Mol Breeding (2021) 41:27
https://doi.org/10.1007/s11032-021-01216-1

Identification of SNP markers associated with soybean fatty acids contents by genome-wide association analyses

Mikyung Sung · Kyujung Van · Sungwoo Lee · Randall Nelson · Jonathan LaMantia · Earl Taliercio · Leah K. McHale · M. A. Rouf Mian

Received: 5 May 2020 / Accepted: 15 February 2021

Abstract Composition of fatty acids (FAs) in soybean seed is important for the quality and uses of soybean oil. Using gas chromatography, we have measured soybean FAs profiles of 621 soybean accessions (maturity groups I through IV) grown in five different environments; Columbus, OH (2015), Wooster, OH (2014 and 2015), Plymouth, NC (2015), and Urbana, IL (2015). Using publicly available SoySNP50K genotypic data and the FA profiles from this study, a genome-wide association analysis was completed with a compressed mixed linear model to identify 43 genomic regions significantly associated with a fatty acid at a genome wide significance threshold of 5%. Among these regions, one and three novel genomic regions associated with palmitic acid and stearic acid, respectively, were identified across all five environments. Additionally, nine novel environment-specific FA-related genomic regions were discovered providing new insights into the genetics of soybean FAs. Previously reported FA-related loci, such as *FATB1a*, *SACPD-C*, and *KASIII*,

Mikyung Sung and Kyujung Van contributed equally to this work.

M. Sung · S. Lee
Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC 27695, USA

K. Van · L. K. McHale (✉)
Department of Horticulture and Crop Science, The Ohio State University, Columbus, OH 43210, USA
e-mail: mchale.21@osu.edu

S. Lee
Department of Crop Science, Chungnam National University, Daejeon 34134, South Korea

R. Nelson
Department of Crop Sciences, University of Illinois and USDA-ARS (retired), Urbana, IL 61801, USA

J. LaMantia
Corn, Soybean, Wheat Quality Research Unit, USDA-ARS, Wooster, OH 44691, USA

J. LaMantia
Germplasm Analytics, TerViva Bioenergy Inc., Fort Pierce, FL 34951, USA

E. Taliercio · M. A. R. Mian
Soybean and Nitrogen Fixation Unit, USDA-ARS, Raleigh, NC 27607, USA

M. A. R. Mian
e-mail: rouf.mian@usda.gov

L. K. McHale
Center for Soybean Research and Center of Applied Plant Sciences, The Ohio State University, Columbus, OH 43210, USA

Published online: 20 March 2021

were also confirmed in this study. Our results will be useful for future functional studies and marker-assisted breeding for soybean FAs.

Keywords Fatty acids · Genome-wide association study · Quantitative trait loci · Soybean · SoySNP50K

Introduction

Soybean [*Glycine max* (L.) Merr.] is a major oilseed crop, accounting for 55% of US vegetable oil consumption in 2018 (SoyStats 2019). Soybean oil normally consists of 16% saturated fatty acids (FAs) [12% palmitic acid (16:0), 4% stearic acid (18:0)], 24% monounsaturated FAs [oleic acid (18:1)], and 60% polyunsaturated FAs [53% linoleic acid (18:2) and 7% linolenic acid (18:3)] (Nwokolo 1996). Specific FA compositions are desirable depending on the end-use of the soybean oil. For example, saturated and monounsaturated FAs are stable for cooking oil, but most polyunsaturated FA molecules react with oxygen to produce off flavors (Clemente and Cahoon 2009). Human consumption of palmitic acid tends to raise low-density lipoprotein (LDL) cholesterol levels, while stearic acid has a neutral effect on LDL cholesterol level in plasma (Liu et al. 2002). In contrast, consumption of unsaturated FAs has the beneficial property of lowering LDL-cholesterol (Mensink and Katan 1992). Linoleic and linolenic acids are recognized as valuable nutritional factors that potentially reduce cardiovascular disease risk and influence cognitive functions and behaviors (Lunn and Theobald 2006). These two essential polyunsaturated FAs are not produced by the human body and need to be acquired through diet. Soybean oil with high amounts of linoleic and linolenic acids is also in demand for industrial drying-oil which is one of the key components in oil-based paints and coatings (Cecil et al. 1988). Although each unique FA profile determines the functional and nutritional qualities of the oil in both food and non-food products, controlling the concentration of each FA in order to attain the varied end use products is made difficult by the complex relationships among individual FAs.

In previous studies, significant negative correlations were observed between oleic acid and palmitic acid, oleic acid and linoleic acid, as well as oleic acid and linolenic acid (Ohlrogge and Browse 1995; Alt et al. 2005; Bachlava et al. 2008). It was also found that seed yield had a negative correlation with oleic acid content, and positive correlations with palmitic acid and polyunsaturated FA contents. Moreover, water stress and high temperature alter seed oil composition. High water stress and high temperature increase oleic acid, but decrease linoleic and linolenic acid contents (Wolf et al. 1982; Bellaloui et al. 2013). Palmitic and stearic acid contents were generally stable under normal soil water potential between −90 and –100 kPa and high temperature changes (36 °C/28 °C, day/night) (Wolf et al. 1982), while palmitic acid increased and stearic acid decreased under severe water stress (i.e., soil water potential between −150 and –200 kPa) or severe high temperature (40 °C/33 °C, day/night) (Bellaloui et al. 2013). Hence, to produce various high-quality soybean oil via breeding soybean cultivars, it is important to investigate FA compositions and the interactions of FAs and soybean growing environments.

FAs in soybean are quantitatively inherited and controlled by both major and minor genes. Many genes or quantitative trait loci (QTL) associated with the five dominant FAs have been identified in numerous studies (Diers and Shoemaker 1992; Li et al. 2002; Hyten et al. 2004b; Panthee et al. 2006; Reinprecht et al. 2006; Pham et al. 2010; Wang et al. 2012). As curated on SoyBase (http://www.soybase.org, March 25, 2020), QTL conditioning soybean FA contents were positioned on nineteen of the twenty chromosomes. Traditional QTL mapping using bi-parental mapping populations and linkage maps were used to identify QTL in most of these studies. Such studies detected genomic loci with relatively large genetic effects on FA contents, in which two parental genotypes of mapping populations had significant differences in FA profiles (Reinprecht et al. 2006; Li et al. 2011; Wang et al. 2012; Xie et al. 2012). In such types of QTL analyses, however, the number of alleles assayed are restricted to those present in the parents and the relatively small population size limits recombination events, often resulting in low mapping resolution.

As an alternative to QTL mapping with bi-parental populations, genome-wide association studies (GWAS) have been recently applied to locate QTL for various traits, including FAs in naturally occurring populations (Li et al. 2015; Fang et al. 2017; Leamy et al. 2017; Zhang et al. 2018; Zhao et al. 2019). Soybean GWAS is more feasible since the SoySNP50K Beadchip was developed and used to genotype nearly 20,000 *G. max* and *G. soja* accessions from the USDA Soybean Germplasm

Collection (GRIN, http://www.ars-grin.gov/cgi-bin/npgs/html/crop.pl?51) (Song et al. 2013, 2015). This SNP dataset has already been used in several GWAS for analyzing resistance against various diseases, such as *Phytophthora sojae* (Schneider et al. 2016; Rolling et al. 2020) and cyst nematode (Vuong et al. 2015), as well as soybean seed compositions (Vaughn et al. 2014; Lee et al. 2019). Using a total of 621 soybean accessions in maturity groups (MGs) I to IV (Lee et al. 2019) and the publicly available SoySNP50K data, we conducted a GWAS to identify QTL controlling FA contents in soybean seeds grown in five environments.

Materials and methods

Germplasms and field trials

Six hundred twenty-one diverse soybean accessions obtained from the USDA Soybean Germplasm Collection were evaluated in three incomplete blocks of 200 entries and a block of 33, each with a set of four checks [Summit (mid-MG II) (McHale et al. 2012), Wyandot (MG II) (Lee et al. 2017; https://mchalelab.cfaes.ohio-state.edu/sites/mchale/files/imce/Wyandot_release_document.pdf), HR09-397 (MG III) (a high protein breeding line from USDA-ARS, Wooster, OH), and Prohio (MG IV) (Mian et al. 2008)] entered at the beginning of the block. The 621 soybean accessions were grown in five environments in Columbus, OH in 2015 (OHC15), Wooster, OH in both 2014 (OHW14) and 2015 (OHW15), Urbana, IL in 2015 (IL15), and Plymouth, NC in 2015 (NC15). According to the information in the Germplasm Resources Information Network (GRIN), these accessions originated from China (358), Europe (8), India (2), Japan (66), Korean Peninsula (106), North America (59), Russia (21), and Taiwan (1). Over 70% of these accessions were from Eastern Asia, the center of origin of soybean, and the detailed information has been reported previously (Lee et al. 2019). In GRIN, the accessions were classified into MGs I (7), II (267), III (187), and IV (160), which are the major maturity groups of soybean cultivated in the north central region of the USA. Only accessions with yellow seed coat color, and less than 4.0 scores (on a 1 to 5 scale) for lodging, pod shattering, and seed mottling were used in this study (with few exceptions). Test plots were managed according to local practices and all the fields were kept mostly free of diseases and weeds.

Determination of contents of five major FAs in soybean seeds

For extraction of fatty acids, approximately 20 seeds were randomly sampled from harvested seeds of each genotype and the pooled seeds were ground using a Laboratory Mill 3610 grinder (Perten Instruments, Inc., Hägersten, Sweden) attached to Mill Feeder 3170 for improving homogeneity and ease of operation. Approximately, 0.3 g of the ground seed powder from each sample was used for overnight extraction of total lipids with 1 ml of chloroform-hexane-methanol (8:5:2, v/v/v, Fisher Scientific, Fair Lawn, NJ, USA). For derivatization, 75 μl of methylating reagent was added to 150 μl of extract, and then 1 ml of hexane was added to dilute. FAs measurements followed methods of Byfield and Upchurch (2007), with standard FA mixtures (Animal and Vegetable Oil Reference Mixture 6, AOACS, Matreya, LLC, State College, PA, USA) used as a reference. The content of each FA was determined by the effective area calculated by multiplying peak height and width at the half of peak height. Each FA content was normalized by the total content and presented as percentage of total seed oil.

Statistical analyses of phenotypic data

We calculated the best linear unbiased predictor (BLUP) values of each FA content by using PROC MIXED (SAS Institute, Cary NC, USA) to normalize phenotypic data collected from five environments. The statistical model for the calculation of BLUP values was as follows:

$$Y_{ijklm} = \mu + E_i + B(E)_{ij} + C_k + G(C)_{kl} + G_l \times E_i + \varepsilon_{ijkl}$$

where μ is overall mean, E_i is effect of ith environment, $B(E)_{ij}$ is effect of jth block in ith environment, C_k is effect of kth class of entry (k = 1, 2, 3, 4, and 5 for four checks, Summit, Wyandot, HR09-397, Prohio, and germplasm accessions, respectively), $G(C)_{kl}$ is effect of lth entry within the kth class and was equivalent to σ^2_G, G_l x E_i is effect of interaction between lth genotype and ith environment, and ε_{ijkl} is experimental error. Class of entry was treated as a fixed effect and all other terms were treated as random effects.

Variance components were estimated by the mixed model above using the restricted maximum likelihood (REML) method (Patterson and Thompson 1971). The

broad-sense heritability (H^2) on an entry-mean basis was calculated based on Lee et al. (2019) with the variance components as follows: $H^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_{GE}/e + \sigma^2/r)$, where e is the number of environments per genotype and r is the total number of replications per genotype (i.e., 5).

Population structure and GWAS analyses

From a total of 52,041 SNPs obtained from SoyBase (https://www.soybase.org/snps/), 18,027 SNPs were excluded: where markers were monomorphic, showed less than 0.05 minor allele frequency (MAF), or were missing greater than 10% of genotypic data. A total of 34,014 SNPs were used for this GWAS.

GWAS analysis was conducted using Genomic Association and Prediction Integrated Tool (GAPIT, http://zzlab.net/GAPIT), a package for identifying the genetic potential of individuals (Lipka et al. 2012). As an analytical method, we used compressed mixed linear model (CMLM), which functions between the generalized linear model and mixed linear model by decreasing the effective size of the samples by grouping individual samples (Zhang et al. 2010). Kinship matrix was constructed by the genomic relationship method (VanRaden 2008), which uses the default clustering algorithm (average) and group kinship type (Mmean) (Supplemental Fig. S1). Two subgroups were determined by conducting population structure and principal component analysis (by origin of the accessions) in our previous work (Lee et al. 2019). A modified Bonferroni adjustment for multiple testing was used to determine the appropriate significance threshold level of $\alpha = 5\%$ and $\alpha = 25\%$; i.e., $-\log_{10}(P) > 5.44$ and $-\log_{10}(P) > 4.74$, respectively, as described in Lee et al. (2019). Significant SNP markers were compared with previously reported FA-associated QTL.

Linkage disequilibrium and identification of candidate genes

We used Haploview 4.2 (Barrett et al. 2005) to obtain the linkage disequilibrium (LD) matrix from the list of SNPs and LD blocks as described in Lee et al. (2019). A maximum distance of 1000 kb and the minimum minor allele frequency of 0.05 was applied in the four-gamete method (Wang et al. 2002) with Hardy-Weinberg cutoff of p value < 0.01 to determine haplotype blocks. Adjacent blocks were combined if each block was separated by less than 10 kb (Schneider et al. 2016). In LD blocks where significant SNPs were identified by genome-wide association analyses, genes located in the LD block encompassing the SNPs were considered to be positional candidate genes for the target trait. *G. max* genome assembly version Glyma.Wm82.a2.v1 (Gmax2.0) was used to identify positional candidate genes. Detailed information of each candidate gene was retrieved from SoyBase (http://soybase.org).

Results and discussion

Phenotypic analysis

The BLUP values of each trait across all environments (ALL) showed continuous variation, ranging from 9.1 to 13.7% for palmitic acid, 3.0 to 6.2% for stearic acid, 17.0 to 33.2% for oleic acid, 46.0 to 58.3% for linoleic acid, and 6.4 to 11.7% for linolenic acid (Fig. 1a and Supplemental Table S1). The broad-sense heritability of each FA was high, 0.94, 0.91, 0.87, 0.86, and 0.94 for palmitic, stearic, oleic, linoleic, and linolenic acids, respectively (Fig. 1a). Palmitic, stearic, oleic, and linolenic acids were skewed to lower than the median with the skewness values of 0.24, 1.58, 0.63, and 0.75, respectively, whereas the distribution of linoleic acid had a negative skewness of −0.68 (data not shown). Moderate positive correlation between linoleic acid and linolenic acid ($r = 0.42$, $p < 0.001$) and strong negative correlations between oleic acid and linoleic acid ($r = -0.93$, $p < 0.001$), and oleic acid and linolenic acid ($r = -0.65$, $p < 0.001$) were observed (Fig. 1). Palmitic acid had a weak negative correlation with stearic acid ($r = -0.18$, $p < 0.001$), a moderate negative correlation with oleic acid ($r = -0.32$, $p < 0.001$), and a moderate positive correlation with linolenic acid ($r = 0.27$, $p < 0.001$). Stearic acid had a weak negative correlation with both linoleic ($r = -0.13$, $p < 0.001$) and linolenic acids ($r = -0.23$, $p < 0.001$) (Fig. 1a). These strong correlations (r value of $\geq |\pm 0.3|$) were in agreement with the previous report (Li et al. 2015). Correlation network analysis explicitly exhibited that stearic acid was not clustered with any other FAs because minimal correlations (r value of $< |\pm 0.3|$) were observed (Fig. 1b). In contrast, oleic, linoleic, and linolenic acids were tightly clustered together, and this proximity illustrates high-magnitude correlations among the three unsaturated FAs, suggesting that genes associated with the three unsaturated FAs are not independent

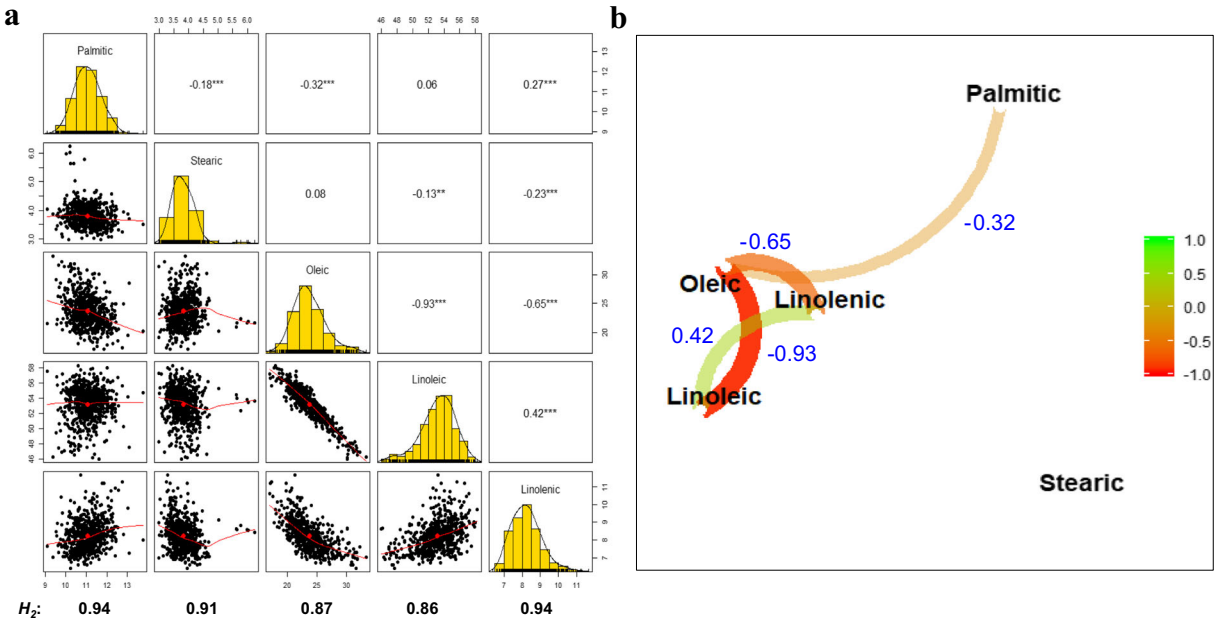

Fig. 1 Correlations for pairs of each trait in soybean seed. **a** The pairwise comparison matrix contents based on the Pearson's correlation coefficient. The numeric values on the *X*-axis and *Y*-axis showed the contents (%) of given variables. Scatter plots of the correlation and values of the correlation and their significance levels were placed on the bottom and top of the chart, respectively. The significance levels were marked with a single (*), a double (**), and a triple (***) asterisks, if $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively. The distributions of frequency for each fatty acid were shown on the diagonal of the panel and the general density curves were adjusted over the histograms. The broad-sense heritability (H_2) for each trait was added at the bottom of matrix. **b** Correlation network plot for each fatty acid. The strength of the correlation was illustrated by both width and transparency of path. A solid colored, wide path denotes the strong relationship and values of correlation coefficient were indicated in blue. The value of correlations set as low as ≥ |± 0.3| (Toubiana et al. 2012)

of one another. For example, genes highly associated with oleic acid are likely to control linoleic acid as well, due to the strong negative correlation between them (Martin and Rinne 1986; Kanobe et al. 2015).

Genome-wide association analysis and identification of candidate genes for FA composition

Under ALL, 14, 146, 2, and 4 SNPs were significantly (5%, $-\log_{10}(P) > 5.44$) associated with palmitic, stearic, oleic, and linoleic acids, respectively (data not shown). Since many of these SNPs follow the same inheritance pattern, only the most significant SNP within an LD block is reported if the LD block contained multiple significant SNPs. This consolidated the reported data into the most significant SNP within 6, 30, 1, and 1 LD blocks for the palmitic, stearic, oleic, and linoleic acids, respectively (Supplemental Table S2). An additional 2, 4, 1, and 1 LD blocks for the palmitic, stearic, oleic, and linoleic acids, respectively, encompass SNPs in which the most significant association is at the suggestive level (25%, $-\log_{10}(P) > 4.74$) (Supplemental Table S2). No significant SNP associated with linolenic acid was identified at either genome-wide or suggestive significant thresholds under ALL.

We also carried out GWAS within each individual environment: OHW14, OHW15, OHC15, IL15, and NC15. A total of 66 LD blocks on 14 different chromosomes were significantly identified as FA-associated QTL (11 for palmitic, 42 for stearic, 5 for oleic, 4 for linoleic, and 4 for linolenic acids) at a significant or suggestive threshold (Supplemental Table S3 and Supplemental Fig. S2-S6).

Most significant LD blocks for palmitic acid located on Chr 5

All identified SNPs under ALL ($\alpha = 5\%$) for palmitic acid were clustered within six LD blocks in a genomic region on Chr 5 (Fig. 2a and Supplemental Table S2). Five out of these six LD blocks were highly significant

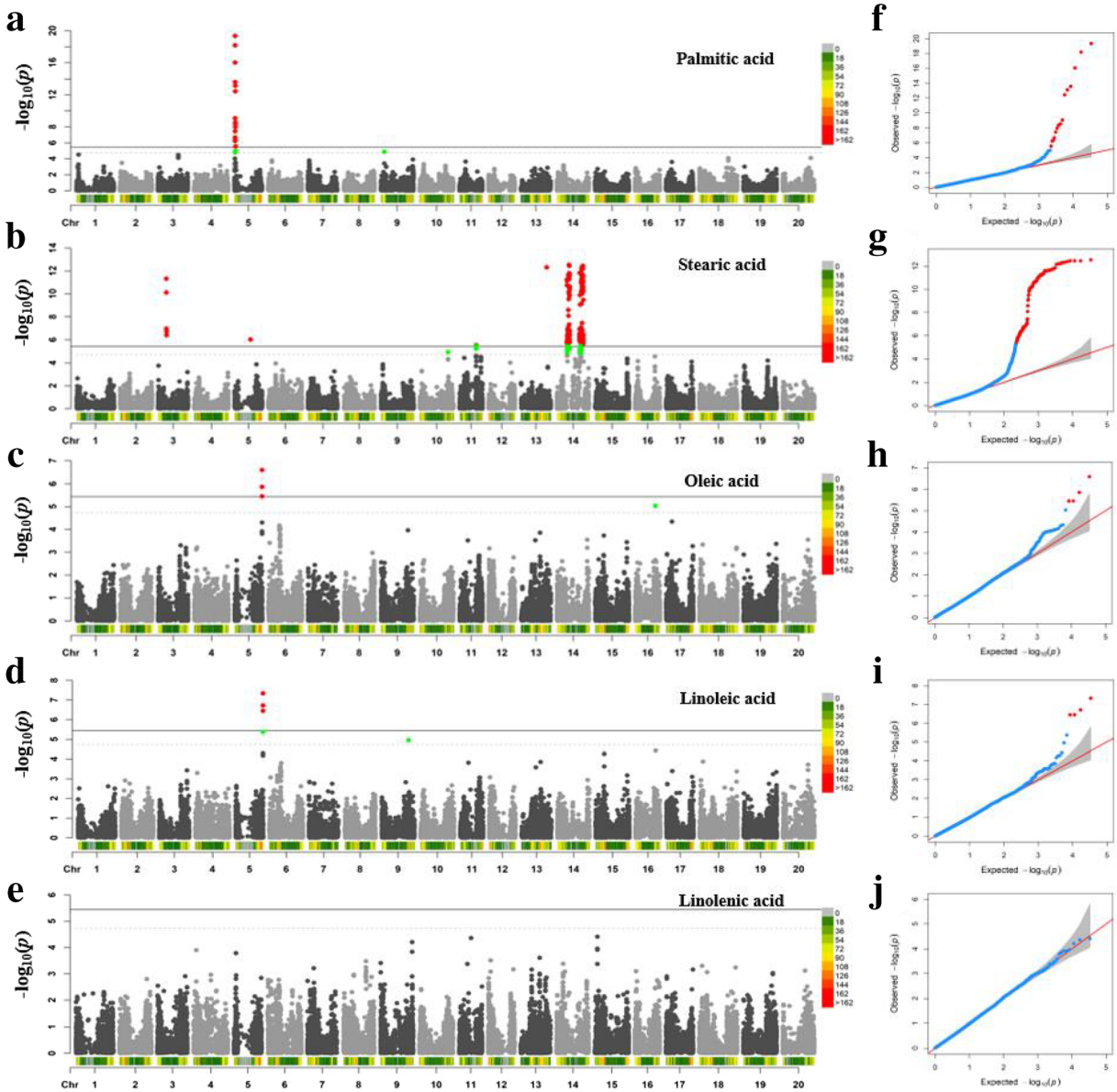

Fig. 2 Manhattan plots (left) and QQ-plots (right) of GWAS for palmitic (**a**, **f**), stearic (**b**, **g**), oleic (**c**, **h**), linoleic (**d**, **i**), and linolenic (**e**, **j**) acids across environments (ALL) using compressed mixed linear model. Red and green dots in Manhattan plots represent the significant SNP markers at genome-wide significance threshold level of $\alpha = 5\%$ (line) and suggestive significance threshold level of $\alpha = 25\%$ (dashed line), respectively. SNP density is presented over the *X*-axis with the corresponding heat map to the right of each Manhattan plot. Red lines in the QQ-plots denote the expected distribution of -$\log_{10}$-transformed *P* values and the region filled in gray indicates the 95% confidence interval. The SNP markers over the significance threshold level of $\alpha = 5\%$ are represented by red dots.

in the five individual environments and ALL, while one LD block with the range of 1,496,131–1,582,267 bp was significant at only OHC15 and ALL (Supplemental Table S3). An additional LD block (802,967–939,220 bp) was significantly associated with palmitic acid at genome-wide significant threshold only under OHW15 (Supplemental Table S3 and Fig. S3). *Glyma.05g011100* annotated as 3-oxoacyl-[acyl-carrier-protein (ACP)] synthase III (*KASIII*), which is one of the key enzymes in FA biosynthesis (Abbadi et al. 2000), was placed within the LD block (884,960–1,049,939 bp) and this candidate gene is discussed in

detail in the section of Palmitic acid vs. Stearic acid (see below). The region between 1,070,290 and 1,274,586 bp on Chr 5 contained *FATB1a* (*Glyma.05g012300*) encoding FA acyl carrier protein (ACP) thioesterase B. The most significant SNP marker (ss715592495) from this LD block was located 49.3 kb upstream from *FATB1a* gene exhibiting an association with seed palmitic acid content. The allelic effect of ss715592495 under ALL accounted for −0.32% variation in the seed palmitic acid content (Supplemental Table S2) and their allelic effects ranged from −0.27 to −0.41 under five different environments (Table 1). De Vries et al. (2011) discovered that a SNP on the *GmFATB1a* resulted in reduction of palmitic acid concentration. Also, the mutant alleles of *FATB1a* have been shown to decrease palmitic acid concentration to approximately 4%, which is a 30% reduction compared to wildtype (Thapa et al. 2016). Further, a large deletion (254 kb) of the *FATB1a* in N0304-303-3 (Goettel et al. 2016) and N87-2122-4 (Bachleda et al. 2016) also led to a reduction of palmitic acid in soybean seed. SNPs have been associated with stearic acid in other GWAS of soybean seed FA as well (Fang et al. 2017; Zhang et al. 2018). Zhang et al. (2018) reported the association of the SNP markers, ss715592503 and ss715592495, with palmitic acid, and *Glyma.05g012300* was located between these two markers. Thus, we confirmed both SNPs for strong association with palmitic acid and *Glyma.05g012300* resided in the LD block with ss715592495.

On Chr 9, ss715605172 within the LD block (5,606,447–5,690,163 bp) was identified as the most significant SNP in NC15 (α = 5%) and ALL (α = 25%) (Supplemental Tables S2, S3 and Fig. S6). This marker exhibited −0.18% (NC15) and −0.12% (ALL) of allelic effects for palmitic acid content and was positioned within the previously identified protein and oil meta-QTL, mPO9-2 (Van and McHale 2017). Since soybean oil content consists of relative proportions of five FA concentrations, ss715605172 could contribute to regulate quantity of seed oil and palmitic acid.

SNPs ss715578879 on Chr 1 (OHW15, OHC15, and IL15), ss715581672 on Chr 2 (OHW15), and ss715630888 on Chr 18 (OHW14) were additionally identified as the most significant SNPs for palmitic acid (Supplemental Table S3 and Fig. S2-S5). The Chr 18 significant LD block was placed within the previously known QTL, seed oil 24-12 (Qi et al. 2011).

Most significant LD blocks for stearic acid located on Chr 14

Under ALL, 30 LD blocks at a 5% level of genome-wide significance were identified as significantly associated with seed stearic acid content on five chromosomes (Chrs 3, 5, 11, 13, and 14) (Fig. 2b and Supplemental Table S2). Nine more environment-specific LD blocks were detected on four chromosomes (2, 1, 1, and 5 LD block(s) on Chrs 6, 10, 12, and 14, respectively) and these were significant under only one environment (Supplemental Table S3 and Fig. S3-S5).

On Chr 3, ss715618536 was highly significant in each environment and ALL (Supplemental Tables S2 and S3). The allelic effects were 0.24% for ALL (Supplemental Table S2) and the individual environment ranged from 0.21 to 0.27% in its allelic effect (data not shown). Fang et al. (2017) identified this genomic region associated with three traits, 18-carbon FA (FA18) content, FA18 ratio (FA18 content/total FA content), and the ratio of FA18 to 16-carbon FA (FA16) (FA18 content/FA16 content). In addition, this region overlapped with a previously reported QTL, cqSeed oil-005 (Pathan et al. 2013). Within this LD block, the candidate gene *Glyma.03g066200* is annotated as a polyketide cyclase/dehydrase and lipid transport (Iyer et al. 2001) and is located 30.9 kb from ss715618536.

Glyma.05g095000 encoding pyruvate kinase was placed within a LD block (24,314,289–25,462,356 bp), which is significant under ALL and NC15 at genome-wide and under OHW15 at suggestive significance threshold (Fig. S3, S6 and Supplemental Tables S2, S3). This gene represents a promising candidate for the regulation of stearic acid content by glycolysis and co-localized with QTL for seed oil 3-3 (Mansur et al. 1996), 38-1 (Rossi et al. 2013), 42-2, and 42-4 (Han et al. 2015). Two significant LD blocks on Chr 6 were identified only under NC15 at genome-wide significance level and the *KASIII* gene of FA biosynthesis process (*Glyma.06g214800*) was in one of the LD blocks (22,036,914–22,733,239 bp) (Supplemental Table S3). A previously known QTL, seed oil 23-1 (Hyten et al. 2004a), covers these two significant LD blocks on Chr 6. Thus, these Chr 6 LD blocks are potential candidate genomic regions for controlling FA and oil contents, although these genomic regions were identified only under NC15.

Among two significant LD blocks on Chr 10, the first LD block (2,789,429–2,924,705 bp) was significant

Table 1 List of candidate genes, the significant SNP, and environment (allelic effect of significant SNPs in percentage) by the associated traits and by the major enzymes related to FA biosynthesis and carbohydrate metabolic process

Trait[a]	FATB[b]	SACPD-C[b]	KASIII[b]	FBPase[b]
Pal	*Glyma.05g012300* ss715592495 ALL** (−0.32%), OHW14 (−0.27%), OHW15** (−0.32%), OHC15** (−0.36%), IL15** (−0.34%), NC15** (−0.41%)		*Glyma.05g011100* ss715592510 ALL** (−0.21%), OHW14** (−0.20%), OHW15** (−0.19%), OHC15** (−0.25%), IL15** (−0.21%), NC15** (−0.28%)	
Ste		*Glyma.14g121400* ss715618427 ALL** (0.24%), OHW14 (0.20%), OHW15** (0.27%), OHC15** (0.26%), IL15** (0.25%), NC15** (0.27%)	*Glyma.06g214800* ss715593906 NC15** (−0.13%)	*Glyma.11g226900*, *Glyma.11g227100* ss715610315 OHC15* (−0.16%)
Lio			*Glyma.05g011100* ss715592508 IL15** (0.73%)	
Lin				*Glyma.15g050100* ss715621816 OHC15** (0.26%)

Estimated allelic effect of alternative allele relative to Williams 82

* Suggestive significance threshold (25%); ** genome-wide significance threshold (5%)

[a] *Pal* palmitic acid, *Ste* stearic acid, *Lio* linoleic acid, *Lin* linolenic acid

[b] *FATB* acyl-ACP thioesterase, *SACPD-C* Δ9-stearoyl-acyl carrier protein-desaturase, *KASIII* 3-oxoacyl-[acyl-carrier-protein (ACP)] synthase III, *FBPase* fructose-1, 6-bisphosphatase

only under OHC15 at 5%, whereas the second LD block (43,186,362–43,479,482 bp) was significant under both ALL and OHC15 at 25% (Supplemental Tables S2, S3 and Fig. S4). The first LD block was located within a previously known QTL, seed stearic 8-9 (Fan et al. 2015), indicating that this LD block is not novel QTL for stearic acid. *Glyma.10g201000* encoding an enzyme in pyruvate decarboxylation was within the second significant LD block on Chr 10. While seed oil 43-33, 43-34, and 43-35 (Mao et al. 2013) were located between these two significant LD blocks, no previously known QTL for FA or oil content overlapped with the second LD block.

Candidate genes related to metabolic process were present in each LD block on Chr 11. Under ALL, *Glyma.11g183200* encoding glucan endo-1,3-β-D-glucosidase and three genes for alcohol-forming fatty acyl-CoA reductase were candidate genes for controlling seed stearic acid content (Supplemental Table S2). The QTL seed oil 24-14 (Qi et al. 2011) covers these two LD blocks (25,024,963–25,112,193 bp and 25,290,358–25,423,069 bp), which are significant only under ALL. One significant LD block on Chr 12 was associated with palmitic acid only under IL15 and *Glyma.12g174300* involving in sphingolipid biosynthesis as a part of lipid biosynthesis (http://soycyc.soybase.org) (Supplemental Table S3). Also, this significant LD block overlapped with the seed oil 44-2 (Leite et al. 2016) and seed linolenic 12-12 QTLs (Ha et al. 2014).

On Chr 13, ss715616084 was significantly associated with stearic acid under each individual and ALL environments (Supplemental Tables S2 and S3). The allelic effects of this SNP ranged from 0.23 to 0.27% (data not shown). The LD block (39,293,625–39,311,323 bp) containing this SNP was located 1.4 Mb away from a previously reported seed stearic 3-1 QTL (Song et al. 2004; Reinprecht et al. 2006).

Out of the significant 30 LD blocks related to stearic acid, Chr 14 contained 26 significant LD blocks under ALL (Supplemental Table S2). These LD blocks fell into two large genomic regions of approximately 6.9 and 6.4 Mb (14,847,459–21,700,497 bp and 34,158,025–40,524,189 bp). These large regions were coincident with the identified genome-wide association signals, such as FA18 content, FA18 ratio, and ratio of FA18 and FA16 (Fang et al. 2017). Meta-QTL, mqSeed oil-005 (Qi et al. 2011), covers all Chr 14 significant LD blocks and numerous QTL for palmitic, stearic, linoleic,

linolenic, or oil contents previously reported by using bi-parental mapping populations in these LD blocks (Csanádi et al. 2001; Tajuddin et al. 2003; Spencer et al. 2004; Panthee et al. 2006; Reinprecht et al. 2006; Bachlava et al. 2009; Kim et al. 2010; Liang et al. 2010; Li et al. 2011; Qi et al. 2011; Eskandari et al. 2013; Mao et al. 2013; Han et al. 2015).

The major stearic acid gene, *SACPD-C* encoding Δ9-stearoyl-acyl carrier protein-desaturase responsible for converting stearic acid to oleic acid, was identified on Chr 14 (Gillman et al. 2014). Soybean mutant lines confirmed that *SACPD-C* gene plays a role to enhance stearic acid levels in soybean seeds by deletion of one *SACPD* isoform encoding gene (*SACPD-C*). Radiation-induced soybean mutants showed moderately increased stearic acid by 10–15% of seed oil (Gillman et al. 2014) and one sodium azide-induced mutant line, A6, displayed remarkably elevated seed stearic acid by up to 28% of the total seed oil (Zhang et al. 2008). This gene has been consistently identified in several recent GWAS reports (Fang et al. 2017; Leamy et al. 2017; Zhang et al. 2018). Also, our study identified the significant ss715618427 within the LD block (17,442,487–17,590,728 bp) and this SNP was placed within *SACPD-C* (*Glyma.14g121400*, 17,499,717–17,502,413 bp). The allelic effect of ss715618427 on stearic acid content was 0.24% in this study. Other candidate genes were also located in these LD blocks on Chr 14. *Glyma.14g121200* encoding alcohol dehydrogenase in pyruvate fermentation is also placed within the same LD block as *SACPD-C*. A gene encoding alcohol dehydrogenase (*Glyma.14g156400*) was identified as one of the candidate genes positioned within the LD block (34,158,025–34,746,851 bp). Thus, ss715618427 could be a good marker for selection of soybean lines modifying stearic acid content in seed.

Among the candidate genes in the Chr 14 LD blocks significantly associated with stearic acid content, many genes were related to metabolic pathways. The significant LD block (16,657,427–17,442,487 bp) contained three candidate genes, *Glyma.14g120200*, *Glyma.14g120300*, and *Glyma.14g120500*, and these genes are involved in lipid degradation, carbohydrate transport, starch biosynthesis, respectively. *Glyma.14g123500* encoding sugar transporter and *Glyma.14g124100* encoding glycosyltransferase for galactolipid biosynthesis I (http://soycyc.soybase.org) are candidate genes located within the significant LD blocks. ss715618492 in the LD block (18,696,303–19,347,299 bp) containing *Glyma.14g124100* showed 0.26% allelic effect, which is the highest effect among the most significant SNPs on Chr 14 under ALL (Supplemental Table S2).

Confirmed and predicted genes controlling contents of palmitic acid vs. stearic acid

FATB1a and *SACPD-C* are the established primary genes that control palmitic acid and stearic acid in soybean, respectively. Our study identified the genomic regions containing these two genes, which are the LD block associated with palmitic acid (Chr 5: 1,070,290–1,274,586 bp) and the LD block related to stearic acid (Chr 14: 17,442,487–17,590,728 bp). Highest negative allelic effect (−0.32%) of ss71559245 on Chr 5 indicated this genomic region is responsible for regulating palmitic acid via *FATB1a* (*Glyma.05g012300*) (Table 1). ss715618427 on Chr 14 was identified as one of the most significant markers related to stearic acid with 0.24% of allelic effect and its position was within the 5' UTR region of *SACPD-C* (*Glyma.14g121400*) (Table 1). Thus, our study confirmed that *FATB1a* and *SACPD-C* are the primary genes for synthesis of these saturated FAs.

In addition, *KASIII* encoding 3-oxoacyl-[acyl-carrier-protein (ACP)] synthase III was identified as a putative soybean gene that increases saturated FA contents. *KASIII* is the major FA condensing enzyme to produce the synthesis of very long-chain fatty acids and typically plays a role in the initiation of FA biosynthesis (Jackowski and Rock 1987; Walsh et al. 1990; Abbadi et al. 2000). Interestingly, our study was able to identify *KASIII* associated with two different FAs in two different genomic regions. One *KASIII* was placed within a LD block each on Chr 5 (884,960–1,049,939 bp) associated with palmitic acid and Chr 6 (22,036,914–22,733,239 bp) associated with stearic acid. In the Chr 5 LD block, ss715592510 was significantly associated with palmitic acid in every individual environment and ALL (Supplemental Tables S2 and S3). *Glyma.05g011100*, putatively encoding a *KASIII*, was located only 10.6 kb away from ss715592510. However, ss715593906 related to stearic acid in the Chr 6 LD block was significantly identified only under NC15 (Table 1 and S3). *Glyma.06g214800*, also annotated as *KASIII* was located 128 kb away from ss715593906. Thus, these *KASIII* genes within our significant LD blocks would be useful for studying genes controlling contents of palmitic acid and stearic acid.

The most significant LD blocks for oleic acid located on Chr 5

Compared to stearic acid, few LD blocks were significant for oleic acid. Under ALL, one LD block on Chr 5 was identified at 5% genome-wide levels (Fig. 2 and Supplementary Table S2). However, additional LD blocks were detected under specific environment(s). Environment-specific significant LD blocks were also identified on Chr 5. One LD block (41,754,397–41,893,109 bp), which is the same LD block under ALL, was significant in every environment except OHW14, whereas the second Chr 5 LD block (41,883,826–42,071,794 bp) was significant only under IL15 (Supplemental Table S3 and Fig. S5). *Glyma.05g245000* was suggested as the candidate gene for regulating oleic acid within the second LD block. This candidate gene was annotated as 3-oxo-5-α-steroid 4-dehydrogenase in lipid metabolic processes and this enzyme may be involved in catalytic action for very-long-chain FA elongation with its enoyl-CoA reductase activity. Arabidopsis mutants disrupted in the gene coding for enoyl-CoA reductase showed changes in very-long-chain FA composition of seed triacylglycerols and sphingolipids (Zheng et al. 2005). A candidate gene is involved in sphingolipid biosynthesis residues within the Chr 12 LD block (33,067,511–33,216,359 bp) associated with stearic acid. Interestingly, both the second Chr 5 LD block for oleic acid and the Chr 12 LD block for stearic acid were significantly identified only under IL15 at genome-wide significance threshold. Zhang et al. (2018) reported major-effect QTL for oil content on Chr 5, including *Glyma.05g245000* as a candidate gene. Also, these two Chr 5 LD blocks were covered by seed oleic 1-1 (Diers and Shoemaker 1992), seed linoleic 1-1 (Diers and Shoemaker 1992), seed stearic 6-2 (Wang et al. 2012), and cqSeed oil-008, 012, and 015 (Pathan et al. 2013). Thus, the second LD block of Chr 5 (41,883,826–42,071,794 bp) may be an environment-specific genomic region associated with both FAs and oil contents.

While the significant LD block on Chr 16 was identified under ALL only at suggestive significance level, it remains interesting because *Glyma.16g146500* within this LD block is putatively involved in sphingolipid biosynthesis (Supplemental Table S2), similarly to the candidate gene for the stearic acid Chr 12 LD block. The −0.79% allelic effect of ss715624331 on Chr 16 was opposite to the allelic effect of ss715591644 (1.27%) on Chr 5. Like two significant Chr 5 LD blocks, this Chr 16 LD block was positioned near QTLs related to FA (Seed linolenic 6-4, Li et al. 2011) and oil contents (Seed oil 5-2, Lee et al. 1996; Seed oil 39-12, Wang et al. 2014; Seed oil 43-19, Mao et al. 2013).

Linoleic acid and its relationship with other FAs

Under ALL, a single LD block (Chr 5: 41,754,397–41,893,109 bp) was significantly associated with linoleic acid (Supplemental Table S2). Additionally, an LD block on Chr 5 (884,960–1,049,939 bp) that was significant only under IL15 (Supplemental Table S3) had also been identified for palmitic acid across all environments and encompasses *KASIII* (Supplemental Table S2). Although the most significant SNP within this LD block is different under each environment, allelic effects for palmitic and linoleic acids were compared. Allelic effect for linoleic acid under IL15 was 0.73%, whereas both ALL and IL15 showed −0.21% of allelic effect for palmitic acid (Table 1). This negative relationship between palmitic acid and linoleic acid is also supported by our study (Fig. 1) and other previous studies (Bachlava et al. 2008; Abdelghany et al. 2019).

One block each on Chrs 5, 11, and 19, which had environment-specific association with both oleic acid and linoleic acid, showed the same significance pattern by environment, regardless of the FA-association. For example, the Chr 5 block (41,754,397–41,893,109 bp) was significant for OHW15, OHC15, IL15, and NC15, the Chr 11 block (11,983,839–12,723,115 bp was significant only for OHW14, and the Chr 19 block (49,750,346–49,887,308 bp) was significant only for NC15 (Supplemental Table S3). Our results indicated a relationship between oleic acid and linoleic acid. Allelic effects of the seven SNP markers (ss715591649, ss715591647, ss715591644, ss715591642, and ss715591641 on Chr 5, ss715604038 on Chr 9, and ss715624331 on Chr 16) associated with oleic, linoleic acids or both traits indicated a negative relationship between oleic and linoleic acids, i.e., an allele was associated with increased oleic acid content and decreasing linoleic acid content or vice versa (Table 2). These findings are supported by previous studies (Heppard et al. 1996; Zhang et al. 2015; Matei et al. 2018) suggesting a negative relationship between oleic acid and linoleic acid as well as by the network association showing a high correlation between oleic linoleic FA concentrations (Fig. 1). Among the five SNPs within the

Table 2 Common SNP markers, allelic effect, and adjacent candidate genes significantly associated with both oleic acid and linoleic acid identified by compressed mixed linear model under ALL

Chr	LD block (bp)[a]	SNP[b]	Position (bp)	-log$_{10}$(P)		Allelic effect (%)[c]		Candidate gene[d]	Annotation[e]
				Oleic	Linoleic	Oleic	Linoleic		
5	41,754,397..41,893,109	ss715591649	41,780,982	4.30	5.36*	1.03	−0.95		
		ss715591647	41,807,117	5.45**	6.45**	1.20	−1.07		
		ss715591644	41,831,154	6.58**	7.33**	1.27	−1.09		
		ss715591642	41,854,786	5.86**	6.71**	1.24	−1.08		
		ss715591641	41,855,235	5.45**	6.45**	1.20	−1.07		
9	41,314,026..41,408,313	ss715604038	41,330,799	3.96	4.96*	0.84	−0.77	*Glyma.09g189200*	Aldehyde dehydrogenase-related
								Glyma.09g189200	Aldehyde dehydrogenase-related
16	30,680,872..30,862,012	ss715624331	30,774,077	5.03*	4.43	−0.79	0.60	*Glyma.16g146500*	Serine *C*-palmitoyltransferase
								Glyma.16g147000	Aspartate kinase

Where no SNPs were in LD with significant SNP, LD block was defined by positions of adjacent markers

* Suggestive significance threshold (25%); ** Genome-wide significance threshold (5%)

[a] Linkage disequilibrium (LD) blocks were constructed by Haploview 4.2 with four gamete method

[b] The most significant SNPs within the LD block

[c] Estimated allelic effect of alternative allele relative to Williams 82

[d] The potential candidate genes significantly associated with soybean metabolic process within the LD block

[e] The candidate genes associated with each trait are anticipated based on PFAM or SoyCyc9.0 annotation in SoyBase (www.soybase.org)

single LD block (41,754,397–41,893,109 bp) on Chr 5, ss715591644 showed the most significant marker for both traits (Supplemental Table S2). The allelic effects of these five SNPs exhibited pleiotropy, with the allelic effect for oleic acid ranging from 1.03 to 1.27% and the allelic effect for linoleic acid ranging from −0.95 to −1.09% under ALL. Previously, we reported that the same five SNPs on Chr 5 were significantly associated with seed oil content with a −0.33% allelic effect (Lee et al. 2019). Instead of showing the typical negative relationship between protein and oil contents, these five SNPs had near-zero allelic effects for protein content (−0.03%) under ALL. Moreover, ss715591638, which is one of the SNPs placed within the same LD block (Chr 5: 41,754,397–41,893,109 bp), was identified as the lead SNP associated with oil by Zhang et al. (2018). Lee et al. (2019) and this study demonstrated the negative relationships of oleic acid vs. linoleic acid and oleic acid vs. oil content and the positive relationship between linoleic acid and oil content, agreeing with Bachlava et al. (2008). Thus, this LD block is the important genomic region for not only oleic and linoleic acid contents, but also oil content in soybean.

Only environment-specific assocations were observed for linolenic acid

For linolenic acid, all associations, suggestive or significant, were environment-specific (Supplemental Table S3). On Chr 15, ss715621816 was significantly associated with linolenic acid under OHC15 (Supplemental Table S3). The significant LD block was located near reported QTL (cqSeed oil-007 and cqSeed oil-010), which inversely affect seed protein and oil content (Pathan et al. 2013) and a meta-QTL, mPO15-2 (3,846,538–3,964,389 bp, Van and McHale 2017). The LD block with ss715621816 overlapped with the reported QTL, seed linolenic 12-4 (Ha et al. 2014). This SNP may contribute to alteration of linolenic acid content by limiting the carbon availability toward oil biosynthesis. Within this LD block (3,937,899–4,110,122 bp) on Chr 15, *Glyma.15g050100* and *Glyma.15g051100* reside. *Glyma.15g050100* putatively encodes an FBPase, a carbon-demanding glycolysis and sucrose biosynthesis gene. *Glyma.15g050100* was previously suggested as a candidate gene by meta-QTL for both protein and oil traits (Van and McHale 2017) and GWAS for both protein and oil contents (Lee et al. 2019). *Glyma.15g051100* encodes phosphatidylserine decarboxylase involved in lipid metabolic process. Thus, this LD block associated with linolenic acid is a candidate genomic region for studying various metabolic processes.

Conclusions

We discovered a total of four novel genomic regions across all environments (ALL) at genome-wide significance threshold. One novel genomic region on Chr 5 for palmitic acid and one novel genomic region each on Chrs 5, 11, and 13 were associated with stearic acid (Fig. 3). Since SNP markers in these four novel genomic regions can be useful for marker-assisted selection, haplotype distributions of these four novel genomic regions were reported (Supplemental Table S4). One of the assayed SNPs (ss715626084 on Chr 13, Fig. 3 and Supplemental Table S2) resided within the coding region of *Glyma.13g293100* encoding Rho family GTPase, participating in the cellular processes, such as signaling to the cytoskeleton and vesicular trafficking (Nagawa et al. 2010). Additionally, nine environment-specific novel genomic regions were identified at a stringent genome-wide significance level. For palmitic acid, one novel genomic region each was detected only in OHW15 (Chr 2) and NC15 (Chr 9). For stearic acid, three novel genomic regions were identified, two specifically identified in NC15 (Chr 6, shown as a combined block in Fig. 3), and one specifically identified in IL15 (Chr 12). One (Chr 11), one (Chr 5), and two (Chrs 9 and 15) novel genomic regions were identified for oleic, linoleic, and linolenic acids, respectively (Fig. 3).

Our study not only identified novel QTL for FAs, but also confirmed many known QTL associated with FAs and oil contents. We were able to locate *FATB1a* (*Glyma.05g012300*) and *KAS III* (*Glyma.05g011100*), which are the major genes for regulating the amount of palmitic acid, within two different LD blocks (Table 2). *Glyma.06g214800* and *Glyma.05g011100*, both putatively encoding *KASIII* enzymes, were identified within QTL controlling stearic acid and palmitic/linoleic acid, respectively, with the latter possessing a negative allelic effect on palmitic acid in all environments and a positive allelic effect on linoleic acid only in IL15 (Supplemental Table S2). Our study confirmed that SNP ss715618427 was significantly associated with the stearic acid gene, *SACPD-C* (*Glyma.14g121400*). Also, genes putatively encoding for FBPase (*Glyma.11g226900/Glyma.11g227100* and *Glyma.15g050100*), one of the major key enzymes in gluconeogenesis, were identified within the significant LD blocks associated with stearic

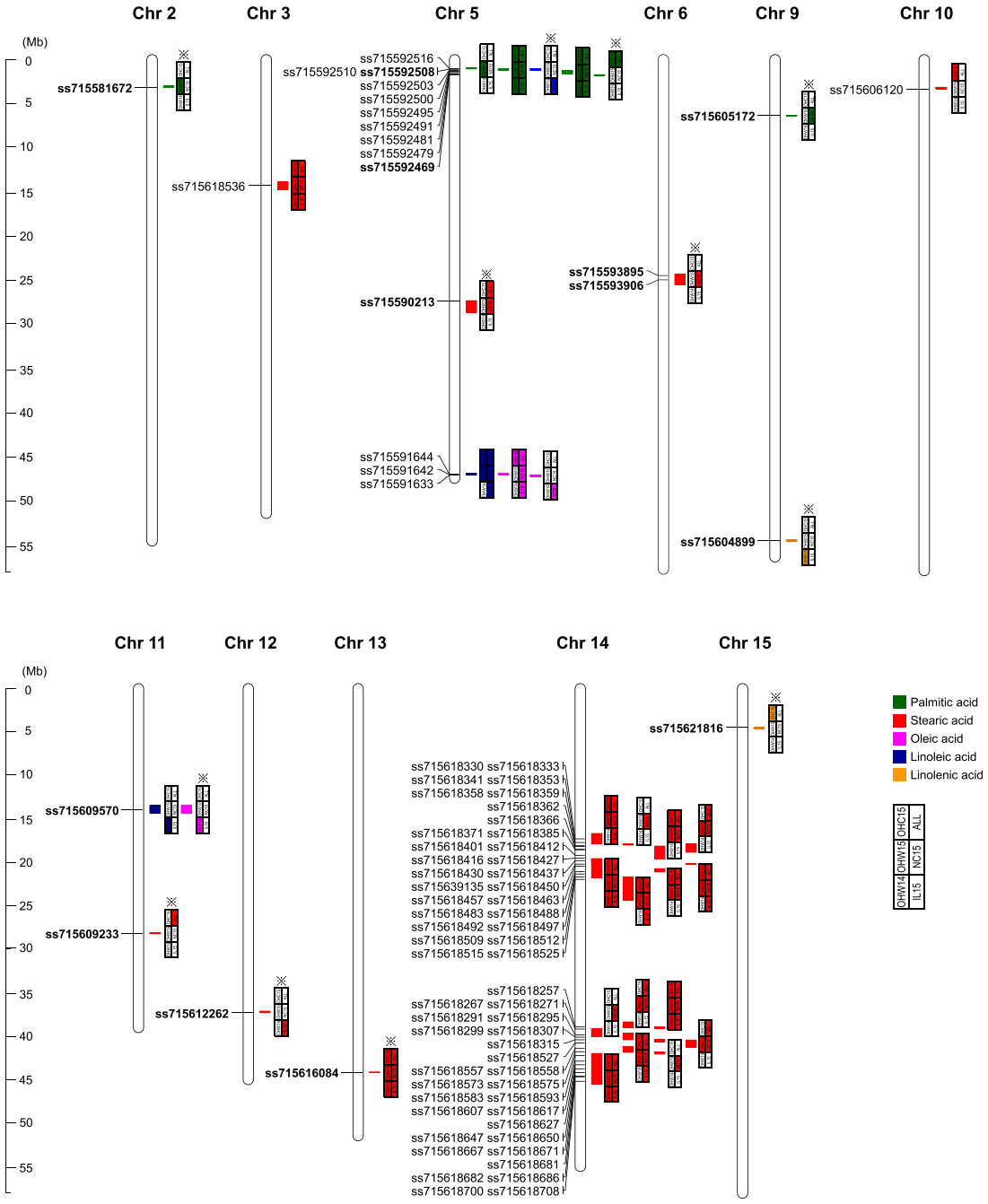

Fig. 3 Mapping of the significant SNP markers and QTL identified by compressed mixed linear model at the genome-wide significance threshold ($-\log_{10}(P) > 5.44$). The chromosomes and the positions of the significant SNPs are schematically presented with the associated fatty acid traits. LD blocks were combined, if ranges of LD block were overlapped and significant under the same environment. Box representing environment (see the enlarged box for key) next to colored bar (QTL) is filled with colors according to fatty acid traits (see color key on the Figure) in the environment-specific matter. " " on the top of the box represents novel genomic regions and bold SNP marker indicates the significant SNP within a novel genomic region. Chromosomes are aligned at approximate positions (Mb)

acid and linolenic acid, respectively, under only specific environments. Our results are in agreement with previous studies reporting a complex network among genes controlling FA biosynthesis and interactions of FAs with the growing environments of the soybean crops.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-021-01216-1.

Acknowledgement We appreciate Layne Connolly and Brandon McNeece for grammatical editing.

Author contribution MS conducted genetic and field experiments, analyzed data, and drafted and edited manuscript. KV conducted field experiments and wrote and edited manuscript. SL, KV, JL, and ET conducted field experiments and edited manuscript. RN contributed to selecting soybean accessions, conducted field experiments, and edited manuscript. LKM and MARM designed and organized the project and edited manuscript. All authors read and approved the final manuscript.

Funding The authors gratefully acknowledge the financial support for this study by United Soybean Board to LKM (Project# 1720-162-0111) and MARM (Project# 1720-152-0106). Salaries and research support for this project were also provided in part by the USDA National Institute of Food and Agriculture, Hatch project OHO01279 to LKM.

Declarations

Conflict of interest The authors declared no conflict of interest.

References

Abbadi A, Brummel M, Schütt BS, Slabaugh MB, Schuch R, Spener F (2000) Reaction mechanism of recombinant 3-oxoacyl-(acyl-carrier-protein) synthase III from *Cuphea wrightii* embryo, a fatty acid synthase type II condensing enzyme. Biochem J 345:153–160

Abdelghany AM, Zhang S, Azam M, Shaibu AS, Feng Y, Li Y, Tian Y, Hong H, Li B, Sun J (2019) Profiling of seed fatty acid composition in 1025 Chinese soybean accessions from diverse ecoregions. Crop J 8:635–644. https://doi.org/10.1016/j.cj.2019.11.002

Alt JL, Fehr WR, Welke GA, Sandhu D (2005) Phenotypic and molecular analysis of oleate content in the mutant soybean line M23. Crop Sci 45:1997–2000

Bachlava E, Burton JW, Brownie C, Wang S, Auclair J, Cardinal AJ (2008) Heritability of oleic acid content in soybean seed oil and its genetic correlation with fatty acid and agronomic traits. Crop Sci 48:1764–1772

Bachlava E, Dewey RE, Burton JW, Cardinal AJ (2009) Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. Crop Sci 49: 433–442

Bachleda N, Pham A, Li Z (2016) Identifying *FATB1a* deletion that causes reduced palmitic acid content in soybean N87-2122-4 to develop a functional marker for marker-assisted selection. Mol Breed 36:45

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265

Bellaloui N, Mengistu A, Kassem MA (2013) Effects of genetics and environment on fatty acid stability in soybean seed. Food Nutr Sci 4:165–175

Byfield GE, Upchurch RG (2007) Effect of temperature on microsomal omega-3 linoleate desaturase gene expression and linolenic acid contetn in developing soybean seeds. Crop Sci 47:2445–2452

Cecil JL, Kurnik WJ, Babcock D (1988) Epoxy compositions containing glycidyl ethers of fatty esters. U.S. Patent U.S. Patent 4,786,666

Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol 151:1030–1040

Csanádi GY, Vollmann J, Stift G, Lelley T (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. Theor Appl Genet 103:912–919

De Vries BD, Fehr WR, Welke GA, Dewey RE (2011) Molecular characterization of the mutant *fap3* (A22) allele for reduced palmitate concentration in soybean. Crop Sci 51:1611–1616

Diers BW, Shoemaker RC (1992) Restriction fragment length polymorphism analysis of soybean fatty acid content. J Am Oil Chem Soc 69:1242–1244

Eskandari M, Cober ER, Rajcan I (2013) Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. Theor Appl Genet 126:483–495

Fan S, Li B, Yu F, Han F, Yan S, Wang L, Sun J (2015) Anlaysis of additive and epistatic qiantitative trait loci underlying fatty acid concentrations in soybean seeds across multiple environments. Ephytica 206:689–700

Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Hu G, Zhou Z, Yu H, Zhang M, Pan Y, Zhou G, Ren X, Du W, Yan H, Wang Y, Han D, Shen Y, Liu S, Liu T, Zhang J, Qin H, Yuan J, Yuan X, Kong F, Liu B, Li J, Zhang Z, Wang G, Zhu B, Tian Z (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol 18:161

Gillman JD, Stacey MG, Cui Y, Berg HR, Stacey G (2014) Deletions of the *SACPD-C* locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biol 14:143

Goettel W, Ramirez M, Upchurch RG, An YQ (2016) Identification and characterization of large DNA deletions affecting oil quality traits in soybean seeds through transcriptome sequencing analysis. Theor Appl Genet 129: 1577–1593

Ha B-K, Kim H-J, Velusamy V, Vuong TD, Nguyen HT, Shannon JG, Lee J-D (2014) Identification of quatitative trait loci

controlling linolenic acid concentration in PI 483463 (*Glycine soja*). Theor Appl Genet 127:1501–1512

Han Y, Teng W, Wang Y, Zhao X, Wu L, Li D, Li W (2015) Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. Plant Breed 134:300–309

Heppard EP, Kinney AJ, Stecca KL, Miao GH (1996) Developmental and growth temperature regulation of two different microsomal ω-6 desaturase genes in soybeans. Plant Physiol 110:311–319

Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Steganiak TR, Schmidt ME (2004a) Seed quality QTL in a prominet soybean population. Theor Appl Genet 109:552–561

Hyten DL, Pantalone VR, Saxton AM, Schmidit ME, Sam CE (2004b) Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J Am Oil Chem Soc 81:1115–1118

Iyer LM, Koonin EV, Aravind L (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43:134–144

Jackowski S, Rock CO (1987) Altered molecular form of acyl carrier protein associated with beta-ketoacyl-acyl carrier protein synthase II (*fabF*) mutants. J Bacteriol 169:1469–1473

Kanobe C, McCarville MT, O'Neal ME, Tylka GL, MacIntosh GC (2015) Soybean aphid infestation induces changes in fatty acid metabolism in soybean. PLoS One 10:e0145660

Kim H-K, Kim Y-C, Kim S-T, Son B-G, Choi Y-W, Kang J-S, Park Y-H, Cho Y-S, Choi I-S (2010) Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. J Life Sci 20: 1186–1192

Leamy LJ, Zhang H, Li C, Chen CY, Song B-H (2017) A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). BMC Genomics 18:18

Lee SH, Bailey MA, Mian MAR, Carter TE Jr, Shipe ER, Ashley DA, Parrott WA, Hussey RS, Boerma HR (1996) RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor Appl Genent 93:649–657

Lee S, Jun T-H, McHale LK, Michel AP, Dorrance AE, Song Q, Mian MA (2017) Registration of Wyandot × PI 567301B soybean recombinant inbred line population. J Plant Reg 11: 324–327

Lee S, Van K, Sung M, Nelson R, LaMantia J, McHale LK, Mian MAR (2019) Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. Theor Appl Genet 132:1639–1659

Leite DC, Pinheiro JB, Campos JB, Di Mauro AO, Uneda-Trevisoli SH (2016) QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. Genet Mol Res 15:1–11

Li Z, Wilson RF, Rayford WE, Boerma HR (2002) Molecular mapping genes conditioning reduced palmitic acid content in N87-2122-4 soybean. Crop Sci 42:373–378

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

Li Y-H, Reif JC, Ma Y-S, Hong H-L, Liu Z-X, Chang R-Z, Qiu L-J (2015) Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. BMC Genomics 16:841

Liang HZ, Yu YL, Wang SF, Yun LI, Wang TF, Wei YL, Gong PT, Liu XY, Fang XJ, Zhang MC (2010) QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). Agric Sci China 9:1108–1116

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399

Liu Q, Singh S, Green A (2002) High-oleic and high-stearic cottonseed oils: nutritionally improved cooking oils developed using gene silencing. J Am Coll Nutr 21:205S–211S

Lunn J, Theobald HE (2006) The health effects of dietary unsaturated fatty acids. Nutr Bull 31:178–224

Mansur LM, Orf JH, Chase K, Jarvik T, Cregan PB, Lark KG (1996) Genetic mapping of agrnomic traits using recombinant inbred lines of soybean. Crop Sci 36:1327–1336

Mao T, Jiang Z, Han Y, Teng W, Zhao X, Li W (2013) Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. Plant Breed 132:630–641

Martin BA, Rinne RW (1986) A comparison of oleic acid metabolism in the soybean (*Glycine max* [L.] Merr.) genotypes Williams and A5, a mutant with decreased linoleic acid in the seed. Plant Physiol 81:41–44

Matei G, Meneguzzi C, Woyann LG, Todeschini MH, Trevizan DM, Conte J, Bozi AH, Benin G (2018) Oil, protein and fatty acid profiles of Brazilian soybean cultivars in multi-environmental trials. Aust J Crop Sci 12:686–698

McHale LK, Feller MK, McIntyre SA, Berry SA, St. Martin SK, Dorrance AE (2012) Registration of 'Summit,' a high-yielding soybean with race-specific resistance to *Phytophthora sojae*. J Plant Reg 7:36–41

Mensink RP, Katan MB (1992) Effect of dietary fatty acids on serum lipids and lipoproteins. Arterioscler Thromb Vasc Biol 12:911–919

Mian MAR, Cooper RL, Dorrance AE (2008) Registration of 'Prohio' soybean. J Plant Reg 2:208–210

Nagawa S, Xu T, Yang Z (2010) RHO GTPase in plants: Conservation and inventions of regulators and effectors. Samll GTPases 1:78–88

Nwokolo E (1996) Soybean (*Glycine max* (L.) Merr.). In: Nwokolo E, Smartt J (eds) Food and Feed from Legumes and Oilseeds. Springer, Boston, MA

Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7:957–970

Panthee DR, Pantalone VR, Saxton AM (2006) Modifier QTL for fatty acid composition in soybean oil. Euphytica 152:67–73

Pathan SM, Vuong T, Clark K, Lee J-D, Shannon JG, Roberts CA, Ellersieck MR, Burton JW, Cregan PB, Hyten DL, Nguyen HT, Sleper DA (2013) Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci 53:765–774

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

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

Qi Z-M, Han X, Sun YN, Qiong WU, Shan DP, Du XY, Liu CY, Jiang HW, Hu GH, Chen QS (2011) An integrated

quantitative trait locus map of oil content in soybean, *Glycine max* (L.) Merr., generated using a meta-analysis method for mining genes. Agric Sci China 10:1681–1692

Reinprecht Y, Poysa VW, Yu K, Rajcan R, Ablett GR, Pauls KP (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome 49:1510–1527

Rolling W, Lake R, Dorrance AE, McHale LK (2020) Genome-wide association analyses of quantitative disease resistance in diverse sets of soybean [*Glycine max* (L.) Merr.]. PLoS One 15:e02227710

Rossi ME, Orf JH, Liu L-J, Dong Z, Rajcan I (2013) Genetic basis of soybena adaptation to North American vs. Asian mega-environments in two independent populations from Canadian x Chinese crosses. Theor Appl Genet 126:1809–1823

Schneider R, Rolling W, Song Q, Cregan P, Dorrance AE, McHale LK (2016) Genome-wide association mapping of partial resistance to *Phytophthora sojae* in soybean plant introductions from the Republic of Korea. BMC Genomics 17:607

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109: 122–128

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, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2015) Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes Genom Genet 5: 1999–2006

SoyStats (2019) http://soystats.com. Accessed 20 Nov 2019

Spencer MM, Landau-Ellis D, Meyer EJ, Pantalone VR (2004) Molecular markers associated with linolenic acid content in soybean. J Am Oil Chem Soc 81:559–562

Tajuddin T, Watanabe S, Yamanaka N, Harada K (2003) Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. Breed Sci 53:133–140

Thapa R, Carrero-Colón M, Hudson KA (2016) New alleles of *FATB1A* to reduce palmitic acid levels in soybean. Crop Sci 56:1076–1080

Toubiana D, Semel Y, Tohge T, Beleggia R, Cattivelli L, Rosental L, Nikoloski Z, Zamir D, Fernie AR, Fait A (2012) Metabolic profiling of a mapping population exposes new insights in the regulation of seed metabolism and seed, fruit, and plant relations. PLoS Genet 8:e1002612

Van K, McHale LK (2017) Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed. Int J Mol Sci 18:1180

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Vaughn JN, Nelson RL, Song Q, Cregan PB, Li Z (2014) The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3: Genes Genom Genet 4:2283–2294

Vuong TD, Sonah H, Meinhardt CG, Deshmukh R, Kadam S, Nelson RL, Shannon JG, Nguyen HT (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. BMC Genomics 16:593

Walsh MC, Klopfenstein WE, Harwood JL (1990) The short chain condensing enzyme has a widespread occurrence in the fatty acid synthetases from higher plants. Phytochem 29:3797–3799

Wang N, Akey JM, Zhang K, Chakraborty R, Jin L (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. Am J Hum Genet 71:1227–1234

Wang X, Jiang GL, Green M, Scott RA, Hyten DL, Cregan PB (2012) Quantitative trait locus analysis of saturated fatty acids in a population of recombinant inbred lines of soybean. Mol Breed 30:1163–1179

Wang X, Jiang GL, Green M, Scott RA, Song Q, Hyten DL, Cregan PB (2014) Identification and validation of quantitative trait loci for seed yield, oil and protein contetns in two recombinant inbred line populations of soybean. Mol Genet Genomics 289:935–949

Wolf RB, Cavins JF, Kleiman R, Black LT (1982) Effect of temperature on soybean seed constituents oil, protein, moisture, fatty acids, amino acids and sugars. J Am Oil Chem Soc 59:230–232

Xie D, Han Y, Zeng Y, Chang W, Teng W, Li W (2012) SSR- and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments. Mol Breed 30:169–179

Zhang P, Burton JW, Upchurch RG, Whittle E, Shanklin J, Dewey RE (2008) Mutations in a Δ^9-stearoyl-ACP-desaturase gene are associated with enhanced stearic acid levels in soybean seeds. Crop Sci 48:2305–2313

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zhang A, Liu X, Wang G, Wang H, Liu J, Zhao W, Zhang Y (2015) Crude fat content and fatty acid profile and their correlations in foxtail millet. Cereal Chem 92:455–459

Zhang J, Wang X, Lu Y, Bhusal SJ, Song Q, Cregan PB, Yen Y, Brown M, Jiang GL (2018) Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. Mol Plant 11:460–472

Zhao X, Chang H, Feng L, Jing Y, Teng W, Qiu L, Zheng H, Han Y, Li W (2019) Genome-wide association mapping and candidate gene analysis for saturated fatty acid content in soybean seed. Plant Breed 138:588–598

Zheng H, Rowland O, Kunst L (2005) Disruptions of the arabidopsis enoyl-CoA reductase gene reveal an essential role for very-long-chain fatty acid synthesis in cell expansion during plant morphogenesis. Plant Cell 17:1467–1481