Theor Appl Genet (2015) 128:839–850
DOI 10.1007/s00122-015-2474-4

ORIGINAL PAPER

QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population

C. V. Warrington · H. Abdel-Haleem · D. L. Hyten · P. B. Cregan · J. H. Orf · A. S. Killam · N. Bajjalieh · Z. Li · H. R. Boerma

Received: 2 December 2013 / Accepted: 31 January 2015 / Published online: 12 February 2015

Abstract
Key message **We identified QTL associated with protein and amino acids in a soybean mapping population that was grown in five environments**. **These QTL could be used in MAS to improve these traits**.
Abstract Soybean, rather than nitrogen-containing forages, is the primary source of quality protein in feed formulations for domestic swine, poultry, and dairy industries. As a sole dietary source of protein, soybean is deficient in the amino acids lysine (Lys), threonine (Thr), methionine (Met), and cysteine (Cys). Increasing these amino acids would benefit the feed industry. The objective of the present study was to identify quantitative trait loci (QTL) associated with crude protein (cp) and amino acids in the 'Benning' × 'Danbaekkong' population. The population was grown in five southern USA environments. Amino acid concentrations as a fraction of cp (Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met + Cys/cp) were determined by near-infrared reflectance spectroscopy. Four QTL associated with the variation in crude protein were detected on chromosomes (Chr) 14, 15, 17, and 20, of which, a QTL on Chr 20 explained 55 % of the phenotypic variation. In the same chromosomal region, QTL for Lys/cp, Thr/cp, Met/cp, Cys/cp and Met + Cys/cp were detected. At these QTL, the Danbaekkong allele resulted in reduced levels of these amino acids and increased protein concentration. Two additional QTL for Lys/cp were detected on Chr 08 and 20, and three QTL for Thr/cp on Chr 01, 09, and 17. Three QTL were identified on Chr 06, 09 and 10 for Met/cp, and one QTL was found for Cys/cp on Chr 10. The study provides information concerning the relationship between crude protein and levels of essential amino acids and may allow for the improvement of these traits in soybean using marker-assisted selection.

Communicated by A. Charcosset.

C. V. Warrington · H. Abdel-Haleem (✉) · Z. Li · H. R. Boerma
Institute of Plant Breeding, Genetics and Genomics, University of Georgia, Athens, GA 30602, USA
e-mail: hussein@uga.edu

Present Address:
C. V. Warrington
Schillinger Genetics, Inc., 200 Bennett's Outlet Lane, Queenstown, MD 21658, USA

Present Address:
H. Abdel-Haleem
Arid Land Agricultural Research Center, US Department of Agriculture, Agricultural Research Service, Maricopa, AZ 85238, USA

D. L. Hyten · P. B. Cregan
Soybean Genomics and Improvement Laboratory, US Department of Agriculture, Agricultural Research Service, Beltsville, MD 20705, USA

D. L. Hyten
DuPont Pioneer, 8305 NW 62nd Ave., PO Box 7060, Johnston, IA 50131, USA

J. H. Orf · A. S. Killam
Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN 55108, USA

N. Bajjalieh
Integrative Nutrition, Inc., Decatur, IL, USA

Introduction

As the world population increases, so will be the demand for animal protein. It is projected that global meat production and consumption will increase from 233 million tons in 2000 to 300 million tons by the year 2020 (Annicchiarico 2002). Soybean [*Glycine max* (L.) Merr.] is one of the world's leading oilseed crops and a provider of high-quality protein meal. Soybean meal, a by-product of the oil extraction, drives the soybean market mainly due to its use as a source of amino acids in livestock rations. The animal feed industry uses roughly 77 % of soybean meal as a source of protein and amino acids (Kerley and Allee 2003). With this in mind, its importance to the US agriculture industry cannot be overstated.

Legume seed proteins are categorized as either albumins or globulins based on their solubility patterns. It has been found that most proteins in soybean are globulins, which can be divided into 7S vicilin-type and 11S leguminin-type (Clarke and Wiseman 2000). Glycinin and β-conglycinin represent the 11S and 7S fractions, respectively, based on their sedimentation properties (Danielsson 1949). In combination, the glycinin and β-conglycinin fractions account for roughly 70 % of the storage proteins in soybean seeds (Yaklich et al. 1999). Both have been found to be deficient in the sulfur-containing amino acids cysteine (Cys) and methionine (Met), with the 11S globulins generally higher than the 7S type (Rajcan et al. 2005; Shewry et al. 1995). It has been found that Met and Cys comprise 3.0–4.5 % of the 11S glycinin amino acid residues and less than 1.0 % of the 7S β-conglycinin fractions (Sexton et al. 1998a, b). The β-conglycinin fragment is composed of three subunits: α-, α′-, and β-subunit. The β-subunit lacks both Met and Cys, thus this subunit is primarily responsible for the low concentration of sulfur-containing amino acids in the β-conglycinin fraction. It is likely that a soybean cultivar with a high 11S:7S ratio will have a higher concentration of the S-containing amino acids.

Protein, per se, is not of great importance in terms of animal nutrition. Yet the balance and composition of the amino acid constituents which comprise the protein is likely the most crucial nutritional aspect of meal rations. The major function of protein in nutrition is to supply adequate amounts of required amino acids (Friedman and Brandon 2001). As a sole dietary source of protein, soybean is deficient in the amino acids Met, Cys, threonine (Thr), and lysine (Lys). Each of these amino acids is considered essential, as monogastric animals (e.g., swine and poultry) cannot synthesize these amino acids and therefore each must be obtained solely from the diet. Any deficiency in the amino acid balance must be supplemented in the diet at additional costs to the animal producer. To overcome deficiencies, poultry and swine producers supplement soybean-based rations with synthetically-produced amino acids, which costs approximately $100 million annually (Imsande 2001). Clarke and Wiseman (2000) speculated that a 10 % increase in Lys, Met, and Thr concentrations would yield a $4.5 to $9.5, $2.7, and $5.9/T increase in commercial meal value, respectively. Moreover, according to George and De Lumen (1991), Met supplementation may cause problems such as leaching during soybean meal processing and bacterial degradation leading to the formation of undesirable volatile sulfides. Therefore, the development of soybean cultivars with enhanced amino acid balance would increase their economic value along the entire soybean value chain, from growers to end-users, and reduce any negative environmental effects associated with supplementation.

Due to the aforementioned issues, it is not surprising that the development of soybean cultivars with increased concentrations of essential amino acids has been an objective in the soybean breeding community for some time. The negative correlation between protein and yield has undermined attempts to release cultivars with higher levels of essential amino acids (Wilson 2004). Until recently, the cost of wet lab techniques necessary to evaluate amino acid concentrations made it costly for breeders to thoroughly evaluate protein and amino acid concentrations in large seed samples. Plant breeders have been successful in increasing protein concentrations in soybean (Weber and Fehr 1970; Burton and Wilson 1998), though the concentrations of sulfur-containing amino acids have remained the same (Wilcox and Shibles 2001). Difficulty in breeding for increased amino acid concentrations stemmed from the lack of genetic variability for these traits (Krober 1956). Some high-protein lines have been shown to improve nutritional value (Edwards et al. 2000), but others were unable to detect consistency in increased amino acid concentrations vs. controls (Serretti et al. 1994). Yaklich (2001) found that high protein soybean lines had increased glycinin and β-conglycinin fractions, with some lines having a greater proportion of glycinin polypeptides, suggesting that it is feasible to improve both the quantity and quality of soybean protein. Using ethyl methanesulfonate (EMS) to mutate soybean seeds, Imsande (2001) was able to select several lines which overproduced Met and Cys at approximately a 20 % higher level than the parental lines.

There have been few studies aimed at elucidating the genetic factors underlying amino acid concentration in soybean. Using 101 F_6-derived recombinant inbred lines (RILs), Panthee et al. (2006a) identified QTL associated with Cys (Chr 1, 13, and 18), Met (Chr 7, 13, and 18), and Met + Cys (Chr 7 and 13) concentrations. Panthee et al. (2006b) also identified genomic regions associated with Lys (Chr 1, 15, and 18) and Thr (Chr 2, 5, 9, and 19). Fallen et al. (2013) reported QTL associated with Cys (Chr 20), Met (Chr 13) and Thr (Chr 13). In a similar study, Panthee et al. (2004) used the same RILs to map QTL associated with the 7S and 11S fractions of soybean storage proteins. Since the glycinin fraction contains higher levels of S-containing amino acids

than the β-conglycinin fraction, the identification of genomic regions governing these storage protein constituents would be beneficial in marker assisted selection (MAS) regimes for improved S-containing amino acids. They found three QTL for glycinin (Chr 17, 19, and 20) and two for β-conglycinin concentration (Chr 16 and 17).

To efficiently develop soybean cultivars with improved amino acid profiles, the genetic basis of amino acid content should be explored thereby allowing for the selection of individual components conditioning improved protein quality. The objective of this study was to identify QTL associated with crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met + Cys/cp in a RIL population derived from a cross of 'Benning' and 'Danbaekkong'.

Materials and methods

Plant materials

A population of 140 F_5-derived recombinant-inbred lines (RILs) was developed from a cross of Benning (PI595645) (Boerma 1997) and Danbaekkong (PI619083) (Kim et al. 1996). The parents were chosen based on their protein levels, with Benning averaging approximately 42 % and Danbaekkong at 51 % on a dry-weight basis. Benning is a high-yielding maturity group VII cultivar that is adapted to the southeastern USA and Danbaekkong is a South Korean maturity group IV, tofu cultivar.

The F_1 plants were grown in the greenhouse in Athens GA and harvested individually. The F_2 plants were grown at the Univ. of Georgia Plant Sciences Farm near Athens, GA. Seeds from individual F_2 plants were advanced to the F_5 generation in Athens, GA and Puerto Rico using the single seed descent method (Brim 1966), where the F_3 and F_4 generations were grown in Puerto Rico and the F_5 generation was grown at the Univ. of Georgia Plant Sciences Farm. At maturity individual F_5 plants were single-plant threshed to create F_5-derived recombinant inbred lines (RILs). Approximately 200 RILs were grown in 2003 and 150 RILs were selected for uniform maturity. Based on their relative maturity, RILs were planted in three sets of 50 RILs plus Danbaekkong, and three elite cultivars, 'NC Roy' (Burton et al. 2005), 'AG6202', and 'Boggs-RR' (Boerma et al. 2000) were included in each set in a randomized complete block with two replications. In 2005 and 2006, the experiment was planted at the Univ. of Georgia Plant Sciences Farm near Athens, GA. The soil types were an Appling coarse sandy loam (2005) and a Cecil coarse sandy loam (fine, kaolinitic, thermic Typic Kanhapludults) (2006), and the experiments were irrigated. The experimental plot unit was two 6.1-m long rows with 76-cm between rows, the plot was end-trimmed to 3.7 m prior to harvesting. In 2006, the experiment was planted in Bay AR, in a mixture of Mhoon (Fine-silty, mixed, superactive, nonacid, thermic Fluvaquentic Endoaquepts) and Dundee (Fine-silty, mixed, active, thermic Typic Endoaqualf) fine sandy loam soil type under irrigation in a two rows plot with 76 cm between rows. At Stuttgart AR, (2006) the experiment was planted in a Stuttgart silt loam (Fine, smectitic, thermic Albaquultic Hapludalfsoil) soil type under irrigation. The plot in Stuttgart AR, consisted of two 6.1-m rows with 76 cm between rows. At Kinston NC (2006), the experiment was planted in a Portsmouth soil type (loam with ~4 % organic matter) (Fine-loamy over sandy or sandy-skeletal, mixed, semiactive, thermic Typic Umbraquults) under rain-fed conditions. The plot consisted of three rows spaced 96 cm apart. The plots were end-trimmed to 4.26 m prior harvesting the middle row of each plot.

Crude protein and amino acid determination

The crude protein and amino acid analyses were conducted on 25-g whole seed samples with near-infrared reflectance (NIR) procedure at the Univ. of Minnesota's Soybean Breeding Laboratory and Integrative Nutrition Inc., Decatur, IL. Briefly, whole soybean samples were first ground using a Perten LM 3600 grinder and then scanned on the Perten 7200 diode array NIR Instrument. The NIR equation was developed using an AOSCA approved method for amino acid analysis on the HPLC from the University of Missouri analytical laboratory. The spectra from the NIR was used to predict the amino acid content of the individual lines using ISI Predict Software version 1.10.2.4842 and verified by HPLC. A total of 900 samples were selected from North American soybean cultivars and breeding lines for calibration. The results were reported in g kg^{-1} on a moisture-free basis. Each amino acid was corrected as a percentage of overall crude protein content in g kg^{-1}.

Statistical analyses

Variance-component heritability estimates were calculated on an entry-mean basis (Nyquist and Baker 1991) using the following equation: $h^2 = \sigma^2_{\mathrm{RIL}} / \left[\left(\sigma^2_{\mathrm{RIL}}\right) + \left(\sigma^2_{\mathrm{RIL*E}} / e\right) + \left(\sigma^2_e / re\right)\right]$, where h^2 represents broad-sense heritability, σ^2_{RIL} is genotypic variance, $\sigma^2_{\mathrm{RIL*E}}$ is genotype × environment variance, σ^2_e is error variance, r is the number of replications, and e is the number of environments. Restricted maximum likelihood (REML) method using PROC MIXED procedure of SAS version 9.2 (Statistical Analysis System, SAS Institute 2001) was used to generate components of variance and covariance for calculating genetic correlations. Genetic correlations were based on the formula (Falconer and Mackay 1996): $r_{\mathrm{G}} = \mathrm{Cov}_{xy} \Big/ \left(\sqrt{\sigma^2_x * \sigma^2_y}\right)$, where r_{G} is the

genetic correlation, Cov_{xy} is the genetic covariance of traits x and y, and σ^2 is the genetic variance. Phenotypic correlation coefficients were determined using the PROC CORR procedure of SAS.

Map construction and segregation analysis

The RIL population and parents were grown in the greenhouse, and leaf tissues were collected from 10 unexpanded trifoliate leaf samples. DNA was extracted using a modified CTAB (Hexadecyltrimethylammonium acid) procedure previously described by Keim et al. (1989). RILs and parents were genotyped with 421 polymorphic markers, including 98 simple sequence repeats (SSR) and 323 single nucleotide polymorphisms (SNP). For the SSR marker assays, PCR reactions were based on the protocol of Diwan and Cregan (1997) with slight modifications. The reactions were performed in a dual 384-well or a 96-well GeneAmp PCR System 9700 (PE-ABI, Foster City, CA). The 10 µl PCR reaction mix contained 20 ng of genomic DNA, 0.5 µM of forward and reverse primers (Grant et al. 2002), 2 mM of each dNTP, 2.5 mM Mg^{2+}, 1X PCR buffer (Promega Corp., Madison, WI), and 0.5 units *Taq* polymerase (Promega Corp., Madison, WI). The PCR amplification consisted of an initial denaturation at 94 °C for 1 min, followed by 32 cycles of three steps: denaturation at 94 °C for 30 s, annealing at 46 °C for 30 s, and elongation at 68 °C for 30 s.

Amplification products were separated either on 4.8 % polyacrylamide gels on an ABI PRISM 377 DNA Sequencer (PE ABI, Foster City, CA) or capillary gel electrophoresis using an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA). For ABI PRISM 377, a loading sample for each lane of a gel was prepared with loading mixture (2.0 µl of deionized formamide, 2.0 µl of loading buffer, 0.3 µl of Genescan Rox-500 DNA size standard) and 1–2 µl of the PCR products. DNA was denatured at 95 °C for 5 min, prior to loading. The marker fragments were analyzed with GeneScan and scored with Genotyper software (PE-ABI, Foster City, CA). For ABI 3730 genotyping, 8 µl of a master mix of ROX size standard, water, and formamide was added to 2 µl of DNA, the samples were denatured for 5 min at 95 °C, and then loaded into the capillary system. The fragments were scored with GenMarker software (SoftGenetics, LLC. State College, PA USA). SNP marker genotyping was conducted using the Illumina GoldenGate Assay (Hyten et al. 2008) at the USDA Agricultural Research Center (USDA-ARS) in Beltsville, MD, where a total of 1,536 SNP markers were assayed on the 150 RILs and parents using the Universal Soybean Linkage Panel 1.0 (USLP 1.0) (Hyten et al. 2010b). The Illumina BeadStation 500G (Illumina Inc. San Diego, CA) was used for genotyping the GoldenGate assay and allele calling was completed using the Illumina GenomeStudio software. To overcome the presence of distorted markers, MapDisto v 1.7 mapping software (Lorieux 2012) was used to adjust these distorted markers by estimating the recombination fraction and developing a constant based on the segregation distortion. Polymorphic SSR and SNP markers were used for segregation and linkage analyses using MapDisto v 1.7, or each marker, a χ^2 test was used to identify the markers showing segregation distortion. A LOD threshold of 3.0 was applied to identify initial linkage groups, followed by a conservative logarithm of the likelihood-of-odds (LOD) score of 1.5 to group each LG individually. The classical recombination fraction setting was selected according to Martin (2006). The Kosambi (1944) mapping function was used to address interference. Based on recombination frequencies, 28 linkage groups (LG) were created, which represented the 20 haploid chromosomes in the soybean genome.

QTL analysis

QTL analysis was performed with QTL Cartographer 2.5 (Wang et al. 2012) using genotypic least square means across environments and the Composite Interval Mapping (CIM) method. CIM was employed to detect QTLs and estimate the magnitude of their effects (Jansen and Stam 1994; Zeng et al. 1999), using Model 6 of the Zmapqtl program. The genome was scanned at 2-cM intervals with the window size at 10 cM. Cofactors were chosen using the forward–backward method of the stepwise regression. Series of 1,000 permutations were run to determine the experiment-wise significance level at $P = 0.05$ of LOD for the trait (Churchill and Doerge 1994). The percentage of phenotypic variance explained by specific QTL (R^2) was taken for the peak QTL position as determined by QTL cartographer.

Results

Crude protein and amino acid concentrations variation in the RIL population of Danbaekkong × Benning

Danbaekkong, and three elite cultivars, NC Roy, AG6202, and Boggs-RR, were used as checks in the five environments. Due to its late maturity, Benning (Maturity group 7.8) was not included in field trials in this study. Danbaekkong is a late maturity group IV cultivar, and the RILs averaged 47 days in maturity (after 31 August) compared to 50, 44, and 48 days for AG6202, Boggs-RR, and NC Roy, respectively. Across the five environments, Danbaekkong averaged 510 g kg^{-1} seed protein content while the three elite checks averaged 433 g kg^{-1} (Table 1). The range of

Table 1 Means of crude protein and lysine (Lys), threonine (Thr), methionine (Met), cysteine (Cys) seed amino acids concentrations as fraction of crude protein for Danbaekkong, three check cultivars, and the means and ranges for 140 F_5-derived RILs Benning × Danbaekkong in five environments

Trait (g kg^{-1})	Danbaekkong[a]	Checks[b]	F_5-derived RILs			$LSD_{(0.05)}$
	Mean	Mean	Mean	Min	Max	
Crude protein (cp)	510	433	468.0	425.0	507.0	13.0
Lys/cp[c]	61.2	64.8	063.3	061.2	065.0	1.1
Thr/cp	34.4	37.9	036.1	033.8	038.0	0.7
Met/cp	14.2	14.4	014.2	013.8	014.7	0.4
Cys/cp	15.7	15.3	015.3	014.7	016.2	0.4
Met + Cys/cp	29.9	29.7	029.6	028.5	030.8	0.7

[a] Danbaekkong

[b] Mean of three checks (NC Roy, AG6202, Boggs-RR)

[c] Amino acid was corrected as a percentage of overall crude protein content in g kg^{-1}

seed crude protein content in the RIL population was from 425 to 507 g kg^{-1} with an average of 468 g kg^{-1}. There was no RIL that exceeded Danbaekkong in the crude protein content. The heritability estimate for protein on a five-environment mean basis was high (0.93).

Because the concentrations of amino acids are dependent on the crude protein amount, the amino acid data in the present study are presented as the amount of a specific amino acid per kg of crude protein (cp). For Lysine, the mean Lys/cp value for Danbaekkong was 61.2 g kg^{-1} compared to 64.8 g kg^{-1} for the mean of the three elite checks (Table 1). The population ranged from 61.2 to 65.0 g kg^{-1} with a mean of 63.3 g kg^{-1}. The population was normally distributed for Lys/cp content (Fig. 1a), with no significant ($P = 0.05$) transgressive segregants lower than Danbaekkong or higher than the mean of the three elite check cultivars (Fig. 1a). The heritability estimate for Lys/cp based on the Lys/cp means across five environments was moderately high (0.69). Danbaekkong had a lower Thr/cp value (34.4 g kg^{-1} cp) compared to the average of elite checks (37.9 g kg^{-1} cp). The RILs averaged 36.1 g kg^{-1} for the Thr/cp trait and ranged from 33.8 to 38.0 g kg^{-1} (Table 1). No significant ($P = 0.05$) transgressive segregation was observed (Fig. 1c; Table 1). The heritability estimate for Thr/cp (0.86) trait was higher than Lys/cp based on the aforementioned estimation criteria.

For Met/cp, both Danbaekkong and the mean of the elite checks were not significantly different ($P = 0.05$), where Danbaekkong was 14.2 g kg^{-1} and the elite checks averaged 14.4 g kg^{-1} (Table 1; Fig. 1b). The mean of the RILs was 14.2 g kg^{-1} for Met/cp trait, equal to the parent, Danbaekkong. Transgressive segregation was evident, as the lines varied from 13.8 to 14.7 g kg^{-1} (Fig. 1d; Table 1). Heritability for Met/cp (0.45) was much lower than crude protein or Thr/cp.

The Cys/cp in the three checks averaged 15.3 g kg^{-1}compared to 15.7 g kg^{-1} for Danbaekkong (Table 1). The mean of the RILs was equal to the check means, and the highest value for any RIL was 0.5 g kg^{-1} greater than the Danbaekkong (Fig. 1d). The 140 RILs ranged from 14.7 to 16.2 g kg^{-1}. The heritability for Cys/cp (0.59) was slightly higher than that of Met/cp. Even though both Danbaekkong (29.9 g kg^{-1} cp) and checks (29.7 g kg^{-1}) were not significantly different for Met + Cys/cp, there were transgressive segregants that exceeded the Met + Cys/cp of both Danbaekkong and the checks (Table 1; Fig. 2). The RILs averaged 29.6 g kg^{-1} with a range from 28.5 to 30.8 g kg^{-1} (Table 1). Heritability for this trait (0.53) was intermediate between both of its constituents, Met/cp and Cys/cp, alone.

Based on phenotypic correlation coefficients among traits estimated for RIL means across the five environments, crude protein content was negatively correlated with Lys/cp ($r = -0.63$**), Thr/cp ($r = -0.85$**), Met/cp ($r = -0.19$*), and Cys/cp ($r = -0.16$). Phenotypic correlation coefficients among amino acids were positive, ranging from 0.41 (Thr/cp vs. Cys/cp) to 0.78 (Lys/cp vs. Thr/cp) (Table 2). Met/cp was found to be highly correlated with both Lys/cp ($r = 0.70$**) and Cys/cp ($r = 0.57$**) (Table 2). Genetic correlations, which indicate the direction and magnitude of correlated responses to selection (Falconer and Mackay 1996) were also calculated. Genotypic correlations for crude protein vs. Lys/cp and Thr/cp were −0.82 and −0.91, respectively. The genetic correlation coefficients for crude protein vs. Met/cp were −0.36 and −0.06 for crude protein vs. Cys/cp. For amino acids, genetic correlations ranged from 0.17 (Lys/cp vs. Cys/cp) to 0.92 (Lys/cp vs. Thr/cp) (Table 2).

Linkage map construction and QTL analyses

The markers on linkage groups created from the Benning × Danbaekkong population aligned with the same

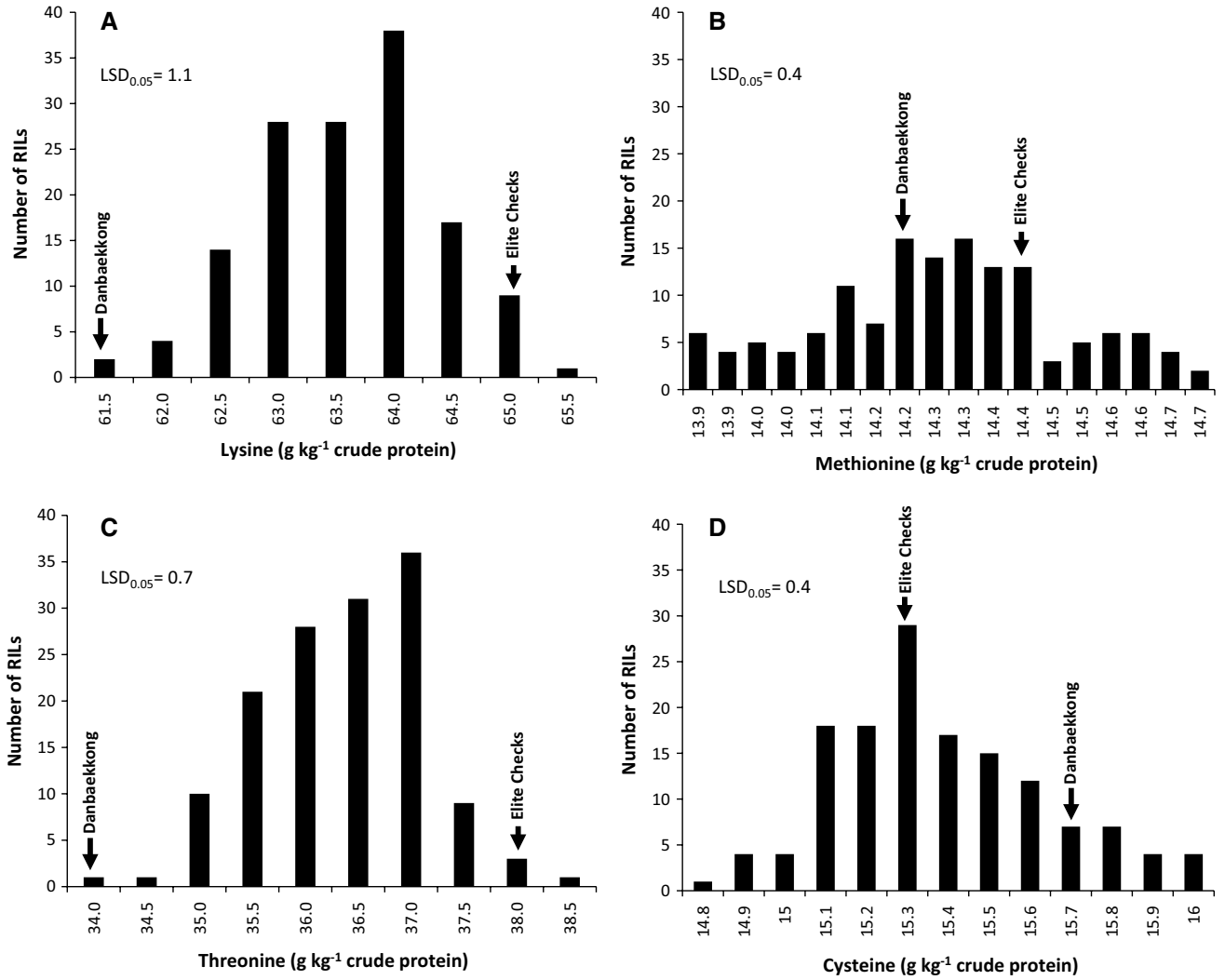

Fig. 1 Distribution of amino acids in the Benning × Danbaekkong RIL population: **a** lysine (Lys/cp), **b** methionine (Met/cp), **c** threonine (Thr/cp), **c**, **d** cysteine (Cys/cp) as fraction of crude protein content (g kg^{-1} crude protein)

markers on the consensus map 4.0 (Hyten et al. 2010a). The marker order on the LGs and the subgroups are in general agreement with those of the integrated genetic linkage map (Consensus Map 4.0). The 421 polymorphic markers (98 SSRs and 323 SNPs) mapped in the RIL population provided a broad coverage of most of the 20 soybean linkage groups with only a few exceptions. Gaps greater than 40 cM (based on the consensus map positions) occurred on Chr 02 (Lg-D1b), Chr 04 (Lg-C1), Chr 06 (Lg-C2), and Chr 14 (Lg-B2). The 28 linkage groups covered 1,124 cM of the genome. Although a fairly large number of markers were used in this study, many SNP markers were clustered in regions of the genome, therefore reducing genome coverage. In the current study, the Benning × Danbaekkong RIL population had four genomic regions with marker distortion skewed towards Benning (Chr 04, 07, 11, and 13) and two regions that were skewed towards Danbaekkong (Chr 16 and 18).

Given that the amino acids were expressed as a percentage of crude protein, it was critical to identify the protein QTLs in this population. Composite interval mapping (CIM) identified four QTL; *qProt_Gm14* (flanked by BARC-018353 and BARC-056587 on Chr 14), *qProt_Gm15* (flanked by Sat_273 and BARC-027786 on Chr 15), *qProt_Gm17* (flanked by BARC-025927 and Satt256 on Chr 17) and *qProt_Gm20* (flanked by GSM0012 and Satt354 on Chr 20) (Table 3; Fig. 3). These QTL explained from 5 to 55 % of the variation in crude protein content with the alleles for increased protein from Danbaekkong (Table 3). The *qProt_Gm20* located on Chr 20 accounted for over five times the variation of the next largest QTL, *qProt_Gm15*.

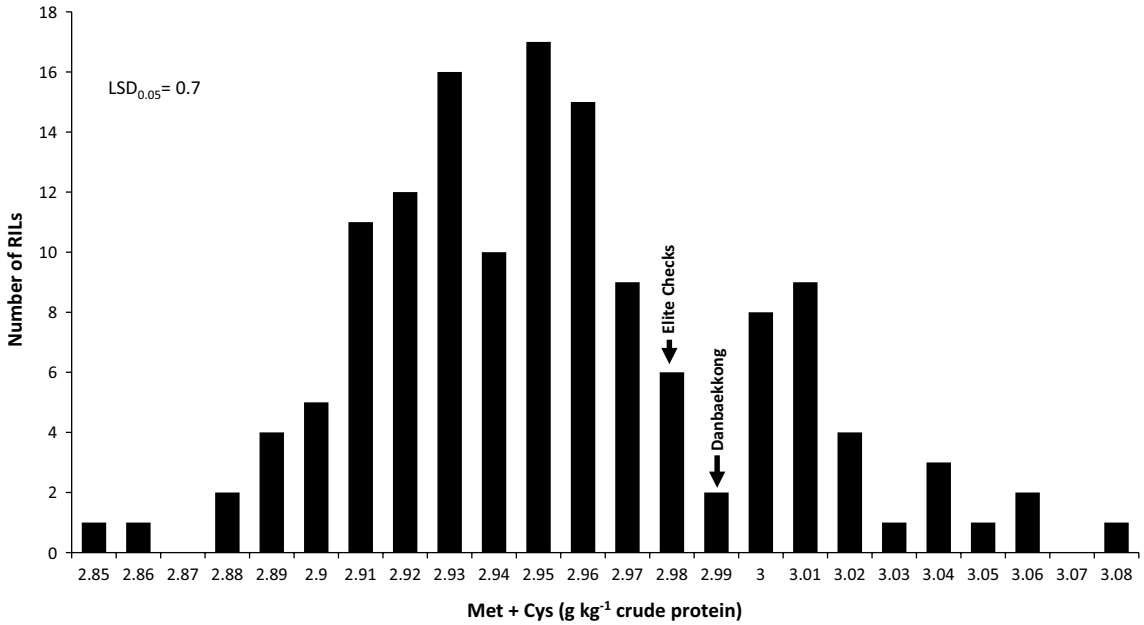

Fig. 2 Distribution of *Met* + *Cys*/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population

Table 2 Phenotypic (upper diagonal in bold font) and genetic correlation (lower diagonal) coefficients among crude protein and lysine (Lys), threonine (Thr), methionine (Met), cysteine (Cys) amino acids as fraction of crude protein for the Benning × Danbaekkong RIL population

Trait	Prot	Lys/cp	Thr/cp	Met/cp	Cys/cp	Met + Cys/cp
Crude protein (prot)		**−0.63****	**−0.85****	**−0.19***	**−0.16**	**−0.20****
Lys/cp	−0.82		**0.78****	**0.70****	**0.41****	**0.63****
Thr/cp	−0.91	0.92		**0.44****	**0.41****	**0.47****
Met/cp	−0.36	0.47	0.57		**0.57****	**0.88****
Cys/cp	−0.06	0.17	0.36	0.76		**0.89****
Met + Cys/cp	−0.21	0.32	0.48	0.92	0.95	

*, ** Significant at the $P \leq 0.01$ and $P \leq 0.001$ level of significance, respectively

Two chromosomal regions associated with Lys/cp concentration were identified on Chr 08 and Chr 20, which explained 6 and 48 % of the phenotypic variation, respectively, and the positive alleles that increased Lys/cp were inherited from Benning (Table 3). The major QTL, *qLys_Gm20* on Chr 20 that was flanked by GSM0012 and Satt354 had the Benning allele that increased Lys/cp by 0.546 g kg^{-1}, while the Benning allele at *qLys_Gm08* QTL increased Lys/cp by 0.187 g kg^{-1} in the same allelic state.

Four QTL were identified for Thr/cp using CIM. Similar to crude protein and Lys/cp a major QTL on Chr 20, *qThr_Gm20,* has the greatest impact and accounted for nine times greater phenotypic variation than the other QTL and explained 52 % of the variation. The Benning allele increased Thr/cp by 0.58 g kg^{-1} at this QTL (Table 3). Each of the other three QTL, on Chr 01 (*qThr_Gm01*), Chr 09 (*qThr_Gm09*), and Chr 17 (*qThr_Gm17*) with favorable alleles from Benning, accounted for 6 % or less of the variation for this trait. The *qMet_Gm20* QTL was responsible for increased Met/cp and accounted for 19 % of the variation (Table 3; Fig. 3). Three other QTL were detected on Chr 06 (*qMet_Gm06* flanked by BARC-055889 and BARC-048217), Chr 09 (*qMet_Gm09* flanked by BARC-31967 and BARC-042449) and Chr10 (*qMet_Gm10* flanked by Satt592 and BARC-043247) and explained 8, 9, and 14 % of the phenotypic variation in Met/cp, respectively (Table 3; Fig. 3). The positive alleles of these QTL contributed to the increases in Met/cp between 0.061 and 0.11 g kg^{-1} (Table 3).

The *qCys_Gm10* QTL on Chr 10 was QTL identified by CIM for Cys/cp and explained 10 % of the variation in cysteine content (Fig. 3). The *qCys_Gm10*

Table 3 QTL for crude protein and lysine (Lys), threonine (Thr), methionine (Met), and cysteine (Cys) amino acids as fraction of crude protein (g kg^{-1}) identified using composite interval mapping (CIM) for the mean of 140 RILs grown in five environments

Trait	QTL name[a]	Chr (Lg)	Position (cM)	Marker interval[c]	CI (cM)[d]	LOD score[e]	a[f]	R^2 (%)[g]	Source of favorable alleles
Protein(cp)	*qProt_Gm14*	14 (B2)	2.0	**BARC-018353**–BARC-056587	0.6–7.9	3.8	4.15	5	Danbaekkong
	qProt_Gm15	15 (E)	33.9	Sat_273–**BARC-027786**	29.2–38.7	4.4	5.57	10	Danbaekkong
	qProt_Gm17	17 (D2)	56.4	BARC-025927–**Satt256**	52.6–60.4	5.1	5.37	9	Danbaekkong
	qProt_Gm20	20 (I)	14.9	**GSM0012**–Satt354	14.7–20.7	29.3	13.64	55	Danbaekkong
Lys/cp	*qLys_Gm08*	08 (A2)	34.6	**BARC-055265**–Satt089	31.7–35.6	3.4	−0.19	6	Benning
	qLys_Gm20	20 (I)	14.9	**GSM0012**-Satt354	14.7–20.6	23.3	−0.54	48	Benning
Thr/cp	*qThr_Gm01*	01 (D1a)	58.7	**BARC-035219**–BARC-018211	53.1–64.4	3.3	−0.19	6	Benning
	qThr_Gm09	09 (K)	31.4	BARC-31967–**BARC-042449**	30.8–32.0	4	−0.18	5	Benning
	qThr_Gm17	17 (D2)	60.0	**Satt256**-BARC–039151	51.8–61	3.7	−0.19	6	Benning
	qThr_Gm20	20 (I)	14.9	**GSM0012**–BARC-020713	14.6–16.1	26.2	−0.58	52	Benning
Met/cp	*qMet_Gm06*	06 (C2)	31.7	**BARC-055889**–BARC-048217	28.2–33.0	3.6	0.06	8	Danbaekkong
	qMet_Gm09	09 (K)	31.4	BARC-31967–**BARC-042449**	30.8–32.0	4.1	−0.11	9	Benning
	qMet_Gm10	10 (O)	56.8	**Satt592**–BARC_043247	52.9–57.8	5.9	−0.08	14	Benning
	qMet_Gm20	20 (I)	15.9	GSM0012–**BARC-020713**	14.6–17.0	7.7	−0.09	19	Benning
Cys/cp	*qCys_Gm10*	10 (O)	56.8	**Satt592**–BARC_043247	52.7–57.8	3.5	−0.09	10	Benning
Met + Cys/cp	*qMet + Cys_Gm10*	10 (O)	55.8	**Satt592**–BARC_043247	52.0–57.8	4.3	−0.15	11	Benning
	qMet + Cys_Gm20	20 (I)	16.9	GSM0012–**BARC-020713**	14.6–19.5	4.4	−0.15	12	Benning

[a] QTL nomenclature is in the form of qTrait name_chromosome name

[b] QTL position in cM starts from the top of the designated linkage group

[c] Marker interval with nearest marker to the QTL peak in bold

[d] Confidence interval based on ±1 LOD

[e] LOD is the log-likelihood at QTL peak position

[f] Additive effect of the difference in the studied trait at a marker locus homozygous for Danbaekkong vs. homozygous for Benning. A positive value indicates the allele for the increased trait value is inherited from Danbaekkong

[g] Proportion of phenotypic variance explained by specific QTL of the CIM model

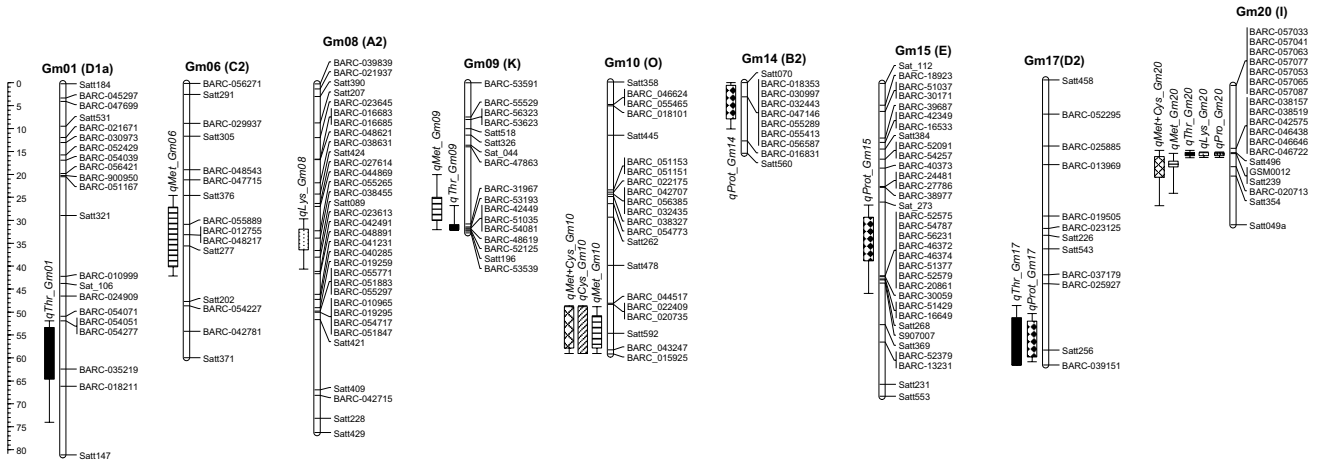

Fig. 3 Composite interval mapping for crude protein (prot) (*diamond bar*), lysine (Lys) (*dotted bar*), threonine (Thr) (*solid bar*), methionine (Met) (*horizontal strip bar*), cysteine (Cys) (*diagonal strip bar*) and methionine + cysteine (Met + Cys) (*lattice bar*) QTL in the Benning × Danbaekkong RIL population based on multi-environments data. QTL nomenclature is in the form of qTrait name_chromosome name. Length of *QTL bar* indicates the LOD-1 confidence interval and QTL *error bar* is extended to LOD-2 confidence interval from the maximum likelihood value

inherited its positive allele from Benning, which provided a 0.088 g kg^{-1} increase in Cys/cp production. Two QTL, *qMet + Cys_Gm10* and *qMet + Cys_Gm20,* were discovered for Met + Cys/cp and explained 11 % and 12 % of the phenotypic variation in Met + Cys/cp, respectively (Table 3). The positive alleles of these QTL were inherited from Benning.

Discussion

In the present study, soybean recombinant inbred lines were discovered with crude protein greater than 500 g kg^{-1}, whereas, the average protein content of most of the high seed protein lines is <500 g kg^{-1} (Yaklich 2001). This finding was not surprising as the Danbaekkong parent averages 510 g $kg^{-1}$of protein. The amino acid values as a percentage of crude protein from this study are similar to those set forth by the National Research Council (1994, 1998) for poultry and swine nutrition, though far from the trait end points proposed by the United Soybean Board's Better Bean Initiative (Sallstrom 2002; Bajjalieh 2004).

Since Benning was not grown in this study, the amount of transgressive segregation can only be interpreted in relation to the Danbaekkong parent. Transgressive segregation is proof that favorable QTL alleles are inherited from both parental genotypes. Otherwise, if all the positive alleles came from just one of the parents, the highest value for the progeny would be similar to that of the highest parent. This was the case for Lys/cp and Thr/cp, for which no significant transgressive segregation was observed. Our findings show that Danbaekkong contributed all of the favorable alleles for crude protein (Table 3). Conversely, at detected QTL, only Benning alleles conditioned increased levels of Lys/cp, Thr/cp, Cys/cp, and Met + Cys/cp (Table 3). Mostly Benning alleles and a single Danbaekkong allele conditioned increased levels of Met/cp (Table 3).

Segregation distortion has been observed in populations of a wide range of organisms including fungi, plants, insects, and mammals (Lyttle 1991). In plants, segregation distortion is a common phenomenon in different mapping populations (F2, DH, RIL), and RIL populations have the highest probability of distortions due to continued selfing for 5 to 6 generations (Singh et al. 2007; Xu et al. 1997). In our study, segregation distortion showed some of systematic patterns along parts of the genome, where the skewed loci tend to cluster in the distorted regions on the chromosomes. Several reports found dual directions of distorted loci toward both parents (Sirithunya et al. 2002; Xu et al. 1997; Cuesta-Marcos et al. 2009). Despite the distribution of distorted loci clusters in Benning × Danbaekkong population on Chr04, 07, 11, 13, 16 and 18, none of the linked markers to the QTL was significantly distorted. As well none of these distorted markers were associated with maturity QTL. That suggests the selection for uniform maturity dates did not impact the mapping of QTL associated with protein and amino acids in this population.

The *qProt_Gm20* has the largest effect conditioning crude protein, and has been identified previously in other soybean populations (Brummer et al. (1997), Diers et al. (1992), Sebolt et al. (2000), Chung et al. (2003), and Mansur et al. (1993). Nichols et al. (2006) fine-mapped this QTL using two sets of backcross lines to the region between SSR marker Satt239 and AFLP marker ACG9b.

In our study, we detected a SNP marker (BARC_016899) within this interval which is associated with crude protein, Lys/cp, Thr/cp, Met/cp, and Met + Cys/cp contents. SoyBase (http://soybase.org/) reports a number of protein and protein-related QTL within 10 cM upstream or downstream of those identified in our study. Eleven QTL associated with protein are reported in SoyBase near the QTL reported in the present study. These include Prot 1-1, 1-2, 1-3, 1-4, 3-12, 11-1, 15-1, 17-1, and the previously mentioned cqPro-003 QTL. Two QTL (Prot 4-5 and 4-6) were detected on Chr 15 approximately 6 cM from our QTL for crude protein. Interestingly, three QTL relevant to protein quality (Glycinin 1-1, Acidic fraction 1-1, and Conglycinin 1-1) reported in SoyBase were located at approximately 8 cM from *qProt_Gm17* and *qThr_Gm17*, the QTL detected in the region for crude protein and Thr/cp in the current study. Protein quality QTL (Acidic fraction 1-3) and a protein QTL (Prot 13-4) were reported to be located at 2 and 4 cM from *qThr_Gm09* and *qMet_Gm09* QTL, respectively. Another QTL for protein (Prot 24-1) was mapped to the similar region of *qProt_Gm06* in our study. Few studies have undertaken the task of elucidating the genomic regions associated with amino acid concentration in soybean. Previously, Panthee et al. (2006a, b) reported QTL associated with Cys (Chr 01, 13, and 18), Met (Chr 07, 13, and 18), Lys (Chr 01, 15, and 18), Thr (Chr 02, 05, 09, and 19) and Met + Cys (Chr 07 and 13) on a dry weight basis. For the sulfur-containing amino acids, no QTL reported in these studies were at the same chromosome regions identified in our mapping population. Panthee et al. (2004) mapped QTL associated with glycinin (Chr 17, 19, and 20) and β-conglycinin (Chr 16 and 17) concentrations. Since the glycinin fraction contains higher levels of S-containing amino acids than the β-conglycinin fraction, the identification of genomic regions governing these storage protein constituents would be beneficial for improved S-containing amino acids. The *qProt_Gm20* in the present study is located 45 cM downstream from glycinin QTL and *qProt_Gm17* and *qThr_Gm17* were located at approximate 8 and 43 cM from the β-conglycinin QTL detected in their study, respectively. The *qMet_Gm20 and qMet + Cys_Gm20* were detected and overlapped in the same chromosomal region (Fig. 3). These QTL had the highest effect for Cys and Met + Cys that explained 19 and 20 % of the phenotypic variation in these traits, respectively. No QTL was detected for Cys alone in the same region. *qMet_Gm10, qCys_Gm10 and qMet + Cys_Gm10* were detected and overlapped in the same region of chromosome 10. These QTL were among the highest QTL for these traits. This suggests that analyzing sulfur-containing amino acids together could detect more significant QTL compare to conducting the QTL analysis for each trait separately.

We observed high negative genetic correlation between protein-Lys/cp and protein-Thr/cp, and a high positive genetic correlation for Lys/cp–Thr/cp. Our QTL mapping results provides a genetic explanation for these associations and also points to breeding issues that should be considered when attempting to modulate these traits. The major protein QTL, *qProt_Gm20* is at the same genomic location as a major QTL for Lys/cp, *qLys_Gm20*, and Thr/cp, *qThr_Gm20*. Each of these QTL explain 48 % or more of the phenotypic variation in their respective traits and one of them, *qProt_Gm20,* obtains its positive allele from Benning, while at the other two QTL the allele from Danbaekkong increases the amino acid content. This provides a genetic explanation for both the negative association between protein content and Lys/cp or Thr/cp and the positive association between Lys/cp and Thr/cp.

It is evident that the crude protein QTL detected in the present study on Chr 20 also plays some role in the sequestration of the amino acids within the seed. It is clear from both phenotypic and genotypic correlations that an increase in crude protein results in decreased values of these amino acids as a proportion of crude protein (Table 3). The negative correlation between crude protein and Thr is especially strong. In terms of the sulfur-containing amino acids, it is known that the synthesis of storage protein within the developing soybean seed is sensitive to Met concentration, therefore, the presence of Met during this process prevents the synthesis of low quality proteins; in other words if the plants are grown in sulfur deficient soils, then poor quality seed storage proteins will be synthesized (Sexton et al. 2002).

In plants, Met, Lys, and Thr are part of the aspartate family of amino acids, thus synthesized from the same precursor, aspartate. Aspartate is the carboxylate ion, or ester, of the non-essential amino acid aspartic acid which and is vital in the biosynthesis of these amino acids (Shen et al. 2002). Therefore, it is not surprising that these three amino acids were correlated. It is of interest to look at aspartate kinase and aspartate semialdehyde dehydrogenase, as they are the first two enzymes which function in the pathway and could therefore be responsible for increased or reduced levels of each of these amino acids.

These results provide direction for various breeding strategies to improve protein quantity and protein quality of existing southern US soybean cultivars. For example, if the breeding goal is to increase overall protein content the introgression of Danbaekkong allele at *qProt_Gm20* into an existing an elite soybean cultivar background would likely enhance crude protein content by approximately 27 g kg^{-1} in a homozygous line. Of course, on the average this would result in a reduction of 1.1, 1.2, and 0.3 g kg^{-1} in Lys/cp, Thr/cp, and Met + Cys/cp, respectively. If the breeding goal is to enhance the protein quality (increase the

amino acid content of Lys/cp, Thr/cp, or Met + Cys/cp), the options are limited in this population. On could slightly increase the Met/cp by selection of the Danbaekkong allele at *qMet_Gm06*. On the average this would only increase Met/cp by 0.12 g kg^{-1} in a homozygous line when compared to existing elite southern cultivars.

This study reinforces the fact that breeding efforts for soybean protein quality should not focus completely on increasing protein concentration alone. Simply increasing crude protein may not increase essential amino acid concentrations. Mapping both crude protein and amino acids concurrently within the same population allows a better understanding of the interaction between alleles conditioning protein and amino acids and how to best proceed with marker-assisted selection (MAS). At the present time, not a single commercial soybean cultivar with the FAO standard for total sulfur containing amino acids has been developed. This is primarily due to the pitfalls described in this paper. With continued improvements in breeding technologies in conjunction with continued elucidation of quantitative traits, it is foreseeable in the future that a soybean cultivar with high yield, high protein, and improved levels of essential sulfur-containing amino acids will be developed.

Author contribution statement HRB and CVW were responsible for population development and generating the field and SSR data. JHO, ASK, and NB conducted the phenotypic amino and protein assays. PBC and DLH were responsible for generating the SNP data. CVW, HA, and ZL were responsible for the linkage mapping and QTL analyses. CVW prepared the initial draft of the manuscript. All authors edited and approved the final draft of the manuscript.

Acknowledgments This research was supported by funds allocated to the Georgia Agricultural Experiment Stations and grants from the United Soybean Board.

Conflict of interest The authors declare that they have no conflicts of interests.

Ethical standards All experiments described in this manuscript comply with the current laws of the country in which they were performed.

References

Annicchiarico P (2002) Genotype × environment interactions: challenges and opportunities for plant breeding and cultivar recommendations. Food & Agriculture Org

Bajjalieh N (2004) Proteins from oilseeds. In: Proceedings on Protein sources for the animal feed industry. Expert consultation and workshop, Bangkok

Boerma HR, Hussey RS, Phillips DV, Wood ED, Rowan GB, Finnerty SL (1997) Registration of 'Benning' soybean. Crop Sci 37:1982

Boerma H, Hussey R, Phillips D, Wood E, Rowan G, Finnerty S (2000) Registration of 'Boggs' soybean. Crop Sci 40(1):294–295

Brim C (1966) A modified pedigree system of selection in soybeans. Crop Sci 6:220

Brummer E, Graef G, Orf J, Wilcox J, Shoemaker R (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37(2):370–377

Burton J, Wilson T (1998) Registration of 'Prolina'soybean. Crop Sci 39(1):294

Burton J, Carter T Jr, Bowman D (2005) Registration of NC-Roy soybean. Crop Sci 45:2654

Chung J, Babka H, Graef G, Staswick P, Lee D, Cregan P, Shoemaker R, Specht J (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci 43(3):1053

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138(3):963

Clarke E, Wiseman J (2000) Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. The Journal of Agricultural Science 134(02):111–124

Council NR (1994) Nutrient Requirements of Poultry. National Academy Press, Washington

Council NR (ed) (1998) Nutrient Requirements of Swine, 10th edn. National Academy Press, Washington

Cuesta-Marcos A, Casas AM, Hayes PM, Gracia MP, Lasa JM, Ciudad F, Codesal P, Molina-Cano JL, Igartua E (2009) Yield QTL affected by heading date in Mediterranean grown barley. Plant Breeding 128(1):46–53

Danielsson C (1949) Seed globulins of the Gramineae and Leguminosae. Biochem J 44(4):387

Diers B, Keim P, Fehr W, Shoemaker R (1992) RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83(5):608–612

Diwan N, Cregan P (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. TheorApplGen 95(5):723–733

Edwards H 3rd, Douglas M, Parsons C, Baker D (2000) Protein and energy evaluation of soybean meals processed from genetically modified high-protein soybeans. Poult Sci 79(4):525

Falconer D, Mackay T (1996) Introduction to quantitative genetics (1996). Longman, Essex

Fallen B, Hatcher C, Allen F, Kopsell D, Saxton A, Chen P, Kantartzi S, Cregan P, Hyten D, Pantalone V (2013) Soybean Seed Amino Acid Content QTL Detected Using the Universal Soy Linkage Panel 1.0 with 1,536 SNPs. J Plant Genome Sci 1(3):68–79

Friedman M, Brandon D (2001) Nutritional and Health Benefits of Soy Proteins. J Agric Food Chem 49(3):1069–1086

George AA, De Lumen BO (1991) A novel methionine-rich protein in soybean seed: identification, amino acid composition, and N-terminal sequence. J Agric Food Chem 39(1):224–227

Grant D, Imsande MI, Shoemaker RC (2002) SoyBase, the USDA-ARS soybean genome database. http://soybase.org. Accessed 8 Feb 2015

Hyten D, Song Q, Choi I, Yoon M, Specht J, Matukumalli L, Nelson R, Shoemaker R, Young N, Cregan P (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116(7):945–952

Hyten D, Choi I, Song Q, Specht J, Carter Jr T, Shoemaker R, Hwang E, Matukumalli L, Cregan P (2010a) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping

Hyten DL, Choi IY, Song Q, Specht JE, Carter TE Jr, Shoemaker RC, Hwang EY, Matukumalli LK, Cregan PB (2010b) A high density integrated genetic linkage map of soybean and the development

of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci 50(3):960–968
Imsande J (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. Crop Sci 41(2):510–515
Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136(4):1447
Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theorical and Applied Genetics 77:786–792
Kerley MS, Allee GL (2003) Modifications in soybean seed composition to enhance animal feed use and value: moving from a dietary ingredient to a functional dietary component. AgBioForum 6(1&2):14–17. http://www.agbioforum.org. Accessed 8 Feb 2015
Kim S-D, Hong E-H, Kim Y-H, Lee S-H, Seong Y-K, Park K-Y, Lee Y-H, Hwang Y-H, Park E-H, Kim H-S, Ryu Y-H, Park R-K, Kim Y-S (1996) A new high protein and good seed quality soybean variety "Danbaegkong". RDA J Agri Sci Upl Ind Crops 38:228–232
Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12(1944):172–175
Krober O (1956) Nutritive quality of crops, methionine content of soybeans as influenced by location and season. J Agric Food Chem 4(3):254–257
Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. Mol Breed 30:1231–1235. doi:10.1007/s11032-012-9706-y
Lyttle TW (1991) Segregation distorters. Annu Rev Genet 25(1):511–581
Mansur L, Lark K, Kross H, Oliveira A (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). TAG. Theor Appl Genet 86(8):907–913
Martin OC (2006) Two-and three-locus tests for linkage analysis using recombinant inbred lines. Genetics 173(1):451
Nichols D, Glover K, Carlson S, Specht J, Diers B (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Sci 46(2):834
Nyquist W, Baker R (1991) Estimation of heritability and prediction of selection response in plant populations. Crit Rev Plant Sci 10(3):235–322
Panthee D, Kwanyuen P, Sams C, West D, Saxton A, Pantalone V (2004) Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. J Am Oil Chem Soc 81(11):1005–1012
Panthee D, Pantalone V, Sams C, Saxton A, West D, Orf J, Killam A (2006a) Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor Appl Genet 112(3):546–553
Panthee D, Pantalone V, Saxton A, West D, Sams C (2006b) Genomic regions associated with amino acid composition in soybean. Mol Breeding 17(1):79–89
Rajcan I, Hou G, Weir A (2005) Advances in breeding of seed-quality traits in soybean. J Crop Improv 14(1):145–174
Sallstrom JR (2002) Better Bean Initiative (BBI)—a tool to enhance competitiveness for the US soybean producer. In: Biennial Conference of the Cellular and Molecular Biology of the Soybean, Urbana-Champaign, 11–14 Aug 2002
Sebolt A, Shoemaker R, Diers B (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci 40(5):1438
Serretti C, Schapaugh W Jr, Leffel R (1994) Amino acid profile of high seed protein soybean. Crop Sci 34(1):207
Sexton P, Naeve S, Paek N, Shibles R (1998a) Sulfur availability, cotyledon nitrogen: sulfur ratio, and relative abundance of seed storage proteins of soybean. Crop Sci 38(4):983–986
Sexton PJ, Paek NC, Shibles RM (1998b) Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. Field Crops Res 59(1):1–8
Sexton P, Paek N, Naeve S, Shibles R (2002) Sulfur metabolism and protein quality of soybean. J Crop Prod 5(1):285–308
Shen B, Li C, Tarczynski MC (2002) High free methionine and decreased lignin content result from a mutation in the Arabidopsis S-adenosyl L-methionine synthetase 3 gene. Plant J 29(3):371–380
Shewry P, Napier J, Tatham A (1995) Seed storage proteins: structures and biosynthesis. The Plant Cell Online 7(7):945
Singh K, Ghai M, Garg M, Chhuneja P, Kaur P, Schnurbusch T, Keller B, Dhaliwal H (2007) An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* × *T. monococcum* RIL population. Theor Appl Genet 115(3):301–312
Sirithunya P, Tragoonrung S, Vanavichit A, Pa-In N, Vongsaprom C, Toojinda T (2002) Quantitative trait loci associated with leaf and neck blast resistance in recombinant inbred line population of rice (*Oryza sativa*). DNA Res 9(3):79–88
Wang S, Basten CJ, Zeng Z-B (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm
Weber C, Fehr W (1970) Registration of Provar soybeans. Crop Sci 10:728
Wilcox J, Shibles R (2001) Interrelationships among seed quality attributes in soybean. Crop Sci 41(1):11
Wilson R (2004) Seed composition. Soybeans Improv Prod Uses 3:621–677
Xu Y, Zhu L, Xiao J, Huang N, McCouch S (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). Mol Gen Genet MGG 253(5):535–545
Yaklich R (2001) [beta]-Conglycinin and glycinin in high-protein soybean seeds. J Agric Food Chem 49(2):729–735
Yaklich R, Helm R, Cockrell G, Herman E (1999) Analysis of the distribution of the major soybean seed allergens in a core collection of Glycine max accessions. Crop Sci 39(5):1444
Zeng ZB, Kao CH, Basten CJ (1999) Estimating the genetic architecture of quantitative traits. Genet Res 74(03):279–289