, Kwanyuen P, Sams CE, West DR, Saxton AM, Pantalone VR (2004) Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. J Am Oil Chem Soc 81:1005–1012

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

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

Prak K, Nakatani K, Katsube T, Adachi M, Maruyama N, Utsumi S (2005) Structure–function relationships of soybean proglycinins at subunit levels. J Agric Food Chem 53:3650–3657

Scallon BJ, Dickinson CD, Nielsen NC (1986) Characterization of a null-allele for the *Gy4* glycinin gene from soybean. Mol Gen Genet 208:107–113

Staswick PE, Hermodson MA, Nielsen NC (1984) The amino acid sequence of the A_2B_{1a} subunit of glycinin. J Biol Chem 259:13424–13430

Takahashi M, Uematsu Y, Kashiwaba Y, 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 217:577–586

Till BJ, Zerr T, Comai L, Henikoff S (2006) A protocol for TILLING and Ecotilling in plant and animals. Nat Protoc 1:2465–2477

Tsubokura Y, Hajika M, Harada K (2006) Molecular marker associated with β-conglycinin deficiency in soybean. Breed Sci 56:113–117

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

Yu K, Poysa V, Haffner M, Zhang B, Woodrow L (2005) Absence of the A4 peptide in the G4 glycinin subunit of soybean cultivar Enrei is caused by a point mutation in the *Gy4* gene. Genet Mol Biol 28:440–443

Zarkadas CG, Gagnon C, Poysa V, Khanizadeh S, Cober ER, Chang V, Gleddie S (2007) Protein quality and identification of the storage protein subunits of tofu and null soybean genotypes, using amino acid analysis, one- and two-dimensional gel electrophoresis, and tandem mass spectrometry. Food Res Int 40:111–128