

RESEARCH

Open Access



Multi-omics analysis reveals novel loci and a candidate regulatory gene of unsaturated fatty acids in soybean (*Glycine max* (L.) Merr)

Xunchao Zhao¹, Yuhang Zhan¹, Kaiming Li¹, Yan Zhang¹, Changjun Zhou², Ming Yuan³, Miao Liu⁴, Yongguang Li¹, Peng Zuo^{1*}, Yingpeng Han^{1*} and Xue Zhao^{1*}

Abstract

Background Soybean is a major oil crop; the nutritional components of soybean oil are mainly controlled by unsaturated fatty acids (FA). Unsaturated FAs mainly include oleic acid (OA, 18:1), linoleic acid (LLA, 18:2), and linolenic acid (LNA, 18:3). The genetic architecture of unsaturated FAs in soybean seeds has not been fully elucidated, although many independent studies have been conducted. A 3 V multi-locus random single nucleotide polymorphism (SNP)-effect mixed linear model (3VmrMLM) was established to identify quantitative trait loci (QTLs) and QTL-by-environment interactions (QEIs) for complex traits.

Results In this study, 194 soybean accessions with 36,981 SNPs were calculated using the 3VmrMLM model. As a result, 94 quantitative trait nucleotides (QTNs) and 19 QEIs were detected using single-environment (QTN) and multi-environment (QEI) methods. Three significant QEIs, namely rs4633292, rs39216169, and rs14264702, overlapped with a significant single-environment QTN.

Conclusions For QTNs and QEIs, further haplotype analysis of candidate genes revealed that the *Glyma.03G040400* and *Glyma.17G236700* genes were beneficial haplotypes that may be associated with unsaturated FAs. This result provides ideas for the identification of soybean lipid-related genes and provides insights for breeding high oil soybean varieties in the future.

Keywords Soybean, 3VmrMLM, Unsaturated fatty acids, Multi-omics analysis, Expression profile

*Correspondence:

Peng Zuo
634905@qq.com
Yingpeng Han
hyp234286@aliyun.com
Xue Zhao
xuezhao@neau.edu.cn

¹ Key Laboratory of Soybean Biology in Chinese Ministry of Education (Key Laboratory of Soybean Biology and Breeding/Genetics of Chinese Agriculture Ministry), Northeast Agricultural University, Harbin 150030, China

² Daqing Branch, Heilongjiang Academy of Agricultural Science, Daqing, China

³ Qiqihar Branch, Heilongjiang Academy of Agricultural Science, Qiqihar, China

⁴ Crop Tillage and Cultivation Institute, Heilongjiang Academy of Agricultural Sciences, Harbin, China

Background

Soybean [*Glycine max* (L.) Merr.], a major oil crop, is commonly used in cooking oil [1]. Soybean oil is mainly composed of saturated fatty acids (FAs) and unsaturated FAs. Among them, saturated FAs include palmitic and stearic acids, and unsaturated FAs include oleic (OA), linoleic acid (LLA), and linolenic acids (LNA) [2, 3]. Unsaturated FA is the main component of vegetable oil, accounting for more than 80% [4]. The increase in the content of OA, a monounsaturated FA, can improve oxidative stability and prevent oxidation [4]. LLA and LNA are polyunsaturated FAs and are very beneficial to human health [5]. However, the LLA and LNA show poor stability at a high temperature and are easily oxidized



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

[6]. Thus, an important goal of soybean breeders is to increase the OA level and reduce the LLA and LNA content [7, 8].

Genome-wide association study (GWAS) mapping can identify the genetic basis of a variety of complex traits [9]. To date, the single-locus GWAS method has been widely applied to mine genetic loci underlying important agronomic traits, including 100-seed weight in soybean and oil content and yield-related traits in maize [2, 10, 11]. However, quantitative trait nucleotides (QTNs) have been detected using the single-locus GWAS method, which has limited ability to detect QTNs because quantitative traits are affected by a polygenic background [12].

Currently, the mixed linear model (MLM) method can correct population structure and family relationships and is widely used [13]. Based on the MLM method, single-locus GWAS methods have been widely proposed, including EMMAX, FaST-LMM and GEMMA [12, 14, 15]. However, single-locus GWAS methods generally require Bonferroni correction and can be affected by a polygenic background. To overcome this problem in single-locus GWAS methods, multi-locus GWAS methods have been applied, in which statistics are applied to all loci [16]. These multi-locus GWAS methods mainly include FASTmrEMMA, FASTmrMLM, FarmCPU, and pLARmEB [17–20]. However, these methods have a high calculation burden, and the advantages of QTN-by-environment interactions (QEIs) have not been fully considered.

To address this, a new multi-locus GWAS model, the 3 V multi-locus random-SNP-effect mixed linear model (3VmrMLM), has been presented [21]. This method improves the QTL detection capability and can analyze the genetic variation of complex traits. It provides a new method for the gene discovery of complex traits.

In this study, a dataset of 194 soybean accessions with 36,981 SNPs was applied [2]. We analyzed the unsaturated FA content in this population of 194 soybean accessions based on the multi-locus GWAS model (3VmrMLM). Our aim was to detect significant QEIs and stable QTNs compared with the results of our previous study and other independent studies and to further identify candidate genes related to unsaturated FA content.

Results

Phenotypic variation of three soybean unsaturated FA compositions

The distribution of unsaturated FA content (including OA, LLA, and LNA) in the 194 soybean accessions is shown in Table 1. The coefficient of variation (CV%) differed among the three years. In 2013, the unsaturated FA content had the highest CV at 51% (OA), 48% (LLA), and 52% (LNA). In 2014, the CVs of OA, LLA, and LNA were relatively consistent at 28%, 25%, and 29%, respectively. In 2015, the CV of unsaturated FAs was basically the same as in 2014 (Table 1). The heritabilities of OA, LLA, and LNA were 0.41, 0.36, and 0.35, respectively (Table 1). The above results showed that the content of unsaturated FAs was affected by the environment.

The correlation coefficient of the unsaturated FA content was calculated. As shown in Fig. 1, OA, LLA, and LNA content had a high correlation within the same year. However, the OA, LLA, and LNA content was not high between different years. In 2013, OA was positively correlated with LLA and LNA (0.92 and 0.83, respectively). In 2014, OA was positively correlated with LLA and LNA (0.84 and 0.65, respectively). In 2015, OA was positively correlated with LLA and LNA (0.79 and 0.64, respectively). These results show that unsaturated FAs affect soybean oil accumulation.

Table 1 Statistical analysis of oleic, linoleic, and linolenic acid traits

Trait	Years	Min	Max	Mean	SD	CV (%)	Heritability
Oleic acid content	2013	10.96	29.91	17.46	8.89	51	0.41
	2014	15.26	31.51	20.09	5.73	28	
	2015	14.86	34.81	18.19	4.9	26	
Linoleic acid content	2013	47.08	63.29	44.52	21.76	48	0.36
	2014	47.36	58.6	50.89	13.21	25	
	2015	44.06	60.68	53.35	12.59	23	
Linolenic acid content	2013	4.4	12.91	6.38	3.34	52	0.35
	2014	4.44	11.46	7.62	2.22	29	
	2015	4.23	13.34	8.72	2.29	26	

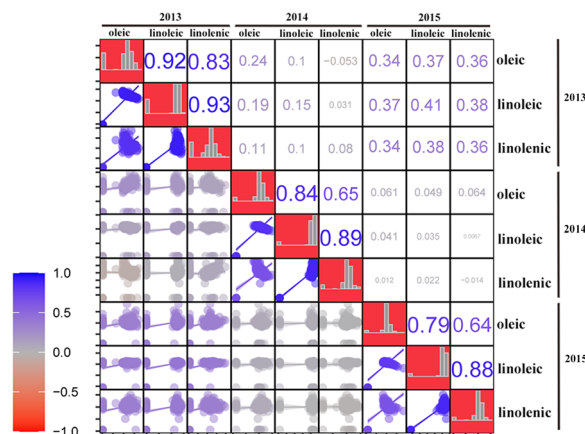


Fig. 1 Distribution of oleic, linoleic, and linolenic traits in soybean and Pearson coefficients

Identification of QTNs for unsaturated FA-related traits using 3VmrMLM

In this study, the unsaturated FA content was reanalyzed using the single-environment QTN model (3VmrMLM). A total of 94 significant QTNs were associated with the unsaturated FA content (LOD score ≥ 3.0). Among them, 30, 34, and 30 QTNs were associated with OA, LLA, and LNA content, respectively (Table 2).

In 2013, 2014, and 2015, 12, 10, and eight QTNs were associated with OA content, with LOD scores of 4.55–25.92, 6.25–23.29, and 3.37–9.57, respectively. A total of 17, 10, and seven QTNs associated with LLA content were identified with LOD scores of 4.43–24.14, 3.42–20.76, and 4.98–17.08 in 2013, 2014, and 2015, respectively. In 3 years (2013, 2014, and 2015), 10, eight, and 12 QTNs associated with LNA content were detected with LOD scores of 4.17–22.53, 5.07–10.72, and 6.46–20.27, respectively (Table 2, Additional file 1: Fig. S1).

Detection of QEIs for unsaturated FA content using 3VmrMLM with multiple environments

The unsaturated FA content was reanalyzed in 3 years (2013, 2014, and 2015) using the multiple-environment QEI model (3VmrMLM) for identifying QEIs. A total of 19 significant/suggested QEIs were identified (Table 3, Fig. 2). Three significant QEIs overlapped with the above QTNs. In these QEIs, the r^2 value was between 2.01 and 14.67, and the variance value was between 0.03 and 1.33 (Table 3).

Candidate gene prediction of significant QTNs associated with unsaturated FA in soybean

There were 1246 genes identified in the flanking genomic region of each significant QTN using the 3VmrMLM method (Additional file 1: Table S1). We

further conducted the Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. As shown in Additional file 1: Fig. S2A, 201 genes were significantly enriched in metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems, including lipid metabolism, amino acid metabolism, energy metabolism, transport, and catabolism. The results of the above enrichment analysis showed that some candidate genes around QTN were found in different processes.

The same methods mentioned above were used to analyze candidate genes in the flanking regions of the QEIs. A total of 301 candidate genes were found in the linked regions of significant QEIs (Additional file 1: Table S2). KEGG analysis found that 53 genes were significantly enriched in metabolism, genetic information processing, environmental information processing, and organismal systems, including carbohydrate metabolism and lipid metabolism (Additional file 1: Fig. S2B). In the multiple-environment QEI model, five known SNP markers were identified. In addition, some new SNP markers, including rs6528670, rs1902760, rs4633292, rs2457629, and rs48948953, were related to FA synthesis. Moreover, some known markers were identified in the multiple-environment QEI model, including rs14264702, rs34595703, rs44492166, rs23852645, and rs26951255. Three significant QEIs, namely rs4633292, rs39216169, and rs14264702, overlapped with significant QTN in a single year, of which rs14264702 has been reported [22].

Transcriptomic analysis of HUFA and LUFA soybean seeds

RNA-seq analysis was conducted to reveal the transcriptional regulation of unsaturated FA metabolism in HUFA (high unsaturated fatty acid) and LUFA (low unsaturated fatty acid) soybean seeds. Three comparison groups were analyzed: a comparison group of five HUFA and five LUFA varieties (FHUFA vs. FLUFA); a comparison of 10 HUFA and 10 LUFA varieties (THUFA vs. TLUFA); and a comparison of 15 HUFA and 15 LUFA varieties (HUFA vs. LUFA) (Additional file 1: Table S3).

There were 4013, 3504, and 2546 DEGs in the FHUFA vs. FLUFA, THUFA vs. TLUFA, and HUFA vs. LUFA groups, respectively (Fig. 3A, B). In each comparison group, the number of upregulated DEGs was higher than the number of downregulated DEGs. As shown in Fig. 3C, 1160 common DEGs were upregulated, while 183 common DEGs were downregulated.

Identification of candidate genes by integrating GWAS and RNA-seq analysis

To further identify candidate genes, DEGs were identified by integrating GWAS and RNA-seq analysis and by analyzing potential candidate genes. In

Table 2 QTNs identified for unsaturated fatty acid content using the QTN detection model in 3VmrMLM

Trait	Year	SNP	Chr	Position (bp)	LOD	Additive effect	Dominant effect	Variance	r ² (%)	p-value	Known QTL	References	
Oleic	2013	rs25094202	2	25094202	25.92	-1.53	-0.54	0.35	3.88	1.2E-26			
		rs6124789	4	6124789	19.04	-1.22	-0.43	0.31	3.45	9E-20			
		rs40652518	4	40652518	19.99	-1.23		0.76	8.48	8.5E-22			
		2014	rs14903097	6	14903097	6.23	-0.39	-1.78	0.14	1.59	5.9E-07		
	rs29444284		6	29444284	11.36	0.68	-2.62	0.58	6.50	4.4E-12			
	rs32503987		6	32503987	17.07	-1.11		0.57	6.41	7.6E-19			
		2015	rs545702	7	545702	28.94	1.61	1.24	0.84	9.37	1.2E-29		
	rs29716728		7	29716728	3.32	0.44	-0.06	0.17	1.96	0.00048			
	rs19742379		8	19742379	11.13	0.01	7.58	0.72	8.10	7.5E-12			
	rs2762447		9	2762447	13.18	-0.17	-3.02	0.70	7.87	6.6E-14			
	rs18612544		9	18612544	4.55	0.04	3.75	0.26	2.95	2.8E-05			
	rs15583833		11	15583833	6.67	0.64		0.28	3.12	3E-08			
	rs25962449		3	25962449	13.08	-0.98	-0.06	0.21	3.10	8.3E-14			
	2014	rs9480030	4	9480030	23.29	1.37		0.74	10.93	3.9E-25			
rs24966112		4	24966112	10.53	-0.91	0.26	0.22	3.23	2.9E-11				
rs37372282		7	37372282	6.50	-0.65		0.27	3.95	4.5E-08				
	2015	rs13246407	8	13246407	12.51	-0.79	4.87	0.45	6.75	3.1E-13			
rs4633292		9	4633292	6.26	0.63	0.60	0.36	5.41	5.5E-07				
rs7608184		11	7608184	7.75	-0.69	2.38	0.33	4.96	1.8E-08				
	2015	rs86980	14	86980	6.96	0.67		0.36	5.27	1.5E-08			
rs21488603		16	21488603	7.07	-0.57	1.80	0.35	5.25	8.6E-08				
rs42357728		19	42357728	9.79	0.78	2.22	0.23	3.48	1.6E-10				
	2015	rs14131618	2	14131618	9.20	0.62	5.83	0.37	5.75	6.3E-10			
rs648720		3	648720	7.62	-0.21	7.18	0.59	9.19	2.4E-08				
rs1225280		12	1225280	4.41	-0.58	-0.73	0.32	4.92	3.9E-05				
	2015	rs16802809	13	16802809	8.65	-0.83		0.30	4.73	2.8E-10			
rs19292381		13	19292381	9.58	-0.47	5.11	0.70	10.91	2.7E-10				
rs30304375		13	30304375	3.38	0.07	3.91	0.25	3.81	0.00042				

Table 2 (continued)

Trait	Year	SNP	Chr	Position (bp)	LOD	Additive effect	Dominant effect	Variance	r ² (%)	p-value	Known QTL	References	
Linoleic	2013	rs23631249	17	23631249	6.54	0.72		0.45	6.94	4.1E-08	Seed oleic 6-9	[22]	
		rs37263037	17	37263037	7.38	0.74	-1.42		0.28	4.36	4.2E-08	Seed oleic 6-9	[22]
		rs39587481	2	39587481	16.48	-0.83	-0.77		0.15	1.93	3.3E-17		
			rs28569907	3	28569907	24.14	-1.09	-0.06	0.26	3.36	7.3E-25		
			rs41402037	4	41402037	12.32	0.70	0.45	0.37	4.68	4.8E-13		
			rs48166843	4	48166843	11.78	-0.68	0.44	0.17	2.22	1.7E-12		
			rs4602961	9	4602961	12.39	0.69	-1.08	0.45	5.71	4.1E-13		
			rs11516441	9	11516441	15.27	-0.80		0.37	4.75	5E-17		
			rs38142875	9	38142875	12.93	0.54	-4.23	0.31	3.99	1.2E-13		
			rs9167160	10	9167160	8.83	0.58		0.32	4.05	1.8E-10		
			rs46493899	10	46493899	7.06	0.47	-1.85	0.25	3.26	8.8E-08	Seed oleic 6-10	[22]
			rs34135970	12	34135970	11.07	-0.36	2.82	0.42	5.43	8.6E-12		
			rs1501986	14	1501986	6.23	0.47		0.20	2.59	8.6E-08		
			rs48109369	14	48109369	5.18	-0.07	2.65	0.18	2.35	6.7E-06		
			rs6755025	15	6755025	5.60	-0.44	-0.98	0.17	2.16	2.5E-06		
			rs32005810	16	32005810	4.43	0.39	-0.66	0.14	1.80	3.7E-05		
			rs30412910	17	30412910	11.47	0.64	-2.71	0.40	5.17	3.4E-12		
			rs9070866	18	9070866	6.15	0.28	-3.37	0.18	2.30	7.1E-07		
			rs4876597	19	4876597	7.75	0.22	1.86	0.18	2.28	1.8E-08		
2014		rs4953186	2	4953186	20.77	-1.01	0.66	0.24	6.07	1.7E-21	rs4953186	[2]	
		rs14587513	2	14587513	5.80	0.47		0.22	5.71	2.4E-07			
		rs6481810	3	6481810	17.03	0.86	2.05	0.36	9.28	9.4E-18	rs6481810	[2]	
		rs24966112	4	24966112	18.84	1.00	-0.08	0.20	5.21	1.5E-19			
		rs17451809	6	17451809	3.43	-0.36		0.12	3.16	7.1E-05			
		rs14264702	7	14264702	14.81	-0.80		0.20	5.28	1.5E-16			
		rs13246407	8	13246407	4.89	0.28	-3.14	0.14	3.59	1.3E-05			
		rs37217687	12	37217687	3.61	-0.22	-2.31	0.12	2.99	0.00025			

Table 2 (continued)

Trait	Year	SNP	Chr	Position (bp)	LOD	Additive effect	Dominant effect	Variance	r ² (%)	p-value	Known QTL	References
Linolenic	2015	rs9446804	15	9446804	6.72	-0.12	-2.23	0.22	5.54	1.9E-07	Seed oleic 4-1	[36]
		rs39216169	17	39216169	5.23	-0.45		0.12	3.11	9.2E-07		
		rs25887810	1	25887810	17.08	1.04		0.26	6.29	7.3E-19		
		rs46294440	6	46294440	4.98	-0.51	-0.85	0.27	6.39	1E-05		
		rs36974795	12	36974795	7.52	0.46	-2.09	0.33	7.85	3E-08		
		rs16802809	13	16802809	7.43	0.64		0.18	4.33	4.9E-09		
		rs25691662	13	25691662	8.76	-0.27	-6.21	0.39	9.48	1.7E-09		
		rs21353920	15	21353920	17.96	1.07	0.40	0.28	6.83	1.1E-18		
		rs14234658	18	14234658	6.51	-0.18	-3.45	0.29	6.96	3.1E-07		
		rs7344385	2	7344385	9.69	0.44		0.13	6.80	2.4E-11		
		rs5009451	3	5009451	7.39	-0.38		0.09	4.47	5.4E-09		
		rs15602593	4	15602593	18.65	0.64	-0.80	0.14	6.99	2.2E-19		
	rs35874955	4	35874955	4.18	-0.13	1.36	0.07	3.69	6.6E-05			
	rs45800101	14	45800101	6.16	-0.34	0.01	0.08	4.13	6.9E-07			
	rs30575841	15	30575841	7.74	0.13	1.86	0.12	5.96	1.8E-08			
	rs4718078	17	4718078	22.54	-0.74		0.15	7.90	2.3E-24			
	rs41437124	17	41437124	4.82	0.06	2.10	0.08	4.14	1.5E-05			
	rs47611415	18	47611415	7.11	0.29	-1.06	0.10	5.10	7.7E-08			
rs18173239	20	18173239	7.13	0.37	-0.68	0.05	2.39	7.5E-08				
rs422280985	4	422280985	7.76	0.03	1.43	0.08	6.95	1.8E-08				
rs42904836	4	42904836	9.14	0.33	0.27	0.03	2.59	7.3E-10				
rs4633296	9	4633296	10.72	-0.36	0.83	0.12	10.00	1.9E-11				
rs21698230	14	21698230	5.68	-0.25		0.04	3.36	3.2E-07	Seed oleic 7-3	[25]		
rs5443771	16	5443771	5.07	-0.24	0.31	0.06	4.72	8.5E-06				
rs39098643	17	39098643	6.31	-0.28	-0.15	0.06	5.16	4.9E-07				
rs41268049	17	41268049	5.73	0.10	-1.02	0.06	5.41	1.8E-06				
rs52833743	18	52833743	10.50	-0.36		0.06	5.04	3.5E-12		rs52833743; Seed oleic 3-3; Seed oleic 5-2	[2, 24]	
rs6099191	2	6099191	8.39	0.14	2.49	0.06	5.41	4.1E-09				

Table 2 (continued)

Trait	Year	SNP	Chr	Position (bp)	LOD	Additive effect	Dominant effect	Variance	r ² (%)	p-value	Known QTL	References
		rs5660143	3	5660143	9.24	0.12	2.23	0.08	6.82	5.8E-10		
		rs32069131	6	32069131	20.28	0.51	-0.14	0.06	4.86	5.3E-21		
		rs38597109	8	38597109	18.99	-0.47		0.07	5.93	8.6E-21		
		rs18435729	10	18435729	6.46	-0.01	2.43	0.06	5.37	3.5E-07		
		rs39609078	12	39609078	6.66	-0.26		0.05	4.29	3.1E-08		
		rs2599191	14	2599191	10.61	0.33	-0.06	0.09	7.78	2.5E-11		
		rs9834940	14	9834940	9.71	0.32	-0.21	0.04	3.58	2E-10		
		rs26951252	14	26951252	10.54	0.30	-0.66	0.08	6.90	2.9E-11	Seed oleic 6-10; Seed oleic 7-3	[22, 25]
		rs21822969	15	21822969	7.38	0.05	2.56	0.07	5.98	4.2E-08		
		rs30575841	15	30575841	15.05	-0.40	0.58	0.07	5.56	9E-16		
		rs35024325	17	35024325	12.77	0.37	-0.18	0.03	2.78	1.7E-13	rs35024325; Seed oleic 6-9	[2, 22]

the single-environment QTN model, 91, 85, and 61 DEGs were found in the FHUFA vs. FLUFA, THUFA vs. TLUFA, and HUFA vs. LUFA groups, respectively (Additional file 1: Table S1). A total of 30 DEGs were found in all three comparison groups. Among them, *Glyma.10G079500*, *Glyma.19G163600*, *Glyma.09G033600*, and *Glyma.02G068900* genes were upregulated (Log2FC > 3), and *Glyma.07G205400*, *Glyma.09G053700*, and *Glyma.06G175100* genes were downregulated (Log2FC < -1) (Table 4, Additional file 1: Fig. S3A).

In the multiple-environment QEI model, 26, 28, and 20 DEGs were found in the FHUFA vs. FLUFA, THUFA vs. TLUFA, and HUFA vs. LUFA groups, respectively (Additional file 1: Table S2). Among these candidate genes, nine were simultaneously detected by GWAS and common DEGs in all three comparison groups. These nine genes included those encoding an Acyl-CoA-binding protein (*Glyma.17G236700*), Ankyrin repeat family protein (*Glyma.09G053700*), Nodulin MtN3 family protein (*Glyma.08G025100*), Integrase-type DNA-binding superfamily protein (*Glyma.18G206600*), Calmodulin-domain protein kinase 9 (*Glyma.14G023500*), ARM repeat superfamily protein (*Glyma.03G036700*), Protein kinase superfamily protein (*Glyma.03G036000*), BRI1 kinase inhibitor 1 (*Glyma.06G039100*), NAC domain-containing protein 73 (*Glyma.13G234700*), and unknown function protein (*Glyma.18G205700*, *Glyma.18G205400*) (Table 5, Additional file 1: Fig. S3B). The expression of these genes was further determined by qRT-PCR and was basically consistent with that of the transcriptome data (Additional file 1: Fig. S4).

Metabolic profiling analysis of MHUFA and MLUFA soybean seeds

To determine the unsaturated FA regulatory network at the seed development stage, a non-targeted metabolic profiling analysis was applied. There were 15 high unsaturated FA (HUFA) and 15 low unsaturated FA (LUFA) soybean varieties applied in this study (Additional file 1: Table S3). Multiple metabolites were detected using non-targeted metabolomics, including secondary metabolites, lipids, amino acids, and flavonoids.

To explore the differences in metabolites between different varieties, three comparison groups were studied: five high-unsaturated FA (FMHUFA) and five low-unsaturated FA (FMLUFA) varieties (FMHUFA vs. FMLUFA); 10 high-unsaturated FA (TMHUFA) and 10 low-unsaturated FA (TMLUFA) varieties (TMHUFA vs. TMLUFA); and 15 high-unsaturated FA (MHUFA) and 15 low-unsaturated FA (MLUFA) varieties (MHUFA vs. MLUFA). The OPLS-DA analysis showed that the model accurately

Table 3 Significant/suggested QELs for soybean unsaturated fatty acid content in three environments detected using the QTN-by-environment detection model in 3YmrMLM

Trait	SNP	Chr	Position (bp)	LOD (QE)	add* env1	dom* env1	add* env2	dom* env2	add* env3	dom* env3	variance	r ² (%)	p-value	Known QTL	References
Oleic	rs6528670	7	6528670	11.36	0.58	-4.19	0.08	0.21	-0.66	3.97	0.43	4.71	1.10E-10		
Oleic	rs1902760	8	1902760	13.68	0.66	4.37	-0.45	0.46	-0.22	-4.83	0.55	6.09	6.70E-13		
Oleic	rs4633292	9	4633292	5.72	0.01	0.19	0.53	-1.73	-0.54	1.55	0.21	2.31	2.60E-05		
Oleic	rs39216169	17	39216169	11.67	0.20		0.69		-0.89		0.44	4.82	2.10E-12		
Oleic	rs31881460	19	31881460	17.04	-0.52	-0.57	-0.63	-0.33	1.15	0.91	0.66	7.32	3.60E-16		
Oleic	rs44492166	19	44492166	30.20	-1.49		0.18		1.31		1.33	14.67	6.30E-31		
Linoleic	rs12727785	5	12727785	11.91	-0.82	-0.76	0.04	1.00	0.78	-0.24	0.43	6.28	3.40E-11		
Linoleic	rs14264702	7	14264702	11.60	0.04		-0.78		0.74		0.39	5.68	2.40E-12	Seed oleic 6-7	[22]
Linoleic	rs34595703	13	34595703	14.71	-0.94	0.03	0.08	-0.16	0.86	0.13	0.54	7.87	6.80E-14	Seed oleic 6-4	[22]
Linoleic	rs2457629	14	2457629	10.39	0.25		0.56		-0.81		0.35	5.09	4.10E-11		
Linoleic	rs44492166	19	44492166	22.99	1.17		-0.05		-1.12		0.87	12.75	1.01E-23	Seed linoleic 2-2; Seed linolenic 2-3	[26]
Linolenic	rs8152225	1	8152225	26.67	0.39	0.39	0.23	0.12	-0.61	-0.51	0.19	10.79	1.30E-25		
Linolenic	rs23852645	2	23852645	16.14	-0.10	0.98	-0.33	-1.08	0.43	0.11	0.11	6.04	2.70E-15	Seed linoleic 7-1	[25]
Linolenic	rs4406682	3	4406682	5.22	-0.24		0.04		0.21		0.03	2.01	6.09E-06		
Linolenic	rs2988645	6	2988645	5.95	-0.03	2.02	-0.07	0.16	0.10	-2.17	0.04	2.21	1.60E-05		
Linolenic	rs1701848	14	1701848	17.78	-0.40		-0.05		0.45		0.12	7.09	1.60E-18		
Linolenic	rs26951255	14	26951255	19.22	0.35	-1.22	0.13	1.19	-0.48	0.03	0.13	7.64	2.70E-18	Seed linolenic 10-1; Seed linolenic 7-6	[22]
Linolenic	rs42367957	18	42367957	8.55	-0.09	-1.31	0.00	-0.79	0.10	2.10	0.05	3.06	5.80E-08		
Linolenic	rs48948953	18	48948953	15.02	0.11		-0.43		0.32		0.10	5.60	9.60E-16		

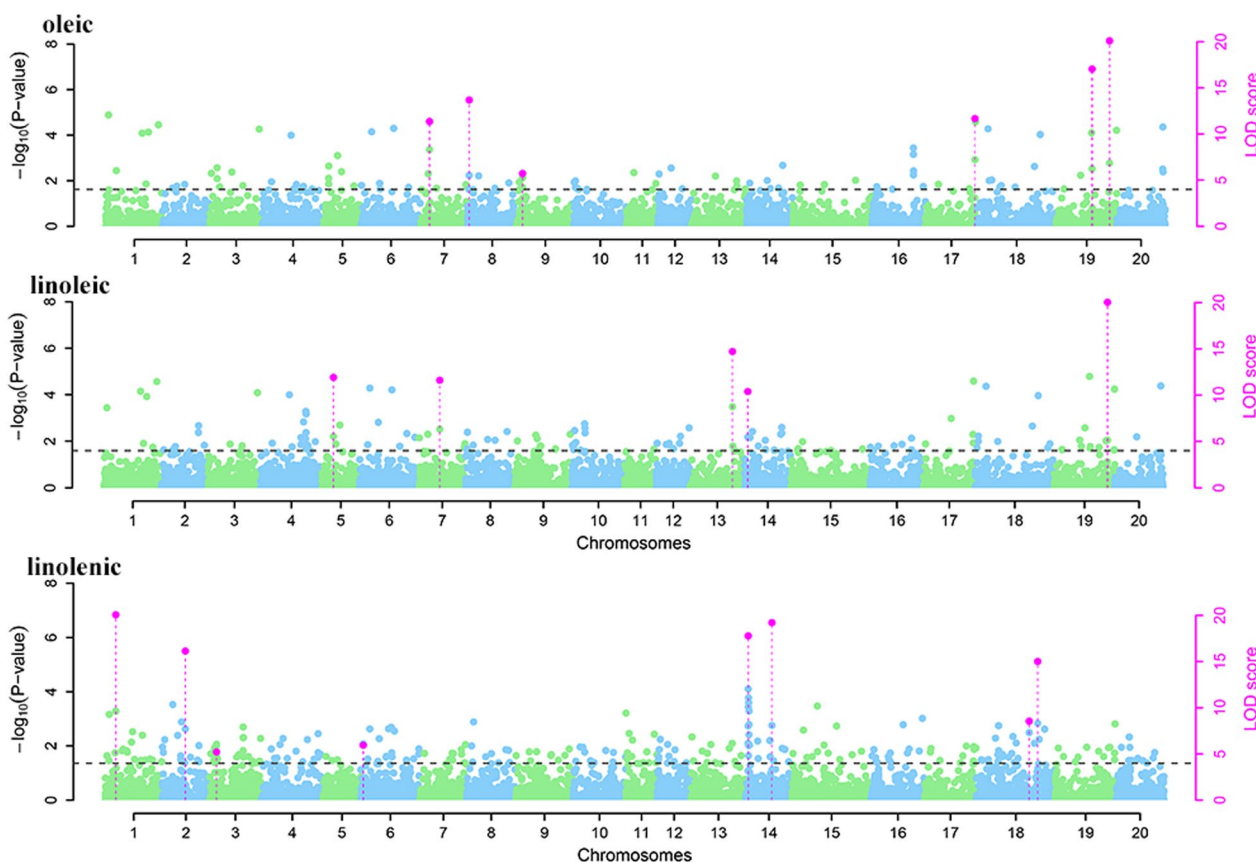


Fig. 2 Manhattan plots of the multi-environment analysis for the oleic, linoleic, and linolenic acid content in soybean seeds

described each sample and was suitable for subsequent analysis (Fig. 4A). According to the OPLS-DA model, 70 differentially abundant metabolites (DAMs) were upregulated, and 291 DAMs were downregulated in FMHUFA vs. FMLUFA (Fig. 4B). These metabolites included lipids, secondary metabolites, and unknown metabolites. In addition, 202 and 322 DAMs were identified in TMHUFA vs. TMLUFA and MHUFA vs. MLUFA, respectively (Fig. 4B). As shown in Fig. 4C, four common upregulated DAMs and 29 common downregulated DAMs were identified.

Differential accumulation of metabolites with MHUFA and MLUFA content

In this study, the metabolic changes of high and low unsaturated FA content in 30 soybean varieties during the R6 period were studied. In FMHUFA vs. FMLUFA, 29 DAMs were annotated into the KEGG pathway. Among them, the isoflavone pathway had the most DAMs, including Genistein, 8-C-glucosylnaringenin, genistin, and biochanin A (Additional file 1: Fig. S5A). In TMHUFA vs. TMLUFA, 16 DAMs were annotated into the KEGG pathway; among them, the TCA

cycle had the most DAMs. 1-Pyrroline-4-hydroxy-2-carboxylate, 5-amino-6-ribitylamino uracil, and 2-(acetamidomethylene) succinate were differentially accumulated in TMHUFA vs. TMLUFA (Additional file 1: Fig. S5B). In MHUFA vs. MLUFA, 39 DAMs were annotated into the KEGG pathway, including the TCA cycle, LLA metabolism, and biosynthesis of amino acids. LysoPC (22:2(13Z,16Z)), (2S,5S)-trans-carboxymethylproline, and quercetin 3-sambubioside were differentially accumulated in MHUFA vs. MLUFA (Additional file 1: Fig. S5C).

Co-expression analysis of candidate genes and DAM metabolites

Candidate genes and metabolite networks were analyzed. In the single-environment model, the co-expression network of 30 candidate genes and DAMs in three comparison groups was constructed. In the FHUFA vs. FLUFA network, the results indicated that the 75 subnetworks were significantly correlated ($|r| > 0.5$, $p < 0.05$). PE (22:5) was positively associated with *Glyma.14G216100* ($r > 0.51$, $p < 0.02$), *Glyma.03G040000* ($r > 0.51$, $p < 0.01$), *Glyma.02G210300*

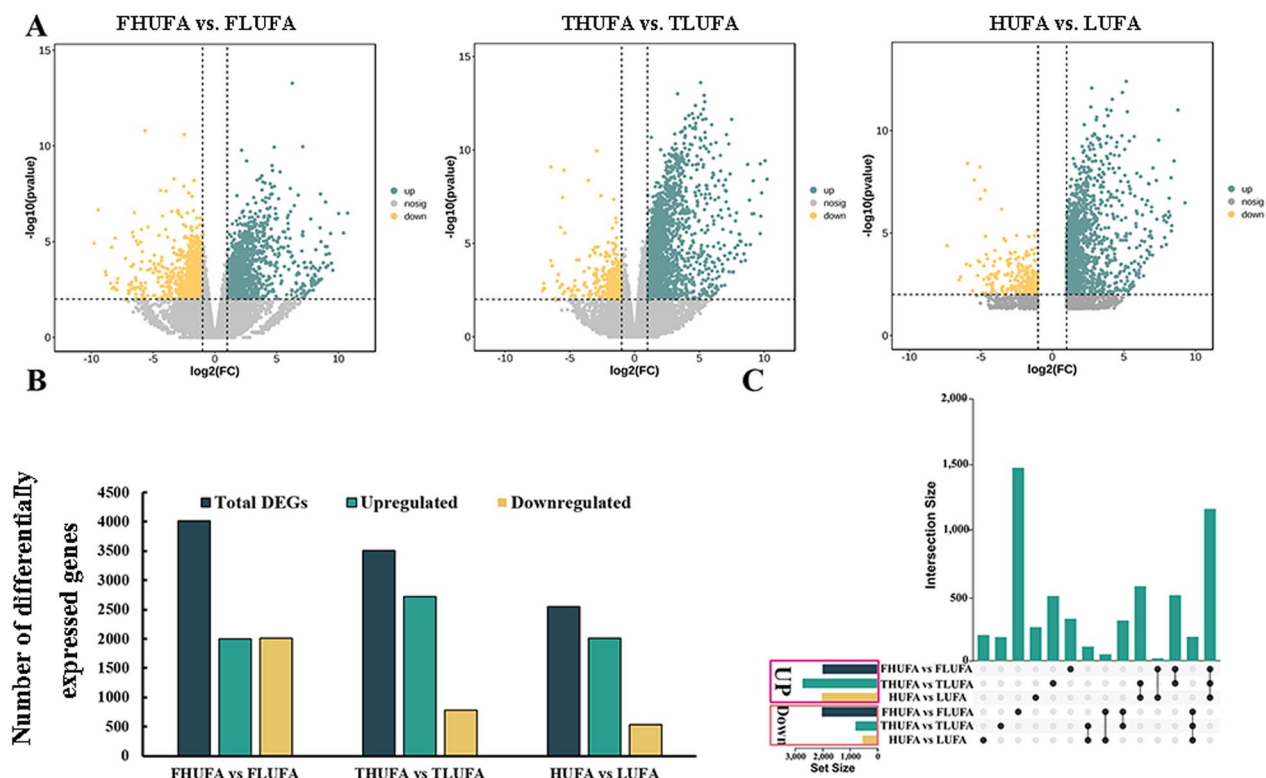


Fig. 3 Multivariate statistical analysis of the transcriptome data in the soybean samples. **A** Volcano maps in different comparison groups. **B** Number of differentially expressed genes. The green and yellow columns represent the numbers of genes with upregulated and downregulated expression, respectively. **C** Upsetplot diagram showing the overlapping DEGs in the three comparison groups

($r > 0.51$, $p < 0.02$), *Glyma.17G236200* ($r > 0.52$, $p < 0.01$), *Glyma.11G100600* ($r > 0.59$, $p < 0.006$), and *Glyma.09G157500* ($r > 0.68$, $p < 0.0007$). Quinone was positively associated with *Glyma.17G236700* ($r > 0.53$, $p < 0.01$) and *Glyma.03G040400* ($r > 0.67$, $p < 0.001$) (Fig. 5A). In the THUFA vs. TLUFA network, 1-pyrroline-4-hydroxy-2-carboxylate was positively associated with *Glyma.09G051900* ($r > 0.88$, $p < 5.53E-14$) and *Glyma.03040400* ($r > 0.88$, $p < 2.35E-14$). *Glyma.17G236700* was positively associated with 2-(acetamidomethylene) succinate ($r > 0.51$, $p < 0.0007$), 5-amino-6-ribitylamino uracil ($r > 0.69$, $p < 6.85E-07$), and 1-pyrroline-4-hydroxy-2-carboxylate ($r > 0.69$, $p < 6.09E-07$) (Fig. 5B). In the HUFUFA vs. LUFUFA network, 1-pyrroline-4-hydroxy-2-carboxylate B was significantly associated with *Glyma.03G040400* ($r > 0.84$, $p < 3.69E-17$) and *Glyma.09G051900* ($r > 0.85$, $p < 8.61E-18$). *Glyma.17G236700* was significantly associated with LysoPC(22:2(13Z,16Z)) ($r > 0.64$, $p < 3.04E-08$) and 1-pyrroline-4-hydroxy-2-carboxylate B ($r > 0.64$, $p < 2.01E-08$) (Fig. 5C).

In the multiple-environment QEI model, a co-expression network of nine candidate genes and DAMs was constructed for three comparison groups. In the

FHUFA vs. FLUFA network, the 22 subnetworks were significantly correlated ($|r| > 0.5$, $p < 0.05$). Quinone was positively associated with *Glyma.08G025100* ($r > 0.88$, $p < 2.14E-07$) and *Glyma.17G236700* ($r > 0.53$, $p < 0.01$) (Fig. 6A). In the THUFA vs. TLUFA network, *Glyma.18G205400* was positively associated with 5-amino-6-ribitylamino uracil ($r > 0.90$, $p < 9.27E-16$) and 1-Pyrroline-4-hydroxy-2-carboxylate ($r > 0.92$, $p < 2.96E-17$). *Glyma.17G236700* was significantly associated with 2-(acetamidomethylene) succinate ($r > 0.51$, $p < 0.0007$), 1-pyrroline-4-hydroxy-2-carboxylate ($r > 0.69$, $p < 6.09E-07$), and 5-Amino-6-ribitylamino uracil ($r > 0.69$, $p < 6.85E-07$) (Fig. 6B). In the HUFUFA vs. LUFUFA network, *Glyma.17G236700* was significantly associated with LysoPC (22:2(13Z,16Z)) ($r > 0.64$, $p < 3.04E-08$) and 1-pyrroline-4-hydroxy-2-carboxylate B ($r > 0.649$, $p < 2.01E-08$) (Fig. 6C).

Gene-based association and haplotype analysis of candidate genes

To further determine the relationship between candidate genes and traits, the SNPs of the candidate genes were applied for the gene-based association and haplotype analysis of the candidate genes. According to

Table 4 Candidate genes are identified in the transcriptome and QTN detection model

Gene ID	Gene function	Arabidopsis homologs	FHUFA vs. FLUFA	THUFA vs. TLUFA	HUFA vs. LUFA
<i>Glyma.02G068900</i>	Xyloglucan endotransglucosylase/hydrolase 5	AT5G13870.1	4.86	6.48	3.93
<i>Glyma.02G210300</i>	Unknown protein	AT2G14095.1	2.84	3.01	2.75
<i>Glyma.03G006700</i>	cysteine synthase 26	AT3G03630.1	1.26	1.32	1.02
<i>Glyma.03G040000</i>	lipid transfer protein 2	AT2G38530.1	4.86	3.01	1.93
<i>Glyma.03G040100</i>	lipid transfer protein 1	AT2G38540.1	4.01	4.67	2.63
<i>Glyma.15G087200</i>	Aldolase-type TIM barrel family protein	AT5G01410.1	1.54	1.59	1.33
<i>Glyma.03G040500</i>	Unknown protein	AT2G40435.1	2.24	1.45	1.16
<i>Glyma.04G102400</i>	Unknown protein	AT1G78170.1	2.19	2.37	2.11
<i>Glyma.04G102700</i>	Major facilitator superfamily protein	AT1G34580.1	2.73	2.84	1.71
<i>Glyma.04G209200</i>	Amino acid permease 2	AT5G09220.1	1.05	2.19	1.61
<i>Glyma.06G175100</i>	Leucine-rich repeat protein kinase family protein	AT2G31880.1	-1.51	-1.26	-1.11
<i>Glyma.07G205400</i>	Cysteine proteinases superfamily protein	AT3G49340.1	-3.61	-2.88	-2.75
<i>Glyma.09G032100</i>	myb domain protein 78	AT5G49620.1	2.11	2.40	1.72
<i>Glyma.09G033500</i>	Unknown protein	AT5G49525.1	3.02	2.50	1.58
<i>Glyma.09G033600</i>	Unknown protein	no	9.01	5.17	5.18
<i>Glyma.09G051900</i>	VQ motif-containing protein	AT4G20000.1	2.70	3.95	3.23
<i>Glyma.17G236700</i>	Acyl-CoA-binding domain 3	AT4G24230.6	1.62	1.96	1.24
<i>Glyma.09G053700</i>	Ankyrin repeat family protein	AT3G54070.1	-6.18	-5.18	-4.86
<i>Glyma.09G157500</i>	Unknown protein	no	1.94	2.03	1.39
<i>Glyma.10G079500</i>	Unknown protein	AT1G32120.1	8.32	6.32	6.43
<i>Glyma.11G100600</i>	Peptidoglycan-binding LysM domain-containing protein	AT5G23130.1	2.51	1.94	1.16
<i>Glyma.14G216100</i>	Protein kinase superfamily protein	AT5G37790.1	2.54	2.18	1.50
<i>Glyma.03G040400</i>	Lipid transfer protein 1	AT2G38540.1	1.26	1.94	1.42
<i>Glyma.15G088900</i>	GDSL-like lipase/acylhydrolase superfamily protein	AT1G29670.1	2.01	1.95	1.58
<i>Glyma.16G161500</i>	DNAse I-like superfamily protein	AT1G71710.1	1.61	1.77	1.20
<i>Glyma.17G220100</i>	Pentatricopeptide repeat (PPR) superfamily protein	AT2G13600.1	-1.59	-1.41	-1.08
<i>Glyma.17G236200</i>	Salt tolerance zinc finger	AT1G27730.1	2.77	3.12	1.19
<i>Glyma.18G239700</i>	Wall-associated kinase-like 2	AT1G16130.1	3.43	3.05	3.28
<i>Glyma.18G239900</i>	Cytochrome P450, family 97, subfamily A, polypeptide 3	AT1G31800.1	1.67	1.49	1.11
<i>Glyma.19G163600</i>	RING/U-box superfamily protein	AT1G04360.1	5.88	7.34	5.37

Table 5 Candidate genes are identified in the transcriptome and QEIs detection model

Gene ID	Gene function	Arabidopsis homologs	FHUFA vs. FLUFA	THUFA vs. TLUFA	HUFA vs. LUFA
<i>Glyma.09G053700</i>	Ankyrin repeat family protein	AT3G54070.1	-6.18	-5.18	-4.86
<i>Glyma.13G234700</i>	NAC domain containing protein 73	AT4G28500.1	-1.56	-1.14	-1.08
<i>Glyma.18G206600</i>	Integrase-type DNA-binding superfamily protein	AT2G40340.1	1.21	1.85	1.24
<i>Glyma.17G236700</i>	Acyl-CoA-binding domain 3	AT4G24230.6	1.62	1.96	1.24
<i>Glyma.14G023500</i>	Calmodulin-domain protein kinase 9	AT3G20410.1	1.66	1.86	1.32
<i>Glyma.03G036000</i>	Protein kinase superfamily protein	AT5G01850.1	1.85	1.90	1.35
<i>Glyma.03G036700</i>	ARM repeat superfamily protein	AT5G01830.1	2.21	1.70	1.37
<i>Glyma.08G025100</i>	Nodulin MtN3 family protein	AT4G10850.1	2.43	1.67	1.24
<i>Glyma.18G205400</i>	Unknown protein	AT3G51750.1	2.43	4.03	6.30

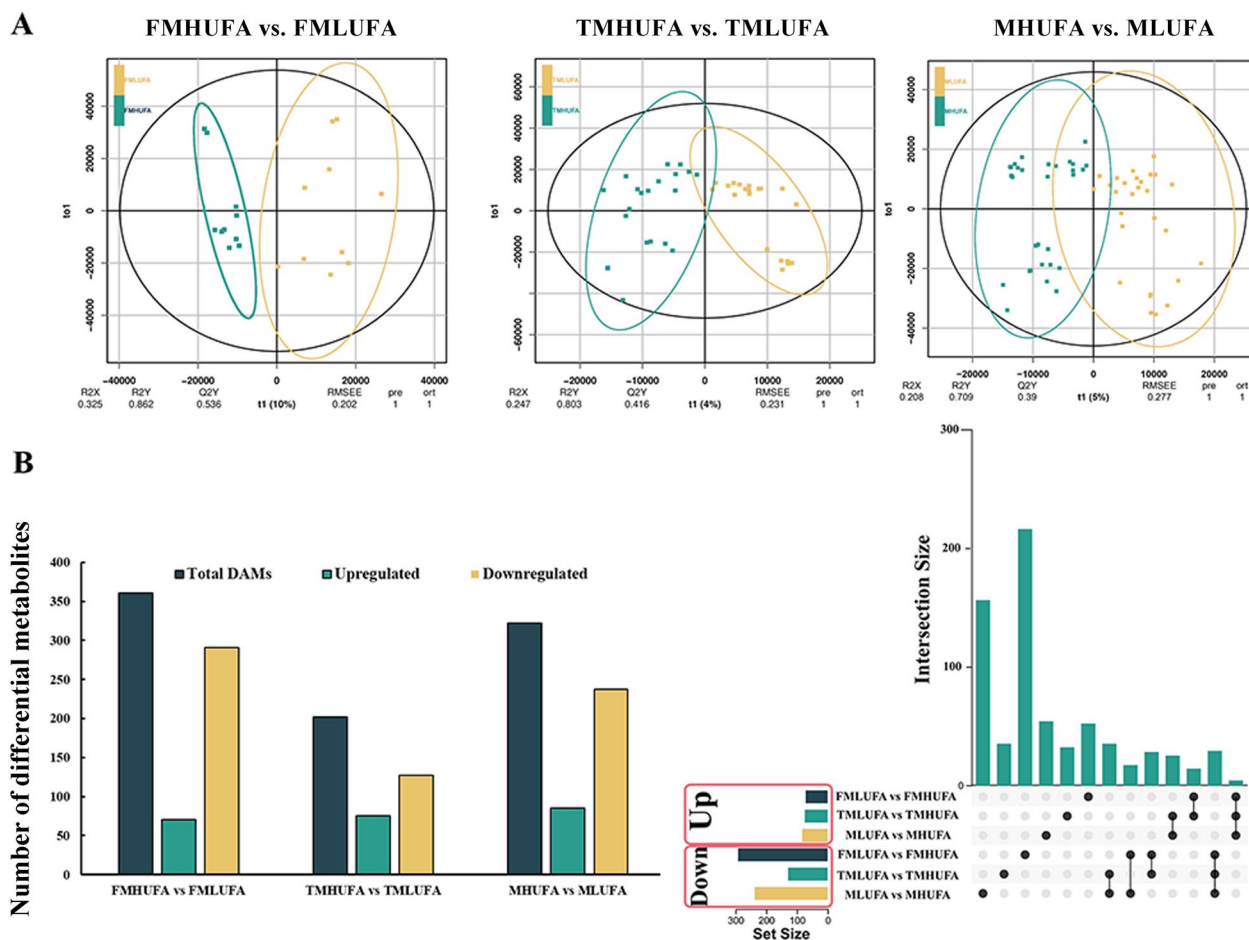


Fig. 4 Multivariate statistical analysis of the metabolome data in the soybean samples. **A** OPLS-DA model analysis. **B** Number of differential metabolites. The green and yellow columns represent the number of genes that were upregulated and downregulated, respectively. **C** Upsetplot diagram showing the overlapping DAMs in the three comparison groups

the results of candidate gene screening based on gene expression data from qRT-PCR and transcriptomics, *Glyma.03G040400* and *Glyma.17G236700*, as the candidate genes of QTNs and overlapping SNPs of QTNs and QEIs, were studied to understand the gene variations affecting soybean unsaturated FAs and to further determine beneficial haplotypes. Three SNPs were found in the promoter and CDS regions of *Glyma.03G040400* (Additional file 1: Table S4). SNP markers 39,188,954, 39,189,172, and 39,190,333 showed an association with LLA (Fig. 7A, Additional file 1: Table S4). Among the three haplotypes of *Glyma.03G040400*, Hap 3 and Hap 2 had a significantly higher LNA content than Hap 1 in 2013 and 2014 (Fig. 7B and C).

For candidate gene *Glyma.17G236700*, seven SNPs were found in the promoter and CDS region. Of these, SNP markers 5,048,564, 5,048,670, 5,048,842, 5,049,415, 5,049,422, 5,049,438, and 5,050,845 were significantly associated with LLA content in 2013 and 2014

($-\log_{10}(P) \geq 2$) (Additional file 1: Table S4, Fig. 7D). Four haplotypes of *Glyma.17G236700* were defined by the seven SNPs (Fig. 7E and F). Among the four haplotypes, Hap 3 and Hap 4 had a significantly higher LNA content (2013 and 2014) than Hap 1 and Hap 2.

Discussion

Soybean is an important oil crop. However, different proportions of FAs may play an important role in soybean oil. Therefore, it is of great significance to improve the content and quality of soybean oil. The single locus method has been widely used to detect genetic variation in crops, including GLM and MLM [23, 24]. However, single-locus GWAS methods generally need Bonferroni correction and can be affected by a polygenic background. In this study, 194 soybean accessions were analyzed using the 3VmrMLM method (Additional file 1: Figure S1, Table 2). We identified 12, 10, and eight significant/suggested SNPs

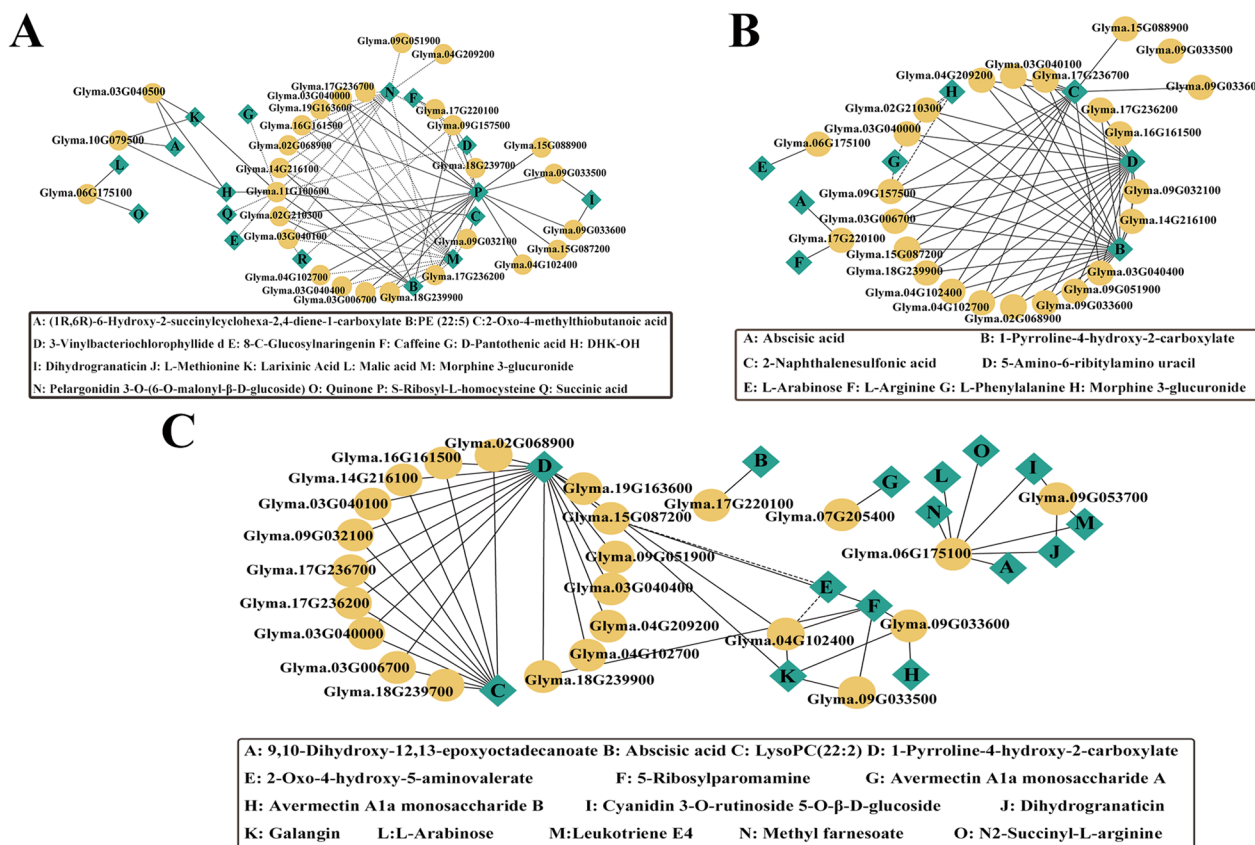


Fig. 5 In the QTN detection model, network analysis of the candidate DEGs and DAMs in the three comparison groups. **A** FHUFA vs. FLUFA, **B** THUFA vs. TLUFA, and **C** HUFA vs. LUFA. Yellow circles represent genes. Green squares represent metabolites. The solid line represents a positive correlation, while the dashed line represents a negative correlation

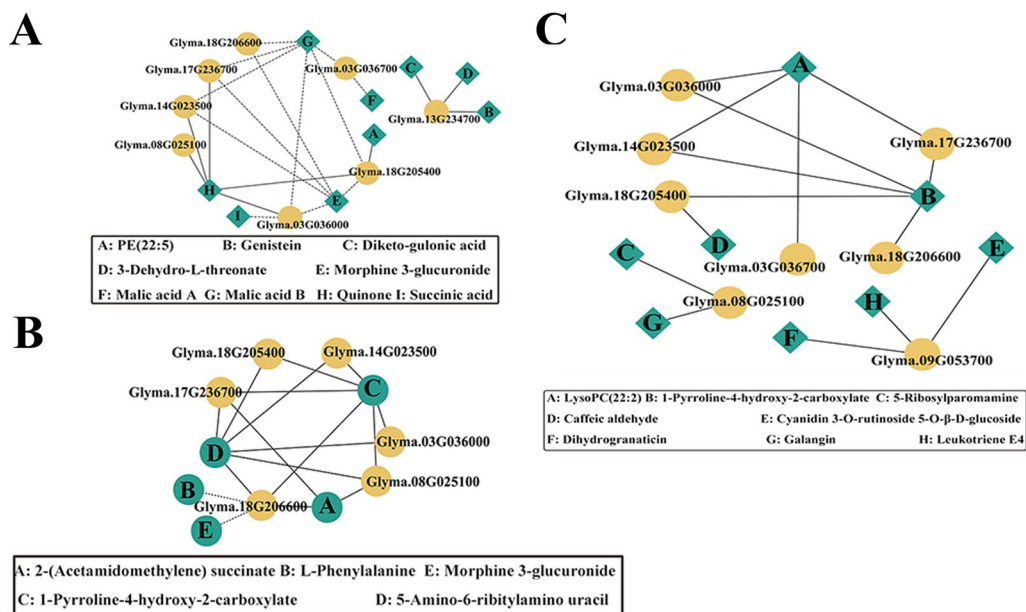


Fig. 6 QTN-by-environment detection model and network analysis of candidate DEGs and DAMs in the three comparison groups. **A** FHUFA vs. FLUFA, **B** THUFA vs. TLUFA, and **C** HUFA vs. LUFA. Yellow circles represent genes. Green squares represent metabolites. The solid line represents a positive correlation, while the dashed line represents a negative correlation

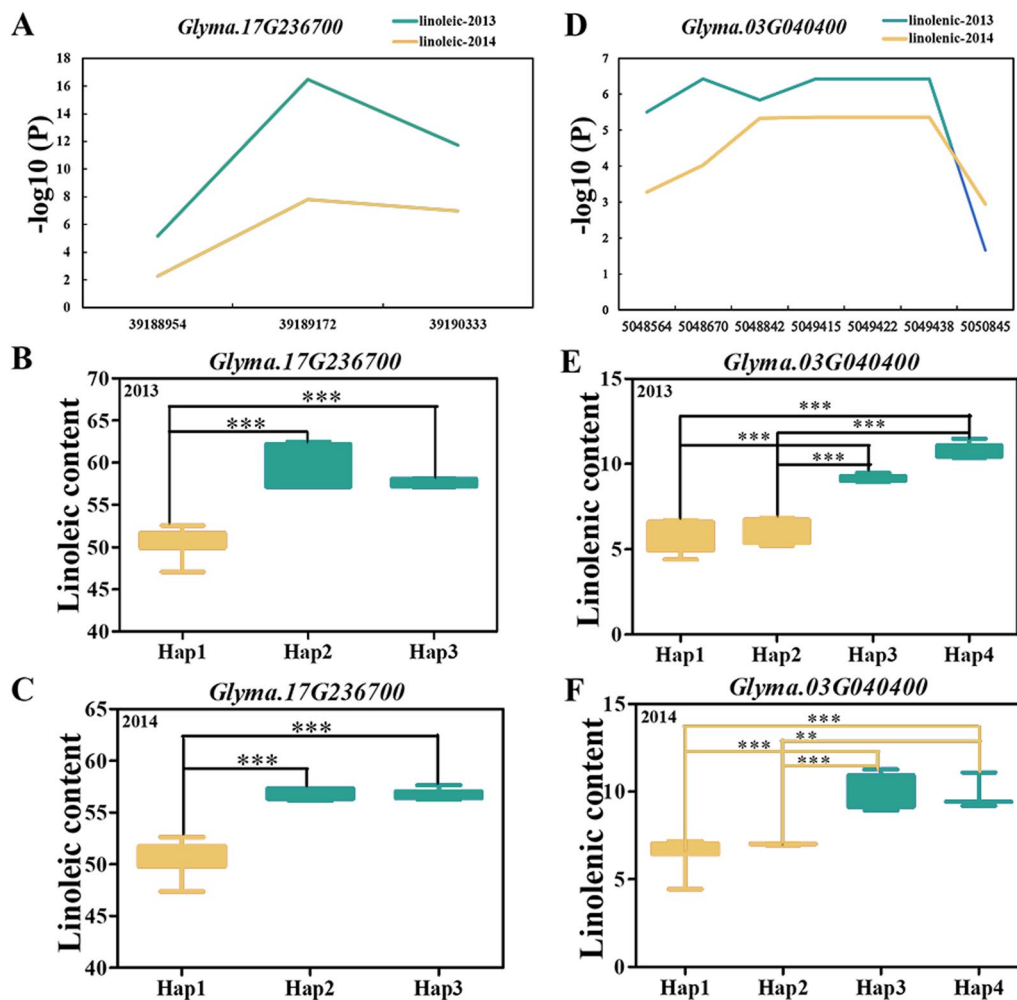


Fig. 7 Gene-based association analysis and haplotype analysis. **A** Gene-based association analysis of *Glyma.17G236700* related to linoleic content. **B, C** The relationship between haplotypes and linoleic content analysis of *Glyma.17G236700* in 2013 and 2014, respectively. **D** Gene-based association analysis of *Glyma.03G040400* related to linolenic content. **E, F** The relationship between haplotypes and linolenic content analysis of *Glyma.03G040400* in 2013 and 2014, respectively. * and ** indicate significance at $p < 0.05$ and $p < 0.01$, respectively

for OA, 17, 10, and seven significant/suggested SNPs for LA, and 10, eight, and 12 significant/suggested SNPs for LLA in 2013, 2014, and 2015, respectively (Table 2). In addition, we compared 3VmrMLM with a single-locus MLM method by Zhao et al. We detected 63 SNPs using the MLM method. Hence, the 3VmrMLM method detected more significant SNPs than the MLM method. Among these SNPs, four SNPs were found using the MLM and 3VmrMLM methods simultaneously, including rs4953186 rs52833743, rs35024325, and rs6481810, and the discovery of rs35024325 and rs6481810 SNPs has been reported [2, 22].

Environmental changes have an important impact on the quality and yield of crops; analysis of multiple

environments can increase the detection capability of SNPs. In this study, six, five, and eight QEIs were found for OA, LA, and LLA, respectively (Fig. 2, Table 3). Among these SNPs, five have been reported [22, 25, 26]. A total of 1246 genes around the significant/suggested QTNs were predicted in this study; of them, 40 genes were involved in lipid synthesis (Additional file 1: Table S1). For example, the MYB transcription factor has been reported to affect oil accumulation [27]. The *OsLTP* gene is involved in the transport of lipid molecules in rice [28]. In this study, *Glyma.03G040400* (*GmLTP1*), located on chromosome 3, was significantly related to LNA using the GLM method based on gene-based association (Additional file 1: Table S4). In

addition, the *GmLTP1* gene was a beneficial haplotype (Fig. 7).

A total of 301 genes around the significant/suggested QEIs were detected in this study (Additional file 1: Table S2). Three significant QEIs, namely rs4633292, rs39216169, and rs14264702, overlapped with significant single-environment QTNs. Among the overlapping SNPs, genes related to FA synthesis and seed development, such as ACBP and FTSH, were identified. ACBPs can play an important role in maintaining lipid homeostasis [29]. In addition, we found that the *Glyma.17G236700* (ACBP) gene had a beneficial haplotype (Fig. 7).

Conclusion

The 3VmrMLM method was more comprehensive for GWAS. This method overcame the huge computational burden of traditional models. In this study, 94 QTNs and 19 QEIs were identified. Five major candidate genes were found. The gene expression data from different soybean tissues and transcriptome data were used to identify *Glyma.03G040400* and *Glyma.17G236700* as key candidate genes around the SNPs. The beneficial haplotypes of *Glyma.03G040400* and *Glyma.17G236700* may be helpful for further application in soybean breeding.

Methods

Plant materials, field trials, and phenotypic evaluation

An association panel of 194 soybean germplasm resources was planted at Harbin (162.41° E, 45.45° N) in 2013, 2014, and 2015. Field trials were conducted using single-row plots (2 m long and 0.65 m between rows) and a randomized complete block design with three replicates per experimental site. The unsaturated FA content of each sample was determined using gas chromatography (GC-14C, Shimadzu Company, Japan), according to our previous method [30]. The OA, LLA, and LNA content were applied in single-environment (QTN) and multi-environment (QEI) analyses.

Genotypic data

A genotypic dataset consisting of 36,981 SNPs from 194 soybean germplasm resources was generated by Specific-Locus Amplified Fragment Sequencing (SLAF-seq), which was reported in Han et al. and Zhao et al. [2, 31]. The 36,981 SNPs were distributed on 20 soybean chromosomes, with minor allele frequencies > 0.04 and missing data of < 10% (Fig. 8).

GWAS

The 36,981 SNPs and unsaturated FA content of 194 soybean accessions were used for association analysis

via the 3VmrMLM method in 3VmrMLM software [21]. QTNs for OA, LLA, and LNA content were calculated from a single environment (3 years of phenotypic data from 2013, 2014, and 2015). QEIs for OA, LLA, and LNA content were calculated using a joint analysis of multiple environments. The threshold of significance for QTNs and QEIs was set at $p=0.05$ and LOD score ≥ 3.0 .

Differential expression analysis based on RNA-seq

At the R6 stage, 30 soybean varieties with a high content of three unsaturated FAs and a low content of three saturated FAs were collected for RNA sequencing (RNA-seq) with two biological replicates. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). The cDNA sequencing libraries of 18 RNA samples were constructed and sequenced, and RNA-seq data were generated using the Illumina platform. Differentially expressed genes (DEGs) were identified using the edgeR package in R software [32]. The significance level was set as follows: $|\log_2(\text{fold change})| \geq 1$.

Identification of candidate genes

The 100-kb flanking region of each identified QTN and QEI was defined to search for candidate genes according to linkage disequilibrium decay analysis, as described in Zhao et al. [2]. Candidate genes for unsaturated FAs were extracted in the following steps. According to previous reports, known genes related to FA content in *Arabidopsis* were considered references to screen their homologous genes in the soybean genome. The new candidate genes were identified using DEGs for unsaturated FAs.

Metabolite profiling

The non-targeted metabolome was completed by Bioacme Biotechnology Co., Ltd. (Wuhan, China). Briefly, a 100 mg soybean sample was loaded into a 2-mL centrifuge tube, and 300 μL 75% methanol/water was added. The tubes were centrifuged at 12,000 rpm for 15 min at 4 °C. Metabolites were screened and identified using the Metlin database. The differential metabolites were calculated using an orthogonal partial least squares-discriminant analysis (OPLS-DA) model, with a variable importance in the projection (VIP) score of ≥ 1 and a $|\log_2(\text{fold change})|$ of ≥ 1 .

Haplotype analysis and gene-based association analysis of candidate genes

The SNP variation of candidate genes was analyzed based on genome sequencing data. These SNPs were located at the full length of the gene, including exons, intronic regions, and upstream and downstream of the gene. Therefore, in this study, phenotypic data, including high

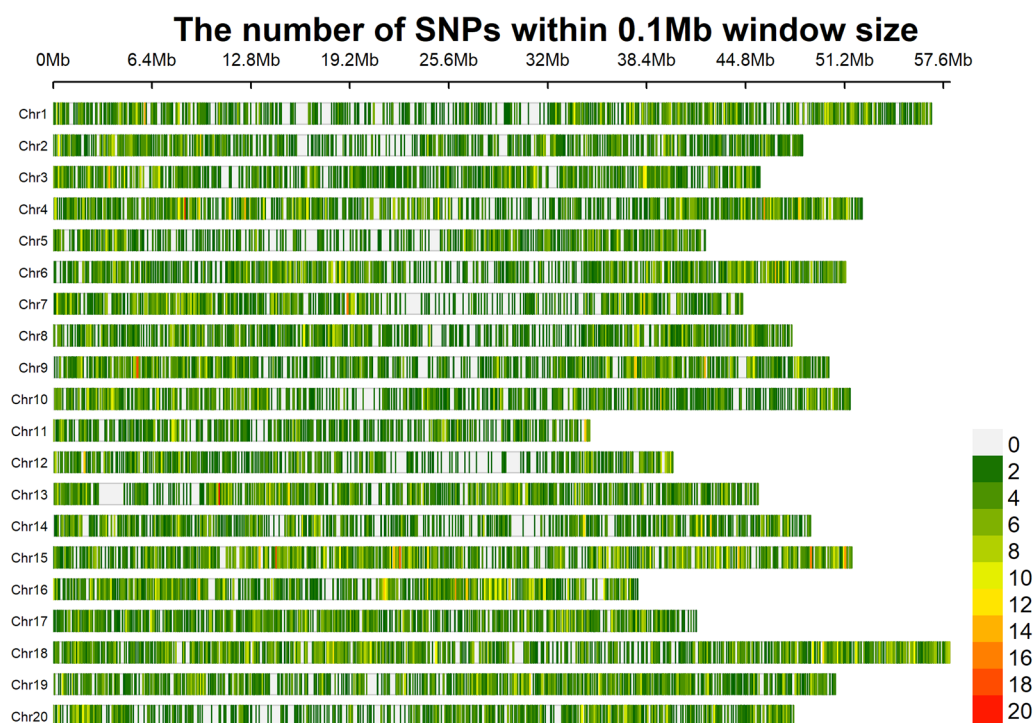


Fig. 8 Density distribution of single nucleotide polymorphisms (SNPs)

and low total unsaturated FA content, from 50 soybean germplasm resources were used over 3 years to conduct an association analysis. A general linear model (GLM) was used to further determine the association between the SNP variation of candidate genes and unsaturated FA content using TASSEL software [33]. Significant SNP variation in candidate genes was considered when the P value was less than 0.01.

Quantitative real-time PCR (qRT-PCR)

Soybean seeds with high and low unsaturated FA content were collected at the R6 stage. Total RNA was extracted using the TRIzol method, and cDNA was generated using the ReverTra Ace qPCR RT Master Mix (TOYOBO, Osaka, Japan). Real-time quantitative PCR (qRT-PCR) was performed on an ABI 7500 fast real-time PCR platform with SYBR Green (TOYOBO, Osaka, Japan). *GmACTIN4* was used as an internal control, and the primer sequences for candidate genes are listed in Additional file 1: Table S5. The L-13 soybean seed samples were used as a calibrator. The results of qRT-PCR were calculated using the $2^{-\Delta\Delta CT}$ method [34].

Co-expression analysis

The correlation coefficient was calculated between candidate genes and DAM metabolites, and a Pearson

correlation cutoff value of 0.5 was generated. Data were visualized using the Cytoscape package [35].

Statistical analysis

Statistical significance was evaluated using Student's t -test performed with SPSS 22.0 software (IBM Corp., Armonk, NY, USA). “*” and “***” represent a significance level of $p < 0.05$ and $p < 0.01$, respectively. The mean and standard deviation (mean \pm SD) were calculated using the data from three biological replicates.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13068-024-02489-2>.

Additional file 1: Table S1. QTNs identified for unsaturated fatty acids content using the QTN detection model in 3VmrMLM. **Table S2.** QEIs identified for unsaturated fatty acids content three environments detected using the QTN-by-environment detection model in 3VmrMLM. **Table S3.** Unsaturated fatty acids content of the 30 soybean varieties. **Table S4.** The association between SNP in *Glyma.17G23670* and *Glyma.03G040400* gene and soybean unsaturated fatty acids content based on 50 soybean germplasms. **Table S5.** Primers used for qRT-PCR. **Figure S1.** Manhattan plots of the single-environment analysis for the oleic, linoleic and linolenic traits in 2013, 2014 and 2015 of soybean. **Figure S2.** A and B KEGG pathway annotation around QTN and QEI candidate genes, respectively. **Figure S3.** Candidate genes are identified in the transcriptome and (A): QTN detection model, (B): QTN-by-environment detection model. **Figure S4.** Analysis of candidate genes by qRT-PCR. **Figure S5.** Differential accumulation of metabolite in the three comparison groups.

Acknowledgements

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Author contributions

XZ (Xunchao Zhao) performed the experiments and prepared the manuscript. YZ (Yuhang Zhan), KL and YZ (Yan Zhang) prepared the samples. CZ, MY, ML and YL analyzed the data. PZ, YH and XZ (Xue Zhao) conceived the experiments and revised the manuscript.

Funding

This study was financially supported by Heilongjiang Provincial Project (ZD2022C002, JD22A015), the National Key Research and Development Project of China (2021YFD1201604, 2021YFF1001204, 2021YFD1201103), the Chinese National Natural Science Foundation (U22A20473), the Youth Leading Talent Project of the Ministry of Science and Technology in China (2015RA228), the National Ten-thousand Talents Program, The national project (CARS-04-PS07). Young leading talents of Northeast Agricultural University (NEAU2023QNLJ-003). The funding bodies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The authors have consented for publication.

Competing interests

The authors declare that they have no competing interests. All authors agree to authorship and approved the final manuscript.

Received: 2 January 2024 Accepted: 7 March 2024

Published online: 16 March 2024

References

- Tunde-Akintunde T, Olajide J, Akintunde B. Mass-volume-area. Related and mechanical properties of soybean as a function of moisture and variety. *Int J Food Prop.* 2005;8:449–56. <https://doi.org/10.1080/10942910500267513>.
- Zhao X, Jiang H, Feng L, Qu Y, Teng W, Qiu L, Zheng H, Han Y, Li W. Genome-wide association and transcriptional studies reveal novel genes for unsaturated fatty acid synthesis in a panel of soybean accessions. *BMC Genomics.* 2019;20:68. <https://doi.org/10.1186/s12864-019-5449-z>.
- Liu X, Qin D, Piersanti A, Zhang Q, Miceli C, Wang P. Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. *BMC Plant Biol.* 2020;20:399. <https://doi.org/10.1186/s12870-020-02607-w>.
- Ascherio A, Willett W. Health effects of trans fatty acids. *Am J Clin Nutr.* 1997;66:1006–10. <https://doi.org/10.1093/ajcn/66.4.1006S>.
- Di Q, Piersanti A, Zhang Q, Miceli C, Li H, Liu X. Genome-wide association study identifies candidate genes related to the linoleic acid content in soybean seeds. *Int J Mol Sci.* 2021;23:454. <https://doi.org/10.3390/ijms23010454>.
- Marangoni F, Agostoni C, Borgh C, Catapano A, Cena H, Ghiselli A, La Vecchia C, Lercker G, Manzato E, Pirillo A, Riccardi G, Risé P, Visioli F, Poli A. Dietary linoleic acid and human health: focus on cardiovascular and cardiometabolic effects. *Atherosclerosis.* 2020;292:90–8. <https://doi.org/10.1016/j.atherosclerosis.2019.11.018>.
- Flores T, Karpova O, Su X, Zeng P, Bilyeu K, Slepser D, Nguyen H, Zhang Z. Silencing of GmFAD3 gene by siRNA leads to low alpha-linolenic acids (18:3) of fad3-mutant phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Res.* 2008;17:839–50. <https://doi.org/10.1007/s11248-008-9167-6>.
- Al Amin N, Ahmad N, Wu N, Pu X, Ma T, Du Y, Bo X, Wang N, Sharif R, Wang P. CRISPR-Cas9 mediated targeted disruption of FAD2–2 microsomal omega-6 desaturase in soybean (*Glycine max* L.). *BMC Biotechnol.* 2019;19:9. <https://doi.org/10.1186/s12896-019-0501-2>.
- Zhu C, Gore M, Buckler E, Yu J. Status and prospects of association mapping in plants. *Plant Genome.* 2008;1:5–20. <https://doi.org/10.3835/plantgenome2008.02.0089>.
- Zhang X, Guan Z, Li Z, Liu P, Ma L, Zhang Y, Pan L, He S, Zhang Y, Li P, Ge F, Zou C, He Y, Gao S, Pan G, Shen Y. Combination of linkage mapping and GWAS brings new elements on the genetic basis of yield-related traits in maize across multiple environments. *Theor Appl Genet.* 2020;133:2881–95. <https://doi.org/10.1007/s00122-020-03639-4>.
- Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton M, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet.* 2013;45:43–50. <https://doi.org/10.1038/ng.2484>.
- Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012;44:821–4. <https://doi.org/10.1038/ng.2310>.
- Yu J, Pressoir G, Briggs W, Vroh B, Yamasaki M, Doebley J, McMullen M, Gaut B, Nielsen D, Holland J, Kresovich S, Buckler E. Unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38:203–8. <https://doi.org/10.1038/ng1702>.
- Kang H, Zaitlen N, Wade C, Kirby A, Heckerman D, Daly M, Eskin E. Efficient control of population structure in model organism association mapping. *Genetics.* 2008;178:1709–23. <https://doi.org/10.1534/genetics.107.080101>.
- Lippert C, Listgarten J, Liu Y, Kadie C, Davidson R, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods.* 2011;8:833–94. <https://doi.org/10.1038/Nmeth.1681>.
- Wang S, Feng J, Ren W, Huang B, Zhou L, Wen Y, Zhang J, Dunwell J, Xu S, Zhang Y. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci Rep.* 2016;6:19444. <https://doi.org/10.1038/srep19444>.
- Liu X, Huang M, Fan B, Buckler E, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 2016;12: e1005767. <https://doi.org/10.1371/journal.pgen.1005767>.
- Ren W, Wen Y, Dunwell J, Zhang Y. pKwMB: integration of Kruskal-Wallis test with empirical bayes under polygenic background control for multi-locus genome-wide association study. *Heredity (Edinb).* 2018;120:208–18. <https://doi.org/10.1038/s41437-017-0007-4>.
- Wen Y, Zhang H, Ni Y, Huang B, Zhang J, Feng J, Wang S, Dunwell J, Zhang Y, Wu R. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief Bioinform.* 2018;19:700–12. <https://doi.org/10.1093/bib/bbw145>.
- Zhang J, Feng J, Ni Y, Wen Y, Niu Y, Tamba C, Yue C, Song Q, Zhang Y. pLARmEB: integration of least angle regression with empirical bayes for multilocus genome-wide association studies. *Heredity (Edinb).* 2017;118:517–24. <https://doi.org/10.1038/hdy.2017.8>.
- Li M, Zhang Y, Xiang Y, Liu M, Zhang Y. IIIvMrMLM: the R and C++ tools associated with 3VmrMLM, a comprehensive GWAS method for dissecting quantitative traits. *Mol Plant.* 2022;15:1251–3. <https://doi.org/10.1016/j.molp.2022.06.002>.
- Bachlava E, Dewey R, Burton J, Cardinal A. Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Sci.* 2009;49:433–42.
- Zhong H, Liu S, Sun T, Kong W, Deng X, Peng Z, Li Y. Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* 2021;21:364. <https://doi.org/10.1186/s12870-021-03146-8>.
- Reinprecht Y, Poysa V, Yu K, Rajcan I, Ablett G, Pauls K. Seed and agronomic QTL in low linolenic acid, lipoxigenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome.* 2006;49:1510–27.
- Kim H, Kim Y, Kim S, Son B, Choi I. Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. *J Life Sci.* 2010;20:1186–92. <https://doi.org/10.5352/JLS.2010.20.8.1186>.

26. Hyten D, Pantalone V, Saxton A, Schmidt M, Sams C. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. *J Amer Oil Chem Soc.* 2004;81:1115–8. <https://doi.org/10.1007/s11746-004-1027-z>.
27. Shi M, Yu L, Shi J, Liu J. A conserved MYB transcription factor is involved in regulating lipid metabolic pathways for oil biosynthesis in green algae. *New Phytol.* 2022;235:576–94. <https://doi.org/10.1111/nph.18119>.
28. Chen L, Ji C, Zhou D, Gou X, Tang J, Jiang Y, Han J, Liu Y, Chen L, Xie Y. OsLTP47 may function in a lipid transfer relay essential for pollen wall development in rice. *J Genet Genomics.* 2022;49:481–91. <https://doi.org/10.1016/j.jgg.2022.03.003>.
29. Hamdan M, Lung S, Guo Z, Chye M. Roles of acyl-coA-binding proteins in plant reproduction. *J Exp Bot.* 2022;73:2918–36. <https://doi.org/10.1093/jxb/erab499>.
30. Xie D, Han Y, Zeng Y, Chang W, Teng W, Li W. SSR- and SNP-related QTL underlying linolenic acid and other fatty acid contents in soybean seeds across multiple environments. *Mol Breed.* 2012;30:169–79.
31. Han Y, Zhao X, Liu D, Li Y, Lightfoot D, Yang Z, Zhao L, Zhou G, Wang Z, Huang L, Zhang Z, Qiu L, Zheng H, Li W. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytol.* 2016;209:871–84. <https://doi.org/10.1111/nph.13626>.
32. Robinson M, McCarthy D, Smyth G. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
33. Bradbury P, Zhang Z, Kroon D, Casstevens T, Ramdoss Y, Buckler E. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23:2633–5. <https://doi.org/10.1093/bioinformatics/btm308>.
34. Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–9. <https://doi.org/10.1038/ng1847>.
35. He L, Xiao J, Rashid K, Yao Z, Li P, Jia G, Wang X, Cloutier S, You F. Genome-wide association studies for pasmo resistance in Flax (*Linum usitatissimum* L.). *Front Plant Sci.* 2019;9:1982. <https://doi.org/10.3389/fpls.2018.01982>.
36. Panthee D, Pantalone V, Saxton A. Modifier QTL for fatty acid composition in soybean oil. *Euphytica.* 2006;152:67–73. <https://doi.org/10.1007/s10681-006-9179-3>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.