BMC Plant Biology

RESEARCH

Open Access

Genome-wide association analysis in peanut accessions uncovers the genetic basis regulating oil and fatty acid variation

Jing Xu[1], Xiao Jiang[1], Xiangzhen Yin[1], Xuhong Zhao[1], Na Chen[1], Lijuan Pan[1], Chun Fu[2], Yanlin Jiao[3], Junqing Ma[1], Mei Yuan[1*] and Xiaoyuan Chi[1*]

Abstract

Background The cultivated peanut, *Arachis hypogaea* L., is a critical oil and food crop worldwide. Improving seed oil quality in peanut has long been an aim of breeders. However, our knowledge of the genetic basis of selecting for seed nutritional traits is limited. Based on *AhFAD2A* and *AhFAD2B*, scientists have now developed higher oleic acid (80–84%) in peanut. Decoding the genetic makeup behind natural variation in kernel oil and fatty acid concentrations is crucial for molecular breeding-based nutrient quantity and quality manipulation.

Results Herein, we recognized 87 quantitative trait loci (QTLs) in 45 genomic regions for the concentrations of oil, oleic acid, and linoleic acid, as well as the oleic acid to linoleic acid (O/L) ratio via a genome-wide association study (GWAS) involving 499 peanut accessions. Eight QTLs explained more than 15% of the phenotypic variation in peanut accessions. Among the 45 potential genes significantly related to the four traits, only three genes displayed annotation to the fatty acid pathway. Furthermore, on the basis of pleiotropism or linkage data belonging to the identified singular QTLs, we generated a trait-locus axis to better elucidate the genetic background behind the observed oil and fatty acid concentration association. Expression analysis indicated that *arahy.AV6GAN* and *arahy.NNA8KD* have higher expressions in the seeds.

Conclusion This natural population consisting of 499 peanut accessions combined with high-density SNPs will provide a better choice for identifying peanut QTLs/genes in the future. Together, our results provide strong evidence for the genetic mechanism behind oil biosynthesis in peanut, facilitating future advances in multiple fatty acid component generation via pyramiding of desirable QTLs.

Keywords Peanut, GWAS, Oil, Fatty acid, Genetic basis

*Correspondence:
Mei Yuan
yuanbeauty@126.com
Xiaoyuan Chi
chi000@126.com
[1]Shandong Peanut Research Institute, Qingdao CN-266100, China
[2]Weifang Academy of Agricultural Sciences, Weifang CN-261071, China
[3]Yantai Academy of Agricultural Sciences, Yantai CN-265500, China

Background

Cultivated peanuts (*Arachis hypogaea* L.) are an allotetraploid oilseed crop that originated from an abiogenetic event in which the diploid ancestor *Arachis duranensis* (AA) was crossed with *Arachis ipaensis* (BB) in South America [1, 2]. At present, peanuts are being produced in over 115 countries, covering 26 million hectares and achieving global yields of 41 million tons respectively [3]. However, this remains inadequate in meeting the rising demand for edible peanut oil. Peanut seeds contain approximately 50% oil, 25% protein, 15% carbohydrates, 2% fiber, 2% ash, and 6% water, as well as minor bioactive compounds, namely, folates, minerals, vitamin E, resveratrol, and flavonoids, which possess robust antioxidant properties. A 1% rise in peanut oil content can potentially bring about an economic surge of 7% in oil processing [4]. Since oil is a major peanut seed component, the fatty acid configuration critically determines seed quality. Recently, monounsaturated oleic acid-based was reported to be highly beneficial to human health, for example, it lowered cholesterol contents and diminished coronary heart and inflammatory disease risks [5, 6]. Owing to its inherent antioxidant property that protects against harmful substances [7], peanut oil with augmented oleic acid levels possesses the optimal balance of fatty acid content, thereby facilitating remarkable stability, and prolonging oil storage duration. There is increasing demand to breed peanut varieties with high oil content and elevated oleic acid levels. Hence, identifying genes/loci controlling these traits is critical and optimizing selection strategies will improve peanut yield and quality.

Most plant seeds contain triacylglycerols. Till date, there has been extensive studies on the model plant *Arabidopsis thaliana*, which has extended our comprehension of storage-oil biosynthetic networks, as well as the genes and gene regulators [8, 9]. A number of lipid metabolism -related genes has been identified, namely, *DGAT* [10, 11], *FAD2* [12, 13], *PEPC* [14], *LEC1* [15], *FAD3* [16], *GPAT9* [17], *KCS* [18], *MAGL3b* [19]. These genes lay down the molecular foundation to enhancing fatty acid composition. Oil, oleic acid, and linoleic acid concentrations in peanut seeds are intricate traits regulated by genes with significant environmental influences, which can be improved by QTL localization analysis and the polymerization of different QTLs. Scientists have employed molecular biomarkers for peanut seed oil-, oleic acid-, and linoleic acid-related QTLs or chromosomal region discoveries. Till now, several QTL-based evaluations utilized early-generation markers namely SSRs [13, 20–23], but few polymorphic markers and relatively low frequency of target trait-related genes markedly diminish application of SSR markers. More recently, with elucidation of the peanut genome, as well as the advent of high-throughput sequencing technologies [1, 24–28], multiple genotyping platforms, and array-based genotyping protocols [29], it is possible to generate massive quantity of data for high-resolution trait mapping [30]. Therefore, an increasing number of QTLs with detailed locations and SNP markers robustly associated with target traits have been reported [4, 31–38]. Nevertheless, even with extensive knowledge of the plant fatty acid biosynthetic axis and associated genes, the molecular mechanism behind the natural fatty acid biosynthesis variation remains elusive in peanuts because of the limited quantity of parental lines employed in prior investigations.

Genome wide association study (GWAS) is typically used to examine the genetic makeup of complicated traits in peanuts. It is particularly beneficial due to its broad diversity and rapid linkage disequilibrium (LD) decay in a given species [39–42]. Some progress has been made in peanut quality traits recently [42–45], for instance, 1,224 SNPs were associated with at least two of four seed quality-related traits (OA, LA, ROLA and PA) which included two loci closed to *FAD2A* and *FAD2B* [45]. Zhang et al. (2021) conducted a GWAS using 13,382 single nucleotide polymorphisms (SNPs) to uncover the genetic makeup behind oil, oleic acid, linoleic acid, and the O/L ratio via a miscellaneous panel of 120 peanut accessions and detected 18, 20, 4, and 21 SNPs, respectively [42]. However, GWAS-based genetic dissection with high throughput genotyping data in peanuts quality traits is limited. Given the large genome size and limited genetic diversity of cultivated peanuts, the development of SNP array chips is essential for enabling high-throughput genotyping [46]. Overall, the first version of a peanut large-scale (58 K) SNP genotyping array 'Axiom_*Arachis*-v1' and an improved version of the peanut SNP array 'Axiom_*Arachis*-v2' were successfully used to dissect the fresh seed dormancy [47, 48], leaf chlorophyll content [49], growth habit-related traits [50], late leaf spot [51], tomato spotted wilt virus [52], seed weight and shelling percentage [29, 53–55]. This was done for the purposes of accelerating high-resolution mapping and molecular breeding of peanuts. In this report, the data were analyzed in an association study for the concentrations of oil, oleic acid, and linoleic acid and the O/L ratio. We used a diverse association panel, containing numerous phenotypic variations for the studied traits. Next, we developed a trait–QTL axis for evaluation of the genetic associations among the 4 traits.

Results

Phenotypic variation in oil-associated traits among AMP

Near-infrared reflectance spectroscopy (NIR) is a widely used quick and non-invasive technique for oil and fatty acid composition detection in peanuts [56–60]. Herein, the 4 oil-related traits were demonstrated to have plentiful diversity, with most traits appearing to be continuous

and with somewhat normal distribution (Fig. 1). The variation exhibited a 1.18-fold alteration in oil composition to a 29.78-fold alteration in oleic acid composition. The oil composition was between 46.25 and 54.35, with a mean of 51.81. The linoleic acid content also varied, with an average of 31.78, ranging from 4.47 ~ 44.17. 47 cultivars (lines) were identified as high oleic acid peanuts and harbored more than 75% oleic acid contents. ANOVA suggested that the genotypic and environmental influences significantly affected individual traits (Additional file 2: Table S2), while the phenotypic variation was controlled primarily by genetic factors, with broad-sense heritability ranging between 0.64 and 0.97 (Additional file 2: Table S2). In terms of pairwise correlations, we observed marked association among oleic acid, linoleic acid, and O/L, likely due to their physiologically correlation (Fig. 1).

SNPs profiling and dispersal within the peanut genome

The SNP array (48 K) developed by Clevenger et al. (2018) [61] identified 47,837 polymorphic loci, with 19,554 (40.8%) derived from subgenome A, 21,876 (45.7%) from subgenome B, and 6,407 (13.4%) originating from scaffold regions. (Fig. 2a). A total of 34,588 polymorphic SNPs whose incorrect alignments and multiple alignments were removed underwent alignment to the reference genome. Approximately 10,948 polymorphisms with MAF ≥ 0.05 and missing information ≤ 20% were selected for a GWAS via a SNP array. Chromosome Arahy.01 had the largest SNPs quantity (8.07%; 884 of 10948), whereas chromosome Arahy.08, the shortest chromosome at 51.53 Mb, had the minimum (3.15%; 345 of 10948). However, the smallest and largest average distances between two adjacent SNPs occurred in Arahy.01 (127.17 kb/SNP) and Arahy.16 (388.80 kb/SNP) (Fig. 2b, Additional file 2: Table S3). The peanut genome contained a mean SNP density of 4.31 SNPs/Mb, with the highest density at Arahy.01 (7.86 SNPs/Mb) and the lowest density at Arahy.16 (2.57 SNPs/Mb).

GWAS of oil, oleic, linoleic, and O/L contents in peanut kernels

To explore the genetic makeup of oil, oleic acid, and linoleic acid concentrations, we conducted GWAS with an association panel of 499 peanut accessions and 10,948 high-quality SNPs with an MAF > 0.05. Association

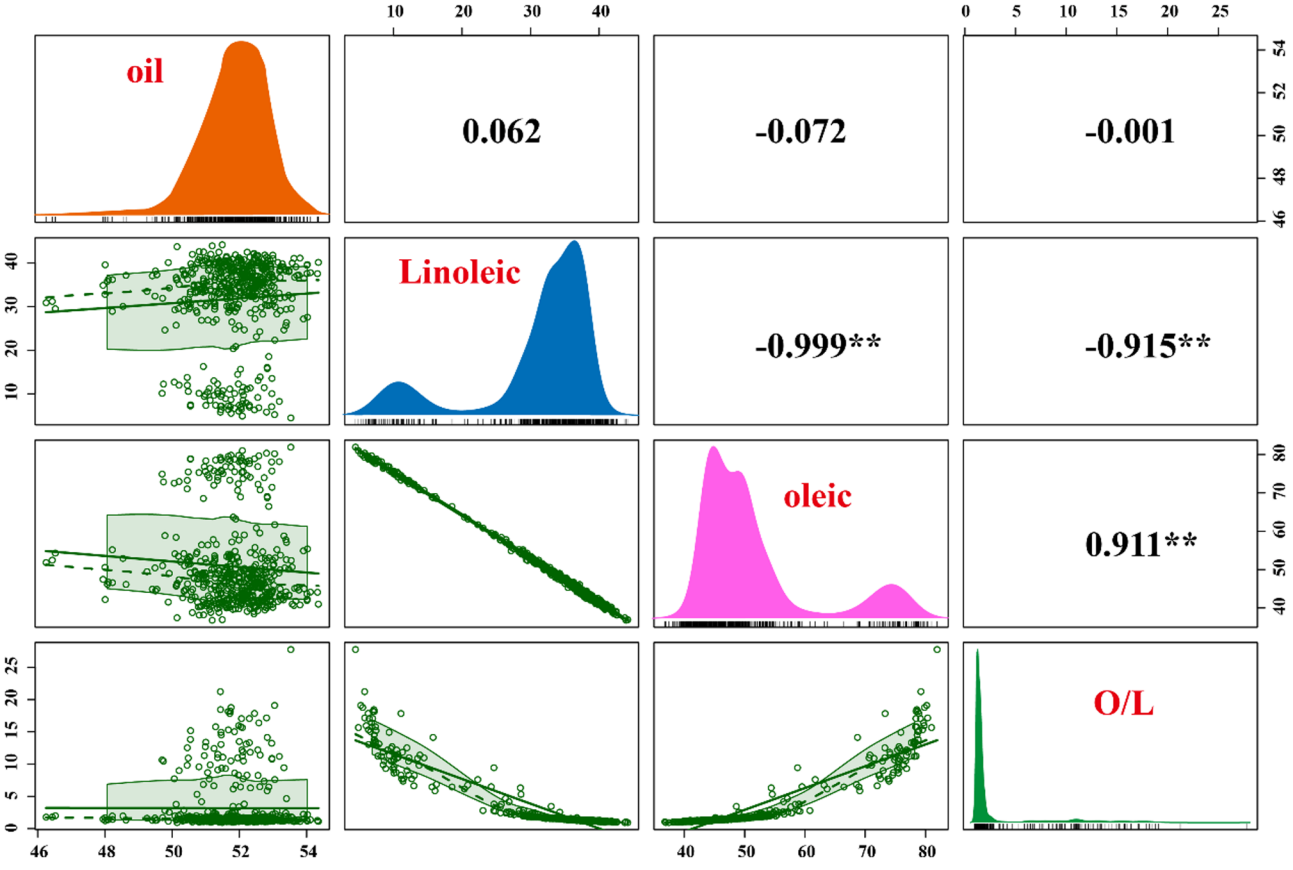

Fig. 1 The phenotypic distribution and correlation among oil, linoleic, oleic, and O/L in peanut accessions. The diagonal area displayed the phenotypic variation of each trait. The lower left part showed scatter plots of every two traits. The upper right part was correlation coefficient between the composition traits, **, $P < 0.01$

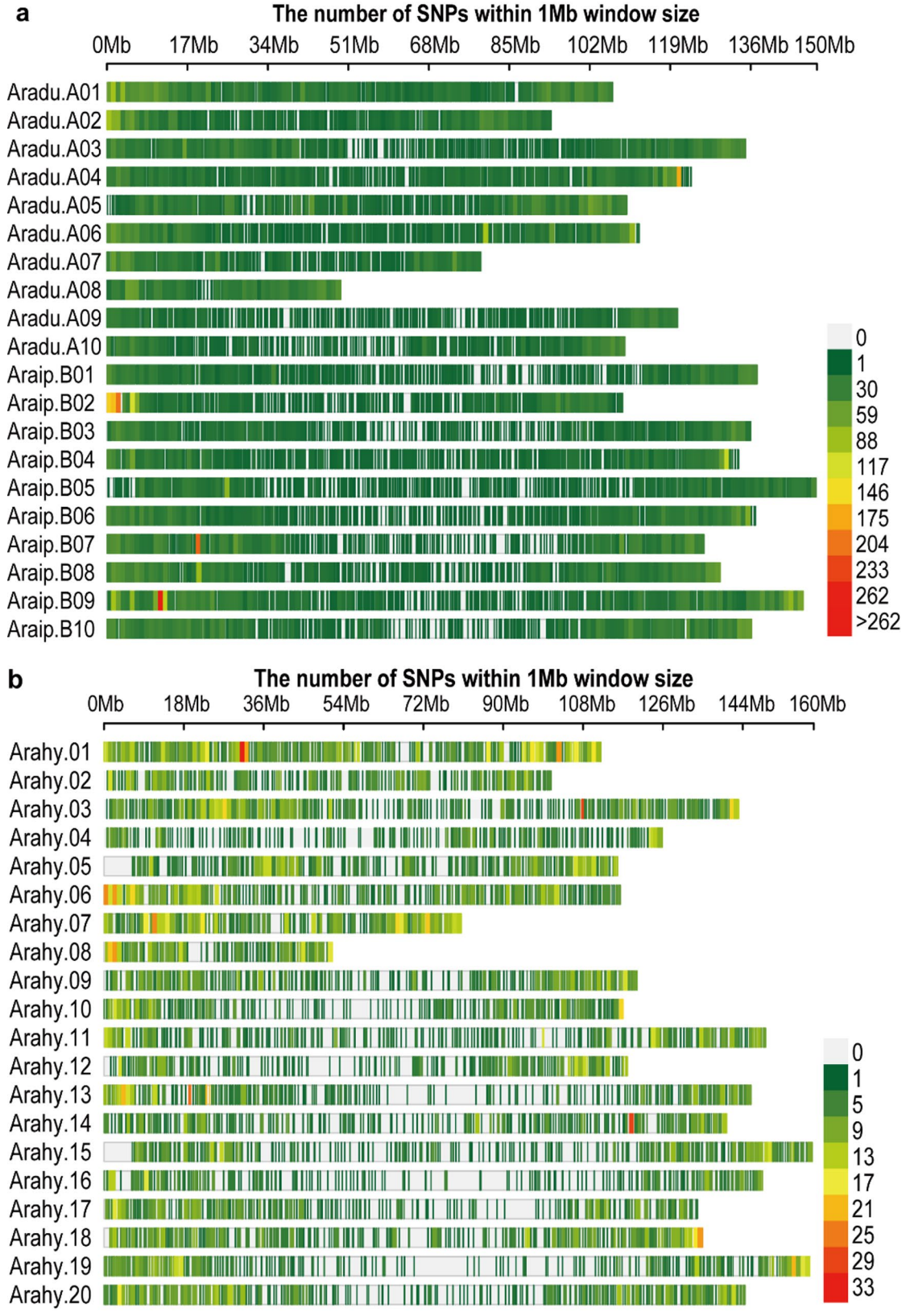

Fig. 2 Single nucleotide polymorphisms (SNP) dispersion among the 20 chromosomes of cultivated peanut. Horizontal axis depicting chromosomal length (Mb), Green and red shades representing SNP density (SNPs quantity per window). Vertical axis depicting the 20 chromosomes. (**a**) Polymorphic SNPs apart from the scaffold biomarkers; (**b**) Polymorphic SNPs (apart from the scaffold biomarkers) following filter

analysis of these polymorphisms revealed 45 loci linked to oil, oleic, and linoleic acid concentrations, and O/L traits at $P < 9.13 \times 10^{-5}$ in the peanut association, with loci quantity per trait between 3 and 36 (Fig. 3a **and** Fig. 3c). As depicted in the quantile-quantile and Manhattan plots depicting the four traits, we observed marked direct correlations following usage of the mixed linear model that takes into account both population structure and familial relatedness (Fig. 3a **and** Fig. 3b). Thirteen of the four trait-associated loci resided on chromosome Arahy.13, seven on chromosome Arahy.8 and chromosome Arahy.19, five on chromosome Arahy.6, three on chromosome Arahy.9, two on chromosome Arahy.2, chromosome Arahy.3, and chromosome Arahy.12, and one on chromosome Arahy.7, chromosome Arahy.10, chromosome Arahy.15, and chromosome Arahy.18 (Fig. 3d). Twenty identified loci were independently associated with only one trait, three for the oil trait and 17 for the O/L trait. The other 25 loci were pleiotropic, including 17 loci associated with oleic, linoleic, and O/L traits simultaneously (Additional

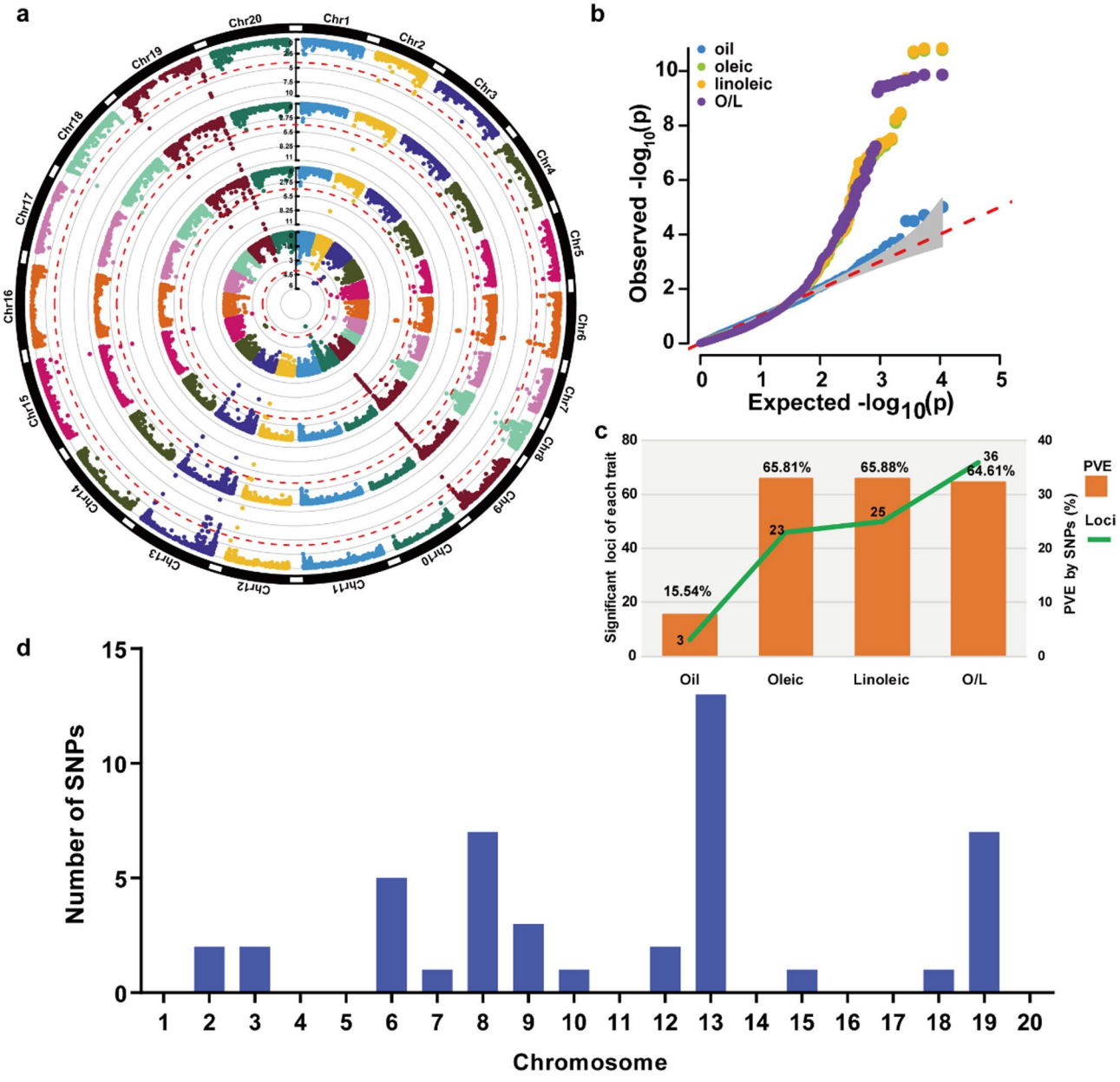

Fig. 3 Overview of GWAS-identified SNPs for oil, linoleic, oleic and O/L traits. (**a**) Manhattan plot depicting GWAS data for 4 traits. Red dashed line representing the likelihood ratio test threshold (LRT = 4.04), subjected to $\log_2$ transformation. The circles from the inside out are oil, oleic, linoleic and O/L, respectively. (**b**) Quantile-quantile plot for 4 traits. (**c**) The number of significant loci and total phenotypic variation explaining all identified loci by GWAS analysis. (d) The distribution of significant associated SNPs on 20 peanut chromosomes

file 1: Fig. S1). The explained phenotypic variation (PVE) of individual loci was between 0.65% and 37.83%, and the total PVE of all recognized loci for individual traits was between 15.54% (oil) to 65.88% (linoleic) (Fig. 3c). Among these loci, 17.78% (8/45) exhibited substantial influences, with a PVE ≥ 15%. Majority loci affected oleic and linoleic acids, which are fatty acids that are probably regulated by a limited number of marked-influence loci given their simplicity as intermediate products in a specific metabolic network.

Genetic basis of the correlations among oil, oleic acid, linoleic acid, and O/L

Pleiotropic-related traits or strongly associated genetic factors are frequently linked [46–48]. In case of the 4 analyzed traits, diverse associations, including strong and weak, direct and inverse, became apparent (Fig. 1). To elucidate the genetic link among various traits, we generated a trait–locus axis using data from locus–locus linkages and locus–trait associations (Fig. 4). We revealed that all traits consisted of ≥ 1 locus that served as a bridge connecting the traits. Unsurprisingly, traits belonging to the same metabolic network were closely associated within the axis, and exhibited augmented connectivity thereby suggesting the presence of multiple common loci. This is likely due to the fact that a relevant enzyme-encoding gene alteration within a specific metabolic axis can bring about changes in metabolite synthesis either downstream or upstream of the enzyme-catalyzed reaction due to a feedforward and/or feedback modulation [49]. For example, oleic acid and its associated loci on chromosome Arahy.09 were strongly linked to linoleic acid and its ratios (Fig. 4; Table 1). This finding was further corroborated by a prior QTL cluster cloning showing *AhFAD2A*, which encodes the enzyme fatty acid desaturase (*AhFAD2A*) which catalyzes the oleic acid to linoleic acid transition [12, 62]. In terms of the trait–locus axis, we also identified 20 loci specific to a singular trait and 25 loci associated with multiple traits (Fig. 4; Table 1). This trait–locus axis is highly informative in elucidating the genetic makeup of associations among various acid traits. This will unequivocally assist in enhancing multiple traits via pyramiding desirable QTLs.

Functional annotation and prediction of candidate genes

Improving oil content and quality represents a critical objective in peanut breeding programs. Thus, GWAS was also performed to identify candidate genes for oil content and quality. A total of 3, 23, 25 and 36 candidate genes were identified for oil, oleic acid, linoleic acid, and the O/L, respectively. Based on our functional category annotation of the 45 potential genes, 3 genes were positively associated with enzymes involved in fatty acid metabolism. The 42 remaining genes were estimated to encode enzymes involved in other metabolic axes, transcription factors, transporters, and other proteins (Table 1, Additional file 1: Fig. S2). Among them, one pleiotropic SNP (AX-147234396) falled on *arahy.42CZAS* (named *AhFAD2A*) gene was associated with oleic acid, linoleic acid, and the O/L, encoding the fatty acid desaturase enzyme, played a crucial role in regulating both oleic acid and linoleic acid levels [13]. The *arahy.D9C186* (named *AhFUS3*), a B3 domain-containing transcription factor, was also found to be identified for linoleic acid, and the O/L. Additionally, the *arahy.PB1D04* (named *AhKCS*) gene, which encodes ketoacyl-CoA synthase, was linked to O/L. *arahy.HGAZ7D* gene is located 10.9 kb up-stream of AX-176,806,119 (chr13) was predicted for oleic acid, linoleic acid, and the O/L traits, and this gene encodes a member of the ERF (ethylene response factor) subfamily which have been reported to influence the accumulation of aliphatic glucosinolates [63]. In addition, four genes (*arahy.AV6GAN*, *arahy.WHSM29*, *arahy.7US8C9* and *arahy.NNA8KD* predicted for fatty acid component) are involved in phosphoglycerate dehydrogenase, late embryogenesis abundant (LEA) protein, sucrose transporter, light-sensor protein kinase, respectively (Table 1). It is highly likely that these genes have no direct involvement in the relevant network, rather, the polymorphic markers are associated with causal polymorphisms present in nearby genes. In addition, a large number of potential genes were estimated to function outside of the fatty acid axis, even though the loci-harboring axis-related genes contributed similarly to the genetic makeup of phenotypic variation seen with linoleic and oleic traits (Additional file 1: Fig. S3). Based on these findings, the pathway gene-specific loci are critical components of the genetics of fatty acid variation. Compared with the unfavorable alleles, the lines carrying favorable alleles of the lead SNPs in the three fatty acid pathway genes presented augmented oleic acid levels or diminished linoleic acid levels at different levels (Fig. 5; Table 2). This study identified a variety of enzymes involved in metabolic processes, suggesting that the regulation of fatty acids is related to other metabolic pathways and that there may be the same metabolic intermediates between different metabolic pathways, so they can be mutually regulated. In addition, we infer that the regulation of oil-related traits was not only related to the enzymes in the metabolic pathway but also indirectly regulated by many transcription factors and proteins.

Expression profiling of identified candidate genes

In this study, we leveraged previously published transcriptome data from peanutbase.org (https://www.peanutbase.org/expression/expr_tissue_Hyp.html) [64] and unpublished transcriptome data from our group to delve deeper into the expression profiles of selected candidate

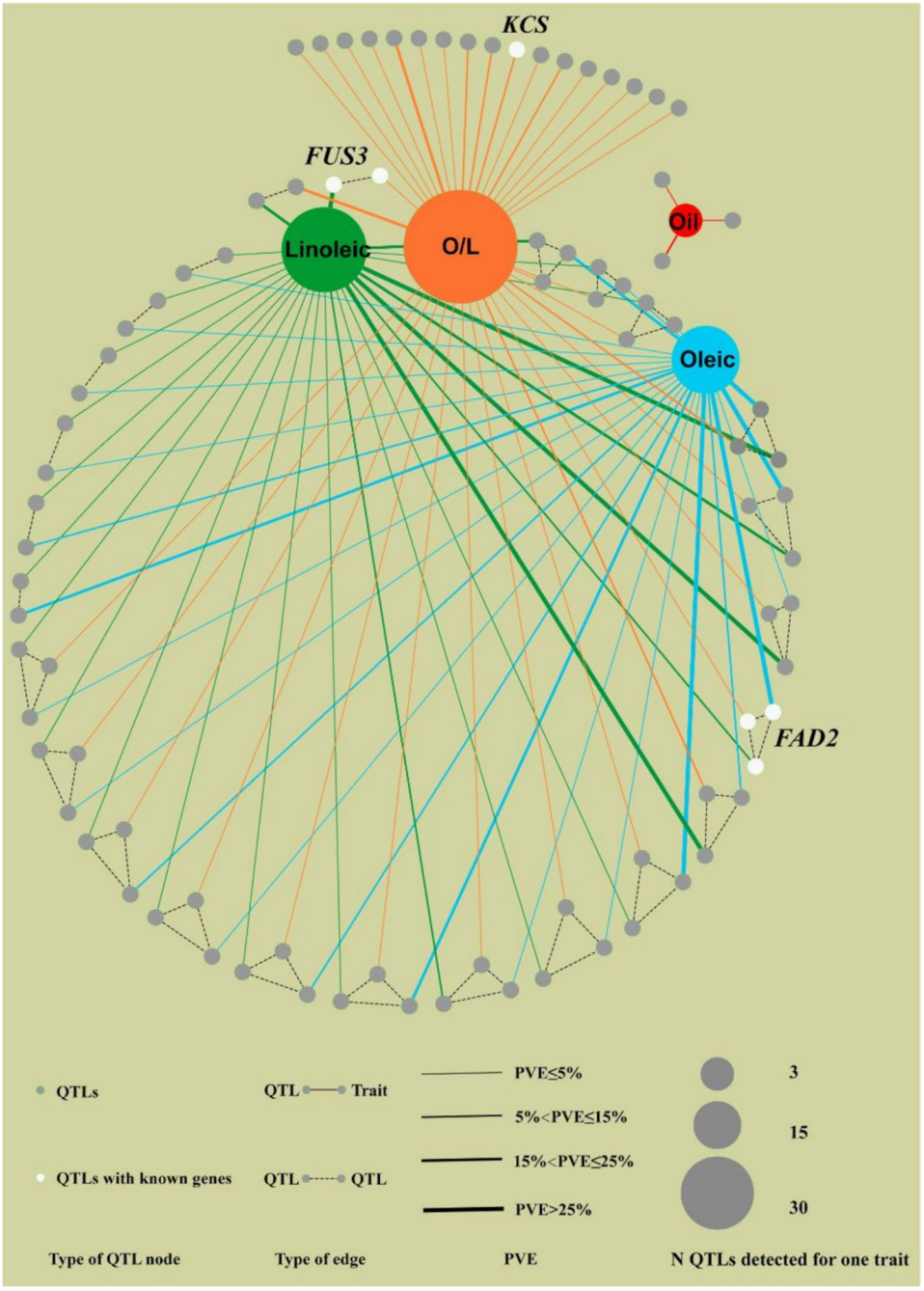

Fig. 4 Trait–QTLs axis for 4 traits and 87 single QTLs. The trait–QTL axis generated according to the QTL findings involving 4 traits and co-localization data of all 87 QTLs. Traits and loci linked by solid lines within the axis were strongly linked to the traits. Loci joined by dashed lines within the axis if two QTLs co-localized at the same loci. Table 1 summarizes the trait names in detail

genes potentially involved in seed development in peanuts. This dataset includes transcriptome-based expressions at 22 different developmental stages, as detailed in Clevenger et al. (2016) for the full development of peanut [64] and our group (unpublished) for the 6 different developmental stages of seed (Fig. 6b). Clearly, the expression levels of *AhFAD2A* and *AhFUS3* were significantly elevated during grain development compared to other stages, highlighting their crucial roles in seed development. Additionally, KCS exhibited high expression during the perianth period. Among the remaining candidate genes, *arahy.AV6GAN* and *arahy.NNA8KD* demonstrated significantly higher expression levels than the others during grain development, suggesting their crucial role in pod swelling processes (Fig. 6a). Furthermore, our investigation extended to the transcriptome

Table 1 SNPs and potential genes strongly linked to oil, linoleic, oleic Concent and O/L

Locus	QTL_ID	Candidate gene[a]	Chr	Position[b]	SNPs	Allele[c]	Trait name	*P* value[d]	Annotation[e]	Class of function[f]
L1	Q1	*arahy.BGTX42*	2	20,628,343	AX-176,805,166	C/T	Linoleic	1.302E-07	ATP binding ABC transporter	transporter
	Q65						oleic	6.134E-08		
	Q26						O/L	2.342E-05		
L2	Q27	*arahy.CA7JWP*	2	96,352,459	AX-176,812,849	G/A	O/L	7.175E-05	Disease resistance protein	stress
L3	Q62	*arahy.XRK6MC*	3	65,329,353	AX-176,815,977	T/C	oil	1.967E-05	cysteine synthase	protein modification
L4	Q63	*arahy.RMNX5H*	3	107,856,306	AX-176,808,118	A/G	oil	3.232E-05	transferring glycosyl group transferase	transferase
L5	Q28	*arahy.4LV1NS*	6	95,323,986	AX-176,810,806	T/C	O/L	1.685E-05	apyrase 2	hydrolase activity
L6	Q2	*arahy.LUT2QN*	6	96,089,127	AX-176,802,651	A/G	Linoleic	1.529E-07	flowering locus protein	protein modification
	Q66						oleic	2.865E-07		
	Q29						O/L	3.433E-10		
L7	Q3	*arahy.D2DMXG*	6	97,268,221	AX-176,817,478	G/A	Linoleic	1.956E-07	glyceraldehyde-3-phosphate dehydrogenase	catalytic activity
	Q67						oleic	3.5E-07		
	Q30						O/L	3.284E-10		
L8	Q4	*arahy.Z9K4G4*	6	98,229,148	AX-147,225,829	A/C	Linoleic	7.226E-08	calcium-binding mitochondrial carrier protein	transporter
	Q68						oleic	1.385E-07		
	Q31						O/L	1.378E-10		
L9	Q5	*arahy.D9C186*	6	98,966,601	AX-176,806,463	T/C	Linoleic	8.454E-05	B3 domain-containing transcription factor, *FUS3*	fatty acid pathway
	Q32						O/L	1.755E-05		
L10	Q33	*arahy.5GJZ36*	7	53,641,722	AX-147,256,961	G/A	O/L	2.027E-05	early nodulin-like protein	metal ion binding
L11	Q34	*arahy.6TI2WT*	8	11,822,331	AX-176,799,876	T/C	O/L	3.901E-07	chitinase-like protein	unknown
L12	Q35	*arahy.1ZYA2W*	8	15,756,459	AX-176,800,487	T/C	O/L	7.556E-05	Ribosomal-protein-alanine N-acetyltransferase	transferase
L13	Q36	*arahy.Q8CPXC*	8	25,513,322	AX-147,230,326	G/A	O/L	5.164E-05	Disease resistance protein	stress
L14	Q37	*arahy.ZFP9XM*	8	27,305,454	AX-147,230,428	C/T	O/L	8.132E-05	telomere repeat-binding protein	protein modification
L15	Q38	*arahy.P5SVLS*	8	27,743,726	AX-177,640,378	C/T	O/L	6.364E-05	Ankyrin repeat family protein	protein modification
L16	Q6	*arahy.PE2UWG*	8	29,492,439	AX-176,815,346	G/A	Linoleic	1.474E-05	shugoshin protein	unknown
	Q69						oleic	2.616E-05		
	Q39						O/L	1.986E-06		
L17	Q40	*arahy.JJEK9U*	8	29,590,776	AX-147,230,536	C/T	O/L	4.028E-05	zinc finger	transcription factor
L18	Q7	*arahy.ZNL5N1*	9	45,597,689	AX-147,233,346	A/G	Linoleic	5.285E-05	Proteasome	protein modification
	Q70						oleic	5.992E-05		
L19	Q8	*arahy.J0DG7J*	9	113,116,605	AX-176,800,330	T/C	Linoleic	5.866E-09	ethylene-responsive transcription factor	transcription factor
	Q71						oleic	7.886E-09		
	Q41						O/L	2.043E-07		
L20	Q9	*arahy.42CZAS*	9	114,175,008	AX-147,234,396	G/A	Linoleic	9.587E-08	fatty acid desaturase 2, *FAD2*	fatty acid pathway
	Q72						oleic	9.126E-08		
	Q42						O/L	1.912E-05		
L21	Q64	*arahy.I0F6BN*	10	108,757,825	AX-177,638,445	A/G	oil	9.924E-06	beta-glucosidase	hydrolase activity
L22	Q10	*arahy.769USI*	12	111,139,624	AX-176,808,914	A/G	Linoleic	2.186E-07	40 S ribosomal protein	cell related
	Q73						oleic	4.07E-07		
	Q43						O/L	3.727E-10		

Table 1 (continued)

Locus	QTL_ID	Candidate gene[a]	Chr	Position[b]	SNPs	Allele[c]	Trait name	*P* value[d]	Annotation[e]	Class of function[f]
L23	Q11	*arahy.7I6F51*	12	111,582,358	AX-176,799,008	C/A	Linoleic	2.188E-07	Unknown	Unknown
	Q74						oleic	4.063E-07		
	Q44						O/L	3.634E-10		
L24	Q45	*arahy.PB1D04*	13	4,285,230	AX-176,819,738	C/T	O/L	4.01E-06	Ketoacyl–CoA Synthase, *KCS*	fatty acid pathway
L25	Q12	*arahy.HGAZ7D*	13	23,102,127	AX-176,806,119	A/C	Linoleic	1.389E-07	ethylene-responsive transcription factor	transcription factor
	Q75						oleic	2.749E-07		
	Q46						O/L	2.409E-10		
L26	Q47	*arahy.AV6GAN*	13	26,392,719	AX-176,821,622	G/T	O/L	1.177E-06	phosphoglycerate dehydrogenase	oxidoreductase activity
L27	Q48	*arahy.WHSM29*	13	29,541,641	AX-176,816,134	C/T	O/L	4.724E-05	Late embryogenesis abundant (LEA) protein	protein modification
L28	Q49	*arahy.MM5I5L*	13	30,457,558	AX-176,794,990	A/C	O/L	5.398E-05	Pentatricopeptide repeat (PPR) superfamily protein	protein modification
L29	Q50	*arahy.5E18AM*	13	31,038,548	AX-147,244,306	C/T	O/L	2.643E-05	polynucleotide phosphatase/kinase	protein modification
L30	Q13	*arahy.HFL4GJ*	13	31,590,472	AX-176,814,627	G/A	Linoleic	8.753E-05	adenylate kinase family protein	protein modification
	Q51						O/L	9.772E-07		
L31	Q14	*arahy.TD4WZI*	13	34,053,349	AX-176,801,965	T/C	Linoleic	6.654E-05	transmembrane protein	unknown
	Q76						oleic	8.789E-05		
	Q52						O/L	2.09E-06		
L32	Q53	*arahy.V4Y3U4*	13	35,888,481	AX-176,796,285	C/T	O/L	1.093E-06	pyruvate phosphate dikinase	protein modification
L33	Q54	*arahy.F6ALXS*	13	36,327,157	AX-147,244,415	G/A	O/L	2.7E-05	ABC transporter family protein	transporter
L34	Q15	*arahy.LC4GJT*	13	138,669,899	AX-176,796,664	A/G	Linoleic	5.184E-06	acetylglucosamine transferase	transporter
	Q77						oleic	8.812E-06		
	Q55						O/L	1.785E-06		
L35	Q16	*arahy.7VS6FJ*	13	139,463,301	AX-147,218,189	A/G	Linoleic	4.638E-08	Glycine cleavage T-protein family	protein modification
	Q78						oleic	8.908E-08		
	Q56						O/L	2.218E-10		
L36	Q17	*arahy.X4FXIQ*	13	140,221,764	AX-147,218,261	T/G	Linoleic	4.07E-06	ATP-binding	ATP-related
	Q79						oleic	4.409E-06		
L37	Q18	*arahy.6QT9XW*	15	149,071,412	AX-176,819,580	T/C	Linoleic	6.596E-07	ninja-family protein	unknown
	Q80						oleic	1.156E-06		
	Q57						O/L	5.871E-10		
L38	Q19	*arahy.7US8C9*	18	4,015,065	AX-176,815,348	G/T	Linoleic	6.74E-05	sucrose transporter	transporter
	Q81						oleic	8.695E-05		
	Q58						O/L	1.374E-06		
L39	Q20	*arahy.X7AYSP*	19	4,487,043	AX-176,805,249	G/A	Linoleic	2.437E-07	F-box protein	metal ion binding
	Q82						oleic	1.297E-07		
	Q59						O/L	4.454E-05		
L40	Q60	*arahy.6D5HXJ*	19	7,357,800	AX-177,643,058	G/A	O/L	4.408E-07	P-loop containing nucleoside triphosphate hydrolase	hydrolase activity
L41	Q21	*arahy.KXR6N9*	19	11,483,770	AX-177,643,206	A/G	Linoleic	4.522E-05	3-hydroxyisobutyryl-CoA hydrolase-like protein	hydrolase activity
	Q83						oleic	5.619E-05		

Table 1 (continued)

Locus	QTL_ID	Candidate gene[a]	Chr	Position[b]	SNPs	Allele[c]	Trait name	*P* value[d]	Annotation[e]	Class of function[f]
L42	Q22	*arahy.NNA8KD*	19	27,580,881	AX-176,803,212	G/A	Linoleic	2.934E-05	Light-sensor Protein kinase	protein modification
	Q84						oleic	2.84E-05		
L43	Q23	*arahy.16ZACY*	19	42,828,965	AX-176,795,908	G/A	Linoleic	3.845E-05	Cyclin-dependent protein kinase regulator	protein modification
	Q85						oleic	3.794E-05		
L44	Q24	*arahy.X7PJ8H*	19	155,907,786	AX-147,262,366	G/A	Linoleic	1.922E-11	Protein kinase superfamily protein	protein modification
	Q86						oleic	2.214E-11		
	Q61						O/L	9.912E-08		
L45	Q25	*arahy.HR8ZWH*	19	157,712,615	AX-176,798,303	G/A	Linoleic	6.055E-05	Leucine-rich repeat protein	protein modification
	Q87						oleic	6.695E-05		

[a] A possible physiological gene within the locus or nearest annotated gene to the leading SNP

[b] Base pair position for the leading SNP, depending on the *Arachis hypogaea* Tifrunner 2.0 reference genome

[c] Major allele, minor allele

[d] *P* value of the 4 traits

[e] Each potential gene was annotated based on the Peanut Base (https://www.peanutbase.org/genome/)

[f] Gene functional annotations as per InterProScan (http://www.ebi.ac.uk/interpro/)

data of two peanut cultivars, H57 (P19-57) and L61 (P19-61), characterized by varying oleic acid contents in their mature seeds—over 75% in P19-57 and less than 45% in P19-61. Analysis of expression patterns at six seed developmental stages revealed that *AhFAD2A* and *AhFUS3* exhibits consistently higher expression levels in P19-57 than P19-61. The expression of *AhKCS* was predominantly observed during the initial phase of grain development. In addition, *arahy.AV6GAN* and *arahy.WHSM29* primarily regulating the oleic acid to linoleic acid ratio (O/L), elevated expression of L61 compared to H57. In contrast, the H57 cultivar exhibited higher expression levels of *arahy.NNA8KD* (Fig. 6c). Based on these observations, we hypothesize that *arahy.AV6GAN* and *arahy.NNA8KD* are the key contributor to peanut fatty acid metabolism. Given the significant roles that *arahy.AV6GAN* and *arahy.NNA8KD* appear to play, further research is essential. Furthermore, candidate gene *arahy.AV6GAN* showed higher expression in L61 than H57 stages in two peanut cultivars with varying oleic acid contents that may be attributed to its role in promoting linoleic acid synthesis during lipid metabolism. On the contrary, candidate gene *arahy.NNA8KD* had higher expression in peanut H57. Moving forward, our research will focus on functional verification of these two candidate genes, including qRT-PCR assays and gene silencing techniques to further confirm the functionality of the candidate gene.

Discussion

The cultivated peanut is an allopolyploid formed by the hybridization of two wild diploid species and chromosome doubling [65]. Relative to other recurrent polyploid species, namely soybeans, which have an ancient tetraploid origin, the cultivated peanut employs a restricted genetic network with scarce amount of genetic variations among cultivars. Unfortunately, this negatively impacts polymorphic biomarker development for SNP based genetic linkage map revealed very high-density map. Predictably, only a few polymorphic biomarkers have been reported in peanut chromosomes. For example, the 2334 polymorphic peanut biomarkers reported by Hu et al. (2018) were not evenly dispersed across 20 chromosomes, and the polymorphic biomarker quantity per LG were between 13 and 369 [35]. This distribution pattern was also reported by other studies involving peanuts [66–68]. Our study extends the current understanding of the genetic architecture of oil and fatty acid traits in peanuts by employing a large and diverse association panel of 499 accessions. This is in contrast to previous studies that have utilized smaller panels or biparental mapping populations [21, 22, 35]. By using a high-density SNP array, we were able to identify more precise QTL locations compared to earlier investigations, which employed fewer polymorphic markers [42, 55]. The GWAS SNP density was 4.31 SNPs/Mb, which corroborated with earlier results and present as satisfactory value for GWAS assessment in peanuts [39, 40, 55]. However, this investigation faced an additional challenge: we detected several notable gaps between SNPs on some chromosomes (Fig. 1b and Additional file 1: Fig. S4). The peanut genome exhibited an overall gap between SNPs of 0.24 Mb, with the Arahy.01 and 07 (0.13 Mb) and Arahy.16 (0.39 Mb) gaps between SNPs being the lowest and highest, respectively (Additional file 1: Fig. S4 and Additional file 2: Table S4). Notably, seven gaps larger than 5 Mb were detected between SNPs on various chromosomes with the largest gap appeared on Arahy.19 (approximately 12 Mb). Wide inter-marker intervals compromise linkage disequilibrium (LD)-based signal detection, resulting

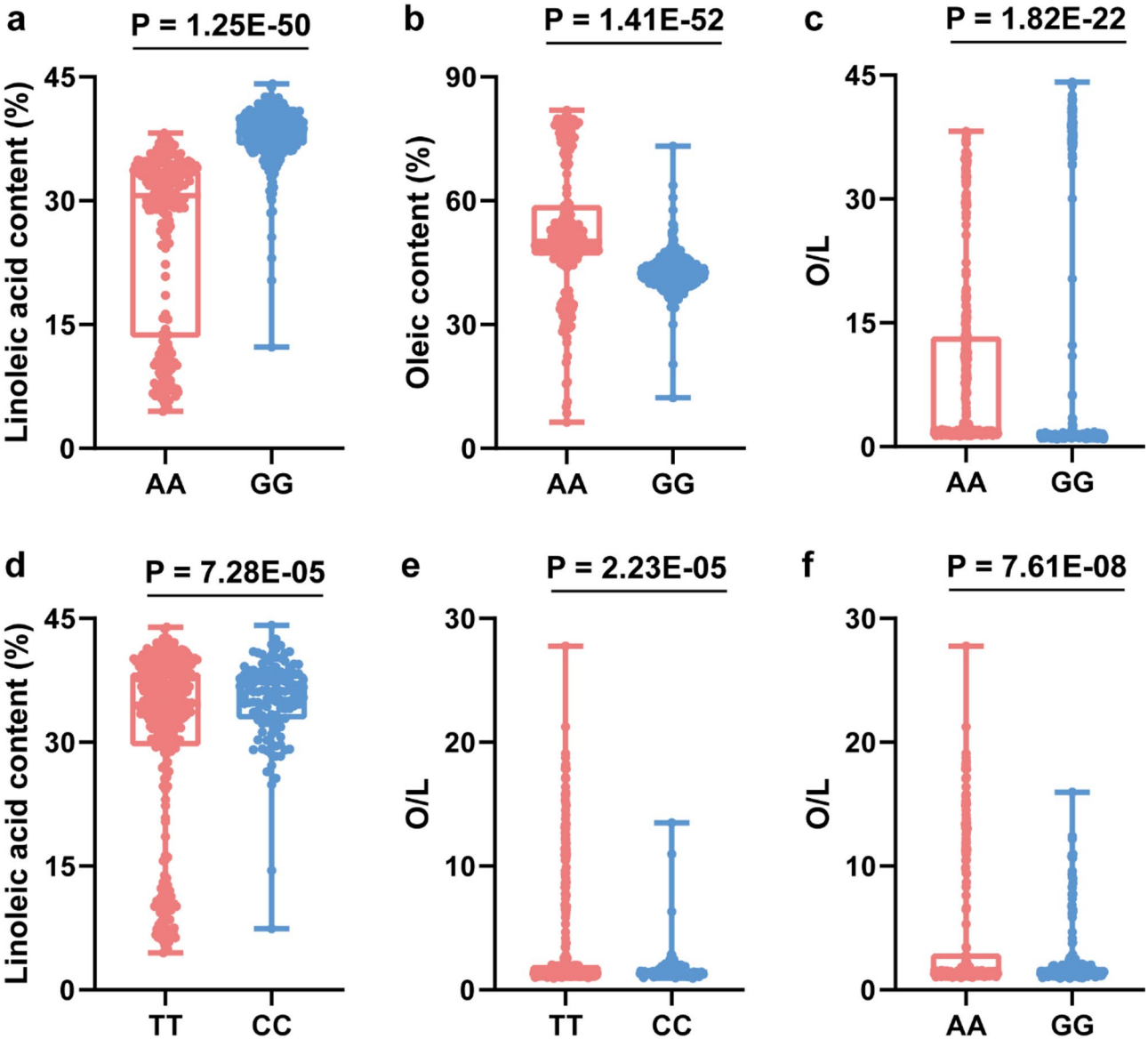

Fig. 5 Genetic impacts of the leading SNPs at the candidate genes of fatty acid pathway on linoleic, oleic content and O/L. (**a**-**c**) Genetic effects of *AhFAD2A* locus on linoleic, oleic content and O/L, respectively. (**d**-**e**) Genetic effects of *FUS3* locus on linoleic, oleic content, respectively. (**f**) Genetic effects of *KCS* locus on O/L. The *P*-values were computed according to a two-tailed Student's *t*-test

Table 2 Genetic impacts of the leading SNPs at the *AhFAD2A*, *AhFUS3* and *AhKCS* loci on fatty acid content

Candidate genes	Lead SNP	Trait name	Allele[a]	PVE (%)[b]	Increased or decreased fatty acid content (%)[c]
AhFAD2A	AX-147,234,396	Linoleic	G/**A**	36.33	32.05
		Oleic		37.44	31.98
		O/L		14.11	255.53
AhFUS3	AX-176,806,463	Linoleic	**T**/C	4.43	11.88
		O/L		5.13	107.37
AhKCS	AX-176,819,738	O/L	**A**/G	5.74	89.34

[a]The bold font underlined nucleotide represents the desirable allele for individual SNP

[b]Percentage of phenotypic variation elucidated by the recognized relevant SNPs

[c]The desirable alleles at the leading SNPs augmented the oleic levels, O/L or decreased linoleic acid content in comparison with unfavourable alleles

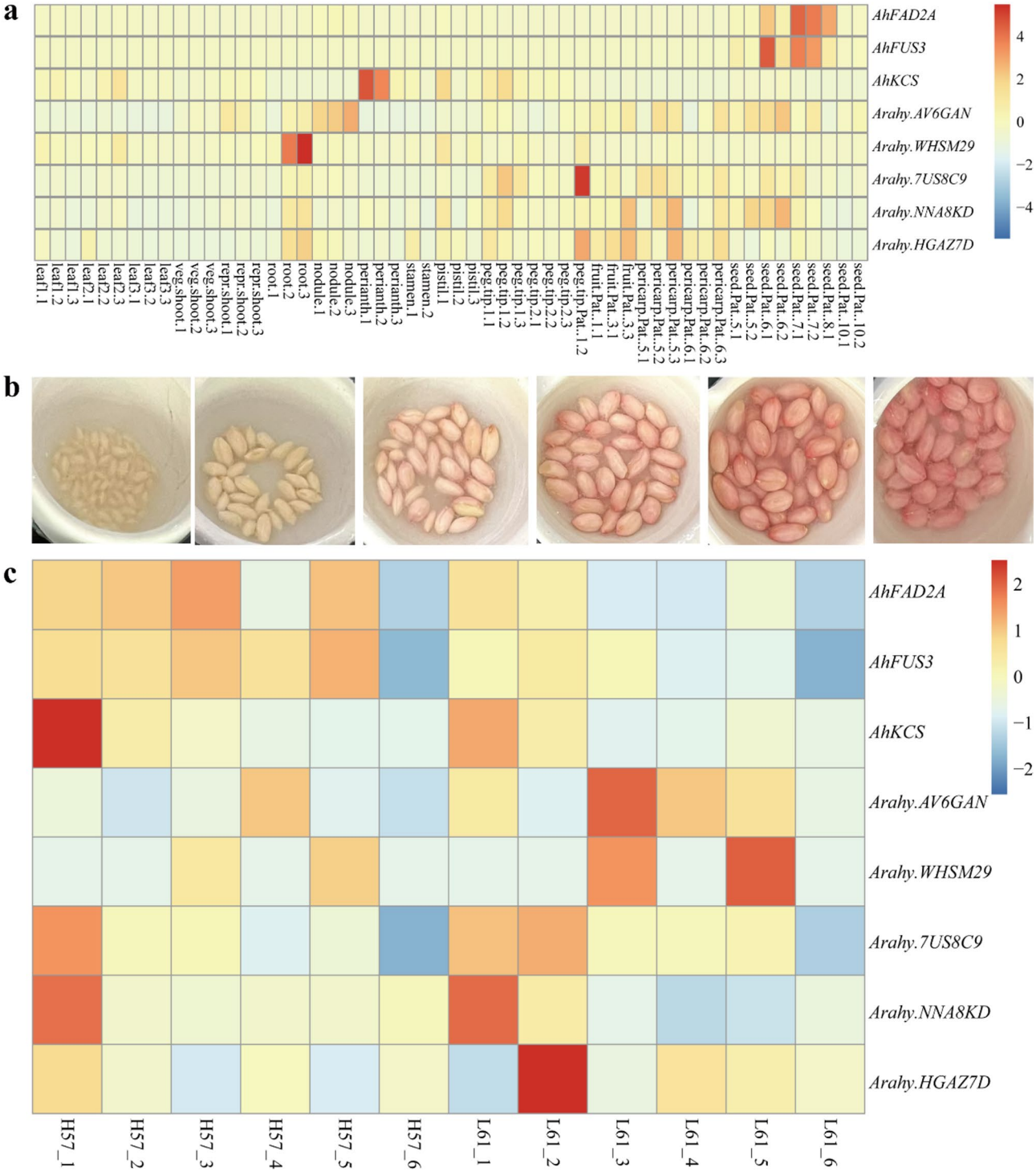

Fig. 6 Expression profiling of candidate genes at different development stages. (**a**) RNA-seq for the full development at 22 different tissue types and ontogenies, respectively [64]. (**b**-**c**) Peanut seeds at different developmental stages and RNA-seq for seed in high and low oleic acid content peanut (our group, unpublished). H57 is a high-oleic acid peanut, and L61 is a low-oleic acid peanut, both undergoing six seed developmental stages (1, 2, 3, 4, 5, 6)

in inadequate genomic resolution to capture trait-associated causal variants [69]. This sparse marker density restricts the precision of QTL mapping, expanding the physical intervals containing potential candidate genes and imposing substantial challenges for downstream functional validation through experimental approaches [70]. We concluded that undetectable genetic factors may drive fatty acid biosynthesis in genomic gaps, as current

SNP markers lack these variants. Future studies employing more comprehensive genotyping approaches, such as whole-genome sequencing, may help uncover these missing factors and provide a more complete picture of the genetic landscape of peanut oil related traits.

Our prediction of the broad-sense heritability of the 4 seed composition traits were exceptionally elevated, indicating the large dependent of peanut seed quality on genetics. Oil and fatty acids form the primary constituents of peanut seeds, and upregulating the seed oil and fatty acid contents is absolutely crucial; of note, oleic or linoleic acid levels are generally not linked to oil concentration [42, 71]. Hence, identifying trait-specific QTLs aid in genotype selection that carry both desirable genes or QTLs for oil and fatty acid concentrations. An increase in oleic acid content does not markedly alter the oil content, even though a drastic downregulation of linoleic acid has been observed [72], such as AX-176,817,478 on Arahy.06, AX-147,234,396 on Arahy.09, and AX-147,262,366 on Arahy.19. This finding indicates that oleic acid-related QTLs can assist in biomarker-guided selection for augmented oleic acid varieties without substantially affecting the oil concentration in peanut breeding. Furthermore, constructing an interaction genetic network of oil-related traits holds great value as it allows for the exploration of the underlying genetic mechanisms of lipid metabolism from multiple perspectives. In this study, we found that 44.44% of loci were significantly associated with only one trait, and 8 and 17 loci were significantly associated with two and three traits, respectively. It is vital to acknowledge the interactions among multiple phenotypes that aim to optimize metabolic processes or biological processes [73, 74]. These interactions not only can effectively capture the developmental features of complex traits but also aids in the selection of QTLs for future breeding programs aiming to enhance multiple traits simultaneously [75].

With rapidly advancing next-generation sequencing technologies and the continued decline of related costs, GWAS is quickly emerging as a gold standard for natural variation detection explaining complex quantitative phenotypes in plants [76–78]. GWAS is frequently employed in multiple biological research areas with advantages of identifying several allele loci without establishing a mapping population [79]. Herein, 45 loci on 12 chromosomes were detected for oil, oleic acid, linoleic acid, and O/L traits in 499 peanut accessions. Relative to the previously identified QTLs associated with oil-associated traits via biparental linkage populations [21, 22, 35], more precise QTLs locations were revealed in this investigation. One possible reason for this difference is that we used a SNP array chip harboring more high-density SNP markers than previous studies. Furthermore, we demonstrated copious variation in oil-associated traits in the association panel of 499 peanut accessions, including augmented-oil and high-oleic acid lines. Employing GWAS, we uncovered 45 oil-, oleic acid-, and linoleic acid concentration- and O/L-related loci, as well as 3 previously cloned genes that contribute to oil biosynthesis (*AhFUS3*, *AhFAD2A*, and *AhKCS*). FUS3 transcription factor of B3 family is considered to be a key regulator of seed development stage. Loss-of-function mutations in the *BnFUS3* transcription factor impair lipid biosynthesis by suppressing sucrose transport and glycolytic flux, thereby limiting carbon allocation and acetyl-CoA availability for fatty acid biosynthesis [80]. FAD2A, the fatty acid desaturase, were known to regulate the conversion from oleic acid to linoleic acid [13]. KCS, a β-ketoacyl-CoA synthase, regulated the biosynthesis of very long chain fatty acids in plants. CRISPR-Cas9 editing of conserved *AhKCS* loci enables engineered enhancement of peanut nutritional quality [18]. *AhFAD2B*, located on Arahy.19, was not detected, as there were no polymorphic SNP markers on this gene. Majority QTLs were first shown to be linked to oil, oleic acid, and linoleic acid contents and O/L. Overall, 45 relevant loci were detected, which included 3, 23, 25, and 36 loci in oil, oleic acid, linoleic acid, and O/L, respectively. Among them, 25 were pleiotropic for different traits. For example, 17 significant SNPs (AX-176805166, AX-176802651, AX-176817478, AX-147225829, AX-176815346, AX-176800330, AX-147234396, AX-176808914, AX-176799008, AX-176806119, AX-176801965, AX-176796664, AX-147218189, AX-176819580, AX-176815348, AX-176805249, and AX-147262366) associated with oleic acid content, linoleic acid content, and O/L were detected at the same time. The relevant QTL regions were described as the genomic region within 0.5 Mb upstream and downstream of the relevant SNPs [41, 81].

In all, 45 genes whose functions included fatty acid metabolism (*AhFUS3*, *AhFAD2A*, and *AhKCS*), which strictly control natural variation in fatty acid contents in peanuts. Using the arahy.Tifrunner.gnm2 reference genome, we identified the polymorphic marker AX-147234396 within the coding region of the *AhFAD2A*. This marker corresponds to a G-to-A substitution at position 448, resulting in an aspartic acid-to-asparagine amino acid change (D150N) that significantly increases oleic acid content, consistent with prior functional studies of *AhFAD2A* in peanut lipid biosynthesis [82]. The markers AX-176806463 and AX-176819738 were localized to intergenic regions, with *AhFUS3* and *AhKCS* identified as candidate genes within the 0.5 Mb flanking regions of these loci. In order to further study the functional sites of these two genes, further research is necessary, like increase marker density or haplotype analysis across the *AhFUS3* and *AhKCS* genomic regions to characterize novel allelic variants influencing fatty acid

profiles. In addition, our analysis revealed that a substantial number of the identified loci were associated with non-fatty acid pathway genes, many of which encode enzymes involved in other metabolic processes such as oxidation-reduction reactions, hydrolase activity, glucose metabolism, and transport complexes. The influence of these non-pathway genes on oil traits could be mediated through pleiotropic effects or indirect regulation via metabolic intermediates shared between different pathways. For example, genes involved in glucose metabolism may affect the availability of precursors for fatty acid synthesis. Based on our observations, a substantial population of the identified loci were strongly associated with non-pathway genes and exhibited comparable activities, with 6 loci elucidating more than 20% of the phenotypic variation in 1–3 fatty acid traits. Five genes, *arahy.ZNL5N1*, *arahy.J0DG7J*, *arahy.X7PJ8H*, *arahy.MZJT69*, and *arahy.HR8ZWH* received annotation as non-fatty acid network genes, and the favorable alleles at the leading SNPs elevated the oleic acid levels by 35.50% (AX-147233346), 26.64% (AX-176800330), 35.62% (AX-176795908), 30.26% (AX-147262366), and 47.14% (AX-176798303), in comparison with unfavorable alleles (Additional file 1: Fig. S5 and Additional file 2: Table S5). This phenomenon may stem from the oligogenic regulation of fatty acid profiles in peanut seeds, involving major-effect quantitative trait loci (QTLs) interacting with polygenic modifiers. Non-pathway genes pyramiding might be an effective way to increase oleic acid content in peanut breeding. Further research is needed to elucidate the specific biological mechanisms by which these non-pathway genes impact oil and fatty acid contents in peanuts.

Conclusion

This GWAS was conducted on the seed composition traits of an association panel involving 499 diverse inbred lines, and several relevant SNPs were discovered. Overall, 45 potential genes significantly associated with oil, oleic acid, linoleic acid, and O/L were obtained via GWAS. The QTLs and genes identified in our GWAS can potentially assist breeders to enhance the oil concentration or alter the fatty acid ratios in peanut seeds. It is imperative to continue research in this area to validate the identified potential genes. Elucidation of desirable genes and alleles with substantial impact is critical for selection of specific traits among a more diverse population. Our discovery of significant loci provides essential genetic resources and biomarkers for efficient oil improvement in peanuts and other crops.

Materials and methods

Sample collection, DNA isolation, and genotyping

In all, we employed 499 peanut accessions in this investigation which conserved at the Shandong Peanut Research Institute (Additional file 2: Table S1). Among them, 238 lines were from China, 259 lines were from the USA, one was from Japan and one was from Australia, among which 350 belonged to *A. hypogaea* var. *hypogaea*, 6 to var. *hirsuta*, 46 to var. *vulgaris*, and 26 to var. *fastigiata* plus 71 irregular types (Additional file 2: Table S1). The irregular categories did not belong to any previously described peanut botanical varieties; rather, they are hybrids of the four botanical varieties [83]. To ensure the genetic purity of each accession, we first cultivated all accessions in 2016 and then harvested pods of each accession from individual plants. After that the 499 lines were seeded in 1-row plots in a random block design in the experimental fields of Laixi, Weifang, and Yantai from 2017 to 2020. The oil, oleic acid, and linoleic acid contents of mature seeds were measured via near-infrared reflectance spectroscopy (NIR) [57, 59, 84]. The seed quality NIR estimation was performed via a Fourier transform NIR spectrometer (Bruker, Germany). For every individual sample, we assessed ten seeds, with two replicates. In addition, one ratio trait (O/L) was derived from the oleic and linoleic contents.

Young healthy leaves from each accession were collected, with extraction of total genomic DNA via a Plant Genomic DNA Kit (Tiangen Biotech (Beijing) Co., Ltd.). The extracted DNA quality and concentration were determined via 0.8% agarose gel electrophoresis using standard lambda DNA and a NanoDrop ND-1000 (Thermo Fisher Scientific Inc., Wilmington, DE, USA).

The 48 K SNP array incorporates 11,516 SNPs from the first version, 1,674 haplotype-based SNPs from contrasting sub-genome-specific sequences, and 28,218 newly identified SNPs from the re-sequencing of 21 tetraploid accessions including three parents of the "T" and the "S" populations [85]. All the lines were genotyped via a high-density SNP array Axiom_Arachis with 48 K SNPs, which consists of 47837 SNPs [29] mapped onto a peanut reference genome (Tifrunner.gnm2. (https://www.peanutbase.org/). Whole-genome sequencing was performed at CapitalBio Technology (Beijing) using the Affymetrix GeneTitan platform, where DNA samples from 499 accessions were genotyped following established protocols [86]. This process generated standardized.CEL files for each sample, which were securely archived. In all, 10948 SNPs with minor allelic frequencies (MAFs) ≥ 0.05 and missing data of less than 20% were used for the genotypic data analyses.

Assessment of phenotypic information

All data analyses were conducted in R version 3.1.1 (www.R-project.org). Using the R function 'aov', Analysis of variance (ANOVA) was conducted for individual traits to assess the genotype and environmental effects and broad-sense heritability of each trait. To minimize

environmental influence, we computed the best linear unbiased predictor (BLUP) value for all lines using the linear mixed model, and both genotype and environment as random effects in the R function 'lme4'. Phenotype description statistics was conducted for the BLUP values for all lines, followed by computation of the Pearson correlation coefficients.

GWAS

We conducted GWAS on peanut kernel oil-related traits to elucidate links between genotype and phenotype. To do this, a mixed linear model (MLM) in the TASSEL software was employed, whereby the population structure (Q) and the kinship matrix (K) were regulated to eliminate false-positive correlations [40]. SNPs with MAF ≥ 5% were entered into association analysis. Additionally, we computed the *P* value of all SNPs, and significance was set at 9.13×10^{-5} (i.e., $P \leq 1/n$, n = total markers used).

Linkage disequilibrium analysis

To identify genotype and phenotype association-related genes, we conducted LD analysis on SNPs from the same chromosome that showed strong significance, and threshold was set at < 0.2 for the LD statistic r^2 via TASSEL version 5.2 [87]. Among the identified association signals, we validated multiple genes both in or near (within 500 kb up- and down-stream of the lead SNP) reported oil metabolism–related genes. Among those that were not in or near oil metabolism–related genes, we considered that the link may be with a more distant gene, the closest among which was regarded as the optimal candidate gene. The physical SNP location was screened on the basis of the peanut genomic sequence version arahy.Tifrunner.gnm2 (https://www.peanutbase.org/peanut_genome).

Generation of a trait–locus axis

Using the QTL mapping results, we next generated a trait-locus axis via Cytoscape 3.2.0 [88]. Traits and associated QTLs were regarded as nodes, and the links between traits and QTLs (overlapping on the basis of physical location) were taken as edges. The link thickness between the traits and QTLs represented the QTL PVE value, and the node size indicated QTLs quantity.

Expression analysis of candidate genes

Two RNA-seq datasets, one published [64] and one unpublished (provided by our group), were used for preliminary verification of the expression levels of candidate genes. Expression dynamics of candidate genes were studied in seeds of different peanut varieties and at different developmental stages to gain deeper insights into the role of identified genes.

Abbreviations

BLUP	Best linear unbiased predictor
GWAS	Genome-wide association study
LD	Linkage disequilibrium
MAF	Minor allelic frequency
MLM	Mixed linear model
NIR	Near-infrared reflectance spectroscopy
O/L	The oleic acid to linoleic acid ratio
PVE	Explained phenotypic variation
QTLs	Quantitative trait loci
SNP	Single nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06690-9.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

Not applicable.

Author contributions

XC, MY and JX conceived the study, designed the experiments and finished the final version of the manuscript. JX drafted the manuscript. JX and XJ analyzed the data. XY, XZ, NC, LP, CF, YJ and JM participated in phenotypic data collection and prepared the genomic DNA for 48 K SNP array chip sequencing. All authors reviewed the manuscript. All the authors read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of Shangdong Province (ZR2021QC172, ZR2023QC146); Major scientific and technological project in Xinjiang (2022A02008-3); Taishan Scholars Program (tstp20240523, tsqn202312292); China Agriculture Research System of MOF and MARA (CARS-13); Key R&D Program of Shandong Province (2024LZGC035); Key research and development plan of Shandong Province (2018GNC110036, 2022TZXD0031); the innovation Project of SAAS (CXGC2023F20, CXGC2024F20, CXGC2024G20); Open Project of Key Laboratory of Digital Upland Crops of Zhejiang Province (2022E10012); Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, P. R. China (KF2024007).

Data availability

The data that support the findings of this study are included in the manuscript. The all sequence data was uploaded to NCBI with accession code GSE289195.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 January 2025 / Accepted: 7 May 2025
Published online: 16 May 2025

References

1. Chen X, Li H, Pandey MK, Yang Q, Wang X, Garg V, Li H, Chi X, Doddamani D, Hong Y. Draft genome of the peanut A-genome progenitor (*Arachis*

duranensis) provides insights into geocarpy, oil biosynthesis, and allergens. PNAS. 2016;113(24):6785–90.
2. Zhang L, Yang X, Tian L, Chen L, Yu W. Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. New Phytol. 2016;211(4):1424–39.
3. Zhao Y, Ma J, Li M, Deng L, Li G, Xia H, Zhao S, Hou L, Li P, Ma C. Whole-genome resequencing-based QTL-seq identified *AhTc1* gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. Plant Biotechnol J. 2020;18(1):96–105.
4. Shasidhar Y, Vishwakarma MK, Pandey MK, Janila P, Variath MT, Manohar SS, Nigam SN, Guo B, Varshney RK. Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis Hypogaea* L). Front Plant Sci. 2017;8:794.
5. Isleib TG, Pattee HE, Sanders TH, Hendrix KW, Dean LO. Compositional and sensory comparisons between normal-and high-oleic peanuts. J Agric Food Chem. 2006;54(5):1759–63.
6. Vassiliou EK, Gonzalez A, Garcia C, Tadros JH, Toney JH. Oleic acid and peanut oil high in oleic acid reverse the inhibitory effect of insulin production of the inflammatory cytokine TNF-α both in vitro and in vivo systems. Lipids Health Dis. 2009;8(1):25–25.
7. Pattee HE, Isleib TG, Moore KM, Gorbet DW, Giesbrecht FG. Effect of high-oleic trait and paste storage variables on sensory attribute stability of roasted peanuts. J Agric Food Chem. 2002;50(25):7366–70.
8. Baud S, Lepiniec L. Physiological and developmental regulation of seed oil production. Prog Lipid Res. 2010;49(3):235–49.
9. Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP. Acyl-lipid metabolism. Arabidopsis Book. 2013;11:e0161.
10. Saha S, Enugutti B, Rajakumari S, Rajasekharan R. Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic Diacylglycerol acyltransferase. Plant Physiol. 2006;141(4):1533–43.
11. Chi X, Hu R, Zhang X, Chen M, Chen N, Pan L, Wang T, Wang M, Yang Z, Wang Q. Cloning and functional analysis of three diacylglycerol acyltransferase genes from peanut (*Arachis Hypogaea* L). PLoS ONE. 2014;9(9):e105834.
12. Jung S, Powell G, Moore K, Abbott A. The high oleate trait in the cultivated peanut [*Arachis hypogaea* L.]. II. molecular basis and genetics of the trait. Mol Gen Genet. 2000;263(5):806–11.
13. Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, Tonnis B, Barkley NA, Wang J, Holbrook CC. Identification of QTLs associated with oil content and mapping *FAD2* genes and their relative contribution to oil quality in peanut (*Arachis Hypogaea* L). BMC Genet. 2014;15(1):133.
14. Pan L, Zhang J, Chi X, Chen N, Chen M, Wang M, Wang T, Yang Z, Zhang Z, Wan Y. The antisense expression of *AhPEPC1* increases seed oil production in peanuts (*Arachis Hypogaea* L). Grasas Aceite. 2016;67(4):e164.
15. Tang G, Xu P, Ma W, Wang F, Liu Z, Wan S, Shan L. Seed-specific expression of *AtLEC1* increased oil content and altered fatty acid composition in seeds of peanut (*Arachis Hypogaea* L). Front Plant Sci. 2018;9:260.
16. Li X, Xue L, Liu H, Qu P, Zhao H, Luo D, Wang X, Huang B, Zhang M, Li C. *AhFAD3-A01* enhances α-linolenic acid content in Arabidopsis and peanut. Gene. 2025;949:149336.
17. Shen Y, Shen Y, Liu Y, Bai Y, Liang M, Zhang X, Chen Z. Characterization and functional analysis of *AhGPAT9* gene involved in lipid synthesis in peanut (*Arachis Hypogaea* L). Front Plant Sci. 2023;14:1144306.
18. Huai D, Xue X, Wu J, Pandey MK, Liu N, Huang L, Yan L, Chen Y, Wang X, Wang Q. Enhancing peanut nutritional quality by editing *AhKCS* genes lacking natural variation. Plant Biotechnol J. 2024;22(11):3015–17.
19. Zhan Y, Wang J, Zhao X, Zheng Z, Gan Y. *Arachis hypogaea* monoacylglycerol lipase AhMAGL3b participates in lipid metabolism. BMC Plant Biol. 2024;24(1):1278.
20. Jadhav MP, Gangurde SS, Hake AA, Yadawad A, Mahadevaiah SS, Pattanashetti SK, Gowda MVC, Shirasawa K, Varshney RK, Pandey MK. Genotyping-by-sequencing based genetic mapping identified major and consistent genomic regions for productivity and quality traits in peanut. Front Plant Sci. 2021;12:668020.
21. Gomez Selvaraj M, Narayana M, Schubert AM, Ayers JL, Baring MR, Burow MD. Identification of QTLs for pod and kernel traits in cultivated peanut by bulked Segregant analysis. Electron J Biotechnol. 2009;12(2):3–4.
22. Sarvamangala C, Gowda M, Varshney R. Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis Hypogaea* L). Field Crop Res. 2011;122(1):49–59.
23. Wang ML, Khera P, Pandey MK, Wang H, Qiao L, Feng S, Tonnis B, Barkley NA, Pinnow D, Holbrook CC. Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis Hypogaea* L). PLoS ONE. 2015;10(4):e0119454.
24. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK, Liu X, Gao D, Clevenger J, Dash S. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48(4):438–46.
25. Lu Q, Li H, Hong Y, Zhang G, Wen S, Li X, Zhou G, Li S, Liu H, Liu H. Genome sequencing and analysis of the peanut B-genome progenitor (*Arachis ipaensis*). Front Plant Sci. 2018;9:604.
26. Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H, Li H, Wang J, Liu H, Li S. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Mol Plant. 2019;12(7):920–34.
27. Yin D, Ji C, Song Q, Zhang W, Zhang X, Zhao K, Chen CY, Wang C, He G, Liang Z. Comparison of *Arachis monticola* with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. Adv Sci. 2019;1901672.
28. Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, Chang W-C, Zhang L, Zhang X, Tang R. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019;51(5):865–76.
29. Pandey MK, Agarwal G, Kale SM, Clevenger J, Nayak SN, Sriswathi M, Chitikineni A, Chavarro C, Chen X, Upadhyaya HD. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. Sci Rep. 2017;7(1).
30. Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, Chitikineni A, Rathore A, Varshney RK. Emerging genomic tools for legume breeding: current status and future prospects. Front Plant Sci. 2016;7:455.
31. Hu P, Zhang J, Song Y, Zhao X, Jin X, Su Q, Yang Y, Wang J. Identification of putative quantitative trait loci for improved seed oil quality in peanuts. Genes. 2024;15(1):75.
32. Wang F, Miao H, Zhang S, Hu X, Li C, Yang W, Chen J. Identification of a new major oil content Qtl overlapped with *FAD2B* in cultivated peanut (*Arachis Hypogaea* L). Plants. 2025;14(4):615.
33. Yang Y, Li Y, Cheng Z, Su Q, Jin X, Song Y, Wang J. Genetic analysis and exploration of major effect QTLs underlying oil content in peanut. Theor Appl Genet. 2023;136(5):97.
34. Xue X, Li J, Wu J, Hu M, Liu N, Yan L, Chen Y, Wang X, Kang Y, Wang Z, et al. Identification of QTLs associated with very-long chain fatty acid (VLCFA) content via linkage mapping and BSA-seq in peanut. Theor Appl Genet. 2024;137(2):33.
35. Hu X, Zhang S, Miao H, Cui F, Shen Y, Yang W, Xu T, Chen N, Chi X, Zhang Z. High-density genetic map construction and identification of QTLs controlling oleic and linoleic acid in peanut using SLAF-seq and SSRs. Sci Rep. 2018;8(1):5479.
36. Zhou F, Liu Y, Liang C, Wang W, Li C, Guo Y, Ma J, Yu Y, Fan L, Yao Y, et al. Construction of a high-density genetic linkage map and QTL mapping of oleic acid content and three agronomic traits in sunflower (*Helianthus annuus* L.) using specific-locus amplified fragment sequencing (SLAF-seq). Breed Sci. 2018;68(5):596–605.
37. Liu N, Guo J, Zhou X, Wu B, Huang L, Luo H, Chen Y, Chen W, Lei Y, Huang Y. High-resolution mapping of a major and consensus quantitative trait locus for oil content to a ~ 0.8-Mb region on chromosome A08 in peanut (*Arachis Hypogaea* L). Theor Appl Genet. 2020;133(1):37–49.
38. Sun Z, Qi F, Liu H, Qin L, Xu J, Shi L, Zhang Z, Miao L, Huang B, Dong W. QTL mapping of quality related traits in peanut using whole-genome resequencing. Crop J. 2021;10:177–84.
39. Zhang X, Zhang J, He X, Wang Y, Ma X, Yin D. Genome-wide association study of major agronomic traits related to domestication in peanut. Front Plant Sci. 2017;8:1611.
40. Wang J, Yan C, Li Y, Li C, Zhao X, Yuan C, Sun Q, Shan S. GWAS discovery of candidate genes for yield-related traits in peanut and support from earlier QTL mapping studies. Genes. 2019;10(10):803.
41. Zhang H, Wang ML, Schaefer R, Dang P, Jiang T, Chen C. GWAS and coexpression network reveal ionomic variation in cultivated peanut. J Agric Food Chem. 2019;67(43):12026–36.
42. Zhang H, Wang ML, Dang P, Jiang T, Zhao S, Lamb M, Chen C. Identification of potential QTLs and genes associated with seed composition traits in peanut (*Arachis Hypogaea* L.) using GWAS and RNA-Seq analysis. Gene. 2021;769:145215.

43. Wang J, Chen H, Li Y, Shi D, Wang W, Yan C, Yuan M, Sun Q, Chen J, Mou Y. Identification of quantitative trait nucleotides and development of diagnostic markers for nine fatty acids in the peanut. Plants. 2023;13(1).
44. Zhou X, Luo H, Yu B, Huang L, Liu N, Chen W, Liao B, Lei Y, Huai D, Guo P. Genetic dissection of fatty acid components in the Chinese peanut (*Arachis Hypogaea* L.) mini-core collection under multi-environments. PLoS ONE. 2022;17(12):e0279650.
45. Lu Q, Huang L, Liu H, Garg V, Gangurde SS, Li H, Chitikineni A, Guo D, Pandey MK, Li S. A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. Nat Genet. 2024;56(3):530–40.
46. Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimarães P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B. Advances in *Arachis* genomics for peanut improvement. Biotechnol Adv. 2012;30(3):639–51.
47. Wang ML, Wang H, Zhao C, Tonnis B, Tallury S, Wang X, Clevenger J, Guo B. Identification of QTLs for seed dormancy in cultivated peanut using a recombinant inbred line mapping population. Plant Mol Biol Rep. 2021;40(1):1–10.
48. Bomireddy D, Sharma V, Gangurde SS, Mohinuddin DK, Kumar R, Senthil R, Singh K, Reddisekhar M, Bera SK, Pandey MK. Multi-locus genome wide association study uncovers genetics of fresh seed dormancy in groundnut. BMC Plant Biol. 2024;24(1):1258.
49. Zou K, Kim K-S, Kang D, Kim M-C, Ha J, Moon J-K, Jun T-H. Genome-wide association study of leaf chlorophyll content using high-density SNP array in peanuts (*Arachis Hypogaea* L). Agronomy. 2022;12(1):152.
50. Li L, Cui S, Dang P, Yang X, Wei X, Chen K, Liu L, Chen CY. GWAS and bulked segregant analysis reveal the loci controlling growth habit-related traits in cultivated peanut (*Arachis Hypogaea* L). BMC Genomics. 2022;23(1):403.
51. Wankhade AP, Chimote VP, Viswanatha KP, Yadaru S, Deshmukh DB, Gattu S, Sudini HK, Deshmukh MP, Shinde VS, Vemula AK. Genome-wide association mapping for LLS resistance in a MAGIC population of groundnut (*Arachis Hypogaea* L). Theor Appl Genet. 2023;136(3):43.
52. Wu D, Zhao C, Korani W, Thompson EA, Wang H, Agarwal G, Fountain JC, Culbreath A, Holbrook CC, Wang X. High-resolution genetic and physical mapping reveals a peanut spotted wilt disease resistance locus, *PSWDR-1*, to tomato spotted wilt virus (TSWV), within a recombination cold-spot on chromosome A01. BMC Genomics. 2025;26(1):224.
53. Gangurde SS, Pasupuleti J, Parmar S, Variath MT, Bomireddy D, Manohar SS, Varshney RK, Singam P, Guo B, Pandey MK. Genetic mapping identifies genomic regions and candidate genes for seed weight and shelling percentage in groundnut. Front Genet. 2023;14:1128182.
54. Patel JD, Wang ML, Dang P, Butts C, Lamb M, Chen CY. Insights into the genomic architecture of seed and pod quality traits in the U.S. peanut mini-core diversity panel. Plants. 2022;11(7).
55. Zou K, Kim K-S, Kim K, Kang D, Park Y-H, Sun H, Ha B-K, Ha J, Jun T-H. Genetic diversity and genome-wide association study of seed aspect ratio using a high-density SNP array in peanut (*Arachis Hypogaea* L). Genes. 2020;12(1):2.
56. Fox G, Cruickshank A. Near infrared reflectance as a rapid and inexpensive surrogate measure for fatty acid composition and oil content of peanuts (*Arachis Hypogaea* L). J Near Infrared Spectrosc. 2005;13(5):287–91.
57. Tillman BL, Gorbet DW, Person G. Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. Crop Sci. 2006;46(5):2121–26.
58. Nadaf H, Hanchinal R. Near infrared reflectance spectroscopy (NIRS) for large scale screening of fatty acid profile in peanut (*Arachis Hypogaea* L). Legume Res. 2014;37(3):272–80.
59. Wang CT, Wang XZ, Tang YY, Wu Q, Xu JZ, Hu DQ, Qu B. Predicting main fatty acids, oil and protein content in intact single seeds of groundnut by near infrared reflectance spectroscopy. Adv Mater Res. 2014;860–863:490–6.
60. Lee JM, Pae S-B, Choung M-G, Lee M-H, Kim S-U, Oh E-y, Oh K-W, Jung C-S, Oh IS. Determination of fatty acid composition in peanut seed by near infrared reflectance spectroscopy. Korean J Crop Sci. 2016;61(1):64–9.
61. Clevenger JP, Korani W, Ozias-Akins P, Jackson S. Haplotype-based genotyping in polyploids. Front Plant Sci. 2018;9:564.
62. A.Knauft MMK. D. The inheritance of high oleic acid in peanut. J Hered. 1989;80(3).
63. Li B, Tang M, Nelson A, Caligagan H, Zhou X, Clark-Wiest C, Ngo R, Brady SM, Kliebenstein DJ. Network-guided discovery of extensive epistasis between transcription factors involved in aliphatic glucosinolate biosynthesis. Plant Cell. 2018;30(1):178–95.
64. Clevenger J, Chu Y, Scheffler B, Ozias-Akins P. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. Front Plant Sci. 2016;7:1446.
65. Kochert G, Stalker HT, Gimenes M, Galgaro L, Lopes CR, Moore K. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (*Leguminosae*). Am J Bot. 1996;83(10):1282–91.
66. Wang Z, Huai D, Zhang Z, Cheng K, Kang Y, Wan L, Yan L, Jiang H, Lei Y, Liao B. Development of a high-density genetic map based on specific length amplified fragment sequencing and its application in quantitative trait loci analysis for yield-related traits in cultivated peanut. Front Plant Sci. 2018;9:827.
67. Li L, Li L, Yang X, Yang X, Cui S, Meng X, Mu G, Hou M, He M, Zhang H et al. Construction of high-density genetic map and mapping quantitative trait loci for growth habit-related traits of peanut (*Arachis Hypogaea* L). Front Plant Sci. 2019;10.
68. Zhang S, Hu X, Miao H, Chu Y, Cui F, Yang W, Wang C, Shen Y, Xu T, Zhao L. QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (*Arachis Hypogaea* L). BMC Plant Biolo. 2019;19(1):537.
69. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22.
70. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467–84.
71. Sungwoo L, Kyujung V, Mikyung S, Randall N, Jonathan L, K ML, Rouf MMA. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. Theor Appl Genet. 2019;132(6):1639–59.
72. Li WM, Y CC, Brandon T, David P, Jerry D, Charles AY-Q, Phat D. Changes of seed weight, fatty acid composition, and oil and protein contents from different peanut *FAD2* genotypes at different seed developmental and maturation stages. J Agric Food Chem. 2018;66(14):3658–65.
73. Lu K, Zhou Z, Huang Z, Bu C, Gong H, Jiang L, Zhang D, Fang Q, Zhang XY, Song Y. Unveiling the genetic networks: exploring the dynamic interaction of photosynthetic phenotypes in woody plants across varied light gradients. Plant Physiol Biochem. 2025;221:109616.
74. Fang H, Fu X, Wang Y, Xu J, Feng H, Li W, Xu J, Jittham O, Zhang X, Zhang L. Genetic basis of kernel nutritional traits during maize domestication and improvement. Plant J. 2020;101(2):278–92.
75. Dong C, Zhang L, Zhang Q, Yang Y, Li D, Xie Z, Cui G, Chen Y, Wu L, Li Z. Tiller Number1 encodes an ankyrin repeat protein that controls tillering in bread wheat. Nat Commun. 2023;14(1):836.
76. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature. 2010;465(7298):627–31.
77. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z. Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet. 2010;42(11):961–67.
78. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013;45(1):43–50.
79. Liu H, Cao K, Zhu G, Fang W, Chen C, Wang X, Wang L. Genome-wide association analysis of red flesh character based on resequencing approach in Peach. J Am Soc Hortic Sci. 2019;144(3):209–16.
80. Elahi N, Duncan RW, Stasolla C. Decreased seed oil production in *FUSCA3 Brassica napus* mutant plants. Plant Physiol Biochem. 2015;96:222–30.
81. Sharma V, Mahadevaiah SS, Latha P, Gowda SA, Manohar SS, Jadhav K, Bajaj P, Joshi P, Anitha T, Jadhav MP. Dissecting genomic regions and underlying candidate genes in groundnut MAGIC population for drought tolerance. BMC Plant Biol. 2024;24(1):1044.
82. López Y, Nadaf HL, Smith OD, Simpson CE, Fritz AK. Expressed variants of Δ12-fatty acid desaturase for the high oleate trait in Spanish market-type peanut lines. Mol Breed. 2002;9(3):183–92.
83. Zhang H. Construction of core collections and genetic diversity analysis of *Arachis hypogaea* L. cultivars in China. China: ZhongKai University of Agriculture and Engineering Guangzhou; 2013.
84. Kandala CV, Sundaram J, Puppala N. Analysis of moisture content, total oil and fatty acid composition by NIR reflectance spectroscopy: a review. LNEE. 2012;146:59–80.
85. Qin H, Feng S, Chen C, Guo Y, Knapp S, Culbreath A, He G, Wang ML, Zhang X, Holbrook CC. An integrated genetic linkage map of cultivated peanut (*Arachis Hypogaea* L.) constructed from two RIL populations. Theor Appl Genet. 2012;124(4):653–64.
86. Gangurde SS, Wang H, Yaduru S, Pandey MK, Fountain JC, Chu Y, Isleib T, Holbrook CC, Xavier A, Culbreath AK. Nested-association mapping (NAM)-based

genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). Plant Biotechnol J. 2020;18(6):1457–71.
87. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60.
88. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504.

Publisher's note