Detection of a major QTL and development of KASP markers for seed weight by combining QTL-seq, QTL-mapping and RNA-seq in peanut

Zhihui Wang1,2 · Liying Yan1 · Yuning Chen1 · Xin Wang1 · Dongxin Huai1 · Yanping Kang1 · Huifang Jiang1 · Kede Liu2 · Yong Lei1 · Boshou Liao1

Abstract

Key message Combining QTL-seq, QTL-mapping and RNA-seq identified a major QTL and candidate genes, which contributed to the development of KASP markers and understanding of molecular mechanisms associated with seed weight in peanut.

Abstract Seed weight, as an important component of seed yield, is a significant target of peanut breeding. However, relatively little is known about the quantitative trait loci (QTLs) and candidate genes associated with seed weight in peanut. In this study, three major QTLs on chromosomes A05, B02, and B06 were determined by applying the QTL-seq approach in a recombinant inbred line (RIL) population. Based on conventional QTL-mapping, these three QTL regions were successfully narrowed down through newly developed single nucleotide polymorphism (SNP) and simple sequence repeat markers. Among these three QTL regions, *qSWB06.3* exhibited stable expression, contributing mainly to phenotypic variance across environments. Furthermore, differentially expressed genes (DEGs) were identified at the three seed developmental stages between the two parents of the RIL population. It was found that the DEGs were widely distributed in the ubiquitin–proteasome pathway, the serine/threonine-protein pathway, signal transduction of hormones and transcription factors. Notably, DEGs at the early stage were mostly involved in regulating cell division, whereas DEGs at the middle and late stages were primarily involved in cell expansion during seed development. The expression patterns of candidate genes related to seed weight in *qSWB06.3* were investigated using quantitative real-time PCR. In addition, the allelic diversity of *qSWB06.3* was investigated in peanut germplasm accessions. The marker Ah011475 has higher efficiency for discriminating accessions with different seed weights, and it would be useful as a diagnostic marker in marker-assisted breeding. This study provided insights into the genetic and molecular mechanisms of seed weight in peanut.

Introduction

Peanut (*Arachis hypogaea* L.) is an important oilseed crop grown in over 100 countries. It is mainly grown in Asia, Africa, and other developing countries and semi-arid tropical regions. It also provides high-quality edible oil and protein for humans. The global annual production increased significantly in recent years, reaching 45.95 million tons from 28.5 million hectares in 2018 (http://faostat.fao.org). It is crucial to improve worldwide peanut production to meet the nutritional demands of a growing global population. Seed weight is considered as a crucial seed trait and positively correlated with yield and oil and protein contents (Shirasawa et al. 2012b; Chen et al. 2016a, 2017). Large-seeded peanuts have been reflected as the subject of many conventional breeding initiatives. Thus, identifying quantitative trait loci (QTLs) controlling seed weight and
characterization of the underlying candidate genes would benefit peanut molecular breeding and yield improvement.

Seed weight is a typical quantitative trait regulated by many genes and significantly influenced by the environment (Li and Li 2016; Li et al. 2018). Many QTLs and genes related to seed weight have been discovered in many crops, such as rice (Daware et al. 2016; Hu et al. 2018), soybean (Han et al. 2012; Zhang et al. 2016; Karikari et al. 2019), wheat (Zhang et al. 2018), maize (Liu et al. 2014; Zhang et al. 2020), chickpea (Singh et al. 2016), and oilseed rape (Liu et al. 2015). In peanut, several QTLs linked to seed weight have been identified using simple sequence repeats (SSR) markers and conventional QTL-mapping in previous studies (Ravi et al. 2011; Fonceka et al. 2012; Huang et al. 2015; Chen et al. 2016a). These QTLs were typically distributed across different chromosomes, and several were main-effect QTLs. Genome-wide association analysis of seed weight found several loci based on genotyping data with SSR markers and diversity arrays technology (DArT) marker panel in 300 germplasms (Pandey et al. 2014). Gangurde et al. (2019) discovered 8 and 11 major QTLs regulating seed weight in 2 nested-association mapping (NAM) population. In addition, previous studies have conducted QTL analyses for pod- and seed-related traits in bi-parental populations (Chen et al. 2017; Fonceka et al. 2012; Huang et al. 2015; Luo et al. 2017a, b, 2018; Shirasawa et al. 2012b). Particularly, our previous study (Wang et al. 2018) and Zhang et al. (2019) reported several QTLs for seed-related traits based on specific length amplified fragment sequencing (SLAF) markers and high-density genetic map. Chavarro et al. (2020) reported 49 QTLs for pod and seed traits, and 7 major QTLs for pod area, pod weight, and seed weight demonstrated phenotypic variability explained greater than 20%. However, the regions covered by the QTL are still quite large in terms of physical distance. Additional study is necessary to narrow down the major QTLs, identify the annotated genes, and develop molecular markers in these regions.

Many genes coordinately regulated seed weight during seed development via signaling pathways involved in cell proliferation and cell expansion (Li and Li 2016; Li et al. 2018). Extensive cell division occurs during the early stage of seed development to increase cell number. As the seed develops, cell expansion plays a predominant role and increases cell size. Finally, seed weight was strongly influenced by the number of cells and their size in the various dimensions of mature seed. Many genes associated with seed weight and length have been identified in a variety of signaling pathways (Li and Li 2016; Li et al. 2018), including the ubiquitin–proteasome pathway, phytohormones, G-protein signaling, the HAIKU (IKU) pathway, the mitogen-activated protein kinase (MAPK) signaling pathway, and transcriptional regulatory factors. Grain Width 2 (GW2) and Seed Width on Chromosome 5 (GW5), both of the genes that function in the ubiquitin–proteasome pathway, have been shown to regulate cell division and seed size during grain development (Song et al. 2007; Chen et al. 2017). Grain Size 3 (GS3 gene), which plays a role in the G-protein signaling pathway, significantly promotes cell proliferation and increases grain size (Shirasawa et al. 2012b). Grain Size 5 (GS5 gene) positively regulates cell division and increases grain weight (Chen et al. 2016a). In Brassica napus, Auxin Response Factor 18 (ARF18) gene was found to increase the seed weight and silique length (Liu et al. 2015). Auxin Response Factor 2 (ARF2) regulates cell division and influences seed and organ size via auxin signaling (Hu et al. 2018). In addition, numerous transcriptional regulators, such as APETALA 2 (AP2), Basic helix–loop–helix (bHLH) and myeloblastosis (MYB) transcriptional factors have been demonstrated to regulate seed weight during seed development (Han et al. 2012; Heang and Sassa, 2012b; Liu et al. 2014; Karikari et al. 2019).

QTL-sequencing (QTL-seq) is an efficient approach to identify QTLs and candidate genes associated with traits. It integrated data from bulked segregant analysis (BSA), next-generation sequencing (NGS), and bioinformatics analysis. Initially, two bulked DNA pools are produced from progenies with extreme phenotypic values and genotyped using whole-genome re-sequencing. Then, the candidate regions or genes associated with traits could be found by comparing the distribution of index single nucleotide polymorphisms (SNPs) between two bulks (Takagi et al. 2013). QTL-seq is effective for rapidly mapping QTL and identifying candidate gene in many crops (Takagi et al. 2015; Huo et al. 2016; Singh et al. 2016; Shu et al. 2018; Ariket et al. 2019; Deokar et al. 2019). Remarkably, this technology was successfully used to map QTLs of disease resistance (Pandey et al. 2017; Clevenger et al. 2018; Luo et al. 2019a), shelling percentage (Luo et al. 2019b), and purple testa color (Zhao et al. 2020) in peanut. However, there has been no report on using the QTL-seq approach to determine QTLs responsible for seed weight in peanut.

After significant QTLs for seed weight are identified, it is important in determining the functional genes underlying candidate QTLs. At present, the integration of data from QTLs and RNA-sequencing (RNA-seq) is regarded as a promising strategy for identifying candidate genes. The combination of QTL-mapping and RNA-seq has been successfully used to identify major QTLs and candidate genes rapidly for capsacinoid biosynthesis (Park et al. 2019), heat tolerance (Wen et al. 2019), heading type (Gu et al. 2017) and salt tolerance (Lei et al. 2020). However, no investigation of seed weight using QTL-seq combined with RNA-seq in peanut has been conducted. In addition, previous studies demonstrated that many genes showed distinct expression patterns at different stages during seed development (Jones
and Vodkin, 2013; Chen et al. 2016b; Qu et al. 2016; Wan et al. 2017). Thus, using RNA-seq at various stages of seed development would enable the investigation of the dynamic gene expression patterns associated with seed weight in peanut.

In this study, we used an NGS-based QTL-seq technique to detect major QTLs associated with seed weight in peanut. Through traditional QTL-mapping, three identified major QTLs were narrowed down using newly developed SNP and SSR markers. Among these three QTL regions, qSWB06.3 exhibited stable expression across environments, contributing largely to phenotypic variance. Subsequently, we used RNA-seq to conduct differential expression analyses in the early, middle, and late stages of seed development. The candidate genes controlling seed weight in qSWB06.3 were predicted by integrating QTL-seq, RNA-seq, and quantitative real-time polymerase chain reaction (qRT-PCR) experiment. Finally, competitive Allele-specific PCR (KASP) markers in qSWB06.3 were successfully validated in diverse peanut varieties, suggesting that they could be deployed in marker-assisted breeding to enhance seed weight. This study provided insights into understanding the genetic and molecular mechanisms of seed weight in peanut.

Materials and methods

The phenotyping of seed weight in the RIL population

A mapping population of 242 recombinant inbred lines (RILs, F_8 generation) was constructed through the single seed descent method by the cross between ‘Zhonghua16’ with high seed weight and ‘sd-H1’ with low seed weight (Wang et al. 2018). The segregating population and parental genotypes were grown for three consecutive years (2016–18) at two geographical locations (Wuchang and Yangluo) of Hubei province, China. Randomized blocks with three replications were arranged, and 10–12 representative plants for each line were selected for phenotype investigation. Mature seeds were measured for 100 seed weight which was taken on an electrical scale. The average weight of 100-matured seeds was used for phenotypic characterization. The distribution of seed weight in the mapping population across six environments is shown in Supplementary Fig. S1.

The whole-genome re-sequencing of four bulks and QTL-seq analysis

According to the seed weight distribution in the RIL population, each of the 20 homozygous individuals with extreme phenotypes was chosen to construct 2 extreme bulks. LSB (low seed weight bulk) and HSB (high seed weight bulk) were constituted by pooling the same amount of DNA from 20 lines with extreme phenotypes together. The Illumina libraries were constructed following the extraction of high-quality DNA from two bulks and parents. The four libraries were used to generate pair-ended reads with 150 bp lengths using the Illumina HiSeq4000 platform (Illumina Inc., San Diego, CA, USA).

The high-quality reads were extracted with over 95% of bases with Phred quality scores larger than 30. The cultivar tetraploid peanut genome sequences of _Arachis hypogaea_ were downloaded as the reference genome (Zhuang et al. 2019). The BWA software (Li and Durbin 2009) was used to align clean reads to the reference genome, and only uniquely mapped reads were used for the following analysis. The alignment in the regions adjacent to Insertion/Deletion (InDel) was refined by the Coval software (Kosugi et al. 2013). Then, SNPs were called using the Samtools software (Li et al. 2009), and low-quality SNPs were filtered using the Coval scripts (Kosugi et al. 2013). Finally, the reference-based assembly for the parents was constructed by replacing the reference bases with alternative bases. Subsequently, SNPs from HSB and LSB were identified after sequence alignment and variation calling.

The SNP-index and ΔSNP-index values were calculated based on the well-documented QTL-seq approach described previously (Abe et al. 2012; Takagi et al. 2013). First, the SNPs were considered low-quality and filtered out when the SNP-index value was less than 0.3 or its read depth was less than 10 in both bulks. Then, the SNP-index and ΔSNP-index values were calculated in one megabase (Mb) sliding window with a ten kilobase (kb) increment. Finally, the statistical tests were conducted under the assumption of no QTL at the \( P < 0.05 \) level. These genomic regions with an SNP-index significantly deviated from 0.5 in both bulks were identified as candidate QTLs controlling seed weight.

Narrowing down major QTLs by traditional QTL-mapping

For three major QTLs on chromosomes A05, B02, and B06, novel SSR and SNP markers were developed for narrowing down the QTL intervals through the traditional QTL-mapping method. To develop KASP-SNP markers, we retrieved the upstream and downstream sequences of selected SNPs and used BLASTN to determine their specificity in genomic sequences (Semagn et al. 2014). The primers for KASP-SNP markers were designed and further validated in two parents. Subsequently, these KASP markers were applied to genotype the RIL population. SSR markers were developed based on the information from previous studies (Ferguson et al. 2004; Moretzsohn et al. 2005; Cuc et al. 2008; Shirasawa et al. 2012a; Huang et al. 2016; Zhou et al. 2016; Luo et al. 2017a). The PCR reactions were conducted according
to the method described in Chen et al. 2008. The 6% polyacrylamide gel and silver staining were used to display PCR products. The genetic map for candidate genomic regions was generated using JoinMap 4.0 software (Voorrips et al. 2006). Kosambi mapping function was used to estimate the genetic distance and LOD values (Kosambi 1944). QTL Cartographer 2.5 software (Wang et al. 2012) was used to identify QTLs using composite interval mapping function.

**RNA extraction, RNA-seq analysis, and qRT-PCR experiment**

For determining the early, middle and late stages of seed development, we measured the seed weight of the two parents at a series of time points during the whole seed development, including 10, 20, 30, 40, 50, 60, and 70 days after flowering (DAF). According to their growth curve of seed weight during seed development (Supplementary Fig. S2), 20, 40, and 60 days after flowering (DAF) in parents ‘Zhonghua16’ and ‘sd-H1’, were determined to represent the early, middle and late stages of seed development, respectively, which is also consistent with previous studies (Wan et al. 2016, 2017). The developing seeds were collected at three developmental stages in two parents and three replicates were set for each stage. Total RNA of each seed sample was extracted using RNAprep pure plant kit (DP441, TIANGEN, China). For ensuring consistency, the same extracted RNA samples were both used for RNA-seq and qRT-PCR in the following experiments. The libraries were sequenced on an Illumina HiSeq 4000 to produce 150 bp paired-end reads. The quality of RNA-seq reads was checked using Trimmomatic (Bolger et al. 2014), and the reads with low quality were filtered out. Then, using Hisat 2 (Kim et al. 2015), clean RNA-seq reads were aligned to the reference genome. The RSEM program was used to calculate gene expression values (Li and Dewey 2011). The DESeq2 package (Anders and Huber 2010) was used for identifying differentially expressed genes. Total RNA was extracted for the qRT-PCR experiment according to the protocol of RNAprep Pure Plant Kit (TIANGEN, China). The reverse transcription from RNA to cDNA was performed using the cDNA Synthesis Kit (TIANGEN, China). The Peanut Actin gene was used as the internal control, and relative gene expression was calculated using the 2−ΔΔCt method.

**Results**

**Extreme bulks for seed weight**

In our previous study, a recombinant inbred line (RIL) population was developed by crossing ‘Zhonghua16’ with ‘sd-H1’ (Wang et al. 2018). The phenotyping data were previously generated for the parents and mapping population across six environments (Supplementary Fig. S1). The 100-seed weight (SW) of the two parents was significantly different (P < 0.001) between ‘Zhonghua16’ (91.16 g ± 10.19) and ‘sd-H1’ (31.95 g ± 3.21) (Fig. 1a). Seed weight showed continuous variation in the RIL population and displayed a normal frequency distribution, with the trait values of the RILs ranging from 26.53 g to 104.69 g (Fig. 1a). The transgressive segregation of seed weight was observed in many individuals with an extremely low or high phenotype in the RIL population. Two extreme bulks for SW were prepared and subjected to the QTL-seq pipeline. Furthermore, 20 of each low and high individual with extremes for seed weight in the RIL population constituted LSB (low seed weight bulk) and HSB (high seed weight bulk) (Fig. 1b). The average values of SW in LSB and HSB were 34.98 g ± 3.89 and 78.82 g ± 12.19, respectively. The strategy and pipeline of combining QTL-seq, QTL-mapping, and RNA-seq for identifying major QTLs, developing diagnostic markers, and predicting candidate genes are shown in Fig. 1c.

**QTL-seq predicted candidate genomic regions controlling seed weight**

The two parents and two extreme bulks (LSB and HSB) were applied for NGS-based high-throughput whole-genome re-sequencing using Illumina Hiseq 4000 platform. After alignment of clean reads to the reference genome, uniquely mapped reads were used to calculate genome coverage and read depth along chromosomes. A total of 875.5 million and 841.6 million high-quality paired-end (PE) reads (150 bp in length) were generated for ‘Zhonghua16’ and ‘sd-H1’, which provided 34.4× and 27.4× average read depth, respectively (Table 1). For LSB and HSB, 1.07 billion and 1.26 billion PE reads were generated, achieving an average depth of 43.0× and 49.2×, respectively (Table 1). The sequence reads from parents and extreme bulks covered 84.4–86.1% of the reference genome.

After quality control and strict filtering, variant calling resulted in 1,684,862, 1,647,238, 1,835,651, 1,777,715 SNPs in ‘Zhonghua16’, ‘sd-H1’, LSB, and HSB, respectively (Table 1). The SNP densities were 0.64–0.71 SNP per kb for two parents and extreme bulks, and varied across different chromosomes in A and B subgenomes of peanut genome (Supplementary Table S1). The heterozygous SNPs and SNPs with low read depth were first filtered out. Then, those SNPs with SNP-index values less than 0.3 were further discarded in both bulks. Finally, 541,282 reliable and high-quality SNPs were identified for the following analyses. To confirm the authenticity of the identified SNPs, 10 SNPs were randomly selected for performing Sanger sequencing. The results showed that nine out of ten randomly selected
SNPs have been proved to be consistent (Supplementary Fig. S3).

Referring to the methods of Takagi et al. (2013), the ΔSNP-index values for the two extreme bulks were calculated and visualized using sliding window analysis along chromosomes (Fig. 2). If the ΔSNP-index value of a genomic region was significantly different from 0 at statistical confidence of $P < 0.01$, a major QTL harboring candidate genes in this genomic region could be considered. Following this principle of QTL-seq, a total of 17 QTLs were found to be associated with seed weight, which was distributed on chromosomes A01, A02, A05, A06, A07, B01, B02, B06, B08, and B10. Their significant signals suggested that these candidate regions were major QTLs associated with seed weight (Fig. 2 and Supplementary Table S2). The same orthologous candidate genomic regions were also identified in two diploid genomes (Supplementary Fig. S4), which confirmed the good collinearity between diploid and tetraploid genomes.

### Table 1 Sequencing of parental lines and bulks and mapping of sequence reads

|                     | Zhonghua16 | sd-H1 | LSB      | HSB      |
|---------------------|------------|-------|----------|----------|
| Raw data (Gb)       | 122.3      | 117.6 | 149.9    | 176.5    |
| High-quality reads  | 875,485,280| 841,560,206| 1,073,006,646| 1,263,247,608|
| Mapped reads        | 868,791,396| 831,141,660| 1,065,307,644| 1,253,763,058|
| Uniquely mapped reads| 601,928,685| 477,074,337| 758,275,502| 866,349,276|
| Genomic coverage    | 85.2%      | 84.4% | 85.9%    | 86.1%    |
| Average depth       | 34.4       | 27.4  | 43.0     | 49.2     |
| Identified SNPs     | 1,684,862  | 1,647,238| 1,835,651| 1,777,715|

Fig. 1 The frequency distribution of 100-seed weight (g) in RILs crossed by ‘Zhonghua16’ and ‘sd-H1’. a The LSB (low seed weight bulk) and HSB (high seed weight bulk) were constructed by selecting 20 individuals with extreme seed weight in RILs. The 100-seed weight varied from 34.98 g ± 3.89 and 78.82 g ± 12.19 for LSB and HSB, respectively. b The phenotype of the RILs was used for the development of two extreme bulks for seed weight. c The strategic pipeline was constructed by the combination of QTL-seq, QTL-mapping, and RNA-seq, which was used for identifying major QTLs, developing diagnostic markers, and predicting candidate genes for seed weight in this study.
In the tetraploid peanut genome, a total of 45,636 SNPs were found to be located in these candidate regions of the tetraploid genome. Among these SNPs, 28,936 were intergenic, 2866 intronic, 7452 5' untranslated region (UTR), 4,381 3'UTR. Notably, 1269 SNPs distributed in 839 genes caused affection of missense, stop gain, and stop lost. Among these candidate genomic regions, three regions exhibited high ΔSNP-index values of 0.65, 0.53, and 0.70 corresponding to 101.70−111.64 Mb on A05, 103.90−111.75 Mb on B02, and 0.30−50.22 Mb on B06, respectively. Especially for the candidate region on B06 (0.30−50.22 Mb), the average SNP-index values were 0.16 and 0.86 for two bulks LSB and HSB, respectively, indicating a strong signal and a powerful QTL located in this genomic region controlling seed weight. Therefore, we developed SNP and SSR markers in the above three candidate genomic regions to narrow down the localization interval through traditional QTL-mapping.

**Narrowing down candidate genomic regions of seed weight by traditional QTL-mapping**

To narrow down the location interval of candidate genomic regions, we performed traditional QTL analysis by further developing SNP and SSR markers in the above three candidate regions on A05, B02, and B06. A total of 55 SNPs within the potential candidate genes were selected for KASP-PCR, out of which 41 KASP markers were successfully developed (Supplementary Table S3) and could distinguish SNP alleles between ‘Zhonghua16’ and ‘sd-H1’ in the RIL population. In addition, 21 SSR markers demonstrating polymorphism between 2 parental genotypes were developed.

![Fig. 2](image-url)
(Supplementary Table S4). Ultimately, 41 KASP markers and 21 SSR markers were applied for genotyping 242 individuals from the RIL population.

The CIM mapping was used to identify major QTLs for seed weight in the RIL population. For the three candidate regions on A05, B02, and B06, nine QTLs were detected to explain 4.92–14.89% of the phenotypic variance across different environments (Fig. 3 and Supplementary Table S5). Among them, qSWA05.1 was stably detected in four environments, while qSWB06.3 was detected in six environments. The QTL qSWA05.1 was located in the 104.85–105.92 Mb region by the nearest flanking markers AD05A20650 and A010637. The QTL qSWB06.3 was located in the 14.21–17.65 Mb region by the nearest flanking markers A011476 and A011478 (Fig. 3). This showed that the application of traditional QTL-mapping successfully narrowed down the candidate regions of QTL-seq using newly developed markers. Since qSWB06.3 displayed stable expression with a large contribution to phenotypic variance across all environments, the candidate genes in the genomic region of qSWB06.3 were further determined in the following analysis by combining the data from RNA-seq.

**Differential expression analysis of candidate genes based on RNA-seq during seed development**

To investigate the candidate genes related to seed weight in peanut, seeds from ‘Zhonghua16’ and ‘sd-H1’ at three different developmental stages were collected for RNA-seq. RNA-seq was performed on seeds at 20, 40, and 60 days after flowering (DAF), corresponding to the early, middle, and late stages of seed development, respectively (Wan et al. 2017). Three independent biological replicates were set up for each developmental stage and subjected to RNA-seq. Each sample generated an average of 12 Gb of clean data, and the average ratio of uniquely mapped reads was 79% (Supplementary Table S7). There were 28,077–32,184 genes that were detected as expressed at different stages of seed development, and a total of 42,106 genes were cumulatively detected as expressed. Among them, 9294 genes and 7983 genes were expressed constitutively during the 3 developmental stages of ‘Zhonghua16’ and ‘sd-H1’, respectively (Fig. 4a).

![Fig. 3](image)

**Fig. 3** Three candidate regions on A05, B02, and B06 were narrowed down using newly developed SSR and SNP markers based on the traditional QTL-mapping method. The phenotype data from six environments were used to identify QTLs which were denoted by different colors (color figure online).
demonstrated stage-specific expression, whereas 212 up-regulated and 64 down-regulated genes showed differential expression at DAF 40 and DAF 60 (Fig. 4c and d). To verify the accuracy of the RNA-seq, we selected 15 DEGs for qRT-PCR. The expression patterns from RNA-seq and qRT-PCR were highly similar and correlated, suggesting that the expression quantification based on RNA-seq was reliable (Supplementary Fig. S5).

The DEGs were associated with cell division and cell expansion during seed development

The Gene Ontology (GO) enrichment analysis showed that DEGs at DAF 20 were enriched in many GO terms relating to cell division such as minichromosome maintenance (MCM) complex, microtubule binding, motor activity, movement, DNA replication initiation, DNA helicase activity, regulation of cell cycle, and mitotic nuclear division.
(Fig. 5a). Moreover, a considerable number of DEGs at DAF 20 were correlated with cell wall organization and regulation of protein serine/threonine kinase activity, protein kinase binding, and so on. The DEGs at DAF 40 and DAF 60 were preferentially enriched in metabolic pathways involved in cell expansion, including nitrogen metabolism, fatty acid elongation, carbon metabolism, and linoleic acid metabolism (Fig. 5a). A substantial number of DEGs involved in cell division were found to be differentially expressed at the early stage, whereas DEGs at the middle and late stages were mostly connected with the regulation of cell expansion. Previous studies have revealed that several metabolic pathways, including the cell division pathway, ubiquitin–proteasome pathway, and serine/threonine-protein pathway, are critical in regulating seed weight (Li and Li 2016; Li et al. 2018). Thus, DEGs in these pathways at three different developmental stages were examined. We identified a total of 27 DEGs in the ATP-binding microtubule motor family, 8 DEGs encoding microtubule protein, 7 DEGs encoding tubulin, and 8 DEGs encoding MCM complex (Fig. 5b). It was found that most of these DEGs involved in regulating cell division were significantly differentially expressed at the early stage of seed development (Fig. 5b). In the ubiquitin–proteasome pathway and the serine/threonine-protein pathway, DEGs encoding E3 ubiquitin-protein ligase, ubiquitin, ubiquitin-conjugating enzyme, serine carboxypeptidase, cyclin, and serine/threonine-protein enzyme demonstrated a complex regulatory pattern, and many up-regulated and down-regulated genes both appeared in different processes (Fig. 5c and d). This result is consistent with previous observations that some genes play a positive regulatory role in these pathways, while others play a negative regulatory role (Li and Li 2016; Li et al. 2018). In addition, previous studies have also shown that the signal transduction of hormones and transcription factors is vital in regulating seed weight (Li et al. 2018). We identified six small auxin-up RNA (SAUR) genes, seven indole-3-acetic acid (IAA) genes, nine auxin-responsive Gretchen Hagen3 (GH3) genes, seven auxin (AUX) genes, and eight auxin response factor (ARF) genes that were differentially expressed at the three different developing stages (Fig. 5e). Simultaneously, 26 bHLH genes and 6 MYB genes were found to differ in expression during all 3 developmental stages (Fig. 5f).

**The candidate genes in qSWB06.3 deduced from QTL-seq and RNA-seq**

The genomic region of qSWB06.3, spanning 2.07 Mb on chromosome B06, had 705 effective SNPs. A total of 211 annotated genes were located in qSWB06.3. Functional annotation analysis of the 705 SNPs found that 311 SNPs were intergenic, 29 were intronic and 334 in UTRs. Several candidate genes related to seed weight in qSWB06.3 were predicted by systematically investigating SNP variation, gene expression, and functional annotation. We further analyzed and validated the expression patterns of these candidate genes across different developmental stages using qRT-PCR (Fig. 6 and Supplementary Table S9).

Among these candidate genes, two genes (AH16G10100 and AH16G09300) showed a similar expression pattern and both were differentially expressed between the two parents across all three stages. Their expression level was highest at the early stage (DAF 20) but decreased at the middle and late stages (DAF 40 and DAF 60). According to a previous study (Kurepa et al. 2009), AH16G10100 is homologous to regulatory particle AAA-ATPase 2A (RPT2a), which regulates 26S proteasome particle AAA-ATPase and influences seed size in *Arabidopsis*, while AH16G09300 is homologous to Glycogen Synthase Kinase 3/SHAGGY-like kinase 2, which is a negative regulator of brassinosteroid signaling involved in GS2-mediated grain size control (Che et al. 2015). Both genes were differentially expressed throughout the entire seed development process, implying a potential role in regulating peanut seed weight in qSWB06.3.

Another two genes (AH16G09020 and AH16G08270) showed high expression and were differentially expressed only at the late stage (DAF 60) in ‘Zhonghua16’. AH16G09020 is homologous to the positive regulator of grain length 1 (PGL1) in rice, which is a positive regulator of grain length that acts by controlling cell elongation (Heang and Sassa, 2012a). AH16G08270 has been annotated to encode chaperone DnaJ-domain superfamily protein, but its exact function has not been explored. In addition, three candidate genes, AH16G08940, AH16G08240, and AH16G08220, were found that code for phytochrome A, squamosa promoter-binding-like protein 1 and transducin family protein, respectively. Their expression patterns were complex, and the differences between the two parents were not consistent at different stages. AH16G08940 was highly expressed in the middle and late stages but lowly expressed at the early stage in ‘Zhonghua16’, whereas AH16G08240 and AH16G08220 were specifically and highly expressed in the late stage and early stage in ‘Zhonghua16’.

The above expression patterns of candidate genes in qSWB06.3 showed that two genes were significantly more highly expressed in the parent ‘Zhonghua16’ at the three developmental stages, and two genes were highly expressed in the early stage and one gene was highly expressed in the late stage in the parent ‘Zhonghua16’. The majority of candidate genes in qSWB06.3 were significantly more expressed in the parent ‘Zhonghua16’ at all three stages or a specific stage, demonstrating the contribution of this QTL to seed weight. In addition, among these candidate genes, the genomic position of the AH16G10100 and AH16G09300 genes was most adjacent to the marker Ah011475, which was found to be effective at distinguishing germplasm
accessions with different seed weights in the following analysis. Since these two genes displayed significant expression differences across three seed developmental stages, they could be considered as candidate genes for functional studies. However, the true casual genes in qSWB06.3 need to be further identified, which typically requires further investigation with fine-mapping and functional studies for definitive identification.

The frequency distribution of qSWB06.3 alleles in peanut germplasm accessions

To investigate the allelic diversity of qSWB06.3 in peanut varieties, we detected the genotype of the KASP-SNP markers in qSWB06.3 in peanut accessions with high and low seed weights. These germplasm accessions were collected from the medium-term gene bank of the Oil Crops Research Institute germplasm center, as previously reported (Fan et al. 2020). Sixty peanut accessions, including 30 accessions with high seed weight (119.10 g ± 13.41, HSW group) and 30 accessions with low seed weight (39.95 g ± 5.45, LSW group), were selected from diverse lines. The qSWB06.3 was located at 49.01 cM (position of peak correlated region) and flanked by the three markers Ah011475 (46.0 cM), Ah011476 (48.0 cM), and Ah011478 (50.3 cM), as shown in Fig. 3. Therefore, these three flanking KASP markers of qSWB06.3 (Ah011475, Ah011476, and Ah011478) which were nearest to the position of peak correlated region, were used to detect qSWB06.3 among different varieties (Supplementary Table S6).

The three markers Ah011475, Ah011476, and Ah011478 amplified the CC, GG, and AA alleles, respectively, in the parent ‘Zhonghua16’ with high seed weight; the AA, TT, and TT alleles, respectively, were amplified in parent ‘sd-H1’ with low seed weight. After genotyping the germplasm accessions, we found that 20 out of the 30 total accessions in the LSW group had the AA genotype for marker Ah011475. These 20 accessions included 6 accessions with AA-TT-TT genotypes and 14 accessions with AA-GG-GG genotypes for Ah011475-Ah011476-Ah011478 (Fig. 7a). Another 10 accessions in the LSW group and all 30 accessions in the HSW group had the CC genotype for marker Ah011475 and CC-GG-GG genotypes for Ah011475-Ah011476-Ah011478 (Fig. 7b). In the HSW group, all accessions shared the same genotype as the parent ‘Zhonghua16’ for three markers. In the LSW group, 66% of accessions had the same genotype as the parent ‘sd-H1’ for marker Ah011475, whereas only 20% shared the same genotype for markers Ah011476-Ah011478 (Fig. 7). The results suggested that compared to markers Ah011476-Ah01147, marker Ah011475 has higher efficiency for discriminating between the HSW and LSW groups. It was deduced that marker Ah011475 is more adjacent to the truly casual gene in qSWB06.3, resulting in a stronger correlation between marker genotype and seed weight. Based on these results, Ah011475 would be useful as a diagnostic marker for selecting elite genomic segments of qSWB06.3 in marker-assisted breeding to enhance seed weight in peanut.

Discussion

Cultivated peanut is an allotetraploid (2n = 4x = 40) comprised of two subgenomes A and B, which is believed to be derived from two diploids, A. duranensis and A. ipaensis (Samoluk et al. 2015). The A and B subgenome sequences exhibit a significant degree of collinearity and sequence similarity (Bertioli et al. 2016, 2019; Chen et al. 2019; Zhuang et al. 2019), which complicates QTL-mapping studies since a considerable number of sequencing reads will be mapped to multiple homoeologous locations in A and B subgenomes (Chavarro et al. 2020). In this study, we successfully overcame this difficulty by adopting strict filtering standards, allowing for only uniquely mapped reads to be used for the QTL-seq analysis. This ensured the accuracy of the SNP identification in the QTL-seq analysis. In addition, a large sequencing quantity was used to obtain sufficient data for further analysis, which finally produced approximately 30× depth for each parent and 45× depth for each bulk. These re-sequencing data covered over 84% of the genomic region, ensuring the generation of sufficient SNPs at the entire genome level. This is especially beneficial for overcoming the challenge in which sufficient SNPs for QTL identification are difficult to obtain due to very low DNA polymorphisms in peanut (Jiang et al. 2010; Wang et al. 2011; Mukri et al. 2014). In previous studies, the low polymorphic rate for a range of markers of peanut was found, including 6.6% for RAPD (Subramanian et al. 2000), 3.6% for AFLP (He and Prakash 1997), 10.4% for EST-SSR (Liang et al. 2009), 14.5% for SSR (Zhao et al. 2012), and 7.6% for SNPs (Wang et al. 2018). In our study, over 1.5 million SNPs were initially identified in each sample, resulting in SNP densities of 0.64–0.71 SNPs per kb. The high-density SNPs allowed for the comparison of the delta SNP-index
in consecutive windows along the chromosomes, leading to the effective identification of the QTLs underlying seed weight. These results revealed that QTL-seq is a quick and efficient method that can be used to scan and identify major QTLs at a genome-wide scale in peanut.

Many studies have reported QTLs for peanut seed weight, and these QTLs may overlap in their physical location. The major QTL (qSWB06.3) of seed weight on B06 was found to overlap with QTLs related to seed weight in our previous study (Wang et al. 2018) using a SLAF high-density genetic map and traditional QTL analysis, as also found by Chavarro et al. (2020). The QTLs related to seed weight, pod weight, and pod area were co-localized on B06 in Chavarro et al. (2020). We further examined whether the three major QTLs identified in this study overlapped with earlier reported QTLs for seed weight in previous QTL-mapping studies (Chen et al. 2017; Huang et al. 2015; Luo et al. 2017b, 2018; Zhang et al. 2019). After comparison, the QTLs on A05 from 101.70 to 111.64 Mb overlapped with those reported in Luo et al. (2017b) and Luo et al. (2018). Huang et al. (2015) identified three QTLs on A08, B02, and B03 that explained 1.69–17.88% of the phenotypic variance for seed weight. Chen et al. (2017) found nine QTLs explaining 5.2–10.8% of the phenotypic variation for seed weight on A02, A07, B04, B06, and B08. No overlapping QTL was found in Huang et al. (2015) and Chen et al. (2017) after a comparison of their physical positions, which were determined by aligning marker sequences to the reference genome. However, one QTL of seed weight in close proximity to qSWB06.3 on chromosome B06 was found in Zhang et al. (2019), which explained over 29% of the phenotypic variation and was consistently detected in at least three environments. The overlapped QTLs on B06 in Zhang et al. (2019) were identified based on a RIL population crossed between ‘Huayu36’ and a germplasm line ‘6–13’, although we did not find the literature to confirm the overlap on the source of their parents and their pedigree between the two studies. However, it was deduced that the QTL on B06 related to seed weight in two studies may be derived from the same germplasms during their breeding history as they shared the same QTL in different genetic backgrounds. These findings revealed that the same QTL may be contained in different genetic
Seed weight is coordinately regulated by cell division and cell expansion during seed development (Li and Li 2016; Li et al. 2018). Cell division occurs predominantly during the early stages of seed development, increasing in cell number (Dante et al. 2014). Cell division slows gradually as the seed matures, and cell expansion occurs at the middle and late stages of seed development to enhance cell size (Mizukami 2001). In this study, we observed that the DEGs between the two parents during the early developmental stage were primarily concentrated in metabolic pathways involved in cell division, whereas the DEGs during the middle and late developmental stages were primarily concentrated in metabolic pathways involved in cell expansion. This is consistent with previous studies that focused on distinct developmental stages of cell division and cell expansion occurring during seed formation (Li and Li 2016; Li et al. 2018). In addition, we observed that the number of DEGs involved in cell division at the early stage was significantly greater than that involved in cell expansion at the middle and late stages of development. This is consistent with prior studies indicating that cell division has a greater effect on seed size and weight than cell expansion, even though both have an influence on seed weight (Breuninger and Lenhard 2010). Notably, a substantial number of DEGs was found from metabolic pathways previously reported in the regulation of seed weight, including the ubiquitin–proteasome pathway, the serine/threonine-protein pathway, and the signal transduction of hormones and transcription factors (Li and Li 2016; Li et al. 2018). These results indicated that seed weight is controlled by a variety of genes and pathways acting explicitly in different stages during seed development, which is a typical feature of quantitative and complex traits (Chavarro et al. 2020; Luo et al. 2017a, b; Ravi et al. 2011; Varshney et al. 2009).

By examining DEGs within the qSWB06.3 region, the expression patterns of some candidate genes associated with seed weight were analyzed. These candidate genes were differentially expressed between two parents across the three developmental stages or only in a specific developmental stage. This is consistent with previous studies whereby many genes regulating seed weight were expressed during a certain seed developmental stage (Li and Li 2016; Li et al. 2018). In qSWB06.3, three well-studied genes related to seed weight, namely RPT2a, Synthase Kinase 3, and PGL1, were highly expressed in the seeds of ‘Zhonghua16’ and were differentially expressed between parents. Among them, RPT2a and Synthase Kinase 3 were closest to a diagnostic marker that has been shown to be effective in distinguishing germplasm resources with different seed weights. Therefore, these
The allele distribution of qSWB06.3 in germplasm accessions demonstrated that the elite allele of qSWB06.3 was also present in other accessions with different genetic backgrounds. It was deduced that the genomic fragments carrying the elite allele of qSWB06.3 were introgressed into ‘Zhonghua 16’ and other varieties with high seed weight during human selection and breeding. In addition, it was observed that some accessions carrying the elite ‘Zhonghua 16’ allele of qSWB06.3 still exhibited a low seed weight. The most likely explanation for this phenomenon, according to the literature and earlier research (Assefa et al. 2019; Han et al. 2008; Li et al. 2020), is the epistatic effect of QTLs. Epistasis is defined as the influence of the genotype at one locus on the effect of a mutation at another locus (Doust et al. 2014). In these accessions caring the elite ‘Zhonghua16’ allele of qSWB06.3, another epistatic gene that inhibited the phenotypic effect of elite qSWB06.3 allele has existed in these genetic backgrounds. In previous studies, marker-assisted breeding has aided in the development of elite peanut cultivars with high oleic acid content (Chu et al. 2011; Janila et al. 2016), desirable fatty acid composition (Bera et al. 2018), and disease resistance (Pandey et al. 2012; Varshney et al. 2014). A combination of genotypic and phenotypic selection was previously found to be efficient in identifying recombinant lines with desired traits (Pandey et al. 2020; Varshney 2016). Therefore, the deployment of newly developed diagnostic markers for seed weight in this study could facilitate the screening of early generation populations as well as marker-assisted backcrossing breeding, leading to the accelerated development of improved cultivars.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04069-0.

Acknowledgements This work was supported by the National Key Research and Development Program of China (2018YFD1000900), the Key Research and Development Program of Hubei Province (2021BBA077), the Key Area Research and Development Program of Hubei Province (21326316D), the Central Public-interest Scientific Institution Basel Research Fund (No. Y2021CG05).

Author contribution statement ZW, YL, HJ, and BL designed the experiments. ZW analyzed the data and performed the bioinformatics analysis. ZW, YL, KL and BL wrote the manuscript. LY, YC, XW, DH, and YK performed the experiments. All the authors read and approved the manuscript.

Funding This research was funded by the National Key Research and Development Program of China, Grant no [2018YFD1000900], the Key Research and Development Program of Hubei Province, Grant no [2021BBA077], the Key Area Research and Development Program of Hebei Province, Grant no [21326316D], the Central Public-interest Scientific Institution Basel Research Fund, Grant no [Y2021CG05].

Availability of data and material The raw sequencing data of QTL-seq have been submitted in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA718683. The RNA-seq data in this study have been deposited in the NCBI SRA under BioProject accession number PRJNA718692.

Declarations

Conflict of interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

Ethical standards The authors state that all experiments in the study comply with ethical standards.

References

Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals genomically important loci in rice using MutMap. Nat Biotechnol 30:174–178

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106

Arikit S, Wanchana S, Khanthong S, Saensuk C, Thianhavon T, Vanavichit A, Toojinda T (2019) QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (Oryza sativa L.). Sci Rep 9:8328

Assefa T, Oyama PF, Brown AV, Kalberer SR, Kulkarni RS, Cannon SB (2019) Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. BMC Genom 20(1):527

Bera SK, Kamdar JH, Kasundra SV et al (2018) Improving oil quality by altering levels of fatty acids through marker-assisted selection of ahfad2 alleles in peanut (Arachis hypogaea L.). Euphytica 214:162

Bertioli DJ, Cannon SB, Froenicke L, Huang G, Farmer AD, Cannon EK et al (2016) The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. Nat Genet 48:438–446

Bertioli DJ, Jenkins K, Clevenger J, Dudchenko O, Gao D, Seijo G et al (2019) The genome sequence of segmental allotetraploid peanut Arachis hypogaea. Nat Genet 51:877–884

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Breuninger H, Lenhard M (2010) Control of tissue and organ growth in plants. Curr Top Dev Biol 91:185

Chavarro C, Chu Y, Holbrook C, Isleib T, Bertioli D, Hovav R et al (2020) Pod and seed trait QTL identification to assist breeding for peanut market preferences. G3 (bethesda) 10(7):2297–2315

Che R, Tong H, Shi B, Liu Y, Fang S, Liu D, Xiao Y, Hu B, Liu L, Wang H, Zhao M, Chu C (2015) Control of grain size and rice yield by GL2-mediated brassinosteroid responses. Nat Plants 2:15195

Chen J, Xiao-Hui HU, Miao HR, Cui FG, Shan-Lin YU (2008) Genome DNA extracted with CTAB method and its use for SSR and SRAF. J Peanut Sci 37(1):29–31

Chen W, Jiao Y, Cheng L, Huang L, Liao B, Tang M, Ren X, Zhou X, Chen Y, Jiang H (2016a) Quantitative trait locus analysis for
Li N, Li Y (2016) Signaling pathways of seed size control in plants. Curr Opin Plant Biol 33:23–32
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
Li N, Xu R, Duan P, Li Y (2018) Control of grain size in rice. Plant Reprod 31:237–251
Li M, Chen L, Zeng J, Razaqq MK, Xu X, Xu Y, Wang W, He J, Xing G, Gai J (2020) Identification of additive–epistatic QTLs conferring seed traits in soybean using recombinant inbred lines. Front Plant Sci 11:566056
Liang X, Chen X, Hong Y, Liu H, Zhou G, Li S et al (2009) Utility of EST-derived SSR in cultivated peanut (Arachis hypogaea L.) and Arachis wild species. BMC Plant Biol 9:35
Liu Y, Wang L, Sun C, Zhang Z, Zheng Y, Qiu F (2014) Genetic analysis and major QTL detection for maize kernel size and weight in multi-environments. Theor Appl Genet 127:1019–1037
Liu J, Hua W, Hu Z, Yang H, Zhang L, Li R, Deng L, Sun X, Wang X, Wang H (2015) Natural variation in ARF8 gene simultaneously affects seed weight and silique length in polyploid rapeseed. Proc Natl Acad Sci USA 112:E5123–5132
Luo H, Xu Z, Li Z, Li X, Lv J, Ren X, Huang L, Zhou X, Chen Y, Yu J, Chen W, Lei Y, Liao B, Jiang H (2017a) Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (Arachis hypogaea L.). Theor Appl Genet 130:1635–1648
Luo H, Ren X, Li Z, Xu Z, Li X, Huang L, Zhou X, Chen Y, Chen W, Lei Y, Liao B, Pandey MK, Varshney RK, Guo B, Jiang X, Liu F, Jiang H (2017b) Co-localization of major quantitative trait loci for pod size and weight to a 3.7 CM interval on chromosome A05 in cultivated peanut (Arachis hypogaea L.). BMC Genom 18:58
Luo H, Guo J, Ren X, Chen W, Huang L, Zhou X, Chen Y, Liu N, Xiong F, Lei Y, Liao B, Jiang H (2018) Chromosomes A07 and A05 associated with stable and major QTLs for pod weight and size in cultivated peanut (Arachis hypogaea L.). Theor Appl Genet 131:267–282
Luo H, Pandey MK, Khan AW, Wu B, Guo J, Ren X, Zhou X, Chen Y, Chen W, Huang L, Liu N, Lei Y, Liao B, Varshney RK, Jiang H (2019a) Next-generation sequencing identified genomic region and diagnostic markers for resistance to bacterial wilt on chromosome B02 in peanut (Arachis hypogaea L.). Plant Biotechnol J 17:2356–2369
Luo H, Pandey MK, Khan AW, Guo J, Wu B, Cai Y, Huang L, Zhou X, Chen Y, Chen W, Liu N, Lei Y, Liao B, Varshney RK, Jiang H (2019b) Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (Arachis hypogaea L.). Plant Biotechnol J 17:1248–1260
Mizukami Y (2001) A matter of size: developmental control of organ size in plants. Curr Opin Plant Biol 4:533–539
Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertolli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of Arachis (Fabaceae). Theor Appl Genet 111:1060–1071
Mukri G, Nadaf HL, Bhat RS, Gowda MVC, Upadhyaya HD, Sujay V (2014) Phenotypic and molecular dissection of ICRISAT mini core collection of peanut. Plant Breed 131:418–422
Pandey MK, Monyo E, Ozias-Akins P, Liang X, Guimaraes P, Nigam SN, Upadhyaya HD, Janila P, Zhang X, Guo B, Cook DR, Bertolli DJ, Michelmore R, Varshney RK (2012) Advances in Arachis genomics for peanut improvement. Biotechnol Adv 30(3):639–651
Pandey MK, Upadhyaya HD, Rathore A, Vadez V, Sheshshayee MS, Sriswathi M et al (2014) Genome-wide 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 9:e105228
for cultivated groundnut (*Arachis hypogaea* L.). Theor Appl Genet 118:729–739
Varshney RK, Pandey MK, Janila SN, Nigam P, Sudini H, Gowda MVC, Srivath M, Radhakrishnan T, Manohar SS, Nagesh P (2014) Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor Appl Genet 127(8):1771–1781
Voorrips R, Van D, Van Den Heuvel LPWJ, Ooijen J, Van JW (2006) JoinMap® 4.0: software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen
Wan L, Bei L, Pandey MK, Wu Y, Yong L, Yan L, Dai X, Jiang H, Zhang J, Guo W, Varshney RK, Liao BS (2016) Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. Front Plant Sci 7:1491
Wan L, Li B, Lei Y, Yan L, Ren X, Chen Y, Dai X, Jiang H, Zhang J, Guo W, Chen A, Liao B (2017) Mutant transcriptome sequencing provides insights into pod development in peanut (*Arachis hypogaea* L.). Front Plant Sci 8:1900
Wang ML, Sukumaran S, Barkley NA, Chen Z, Chen CY, Guo B, Pittman RN, Stalker HT, Holbrook CC, Pederson GA, Yu J (2011) Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. Theor Appl Genet 123:1307–1317
Wang S, Basten C, Zeng Z (2012) Windows QTL Cartographer v.25. Department of Statistics North Carolina State University, Raleigh
Wang Z, Huai D, Zhang Z, Cheng K, Kang Y, Wan L, Yan L, Jiang H, Lei Y, Liao B (2018) Development of a high-density genetic map based on specific length amplified fragment sequencing and its application in quantitative trait loci analysis for yield-related traits in cultivated peanut. Front Plant Sci 9:827
Wen J, Jiang F, Weng Y, Sun M, Shi X, Zhou Y, Yu L, Wu Z (2019) Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol 19:398
Zhang J, Song Q, Cregan PB, Jiang GL (2016) Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). Theor Appl Genet 127:119–130
Zhang Y, Li D, Zhang D, Zhao X, Cao X, Dong L, Liu J, Chen K, Zhang H, Gao C, Wang D (2018) Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. Plant J 94:857–866
Zhang S, Hu X, Miao H, Chu Y, Cui F, Yang W, Wang C, Shen Y, Xu T, Zhao L, Zhang J, Chen J (2019) QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (*Arachis hypogaea* L.). BMC Plant Biol 19:537
Zhang X, Guan Z, Wang L, Fu J, Zhang Y, Li Z et al (2020) Combined GWAS and QTL analysis for dissecting the genetic architecture of kernel test weight in maize. Mol Genet Genom 295:409–420
Zhao Y, Prakash CS, He G (2012) Characterization and compilation of polymorphic simple sequence repeat (SSR) markers of peanut from public database. BMC Res Notes 5:362
Zhao Y, Ma J, Li M, Deng L, Li G, Xia H, Zhao S, Hou L, Li P, Ma C, Yuan M, Ren L, Gu J, Guo B, Zhao C, Wang X (2020) Whole-genome resequencing-based QTL-seq identified AhTc1 gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. Plant Biotechnol J 18:99–105
Zhou X, Dong Y, Zhao J, Huang L, Ren X, Chen Y, Huang S, Liao B, Lei Y, Yan L, Jiang H (2016) Genomic survey sequencing for development and validation of single-locus SSR markers in peanut (*Arachis hypogaea* L.). BMC Genom 17:420
Zhuang W, Chen H, Yang M, Wang J, Pandey MK, Zhang C et al (2019) The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51:865–876

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