J. Crop Sci. Biotech. 2017 (December) 20 (4) : 243 ~ 253
DOI No. 10.1007/s12892-017-0020-0

RESEARCH ARTICLE

Identifying and Exploring Significant Genomic Regions Associated with Soybean Yield, Seed Fatty Acids, Protein and Oil

Christopher J Smallwood[1]*, Jason D Gillman[2], Arnold M Saxton[3], Hem S Bhandari[1], Phillip A Wadl[4], Benjamin D Fallen[5], David L Hyten[6], Qijian Song[1], Vincent R Pantalone[1]

[1]*Department of Plant Sciences, University of Tennessee, 2431 Joe Johnson Drive, Knoxville, TN 37996, USA*
[2]*Plant Genetic Research Unit, USDA-ARS, University of Missouri, 110 Waters Hall, Columbia, MO 65211, USA*
[3]*Department of Animal Science, University of Tennessee, 2506 River Drive, Knoxville, TN 37996, USA*
[4]*U.S. Vegetable Laboratory, USDA-ARS, 2700 Savannah Highway, Charleston, SC 29414, USA*
[5]*Clemson University, Advanced Plant Technology, Clemson Pee Dee REC , 2200 Pocket Road, Florence, SC 29506, USA*
[6]*Department of Agronomy & Horticulture, University of Nebraska-Lincoln, 322 Keim Hall, Lincoln, NE 68583, USA*

Received: February 10, 2017 / Revised: October 20, 2017 / Accepted: October 26, 2017

Abstract

Soybean [*Glycine max* (L.) Merrill] yield and seed fatty acids, protein, and oil content are important traits for which an improved understanding of significant genomic regions would be useful. To accomplish this, a soybean population consisting of 203 F_5 derived recombinant inbred lines (RILs) was developed and genotyped with 11,633 polymorphic single nucleotide polymorphisms (SNPs). Each RIL was grown in a single plot at Knoxville, TN in 2010; followed by replicated, multi-location field trials in 2013 and 2014. The data from 2010, 2013, and 2014 were analyzed together in order to detect quantitative trait loci (QTL) for these traits, and 30 total QTLs were detected. Five QTLs are candidates for confirmed status and one QTL is a candidate for positional confirmation. Many of the genes with mutations in close proximity to the fatty acid QTLs are involved in biological processes for fatty acids and/or lipids and could be considered possible candidate genes. Similarly, genes with mutations in genomic regions near yield, protein, and oil QTLs were plentiful and may contribute to the variation observed in these traits. Except for yield and stearic acid, each trait displayed pleiotropic effects with other traits in this study. Notable are the pleiotropic effects for oleic and linolenic acid on chromosomes 9, 13, and 19. Overall, the findings from this research contribute new information to the genetic understanding of soybean yield and seed fatty acids, protein and oil content. This understanding will be useful in making trait improvements.

Key words : Candidate gene, molecular breeding, pleiotropy, QTL, soybean

Introduction

Soybean [*Glycine max* (L.) Merrill] is a prominent crop grown throughout much of the world for many purposes. Seed protein (~400 g kg^{-1}) and oil (~200 g kg^{-1}) are primary components of soybean that contribute to its high value. These traits are common targets for research efforts to improve the value of soybean. Additionally, with the Food and Drug Administration (FDA) removal of "generally recognized as safe" status for partially hydrogenated oils (PHOs) (www.federalregister.gov, "Final Determination Regarding Partially Hydrogenated Oils", accessed 7/24/2015), improving the fatty acid profile of soybean has become a major breeding objective.

The five primary fatty acids in soybean seeds are palmitic 16:0), stearic (18:0), oleic (18:1), linoleic (18:2), and linolenic (18:3); typically occurring in relative concentrations of 100, 40, 220, 540, and 100 g kg^{-1} of total lipids, respectively (Wilson 2004). Due to the FDA ban on PHOs, a major initiative in

Christopher J Smallwood (✉)
Email: csmallwood@utk.edu
Tel: +865 9745215 Fax: +8659741947

fatty acid improvement is to reduce linolenic acid (< 30 g kg^{-1}), thereby increasing oxidative stability of soybean oil and reducing the need for hydrogenation, a process that creates partially hydrogenated soybean oil. Increasing the monounsaturated oleic acid (> 800 g kg^{-1}) is another major goal, as the oxidative stability of soybean oil is improved by increasing the concentration of oleic acid, leading to increased shelf life for soybean oil food products (Kinney 1996) and biodiesel (Fallen et al. 2012; Kinney and Clemente 2005). Further, consumption of oleic acid has been shown to lower cholesterol in comparison with saturated fatty acids when consumed by humans (Kris-Etherton and Yu 1997). When comparing saturated fatty acids, there is evidence that stearic acid is neutral with respect to cholesterol in humans, which differs from the hypercholesteromic effect of palmitic acid (Kris-Etherton and Yu 1997). Thus, aside from specialty uses, improving soybean saturated fatty acid involves increasing stearic acid while reducing palmitic acid.

Many studies have focused on improving soybean fatty acids (Bilyeu et al. 2011; Boersma et al. 2012; Gillman et al. 2014; Pantalone et al. 2002; Pham et al. 2010). However, there is still a need for further improvement. For stearic acid, competitively yielding breeding lines have yet to achieve the targeted goal of (> 200 g kg^{-1}) of total oil (Gillman et al. 2014). For oleic acid there is concern that environmental variation may result in levels below 800 g kg^{-1} (Fallen et al. 2012; Lee et al. 2012). Breeding with small effect modifier quantitative trait loci (QTLs) for fatty acids in addition to major QTLs may facilitate greater trait stability by capturing a greater portion of total genetic variation (Hyten et al. 2004b). Additionally, determining the genetic origins and pleiotropic effects of such QTLs would be useful for making breeding improvements (Cardinal et al. 2014). Thus, our objectives were to identify QTLs for soybean yield and seed fatty acids, protein, and oil content, and to examine these QTLs for pleiotropic effects and candidate genes.

Materials and Methods

Plant materials

A population of 860 recombinant inbred lines (RILs) with genotypic and phenotypic data derived from parental lines Essex and Williams 82 (hereafter known as E×W-50K) was created for QTL detection. Essex is a maturity group (MG) V cultivar with a determinate growth habit, purple flower and gray pubescence (Smith and Camper 1973); while Williams 82 is a MG III cultivar with indeterminate growth habit, white flower and tawny pubescence (Bernard and Cremeens 1988). Seed for both Essex and Williams 82 were obtained from the USDA Soybean Germplasm Collection (www.ars-grin.gov). In order to provide highly homozygous parental lines for RIL development, a random single plant of each parental line was intentionally selfed for two additional generations.

Each of the 860 F_5 derived RILs developed through single seed descent (Brim 1966) were grown in single replicate plots in 2010 in Knoxville, TN (35° 54' 15"N, 83° 57' 13"W). Along with the RILs and the parents, four checks with relevant maturities were included in the 2010 field test; LD00-3309 (MG early-IV) (Diers et al. 2006), IA4004 (MG early-IV), 5002T (MG early-V) (Pantalone et al. 2004), and 5601T (MG mid-V) (Pantalone et al. 2003). A subset of this population (276 RILs) ranging in maturity from MG mid-IV to late-IV was selected for advancement into replicated field trials planted in 2013. The 2013 field test design was a randomized complete block design (RCBD) with three replications per environment at three environments [Knoxville, TN (35° 54' 15"N, 83° 57' 13"W); Springfield, TN (36° 28' 12"N, 86° 50' 31"W); and Milan, TN (35° 56' 3"N, 88° 43' 44"W)], representative of the eco-geographic regions of East, Middle, and West Tennessee, respectively. In addition to the RILs and parents, three maturity checks were included; LD00-3309 (MG early-IV), LD00-2817P (MG mid-IV) (Diers et al. 2010), and Ellis (MG late-IV). Using data from the combined 2010 and 2013 growing seasons, a new subset of 203 RILs ranging in maturity from MG mid-IV to late-IV was selected for advancement into replicated field trials planted in 2014. The field design and locations were consistent from 2013 to 2014, with IA4005 (MG early-IV) included as an additional maturity check in 2014. For each location, plots consisted of two adjacent 6.1 m rows (end-trimmed to 4.9 m at ~R6 growth stage), with the rows spaced 0.8 m apart. Soil type was primarily Shady Loam and Shady-Whitwell Complex in Knoxville, TN, Dickson and Staser Silt Loams in Springfield, TN, and Loring and Routon Silt Loams in Milan, TN. For each field season plant height and maturity were determined at the R8 growth stage (Fehr and Caviness 1977), with plots harvested at maturity. Yield was measured in kg ha^{-1} after adjusting the plot weight to 13% moisture.

Seed quality trait detection

Determination of fatty acid content for palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid were performed by gas chromatography following the procedures of Spencer et al. (2004). Seeds from each plot from the 2010, 2013, and 2014 field tests were analyzed using a Hewlett Packard HP 6890 series gas chromatograph (Agilent Technologies, Santa Clara, CA). Fatty acids were initially estimated as a percentage of seed oil, but were converted to g kg^{-1} seed oil.

To obtain estimates of protein and oil content, near infrared reflectance spectroscopy (NIRS) was used. Approximately 25 g seed samples from each plot in 2010 were uniformly ground for 20 sec in a Knifetec 1095 Sample Mill (FOSS Tecator, Hoganas, Sweden). The NIRS instrument (NIR 6500, FOSS North America, Eden Prairie, MN) was used for analysis as described by Panthee et al. (2006a), except that samples in this study were scanned using updated ISIscan software v. 2.85. Seed from each plot from the 2013 season was scanned as a whole bean sample using a Perten DA 7200 Diode Array (Perten, Hägersten, Sweden) NIRS instrument at the University of Minnesota, with calibration equations developed through a cooperative effort between Perten and the University

of Minnesota (Bolon et al. 2011). Seed from plots from the 2014 growing season were scanned using the same procedure as the 2013 plots with a Perten DA 7250 Diode Array NIRS instrument at the University of Tennessee. Values for protein and oil concentration were adjusted to g kg^{-1} of total seed on a dry weight basis for each NIRS analysis.

SNP genotyping

DNA was isolated from leaf tissue of each F_5 plant used for RIL derivation in 2009 and analyzed with the Illumina Infinium beadchip SoySNP50K (Song et al. 2013) at the Soybean Genomics Improvement Laboratory at the USDA Beltsville Agricultural Research Center (USDA-ARS) in Beltsville, MD. Marker positions from this analysis were obtained from the genetic map estimated in Song et al. (2016). Imputations using default settings in the Beagle Genetic Analysis Software Package v. 3.3.1 (Browning and Browning 2007, 2009) via the 'synbreed' package (Wimmer et al. 2012) in R were used to address missing marker data. Finally, the 'calc.errorlod' function within the 'qtl' package (Broman et al. 2003) in R was used to screen for potential genotyping errors.

Statistical analysis and QTL detection

A mixed model analysis was used in SAS PROC GLIMMIX (SAS Institute Inc., Cary, NC, USA, SAS 9.4, 2002-2012) in order to estimate least squares means (LSMEANS) for the 203 RILs from the combined 2010, 2013, and 2014 datasets. The different years for the analysis were treated as separate environments, resulting in seven total environments and 19 total reps. The fixed term for the model was RIL, with environment, rep(environment), and RIL × environment as random terms, and denominator degrees of freedom method set to residual. Covariance parameter estimates and Wald Z tests were obtained in order to determine the significance of the RIL× environment term in the model. Estimate statements for indeterminate (93 RILs), segregating (11 RILs) and determinate (99 RILs) genotypes, as well as contrast statements for comparison of each stem type were included in this analysis. Pearson correlations were performed between the LSMEANs for each trait using the 'Hmisc' package (Harrell 2015) in the R language and environment for statistical computing (R Core Team 2015).

An additional model with no fixed terms and RIL, environment, rep(environment), and RIL × environment as random terms was run for all analyses in order to obtain the variance for each term. These variances were then used to estimate broad-sense heritability on an entry means basis (Nyquist 1991).

The LSMEANS from the 203 RILs for yield, fatty acids, protein and oil were combined with the 11,633 polymorphic SNPs and used to detect QTLs for these traits with the 'qtl' package (Broman et al. 2003) in the R language and environment for statistical computing (R Core Team, 2015). Since this population segregates at the Dt1 (growth habit) locus, SNPs located adjacent to (< 6 kilo base-pairs) the Dt1 (ss715635422 and ss715635423, confirmed by field calls) locus based on the Wm82.a2.v1 genome sequence were used to predict the parental allele. Interval mapping (IM) was the primary method of detection so that the Dt1 locus could be included as an additive covariate. A chi-squared test was performed in the R language and environment for statistical computing (R Core Team 2015) to test for segregation distortion at the Dt1 locus.

In addition to IM, QTL estimates with composite interval mapping (CIM) were used as a safe guard against 'ghost' QTL (Martinez and Curnow 1992), and multiple interval mapping (MIM) to screen for QTL interactions (Zeng et al. 1999). The CIM analysis was not used as the primary method because it did not include an option to add the Dt1 locus as a covariate in the 'qtl' package (Broman 2003). For both IM and CIM, step size was one cM and method was the Haley-Knott regression; otherwise default settings were used. In the IM procedure, 10,000 permutations were performed for each trait in order to determine significance thresholds of 1 and 5% (Churchill and Doerge 1994). Each QTL from the IM procedure was used for 'makeqtl' (Broman 2003) assignments. Then the R^2 and estimated effect for each QTL were determined with 'fitqtl' (Broman 2003) with the Dt1 locus included as a covariate and the method set to Haley-Knott regression. The MIM analysis was performed with a reduced set of genotypic data containing 2602 SNPs by using the 'findDupMarkers' and 'drop.markers' functions. The QTLs detected with IM were used to establish initial models for the MIM analysis, with a step size of five cM and 1000 permutations to determine a 1% significant threshold. As with IM, Dt1 was used as a covariate and method was the Haley-Knott regression for MIM.

DNA sequencing and candidate gene search

DNA was isolated from ~40 mg of lyophilized seedling leaf tissue from a single plant each for Essex and Williams 82 using the DNeasy Plant Mini Kit (Qiagen). Isolated DNA was randomly sheared using a Covaris S2 Adaptive Focused Acoustic Disruptor instrument (Covaris, Inc., Woburn, MA, USA) to an average fragmentation size of 200 basepairs. Sheared DNA was then used to prepare indexed Illumina libraries by Global Biologics, LLC (Columbia, MO) using standard Illumina adapters. Libraries were then paired-end sequenced (2x 100 bps) using a HiSeq 2000 instrument at the DNAcore facility at the University of Missouri.

All resequencing analyses used the Wm82.a2.v1 genome sequence and annotation build, which was downloaded from (http://phytozome.jgi.doe.gov/). Read mapping and variant calling were done using CLC Genomics Workbench software Version 8.0 (Qiagen/CLC Biotech, Cambridge, MA). Read mapping used the following settings: insertion cost of 3, deletion cost of 3, a similarity fraction setting of 0.8, automatic detection of paired-end distances, and non-specific matching handling was set to ignore. For variant calling, the fixed ploidy program was used with the settings: minimum coverage of 9, variant probability cutoff of 90.0, maximum variant count of 2, and non-specific read matches were ignored. Total variant

Table 1. Simple statistics taken from combined analysis using data from 2010, 2013, and 2014 field seasons for soybean population E×W-50K subset consisting of 203 F_5 derived RILs. Trait estimates for indeterminate (93 RILs), segregating (11 RILs), and determinate (99 RILs) genotypes included for comparison.

Trait	Genotype P value	GeE Z value	Essex	Williams 82	min	mean	max	Ind.	Seg.	Det.	std. dev.[a]	LSD value	h^{2-b}
Maturity (Julian)	***	***	273.8	264.7	269.7	272.5	277.1	272.9	272.9	272.1	1.5	1.8	0.81
Height (cm)	***	***	75.3	85.7	57.4	90.3	125.1	106.7	90.5	74.9	16.9	7.9	0.97
Yidld (kg ha^{-1})	***	***	3296.9	2463.9	2298.0	3048.3	3706.1	2958.8	2948.7	3143.4	287.6	480.9	0.64
Palmitic (g kg^{-1} seed oil)	***	***	108.3	101.5	93.7	105.4	116.0	103.1	103.6	107.7	4.5	2.6	0.96
Stearic (g kg^{-1} Seed oil)	***	***	40.8	37.9	32.7	38.7	47.1	40.5	40.0	37.0	3.1	1.9	0.96
Oleic (g kg^{-1} seed oil)	***	***	226.5	245.9	177.7	233.9	346.5	258.3	257.2	208.4	31.4	16.3	0.97
Linolenic (g kg^{-1} seed oil)	***	***	550.9	546.7	458.9	551.6	592.9	532.1	532.9	571.9	25.1	13.7	0.96
Linolenic (g kg^{-1} seed oil)	***	***	73.5	68.0	57.7	70.5	86.3	66.2	66.2	75.0	6.2	3.5	0.96
Protein (g kg^{-1} seed dry weight)	***	***	432.6	429.6	391.0	420.3	449.0	423.5	419.1	417.5	10.7	8.1	0.93
Oil (g kg^{-1} seed dry weight)	***	***	210.1	225.9	198.0	216.7	234.7	218.4	219.1	214.8	5.4	3.8	0.94

***Significant at 0.001 probability level.
E×W-50K = soybean population with parental lines Essex and Williams 82.
a) std. deviation of LSMEANs.
b) heritability calculated using entry means basis (Nyquist, 1991).

calls were filtered to call genes with mutations resulting in amino acid changes.

A candidate gene search for major genes affecting soybean fatty acids in close proximity to QTLs detected in this study was performed using selected genes listed in Gillman and Bilyeu (2012). Given the relatively normal range of the fatty acids in this population (Table 1) when compared with other studies (Bilyeu et al. 2011; Boersma et al. 2012; Gillman et al. 2014; Pham et al. 2010; Pantalone et al. 2002), it seems reasonable to assume that fatty acid QTLs from this study would be for small effect modifier genes rather than large effect genes. Thus, a candidate search for genes with sequence differences leading to amino acid changes between the parent lines was performed for these small effect fatty acid QTLs; as well as for yield, protein, and oil QTLs. Genes with parental mutations located between terminal SNPs of LOD – 1.5 support intervals (Dupuis and Siegmund 1999) for each QTL were processed to determine candidate genes based on gene ontologies (GO) (www.SoyBase.org, "SoyBase Gene Model Data Mining and Analysis", accessed 7/26/2016).

Results

Each of the traits studied for QTLs (yield and seed fatty acids, protein, and oil content) displayed a significant difference ($P < 0.001$) among RILs from the combined 2010, 2013, and 2014 field seasons (Table 1). Also, significant differences ($P < 0.001$) were observed between indeterminate and determinate genotypes for all of the traits in this study, with stem height termination estimates provided in Table 1. The Dt1 locus displayed the expected segregation pattern ($P > 0.75$) for an F_5 derived RIL population. This was not the case for the E1 maturity locus, which displayed extreme segregation distortion ($P < 0.0001$), with the Essex E1 allele primarily represented. A possible explanation for this is the continued narrowing of the population through maturity selection, leading to a greater uniformity at the E1 locus among the 203 remaining RILs.

Transgressive segregation, with LSMEANs for high and low RILs significantly different ($P < 0.05$) from both parents, was observed for the fatty acids, protein, and oil (Table 1). This finding indicates that both parents carry different alleles for seed fatty acids, protein, and oil content. The high heritability estimates, ranging from 0.64 (yield) to 0.97 (oleic) provide further evidence for the genetic variation in this study useful for detecting significant genomic regions (Table 1). These estimates exceed those from previous studies using the same formula for heritability calculation (Hyten et al. 2004a, 2004b; Panthee et al. 2005, 2006b; Wiggins, 2012).

Pearson correlations between the LSMEANs for each trait are provided in Table 2. Maturity was significantly correlated ($P < 0.05$) with yield and each fatty acid; however, these correlations were quite low, ranging from -0.24 (maturity: palmitic) to 0.25 (maturity:stearic). Modest significant correlations ($P < 0.05$) were observed between yield and each fatty acid, with values ranging from -0.34 (yield:oleic) to 0.35 (yield:linoleic). Each of the fatty acids were significantly

Table 2. Pearson correlations between LSMEAN estimates for soybean maturity, yield, fatty acids, protein, and oil using data from 2010, 2013, and 2014 field seasons for soybean population E×W-50K subset consisting of 203 F_5 derived RILs.

	Maturity	Yield	Palmitic	Stearic	Oleic	Linoleic	Linolenic	Protein	Oil
Maturity	\	0.20	-0.24	0.25	0.18	-0.16	-0.20	-0.02	0.05
Yield	**	\	0.22	-0.22	-0.34	0.35	0.29	-0.07	-0.13
Palmitic	***	***	\	-0.37	-0.65	0.52	0.66	-0.14	-0.42
Stearic	***	**	***	\	0.53	-0.60	-0.50	0.21	0.22
Oleic	*	***	***	***	\	-0.98	-0.90	-0.27	0.35
Linoleic	*	***	***	***	***	\	0.83	-0.28	-0.28
Linolenic	**	***	***	***	***	***	\	-0.25	-0.47
Protein	NS	NS	*	**	***	***	***	\	-0.43
Oil	NS	NS	***	**	***	***	***	***	\

*Siginficant at 0.05 probability level.
**Siginficant at 0.01 probability level.
***Siginficant at 0.001 probability level.
NS, Not significant at 0.05 probability level.
E×W-50K = soybean population with parental Lines Essex and Williams 82.

correlated ($P < 0.05$) with each other and with protein and oil. Among the fatty acids, stearic and oleic were positively correlated with each other and negatively correlated with palmitic, linoleic, and linolenic. Overall, the fatty acid correlations ranged from -0.98 (oleic:linoleic) to 0.83 (linoleic: linolenic). These very strong correlations could be due to genetic differences affecting fatty acid biosynthesis and/or accumulation, with monounsaturated and polyunsaturated fatty acids trending very strongly in opposing directions (Table 2). This result is expected, given that oleic acid is a precursor to linoleic acid in fatty acid biosynthesis. In agreement with previous research, protein and oil were significantly ($P < 0.001$) negatively correlated with each other (Yaklich et al. 2002).

The approximate sequence coverage with respect to the Wm82.a2.v1 reference genome was 93% (~16.6 Giga base-pairs) for Essex and 94% (~15.6 Giga base-pairs) for Williams 82, with an average depth of 15.76x and 14.84x, respectively. Excluding genomic regions with no coverage, the average depth increases to 16.92x for Essex and 15.73x for Williams 82. With respect to the Wm82.a2.v1 reference genome, Essex had approximately 831,000 total nucleotide differences, with 14,000 resulting in amino acid changes, whereas, Williams 82 had approximately 69,000 total nucleotide differences and 1,800 amino acid changes. The nucleotide and amino acid differences between this strain of Williams 82 and the one used for the Wm82.a2.v1 reference genome can possibly be accounted for by intracultivar genetic heterogeneity (Haun et al. 2011). The mutations resulting in amino acid differences (> 12,000) between parents are indicative of the gene pools for origination; Essex in the southern pool and Williams 82 in the northern pool. The large genomic differences between parents provide plentiful opportunities for detecting significant genomic regions for yield and seed fatty acids, protein, and oil content.

Overall, there were 30 QTLs detected in this study for soybean yield, fatty acids, protein and oil (Table 3). No significant QTL interactions ($P < 0.01$) were detected for any of these traits. For clarity, each QTL significant at the 1% LOD threshold will be presented with bold text. Of the additive QTLs, only one, located on chromosome 14, was detected for yield. This QTL, designated as **Yld14**, was highly significant ($P < 0.01$), with a LOD score of 13.3, an R^2 of 0.24, and an effect of 147.27kg ha^{-1} (Table 3). There were 18 unique genes with mutations resulting in parental amino acid differences located between terminal SNPs of the **Yld14** LOD – 1.5 support interval (Table 3).

For soybean seed fatty acids, a total of 18 QTLs were detected, with six for palmitic, one for stearic, three for oleic, three for linoleic, and five for linolenic (Table 3). The palmitic acid QTLs were designated **Pal1** ($R^2 = 0.04$), **Pal4** ($R^2 = 0.05$), **Pal9** ($R^2 = 0.12$), **Pal17.1** ($R^2 = 0.03$), **Pal17.2** ($R^2 = 0.03$), and Pal19 ($R^2 = 0.04$), which collectively explain 31% of the palmitic variation (Table 3). With regard to palmitic acid candidate gene search, numerous genes with mutations resulting in parental amino acid differences were located between terminal SNPs within the LOD – 1.5 support interval for **Pal1** (23), **Pal4** (69), **Pal9** (12), **Pal17.1** (9), **Pal17.2** (12), and Pal19 (28) (Table 3).

The stearic acid QTL, located on chromosome 14, was designated as **Ste14** ($R^2 = 0.14$) (Table 3). None of the 13 genes with mutations resulting in parental amino acid differences in the **Ste14** support interval are involved with lipid or fatty acid biological processes.

Three QTLs on chromosomes 9, 13, and 19 were detected for oleic acid (Table 3). These QTLs were designated as Ole9 ($R^2 = 0.02$), Ole13 ($R^2 = 0.02$), and **Ole19** ($R^2 = 0.03$), respectively, and collectively explained 7% of the variation for oleic acid in this study (Table 3). Several genes with mutations leading to parental amino acid differences were located between terminal SNPs within the LOD – 1.5 support intervals for Ole9 (72), Ole13 (21), and **Ole19** (9).

The QTLs for linoleic acid detected in this study, designated as Lin13.1 ($R^2 = 0.02$), **Lin13.2** ($R^2 = 0.02$), and Lin19 ($R^2 = 0.03$) were located on chromosomes 13, 13, and

Table 3. Quantitative trait loci (QTL) for yield, fatty acids, protein, and oil detected in soybean population E×W-50K subset consisting of 203 F_5 derived RILs using 11,633 SNPs. Phenotypic data was estimated from the combined analysis using data from 2010, 2013, and 2014 field seasons. Significant QTL at the 1% threshold (based on 10,000 permutations) are displayed in bold red text.

Trait	LOD Threshold	QTL	Chr	Position (cM)	LOD – 1.5 Interval (cM)	LOD	R^{2-a}	Effect[b]	Closest SNP	SNP Position (Wm82.a2.v1)	No. Genes[c]
Yield	1% 4.2 5% 3.4	YLD14	14	6.7	6.5-14	13.3	0.24	147.27	ss715617871	1,736.967	18
Palmitic	1% 4.1	Pal1	1	120.0	117.6-121.4	4.4	0.04	0.88	ss715580677	56,143,364	23
	5% 3.4	Pal4	4	78.4	72.6-97.0	4.6	0.05	1.06	ss715587857	40,276,263	69
		Pal9	9	80.0	79.4-85.0	9.1	0.12	-1.60	ss715603983	40,608,710	12
		Pal17.1	17	4.0	0-14.0	4.3	0.03	-0.94	ss7155626513	2,485,630	9
		Pal17.2	17	37.6	22.0-48.0	4.6	0.03	-0.87	ss7155628181	7,523,398	12
		Pal19	19	83.9	83.0-84.7	3.6	0.04	-1.77	ss7155635339	44,515,446	28
Stearic	1% 4.1 5% 3.4	Ste14	14	69.9	68.0-70.1	9.4	0.14	-1.20	ss715617435	10,422,143	13
Oleic	1% 4.0	Ole9	9	91.1	85.0-107.0	3.6	0.02	4.86	ss715604287	42,611,042	72
	5% 3.3	Ole13	13	8.1	4.0-12.0	3.5	0.02	-4.84	ss715617125	14,520,516	21
		Ole19	19	11.4	5.9-18.0	5.3	0.03	5.65	ss715633470	2,406,009	9
Linoleic	1% 4.1	Lin13.1	13	99.0	95.0-105.0	4.0	0.02	4.10	ss715615227	31,950,419	23
	5% 3.3	Lin13.2	13	136.0	111.0-139.0	4.4	0.02	3.88	ss715616006	38,026,151	51
		Lin19	19	11.4	5.0-19.0	4.0	0.03	-4.85	ss715633470	2,406,009	11
Linolenic	1% 4.1	Len9.1	9	19.5	17.0-33.0	4.7	0.04	-1.25	ss715603590	3,209,966	24
	5% 3.4	Len9.2	9	91.1	87.3-97.2	8.0	0.06	-1.55	ss715604287	42,611,042	58
		Len13	13	6.0	1.4-10.6	7.1	0.06	1.61	ss715617125	14,520,516	63
		Len17	17	1.3	0-7.0	3.4	0.01	-0.63	ss715626513	2,485,630	7
		Len19	19	11.5	7.9-22.0	5.4	0.02	-1.01	ss715633481	2,435,311	6
Protein	1% 4.0	Prot6.1	6	27.0	16.0-39.0	3.8	0.04	-2.43	ss715595361	5,729,763	1
	5% 3.4	Prot6.2	6	145.0	140.05-147.5	3.8	0.07	3.24	ss715594897	48,464,349	40
		Prot7	7	50.7	44.0-53.2	3.8	0.04	-2.31	ss715598793	8,309,503	49
		Prot9.1	9	4.5	0-11.0	3.9	0.07	-2.96	ss715605440	888,248	44
		Prot9.2	9	79.1	75.5-80.0	5.7	0.07	2.85	ss715603959	40,376,477	24
		Prot13.1	13	114.0	107.0-123.0	6.0	0.02	-1.73	ss715615584	34,821,865	1
		Prot13.2	13	144.0	129.0-153.3	5.6	0.06	-3.06	ss715616094	39,427,301	82
Oil	1% 4.1	Oil1	1	43.0	32.3-50.6	3.8	0.05	1.48	ss715580879	9,763,709	11
	5% 3.4	Oil6	6	138.7	133.0-141.4	6.4	0.10	-1.96	ss715594713	47,468,779	44
		Oil11	11	134.0	124.2-139.0	3.7	0.06	1.49	ss715610403	32,861,170	10
		Oil17	17	47.0	24.0-59.0	4.0	0.04	1.41	ss715628177	7,488,569	15

a) estimated variance in trait captured by QTL, with range between 0 (none) and 1 (all).
b) Estimated effect in kg ha-1 (yield) or g kg-1 (all other traits) with respect to the Williams 82 allele.
c) Number of genes with parental amino acid differences located between terminal SNPs of LOD – 1.5 support interval.
E×W-50K = soybean population with parental lines Essex and Williams 82.

19, respectively (Table 3), and collectively explained 7% of the variation for linoleic acid. Numerous genes with mutations resulting in parental amino acid differences were located within the LOD – 1.5 support interval for Lin13.1 (23), **Lin13.2** (51), and Lin19 (11) (Table 3).

Five QTLs on chromosomes 9, 13, 17, and 19, designated as **Len9.1** ($R^2 = 0.04$), **Len9.2** ($R^2 = 0.06$), **Len13** ($R^2 = 0.06$), Len17 ($R^2 = 0.01$), and **Len19** ($R^2 = 0.02$), respectively, were detected for linolenic acid (Table 3). Together, these QTLs explain 19% of the variation for linolenic acid detected in this study. None of the QTLs for linolenic acid detected in this study were located on chromosomes 2, 14, or 18, and so could not be associated with the major linolenic acid genes (Glyma.02g227200, Glyma.14g194300, and Glyma.18g062000, respectively) listed in Gillman and Bilyeu (2012). However, genes with mutations leading to parental amino acid changes within the LOD – 1.5 intervals for **Len9.1** (24), **Len9.2** (58), **Len13** (63), Len17 (7) and **Len19** (6) were identified (Table 3).

Seven QTLs on chromosomes 6, 7, 9, and 13, designated as Prot6.1 (R2 = 0.04), Prot6.2 (R^2 = 0.07), Prot7 (R^2 = 0.04), Prot9.1 (R^2 = 0.07), **Prot9.2** (R^2 = 0.07), **Prot13.1** (R^2 = 0.02), and **Prot13.2** (R^2 = 0.06), were detected for seed protein in this study (Table 3). These QTLs collectively explain 36% of the protein variation observed in this study. Numerous genes with mutations leading to parental amino acid differences and a wide variety of associated biological processes are within the LOD – 1.5 support intervals for Prot6.1 (1), Prot6.2 (40), Prot7 (49), Prot9.1 (44), **Prot9.2** (24), **Prot13.1** (1), and **Prot13.2** (82) (Table 3).

The four QTLs for oil detected in this population are located on chromosomes 1, 6, 11 and 17. These QTLs are designated as Oil1 (R^2 = 0.05), **Oil6** (R^2 = 0.10), Oil11 (R^2 = 0.06), and Oil17 (R^2 = 0.04), respectively, and collectively explain 25% of the variation detected for oil in this study (Table 3). There were 11, 44, 10, and 15 genes with mutations resulting in parental amino acid differences located within the LOD – 1.5 support intervals for Oil1, **Oil6**, Oil11, and Oil17, respectively.

Discussion

Due to the importance of soybean for various uses throughout the world, it is necessary to study methods for improving valuable traits, such as yield, seed fatty acids, protein and oil. Toward this end, we sought to identify novel significant genomic regions for these traits and to explore for possible candidate genes, as well as to confirm QTLs from previous research.

None of the previously reported markers associated with QTLs for seed yield listed in SoyBase were within the LOD – 1.5 support interval of the **Yld14** QTL (www.SoyBase.org, "SoyBase browser", accessed 2/15/16). However, a major yield QTL (R^2 = 0.16) linked to Satt168 detected by Kabelka et al. (2004) was also located on chromosome 14, but outside of our support interval. Further efforts using a fine-mapping approach similar to Pham et al. (2015) may be beneficial for narrowing the list of candidate genes for **Yld14**, which could provide greater insight into the genetic causes of soybean yield.

While **Pal4** is located on chromosome 4, no other QTL for seed palmitic acid listed in SoyBase has previously been identified on chromosome 4 (www.SoyBase.org, "SoyBase browser", accessed 2/15/2016). Notably, a previous study using nearly the same parent lines for RIL development detected palmitic acid QTLs with overlapping marker support intervals based on the Wm82.a2.v1 sequence assembly for **Pal9**, **Pal17.2**, and Pal19 (Hyten 2002). While Essex was used in both studies, Hyten (2002) used Williams (Bernard and Lindahl 1972) as a parent rather than Williams 82: Williams is the recurrent parent of Williams 82 (Bernard and Cremeens 1988). **Pal9** is significant at the 1% threshold, shares a common parent (Essex), and shares a support interval with the flanking markers (Satt273 and Satt260) for the combined location palmitic acid QTL on chromosome 9 listed in Hyten (2002). Therefore, **Pal9** is a strong candidate for a confirmed QTL (http://www.soybase.org/resources/QTL.php, accessed 11/10/2015), and we propose the confirmed QTL symbol cqSeed palmitic-001. Further, **Pal17.2** is significant at the 1% threshold, shares a common parent (Essex), and shares a support interval with the flanking markers Satt458 and Satt154) for the combined location palmitic acid QTL on chromosome 17 from Hyten (2002). Thus, we propose the confirmed QTL symbol cqSeed palmitic-002 for **Pal17.2**.

A major gene for palmitic acid (Glyma.17g047000) is located near (< 1.5 Mbp) **Pal17.1**. However, no genetic differences were detected between the parents in this population at Glyma.17g047000. One of the genes in the support interval for **Pal4**, Glyma.04g131700, is involved in lipid storage (GO:0019915), and thus may be significant in the genetic variation associated with palmitic acid. In the support interval for **Pal17.1**, Glyma.17g034100 (GO:0006635) is involved in a fatty acid biological process, and may be associated with the palmitic acid genetic variation.

In addition to this study, several stearic acid QTLs have been identified on chromosome 14 (Li et al. 2011; Panthee et al. 2006b). The major stearic acid gene Glyma.14g121400 is located on chromosome 14 from 17,499,717-17,502,413 on the Wm82.a2.v1 reference genome. However, the terminal marker (ss715617588) in the genetic map used in this study (Song et al. 2016) for chromosome 14 is located from 12,437,074-12,437,194. However, **Ste14** is located at the far end of chromosome 14 in this study, but may be actually be located further downstream than would be possible to detect with this map.

Previous QTLs for oleic acid have been detected on chromosomes 9, 13, and 19 (www.SoyBase.org, "SoyBase browser", accessed 2/15/2016). Of these, Hyten (2002) detected a QTL on chromosome 19 linked to Satt182, which is located within the LOD – 1.5 support interval for **Ole19**. However, As the Hyten (2002) QTL was only significant at one location in that study, it will not be considered for confirmation. The QTLs detected for oleic acid (Ole9, Ole13, and **Ole19**) in this study could be useful to plant breeders seeking to stabilize high concentrations (> 800 g kg^{-1}) for that trait (Fallen et al. 2012; Lee et al. 2012;).

None of the genes associated with Ole13 and **Ole19** were associated with lipid or fatty acid biological processes. For Ole9, five genes within the LOD – 1.5 support interval emerge as leading candidates due to their involvement in biological processes involving fatty acids and/or lipids. These genes are Glyma.09g191400 (GO:0006629), Glyma.09g191700 (GO:0006631 and GO:0006635), Glyma.09g200500 (GO:0006636 and GO:0019216), Glyma.09g207900 (GO:0006636), and Glyma.09g209400 (GO:0006629 and GO:0016042).

While few efforts seeking to improve soybean fatty acids are focused on linoleic, it is possible that the QTLs listed in this study could be used to decrease linoleic acid. Such a decrease would allow other fatty acids, primarily oleic as the second most abundant fatty acid, to fill this void. By using

such an approach, breeders may be able to maintain higher levels of stability in cultivars with high levels (> 800 g kg^{-1}) of oleic acid (Fallen et al. 2012; Lee et al. 2012). Hyten (2002) detected a QTL for linoleic acid on chromosome 13 linked to Satt144, which is located between the terminal SNPs for the **Lin13.2** LOD – 1.5 support interval. Because **Lin13.2** is significant at the 1% threshold, shares a common parent (Essex), and shares a support interval with the flanking markers (Satt144 and Satt522) for the combined location palmitic acid QTL on chromosome 13 listed in Hyten (2002), we propose the confirmed QTL symbol cqSeed linoleic-001.

None of the genes with mutations leading to parental amino acid differences associated with **Lin19** are involved in fatty acid or lipid biological processes. For Lin13.1, Glyma. 13g213500 (GO:0006636 and GO:0019216) and Glyma. 13g215400 (GO:0006636) are involved in fatty acid and/or lipid biological processes, and may be responsible for some of the variation of linoleic acid detected in this study. A potential candidate gene for **Lin13.2** is Glyma.13g256100 (GO:0019915), as it is involved in lipid storage.

While Len17 is located on chromosome 17, no other QTLs for seed linolenic acid listed in SoyBase have previously been located on chromosome 17 (www.SoyBase. org, "SoyBase browser", accessed 2/15/2016). However, a previous QTL has been detected on chromosome 13 between Satt146 and Satt269, which overlaps with the LOD – 1.5 support threshold for **Len13** (Hyten, 2002). As **Len13** meets the criteria for confirmed status (http://www.soybase.org/ resources/QTL.php, accessed 11/10/2015), we propose cqSeed linolenic-001 as the confirmed QTL symbol.

Also, a previous study has detected a QTL for linolenic acid on chromosome 19 linked to Satt238 (Kim et al. 2010). The terminal SNP within the LOD – 1.5 support threshold downstream from **cq,** ss715633497, is located at 12.2 cM on chromosome nineteen, while the support threshold continues until 22 cM (Table 3). Using all of the SNPs within the support interval for **Len19** +/- 3 cM (4.9-25 cM), a stepwise approach in SAS PROC REG (SAS Institute Inc., Cary, NC, USA, SAS 9.4, 2002-2012) was run to determine the best polynomial regression model between linkage map position (x-axis) and Wm82.a2.v1 position (y-axis). Using the selected model, $Y = b_0 + b_2x^2 + b_3x^3 + b_6x^6$ ($R^2 > 0.99$), the LOD – 1.5 support interval was estimated to range from 2.03-3.80 Mbp, which contains both **Len19** and Satt238. While this study used different parents from Kim et al. (2010), the significance ($P < 0.01$) and location of **Len19** make it a strong candidate as a positional QTL confirmation (Smallwood et al., 2014). The QTLs detected for linolenic acid (**Len9.1**, **Len9.2**, **Len13**, Len17, and **Len19**) in this study could be useful to plant breeders seeking to meet the dual goal of high oleic (> 800 g kg^{-1}), low linolenic (< 30 g kg^{-1}) soybeans.

Candidate genes with mutations resulting in amino acid changes near **Len9.1** with involvement in fatty acid or lipid biological processes include Glyma.09g041200 (GO:0006629 and GO:0016042), Glyma.09g043700 (GO:0000038, GO:0006633, and GO:0008610), and Glyma.09g043800 (GO:0019915). For **Len9.2**, the gene candidates with gene ontologies in fatty acid or lipid biological processes were Glyma.09g200500 (GO:0006636 and GO:0019216), Glyma.09g207900 (GO: 0006636), and Glyma.09g209400 (GO:0006629 and GO: 0016042). For **Len13**, Glyma.13g033300 (GO:0006629), Glyma.13g041000 (GO:0006635), Glyma.13g043000 (GO:0015245 and GO:0015908), Glyma.13g043500 (GO:0015245 and GO: 0015908), and Glyma.13g043600 (GO:0015245 and GO: 0015908) are possible gene candidates as they are each involved in fatty acid and/or lipid biological processes.

Many protein QTLs have been previously reported on each soybean chromosome (www.SoyBase.org, "SoyBase browser", accessed 2/15/2016). Of particular note are those detected by Hyten et al. (2004a), which used nearly the same parents for RIL development as this study. While Essex was used in both studies, Hyten et al. (2004a) had Williams (Bernard and Lindahl 1972) rather than Williams 82. The Hyten et al. (2004a) protein QTLs on chromosomes 7 (Satt463), 9 (Satt539), and 13 (Satt144) are located within the LOD – 1.5 support interval ranges for Prot7, Prot 9.1, and **Prot13.2**, respectively. Since **Prot13.2** is highly significant ($P < 0.01$) and closely associated with the QTL on chromosome 13 from Hyten et al. (2004a), it is an excellent candidate to be a confirmed QTL (http://www.soybase.org/resources/QTL.php, accessed 7/26/2015). Thus, the confirmed QTL symbol of cqSeed protein-017 is proposed for **Prot13.2**. While further research is needed to fully understand what genes are responsible for these QTLs, it is still possible for breeders to use this information in making protein improvements.

Numerous studies have previously reported seed oil QTLs on chromosomes 1, 6, 11, and 17 (www.SoyBase.org, "SoyBase browser", accessed 2/15/2016). Notably, A QTL linked to Satt154 (Hyten et al. 2004a) for seed oil on chromosome 17 is located within the LOD – 1.5 support interval for Oil17 (Table 3). However, as Oil17 does not meet the 1% significance threshold, QTL confirmation will not be pursued. None of the genes with mutations leading to parental amino acid differences associated with the oil QTLs are involved with lipid biological processes. As with protein, further research is needed to fully understand which genes are responsible for the effects seen by these QTLs

As many of the QTLs detected in this study for different traits share LOD – 1.5 support intervals (Table 3), it is worth considering the possibility of pleiotropy. On chromosome 6, the support intervals for Prot6.2 and **Oil6** overlap with each other (Table 3). l Given the historical evidence that oil and protein in soybean seed are negatively correlated (Yaklich et al. 2002) as well as the negative correlation between these traits detected in this study (Table 2), it seems reasonable to think that whatever causative gene or genes in this region are affecting seed accumulation for protein and oil in opposite directions. Further evidence is provided with the opposing direction of effects in this QTL region for protein and oil (Table 3).

On chromosome 9, Ole9 and **Len9.2** had overlapping

support intervals with opposite effects (Table 3) and a significant ($P < 0.05$) negative correlation (Table 2). Similar pleiotropic effects with opposing effects for oleic acid and linolenic acid occur on chromosome 13 with Ole13 and **Len13**, and chromosome 19 with **Ole19** and **Len19** (Table 3). For the pleiotropic effects on chromosome 19 between **Ole19** and **Len19**, linoleic acid (Lin19) was also involved (Table 3). Other pleiotropic effects were found on chromosome 9 for palmitic acid and protein between **Pal9** and **Prot9.2**, chromosome 13 for linoleic acid and protein between **Lin13.2** and both **Prot13.1** and **Prot13.2**, and on chromosome 17 for palmitic acid and linolenic acid between **Pal17.1** and Len17 (Table 3). These findings could be very useful for breeders seeking to adjust multiple traits simultaneously, such as achieving the dual goal of high oleic (> 800 g kg^{-1}), low linolenic (< 30 g kg^{-1}) soybeans.

Conclusions

Due to the importance of yield and seed fatty acids, protein, and oil content in soybean production, it is critical to develop an improved understanding of significant genomic regions governing these traits. In this study, 30 QTLs were detected for yield (1), palmitic acid (6), stearic acid (1), oleic acid (3), linoleic acid (3), linolenic acid (5), protein (7), and oil (4) (Table 3). Of these, **Pal9**, **Pal17.2, Lin13.2**, **Len13**, and **Prot13.2** are excellent candidates for confirmed QTLs (http://www.soybase.org/resources/QTL.php, accessed 11/10/2015), while **Len19** is a strong candidate for a positional QTL confirmation (Smallwood et al. 2014).

Since none of the major fatty acid genes listed in Gillman and Bilyeu (2012) had mutations with amino acid differences in this population, the QTLs detected for fatty acids are likely the result of small effect modifier genes (Hyten et al. 2004b) or expression differences in causative genes. Many of the genes with mutations resulting in amino acid changes in close proximity to the fatty acid QTLs are involved in biological processes for fatty acids and/or lipids. Such genes may be useful in breeding for fatty acid improvement (Fallen et al. 2012; Gillman et al. 2014; Lee et al. 2012). Similarly, genes with mutations leading to amino acid changes in genomic regions near yield, protein, and oil were plentiful, and may contribute to some of the variation observed in these traits (Table 1).

All of the traits except yield and stearic acid were involved in probable pleiotropic relationships with other traits in this study. Of note is the pleiotropic effect between protein and oil on chromosome 6 (Table 3). Given the well-established negative relationship between protein and oil (Yaklich et al. 2002), it is likely that whatever causative gene or genes in this region are affecting seed accumulation of both protein and oil in a negatively correlated manner. While the E1 maturity gene does segregate in this population, it is not likely to be the causative gene for this pleiotropic effect between protein and oil, as it is outside the LOD – 1.5 support intervals for both QTLs. Also, the pleiotropic effects between oleic acid and linolenic acid on chromosomes 9, 13, and 19 are worthy of further research, as these traits are the subject of much interest in soybean improvement.

Overall, the findings from this research contribute new information to the genetic understanding of soybean yield and seed fatty acids, protein and oil content. This understanding will be useful in making trait improvements. Further research seeking to narrow the list of candidate genes for these QTLs would be beneficial.

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Jim Orf and Art Killam of the University of Minnesota for performing and assisting with Near Infrared Reflectance Spectroscopy analyses. We would also like to thank Tiffany Langewisch and Kristin Bilyeu of the USDA-ARS, Columbia, MO, for identifying linked markers to the E1, E3, and Dt1 loci. We are grateful for assistance provided by members of the University of Tennessee soybean breeding team and the University of Tennessee Institute of Agriculture: without your help this research would not have been possible. Generous funding for this research was provided by the United Soybean Board, the Tennessee Soybean Promotion Board, and the University of Tennessee Institute of Agriculture.

References

Bernard RL, Cremeens CR. 1988. Registration of 'Williams 82' soybean. Crop Sci. 28: 1027-1028

Bernard RL, Lindahl DA. 1972. Registration of 'Williams' soybean. Crop Sci. 12: 716

Bilyeu K, Gillman JD, LeRoy AR. 2011. Novel FAD3 mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. Crop Sci. 51: 259-264

Boersma JG, Gillman JD, Bilyeu K, Ablett GR, Grainger C, Rajcan I. 2012. New mutations in a delta-9-stearoyl-acyl carrier protein desaturase gene associated with enhanced stearic acid levels in soybean seed. Crop Sci. 52: 1736-1742

Bolon Y, Haun WJ, Xu WW, Grant D, Stacey MG, Nelson RT, Gerhardt DJ, Jeddeloh JA, Stacey G, Muehlbauer GJ, Orf JH, Naeve SL, Stupar RM, Vance CP. 2011. Phenotypic and genomic analyses of fast neutron mutant population resource in soybean. Plant Physiol. 156: 240-253

Brim CA. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6: 220

Broman KW, Wu H, Sen Ś, Churchill GA. 2003. R/qtl: QTL

mapping in experimental crosses. Bioinformatics 19: 889-890
Browning BL, Browning SR. 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84: 210-223
Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. Am. J. Hum. Genet. 81: 1084-1097
Cardinal AJ, Whetten R, Wang S, Auclair J, Hyten D, Cregan P, Bachlava E, Gillman J, Ramirez M, Dewey R, Upchurch G, Miranda L, Burton JW. 2014. Mapping the low palmitate fap1 mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations. Theor. Appl. Genet. 127: 97-111
Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971
Diers BW, Cary TR, Thomas DJ, Colgrove A, Niblack T. 2010. Registration of 'LD00-2817P' germplasm line with resistance to soybean cyst nematode from PI 437654. J. Plant Regist. 4: 141-144
Diers BW, Cary TR, Thomas DJ, Nickell CD. 2006. Registration of 'LD00-3309' soybean. Crop Sci. 46: 1384
Dupuis J, Siegmun D. 1999. Statistical methods for mapping quantitative trait loci from a dense set of markers. Genetics 151: 373-386
Fallen BD, Rainey K, Sams CE, Kopsell DA, Pantalone VR. 2012. Evaluation of agronomic and seed characteristics in elevated oleic acid soybean lines in the south-eastern US. J. Am. Oil Chem. Soc. 89: 1333-1343
Federal Register. 2015. Final determination regarding partially hydrogenated oils. https://www.federalregister.gov/articles/2015/06/17/ 2015-14883/final-determination-regarding-partially-hydrogenated-oils (accessed 24 July 2015)
Fehr WR, Caviness CE. 1977. Stages of soybean development. Special Report, Agriculture and Home Economics Experiment Station, Iowa State University, 1977, issue 80, p 11
Gillman JD, Bilyeu KD. 2012. Genes and alleles for quality traits on the soybean genetic/physical map. In: R.F. Wilson (ed) Designing soybean for 21st century markets. AOCS Press, Urbana, IL, pp 67-96
Gillman JD, Stacy MG, Cui Y, Berg HR, Stacey G. 2014. Deletions of the SACPD-C locus elevate seed stearic acid but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol. 14: 143
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. Nucl. Acids Res. 40: D1178-D1186
Grant D, Nelson RT, Cannon SB, Shoemaker RC. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucl. Acids Res. 38: D843-D846
Harrell Jr FE. 2015. Package 'Hmisc'. http://biostat.mc.vanderbilt.edu/Hmisc
Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM. 2011. Plant Physiol. 155: 645-655
Hyten DL. 2002. QTL mapping and identification of GxE interactions of agronomic and seed quality traits in soybean. Thesis, University of Tennessee
Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, Schmidt ME. 2004a. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552-561
Hyten DL, Pantalone VR, Saxton AM, Schmidt ME, Sams CE. 2004b. MoleculaR mapping and identification of soybean fatty acid modifier quantitative trait loci. J. Am. Oil Chem. Soc. 81: 1115-1118
Kabelka EA, Diers BW, Fehr WR, LeRoy AR, Baianu IC, You T, Neece DJ, Nelson RL. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. 44: 784-791
Kim H, Kim Y, Kim S, Son B, Choi Y, Kang J, Park Y, Cho Y, Cho I. 2010. Analysis of quantitiative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. J. Life Sci. 20: 1186-1192
Kinney AJ. 1996. Development of genetically engineered soybean oils for food application. J. Food Lipids 3: 273-292
Kinney AJ, Clemente TE. 2005. Modifying soybean oil for enhanced performance in biodiesel blends. Fuel Pro. Technol. 86: 1137-1147
Kris-Etherton PM, Yu S. 1997. Individual fatty acid effects on plasma lipids and lipoproteins: Human studies. Am. J. Clin. Nutr. 65: S1628-S1644
Lee JD, Bilyeu KD, Pantalone VR, Gillen AM, So YS, Shannon JG. 2012. Environmental stability of oleic acid concentration in seed oil for soybean lines with FAD2-1A and FAD2-1B mutant genes. Crop Sci. 52: 1290-1297
Li H, Zhao T, Wang Y, Yu D, Chen S, Zhou R, Gai J. 2011. Genetic structure composed of additive QTL, epistatic QTL pairs and collective unmapped minor QTL conferring oil content and fatty acid components of soybeans. Euphytica 182: 117-132
Martínez O, Curnow RN. 1992. Estimating the locations and sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85: 480-488
Nyquist WE. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci.10: 235-322
Pantalone VR, Allen FL, Landau-Ellis D. 2003. Registration of '5601T' soybean. Crop Sci. 43: 1123-1124
Pantalone VR, Allen FL, Landau-Ellis D. 2004. Registration of '5002T' soybean. Crop Sci. 44: 1483-1484
Pantalone VR, Wilson RF, Novitzky WP, Burton JW. 2002. Genetic regulation of elevated stearic acid concentration in soybean oil. J. Am. Oil Chem. Soc. 79: 543-553
Panthee DR, Pantalone VR, Sams CE, Saxton AM, West DR, Orf JH,Killam AS. 2006a. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor. Appl. Genet. 112: 546-553
Panthee DR, Pantalone VR, Saxton AM. 2006b. Modifier QTL for fatty acid composition in soybean oil. Euphytica 152: 67-73

Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE. 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci. 45: 2015-2022

Pham AT, Harris DK, Buck J, Hoskins A, Serrano J, Abdel-Haleem H, Cregan P, Song QJ, Boerma HR, Li Z. 2015. Fine mapping and characterization of candidate genes that control resistance to *Cercospora sojina* K. Hara in two soybean germplasm accessions. PLoS ONE 10: e0126753

Pham AT, Lee JD, Shannon JG, Bilyeu KD. 2010. Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biol. 10: 195

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R- project.org/

SAS Institute Inc. 2002-2012. Cary, NC, USA. SAS 9.4

Smallwood CJ, Nyinyi CN, Kopsell DA, Sams CE, West DR, Chen P,Kantartzi SK, Cregan PB, Hyten DL, Pantalone VR. 2014. Detection and confirmation of quantitative trait loci for soybean seed isoflavones. Crop Sci. 54: 1-12

Smith TJ, Camper HM. 1973. Registration of Essex Soybean (Reg. No. 97). Crop Sci. 13: 495

Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL,Cregan PB. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8: e54985

Song Q, Jenkins J, Jia G, Hyten DL, Pantalone V, Jackson SA, Schmutz J, Cregan PB. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. BMC Genomics 17: 33

SoyBase and the Soybean Breeder's Toolbox. 2007. QTL nomenclature. http://www.soybase.org/resources/QTL.php. accessed 26 July 2015

Spencer MM, Landau-Ellis D, Meyer EJ, Pantalone VR. 2004. Molecular markers associated with linolenic acid content in soybean. J. Am. Oil Chem. Soc. 81: 559562

USDA, ARS, National Genetic Resources Program. Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, Maryland. URL: http://www.ars-grin.gov.4/cgi-bin/npgs/html/index.pl?language=en (24 July 2015)

Wiggins BT. 2012. Heritability and genetic gain of seed protein, oil, and yield among RIL of soybean. M.S. thesis. Univ. of Tennessee, Knoxville, TN, USA

Wilson RF. 2004. Seed composition. In HR Boerma HR, JE Specht. eds. Soybeans: Improvement, production, and uses. 3rd ed. ASA, CSSA, and SSSA, Madison, WI pp 621-678

Wimmer V, Albrecht T, Auinger HJ, Schön CC. 2012. Synbreed: A framework for the analysis of genomic prediction using R. Bioinformatics. 28: 2086-2087

Yaklich RW, Vinyard B, Camp M, Douglass S. 2002. Analysis of seed protein and oil from soybean northern and southern region uniform tests. Crop Sci. 42: 1504-1515

Zeng Z, Kao C, Basten C. 1999. Estimating the genetic architecture of quantitative traits. Genet. Res. 74: 279-289