BMC Genetics

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

Nian Liu[1], Li Huang[1], Weigang Chen[1], Bei Wu[1], Manish K. Pandey[2], Huaiyong Luo[1], Xiaojing Zhou[1], Jianbin Guo[1], Haiwen Chen[1], Dongxin Huai[1], Yuning Chen[1], Yong Lei[1], Boshou Liao[1], Xiaoping Ren[1], Rajeev K. Varshney[2] and Huifang Jiang[1*]

Abstract

Background: Peanut is one of the primary sources for vegetable oil worldwide, and enhancing oil content is the main objective in several peanut breeding programs of the world. Tightly linked markers are required for faster development of high oil content peanut varieties through genomics-assisted breeding (GAB), and association mapping is one of the promising approaches for discovery of such associated markers.

Results: An association mapping panel consisting of 292 peanut varieties extensively distributed in China was 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 structure, phylogenetic relationship, and principal component analysis (PCA) indicated two subgroups majorly differentiating based on geographic regions. Genome-wide association analysis 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 and also indicated selection during the breeding program.

Conclusion: This study provided insights into the genetic basis of oil content in peanut and verified highly associated two SSR markers to facilitate marker-assisted selection for developing high-oil content breeding peanut varieties.

Keywords: Peanut, Oil content, Association analysis, Geographic region

Background

Cultivated peanut or groundnut (*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 [1]. The global annual planting area is stands at 28.52 Mha which yielded, with the annual production of 45.95 Mt

of peanut production in 2018 [2]. Despite China being the largest peanut producer in the world, the current production hardly meets the increasing domestic demands for peanut oil. High seed oil content is one of the most significant traits, and increase of just 1% oil content in popularly grown varieties results in 7% increase in economic benefit to the farmers and oil processing units. Thus, enhancing oil content is an important long-term objective in all the peanut breeding programs not only in China but in many other countries such as India.

* Correspondence: peanutlab@oilcrops.cn
[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062, People's Republic of China
Full list of author information is available at the end of the article

Liu *et al. BMC Genetics* (2020) 21:60

Page 2 of 12

Oil content is a polygenetic trait and is significantly influenced by the environment [3, 4]. 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 process. Deployment of well-validated markers could accelerate the process and precision in genetic improvement [5–9]. Dissection of the genetic architecture underlying oil content is a prerequisite to deploying markers in high-oil content breeding programs. Triacylglycerol is the major form of storage oil in most plant seeds including peanut. Its biosynthetic pathway and the relevant genes have been extensively understood in the model plant *Arabidopsis* [10, 11]. But, the genetic mechanism of natural variation for oil biosynthesis remains poorly understood in peanut.

The available significant variation of oil content among peanut germplasm provides an opportunity to identify genomic regions controlling higher oil accumulation [12]. Using the bi-parental 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 [13]. Subsequently, three and eight QTLs were detected in an advanced backcross population and RIL population, respectively [14, 15]. 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) [16]. 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 above mentioned studies. Compared with linkage analysis, association analysis, utilizing historical recombination in natural populations, facilities high-resolution mapping and testing of multiple-alleles in far less time [17, 18]. This genetic method has been successfully used to reveal the genetic basis of complex agronomic traits in multiple crops [19–22], however, this approach has so far been deployed in couple of studies in peanut [23, 24].

Cultivars and elite germplasm lines possessing desirable traits get preference by plant breeders in using them as trait source in almost all the genetic improvement programs as compared to wild relatives. Therefore, discovery of associated genomic regions and potential candidate genes in elite association 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, (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, population structure and linkage disequilibrium analysis

A total of 583 simple sequence repeats (SSR) polymorphic markers randomly distributed on the genome, were used to genotype the association mapping panel of 292 peanut accessions. The polymorphic markers produced 3663 alleles with an average of 6.28 alleles per locus ranging from 2 to 20 (Table 1 and Additional file 1: Table S1). 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 S1). Of the 3663 alleles, 629 were unique alleles (allele frequency < 0.05%), 1471 allele were rare alleles (0.05% ≤ allele frequency < 5%), 1547 allele were polymorphic alleles (5% ≤ allele frequency < 95%), and 15 were fixed alleles (allele frequency ≥ 95%), with corresponding proportions of 17.17, 40.16, 42.23 and 0.41%, respectively (Additional file 1: Table S2).

The population structure analysis was performed using multi-allelic SSR genotyping 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 b). 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 d). 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 (1972) genetic distance

Table 1 Statistic summary for population diversity

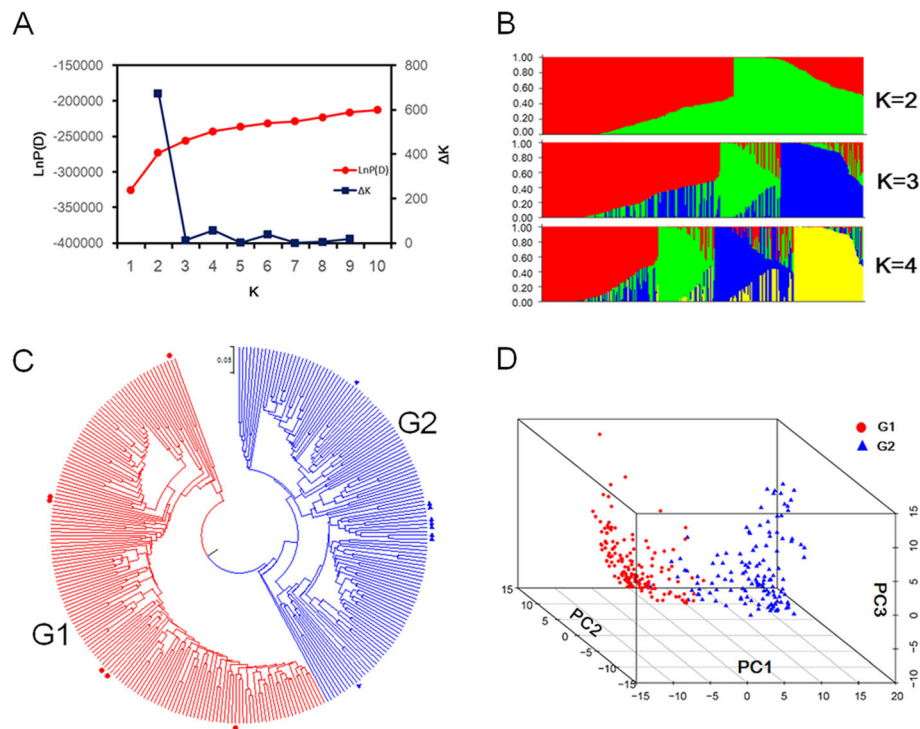| Population	Sample Size	Allele Number	Major Allele Frequence	Genetic Diversity	PIC
G1	161	5.36	0.72	0.40	0.36
G2	131	5.29	0.66	0.47	0.42
Total	292	6.28	0.61	0.51	0.45

Fig. 1 Population structure analysis in 292 peanut accessions. **a** Determination of optimal *K* based on LnP(D) and Δ*K* 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 (1972) 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

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).
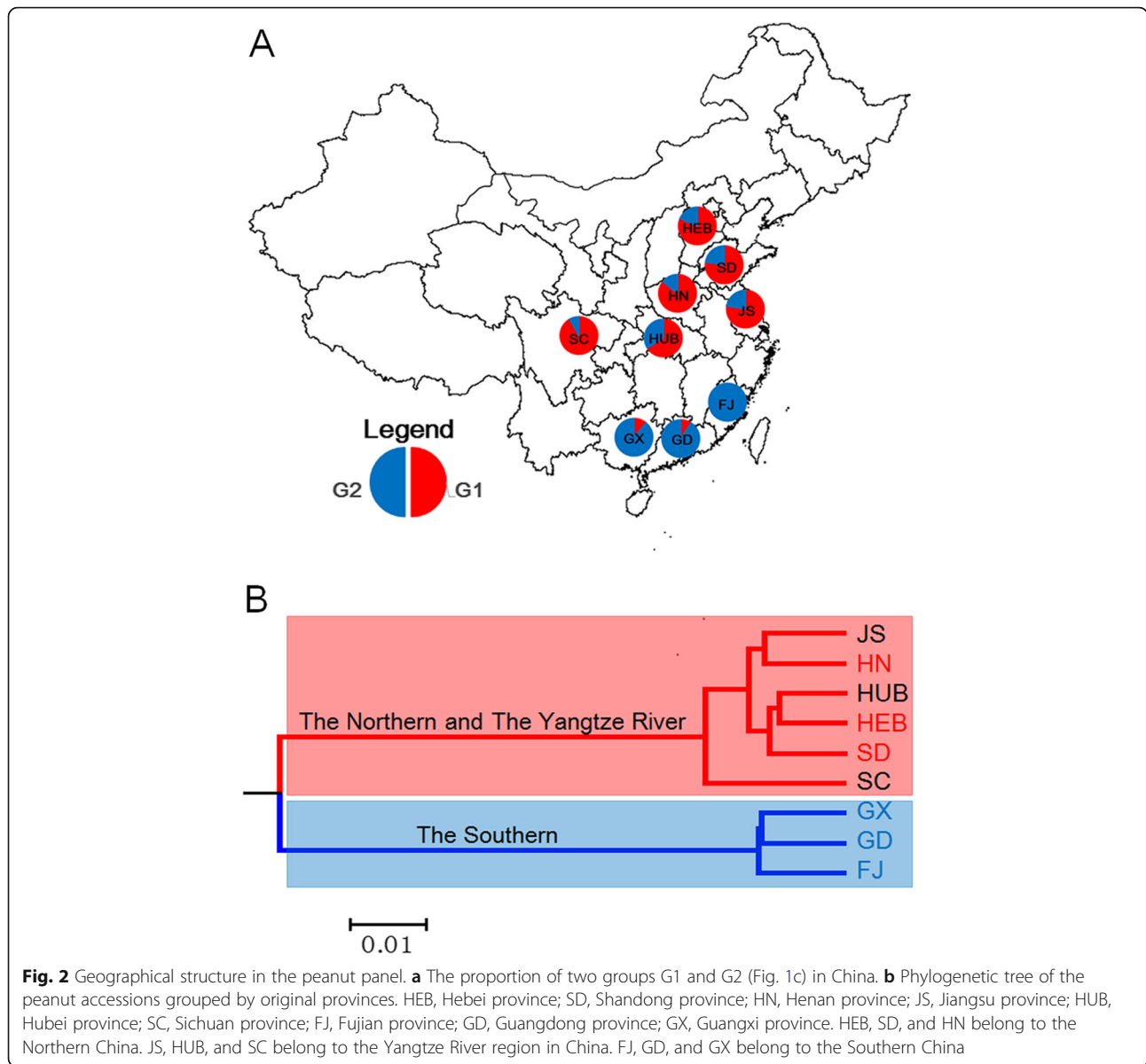
The peanut association mapping panel consisted of cultivars from 17 provinces of China. Most accessions (93.2%) were distributed in nine provinces (Hebei, Shandong, Henan, Sichuan, Hubei, Jiangsu, Fujian, Guangdong, and Guangxi). The proportion of two subgroups in these provinces exhibited obvious differences (Fig. 2a). In Northern China (Hebei, Shandong, and Henan 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 (Sichuan, Hubei, and Jiangsu 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.

The linkage disequilibrium (LD) was estimated using coefficients (r^2) of 280 SSR markers mapped on 20 linkage groups [25]. The average r^2 was 0.11 and almost 53.4% of the coefficients (r^2) showed statistically significant ($P < 0.01$). The 95th percentile of distribution of all r^2 between the unlinked marker-pairs, i.e., $r^2 = 0.28$, was set as the background level. Since the average distance of pair combinations was below 1 cM with the r^2 plot dropping to background level, the estimated LD decay in the peanut panel is 1 cM (Additional file 2: Fig. S1).

Phenotypic variation for oil content among peanut accessions

The analysis was done for the phenotyping data generated on oil content for 292 Chinese peanut accessions from seeds harvested from four environments. The oil content among association panel 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

Fig. 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

the standard deviation of phenotypic data ranged from 1.78 to 2.39 across four environments. Two elite cultivars (Zhonghua 15 and Yuhua 9326) which have superior yield potential, exhibited stably high-oil feature across four-environmental trials (average oil content >

55%). The continuous distributions of phenotypic values for peanut accessions were shown in Additional file 2: Fig. S2. 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 Phenotypic variation for oil content (%) for 292 peanut accessions across four environments

Env	Min (%)	Max (%)	Median (%)	SD	Kurt	Skew	w (Sig)	H^2
2015WH	45.85	59.72	51.89	2.39	−0.04	0.22	1.00 (0.51)	0.76
2016WH	43.82	55.88	50.13	1.78	0.54	0.00	0.99 (0.36)	
2017NC	44.22	54.97	48.47	1.94	0.66	0.06	0.99 (0.04)	
2017WH	45.11	56.69	51.53	1.86	0.32	0.06	1.00 (0.50)	

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

Liu *et al. BMC Genetics*   (2020) 21:60

Page 5 of 12

(Table 2). Variance analysis across four environmental trials showed that genotype, environment, and genotype × environment significantly influenced oil content at the $P < 0.001$ level (Additional file 2: Fig. S2). The broad sense heritability for oil content was evaluated to be 0.76 in the peanut panel.

We further studied phenotypic differences in the genetically highly diverse association mapping panel containing genotypes different geographic distributions in China. The oil content in the 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 had 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 of the four environments. Meanwhile, we also made a comparison among cultivated peanuts released at different times (Additional file 2: Fig. S3). In general, there was no obvious difference in oil content between cultivars released at different times.

Association analysis for oil content

The Mixed linear model with K + Q matrix was used to perform association mapping with SSR-markers and the phenotypic data on oil content generated on 292 peanut accessions in four environments. The marker-trait association analysis 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. S4). Among them,

AGGS1014_2 with up to 9.94% PVE had been repeatedly detected in multiple environments (2016WH, 2017NC, and 2017WH). These markers were widely distributed on nine linkage groups based on previously reported genetic maps (Additional file 1: Table S2). Physical position of associated markers were on 12.7 Mb of B01 (AGGS1014_2), 57.1 Mb of B07 (AGGS1081), 47.4 Mb of A03 (AGGS1149), 124.9 Mb of B06 (AHGS0798), 30.1 Mb of B08 (AHGS1388), 20.8 Mb of B06 (AHGS1431), 36.9 Mb of A04 (AHGS1679), 57.1 Mb of B07 (AHGS2053), 67.8 Mb of B07 (AHS0127), 119.6 Mb of A09 (pPGPseq8D9), 5.1 Mb of A10 (TC11B4_2), 35.5 Mb of A08 (TC9F10_2), respectively.

The allelic number of these associated loci ranged from two (pPGPseq8D9 and AGGS1014_2) to six (TC11B4_2). The most favorable alleles which have the largest effect values included pPGPseq8D9-131 bp, TC9F10_2-256 bp, TC11B4_2-298 bp, AHGS1679-293 bp, AGGS1149-192 bp, AGGS1081-201 bp, AGGS1014_2-215 bp, AHGS2053-256 bp, AHS0127-188 bp, AHGS1431-260 bp, AHGS0798-174 bp, and AHGS1388-304 bp (Table 3, Additional file 1: Table S3). The accessions with different alleles showed statistically significant difference in a four-environment average of oil content (Fig. 3a). Compared with accessions in Southern China (FJ, GD, and GX), the genotypes from Northern China and Yangtze River (SD, HEB, HN, JS, SC, and HUB) carried more alleles with relatively high effect (Fig. 3b). 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 most favorable allele frequency was highest in Northern China, the second-highest in the Yangtze River region, and lowest in Southern China (Fig. 3c). However, the

Table 3 Marker–trait associations across four environments for oil content

Marker	Environment	F-value	P-value	PVE(%)	Favorable allele
pPGPseq8D9	2017NC	13.22	3.29E-04	4.61	pPGPseq8D9-131 bp
TC9F10_2	2017WH	6.48	3.09E-04	7.59	TC9F10_2-256 bp
TC11B4_2	2017WH	4.08	1.43E-03	8.84	TC11B4_2-298 bp
AHGS1679	2017WH	5.95	6.04E-04	6.39	AHGS1679-293 bp
AGGS1149	2016WH	6.54	1.68E-03	4.54	AGGS1149-192 bp
AGGS1081	2016WH	5.47	1.15E-03	5.76	AGGS1081-201 bp
AGGS1014_2	2016WH	23.23	2.50E-06	9.94	AGGS1014_2-215 bp
	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	AHGS2053-256 bp
AHS0127	2016WH	10.04	6.13E-05	6.99	AHS0127-188 bp
AHGS1431	2016WH	9.11	1.52E-04	7.35	AHGS1431-260 bp
AHGS0798	2015WH	9.54	1.03E-04	7.28	AHGS0798-174 bp
AHGS1388	2016WH	8.84	1.94E-04	6.78	AHGS1388-304 bp

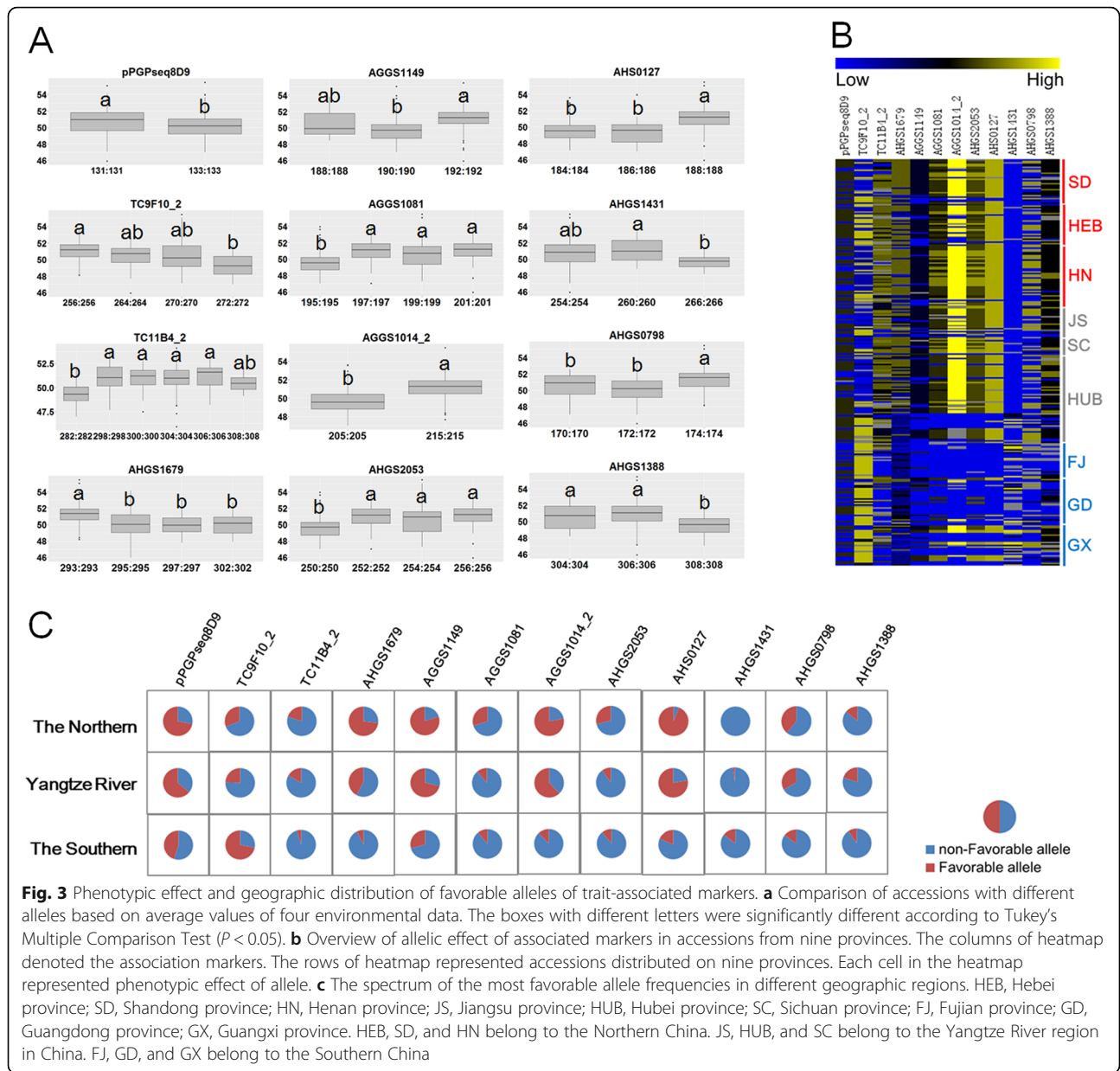*PVE* phenotypic variance explained

Fig. 3 Phenotypic effect and geographic distribution of favorable alleles of trait-associated markers. **a** Comparison of accessions with different alleles 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 effect of associated markers in accessions from nine provinces. The columns of heatmap denoted the association markers. The rows of heatmap represented accessions distributed on nine provinces. Each cell in the heatmap represented phenotypic effect of allele. **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

most 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 makers (AGGS1014_2, AHGS0798, and AHGS1431) showed association with oil content 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 S4). 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. 4a and Additional file 1: Table S4). 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. 4a and Additional file 1: Table S4). Combined allele effect of AGGS1014_2 and
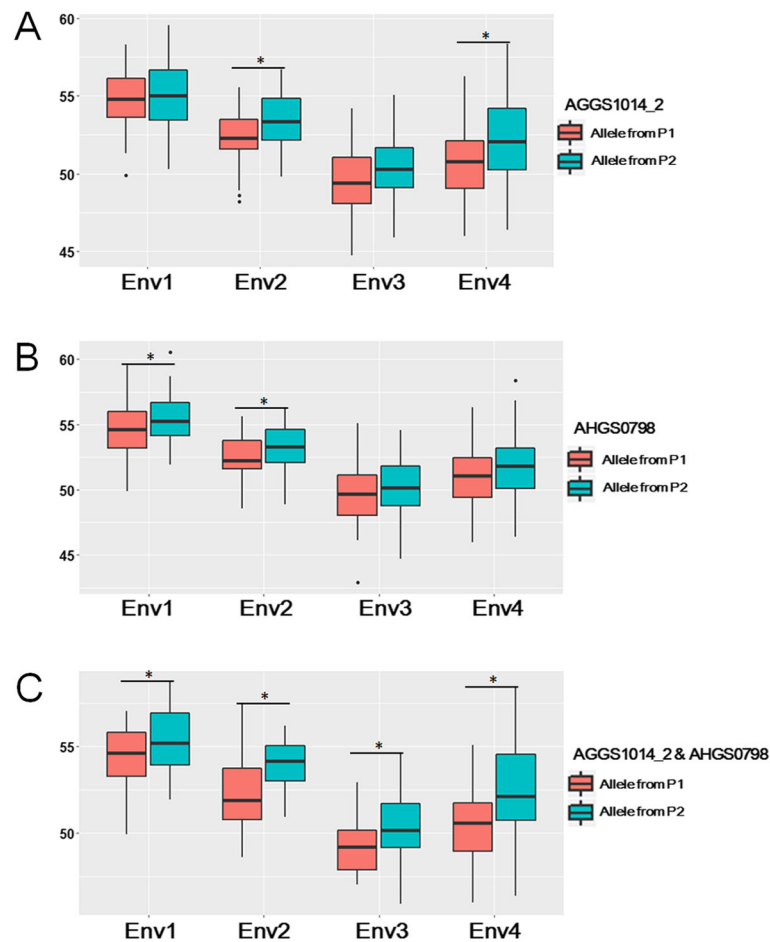
Fig. 4 Confirmation of two trait-associated markers in a RIL population. Env1, 2, 3 and 4 represented field trials in Wuhan (2015), Wuhan (2016), Xiangyang (2017) and Wuhan (2017). P1 female parent, P2 male parent

AHGS0798 showed that oil content of homozygous alleles from P2 was significantly higher than that of the homozygous allele from P1 across environments (Fig. 4c).

Among the 292 peanut accessions, the alleles at AGGS1014_2 (X) locus and AHGS798 (Y) locus formed six combined genotypes, namely X-205 bp/Y-170 bp, X-205 bp/Y-172 bp, X-205 bp/Y-174 bp, X-215 bp/Y-170 bp, X-215 bp/Y-172 bp, X-215 bp/Y-174 bp (Additional file 2: Fig. S5). The oil content is highest in X-215 bp/Y-174 bp (51.49 ± 1.30%), median in X-215 bp/Y-172 bp (51.04 ± 1.11%), X-215 bp/Y-170 bp (50.88 ± 1.17%), X-205 bp/Y-174 bp (50.66 ± 1.38%), and X-205 bp/Y-170 bp (49.72 ± 2.10%), and lowest in X-205 bp/Y-172 bp (49.62 ± 1.19%). The genotypic frequency of X-215 bp/Y-174 bp 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-205 bp/Y-172 bp has an increase from 12.00% in peanut varieties released before 1980 to 32.79% in peanut varieties released after 2000. The frequency of X-205 bp/Y-172 bp

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

Discussion

For performing high resolution mapping using GWAS studies, multilocation phenotyping data for target traits on diverse panel together with genotyping data is necessary for discovery of significantly associated markers. Keeping this in mind, this research effort used the peanut panel which consists of 222 cultivars, 55 breeding lines, and 15 landraces representing 17 provinces of China. The screening of genome-wide SSR marker in the peanut panel produced high number (3663) 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 [26–28]. 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 [23] 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 consisting 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 [29–32]. 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 had also been indicated by the dendrogram tree and PCA analysis (Fig. 1c and d). 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 [23, 27, 33–35]. 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 S5). 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 [29, 36–38]. 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 with polygenic inheritance. In the present study, the association 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, explaining 4.54–9.94% phenotypic variance, were identified for oil content (Table 3). Interestingly, the favorable alleles with relatively higher effects were comparatively abundant in the varieties from Northern China and the Yangtze River region, than the varieties from Southern China (Figs. 3b and 4c). Correspondingly, oil content of varieties showed a geographic difference clearly. The accessions from Northern China and the Yangtze River region had significantly higher oil content than the accessions from Southern China (Fig. 5). It seemed that oil content and its underlying loci may undergo selection during geographical differentiation in China. However, more experimental evidence, such as a multiple-ecological
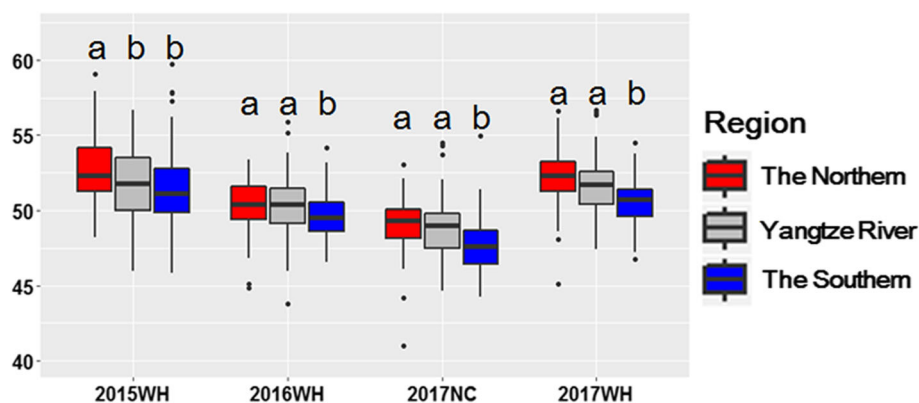

Fig. 5 Comparison of oil content (%) among peanut accessions from different geographic regions. The boxes with different letters were significantly different according to Tukey's Multiple Comparison Test ($P < 0.05$)

Liu et al. BMC Genetics   (2020) 21:60

Page 9 of 12

investigation of phenotype, was needed to verify the hypothesis.

The oil content showed an additive inheritance in crops [15, 39–41], 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 (Additional file 1: Table S1) based on the information from previous linkage maps and the physical location on genome [25, 42–44]. Compared with the previous results, chromosome A03, A04, A08, B06, B07, and B08 were also found to harbor QTLs for oil content [13, 14, 16]. 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 [16]. Another two markers (TC1A02 and AHGS0393), which were highly linked to QTLs for oil content in the early studies [13, 14], 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 –LogP value of 2.50E-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 ~ 0.34% - ~ 1.50% or ~ 0.61% - ~ 0.88% in four environmental trials. Combining favorable alleles for two loci, oil content could increase to ~ 1.11% - ~ 2.06% (Fig. 4 and Additional file 1: Table S4). 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. S5). It is suggested that the selection of favorable alleles of AGGS1014_2 and AHGS0798 has been underway in China during the breeding program.

Conclusions
This study provided a resource to improve our understanding the genetic basis of oil content in peanut and

to reveal a close relationship of the geographical region with population structure. Two associated markers (AGGS1014_2 and AHGS0798) in the present study were verified and can be deployed for use in GAB for enhancing oil content 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 association mapping panel (Additional file 1: Table S5). 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 F_8–F_{10} generations of RIL population were used for analysis of oil content, the F_{11} generation was utilized for generation of genotypic data.

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.

Sixteen accessions with abundance phenotypic variation in peanut panel were used to screen polymorphic markers in 4485 previously reported markers [45–53]. A total of 583 high-quality polymorphic markers were obtained and labeled with fluorescence dyes 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 [54]. The number of subgroups of this peanut panel was estimated using STRUCTURE software V2.1 based on the model-based Bayesian clustering method [55]. 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 50,000 followed by 10,000 iterations were conducted with the admixture and related frequency models. The optimal *K* value was determined by the posterior probability [LnP(D)] and ΔK [56].

Phylogenetic analysis was performed to construct a UPGMA tree based on Nei's (1972) genetic distance. Nei's (1972) genetic distance was calculated using PowerMarker [54] 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/).

SSR markers mapped on a dense genetic map were selected to estimate LD. The pairs of markers located on the same linkage group were treated as linked markers, otherwise as unlinked markers. The r^2 and *p* value was calculated with TASSEL 3.0 [57]. LD decay in the peanut panel with r^2 values were plotted against the genetic distance (cM) between markers.

Evaluation of oil content and phenotypic data analysis

The percentages of Oil and H_2O in seeds were measured using nuclear magnetic resonance (PQ001, Niumag, China). Matured seeds (~ 10 g) with less than 10% moisture content were analyzed for each of the three subsamples per entry. Oil content (%) was calculated based on dry-weight using the formula $\{[oil/(100 – H_2O)] \times 100\}$ [13].

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 $H^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{g \times e}^2/r + \sigma_\varepsilon^2/rn)$, where σ_g^2 is the genotypic variance, $\sigma_{g \times e}^2$ is the genotype × environment interaction variance, σ_ε^2 is the residual variance, r represents the number of environments and n represents the number of replications in each environment.

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 [57]. The population structure (Q) was obtained from model-based program STRUCTURE V2.1 [55]. The pairwise kinship matrix (K) was calculated using SPAGeDi software [58]. To estimate allelic effect, the phenotypic effect of last allele for an associated marker is set to zero and the other allele estimates are relatives to that. The mean of four-trial allelic effect at 12 association locus were used to construct a heatmap to view geographic distribution of allelic effect for associated markers. The software MeV was used to visualize the heatmap [59].

Supplementary information

> **Additional file 1: Table S1.** Statistic summary of 578 polymorphic markers. **Table S2.** Summary of allele frequency in 292 accessions. **Table S3.** Allelic effect of associated loci. **Table S4.** Comparison of oil content between two genotypes of three markers in ZI population. **Table S5.** The detailed information on 292 peanut accessions.
>
> **Additional file 2: Figure S1.** Linkage disequilibrium (LD) decay in 292 peanut accessions. **Figure S2.** Description of phenotypic values for 292 peanut accessions. **Figure S3.** Comparison of oil content (%) among peanut accessions released at different stages. **Figure S4.** Association study for oil content. **Figure S5.** Frequency and phenotypic effect of combined genotypes between AGGS1014_2 and AHGS0798 in the peanut panel.

Abbreviations
Coefficients: r^2; GAB: Genomics-assisted breeding; LD: Linkage disequilibrium; PCA: Principal component analysis; PIC: Polymorphic information content; PVE: Phenotypic variance explained; RIL: Recombinant inbred line; SSR: Simple sequence repeats

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 analyzed the data. NL wrote the manuscript. MP, BL, RV, and HJ revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The raw phenotype data and genotype data are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Liu *et al. BMC Genetics*   (2020) 21:60

Page 11 of 12

Author details
[1]Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan 430062, People's Republic of China. [2]International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), 502324, Hyderabad, India.

References
1. Pandey MK, Pandey AK, Kumar R, Nwosu CV, Guo BZ, Wright GC, et al. Translational genomics for achieving higher genetic gains in groundnut. Theor Appl Genet. 2020;133:1679–702.
2. Food and Agriculture Organization of the United Nations (FAOSTAT). https://www.fao.org/faost.at/en/#data. Accessed 23 Mar 2020.
3. Baring MR, Wilson JN, Burow MD, Simpson CE, Ayers JL, Cason JM. Variability of total oil content in peanut across the state of Texas. J Crop Improv. 2013;27:125–36.
4. Wilson J, Baring M, Burow M, Rooney W, Simpson C. Generation means analysis of oil concentration in peanut. J Crop Improv. 2013;27:85–95.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Baud S, Lepiniec L. Physiological and developmental regulation of seed oil production. Prog Lipid Res. 2010;49:235–49.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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. 2020;133:37–49.
17. 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.
18. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. Annu Rev Plant Biol. 2003;54:357–74.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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. Sci Rep. 2018;8:14500.
34. 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.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.

Liu *et al. BMC Genetics* (2020) 21:60

Page 12 of 12

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

45. Shirasawa K, Koilkonda P, Aoki K, Hirakawa 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.

46. Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli 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.

47. 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.

48. 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.

49. Proite K, Leal-Bertioli S, Bertioli 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.

50. 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.

51. 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.

52. 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 Genomics. 2012;13:90.

53. 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.

54. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics. 2005;21:2128–9.

55. 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.

56. 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.

57. 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.

58. 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.

59. Howe E, Holton K, Nair S, Schlauch D, Sinha R, Quackenbush J. MeV: MultiExperiment Viewer. In: Ochs M, Casagrande J, Davuluri R, editors. Biomedical informatics for Cancer research. Boston: Springer; 2010. p. 267–77.

Publisher's Note

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