



Genetic mapping and validation of the loci controlling 7S α' and 11S A-type storage protein subunits in soybean [*Glycine max* (L.) Merr.]

Jeffrey D. Boehm Jr.¹ · Vi Nguyen² · Rebecca M. Tashiro¹ · Dale Anderson³ · Chun Shi³ · Xiaoguang Wu³ · Lorna Woodrow³ · Kangfu Yu³ · Yuhai Cui² · Zenglu Li¹

Received: 28 August 2017 / Accepted: 21 November 2017 / Published online: 9 December 2017
© Springer-Verlag GmbH Germany, part of Springer Nature 2017

Abstract

Key message Four soybean storage protein subunit QTLs were mapped using bulked segregant analysis and an F₂ population, which were validated with an F₅ RIL population.

Abstract The storage protein globulins β -conglycinin (7S subunit) and glycinin (11S subunits) can affect the quantity and quality of proteins found in soybean seeds and account for more than 70% of the total soybean protein. Manipulating the storage protein subunits to enhance soymeal nutrition and for desirable tofu manufacturing characteristics are two end-use quality goals in soybean breeding programs. To aid in developing soybean cultivars with desired seed composition, an F₂ mapping population ($n = 448$) and an F₅ RIL population ($n = 180$) were developed by crossing high protein cultivar ‘Harovinton’ with the breeding line SQ97-0263_3-1a, which lacks the 7S α' , 11S A₁, 11S A₂, 11S A₃ and 11S A₄ subunits. The storage protein composition of each individual in the F₂ and F₅ populations were profiled using SDS-PAGE. Based on the presence/absence of the subunits, genomic DNA bulks were formed among the F₂ plants to identify genomic regions controlling the 7S α' and 11S protein subunits. By utilizing polymorphic SNPs between the bulks characterized with Illumina SoySNP50K iSelect BeadChips at targeted genomic regions, KASP assays were designed and used to map QTLs causing the loss of the subunits. Soybean storage protein QTLs were identified on Chromosome 3 (11S A₁), Chromosome 10 (7S α' and 11S A₄), and Chromosome 13 (11S A₃), which were also validated in the F₅ RIL population. The results of this research could allow for the deployment of marker-assisted selection for desired storage protein subunits by screening breeding populations using the SNPs linked with the subunits of interest.

Communicated by Brian Diers.

Kangfu Yu, Yuhai Cui and Zenglu Li contributed equally.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00122-017-3027-9>) contains supplementary material, which is available to authorized users.

✉ Yuhai Cui
yuhai.cui@agr.gc.ca

✉ Zenglu Li
zli@uga.edu

¹ Department of Crop and Soil Sciences, Institute for Plant Breeding, Genetics, and Genomics, University of Georgia, Athens, GA 30602, USA

² Agriculture and Agri-Food Canada, London Research and Development Centre, London, ON N5V 4T3, Canada

³ Agriculture and Agri-Food Canada, Harrow Research and Development Centre, Harrow, ON N0R 1G0, Canada

Introduction

Soybean (*Glycine max* L. Merr.) is the number one oilseed crop produced in the world and one of the most widely grown field crops, with over 340 million metric tons produced in 2016 (American Soybean Association 2017). Soybean is grown for both oil production and as a protein source, with the seeds yielding on average 20% oil and 40% protein (Krober and Cartter 1962). With such a high protein concentration, soybean protein products account for roughly two-thirds of the value of soybeans, leading many to call soybean a protein crop, not an oilseed crop (Johnson and Smith 2003). Soybean protein may be found in many food ingredients, such as tofu, but it is mainly used for the production of soymeal, a common feed ingredient of swine (*Sus domesticus*) and poultry (*Gallus domesticus*). Soymeal provides a rich source of digestible amino acids for these two nonruminant species, closely matching their amino acid requirements for survival (Cromwell 1999). Harvested

soybeans must first undergo a solvent extraction process to remove the oil. Afterwards, the remaining residues, known collectively as the soybean meal, are heated to a specific temperature, and the hulls may be blended back in, thereby creating “conventional soybean meal” with ~ 43 to 44% protein, or the hulls may be left out entirely, leaving “dehulled soybean meal” with ~ 47 to 49% protein (Soyatech 2017; Cromwell 1999; Zarkadas et al. 2007b). The prevalence of soybean as a protein staple in the livestock sector has led to much research into the protein components, particularly the storage proteins.

More than 70% of soybean seed protein is comprised of storage proteins made up of 7S and 11S globulins (Thanh and Shibasaki 1976; Tsukada et al. 1986). The percentage that each storage protein contributes to total protein concentration varies, and may be influenced by genotype, and environment (Paek et al. 1997; Fehr et al. 2003; Zarkadas et al. 2007a). The 7S globulin, also known as β -conglycinin or vicilin, makes up about 30% of the protein found in soybean seed (Paek et al. 1997). The 7S globulin protein is a trimer consisting of the subunits α , α' , and β that can combine in various manners to form different isomers, with molecular weights of 76, 72, and 52 kDa, respectively (Thanh and Shibasaki 1978). The β subunit is known to have allergenic effects (Wang et al. 2014; Banaszkiwicz 2011). The α' subunit is encoded by a single gene, while the α and β subunits of 7S globulin are encoded by different gene families, each containing multiple genes (Harada et al. 1989; Li and Zhang 2011). There are two homologous mRNA families differing in size that code for β conglycinin subunits: α/α' are coded by 2.5 kb RNA, whereas subunit β is coded by 1.7 kb RNA (Goldberg et al. 1981; Beachy et al. 1981; Schuler et al. 1982). Harada et al. (1989) reported a total of 15 genes, named *CG-1*–*CG-15* that encode 7S β conglycinin with *CG-1* assigned to the α' subunit [also denoted as *GM7S* using gene nomenclature adopted from NCBI (<http://www.ncbi.nlm.nih.gov/gene/>)], *CG-2* and *CG-3* assigned to the α subunit, and *CG-4* assigned to the β subunit. On the basis of cloned gene fragments that are hybridized to leaf nuclear DNA, Harada et al. (1989) also assigned five other genes to the β subunit (*CG-8*, *11*, *12*, *13*, and *15*) and four other genes to both the α and β subunits (*CG-5*, *6*, *7*, and *14*) that were ordered by DNA gel blots. Genes *CG-8*, *11*, *12*, *13*, and *15* are highly homologous to *CG-4* and code for unique β subunits, whereas *CG-5*, *6*, *7*, and *14* are highly homologous to both the mRNA families (Singh et al. 2013). The 11S globulin, also known as glycinin or legumin, can make up about 50% of the total soybean seed protein (Poysa et al. 2006). The 11S globulin is a hexameric protein consisting of a combination of six subunits that contain both an acidic (A) and basic (B) polypeptide chain (Nielsen et al. 1989). The six genes that encode the 11S globulin subunits comprise three different gene families, with *Gy1* ($A_{1a}B_2$),

Gy2 (A_2B_{1a}), and *Gy3* ($A_{1b}B_{1b}$) making up one gene family (Group I), *Gy4* ($A_5A_4B_3$) and *Gy5* (A_3B_4) making up the second family (Group II), and *Gy7* (polypeptide groups not assigned) and two pseudogenes (*gy6* and *gy8*) creating the third family (Group III) (Nielsen et al. 1989; Beilinson et al. 2002; Li and Zhang 2011). The Group I and Group II *Glycinin* genes have been successfully cloned and sequenced (Fischer and Goldberg 1982; Scallan et al. 1985; Nielsen et al. 1989). Fischer and Goldberg (1982) first reported and characterized the Group I glycinin genes *G₁*, *G₂* and *G₃*, which were later denoted as *Gy1*, *Gy2* and *Gy3* by Davies et al. (1987). Scallan et al. (1985) are credited for first identifying the clones of Group II glycinin genes.

Mutations causing the loss of the 7S and 11S subunits have led to the development of soybean lines lacking some or many of the storage protein subunits. The Japanese soybean cultivar ‘Keburi’ lacks the 7S α' subunit gene (Poysa et al. 2006), and ‘Yumeminori’ is a Japanese cultivar derived by γ -irradiation that lacks the 7S α and α' subunits, and has reduced β subunit concentration (Takahashi et al. 1994, 2004). Complete 7S α and β subunit null lines were found to have a lethal chlorosis phenotype at the seedling stage (Liu et al. 2006), so these null lines are impractical for breeding programs. A recessive gene, *cgdef* that mapped to Chromosome (Chr) 19 [Linkage Group (LG) L], was identified in a γ -ray mutation study to suppress 7S subunit production, but the mutation has other undesirable growth traits not suitable for introgression in a breeding program (Hayashi et al. 2009; Tsubokura et al. 2012). A single dominant gene, *Scg-1* that mapped to Chr 20 (LG I), was found to suppress the transcription of all 7S subunit genes without affecting the rest of the plant (Teraishi et al. 2001). Wang et al. (2014) found that the β subunit was controlled by a single dominant locus *qBSC-1* on Chr 20 (LG I), although Panthee et al. (2004) identified β subunit quantitative trait loci (QTLs) on soybean Chrs 3 (LG N), 16 (LG J), and 17 (LG D2), indicating that the genetic architecture underlying the 7S β -conglycinin is complicated.

Within the 11S subunit gene family, Beilinson et al. (2002) mapped the Group I genes *Gy1* to Chr 19 (LG L), *Gy2* to Chr 3 (LG N) and *Gy3* to Chr 19 (LG L), but in his populations, *Gy1* and *Gy2* were linked to two independently segregating loci on two different linkage groups. This contradicted the previous reports that the Group I glycinin genes occupy two chromosomal domains, with one domain consisting of a linked *Gy1* and *Gy2*, and the other consisting of *Gy3* (Nielsen et al. 1989). Li and Zhang (2011) confirmed the work of Cho et al. (1989) in concluding that *Gy1* and *Gy2* (with the newly characterized pseudo gene *gy6*) are tandemly linked one after another on Chr 3 (LG N), and also reported that *Gy3* and *Gy7* are located on Chr 19 (LG L). Other studies have successfully used restriction fragment length polymorphism (RFLP) markers to show that the *Gy4*

and *Gy5* glycinin genes are located on Chrs 10 (LG O) and 13 (LG F), respectively (Diers et al. 1993; Chen and Shoemaker 1998). There has been a concentrated breeding effort that focused on deleting the *Gy4* gene, because presence of the *A₄* Group II subunit was shown to have a negative effect on tofu quality and coagulation (Jegadeesan et al. 2012). Roughly 20% of all Japanese soybean lines have mutations in the *Gy4* gene (Yu et al. 2005; Jenkinson and Fehr 2010). There is a point mutation in *Gy4* leading to the alteration of the initiation codon from ATG to ATA, which is found in the cultivars ‘Enrei’ and ‘Raiden’ (Yu et al. 2005), and eight North American *Gy4* mutant accessions (Kim et al. 2008). Three soybean breeding lines with high sulfur concentration were recently released by Krishnan et al. (2015) with each lacking the *Gy4* gene. A Japanese line ‘Nanohare’ was also developed, that is missing all the 11S subunits (Tsubokura et al. 2012).

Herein, breeding line SQ97-0263_3-1a, lacking the 7S α' subunit (*Gm7S*) and each of the 11S subunits [11S *A₁* (*Gy1*), 11S *A₂* (*Gy2*), 11S *A₃* (*Gy5*) and 11S *A₄* (*Gy4*)], as determined using SDS-PAGE analysis (Zarkadas et al. 2007b), was crossed with ‘Harovinton’, a high protein tofu-type cultivar with a complete complement of the 7S β -conglycinin and 11S glycinin subunits (Buzzell et al. 1991), to locate the genomic regions responsible for the storage protein subunits. Breeding efforts conducted to alter the 7S and 11S storage proteins in soybean are driven mostly by the livestock and food processing industries. In the livestock feed sector, soymeal has often been referred to as the “gold standard” among protein sources used in the feed industry for its outstanding amino acid profile (Cromwell 1999). Yet, a limiting factor to that profile are the two sulfur-containing amino acids, cysteine and methionine, respectively (Banaszkiewicz 2011; Dozier and Hess 2011; Panthee et al. 2004). The methionine plus cysteine concentration of soybean protein is approximately half that of egg protein (Imsande 2001). Soybean 7S storage proteins have very low cysteine and methionine concentration (Paek et al. 1997; Krishnan 2005), while the 11S storage proteins have up to four times more of these amino acids per gram of protein (Koshiyama 1968). Numerous studies have demonstrated a significant negative correlation existing between the 7S β -conglycinin subunit and the 11S glycinin subunits, with accumulation of the β -subunit of β -conglycinin in the seed, causing as much as a 40% decrease in 11S glycinin, due to sulfur deficiency (Wang et al. 2014; Gayler and Sykes 1985; Paek et al. 1997, 2000; Imsande and Schmidt 1998; Krishnan 2005; Imsande 2001) or excess nitrogen (Krishnan 2005). Hence, the nutritional quality of soybean protein is often compromised by the accumulation of β -conglycinin in the seed, which results in a concomitant decrease in the 11S glycinins, thereby lowering cysteine and methionine concentration (Krishnan 2005). To offset the low amounts of cysteine and methionine

found in soymeal, synthetic forms of these amino acids, accounting for up to \$100 million annually, are commonly added as supplements at feed mills to obtain optimal dietary requirements for animal growth and skeletal development (Imsande 2001; Han and Lee 2000; Krishnan 2005). Therefore, developing soybeans with increased sulfur-containing amino acids is a priority for breeders, as it would lead to the production of a higher quality, more nutritious soymeal and at the same time save livestock producers millions of dollars in synthetic amino acid amendments to feed stocks.

In addition, in the commercial food sector, soybeans with altered 7S and/or 11S subunits were found to have better gelling properties when used to make tofu (Teraishi et al. 2001; Poysa et al. 2006; Zarkadas et al. 2007b). Tofu is made from soybean by boiling soymilk and adding a coagulant to make the denatured proteins form a gel matrix (Kohyama et al. 1995). Consequently, the amount and type of proteins in the soybean seeds affects both the quality and yield of the tofu produced (Poysa and Woodrow 2002). Poysa et al. (2006) used a series of mutant genotypes with differing 7S and 11S subunits to demonstrate that tofu firmness increased with the absence of the 7S α' glycinin subunit and 11S *A₄* conglycinin subunit, and also found that the 11S *A₃* subunit contributed to firmness with any coagulant tested (Poysa et al. 2006). The identification of storage protein subunits that alter tofu texture allows for breeding and selection of soybean lines with specific storage protein subunit concentration that can produce the desired commercial texture with minimal total protein input during the manufacturing process. Since selection of seed-storage proteins needs to use the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with protein samples from the seeds, necessitating that breeders must wait until soybean seeds are harvested, which is time-consuming and costly, the development of an efficient SNP marker assay, eg., the KASP-SNP assay, is required to speed up the selection process in breeding.

The objectives of this research are to: (1) use a bulked segregant analysis (BSA) strategy to identify putative genomic regions controlling the 7S and 11S storage protein subunits in an *F₂* mapping population ($n = 448$) genotyped using Illumina SoySNP50K iSelect Beadchips; (2) develop kompetitive allele specific polymerase chain reaction (KASP) SNP assays using the polymorphic markers previously identified between the bulks at the candidate genomic regions for each of the protein storage subunits to improve selection efficiency in cultivar development; (3) identify QTLs controlling each of the protein storage subunits in the *F₂* mapping population using genotyping data from the newly developed KASP markers and phenotyping data of the storage protein profiles using SDS-PAGE; and (4) validate those QTL positions using an *F₅* recombinant inbred line (RIL) population ($n = 180$) with the flanking KASP markers of the QTLs previously identified. The results of

this work will elucidate the genetic architecture underlying the complex 7S and 11S storage protein subunit multi-gene families in this breeding line and allow breeders to develop new breeding lines with desirable commercial qualities by making selections using the new KASP markers linked with the storage protein subunits of interest.

Materials and methods

Plant materials and development of the mapping populations

An F_2 mapping population ($n = 448$), derived from the cross of Harovinton $\times$ SQ97-0263_3-1a, was used for the initial mapping of the protein storage subunits. Harovinton is a soybean cultivar developed by Agriculture and Agri-Food Canada (AAFC) from the cross of 'Vinton' $\times$ 'Vinton 81' (Buzzell et al. 1991), which has a complete complement of the 11S glycinin and 7S β -conglycinin subunits (Poysa et al. 2006). Harovinton is a maturity group II, large seeded tofu-type cultivar with high protein concentration and yellow seed coat and hilum (Buzzell et al. 1991). SQ97-0263_3-1a is a soybean breeding line developed by AAFC-Harrow Research and Development Centre (Harrow-RDC) (Harrow, Ontario, Canada), derived from a cross between Harovinton and 'Iwate-3'. Iwate-3 lacks the α' subunit of 7S β -conglycinin and all the 11S glycinin subunits as determined by SDS-PAGE analysis (Zarkadas et al. 2007b). The cross was made in the greenhouse at Harrow-RDC in the winter of 2012. The F_1 plants were grown in the field at Harrow, Ontario in the summer of 2013 to advance to the F_2 generation and were confirmed to be true hybrids using molecular markers. F_2 seeds were collected and chipped with a razor blade into two pieces. The 1/4 piece containing the seed cotyledon was used for protein extraction and protein profile analysis was done using SDS-PAGE, while the remaining piece containing the embryo was grown in the greenhouse in the fall of 2013 under 14 h day at 25 °C and 10 h night at 20 °C, with a relative humidity of 70%. Leaf material was collected off the F_2 plants in the greenhouse for DNA extraction. DNA was extracted from lyophilized material using a modified cetyltrimethyl ammonium bromide (CTAB) method (Keim et al. 1988).

The F_2 soybean plants used in the mapping study were advanced to the F_5 generation by single seed descent (SSD) at Harrow-RDC, Ontario, Canada. The 180 RILs were chosen based on the order of line IDs to validate the initial mapping results. A single seed randomly chosen from each of the 180 F_5 RILs was chipped for protein extraction and subsequently used in SDS-PAGE, as described previously, to determine the storage protein profile. The rest of the seed

sample was processed for DNA extraction, as described above.

SDS-PAGE profiles of seed protein and χ^2 analysis

Protein subunits of F_2 and F_5 soybean seeds were determined by SDS-PAGE using a PhastSystem (Pharmacia Biotech, Uppsala, Sweden) apparatus and pre-cast gels following the procedure described by Poysa et al. (2006). All the chemicals used in this study were of molecular biology or reagent grade (Sigma-Aldrich, St. Louis, MO, USA). The seed chip described above was used for protein extraction by grinding the chip into powder in liquid nitrogen. Proteins were extracted from the seed powder in a buffer containing 10 mM Tris-HCl, pH 8.0, 2.5% (w/v) sodium dodecyl sulfate, 10 mM β -mercaptoethanol, and 0.01% (w/v) bromophenol blue. Pre-cast PhastGel Gradient™ 10–15% polyacrylamide gradient gels were pre-treated by soaking for 1 h in 112 mM tris-acetate buffer, pH 6.4, and containing 4 M urea to modify protein mobility. This protocol facilitated separation of the α and α' subunits of the 7S β -conglycinin and also the 11S A_4 subunit from the 11S A_1 and A_2 subunits. The gels were blotted on both sides with lint-free tissue and air-dried for 15 min. Aliquots of extract containing 2 μ g proteins were applied to each lane. Gels were run using PhastGel SDS Buffer Strips in the PhastSystem apparatus containing electrophoresis buffer (25 mM Tris, 192 mM glycine, 0.1% glycine, 0.1% SDS, pH 8.3) under the following run settings: 250 V, 10 mA, 3.0 W, 65 Vh. The gels were stained with Brilliant Blue R to visualize the protein subunit bands. The presence or absence of protein subunit bands was assessed qualitatively by visually comparing the band pattern with those documented by Fontes et al. (1984) for soybeans. This protocol permitted efficient and highly reproducible identification of genotypes deficient in 11S glycinin subunits A_1 and A_2 (group I), A_3 (group IIb), A_4 (group IIa), and the α , α' and β subunits of 7S β -conglycinin (Poysa et al. 2006).

Chi square test probabilities were calculated against the predicted segregation ratios in the F_2 population for each of the storage protein subunits as follows: 7S α' (3:1), 11S A_1 (3:1), 11S A_2 (3:1), 11S A_3 (3:1) and 11S A_4 (3:1). Chi square test probabilities were also calculated for the F_5 RIL population against the predicted 1:1 segregation ratio for each of the storage protein subunit profiles.

Bulked segregant analysis

Leaf materials from the parents and F_2 population were collected from the plants grown in the greenhouse. DNA was extracted from lyophilized material using a modified CTAB method (Keim et al. 1988). Based on the presence/absence of each of subunits identified with the SDS-PAGE profiling method, DNA bulks were formed. Equal amounts of DNA

from 31 F_2 plants deficient only in the 11S A_1 or A_2 subunits were pooled and named as “ A_1 -bulk 1” or “ A_2 -bulk 1”. DNA from 50 F_2 plants that had the 11S A_1 or A_2 subunit were pooled and designated as “ A_1 -bulk 2” or “ A_2 -bulk 2”. Similar bulks were created for the 11S A_3 , 11S A_4 and 7S α' subunits, except that bulk 1 consisted of DNA from 50 F_2 plants instead of 31 plants. Bulk DNA samples, along with the parents, Harovinton and SQ97-0263_3-1a, were genotyped at Michigan State University using the Illumina SoySNP50K iSelect BeadChips (Song et al. 2013).

BSA (Michelmore et al. 1991; Quarrie et al. 1999) was subsequently used to identify the putative genomic regions that control the presence of the four protein subunits in SQ97-0263_3-1a. First, putative genomic regions for each subunit of storage protein were identified by making comparisons between the parents, and then by differentiating between the two bulks for each protein subunit after removing the missing data. Second, polymorphic single nucleotide polymorphisms (SNPs) resulting from all the comparisons were extracted for further analysis. A putative genomic region using BSA was targeted, if the haplotype SNP alleles of the protein subunit absent bulk (bulk 1 group) matched with the alleles of SQ97-0263_3-1a and were different from the Harovinton alleles. Genomic regions possessing more than 20 positive hits were considered to be candidate regions for genes controlling the storage protein subunits.

Development of KASP–SNP assays and genotyping

Polymorphic SNPs between the parents were first identified using Illumina SoySNP50K iSelect BeadChips, and then chosen to design KASP–SNP assays, if they fell within the putative genomic regions identified with BSA. In addition, the point mutation leading to the *gy4* null phenotype in several soybean lines (Yu et al. 2005; Kim et al. 2008) was also used to design a KASP–SNP assay (GSM0587). The primers are listed in the Supplementary Table 1. DNA from a subset of 236 individuals out of 448 in the F_2 population

was selected for genotyping. KASP assays were performed in a 4 μ L reaction, which included 1.894 μ L low ROX KASP master mix (KBiosciences, Herts, England), 0.106 μ L of primer mix (containing 0.318 μ M of each primer) and 2 μ L of genomic DNA (10–25 ng/ μ L). The PCR program for KASP was 94 °C for 15 min, followed by 10 cycles of touchdown PCR from 65 °C to 57 °C with a 0.8 °C decrease per cycle, then followed by 30 cycles of 94 °C for 20 s and 57 °C for 1 min. Allele calls for each SNP were performed in a Roche LightCycler 480 II with LightCycler® Software (Roche Diagnostics Corporation) and exported to an Excel file. Genotype calls for each F_2 individual were determined based upon the parental allele calls and clusters for each KASP marker.

QTL identification in the F_2 population

The genotyping data from the KASP–SNP assays of the F_2 mapping population ($n = 236$), along with protein profiling data from the SDS-PAGE gels, were used to map the storage protein subunits onto linkage groups using JoinMap 4.1 software (Van Ooijen 2006). The genotypic and phenotypic data were converted into the proper classification type, where “a” represented data derived from the Harovinton parent and “b” represented data derived from the SQ97-0263_3-1a parent that lacked each of the protein storage subunits. The linkage maps were created using the Kosambi mapping function in the regression mapping algorithm and linkages were selected at an LOD ≥ 5.0 . Polymorphic SNPs located at each end of the candidate genomic regions and in the center of the putative QTLs were tested first, and the initial mapping results were used to select additional SNPs for genetic map saturation. The linkage maps were further refined, until markers closely linked to the traits were identified. The final linkage maps were drawn in MapChart version 2.30 (Voorrips 2002) for illustration purposes using data derived from JoinMap 4.1.

Table 1 Parental allele frequencies for storage protein subunits in the F_2 mapping population ($n = 448$) and F_5 RIL population ($n = 180$) derived from the cross of ‘Harovinton’ $\times$ SQ97-0263_3-1a

Mapping population	Subunit profiled	Harovinton	SQ97-0263_31a	Expected ratio	χ^2	P
F_2	7S α'	343	105	3:1	0.58	0.44
	11S A_1	257	31	3:1	31.10	< 0.001
	11S A_2	257	31	3:1	31.10	< 0.001
	11S A_3	338	110	3:1	0.04	0.82
	11S A_4	348	100	3:1	1.71	0.19
F_5	7S α'	95	85	1:1	0.55	0.45
	11S A_1	164	16	1:1	121	< 0.001
	11S A_2	143	37	1:1	62.40	< 0.001
	11S A_3	82	98	1:1	1.42	0.23
	11S A_4	151	29	1:1	82.70	< 0.001

QTL validation in the F₅ RIL population

The F₅ RILs ($n = 180$) were genotyped using the flanking KASP markers of the previously identified QTLs in the F₂ mapping population and phenotyped using SDS-PAGE to obtain new storage protein profiles for the subunits of 7S α' , 11S A₁, 11S A₃ and 11S A₄. The genotypic and phenotypic data were again converted into the proper classification type, where “a” represented data derived from the Harovinton parent and “b” represented data derived from the SQ97-0263_3-1a parent that lacked each of the protein storage subunits. Linkage groups were constructed using MapDisto version 1.8.1 (Lorieux 2012). Linkage groups were established using a LOD score of 5, and the “order” and “ripple” functions were used to determine the best arrangement of the markers for each linkage group. The recombination fractions were converted into centiMorgans (cM) using the Kosambi mapping function.

Comparison of mapped protein subunit QTL with reported storage protein genes

Previous mapping studies have placed the *GM7S* (*CG-1*) and *Gyl-7* genes on Chrs 3, 10, 13, and 19 (Nielsen et al. 1989; Diers et al. 1993; Beilinson et al. 2002; Li and Zhang 2011). To compare our mapping results of protein subunits from SQ97-0263_3-1a with reported storage protein genes, physical locations of the annotated storage protein genes were determined using the Williams 82 reference genome sequence (Schmutz et al. 2010) version 2 on NCBI (<http://www.ncbi.nlm.nih.gov>) and SoyBase (<http://www.soybase.org>). Physical locations of the polymorphic SNPs used for mapping of protein subunit QTLs in this study were confirmed using the dbSNP identification number (ss#) on the “SoySNP50K” track of the soybean sequence browser in SoyBase (Grant et al. 2009). The locations of the genes in relation to the SNP markers used to map the protein subunits were determined based on physical locations of both the genes and interval of each QTL flanked by SNP markers.

Results

SDS-PAGE profiling

To profile the F₂ population and the F₅ RILs segregating for the 7S α' or the 11S storage protein subunits, SDS-PAGE gels were run for each of the F₂ individuals ($n = 448$) and each of the F₅ RILs ($n = 180$). As shown in Fig. 1, each F₂ seed was evaluated for band presence or absence at the subunit location using a 0/1 binomial system (0 = no band, 1 = band present). The phenotypic scores from the SDS-PAGE gels were then converted into genotype calls, with 0

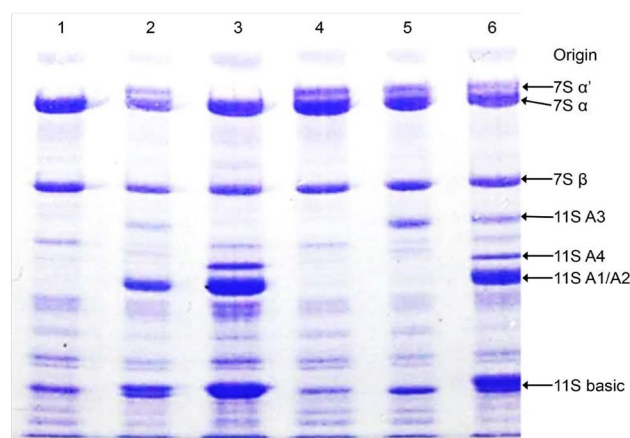


Fig. 1 One-dimensional polyacrylamide gel electrophoretic analysis of soybean protein variants illustrating the presence or absence of storage protein subunits. Lane 1, parent SQ97-0263_3-1a (complete 11S null, 7S α' null); Lane 2, 11S A₄ null variant; Lane 3, 11S A₃ null, 7S α' null variant; Lane 4, complete 11S null variant; Lane 5, 11S A₃, A₁, A₂ null variant; Lane 6, ‘Harovinton’ parent (all protein subunits present). Note that in this electrophoretic protocol, lower molecular weight proteins including the 11S basic subunits are not separated

and 1 scores representing homozygous null and presence of wild type parental alleles, respectively. Table 1 shows the parental allelic frequency for each storage protein subunit evaluated in this study. The observed ratios for the 7S α' , 11S A₃, and 11S A₄ subunits all fit the expected 3:1 segregation ratios for single gene inheritance (Table 1). Due to the close proximity of the 11S A₁ and 11S A₂ subunit bands on the gel (Fig. 1), ambiguous results were not scored, leading to a reduced number of genotypes with SQ97-0263_3-1a alleles for these two subunits. The 11S A₁ and 11S A₂ traits did not fit the expected 3:1 segregation ratio for single gene inheritance at the 0.01 significance level likely due to difficulty in scoring the presence/absence of the trait on the SDS-PAGE gels (Table 1).

As for the F₅ RILs, Chi square test probabilities for segregation distortion were calculated against the predicted 1:1 segregation ratio and only the storage protein profiles of the 11S A₁ and 11S A₄ subunits fit the expected 1:1 ratio (Table 1). The 7S α' and the 11S A₃ subunits each deviated from the expected 1:1 ratio ($P < 0.05$). The 11S A₂ subunit trait deviated significantly from the expected 1:1 ratio ($P < 0.0001$), again likely due to difficulty in scoring the presence/absence of the bands on the SDS-PAGE gels, as described above.

Bulked segregant analysis

Using the phenotypic data derived from the SDS-PAGE gels, bulked samples were formed (from the corresponding F₂ plants) for both the presence and absence of each

storage protein subunit evaluated in this study. The bulked samples and the two parents were first genotyped using the Illumina SoySNP50K iSelect BeadChips (Song et al. 2013). A total of 13,997 SNPs from the SoySNP50K BeadChips were identified as being polymorphic between the two parents. Using the selection criteria of distinguishing at least 20 polymorphic SNPs between both the trait bulks and the two parents per chromosome, potential genomic regions responsible for the loss of the storage protein subunits were located (Table 2). There were a total of 58 polymorphic SNPs between the two 7S α' subunit bulk groups and between the parents. The putative genomic region for the 7S α' subunit spans a 2.8 Mb region of Chr 10 (LG O), which contains 21 SNPs from the SoySNP50K BeadChips. A total of 86 polymorphic SNPs were found between the two 11S A₁ subunit bulks that were also polymorphic between the two parents. There are two potential genomic regions associated with the 11S A₁ subunit based on the BSA results: one is located on Chr 3 (LG N), which had 20 polymorphic SNPs spanning 2.3 Mb, while the second is located on Chr 19 (LG L), which had 50 polymorphic SNPs spanning 5.3 Mb. The 11S A₂ subunit groups had only 25 polymorphic SNPs between the two bulk groups. Of the 25 polymorphic SNPs identified, only 12 SNPs were also polymorphic between the two parents, so QTL analysis of the 11S A₂ subunit was not pursued. There were 251 polymorphic SNPs between the 11S A₃ subunit bulks and the parents. The 11S A₃ subunit putative genomic region contains 206 SNPs and spans a 24.3 Mb region of Chr 13 (LG F). The 11S A₄ subunit bulks had 44 polymorphic SNPs that were also polymorphic between the parents. The genomic region of interest for the 11S A₄ subunit spans 3.8 Mb on Chr 10 (LG O) which contains 36 SNPs from the SoySNP50K BeadChips.

Table 2 Putative genomic regions controlling the storage protein subunits that were identified by bulked segregant analysis (BSA) using Illumina SoySNP50K iSelect BeadChips in the F₂ mapping population derived from the cross of ‘Harovinton’ × SQ97-0263_3-1a

Subunit profiled	Polymorphic SNPs between bulks	Gene	Chr	Polymorphic SNPs at putative QTL region
7S α'	99	<i>GM7S</i> ^a	10	21
11S A ₁	122	<i>Gy1</i>	3, 19	20, 50
11S A ₂	25	<i>Gy2</i>	3	N/A
11S A ₃	251	<i>Gy5</i>	13	206
11S A ₄	49	<i>Gy4</i>	10	36

^aGene nomenclature for the 7S α' subunit as listed in NCBI (<http://www.ncbi.nlm.nih.gov/gene/>)

QTL results in the F₂ population

To map the QTL to the genomic regions identified with BSA, KASP assays were designed for the selected SNPs in the genomic regions for each of the storage protein subunits, and subsequently used to genotype a subset of 236 individuals from the population. Due to a large population size, only part of the F₂ population was chosen for QTL mapping. Selection of these F₂ individuals was based on the order of sample IDs. The genotypic and phenotypic data from the population were used to map the storage protein subunits onto linkage groups. As shown in Fig. 2a, the 7S α' subunit was placed between markers GSM0577 and GSM0500 on Chr 10 (LG O). The 11S A₁ subunit was mapped to a region on Chr 3 (LG N) between markers GSM0481 and GSM0582 (Fig. 2b), but did not map within the putative genomic region on Chr 19 (LG L) that was identified by BSA (data not shown). The 11S A₃ subunit was mapped to a region on Chr13 (LG F) between markers GSM0586 and GSM0585 (Fig. 2c). The 11S A₄ subunit was mapped between the markers GSM0497 and GSM0588 on Chr 10 (LG O) (Fig. 2d). Marker GSM0587 that was developed using the *gy4* point mutation was identified in several 11S A₄ null lines mapped 2.8 cM away from the 11S A₄ trait on Chr 10.

QTL validation in the F₅ RIL population

The following QTLs were validated with the F₅ RILs using genotyping data of the flanking KASP markers of each QTL from the F₂ mapping population and phenotyping data of the storage protein profiles from SDS-PAGE as follows: the 7S α' β -conglycinin subunit to Chr 10 (LG O), 1.6 cM from marker GSM0577 and 1.2 cM from GSM0500 (Fig. 3a); the 11S A₁ subunit to Chr 3 (LG N), 0.3 cM from GSM0481 and 1.2 cM to GSM0582 (Fig. 3b); and the 11S A₃ subunit to Chr 13 (LG F), 0.1 cM from marker GSM0586 and 1.1 cM to GSM0585 (Fig. 3c). The 11S A₄ subunit was located 0.3 cM from GSM0588 on Chr 10 and just 1.9 cM from GSM0497, but was instead flanked by GSM0591 (Fig. 3d). The QTL positions that were mapped in the F₅ RIL population were generally consistent and validated the storage protein subunit positions from the initial mapping results using the F₂ population. In the case of the 11S A₄ QTL mapped in the F₅ population, the marker order of the linkage group did change, likely due to the errors associated with genotyping or different recombinations between two populations.

Comparison of mapped protein subunit QTL with reported storage protein genes

The physical locations of the storage protein subunit genes were placed using the Williams 82 genome (Schmutz et al.

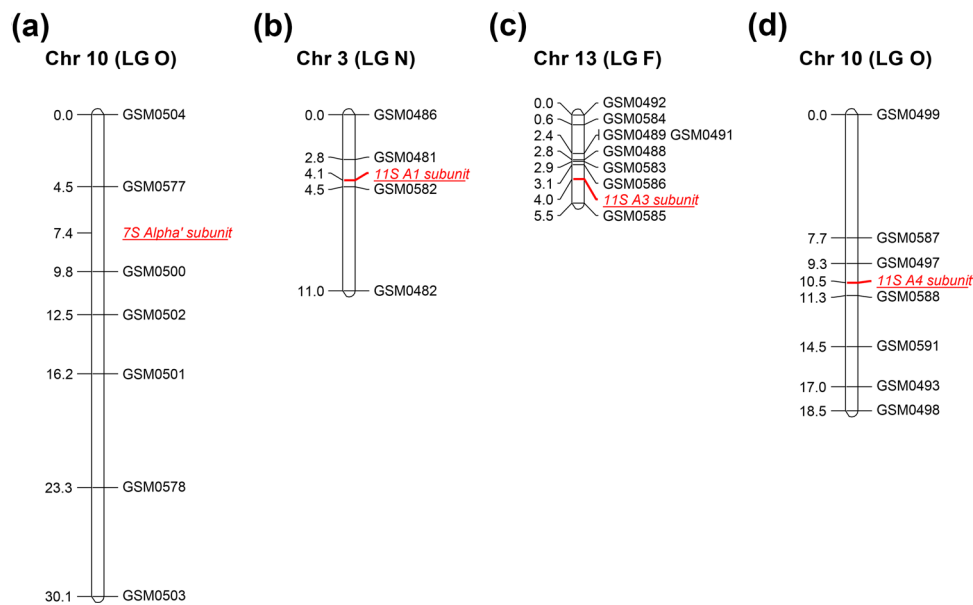


Fig. 2 Linkage map of the 7S β -conglycinin and 11S A_1 , 11S A_3 and 11S A_4 glycinin storage protein subunits in an F_2 population ($n = 236$) derived from the cross of ‘Harovinton’ $\times$ SQ97-0263_3-1a. The numbers on the left side represent the distances in centiMorgan to the corresponding kompetitive allele specific polymerase chain reaction (KASP) markers displayed on the right side. The phenotypic scores from the sodium dodecyl sulfate polyacrylamide gel electro-

phoresis (SDS-PAGE) gels were converted into genotypes, allowing for the concurrent construction of linkage groups and mapping of traits. Each of the storage protein subunit QTLs were mapped as follows: **a** 7S α' subunit to Chromosome (Chr) 10 (LG O); **b** 11S A_1 subunit to Chr 3 (LG N); **c** 11S A_3 subunit to Chr 13 (LG F); and **d** 11S A_4 subunit to Chr 10 (LG O)

2010) version 2 sequence browser in Soybase (<http://www.soybase.org>) using results from other mapping studies, described previously. As shown in Table 3, all of the storage protein subunit genes investigated here are located within the mapped QTL regions identified. The 7S α' subunit gene is 2.7 kb long and located on Chr 10 (LG O), starting at the position 47,483,367. The 11S A_3 subunit gene (*Gy5*) is 3.0 kb long starting at the position 23,665,664 on Chr 13 (LG F). The 11S A_4 subunit gene (*Gy4*) is 2.8 kb long starting at position 3,256,351 on Chr10 (LG O). There are SNPs from the SoySNP50K BeadChips located within the coding regions of the following genes that were not identified as polymorphic between the parents: ss715607694 (7S α'), ss715585857 (*Gy1*), and ss715606240 (*Gy4*) (data not shown). There was not a SNP from the SoySNP50K BeadChips within the coding region of the *Gy5* gene, and the two closest SNPs (ss715614090 and ss715614091) located on either side of the gene were not polymorphic between the parents (data not shown).

Discussion

To aid in the development of soybean cultivars lacking certain storage protein subunits, an F_2 mapping population and an F_5 RIL population were developed by crossing the

cultivar Harovinton with the breeding line SQ97-0263_3-1a, which lacks the 7S α' β -conglycinin and all 11S glycinin storage protein subunits. The source of the mutations that led to the loss of expression for the storage proteins in Iwate 3 is believed to be γ -irradiation (V. Poysa, personal communication), so the physical location of the mutations were unknown at the initiation of this study. As such, the storage protein subunits in this population were treated as uncharacterized traits, even though physical locations of the genes were previously published (Harada et al. 1989; Nielsen et al. 1989; Diers et al. 1993; Chen and Shoemaker 1998; Beilinson et al. 2002; Li and Zhang 2011). Additionally, there have been conflicting reports concerning the linkage of *Gy1* and *Gy2* on Chr 3 (LG N), with Beilinson et al. (2002) concluding that the linkage of the Group I genes could be determined upon the accumulation of a gene product rather than identifying the actual glycinin genes, and that γ -irradiation-induced mutation may uncover trans-acting modifier genes controlling seed-storage protein accumulation. Because it is thought that the mutations in the current populations used herein stem from γ -irradiation, it was of further interest to map the Group I glycinin genes in the F_2 and F_5 populations.

The 448 individuals in the F_2 mapping population were analyzed by SDS-PAGE for subunit composition. The close proximity of the 11S A_1 and 11S A_2 subunits on the protein gel (Fig. 1) led to ambiguous results for these subunits in

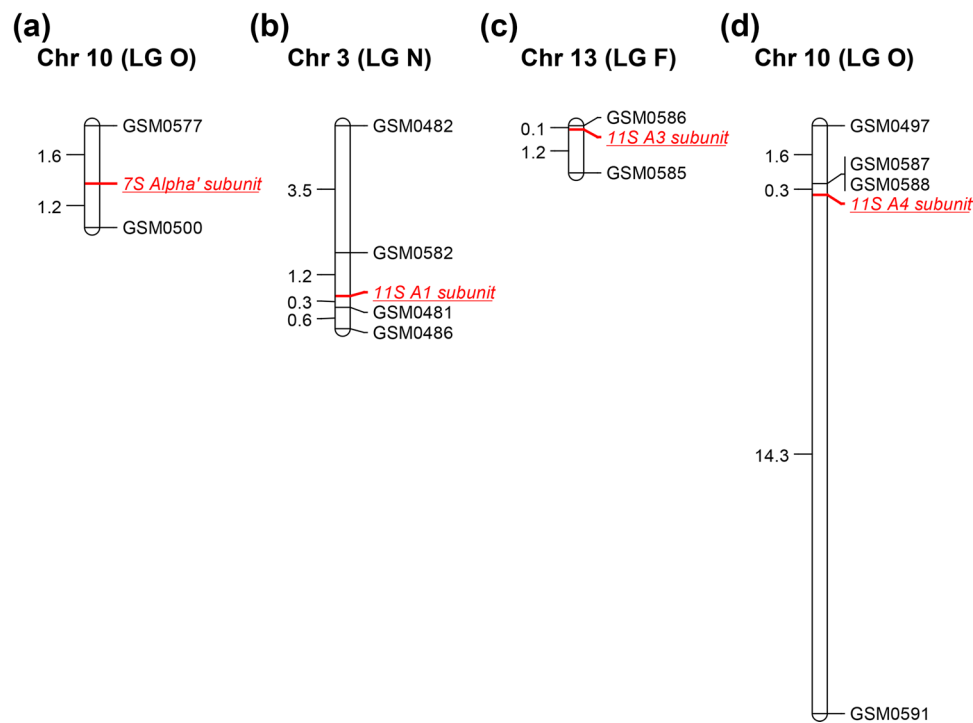


Fig. 3 Linkage map of the 7S β -conglycinin and 11S A₁, 11S A₃ and 11S A₄ glycinin storage protein subunits in an F₅ RIL population ($n = 180$) derived from the cross of ‘Harovinton’ $\times$ SQ97-0263_3-1a. The numbers on the left side represent the interval distances in centiMorgan between the kompetitive allele specific polymerase chain reaction (KASP) markers and QTLs displayed on the right side. The phenotypic scores from the sodium dodecyl sulfate polyacrylamide

gel electrophoresis (SDS-PAGE) gels were converted into genotypes, allowing for the concurrent construction of linkage groups and mapping of traits using the flanking KASP markers of the QTL identified in the F₂ mapping population. Each of the storage protein subunit QTL were validated as follows: **a** 7S α' subunit to Chromosome (Chr) 10 (LG O); **b** 11S A₁ subunit to Chr 3 (LG N); **c** 11S A₃ subunit to Chr 13 (LG F); and **d** 11S A₄ subunit to Chr 10 (LG O)

Table 3 Comparison of physical position of reported storage protein subunit genes with mapped QTL in this study

Subunit	Chromosome	Flanking SNP markers	Marker locations (bp) ^a	Gene	Reported gene location (bp) ^a
7S α'	10 (LG O)	GSM0500 GSM0577	Gm10: 46,729,408 Gm10: 47,099,455	<i>GM7S</i> ^{b,c}	47,483,367–47,486,034
11S A ₁	3 (LG N)	GSM0481 GSM0582	Gm03: 37,700,650 Gm03: 37,832,346	<i>Gy1</i> ^d	37,823,096–37,825,738
11S A ₃	13 (LG F)	GSM0583 GSM0585	Gm13: 23,464,228 Gm13: 24,167,615	<i>Gy5</i> ^e	23,665,664–23,668,723
11S A ₄	10 (LG O)	GSM0497 GSM0588	Gm10: 2,742,385 Gm10: 3,285,126	<i>Gy4</i> ^e	3,256,351–3,259,176

The flanking kompetitive allele specific polymerase chain reaction (KASP) markers were used to map storage protein subunit quantitative trait loci (QTL) in an F₂ population, which were also validated in an F₅ RIL population. The mapping populations were derived from the cross of ‘Harovinton’ $\times$ SQ97-0263_3-1a

^aPhysical location of markers and genes based upon the Williams 82 reference genome sequence (Schmutz et al. 2010) in Soybase (<http://www.soybase.org>)

^bGene nomenclature for the 7S α' subunit as listed in NCBI (www.ncbi.nlm.nih.gov/gene/)

^cHarada et al. (1989)

^dBeilinson et al. (2002)

^eDiers et al. (1993); Chen and Shoemaker (1998)

some individuals (160 individuals, 36% of the population). The ambiguous scoring results were not included in the protein profile data (Table 1), leading to a reduced number of individuals available to be used for mapping these two subunits. Based on the protein profile data from the mapping

population, groups of individuals were formed for BSA. The groups were designated by the presence or absence of one of the storage protein subunits (7S α' , 11S A₁, 11S A₂, 11S A₃ and 11S A₄), which led to ten bulked groups and the two parents being genotyped using the Illumina SoySNP50K iSelect

BeadChips (Song et al. 2013). BSA allows for rapid identification of potential genomic regions by pooling individuals that should have similar genotypes for a given trait, while the rest of the genome is presumed heterozygous (Michelmore et al. 1991). By first filtering the SNP data for polymorphisms between the two parents, the number of SNPs was reduced from the original 52,041 to 13,997. Of those SNPs, polymorphisms between the two bulks for each trait were then identified. This led to a reduced number of SNPs with polymorphic calls after the data filtering process (Table 2), which allowed for rapid identification of genomic regions that may be associated with the trait of interest. Use of the BSA method in the present study identified five putative genomic regions for four of the storage protein subunits: 7S α' , 11S A₁, 11S A₃ and 11S A₄.

Two potential QTLs were associated with the 11S A₁ subunit. The 11S A₁ subunit QTL that was not successfully mapped in this study likely resides on Chr 19 (LG L), where a large region of SNPs from the Illumina SoySNP50K iSelect SNP BeadChips (Song et al. 2013) were homozygous within the 11S A₁ subunit bulked group (Table 2). This hypothesis is in agreement with Li and Zhang (2011) who found the *Gy3* gene (A_{1b}B_{1b}) to be linked with *Gy7* on Chr 19 (LG L), where large regions of glycinin gene syntenic segments occur over 13 megabases (MB), stemming from whole genome duplication events that occurred ~ 50 to 60 and ~ 10 to 15 mya. Markers spanning the potential 11S A₁ subunit QTL region on Chr 19 (LG L) were used to genotype the mapping population, but were not found to be linked with the trait (data not shown).

The 11S A₂ subunit QTL was not pursued in this study, due to the low number of confirmed parental polymorphisms at the putative genomic region controlling the trait identified by BSA and due to difficulty in scoring the 11S A₁ and 11S A₂ subunits when run on a protein gel, the latter leading to ambiguous results. Additionally, it was reported that a gene encoding the 11S A₁ subunit (*Gy1*: A_{1a}B₂) underwent a tandem duplication event (Li and Zhang 2011), which led to the origin of the *Gy2* gene (A₂B_{1a}), with the two genes being separated by only ~3 kb on Chr 3 (LG N) (Beilinson et al. 2002). The tight linkage of the 11S A₁ and A₂ subunits, both phenotypic and genomic, would preclude independent QTL identification for one of the traits. Alternatively, the close association of the two subunits does suggest that the markers identified for mapping the 11S A₁ subunit trait may be useful for identification of the 11S A₂ subunit, as well. The fact that the 11S A₁ trait mapped to Chr 3 (LG N) in the present study supports the conclusion that the 11S A₁ and 11S A₂ traits are linked on Chr 3 (LG N). Further work needs to be done to confirm the tight linkage of these two 11S storage protein subunits and to develop markers associated with the 11S A₂ subunit to identify possible recombination events between the two genes.

Using the Illumina SoySNP50K iSelect BeadChips (Song et al. 2013) genotyping data in conjunction with the BSA strategy not only made the identification of potential genomic regions of interest for the storage protein subunits efficient, but also provided the genetic data required to develop marker assays within the putative genomic regions to allow for QTL mapping. The number of SNPs within each putative genomic region was dependent on both the size and location of the identified region of interest. The SoySNP50K BeadChips have an average SNP distance of 9.1 and 49.1 kb ($\pm$ 9.1, 33.3 SD) for euchromatic and heterochromatic regions, respectively (Song et al. 2013). The putative genomic regions of interest identified herein had an average of one polymorphic SNP per 116 kb within each region. Most of the QTL regions identified by BSA in this study averaged 32 polymorphisms and spanned an average region of 3.6 Mb. The putative genomic region of 11S A₃ subunit had the largest number of polymorphic SNPs (206) and consequently spanned the largest distance (24.3 Mb). This suggests that there is reduced recombination frequency for this region of the soybean genome. As the putative genomic region for the 11S A₃ trait is not located near the centromere or telomeres on Chr 13 (LG F), the low recombination rate of the putative 11S A₃ QTL is unlikely due to a heterochromatic region, but rather to another as yet undetermined cause.

Using SNPs identified with the BSA strategy segregating for the storage protein subunits of interest made mapping the traits a straightforward process. All subunits evaluated in the present study are controlled by single genes (Beilinson et al. 2002; Li and Zhang 2011). This enabled us to treat the storage protein traits qualitatively, rather than quantitatively, which allowed for the concurrent mapping of subunits and construction of the SNP-based linkage maps (Fig. 2). The close association of the protein storage subunits and the KASP markers developed in the present study should allow breeders to utilize the Georgia Soybean Markers (GSM) in their breeding programs using marker-assisted selection (MAS). Furthermore, the QTL positions for each of the storage protein subunits were validated using the F₅ RILs (Fig. 3).

Table 3 lists the GSM–KASP markers developed herein that are suggested for use in breeding programs interested in deploying MAS for the 7S α' , 11S A₁/A₂, 11S A₃, and 11S A₄ storage protein subunit traits (KASP assay primers are also presented in Supplementary Table 1). As with any marker intended for selection purposes, the markers developed in this work need to be tested for polymorphisms within a researcher's breeding program. The markers for the 7S α' subunit, GSM0500 and GSM0577, are both located upstream of the physical location of the 7S α' gene. In this population, the 7S α' trait was more closely linked to all of the markers whose physical locations were upstream of the

7S α' gene (Fig. 2a; Supplementary Table 1). As the cause of the null phenotypes is unknown for all of the storage proteins studied in this population, the 7S α' null phenotype observed in this population may be due to a mutation of a regulatory gene or element rather than the gene itself. The SNP marker with the closest physical location to the 7S α' subunit physical location is GSM0577, which is 384 kb upstream.

The markers for the 11S A₁ storage protein subunit listed in Table 3 are located upstream and downstream of the 11S A₁ gene (*Gy1*) on Chr 3 (LG N), with marker GSM0582, located 6.6 kb downstream of the physical location of *Gy1*, being the closest marker in physical location to the gene. The markers listed in Table 3, if used for selection of the 11S A₃ trait, have physical locations flanking the *Gy5* gene on Chr 13 (LG F). Marker GSM0583 is closest in physical location to the *Gy5* gene on Chr 13 (LG F), located 201 kb upstream from the gene. As discussed previously, there was low recombination amongst the markers used to map this trait in the population, and marker GSM0586 that mapped between marker GSM0583 and the 11S A₃ trait (Fig. 2c) could be an artifact of the mapping algorithm and low recombination frequency. The physical locations of markers GSM0497 and GSM0588 are upstream and downstream of the physical location of *Gy4* on Chr 10 (LG O). Marker GSM0588 is closest in physical location to the *Gy4* gene, located 26 kb downstream from the gene. Marker GSM0587, which was developed using the point mutation found in many *Gy4* null lines, was not the most tightly linked marker in this study. As shown in Fig. 2d, marker GSM0587 mapped 2.8 cM away from the 11S A₄ trait, indicating that recombination occurred between the marker and the trait in the population used in this study. This further suggests that the 11S A₄ null phenotype observed in this population is not caused by the point mutation found in other soybean lines (Yu et al. 2005; Kim et al. 2008). However, marker GSM0587 may be useful for soybean breeders using *Gy4* null lines, whose pedigree can be traced to cultivars known to have the *gy4* point mutation or *Gy4* mutation caused by γ -irradiation. Further work needs to be done to determine the cause of the null phenotypes for the storage protein subunits derived from the SQ97-0263_3-1a parent used in this study.

It has been suggested that soybean benefits from the presence of duplicated genes in the genome, because more protein products are produced, which have enabled those genes to withstand purifying selection and positive selection forces (Li and Zhang 2011). In the case of the Group I glycinin genes, this type of selection pressure offers an explanation for mutations to *Gy1* and *Gy2* and the resulting higher expression of sulfur-containing amino acids cysteine and methionine (Li and Zhang 2011). A common goal of soybean breeders is to be able to take advantage of these mutations by developing soybeans with increased cysteine and methionine concentration, two limiting nutritional

components in the soybean amino acid profile. The ability to use the KASP assays designed herein for breeding line selection of storage protein subunits derived from SQ97-0263_3-1a will give breeders new tools to achieve that goal. Further research needs to be done to identify those breeding lines with the ideal combination of storage protein subunits with the highest cysteine and methionine seed concentration.

Previous research has found that tofu made from soybean lacking the 7S α' and 11S A₄ subunits lead to a firmer tofu texture (Poysa et al. 2006). In a given F₂ population, when selecting for the loss of two independently inherited genes, 0.25² (or 6.3%) of the population is expected to be observed lacking the two genes. In the mapping population developed for this research, seven lines (2.9% of population) lacked the 7S α' and 11S A₄ subunits, but each one had all the other 11S subunits present. This number is skewed by the large number of lines that were not scored for the 11S A₁ and 11S A₂ traits, due to their close proximity on the SDS-PAGE gel. If we ignore the other subunits and select only the lines lacking the 7S α' and 11S A₄ subunits, the number of lines in the mapping population increases to 21 (8.8% of population). Overall, the number of soybean lines lacking the 7S α' and 11S A₄ subunits in the population is still quite low, and by utilizing the markers developed as part of this research, soybean breeders now have a suite of KASP markers available for MAS to quickly identify the breeding lines containing the storage protein subunit profile desired for firmer tofu texture. The application of these markers in a breeding program should lead to soybean cultivars that have better gelling properties without the need for additional protein input during the tofu manufacturing process. Again, future breeding efforts will determine if the use of these markers will lead to the selection of breeding lines with a more favorable amino acid profile containing the desired ratio of glycinin to conglycinin globulins.

Author contribution statement JDB wrote the manuscript and validated QTL positions. KY, YC and ZL conceived and designed the project, interpreted data and supervised and reviewed the manuscript. RT assisted with initial marker design, genotyping, QTL analysis and partial writing. DA and LW performed SDS-PAGE analysis. CS, XW and VN assisted in population development and leaf and seed sample collection. VN collected leaf samples and extracted genomic DNA for genotyping.

Acknowledgments The authors would like to thank Michael Bissonnette from Agriculture and Agri-Food Canada for his assistance in planting and sample collection, Tatyana Nienow, Leon Chen, and Zi Shi from the University of Georgia for their technical assistance in marker design or genotyping. The research was funded by Agriculture and Agri-Food Canada A-Base, the Growing Forward II (GFII) Science Cluster projects, and the United Soybean Board.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- American Soybean Association (2017) Soystats. <http://soystats.com/international-world-soybean-production/>. Accessed 20 Jun 2017
- Banaszkiewicz T (2011) Nutritional Value of Soybean Meal. In: El-Shemy H (ed) Soybean and Nutrition, ISBN: 978-953-307-536-5. InTech. <http://www.intechopen.com/books/soybean-and-nutrition/nutritional-value-of-soybean-meal>
- Beachy RN, Jarvis NP, Barton KA (1981) Biosynthesis of subunits of the soybean 7S storage protein. *J Mol Appl Genet* 1:19–27
- Beilinson V, Chen Z, Shoemaker R, Fischer R, Goldberg R, Nielsen N (2002) Genomic organization of glycinin genes in soybean. *Theor Appl Genet* 104:1132–1140
- Buzzell RI, Anderson TR, Hamill AS, Welacky TW (1991) Harovinton soybean. *Can J Plant Sci* 71:525–526
- Chen Z, Shoemaker RC (1998) Four genes affecting seed traits in soybean map to linkage group F. *J Hered* 89:211–215
- Cho TJ, Davies CS, Nielsen NC (1989) Inheritance and organization of glycinin genes in soybean. *Plant Cell* 1:329–337
- Cromwell G (1999) Soybean meal—an exceptional protein source. Soybean Meal INFOcenter. <http://www.soymeal.org/reviewpapers.html>. Accessed 20 Jun 2017
- Davies CS, Cho T-J, Nielsen NC (1987) Tests for linkage of seed protein genes *Lx₁*, *Lx₃*, and *Gcy₁* to markers on known linkage groups. *Soybean Genet Newsl* 14:248–252
- Diers BW, Beilinson V, Nielsen NC, Shoemaker RC (1993) Genetic mapping of the *Gy4* and *Gy5* glycinin genes in soybean and the analysis of variant *Gy4*. *Theor Appl Genet* 89:297–304
- Dozier WA, Hess JB (2011) Soybean meal quality and analytical techniques. In: El-Shemy H (ed) Soybean and nutrition, ISBN: 978-953-307-536-5. InTech. <http://www.intechopen.com/books/soybean-and-nutrition/soybean-meal-quality-and-analytical-techniques>
- Fehr WR, Hoeck JA, Johnson SL, Murphy PA, Nott JD, Padilla GI, Welke GA (2003) Genotype and environment influence on protein components of soybean. *Crop Sci* 43:511–514
- Fischer RL, Goldberg RB (1982) Structure and flanking regions of soybean seed protein genes. *Cell* 29:651–660
- Fontes EPB, Moreira MA, Davies CS, Nielsen NC (1984) Urea-elicited changes in relative electrophoretic mobility of certain glycinin and beta-conglycinin subunits. *Plant Physiol* 76:840–842
- Gayler KR, Sykes GE (1985) Effects of nutritional stress on the storage proteins of soybeans. *Plant Physiol* 78:582–585
- Goldberg RB, Hoscheck G, Ditta GS, Breidenbach RW (1981) Developmental regulation of clone superabundant embryo mRNAs in soybean. *Dev Biol* 83:14
- Grant D, Nelson RT, Cannon SB, Shoemaker RC (2009) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res* 38(suppl_1):D843–D846
- Han K, Lee J (2000) The role of synthetic amino acids in monogastric animal production. *Asian Austral J Anim* 13:543–560
- Harada JJ, Barker SJ, Goldberg RB (1989) Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* 1:415–425
- Hayashi M, Kitamura K, Harada K (2009) Genetic mapping of *Cgdef* Gene controlling accumulation of 7S globulin (β -conglycinin) subunits in soybean seeds. *J Hered* 100:802–806
- Imsande J (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. *Crop Sci* 41:510–515
- Imsande J, Schmidt JM (1998) Effect of N source during soybean pod filling on nitrogen and sulfur assimilation and remobilization. *Plant Soil* 202:41–47
- Jegadeesan S, Yu K, Woodrow L, Wang Y, Chun S, Poysa V (2012) Molecular analysis of glycinin genes in soybean mutants for development of gene-specific markers. *Theor Appl Genet* 124:365–372
- Jenkinson JE, Fehr WR (2010) Agronomic and seed characteristics of soybean lines with alleles for modified glycinin concentration. *Crop Sci* 50:1896–1903
- Johnson L, Smith K (2003) Soybean processing fact sheet. Soybean Meal INFOcenter. <http://www.soymeal.org/FactSheets/process-ing3.pdf>. Accessed 20 Jun 2017
- Keim P, Olson TC, Shoemaker RC (1988) A rapid protocol for isolating soybean DNA. *Soybean Genet Newsl* 15:150–152
- 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* 56:11320–11326
- Kohyama K, Sano Y, Doi E (1995) Rheological characteristics and gelation mechanism of tofu (soybean curd). *J Agric Food Chem* 43:1808–1812
- Koshiyama I (1968) Chemical and physical properties of a 7S protein in soybean globulins. *Cereal Chem* 45:394–404
- Krishnan HB (2005) Engineering soybean for enhanced sulfur amino acid content. *Crop Sci* 45:454–461
- Krishnan HB, Kim W-S, Oehrle NW, Alaswad AA, Baxter I, Wiebold WJ, Nelson RL (2015) Introgression of Leginsulin, A Cysteine-rich protein, and high-protein trait from an Asian soybean plant introduction genotype into a North American Experimental Soybean Line. *J Agric Food Chem* 63:2862–2869
- Krober OA, Cartter JL (1962) Quantitative interrelations of protein and nonprotein constituents of soybeans. *Crop Sci* 2:171–172
- Li C, Zhang YM (2011) Molecular evolution of glycinin and beta-conglycinin gene families in soybean (*Glycine max* L. Merr.). *Heredity* 106:633–641
- Liu S, Ohta K, Dong C, Thanh VC, Ishimoto M, Qin Z, Hirata Y (2006) Genetic diversity of soybean (*Glycine Max* (L.) Merrill) 7S globulin protein subunits. *Genet Resour Crop Evol* 53:1209–1219
- Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. *Mol Breed* 30:1231–1235
- Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease resistance genes by bulked segregant analysis—a rapid method to detect markers in specific genomic regions by using segregating populations. *P Natl Acad Sci USA* 88:9828–9832
- Nielsen NC, Dickinson CD, Cho T-J, Thanh VH, Scallan BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB (1989) Characterization of the glycinin gene family in soybean. *Plant Cell* 1:313–328
- Paek NC, Imsande J, Shoemaker RC, Shibles R (1997) Nutritional control of soybean seed storage protein. *Crop Sci* 37:498–503
- Paek NC, Sexton PJ, Naeve SL, Shibles R (2000) Differential accumulation of soybean seed storage protein subunits in response to sulfur and nitrogen nutritional sources. *Plant Prod Sci* 3:268–274
- 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

- Quarrie SA, Lazic-Jancic V, Kovacevic D, Steed A, Pekic S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *J Exp Bot* 50:1299–1306
- Scallan B, Thanh VH, Floener LA, Nielsen NC (1985) Identification and characterization of DNA clones encoding group-II glycinin subunits. *Theor Appl Genet* 70:510–519
- Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X-C, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183
- Schuler MA, Ladin BF, Pollaco JC, Freyer G, Beachy RN (1982) Structural sequences are conserved in the genes coding for the α , α' and β -subunits of the soybean 7S seed storage protein. *Nucleic Acids Res* 10:8245–8261
- Singh A, Meena M, Kumar D, Dubey AK, Hassan MI (2013) Structural and Functional Analysis of Various Globulin Proteins from Soy Seed. *Crit Rev Food Sci Nutr* 55:1491–1502. <https://doi.org/10.1080/10408398.2012.700340>
- Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8
- Soyatech (2017) Soy facts. http://www.soyatech.com/soy_facts.htm. Accessed 20 Jun 2017
- Takahashi K, Banba H, Kikuchi A, Ito M, Nakamura S (1994) An induced mutant line lacking the alpha-subunit of beta-conglycinin in soybean *Glycine max* (L.) Merrill. *Breed Sci* 44:65–66
- Takahashi K, Shimada S, Shimada H, Takada Y, Sakai T, Kono Y, Adachi T, Tabuchi K, Kikuchi A, Yumoto S (2004) A new soybean cultivar Yumeminori with low allergenicity and high content of 11S globulin. *B Nat Agric Res Cent Tohoku Reg (Jpn)*
- Teraishi M, Takahashi M, Hajika M, Matsunaga R, Uematsu Y, Ishimoto M (2001) Suppression of soybean β -conglycinin genes by a dominant gene, Scg-1. *Theor Appl Genet* 103:1266–1272
- Thanh VH, Shibasaki K (1976) Major proteins of soybean seeds—straightforward fractionation and their characterization. *J Agric Food Chem* 24:1117–1121
- Thanh VH, Shibasaki K (1978) Major proteins of soybean seeds—subunit structure of betaconglycinin. *J Agric Food Chem* 26:692–695
- Tsubokura Y, Hajika M, Kanamori H, Xia Z, Watanabe S, Kaga A, Katayose Y, Ishimoto M, Harada K (2012) The β -conglycinin deficiency in wild soybean is associated with the tail-to-tail inverted repeat of the α -subunit genes. *Plant Mol Biol* 78:301–309
- Tsukada YK, Kitamura KH, Kaizuma N (1986) Genetic analysis of subunits of two major storage proteins (β -conglycinin and glycinin) in soybean seeds. *Jpn J Breed* 36:390–400
- Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. In: 4 edn. Kyazma B.V., Wageningen
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78
- Wang J, Liu L, Guo Y, Wang Y, Zhang L, Jin L, Guan R, Liu Z, Wang L, Chang R, Qiu L (2014) A dominant locus *qBSC-1*, controls β subunit content of seed storage protein in soybean. *J Integ Agric* 13:1854–1864
- Yu KF, Poysa V, Haffner M, Zhang BL, 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, Gleddie S, Khanizadeh S, Cober ER, Guillemette RJD (2007a) Assessment of the protein quality of fourteen soybean [*Glycine max* (L.) Merr.] cultivars using amino acid analysis and two-dimensional electrophoresis. *Food Res Int* 40:129–146
- Zarkadas CG, Gagnon C, Poysa V, Khanizadeh S, Cober ER, Chang V, Gleddie S (2007b) 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