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 Wang (wangzhihui0229@126.com)
Oil Crops Research Institute Chinese Academy of Agricultural Sciences https://orcid.org/0000-0002-9967-4464

Liying Yan
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Yuning Chen
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Xin Wang
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Dongxin Huai
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Yanping Kang
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Huifang Jiang
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Yong Lei
Oil Crops Research Institute Chinese Academy of Agricultural Sciences https://orcid.org/0000-0002-3738-4197

Boshou Liao
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Research Article

Keywords: seed weight, QTLs, RNA-seq, candidate genes, peanut

Posted Date: June 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-531536/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on March 9th, 2022. See the published version at https://doi.org/10.1007/s00122-022-04069-0.

Abstract

Seed weight is a major target of peanut breeding as an important component of seed yield. However, relatively little is known about 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 NGS-based QTL-seq approach for a RIL population. These three QTL regions have been successfully narrowed down through newly developed SNP and SSR markers based on traditional QTL mapping. Among these three QTL regions, *qSWB06.3* exhibited stable expression with large contribution to phenotypic variance across all environments. Furthermore, RNA-seq were applied for early, middle and late stages of seed development, and differentially expression genes (DEGs) were identified in ubiquitin-proteasome pathway, serine/threonine protein pathway and signal transduction of hormones and transcription factors. Notably, DEGs at early stage were majorly related to regulating cell division, whereas DEGs at middle and late stages were mainly associated with cell expansion during seed development. Through integrating SNP variation, gene expression and functional annotation, candidate genes related to seed weight in *qSWB06.3* were predicted and distinct expression pattern of those genes were exhibited using qRT-PCR. In addition, KASP-markers in *qSWB06.3* were successfully validated in diverse peanut varieties and the alleles of parent Zhonghua16 in *qSWB06.3* was associated with high seed weight. This suggested that *qSWB06.3* was reliable and the markers in *qSWB06.3* could be deployed in marker-assisted breeding to enhance seed weight. This study provided insights into the understanding of genetic and molecular mechanisms of seed weight in peanut.

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.

Introduction

Peanut (*Arachis hypogaea* L.) is an important oilseed crop that is widely planted in more than 100 countries. It provides high-quality edible oil and protein for humans in the semi-arid tropics. The global annual production increased rapidly in recent years and amounted to be 45.95 million tons from 28.5 million ha area in 2018 (http://faostat.fao.org). It is crucial to increase the world-wide peanut production to fulfill the dietary demand of an increasing global population. Seed weight is considered as a crucial seed trait and generally positively correlated with yield as well as oil and protein contents (Shirasawa et al., 2012b; Chen et al., 2016a; Chen et al., 2017). Large-seeded peanut has always been preferred by consumers and pursued by several conventional breeding programs. Thus, identification of QTLs controlling seed weight and characterization of the underlying candidate genes will benefit molecular breeding and yield enhancement of peanut.

Seed weight is a typical quantitative trait regulated by many genes with significant environmental influences (Li and Li, 2016; Li et al., 2018). Many quantitative trait loci (QTLs) and genes related to seed weight were reported in 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), oilseed rape (Liu et al., 2015). In peanut, several QTLs related to seed weight have been reported based on SSR markers and traditional QTL mapping in previous studies (Ravi et al., 2011; Fonceka et al., 2012; Huang et al., 2015; Chen et al., 2016a). These QTLs were usually distributed on different chromosomes and some of them were main-effect QTLs. Based on genotyping data with SSRs and DArT arrays in 300 germplasms, some loci associated with seed weight were identified from GWAS analysis (Pandey et al., 2014). Gangurde et al. (2019) identified 8 and 11 major effect QTLs controlling seed weight in two nested-association mapping (NAM) population. For pod- and seed-related traits, 12 and 83 QTLs were reported in previous studies (Chen et al., 2017; Shirasawa et al., 2012b), respectively. In addition, Wang et al. (2018) and Zhang et al. (2019) reported several QTLs for seed-related traits based on SLAF-based high density genetic map (HDGM). However, the genetic mechanisms of seed weight in peanut still remain poorly understood.

Seed weight were coordinately regulated by many genes in signaling pathways related to cell proliferation and cell expansion during seed development (Li and Li, 2016; Li et al., 2018). During early stage of seed development, the extensive cell division occurs to increase cell number. As the seed develops, cell expansion starts to play a predominant role and increase cell size. Ultimately, seed weight were highly influenced by cell number and cell size in different dimensions of mature seed. Many genes associated with seed weight and size have been identified in several signaling pathways (Li and Li, 2016; Li et al., 2018), such as ubiquitin–proteasome pathway, phytohormones, G protein signaling, IKU pathway, MAPK signaling pathway and transcriptional regulatory factors. In previous studies, GW2 and GW5, functioning in the ubiquitin-proteasome pathway, both regulate cell division in grain development and control seed size (Song et al., 2007; Chen et al., 2017). GS3, which plays role in G-protein signaling pathway, significantly increase cell proliferation and lead to large grain size (Shirasawa et al., 2012b). GS5 positively regulate cell division and promote the large grain size and increase grain weight (Chen et al., 2016a). In *B. napus*, auxin-response factor 18 (ARF18) genes was found to enlarge seed weight and silique length (Liu et al., 2015). ARF2 in auxin signaling regulate cell division and affect seed and organ size (Hu et al., 2018). In addition, many transcriptional regulators, such as AP2, bHLH, MYB, were shown to regulate seed weight during seed development (Han et al., 2012; Heang and Sassa, 2012b; Liu et al., 2014; Karikari et al., 2019).

QTL-seq is a efficient approach applied for rapidly mapping QTLs of interested traits and identifying genomic regions and candidate genes. It integrated bulked segregant analysis (BSA), next generation sequencing (NGS) and bioinformatics analysis. Initially, two bulked DNA are generated from progenies with extreme phenotypic values and genotyped with whole genome resequencing. Then, the candidate regions or genes associated with target traits could be identified through comparing the distribution of index SNPs between two bulks (Takagi et al., 2013). QTL-seq have been proven successful for rapidly QTL mapping and candidate gene identification in many crops (Takagi et al., 2015; Huo et al., 2016;

Singh et al., 2016; Shu et al., 2018; Arikit et al., 2019; Deokar et al., 2019). Remarkably, this technology was also used in mapping peanut QTLs of disease resistance (Pandey et al., 2017; Clevenger et al., 2018; Luo et al., 2019a), shelling percentage (Luo et al., 2019b), purple testa colour (Zhao et al., 2020). However, so far, there has been no report on the use of QTL-seq approach for QTL localization of seed weight in peanut.

When major QTLs related to seed weight are obtained, it is essential to identify the functional genes underlying candidate QTLs. At present, the integration of data from QTLs and RNA-seq is considered to be a promising method to identify potential candidate genes. The strategy of QTL mapping combined with RNA-seq have been successfully used to rapidly identify major QTLs and candidate genes for capsaicinoid biosynthesis (Park et al., 2019), heat tolerance (Wen et al., 2019), heading type (Gu et al., 2017) and salt tolerance (Lei et al., 2020), but the application of QTL-seq combined with RNA-seq in the study of seed weight in peanut has not been conducted. In addition, previous studies showed that many genes showed distinct expression at different stages during seed development (Jones and Vodkin, 2013; Chen et al., 2016b; Qu et al., 2016; Wan et al., 2017). Therefore, RNA-seq applied for different stages during seed development would enable the investigation of the dynamic patterns of gene expression associated with seed weight in peanut.

In this study, we applied NGS-based QTL-seq approach to identify major QTLs related to seed weight in peanut. The genomic intervals of three major QTLs were further narrowed down by newly developed SNP and SSR markers through traditional QTL mapping. Among these three QTL regions, *qSWB06.3* exhibited stable expression with large contribution to phenotypic variance across all environments. Subsequently, the RNA-seq were further applied to perform differential expression analysis at early, middle and late stages of seed development. By integrating the information from QTL-seq, RNA-seq and qRT-PCR experiments, the candidate genes controlling seed weight in *qSWB06.3* were predicted. Finally, KASP-markers in *qSWB06.3* were successfully validated in diverse peanut varieties and could be deployed in marker-assisted breeding to enhance to enhance seed weight. This study provided insights into the understanding of genetic and molecular mechanisms of seed weight in peanut.

Materials And Methods

The phenotyping of seed weight in the RIL population

Through single seed decent method, a RIL mapping population of 242 lines (F_8 generation) was constructed by the cross ZH16 × sd-H1 (Wang et al., 2019). 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 RIL line were selected for phenotype investigation. Mature seeds were measured for 100 seed weight which was taken on an electrical scale. The average weight (g) of 100-matured seeds were used for phenotypic characterization. The distribution of seed weight in mapping population across six environments were shown in Supplementary Figure S1.

The whole genome resequencing of four bulks and QTL-seq analysis

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

The high-quality reads were extracted when it has over 95% of bases with phred quality scores larger than 30. The cultivar tetraploid peanut genome sequences of *A. hypogaea* were downloaded as the reference genome (Zhuang et al., 2019). Meanwhile, the two diploid progenitor genome sequences of *A. duranensis* and *A. ipaensis* were also concatenated as the reference genome (Bertioli et al., 2016). In most cases, the candidate genomic regions and gene ID in tetraploid and their diploids were both given in results. The two sets of genomic sequences can complete each other for overcoming the incompleteness of genome assembly and annotation, and also help to trace the origination of candidate genomic regions or genes from diploids to tetraploid. The BWA software (Li and Durbin, 2009) was used to align clean reads to reference genome and only uniquely mapped reads were used for following analysis. The Coval software (Kosugi et al., 2013) was applied to refine the alignment in the regions near InDel. Then, Samtools software (Li et al., 2009) was used to call SNPs and the Coval scripts was used to further filter low-quality SNPs (Kosugi et al., 2013). Finally, the reference-based assembly for the parents were constructed by replacing the reference bases with alternative bases. Subsequently, SNPs from HSB and LSB were identified after sequence alignment and variation calling.

SNP-index andΔSNP-index were estimated based on the well-documented QTL-seq approach as described in previous studies (Abe et al., 2012; Takagi et al., 2013). Firstly, the SNPs was considered to be 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 andΔSNP-index in 1 Mb sliding window with 10 kb increment were calculated. Finally, the statistical test analysis was conducted under the assumption of no QTL at $P < 0.05$ level. These genomic regions with SNP-index significantly deviated from 0.5 in both bulks were defined as candidate QTLs controlling seed weight.

Narrowing down major QTLs through traditional QTL mapping

For three major QTLs on chromosomes A05, B02 and B06, new SSR and SNP markers were developed for narrowing down the intervals through traditional QTL mapping method. For developing KASP-SNP markers, the up- and down-stream sequences of selected SNPs were extracted for checking their specificity in genomic sequences by BLASTN (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 were generated using JoinMap 4.0 software (Van Ooijen, 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

The developing seeds from early, middle and late stages were collected at 20, 40, and 60 days after flowering (DAF) in parents Zhonghua 16 and sd-H1. Three replicates were set for each stage. Total RNA of each seed sample was extracted using RNAprep pure plant kit (DP441, TIANG EN, China). The libraries were sequenced on a HiSeq 4000 (Illumina) to produce paired-end reads with the length of each 150 bp. The quality of RNA-seq reads were checked using Trimmomatic (Bolger et al., 2014) and the low quality of reads were filter out. Then, Hisat 2 was used to align clean RNA-seq reads to reference genome. The RSEM software (Li and Dewey, 2011) was used to calculate gene expression values. The DESeq2 package (Anders and Huber, 2010) was used for identify differential expressed genes. Total RNA was extracted for qRT-PCR experiment according to the protocol of RNAprep Pure Plant Kit (TIANGEN, China). The reverse transcription from RNA to cDNA was performed using cDNA Synthesis Kit (TIANGEN, China). Peanut *Actin* gene was used as the internal control and relative gene expression was calculated by the 2-ΔΔCt method.

Results

Extreme bulks for seed weight

A RIL population have been developed by crossing Zhonghua 16 and sd-H1 in our previous study (Wang et al., 2018). The phenotyping data were generated earlier for parents and mapping population across six environments (Supplementary Figure S1). The 100-seed weight (SW) of two parents were significantly different ($P < 0.001$) between Zhonghua 16 (91.16 g ± 10.19) and sd-H1 (31.95 g ± 3.21) (Figure 1a). Seed weight showed continuous variation in RIL population and display normal frequency distribution, with trait values of RILs ranging between 26.53 g and 104.69 g (Figure 1a). The transgressive segregation of seed weight were observed in many individuals with extreme low or high phenotype in RIL population. Two extreme bulks for SW were prepared and subjected to the QTL-seq pipeline. Twenty of each low and high individuals with extremes for seed weight in RIL population were used to constitute LSB (low seed weight bulk) and HSB (high seed weight bulk), respectively (Figure 1b). The average values of SW in the LSB and HSB were 34.98 g ± 3.89 and 78.82 g ± 12.19, respectively.

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 based on 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, respectively, 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 and achieved 43.0 × and 49.2 × average depth, respectively (Table 1). The sequence reads from parents and extreme bulks covered 84.4–86.1% of the reference genome.

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

	Zhonghua 16	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

After quality control and strict filtering, variant calling resulted in 1,684,862, 1,647,238, 1,835,651, 1,777,715 single nucleotide polymorphisms (SNPs) in Zhonghua 16, 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 the density of SNP was obviously higher in B subgenome, rather than A subgenome (Supplementary Table S1). The heterozygous SNPs and SNPs with low read depth were firstly filtered out. Then, those SNPs with SNP-index value less than 0.3 in both bulks were further discarded. Finally, 541,282 reliable and high-quality SNPs along chromosomes were identified for the following analysis (Supplementary Table 2). In order to confirm the authenticity of the identified SNPs, ten SNPs were randomly selected for performing Sanger sequencing. The results showed that 9 out of 10 randomly selected SNPs have been proved to be consistent (Supplementary Figure S2).

Referring to the methods of Takagi et al. (2013) and Lu et al. (2014), theΔSNP−index values for the two extreme bulks were calculated and showed by sliding window analysis along chromosomes (Figure 2). If ΔSNP−index value of a genomic region was significantly different from 0 at a statistical confidence of $P < 0.01$, a major QTL harboring candidate genes in this region could be considered. Following this principle of QTL-seq, the regions on chromosome A02 from 0.58 Mb to 0.65 Mb, chromosome A05 from 101.70 Mb to 111.64 Mb, chromosome B02 from 103.90 Mb to 111.75 Mb, chromosome B06 from 0.30 to 50.22 Mb exhibited significant signals (Figure 2), suggesting that these candidate regions were major QTLs associated with seed weight. The same orthologous candidate genomic regions were also identified in two diploid genomes (Supplementary Figure S3), confirmed the good collinearity between diploid and tetraploid genomes. Totally, 45,636 SNPs were found to be located in these candidate regions of tetraploid genome. Among these SNPs, 28,936 were intergenic, 2,866 intronic, 7,452 5’ UTR, 4,381 3’UTR. Importantly, 1,269 SNPs distributed in 839 genes caused affection of missense, stop gain, stop

lost. Among these candidate genomic regions from QTL-seq, three regions including 101.70–111.64 Mb on A05, 103.90–111.75 Mb on B02 and 0.30–50.22 Mb on B06 exhibited high ΔSNP-index value of 0.65, 0.53, and 0.70, respectively. Especially for candidate region on B06 (0.30–50.22 Mb), the average SNP-index were 0.16 and 0.86 for two bulks LSB and HSB, respectively, showing a strong signal and a powerful QTL controlling seed weight in the region. Therefore, we further developed SNP and SSR markers in above three candidate regions for narrowing 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, the traditional QTL analysis was performed by further developing SNP and SSR markers in above three candidate regions on A05, B02 and B06. Totally 55 SNPs within the potential candidate genes were selected for competitive allele-specific PCR (KASP). Among them, 41 KASP markers were successfully developed (Supplementary Table S2) and could distinguish SNP alleles from Zhonghua16 and sd-H1 in the RILs population. In addition, 21 SSR markers showing polymorphic between two parental genotypes were also developed (Supplementary Table S3). Finally, 41 KASP markers together with 21 SSR markers were applied for genotyping 242 individuals of the RIL population.

The CIM mapping were used to identify major QTLs for seed weight in the RIL population. For the above 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 (Figure 3 and Supplementary Table S4). Among them, *qSWA05.1* was stably detected in four environments and *qSWB06.3* was detected in six environments. The *qSWA05.1* was located in the region of 104.85-105.92 Mb by the nearest flanking markers AD05A20650 and A010637. The *qSWB06.3* was located in the region of 14.21-17.65 Mb by the nearest flanking markers A011476 and A011478 (Figure 3). It showed that the application of traditional QTL mapping successfully narrowed down the candidate regions of QTL-seq through newly developed markers. Because the *qSWB06.3* exhibited stable expression with large contribution to phenotypic variance across all environments, the candidate genes in the genomic region of *qSWB06.3* would be further deduced by combining the data from RNA-seq in the following analysis.

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

In order to investigate 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) which corresponds to 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 subject to RNA-seq. The sequencing generated ~12 Gb of clean data on average in each sample, and the average ratio of uniquely mapped reads was 79% (Supplementary Table S6). 28,077–32,184 genes were detected as expressed in different stages during seed development and totally 42,106 genes were cumulatively to be detected as expressed. Among them, 9,294 genes and 7,983

genes were constitutively expressed in three developmental stages of Zhonghua16 and sd-H1, respectively (Figure 4a).

We identified differentially expressed genes (DEGs) between Zhonghua16 and sd-H1 at each developmental stage as well as DEGs between different developmental stages in each parent (Figure 4b). A total of 4,539, 2,489 and 3,966 DEGs were identified at DAF 20, DAF 40, and DAF 60 between Zhonghua 16 and sd-H1, respectively. We identified 3,422, 1,712, 2,502 up-regulated genes and 1,117, 777, 1,464 down-regulated genes at DAF 20, DAF 40 and DAF 60 between two parents, respectively. Particularly, there are clearly more DEGs in early stage (4,539), rather than that in middle (2,489) and late stages (3,966). Meanwhile, the number of DEGs (6,443) between DAF 20 and DAF 40 in sd-H1 was significantly more than that (2,243) between DAF 40 and DAF 60. The results suggested that a large number of genes showed differential expression between two parents at the initial stage of seed development, which may ultimately determined the seed weight. Further comparison showed that 1,457 up-regulated and 653 down-regulated DEGs at DAF 20 exhibited stage-specific, whereas 212 up-regulated and 64 down-regulated genes showed differential expression at DAF 40 and DAF 60 (Figures 4c and 4d). In order to verify the accuracy of RNA-seq, we selected 15 DE genes for qRT-PCR experiment. The expression patterns from RNA-seq and qRT-PCR were highly similar and correlated, suggested good reliability of expression quantification based on RNA-seq (Supplementary Figure S4).

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

The GO enrichment analysis showed that DEGs at DAF 20 were enriched in many GO terms relating to cell division such as MCM complex, microtubule binding, motor activity, movement, DNA replication initiation, DNA helicase activity, regulation of cell cycle and mitotic nuclear division (Figure 5a). Moreover, an additional very considerable number of DEGs at DAF 20 were correlated with cell wall organization and regulation of protein serine/threonine kinase activity, protein kinase binding *etc*. For DEGs at DAF 40 and DAF 60, they were preferentially enriched in metabolisms associated with regulation of cell expansion such as nitrogen metabolism, fatty acid elongation, carbon metabolism, linoleic acid metabolism (Figure 5a). The results showed that a large number of genes involved in cell division differentially expressed at early stage, nevertheless, DEGs at middle and late stages were mainly associated with the regulation of cell expansion.

Previous studies have also shown that several metabolic pathways such as cell division pathway, ubiquitin-proteasome pathway and serine/threonine protein pathway is important in regulating seed weight (Li and Li, 2016; Li et al., 2018). Thus, DEGs in these pathways at three different developing stages were studied. Totally, we identified 27 DEGs in ATP-binding microtubele motor family, 8 DEGs encoding microtubule protein, 7 DEGs encoding tublin, 8 DEGs encoding MCM complex (Figure 5b). It was found that most of these DEGs involved in regulating cell division were significantly differentially expressed at early stage of developing seeds (Figure 5b). In ubiquitin-proteasome pathway and ubiquitin-proteasome pathway and 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 (Figures 5c and 5d). This is consistent with previous reports that some genes play positively regulated while other play negatively regulated genes in these pathways (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 important in regulating seed weight (Li et al., 2018). We identified 6 SAUR genes, 7 IAA genes, 9 GH3 genes, 7 AUX genes, 8 ARF genes were differentially expressed at the three different developing stages (Figure 5e). Simultaneously, 26 bHLH genes and 6 MYB genes were found to have differences in expression during all three developmental stages (Figure 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*. Function annotation analysis of the 705 SNPs found that 311 SNPs were intergenic, 29 intronic, 334 in UTR. Through systematically investigating SNP variation, gene expression and functional annotation, several candidate genes related to seed weight in *qSWB06.3* were predicted. For these candidate genes, we further analyzed and validated their expression pattern across different developmental stages (DAF 20, DAF 40 and DAF 60) using qRT-PCR (Figure 6).

Among these candidate genes, two genes (AH16G10100 and AH16G09300) showed similar expression pattern and both differentially expressed between two parents across all three stages. Their expression level were highest at early stage (DAF 20) but decreased at middle and late stages (DAF 40 and DAF 60). According to previous study (Kurepa et al., 2009), AH16G10100 was homologous to RPT2A which was known to regulate 26S proteasome particle AAA-ATPase and influence seed size in *Arabidopsis*. While, AH16G09300 was homologous to GSK3/SHAGGY-like kinase 2 which is a negative regulator of BR signaling involved in GS2-mediated grain size control (Che et al., 2015). Both two genes differentially expressed throughout the whole seed development, suggesting that their potential role in regulating peanut seed weight in *qSWB06.3*.

Another two genes (AH16G09020 and AH16G08270) showed high expression and differentially expressed only at late stage (DAF 60) in Zhonghua16. AH16G09020 was homologous to PGL1 in rice which is a positive regulator of grain length by controlling cell elongation (Heang and Sassa, 2012a). AH16G08270 have been annotated to encode chaperone DnaJ-domain superfamily protein, but its specific function was not studied. In addition, three candidate genes AH16G08940, AH16G08240 and AH16G08220 were identified which code for phytochrome A, squamosa promoter-binding-like protein 1 and transducin family protein. Their expression patterns were complex, and the differences between two parents were not consistent at different stages. AH16G08940 were highly expressed in middle and late stages but lowly expressed at early stage in Zhonghua16. AH16G08240 and AH16G08220 were specifically and highly expressed in late stage and early stage in Zhonghua16. However, these candidate genes related to seed weight will need to be confirmed by further functional genomics studies.

The validation of KASP markers of *qSWB06.3* in natural peanut varieties

To validate the effectiveness of markers in *qSWB06.3* for deploying in marker-assisted breeding, the genotype of KASP-SNP markers in *qSWB06.3* were detected in diverse peanut varieties. 60 natural peanut varieties including 30 varieties with high seed weight (119.10 g ± 13.41, HSW group) and 30 varieties with low seed weight (39.95 g ± 5.45, LSW group) were selected from diverse lines. Three KASP markers, Ah011475, Ah011476 and Ah011478 in *qSWB06.3*, were used to detect genotypes among different varieties (Supplementary Table S5). They amplified the AA, TT and TT alleles in the parent Zhonghua16 with high seed weight while the CC, GG and GG in parent sd-H1 with low seed weight, respectively. For marker Ah011475, 20 accessions from LSW group showed CC alleles (Figure 7a). For markers Ah011476 and Ah011478, six accessions from LSW group showed TT and TT alleles, respectively (Figures 7b and 7c). The above natural peanut varieties which have same alleles from parent sdH1 were all came from LSW group, suggesting a correlation between the allele of parent sdH1 in *qSWB06.3* and the low seed weight. The natural peanut varieties with AA-TT-TT alleles showed significantly lower seed weight, compared with those with CC-GG-GG alleles (Figure 7d). The results suggested that these natural peanut varieties from HSW group have acquired the same alleles of Zhonghua16 in *qSWB06.3* during peanut breeding and improvement. The above results indicated that markers, Ah011475, Ah011476 and Ah011478, could be used as diagnostic markers for selecting breeding lines with high or low seed weight.

Discussion

In this study, we successfully identified the candidate major QTLs associated with seed weight using NGS-based QTL-seq approach. Among these candidate QTLs, three of them were further validated and narrowed down by SNP and SSR marker-based traditional QTL mapping. These results indicated that QTL-seq is a quick and efficient method, which can be used to scan and identify major QTLs at a genome-wide scale in peanut. In particular, it can overcome the difficulty that there are usually not enough markers for genetic mapping construction due to the low levels of marker polymorphism in peanut varieties (Jiang et al., 2010; Wang et al., 2011; Mukri et al., 2014). The underlying reason for the success should be attributed to the combination of NGS-based SNP discovery and bulked sample segregant analysis as well as the completion of peanut whole-genome sequencing projects (Bertioli et al., 2019; Chen et al., 2019; Zhuang et al., 2019). In addition, because A and B subgenomes in tetraploid peanut have high level of sequence similarity, a considerable number of reads could map to multiple homeologous locations in A and B subgenomes. In order to ensure the accuracy of SNP identification, we adopted strict filtering standards and only the uniquely mapped reads were used for the QTL-seq analysis. For ensuring enough data for subsequent analysis, a very high sequencing amount is used in this study which finally provided approximately 30 × and 45 × for two parents and bulks, respectively. Moreover, the panel of 242 RILs were phenotyped in multiple environments and years, which also guaranteed the reliability of identified QTLs.

Seed weight is an important determinant of seed yield and is controlled by multiple genes with different effects. Many QTLs related to seed weight have been reported in crops (Li and Li, 2016; Li et al., 2018), however, there are few studies on revealing the underlying genetic mechanism of seed weight in peanut. In the present study, a total of 17 QTLs associated with seed weight were found to be distributed on A01,

A02, A05, A07, B02, B06, B08 and B10. Among them, three QTLs on A02, A05 and B02 showed largestΔSNP-index value. By comparison with previously reported QTLs, the QTLs on A05 from 101.70 Mb to 111.64 Mb is overlapped with the QTLs reported in Luo et al., 2017b; Luo et al., 2018. Other QTLs were not found to overlapped with previously reported QTLs, indicating that they may be specific to peanut varieties used in this study. Subsequently, we narrowed down above three QTLs by developing new SNP and SSR markers through traditional QTL method. Among these newly detected QTLs, *qSWB06.3* located in the region of 104.85-105.92 Mb on A06 explained largest phenotypic variation and showed stable expression across different environments. The investigation of *qSWB06.3* genotype in natural peanut varieties also confirmed the correlation between genotype and phenotype. The genotype of *qSWB06.3* in sdH1 only appeared in peanut varieties with low seed weight, suggesting this genotype has a negative effect on the improvement of seed weight. On the other hand, it also showed that elite alleles have been introgressed into Zhonghua 16 and other cultivars with high seed weight in the process of modern breeding.

Seed weight is a complex trait which is regulated by many signaling pathways and genes. In this study, we investigated DEGs across different stages of seed development particularly distributed in ubiquitin-proteasome pathway, serine/threonine protein pathway and the signal transduction of hormones and transcription factors. The results showed that several genes encoding E3 ubiquitin-protein ligase, ubiquitin, ubiquitin-conjugating enzyme in ubiquitin-proteasome pathway showed obviously differential expression during seed development. As these genes in ubiquitin-proteasome pathway were reported to regulate cell division and increase the cell number in seed weight control (Li and Li, 2016; Li et al., 2018), therefore these DEGs in this pathway could be considered as important candidates regulating seed weight in peanut. Furthermore, a complex regulatory pattern including up-regulated and down-regulated genes were found for genes encoding cyclin and serine/threonine protein enzymes in serine/threonine protein pathway. These genes in serine/threonine protein pathway have been known to regulate cell division during seed development (Li and Li, 2016; Li et al., 2018). In addition, the DEGs belonging to TFs and hormone signaling pathways have been identified, such as SAUR, IAA, GH3, AUX, ARF, bHLH and MYB gene copies. In previous studies, signal transduction of hormones and transcription factors have been revealed to play an important role in regulating seed weight. Notably, DEGs in early stage were enriched in GO categories related to cell division and distributed in ATP-binding microtubele motor family, microtubule protein, tublin and MCM complex *etc*. Whereas, the DEGs in middle and late stages were concentrated on metabolisms related to regulating cell expansion such as nitrogen metabolism, fatty acid elongation, carbon metabolism, linoleic acid metabolism. These results indicated that seed weight is controlled by a series of genes and pathways specifically acted in different stages during seed development, which is a typical feature of quantitative and complex traits.

In this study, three major candidate regions associated with seed weight have been detected on chromosomes A02, B02 and B06 through QTL-sEq. Then, we further developed SNP and SSR markers in above three candidate regions for narrowing down the localization interval through traditional QTL mapping. Among these three QTL regions, *qSWB06.3* exhibited stable expression with large contribution to phenotypic variance across all environments. In the following analysis, we analyzed the genotypes of

qSWB06.3 in sixty natural peanut varieties. The results showed that the group of natural peanut varieties with high seed weight have the same alleles of parent Zhonghua16 in *qSWB06.3*, indicating that this QTL was reliable to be responsible for controlling seed weight. Simultaneously, these KASP-SNP markers in *qSWB06.3* will be useful as diagnostic markers for marker-assisted breeding to enhance seed weight in peanut. Finally, through systematically investigating SNP variation, gene expression and functional annotation, several candidate genes related to seed weight in *qSWB06.3* were predicted. Their expression in different stages during seed development were examined using qRT-PCR, and distinct pattern of expression differences were shown for different candidate genes. Subsequently, fining mapping and candidate gene identification of seed weight in *qSWB06.3* could be further explored through construction of backcross inbred population. In sum, this study provided insights into the understanding of genetic and molecular mechanisms of seed weight in peanut.

Declarations

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFD1000900).

Author contribution statement

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

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 Sequence Read Archive (SRA) under BioProject accession number PRJNA718692.

Conflict interests On behalf of all 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 the ethical standards.

References

1. 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 agronomically important loci in rice using MutMap. Nat. Biotechnol. 30:174-178.
2. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol. 11:R106.

3. Arikit S, Wanchana S, Khanthong S, Saensuk C, Thianthavon 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.
4. 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.
5. Bertioli DJ, Jenkins J, 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.
6. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.
7. 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.
8. Chen J, Xiao-Hui HU, Miao HR, Cui FG, and Shan-Lin, Y.U. (2008) Genome DNA extracted with CTAB method and its use for SSR and SRAP. J. Peanut Sci. 37(1): 29-31.
9. 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 pod- and kernel-related traits in the cultivated peanut (*Arachis hypogaea L.*). BMC Genet. 17:25.
10. Chen X, Yang Q, Li H, Li H, Hong Y, Pan L et al (2016b) Transcriptome-wide sequencing provides insights into geocarpy in peanut (*Arachis hypogaea L.*). Plant Biotechnol. J. 14:1215-1224.
11. Chen X, Lu Q, Liu H, Zhang J, Hong Y, Lan H et al (2019) Sequencing of Cultivated Peanut, *Arachis hypogaea*, Yields insights into genome evolution and oil improvement. Mol. Plant 12:920-934.
12. Chen Y, Ren X, Zheng Y, Zhou X, Huang L, Yan L et al (2017) Genetic mapping of yield traits using RIL population derived from Fuchuan Dahuasheng and ICG6375 of peanut (*Arachis hypogaea L.*). Mol. Breeding 37:17.
13. Clevenger J, Chu Y, Chavarro C, Botton S, Culbreath A, Isleib TG, Holbrook CC, Ozias-Akins P (2018) Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker-assisted selection. Front. Plant Sci. 9:83.
14. Cuc LM, Mace ES, Crouch JH, Quang VD, Long TD, Varshney RK (2008) Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol. 8:55.
15. Daware A, Das S, Srivastava R, Badoni S, Singh AK, Agarwal P, Parida SK, Tyagi AK (2016) An efficient strategy combining SSR markers- and advanced QTL-seq-driven QTL mapping unravels candidate genes regulating grain weight in rice. Front. Plant Sci. 7:1535.
16. Deokar A, Sagi M, Daba K, Tar'an B (2019) QTL sequencing strategy to map genomic regions associated with resistance to ascochyta blight in chickpea. Plant Biotechnol. J. 17:275-288.
17. Ferguson ME, Burow MD, Schulze SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*A. hypogaea L.*). Theor. Appl. Genet. 108:1064-1070.

18. Fonceka D, Tossim HA, Rivallan R, Vignes H, Faye I, Ndoye O, Moretzsohn MC, Bertioli DJ, Glaszmann JC, Courtois B, Rami JF (2012) Fostered and left behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. BMC Plant Biol. 12:26.
19. Gangurde SS, Wang H, Yaduru S, Pandey MK, Fountain JC, Chu Y et al (2019) Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). Plant Biotechnol. J. 18:1457-1471.
20. Gu A, Meng C, Chen Y, Wei L, Dong H, Lu Y, Wang Y, Chen X, Zhao J, Shen S (2017) Coupling Seq-BSA and RNA-Seq analyses reveal the molecular pathway and genes associated with heading type in Chinese cabbage. Front. Genet. 8:176.
21. Han Y, Li D, Zhu D, Li H, Li X, Teng W, Li W (2012) QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. Theor. Appl. Genet. 125:671-683.
22. Heang D, Sassa H (2012a) Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. PloS One 7:e31325.
23. Heang D, Sassa H (2012b) An atypical bHLH protein encoded by POSITIVE REGULATOR OF GRAIN LENGTH 2 is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein APG. Breeding Sci. 62:133-141.
24. Hu Z, Lu SJ, Wang MJ, He H, Sun L, Wang H et al (2018) A Novel QTL qTGW3 Encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that Interacts with OsARF4 to Negatively Regulate Grain Size and Weight in Rice. Mol. Plant 11:736-749.
25. Huang L, He H, Chen W, Ren X, Chen Y, Zhou X, Xia Y, Wang X, Jiang X, Liao B, Jiang H (2015) Quantitative trait locus analysis of agronomic and quality-related traits in cultivated peanut (*Arachis hypogaea L.*). Theor. Appl. Genet. 128:1103-1115.
26. Huang L, Wu B, Zhao J, Li H, Chen W, Zheng Y, Ren X, Chen Y, Zhou X, Lei Y, Liao B, Jiang H (2016) Characterization and Transferable Utility of Microsatellite Markers in the Wild and Cultivated Arachis Species. PloS One 11:e0156633.
27. Huo H, Henry IM, Coppoolse ER, Verhoef-Post M, Schut JW, de Rooij H, Vogelaar A, Joosen RV, Woudenberg L, Comai L, Bradford KJ (2016) Rapid identification of lettuce seed germination mutants by bulked segregant analysis and whole genome sequencing. Plant J. 88:345-360.
28. Jiang HF, Ren XP, Zhang XJ, Huang JQ, Lei Y, Yan LY, Liao BS, Upadhyaya HD (2010) Comparison of genetic diversity between peanut mini core collections from China and ICRISAT by SSR markers. Acta Agron. Sin. 36:1084-1091.
29. Jones SI, Vodkin LO (2013) Using RNA-Seq to profile soybean seed development from fertilization to maturity. PloS One 8:e59270.
30. Karikari B, Chen S, Xiao Y, Chang F, Zhou Y, Kong J, Bhat JA, Zhao T (2019) Utilization of interspecific high-density genetic map of RIL population for the QTL detection and candidate gene mining for 100-seed weight in soybean. Front. Plant Sci. 10:1001.
31. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12(4):357.

32. Kosambi DD (1944) The estimation of map distance from recombination values. Ann. Eugen 12(1):172-175.
33. Kosugi S, Natsume S, Yoshida K, MacLean D, Cano L, Kamoun S, Terauchi R (2013) Coval: improving alignment quality and variant calling accuracy for next-generation sequencing data. PloS One 8:e75402.
34. Kurepa J, Wang S, Li Y, Zaitlin D, Pierce AJ, Smalle JA (2009) Loss of 26S proteasome function leads to increased cell size and decreased cell number in *Arabidopsis* shoot organs. Plant Physiol. 150:178-189.
35. Lei L, Zheng H, Bi Y, Yang L, Liu H, Wang J, Sun J, Zhao H, Li X, Li J, Lai Y, Zou D (2020) Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa L.*) using QTL-Seq and RNA-Seq. Rice 13, 55.
36. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.
37. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760.
38. 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.
39. Li N, Li Y (2016) Signaling pathways of seed size control in plants. Curr. Opin. Plant Biol. 33:23-32.
40. Li N, Xu R, Duan P, Li Y (2018) Control of grain size in rice. Plant Reproduction 31:237-251.
41. Liu J, Hua W, Hu Z, Yang H, Zhang L, Li R, Deng L, Sun X, Wang X, Wang, H (2015) Natural variation in ARF18 gene simultaneously affects seed weight and silique length in polyploid rapeseed. Proc. Nati. Acad. Sci. USA 112:E5123-5132.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.

47. 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 Genomics 18:58.
48. Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SC, Gimenes MA, Martins WS, Valls JF, Grattapaglia D, Bertioli DJ (2005) A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). Theor. Appl. Genet. 111:1060-1071.
49. 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 Breeding 131:418-422.
50. Pandey MK, Khan AW, Singh VK, Vishwakarma MK, Shasidhar Y, Kumar V, Garg V, Bhat RS, Chitikineni A, Janila P, Guo B, Varshney RK (2017) QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea L.*). Plant Biotechnol. J. 15:927-941.
51. 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.
52. Park M, Lee JH, Han K, Jang S, Han J, Lim JH, Jung JW, Kang BC (2019) A major QTL and candidate genes for capsaicinoid biosynthesis in the pericarp of Capsicum chinense revealed using QTL-seq and RNA-seq. Theor. Appl. Genet. 132:515-529.
53. Qu J, Ma C, Feng J, Xu S, Wang L, Li F, Li Y, Zhang R, Zhang X, Xue J, Guo D (2016) Transcriptome dynamics during maize endosperm development. PloS One 11: e0163814.
54. Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MV, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 122:1119-1132.
55. Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol. Breeding 33:1-14.
56. Shirasawa K, Hirakawa H, Tabata S, Hasegawa M, Kiyoshima H, Suzuki S, Sasamoto S, Watanabe A, Fujishiro T, Isobe S (2012a) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor. Appl. Genet. 124:1429-1438.
57. Shirasawa K, Koilkonda P, Aoki K, Hirakawa H, Tabata S, Watanabe M (2012b) 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. 12:80.
58. Shu J, Liu Y, Zhang L, Li Z, Fang Z, Yang L, Zhuang M, Zhang Y, Lv H (2018) QTL-seq for rapid identification of candidate genes for flowering time in broccoli x cabbage. Theor. Appl. Genet. 131:917-928.

59. Singh VK, Khan AW, Jaganathan D, Thudi M, Roorkiwal M, Takagi H, Garg V, Kumar V, Chitikineni A, Gaur PM, Sutton T, Terauchi R, Varshney RK (2016) QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. Plant Biotechnol. J. 14:2110-2119.
60. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat. Genet. 39:623-630.
61. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C et al (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74:174-183.
62. Takagi H, Tamiru M, Abe A, Yoshida K, Uemura A, Yaegashi H et al (2015). MutMap accelerates breeding of a salt-tolerant rice cultivar. Nat Biotechnol. 33:445-449.
63. 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.
64. 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.
65. 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.
66. Wang S, Basten C, Zeng Z (2012) Windows QTL Cartographer v2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (https://brcwebportal.cos. ncsu.edu/qtlcart/WQTLCart.htm)
67. 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.
68. 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.
69. 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. 129:117-130.
70. 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.
71. 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. Genomics 295:409-420.
72. 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.

73. 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. 18:96-105.
74. 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 genomics 17:420.
75. 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.

Figures

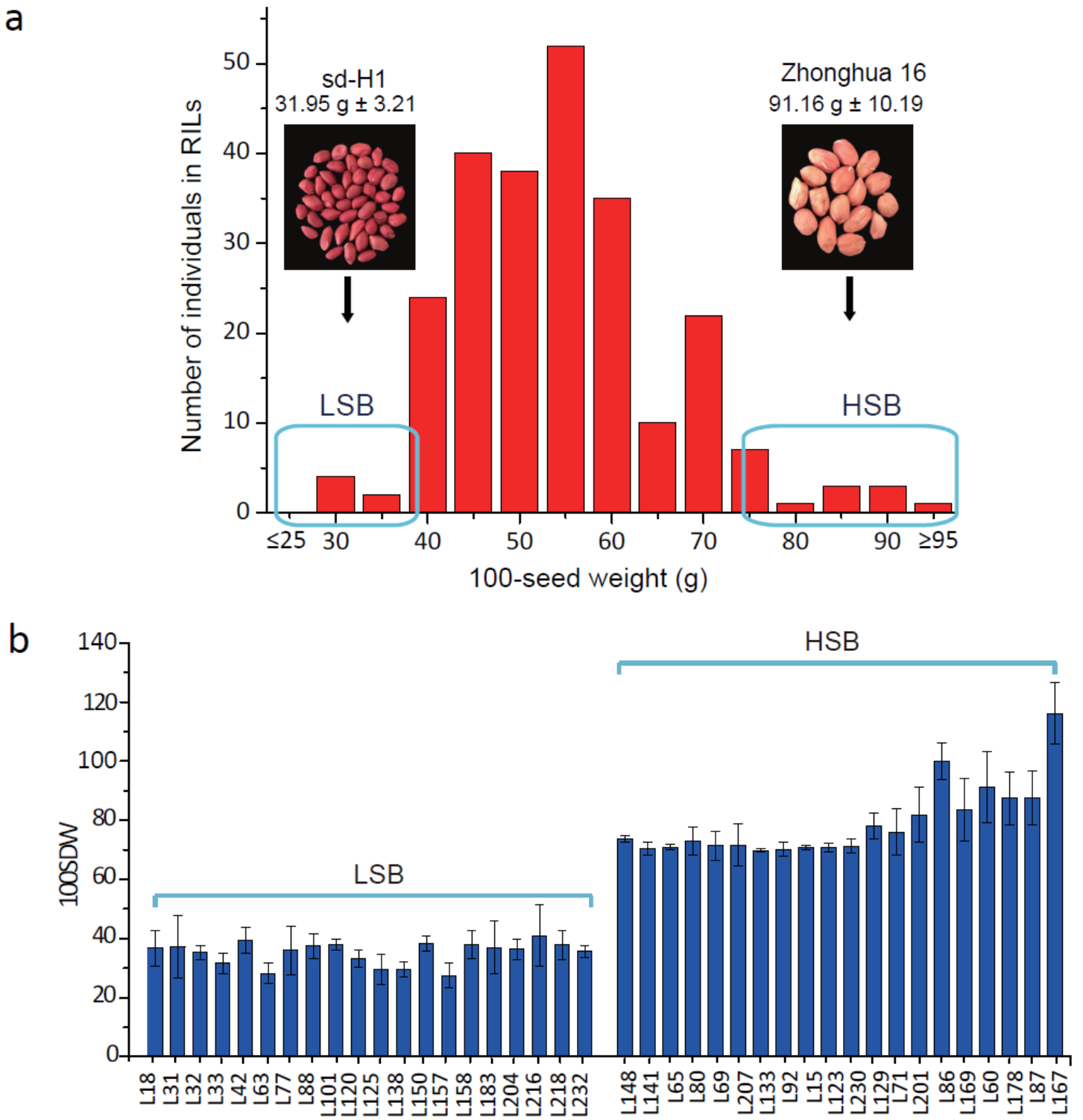

Figure 1

Frequency distribution of 100-seed weight (g) in RILs crossed by Zhonghua 16 and sd-H1. (a) The LSB (low seed weight bulk) and HSB (high seed weight bulk) were constructed by selecting twenty 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 phenotypic variability among the RILs selected for development of two extreme bulks for seed weight.

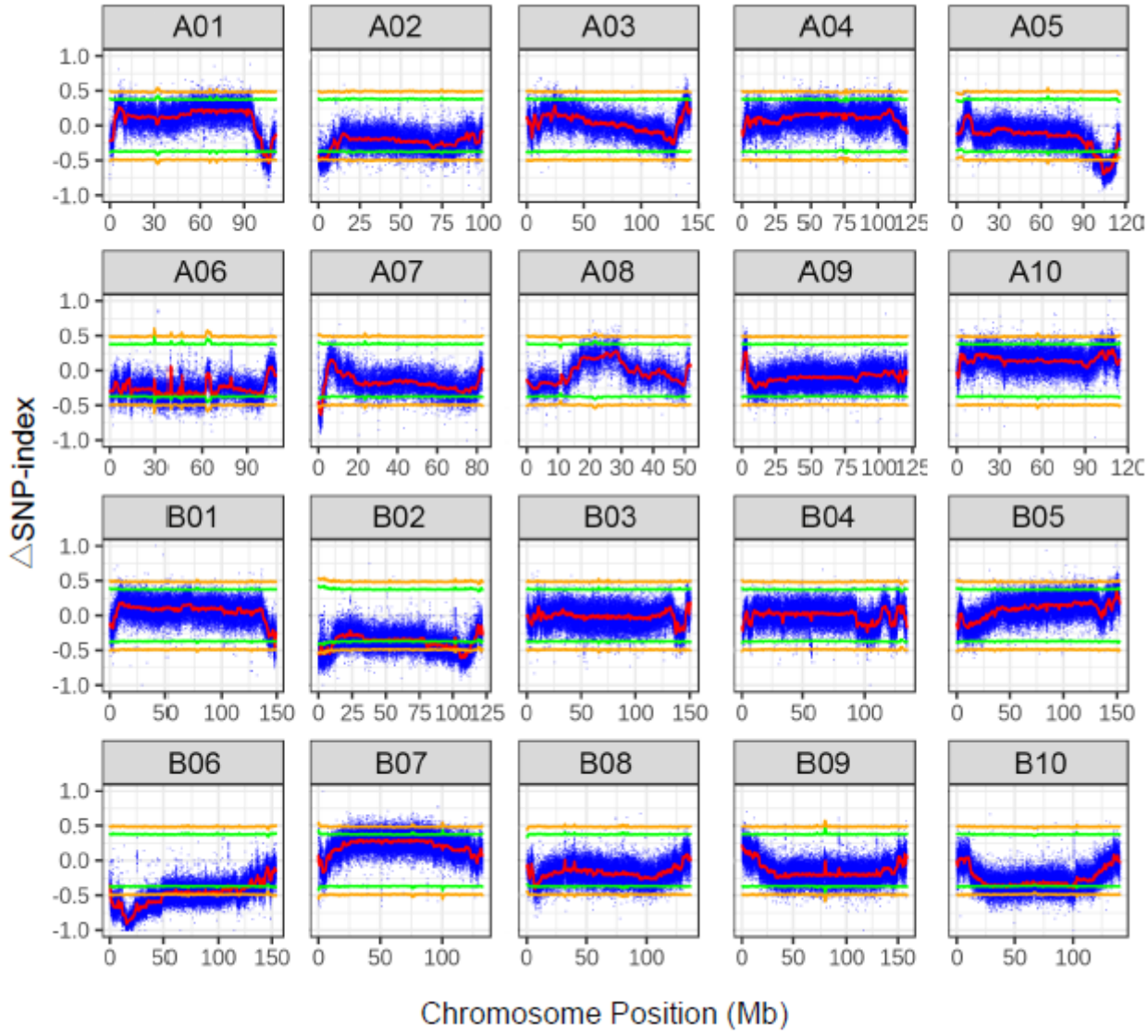

Figure 2

The distribution of ΔSNP-index values across all chromosomes. ΔSNP-index was estimated based on 1 Mb physical interval with a 10 kb sliding window. The blue dots and red lines indicated theΔSNP-index value for SNPs and windows. The orange and green lines showed the statistical confidence interval representing significant levels at 0.01 and 0.05.

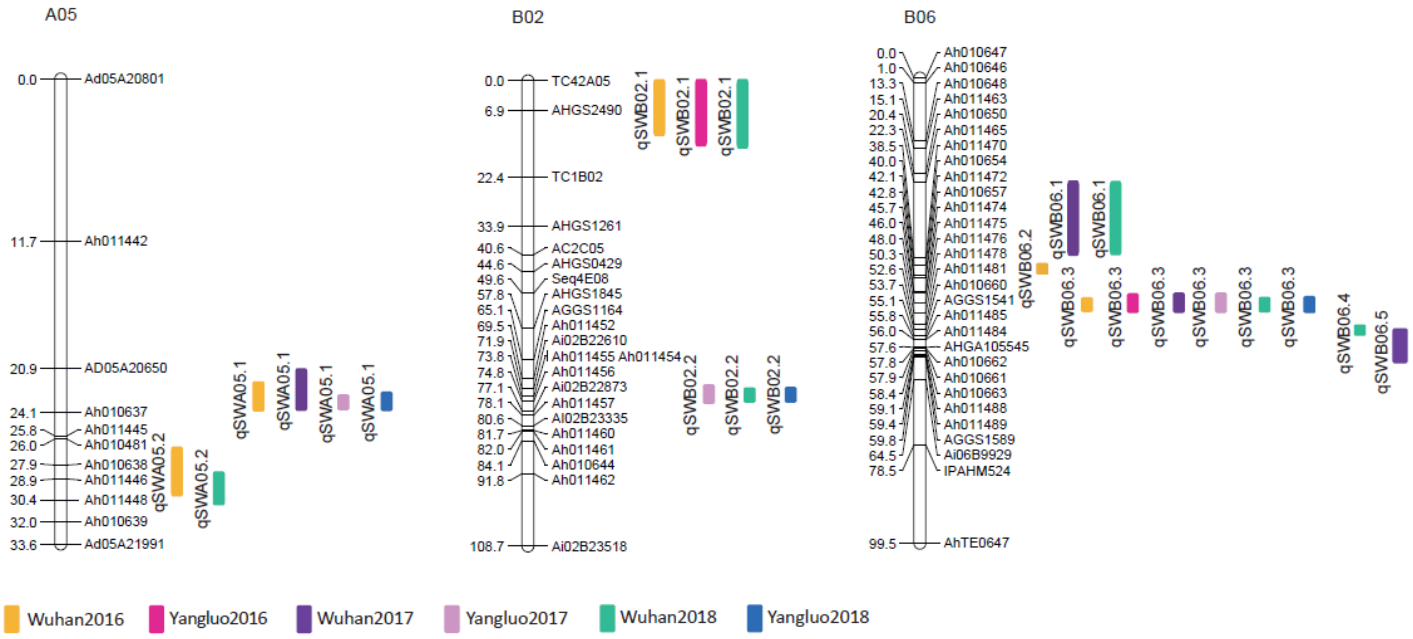

Figure 3

Three candidate regions on A05, B02 and B06 were narrowed down through newly developed SSR and SNP markers based on traditional QTL mapping method. The phenotype from six environments were used to identified QTLs and denoted as different colors.

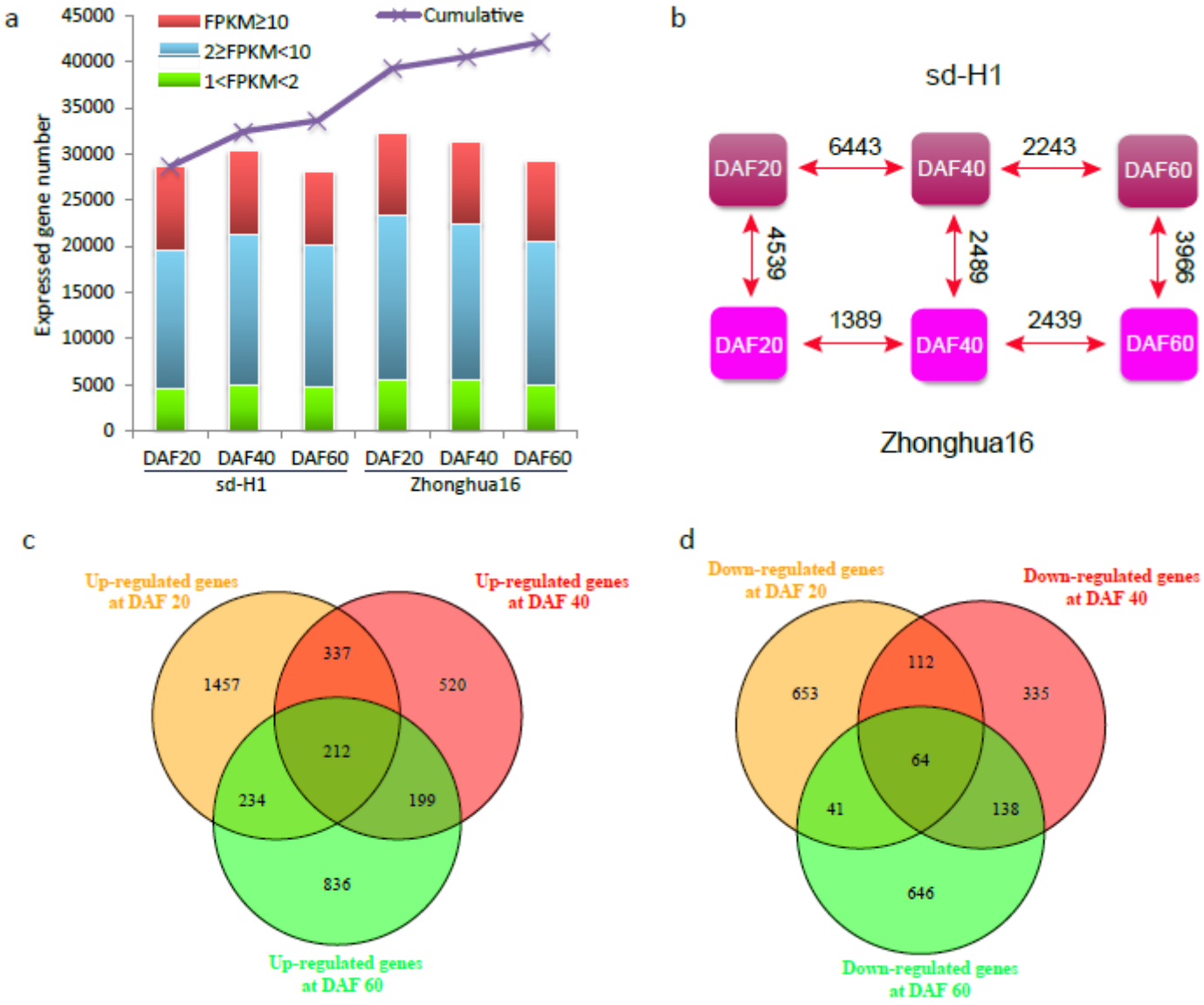

Figure 4

The differently expression analysis at three stages of seed development based on RNA-seq data. (a) The number of expressed gene at three stages (DAF 20, DAF 40 and DAF 60) of seed development in two parents Zhonghua 16 and sd-H1. Three different groups were divided based on their gene expression value (FPKM), denoted as red, blue and green colors. The cumulative gene number was also showed calculated based on all samples. (b) The differently expressed genes (DEGs) between Zhonghua16 and sd-H1 at each developmental stage as well as DEGs between different developmental stages in each parent. (c) The shared and specific up-regulated DEGs at three different stages. (d) The shared and specific down-regulated DEGs at three different stages.

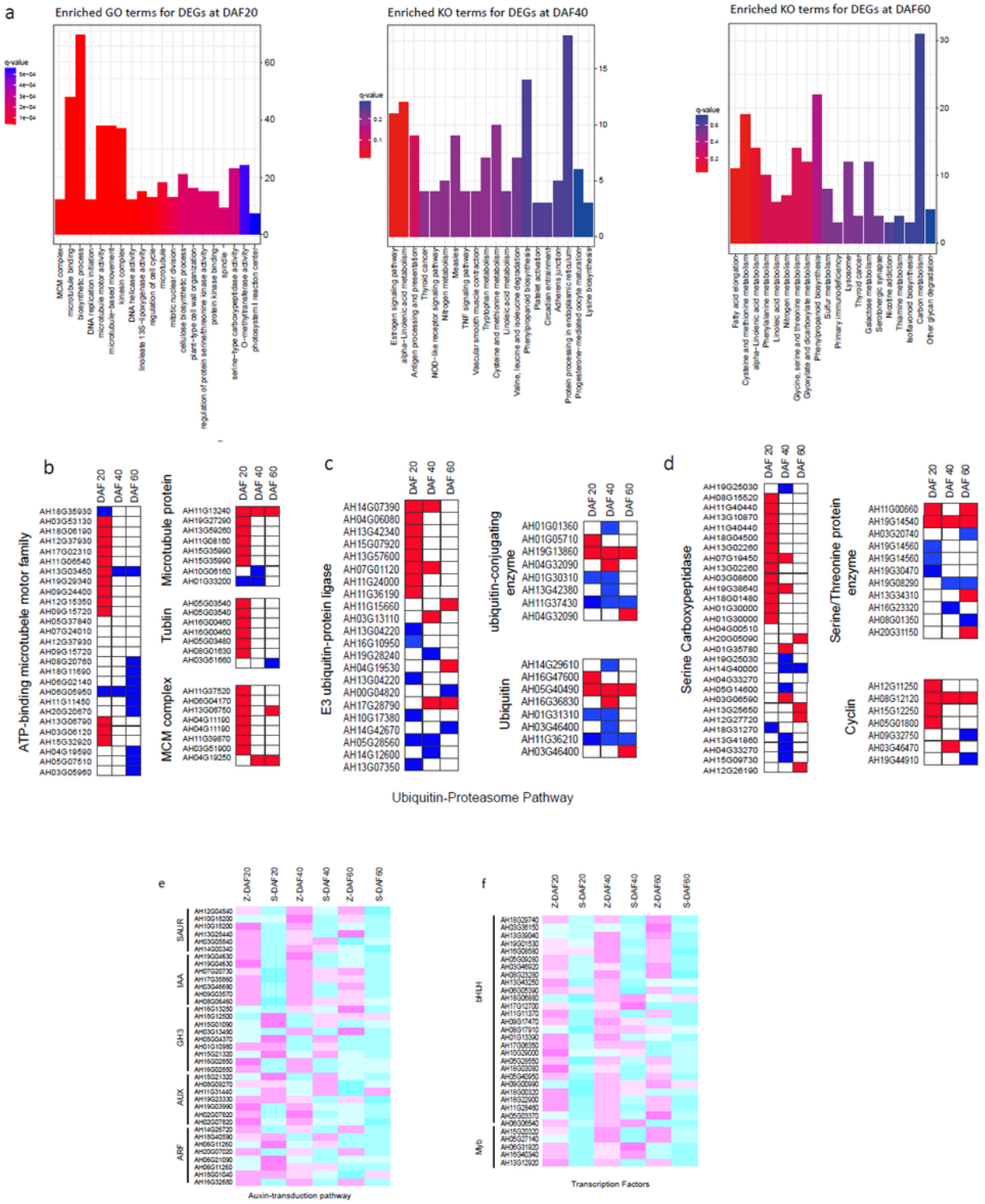

Figure 5

The distribution of DEGs in different pathways and gene families. (a) The enriched GO categories for DEGs at three developing stages DAF 20, DAF 40 and DAF 60, respectively. (b-d) The DEGs at three developing stages were examined in ATP-binding microtubele motor family, Ubiquitin-Proteasome Pathway and serine/threonine protein pathway. The red and blue boxes represented that DEGs were up-regulated and down-regulated, respectively. (e) The DEGs at three developing stages in auxin-transduction

pathway including SAUR, IAA, GH3, AUX, ARF genes. (f) The DEGs at three developing stages in bHLH and MYB genes. The gene expression values were colored, and red and blue represented high and low expression, respectively.

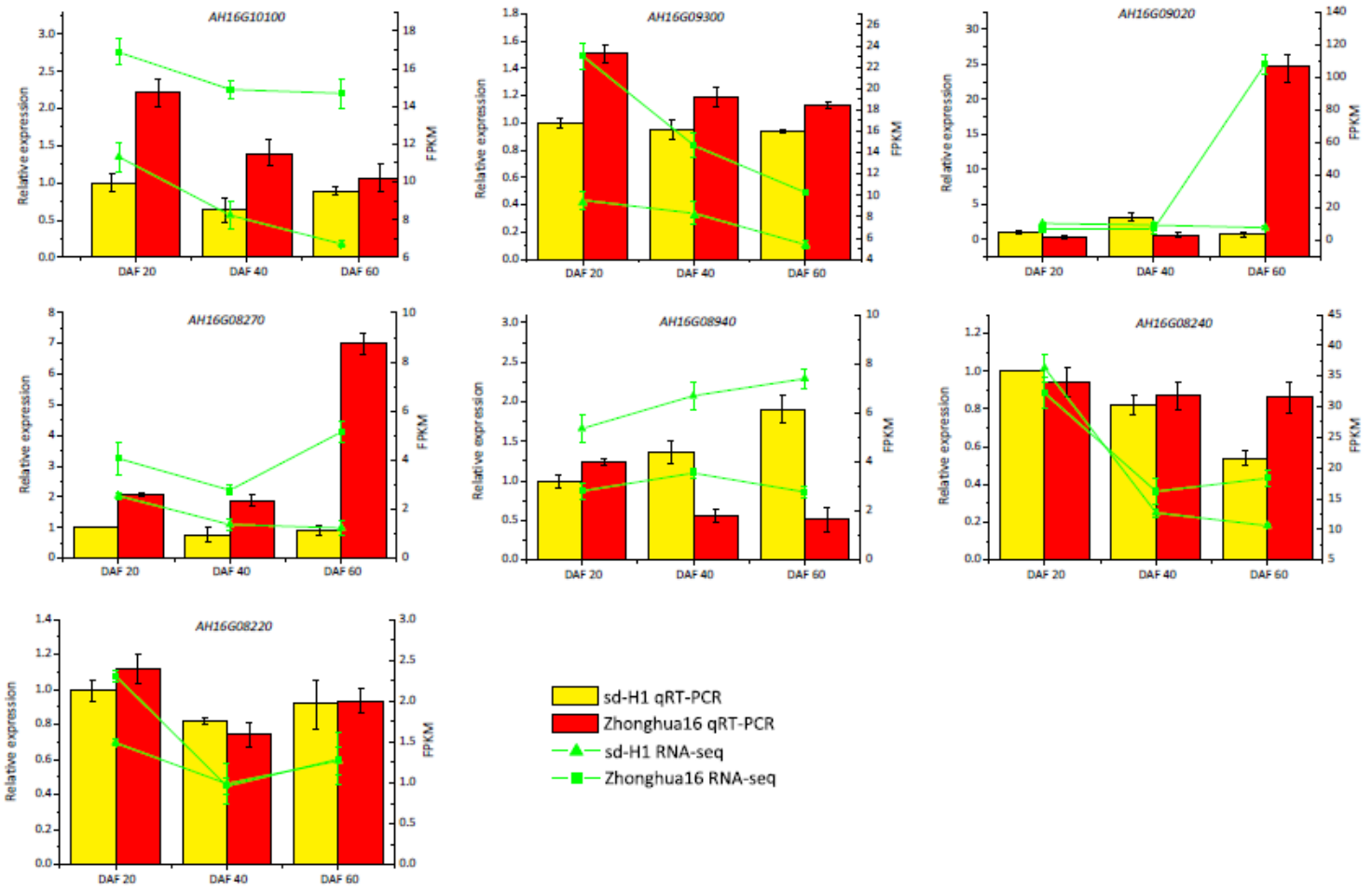

Figure 6

The gene expression pattern of predicted candidate genes in qSWB06.3 determined by RNA-seq and qRT-PCR. The relative gene expression estimated from qRT-PCR experiment and FPKM value calculated based on RNA-seq data were shown in left and right y-axes. The two colored histograms and lines represented the relative gene expression and FPKM values in two parents, respectively.

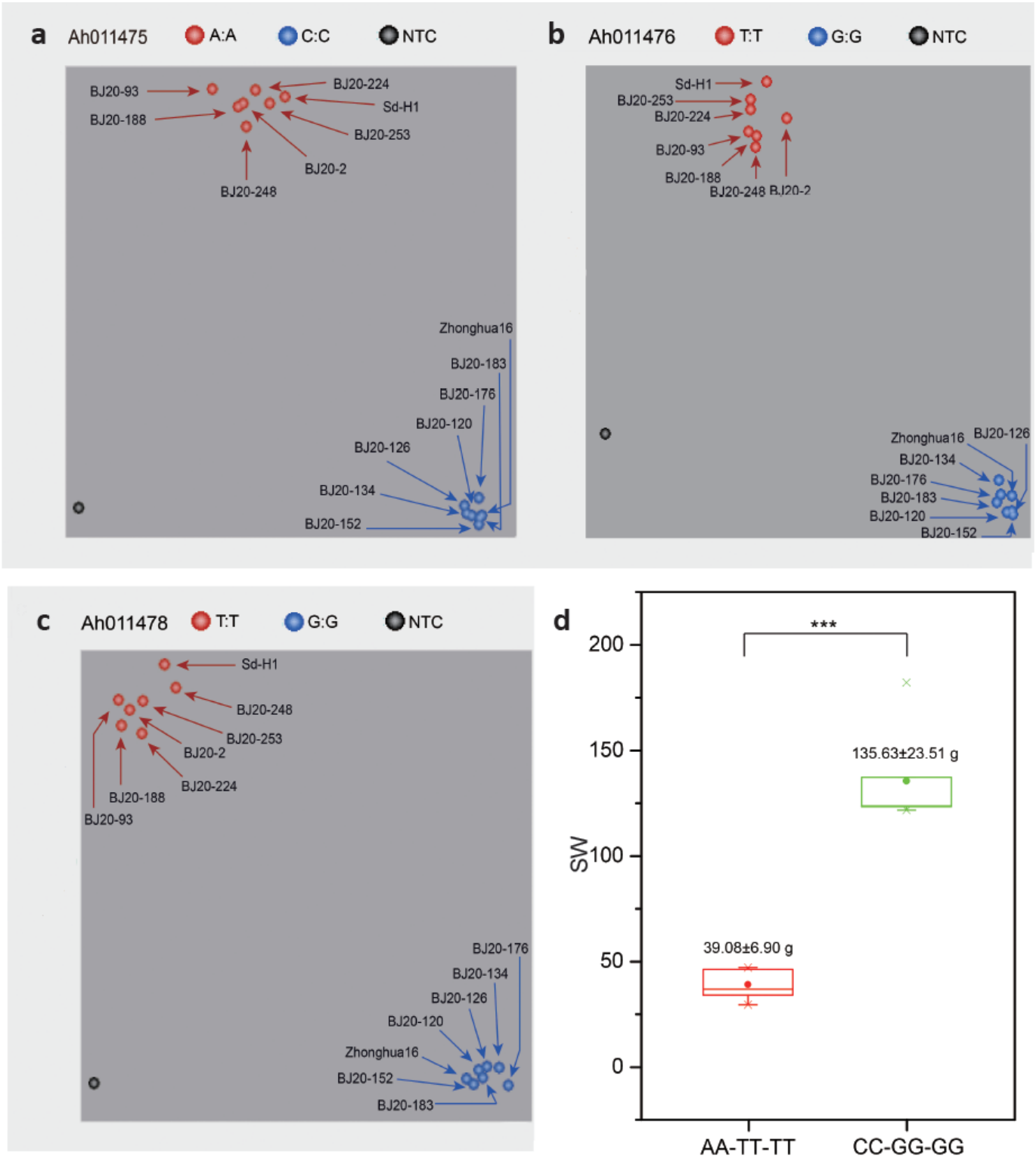

Figure 7

The alleles of three SNP markers in qSWB06.3 were examined in sixty natural peanut varieties with different seed weight. Three KASP-SNP markers Ah011475 (a), Ah011475 (b), Ah011478 (c) were separately examined their genotype among peanut varieties. The red and blue points were corresponding to two different genotypes between Zhonghua 16 and sd-H1. The parent Zhonghua16 with high seed weight has AA, TT and TT alleles for three markers, whereas, parent sd-H1 with low seed weight has CC,

GG and GG alleles for three markers. (d) The natural peanut varieties with AA-TT-TT alleles showed significantly lower seed weight, compared with those with CC-GG-GG alleles.

Supplementary Files

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

- SupplementaryFigureS1.pdf
- SupplementaryFigureS2.pdf
- SupplementaryFigureS3.pdf
- SupplementaryFigureS4.pdf
- SupplementaryTableS1.pdf
- SupplementaryTableS2.pdf
- SupplementaryTableS3.pdf
- SupplementaryTableS4.pdf
- SupplementaryTableS5.pdf
- SupplementaryTableS6.pdf