



OPEN ACCESS

EDITED BY

Jianping Wang,
University of Florida, United States

REVIEWED BY

Pradeep Sharma,
Indian Institute of Wheat and Barley
Research (ICAR), India
Javaid Akhter Bhat,
Nanjing Agricultural University, China

*CORRESPONDENCE

Jun Wang,
✉ wjsoybean_2008@163.com
Lijuan Qiu,
✉ qjulijuan@caas.cn

[†]These authors have contributed equally
to this work

RECEIVED 08 April 2026
REVISED 09 May 2026
ACCEPTED 18 May 2026
PUBLISHED 04 June 2026

CITATION

Xiong Y, Yu H, Mao A, Jadoon S, Liu Z,
Liu K, Zhang Z, Chen Y, Gyebi-Nimako B,
Qiu L and Wang J (2026) Identification of
QTL associated with pod number
in soybean.
Front. Genet. 17:1850332.
doi: 10.3389/fgene.2026.1850332

COPYRIGHT

© 2026 Xiong, Yu, Mao, Jadoon, Liu, Liu,
Zhang, Chen, Gyebi-Nimako, Qiu and
Wang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Identification of QTL associated with pod number in soybean

Yajun Xiong^{1,2†}, Huan Yu^{1,2†}, Alu Mao^{1,2†}, Sawaira Jadoon¹,
Zhiyu Liu^{1,2}, Kanglin Liu¹, Zhiqing Zhang^{1,2}, Yijie Chen^{1,2},
Bernard Gyebi-Nimako^{1,2}, Lijuan Qiu^{3*} and Jun Wang^{1,4*}

¹The Shennong Laboratory, Zhengzhou, Henan, China, ²Yangtze University, Jingzhou, Hubei, China,
³State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese
Academy of Agricultural Sciences, Beijing, China, ⁴Institute of Crop Molecular Breeding, Henan Academy
of Agricultural Sciences, Zhengzhou, Henan, China

Pod number is an important factor influencing soybean yield. In this study, a recombinant inbred line population derived from the cross between improved cultivar (*Glycine max*) and wild soybean (*Glycine soja*) was subjected to multi-environment phenotypic evaluation for pod number. Quantitative trait locus (QTL) mapping associated with pod number was performed utilizing both linkage analysis (LA) and genome-wide association study (GWAS) by integrating the phenotypic and genotypic data. The results showed that a total of 16 QTLs associated with pod number, notably, overlapping genomic intervals were observed between *qPN10-1* (by linkage analysis) and *qPN10-3* (by EMMAX), as well as between *qPN11-2* (by LA) and *qPN11-3* (by EMMAX), and these intervals were regarded as major-effect Quantitative trait locus regions. Furthermore, two candidate genes (*Glyma.10G206300*, and *Glyma.10G207500*) encode a basic Helix-Loop-Helix (bHLH) transcription factor and an armadillo (ARM)-repeat protein respectively, as determined through linkage disequilibrium analysis, SNP variant analysis, haplotype analysis, and gene functional annotation. These two candidates are possibly involved in the regulation of flower number, flowering time, and pod number (siliques number in *Arabidopsis thaliana*). These findings pave the way for gene cloning related to pod number, and provide novel insights into the genetic architecture underlying pod number variation in soybean.

KEYWORDS

candidate genes, genetic analysis, GWAS, pod number, soybean

1 Introduction

Yield in soybean is a comprehensive trait consisting of 100-seed weight, pod number, and seed number per plant. A strong correlation between pod number and yield has been previously established ($r = 0.63$, $P < 0.01$) (Berhanu et al., 2021). Numerous traits have been reported to contribute to pod number in soybean, such as plant height, branch number and flowering time. Research has shown that increased plant height correlates with a higher pod number in soybean (Liu et al., 2011). A significant positive correlation exists between branch number and pod number ($P < 0.001$) (Xu et al., 2021). Furthermore, flowering time is important for soybean pod number, as late-developing flowers are prone to abortion due to substantial assimilate consumption by early-developing pods (Egli and Bruening, 2000). The flower shedding rate also constitutes a key determinant of pod set (Raza et al., 2019). During the terminal flowering phase, soybean plants not only produce limited flowers but also exhibit flower shedding rates as high as 77.96% (Zhao et al., 2013). The physiological determination of pod number in soybean is influenced by numerous factors, including light exposure, defoliation, abscission, and phytohormone distribution. For instance, shading

from the beginning of flowering (R1) to full maturity (R8) stages can reduce pod number (Jiang and Egli, 1993). Conversely, prolonged light exposure can significantly increase pod number; however, a greater impact has been observed at the late vegetative (V5) stage than at the early pod formation (R3) (Mathew et al., 2000). Defoliation of the lower part of soybean plant, usually caused by dense planting, also decreases pod number due to reduced light interception and diminished assimilation capacity (Board and Tan, 1995). Additionally, cytokinin can increase pod number per plant in soybean, presumably by altering assimilate partitioning and enhancing the “sink strength” of pods to prevent their abscission, thereby boosting yield (Nagel et al., 2001).

Due to its significant role in soybean yield, the genetic basis of pod number has been extensively studied. Using raw mean values across four environments, a total of 10 QTLs for the number of four-seeded pods at lower nodes in soybean were detected, accounting for 0.10%–2.94% of the phenotypic variance explained (PVE) (Raja et al., 2026). In addition, utilizing a RIL population, researchers mapped two quantitative trait loci (QTLs) (*qPN11-1* *qPN6-11*) for pod number in soybean, which are located on chromosomes 11 and 12, respectively (Shim et al., 2017). A total of 602 QTLs associated with pod number were identified by a combination of linkage analysis (LA) and genome-wide association study (GWAS), and 11 candidate genes potentially involved in pod development and growth were identified (Song et al., 2020). Across six different environments, six QTLs associated with soybean pod number were identified in a recombinant inbred line (RIL) population (Charleston × Dongnong 594). Among those QTLs, *qpn-Chr34* accounted for 19.72% of the phenotypic variance explained (PVE) (Zhang et al., 2015). Using an RIL population (Qihuang 34 × Dongsheng 16) and a high-density SLAF (Specific-Locus Amplified Fragment)-based genetic map, two QTLs associated with pod number were identified across two environments. In 2024, a major-effect QTL was detected on chromosome 19, exhibiting an LOD score of 18.16 and explaining 23.02% of the PVE. Furthermore, 17 candidate genes were identified within this QTL region (Zhang et al., 2025). The genetic regulation of pod number in soybean involves genes expressed at specific developmental stages. Indeed, studies on pod number QTLs across different stages have demonstrated that some of these QTLs exhibit stage-specific expression patterns (Zhang et al., 2010; Sun et al., 2006). Several genes influencing pod number have been identified in soybean. For instance, the allelic variant *GmCYP78A10b* exerts an indirect negative regulation on pod number in soybean by conferring a positive effect on seed size/weight traits (Wang et al., 2015). The *GmAPI* gene potentially boosts pod yield in soybean through node number proliferation, suggesting a pathway for improving reproductive efficiency (Chen et al., 2020). *GmFULa* significantly increased soybean pod number ($P < 0.01$) by enhancing carbon assimilation to promote vegetative growth and biomass accumulation (Yue et al., 2021). Additionally, knocking out *GmST05* significantly reduced the number of pods per plant ($P < 0.01$) (Duan et al., 2022), while knockout of *GmEID1* resulted in increased pod number (Qin et al., 2023).

Although previous studies have identified multiple QTLs and genes associated with soybean pod number, its genetic architecture remains poorly understood—maybe due to the narrow genetic basis of cultivated soybean parents used in previous studies. In this study,

we constructed a RIL population from a cross between the high-yielding cultivated soybean cultivar Zhongdou 41 and the wild soybean germplasm ZYD02878 to perform QTL mapping and predict candidate genes for pod number. This approach facilitates the discovery of favorable alleles and broadens the genetic diversity for breeding. These findings will provide an important theoretical foundation for subsequent cloning and functional validation of pod number-related regulatory genes and hold significant practical value for enhancing the efficiency of soybean yield genetic improvement.

2 Materials and methods

2.1 Plant materials and phenotyping

This study used the cultivated soybean Zhongdou 41 as the female parent and wild soybean ZYD02878 as the male parent to construct an advanced RIL population F_7 , F_9 , and F_{11} , with a population size of 365 lines. Zhongdou 41, a main summer soybean cultivar in Hubei Province, was provided by the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences. ZYD02878, a wild soybean accession originating from Ningwu County, Shanxi Province, was obtained from the National Soybean Germplasm Bank. The parental lines exhibited significant differences in traits including seed coat color, number of pods per plant, resistance traits, and climbing growth habit. The population was planted in the summer of 2019 at the Yangtze University Science and Technology Industrial Park in Jingzhou City, Hubei Province (F_7 , 112°15'E, 30°36'N), in the summer of 2020 at the same location (F_9), and in the summer of 2021 in Yucheng County, Shangqiu City, Henan Province (F_{11} , 115°86'E, 34°41'N). These environments were labeled as 19JZ, 20JZ and 21SQ, respectively. Each RIL line was planted randomly with only one biological replicate per site. To prevent intertwining of soybean plants, a row spacing of 100 cm was adopted, and bamboo stakes were erected beside each plant during the early vegetative growth stage to guide the plants to climb vertically along the stakes. Individual plants were harvested separately, which facilitated subsequent trait measurement and prevented sample confusion. After the soybeans reached maturity, three plants were randomly selected from each line, and the number of effective pods was investigated in accordance with the Descriptors and Data Standard for Soybean (*Glycine* spp.) (Qiu and Chang, 2006).

Based on the 3 σ principle and utilizing R packages including but not limited to outliers, car, nortest, fBasics, and Hmisc, outliers were eliminated. Subsequently, arithmetic means, Best Linear Unbiased Estimate (BLUE), and Best Linear Unbiased Predictions (BLUP) for each group of lines within individual environments were calculated using the R packages lmerTest, lme4, reshape2, and lsmeans. Descriptive statistics, correlation analysis, and normality tests were then performed with the lme4 and outliers packages in R.

2.2 Genotyping

Genomic DNA was extracted from young leaves collected at the V5 stage using the CTAB method (Doyle, 1990). Genotyping was conducted using the ZDX1 chip, which was designed based on resequencing data from 2,214 representative soybean accessions on

the Illumina platform, with reference to the Wm82.a2.v1 genome and encompassing 158,959 SNPs (Sun et al., 2022). A total of 158,920 SNPs were obtained and after filtration out SNPs with integrity less than 0.8 and minor allele frequency (MAF) less than 0.05 using PLINK (Purcell et al., 2007), 74,805 SNPs were retained.

2.3 QTL analysis

QTL mapping was conducted in three environments (19JZ, 20JZ, 21SQ). Based on the mixed linear model (MLM), QTL mapping analysis was performed separately using Efficient Mixed-Model Association eXpedited (EMMAX) (Kang et al., 2010) and IciMapping 4.2 software. The kinship matrix for EMMAX analysis was calculated using the emmax-kin-inter64 program. The significance thresholds were set to $-\lg$ (1/total SNPs). The QTL interval was defined as the region extending 100 kb both upstream and downstream of the significant locus. Manhattan and Q-Q plots were generated using the CMplot package in R v 4.3.2. In IciMapping 4.2, the map function was used to construct the genetic map. QTL mapping was performed using the inclusive composite interval mapping (ICIM) method, with a mapping step size of 1.00 cM, a PIN of 0.001, and Permutation tests were employed to empirically determine the LOD threshold, with 1,000 iterations performed for the pod number per plant phenotype in each planting environment at a significance level of $P = 0.05$. Consequently, the genome-wide significance LOD threshold was established as 2.5.

2.4 Candidate gene identification

QTL regions mapped across multiple environments (19JZ, 20JZ, and 21SQ) were selected for candidate gene identification. Candidate gene analysis was performed using four different approaches, including linkage disequilibrium (LD) analysis (Shin et al., 2006), SNP variation, haplotype, and gene functional annotation analyses were conducted. First, Using the Wm82.a2.v1 genome as the reference, significant loci were selected, and the flanking 100 kb regions on each side were defined as candidate intervals for LD analysis. The target intervals were refined based on an r^2 threshold (>0.8). Subsequently, SNP variant analysis was performed to further screen and retain SNPs resulting in nonsynonymous mutations, premature termination (stop-gain), readthrough (stop-loss), and alternative splicing. Haplotype analysis was then performed on genes within the narrowed candidate intervals based on the genotypic and phenotypic data of a NAM population consisting of 2,443 RILs (Song et al., 2024) to examine the significance of phenotypic differences among haplotypes. Phenotypic differences between distinct haplotypes were assessed using the Wilcoxon rank-sum test. Haplotype analysis was conducted using the HaploAssistant package (Yu et al., 2023), and haplotypes with frequencies below 30 were excluded from subsequent analysis due to insufficient statistical power. Finally, gene functional annotation was performed on the candidate genes using data from the SoyBase database (<https://www.soybase.org/>) and the Phytozome database (<https://phytozome.jgi.doe.gov/>; assembly Wm82.a2.v1) to identify genes regulating the number of pods per plant in soybean.

3 Results

3.1 Phenotypic analysis of pod number

Among the phenotypic values (mean, BLUE, and BLUP) obtained from the three environments (19JZ, 20JZ, and 21SQ), the wild parent ZYD02878 (P2) exhibited a consistently higher pod number than the cultivated parent Zhongdou 41 (P1) across all environments, with the most pronounced difference observed in the 21SQ environment (1934.00 for P2 vs. 128.00 for P1). The largest range of pod number variation was observed in the 19JZ environment (37.00–1,827.00), indicating the greatest phenotypic divergence among lines, whereas the smallest range occurred in the 20JZ environment (11.00–1,197.33). The highest mean pod number per plant was recorded in the 21SQ environment, while the 20JZ environment exhibited greater within-population variability. Transgressive segregation was observed for all measured values. Skewness values were all positive (ranging from 0.47 to 0.92), indicating right-skewed distributions across all environments, i.e., most lines produced fewer pods, while a small number of lines exhibited extremely high pod numbers. Kurtosis values were all negative (ranging from -0.07 to -0.38), suggesting relatively dispersed data distributions. (Table 1). The frequency distribution of pod number in the 21SQ environment was approximately normal, whereas the distributions for the 19JZ and 20JZ environments, as well as the BLUE and BLUP values, were right-skewed (Figure 1). Correlation analysis revealed that pod number had significant, albeit moderate, positive correlations among the three environments ($r = 0.49$ – 0.59 ; $P < 0.01$) (Table 2).

3.2 QTL mapping of pod number

A total of 11 QTLs associated with pod number were mapped to chromosomes 3, 5, 6, 10, 11, 12, 14, and 15 using ICIM (Supplementary Figure S1). The LOD scores of these QTLs ranged from 2.68 to 16.91, and the phenotypic variance explained (PVE) ranged from 1.79% to 21.99%. Among them, five QTLs ($qPN5-1$, $qPN6-1$, $qPN10-1$, $qPN11-2$, and $qPN12-1$) were consistently detected across multiple environments, while the remaining QTLs were detected only in single environment (Table 3). Specifically, $qPN5-1$, $qPN10-1$ and $qPN11-2$ were detected as major-effect QTL in multiple environments with phenotypic variance explained exceeding 10% (Figure 2).

EMMAX-GWAS identified five QTLs, four of which were repeatedly detected across multiple environments, with the exception of $qPN3-4$, which was observed only in a single environment (20JZ) (Table 4) (Figure 3). In addition, three QTLs ($qPN3-4$, $qPN10-3$, and $qPN11-3$) overlapped with $qPN3-1$, $qPN10-1$, $qPN10-2$, and $qPN11-2$ mapped by ICIM, those QTLs were co-localized.

3.3 Candidate gene analysis

Among the 16 QTLs associated with pod number, four unique QTL regions ($qPN3-3$, $qPN10-3$, $qPN11-3$, and $qPN19-1$) that were consistently mapped across three environments (19JZ, 20JZ, and 21SQ) were selected as candidate QTL for further analysis. A total of 777 genes were annotated within those four QTL regions. The $qPN3-$

TABLE 1 Phenotypic statistics of pod number.

Environments	Parent		Population							
	P1	P2	Min	Max	Mean	SD	CV	Skew	Kurtosis	P value
19JZ	107.00	1727.00	37.00	1827.00	647.78	423.18	0.65	0.73	−0.24	<0.001
20JZ	49.00	1748.00	11.00	1197.33	362.39	280.00	0.77	0.92	−0.07	<0.001
21SQ	128.00	1934.00	58.33	1833.00	793.66	389.53	0.49	0.47	−0.36	<0.001
BLUE	118.62	1048.09	83.05	1578.05	647.20	337.21	0.52	0.61	−0.38	<0.001
BLUP	−429.93	196.71	−455.02	638.46	−24.04	250.04	—	0.56	−0.37	<0.001

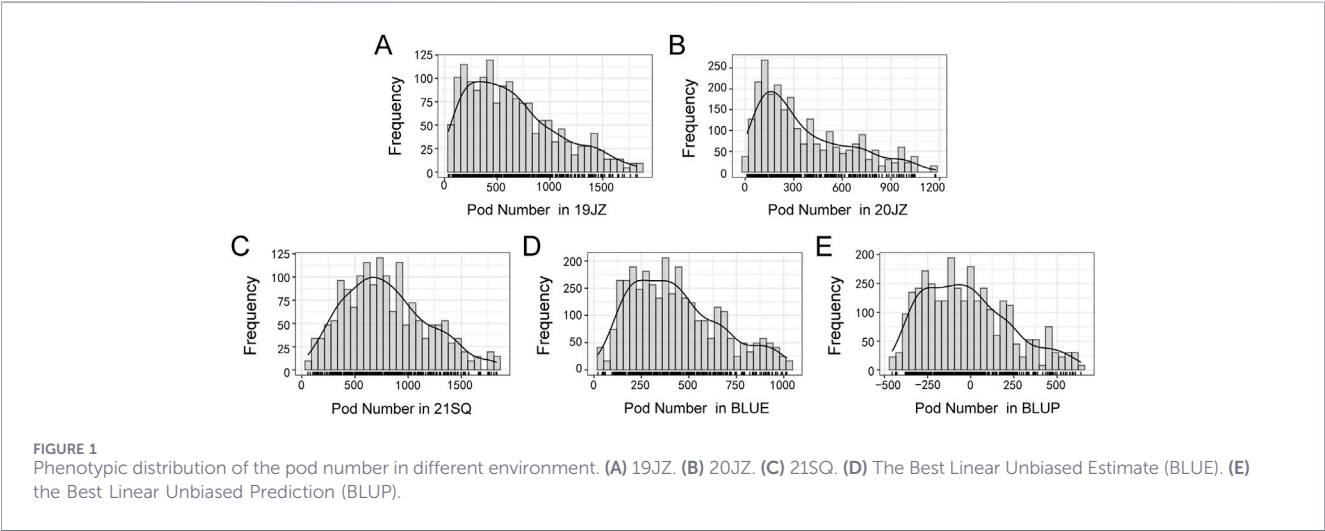


TABLE 2 Correlation analysis of pod number in different environments.

Environment	19JZ	20JZ	21SQ
19JZ	1		
20JZ	0.57**	1	
21SQ	0.49**	0.59**	1

**Represents significance level of 0.01.

3 region was excluded from subsequent analysis because it was defined by only a single SNP that exceeded the significance threshold and therefore failed to satisfy our filtering criteria. The sizes of the remaining three candidate regions (*qPN10-3*, *qPN11-3*, and *qPN19-1*) were 3,640,449 bp, 9,476,310 bp, and 2,101,180 bp, respectively. Following LD analysis, these intervals were refined to 1,697,290 bp, 72,147 bp, and 850,158 bp, which collectively contained 302 genes (Figure 4).

SNP variant analysis identified 86 genes harboring SNP variants, which corresponded to 127 mutation sites within the narrowed intervals. These included 16 nonsynonymous, 41 synonymous, 32 UTR, and 13 upstream/downstream SNPs, respectively (Supplementary Table S1).

In the NAM population, the number of SNPs within the coding sequence (CDS) regions of candidate genes across the three

candidate intervals (*qPN10-3*, *qPN11-3*, and *qPN19-1*) ranged from 1 to 23, encompassing four variant types: nonsynonymous, synonymous, premature stop codon, and UTR mutations (Supplementary Table S2). After filtering low-frequency haplotypes, a maximum of five haplotypes per gene were identified. Hap1 was the predominant haplotype, represented in the highest number of accessions, accounting for 49.57%–95.66% of the total samples. This was followed by Hap2, which accounted for 1.23%–35.2% (Supplementary Table S3). Association analysis between gene haplotypes and pod number revealed significant phenotypic differences among haplotypes for 124 genes (Supplementary Table S4).

Functional annotation identified two key candidate genes potentially regulating pod number (Table 5). *Glyma.10G206300* is an ortholog of *AT1G31050*, encoding a basic helix-loop-helix (bHLH) DNA-binding protein with protein dimerization activity. *Glyma.10G207500* encodes an ARM repeat-containing protein involved in intracellular protein transport, showing orthologs to the *Arabidopsis* gene *AT2G35630*.

The results of the haplotype-phenotype association analysis for the candidate genes showed that both candidate genes, *Glyma.10G206300* and *Glyma.10G207500*, harbored two haplotypes. Highly significant associations (Wilcoxon test, $P < 0.01$) were observed between both Hap1 and Hap2 within each candidate gene and the phenotype (Figures 5A,B).

TABLE 3 QTL mapping of pod number by ICIM.

QTL name	Chromosome	Start	End	ADD	PVE/%	LOD value	Environment
<i>qPN3-1</i>	3	36055334	36093981	−35.37	1.97	2.69	BLUP
<i>qPN3-2</i>	3	36717177	36793579	−65.80	2.79	2.91	21SQ
<i>qPN5-1</i>	5	37167537	37481276	70.92	18.16	3.39	20JZ, BLUE, BLUP
<i>qPN6-1</i>	6	3709181	4138913	51.43	2.32	2.82	21SQ, BLUE, BLUP
<i>qPN10-1</i>	10	45243194	45322107	129.70	21.99	16.91	20JZ,21SQ, BLUE
<i>qPN10-2</i>	10	46156099	46350300	119.69	5.48	5.01	19JZ
<i>qPN11-1</i>	11	31891786	31901505	−148.50	8.49	5.18	19JZ
<i>qPN11-2</i>	11	10904384	11099081	−91.30	13.36	14.83	21SQ,19JZ, BLUE, BLUP
<i>qPN12-1</i>	12	5475077	5817926	6.75	7.03	9.11	21SQ, BLUP
<i>qPN14-1</i>	14	5929565	6943720	−71.89	3.39	2.68	21SQ
<i>qPN15-1</i>	15	10851120	10873304	117.14	7.91	2.75	21SQ

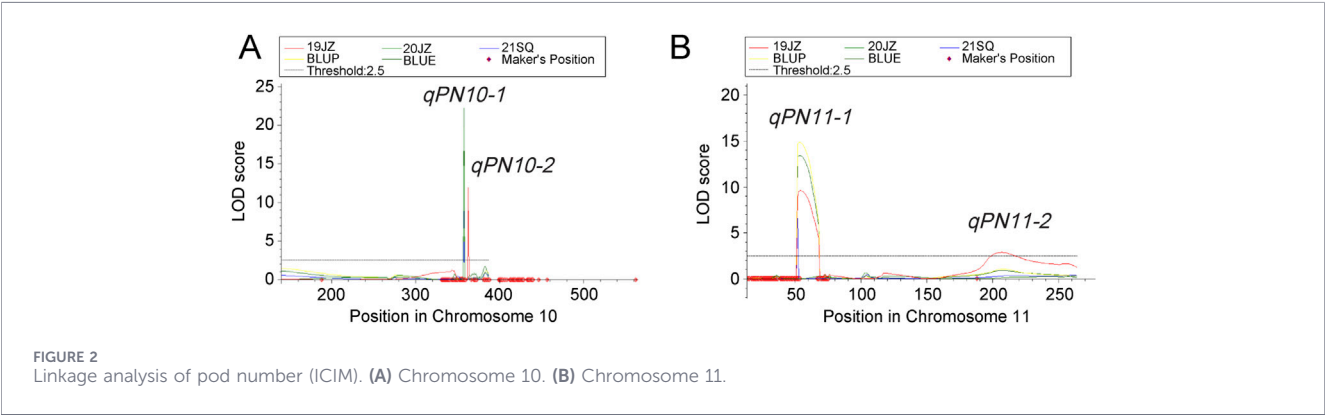


TABLE 4 QTL mapping of pod number by EMMAX.

QTL name	Chromosome	Start/	End	Number of genes	Environment
<i>qPN3-3</i>	3	40420694	40771281	50	19JZ,20JZ,21SQ, BLUE, BLUP
<i>qPN3-4</i>	3	35914456	36409036	49	20JZ
<i>qPN10-3</i>	10	43350192	46990641	391	19JZ,20JZ,21SQ, BLUE, BLUP
<i>qPN11-3</i>	11	10715099	11238789	54	19JZ,20JZ,21SQ, BLUE, BLUP
<i>qPN19-1</i>	19	43837861	46274815	245	19JZ,21SQ, BLUE, BLUP

4 Discussion

In this study, a RIL population derived from a cross between the wild soybean ZYD02878 and the cultivated soybean Zhongdou 41 was subjected to multi-environment phenotypic evaluation. By integrating LA and GWAS, candidate genes were identified, and their potential roles in flower and fruit development were validated through functional analysis of their homologous genes in *Arabidopsis thaliana*. This study broadens the genetic basis of soybean, overcoming the limitations imposed by the narrow genetic background of

cultivated soybean parents used in previous studies. The findings provide a theoretical foundation and practical tools for marker-assisted selection, gene editing, and high-yield soybean breeding targeting pod number, thereby holding significant theoretical and practical value for improving soybean yield and ensuring food security.

Wild soybean and cultivated soybean are closely related without reproductive isolation. Numerous studies have utilized wild soybeans for QTL mapping and gene mining related to agronomic traits (Chen et al., 2023; Li et al., 2022; Wang et al., 2024; Yang et al., 2019; Yang et al., 2025). Previous research has

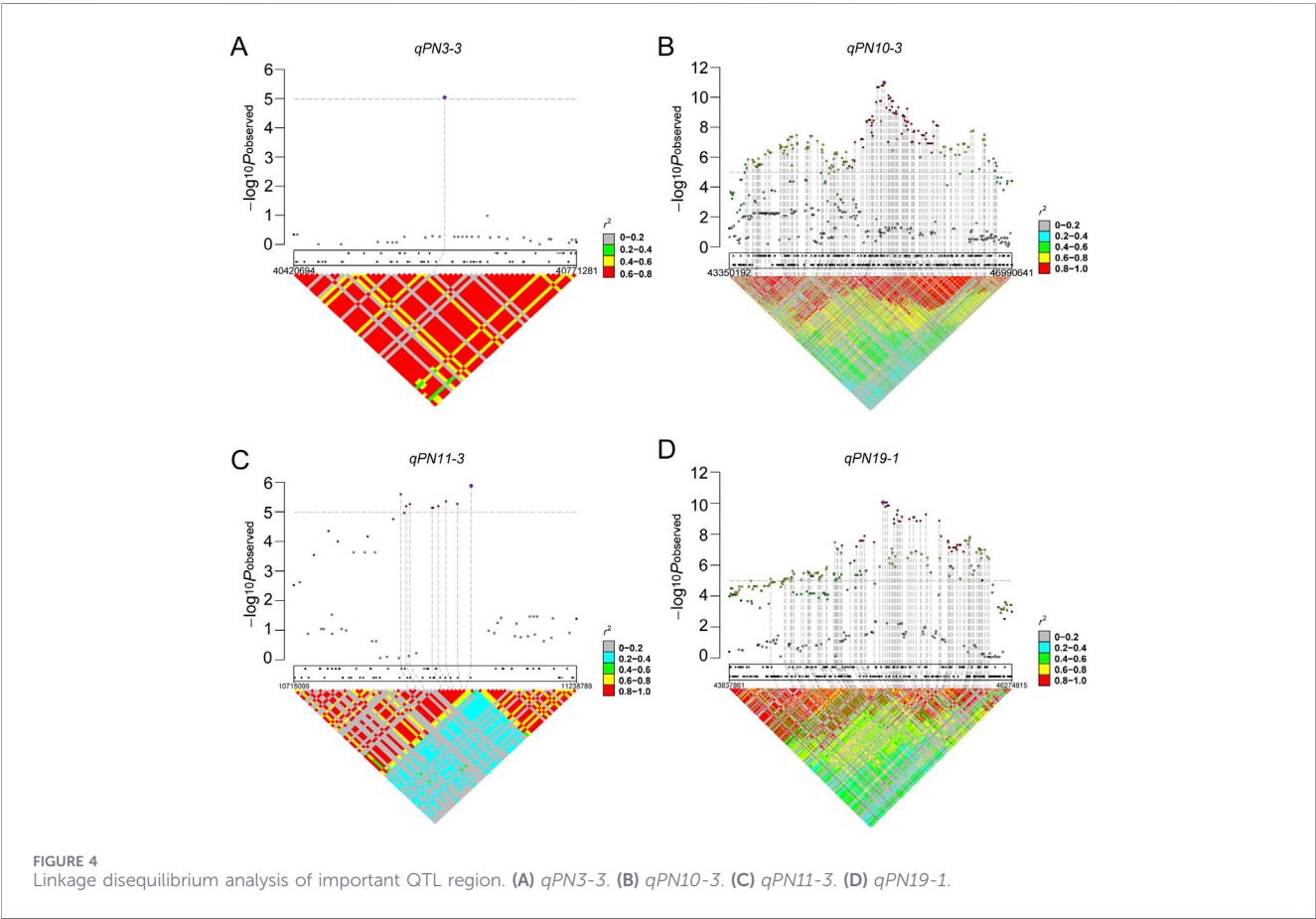
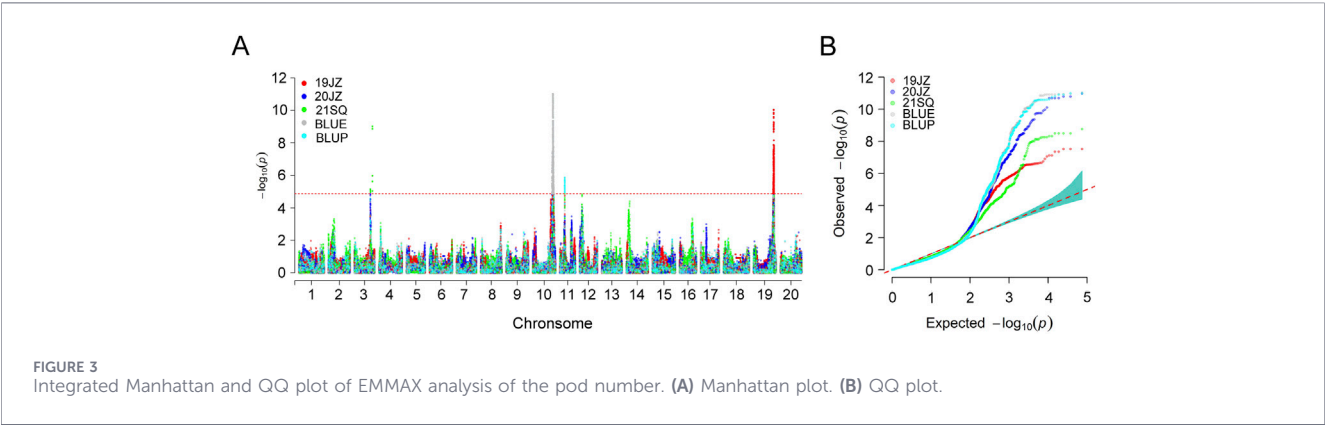


TABLE 5 Functional annotation of candidate genes.

Gene	Homologue	Annotation
<i>Glyma.10G206300</i>	<i>AT1G31050</i>	Basic helix-loop-helix (<i>bHLH</i>) DNA-binding superfamily protein
<i>Glyma.10G207500</i>	<i>AT2G35630</i>	ARM repeat superfamily protein

shown transgressive segregation for quality and yield traits in hybrid offspring populations of wild and cultivated soybeans (Wang et al., 2024; Yang et al., 2025), which is consistent with the findings in this study.

In this study, 16 QTLs associated with pod number were identified. Among these, *qPN10-1* colocalizes with *qPNLD-10-3*, which is responsible for four-seed pods in the lower part of soybean plants. This QTL was identified using a cultivated

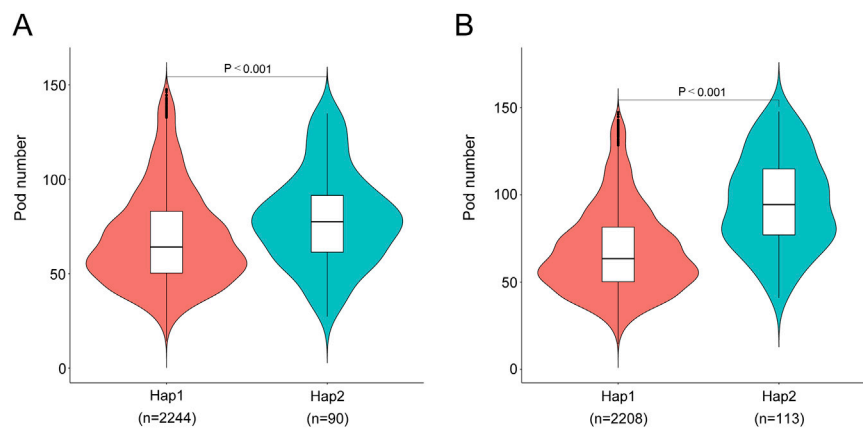


FIGURE 5
Phenotypic comparison of haplotypes of candidate gene. (A) *Glyma.10G206300*. (B) *Glyma.10G207500*.

soybean-derived RIL population for pod number-related QTL analysis (Song et al., 2020). Additionally, *qPN19-1* coincides with a QTL (*qBR19-1*) associated with soybean branching (Shim et al., 2017).

This study identified two candidate genes potentially regulating pod number in soybean. *Glyma.10G206300* encodes a basic helix-loop-helix (bHLH) protein. In *Arabidopsis*, bHLH transcription factors play significant roles in seed germination, flowering time, and abiotic stress response. For example, the *Arabidopsis* gene ortholog *AT1G31050*, which encodes a bHLH transcription factor, has been shown to influence multiple developmental processes: knockout of this gene promotes primary root growth, increases flower number, shortens internode length, and delays flowering (Makkena and Lamb, 2013). *GmbHLH98* is upregulated under drought and salt stress, and its overexpression enhances seed protein content in transgenic plants, demonstrating the potential of bHLH transcription factors in improving both yield and quality in soybean (Wu et al., 2025). Another *Arabidopsis* homologue of *Glyma.10G206300*, *SPATULA* (*SPT*), also supports its potential role in fruit development: loss-of-function *SPT* mutants produce significantly fewer siliques than wild-type plants (Groszmann et al., 2008). These functional insights suggest that *Glyma.10G206300* may be a key regulator of pod number in soybean. *Glyma.10G207500* encodes an ARM repeat protein. In soybean, *GmDFB1*—another ARM repeat protein—was reported to interact with *GmSK17* and participate in siRNA suppression by blocking *GmAGO1*-mediated mRNA cleavage, ultimately leading to reduced miRNA accumulation. Phenotypically, the *gmdfb1* mutant exhibits severe floral structural defects that result in a complete absence of pod formation (Li et al., 2024). The *Arabidopsis* homologue of *Glyma.10G207500*, *MOR1* (*AT2G35630*), shares 85.7% sequence identity. Knockout mutants of *MOR1* exhibit dwarfism, failure to flower, and lethality at temperatures above 28 °C, although normal growth, flowering, and ploidy are restored when plants are returned to optimal temperatures (Whittington et al., 2001). Additionally, *mor1* mutants display temperature-sensitive anther indehiscence and impaired flowering (Liu and Tan,

2022). Therefore, both candidates, *Glyma.10G207500* and *Glyma.10G206300*, were likely to play significant roles in pod number formation, warranted further investigation.

In this study, a recombinant inbred line (RIL) population derived from the cross between the cultivated soybean (Zhongdou 41) and wild soybean (ZYD02878) was constructed. Combined with multi-environment phenotypic evaluation and multiple QTL mapping approaches, a total of 16 QTLs associated with pod number per plant in soybean were identified. Subsequently, by employing LD analysis, SNP variant analysis, haplotype analysis, and gene functional annotation, two candidate genes, *Glyma.10G206300* and *Glyma.10G207500*, were screened out as key regulators of soybean pod number. These two genes encode a bHLH transcription factor and an ARM repeat protein, respectively. This discovery not only provides new genetic resources and molecular markers for high-yield molecular breeding in soybean and broadens the genetic basis for breeding, but also lays a theoretical foundation for the subsequent cloning and functional validation of pod number-related genes. Furthermore, it holds significant theoretical and practical importance for improving the efficiency of genetic enhancement of soybean yield and ensuring the security of soybean production.

5 Conclusion

Utilizing a RIL population derived from Zhongdou 41 × ZYD02878, we performed QTL mapping and candidate gene prediction were performed based on mean pod number per plant across three environments, along with BLUE and BLUP values. In this study, 16 QTLs were identified, among which *qPN3-2*, *qPN10-1*, and *qPN11-1* were consistently detected by both methods. Notably *qPN10-1* and *qPN19-1* co-localized with previously reported loci and were detected across multiple environments. Additionally, two candidate genes, *Glyma.10G206300* and *Glyma.10G207500*, were predicted to indirectly affect soybean pod number by mediating flowering time and flower number. These findings establish a foundation for the cloning and functional characterization of pod number-associated genes in soybean.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found in the Soybean Functional Genomics and Breeding repository at <https://sfgb.rmbreeding.cn/about/data>.

Author contributions

YX: Formal Analysis, Writing – review and editing, Methodology, Writing – original draft. HY: Investigation, Writing – review and editing. AM: Writing – review and editing, Investigation. SJ: Data curation, Writing – review and editing. ZL: Writing – review and editing, Data curation. KL: Writing – review and editing, Data curation. ZZ: Data curation, Writing – review and editing. YC: Data curation, Writing – review and editing. BG-N: Writing – review and editing, Data curation. LQ: Writing – review and editing, Conceptualization, Supervision, Funding acquisition, Resources, Project administration. JW: Project administration, Resources, Writing – review and editing, Supervision, Funding acquisition, Conceptualization.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by Key Research Project of the Shennong Laboratory (SN01-2024-01) and Major Agricultural Science and Technology Project of the Ministry of Agriculture and Rural Affairs (Grant No: NK202308010302). Thanks for the support of China-Uruguay Joint Laboratory on Soybean Research and Innovation, and National Key R&D Program of China (2022YFE0203300).

References

- Berhanu, H., Tesso, B., and Lule, D. (2021). Correlation and path coefficient analysis for seed yield and yield related traits in soybean (*Glycine max* (L.) Merr.) genotypes. *Plant* 9 (4), 106. doi:10.11648/j.plant.20210904.15
- Board, J. E., and Tan, Q. (1995). Assimilatory capacity effects on soybean yield components and pod number. *Crop Sci.* 35 (3), 846–851. doi:10.2135/cropsci1995.0011183X003500030035x
- Chen, L., Nan, H., Kong, L., Yue, L., Yang, H., Zhao, Q., et al. (2020). Soybean *API* homologs control flowering time and plant height. *J. Integr. Plant Biol.* 62 (12), 1868–1879. doi:10.1111/jipb.12988
- Chen, Y., Xiong, Y., Hong, H., Li, G., Gao, J., Guo, Q., et al. (2023). Genetic dissection of and genomic selection for seed weight, pod length, and pod width in soybean. *Crop J.* 11 (3), 832–841. doi:10.1016/j.cj.2022.11.006
- Doyle, J. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12 (1), 13–15.
- Duan, Z., Zhang, M., Zhang, Z., Liang, S., Fan, L., Yang, X., et al. (2022). Natural allelic variation of *GmSTO5* controlling seed size and quality in soybean. *Plant Biotechnol. J.* 20 (9), 1807–1818. doi:10.1111/pbi.13865
- Egli, D. B., and Bruening, W. P. (2000). Potential of early-maturing soybean cultivars in late plantings. *Agron. J.* 92 (3), 532–537. doi:10.2134/agronj2000.923532x
- Groszmann, M., Paicu, T., and Smyth, D. R. (2008). Functional domains of *SPATULA*, a bHLH transcription factor involved in carpel and fruit development in *Arabidopsis*. *Plant J.* 55 (1), 40–52. doi:10.1111/j.1365-313X.2008.03469.x
- Jiang, H., and Egli, D. B. (1993). Shade induced changes in flower and pod number and flower and fruit abscission in soybean. *Agron. J.* 85 (2), 221–225. doi:10.2134/agronj1993.00021962008500020011x
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42 (4), 348–354. doi:10.1038/ng.548
- Li, J., Zhang, Y., Ma, R., Huang, W., Hou, J., Fang, C., et al. (2022). Identification of *ST1* reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnol. J.* 20 (6), 1110–1121. doi:10.1111/pbi.13791
- Li, J., Zhang, W., Lu, Q., Sun, J., Cheng, C., Huang, S., et al. (2024). *GmDFB1*, an ARM-repeat superfamily protein, regulates floral organ identity through repressing siRNA- and miRNA-mediated gene silencing in soybean. *J. Integr. Plant Biol.* 66 (8), 1620–1638. doi:10.1111/jipb.13709
- Liu, X., and Tan, X. (2022). Mutation in *Arabidopsis MOR1* gene impairs endocytosis in stamen filament cells and results in anther indehiscence. *Plant Growth Regul.* 96 (2), 303–314. doi:10.1007/s10725-021-00777-7
- Liu, W., Kim, M. Y., Van, K., Li, Y., Li, B., Liu, X., et al. (2011). QTL identification of yield-related traits and their association with flowering and maturity in soybean. *J. Crop Sci. Biotechnol.* 14, 65–70. doi:10.1007/s12892-010-0115-7
- Makkena, S., and Lamb, R. S. (2013). The bHLH transcription factor *SPATULA* regulates root growth by controlling the size of the root meristem. *BMC Plant Biol.* 13 (1), 1. doi:10.1186/1471-2229-13-1
- Mathew, J. P., Herbert, S. J., Zhang, S., Rautenkranz, A. A. F., and Litchfield, G. V. (2000). Differential response of soybean yield components to the timing of light enrichment. *Agron. J.* 92 (6), 1156–1161. doi:10.2134/agronj2000.9261156x
- Nagel, L., Brewster, R., Riedell, W. E., and Reese, R. N. (2001). Cytokinin regulation of flower and pod set in soybeans (*Glycine max* (L.) Merr.). *Ann. Bot.* 88 (1), 27–31. doi:10.1006/anbo.2001.1423

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2026.1850332/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

Linkage analysis of pod number (ICIM).

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81 (3), 559–575. doi:10.1086/519795
- Qin, C., Li, H., Zhang, S., Lin, X., Jia, Z., Zhao, F., et al. (2023). *GmEID1* modulates light signaling through the evening complex to control flowering time and yield in soybean. *Proc. Natl. Acad. Sci.* 120 (15), e2212468120. doi:10.1073/pnas.2212468120
- Qiu, L., and Chang, R. (2006). *Descriptors and Data Standard for Soybean* (Glycine spp.). Beijing: China Agriculture Press.
- Raja, R., Huang, Y., Ning, S., Hu, B., Siyal, M., Li, W.-X., et al. (2026). Genetic basis analysis for candidate QTLs and functional genes controlling four-seeded pods at lower-node in soybean (*Glycine max*) plant. *Plants* 15 (6), 966. doi:10.3390/plants15060966
- Raza, M., Feng, L., Werf, W., Iqbal, N., Khalid, M., Chen, Y., et al. (2019). Maize leaf-removal: a new agronomic approach to increase dry matter, flower number and seed-yield of soybean in maize soybean relay intercropping system. *Sci. Rep.* 9 (1), 13453. doi:10.1038/s41598-019-49858-8
- Shim, S., Kim, M. Y., Ha, J., Lee, Y.-H., and Lee, S.-H. (2017). Identification of QTLs for branching in soybean (*Glycine max* (L.) Merr.). *Euphytica* 213 (9), 225. doi:10.1007/s10681-017-2016-z
- Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* 16, 1–9. doi:10.18637/jss.v016.c03
- Song, J., Sun, X., Zhang, K., Liu, S., Wang, J., Yang, C., et al. (2020). Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. *Mol. Breed.* 40, 1–14. doi:10.1007/s11032-020-01140-w
- Song, J., Xiong, Y., Chen, Y., Xu, R., Liu, K., Guo, Q., et al. (2024). Genetic analysis of seed coat and flower color based on a soybean nested association mapping population. *Acta Agron. Sin.* 50 (03), 556–575. doi:10.3724/sp.j.1006.2024.34094
- Sun, D., Li, W., Zhang, Z., Chen, Q., Ning, H., Qiu, L., et al. (2006). Quantitative trait loci analysis for the developmental behavior of Soybean (*Glycine max* (L.) Merr.). *Theor. Appl. Genet.* 112 (4), 665–673. doi:10.1007/s00122-005-0169-y
- Sun, R., Sun, B., Tian, Y., Su, S., Zhang, Y., Zhang, W., et al. (2022). Dissection of the practical soybean breeding pipeline by developing ZDX1, a high-throughput functional array. *Theor. Appl. Genet.* 135 (4), 1413–1427. doi:10.1007/s00122-022-04043-w
- Wang, X., Li, Y., Zhang, H., Sun, G., Zhang, W., and Qiu, L. (2015). Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean. *Mol. Biol. Rep.* 42 (2), 489–496. doi:10.1007/s11033-014-3792-3
- Wang, J., Yang, Q., Chen, Y., Liu, K., Zhang, Z., Xiong, Y., et al. (2024). QTL mapping and genomic selection of stem and branch diameter in soybean (*Glycine max* L. Merr.). *Front. Plant Sci.* 15, 1388365. doi:10.3389/fpls.2024.1388365
- Whittington, A. T., Vugrek, O., Wei, K. J., Hasenbein, N. G., Sugimoto, K., Rashbrooke, M. C., et al. (2001). *MOR1* is essential for organizing cortical microtubules in plants. *Nature* 411 (6837), 610–613. doi:10.1038/35079128
- Wu, N., Jiang, T., Feng, Y., and Yuan, M. (2025). Genome-wide identification and expression analysis of soybean bHLH transcription factor and its molecular mechanism on grain protein synthesis. *Front. Plant Sci.* 16, 1481565. doi:10.3389/fpls.2025.1481565
- Xu, C., Li, R., Song, W., Wu, T., Shi, S., Hu, S., et al. (2021). Responses of branch number and yield component of soybean cultivars tested in different planting densities. *Agriculture* 11 (1), 69. doi:10.3390/agriculture11010069
- Yang, H., Wang, W., He, Q., Xiang, S., Tian, D., Zhao, T., et al. (2019). Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. *Theor. Appl. Genet.* 132 (10), 2793–2807. doi:10.1007/s00122-019-03388-z
- Yang, Q., Wang, J., Xiong, Y., Mao, A., Zhang, Z., Chen, Y., et al. (2025). Identification of QTL for branch traits in soybean (*Glycine max* (L.) Merr.) and its application in genomic selection. *Front. Genet.* 16, 1484146. doi:10.3389/fgene.2025.1484146
- Yu, H., Chen, Y. J., Liu, K., and Zhang, F. (2023). Development and application of a haplotype analysis R package (HaploAssistant). *Mol. Plant Breed.* 1–11.
- Yue, Y., Sun, S., Li, J., Yu, H., Wu, H., Sun, B., et al. (2021). *GmFULa* improves soybean yield by enhancing carbon assimilation without altering flowering time or maturity. *Plant Cell Rep.* 40 (10), 1875–1888. doi:10.1007/s00299-021-02752-y
- Zhang, D., Cheng, H., Wang, H., Zhang, H., Liu, C., and Yu, D. (2010). Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* (L.) Merr.). *J. Genet. Genomics* 37 (8), 545–556. doi:10.1016/S1673-8527(09)60074-6
- Zhang, H., Hao, D., Siteo, H. M., Yin, Z., Hu, Z., Zhang, G., et al. (2015). Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max* (L.) Merr.) by association analysis across multiple environments. *Plant Breed.* 134 (5), 564–572. doi:10.1111/pbr.12305
- Zhang, C., Zha, B., Yuan, R., Zhao, K., Sun, J., Liu, X., et al. (2025). Identification of quantitative trait loci for node number, pod number, and seed number in soybean. *Int. J. Mol. Sci.* 26 (5), 2300. doi:10.3390/ijms26052300
- Zhao, S., Tang, X., Zhao, X., Feng, Y., Zhao, C., and Zhang, M. (2013). Observation and research on the temporal and spatial distribution of flowering and flowers dropping of soybean. *Sci. Agric. Sin.* 46 (8), 1543–1554. (in Chinese). doi:10.3864/j.issn.0578-1752.2013.08.003