

Dissection of the genetic basis of oil content in Chinese peanut cultivars by association mapping

CURRENT STATUS: UNDER REVIEW

BMC Genetics  BMC Series

Nian Liu

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Li Huang

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Weigang Chen

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Bei Wu

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Manish K. Pandey

International Crops Research Institute for the Semi-Arid Tropics

Huaiyong Luo

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Xiaojing Zhou

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Jianbin Guo

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Haiwen Chen

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Dongxin Huai

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Yuning Chen

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Yong Lei

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Boshou Liao

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Xiaoping Ren

Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Rajeev K. Varshney

International Crops Research Institute for the Semi-Arid Tropics

Huifang Jiang

✉ peanutlab@oilcrops.cn *Corresponding Author*
ORCID: <https://orcid.org/0000-0003-4087-205X>

DOI:

10.21203/rs.3.rs-17113/v1

SUBJECT AREAS

Population Genetics

KEYWORDS

Peanut, Oil content, Association analysis, Geographic region

Abstract

Background Peanut is one of the primary sources for vegetable oil worldwide, and enhancing oil content is the main objective in these peanut breeding programs. Linked markers for oil content is required for use in genomics-assisted breeding (GAB), and association mapping is one of the promising approaches for discovery of associated markers.

Results An association mapping panel consisting of 292 peanut varieties extensively distributed in China were phenotyped for oil content and genotyped with 583 polymorphic SSR markers. These markers amplified 3663 alleles with an average of 6.28 alleles per locus. The results of structure, phylogenetic relationship, and PCA analyses indicated two subgroups majorly differentiating based on geographic regions. Genome-wide association analysis using genetic and phenotypic data identified 12 associated markers including one (AGGS1014_2) highly stable association controlling up to 9.94% phenotypic variance explained (PVE) across multiple environments. Interestingly, the frequency of the favorable alleles for 12 associated markers showed a geographic difference. Two associated markers AGGS1014_2 and AHGS0798 with 6.90-9.94% PVE were verified to enhance oil content in an independent RIL population. The combined genotypes of AGGS1014_2 and AHGS0798 appeared to experience selection during the breeding program.

Conclusion This study provided insights into the genetic basis of oil content in peanut and verified that two SSR markers were highly associated with oil content. Our results could facilitate marker-assisted selection for high-oil content breeding.

Background

Cultivated peanut (*Arachis hypogaea* L.) is one of the most important oilseed crops worldwide which sooner may gain the status of food crop in the near future because of diverse consumption modes.

The annual planting area is 27.94 Mha, with the annual production of 47.10 Mt (FAOSTAT 2017). China is the largest peanut producer in the world, with over 36% of global peanut production. However, the production in China hardly meets the increasing domestic demands for peanut oil. Oil content of seed is one of the most significant traits determining oil yield, and an increase of 1% in the oil content could result in a 7% increase in economic benefit for oil processing. Thus, enhancing oil content is an

important long-term objective in all the peanut breeding programs in China and also India.

Oil content is a polygenetic trait and is significantly influenced by the environment [1, 2]. In the current breeding programs, high oil content lines are selected based on the phenotyping data generated from multiple environments, which is low efficient and time-consuming. Deployment of well-validated markers could accelerate the process and precision in genetic improvement [3–7]. Dissection of the genetic architecture underlying oil content is a prerequisite to deploying markers in high-oil breeding. Triacylglycerol is the major form of storage oil in most plant seeds. Its biosynthetic pathway and the relevant genes have been extensively understood in the model plant *Arabidopsis* [8, 9]. However, the genetic mechanism of natural variation in oil biosynthesis in peanut remains poorly understood.

The available significant variation of oil content among germplasm in peanut provides an opportunity to identify genomic regions controlling oil accumulation [10]. Using bi-parent populations with various phenotypes, linkage analyses were performed to identify quantitative loci and linked markers in peanut. For instance, six and nine QTLs for oil content have been respectively identified in two different recombinant inbred line (RIL) populations [11]. Subsequently, three and eight QTLs were detected in an advanced backcross population and RIL population, respectively [12, 13]. Most recently, seven QTLs for oil content were detected in a RIL population, including one major and stable QTL with 10.14–27.19% phenotypic variance explained (PVE) [14]. Despite discovery of several QTLs for oil content, linkage analysis of oil content hardly fully reveals genetic variations for oil content in peanut due to the limited number of parental lines used in the studies. Compared with linkage analysis, association analysis, utilizing historical recombination in natural populations, could offer the promise of high-resolution mapping and multiple-alleles evaluation, in far less time [15, 16]. This genetic method has been successfully used to reveal the genetic basis of complex agronomic traits in multiple crops [17–20]. In peanut, the limited works on association analysis were conducted to evaluate genetic variation in diverse germplasm [21, 22].

Cultivars are elite germplasm with desirable traits which gets preference in use as trait source in almost all the genetic improvement programs. Discovery of associated genomic regions in such panel

have potential of faster application in ongoing breeding programs. Therefore, the present study used the Chinese peanut panel which consists of 222 cultivars, 55 breeding lines, and 15 landraces to study (1) genetic architecture of Chinese peanut cultivars, (2) elucidation of genetic basis of natural variation in peanut cultivars for oil content, and (3) development and validation of associated markers which could be used to enhance oil content through genomics-assisted breeding (GAB).

Results

Genetic diversity and population structure analysis

A total of 583 polymorphic markers which were randomly distributed on the genome, were used to genotype the 292 peanut accessions. The polymorphic markers produced 3,663 alleles with an average of 6.28 alleles per locus ranging from 2 to 20 (Table 1 and Additional file 1: Table S2). The major allele frequency ranged from 0.15 to 0.98, with a mean value of 0.60. The average genetic diversity was 0.51 and ranged from 0.03 to 0.90. The polymorphic information content (PIC) ranged from 0.03 to 0.90, with an average of 0.45 (Additional file 1: Table S2). Of the 3,663 alleles, 629 were unique alleles (allele frequency < 0.05%), 1,471 allele were rare alleles ($0.05\% \leq$ allele frequency < 5%), 1,547 allele were polymorphic alleles ($5\% \leq$ allele frequency < 95%), and 15 were fixed alleles (allele frequency \geq 95%), with corresponding proportions of 17.17%, 40.16%, 42.23% and 0.41%, respectively (Additional file 1: Table S3).

The population structure analysis was performed using multi-allelic SSR data. The most significant change of the LnP(D) value was observed when parameter K increased from 1 to 2, and the highest ΔK value was obtained when K=2 (Fig. 1a and 1b). The previously available information suggested two subgroups in the peanut panel and the results of this study on phylogenetic relationship and PCA analysis further proved that the 292 peanut accessions could be clearly divided into two subgroups (G1 and G2), which were consistent with the population structure results (Fig. 1c and 1d). All the landraces in G1 subgroup were subsp. *hypogaea*, while the landraces in G2 subgroup belonged to subsp. *fastigiata* (Fig. 1c). The pairwise F_{ST} value between the two subgroups was 0.16, and Nei's genetic distance was 0.27. Compared with G1 subgroup, G2 had a relatively higher genetic diversity (0.47) and PIC value (0.36). However, the allele number per locus was higher in G1 than G2 (Table 1).

The peanut panel was consisted of cultivars from 17 provinces of China. Most accessions (93.2%) were distributed in nine provinces (HEB, SD, HN, SC, HUB, JS, Fujian, Guangdong, and Guangxi). The proportion of two subgroups in these provinces exhibited obvious differences (Fig. 2a). In Northern China (HEB, SD, and HN provinces), the proportion of G1 subgroup ranged from 77.42% to 85.71%. Similarly, the proportion of G1 ranged from 66.10% to 91.67% in peanut accessions distributed in the Yangtze River region (SC, HUB, and JS provinces). Whereas, the proportion of G1 subgroup were below 11.11% in Southern China (Fujian, Guangdong, and Guangxi provinces). It is suggested that genetic diversity was highly linked to geographic distribution. The phylogenetic tree showed that the peanut-distributed provinces could be clearly clustered into two clades (Fig. 2b). The provinces in Southern China (Fujian, Guangdong, and Guangxi) were clustered together, and the provinces from Northern China and the Yangtze River region were grouped into another clade.

Phenotypic variation for oil content among peanut accessions

The oil content for 292 Chinese peanut accessions was analyzed from seeds harvested from four environments. The phenotypic values ranged from 45.85 to 59.72% in 2015WH, 43.82 to 55.88% in 2016WH, 44.22 to 54.97% in 2017NC, and 45.11 to 56.69% in 2017WH, respectively (Table 2). The median values of oil content in four environments varied from 48.47% to 51.89%, and the standard deviation of phenotypic data ranged from 1.78 to 2.39 across four environments. The continuous distributions of phenotypic values for peanut accessions were shown in Additional file 2: Fig. S1. The phenotypic data of the peanut panel in 2015WH, 2016WH, and 2017WH followed a normal distribution based on the Shapiro-Wilk normality test (Table 2). Based on phenotypic data of 292 peanut accessions across four environments, the broad sense heritability for oil content was evaluated to be 0.76.

Since genetic diversity was present among peanuts from different Chinese regions, we further evaluated whether phenotypic differences existed in the different geographic distributions. The oil content of accessions from Northern China was statistically higher than that from Southern China in all the field trials. Similarly, the accessions from the Yangtze River region have higher oil content than

the accessions from Southern China in 2016WH, 2017NC, and 2017WH. The phenotypic difference between Northern China and the Yangtze River region was not statistically significant in three environments. Meanwhile, we also made a comparison among cultivated peanuts released at different times (Additional file 2: Fig. S2). In general, there was no obvious difference in oil content between cultivars released at different times.

Association analysis for oil content

Based on four environmental phenotypic data of 292 peanut accessions, the MLM model with K+Q matrix was used to execute association mapping of SSR-markers with oil content. The marker-trait analysis for four environmental trials identified two associated loci for 2015 WH environment, eight associated loci for 2016WH environment, three associated loci for 2017NC environment, and five associated loci for 2017WH environment. Twelve significantly associated loci at $P < 0.00186$ explained 4.54-9.94% phenotypic variance across four environments (Table 3 and Additional file 2: Fig. S3). Among them, AGGS1014_2 with up to 9.94% PVE had been repeatedly detected in multiple environments (2016WH, 2017NC, and 2017WH). The allelic number of these associated loci ranged from two (pPGPseq8D9 and AGGS1014_2) to six (TC11B4_2). The accessions with different alleles for these loci had significantly different phenotypes in a four-environment average of oil content (Fig. 4a). The most favorable alleles associated with oil content were pPGPseq8D9-131bp, TC9F10_2-256bp, TC11B4_2-298bp, AHGS1679-293bp, AGGS1149-192bp, AGGS1081-201bp, AGGS1014_2-215bp, AHGS2053-256bp, AHS0127-188bp, AHGS1431-260bp, AHGS0798-174bp, and AHGS1388-304bp (Additional file 1: Table S4). In general, the allelic effects of associated loci were higher in accessions from the Northern China and Yangtze River region than accessions from Southern China (Fig. 4b). Similarly, the frequencies of the most favorable alleles also showed geographic differences. For ten associated loci (pPGPseq8D9, TC11B4_2, AHGS1679, AGGS1149, AGGS1014_2, AHGS2053, AHGS0127, AHGS1431, AGHS0798, and AHGS1388), the favorable allele frequency was highest in Northern China, the second-highest in the Yangtze River region, and lowest in Southern China (Fig. 4c). However, the favorable allele frequencies were highest in Southern China for another two

associated loci (TC9F10_2 and AHGS1431).

Evaluation of RIL population and confirmation of associated markers

To estimate potential values of associated loci in peanut breeding, a RIL population derived from two additional accessions (Zhonghua 10 and ICG12625) was employed as a test population. Oil content of the RIL population across four environments ranged from 47.45% to 60.88% in Env1, 45.30% to 58.96% in Env2, 42.89% to 55.07% in Env3, and 45.98% to 58.37% in Env4, respectively. The oil content of the female parent was 51.88 ± 1.41 %, whereas that of the male parent was 53.32 ± 1.47 %. Three associated makers (AGGS1014_2, AHGS0798, and AHGS1431) showed polymorphism in the RIL population. A significant difference in oil content between homozygous alleles from P1 and P2 at AHGS1431 locus was observed in Env1 (Additional file 1: Table S5). Compared with the homozygous allele from P1 at AGGS1014_2 locus, the homozygous allele from P2 had significantly higher oil content in two environments i.e., Env2 and Env4 (Fig. 5a and Additional file 1: Table S5). For marker AHGS0798, oil content of the homozygous allele from P2 was significantly higher than that of the homozygous allele from P1 in two environments (Fig. 5a and Additional file 1: Table S5).

Combined AGGS1014_2 and AHGS0798, it is obvious to find that oil content of homozygous alleles from P2 was significantly higher than that of the homozygous allele from P1 across environments (Fig. 5c).

In the 292 peanut accessions, the alleles at AGGS1014_2 (X) locus and AHGS798 (Y) locus formed six combined genotypes, namely X-205bp/Y-170bp, X-205bp/Y-172bp, X-205bp/Y-174bp, X-215bp/Y-170bp, X-215bp/Y-172bp, X-215bp/Y-174bp (Additional file 2: Fig. S4). The oil content is highest in X-215bp/Y-174bp (51.49 ± 1.30 %), median in X-215bp/Y-172bp (51.04 ± 1.11 %), X-215bp/Y-170bp (50.88 ± 1.17 %), X-205bp/Y-174bp (50.66 ± 1.38 %), and X-205bp/Y-170bp (49.72 ± 2.10 %), and lowest in X-205bp/Y-172bp (49.62 ± 1.19 %). The genotypic frequency of X-215bp/Y-174bp was 4.00% in peanut varieties released before 1980, and it increased to 22.13% in peanut varieties released after 2000. Similarly, the frequency of X-205bp/Y-172bp has an increase from 12.00% in peanut varieties released before 1980 to 32.79% in peanut varieties after 2000. The frequency of X-205bp/Y-

172bp and X-215bp/Y-170bp were lower in peanut varieties released after 2000 than these in peanut varieties released before 1980.

Discussion

For achieving reliable results from GWAS studies, multilocation phenotyping data on diverse panel together with genotyping data is necessary for discovery of significantly associated markers for traits of interest. Keeping this mind, this research effort used the peanut panel which consist of 222 cultivars, 55 breeding lines, and 15 landraces, which were widely distributed in 17 provinces of China. The genome-wide SSR marker in the peanut panel screening produced high number (3,663) of alleles including 629 unique alleles showing high molecular diversity. For example, this panel of 292 Chinese cultivated accessions showed on par average allele number (2.99 to 8.10 per locus), gene diversity (0.51) and PIC (0.45) as compared to average allele number (6.28 per locus), gene diversity (0.11 to 0.59) and PIC (0.21 to 0.53) in other Chinese germplasm collections or US peanut mini-core collection [23–25]. On the other hand, the higher values for average allele number (22.21), gene diversity (0.74) and PIC (0.72) was observed in the peanut ‘reference set’ of ICRISAT [21] which may due to the diverse genotypes included from 48 countries representing global diversity including wild accessions. From all these comparisons, the Chinese cultivated accessions in the present study represented high molecular diversity comparable to other such collections consist of cultivated genotypes, indicating that this population is suitable for association mapping. Several studies in other crops also reported that the genetic diversity of cultivated species was always lower than the corresponding wild species [26–29]. It is essential to deploy diverse and wild genetic resources into Chinese cultivars to broaden their genetic base of founder parents for enhancing the genetic diversity and achieving higher genetic gains. The information available through genotyping and multilocation phenotyping will further facilitate identification of potential founder parents for the ongoing breeding program.

Population structure is an important component in association mapping analysis and it helps in reducing the detection of false positives among associated markers. The STRUCTURE analysis identified two subpopulations for the 292 accessions (Fig. 1b) which is also been indicated from the dendrogram tree and PCA analysis reached similar results (Fig. 1c and 1d). The peanut germplasm

collections in previous studies could be divided into 2 to 4 subpopulations, which were always associated with the types of botanical varieties [21, 24, 30–32]. In the present study, the landraces in the peanut panel could be clearly divided into subsp. *hypogaea* (G1) and subsp. *fastigiata* (G2), respectively (Fig. 1c). However, most peanut cultivars and breeding lines in this population harbored mixed morphological features from the reciprocal cross between different botanical varieties. Thus, it is hard to distinguish the botanical difference between two subgroups clearly. Most accessions in the G1 group were from the provinces distributed on Northern China and the Yangtze River region. More than half of accessions in the G2 group were from the provinces distributed in Southern China (Fig. 2a and Additional file 1: Table S1). Comparing to Southern China, the varieties from Northern China were more closely related to the varieties from the Yangtze River region (Fig. 2b). It is indicated that the geographic origins of accessions had a significant effect on the population structure. A similar phenomenon was observed in many other crops [26, 33–35]. Different climate condition and their corresponding cropping system among Northern China, Southern China, and the Yangtze River region, might be responsible for genetic differentiation of the peanut population in China, enabling peanut varieties to adapt to various ecological environments.

Oil content is an important trait in peanut breeding, which acts as polygenic inheritance. In the present study, the associated analysis was performed to evaluate the phenotypic effect of multiple alleles in the diverse genetic background across multiple environments. A total of 42 alleles for twelve associated loci, which explained 4.54–9.94% phenotypic variance, were identified for oil content (Table 3). Interestingly, the favorable alleles with relatively higher effects were relatively abundant in the varieties from Northern China and the Yangtze River region, compared with the varieties from Southern China (Fig. 4b and 5c). Correspondingly, oil content of varieties showed a geographic difference clearly. The accessions from Northern China and the Yangtze River region had significantly higher phenotypic values than the accessions from Southern China (Fig. 3). It seemed that the trait of oil content and its underlying loci may undergo selection during geography differentiation in China. However, more experimental evidence, such as a multiple-ecological investigation of phenotype, was needed to verify the hypothesis.

The trait of oil content was inherited in an additive manner for oil crops [13, 36–38], which could facilitate pyramiding associated loci straightforward in the breeding program. In this study, associated markers were widely located on chromosome A03, A04, A08, A09, A10, B01, B06, B07, and B08 based on the information from previous linkage maps and the physical location on genome [39–42]. Compared with the previous results, chromosome A03, A04, A08, B06, B07, and B08 were also found to harbor QTLs for oil content [11, 12, 14]. For instance, the associated marker AHGS0798 on chromosome B06 (124.9 Mb), is close to qOCB06.1 (121.9 Mb-124 Mb) detected in the RIL population derived from Xuhua 13 and Zhonghua 6 [14]. Another two markers (TC1A02 and AHGS0393), which were highly linked to QTLs for oil content in the early studies [11, 12], were located at 127.5 Mb and 139.3 Mb on chromosome B06. These results suggested that AHGS0798 with PVE of 7.28% would be a reliable marker associated with oil content. In addition, three associated markers (AGGS1014_2, pPGPseq8D9, and TC11B4_2) in the present study could not collocate with the previous QTLs, suggesting that they are newly identified loci controlling oil content. Among them, the locus (AGGS1014_2) was repeatedly detected in three environments, with the maximum $-\text{LogP}$ value of $2.50\text{E-}06$ and PVE of 9.94% (Table 3). To evaluate the potential value of these loci in peanut breeding, two associated markers (AHGS0798 and AGGS1014_2) were verified in the RIL population derived from Zhonghua 10 and ICG12625. The favorable allele of single locus (AGGS1014_2 or AHGS0798) could increase oil content by $\sim 0.34\%$ - $\sim 1.50\%$ or $\sim 0.61\%$ - $\sim 0.88\%$ in four environmental trials. Combining favorable alleles for two loci, oil content could increase to $\sim 1.11\%$ - $\sim 2.06\%$ (Fig. 5 and Additional file 1: Table S5). It is indicated that using associated markers to accumulate favorable alleles would be an effective way to increase oil content in peanut breeding. In the present peanut panel, AGGS1014_2-215 bp/AHGS0798-174 bp is one of six combined genotypes between two associated markers, which expressed the highest phenotypic effect. In the varieties released before 1980, one accession (4%) possessed this genotype whereas the frequency increased to 22.13% (27 accessions) in the varieties released after 2000 (Additional file 2: Fig. S4). It is suggested that the selection of favorable alleles of AGGS1014_2 and AHGS0798 has been underway in China during the breeding program.

In conclusion, this study provided insight into the close relationship of the geographical region with population structure and the trait of oil content in China and informed allelic variations of oil content in Chinese cultivars collections. Two associated markers (AGGS1014_2 and AHGS0798) in the present study were verified to be valuable in MAS for oil improvement in peanut.

Methods

Plant materials and field planting

A total of 222 cultivars, 55 breeding lines, and 15 landraces from 17 different provinces in China were selected to constitute the peanut panel (Additional file 1: Table S1). A RIL population was developed from a cross between Zhonghua 10 and ICG12625 using single seed descent method and was later used for performing validation of associated markers.

The peanut panel was planted in the experimental field of the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences in Wuhan from 2015 to 2017 and also in the experimental field of the Institute of Nanchong Agricultural Science and Technology in Nanchong in 2017. The RIL population was planted in the experimental field of the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences in Wuhan from 2015 to 2017 and in the experimental field of the Xiangyang Academy of Agricultural Science in Xiangyang in 2017. Field trials were conducted in a randomized complete block design with three replications. Each replication contained 12 plants at a spacing of 20 cm × 30 cm. Field management followed standard agricultural practice.

DNA isolation and genotyping

The genomic DNA of 292 peanut accessions was extracted from fresh leaves following the modified cetyltrimethylammonium bromide method. The quality and quantity of DNA were checked using 1% agarose gel and NanoDrop (Thermo Fisher Scientific, USA), respectively.

A total of 4485 SSR markers obtained from the published literature were used to screen the diverse panel for identifying polymorphic markers [43-51]. The polymorphic marker primers labeled with fluorescence dyes were used to perform PCR amplification. The PCR production mixed with GeneScan 500 LIZ standard (Applied Biosystems, USA) was loaded to perform capillary electrophoresis using

3730 DNA Analyzer (Applied Biosystems, USA). The output of electrophoretic data was visualized and transferred to allele size using GeneMarker V2.1 software (<https://softgenetics.com/GeneMarker.php>). The SSR allele was numerically coded according to the fragment size.

Genotypic data analysis

The allele number, major allele frequency, genetic diversity and polymorphism information content (PIC) were calculated using PowerMarker V3.25 software [52]. The number of subgroups of this peanut panel was estimated using STRUCTURE software V2.1 based on the model-based Bayesian clustering method [53]. To determine an optimum number of subgroups (K), five independent runs were performed to estimate each K values from 1 to 10 for each accession. For each run, a burn-in length of 50000 followed by 10000 iterations were conducted with the admixture and related frequency models. The optimal K value was determined by the posterior probability [$\ln P(D)$] and ΔK [54].

Phylogenetic analysis was performed to construct a UPGMA tree based on Nei's distance. Nei's distance was calculated using PowerMarker [52] and the tree was formed using MEGA 4.0 (<http://www.megasoftware.net>). Principal component analysis (PCA) was complement using R package "FactoMineR" (<https://cran.r-project.org/web/packages/FactoMineR/index.html>) and three-dimensional scatter plot of PCA was completed using R package "scatterplot3d" (<https://cran.r-project.org/web/packages/scatterplot3d/>).

Evaluation of oil content and phenotypic data analysis

The percentages of Oil and H₂O in seeds were measured using nuclear magnetic resonance (PQ001, Niumag, China). Matured seeds (~10g) with less than 10% moisture content were analyzed for each of the three sub-samples per entry. Oil content (%) was calculated based on dry-weight using the formula $\{[\text{oil}\% / (100 - \text{H}_2\text{O}\%)] \times 100\}$ [11].

The field trials in Wuhan in 2015, 2016 and 2017 were treated as Environment I, II and III, respectively. The field trial in Nanchong in 2017 was treated as Environment IV. The phenotypic data statistical analyses were performed using the IBM SPSS Statistics software (V.22, IBM, USA). The family-based broad-sense heritability for oil content was calculated as $H^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2/n)$, where σ_g^2 is the genotypic variance, σ_e^2 is the residual error, n is the number of environments.

Marker-trait association analysis

Associations between SSR markers and the trait of oil content were performed using TASSEL software based on a Q+K mixed linear model [55]. The population structure (Q) was obtained from model-based program STRUCTURE V2.1 [53]. The pairwise kinship matrix (K) was calculated using SPAGeDi software [56]

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The raw phenotype data and genotype data are available from the corresponding author on reasonable request

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

Founding

This work was financially supported by National Natural Science Foundation of China (no. 31871666, no. 31801403, and no. 31761143005), the National Peanut Industry Technology System Construction (CARS-13), Plant Germplasm Resources Sharing Platform (NICGR2017-36), the National Program for Crop Germplasm Protection of China (2019NWB033), and the Science and technology innovation Project of Chinese Academy of Agricultural Sciences.

Authors' contributions

NL and HJ conceived and designed the experiments. WC, JG, HC, YC, YL and XR conducted the field experiment. NL, LH, BW and XZ conducted the molecular experiment. NL, HL, and DH analysed the data. NL wrote the manuscript. MP, BL, RV, and HJ revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. Baring MR, Wilson JN, Burow MD, Simpson CE, Ayers JL, Cason JM. Variability of total oil content in peanut across the state of Texas. *Journal of Crop Improvement*. 2013;27:125–36.
2. Wilson J, Baring M, Burow M, Rooney W, Simpson C. Generation means analysis of oil concentration in peanut. *Journal of Crop Improvement*. 2013;27:85–95.
3. Chu Y, Wu CL, Holbrook CC, Tillman BL, Person G, Ozias-Akins P. Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome*. 2011;4:110–7.
4. Janila P, Variath MT, Pandey MK, Desmae H, Motagi BN, Okori P, et al. Genomic tools in groundnut breeding program: status and perspectives. *Front Plant Sci*. 2016;7:289.
- 5.

- Varshney RK. Exciting journey of 10 years from genomes to fields and markets: some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 2016;242:98–107.
- 6.
- Pandey MK, Roorkiwal M, Singh VK, Ramalingam A, Kudapa H, Thudi M, et al. Emerging genomic tools for legume breeding: current status and future prospects. *Front Plant Sci.* 2016;7:455.
- 7.
- Anacleto R, Cuevas RP, Jimenez R, Llorente C, Nissila E, Henry R, Sreenivasulu N. Prospects of breeding high-quality rice using post-genomic tools. *Theor Appl Genet.* 2015;128:1449–66.
- 8.
- Beisson F, Koo A, Ruuska S, Schwender J, Pollard M, Thelen JJ, et al. Arabidopsis genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.* 2003;132:681–97.
- 9.
- Baud S, Lepiniec L. Physiological and developmental regulation of seed oil production. *Prog Lipid Res.* 2010;49:235–49.
- 10.
- Yol E, Ustun R, Golukcu M, Uzun B. Oil content, oil yield and fatty acid profile of groundnut germplasm in Mediterranean climates. *J Am Oil Chem Soc.* 2017;94:1–18.
- 11.
- Pandey MK, Wang ML, Qiao L, Feng S, Khera P, Wang H, 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.
- 12.
- Wilson JN, Chopra R, Baring MR, Selvaraj MG, Simpson CE, Chagoya J, et al. Advanced backcross quantitative trait loci (QTL) analysis of oil concentration and oil quality traits in peanut (*Arachis hypogaea* L.). *Trop Plant Biol.* 2017;10:1–17.
- 13.
- Shasidhar Y, Vishwakarma MK, Pandey MK, Janila P, Variath MT, Manohar SS, et al. Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis hypogaea* L.). *Front Plant Sci.* 2017;8:794.
- 14.
- Liu N, Guo J, Zhou X, Wu B, Huang L, Luo H, 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.* 2019;26.
- 15.
- Hall D, Tegstrom C, Ingvarsson PK. Using association mapping to dissect the genetic basis of complex traits in plants. *Brief Funct Genomics.* 2010;9:157–65.
- 16.
- Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol.* 2003;54:357–74.
- 17.
- Sun Z, Wang X, Liu Z, Gu Q, Zhang Y, Li Z, et al. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in *Gossypium hirsutum* L. *Plant Biotechnol J.* 2017;15:982–96.
- 18.
- Wang XL, Wang HW, Liu SX, Ferjani A, Li JS, Yan JB, et al. Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. *Nat Genet.* 2016;48:1233–41.

19.
Si LZ, Chen JY, Huang XH, Gong H, Luo JH, Hou QQ, et al. OsSPL13 controls grain size in cultivated rice. *Nat Genet.* 2016;48:447–56.
20.
Wang B, Wu Z, Li Z, Zhang Q, Hu J, Xiao Y, et al. Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in *Brassica napus*. *Plant Biotechnol J.* 2018;16:1336–48.
21.
Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshayee MS, Sriswathi M, et al. Genomewide association studies for 50 agronomic traits in peanut using the ‘Reference Set’ comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. *PLoS ONE.* 2014;9:e105228.
22.
Wang X, Xu P, Yin L, Ren Y, Li S, Shi Y, et al. Genomic and transcriptomic analysis identified gene clusters and candidate genes for oil content in peanut (*Arachis hypogaea* L.). *Plant Mol Biol Report.* 2018;36:518–29.
23.
Shirasawa K, Koilkonda P, Aoki K, Hidakawa H, Tabata S, Watanabe M, et al. In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. *BMC Plant Biol.* 2012;12:80.
24.
Shirasawa K, Bertoli DJ, Varshney RK, Moretzsohn MC, Leal-Bertoli SC, Thudi M, et al. Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes. *DNA Res.* 2013;20:173–84.
25.
Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, et al. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). *Theor Appl Genet.* 2004;108:1064–70.
26.
He GH, Meng RH, Gao H, Guo BZ, Gao GQ, Newman M, et al. Simple sequence repeat markers for botanical varieties of cultivated peanut (*Arachis hypogaea* L.). *Euphytica.* 2005;142:131–6.
27.
Proite K, Leal-Bertoli S, Bertoli DJ, Moretzsohn MC, da Silva FR, Martins NF, et al. ESTs from a wild *Arachis* species for gene discovery and marker development. *BMC Plant Biol.* 2007;7:7.
28.
Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK. Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). *BMC Plant Biol.* 2008;8:55.
29.
Liang XQ, Chen XP, Hong YB, Liu HY, Zhou GY, Li SX, Guo BZ. Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol.* 2009;9:35.
30.
Zhang J, Shan L, Duan J, Jin W, Chen S, Cheng Z, et al. De novo assembly and characterisation of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genom.* 2012;13:90.
31.
Luo H, Xu Z, Li Z, Li X, Lv J, Ren X, et al. Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.). *Theor Appl Genet.* 2017;130:1635–48.
- 32.

- Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*. 2005;21:2128-9.
- 33.
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*. 2003;164:1567-87.
- 34.
- Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611-20.
- 35.
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23:2633-5.
- 36.
- Hardy OJ, Vekemans X. SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2002;2:618-20.
- 37.
- Ren X, Jiang H, Yan Z, Chen Y, Zhou X, Huang L, et al. Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. *PLoS ONE*. 2014;9:e88091.
- 38.
- Jiang H, Huang L, Ren X, Chen Y, Zhou X, Xia Y, et al. Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) mini-core collection. *J Integr Plant Biol*. 2014;56:159-69.
- 39.
- Zhao J, Huang L, Ren X, Pandey MK, Wu B, Chen Y, et al. Genetic variation and association mapping of seed-related traits in cultivated peanut (*Arachis hypogaea* L.) using single-locus simple sequence repeat markers. *Front Plant Sci*. 2017;8:2105.
- 40.
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol*. 2015;33:408-14.
- 41.
- Varshney RK, Thudi M, Roorkiwal M, He W, Upadhyaya HD, Yang W, et al. Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat Genet*. 2019;51:857-64.
- 42.
- Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature*. 2012;490:497-501.
- 43.
- Hufford MB, Xu X, Heerwaarden J, Pyhäjärvi T, Chia J, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 2012;44:808-11.
- 44.
- Zheng Z, Sun Z, Fang Y, Qi F, Liu H, Miao L, et al. Genetic diversity, population structure, and botanical variety of 320 global peanut accessions revealed through tunable genotyping-by-sequencing. *Scientific reports*. 2018;8:14500.
- 45.
- 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.

46.
Ming LW, Sivakumar S, Noelle AB, Zhenbang C, Charles YC, Baozhu G, et al. Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *Theor Appl Genet.* 2011;123:1307-17.
47.
Wang H, Zhu SS, Dang XJ, Liu EB, Hu XX, Eltahawy MS, Zaid IU, Hong DL. Favorable alleles mining for gelatinization temperature, gel consistency and amylose content in *Oryza sativa* by association mapping. *BMC Genet.* 2019;20:34.
48.
Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, et al. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci U S A.* 2013;110:453-8.
49.
Noble TJ, Tao Y, Mace ES, Williams B, Jordan DR, Douglas CA, et al. Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. *Front Plant Sci.* 2017;8:2102.
50.
Qi Z, Zhang X, Qi H, Xin D, Han X, Jiang H, et al. Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. *Euphytica.* 2017;213:162.
51.
Fu Y, Zhang D, Gleeson M, Zhang Y, Lin B, Hua S, et al. Analysis of QTL for seed oil content in *Brassica napus* by association mapping and QTL mapping. *Euphytica.* 2017;213:17.
52.
Li H, Peng ZY, Yang XH, Wang WD, Fu JJ, Wang JH, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet.* 2013;45:43-50.
53.
Liu N, Chen H, Huai D, Xia F, Huang L, Chen W, et al. Four QTL clusters containing major and stable QTLs for saturated fatty acid contents in a dense genetic map of cultivated peanut (*Arachis hypogaea* L.). *Mol Breed.* 2019;39:23.
54.
Luo H, Guo J, Ren X, Chen W, Huang L, Zhou X, et al. Chromosomes A07 and A05 associated with stable and major QTLs for pod weight and size in cultivated peanut (*Arachis hypogaea* L.). *Theor Appl Genet.* 2017;131:267-82.
55.
Bertioli DJ, Jenkins J, Clevenger J, Dudchenko O, Gao D, Seijo G, et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet.* 2019;51:877-84.
56.
Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C, et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet.* 2019;51:865-76.

Tables

Table 1 Statistic summary for population diversity.

Population	Sample Size	Allele Number	Major Allele Frequency	Genetic Diversity
G1	161	5.36	0.72	0.40
G2	131	5.29	0.66	0.47
Total	292	6.28	0.61	0.51

Table 2 Phenotypic variation for oil content (%) for 292 peanut accessions across four environments.

Env	Min (%)	Max (%)	Median (%)	SD	Kurt
2015WH	45.85	59.72	51.89	2.39	-0.04
2016WH	43.82	55.88	50.13	1.78	0.54
2017NC	44.22	54.97	48.47	1.94	0.66
2017WH	45.11	56.69	51.53	1.86	0.32

Env, environment; SD, standard deviation; Kurt, Kurtosis; Skew, skewness; w, Shariro-Wilk statistic value; Sig, significance

Table 3 Marker-trait associations across four environments for oil content.

Marker	Environment	<i>F</i> -value	<i>P</i> -value	PVE (%)
pPGPseq8D9	2017NC	13.22	3.29E-04	4.61
TC9F10_2	2017WH	6.48	3.09E-04	7.59
TC11B4_2	2017WH	4.08	1.43E-03	8.84
AHGS1679	2017WH	5.95	6.04E-04	6.39
AGGS1149	2016WH	6.54	1.68E-03	4.54
AGGS1081	2016WH	5.47	1.15E-03	5.76
AGGS1014_2	2016WH	23.23	2.50E-06	9.94
	2017NC	14.89	1.45E-04	6.90
	2017WH	21.43	5.90E-06	8.75
AHGS2053	2016WH	6.29	3.81E-04	6.65
AHS0127	2016WH	10.04	6.13E-05	6.99
AHGS1431	2016WH	9.11	1.52E-04	7.35
AHGS0798	2015WH	9.54	1.03E-04	7.28
AHGS1388	2016WH	8.84	1.94E-04	6.78

PVE phenotypic variance explained

Figures

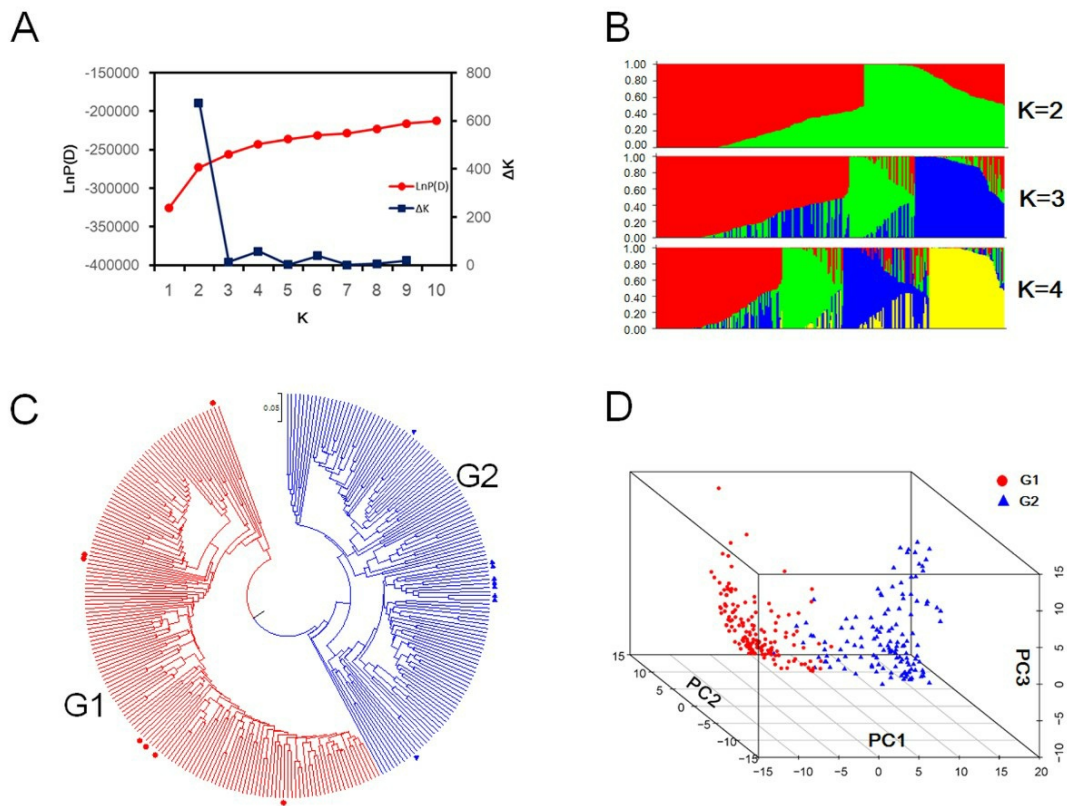
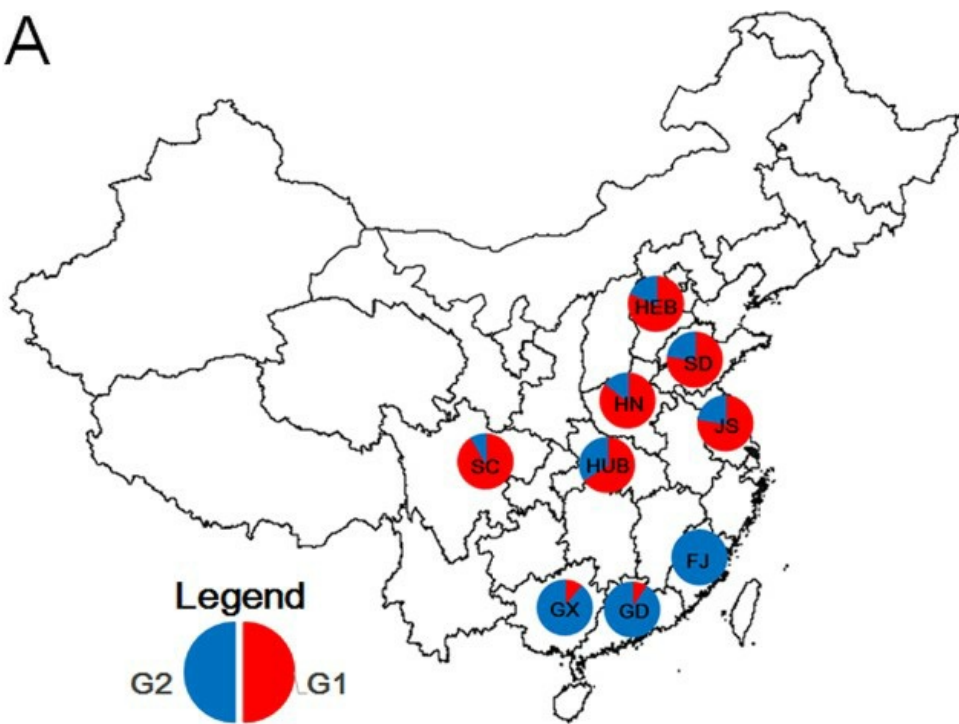


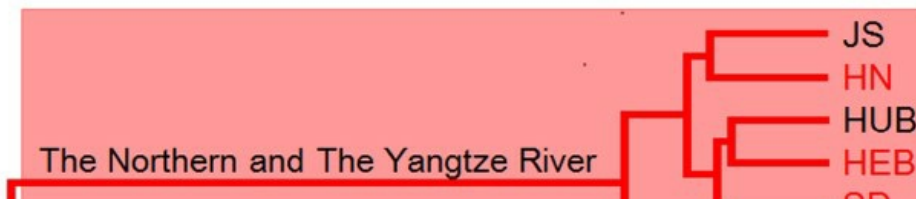
Figure 1

Population structure analysis in 292 peanut accessions. a Determination of optimal K based on LnP(D) and ΔK values. b the population structure in the peanut panel at K=2, 3, 4. c Phylogenetic tree of the peanut panel based on Nei's genetic distance. The 292 peanut accessions were grouped into two clusters G1 (red lines) and G2 (blue lines). The red dots represented the landraces belonging to Subsp. hypogaea, the blue triangles denoted the landraces belonging to Subsp. fastigiata. d Three-dimensional scatter plots of the first three principal components. The red dots represented cluster G1 in the phylogenetic tree. The blue triangles represented cluster G2 in the phylogenetic tree.

A



B



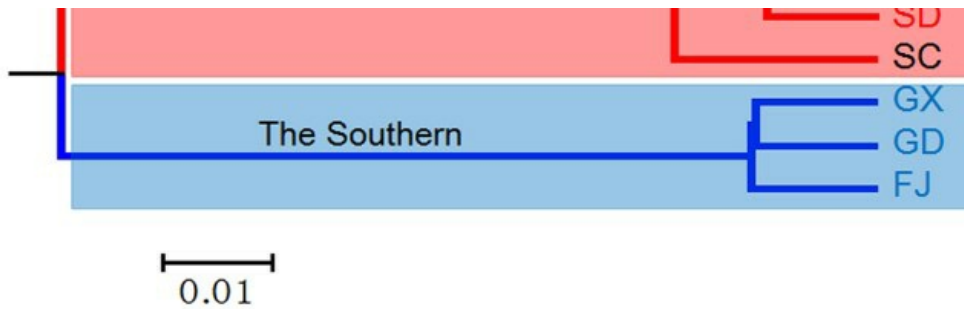


Figure 2

Geographical structure in the peanut panel. a The proportion of two groups G1 and G2 (fig.1c) in China. b Phylogenetic tree of the peanut accessions grouped by original provinces. HEB, Hebei province; SD, Shandong province; HN, Henan province; JS, Jiangsu province; HUB, Hubei province; SC, Sichuan province; FJ, Fujian province; GD, Guangdong province; GX, Guangxi province. HEB, SD, and HN belong to the Northern China. JS, HUB, and SC belong to the Yangtze River region in China. FJ, GD, and GX belong to the Southern China.

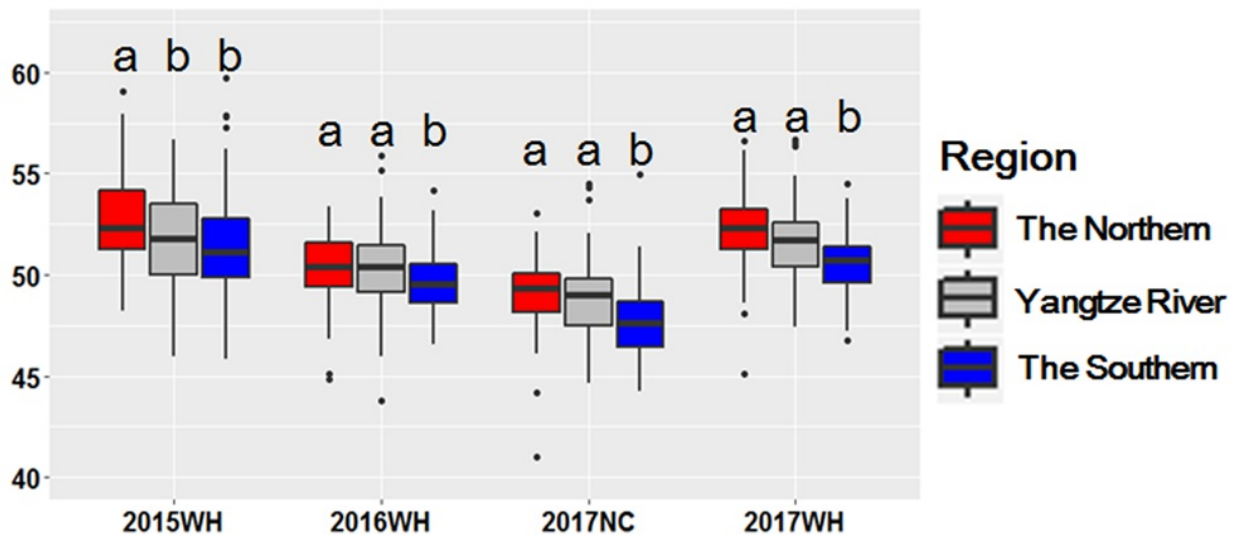


Figure 3

Comparison of oil content (%) among peanut accessions from different geographic regions. The boxes with different letters were significantly different according to Tukey's Multiple Comparison Test ($P < 0.05$).

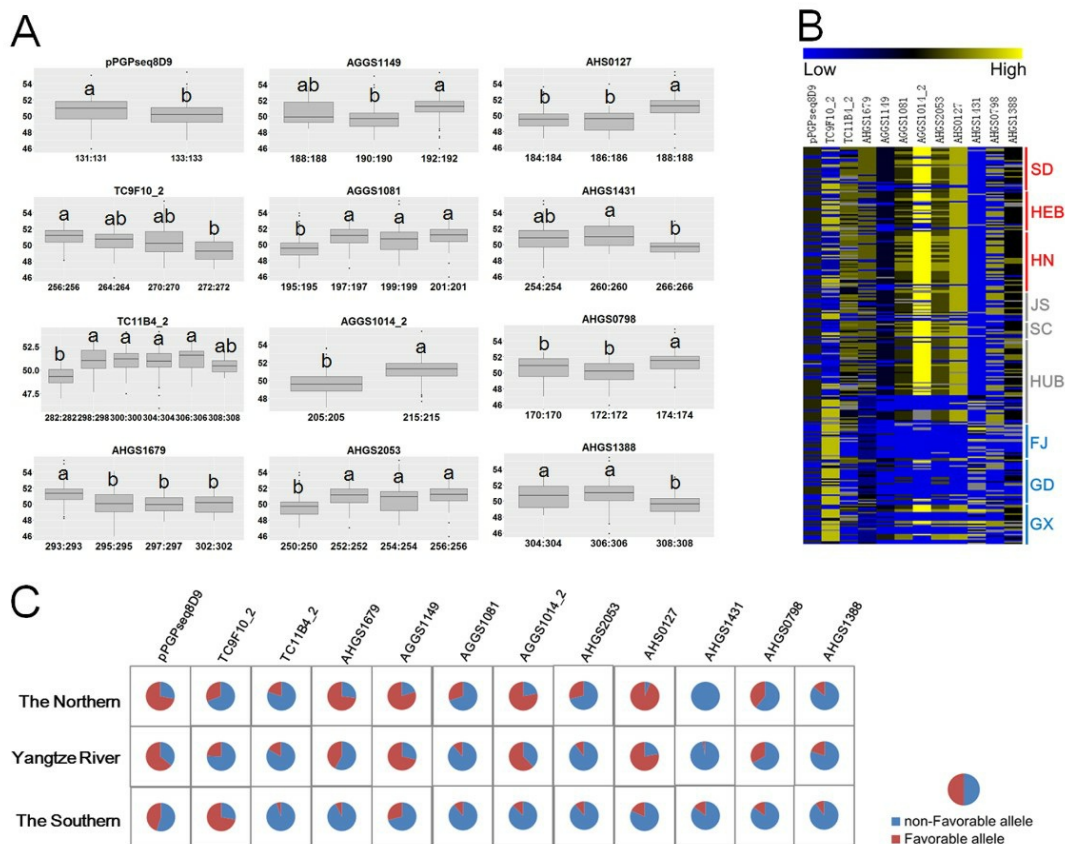


Figure 4

Phenotypic effect and geographic distribution of favorable alleles of trait-associated markers. a Comparison of allelic effects of trait-associated loci based on average values of four environmental data. The boxes with different letters were significantly different according to Tukey's Multiple Comparison Test ($P < 0.05$). b Overview of allelic effects of each trait-associated locus in peanut accessions distributed on nine provinces. c The spectrum of the most favorable allele frequencies in different geographic regions. HEB, Hebei province; SD, Shandong province; HN, Henan province; JS, Jiangsu province; HUB, Hubei province; SC, Sichuan province; FJ, Fujian province; GD, Guangdong province; GX, Guangxi province. HEB, SD, and HN belong to the Northern China. JS, HUB, and SC belong to the Yangtze River region in China. FJ, GD, and GX belong to the Southern China.

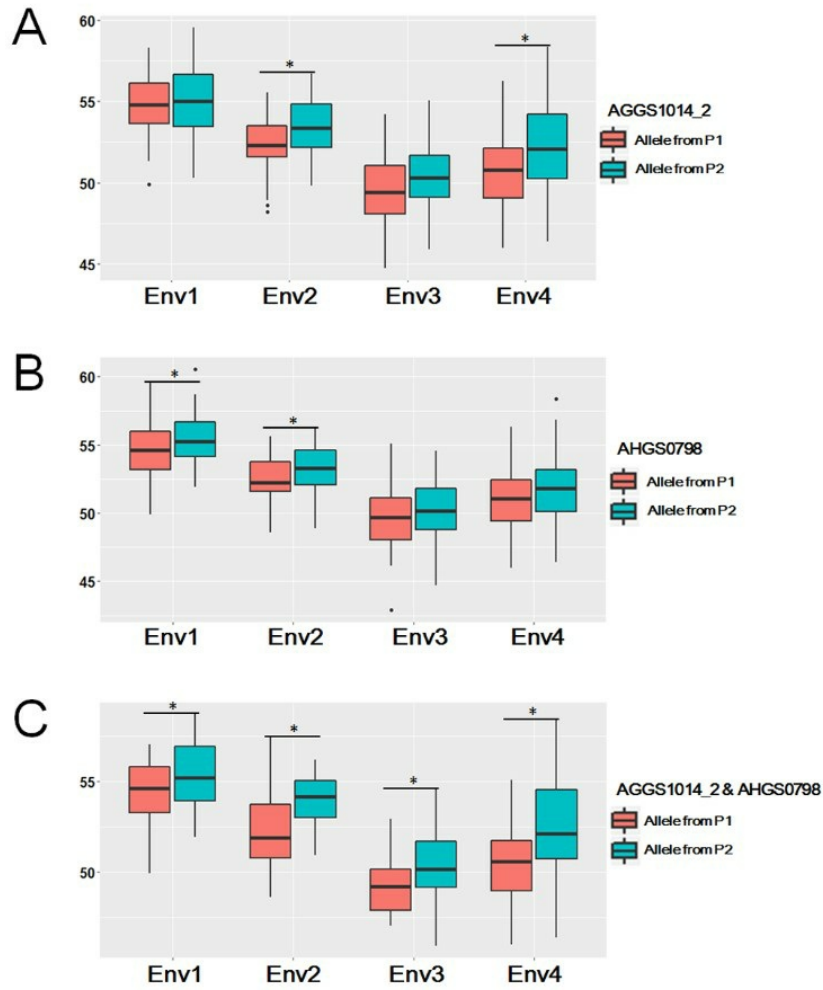


Figure 5

Confirmation of two trait-associated markers in a RIL population. Env1, 2, 3 and 4 represented field trials in Wuhan (2015), Wuhan (2016), Xiangyang (2017) and Wuhan (2017). P1 female parent, P2 male parent.

Supplementary Files

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

[additional file 2.docx](#)

[additional file 1.xlsx](#)