BMC Genomics

RESEARCH

Open Access

Fine mapping and candidate gene analysis of major QTLs for number of seeds per pod in *Arachis hypogaea* L.

Long Li[1], Shunli Cui[1], Xiukun Li[1], Mingyu Hou[1], Yingru Liu[1] and Lifeng Liu[1*]

Abstract

Background Peanut (*Arachis hypogaea* L., 2n = 2x = 20) is an important industrial and oil crop that is widely grown in more than 100 countries. In recent years, breeders have focused on increasing the seed number per pod to improve their yield in addition to other breeding for other key components of yield, including the pod number, seeds per pod, and 100-seed weight.

Results In this study, a secondary population of 1,114 BC_1F_2 lines was derived from a cross between the parents R45 and JNH3. Two stable major-effect quantitative trait loci of *qRMPA09.1* and *qRMPA09.2* were detected simultaneously and mapped within chromosomal intervals of approximately 400 Kb and 600 Kb on chromosome A09. Additionally, combined whole-genome and RNA sequencing analyses showed the differential expression of the *Arahy.04JNDX* gene that belongs to a MYB transcription factor (TF) between the two parents. The *AhMYB51* gene was also inferred to influence the number of seeds per pod in peanuts. An examination of the backcross lines L_2/L_4 showed that *AhMYB51* increases the rate of multiple pods per plant (RMSP) primarily by affecting brassinosteroids in the flowers, while its overexpression promotes the length of siliques in *Arabidopsis thaliana*.

Conclusions Our findings provide valuable insights for the cloning of favorable alleles for RMSP in peanuts. The *qRMSPA09.1* and *qRMSPA09.2* are two novel QTL associated with the RMSP trait, with *AhMYB51* predicted as its candidate gene. Moreover, the closely linked polymorphic SNP markers for loci of two significant QTLs may be useful in accelerating marker-assisted breeding in peanuts.

Keywords *AhMYB51*, MYB transcription factor, Bulked Segregant analysis, Peanut seeds per pod

*Correspondence:
Lifeng Liu
liulifeng@hebau.edu.cn
[1]College of Agronomy, State Key Laboratory of North China for Crop Improvement and Regulation/Key laboratory of Crop Germplasm Resources of Hebei Province, Hebei Agricultural University, Baoding 071001, P.R. China

Background

Peanut (*Arachis hypogaea* L.) is a significant oil crop that is generally grown in more than 100 countries worldwide [1]. Peanuts are also one of the most important and popular nuts and snacks available to consumers, and they have been recognized as healthy and nutritious. Their kernels contain 28% protein, 48% oil, essential vitamins and minerals and can be processed into candies, cookies, and peanut butter [2]. The economic return of peanuts is directly influenced by their yields [3, 4], thus, researchers breed peanuts with an increased number of seeds per pod, pod number and 100-seed weight [5, 6]. The number of seeds per pod is also an important agronomic trait that is highly influenced by environmental factors [7]. In recent years, different populations of peanuts have been utilized to identify the major quantitative trait loci (QTL) associated with seed numbers in peanuts, particularly on chromosomes A05 and B06 [8]. Although the QTL for pod number have been identified, their underlying metabolic and molecular mechanisms remain unclear.

Phytohormones regulate many aspects of plant development, including cell metabolism, apical dominance, flowering time, and seed development [9, 10]. For instance, brassinosteroids and cytokinins play critical roles in pollination, fertilization, and ovule development during the development of pods [11, 12]. Recent studies have increasingly focused on the seed number, particularly in relation to ovule development and pollination in crops, such as soybean (*Glycine max* L.) [13], maize (*Zea mays* L.) [14], oilseed rape (*Brassica napus* L.) [15], and broad bean (*Phaseolus vulgaris* L.). The molecular mechanisms that control embryogenesis and development in plants have been studied extensively [16, 17]. Phytohormones, particularly brassinosteroids [18], and TFs, such as NAC [19], MYB [20], and bHLH [21], have been shown to play a key role in the regulation of seed development, which is a dynamic and complicated process governed by a regulatory mechanism of multiple genes in plants. Although many genes have been reported to be involved in the development of pods across various crops, the regulatory networks that govern metabolism in peanut seeds remain unclear. Nevertheless, the availability of high-quality reference genomes for both wild diploid and cultivated peanuts has accelerated studies on their transcription and the development of multi-omics approaches, as well as molecular breeding programs aimed at improving the yield and quality of their oil [22–24]. Multi-omics technology offers a valuable opportunity to elucidate the regulatory mechanisms of seed development in peanuts [25].

In a previous study, the rate of multiple pods per plant (RMSP) was mapped in two intervals, A09 chromosome 114.00-119.66 Mbp (5.66 Mbp) and 110.90-131.6 Mbp (2.26 Mbp) [26]. In this study, we developed a secondary population and finely mapped it in combination with QTL-seq, whole-genome resequencing, and RNA-seq to analyze the differentially expressed genes (DEGs) related to the seed number per pod in peanuts. We also analyzed the candidate genes in extreme offspring and *Arabidopsis thaliana*. These findings will help to understand the genetic basis of the regulation of RMSP and could ultimately facilitate the development of high-yielding breeds in peanuts.

Methods

Plant materials and field evaluation

In the RIL population, R45 (female parent with RMSP of 77.6%) peanut carries the two target QTL and has a genetic background that is similar that of the multi-pod variety Silihong. Therefore, R45 was backcrossed with the non-multi-pod parent JNH3 (male parent with RMSP of 0%) to construct a secondary population, BC_1F_1. The BC_1F_1 population was then cultivated in Sanya, Hainan Province, China (109.16E; 18.19 N) in December, 2023. The pods were harvested in April 2023 and planted in June at the Xushui Experimental Base of Hebei Agricultural University (Baoding, China) (115.56 E; 38.79 N), and the individual plants were harvested in September 2023. The BC_1F_2 population, which consisted of 1,114 lines, was evaluated for the seed number per pod after harvest. Two extreme lines were selected from the BC_1F_2 population and designated BC_1F_2-L_2 (two pods per plant) and BC_1F_2-L_4 (multiple pods per plant). These lines were then self-pollinated to produce BC_1F_3-L_2, $BC1F_4$-L_2 and BC_1F_3-L_4, BC_1F_4-L_4. The selected lines were grown in a greenhouse at 16 h:8 h light: dark at 28 °C / 20 °C.

Bulked Segregant analysis (BSA)-seq analysis and whole-genome resequencing

The genomic DNA was extracted and tested for quality as previously described [27]. For BSA-seq, the same amounts of DNA from 30 lines with multiple pods per plant and 30 lines with double pods per plant from the BC_1F_2 population were pooled to create a pool of four pods (F-pool) and one of two pods (T-pool), respectively. DNA libraries from R45, JNH3, BC_1F_3-L_4, BC_1F_3-L_2, and the two DNA pools were sequenced using the Illumina NovaSeq6000 platform (Illumina, San Diego, CA, USA) by Shanghai Majorbio Technologies (Shanghai, China).

Raw reads of low quality (mean Phred score < 20), including those that contained adapter contamination or unrecognizable nucleotides (N base > 10), were trimmed using Fastp [28] and mapped to the reference genome using BWA-MEM [29] with the default parameters. After the base quality had been recalibrated, the germline variant calling, which included SNPs and InDels across all the samples, was performed using the Haplotyper and Gvcftyper programs from the Sentieon genomics toolkit

[30]. All the variants were filtered based on the standard hard filtering parameters recommended by the GATK Best Practices pipeline and annotated using SnpEff [31]. The SNPs and InDels were categorized by their chromosomal positions, including their locations to intergenic, exons, and introns. The BSA-seq analysis was conducted using the index-slid method [32], index-loess method, Euclidean distance (ED) algorithm [33], and the Gprime method [34], which were driven by deep learning and the least square method [35], respectively.

RNA sequencing and data analysis

The total RNA from SLH, JNH3, BC_1F_3-L_4, and BC_1F_3-L_2 was isolated using a Plant RNA Extraction Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's instructions. There were 12 libraries that consisted of four samples, each with three biological replicates. They were sequenced on an Illumina NovaSeq 6000 platform and generated 150-nucleotide-long paired-end reads. High-quality sequences were aligned to the peanut reference genome Tifrunner (https://peanutbase.org) [36]. The gene expression was defined by Fragments Per Kilobase Million (FPKM), and the FPKM for each annotated reference gene was calculated using StringTie v. 1.3.4 [37]. The presence of DEGs was confirmed using DESeq2 where transcripts with a false discovery rate (FDR) < 0.05 and|fold change| ≥ 2 were considered to be differentially expressed. Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses were subsequently conducted.

Vector construction for the transformation of *Arabidopsis thaliana*

Appropriate primers were designed by Premier 5.0 software to obtain the cDNA sequence of the TF *AhMYB51* (Table S1). *Arabidopsis thaliana* plants were transformed using the floral dip method [38] with *Agrobacterium tumefaciens* GV3101, which harbors the pCAMBIA-MYB51 vector, with kanamycin resistance gene as a selectable marker (Table S1). *AhMYB51* was ligated into the pCAMBIA-MYB51 vector as described by Liu et al. [39]. The *A. thaliana* plants were dipped with the *AhMYB51* gene using *Agrobacterium tumefaciens* GV3101 and then cultivated on MS solid medium (50 mg/L Basta) at 22 ℃ under a 16 h: 8 h light: dark cycle in a growth chamber to check for the overexpression of *AhMYB51*. After PCR identification and glyphosate screening, T_3 generation homozygous transgenic lines that overexpressed *AhMYB51* (*AhMYB51*-OE) with a single insertion site were obtained. Plant growth and flowering time were recorded for each plant along with the seed number and pod length.

Quantitative real-time PCR (qRT–PCR)

Seven varieties, including SLH, R45, JNH3, BC_1F_3-L_4, BC_1F_3-L_2, BC_1F_4-L_4, and BC_1F_4-L_2, were selected for a qRT–PCR analysis of *AhMYB51*. Samples were collected at the flower and peg states. The total RNA was extracted from the samples using a Plant RNA Extraction Kit (Tiangen Biotech, Beijing, China) according to the manufacturer's instructions. The primers to target *AhMYB51* were designed by Premier 5.0 software (Table S1) using the cDNA sequences retrieved from Peanut Base (https://peanutbase.org). qRT-PCR was performed using SYBR Premix Ex Taq II (TaKaRa, Dalian, China), on an ABI StepOne Plus Real-Time PCR System (Roche, Basel, Switzerland). All the data were analyzed using the $2^{-\Delta\Delta CT}$ method [40].

Statistical analysis

The calculations were performed using SPSS 28.0 (IBM, Inc., Armonk, NY, USA) and Origin 8.0 (OriginLab, Northampton, MA, USA). The results were presented as the mean ± SD of three independent biological replicates. The least signifcant difference (LSD) test was performed to determine the signifcance of differencesbetween different treatment groups. A genetic analysis was conducted using Performance software [41, 42].

Results

Phenotypic variation of the rate of multiple pods per plant in BC_1F_2

The backcross between R45 and JNH3 yielded 1,114 BC_1F_2 lines, and their harvested pods were used to determine the RMSP in each line. R45 had a significantly higher pod number compared to JNH3 (Fig. 1b). The RMSP of SLH (52.72%) and R45 (77.6%) was higher than that of JNH3 (0%).

Identification of QTLs for RMSP using BSA methods

Based on the phenotype of the BC_1F_2, the T-pool, F-pool, and parental lines were bulked. The resequencing of the whole genome and the analysis of BSA-seq were produced for the extreme and parent pools. A total of 24.7 Gb, 27.03 Gb, 75.37 Gb, and 72.35 Gb Clean reads were generated from JNH3, R45, T-pool, and F-pool, respectively. Of these, 99.79%, 99.92%, 99.89%, and 99.85% of the reads were mapped to the reference genome (Table 1).

The low-quality clean reads were removed, and the high-quality reads were then identified in JNH3, R45, the T-pool and F-pool. Among these, high-quality SNPs/InDels were located on ChrA05, ChrA06, ChrA09, ChrB05, ChrB06, which indicated that the main gene that regulates RMSP is probably located on ChrA09 (Table 2; Fig. 2, Fig. S1 and S4).

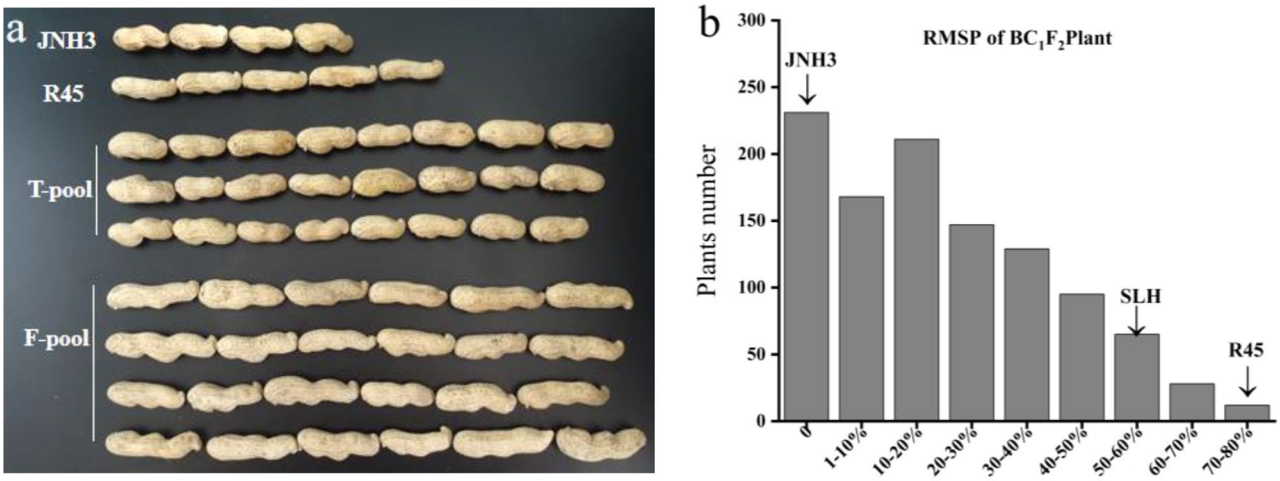

Fig. 1 Phenotypes of JNH3, R45, T-pool and F-pool in the BC_1F_2 population. (**a**) Phenotypes of JNH3 the rate of multiple pods per plant (RMSP = 0) and R45 (RMSP = 77.6%), (**b**) RMSP in the BC_1F_2 population after harvest. Arrows represent R45, JNH3 and SLH

Table 1 Summary of the sequencing data

Name	Clean data	Mapped Ratio (%)	Proper Ratio (%)	Clean Reads	Real Depth	Coverage (%) (≥ 1x)	Clean Q30 (%)
JNH3	24.7G	99.79	99.1	165,003,064	9.88 X	97.55	91.68
R45	27.03G	99.92	99.18	180,629,916	10.79 X	97.88	91.66
T-pool	75.37G	99.89	99.02	503,823,652	29.69 X	99.15	91.33
F-pool	72.35G	99.85	99.09	483,714,140	28.49 X	99.15	90.87

Table 2 Gene mapping results of BSA analysis RMSP

Method	Region (M)	SNPs	Effective SNP	InDels	Genes
Euclidean	ChrB05:16.55:17.13	14	1	1	34
Euclidean	ChrB06:145.1-145.5	14	3	2	31
Euclidean	ChrA05:62.0-62.8	11	0	0	1
Gprime	ChrA05:61.9–62.5	11	0	0	1
Gprime	ChrA06:5.9–6.9	22	0	3	139
Gprime	ChrA09:112.7-115.2	42	2	11	301
index-slid	ChrA05:62.4–62.5	4	0	0	0
index-loess	ChrA09:109.6-115.3	89	4	24	747

The whole-genome sequence analysis revealed a reduction in the number of genes

Based on the localization results of the number of pods in the previous RIL population across two environments, two main QTL, *qRMSPA09.1* and *qRMSPA09.2* were detected on ChrA09 [26]. *qRMSPA09.1* was detected between 110.90 and 113.16 Mb, while *qRMSPA09.2* was in the region between 114.00 and 119.66 Mb, with sizes of 2.3 Mb and 5.2 Mb, respectively, based on Euclidean, Gprime, index-slid and index-loess methods (Fig. 2; Table 2). We also integrated the BSA-seq analysis data for ChrA09:112.76–115.25 Mb and refined the QTL ranges to 400 Kb and 600 Kb, respectively. The *qRMSPA09.1* interval contained 32 genes, while the *qRMSPA09.2* interval contained 86 genes (Table S2).

To further map the candidate genes for *qRMSPA09.1* and *qRMSPA09.2*, we compared the amount of genomic variation between the two parental lines. After stringent filtering, a total of 34 variations were identified, which are unlikely to cause functional deficiencies; they included 20 in the intergenic regions, five in introns, four upstream, three downstream, one in 5′-UTR, while one caused missense mutations (Fig. 3a). Among the variations in introns, four SNPs, and one InDel caused mutations, while 10 SNPs and 24 InDels resulted in significant variation in the sequence of bases, which is likely to disrupt the function of genes (Fig. 3b). The variations were also associated with 34 genes (Table 3), including three uncharacterized protein genes and 31 annotated against the Tifrunner v. 2.0 reference genome (https://peanutbase.org), with none of the annotated genes associated with an increase in seed number. Therefore, we hypothesized that the functional variation associated with the *qRMSPA09.1* and *qRMSPA09.2* loci was probably owing to changes in the patterns of gene expression, which could possibly have arisen from the differential promoter activity.

Validation of the candidate genes by differential gene analysis

To identify the genes associated with *qRMSPA09.1* and *qRMSPA09.2*, we examined the patterns of expression of the candidate genes. A total of 32 genes were annotated for *qRMSPA09.1* and 86 for *qRMSPA09.2*. We then assessed the DEGs of the remaining 34 predicted genes

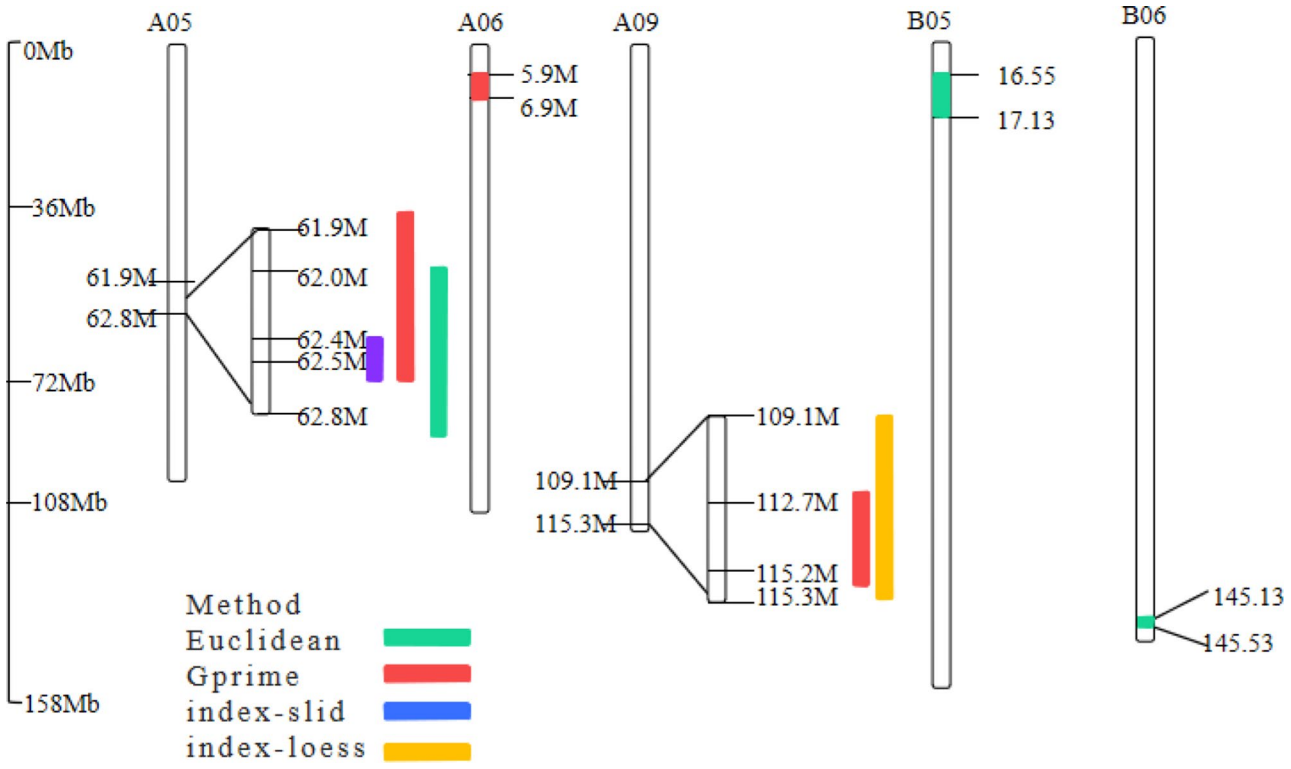

Fig. 2 Quantitative trait loci for the rate of multiple pods per plant traits identified on chromosomes

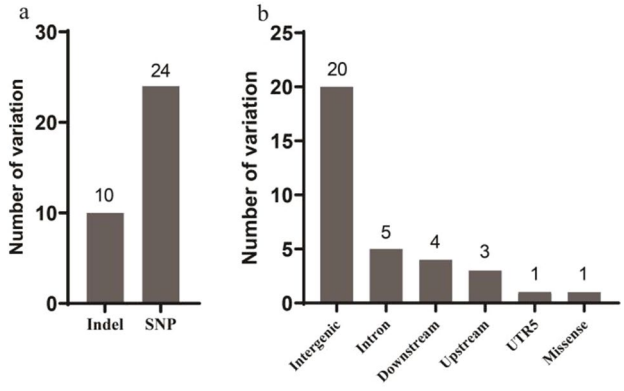

Fig. 3 **(a)** The genomic distribution of variations in two candidate regions, (**b**) type of mutation in their coding region

based on the|log2 fold-change| ≥ 1 and $P<0.001$ and identified five DEGs for *qRMSPA09.1* and seven DEGs for *qRMSPA09.2*. Among these 12 candidate genes, *Arahy.04JNDX*, which encoded an MYB TF was identified as the candidate gene based on our previous study [20]. *Arahy.04JNDX* was expressed at significantly lower level in the two-seeded pods of JNH3 (FPKM = 3.72) than in the multiple pods of SLH during flowering (FPKM = 9.89) (Table 4). Therefore, this SNP was probably induced by the differential expression observed between the two parental lines. The MYB genes have also been well-documented as key regulators of the development of pollen and ovules in plants [43].

The whole-genome sequencing and transcriptome analysis of offspring BC_1F_3

Two allelic lines were developed from the BC_1F_3 population, including L_2, which exhibited the phenotypes of their female parent (JNH3), and L_4 lines that resembled their male parent, R45. The resequencing of L_4 and L_2 lines at a depth of 10 X generated a total of 634,721 SNPs and 168,924 InDels within ChrA09. The intergenic region of *Arahy.04JNDX* was mutated at ChrA09:114,148,416, while its downstream region was stably inherited. The KEGG pathway analysis indicated that the DEGs and variants were predominantly involved in the extreme offspring biosynthetic signaling pathway (Fig. 4a), transport processes, and metabolism throughout the process of flower development (Fig. 4b). We compared the expression of genes in the $BC1F_3$-L_2 and BC_1F_3-L_4 flowers and their genome and identified 4,476 DEGs and 35,283 variable sites, respectively (Fig. 5). A comparative GO analysis of the JNH3 and SLH flowers revealed that the DEGs were primarily involved in pollination, while the KEGG analysis highlighted their significant role in the brassinosteroid biosynthesis signaling pathway (Table S3). The parental whole-genome sequence comparison also indicated an SNP of G-A at Chr09-114145495 in the intergenic region of *Arahy.04JNDX*, which further supports

Table 3 Types of SNP and indel mutations in the candidate genes

Candidate genes	Start position	End position	JNH3	R45	Mutant annotation
Arahy.XW3T4W	112,790,246	112,800,686	C	T	Intergenic region
Arahy.8F881C	112,810,188	112,822,568	A	C	Upstreamgene variant
Arahy.75YJ3Z	112,812,745	112,815,516	A	C	Missense variant
Arahy.3TDE18	112,849,718	112,859,150	C	T	Intergenic region
Arahy.G05VUU	112,870,426	112,875,195	G	T	Intergenic region
Arahy.NHVJ82	112,958,766	112,962,321	C	T	Intergenic region
Arahy.9468DL	112,980,265	112,992,011	G	T	Intergenic region
Arahy.KZCX9C	113,000,931	113,004,034	G	T	Downstreamgene variant
Arahy.ZAU2KF	113,004,585	113,009,189	GA	G	5′UTR variant
Arahy.ELL9VR	113,010,087	113,014,838	GA	G	Downstream gene variant
Arahy.GYF1N1	113,050,568	113,053,769	C	CA	Intergenic region
Arahy.GC5VJY	114,036,957	114,046,905	G	A	Upstream gene variant
Arahy.JNSA3Z	114,051,294	114,053,527	G	A	Intergenic region
Arahy.04JNDX	114,143,849	114,145,495	A	G	Intergenic region
Arahy.DEP6R1	114,158,088	114,158,451	A	G	Intergenic region
Arahy.KL5LPC	114,160,087	114,163,608	T	G	Intron variant
Arahy.42CZAS	114,172,195	114,175,899	T	A	Intron variant
Arahy.FJSM0P	114,212,023	114,212,927	A	T	Downstream gene variant
Arahy.L8YGTT	114,216,085	114,221,210	A	T	Intergenic region
Arahy.N0KYQA	114,257,350	114,262,137	C	T	Intergenic region
Arahy.H5PXMS	114,268,287	114,275,077	C	T	Intergenic region
Arahy.T57XL4	114,593,865	114,608,503	C	T	Intron variant
Arahy.W4LCKL	114,650,557	114,654,924	G	A	Intron variant
Arahy.FI0VBI	114,667,183	114,681,730	CT	C	Intron variant
Arahy.UME9CN	114,736,284	114,739,868	G	GT	Downstream gene variant
Arahy.FXE6P8	114,742,343	114,746,606	G	GT	Intergenic region
Arahy.X9NSEY	114,746,628	114,750,147	C	T	Intergenic region
Arahy.I9L5JE	114,771,575	114,776,586	C	T	Intergenic region
Arahy.VWQQ3X	114,797,559	114,798,968	A	AT	Intergenic region
Arahy.G1TI3A	114,824,454	114,833,063	A	AT	Intergenic region
Arahy.MIS9XN	114,929,088	114,931,838	C	T	Intergenic region
Arahy.JYC97M	114,944,545	114,951,892	C	T	Intergenic region
Arahy.ZLV99U	114,952,416	114,957,915	C	CA	Intergenic region
Arahy4.G0ENY	114,967,166	114,970,904	C	CA	Upstream gene variant

its potential role in regulating the pod number (Fig. 6a). We analyzed peanut transcriptome databases published in different databases tomutually confirm 12 candidate genes. Finally, only *AhMYB51* was highly expressed in these two databases and the transcriptome (Table S4). Therefore, based on these findings, *Arahy.04JNDX* could be the primary regulator for *qRMSPA09.2* (Fig. 6b), but this function requires further validation through additional experiments, including the evaluation of its distribution in offspring varieties and overexpression of the *AhMYB51* gene in *A. thaliana* to confirm its function.

Analysis of *AhMYB51* gene expression in peanut via qRTPCR

The level of expression and RMSP of *AhMYB51* were evaluated in the allelic lines (L_2 and L_4) and their parents, R45 and JNH3, to determine its functions. *AhMYB51* was expressed at higher levels in L_4 and R45 (Fig. 7), which also showed an increase in seed numbers. This suggests that *AhMYB51* enhances the RMSP under field conditions.

Overexpression of *AhMYB51* promotes silique length in *Arabidopsis thaliana*

Since R2R3-type MYB genes are the main transcriptional regulators of seed number in multiple oil crop species, we overexpressed the *AhMYB51* in *A. thaliana* to validate if it regulates the seed number in peanuts. We obtained T_3 generation plants from three batches of transgenic *A. thaliana* and selected three independent lines, including OE-1, OE-2, and OE-3, with similar phenotypic intensities. Line OE-3 exhibited the most significant phenotypic effects (Fig. 8a), with pronounced increases in silique length and seed number compared to lines OE-1, OE-2, and the control (Fig. 8b).

Table 4 Annotation of DEGs in *qRMSPA09.1* and *qRMSPA09.2* candidate intervals

Loci	Candidate genes	Gene family or homologous gene	Gene prediction	SLH-FFPKM	JNH3-FFPKM	SLH-ZFPKM	JNH3-ZFP-KM
qRMSPA09.1	*Arahy.XW3T4W*	MATE	*MATE efflux family protein*	2.41	0.22	0.71	2.23
	Arahy.3TDE18	DExH15	*ATP-dependent RNA helicase*	1.90	0.81	3.76	3.75
	Arahy.G05VUU	DB3S	*2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein*	3.28	1.19	4.37	5.09
	Arahy.ZAU2KF	PI4KA	*lipid phosphate phosphatase*	3.48	1.31	6.01	9.26
	Arahy.GYF1N1	Bem1p	*Octicosapeptide/Phox/Bem1p family protein*	34.61	12.40	31.08	53.93
qRMSPA09.2	*Arahy.04JNDX*	MYB51	*MYB transcription factor MYB51*	9.89	3.72	2.89	3.93
	Arahy.DEP6R1	/	*uncharacterized protein LOC100306658*	68.39	45.80	75.06	62.98
	Arahy.KL5LPC	AMMECR1	*AMMECR1 family*	31.82	9.98	23.98	16.59
	Arahy.42CZAS	FADS2	*fatty acid desaturase2*	14.25	5.55	1.27	0.14
	Arahy.N0KYQA	/	*unknown protein*	96.38	34.12	39.35	49.75
	Arahy.X9NSEY	HSP40	*Heat shock protein DnaJ domain protein*	9.18	3.32	3.20	4.61
	Arahy.G1TI3A	/	*Major facilitator superfamily protein*	2.70	0.27	4.76	10.00

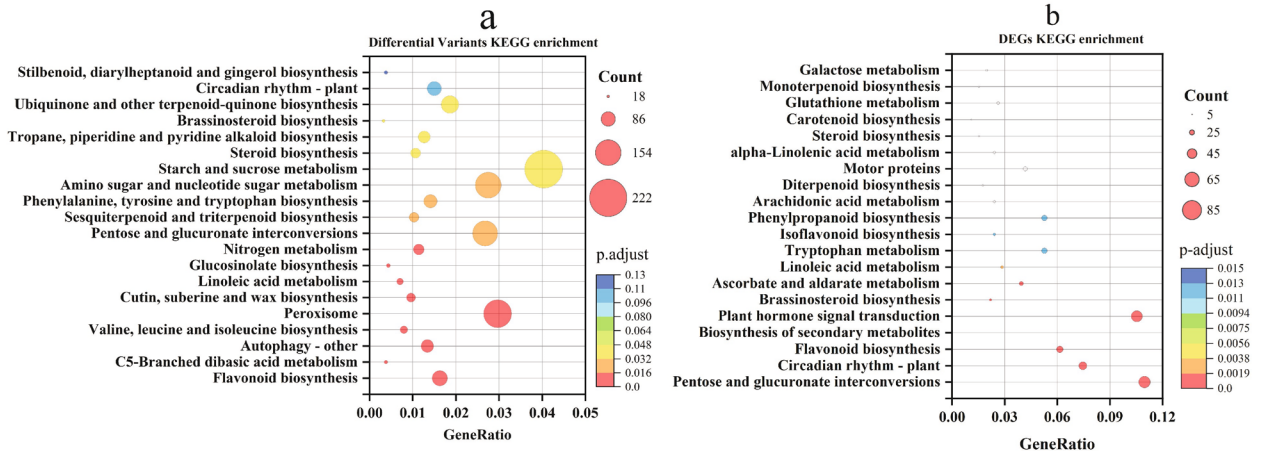

Fig. 4 Differential gene expression and KEGG analysis of the genes related to peanut flowering in BC_1F_3-L_2 and BC_1F_3-L_4. **(a)** Differential variations in KEGG enrichment in the genome of BC_1F_3-L_2 and BC1F_3-L_4. (**b**) KEGG enrichment analysis of the differentially expressed genes in BC_1F_3-L_2 and BC_1F_3-L_4 at flowering

To validate the biological function of *AhMYB51*, we measured various phenotypic traits, including silique length, flowering period, and seed number, in the *AhMYB51-OE* transgenic *A. thaliana* plants. The silique length and seed number in the *AhMYB51*-overexpressing lines of OE-1, OE-2, and OE-3 from the T_3 generation increased by 10% and 20%, respectively, compared to that of wild-type plants (Fig. 8b and c). These results highlight the significant effects of the overexpression of *AhMYB51* on both silique length and seed number in transgenic *A. thaliana* plants (Fig. 9a).However, the transgenic plants had an almost equal flowering period, with similar seed size (Fig. 9b).

Discussion

The regulatory mechanism that controls seed number in peanut

The seed number is a crucial agronomic trait in peanuts, which directly affects their yield, while phytohormones influence various aspects of plant development, including cell metabolism, apical dominance, blooming time, ovule development, pollen germination and seed development [44, 45]. For instance, the development of peanut kernel number involves multiple phytohormones, with brassinosteroids involved in the transport and signaling pathways that play central roles in the generation of seed numbers within pollen cells [46, 47]. In addition to brassinosteroids, cytokinins and gibberellins also regulate

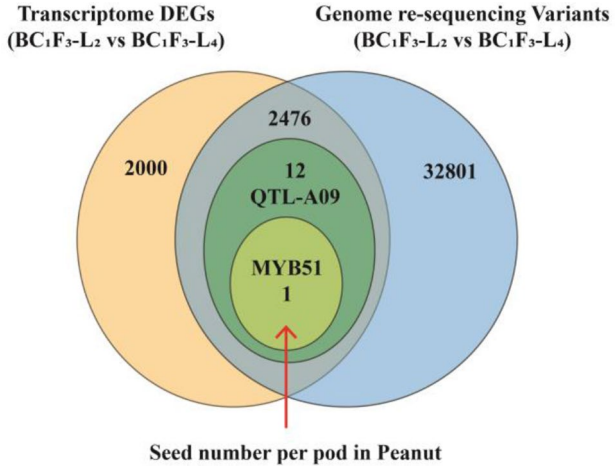

Fig. 5 Venn diagram of the rate of multiple pods per plant differentially expressed RMSP genes and varieties in BC_1F_3-L_2 and BC_1F_3-L_4 in the four comparison groups

the seed number through their effects on pollen activity and ovule number [48]. In this study, the RNA-seq and genome sequencing of the flowers and pegs revealed DEGs and variation in the brassinosteroid pathway, with the DEGs significantly enriched in the metabolic pathways that are known to influence the pod number. This suggests that phytohormones may play a significant role in regulating the seed number in peanuts. Although the regulatory pathways that affect seed number might vary across different modified crops, they share common elements that could be crucial in regulating seed number in various crops. Several TFs involved in the metabolism of brassinosteroids, including MYB, bHLH, NAC, and WRKY, were identified. Notably, the MYB gene family plays a critical role in plant development and metabolism, particularly in the regulation of phytohormones, such as brassinosteroids. These results highlight the potential of multi-omics approaches to improve the agronomic traits, yield, and quality in peanut breeding. Although multi-omics can quickly screen metabolic pathways, the accurate determination of gene function is through reverse genetics. Further experiments are merited to verify the function of this gene in peanuts.

Insights from omics and BSA-seq analysis of the seed number

Map-based cloning is a useful method that can identify the target genes associated with significant agronomic traits and narrow down valuable genes in crops with frequent chromosomal exchanges and rich genomic changes to a specific area [49, 50]. However, in cultivated peanuts with a low amount of genomic variation, its effectiveness is significantly limited. Nevertheless, the availability of high-quality reference genomes for both wild diploid and cultivated peanuts has accelerated transcriptome studies and the development of multi-omics approaches, as well as molecular breeding protocols aimed at improving their yield and quality [51]. These genome sequences are also enabling precise structural and functional genomics research in peanuts. Multi-omics technologies offer an opportunity to elucidate the regulatory mechanisms that underlie seed development in peanuts. Despite this progress, the regulatory networks that govern seed development in peanuts remain unclear. Therefore, this study combined finer transcriptome analysis, bulked segregant analysis sequencing, and backcrossed population analysis to pinpoint the candidate region or genes on the chromosome.

Since the levels of gene expression can influence the efficiency of gene mechanisms, we compared the abundances of gene expression and SNP variations in the 3'-UTR regions of the two candidate regions between the two parental lines and found the *Arahy.04JNDX* gene associated with *qRMSPA09.2*, which encodes a MYB TF. *Arahy.04JNDX* is highly expressed in the multiple-pod-number cultivar SLH, probably owing to an SNP in its 3'-UTR region. This indicates that it could serve as a reliable candidate gene for the RMSP-agronomic characters. We will consider incorporating the dQTG-seq method in future studies to further validate and expand upon our results [52]. Thus, we hypothesize that the candidate interval identified in this study may facilitate this type of research and eventually lead to variation in the seed number. Further studies are merited to verify the function of this candidate interval.

Candidate genes analysis to control the RMSP in peanut

Seed number is a critical factor in plant growth that influences ovule development, pollen germination and architectural formation. Candidate genes are involved in the brassinosteroid signaling pathways. Numerous TFs, including NAC [53], SPL [54], bHLH [55], and MYB [56] and protein kinases, play a significant role in regulating seed number [57]. The MYB gene family is one of the largest gene families in plants that participates in stress resistance, responds to brassinosteroid signals and regulates seed development. For instance, MORE FLORET1, a MYB TF, regulates the development of spikelets in rice (*Oryza sativa* L.) [58]. This study identified 12 candidate genes related to the major QTL *qRMSPA09*, including two uncharacterized genes with missense variant SNPs in their coding sequences and 10 DEGs during flower development. Among these DEGs, *Arahy.04JNDX* was annotated to encode MYB TFs and thus, has emerged as a key candidate.

The level of expression of *Arahy.04JNDX* in SLH was higher than that in JNH3 during the stage of flower development, which suggests that *Arahy.04JNDX* is responsible for seed number in peanuts. Furthermore,

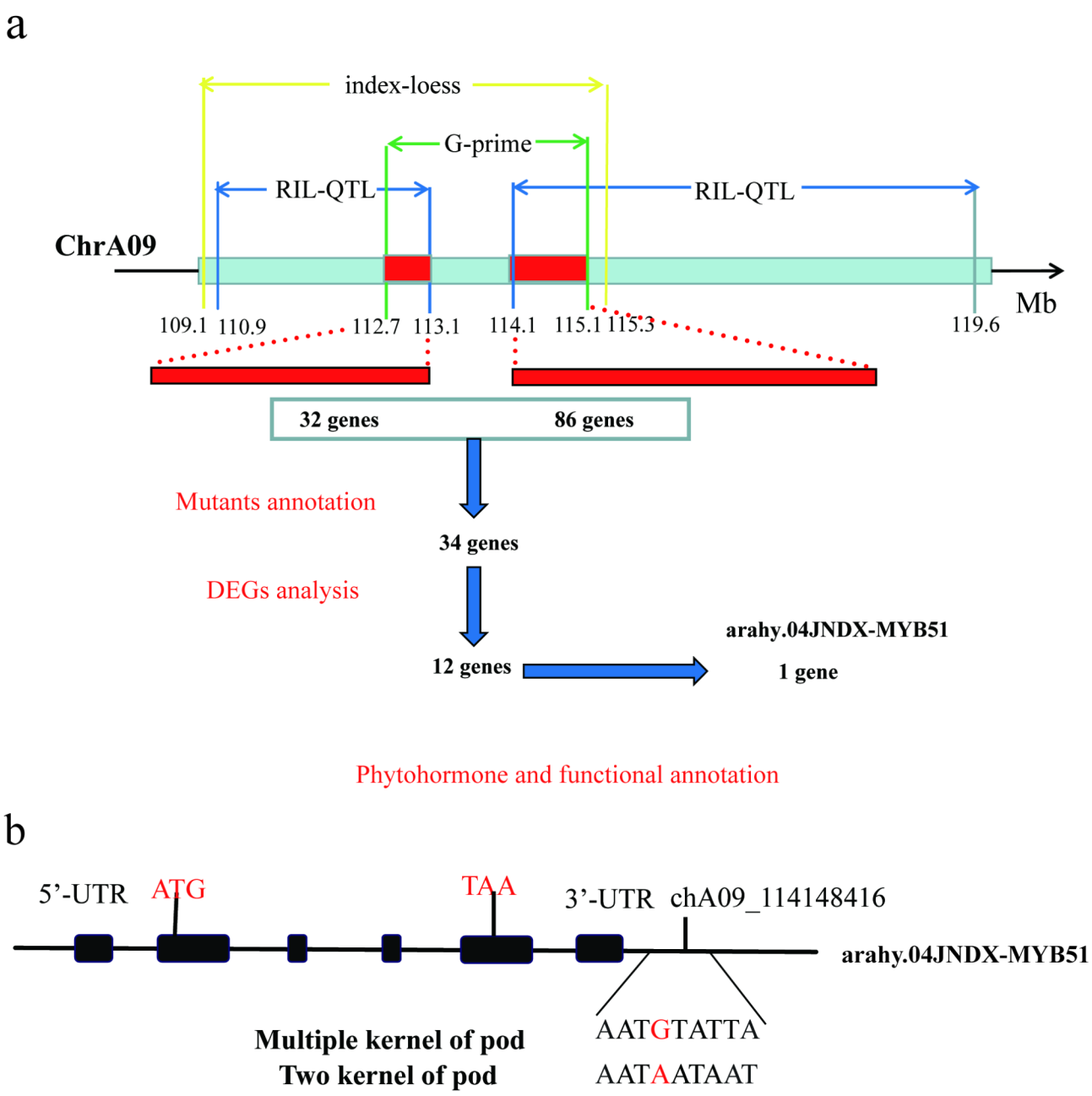

Fig. 6 Identification of the candidate genes by RNA-seq, variation analysis, and co-expression. **(a)** The candidate gene was identified by merging polymorphic loci and differentially expressed genes between SLH and JNH3 analysis. (**b**) The differences in the SNPs of the *Arahy.04JNDX-MYB51* in two and multiple kernels of pods

the high level of expression of *AhMYB51* was consistently maintained in the flowers, which indicates its potential role in pollen activity. This substantially facilitates the mapping of specific QTL loci. Overexpression of the *AhMYB51* gene increased the silique length in transgenic *A. thaliana* and resulted in compact plants. Therefore, *AhMYB51* could possibly affect floral development through the brassinosteroid pathway. However, there is a need to validate the genes through yeast heterozygosity, gene editing, and subcellular localization.

Conclusions

In this study, the secondary population constructed by the cross between line R45 and the parent JNH3 was used to fine map the RMSP of peanut, and the QTL interval length was narrowed from 2.3 Mb and 5.2 Mb to 400 Kb and 600 Kb, respectively. Based on bulked segregant analysis sequencing and multi-omics approaches, 12 DEGs were identified as candidate genes, including three genes with unknown functions and nine genes with known functions. *AhMYB51* was found to regulate the

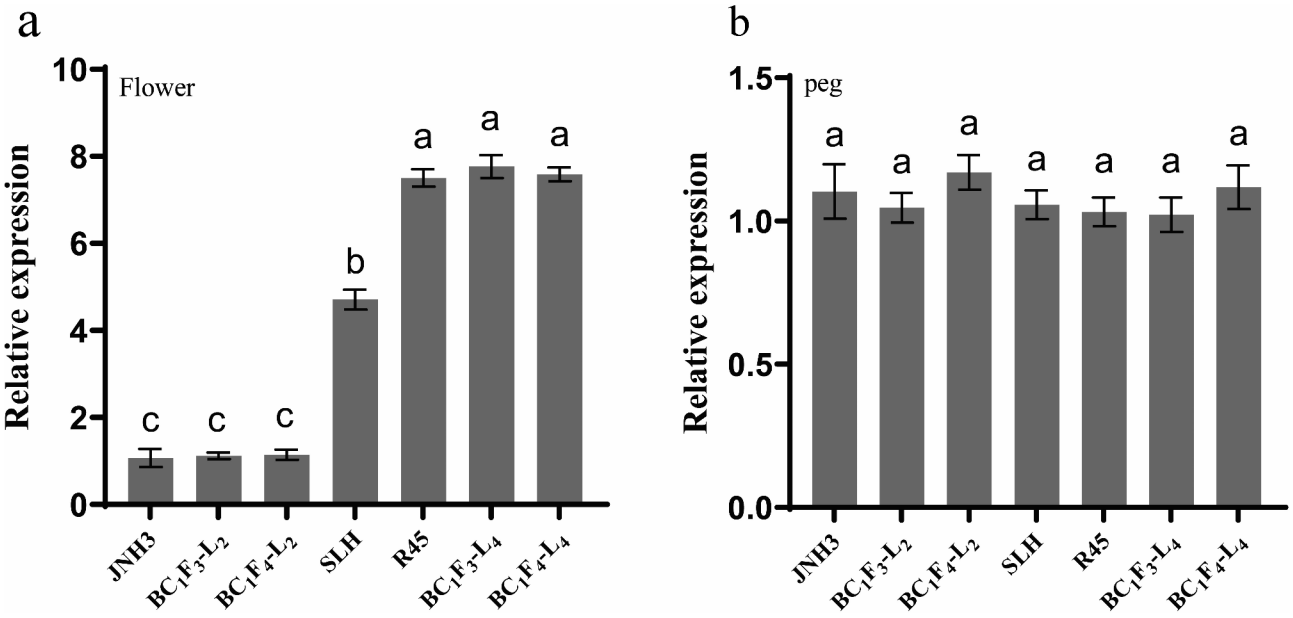

Fig. 7 Quantitative real-time PCR analysis of *AhMYB51* gene in the extreme flowers (**a**) and peg (**b**) of JNH3, BC_1F_3-L_2, BC_1F_4-L_2, SLH, R45, BC_1F_3-L_4 and BC_1F_4-L_4. Error bars ($n = 3$) represent the standard error, while lowercase letters above the bars indicate significant differences ($\alpha = 0.05$, LSD)

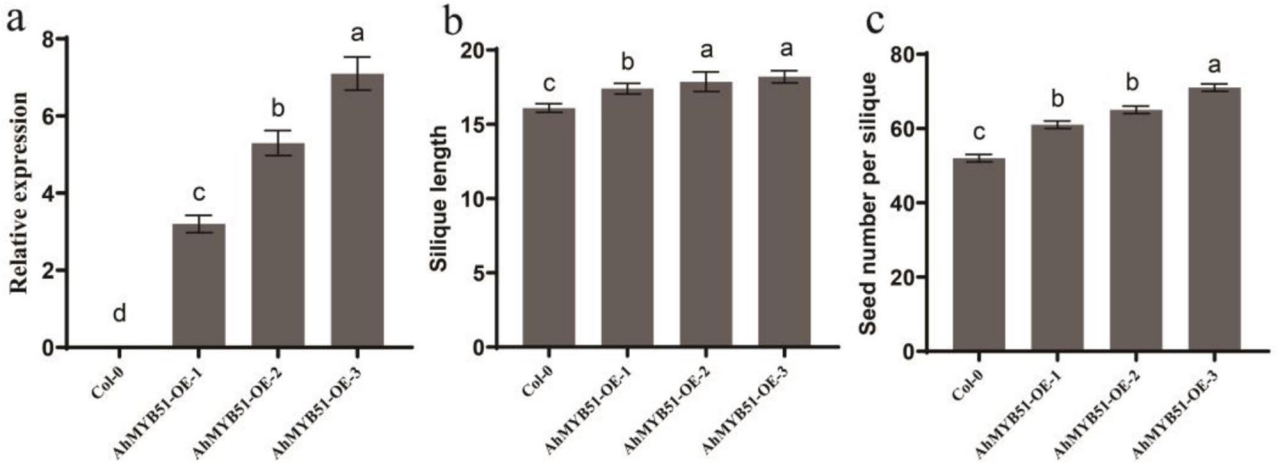

Fig. 8 Relative expression, silique length (SL) and seed number per silique (SNS) of Col-0, OE-1, OE-2, and OE-3. (**a**) The relative expression of *AhMYB51* in the wild type, OE-1, OE-2, and OE-3*Arabidopsis thaliana*. (**b**) The SL of Col-0, OE-1,OE-2 and OE-3. (**c**) The SNS of Col-0,OE-1,OE-2 and OE-3. Error bars ($n = 3$) represent the SD, while lowercase letters above the bars indicate significant differences ($\alpha = 0.05$, LSD) among the treatments

length of *A. thaliana* siliques and seed number and was shown to influence the increased seed number per pod (RMSP) in peanuts.

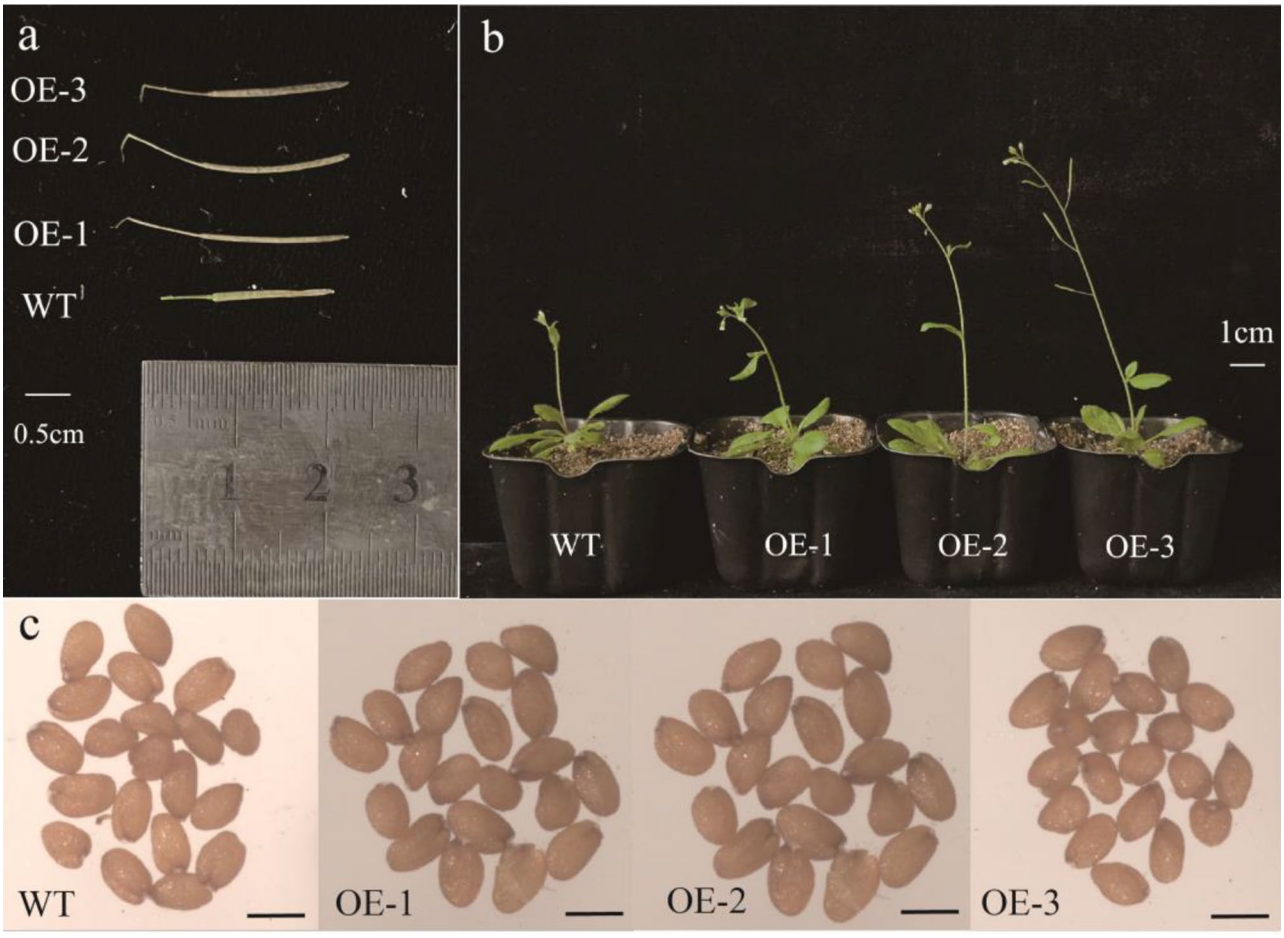

Fig. 9 The growth phenotype of WT and *AhMYB51-OE* transgenic *Arabidopsis thaliana* lines (OE-1, OE-2, OE-3). (**a**) Phenotype in different siliques. The siliques were longer in the OE-1, OE-2 and OE-3 line (Scale bars = 0.5 cm). (**b**) Flowering date and height comparison between the WT and OE. Flower buds were significantly visible at OE-1, OE-2 and OE-3 (Scale bars = 1 cm). (**c**) Comparison of the seed size in WT, OE-1, OE-2 and OE-3. There was an insignificant difference ($P > 0.05$) in seed sizes among the overexpressed lines and WT (Scale bars = 0.8 mm). OE, overexpression; WT, wild type

Abbreviations

BSA	Bulked segregant analysis
WGR	Whole Genome Resequencing
QTL	Quantitative trait locus
PCR	Polymerase chain reaction
qRT-PCR	Quantitative real-time polymerase chain reaction
ORFs	Open reading frames
RMSP	Rate of multiple pods per plant
SL	Seed length
SW	Seed width
PH	Plant height
BRs	Brassinolides
CK	Cytokinin

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11560-7.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
Not applicable.

Author contributions
LFL designed and supervised the project. LL and SLC performed most of the experiments, with the assistance of XKL, YRL, and YHM participated in data analysis. LL drafted the manuscript and revised by LFL. All authors read and approved the final manuscript for publication.

Funding
This study was financially sponsored by the National Natural Science Foundation of China (320720977), the China Agriculture Research System (CARS-13), Hebei Agriculture Research System (HBCT2024040205), S&T Program of Hebei (23567601 H), This research was funded by the Natural Science Foundation of Hebei Province(C2024402049).

Data availability
Sequence data that support the findings of this study have been deposited in the European Nucleotide Archive with the primary accession code GSA: CRA021078.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 14 November 2024 / Accepted: 2 April 2025
Published online: 15 April 2025

References

1. Guo M, Deng L, Gu J, et al. Genome-wide association study and development of molecular markers for yield and quality traits in peanut (*Arachis Hypogaea* L). BMC Plant Biol. 2024;24(1):244.
2. Ma Y, Zhang K, Xu C, et al. Contribution of lipid to the formation of characteristic volatile flavor of peanut oil. Food Chem. 2024;442:138496.
3. Vadez V, Ratnakumar P. High transpiration efficiency increases pod yield under intermittent drought in dry and hot atmospheric conditions but less so under wetter and cooler conditions in groundnut (*Arachis Hypogaea* L). Field Crops Res. 2016;193:16–23.
4. Li G, Guo X, Sun W, et al. Nitrogen application in pod zone improves yield and quality of two peanut cultivars by modulating nitrogen accumulation and metabolism. BMC Plant Biol. 2024;24(1):48.
5. Sun Z, Zheng Z, Qi F, et al. Development and evaluation of the utility of genobaits peanut 40K for a peanut MAGIC population. Mol Breed. 2023;43(10):72.
6. Zhou X, Guo J, Pandey MK, et al. Dissection of the genetic basis of Yield-Related traits in the Chinese peanut Mini-Core collection through Genome-Wide association studies. Front Plant Sci. 2021;12:637284.
7. Chen W, Jiao Y, Cheng L, et al. Quantitative trait locus analysis for pod- and kernel-related traits in the cultivated peanut (*Arachis Hypogaea* L). BMC Genet. 2016;17:25.
8. Chavarro C, Chu Y, Holbrook C, Isleib T, Bertioli D, Hovav R, Butts C, Lamb M, Sorensen RA, Jackson S. OZIAS-AKINS P. Pod and seed trait QTL identification to assist breeding for peanut market preferences. Genes Genomes Genet. 2020;10(7):2297–315.
9. Chen F, Zhang J, Ha X, Ma H. Genome-wide identification and expression analysis of the Auxin-Response factor (*ARF*) gene family in medicago sativa under abiotic stress. BMC Genomics. 2023;24(1):498.
10. Breygina M, Voronkov A, Galin I, et al. Dynamics of endogenous levels and subcellular localization of ABA and cytokinins during pollen germination in Spruce and tobacco. Protoplasma. 2023;260(1):237–48.
11. Jia D, Chen LG, Yin G, et al. Brassinosteroids regulate outer ovule integument growth in part via the control of INNER NO OUTER by BRASSINOZOLE-RESISTANT family transcription factors. J Integr Plant Biol. 2020;62(8):1093–111.
12. Zhou B, Luo Q, Shen Y, et al. Coordinated regulation of vegetative phase change by brassinosteroids and the age pathway in *Arabidopsis*. Nat Commun. 2023;14(1):2608.
13. Kambhampati S, Aznar-Moreno JA, Bailey SR, et al. Temporal changes in metabolism late in seed development affect biomass composition. Plant Physiol. 2021;186(2):874–90.
14. Lv G, Li Y, Wu Z, et al. Maize actin depolymerizing factor 1 (*ZmADF1*) negatively regulates pollen development. Biochem Biophys Res Commun. 2024;703:149637.
15. Braatz J, Harloff HJ, Emrani N, et al. The effect of INDEHISCENT point mutations on silique shatter resistance in oilseed rape (*Brassica napus*). Theor Appl Genet. 2018;131(4):959–71.
16. Feng X, Zhou B, Wu X, et al. Molecular characterization of *SPL* gene family during flower morphogenesis and regulation in blueberry. BMC Plant Biol. 2023;23(1):40.
17. Feng K, Hou XL, Xing GM, et al. Advances in *AP2/ERF* super-family transcription factors in plant. Crit Rev Biotechnol. 2020;40(6):750–76.
18. Zu SH, Jiang YT, Chang JH, Zhang YJ, Xue HW, Lin WH. Interaction of brassinosteroid and cytokinin promotes ovule initiation and increases seed number per silique in *Arabidopsis*. J Integr Plant Biol. 2022;64(3):702–16.
19. Ren Y, Huang Z, Jiang H, et al. A heat stress responsive NAC transcription factor heterodimer plays key roles in rice grain filling. J Exp Bot. 2021;72(8):2947–64.
20. Battat M, Eitan A, Rogachev I, et al. A MYB triad controls primary and phenylpropanoid metabolites for pollen coat patterning. Plant Physiol. 2019;180(1):87–108.
21. Zumajo-Cardona C, Gabrieli F, Anire J, Albertini E, Ezquer I, Colombo L. Evolutionary studies of the bHLH transcription factors belonging to MBW complex: their role in seed development. Ann Bot. 2023;132(3):383–400.
22. Xue H, Zhao K, Zhao K, et al. A near complete genome of Arachis Monticola, an allotetraploid wild peanut. Plant Biotechnol J. 2024;22(8):2110–2.
23. Liu Y, Shao L, Zhou J, et al. Genomic insights into the genetic signatures of selection and seed trait loci in cultivated peanut. J Adv Res. 2022;42:237–48.
24. Zhang X, Pandey MK, Wang J, et al. Chromatin Spatial organization of wild type and mutant peanuts reveals high-resolution genomic architecture and interaction alterations. Genome Biol. 2021;22(1):315.
25. Wang L, Chen H, Zhuang Y, et al. Multiple strategies, including 6 ma methylation, affecting plant alternative splicing in allopolyploid peanut. Plant Biotechnol J. 2024;22(6):1681–702.
26. Hao J, Li K, Cui L, Deng T, Hou Y, Liu R, Yang L, Mu J. Liu F QTL mapping for traits related to seed number per pod in peanut (*Arachis Hypogaea* L). Chin Agricul Sci 2022,55 (13) (In Chinese).
27. Agyenim-Boateng KG, Lu J, Shi Y, Zhang D, Yin X. SRAP analysis of the genetic diversity of wild castor (*Ricinus communis* L.) in South China. PLoS ONE. 2019;14(7):e219667.
28. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
29. Makino J, Ebisuzaki T, Himeno R, Hayashizaki Y. Fast and accurate short-read alignment with hybrid hash-tree data structure. Genomics Inf. 2024;22(1):19.
30. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
31. Cingolani P, Platts A, Wang le. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80–92.
32. Sun G, Zhang C, Shan X, et al. Conjunctive BSA-Seq and BSR-Seq to map the genes of yellow leaf mutations in hot peppers (*Capsicum annuum* L). Genes (Basel). 2024;15(9):1115.
33. Hill JT, Demarest BL, Bisgrove BW, Gorsi B, Su YC, Yost HJ. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. Genome Res. 2013;23(4):687–97.
34. Magwene PM, Willis JH, Kelly JK. The statistics of bulk Segregant analysis using next generation sequencing. PLoS Comput Biol. 2011;7(11):e1002255.
35. Li Z, Chen X, Shi S, et al. DeepBSA: A deep-learning algorithm improves bulked Segregant analysis for dissecting complex traits. Mol Plant. 2022;15(9):1418–27.
36. Rao X. Plant transcriptome analysis with HISAT-StringTie-Ballgown and TopHat-Cufflinks pipelines. Methods Mol Biol. 2024;2812:203–13.
37. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5.
38. Chen S, Lei Y, Xu X, et al. The peanut (*Arachis Hypogaea* L.) gene *AhLPAT2* increases the lipid content of Transgenic *Arabidopsis* seeds. PLoS ONE. 2015;10(8):e0136170.
39. Liu Y, Yang SX, Cheng Y, et al. Production of herbicide-resistant medicinal plant salvia miltiorrhiza transformed with the bar gene. Appl Biochem Biotechnol. 2015;177(7):1456–65.
40. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta delta C(T)) method. Methods. 2001;25(4):402–8.
41. Mansfeld BN, Grumet R. QTLseqr: an R package for bulk Segregant analysis with next-generation sequencing. Plant Genome. 2018;11(2):180006.
42. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, et al. QTL-seq: rapid mapping of quantitativetrait loci in rice by whole genome resequencing of DNA from two bulkedpopulations. Plant J. 2013;74(1):174–83.
43. Huang X, Li Y, Chang Z, et al. Regulation by distinct MYB transcription factors defines the roles of *OsCYP86A9* in anther development and root Suberin deposition. Plant J. 2024;118(6):1972–90.
44. Kawamoto N, Del Carpio DP, Hofmann A, et al. A peptide pair coordinates regular ovule initiation patterns with seed number and fruit size. Curr Biol. 2020;30(22):4352–61.
45. Timofeeva AM, Galyamova MR, Sedykh SE. How do plant Growth-Promoting bacteria use plant hormones to regulate stress reactions?? Plants (Basel). 2024;13(17):2371.
46. Huang HY, Jiang WB, Hu YW, et al. BR signal influences Arabidopsis ovule and seed number through regulating related genes expression by *BZR1*. Mol Plant. 2013;6(2):456–69.
47. Li BF, Yu SX, Hu LQ, et al. Simple culture methods and treatment to study hormonal regulation of ovule development. Front Plant Sci. 2018;9:784.

48. D'Aloia M, Bonhomme D, Bouché F, et al. Cytokinin promotes flowering of Arabidopsis via transcriptional activation of the FT paralogue TSF. Plant J. 2011;65(6):972–9.
49. Ma Z, Miao J, Yu J, et al. The wall-associated kinase GWN1 controls grain weight and grain number in rice. Theor Appl Genet. 2024;137(7):150.
50. Zhang X, Wen H, Wang J, et al. Genetic analysis of QTLs for lysine content in four maize DH populations. BMC Genomics. 2024;25(1):852.
51. Kumar R, Janila P, Vishwakarma MK, et al. Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. Plant Biotechnol J. 2020;18(4):992–1003.
52. Li P, Li G, Zhang YW, Zuo JF, Liu JY, Zhang YM. A combinatorial strategy to identify various types of QTLs for quantitative traits using extreme phenotype individuals in an F2 population. Plant Commun. 2022;3(3):100319.
53. Van Durme M, Olvera-Carrillo Y, Pfeiffer ML, et al. Fertility loss in senescing Arabidopsis ovules is controlled by the maternal sporophyte via a NAC transcription factor triad. Proc Natl Acad Sci U S A. 2023;120(25):e2219868120.
54. Ren Y, Ma R, Fan Y, et al. Genome-wide identification and expression analysis of the SPL transcription factor family and its response to abiotic stress in Quinoa (*Chenopodium quinoa*). BMC Genomics. 2022;23(1):773.
55. Nan GL, Teng C, Fernandes J, O'Connor L, Meyers BC, Walbot V. A cascade of bHLH-regulated pathways programs maize anther development. Plant Cell. 2022;34(4):1207–25.
56. Wang X, Wu R, Shen T, et al. An R2R3-MYB transcription factor OsMYBAS1 promotes seed germination under different sowing depths in Transgenic rice. Plants (Basel). 2022;11(1):139.
57. Xiao W, Hu S, Yu K, et al. Fine-tuning of MPK6/3 phosphorylation by a lectin receptor-like kinase LecRK-VIII.2 regulates seed development. Plant Biotechnol J. 2023;21(12):2414–6.
58. Lu Z, Zhu L, Liang G, et al. *MORE FLORET1* controls anther development by negatively regulating key tapetal genes in both diploid and tetraploid rice. Plant Physiol. 2024;195(3):1981–94.

Publisher's note