Journal of Plant Genetics and Breeding

Bhandari and Kantartzi, J Plant Genet Breed 2024, 8:6

Research Article Open Access

Genome-wide Association Study of Oil Content and Fatty Acid Composition in a Global Soybean Germplasm Panel

Rita Bhandari, Stella K. Kantartzi*

School of Agricultural Sciences, Southern Illinois University, Carbondale, IL, USA

Abstract

Soybean is an essential oilseed crop worldwide, contributing significantly to the food, feed, and pharmaceutical industries. Breeding efforts primarily focus on increasing the oil content while enhancing its composition, which is crucial for flavor, stability, and nutritional value. In this study, a genome-wide association study (GWAS) was performed to identify quantitative trait loci (QTL) and candidate genes associated with oil content and fatty acid composition using single nucleotide polymorphism (SNP) markers in a global panel of 146 soybean accessions. The results revealed 26 significant SNPs associated with oil content and the concentration of palmitic acid, oleic acid, linoleic acid, and linolenic acid. Of these, two SNPs were localized near Glyma.04g192100 on Chr. 4 and Glyma.05g015400 on Chr. 5 that have been found to be related to fatty acid metabolism in Arabidopsis thaliana. Overall, our findings offer valuable insight into the genetic basis of soybean oil traits, facilitating breeders in improving the oil content and composition.

Keywords: Oil content; Fatty acid; SNPs; GWAS

Introduction

Soybean (Glycine max L. Merr.) is an essential oil seed crop that is widely cultivated worldwide [1]. Over 50% of edible seed oil consumed by humans or used for animal feed is derived from soybeans [2]. The primary components of soybean oil are five fatty acids, namely palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) [3,4]. Due to the relatively high concentration of unsaturated fatty acids (oleic acid, 18% average concentration; linoleic acid, 55% average concentration; linolenic acid, 13% average concentration), soybean oil offers several health benefits, such as anti-tumor effects, inhibition of inflammatory processes, prevention of atherosclerosis, and regulation of cholesterol metabolism [5,6]. Oleic acid is a monounsaturated fatty acid with high oxidative stability and, consequently, highly useful for diverse applications as an excipient or solubilizing agent [7,8]. Conversely, linoleic and linolenic acids are polyunsaturated fatty acids (PUFAs) with low oxidative stability and rancidification that reduce storage time and alter oil flavor; however, they are considered essential acids since they cannot be synthesized in the human body [9]. Therefore, genetic studies focused on increasing soybean seed oil content and concurrently improving composition are important to satisfy global requirements in the food, feed, and pharmaceutical industries [10].

Oil content and fatty acid composition are heritable quantitative traits controlled by multiple major and minor quantitative trait loci (QTL). Previous studies have reported 327 QTL related to oil content and 228 QTL related to fatty acid composition across all 20 soybean chromosomes [11-16]. Nevertheless, most of these QTL have been identified using linkage mapping in dual or multi-parental mapping populations, an approach with low genome resolution for examining recombination events and apprehending allelic diversity [17,18]. Genome-wide association study (GWAS) is an alternative method for identifying QTL and mining candidate genes in natural populations [19]. Compared with conventional linkage mapping, GWAS has been shown to increase marker positioning accuracy and lead to efficient QTL mapping [14,20]. Silva LCC, et al. [21] applied GWAS in an early generation segregating population and pinpointed 20 QTL related to oil content and the five fatty acids; nevertheless, they did not report any candidate genes.

In this study, we used 1,242 single nucleotide polymorphism (SNP) markers to detect QTL and candidate genes associated with oil content and fatty acid composition in a global panel of 146 diverse soybean accessions. Our data may help to better understand the underlying molecular mechanisms controlling the traits and identify markers that can be used to accelerate the improvement of oil content and composition in soybean seed.

Materials and Methods

Germplasm panel

We created a global panel of 146 soybean accessions, in which 116 maturity group (MG) IV represented the most genetically diverse accessions as described in previously published studies [22,23], whereas the rest, 15 MG 00–III and 15 MG V–VII, were selected from the USDA-ARS Germplasm Resources Information Network (GRIN)-Global, based on their phenotypic description and various breeding values. The accessions were originated from 15 different countries, including China (88), United States (17), Japan (12), South Korea (8), Serbia (5), France (3), Georgia (3), India (2), Uganda (2), Morocco (1), Nepal (1), South Africa (1), Taiwan (1), Vietnam (1), and Zambia (1). Detailed information of the germplasm panel can be provided upon request.

Phenotypic data and descriptive statistics

Phenotypic data on oil content and fatty acid composition were obtained from USDA GRIN-Global (https://npgsweb.ars-grin.

***Corresponding author:** Stella K. Kantartzi, School of Agricultural Sciences, Southern Illinois University, Carbondale, IL, USA, E-mail: stella.kantartzi@siu.edu

Received: 31-Oct-2024, Manuscript No. jpgb-24-151515; **Editor assigned:** 02-Nov-2024, Pre QC No. jpgb-24-151515 (PQ); **Reviewed:** 14-Nov-2024, QC No. jpgb-24-151515, **Revised:** 22-Nov-2024, Manuscript No. jpgb-24-151515 (R); **Published:** 29-Nov-2024, DOI: 10.4172/jpgb.1000236

Citation: Bhandari R, Kantartzi SK (2024) Genome-wide Association Study of Oil Content and Fatty Acid Composition in a Global Soybean Germplasm Panel. J Plant Genet Breed 8: 236.

gov). Descriptive statistics, univariate distribution, analysis of variance, Tukey-Kramer post-hoc test, and Pearson's correlation were performed using JMP Pro 17 (SAS Institute Inc., Cary, NC, USA).

Genotypic data and population structure

All the accessions were sown at the Horticulture Research Center, Southern Illinois University, Carbondale, IL, in six-inch plastic pots containing Berger BM1 nutrient holding mix (Berger, Saint-Modeste, QC, Canada) and allowed to grow at 27°C ± 2°C with irrigation based on soil moisture. The experimental arrangement was a randomized complete block design with two blocks and three repetitions per block. Two leaf punches were collected at the V1 stage from different leaves of each accession. DNA extraction with the HotSHOT method and SNP genotyping with PlexSeq™ were carried out by AgriPlex Genomics (Cleveland, OH, USA). Marker filtering for missing data, minor allele frequency lower than 5%, and heterozygosity higher than 10% generated 1,242 high-quality SNPs across all 20 chromosomes that were used for further analysis [24,25]. The population structure of the global panel and a Q matrix for GWAS were generated with STRUCTURE 2.3.4, a Bayesian model-based software [26]. The burn-in iteration was 10,000, followed by 100,000 Markov chain Monte Carlo replications after burn-in using an admixture and allele frequencies correlated model. The hypothetical number of subpopulations (k) ranged from 1 to 10, and the statistical value delta K was calculated as described by Evanno et al. [27]. STRUCTURE HARVESTER, a Python-based front-end software, was used to determine the optimal value of K [28]. All 146 soybean accessions were assigned to a subpopulation based on the optimum k (k = 4), and the population structure matrix (Q) was generated for GWAS.

Association analysis

GWAS for all traits (oil content and concentration of palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid) was conducted with TASSEL 5.0, applying three different models, including the single-marker regression (SMR) without structure or kinship, the general linear model with structure (GLM-Q), and the mixed linear model with structure and kinship (MLM-Q + K) [29]. Quantile-quantile (Q-Q) and Manhattan plots were generated to illustrate the results of GWAS for each trait. SNPs with a $-\log_{10}(P)$ (LOD) > 3 were significantly associated at $p < 0.05$, as previously suggested by Churchill & Doerge GA [30].

Candidate gene identification

Using the Glyma.Wm82. a2 reference in SoyBase (https://www.soybase.org), we searched the flanking regions within 10 kb of each significant SNP to discover candidate genes for each trait. Gene description and functional annotations were reported based on the Arabidopsis Information Resource (TAIR) as a primary source of information, followed by PANTHER or KEGG Orthology databases [31-33].

Results and Discussion

Phenotypic analysis

Descriptive statistics (mean, standard deviation, minimum, and maximum) and values of skewness and kurtosis for each trait (oil content and concentration of palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid) are provided in Table 1. We observed a substantially wide amplitude for each trait with oil content ranging from 10.90% of dry seed weight (PI 253665B, China) to 23.10% of dry seed weight (PI 567307, China); palmitic acid ranging from 9.0% of oil content (PI 567477, China) to 14.10% of oil content (PI 578494A, China); stearic acid ranging from 2.70% of oil content (PI 567749B, PI 417345B, China, PI 518664, PI 512039, USA, and PI 548401, Morocco) to 5.50% of oil content (PI 574534, China); oleic acid ranging from 14.90% of oil content (PI 266807D, China) to 32.95% of oil content (PI 548546, USA); linoleic acid ranging from 46.45% of oil content (PI 548546, USA) to 59.70% of oil content (PI 518664, USA); linolenic acid ranging from 5.40% of oil content (PI 423926, Japan, and PI 567583C, China) to 14.40% of oil content (PI 567633, China). The Shapiro-Wilk (w) test indicated normal distribution for oil content and palmitic acid concentration but not for the concentrations of stearic acid ($p < 0.01$), oleic acid ($p < 0.001$), linoleic acid ($p < 0.01$), and linolenic acid ($p < 0.05$). Any deviation from the normal curve could be explained by the presence of low or high outliers; however, all the distributions were unimodal and did not indicate the presence of any sub-populations or major genetic effects that can significantly distort the normality [21,34]. Pearson's coefficients revealed that the oil content was positively correlated with oleic acid ($r = 0.20$, $p < 0.05$) and negatively correlated with linolenic acid ($r = -0.24$, $p < 0.01$) (Figure 1A and Figure 1B). These results suggested the existence of tightly linked genetic factors that control these traits. Additionally, we found that oleic acid was negatively correlated with linoleic acid ($r = -0.72$, $p < 0.001$) and linolenic acid ($r = -0.59$, $p < 0.001$) (Figure 1C and Figure 1D). Previous studies also reported a significant and negative correlation between oleic acid and linoleic acid but a positive correlation between oleic acid and linolenic acid [35,36]. Correlations of oleic acid with linoleic and linolenic acids are expected since the three fatty acids are involved

Table 1: Descriptive statistics (mean, standard deviation, minimum, and maximum), skewness, and kurtosis of oil content and concentration of five fatty acids (palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid) in the panel of 146 soybean accessions.

Traits	Mean	SD	Min	Max	Skewness	Kurtosis
Oil content (% of dry seed weight)	18.19	2.07	10.9	23.1	-0.41	0.74
Palmitic acid (% of oil)	11.63	0.83	9	14.1	-0.03	0.7
Stearic acid (% of oil)	3.81	0.55	2.7	5.5	0.4	0.8
Oleic acid (% of oil)	21	2.77	14.9	32.95	1.02	2.51
Linoleic acid (% of oil)	54.6	2.2	46.45	59.7	-0.58	1.06
Linolenic acid (% of oil)	8.98	1.74	5.4	14.4	0.57	0.43
Traits	Mean	SD	Min	Max	Skewness	Kurtosis
Oil content (% of dry seed weight)	18.19	2.07	10.90	23.10	-0.41	0.74
Palmitic acid (% of oil)	11.63	0.83	9.00	14.10	-0.03	0.70
Stearic acid (% of oil)	3.81	0.55	2.70	5.50	0.40	0.80
Oleic acid (% of oil)	21.00	2.77	14.90	32.95	1.02	2.51
Linoleic acid (% of oil)	54.60	2.20	46.45	59.70	-0.58	1.06
Linolenic acid (% of oil)	8.98	1.74	5.40	14.40	0.57	0.43

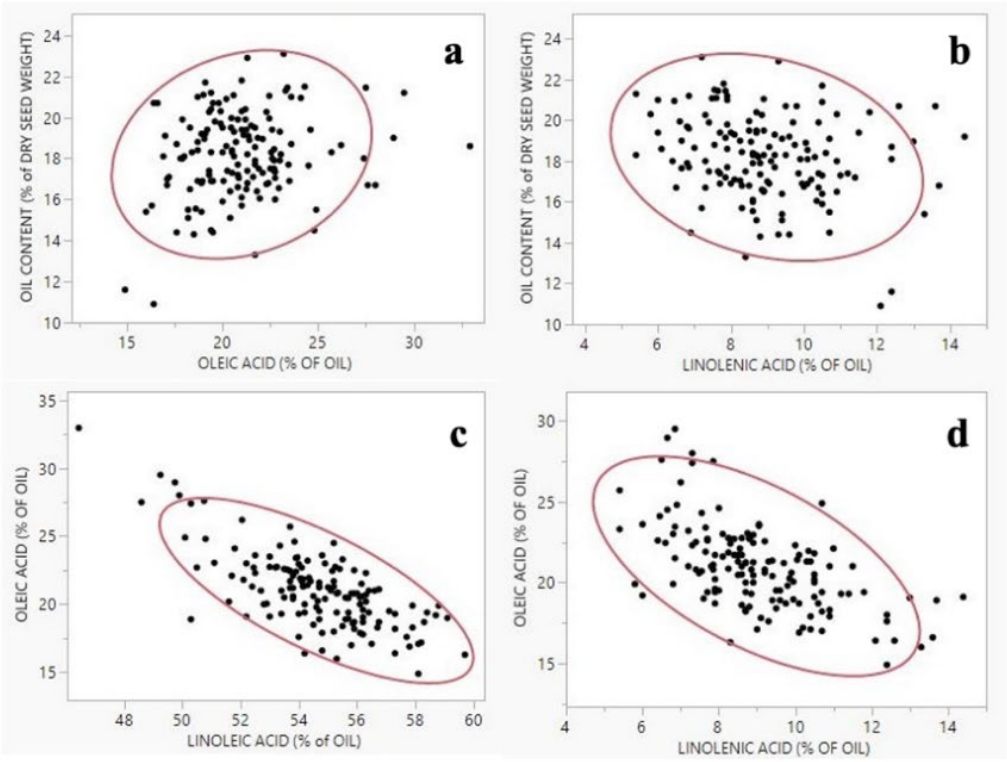

Figure 1: Bivariate correlation of **(A)** oil content and oleic acid concentration (r = 0.20, p < 0.05), **(B)** oil content and linolenic acid (r = -0.24, p < 0.01) **(C)** oleic acid and linoleic acid (r = -0.72, p < 0.001) **(D)** oleic acid and linolenic acid (r = -0.59, p < 0.001).

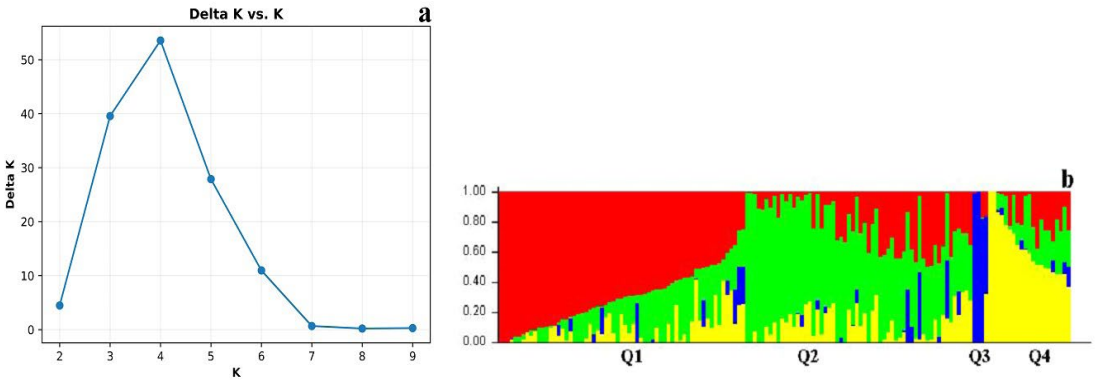

Figure 2: Structure analysis in the panel of 146 soybean accessions: **(A)** optimal delta was identified at K = 4 using STRUCTURE HARVESTER and **(B)** classification into four clusters (Q1, red; Q2, green; Q3, blue; Q4, yellow) using STRUCTURE 2.3.4.

in the same metabolic pathway [31]; nonetheless, any discrepancies could be attributed to the different growth conditions of soybean populations. It is known that environmental temperature modifies the enzymatic activity of oleate and linoleate desaturase, affecting the oleic-linoleic-linolenic acid composition in the soybean seed [37]. This is backed by field and greenhouse experimental data which showed that the concentration of oleic acid increases with temperature while that of linoleic and linolenic acids decreases [38-40].

Population structure

The 146 accessions were grouped into four sub-populations (clusters Q1–Q4), since the peak of delta K was observed at K = 4 (Figures 2A and Figure 2B). Cluster 1 (red) included 61 accessions, of which 33 were from China, six each from Japan, South Korea, and the United States, two each from France and Georgia, and one each from Taiwan, Uganda, and Zambia; Cluster 2 (green) included 58 accessions, of which 40 were from China, six from the United States, five from Japan, two from India, and one each from France, South Korea, Serbia, Uganda, and Morocco; Cluster 3 (blue) included six accessions, of which four were from China and one each from Serbia and the United States; and Cluster 4 (green) included 21 accessions, of which 11 were from China, four from the United States and one each from Vietnam, Georgia, Japan, South Korea, Nepal, and South Africa.

Association analysis and SNP markers identification

Of the three different models used in TASSEL, GLM + Q significantly reduced the false positive rates. The association of SNPs with oil content and the concentration of palmitic acid, oleic acid, linoleic acid, and linolenic acid was depicted by the deviation of the observed *p*-value from the expected distribution in the QQ plots (Figure 3). We identified two SNPs for oil content on Chr. 4 (ss715588323) and Chr. 15 (ss715622616), results that were in accordance with those previously reported in simple and composite internal mapping studies [13,21]. In addition, we located three SNPs for palmitic acid concentration on Chr. 5 (ss715592481), Chr. 10 (ss715607624), and Ch. 12 (ss715613242), which could partially confirm the results of Zhao X, et al. [41] that reported SNPs on all chromosomes except for Chr. 6 across three different environments. For oleic acid concentration, eight SNPs were found on Chr. 8 (ss715601356), Chr. 9 (ss715604069),

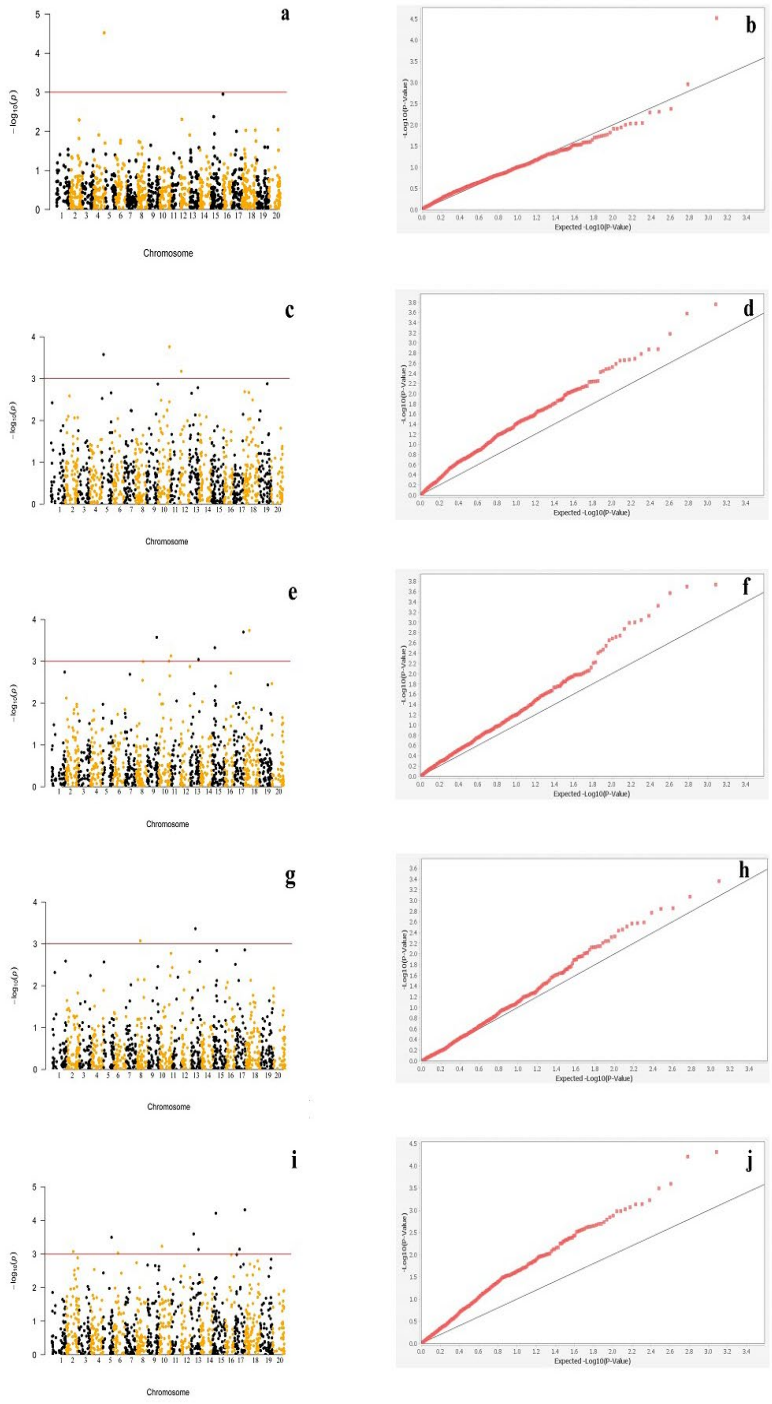

Figure 3: Manhattan plots (left) and quantile-quantile (Q-Q) plots (right) for (**a, b**) oil content, (**c, d**) palmitic acid, (**e, f**) oleic acid, (**g**, **h**) linoleic acid, (**i, j**) and linolenic acid using the generalized linear model with structure (Q). The red line in Manhattan plots indicates the threshold of -log (P) (LOD) = 3 for single nucleotide polymorphism (SNP) significance at $p < 0.05$.

Table 2: Candidate gene models and descriptions within 10-kb flanking regions of single nucleotide polymorphism (SNP) associated with oil content and concentrations of four fatty acids (palmitic acid, oleic acid, linoleic acid, and linolenic acid) using Wm82.a2. v1.

Traits	Chr.	SNP Marker	Major Allele	Minor Allele	GLM+Q LOD	Genomic Location	Gene Name	Gene Annotation
Oil content	4	ss715588323	A	G	4.521	Intergenic	Glyma.04g192000	Relative to early flowering 6
							Glyma.04g192100	Zinc finger C-x8-C-x5-C-x3-H type family protein.
Palmitic acid	15	ss715622616	T	G	2.95	Intergenic	Glyma.15g266300	Myosin-like protein XIF
							Glyma.15g266400	Leucine-rich repeat protein kinase family protein
	5	ss715592481	A	C	3.57	Intergenic	Glyma.05g015200	Predicted RNA binding protein, contains G-patch domain
							Glyma.05g015300	BUCENTAUR RELATED
							Glyma.05g015400	Amino phospholipid ATPase10 (ALA10)
							Glyma.05g015500	EF hand calcium-binding protein
	10	ss715607624	C	T	3.76	Intergenic	Glyma.10g238900	BCL-2-associated athanogene
							Glyma.10g239000	MATE efflux family protein
	12	ss715613242	A	G	3.17	3UTR	Glyma.12g078800	5\'-AMP-activated protein kinase beta-2 subunit protein
							Glyma.12g078900	Nitrate transporter 2
							Glyma.12g079000	Transmembrane protein
							Glyma.12g079100	
							Glyma.12g079200	ATPase, AAA-type, CDC48 protein
Oleic acid	8	ss715601356	T	C	2.99	Intergenic	Glyma.08g264900	Serine/Threonine-protein kinase YPK-related
	9	ss715604069	C	A	3.57	Intergenic	Glyma.09g190100	Succinate Dehydrogenase Assembly Factor 4
							Glyma.09g190400	Regulator of Vps4 activity in the MVB pathway protein
							Glyma.09190200	Transmembrane protein
							Glyma.09g190300	
	10	ss715608078	C	T	3.12	3UTR	Glyma.10g278400	Hemerythrin HHE cation-binding domain protein
							Glyma.10g278500	Ribonuclease H2 Subunit A
							Glyma.10g278600	
							Glyma.10g278700	
							Glyma.10g278800	Nucleic acid-binding, OB-fold-like protein
							Glyma.10g278900	Ferredoxin 3
	10	ss715607155	C	T	2.99	Intergenic	Glyma.10g188500	Signal transduction histidine kinase, hybrid-type, ethylene sensor
	13	ss715615744	C	T	3.04	Intergenic	Glyma.13g259000	
							Glyma.13g258900	GPI transamidase component Gpi16 subunit family protein
	15	ss715620295	G	A	3.32	Intron	Glyma.15g137600	
							Glyma.15g137700	F-box family protein
							Glyma.15g137900	Translation elongation factor EF1B/ribosomal protein S6 family protein
							Glyma.15g137800	Chaperone DnaJ-domain superfamily protein
	17	ss715627247	C	A	3.69	CDS	Glyma.17g222500	WRKY DNA-binding protein 40
							Glyma.17g222600	Formin-related
	18	ss715629786	C	T	3.73	Intergenic		
Linoleic acid	8	ss715599505	T	G	3.07	CDS	Glyma.08g174500	Leucine-rich receptor-like protein kinase family protein
							Glyma.08g174600	Glutathione S-transferase TAU 19
							Glyma.08g174700	Glutathione S-transferase TAU 19
							Glyma.08g174800	Pentatricopeptide repeat (PPR) superfamily protein
							Glyma.08g174900	Glutathione S-transferase TAU 19
							Glyma.08g175000	Glutathione S-transferase TAU 19
							Glyma.08g175100	Glutathione S-transferase TAU 19
	13	ss715613952	A	G	3.36	Intergenic	Glyma.13g027300	Uncharacterized conservative protein
Linolenic acid	2	ss715581761	T	C	3.06	Intron	Glyma.02g165400	ATP-dependent RNA helicase
	5	ss715591200	C	T	3.49	Intron	Glyma.05g167200	Calpain-type cysteine protease DEK1
							Glyma.05g167300	LETM1-like protein
	6	ss715593858	C	A	3.02	Intron	Glyma.06g209800	Actin binding protein
	10	ss715605533	G	A	3.22	Intron		
	13	ss715615506	A	G	3.13	CDS	Glyma.13g231700	Pyruvate decarboxylase 2
							Glyma.13g231800	GDSL-like Lipase/Acyl hydrolase superfamily protein
							Glyma.13g231900	Encodes a signaling peptide influencing lateral organ separation.
							Glyma.13g232000	Tetratricopeptide repeat (TPR)-like superfamily protein
							Glyma.13g232100	Tho complex subunit 7/Mft1p
							Glyma.13g232200	SGNH hydrolase-type esterase superfamily protein
							Glyma.13g232300	
	13	ss715613827	A	G	3.59	Intergenic		
	15	ss715620618	C	T	4.21	Intergenic	Glyma.15g160400	Protein binding
							Glyma.15g160600	Ankyrin repeat family protein
							Glyma.15g160500	Uncharacterized conservative protein
	16	ss715623790	A	C	2.97	Intergenic		
	17	ss715628319	G	A	2.98	Intron	Glyma.17g110700	
							Glyma.17g110800	ENTH/ANTH/VHS superfamily protein
							Glyma.17g110900	BSD domain-containing protein
							Glyma.17g111000	Vesicle-associated membrane protein 713
	17	ss715626369	T	C	3.14	Intergenic	Glyma.17g179400	Disease resistance protein (CC-NBS-LRR class) family
	17	ss715627808	T	C	4.31	5UTR	Glyma.17g255100	Salt tolerance homolog2
							Glyma.17g255200	Glycine-rich protein
							Glyma.17g255400	Alpha/beta-Hydrolases superfamily protein
							Glyma.17g255500	LSD1-like 3
							Glyma.17g255300	Chaperone Dnaj-domain superfamily protein

Chr. 10 (ss715608078 and ss715607155), Chr. 13 (ss715615744), Chr. 15 (ss715620295), Chr. 17 (ss715627247), and Chr. 18 (ss715629786). Zhao X, et al. Liu X, et al. and Silva LCC, et al. [21,36,41] reported significant SNPs for oleic acid on all chromosomes excluding Chr.5 over three, two, and one-year periods, respectively, employing GWAS and simple interval mapping. Besides, we found two SNPs for linoleic acid on Chr. 8 (ss715599505) and Chr. 13 (ss715613952) as well as 11 SNPs for linolenic acid on Chr. 2 (ss715581761), Chr. 5 (ss715591200), Chr. 6 (ss715593858), Chr. 10 (ss715605533), Chr. 13 (ss715615506 and ss715613827), Chr. 15 (ss715620618), Chr. 16 (ss715623790), and Chr. 17 (ss715628319, ss715626369, and ss715627808). Zhao X, et al. [42] conducted a GWAS in a panel of 194 soybean accessions using 3-yr data and reported SNPs for linoleic acid on all chromosomes except for Chr. 12 and for linolenic acid on Chr. 1, Chr. 2, Chr. 4, Chr. 7, Chr. 10, Chr. 12, Chr. 13, Chr. 17, Chr. 19, and Chr. 20. The locations for linoleic acid found in the present study overlapped with those reported by Zhao X, et al. [41], whereas those for linolenic acid on Chr. 5, Chr. 6, and Chr. 15 aligned with those reported by Wang X, et al. and Yao Y, et al. [13,43].

Candidate gene identification

In total, 67 genes identified using Glyma.Wm82. a2 as the reference genome are presented in Table 2. In summary, four genes for oil content were found on Chr. 4 (Glyma.04g192000, Glyma.04g192100) and Chr. 15 (Glyma.15g266300, Glyma.15g266400); 11 genes for palmitic acid concentration on Chr. 5 (Glyma.05g015200, Glyma.05g015300, Glyma.05g015400, Glyma.05g015500), Chr. 10 (Glyma.10g238900, Glyma.10g239000), and Chr. 12 (Glyma.12g078800, Glyma.12g078900, Glyma.12g079000, Glyma.12g079100, and Glyma.12g079200); 20 genes for oleic acid concentration on Chr. 8 (Glyma.08g264900), Chr. 9 (Glyma.09g190100, Glyma.09g190400, Glyma.09190200, Glyma.09g190300), Chr. 10 (Glyma.10g278400, Glyma.10g278500, Glyma.10g278600, Glyma.10g278700, Glyma.10g278800, Glyma.10g278900, Glyma.10g278500), Chr. 13 (Glyma.13g259000, Glyma.13g258900), Chr. 15 (Glyma.15g137600, Glyma.15g137700, Glyma.15g137900, Glyma.15g137800), and Chr. 17 (Glyma.17g222500 Glyma.17g222600); eight genes for linoleic acid concentration on Chr. 8 (Glyma.08g174500, Glyma.08g174600, Glyma.08g174700, Glyma.08g174800, Glyma.08g174900, Glyma.08g175000, Glyma.08g175100) and Chr. 13 (Glyma.13g027300); and 24 genes for linolenic acid concentration on Chr. 2 (Glyma.02g165400), Chr. 5 (Glyma.05g167200, and Glyma.05g167300), Chr. 6 (Glyma.06g209800), Chr. 13 (Glyma.13g231700, Glyma.13g231800, Glyma.13g231900, Glyma.13g232000, Glyma.13g232100, Glyma.13g232200, Glymag0232300), Chr. 15 (Glyma.15g160400, Glyma.15g160500, Glyma.15g160600), and Chr. 17 (Glyma.17g255300, Glyma.17g255500, Glyma.17g255400, Glyma.17g255200, Glyma.17g255100, Glyma.17g179400, Glyma.17g111000, Glyma.17g110900, Glyma.17g110800, Glyma.17g110700). Of all the significant SNPs, 53.85% (14) were located in an intergenic region, 23.07% (six) in an intron region, 11.5% (three) in an untranslated region (UTR), 7.69% (two) in a three prime untranslated region (3' UTR), and 3.84% (one) in a five prime untranslated region (5' UTR). It is worth mentioning that Glyma.04g192100 on Chr. 4 encodes a type of family protein known as zinc finger C-x8-C-x5-C-x3-H that in *Arabidopsis thaliana* is known to increase the concentration of oleic, linoleic, and linolenic acids [44]. Besides, Glyma.05g015400 on Chr. 5 encodes the amino phospholipid ATPase10, which has been reported to be associated with palmitic acid concentration in *A. thaliana* [45].

Conclusion

In this study, a GWAS was conducted for oil content and fatty acid composition using 146 soybean accessions. Twenty-six SNPs for oil content and the concentration of palmitic acid, oleic acid, linoleic acid, and linolenic acid were detected in locations coinciding with previous studies. Additionally, 67 candidate genes within 10 kb of the significant SNPs were found, among which two SNPs were located near Glyma.04g192100 on Chr. 4 and Glyma.05g015400 on Chr. 5, which have been linked to fatty acid metabolism in A. thaliana. These results can enhance marker-assisted breeding and aid in studying the molecular mechanisms underlying soybean oil content and composition.

References

1. Wang Q, Tang J, Han B, Huang X (2020) Advances in genome-wide association studies of complex traits in rice. Theor appl genet 133: 1415-1425.
2. Wilson RF (2008) Soybean: market driven research needs. Genetics and Genomics of Soybean, Springer: 3- 15.
3. Lee JD, Bilyeu KD, Shannon JG (2007) Genetics and breeding for modified fatty acid profile in soybean seed oil. J Crop Sci Biotechnol 10: 201-210.
4. Clemente TE, Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. Plant Physiol 151: 1030–1040.
5. Bu Y, Wang G, Chen Z (2005) Antitumor of transgenic soybean oil. J Chinese Biotechnol 25: 97.
6. Bahrami G, Masoumi M, Rahimi Z (2009) Co-existence of fatty acids changes in aorta artery and adipose tissue; comparison between CAD and non-CAD patients. J Thromb Thrombolysis 27: 185-190.
7. Maheshwari P, Kovalchuk I (2016) Genetic transformation of crops for oil production in industrial oil crops. AOCS Press 379-412.
8. Woyann LG, Meira D, Zdziarski AD, Matei G, Milioli AS, et al. (2019) Multiple-trait selection of soybean for biodiesel production in Brazil. Industrial Crops and Products 140: 111721.
9. Patterson E, Wall R, Fitzgerald GF, Ross RP, Stanton C, et al. (2012) Health implications of high dietary omega-6 polyunsaturated fatty acids. J Nutr Metab.
10. Guschina IA, Everard JD, Kinney AJ, Quant PA, Harwood JL, et al. (2014) Studies on the regulation of lipid biosynthesis in plants: Application of control analysis to soybean. Biochim Biophys Acta 1838: 1488-1500.
11. Qi Z, Wu Q, Han X, Sun Y, Du X, et al. (2011) Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. Euphytica 179: 499-514.
12. Akond M, Liu S, Boney M, Kantartzi SK, Meksem K, et al. (2014) Identification of quantitative trait loci (QTL) underlying protein, oil, and five major fatty acids contents in soybean. Am J Plant Sci 5: 158-167.
13. Yao Y, You Q, Duan G, Ren J, Chu S, et al. (2020) Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. BMC Plant Biology 20: 1-13.
14. Li Q, Lu X, Wang C, Shen L, Dai L, et al. (2022) Genome-wide association study and transcriptome analysis reveal new QTL and candidate genes for nitrogen-deficiency tolerance in rice. The Crop Journal 10: 942-951.
15. Liu A, Cheng SS, Yung WS, Li MW, Lam HM, et al. (2022) Genetic regulations of the oil and protein contents in soybean seeds and strategies for improvement. Adv Botanical Res 102: 259-293.
16. Diers BW, Specht JE, Graef GL, Song Q, Rainey KM, et al. (2023) Genetic architecture of protein and oil content in soybean seed and meal. Plant Genome 16: e20308.
17. Monteros MJ, Burton JW, Boerma HR (2008) Molecular mapping and confirmation of QTLs associated with oleic acid content in N00-3350 Soybean. Crop Science 48: 2223-2234.
18. Li H, Zhao T, Wang Y, Yu D, Chen S, et al. (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.

19. Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5: 107–111.

20. Wang S, Liu S, Wang J, Yokosho K, Zhou B, et al. (2020) Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. National Sci Rev 7: 1776-1786.

21. Silva LCC, Da Matta LB, Pereira GR, Bueno RD, Piovesan ND, et al. (2021) Association studies and QTL mapping for soybean oil content and composition. Euphytica 217: 24.

22. Qin J, Song Q, Shi A, Li S, Zhang M, et al. (2017) Genome-wide association mapping of resistance to Phytophthora sojae in a soybean [*Glycine max* (L.) Merr.] germplasm panel from maturity groups IV and V. PLoS One 12: e0184613.

23. Kaler AS, Ray JD, Schapaugh WT, Asebedo AR, King CA, et al. (2018) Association mapping identifies loci for canopy temperature under drought in diverse soybean genotypes. Euphytica 214: 1-18.

24. Akond M, Liu S, Schoener L, Anderson JA, Kantartzi SK, et al. (2013) A SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium BeadChip genotyping array. Plant Genetics, Genomics, and Biotechnology 1: 80-89.

25. Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, et al. (2015) Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes, Genomes, Genetics 5: 1999-2006.

26. Pritchard JK, Wen X, Falush D (2010) Documentation for structure software: Version 2.3. University of Chicago, Chicago, IL 1-37.

27. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

28. Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.

29. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

30. Churchill GA, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

31. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids Res 28: 27-30.

32. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, et al. (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13: 2129-2141.

33. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, et al. (2015) The Arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. Genesis 53: 474-485.

34. Bilyeu K, Palavalli L, Sleper D, Beuselinck P (2005) Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. Crop Sci 45: 1830-1836.

35. Kurt C (2018) Variation in oil content and fatty acid composition of sesame accessions from different origins. Grasas y aceites 69: e241-e241.

36. Liu X, Qin D, Piersanti A, Zhang Q, Miceli C, et al. (2020) Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. BMC Plant Biol 20: 399.

37. Cheesbrough TM (1989) Changes in the enzymes for fatty acid synthesis and desaturation during acclimation of developing soybean seeds to altered growth temperature. Plant Physiol 90: 760-764.

38. Wrather JA, Sleper DA, Stevens WE, Shannon JG, Wilson RF, et al. (2003) Planting date and cultivar effects on soybean yield, seed quality, and Phomopsis sp. seed infection. Plant Disease 87: 529-532.

39. Hou G, Ablett GR, Pauls KP, Rajcan I (2006) Environmental effects on fatty acid levels in soybean seed oil. J Am oil chem society 83: 759-763.

40. Byfield GE, Upchurch RG (2007) Effect of temperature on delta-9 stearoyl-ACP and microsomal omega-6 desaturase gene expression and fatty acid content in developing soybean seeds. Crop Sci 47: 1698-1704.

41. Zhao X, Chang H, Feng L, Jing Y, Teng W, et al. (2019) Genome-wide association mapping and candidate gene analysis for saturated fatty acid content in soybean seed. Plant Breeding 138: 588-598.

42. Zhao X, Jiang H, Feng L, Qu Y, Teng W, et al. (2019) Genome-wide association and transcriptional studies reveal novel genes for unsaturated fatty acid synthesis in a panel of soybean accessions. BMC Genomics 20: 1-16.

43. Wang W, Zheng H, Wang Y, Han G, Sui N, et al. (2018) Overexpression of CCCH zinc finger protein gene delays flowering time and enhances salt tolerance in Arabidopsis by increasing fatty acid unsaturation. Acta Physiologiae Plantarum 40: 1.

44. Wang X, Jiang GL, Green M, Scott RA, Hyten DL, et al. (2014) Quantitative trait locus analysis of unsaturated fatty acids in a recombinant inbred population of soybean. Molecular Breeding 33: 281-296.

45. Li YH, Reif JC, Ma YS, Hong HL, Liu ZX, et al. (2015) Targeted association mapping demonstrating the complex molecular genetics of fatty acid formation in soybean. BMC Genomics 16: 1-13.