QTL Mapping of Quality Related Traits in Peanut Using Whole-Genome Resequencing

Ziqi Sun
Henan Academy Of Crops Molecular Breeding
Feiyan Qi
Henan Academy of Crops Molecular Breeding
Hua Liu
Henan Academy of Crops Molecular Breeding
Li Qin
Henan Academy of Crops Molecular Breeding
Jing Xu
Henan Academy Of Sciences
Lei Shi
Henan Academy of Crops Molecular Breeding
Zhongxin Zhang
Henan Academy Of Crops Molecular Breeding
Lijuan Miao
Henan Academy of Crops Molecular Breeding
Bingyan Huang
Henan Academy Of Crops Molecular Breeding
Wenzhao Dong
Henan Academy of Crops Molecular Breeding
Xiao Wang
Henan Academy of Crops Molecular Breeding
Mengdi Tian
Henan Academy of Crops Molecular Breeding
Jingjing Feng
Henan Academy of Crops Molecular Breeding
Ruifang Zhao
Henan Academy of Crops Molecular Breeding
Zheng Zheng
Henan Academy of Crops Molecular Breeding
Xinyou Zhang (haasz@126.com)
Henan Academy of Agricultural Sciences

Research article

Keywords: QTL mapping, peanut, oil content, fatty acid, whole-genome resequencing

Posted Date: January 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-138961/v1

Abstract

Background: Oil and protein content, as well as fatty acid composition, are important quality traits in peanut. Elucidating the genetic mechanisms underlying these traits may help researchers to obtain improved cultivars through molecular breeding techniques.

Results: Whole-genome resequencing of an RIL population of 318 lines was performed to construct a high-density linkage map and identify QTLs for peanut quality. The map, containing 4561 bin markers, covered a length of 2032.39 cM with an average marker density of 0.45 cM. A total of 109 QTLs for oil content, protein content, and fatty acid compositions were mapped on the 18 peanut chromosomes. The QTL *qA05.1* was detected in four different environments and exhibited a major phenotypic effect on the content of oil, proteins, and six fatty acids. The genomic region spanned by *qA05.1*, corresponding to a physical interval of approximately 1.50 Mb, contains two polymorphic SNPs between two parents that could cause missense mutations. The two SNP sites were employed as KASP markers and validated using lines with extremely high and low oil contents; these sites may be useful in the marker-assisted breeding of peanut varieties with high oil contents.

Conclusions: A high-density genetic map with 4561 bin markers was constructed, and a major and pleiotropic QTL located on LG05 was stably detected for oil, protein and fatty acids across four different environments.

Background

Peanut (*Arachis hypogaea* L., 2n = 4x = 40) is one of the most important industrial crops worldwide [1]. In China and India, peanut seeds are primarily utilized as a source of vegetable oil, whereas in Western countries, they are also consumed as edible food products [2–4]. Peanut seeds primarily contain oil (approximately 40–56%) along with proteins (approximately 20–30%) and carbohydrates (approximately 10–20%) [4]. Peanut oil is composed of different fatty acids, including palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), arachidic acid (C20:0), behenic acid (C22:0) and arachidonic acid (C20:1). The ratio between saturated (palmitic, stearic, arachidic, and behenic) and unsaturated (oleic, linoleic and arachidonic) fatty acids is approximately 1:4 [5]. Peanut proteins contain essential amino acids that are easily adsorbed by the human body [6].

Depending on final product destinations, protein and oil content are important quality traits in peanut. In addition, fatty acid composition, particularly oleic acid content, is one of the most important quality traits, as oleic acid can increase the shelf life of peanut products and is beneficial for human health [7]. Elucidating the genetic mechanisms underlying the quality traits mentioned above may help researchers to develop improved cultivars through marker-assisted breeding.

Several quantitative trait loci (QTLs) have been reported to be linked with oil content and fatty acid composition in peanut [1–4, 8]. Using two RIL populations and simple sequence repeat (SSR) markers, a total of 78 main-effect QTLs and 10 epistatic QTLs were detected for oil content and oil quality traits [2]. Another SSR-based QTL mapping study identified 12 QTLs for eight quality-related traits with phenotypic variation explained (PVE) values ranging from approximately 1.72–20.20% [3]. Two major QTLs located on chromosomes A02 and A10 and twenty major QTLs on chromosomes A05, A07-A10, B01, B04 and B09 were detected for oil content and

fatty acid compositions, respectively, using two F_2 populations and DArT markers [4]. Finally, Liu et al. [1], utilizing SNP markers from ddRAD sequencing, mapped the major and consensus QTL *qOCA08.1* to an approximately 0.8-Mb genomic region containing two annotated genes predicted to affect oil biosynthesis.

With advances in next-generation sequencing (NGS) technology and the availability of reference genomes for diploid progenitors and cultivated peanut [9–12], high-resolution mapping has been successfully performed for complex traits in peanut, such as yield [13, 14] and disease resistance [15–17]. In peanut, various NGS methods have been employed to generate a large number of SNPs, such as restriction-site-associated DNA sequencing (RAD-seq) [17–18], single nucleotide polymorphism (SNP) array [19–20], specific-locus amplified fragment sequencing (SLAF-seq) [13, 21], diversity array technology (DArT) [4, 22], and whole genome resequencing (WGRS) [15].

In this study, WGRS was applied to an RIL population segregating for peanut quality traits, which enabled the mapping of QTLs for the content of oil, protein, and seven fatty acids. Candidate genes involved in the pathway of fat biosynthesis were predicted, and SNPs covered by these genes were validated using lines with extremely high and low oil contents. The results of this study may help to establish a foundation for further genetic research and for the development of high-oil peanut cultivars.

Results

Phenotypic analysis of the quality related traits

Nine quality-related traits, i.e., the contents of oil, protein, palmitic acid, stearic acid, arachidic acid, behenic acid, oleic acid, linoleic acid, and arachidonic acid, were assessed in four environments (Zhengzhou, 2018 and 2019; Shangqiu and Weifang, 2019) on the male parental line P1 (W1202), the female parental line P2 (Yuhua15), and 329 segregating RILs. ANOVA indicated that genotypic effects significantly affected all the traits (Table 1). P1 exhibited higher contents of oleic, behenic, and arachidonic acids, whereas P2 exhibited higher contents of oil, proteins, palmitic acid, stearic acid, linoleic acid and arachidic acid (Table 1). For all the traits and environments, wide phenotypic variation and transgressive segregation were observed in the RIL population (Table 1 and Fig. 1). The CV ranged from 4.16% for oil content to 20.65% for arachidonic acid content, while broad-sense heritability ranged from 0.74 for linoleic acid content to 0.91 for behenic acid content (Table 1).

Table 1
Basic statistics and genetic heritability of nine quality traits

Trait	P1 (%)	P2 (%)	Range (%)	Mean (%)	SD	CV (%)	Skewness	Kurtosis	Sig	H^2
Oil	51.88 ~ 54.29	53.80 ~ 55.43	45.15 ~ 59.80	52.97	2.20	4.16	-0.08	-0.20	***	0.90
Protein	22.97 ~ 23.59	22.70 ~ 24.13	18.15 ~ 28.96	23.78	1.56	6.57	-0.06	0.02	***	0.86
Palmitic acid	11.62 ~ 12.85	12.10 ~ 12.83	8.3 ~ 14.75	12.42	0.73	5.86	-0.35	2.13	***	0.83
Stearic acid	3.15 ~ 4.22	4.12 ~ 4.44	1.03 ~ 5.84	3.88	0.62	15.86	-0.20	0.31	***	0.89
Oleic acid	37.94 ~ 44.54	37.03 ~ 40.85	27.87 ~ 57.77	40.64	4.06	9.99	0.32	1.22	***	0.78
Linoleic acid	34.55 ~ 39.40	36.62 ~ 41.17	21 ~ 51.15	37.35	3.68	9.86	-0.20	0.83	***	0.74
Arachidic acid	1.38 ~ 1.60	1.50 ~ 1.67	0.96 ~ 1.93	1.48	0.14	9.40	-0.08	0.13	***	0.89
Behenic acid	2.13 ~ 2.28	2.01 ~ 2.27	1.55 ~ 2.88	2.12	0.18	8.58	0.11	-0.20	***	0.91
ArArachidonic acid	0.65 ~ 0.93	0.52 ~ 0.69	0.19 ~ 1.41	0.72	0.15	20.65	0.32	1.11	***	0.88

P1 male parent; *P2* female parent; *SD* standrad deviation; *CV* coefficient of variation; *Sig* significance; H^2 heritability per mean

The oil content displayed a negative correlation with the protein (-0.79), oleic acid (-0.39) and arachidonic acid (-0.53) contents and exhibited a positive correlation with the palmitic acid (0.40), stearic acid (0.75), linoleic acid (0.12), arachidic acid (0.87), and behenic acid (0.85) contents (Table 2). In addition, negative correlations were observed between protein content and behenic acid (-0.83), arachidic acid (-0.62) and palmitic acid (-0.33) contents. Among fatty acids, a strong negative correlation was observed between oleic and linoleic acid (-0.91), as well as between stearic and arachidonic acid (-0.87), whereas a strong positive correlation was observed between stearic and arachidic acid (0.86).

Table 2
Pairwise correlation among the contents of oil, proteins, and six different fatty acids

	Oil	Protein	Palmitic acids	Stearic acids	Oleic acids	Linoleic acids	Arachidic acids	Behenic acids
Protein	-0.79***							
Palmitic acid	0.40***	-0.06						
Stearic acid	0.75***	-0.33***	0.78***					
Oleic acid	-0.39***	0.21***	-0.91***	-0.64***				
Linoleic acid	0.12*	0.01	0.78***	0.41***	-0.91***			
Arachidic acid	0.87***	-0.62***	0.53***	0.86***	-0.50***	0.18**		
Behenic acid	0.85***	-0.83***	0.28***	0.53***	-0.30***	-0.02	0.77***	
Arachidonic acid	-0.53***	0.08	-0.74***	-0.87***	0.66***	-0.50***	-0.78***	-0.27***

*, **, and *** denote significance levels of 0.05, 0.01 and 0.001, respectively.

Snp And Bin Marker Discovery Through Whole-genome Sequencing

Whole-genome resequencing of the two parental lines and 329 RILs generated approximately 700 Gb of clean data (9.10 billion reads). For each sample, the rate of mapped reads and the rate of mapped reads with unique positions were over 96% and 73%, respectively. The effective sequencing depths were 34.42 × and 34.58 × for P1 and P2, respectively, and ranged from 1.20 × to 1.40 × for the RIL population (Table S1). The coverage rate was 99.1% for P1 and 98.49% for P2 and ranged from 52.03–63.99% for the RIL population (Table S1). All the clean sequence data obtained in this study are available from the NCBI database under Sequence Read Archive (SRA) submission SUB8691701. Following alignment and the application of the GATK protocol, 741,564 SNPs were obtained. Further filtering enabled the definition of 213,868 SNPs homozygous and polymorphic between the two parents, which were utilized to identify bin markers.

Construction Of A High-density Genetic Linkage Map

As the RIL population was sequenced at low depth, the SNP dataset was converted into bin markers using a sliding window approach [23]. In total, 7595 bin markers were detected, and eleven lines exceeding 10% of the heterozygosis rate were removed from further analysis (Table S2). After redundant markers were filtered out, 4565 bin markers were finally used to construct a linkage map. Four bin markers remained unlinked, whereas the remaining 4561 bin markers were assigned to 20 linkage groups (LGs), as reported in Fig. 2 and Table 3. As the total map length was 2032.39 cM, the average map density per marker was 0.45 cM (Table 3). The number of

markers per LG ranged from 173 (LG11) to 323 (LG13), the LG length varied from 77.50 cM (LG20) to 170.15 cM (LG06), and the average marker density per marker ranged between 0.37 cM (LG15) and 0.59 cM (LG06) (Table 3). The maximum marker interval was 13.41 cM on LG06, while more than 90% of marker intervals were below 1 cM (Table 3).

Table 3
Summary of the high-density linkage groups (LGs) obtained for the RIL population

ID	Number of Markers	Map Length (cM)	Marker Density (cM)	Max Interval (cM)	Ratio of marker interval < = 1 cM
LG01	233	97.08	0.42	1.96	93.13%
LG02	206	103.91	0.50	3.03	89.81%
LG03	319	148.74	0.47	6.62	93.73%
LG04	198	100.87	0.51	7.84	89.39%
LG05	228	83.94	0.37	2.13	96.05%
LG06	290	170.15	0.59	13.41	91.38%
LG07	189	85.32	0.45	2.65	89.42%
LG08	260	102.03	0.39	2.48	94.62%
LG09	197	94.22	0.48	2.48	88.83%
LG10	189	86.80	0.46	3.91	89.42%
LG11	173	95.53	0.55	3.36	86.13%
LG12	212	106.07	0.50	5.62	90.09%
LG13	323	125.53	0.39	3.37	94.74%
LG14	235	92.18	0.39	1.96	93.19%
LG15	260	95.67	0.37	8.32	94.62%
LG16	189	79.55	0.42	3.91	89.95%
LG17	232	99.30	0.43	2.31	90.95%
LG18	197	91.59	0.46	3.72	90.86%
LG19	237	96.40	0.41	2.30	91.98%
LG20	194	77.50	0.40	2.65	90.72%
Whole Genome	4561	2032.39	0.45		91.45%

Identification Of Qtl For Peanut Quality-related Traits

With a LOD threshold of 3.3 being employed, 109 QTLs were identified for the nine quality-related traits under investigation. QTLs were distributed on all the LGs, except for LG15 and LG19 (Tables 4 and 5 and Table S3).

Table 4
QTLs identified on LG05 for oil, protein and fatty acid content in four environments

QTL	Position	Marker Interval	LOD	PVE (%)	Trait (Environment; Additive Effect)
qA05.1	0.0 ~ 0.5	bin1572 ~ bin1581	4.77 ~ 28.44	0.76 ~ 26.99	Oil (2018ZZ, 2019ZZ, 2019SQ, 2019WF; -0.88~-0.65); Protein (2018ZZ, 2019ZZ, 2019WF; 0.25 ~ 0.28); Palmitic (2018ZZ, 2019ZZ, 2019SQ, 2019WF; -0.22~-0.13); Stearic (2018ZZ, 2019ZZ, 2019SQ, 2019WF; -0.25~-0.19); Oleic (2018ZZ, 2019ZZ; 0.84 ~ 0.95); Arachidic (2018ZZ, 2019ZZ, 2019SQ, 2019WF; -0.05~-0.04); Behenic (2018ZZ, 2019ZZ, 2019SQ, 2019WF; -0.07~-0.05); Arachidonic (2018ZZ, 2019ZZ, 2019SQ; 0.04)
qA05.2	6.0	bin1593 ~ bin1594	52.35	10.39	Linoleic (2018ZZ; 1.832)
qA05.3	6.7	bin1598 ~ bin1600	70.92	16.43	Linoleic (2018ZZ; -2.30)
qA05.4	7.3	bin1601 ~ bin1602	8.25	5.41	Oleic (2019WF; 0.96)
qA05.5	10.2	bin1611 ~ bin1612	5.53 ~ 9.38	5.37 ~ 11.27	Stearic (2019WF; -0.13); Arachidonic (2019WF; 0.05)
qA05.6	42.9	bin1790 ~ bin1791	3.60	2.72	Protein (2019WF; 0.21)
qA05.7	46.3	bin1798 ~ bin1800	3.31	2.42	Behenic (2019SQ; -0.03)
qA05.8	48.2 ~ 48.6	bin1802 ~ bin1803	7.64 ~ 9.77	5.23 ~ 9.84	Oil (2019SQ; -0.48); Protein (2019SQ; 0.42)
qA05.9	51.7 ~ 52.7	bin1808 ~ 1812	4.37 ~ 16.16	2.83 ~ 8.68	Protein (2019ZZ; 0.22); Palmitic (2019SQ; -0.25); Stearic (2019SQ; -0.19); Oleic (2019SQ; 1.25); Arachidonic (2019SQ; 0.05)
qA05.10	55.3 ~ 56.8	bin1817 ~ bin1824	5.05 ~ 39.85	0.29 ~ 15.18	Palmitic (2018ZZ, 2019ZZ, 2019SQ, 2019WF; 0.15 ~ 0.42); Stearic (2018ZZ, 2019SQ, 2019WF; 0.09 ~ 0.27); Oleic (2018ZZ, 2019SQ, 2019WF; -2.08~-0.73); Linoleic (2018ZZ, 2019SQ, 2019WF; 0.77 ~ 1.02); Arachidonic (2018ZZ, 2019SQ; -0.07~-0.03)

ZZ represents Zhengzhou, *SQ* represents Shangqiu and *WF* represents Weifang.

QTL	Position	Marker Interval	LOD	PVE (%)	Trait (Environment; Additive Effect)
qA05.11	66.4	bin1852 ~ bin1853	4.92	3.14	Oleic (2019WF; 0.73)
qA05.12	69.2	bin1860 ~ bin1861	7.24	1.01	Linoleic (2018ZZ; -0.57)

ZZ represents Zhengzhou, *SQ* represents Shangqiu and *WF* represents Weifang.

Table 5
QTLs identified on LG08, LG12 and LG14 for oil, protein and fatty acid content in four environments

QTL	Position	Marker Interval	LOD	PVE(%)	Trait (Environment; Additive Effect)
qA08.1	28.6	bin2648 ~ bin2649	5.76	0.79	Linoleic (2018ZZ; 0.51)
qA08.2	37.7 ~ 37.9	bin2711 ~ bin2718	4.40 ~ 4.83	3.24 ~ 4.08	Arachidic (2019SQ; -0.03); Behenic (2019SQ; -0.03)
qA08.3	59.3	bin2780 ~ bin2781	4.13	3.52	Arachidic (2018ZZ; -0.02)
qA08.4	59.8 ~ 60.5	bin2782 ~ bin2787	6.30 ~ 14.67	3.88 ~ 12.58	Oil (2018ZZ, 2019ZZ, 2019WF; -0.64~-0.42); Protein (2018ZZ, 2019WF; 0.33 ~ 0.37); Behenic (2018ZZ, 2019WF; -0.04~-0.03)
qA08.5	62.0 ~ 62.4	bin2788 ~ bin2789	5.70 ~ 11.27	5.59 ~ 9.59	Oil (2019SQ; -0.52); Protein (2019ZZ, 2019SQ; 0.32 ~ 0.42); Arachidic (2019ZZ; -0.03); Behenic (2019ZZ; -0.05)
qA08.6	76.8	bin2831 ~ bin2832	4.04	3.35	Arachidic (2019WF; -0.03)
qA08.7	97.9 ~ 98.0	bin2884 ~ bin2885	5.57 ~ 6.00	4.22 ~ 4.62	Protein (2018ZZ, 2019WF; 0.26 ~ 0.28)
qA08.8	101.9	bin2896 ~ bin2898	6.10	4.00	Protein (2019ZZ; 0.27)
qA12.1	36.3 ~ 37.0	bin4060 ~ bin4061	4.79 ~ 5.69	3.00 ~ 4.11	Oil (2018ZZ; 0.37); Behenic (2018ZZ; 0.03)
qA12.2	39.7	bin4065 ~ bin4066	7.02	5.45	Protein (2019WF; -0.30)
qA12.3	40.6	bin4067 ~ bin4068	4.60	3.49	Protein (2018ZZ; -0.24)
qA12.4	42.7	bin4071 ~ bin4072	4.09	3.10	Behenic (2019ZZ; 0.03)
qA12.5	43.9	bin4074 ~ bin4075	3.57	3.42	Stearic (2019WF; 0.11)

ZZ represents Zhengzhou, *SQ* represents Shangqiu and *WF* represents Weifang.

QTL	Position	Marker Interval	LOD	PVE(%)	Trait (Environment; Additive Effect)
qA12.6	45.0	bin4076 ~ bin4077	3.78 ~ 5.23	1.34 ~ 4.93	Oleic (2019SQ; -0.70); Linoleic (2019SQ; 0.73)
qA12.7	46.5 ~ 47.2	bin4078 ~ bin4079	3.96 ~ 9.72	2.21 ~ 6.61	Oil (2019ZZ, 2019SQ; 0.38 ~ 0.48); Protein (2019ZZ, 2019SQ; -0.33~-0.34); Stearic (2019SQ; 0.10); Arachidic (2019SQ; 0.02)
qA12.8	51.4	bin4092 ~ bin4096	7.20	5.41	Behenic (2019SQ; 0.04)
qA12.9	54.6	bin4111 ~ bin4112	4.14 ~ 6.80	3.13 ~ 5.13	Oil (2019WF; 0.46); Behenic (2019WF; 0.03)
qA12.10	105.7	bin4491 ~ bin4492	4.55	2.06	Stearic (2019SQ; 0.10)
qA14.1	28.0	bin4971 ~ bin4972	3.53	2.24	Oleic (2019WF; -0.62)
qA14.2	32.3	bin4985 ~ bin4987	4.67	4.50	Stearic (2019WF; 0.12)
qA14.3	38.5	bin5019 ~ bin5020	4.23	4.62	Linoleic (2019WF; 0.68)
qA14.4	39.8	bin5059 ~ bin5069	5.64	3.99	Linoleic (2019ZZ; 0.69)
qA14.5	40.3	bin5139 ~ bin5149	4.26 ~ 5.98	3.76 ~ 3.82	Stearic (2019ZZ; -0.10); Arachidic (2019ZZ; -0.03)
qA14.6	41.2 ~ 41.5	bin5338 ~ bin5404	4.21 ~ 13.74	2.00 ~ 4.91	Oil (2019ZZ; -0.38); Oleic (2018ZZ; -0.65); Linoleic (2018ZZ; 0.82); Behenic (2018ZZ, 2019ZZ, 2019WF; -0.04~-0.03)
qA14.7	42.2	bin5413 ~ bin5416	10.72	7.41	Oil (2019SQ; -0.58)
qA14.8	42.5 ~ 42.6	bin5417 ~ bin5420	4.08 ~ 7.66	1.84 ~ 6.61	Stearic (2019SQ; -0.10); Arachidic (2019SQ; -0.03); Arachidonic (2019SQ; 0.03)
qA14.9	43.1 ~ 43.4	bin5428 ~ bin5438	4.17 ~ 8.30	3.28 ~ 8.21	Oil (2018ZZ, 2019WF; -0.45~-0.39); Stearic (2018ZZ, 2019WF; -0.17~-0.09); Arachidic (2018ZZ, 2019WF; -0.03)

ZZ represents Zhengzhou, *SQ* represents Shangqiu and *WF* represents Weifang.

QTL	Position	Marker Interval	LOD	PVE(%)	Trait (Environment; Additive Effect)
qA14.10	51.2	bin5511 ~ bin5512	7.34	5.52	Behenic (2019SQ; -0.04)
ZZ represents Zhengzhou, *SQ* represents Shangqiu and *WF* represents Weifang.					

Twelve QTLs were mapped on LG05 (Table 4). Among these QTLs, QTL *qA05.1* covered a region of 0.5 cM and was associated with all traits, except for linoleic acid content (Table 4). QTL *qA05.1* showed a negative additive effect on five traits (oil, palmitic acid, stearic acid, arachidic acid, and behenic acid content), which was identified in all four environments. This QTL also exhibited positive additive effects on the protein, oleic acid and arachidonic acid contents, which were detected in two or three environments. The *qA05.1* region flanked by the markers bin1572 and bin1573 exhibited a considerable effect on the traits, being associated with PVE values of approximately 10.44–26.99% and LOD scores of 10.42 to 28.44 for oil content, stearic acid, arachidic acid and behenic acid. Another major QTL on LG05, *qA05.10*, covered a region of 1.5 cM and had a pleiotropic effect on the content of five fatty acids (except for arachidic acids and behenic acids) (Table 4). The LOD score associated with *qA05.10* varied from 5.05 to 39.85, whereas the PVE values were approximately 0.29–15.18%.

Several QTL loci were mapped on LG08, 12 and 14 (Table 5). On LG08, a region of 2.6 cM, covered by QTLs *qA08.4* and *qA08.5*, was associated with oil, protein and behenic acid content, with LOD scores of approximately 5.70-14.67 and PVE values of approximately 3.88–12.58%. Associations with oil and protein content were consistent for all four environments being tested, whereas association with behenic acids was confirmed for three environments (Table 5). A large genomic region containing several QTLs with minor phenotypic effects was identified on LG12 (Table 5). In particular, QTLs for oil, protein, and behenic acid content that were consistent for all four environments were detected in regions spanning 18.30 cM, 7.40 cM, and 17.60 cM, respectively. On LG14, QTLs from *qA14.5* to *qA14.9*, which were included in the interval between 40.3 cM and 43.4 cM, were detected in four environments for oil, stearic acid and arachidic acid content and three environments for behenic acid content (Table 5).

Among the 69 QTLs mapped on LGs different from those mentioned in the previous paragraph, some exhibited pleiotropy on several traits and exhibited consistent effects in more than one environment (Table S3). In particular, a region of 3.4 cM covered by the QTLs *qA06.3* and *qA06.4* was associated with protein content in three environments with LOD ranging from approximately 15.40-18.81 and PVE values of approximately 10.84–15.78%. QTL *qA06.4* was associated with behenic acid in all four environments, with LODs scoring of approximately 13.92 to 24.79 and PVE values of approximately 11.01–17.55%. QTL *qA06.6* showed a major effect on arachidic acid content, exhibiting a PVE value of 10.03% and a LOD score of 11.73.

Annotation of genes and validation of the SNPs in the QTL intervals

The genes in the intervals of *qA05.1*, *qA05.9* and *qA05.10* on LG05, *qA06.3* and *qA06.4* on LG06, *qA08.4* and *qA08.5* on LG08, *qA12.1* to *qA12.7* on LG12, *qA14.5*, *qA14.6*, *qA14.7*, and *qA14.8* on LG14 were extracted and screened for polymorphic SNPs between two parents, and a total of 84 polymorphic SNPs in 71 genes were identified (Table S4). Among these SNPs, 17 resulted in missense mutations (Table 6), whereas the remaining 67 were in introns or resulted in silent mutations. KASP (Kompetitive allele specific PCR) markers were designed

on the 17 SNPs associated with missense mutations (Table S5), and the markers were validated using the two parents and 44 lines of the RIL population displaying contrasting oil content. Two SNPs at sites Arahy05:6599714 and Arahy05:6709559 were closely linked with the oil content (Fig. 4). Specifically, the average oil content was 55.40% in RILs displaying G at Arahy05:6709559, whereas RILs exhibiting nucleotide A at the same loci displayed an oil content of 50.62% (Fig. 5). The two SNPs were included in the genes *Arahy.T0P5W2* and *Arahy.YR3A5K*, encoding a scarecrow-like transcription factor PAT1-like and a galactosyl transferase GMA12/MNN10 family protein, respectively (Table S4).

Table 6
Mutation type of SNPs located in candidate genes

Gene name	Chromosome	Position	P1	P2	Mutation type
Arahy.T0P5W2	Arahy.05	6599714	C	A	P-T
Arahy.YR3A5K	Arahy.05	6709559	A	G	Y-C
Arahy.DH8F8S	Arahy.05	6917629	G	A	V-I
Arahy.USM880	Arahy.05	7033745	C	T	G-R
Arahy.5YK3TE	Arahy.05	7297191	T	A	L-M
Arahy.S4GWG4	Arahy.05	7514948	T	G	N-T
Arahy.925H3N	Arahy.06	6000283	C	T	G-R
Arahy.7E2TSQ	Arahy.06	6317756	A	T	V-D
Arahy.EYYU9K	Arahy.08	37430658	T	A	W-R
Arahy.LF06ED	Arahy.08	37441082	T	C	W-R
Arahy.W9QXGB	Arahy.08	37701095	A	G	H-R
Arahy.W26JNR	Arahy.12	3782327	A	T	F-I
Arahy.V6I7WA	Arahy.12	4236238	G	C	K-N
Arahy.0EHV1A	Arahy.12	4281118	G	A	D-N
Arahy.N0BKZ2	Arahy.12	4908333	T	G	K-Q
Arahy.X5Q10C	Arahy.12	4995776	C	T	S-L
Arahy.C4F96H	Arahy.12	5159247	T	C	H-R

Discussion

NGS technologies have enabled the development of ultrahigh-density linkage maps in various crops; these maps are fundamental for accurate marker-assisted breeding [24]. The genetic map employed in this study consisted of 4561 bin markers and spanned a length of 2032.39 cM with an average genetic distance of 0.45 cM. Both marker number and marker density were larger than those reported for other recent peanut linkage maps [1, 14–15, 19–20]. This finding might be attributable to the large size of the population used in this study and/or the whole genome resequencing strategy adopted, which is more appropriate for tetraploids with respect to other

sequencing technologies. Similar to the linkage map reported by Liu et al. [15], the marker order in our map was consistent with the physical order, except for two translocations between LG3 and LG13, LG6 and LG16.

For oil content, a total of 27 QTLs were mapped on 12 LGs (Tables 4 and 5 and Table S3). Among these QTLs, those mapped on LG05, 08, 12 and 14 were detected in at least two environments. QTL *qA05.1*, detected in all four environments, displayed LOD scores of approximately 13.62–26.94, PVE values of approximately 9.62–22.74% and additive effects of approximately − 0.88 to -0.65 (Table 4). The interval of 0.5 cM spanned by *qA05.1*, corresponding to an approximately 6.3–7.8 Mb physical region of chromosome A05, might be the same as that reported by Pandey et al. [2], flanked by the markers GM1878 and GM1890, which were mapped to the approximately 6.4-10.9-Mb region of A05 [1]. QTL *qA08.4* was detected in three environments, whereas the neighboring QTL *qA08.5* was detected in the remaining environment (Table 5). Taken together, the two QTLs spanned an interval of 2.6 cM, corresponding to a physical region of approximately 37.0-38.2 Mb, close to the hotspot region (approximately 39.9–43.8 Mb) on chromosome A08 for genes controlling oil content, as reported by Liu et al. [1].

qA05.1 was repeatedly detected for eight of nine peanut quality-related traits investigated in this study, thereby indicating that it has pleiotropic effects. In the physical regions of 1.5 Mb spanned by *qA05.1*, six polymorphic SNPs between two parents could cause missense mutations of six genes (Table 6). Additionally, 11 other polymorphic SNPs in *qA06.4*, *qA08.4*, *qA12.1*, *qA12.2*, *qA12.4* and *qA12.5* could affect the genes encoding different proteins (Table 6). The 17 SNPs were designed into KASP markers and validated using the two parents and 44 RIL lines. Of these SNPs, two sites on Arahy 05 (Arahy05:6599714 and Arahy05:6709559) were linked with the oil content (Fig. 4 and Fig. 5). Compared with the reference genome, the bases of the high oil content parent Yuhua15 at the sites Arahy.05:6599714 and Arahy.05:6709559 changed from C to A and A to G, respectively, and the encoded proteins changed from proline (P) to threonine (T) and from tyrosine (Y) to cysteine (C) (Table 6).

A total of 2559 genes were involved in the metabolism of fatty acids and lipid storage, which were unevenly distributed on the 20 peanut chromosomes [12]. The SNP Arahy.05:6709559 was located in the exon of the gene *Arahy. YR3A5K* encodes a galactosyl transferase GMA12/MNN10 family protein. This gene is involved in transferring glycosyl groups and xyloglucan metabolic processes in *Arabidopsis thaliana* [25]. In the high oil-content peanut variety, the average expression level of this gene (1.29) was lower than that (4.74) in the low oil-content peanut variety (from unpublished transcriptome data). The SNP Arahy.05:6599714, located in the gene *Arahy.T0P5W* and encoding the scarecrow-like transcription factor PAT1-like, is involved in phytochrome A (phyA) signal transduction in *Arabidopsis thaliana* [26]. This gene may not be involved in the fatty acid biosynthetic pathway, as its expression level did not differ between the high and low oil-content peanut varieties.

Conclusion

A high-density genetic map with 4561 bin markers was constructed, and a major QTL for the content of oil, protein and several fatty acids, which was located on LG05, was consistently detected across four environments. The SNPs Arahy05:6599714 and Arahy05:6709559 were employed to design KASP markers, which were validated on lines displaying contrasting oil contents. These two markers may facilitate marker-assisted breeding to develop high oil-content peanut cultivars.

Methods

Plant materials

A population of 329 RILs was derived from a cross between Yuhua15 (female parent) and W1202 (male parent), which was undertaken by the corresponding author's lab in 2012. Yuhua15 is an irregular peanut cultivar with high oil content (54.00%) released by the Institute of Industrial Crops, Henan Academy of Agricultural Science, in 2001. W1202 is a breeding line with relatively low oil content (52.60%) developed and preserved in the author's lab. F_{10} RIL lines were obtained through the SSD (single seed descent) method, which was performed in the Chinese provinces of Hainan and Henan to reduce generation time. Single plants from parental lines and RILs were used for DNA isolation and sequencing and propagated by self-pollination for phenotyping.

Field Trails And Phenotyping

The RIL population and the two parental lines were grown in Zhengzhou (Henan Province) in 2018 and three locations (Zhengzhou and Shangqiu, Henan Province; Weifang, Shandong Province) in 2019. Twenty seeds for each line were sown in a 3 m × 0.5 m plot according to a complete randomized block design with two replicates. After harvesting, approximately 10 dried and plump seeds for each genotype were used to estimate the content of oil, protein and fatty acids by near infrared reflectance spectroscopy (NIRS) (Perten 7200). In particular, fatty acids assessed were palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), arachidic acid (C20:0), behenic acid (C22:0), and arachidonic acid (C20:1).

Statistical Analysis Of Phenotypic Data

Statistics from phenotypic data, i.e., the mean, standard deviation (SD), coefficient of variation (CV), skewness, and kurtosis, were obtained using SAS software. Combined analysis of variance (ANOVA) for each trait and a correlation study among traits were performed using the AOV module implemented in QTL IciMapping software [27]. Broad-sense heritability on the basis of the mean across replications and environments (or heritability per mean) was estimated by $H^2 = \frac{V_G}{V_G + \frac{1}{e}V_{GE} + \frac{1}{re}V_\varepsilon}$, where *e* represents the number of environments and *r* represents the number of replicates.

Sequencing And Snp Calling

Genomic DNA was extracted from fresh leaves using the Plant Genomic DNA Kit (TIANGEN). DNA quality, concentration, and integrity were assessed using a NanoDrop-2000 spectrophotometer (Thermo), a Qubit Fluorometer (Thermo), and agarose gel electrophoresis. DNA that passed the quality control step was further randomly sheared by sonication, and fragments of approximately 300 bp DNA were recovered by electrophoresis. DNA fragments with adapters were used to prepare DNA clusters, which were sequenced on the Illumina HiSeq Xten platform with PE151.

Clean data were obtained after filtering for adapters and low-quality reads (i.e., > 50% of bases with a quality score ≤ 20 and ≥ 1% of missing bases) by the SOAPnuke tool [28]. Trimmed reads were aligned to the peanut reference genome [20] using the aln command implemented in bwa-0.7.10 software [29]. Unique mapped reads were used to identify SNPs by the GATK3.3.0 pipeline [30]. After obtaining the raw SNP dataset, low-quality SNPs between the two parents were excluded based on missing values, heterozygosity, sequencing depth < 10 and GQ < 20. Homozygous and polymorphic loci between the two parents were used to genotype the RIL population.

Genetic Map Construction And Qtl Mapping

A sliding window approach for genotype calling and recombination breakpoint determination [23] was applied to convert SNPs into bin markers. These were used to construct a linkage map and QTL mapping using QTL IciMapping software [27]. First, the BIN module was used to filter the redundant markers by missing rate. Second, filtered markers were used to construct the linkage map by the MAP module. For grouping, the group number was set as 20 according to the chromosome number of peanut. For ordering, REC and 2-OptMAP of the k-optimality method were chosen, as they gave a higher proportion of correct order in less time [31]. For QTL mapping, the mapping step was set to 0.10 cM, and the LOD threshold was set as 3.30, as calculated using the equation given by Sun et al. [32].

Abbreviations

QTLs; quantitative trait loci; RIL:recombinant inbred line; LG:linkage group; SNP:single nucleotide polymorphisms; KASP:kompetitive allele-specific PCR; SSR:simple sequence repeat; NGS:next-generation sequencing; RAD-seq:restriction-site-associated DNA sequencing; SLAF-seq:specific-locus amplified fragment sequencing; DArT:diversity array technology; WGRS:whole genome resequencing; PVE:phenotypic variance explained; CV:coefficient of variation; BAM:binary alignment mapping; SSD:single seed descent; NIRS:near infrared reflectance spectroscopy; SD:standard deviation; ANOVA:combined analysis of variance;

Declarations

Acknowledgments

The authors would like to thank Prof. Stefano Pavan (University of Bari Aldo Moro, Italy) for fruitful discussions on QTL analysis and for editing the English text of the manuscript. They also would like to thank the AJE Company for editing the English text of the manuscript.

Authors' contributions

ZS performed field experiments and phenotypic analysis and wrote the manuscript. FQ performed laboratory experiments and genotype analysis. HL and LQ developed the RIL population. JX and ZZ1 provided help in field experiments. LS, LM, WX, MT, JF and RZ provided help in laboratory and field experiments. BH and WD provided help to design the experiments. XZ and ZZ2 conceived and designed the experiments, facilitated the project, and assisted in manuscript preparation. All authors read and approved the final manuscript.

Funding

This work was supported by the Science-Technology Foundation for China Agriculture Research System (CARS-13), Henan Provincial Agriculture Research System, China (S2012-5), Outstanding Young Scientists of Henan Academy of Agricultural Sciences (Grant no. 2020YQ08), Fund for Distinguished Young Scholars from Henan Academy of Agricultural Sciences (2019JQ02), and Henan Provincial Young Talents Supporting Project (2020HYTP044). The funding agencies played no role in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in the manuscript and its Additional file 1, Additional file 2, Additional file 3, Additional file 4, Additional file 5. The clean data relative to resequencing data obtained in this study are available at the BioProject database at NCBI under the Sequence Read Archive (SRA) submission: SUB8691701. Materials used in this study are available from the corresponding authors.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflict of interest statement

The authors declare that they have no competing interests.

Author details

Industrial Crops Research Institute, Henan Academy of Agricultural Sciences/Key Laboratory of Oil Crops in Huang-Huai-Hai Plains, Ministry of Agriculture and Rural Affairs/Henan Provincial Key Laboratory for Genetic Improvement of Oil Crops, Zhengzhou 450002, P.R. China. 3Institute of Plant Protection, Henan Academy of Agricultural Sciences, Zhengzhou 450002, P.R. China.

References

1. Liu N, Guo JB, Zhou XJ, Wu B, Huang L, Luo HY, Chen YN, Chen WG, Lei Y, Huang Y, et al. 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;133:37–49.
2. Pandey MK, Wang ML, Qiao LX, Feng SP, Khera P, Wang H, Tonnis B, Barkley NA, Wang JP, Holbrook CC, et al. 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;15:133.
3. Huang L, He HY, Chen WG, Ren XP, Chen YN, Zhou XJ, Xia YL, Wang XL, Jiang XG, Liao BS, et al. Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea* L.). Theor Appl Genet. 2015;128:1103–15.

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. Kavera B. Oil quality improvement in groundnut (*Arachis hypogaea* L.) through induced mutagenesis. Ph.D. Thesis, University of Agricultural Sciences. 2008.
6. Kholief TS. Chemical composition and protein properties of peanuts. Z Ernährungswiss. 1987;26:56–61.
7. O Keefe SF, Wiley VA, Knauft DA. Comparison of oxidative stability of high- and normal-oleic peanut oils. J Am Oil Chem Soc. 1993;70:489–92.
8. Wang ML, Khera P, Pandey MK, Wang H, Qiao L, Feng S, Tonnis B, Barkley NA, Pinnow D, Holbrook CC, et al. 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:e0119454.
9. Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EKS, Liu X, Gao D, Clevenger J, Dash S, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. Nat Genet. 2016;48:438–46.
10. Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, Leal-Bertioli SCM, Ren L, Farmer AD, Pandey MK, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet. 2019;51:877–84.
11. Zhuang WJ, Chen H, Yang M, Wang JP, Pandey MK, Zhang C, Chang WC, Zhang LS, Zhang XT, Tang RH, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet. 2019;51:865–76.
12. Chen XP, Lu Q, Liu H, Zhang JN, Hong YB, Lan HF, Li HF, Wang JP, Liu HY, Li SX, et al. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement. Mol Plant. 2019;12:920–34.
13. Wang ZH, Huai DX, Zhang ZH, Cheng K, Kang YP, Wan LY, Yan LY, Jiang HF, Lei Y, Liao BS. 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.
14. Gangurde SS, Wang H, Yaduru S, Pandey MK, Fountain JC, Chu Y, Isleib T, Holbrook CC, Xavier A, Culbreath AK, et al. Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). Plant Biotechnol. 2020;18:1457–71.
15. Liu H, Sun ZQ, Zhang XY, Qin L, Qi FY, Wang ZY, Du P, Xu J, Zhang ZX, Han SY, et al. QTL mapping of web blotch resistance in peanut by high-throughput genome-wide sequencing. BMC Plant Biol. 2020;20:249.
16. Agarwal G, Clevenger J, Pandey MK, Wang H, Shasidhar Y, Chu Y, Fountain JC, Choudhary D, Culbreath AK, Liu X, et al. High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. Plant Biotechnol J. 2018;16:1954–67.
17. Wang LF, Zhou XJ, Ren XP, Huang L, Luo HY, Chen YN, Chen WG, Liu N, Liao BS, Lei Y, et al. A major and stable QTL for bacterial wilt resistance on chromosome B02 identified using a high-density SNP-based genetic linkage map in cultivated peanut Yuanza 9102 derived population. Front Genet. 2018;9:652.
18. Han SY, Yuan M, Clevenger JP, Li C, Hagan A, Zhang XY, Chen C, He GH. A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (A*rachis hypogaea* L.). Front Plant Sci. 2018;9:1012.

19. Chavarro C, Chu Y, Holbrook C, Isleib T, Bertioli D, Hovav R, Butts C, Lamb M, Sorensen R, Jackson SA, et al. Pod and seed trait QTL identification to assist breeding for peanut market preferences. G3. 2020;10:2297–2315.
20. Luo ZL, Cui RJ, Chavarro C, Tseng YC, Zhou H, Peng Z, Chu Y, Yang XP, Lopez Y, Tillman B, et al. Mapping quantitative trait loci (QTLs) and estimating the epistasis controlling stem rot resistance in cultivated peanut (*Arachis hypogaea*). Theor Appl Genet. 2020;133:1201–12.
21. Hu XH, Zhang SZ, Miao HR, Cui FG, Shen Y, Yang WQ, Xu TT, Chen N, Chi XY, Zhang ZM, et al. High-density genetic map construction and identification of QTLs controlling oleic and linoleic acid in peanut using SLAF-seq and SSRs. Sci Rep. 2018;8:5479.
22. Vishwakarma MK, Pandey MK, Shasidhar Y, Manohar SS, Nagesh P, Janila P, Varshney RK. Identification of two major quantitative trait locus for fresh seed dormancy using the diversity arrays technology and diversity arrays technology-seq based genetic map in Spanish-type peanuts. Plant Breeding. 2016;135:367–75.
23. Huang XH, Feng Q, Qian Q, Zhao Q, Wang L, Wang AH, Guan JP, Fan DL, Weng QJ, Huang T, et al. High-throughput genotyping by whole-genome resequencing. Genomes Res. 2009;19:1068–76.
24. Jaganathan D, Bohra A, Thudi M, Varshney R. Fine mapping and gene cloning in the post–NGS era: advances and prospects. Theor Appl Genet. 2020;133:1791–810.
25. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, Haberer G, Hollister JD, Ossowski S, Ottilar RP, Salamov AA, Schneeberger K, Spannagl M, Wang X, Guo YL. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet. 2011;43:476–81.
26. Bolle C, Koncz C, Chua NH. PAT1, a new member of the GRAS family, is involved in phytochrome A signal transduction. Genes Dev. 2000;14:1269–78.
27. Meng L, Li LL, Zhang LY, Wang JK. QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. Crop Journal. 2015;3:269–83.
28. Chen YX, Chen YS, Shi CM, Huang ZB, Zhang Y, Li SK, Li Y, Ye J, Yu C, Li Z, et al. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience. 2018;7:1–6.
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 2009;25:1754–60.
30. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky AM, Garimella K, Altshuler D, Gabriel S, Daly MJ, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.
31. Zhang LY, Li HH, Meng L, Wang JK. Ordering of high-density markers by the k-Optimal algorithm for the traveling-salesman problem. Crop journal. 2020;8:701–12.
32. Sun ZQ, Li HH, Zhang LY, Wang JK. Properties of the test statistic under null hypothesis and the calculation of LOD threshold in quantitative trait loci (QTL) mapping. Acta Agronomica sinica. 2013;39:1–11. (in Chinese with English abstract).

Figures

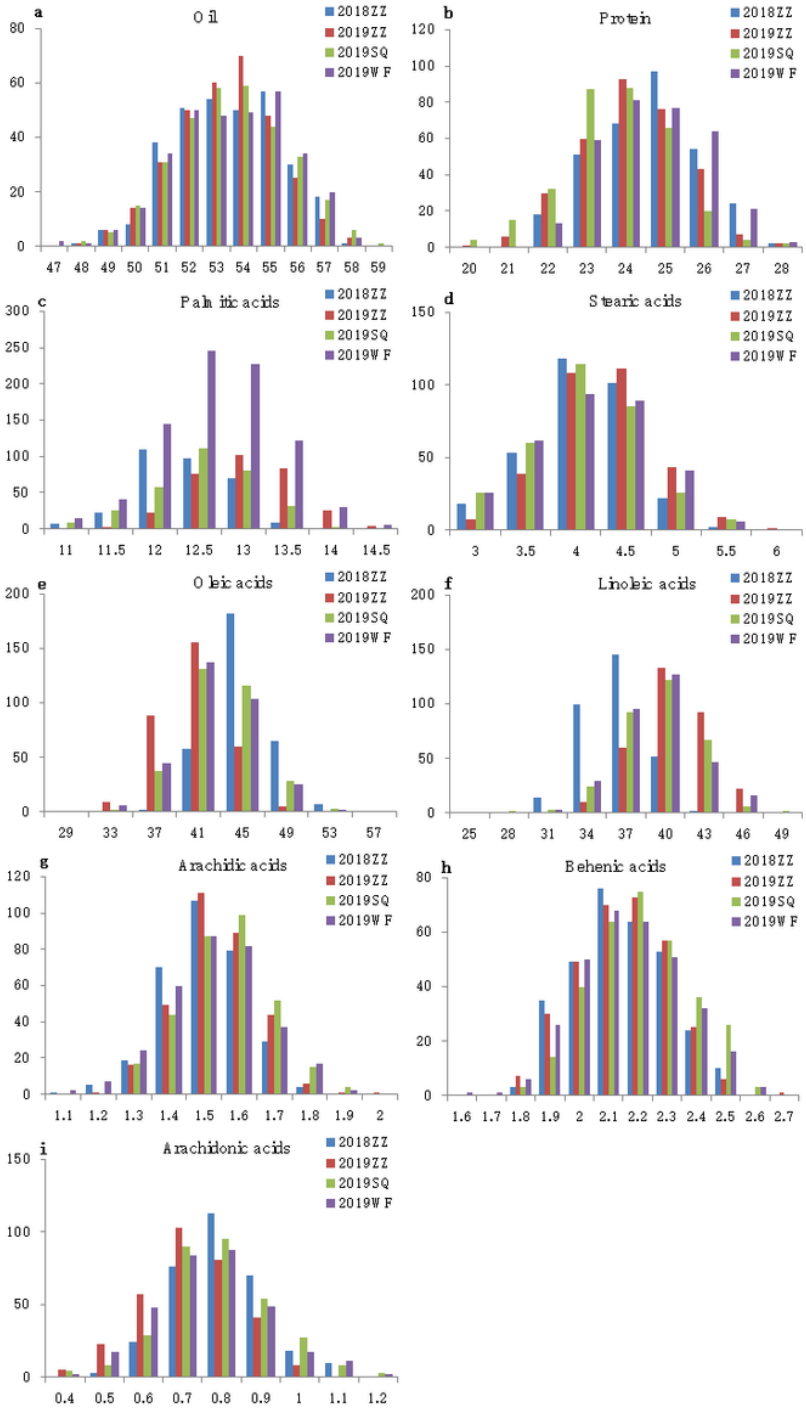

Figure 1

Phenotypic distribution of nine quality-related traits for the RILs in four environments. Figure 1a-1i shows the phenotypic distribution of the contents of oil, protein, palmitic acid, stearic acid, oleic acid, linoleic acid, arachidic acid, behenic acid, and arachidonic acid. ZZ represents Zhengzhou, SQ represents Shangqiu and WF represents Weifang.

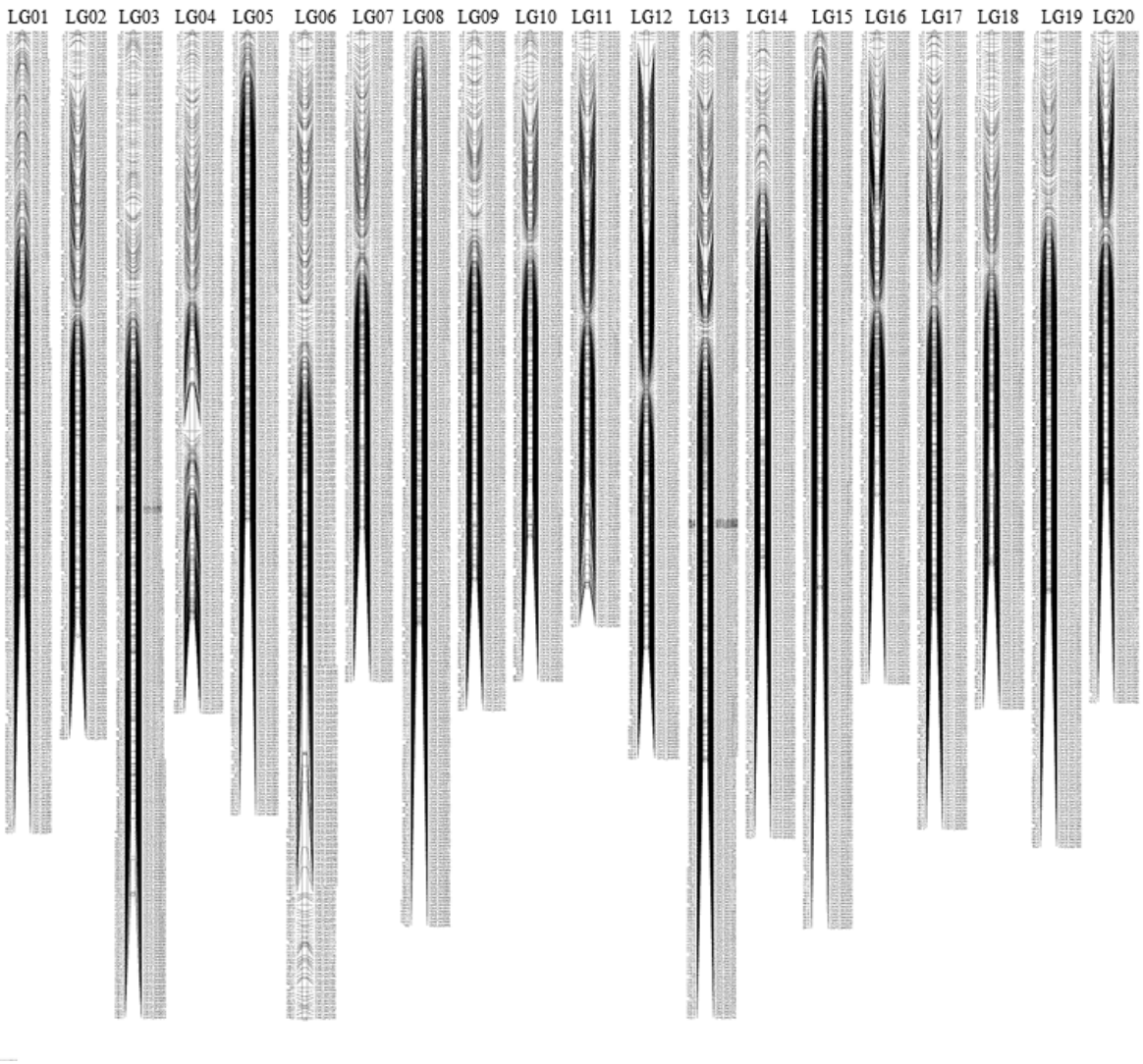

Figure 2

Genetic map obtained from the Yuhua15 × W1202 RIL population.

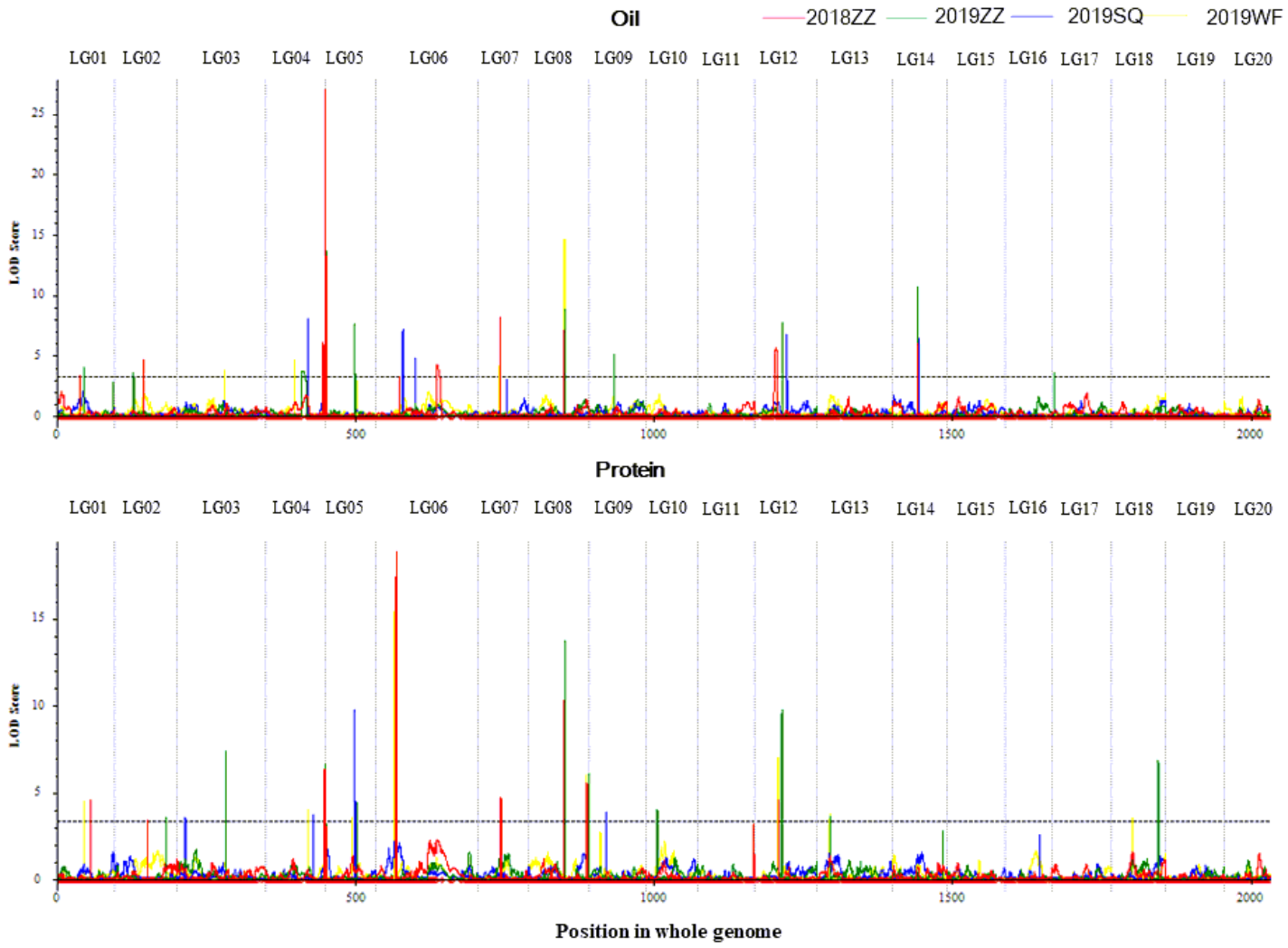

Figure 3

LOD curves of oil and protein content across the whole genome under four environments

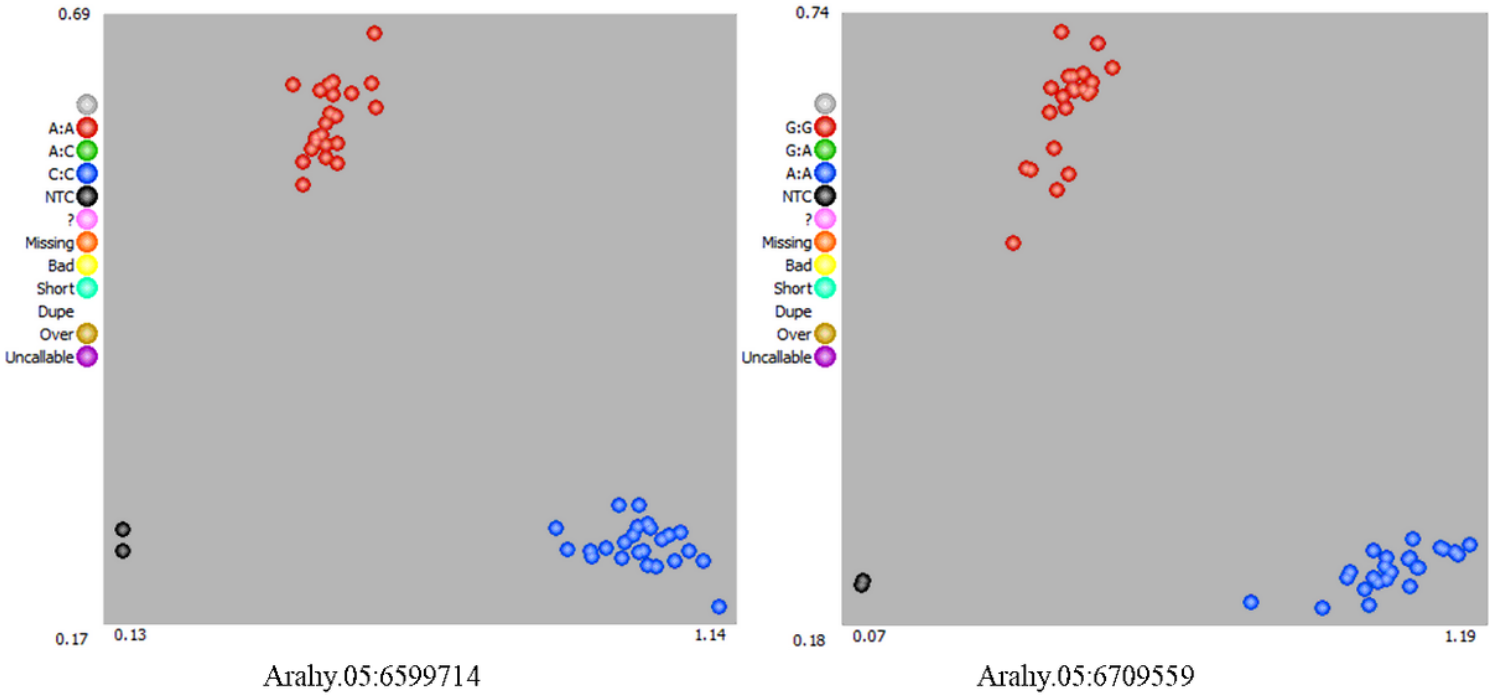

Figure 4

KASP validations for the two SNPs on LG05

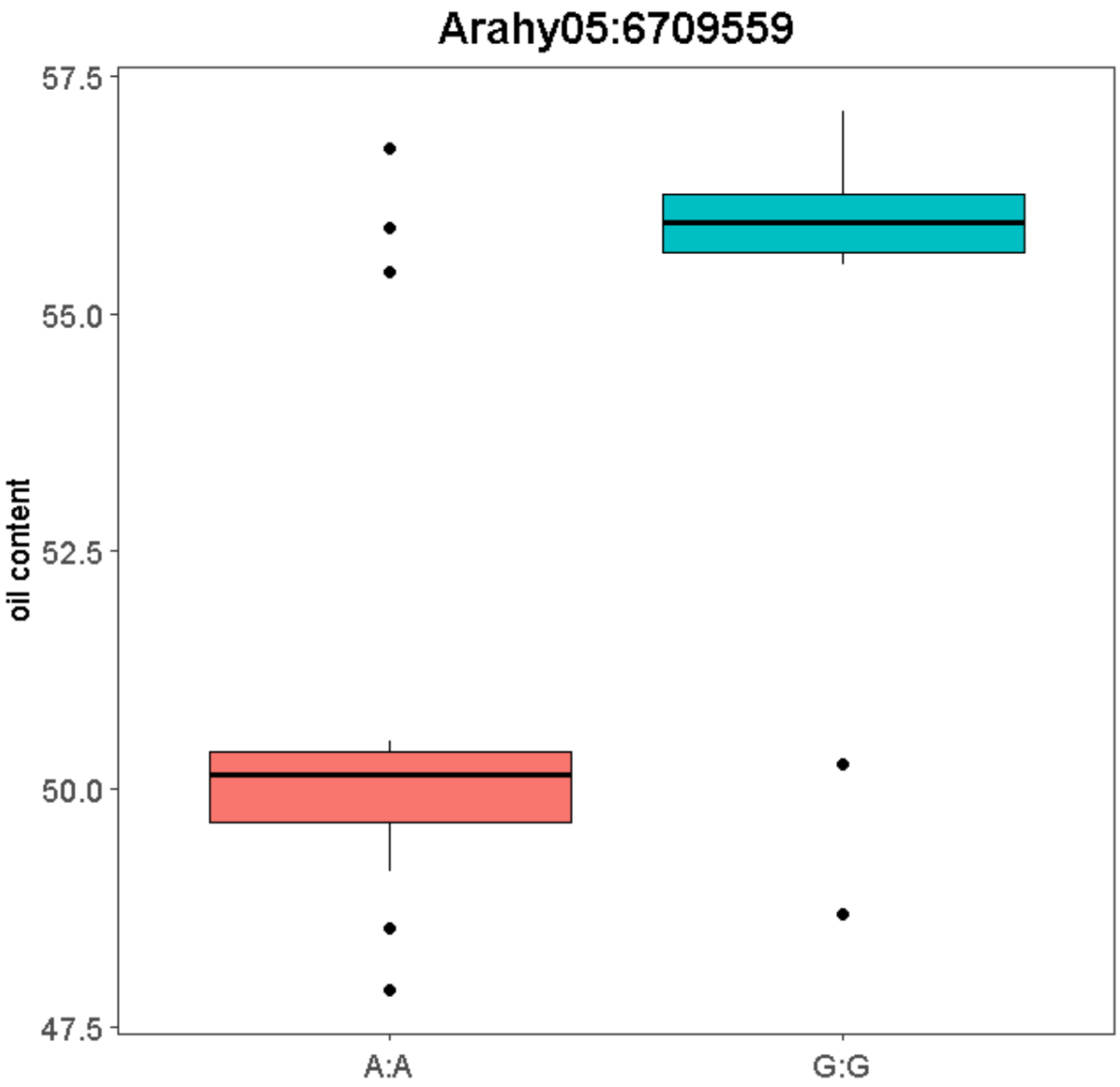

Figure 5

Phenotypic difference between two genotypes at the SNP site Arahy05:6709559 for 44 lines of the RIL population displaying contrasting oil content

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1.xlsx
- SupplementaryTableS2.xlsx
- SupplementaryTableS3.xlsx
- SupplementaryTableS4.xlsx
- SupplementaryTableS5.xlsx