Food and Nutrition Sciences, 2015, 6, 974-984
Published Online August 2015 in SciRes. http://www.scirp.org/journal/fns
http://dx.doi.org/10.4236/fns.2015.611101

Mapping of QTL Associated with Seed Amino Acids Content in "MD96-5722" by "Spencer" RIL Population of Soybean Using SNP Markers

Laila Khandaker[1], Masum Akond[1], Shiming Liu[2], Stella K. Kantartzi[2], Khalid Meksem[2], Nacer Bellaloui[3], David A. Lightfoot[2], My Abdelmajid Kassem[1,4*]

[1]Plant Genomics and Biotechnology Laboratory, Department of Biological Sciences, Fayetteville State University, Fayetteville, NC, USA
[2]Department of Plant, Soil and Agricultural Systems, Southern Illinois University, Carbondale, IL, USA
[3]Crop Genetics Research Unit, Agricultural Research Service, USDA, Stoneville, MS, USA
[4]The School of Arts and Sciences, American University of Ras Al Khaimah, Ras Al Khiamah, United Arab Emirates
Email: [*] my.kassem@aurak.ae

Received 19 February 2015; accepted 8 August 2015; published 13 August 2015

Abstract

Limited information is available on the genetic analysis of amino acid composition in soybean seeds. Previously, quantitative trait loci (QTLs) for seed isoflavones, protein, oil, and fatty acids were identified in the "MD96-5722" by "Spencer" and other RIL populations. There were wide variations for these seed constituents among the RIL populations. Therefore, the objective of this study was to identify QTLs controlling different amino acids content in soybean seeds. To achieve this objective, ninety-two $F_{5:7}$ recombinant inbred lines (RIL), developed from a cross of MD96-5722 and Spencer, using a total 5376 Single Nucleotide Polymorphism (SNPs) markers, were used. The RILs were genotyped by using 537 polymorphic, reliably segregating SNP markers, developed from the Illumina Infinium SoySNP6K BeadChip array. A total of 13 QTLs were identified with three QTLs for threonine on the linkage group (LG) A1, C2, and B2. Two QTLs were identified for each of the amino acids proline on LG D1a and B2, serine on LG A1 and C2, tryptophan on LG K and G, and cysteine on LG A1 and K. One QTL was identified for arginine on LG N and histidine on LG J. The new QTLs findings for seed amino acid will facilitate the development of soybean cultivars with higher protein and amino acid quality to help meet the industry and consumer needs.

[*]Corresponding author.

How to cite this paper: Khandaker, L., Akond, M., Liu, S.M., Kantartzi, S.K., Meksem, K., Bellaloui, N., Lightfoot, D.A. and Kassem, M.A. (2015) Mapping of QTL Associated with Seed Amino Acids Content in "MD96-5722" by "Spencer" RIL Population of Soybean Using SNP Markers. *Food and Nutrition Sciences*, **6**, 974-984. http://dx.doi.org/10.4236/fns.2015.611101

Keywords

Soybean, Amino Acids, SNP, QTLs, Seed Composition, Constituents

1. Introduction

Soybean [*Glycine max* (L.) Merr.] is a legume crop, and its seeds are among few plant-derived food sources that contain useful concentrations of all essential amino acids [1]. Amino acids (AA) are key elements for proper growth and development of animals, and both essential and non-essential amino acids are important for healthy diet [2]. In soybean, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine are essential AA for humans, but histidine is considered semi-essential, whereas arginine, alanine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, proline, serine, and tyrosine are considered non-essential AA [2]. The amounts and proportions of essential AA in animal diets should be optimal because AA are the building blocks of protein, and play fundamental roles in the formation, stability, and function of proteins [3].

The nutrient content in soybean meal depends on the genotype, environment in which the plants are grown, and, to a lesser extent, meal processing [4]. Several biotechnological approaches have been employed with varying degrees of success to enhance the level of certain essential AA in soy protein to meet the needs of desired feed formulations [5]. To date, very few studies are available on the genetic analysis of amino acid composition in soybean in spite of the several quantitative trait loci (QTLs) identified for amino acids in soybean, rice, and maize [6]-[10]. For example, genomic regions associated with amino acid composition in soybean were identified using 101 F6-derived recombinant inbred lines developed from a cross of N87-984-16 x TN93-99 [11]. They found a significant ($p < 0.01$) difference among the RIL for amino acid composition and a total of 94 polymorphic simple sequence repeats (SSRs) molecular markers. They found that at least one QTL for each amino acid in this population, and 3 QTLs for cysteine and 4 QTLs for methionine. Also, they reported that phenotypic variation explained by an individual QTL ranged from 9.4% to 45.3%. Using the same population, QTLs for all AA, except asparagine and glutamine, were identified and mapped on the soybean genome. In a more recent study, a total of ten QTLs were detected in the "Essex" × "Williams" 82 RIL population [11]. They [11] identified QTL controlling soybean seed amino acids by using the Universal Soy Linkage Panel 1.0 with 1536 SNPs. They analyzed essential and non-essential amino acid composition in soybean seed proteins by using 282 F5:9 recombinant inbred lines developed from a cross of Essex × Williams 82 [11]. They used the Universal Soy Linkage Panel (USLP) 1.0 of 1536 single nucleotide polymorphism (SNP) DNA markers to genotype the 282 RILs and identify 480 useful genetic markers [11]. They also found significant differences ($p < 0.05$) in the content of amino acids in all amino acids among genotypes in the population, and a total of ten QTLs were detected on chromosomes 5, 7, 9, 10, 13 and 20 [11]. They [11] reported that these QTLs explained 5% to 14% of the total phenotypic variation for a given amino acid. They concluded that detecting QTLs for seed amino acids in soybean would provide useful information to select genotypes with improved seed amino acid qualities for human nutrition and livestock production [11].

Although soybean seed protein represents an energy-efficient source of amino acids used in livestock feed, protein alone as a dietary source for animals is deficient in the sulfur-containing amino acids methionine, cysteine, and may contain sub-optimum concentrations of threonine, and lysine. Therefore, enhancing these essential amino acids would improve the nutritive value of soybean meal and provide additional value to the animal feed industry. Others [12] studied QTL mapping and resource allocation for improved soybean seed amino acids, especially sulfur-containing amino acids such as methionine, cysteine, which were deficient in soybean seed. It was able to detect QTL for crude protein (cp), lysine/cp, threonine/cp, methionine/cp, cysteine/cp, and methionine + cysteine/cp by using DNA markers, and a major protein QTL was found to be associated with reduced sulfur-containing amino acids [12]. They concluded that selection for amino acid QTLs on other chromosomes may enhance protein quality and maintain high crude protein, and further advance our understanding of genetic basis of amino acids composition. Understanding the genetic basis of seed amino acids composition is critical for soybean breeders to develop cultivars with improved amino acid profiles and improved protein quality using marker assisted selection (MAS). Therefore, the objective of this study was to identify QTLs underlying amino acid contents in the seeds of 92 $F_{5:7}$ recombinant inbred lines developed from a cross between MD96-5722 and

"Spencer" using a total 5376 Single Nucleotide Polymorphism (SNPs) markers.

2. Materials and Methods

2.1. Plant Material

Ninety-two $F_{5:7}$ recombinant inbred lines (RIL) population developed by crossing line MD 96-5722 × Spencer were grown at Fayetteville State University, Fayetteville, NC in 2012 with plant to plant distance 25 cm; no additional fertilizer or insecticide was used. Seeds of MD96-5722 and "Spencer" were obtained from the National Soybean Research Laboratory (NSRL) in 2006 [13]. Details of development of RIL population can be found elsewhere [13].

2.2. Amino Acid Analyses

Harvested seeds were analyzed for AA contents (%) of threonine (THR), serine (SER), proline (PRO), cysteine (CYS), histidine (HIS), arginine (ARG), and tryptophan (TRP). Contents of AA were analyzed by using near-infrared (NIR) reflectance diode array feed analyzer (Perten, Spring Field, IL) as described previously elsewhere [6] [7] [14]. A sample of approximately 25 g of seed from each plot was ground using a Laboratory Mill 3600 (Perten, Springfield, IL) as recommended by others [15] [16]. Primary calibration equations for AA quantifications were developed by the Department of Agronomy and Plant Genetics, University of Minnesota St Paul, MN, using Thermo Galactic Grams PLS IQ software developed by Perten Company (Perten, Springfield, IL). The development of the updated calibration equation was based on the use of the Association of Official Analytical Chemists (AOAC) methods [17] and use of 8540 samples spectra, resulting in sufficiently accurate estimations of AA contents. Determination of AA content was performed on the basis of percent dry matter of the seeds.

2.3. Genetic Map Construction and QTL Identification

The MD96-5722 by Spencer populations were genotyped by using SoySNP6K Illumina Infinium BeadChip array which assayed 5376 SNPs. There were 1465 polymorphic SNPs that fit the 1:1 segregation ratio and had less than 20% of lines with no call (missing data plus heterozygosity) among the RILs. Among the 1456, 657 were polymorphic between the single plants chosen to represent the parents DNAs. Heterogeneity was over 20% in individuals within both parents. The 657 SNPs clearly associated with either of the parents were mapped. The genetic linkage map [13] was constructed through Join Map 4 (Kyazma BV, Wageningen, the Netherlands) [18]. Composite interval mapping (CIM) was used to detect QTL through WinQTLCart 2.5 software (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm) [13] using the genetic linkage map [13] and phenotypic data. The Model 6 with four parameters for forward and backward stepwise regression, 10 cM window size, 1 cM step size and five (5) control markers were chosen for running WinQTLCart [13]. The threshold was determined by permutations in 1000 times. A total of 550 of the 657 reliable SNPs formed 16 linkage groups (LGs) located on 16 of the 20 chromosomes [1]-[6] [9] [10] [12]-[16] [18]-[20] and of the soybean genome. The total map length was 201.57 centiMorgans (cM) with an average marker density of 0.37 cM.

Statistical analyses were conducted by SAS [20]. Analysis of Means (CV, maximum and minimum values, and SD) were carried out using Proc Means in SAS. Correlations were conducted by SAS using PROC REG.

3. Results

3.1. Variation in Amino Acid Composition

Analysis of variance [20], mean values, standard deviation of parental and RIL lines for their AA contents (%) are presented in **Table 1**. Lowest variations among RIL lines were found for PRO (CV = 3.13%) and highest was for CYS (CV = 19.35%) followed by HIS (CV = 13.00%). Generally, THR, SER, CYS, and ARG have lower, similar variabilities and CV ranged from 3.13% - 3.93%, but CYS, HIS, and TRP had higher variabilities and CV ranged from 11.63% - 19.35%. The frequency distributions of different AA contents in the seeds of RILs were generally normally distributed (**Figure 1** and **Figure 2**); this AA had a bimodal distribution although the skewness and kurtosis values for these traits were <1.00 (**Table 1**). Correlation among AA, conducted by SAS [20], showed significant positive and negative correlations. For example, THR had a significant ($p \leq 0.05$) positive correlation with SER (r = 0.76), PRO (r = 0.69), ARG (r = 0.47), and TRP (r = 0.24), but significant

Table 1. Mean values, range, coefficient of variations (CV%), distribution, and correlations of threonine (THR), serine (SER), proline (PRO), cysteine (CYS), histidine (HIS), arginine (ARG), and tryptophan (TRP) contents (%) of the parental materials of MD 96-5722 and Spencer and their RIL $F_{5:7}$ populations.

Traits	Mean Values ± Standard Deviations			Mean Values of RIL and CV		Frequency Distribution	
	MD96-5722	Spencer	RIL	Range	CV (%)	Skewness	Kurtosis
THR	1.70 ± 0.03[a]	1.77 ± 0.02	1.70 ± 0.06	1.60 - 1.91	3.53	0.39	0.43
SER	2.20 ± 0.09	2.41 ± 0.03	2.19 ± 0.08	2.00 - 2.41	3.65	0.24	1.15
PRO	2.21 ± 0.05	2.39 ± 0.01	2.24 ± 0.07	2.10 - 2.42	3.13	0.20	−0.05
CYS	0.30 ± 0.02	0.33 ± 0.01	0.31 ± 0.06	0.20 - 0.41	19.35	0.03	0.13
HIS	0.89 ± 0.09	1.12 ± 0.04	1.00 ± 0.13	0.60 - 1.32	13.00	−0.38	0.49
ARG	3.07 ± 0.01	3.42 ± 0.01	3.05 ± 0.12	2.80 - 3.33	3.93	0.11	−0.14
TRP	0.43 ± 0.01	0.44 ± 0.01	0.43 ± 0.05	0.40 - 0.51	11.63	0.70	−1.54

[a]Mean ± SD. SD = Standard deviation.

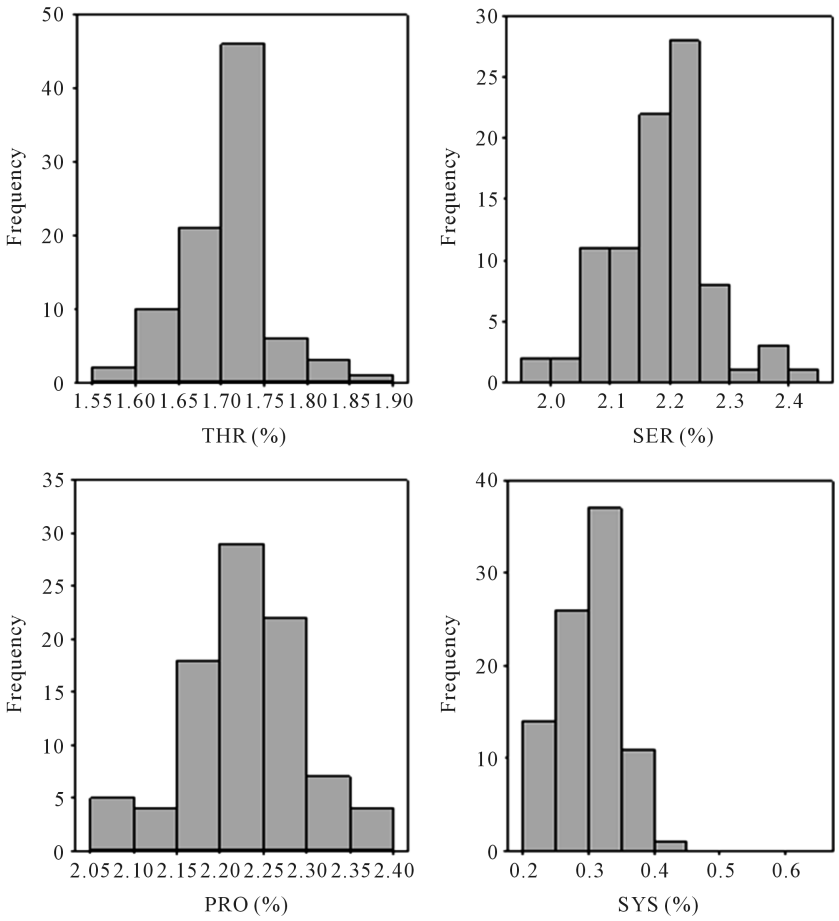

Figure 1. Frequency distribution for seed amino acids threonine (THR), serine (SER), proline (PRO), and cysteine (CYS) in the MD96-5722 by "Spencer" RIL population in soybean.

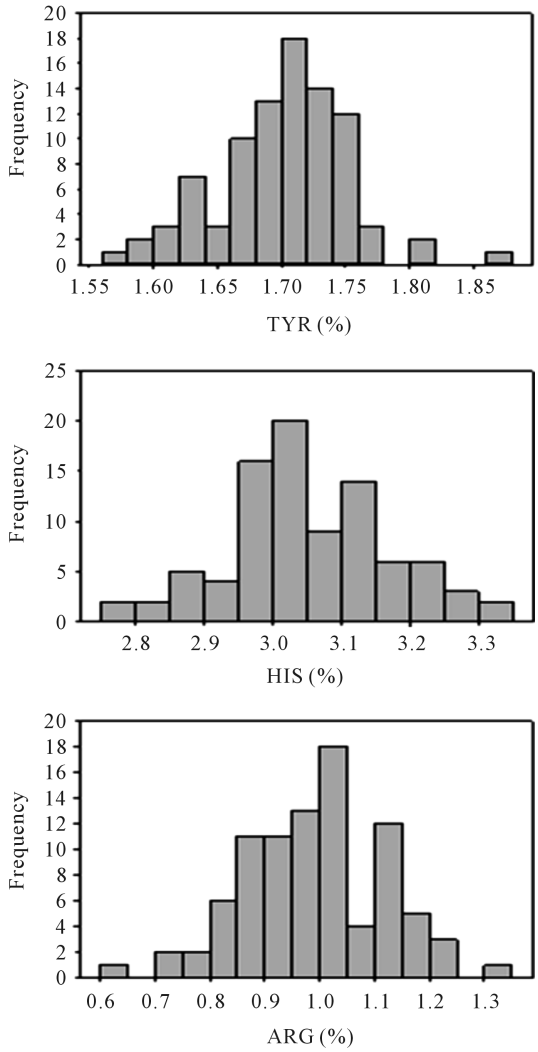

Figure 2. Frequency distribution for seed amino acids histidine (HIS), arginine (ARG), and tryptophan (TRP) in the MD96-5722 by "Spencer" RIL population in soybean.

($p \leq 0.05$) negative correlation with CYS ($r = -35$) (**Table 2**). Amino acid SER had a significant ($p \leq 0.05$) positive correlation with PRO ($r = 0.71$), ARG ($r = 41$), but significant ($p \leq 0.05$) negative correlation with CYST ($r = -0.40$). Positive correlations were found between CYS and HIS, between CYS and ARG, and between CYS and TRP; between HIS and ARG, and between HIS and TRP; between ARG and TRP (**Table 2**). The only obvious significant ($p \leq 0.05$) negative correlation in AA were between CYS and THR ($r = -0.35$), between CYS and SER ($r = -0.40$), and low correlation (non-significant correlation at $p \leq 0.05$) between CYS and PRO ($r = -014$), and between HIS and THR ($r = -0.03$).

3.2. Detection of QTLs Associated with Seed Amino Acids

Composite Interval Mapping (CIM) identified 13 candidate QTLs for different AA contents on eight of the 16 different chromosomes (Chr) represented by short linkage groups (LG) (**Table 3**; **Figure 3** and **Figure 4**). One QTL for PRO (*qPRO*001) was identified on LG D1a/Chr1 at 3.30 - 4.50 LOD support interval with a LOD score of 3.80. Another QTL was identified for seed ARG (*qARG*001) on LG N/Chr3 with a LOD score of 3.74 at 3.20 - 4.10 LOD support interval. Three QTL were identified in LG A1/Chr5; one for seed THR (*qTHR*001), one for

Table 2. Simple phenotypic correlation coefficients between seed amino acids (%) in 92 $F_{5:7}$ derived RILs of MD96-5722, and Spencer.

Traits	THR	SER	PRO	CYS	HIS	ARG	TRP
THR[a]	1						
SER	0.76	1					
PRO	0.69	0.71	1				
CYS	−0.35	−0.40	−0.14	1			
HIS	−0.03	0.06	0.17	0.42	1		
ARG	0.47	0.41	0.70	0.28	0.26	1	
TRP	0.24	0.14	0.51	0.20	0.26	0.53	1

[a]Threonine = THR; serine = SER; proline = PRO; cysteine = CYS; histidine = HIS; arginine = ARG; tryptophan = TRP.

Table 3. Chromosomal locations and parameters associated with the quantitative trait loci (QTL) of components of seed amino acid contents (%) in MD96-5722 and Spencer recombinant inbred linepopulation of soybean.

Trait	QTL (Env.)	LG/Chr	[†]Peak Position (cM)	[‡]2-LOD Support Interval (cM)	Markers Interval	[§]Peak LOD	[#]R^2	[††]Additive Effects
PRO	*qPRO*001	D1a/Chr1	3.80	3.30 - 4.50	ss244519430 - ss244503611	3.64	0.10	0.01
ARG	*qARG*001	LG N/Chr3	4.00	3.20 - 4.10	ss244936977 - ss244942217	3.74	0.60	−0.06
THR	*qTHR*001	LG A1/Chr5	9.50	8.40 - 9.60	ss245747167 - ss245786667	30.57	0.06	20.30
SER	*qSER*001	LG A1/Chr5	9.50	8.50 - 10.80	ss245747167 - ss245786667	7.24	0.09	0.13
CYS	*qCYS*001	LG A1/Chr5	9.50	8.50 - 10.20	ss245747167 - ss245786667	2.80	0.06	0.02
THR	*qTHR*001	LG C2/Chr6	3.50	3.30 - 3.70	ss246091245 - ss246092064	2.82	0.12	−0.06
SER	*qSER*001	LG C2/Chr6	7.20	7.20 - 7.30	ss246087580 - ss246091735	5.00	0.01	0.13
CYS	*qCYS*001	LG K/Chr9	0.50	0.05 - 0.06	ss246894176 - ss246871129	3.98	0.05	0.05
TRY	*qTRP*001	LG K/Chr9	0.50	0.40 - 0.50	ss246871129 - ss246860974	11.98	0.03	0.05
THR	*qTHR*003	LG B2/Chr14	9.00	7.90 - 10.10	ss24829340 - ss248275088	29.52	0.26	0.06
PRO	*qPRO*001	LG B2/Chr14	8.90	6.90 - 11.10	ss24829340 - ss248275088	7.32	0.08	−0.10
HIS	*qHIS*001	LG J/Chr16	28.80	27.70 - 29.80	ss248982430 - ss249037210	3.00	0.05	−0.05
TRY	*qTRP*001	LG G/Chr18	2.60	2.50 - 2.70	ss249599552 - ss249581714	10.63	0.04	0.05

[†]Position of peak LOD value on composite maps described previously (Coles *et al*., 2010); [‡]The positions that define the two LOD intervals around the position of peak likelihood for the QTL; [§]The log of odds (LOD) value at the position of peak likelihood of the QTL; [#]R^2 estimates the proportion of RIL mean variance (%) explained by the detected QTL; [††]A positive number in additive effect of the QTL indicates that the allele for susceptibility was derived from the line indicated and a negative number means that the allele for resistance was derived from the line indicated. Threonine = THR; serine = SER; proline = PRO; cysteine = CYS; histidine = HIS; arginine = ARG; tryptophan = TRP.

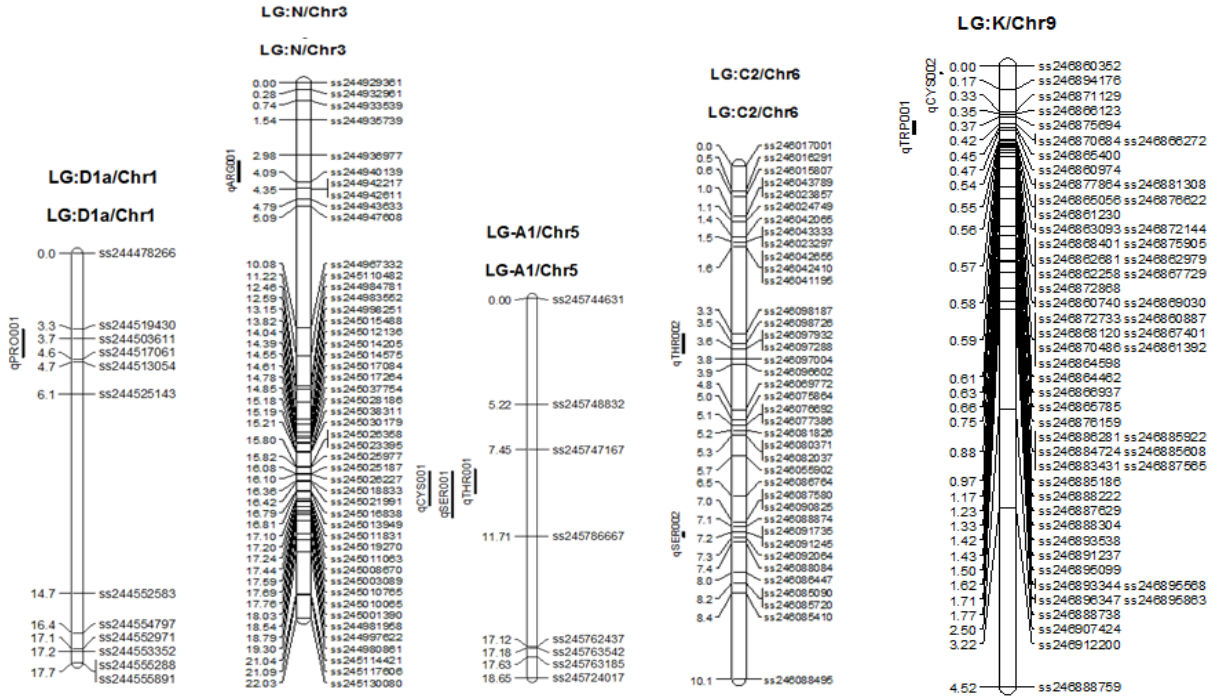

Figure 3. Locations of SNP markers and the QTLs that underlie different amino acids: proline (qPRO), arginine (qARG), cysteine (qCYS), serine (qSER), threonine (qTHR), tryptophan (qTRP) contents (% on seed dry-base) in the "Maryland" by Spencer RIL populations on molecular linkage groups (LGs) D1a, N, A1, C2, K using 5376 Single Nucleotide Polymorphism (SNP) markers.

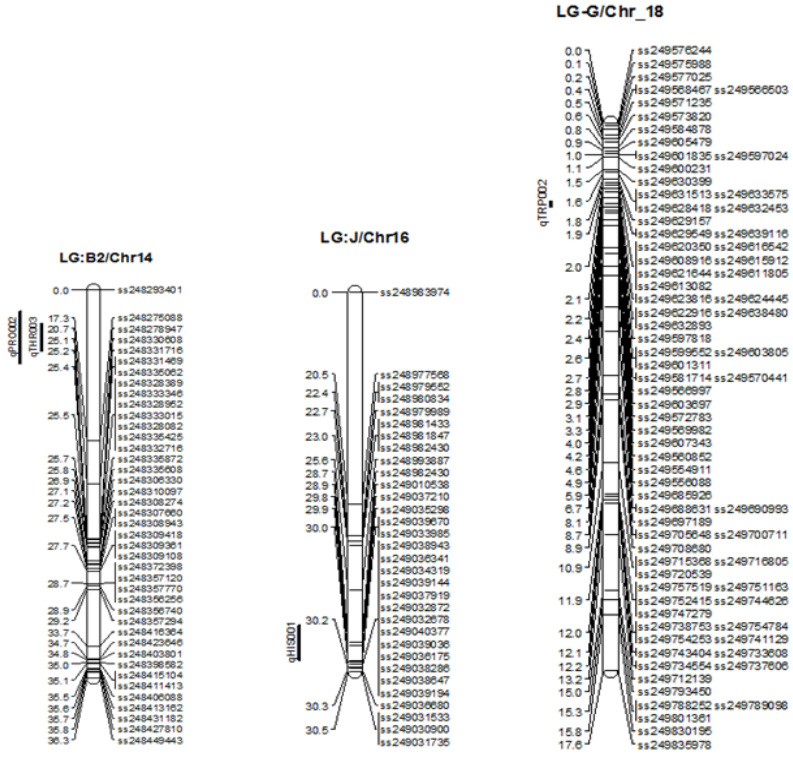

Figure 4. Locations of SNP markers and the QTL that underlie different amino acids: proline (qPRO), threonine (qTHR), histidine (qHIS), and tryptophan (qTRP) contents (% on dry-base) in the "Maryland" by Spencer RIL populations on molecular linkage groups (LGs) B2, J, G using 5376 Single Nucleotide Polymorphism (SNP) markers.

SER (*qSER*001), and one for CYS (*qCYS*001), and the LOD scores for these three QTL were 30.57, 7.24, and 2.80, respectively. On LG C2/Chr6 two QTL were identified, one was for seed THR (*qTHR*002) and another for seed SER (*qSER*002), with LOD scores of 2.82 and 5.00, respectively. LG K/Chr9 had two QTL, one was for CYS (*qCYS*002 *with* LOD score of 3.98) and another for TRP (*qTRP*001 *with* LOD score of 11.98). One QTL of each for THR (*qTHR*002 *with* LOD score of 29.52) and PRO (*qPRO*002 *with* LOD score of 7.32) were identified on LG B2/Chr14. Linkage group J/Chr16 and LG G/Chr18 had one QTL on each for seed histidine (*qHIS*001 *with* score LOD of 3.00) and TRP (*qTRP*002 *with* score LOD of 10.63). Of 13 QTL in the "MD96-5722" by Spencer RIL population, two of each were for PRO, SER, TRP, and CYS; three for THR, and one of each for ARG and HIS contents.

4. Discussion

4.1. Amino Acids Variation in RIL Population

There were narrow differences among the RIL for some AA and wider ranges in other AA in the soybean population we studied. For example, the range was low for some of the AA such as THR (CV = 3.53%), SER (CV = 3.65%), and PRO (CV = 3.13). However, there were reasonable level of variations found for CYS (CV = 19.35%), HIS (CV = 13.0) and TRP (CV = 11.63). Similar results were observed by [6] in the soybean inbred lines (RIL) developed from a cross of N87-984-16 x TN93-99. The low phenotypic variation found in the present population for some of the AA may be the result of conserved areas of the genome in the two parents. Limited information is available on genetic analysis of the soybean seed AA composition we studied herein. The findings from the present study are expected to advance our understanding of the genetic basis for future research in genetic improvement of soybean for higher seed AA composition. The narrow variation (**Table 1**) of the AA contents among the RILs suggests that little less genetic gain can be made in soybean, although modest genetic gains can be achieved for CYS, HIS, and TRP. Cysteine is a sulfur-containing amino acid that is generally deficient in soybean seed cultivars, and therefore higher CYS is a desirable trait, and variation among our RIL may provide an opportunity for further research to select for higher sulfur-containing AA (CYS).

4.2. Correlations among Amino Acid Composition

The relationship among seven AA were determined and almost all were positively correlated (r = 0.14 to 0.76) (**Table 2**). However, CYS was shown to have negative correlations with three others amino acids THR, SER, PRO, and a low correlation (r = −0.03) between HIS and THR. Amino acid SER had a positive correlation with PRO (r = 0.71), ARG (r = 41), but negative correlation with CYST (r = −0.40). Positive correlations were found between CYS and HIS, CYS and ARG, and CYS and TRP; between HIS and ARG, and HIS and TRP; between ARG and TRP (**Table 2**). The only obvious negative correlation in AA were between CYS and THR (r = −0.35), between CYS and SER (r = −0.40), and low correlation between CYS and PRO (r = −014), and between HIS and THR (r = −0.03). These results are partially in agreement with those of others [6] in that positive correlations were found between THR and SER (r = 0.94), THR and PRO (r = 0.62), THR and ARG (r = 0.44), THR and TRP (r = 0.77), and THR and CYS (r = 0.58). However, the findings of others [6] did not agree with our findings in that they found positive correlations, rather than negative correlations, between CYS and SER (r = 0.60), and between CYS and PRO (r = 0.13), and between HIS and THR (r = 0.80). A negative correlations was found between THR and GLU (r = −0.47), between THR and LEU (r = −0.54), between THR and LYS (r = −0.46), and between SER and GLU (r = −0.47); SER and LEU (r = −0.54), and between CYS and LEU (r = −0.50) [6]. The results were similar to those of Fallen *et al.* (2013) in that they found that almost all AA were positively correlated (r = 0.17 to 0.97), but also reported that LYS had weak to moderate negative correlation with ten amino acids and a weak to moderate positive correlation with three amino acids. For example, LYS showed a weak negative correlation with CYS (r = −0.23), MET (r = −0.29), PRO (r = −0.27), TRP (r = −0.16) and TYR (r = −0.30). In our experiment, negative correlation between some AA was found and ranged from week (r = −0.003 to −0.40), agreeing with others [11].

4.3. Amino Acids QTL Detection

QTLs underlying AA contents were identified by composite interval mapping (CIM), using QTL Cartographer. Our research showed that two QTLs were detected for PRO on LG D1a (Chr1) and LG B2 (Chr14). A PRO

QTL linked to SNPs on LG F (Chr13) [11], and [6] also reported three QTLs for PRO linked microsatellites markers on three different LGs [(LG A2 (Chr8), LG G (Chr18), and LG L (Chr19)]. Therefore, the two QTLs reported in this study for PRO are different from those reported by others [6] [11]). Both identified one QTL for ARG on LG F (Chr13) and LG D1b (Chr1), respectively [6] [11], whereas, in this study QTLs were identified for ARG on LG N (Chr3). The three QTL that were identified in this study for THR on LGs A1 (Chr5), C2 (Chr6), and B2 (Chr14) were also identified by others [6] [11], but on different LGs. Also, Warrington (2011) detected a QTL associated with THR linked to SSR marker BARC-048619 (79.06 cM) on Chr9, making the QTLs identified in this study for THR are different from those reported by others [6] [11] [12]. Similarly, QTL for SER, CYS, and TRP identified here were also identified by others [6] [11], but on different LGs, making QTLs described in this study for SER, CYS, and TRP new addition. Only one QTL for HIS was identified by others [6] and herein on the same LG J (Chr16), which was only 2 cM from the QTL identified herein. Four QTL were identified for CYS on LG I (Chr 20), SER on LG F (Chr 13), and TRP on LG F (Chr 13), and HIS on LG O (Chr10) [11]. Recently, [21] [22] investigated amino acids, including LYS, THR, and METH, and the use of genome-wide association scans across multiple populations for QTLs associated with soybean seed composition, including seed methionine (MET), THR, CYS, and LYS content [22]. They reported that multiple QTLs found have not been observed in family-based mapping studies, and found that chromosomes 1 and 8 contain candidate alleles for essential amino acid increases. They found multiple closely spaced markers on Chr8 associate with CYS, LYS, and THR, and genetic overlap between essential amino acid traits, reflecting possible common biochemical pathways for these amino acids synthesis.

Amino acids SER, CYS, and THR were clustered together on LG A1 (Chr 5) and amino acids PRO and TRP were clustered together on LG B2 (Chr14), suggesting possible common metabolic pathways. Previous research suggested interrelations between different amino acids existed [6] [7] [14] [15]. For example, it was found that the essential amino acids such as isoleucine (ILE), Leucine (LEU) and Valine (VAL) were found to have a common biosynthesis pathway [15]. One common QTL was detected between ILE and LEU, and three between ILE and VAL [6], and it was suggested that these amino acids have a common biosynthesis pathway because of interdependent relationships [17]. A major common genomic region was also found on LG L for these amino acids, and a common QTL on LG F for the two sulfur-containing amino acids, CYS and MET [6], and suggested a common pathway regulating these two sets of amino acids. It was explained that in the biochemical pathway of MET biosynthesis, CYS is a precursor for sulfur assimilation [18] [19]. Another example is the aromatic AA such as PHE, TRP and TYR, which have the common shikimate pathway, responsible for the chorismate compound production, which is subsequently used in the biosynthesis of PHE, TRP, and TYR [6] [14]. It was reported that there are seven different enzymes involved in the shikimate pathway, and three are involved with PHE and TYR, and five with TRP, and indicated that more QTLs related with PHE, TRP, and TYR will be identified [6]. Further, QTLs associated with sulfur-containing amino acids were found, and most of the QTL associated with MET were also associated with CYS [7], and this is due to the fact that CYS is the intermediate product in the sulfur assimilation process [18] [19]. It was concluded that since one amino acid depends on the concentration of others, it is possible to detect the same QTL for both MET and CYS [7].

In spite of the unknown molecular mechanisms of this overlapping, the identification of high-resolution MET-specific QTL is still promising for fine-mapping genes responsible for relative seed amino acid content. They added that, using family-based mapping studies of MET and CYS content, they were able to identify only two QTLs. They also reported that they did not find a significant association between either trait and any marker on Chr13, and MET had a weak association with a SNP on Chr18. Although this is a departure from the QTLs markers located on Chr13 and Chr18 found by others [6], other researchers reported that it is common for family-based studies to have distinct results from genome-wide association scans [13]. They concluded that nested-association mapping populations in soybean may address issues in spite of the possible reduced resolution of the populations and allelic diversity relative to genome-wide association scans using large, heterogeneous panels [13].

Therefore, in the current research, several QTLs were found that did not coincide with other researchers groups, but some QTLs were detected previously using the same population MD96-5722 x Spencer [23] [24]. For example, a linoleic acid QTL was detected by others elsewhere [7] on LG N (Chr3) at 4.80 cM position was very close to the QTL for ARG at 4.00 cM; the QTL we identified for SER on LG C2 (Chr6) has been identified for raffinose at the same position (7.20 cM) (unpublished data)and the QTL identified in LG K (Chr 9) for SER and TRP (0.50 cM) were very close to QTL identified in a previous study for the seed content of the isoflavone

glycitein (0.60 cM; [20]. Here, novel moderate to major QTL were detected and associated with ARG (60% of variation) on LG D1a (Chr 3), SER (93%) on LG A1 (Chr 5), THR (26%) on LG B2 (Chr 14), and PRO (87%) on LG B2 (**Table 3**), reflecting significant contributions of the current study to the previous findings. To date, there are only few studies available on the genetic analysis of AA composition in soybean. The different QTL, reported among the limited literature, detected on different LG on different chromosomes could be due to the use of different molecular markers, genotypes, and populations. The limited research that has been done in this area will open opportunities to advance understanding of the genetic basis of AA QTL on molecular linkage groups. The molecular markers associated with them will be useful for selection of AA contents in different types of soybean seeds. The information contained in the current research would be help soybean breeders to develop desired soybean cultivars with higher protein quality with desirable AA profiles using molecular markers associated with QTL. Additional information on QTLs associated with soybean seed amino acids will be useful for breeding strategies for the continued soybean seed nutrition improvement [13].

5. Conclusion

Here, 13 QTL for AA were identified in 92 $F_{5:7}$ recombinant inbred lines (RIL), developed from a cross of MD96-5722 × "Spencer", using Single Nucleotide Polymorphism (SNP) markers, which produced total 5376 SNPs. Most of the QTL identified here are not previously reported, and to our knowledge these QTLs are novel. Since CYS is a sulfur-containing amino acid, which is generally deficient in soybean seed cultivars, and higher CYS is a desirable trait, the detection of QTLs related to this amino acid and variation among the RILs studied here may provide an opportunity for further research to select for higher sulfur-containing AA (CYS). The clustering of QTLs for amino acids SER, CYS, and THR on the same LG A1 (Chr5) suggests possible common metabolic pathways and interrelations among different amino acids, proving researchers further physiological and biochemical understanding for further molecular and genetic research. This information may be significant in which the new QTL for amino acids in soybean could facilitate developing soybean cultivars with higher protein quality and desired amino acid profiles to help meet the needs of the seed industry and soy-based products for human nutrition and livestock feed.

Acknowledgements

The authors would like to thank the Department of Defense (DoD) for funding this work through the grant# W911NF-11-1-0178 to M.A.K and S.K. We thank Ms. Pam Ratcliff and the rest of the undergrad students crew at FSU for taking care of the plants in the greenhouse and field, and Sandra Mosley at USDA-ARS, Stoneville, MS, for lab assistance on isoflavones analysis. This research Is partially funded by United States Department of Agriculture, Agricultural Research Service project number 6402-21220-012-00D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the Unites States Department of Agriculture (USDA). USDA is an equal opportunity provider and employer.

References

[1] Berk, Z. (1992) Technology of Production of Edible Flours and Protein Products from Soybeans. FAO Agricultural Services Bulletin No. 97, Rome, 178.

[2] Ufaz, S. and Galili, G. (2008) Improving the Content of Essential Amino Acids in Crop Plants: Goals and Opportunities. *Plant Physiology*, **147**, 954-961. http://dx.doi.org/10.1104/pp.108.118091

[3] National Research Council (1998) Nutrient Requirements of Swine: 10th Revised Edition. Subcommittee on Swine Nutrition, Committee on Animal Nutrition, The National Academies Press, Washington DC, USA, 210.

[4] Wilson, R.F. (1987) Seed Metabolism. In: Wilcox, J.R., Ed., *Soybeans*: *Improvement*, *Production and Uses*, 2nd Edition, American Society of Agronomy, Madison, 643-687.

[5] Kwanyuen, P., Pantalone, V.R., Burton, J.W. and Wilson, R.F. (1997) A New Approach to Genetic Alteration of Soybean Protein Composition and Quality. *Journal of the American Oil Chemists Society*, **74**, 983-987. http://dx.doi.org/10.1007/s11746-997-0015-2

[6] Panthee, D.R., Pantalone, V.R., Saxton, A.M., West, D.R. and Sams, C.E. (2006) Genomic Regions Associated with Amino Acid Composition in Soybean. *Molecular Breeding*, **17**, 79-89. http://dx.doi.org/10.1007/s11032-005-2519-5

[7] Panthee, D.R., Pantalone, V.R., Saxton, A.M., West, D.R., Sams, C.E., Orf, J.H. and Killiam, A.S. (2006) Quantitative Trait Loci Controlling Sulfur Containing Amino Acids, Methionine and Cysteine, in Soybean Seeds. *Theoretical and Applied Genetics*, **112**, 546-553. http://dx.doi.org/10.1007/s00122-005-0161-6

[8] SoyBase (2014) SoyBase—A Genome Database for Glycine. Administered Online by USDA-ARS and Iowa State University. http://soybeanbreederstoolbox.org

[9] Wang, X.L. and Larkins, B.A. (2001) Genetic Analysis of Amino Acid Accumulation in *Opaque*-2 Maize Endosperm. *Plant Physiology*, **125**, 1766-1777. http://dx.doi.org/10.1104/pp.125.4.1766

[10] Zhao, W.G., Park, E.J., Chung, J.W., Park, Y.J., Chung, I.M., Ahn, J.K. and Kim, G.H. (2009) Association Analysis of the Amino Acid Contents in Rice. *Journal of Integrative Plant Biology*, **51**, 1126-1137. http://dx.doi.org/10.1111/j.1744-7909.2009.00883.x

[11] Fallen, B.D., Hatcher, C.N., Allen, F.L., Kopsell, D.A., Saxton, A.M., Chen, P., Kantartzi, S.K., Cregan, P.B., Hyten, D.L. and Pantalone, V.R. (2013) Soybean Seed Amino Acid Content QTL Detected Using the Universal Soy Linkage Panel 1.0 of 1536 SNPs. *Journal of Plant Genome Science*, **1**, 68-79.

[12] Warrington, C.V. (2011) QTL Mapping and Optimum Resource Allocation for Enhancing Amino Acid Content in Soybean. Ph.D. Dissertation, University of Georgia, Athens, GA.

[13] Akond, M., Kantartzi, S.K., Meksem, K., Song, Q., Wang, D., Wen, Z., Lightfoot, D.A. and Kassem, M.A. (2013) A SNP-Based Genetic Linkage Map of Soybean Using the SoySNP6K Illumina Infinium BeadChip Genotyping Array. *Journal of Plant Genome Science*, **1**, 80-89.

[14] Siehl, D.L. (1999) The Biosynthesis of Tryptophan, Tyrosine, and Phenylalanine from Chorismate. In: Singh, B.K., Ed., *Plant Amino Acids*: *Biochemistry and Biotechnology*, Marcel Dekker Inc., New York, 171-204.

[15] Singh, B.K. (1999) Biosynthesis of Valine, Leucine, and Isoleucine. In: Singh, B.K., Ed., *Plant Amino Acids*: *Biochemistry and Biotechnology*, Marcel Dekker Inc., New York, 227-247.

[16] Fontaine, J. (2003) Amino Acid Analysis of Feeds. In: D'Mello, J.P.F., Ed., *Amino Acids in Animal Nutrition*, 2nd Edition, CABI Publishing, Cambridge, MA, 15-40. http://dx.doi.org/10.1079/9780851996547.0015

[17] AOAC (1990) Official Methods of Analysis of the Association of Official Analytical Chemists, 982.30E (a,b,c) Ch. 45.3.05. 15th Edition, AOAC International, Arlington, VA.

[18] Ooijen, J.W.V. (2006) Join Map 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma, B.V., Wageningen.

[19] Saito, K. (1999) Biosynthesis of Cysteine. In: Singh, B.K., Ed., *Plant Amino Acids*: *Biochemistry and Biotechnology*, Marcel Dekker Inc., New York, 267-291.

[20] SAS (2002-2010) Statistical Analysis System, Cary, NC.

[21] Matthews, B.F. (1999) Lysine, Threonine, and Methionine Biosynthesis. In: Singh, B.K., Ed., *Plant Amino Acids*: *Biochemistry and Biotechnology*, Marcel Dekker Inc., New York, 205-225.

[22] Vaughn, J.N., Nelson, R.L., Song, O., Cregan, P.B. and Li, Z. (2014) The Genetic Architecture of Seed Composition in Soybean is Refined by Genome-Wide Association Scans across Multiple Populations. *Genes*, *Genomes*, *Genetics*, **4**, 2283-2294. http://dx.doi.org/10.1534/g3.114.013433

[23] Akond, M., Liu, S., Kantartzi, S.K., Khalid, M., Bellaloui, N., Lightfoot, D.A., Yuan, J., Wang, D. and Kassem, M.A. (2014) Quantitative Trait Loci for Seed Isoflavone Contents in "MD96-5722" by "Spencer" Recombinant Inbred Lines of Soybean. *Journal of Agricultural and Food Chemistry*, **62**, 1464-1468. http://dx.doi.org/10.1021/jf4040173

[24] Akond, M., Liu, S., Kantartzi, S.K., Meksem, K., Bellaloui, N., Lightfoot, D.A., Yuan, J., Wang, D. and Kassem, M.A. (2014) Identification of QTLs Controlling Protein, Oil and Fatty Acids' Contents in Soybean. *American Journal of Plant Sciences*, **5**, 158-167. http://dx.doi.org/10.4236/ajps.2014.51021