Canadian Journal of Plant Science
Revue canadienne de phytotechnie

Detection of QTL Underlying Seed Quality Components in Soybean [*Glycine max* (L.) Merr.]

Journal:	<i>Canadian Journal of Plant Science</i>
Manuscript ID	CJPS-2017-0204.R4
Manuscript Type:	Article
Date Submitted by the Author:	13-Dec-2017
Complete List of Authors:	Akond, Masum; Fayetteville State University, Biological Sciences Yuan, Jiazheng (John); Fayetteville State University, Biological Sciences Liu, Shiming; Southern Illinois University, Plant, Soil, and Agricultural Systems Kantartzi, Stella; Southern Illinois University, PSAS Meksem, Khalid; Southern Illinois University, Plant, Soil, and Agricultural Systems Bellaloui, Nacer; USDA-ARS, Crop Genetics Research Unit, 141 Experiment Station Road, P.O. Box 345 Lightfoot, David; Southern Illinois University, Plant, Soil, and Agricultural Systems Kassem, My Abdelmajid; Fayetteville State University, Biological Sciences
Keywords:	Soybean, SNP Linkage Map, QTL, RIL, Protein

SCHOLARONE™
Manuscripts

Detection of QTL Underlying Seed Quality Components in Soybean [*Glycine max* (L.) Merr.]

Masum Akond¹, Jiazheng Yuan¹, Shiming Liu², Stella K. Kantartzi², Kkalid Meksem², Nacer Bellaloui³, David A. Lightfoot², and My Abdelmajid Kassem^{1*}

¹ Plant Genomics and Biotechnology Lab, Department of Biological Sciences, Fayetteville State University, Fayetteville, NC 28301, USA; ² Department of Plant, Soil and Agricultural Systems, Southern Illinois University, Carbondale, IL 62901, USA; ³ USDA-ARS, Crop Genetics Research Unit, 141 Experiment Station Road, P.O. Box 345, Stoneville, MS 38776, USA.

*Corresponding author: Email: mkassem@uncfsu.edu. Tel: +1(910) 672 1692. Fax: +1(910) 672 1159.

Canadian Journal of Plant Science

Received: June 28, 2017

Accepted: December 20, 2017

Abstract

Improving seed composition and quality, including protein, oil, fatty acid, and amino acid contents is an important goal of soybean farmers and breeders. The aim of this study was to map the QTL underlying the contents of protein, oil, fatty acids, and amino acids with 1510 single nucleotide polymorphism (SNP) markers using the ‘Hamilton’ by ‘Spencer’ recombinant inbred line (RIL) population ($H \times S$; $n = 93$). A total of 13 QTL for the traits studied have been mapped on 3 chromosomes (Chr.) of the soybean genome. Three major QTL have been mapped to a 7–13 cM region on Chr 6. One major QTL for oil content (qOIL001) explained approximately 76% of the total phenotypic variation in this population; the second major QTL for amino acid Alanine (Ala; qALA001) explained approximately 74% of the total variation in Ala content; moreover, two major QTL for palmitic acid (qPAL001 and qPAL002) were identified on Chr. 6 and explained approximately 21% phenotypic variation in this population. The SNP markers flanking the QTL identified here will be very useful for soybean breeders to develop and select soybean lines with higher seed composition qualities, using marker-assisted selection.

Keywords: Soybean – SNP Linkage Map – QTL – RIL – Protein – Oil – fatty acids – Hamilton – Spencer

Abbreviations: FAME, fatty acid methyl esters; BC, backcross; QTL, quantitative trait loci; RIL, recombinant inbred line; NSRL, National Soybean Research Laboratory; CIM, composite Interval Mapping; NIR, near-infrared; SNP, single-nucleotide polymorphism; CV, coefficient of variation; LOD, logarithm of the odds; DOD, Department of Defense.

Introduction

Soybean [*Glycine max* (L.) Merr.] seeds are rich in protein, oil, isoflavones, minerals, sugars, fatty acids, amino acids, and other bioactive compounds (Bellaloui et al., 2015). Soybean seeds contain, on a dry weight basis, about 380 to 420 g kg⁻¹ protein (40%), 190 to 230 g kg⁻¹ oil (18–22%), and on the total oil basis, soybean seeds contain about 120 to 130 g kg⁻¹ palmitic acid, 200 to 300 g kg⁻¹ oleic acid, 480 to 580 g kg⁻¹ linoleic acid, and 50 to 80 g kg⁻¹ linolenic acid (Bellaloui et al., 2015) and up to 35% carbohydrates (Mateos-Aparicio et al., 2008). Soybean oil quality depends on its composition of fatty acids that can affect its stability, flavor, and nutritional value. For edible oils, high contents of oleic acid, and low contents of palmitic acid are desirable because unsaturated fatty acids (oleic acid) can reduce the risk of coronary diseases (Fehr, 2007) and lower cholesterol can decrease the risk of heart disease and arteriosclerosis (Shannon, 2012). On the other hand, high amounts of monounsaturated fatty acid methyl esters (FAME) enhance the oxidation stability of biodiesel and are desirable for the biodiesel industry (Fallen et al., 2011).

Developing cultivars with high seed yield, protein, oil, unsaturated fatty acids, and amino acids contents has been the primary objective for most soybean breeding programs; however, many studies showed that some of these traits including seed protein contents are negatively correlated with seed yield (Wilcox and Cavins, 1995). These traits are complex multifactorial traits that are affected by genotype and genotype by environment (G X E) interactions (Wolf et al., 1982; Dornbos and Mullen, 1992; Wilcox and Cavins, 1995; Vollmann et al., 2000; Yaklich et al., 2002). A study showed that drought and high temperatures during the seed filling stage increased protein contents but decreased oil content and seed yield (Wolf et al., 1982). Another similar study noticed that the drought stress and high temperature on several soybean genotypes

during the seed filling period increased protein contents about 4.4%, decreased oil content about 2.9%, and showed no effect on fatty acid compositions. Moreover, the proportion of polyunsaturated fatty acids was reduced by high temperatures (Dornbos and Mullen, 1992). Wilcox and Cavins (1995) investigated seed yield and protein contents in three backcross (BC) populations and found that these traits were negatively correlated in all. Vollmann et al. (2000) studied the effects of environmental factors on seed protein contents in several European soybean cultivars and found significant differences among genotypes in seed protein contents. The high temperatures and moderate rainfall promoted high protein contents, and insufficient nitrogen fixation reduced seed protein contents drastically (Vollmann et al., 2000). Amino acids are essential elements for living organism's growth. Amino acids leucine (Leu), isoleucine (Ile), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp), and valine (Val) are essential for humans while His is semi-essential, and arginine (Arg), alanine (Ala), aspartic acid (Asp), asparagine (Asn), cysteine (Cys), glutamic acid (Glu), glutamine (Gln), glycine (Gly), proline (Pro), serine (Ser), tyrosine (Tyr) are non-essential amino acids (Ufaz and Galili, 2008).

Several breeding attempts were made to develop soybean cultivars with increased seed protein and oil content (Burton and Brim, 1981; Grant et al., 2010). Through marker-assisted selection, it was also possible to develop cultivars with enhanced levels of protein, oil, fatty acids and amino acids. (Burton and Brim, 1981; Grant et al., 2010). Other studies were successful in developing cultivars with high oleic acid contents for human consumption and low palmitic acid contents for biodiesel industry (Fehr, 2007; Grant et al., 2010).

Sulfur containing essential amino acids such as methionine and cysteine are low in soybean seeds which limit its nutritional value. Traditional breeding attempts have been made to increase total seed protein contents but these essential sulfur-containing amino acids' contents,

especially methionine, remained constant even when protein contents were highly increased (Wilcox and Shibles, 2001). Genetically engineered soybean plants where genes encoding for methionine-rich heterologous proteins were introduced showed a modest increase in methionine content (Krishnan, 2005). In another study, mutagenizing soybean seeds via methanesulfonate (EMS) efficiently created a soybean mutant with a 20% increase in sulfur amino acids (Imsande, 2001).

Quantitative trait loci (QTL) for protein, oil, fatty acid, and amino acid contents have been mapped in most soybean chromosomes (Mao et al., 2013; Akond et al., 2014; Khandaker et al., 2015; Grant et al., 2010). SoyBase contains hundreds of QTL that underlie seed protein, amino acids, oil, and fatty acids content mapped in different populations and distributed across all the 20 chromosomes of the soybean genome (Grant et al., 2010). More QTL for these and other traits need to be identified in different genetic backgrounds to further encompass more of the genetic variation in the soybean genome. Therefore, our aim was to use the 'Hamilton' X 'Spencer' recombinant inbred line (RIL) population to map QTL for protein, oil, fatty acid and amino acid contents.

Materials and Methods

Plant Material and Seed Analysis for Protein, Amino Acids, Oil, and Fatty Acids

Seeds of F_5 derived RILs (HxS, $n=93$) and their parents 'Hamilton' (maturity group IV), 'Spencer' (maturity group IV; Akond et al., 2015) were collected from the National Soybean Research Laboratory (NSRL) and were grown at Fayetteville State University campus, Fayetteville, in 2012 with a row-spacing of 25 cm and plant density of 160,000 plants/ha with four replicates per RIL using randomized block design and no additional fertilizer or insecticide was used. Seeds at harvest maturity stage were analyzed for protein, oil, and fatty acids. About

25 g of seed from each line was ground using a Laboratory Mill 3600 (Perten, Springfield, IL). The ground samples were analyzed by near infrared reflectance using a diode array feed analyzer AD 7200 (Perten, Springfield, IL; Bellaloui et al., 2009). Perten's Thermo Galactic Grams PLS IQ software, initially developed by the University of Minnesota, was used for calibrations. Protein and oil analyses were based on a seed dry matter basis, and fatty acids were analyzed based on total oil (Bellaloui et al., 2009). Seed amino acids were analyzed in 2012–2013 using near-infrared (NIR) reflectance diode array feed analyzer (Perten, Springfield, IL; Siehl, 1999). Calibrations were developed by Perten using Thermo Galactic Grams PLS IQ. The calibration curve was regularly updated from six months to one year for unique samples. The analysis was performed based on percent dry matter.

Genetic Map Construction, QTL Identification, Statistical Analysis

The 'Hamilton' by 'Spencer' (HxS) RIL population was genotyped using 5,376 SNPs through SoySNP6K Illumina Infinium BeadChip array (Illumina, San Diego, CA) with 24 samples per chip. A total of 4 chips were used to conduct the soybean genotyping on HxS RILs and each sample was genotyped by 5,376 SNPs in the array. A genetic linkage map (Akond et al., 2015) was constructed through JoinMap 4.0 (Kyazma BV, Wageningen, Netherlands) (Van Ooijen et al., 2006). WinQTLCart 2.5 (<https://brcwebportal.cos.ncsu.edu/qtlcart/WQTLCart.htm>) was used to identify QTL from genotypic and phenotypic data via Composite Interval Mapping (CIM) option (Wang et al., 2012). During analysis by WinQTLCart, the Model 6 with four parameters were chosen; forward and backward stepwise regression, 10 cM window size, 1 cM step size and five control markers. The threshold was determined by permutation tests with 1,000 iterations. The R program (www.r-project.org) was used to analyze the data for ANOVA, mean, range, standard deviation, and CV.

Results

The parental lines of HxS RIL population had wide ranges of phenotypic variations especially for protein and oil (Table 1). The mean of the phenotypic values in the RILs showed either nearer or higher than to that of parental lines. Ten out of twenty-five traits displayed higher population mean values than that of parental lines. The highest value of population mean is the oleic acid (25.52 vs P1 (Hamilton)=19.45 and P2 (Spencer)=18.35, range=22). The phenotypic values between two parental lines were significantly different for protein and oil traits and however, the population mean of protein content was deviated toward P1 while the population mean for oil was located in between that of two parental lines (Table 1). Moreover, the coefficient of variation (CV) of protein content (4.42%), and oil content (6.09%) among RILs were relatively narrow. Coefficient of variation of glutamic acid (7.06%), cysteine (15.41%), valine (6.58%) and histidine (13.55%) were wider than protein but alanine (4.39%) was narrower. Variations in fatty acids palmitic (10.87%), oleic (16.82%) and linolenic (22.5%) acids were wider than oil content. Protein, oil, amino acid and the three major fatty acid contents were normally distributed in the RIL as the skewness and kurtosis values for these traits were <1.00 (Table 1).

A total of 1510 high-quality SNP marker from the beadchips were obtained using GenomeStudio (Illumina, San Diego, CA V2011.1) and showed segregation in the HxS RIL population. The remaining SNP markers were discarded due to being either monophorphic between the parents or false calls based on GenomeStudio software analysis (Illumina, San Diego, CA). All of the 1510 polymorphic markers were mapped to unique positions on 20 soybean linkage groups of the HxS RIL population using the JoinMap 4 program (Van Ooijen, 2006). Specifically, a total of 89, 114, and 113 SNP markers were assigned to chromosome 3, 6,

and 18 with 70, 72, and 67 cM genetic length, respectively. This SNP-based genetic map is highly dense and more suitable for accurate QTL mapping than genetic maps based on SSR markers.

Details of the QTL identified for protein, oil, fatty acids (palmitic, oleic, and linoleic), amino acids (Glu, Ala, Cys, Val, and His) content including the chromosome locations, marker intervals, LOD scores, R^2 , and QTL additive effects are shown in Table 2. The chromosomes (Chr) and the location of the QTL for these traits are also shown in Figure 1.

One QTL for protein content (qPRO001; LOD=2.5, R^2 =10%, additive effect of 4.32) was identified and mapped on Chr 18. One major QTL for oil content (qOIL001) was identified and mapped on Chr. 6 defined by SNP markers ss246100375 and ss245879277. The LOD score of 39.47 explained approximately 76% of the total variation with additive effect of 10.09 for oil content. Two QTL for palmitic acid content (qPAL001; LOD = 2.71, R^2 = 21%, additive effect of 1.87 and qPAL002, LOD = 2.98, R^2 = 21%, additive effect of 1.81). These two QTL were identified and mapped on two different regions of Chr. 6 (7–13 cM and 26.3–40 cM) and defined by SNP markers ss246100375-ss245879277 (qPAL001) and ss245914593-ss245908292 (qPAL002), respectively. Three QTL for oleic acid content (qOLE001; LOD = 2.65, R^2 = 9%, additive effect of 2.79; qOLE002; LOD = 2.83, R^2 = 15%, additive effect of 4.35; qOLE003; LOD = 5.0, R^2 = 12%, additive effect of 2.96) were identified and mapped on Chr. 3, Chr. 6, and Chr. 18, respectively. A QTL for linolenic acid content (qLIN001; LOD = 2.66, R^2 = 13%, additive effect of -0.82) was identified and mapped on Chr. 6.

For amino acids contents, a QTL for each of Ala (qALA001; LOD = 50.24, R^2 = 74%, additive effect of 1.05), Glu (qGLU001; LOD = 2.51, R^2 = 14%, additive effect of 1.07), Cys (qCYS001; LOD=2.53, R^2 =11%, additive effect of 0.06), Val (qVAL001; LOD = 2.62, R^2 =

11%, additive effect of 0.25), and His (qHIS001; LOD = 2.70, $R^2 = 12\%$, 0.19) were identified by the SNP markers and mapped on Chr. 6 and Chr. 18. The QTL for Ala was a major QTL with a LOD score of 50.24 that explained approximately 74% of the total variation in Ala content and identified in the same locus as that of qOIL001 by SNP markers ss246100375 and ss245879277 (Table 2, Figure 1). The favourable alleles to increase soybean quality traits were contributed by each of parental lines depending on the traits in HxS RIL population. There was no significant epistatic effect detected based on the analysis of WinQTLCart 2.5. As expected, the QTL underlying these nutritional traits showed a pleiotropic effect in which the QTL related to oleic acid qOLE002 shared the same loci with the palmitic acid (qPAL002) and amino acids QTL on chromosome 6. Moreover, the QTL for protein content was defined in the same interval as that of the QTL for oleic acid (qOLE003) on chromosome 18.

Discussion

Developing cultivars with high seed yield and better-quality traits are the ultimate goals of plant breeding programs. Protein and oil content in soybean seeds are usually negatively correlated and significant variation of these two traits due to changes of environments have been reported (Rotundo et al., 2009). Therefore, it is generally difficult to enhance both traits in the breeding efforts. Several QTL for soybean seed protein and oil content and other components of soybean seed have been identified and some of the molecular markers have been verified the association with these traits in several populations (Mao et al., 2013; Yesudas et al., 2013; Warrington et al., 2015; Patil et al., 2017; Grant et al., 2010). The mean of the phenotypic values in the RILs was close or higher than to the mean value of parental lines and 10 traits displayed higher population mean values than that of parental lines suggesting a transgressive segregating pattern for these traits in the RIL population.

Multiple QTL for protein, oil, palmitic, oleic, and linolenic acids contents were identified in the same intervals of Chr. 3, Chr. 6, and Chr. 18, respectively. The Chr.6 had the QTL for amino acids contents, glutamine, alanine, cystine, valine and histidine. These results suggest the pleiotropic effect of these QTL (Mao et al., 2013; Yesudas et al., 2013; Warrington et al., 2015; Patil et al., 2017; Grant et al., 2010). Chr 6 had been widely reported to contain many QTL for protein, oil, and fatty acids contents (Grant et al., 2010). However, the common QTL on Chr 20 appeared not to segregate in HxS. Furthermore, the QTL for percent oil content may also be correlated with a QTL for flower development as neighbouring SSR markers such as SATT316, have been mapped to QTL associated with the photoperiod insensitivity and flowering time in soybean (Tasma et al., 2001). The QTL (qPAL002), (qGLU001), and (qLIN006) that were identified in our study were very close to (qMet_Gm06) identified by (Warrington et al., 2015). While several previous studies identified QTL for protein content (Mao et al., 2013; Warrington et al., 2015; Patil et al., 2017), oil content (Mao et al., 2013; Grant et al., 2010), stearic, oleic and linolenic acid contents (Bachlava et al., 2009), the QTL for oleic acid (qOLE001) was also identified in the same interval on Chr. 3 in our study. Additionally, other studies reported QTL for protein content (Pathan et al., 2013), oil content (Grant et al., 2010), and glycitein content (Gutierrez-Gonzalez et al., 2009) in the same genomic region containing the cluster of QTL for oil (qOIL001), Ala (qALA001), oleic acid (qOLE002), palmitic acid (qPAL002), linolenic acid (qLIN001), and the amino acids His (qHIS001), Cys (qCYS001), and Glu (qGLU001) contents identified in the present study. Approximately 13 cM downstream of the same region and between 26 cM and 40 cM, previous studies identified QTL for protein (Pathan et al., 2013; Mao et al., 2013), oil (Kim et al., 2010), stearic, palmitic and linolenic acid contents (Grant et al., 2010), as well as various agronomic traits (Grant et al., 2010). In this study, although the

population mean value of oil was positioned between P1 and P2, the genetic variation within population was not typically narrow because the range of oil content within the RIL population was 6 (Table 1). Furthermore, the oil difference between two parental lines based on mean of four replicates was also significant. In QTL mapping, the associations between marker and trait can be easily detected as long as the phenotypic variance within the population is large and segregating in the RIL population. Except linolenic acid, all of analyzed traits showed a various number of QTL with positive additive effects indicating the possibility of a unique selection scheme for these traits.

The genomic region on Chr. 18 that spans from 0–5.5 cM and contained QTL for protein (qPRO001), oleic acid (qOLE003), and Valine (qVAL001) contents was a newly discovered region, in which no QTL have been previously reported for any agronomic traits (Grant et al., 2010); however, QTL for SDS and SCN resistance have been identified and mapped approximately 2–3 cM downstream of the region (Kassem et al., 2006; Concibido et al., 2004).

In a recent study, we used the ‘MD96-5722’ by ‘Spencer’ RIL population and mapped three QTL for Thr on Chr. 5, Chr. 6, and Chr. 14; two QTL for Pro on Chr. 1 and Chr. 14, Ser on Chr. 5 and Chr. 6; Cys on Chr. 5 and Chr. 9; and Trp on Chr. 9 (Khandaker et al., 2015). A QTL for Arg and another for His were mapped on Chr. 7 and Chr. 16, respectively (Khandaker et al., 2015). The two QTL for Thr and Ser identified previously (Khandaker et al., 2015) are located in the same region containing qVAL001 on Chr. 6 in the present study using a different genetic background (H x S RILs), which may indicate that this region contains genes involved in amino acid biosynthetic pathways. Interestingly, the QTL for palmitic acid (qPAL002), glutamic acid (qGLU001), and linolenic acid (qLIN006) on chr. 6 were mapped 0.9 and 3.3 cM, respectively

from the QTL for Met and crude protein (*qMet_Gm06*) identified previously in the ‘Benning’ by ‘Danbaekkong’ RIL population (Warrington et al., 2015).

By using the high-throughput genotyping platform, more than 1500 high-quality SNP markers were obtained using Illumina Beadchip Array and a high-density genetic map was constructed using these markers. However, the cost of utility of the platform prevents us to repeat the genotyping in a validation population. Combining high density genetic map with the powerful phenotyping approaches on these quality traits, the QTL and SNP markers associated with the soybean quality traits for protein, oil, fatty acid, and amino acid identified in the present study will be useful in breeding programs to develop soybean cultivars with high seed quality, improved composition, and disease resistance.

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 protein, oil and fatty acids analysis. This research was partially funded by United States Department of Agriculture, Agricultural Research Service project number 6402-21220-012-00D.

References

- Akond, M., S. Liu, M. Boney, S. K. Kantartzi, K. Meksem, N. Bellaloui, and M. A. Kassem, 2014: Identification of Quantitative Trait Loci (QTL) Underlying Protein, Oil, and Five Major Fatty Acids’ Contents in Soybean. *American Journal of Plant Sciences* **5**, 158-167.
- Akond, M., S. Liu, S. K. Kantartzi, K. Meksem, N. Bellaloui, D. A. Lightfoot, and M. A. Kassem, 2015: Quantitative Trait Loci Underlying Seed Sugars Content in “MD96-5722”

- by “Spencer” Recombinant Inbred Line Population of Soybean. *Food and Nutrition Sciences* **06**, 964-973.
- Bachlava, E., R. E. Dewey, J. W. Burton, and A. J. Cardinal, 2009: Mapping and Comparison of Quantitative Trait Loci for Oleic Acid Seed Content in Two Segregating Soybean Populations. *Crop Science* **49**, 433 -442.
- Bellaloui, N., J. R. Smith, J. D. Ray, and A. M. Gillen, 2009: Effect of Maturity on Seed Composition in the Early Soybean Production System as Measured on Near-Isogenic Soybean Lines. *Crop Science* **49**, 608-620.
- Bellaloui, N., R. B. Turley, and S. R. Stetina, 2015: Water Stress and Foliar Boron Application Altered Cell Wall Boron and Seed Nutrition in Near-Isogenic Cotton Lines Expressing Fuzzy and Fuzzless Seed Phenotypes. *PLoS ONE* **10**, e0130759.
- Brouwer, I.A., A.J. Wanders, and M.B. Katan, 2010: Effect of Animal and Industrial Trans Fatty Acids on HDL and LDL Cholesterol Levels in Humans – A Quantitative Review. *PLoS ONE*, **5**, e9434.
- Burton, J. W., and C. A. Brim, 1981: Recurrent Selection in Soybeans. III. Selection for Increased Percent Oil in Seeds. *Crop Science* **21**, 31-34.
- Concibido, V. C., B. W. Diers, and P. R. Arelli, 2004: A Decade of QTL Mapping for Cyst Nematode Resistance in Soybean. *Crop Science* **44**, 1121-1131.
- Dornbos, D. L., and R. E. Mullen, 1992: Soybean Seed Protein and Oil Contents and Fatty Acid Composition Adjustments by Drought and Temperature. *Journal of the American Oil Chemists Society* **69**, 228-231.

- Fallen, B. D., V. R. Pantalone, C. E. Sams, D. A. Kopsell, S. F. Vaughn, and B. R. Moser, 2011: Effect of Soybean Oil Fatty Acid Composition and Selenium Application on Biodiesel Properties. *Journal of the American Oil Chemists' Society* **88**, 1019-1028.
- Fehr, W. R., 2007: Breeding for Modified Fatty Acid Composition in Soybean. *Crop Science* **47**, S-72-S-87.
- Grant, D., R. T. Nelson, S. B. Cannon, and R. C. Shoemaker (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucl. Acids Res.* (2010) 38 (suppl 1): D843-D846. doi: 10.1093/nar/gkp798.
- Gutierrez-Gonzalez, J. J., X. Wu, J. Zhang, J. Lee, M. Ellersieck, J. G. Shannon and D. A. Sleper, 2009: Genetic Control of Soybean Seed Isoflavone Content: Importance of Statistical Model and Epistasis in Complex Traits. *Theoretical and Applied Genetics* **119**, 1069-1083.
- Imssande, J., 2001: Selection of Soybean Mutants with Increased Concentrations of Seed Methionine and Cysteine. *Crop Science* **41**, 510-515.
- Kassem, M. A., J. Shultz, K. Meksem, Y. Cho, A. J. Wood, M. J. Iqbal, and D. A. Lightfoot, 2006: An Updated 'Essex' by 'Forrest' Linkage Map and First Composite Map of QTL Underlying Six Soybean Traits. *Theoretical and Applied Genetics* **113** (6), 1015-1026.
- Khandaker, L., M. Akond, S. Liu, S. K. Kantartzi, K. Meksem, N. Bellaloui, and M. A. Kassem, 2015: Mapping of QTL Associated with Seed Amino Acids Content in "MD96-by "Spencer" RIL Population of Soybean Using SNP Markers. *Food and Nutrition Sciences* **06**, 974-984.
- Kim, H. K., S. T. Kim, B. G. Son, Y. W. Choi, J. S. Kang, Y. H. Park, Y. S. Cho, and I. S. Choi, 2010: Analysis of Quantitative Trait Loci (QTL) for Seed Size and Fatty Acid

- Composition Using Recombinant Inbred Lines in Soybean. *Journal of Life Science* **20**, 1186-1192.
- Krishnan, H. B., 2005: Engineering Soybean for Enhanced Sulfur Amino Acid Content. *Crop Science* **45**, 454-461.
- Mao, T., Z. Jiang, Y. Han, W. Teng, X. Zhao and W. Li, 2013: Identification of Quantitative Trait Loci Underlying Seed Protein and Oil Contents of Soybean Across Multi-genetic Backgrounds and Environments. *Plant Breeding* **132**, 630-641.
- Mateos-Aparicio, I., A. R. Cuenca, M. J. Villanueva-Soares, and M. A. Zapata-Revilla, 2008: Soybean, a Promising Health Source. *Nutricion Hospitalaria* **23**, 305-312.
- Pathan, S. M., T. Vuong, K. Clark, J.D. Lee, J.G. Shannon, and C.A. Roberts, et al, 2013: Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Science* **53**, 765-774. 10.2135/cropsci2012.03.0153.
- Patil, G. Mian, R. Vuong, T. Pantalone, V. Song, Q. Chen, P. Shannon, G. S. Carter, T. C. and Nguyen H. T. 2017. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor Appl Genet*. DOI 10.1007/s00122-017-2955-8.
- Rotundo, J. L., L. Borrás, M. E. Westgate, and J. H. Orf, 2009: Relationship between Assimilates Supply per Seed during Seed Filling and Soybean Seed Composition. *Field Crops Research* **112**, 90-96.
- Shannon, G., 2012: High Oleic Acid Soybeans for Wider Use of Oil in Food, Fuel and Other Products. In Fifteenth Annual National Conservation Systems Cotton & Rice Conference, pp. 43.

- 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.
- Tasma, I. Lorenzen, D.E. Green, and Shoemaker R, 2001: Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Mol. Breed* **8**, 25-35.
- Ufaz, S., and G. Galili, 2008: Improving the Content of Essential Amino Acids in Crop Plants: Goals and Opportunities. *Plant Physiology* **147**, 954-961.
- Van Ooijen, J. W. 2006: JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations,” Kyazma BV, Wageningen, Netherlands.
- Vollmann, J., C. N. Fritz, H. Wagentristl, P. Ruckebauer, 2000: Environmental and Genetic Variation of Soybean Seed Protein Content under Central European Growing Conditions. *Journal of the Science of Food and Agriculture* **80**, 1300-1306.
- Wang, S., C. J. Basten, and Z. B. Zeng, 2012: Windows QTL Cartographer 2.5. Department of Statistics, NCSU, Raleigh, NC (<https://brcwebportal.cos.ncsu.edu/qtlcart/WQTLCart.htm>).
- Warrington, C. V. Abdel-Haleem, H. Hyten, D. L. Cregan, P. B. Orf, J. H. Killam, A. S. Bajjalieh, N. Li, Z. and Boerma, H. R. 2015. QTL for seed protein and amino acids in the Benning X Danbekkong soybean population. *Theor Appl Genet* **128**: 839–850.
- Wilcox, J. R., and J. F. Cavins, 1995: Backcrossing High Seed Protein to a Soybean Cultivar. *Crop Science* **35**, 1036-1041.
- Wilcox, J. R., and R. M. Shibles, 2001: Interrelationships among Seed Quality Attributes in Soybean. *Crop Science* **41**, 11-14.

- Wolf, R. B., J. F. Cavins, R. Kleiman, and L. T. Black, 1982: Effect of Temperature on Soybean Seed Constituents: Oil, Protein, Moisture, Fatty Acids, Amino Acids and Sugars. *Journal of the American Oil Chemists' Society* **59**, 230-232.
- Yaklich, R. W., B. Vinyard, M. Camp, and S. Douglass, 2002: Analysis of Seed Protein and Oil from Soybean Northern and Southern Region Uniform Tests. *Crop Science* **42**, 1504-1515.
- Yesudas, C. R., R. Bashir, M. Geisler, and D. A. Lightfoot, 2013: Identification of germplasm with stacked QTL underlying seed traits in an inbred soybean population from cultivars Essex and Forrest. *Molecular Breeding* **31**: 693- 703.

Table 1. Mean, range, standard deviation and coefficient of variation (CV) for seed protein, oil, fatty and amino acids contents based on % dry weight of soybean cultivars ‘Hamilton’ and ‘Spencer’ and their F_{5:7} RILs in 2012 in North Carolina.

Traits	Hamilton x Spencer RIL (F _{5:7})				Parental lines		
	Mean	Range	SD	CV (%)	P1	P2	P-value
Protein_DB (%)	44.53	9	1.97	4.42	40.55	45.55	<0.0001
Oil_DB (%)	20.42	6.6	1.24	6.09	21.03	19.28	<0.0001
Palmitic_DM (%)	11.11	6.6	1.2	10.78	12.08	12.33	<0.0001
Stearic_DM (%)	3.91	1.7	0.37	9.41	4.20	3.50	0.13
Oleic_DM (%)	25.52	22.2	4.27	16.72	19.45	18.35	<0.0001
Linoleic_DM (%)	52.66	14.7	3.52	6.7	55.65	52.78	0.003
Linolenic_DM (%)	6.69	8.7	1.51	22.55	8.88	13.00	<0.0001
Aspartic_DB (%)	5.05	1.2	0.28	5.59	4.60	5.20	<0.0001
Threonine_DB (%)	1.75	0.3	0.06	3.69	1.68	1.80	0.002
Serine_DB (%)	2.38	0.6	0.13	5.54	2.18	2.43	<0.0001
Glutamic_DB (%)	7.25	2.1	0.51	7	6.33	7.43	<0.0001
Proline_DB (%)	2.41	0.5	0.12	4.83	2.25	2.40	0.002
Glycine_DB (%)	2.27	1	0.19	8.37	2.15	2.15	1
Alanine_DB (%)	2.13	0.4	0.09	4.36	1.98	2.10	0.35
Cysteine_DB (%)	0.36	0.3	0.06	15.72	0.30	0.33	0.35
Valine_DB (%)	2.54	0.8	1.66	6.51	2.38	2.55	0.003
Methion_DB (%)	0.64	0.4	0.06	9.7	0.60	0.60	NA
Isoleuc_DB (%)	2.28	0.4	0.1	4.24	2.15	2.30	0.002
Leucine_DB (%)	3.47	0.9	0.18	5.41	3.28	3.60	0.0005
Tyrosine_DB (%)	1.81	0.4	0.09	4.89	1.73	1.88	0.005
Phenylal_DB (%)	2.18	0.4	0.1	4.7	1.93	2.15	0.001
Lysine_DB (%)	2.54	0.7	0.15	6.02	2.43	2.60	<0.0001
Histidine_DB (%)	1.17	0.9	0.16	13.58	0.90	1.13	0.003
Arginine_DB (%)	3.4	1.5	0.22	6.56	3.10	3.40	<0.0001
Tryptoph_DB (%)	0.49	0.1	0.02	4.28	0.40	0.43	0.355

Table 2. Chromosomal locations and parameters associated with the quantitative trait loci (QTL) of protein, oil, fatty and amino acids contents in the ‘Hamilton’ by ‘Spencer’ RIL population of soybean.

Trait	QTL	Chr/LG	Peak Position (cM) ^a	2-LOD Support Interval (cM) ^b	Markers Interval	Peak LOD ^c	R ² (%) ^d	Additive Effect ^e
Protein	<i>qPRO001</i>	18/G	1.40	0.00-3.90	ss249909538- ss249919445	2.50	0.10	4.32
Oil	<i>qOIL001</i>	6/C2	9.10	7.00-11.70	ss246100375- ss245879277	39.47	0.76	10.09
Palmitic	<i>qPAL001</i>	6/C2	11.20	7.70-12.90	ss246100375- ss245879277	2.71	0.21	1.87
Palmitic	<i>qPAL002</i>	6/C2	26.70	26.50-27.30	ss245914593- ss245908292	2.98	0.21	1.81
Oleic	<i>qOLE001</i>	3/N	62.30	59.10-67.80	ss245246964- ss245259860	2.65	0.09	2.79
Oleic	<i>qOLE002</i>	6/C2	26.70	26.30-27.30	ss245914593- ss245908292	2.83	0.15	4.35
Oleic	<i>qOLE003</i>	18/G	3.10	0.40-5.40	ss249909538- ss249506152	5.00	0.12	2.96
Linolenic	<i>qLIN001</i>	6/C2	39.40	36.30-40.00	ss245914593- ss245790648	2.66	0.13	-0.82
Glutamic	<i>qGLU001</i>	6/C2	26.70	26.30-27.30	ss245914593- ss245908292	2.51	0.14	1.07
Alanine	<i>qALA001</i>	6/C2	9.20	7.00-9.40	ss246100375- ss245879277	50.24	0.74	1.05
Cysteine	<i>qCYS001</i>	6/C2	26.60	25.90-27.30	ss245898080- ss245908292	2.53	0.11	0.06
Valine	<i>qVAL001</i>	18 /G	1.40	0.00-4.00	ss249909538- ss249919445	2.62	0.11	0.25
Histidine	<i>qHIS001</i>	6/C2	26.70	26.00-27.30	ss245898080- ss245908292	2.70	0.12	0.19

Note: ^a Position of peak LOD value on composite maps described previously (Akond et al., 2015).

^b The positions that define the two LOD intervals around the position of peak likelihood for the QTL.

^c The log of odds (LOD) value at the position of peak likelihood of the QTL.

^d R² estimates the proportion of RIL mean variance (%) explained by the detected QTL.

^e A positive number in additive effect of the QTL indicates that the allele has positive effect on the quality trait and a negative number means that the allele has a negative effect on the trait.

Figure 1. Locations of SNP markers and the QTL that underlie soybean seed protein (qPRO), oil (qOIL), palmitic (qPAL), stearic (qSTEL), Oleic (qOLE), Linoleic (qLINL) and Linolenic (qLINN) acid contents (% on dry based) in the 'Hamilton' by 'Spencer' RIL populations. Chr./LGs were drawn using MapChart according to the 'Hamilton' by 'Spencer' genetic linkage map (Akond et al., 2015). The numbers in squared parentheses indicate the number of intervals inferred.

For Review Only

Figure 1.

LG-C2(Chr_6)[2]

LG-G(Chr_18)