agronomy

MDPI

Article

Multi-Environment Genome-Wide Association Analysis Reveals Stable Genetic Loci for Soybean Yield Component Traits

Ruonan Du [1,†], Qiang Gao [2,†], Hui Jin [1], Jumei Zhang [1], Yordan Dimitrov [1], Haibin Zhao [1], Yu-E Wu [1], Danna Chang [3], Chunwei Zhou [4], Zhuo Li [5], Xue Yang [1,*] and Rui Zhang [1,*]

[1] Institute of Forage and Grassland Sciences, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China; drn0713@126.com (R.D.); jinhuicaas@126.com (H.J.); zjm312@aliyun.com (J.Z.); ydtsvetkov@163.com (Y.D.); hbzhao0617@163.com (H.Z.); lunawye@163.com (Y.-E.W.)
[2] Horticultural Branch, Heilongjiang Academy of Agricultural Sciences, Harbin 150040, China; 18500654770@163.com
[3] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China; changdanna@caas.cn
[4] Institute of Rural Revitalization Science and Technology, Heilongjiang Academy of Agricultural Sciences, Harbin 154005, China; zhouchunwei_0405@163.com
[5] Department of Architectural Engineering, Heilongjiang Oriental University, Harbin 150076, China; 2022112206@nefu.edu.cn
* Correspondence: yxflax@126.com (X.Y.); zr0705@126.com (R.Z.)
† These authors contributed equally to this work.

Abstract

Soybean yield is governed by key agronomic traits such as main stem node number, lowest pod height, and branch number. These polygenic traits exhibit substantial environmental variation, and the instability of associated genetic loci identified in single-environment studies constrains their application in molecular breeding. To uncover their stable genetic basis, we conducted multi-environment phenotyping over three years and six locations using a panel of 320 soybean accessions. The Best Linear Unbiased Prediction (BLUP) model was employed to integrate phenotypic data and control for genotype-by-environment interactions. Subsequently, genome-wide association studies (GWAS) were performed using BLUP-integrated phenotypic values to capture stable genetic effects, with optimal model using the Mixed Linear Model (MLM). As a result, a total of 22 significant single-nucleotide polymorphism (SNP) loci were identified. Among these, 10, 7, and 5 significant loci were associated with main stem node number, lowest pod height, and branch number, respectively, representing stable genetic signals across multiple environments. These loci primarily clustered on chromosomes 19, 8, and 18. Within these associated regions, we predicted eight high-confidence candidate genes. Functional annotation revealed that these genes are significantly enriched in pathways related to cell wall biosynthesis, energy metabolism, and stress response. This study provides effective genomic resources derived from a multi-environment GWAS framework. These stable loci and candidate genes directly facilitate the molecular breeding of soybean varieties with improved yield-related traits and environmental adaptability.

Keywords: G×E interaction; BLUP; plant architecture; marker-assisted selection; multi-environment trial

Academic Editor: Dilip R. Panthee

Received: 13 April 2026
Revised: 30 April 2026
Accepted: 5 May 2026
Published: 7 May 2026

1. Introduction

Soybean (*Glycine max* L.), a legume of Asian origin, is a globally important economic crop and oilseed; many cultivars exhibit short-day photoperiod sensitivity, although

photoperiod-insensitive accessions also exist [1]. Domesticated in China approximately 3000–5000 years ago [2,3], this crop requires sustained yield improvement through genetic breeding to meet rising global demand. Soybean yield is a complex integrated trait governed not by a single factor but by the synergistic action of multiple key agronomic components under polygenic control. Among these, main stem node number governs potential pod-bearing sites, lowest pod height determines mechanical harvest efficiency, and branch number modulates canopy architecture and light interception, thereby collectively determining final yield [4,5]. Therefore, the targeted genetic improvement of these components represents an effective strategy to improve yield. However, this endeavor faces major genetic challenges. These traits are classical complex quantitative traits, whose phenotypic variation is controlled by numerous small-effect loci, modulated by epistasis, and profoundly influenced by the environment [6]. This polygenic and context-dependent architecture makes identifying stable, major effect loci for breeding exceptionally difficult. Importantly, these traits exhibit high phenotypic plasticity due to significant genotype-by-environment (G×E) interactions, which can mask consistent genetic effects across environments. Consequently, G×E not only reduces heritability and phenotypic selection efficiency [7,8] but, more importantly, severely compromises the stability and portability of identified genetic loci, limiting their utility in molecular breeding. This limitation is evident in soybean research. For example, while early GWAS on traits like plant height identified numerous marker-trait associations, many proved unstable across different environments or populations, highlighting the inherent limitation of single-environment studies [9]. Thus, deciphering the stable genetic architecture of yield components across variable environments—by identifying environmentally robust loci—is a fundamental prerequisite for their effective application in modern breeding programs.

Genome-wide association study (GWAS), as a genetic analysis method based on natural populations, provides a powerful tool to dissect the genetic architecture of such complex traits without the need for constructing specific mapping populations. This strategy enables efficient identification of genetic loci significantly associated with target traits through genome-wide scanning of high-density single nucleotide polymorphism (SNP) markers [10,11]. GWAS has been successfully applied to dissect various important agronomic traits, including seed quality (protein and oil content), phenology (flowering time), stress resistance, and plant architecture (plant height, pod height) [12–14] in soybean. Notably, recent studies have highlighted the advantages of multi-environment experimental designs. For example, multi-environment GWAS on soybean hypocotyl length and yield-related traits in wild soybean (*Glycine soja*) have effectively dissected G×E interactions and identified quantitative trait loci (QTL) with stable expression across diverse environments, providing methodological insights for conducting stable genetic analyses of key yield traits in cultivated soybean [15]. However, the aforementioned studies have primarily focused on single traits or utilized wild soybean germplasm. For key yield component traits that directly influence single-plant yield of cultivated soybean, such as main stem node number, lowest pod height, and branch number, systematic identification of stable genetic loci using multi-year and multi-location trials remains to be thoroughly explored. The main reason for this gap is that most existing GWAS rely on phenotypic data collected from single or limited environments. Genetic loci identified under such conditions are often highly dependent on specific environmental factors, resulting in poor reproducibility and stability. Consequently, these findings are difficult to apply directly in molecular breeding practices across diverse ecological regions [16,17]. To discover robust genetic markers with broad applicability, it is imperative to adopt multi-environment trial designs encompassing multiple years and locations. Such approaches should be integrated with statistical models such as Best Linear Unbiased Prediction (BLUP) to correct for environmental noise and estimate stable

genetic potential of genotypes, thereby enhancing the power and reliability of association analysis [18].

Based on this rationale, the present study systematically phenotyped a panel of 320 representative soybean accessions across six locations over three consecutive years. The specific objectives of this study were:

(1) to integrate multi-environment phenotypic data using the Best Linear Unbiased Prediction (BLUP) model to minimize environmental noise;
(2) to identify stable SNP loci associated with main stem node number, lowest pod height, and branch number via MLM-based GWAS;
(3) to predict candidate genes within the significant genomic regions and annotate their potential functions;
(4) to provide reliable molecular markers for breeding of soybean varieties with optimized plant architecture and enhanced environmental adaptability.

2. Materials and Methods

2.1. Plant Materials and Field Trials

A diverse panel of 320 soybean accessions (*Glycine max* L.) was assembled for this study, comprising representative germplasm from major soybean-producing regions (Table A1). Multi-environment field trials were conducted across three years (2018, 2019, and 2021) at six representative locations within the spring soybean region of Northeast China.

The experimental design employed a randomized complete block design (RCBD) with three replications at each location. Individual plots consisted of 7.8 m^2 growing areas, with row spacing and planting density following local high-yield cultivation standards. Standard agronomic management practices, including fertilization, irrigation, and pest control, were applied according to regional recommendations at each site.

Trial locations were strategically selected to capture the environmental diversity of the target ecological zone, encompassing variations in latitude, altitude, and soil types (Table 1). Specifically, trials were conducted as follows: in 2018, three locations including Harbin (E1), Mudanjiang (E2), and Qing'an (E3); in 2019, Yilan (E4); and in 2021, Xiangyang (E5) and Minzhu (E6). This design facilitated systematic evaluation of genotype-by-environment (G×E) interactions across diverse climatic and edaphic conditions prevalent in Northeast China. Note: Due to the interruption of field management caused by the COVID-19 pandemic in 2020, the data for that year was not included in this analysis.

Table 1. Information on multi-environment trial locations.

Environment Code	Year	Location	Latitude (N)	Longitude (E)	UTM Easting (m)	UTM Northing (m)	UTM Zone	Altitude (m)	Soil Type	Mean Annual Temp. (°C)	Annual Precipitation (mm)
E1	2018	Harbin City, Heilongjiang province	45.8°	126.5°	318,708	5,066,491	52 N	145	Phaeozem (Black soil)	4.5	550
E2	2018	Mudanjiang City, Heilongjiang province	44.6°	129.6°	705,916	4,932,459	52 N	230	Dark Brown Soil	4.5	520
E3	2018	Qing'an County, Heilongjiang province	46.9°	127.5°	387,728	5,192,079	52 N	180	Phaeozem (Black soil)	2.9	580
E4	2019	Yilan County, Heilongjiang province	40.5°	130.1°	703,894	5,128,427	52 N	153	Phaeozem (Black soil)	3.6	556
E5	2021	Xiangyang County, Heilongjiang province	45.7°	126.6°	332,798	5,066,317	52 N	129	Phaeozem (Black soil)	4.2	530
E6	2021	Minzhu County, Heilongjiang province	44.7°	129.5°	322,761	5,078,246	52 N	145	Phaeozem (Black soil)	4.8	436

Climatic data are long-term averages from meteorological stations in the counties/cities where the trial sites are located. UTM coordinates are based on the WGS84 datum.

2.2. Trait Investigation and Phenotypic Data Analysis

At physiological maturity, ten representative plants were randomly sampled from each plot for trait measurement. Three key agronomic traits were recorded according to standardized procedures: main stem node number was counted from the cotyledonary node upward; lowest pod height was measured as the distance from the cotyledonary node to the lowest well-developed pod; and branch number was defined as the count of lateral branches bearing at least two well-developed pods. The mean value of each trait from the three replications was calculated for subsequent statistical analysis.

Descriptive statistics, including mean, standard deviation (SD), standard error (SE), maximum, and minimum values, were computed using R v4.2.1 with agricolae package (version 1.3-6). Multi-environment analysis of variance (ANOVA) was conducted using the following linear mixed model:

$$Y\_ijk = \mu + G\_i + E\_j + (GE)\_ij + R(E)\_jk + \varepsilon\_ijk$$

where μ is the overall mean, G_i is the random effect of the i-th genotype, E_j is the fixed effect of the j-th environment, (GE)_ij is the random genotype-by-environment interaction effect, R(E)_jk is the effect of the k-th replication nested within environment j, and ε_ijk is the residual error. Coefficient of variation (CV) was calculated as (SD/Mean) $\times$ 100%. Significant differences among environments for each trait were assessed using one-way ANOVA followed by the Least Significant Difference (LSD) test for multiple comparisons. Pearson correlation coefficients were calculated to examine relationships among traits. Violin plots were generated to visualize trait distributions and correlations. Normality of phenotypic distributions was tested using Shapiro–Wilk test; non-normal traits were Box–Cox transformed before GWAS.

2.3. Genotyping and Population Structure Analysis

2.3.1. SNP Genotyping

Total genomic DNA was extracted from young leaves (collected from two-week-old seedlings) using a standard cetyltrimethylammonium bromide (CTAB) method. DNA quality and concentration were assessed by 1% agarose gel electrophoresis and a Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) [19]. Subsequently, DNA samples extracted from 320 diverse soybean accessions were genotyped by re-sequencing using the Illumina HiSeq 2500 platform (Illumina, Inc., San Diego, CA, USA) by Biomarker Biotechnology Co., Ltd. (Beijing, China). Quality control filtering was applied to remove SNPs with minor allele frequency (MAF) < 5% or missing rate > 10%. A genome-wide significance threshold of $p < 1.0 \times 10^{-7}$ was used, corresponding to the Bonferroni-corrected threshold (0.05/486,256). SNPs deviating from Hardy–Weinberg equilibrium ($p < 1 \times 10^{-6}$) were excluded. All SNP positions and gene annotations were based on the soybean reference genome *Glycine max* cv. Williams 82 v4.0 (Wm82.a4.v1; Glycine_max_v4.0) from Phytozome v13 [20].

2.3.2. Genome-Wide Association Study and Candidate Gene Identification

To account for genotype-by-environment interactions and enhance the detection of stable genetic signals, we first calculated the Best Linear Unbiased Prediction (BLUP) values for each trait across all environments. This was done by fitting a linear mixed model using the lme4 package (version 1.1-31) in R, with genotype as a random effect and environment as a fixed effect. The BLUP values for each genotype, representing their estimated genetic potential with environmental noise minimized, were then extracted using the ranef() function. These BLUP-adjusted phenotypic values served as the input for all subsequent genome-wide association analyses.

Genome-wide association studies (GWAS) were conducted using the Mixed Linear Model (MLM) as implemented in GAPIT v3. This model was chosen for its ability to control both population structure and cryptic relatedness, thereby reducing false positives. Specifically, we incorporated a kinship matrix (K), calculated using the EMMA method, and the first three principal components (PCA1-3, accounting for >12% of genetic variance) as covariates in the MLM. GWAS was performed using GAPIT v3. Genome-wide significance was determined by Bonferroni correction ($p < 1.0 \times 10^{-7}$), corresponding to 0.05/486,256 SNPs. Candidate genes were identified within ±50 kb of lead SNPs based on linkage disequilibrium ($r^2 > 0.6$). Manhattan and quantile-quantile (Q-Q) plots were generated to visualize association signals and assess model fit, respectively.

2.3.3. Candidate Gene Definition and Identification

A candidate gene is defined as a gene located within the linkage disequilibrium (LD) block of a significant SNP that may be functionally responsible for the observed phenotypic variation. In this study, for each significant SNP ($p < 1 \times 10^{-7}$), we extended the genomic interval to ±50 kb on each side. SNPs in strong LD with the lead SNP ($r^2 > 0.6$) were used to define the candidate interval. Genes annotated within these intervals were retrieved from the Phytozome v13 soybean genome database based on the reference genome Glycine max cv. Williams 82 v4.0. Functional annotations were assigned using Gene Ontology (GO) terms and KEGG pathway enrichment analyses. Only genes with plausible biological relevance to plant architecture, hormone regulation, cell wall biosynthesis, or yield-related traits were retained as high-confidence candidates.

3. Results

3.1. Phenotypic Evaluation

Substantial genetic variation for agronomic traits was observed among the 320 soybean accessions, with significant environmental effects. The best linear unbiased predictors (BLUPs) for genotypes showed that the mean number of main stem nodes ranged from 8.84 to 19.96 (CV = 19%). The lowest pod height varied from 2.25 to 29.33 cm (CV = 75%), while branch number ranged from 0 to 15.9 (CV = 58%) (Table 2). The standard deviations of environment means across all test locations were 0.78, 3.19, and 0.44 for node number, pod height, and branch number, respectively, confirming substantial environmental influence on these traits.

Strong positive correlations were detected among all three traits: node number with pod height ($r = 0.78$, $p < 0.001$) and branch number ($r = 0.82$, $p < 0.001$), and pod height with branch number ($r = 0.65$, $p < 0.001$) (Figure A2). Analysis of variance revealed that genotype, environment, and genotype × environment interaction effects were all highly significant ($p < 0.01$) for each trait. Despite significant GE interactions, the SNP-based heritability estimates were high: 0.77 for node number, 0.75 for pod height, and 0.78 for branch number, indicating strong genetic control of these traits.

3.2. Genome-Wide Association Study (GWAS)

3.2.1. Nodes on Main Stem

Ten SNPs exceeded the significance threshold for main stem node number (Figure 1; Table 3). Manhattan plots revealed a prominent signal cluster on chromosome 19. The Q-Q plot exhibited deflection from the null distribution, confirming robust detection of true associations. Among the associated regions, three high-confidence candidate genes were prioritized for main stem node number: *Glyma.19G186000* (encoding a phosphatidylinositol-4-phosphate 5-kinase), *Glyma.19G190100* (a putative auxin response factor), and *Glyma.19G033500* (a cellulose-synthase-interactive protein). For instance, acces-

sions carrying the favorable allele at *Chr19.44576961* (near *Glyma.19G190100*) had an average of 16.8 main stem nodes, whereas those with the alternative allele averaged 12.3 nodes (Table 2). This 36.6% difference suggests that this SNP may contribute to yield potential through increased node number, though direct yield data are needed for confirmation.

Table 2. Descriptive Statistical Analysis of Soybean Main Stem Node Number, Pod Height at the Lowest Node, and Branch Number Across Six Environments.

1. Descriptive Statistical Analysis of Soybean Main Stem Node Number Across Six Environments						
Environment Code	**Mean**	**SD**	**SE**	**Min**	**Max**	**CV%**
E1	14.43	2.22	0.12	8.84	19.96	15
E2	15.11	2.97	0.17	5.17	23.95	20
E3	15.49	3.02	0.17	7.67	23.05	19
E4	13.55	2.24	0.13	7.50	20.90	17
E5	14.32	2.27	0.13	9.00	20.00	16
E6	14.36	3.03	0.17	5.00	25.72	21
Total	14.54	2.72	0.06	5.00	25.72	19
2. Descriptive Statistical Analysis of Pod Height at the Lowest Node Across Six Environments						
Environment Code	**Mean (m)**	**SD (cm)**	**SE (cm)**	**Min (cm)**	**Max (cm)**	**CV%**
E1	9.98	5.08	0.28	2.65	29.15	51
E2	5.83	2.33	0.13	2.37	15.25	40
E3	8.80	3.15	0.18	2.95	18.80	36
E4	4.89	1.52	0.08	2.38	17.17	31
E5	10.92	12.50	0.70	2.25	20.52	98
E6	13.63	4.56	0.26	3.37	29.33	33
Total	9.01	6.75	0.15	2.25	29.33	75
3. Descriptive Statistical Analysis of Branch Number Across Six Environments						
Environment Code	**Mean**	**SD**	**SE**	**Min**	**Max**	**CV%**
E1	2.20	1.36	0.08	0.00	8.14	62
E2	3.25	1.44	0.08	0.00	8.34	44
E3	2.53	1.35	0.08	0.00	7.95	53
E4	3.27	1.54	0.09	0.00	15.90	47
E5	2.52	1.75	0.10	0.00	9.67	70
E6	2.65	1.82	0.10	0.00	13.67	69
Total	2.74	1.60	0.04	0.00	15.90	58

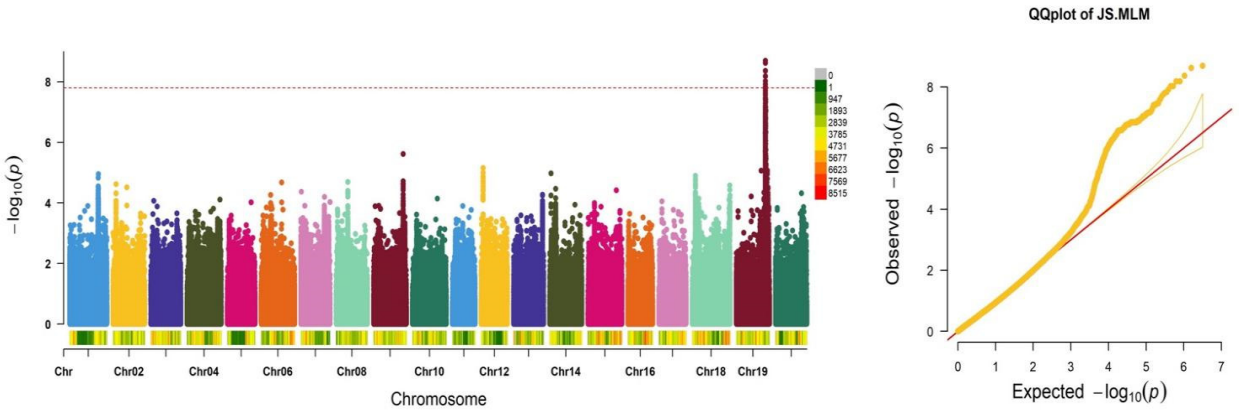


Figure 1. Manhattan plot and Q-Q plot for main stem node number after BLUP processing. In the Manhattan plot, different colors represent different chromosomes; the red horizontal line indicates the genome-wide significance threshold ($p < 1.0 \times 10^{-7}$). The Q-Q plot shows observed versus expected $-\log_{10}(p$-values).

Table 3. SNP loci identified for main stem node number using the MLM method after BLUP processing.

SNP ID	Chromosome	Position (bp)	Effect	SE	*p*-Value	MAF	Gene ID	Functional Annotation
Chr19.44493180	Chr19	44493180	−0.589447392	0.099974408	1.20×10^{-8}	0.15	*Glyma.19G186000*	Phosphatidylinositol-4-phosphate 5-kinase
Chr19.44493211	Chr19	44493211	−0.650518799	0.105366607	8.50×10^{-9}	0.14	*Glyma.19G186000*	Phosphatidylinositol-4-phosphate 5-kinase
Chr19.44514460	Chr19	44514460	−0.610616967	0.102397487	3.10×10^{-8}	0.13	*Glyma.19G186000*	Phosphatidylinositol-4-phosphate 5-kinase
Chr19.44544551	Chr19	44544551	−0.821986617	0.141567426	7.20×10^{-9}	0.12	*Glyma.19G190100*	Auxin response factor
Chr19.44568904	Chr19	44568904	−0.759068587	0.125683236	2.50×10^{-8}	0.11	*Glyma.19G190100*	Auxin response factor
Chr19.44576961	Chr19	44576961	−0.845134981	0.137562057	1.10×10^{-9}	0.10	*Glyma.19G190100*	Auxin response factor
Chr19.44577127	Chr19	44577127	−0.829278694	0.140631027	3.40×10^{-9}	0.10	*Glyma.19G190100*	Auxin response factor
Chr19.44625777	Chr19	44625777	−0.760768861	0.130775894	1.80×10^{-8}	0.12	*Glyma.19G190100*	Auxin response factor
Chr19.44652965	Chr19	44652965	−0.750686797	0.125871383	2.70×10^{-8}	0.11	*Glyma.19G190100*	Auxin response factor
Chr19.44736902	Chr19	44736902	−0.737821833	0.126056516	3.90×10^{-8}	0.11	*Glyma.19G033500*	Cellulose-synthase-interactive protein

SNP positions and gene annotations are based on the *Glycine max* cv. Williams 82 v4.0 reference genome (Phytozome v13). *p*-values and minor allele frequencies (MAF) were extracted from GAPIT v3 output.

3.2.2. Lowest Pod Height

Seven SNPs exceeded the genome-wide significance threshold for lowest pod height (Figure 2; Table 4). The Manhattan plot revealed a clear signal cluster on chromosome 8, with the most significant SNPs located within the 44.0–44.3 Mb interval. The Q-Q plot showed appropriate deviation from the null distribution, indicating effective control of false positives. Within the associated genomic regions, three high-confidence candidate genes were prioritized for lowest pod height: *Glyma.08G325000* (encoding a protein involved in cytokinin biosynthesis), *Glyma.08G321800* (a transducin family protein with WD-40 repeats), and *Glyma.08G322900* (a V-type proton ATPase subunit e1).

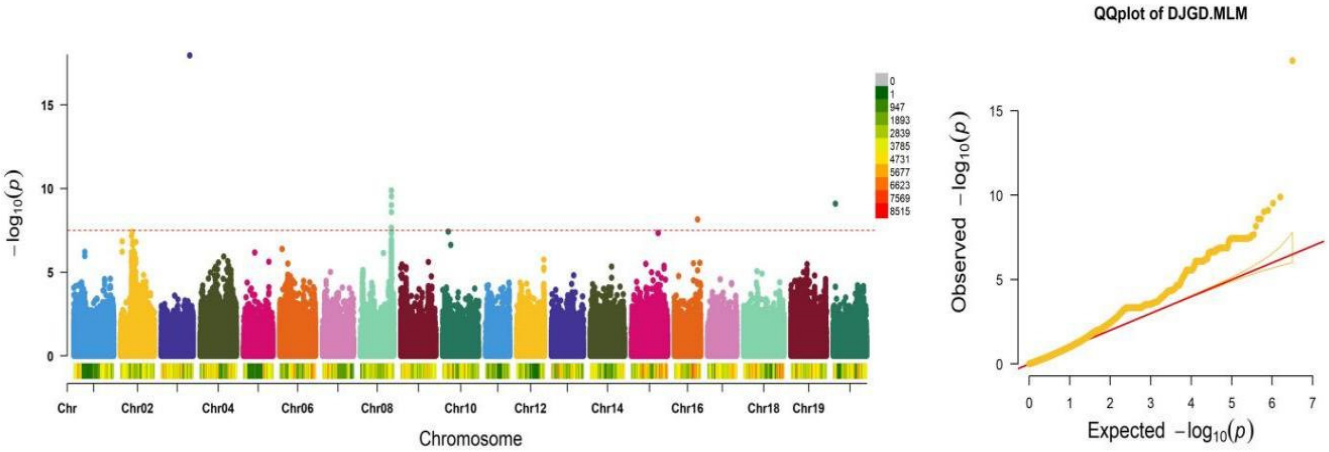


Figure 2. Manhattan plot and Q-Q plot for pod height at the lowest node after BLUP processing. Manhattan plot and Q-Q plot for pod height at the lowest node after BLUP processing. Color coding and significance threshold are the same as in Figure 1.

Table 4. SNP loci identified for pod height at the lowest node using the MLM method after BLUP processing.

SNP	CHROM	POS	Effect	SE	*p*-Value	MAF	Gene ID	Functional Annotation
Chr08.44018897	Chr08	44018897	1.915270052	0.303979909	6.50×10^{-9}	0.23	*Glyma.08G325000*	Cytokinin biosynthetic process
Chr08.44157110	Chr08	44157110	1.825000213	0.280452215	2.10×10^{-8}	0.21	*Glyma.08G325000*	Cytokinin biosynthetic process
Chr08.44159716	Chr08	44159716	1.666579917	0.290141236	5.20×10^{-8}	0.20	*Glyma.08G321800*	Transducin/WD-40 repeat protein

Table 4. *Cont.*

SNP	CHROM	POS	Effect	SE	*p*-Value	MAF	Gene ID	Functional Annotation
Chr08.44226650	Chr08	44226650	1.68015495	0.293573039	4.40×10^{-8}	0.19	*Glyma.08G321800*	Transducin/WD-40 repeat protein
Chr08.44289397	Chr08	44289397	2.011206389	0.30244952	5.80×10^{-9}	0.18	*Glyma.08G322900*	V-type proton ATPase subunit e1
Chr08.44290786	Chr08	44290786	1.88195646	0.306746582	1.40×10^{-8}	0.18	*Glyma.08G322900*	V-type proton ATPase subunit e1
Chr08.44290823	Chr08	44290823	1.88195646	0.306746582	1.40×10^{-8}	0.18	*Glyma.08G322900*	V-type proton ATPase subunit e1

Same as Table 3.

3.2.3. Branch Number

GWAS identified five significant SNPs for branch number (Figure 3; Table 5). The strongest signals localized to chromosome 18. The Q-Q plot showed deflection from the null distribution. Candidate gene analysis highlighted *Glyma.18G274400* (encoding a chalcone-synthase-like enzyme involved in flavonoid biosynthesis) and *Glyma.18G274000* (a zinc-finger CCCH domain protein implicated in zinc ion homeostasis and auxin biosynthesis).

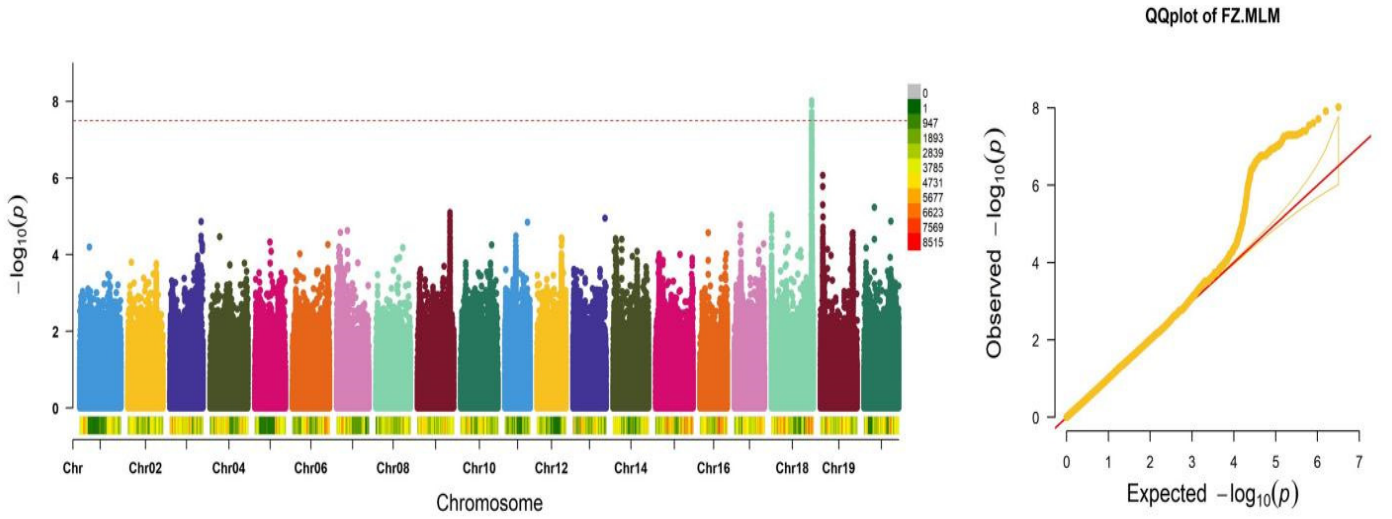


Figure 3. Manhattan plot and Q-Q plot for branch number after BLUP processing. Manhattan plot and Q-Q plot for branch number after BLUP processing. Color coding and significance threshold are the same as in Figure 1.

Table 5. SNP loci identified for branch number using the MLM method after BLUP processing.

SNP	CHROM	POS	Effect	SE	*p*-Value	MAF	Gene ID	Functional Annotation
Chr18.55620032	Chr18	55620032	−0.325549081	0.057198073	2.20×10^{-8}	0.24	*Glyma.18G274400*	Chalcone synthase-like
Chr18.55643896	Chr18	55643896	−0.322998227	0.055221188	3.50×10^{-8}	0.23	*Glyma.18G274400*	Chalcone synthase-like
Chr18.55654817	Chr18	55654817	−0.314161324	0.054494017	5.50×10^{-8}	0.22	*Glyma.18G274000*	Zinc-finger CCCH domain protein
Chr18.55654834	Chr18	55654834	−0.316146516	0.054494017	5.10×10^{-8}	0.22	*Glyma.18G274000*	Zinc-finger CCCH domain protein
Chr18.55654882	Chr18	55654882	−0.330405515	0.056054222	1.90×10^{-8}	0.23	*Glyma.18G274000*	Zinc-finger CCCH domain protein

Same as Table 3.

3.3. Candidate Gene Prediction and Functional Annotation

Significant GWAS loci for node number, lowest pod height and branch number were annotated using Phytozome (±50 kb, $r^2 > 0.6$) and KEGG pathways. For each trait, we prioritized high-confidence candidate genes located within the associated genomic intervals (±50 kb of lead SNPs) based on their functional annotations and expression evidence.

The key candidates for each trait are summarized in Table 6, with detailed functional information provided in Tables A6–A8.

Table 6. Functional Annotation of Candidate Genes.

Lead SNP	CHROM	POS (bp)	GO Term Accession	Functional Annotation	Reference
Glyma.08G325000	Chr08	44,600,528	GO:0009691	Promotes ribosome production	Phytozome v13; [21]
Glyma.08G321800	Chr08	43,445,972	GO:0005515	The endoplasmic reticulum	Phytozome v13; [9]
Glyma.08G322900	Chr08	43,530,972	GO:1902600	Oxidative phosphorylation	Phytozome v13; [18]
Glyma.18G274400	Chr18	55,685,303	GO:0008194	Secondary metabolites	Phytozome v13; [22]
Glyma.18G274000	Chr18	55,626,229	GO:0016567	Promotes binding with Zn+	Phytozome v13; [21,22]
Glyma.19G186000	Chr19	45,213,000	GO:0046166	Controls phosphoinositol production; Affect the carbon fixation process	Phytozome v13; [9]
Glyma.19G190100	Chr19	45,213,327	GO:0019752	Promotes glycolysis; Promotes glucose production	Phytozome v13; [22]
Glyma.19G033500	Chr19	4,309,010	GO:0010215	Promotes cellulose production	Phytozome v13; [23]

References are from the manuscript's reference list: [9] Fang et al. (2017) *Genome Biol.*; [18] Reena et al. (2023) *Front. Plant Sci.*; [24] Zhao et al. (2024) *Int. J. Mol. Sci.*; [23] Cai et al. (2024) *Agriculture*; [25] Borgohain et al. (2026) *Funct. Integr. Genom.* "Phytozome v13" refers to the database (https://phytozome-next.jgi.doe.gov/ (accessed on 7 April 2026)).

For main stem node number, three candidate genes on chromosome 19 were prioritized: *Glyma.19G186000* (encoding a phosphatidylinositol-4-phosphate 5-kinase), *Glyma.19G190100* (a putative auxin response factor), and Glyma.19G033500 (a cellulose-synthase-interactive protein).

For lowest pod height, three candidate genes on chromosome 8 were prioritized: *Glyma.08G325000* (involved in cytokinin biosynthesis), *Glyma.08G321800* (a transducin family protein with WD-40 repeats), and Glyma.08G322900 (a V-type proton ATPase subunit e1).

For branch number, two candidate genes on chromosome 18 were prioritized: *Glyma.18G274400* (encoding a chalcone-synthase-like enzyme involved in flavonoid biosynthesis)and *Glyma.18G274000* (a zinc-finger CCCH domain protein implicated in zinc ion homeostasis and auxin biosynthesis).

4. Discussion

4.1. Analysis of Yield-Related Agronomic Traits in Soybean

Key yield-related agronomic traits in soybean, such as the number of nodes on the main stem, the height of the lowest pod, and the number of branches, are all quantitative traits controlled by multiple genes and susceptible to environmental influences [26–28], which poses a major challenge for their genetic dissection and breeding improvement. Therefore, this study conducted multi-year, multi-location phenotypic evaluation and genome-wide association analysis (GWAS) on a panel of 320 soybean accessions, aiming to elucidate the genetic basis of these key yield-component traits.

Phenotypic analysis showed that the coefficient of variation (CV) for the number of nodes on the main stem across different environments (average 19%) was significantly lower than that for the height of the lowest pod (75%) and the number of branches (58%), suggesting that node number has the highest genetic stability and is relatively less affected by environmental fluctuations. This characteristic makes it an ideal target trait for achieving stable yield improvement through genetic breeding [29].

While grain weight (e.g., 100-seed weight) is considered the most direct and important determinant of soybean yield [18], the present study focused on three yield-related architectural traits (main stem node number, lowest pod height, and branch number) that indirectly influence yield potential. We acknowledge that grain weight was not measured in our multi-environment trials; consequently, we did not identify SNPs associated with grain weight. Nonetheless, the stable genetic loci detected for architectural traits provide valuable resources for marker-assisted selection, and future studies should integrate grain weight and other direct yield components to achieve comprehensive genetic improvement [6].

In contrast, the height of the lowest pod and the number of branches exhibited higher phenotypic plasticity. On one hand, this reveals that they are more sensitive to environmental signals, with their expression more easily influenced by factors such as water and fertilizer management, planting density, and light and temperature conditions; on the other hand, it also implies significant potential breeding value. Through precise cultivation management (e.g., optimizing planting density to regulate branching) or leveraging positive genotype-by-environment interactions (G×E), yield potential within specific ecological regions can be more fully exploited. Such a plasticity-based "management-genotype" co-optimization strategy has been proven effective in high-yield practices of crops such as maize and wheat [30,31].

The relatively small standard errors for each trait further confirmed that the phenotypic data obtained in this study are reliable, and the observed variation is primarily driven by genotypic differences and genotype-by-environment interactions [22,32].

Additionally, correlation analysis revealed highly significant positive correlations among the number of nodes on the main stem, the height of the lowest pod, and the number of branches. This finding holds important practical value for breeding: it suggests that during genetic improvement of these traits, there may exist pleiotropic regulatory hubs or tightly linked gene clusters [21,24]. For example, key genes regulating the synthesis or signaling of plant hormones (such as gibberellins and strigolactones) are often reported to simultaneously influence multiple plant architecture-related traits [23]. Such genetic synergy makes it possible to synchronously improve multiple yield-component traits by selecting a single major locus or haplotype [33], thereby potentially breaking the common negative trade-offs in conventional breeding (such as between yield and lodging resistance) [25,34] and enabling more efficient development of improved cultivars with superior comprehensive traits.

4.2. Analysis of Results Following BLUP Processing

To eliminate noise arising from genotype-by-environment (G×E) interactions in multi-environment phenotypic data, this study employed the Best Linear Unbiased Prediction (BLUP) model to optimize the data, thereby obtaining phenotypic values that represent the stable genetic potential of each individual. The core principle of this strategy lies in decomposing the observed phenotypic values from specific environments into genotypic effects, environmental effects, G×E interaction effects, and random error. The BLUP model optimally predicts random effects—here, the genotypic effects—through shrinkage estimation, which reduces the influence of extreme environmental values and yields more robust estimates of the intrinsic genetic potential of genotypes [35]. This critical step effectively addresses the issues of weak association signals and high false-positive rates observed when using raw data directly, ensuring that subsequent genome-wide association studies (GWAS) can accurately identify genetic loci that are stable across multiple environments.

Specifically, when GWAS is conducted using raw single-environment phenotypic data, significant signals are often strongly coupled with specific environmental conditions, leading to poor reproducibility of loci. In contrast, GWAS based on BLUP-estimated

values captures trait variation driven by genetic structure, free from interference by specific environments. The results of this study show that after BLUP adjustment, the significance peaks in the Manhattan plots become sharper, and background noise is substantially reduced (Figures 1–3). This directly demonstrates the efficacy of the model in filtering environmental noise and enhancing genetic signals [36–38]. This approach has also been validated as effective in multi-environment GWAS studies of other crops [39–41].

4.3. Validation of the MLM Approach for Association Mapping

The Mixed Linear Model (MLM) was selected as the optimal analytical approach for this study due to its robust control of population stratification and cryptic relatedness through the incorporation of both principal components (Q matrix) and kinship matrix (K matrix). While alternative methods such as the General Linear Model (GLM) and FarmCPU exist, MLM has been established as the mainstream choice for GWAS of complex quantitative traits in plant science, particularly when population structure is present [42–45].

In this study, the Q-Q plot from MLM analysis showed appropriate adherence to the expected null distribution with minimal deviation at the tail, indicating effective control of false positives (Figures 1–3). The genomic inflation factors (λ) ranged from 0.98 to 1.05 across the three traits, confirming that the model successfully accounted for confounding genetic structure. The Manhattan plots exhibited clear, distinct peaks above the significance threshold with low background noise, further validating the reliability of the detected associations. Therefore, all significant loci reported in this study were identified using the MLM framework.

4.4. Comparison of GWAS Results with Previous Studies

This study conducted phenotypic evaluations across three years and six locations on a panel of 320 domestic and international soybean cultivars, followed by genotyping and genome-wide association analysis, which identified SNP loci associated with the number of nodes on the main stem, the height of the lowest pod, and the number of branches. Placing our findings within the context of existing knowledge for cross-validation is crucial for assessing their reliability and uncovering novel insights. By comparing our results with previously reported important quantitative trait loci (QTL) or association signals, we can not only confirm known genetic "hotspot" regions but also potentially reveal new regulatory loci, thereby progressively refining the genetic map of soybean yield-related traits.

Through GWAS, we identified 10 SNP loci located within the Chr19:45145714 and Chr19:45205190 intervals. The signal cluster pinpointed in our study falls precisely within two broader intervals reported by Wang et al. (2016) [46]. This high degree of overlap is not coincidental; it strongly supports the notion that this specific genomic region on chromosome 19 represents a conserved and important genetic hotspot for regulating soybean stem development, including the number of nodes on the main stem. The application of a multi-environment BLUP strategy likely enabled us to further narrow down the candidate region within these broad intervals, pointing towards more precise candidate loci.

Jiaoping Zhang et al. (2015) [47] pinpointed a candidate gene, Chr08.44265646, associated with the height of the lowest pod on chromosome 8. Our GWAS analysis identified four SNP loci located near this position: Chr08.44226650, Chr08.44289397, Chr08.44290786, and Chr08.44290823. Although these four loci do not exactly coincide with the one reported by Jiaoping Zhang et al., their physical positions are extremely close. Their tight physical proximity (all within the ~44.2–44.3 Mb interval) strongly suggests that this region may harbor a gene cluster or regulatory module influencing the height of the lowest pod (or more broadly, basal stem elongation). Differences in populations used, marker density, and statistical models across studies can lead to the detection of different linked SNPs

derived from the same causal variant [48]. Consequently, our results do not contradict the previous findings; instead, from the perspective of independent data and a different analytical method, they collectively reinforce the critical role of this region on chromosome 8 in controlling the height of the lowest pod in soybean.

Fang et al. (2017) [9] identified a gene locus associated with branch number within the Chr18:55626229–55628880 interval. Our GWAS analysis identified one SNP locus at Chr18.55620032. While this locus is not within the 55626229–55628880 interval, it is located in close proximity. Despite an approximate 6 kb offset in physical position, on a genomic scale, these two signals can be considered as pointing to the same genetic interval. Such minor offsets are common in GWAS and may arise from differences in population genetic structure, linkage disequilibrium patterns, or statistical fluctuations [49]. Importantly, both studies independently associate branch number with signals in the same narrow region of chromosome 18 (~55.62–55.63 Mb), which significantly increases the credibility that a genuine branch number regulatory gene resides in this area.

The comparisons with previous studies outlined above provide a relevant research basis for subsequent analyses and for the improvement and regulation of traits such as the number of nodes on the main stem, the height of the lowest pod, and the number of branches in soybean, ultimately aiding in the development cultivars with improved yield potential.

In summary, the stable association loci identified in this study show a high degree of consistency or close physical proximity with important QTL intervals reported in multiple previous studies. This cross-validation across different populations and methodologies not only confirms the reliability of our GWAS results but also highlights the significance of these specific regions on soybean chromosomes 19, 8, and 18 as key targets for the genetic improvement of plant architecture and yield-related traits. Future functional studies should prioritize these "consensus regions," employing strategies such as fine mapping, haplotype analysis, and gene editing to ultimately clone the key genes regulating these complex traits.

4.5. Functional Classification of Candidate Genes

The eight high-confidence candidate genes identified in this study can be categorized into three major functional groups based on their Gene Ontology annotations and predicted biological roles.

Hormone-related genes: *Glyma.08G325000* is involved in cytokinin biosynthesis, a key regulator of cell division and shoot architecture; *Glyma.19G190100* encodes a putative auxin response factor, which mediates auxin signaling and influences stem elongation and node development.

Cell wall and carbohydrate metabolism: *Glyma.19G033500* is a cellulose-synthase-interactive protein, potentially affecting cell wall formation and stem strength; *Glyma.19G186000* encodes a phosphatidylinositol-4-phosphate 5-kinase involved in carbon fixation and glyceraldehyde-3-phosphate biosynthesis.

Stress response and energy metabolism: *Glyma.08G322900* codes for a V-type proton ATPase subunit e1, essential for transmembrane proton transport and oxidative phosphorylation; *Glyma.08G321800* is a transducin/WD-40 repeat protein involved in protein binding and signaling; *Glyma.18G274400* is a chalcone-synthase-like enzyme participating in flavonoid (secondary metabolite) biosynthesis; *Glyma.18G274000* is a zinc-finger CCCH domain protein implicated in zinc ion homeostasis and auxin biosynthesis, linking nutrient sensing with hormone regulation.

This functional diversification suggests that soybean yield-related architectural traits are controlled by a complex network involving hormone signaling, cell wall remodeling, pri-

mary and secondary metabolism, and stress-responsive pathways. The stable loci identified here provide promising targets for further functional validation and molecular breeding.

4.6. Candidate Genes Harboring Significant SNPs and Their Potential Functional Mechanisms

In this study, each of the prioritized candidate genes contains at least one significant SNP ($p < 1 \times 10^{-7}$) identified by multi-environment GWAS (Tables 3–5). Based on their functional annotations, we propose the following plausible mechanisms contributing to the observed trait variation.

For main stem node number: The significantly associated SNPs on chromosome 19 are located within or near *Glyma.19G186000* (phosphatidylinositol-4-phosphate 5-kinase), *Glyma.19G190100* (auxin response factor), and *Glyma.19G033500* (cellulose-synthase-interactive protein). The auxin response factor may regulate meristem activity and internode elongation, directly influencing node number; the phosphatidylinositol kinase is involved in carbon fixation and glyceraldehyde-3-phosphate biosynthesis, potentially affecting energy supply for stem growth; the cellulose-synthase-interactive protein may impact cell wall formation and stem mechanical properties, indirectly modulating node development [33].

For lowest pod height: Significant SNPs on chromosome 8 map to *Glyma.08G325000* (cytokinin biosynthesis), *Glyma.08G321800* (WD-40 repeat protein), and *Glyma.08G322900* (V-type proton ATPase subunit e1). Cytokinin is known to regulate cell division and elongation in stem internodes [24]; variation in *Glyma.08G325000* may affect basal internode elongation, thereby altering pod height. The V-type ATPase is critical for proton gradient maintenance and energy metabolism, which could influence cell expansion. The WD-40 protein may act as a scaffold in signaling pathways controlling stem development [9].

For branch number: Significant SNPs on chromosome 18 are associated with *Glyma.18G274400* (chalcone-synthase-like) and *Glyma.18G274000* (zinc-finger CCCH domain protein). Chalcone synthase is a key enzyme in flavonoid biosynthesis, which can affect auxin transport and lateral bud outgrowth; the zinc-finger CCCH protein has been implicated in auxin biosynthesis and zinc homeostasis, both of which regulate branching [21,24].

The identification of these SNP–gene associations provides a solid foundation for future functional validation. As suggested in Section 4.4, fine mapping within these candidate regions, haplotype analysis in natural populations, and gene editing (e.g., CRISPR/Cas9) of *Glyma.19G190100*, *Glyma.08G325000*, or *Glyma.18G274000* could definitively establish their causal roles in soybean plant architecture and yield-related traits [33].

5. Conclusions

This study systematically identified stable genetic loci and candidate genes associated with three key yield-related traits in soybean—namely, the number of nodes on the main stem, the height of the lowest pod, and the number of branches—through the integration of multi-environment phenotypic data collected over three years and six locations with genome-wide association analysis (GWAS). Following environmental noise correction using the Best Linear Unbiased Prediction (BLUP) model, a total of 22 SNPs that were significant and stable across multiple environments were identified. These loci were not randomly distributed but were concentrated and enriched within specific regions of chromosomes 19, 8, and 18, suggesting that these regions may represent genetic hotspots for the regulation of soybean plant architecture.

Within the associated intervals, eight high-confidence candidate genes were further predicted (Table 6). Functional annotation revealed their significant involvement in key pathways such as cell wall biosynthesis, energy metabolism, and stress response, providing

new insights into the molecular mechanisms underlying these traits. This study demonstrates that a multi-environment trial design combined with a BLUP-based correction strategy effectively overcomes interference caused by genotype-by-environment interactions, thereby enabling the identification of robust genetic markers with broad adaptability.

These findings not only provide reliable target resources for marker-assisted selection aimed at developing stable yield-related traits and stable soybean varieties but also lay an important theoretical and material foundation for future precision improvement of plant architecture through gene editing or allele pyramiding.

Author Contributions: R.D. and Q.G. designed the research, analyzed the data and drafted the manuscript. J.Z., Y.D., X.Y., H.Z. and R.Z. provided the plant materials. H.J., Y.-E.W., D.C., C.Z. and Z.L. performed the experiment. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Heilongjiang Province Seed Industry Innovation and Development Project (Precise Identification, Evaluation, and Utilization of Excellent Crop Germplasm Resources, 2024HZYCXNK01), CARS-Green Manure (GARS-22), China National Crop Germplasm Resources Platform for Green Manure (NCCGR-2025-19) and Safe Preservation of Green Manure Germplasm Resources (22250404).

Data Availability Statement: All datasets generated for this study are included in the article; further inquiries can be directed to the first author.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Full name
MAF	Minor allele frequency
CV	Coefficient of variation
BLUP	Best linear unbiased prediction
RCBD	Randomized complete block design
GLM	General linear model
MLM	Mixed linear model
FarmCPU	Fixed and random model Circulating Probability Unification
QTL	Quantitative trait locus
SNP	Single-nucleotide polymorphism
GWAS	Genome-wide association study
G×E	Genotype-by-environment interaction

Appendix A

Table A1. Source and grouping information of 320 soybean genotypes and population structure.

Number	Name	Resource ID	Source	Number	Name	Resource ID	Source
1	C001	2016D3994 HZDD3867	America	161	C207	2016D3680 HZDD3435	Germany
2	C003	2016D0494 HZDD2816	Austria	162	C208	2016D3785 HZDD3614	Canada
3	C005	2016D4088 HZDD3990	America	163	C213	Hefeng55	China
4	C007	2016D4417 HZDD4440	Russia	164	C215	Henong65	China
5	C013	2016D4461 HZDD4485	Canada	165	C222	Suizhongzuo40	China
6	C014	2016D3995 HZDD3868	America	166	C237	2016D2675 HZDD2012	Ukraine
7	C017	2016D0799 HZDD3283	Russia	167	C240	2016D2891 HZDD2312	Russia
8	C033	Heihe49	China	168	C241	2016D3074 HZDD2596	Ukraine
9	C039	2016D1261 HZDD4280	Russia	169	C242	2016D4039 HZDD3932	America
10	C049	2016D0359 HZDD1818	Ukraine	170	C244	2016D4061 HZDD3958	Poland

Table A1. *Cont.*

Number	Name	Resource ID	Source	Number	Name	Resource ID	Source
11	C009	2016D0446 HZDD2271	Belarus	171	C251	2016D2840 HZDD2246	Moldova
12	C011	2016D4544 HZDD4556	Ukraine	172	C252	2016D2880 HZDD2299	Canada
13	C015	2016D4568 HZDD4529	France	173	C256	2016D3173 HZDD2723	Russia
14	C021	2016D4070 HZDD3968	America	174	C265	2016D4528 HZDD4573	Russia
15	C025	2016D4495 HZDD4519	Hungary	175	C280	2016D2882 HZDD2301	France
16	C027	2016D0133 HZDD1562	Hungary	176	C291	Heinong48	China
17	C035	2016D4498 HZDD4522	Hungary	177	C296	Suinong35	China
18	C037	2016D0417 HZDD2088	America	178	C298	2016D1222 HZDD767	Czech Republic
19	C041	2016D4569 HZDD4528	Germany	179	C299	2016D1742 HZDD558	Moldova
20	C047	Heihe44	China	180	C084	2016D2532 HZDD1794	Romania
21	C061	2016D0431 HZDD2175	Serbia	181	C034	2016D3105 HZDD2640	Canada
22	C019	2016D3160 HZDD2705	Belarus	182	C046	2016D1913 HZDD754	Moldova
23	C023	2016D4089 HZDD3991	America	183	C052	2016D1533 HZDD296	Moldova
24	C031	2016D1044 HZDD847	Czech Republic	184	C118	2016D2070 HZDD967	China
25	C043	2016D0516 HZDD3320	Russia	185	C130	2016D2811 HZDD2209	China
26	C045	2016D0842 HZDD4117	America	186	C175	2016D0185 HZDD1957	America
27	C051	2016D0355 HZDD1789	Poland	187	C183	2016D0358 HZDD1802	Romania
28	C055	2016D2047 HZDD929	Algeria	188	C203	2016D1243 HZDD2851	China
29	C063	2016D0452 HZDD2391	Canada	189	C206	2016D3569 HZDD3260	China
30	C065	2016D0134 HZDD1625	Czech Republic	190	C210	Dongsheng77	China
31	C075	2016D4479 HZDD4503	Czech Republic	191	C217	Kangxianchong9	China
32	C079	2016D3943 HZDD3810	Canada	192	C218	Kenfeng13	China
33	C083	2016D4397 HZDD4417	Germany	193	C219	Longheidadou2	China
34	C093	2016D4225 HZDD4205	Russia	194	C221	Qinong5	China
35	C113	2016D1808 HZDD635	Belarus	195	C223	2016D0337 HZDD1637	Czech Republic
36	C119	2016D4045 HZDD3940	Germany	196	C227	2016D1964 HZDD815	Czech Republic
37	C145	Heihe43	China	197	C235	2016D1029 HZDD620	Czech Republic
38	C163	2016D0768 HZDD2490	Germany	198	C239	2016D2855 HZDD2273	Moldova
39	C167	2016D1437 HZDD175	America	199	C253	2016D2907 HZDD2329	Poland
40	C029	2016D0430 HZDD2166	America	200	C254	2016D2910 HZDD2338	Korea
41	C057	2016D2850 HZDD2266	Belarus	201	C261	2016D4002 HZDD3878	Ukraine
42	C067	2016D0144 HZDD2556	Russia	202	C262	2016D4011 HZDD3892	Hungary
43	C069	2016D0468 HZDD2464	Sweden	203	C278	2016D1629 HZDD429	Ukraine
44	C089	2016D1547 HZDD313	Czech Republic	204	C283	2016D3267 HZDD2833	Ukraine
45	C141	2016D4444 HZDD4467	Ukraine	205	C286	2016D2999 HZDD2456	Russia
46	C059	2016D3576 HZDD3269	Moldova	206	C287	2016D3000 HZDD2457	Russia
47	C088	2016D3063 HZDD2575	Hungary	207	C297	2016D4574 HZDD4525	Sweden
48	C131	2016D1269 HZDD3603	Russia	208	C310	2016D2632 HZDD1947	Canada
49	C136	2016D2044 HZDD926	Africa	209	C216	Longdou2	China
50	C137	2016D2842 HZDD2250	America	210	C248	2016D2431 HZDD1634	Czech Republic
51	C147	2016D3773 HZDD3588	Moldova	211	C257	2016D3178 HZDD2729	Hungary
52	C148	2016D0751 HZDD1519	Russia	212	C279	2016D2749 HZDD2113	America
53	C149	2016D0834 HZDD4070	Russia	213	C281	2016D3201 HZDD2757	France
54	C151	2016D4549 HZDD4551	Czech Republic	214	C282	2016D3224 HZDD2781	Russia
55	C153	2016D0405 HZDD2032	Czech Republic	215	C236	2016D1196 HZDD4180	Russia
56	C157	2016D1734 HZDD550	Hungary	216	C238	2016D2834 HZDD2237	Moldova
57	C158	2016D2984 HZDD2439	Austria	217	C267	2016D0379 HZDD1889	Canada

Table A1. *Cont.*

Number	Name	Resource ID	Source	Number	Name	Resource ID	Source
58	C174	2016D3930 HZDD3796	Canada	218	C268	2016D0771 HZDD2516	Lithuania
59	C226	2016D1570 HZDD351	Hungary	219	C269	2016D1633 HZDD433	America
60	C228	2016D2671 HZDD2007	Hungary	220	C271	2016D0492 HZDD2730	Hungary
61	C245	2016D2665 HZDD1999	Ukraine	221	C290	2016D3962 HZDD3831	Canada
62	C247	2016D1855 HZDD686	Moldova	222	C301	2016D0569 HZDD932	Austria
63	C260	2016D3858 HZDD3712	Czech Republic	223	C311	2016D2674 HZDD2010	Hungary
64	C263	2016D3186 HZDD2740	China	224	C315	2016D1372 HZDD77	Russia
65	C053	2016D1199 HZDD4187	Russia	225	C317	2016D1624 HZDD424	Moldova
66	C071	2016D0559 HZDD172	Canada	226	C096	2016D1035 HZDD705	Russia
67	C073	2016D4349 HZDD4364	China	227	C106	2016D0695 HZDD348	Moldova
68	C081	2016D0400 HZDD1989	Ukraine	228	C212	Hefeng51	China
69	C085	2016D0526 HZDD3696	Romania	229	C038	2016D2499 HZDD1736	Russia
70	C091	2016D1577 HZDD362	Germany	230	C066	2016D0730 HZDD1018	Russia
71	C095	2016D0418 HZDD2103	Czech Republic	231	C104	2016D0700 HZDD380	Moldova
72	C097	2016D3974 HZDD3846	America	232	C202	2016D1200 HZDD4198	America
73	C099	2016D0415 HZDD2074	Germany	233	C209	2016D4450 HZDD4473	Japan
74	C101	2016D1416 HZDD136	Russia	234	C224	2016D0791 HZDD3185	Moldova
75	C103	2016D1412 HZDD129	Japan	235	C230	2016D3038 HZDD2534	Romania
76	C115	2016D1003 HZDD316	Ukraine	236	C234	2016D0841 HZDD4107	Moldova
77	C121	2016D1725 HZDD539	Hungary	237	C249	2016D2459 HZDD1673	Ukraine
78	C125	2016D0829 HZDD4022	France	238	C255	2016D2928 HZDD2357	America
79	C127	2016D1022 HZDD538	Czech Republic	239	C259	2016D3629 HZDD3362	Canada
80	C129	2016D1090 HZDD1813	Romania	240	C266	2016D3581 HZDD3275	Russia
81	C133	2016D1665 HZDD469	Hungary	241	C270	2016D0480 HZDD2592	Ukraine
82	C135	2016D1959 HZDD808	Russia	242	C272	2016D1072 HZDD1135	Russia
83	C138	2016D3166 HZDD2716	Serbia	243	C284	2016D2350 HZDD1501	Sweden
84	C139	2016D3229 HZDD2786	Russia	244	C288	2016D3000 HZDD2457-2	Russia
85	C142	2016D4570 HZDD4527	Algeria	245	C293	Heinong64	China
86	C144	2016D0863 HZDD564	Germany	246	C294	Heinong67	China
87	C159	2016D3176 HZDD2727	Hungary	247	C308	Mudou8	China
88	C160	Heihe45	China	248	C054	2016D1247 HZDD2906	Russia
89	C173	2016D0126 HZDD1286	Moldova	249	C072	2016D1069 HZDD1043	China
90	C176	2016D1392 HZDD96	China	250	C074	2016D1112 HZDD2574	Hungary
91	C177	2016D3144 HZDD2685	Germany	251	C076	2016D1183 HZDD3707	Romania
92	C181	2016D0857 HZDD124	Japan	252	C080	2016D2118 HZDD1086	China
93	C186	2016D1255 HZDD3592	Russia	253	C098	2016D1063 HZDD1015	Russia
94	C187	2016D1434 HZDD157	America	254	C100	2016D0314 HZDD1424	Russia
95	C188	2016D1554 HZDD326	Moldova	255	C102	2016D0696 HZDD357	America
96	C190	2016D2661 HZDD1995	Moldova	256	C126	2016D3730 HZDD3522	Czech Republic
97	C193	2016D2847 HZDD2258	Ukraine	257	C274	2016D1315 HZDD22	Russia
98	C196	2016D3254 HZDD2820	Belarus	258	C036	2016D2886 HZDD2305	France
99	C201	2016D0580 HZDD933	Austria	259	C048	2016D2137 HZDD1112	China
100	C204	2016D1774 HZDD596	Poland	260	C062	2016D1019 HZDD497	America
101	C225	2016D0857 HZDD124	Japan	261	C082	2016D2128 HZDD1100	China
102	C232	2016D3156 HZDD2701	Moldova	262	C273	2016D1184 HZDD3825	America
103	C243	2016D4047 HZDD3942	France	263	C275	Longhuang2	China
104	C090	2016D3133 HZDD2673	Ukraine	264	C289	2016D3031 HZDD2521	Russia
105	C092	2016D0813 HZDD3490	Algeria	265	C292	Heinong50	China
106	C107	2016D0954 HZDD3308	Russia	266	C295	Nongqingdou4	China
107	C108	2016D0302 HZDD1325	Hungary	267	C300	2016D0491 HZDD2725	Russia
108	C109	2016D1033 HZDD689	Moldova	268	C303	2016D3230 HZDD2787	Russia
109	C110	2016D0293 HZDD1308	Moldova	269	C309	Suinong29	China
110	C112	2016D0760 HZDD1735	Russia	270	C325	2016D1366 HZDD71	Romania
111	C116	2016D1576 HZDD361	Moldova	271	C030	2016D1981 HZDD838	Germany

Table A1. *Cont.*

Number	Name	Resource ID	Source	Number	Name	Resource ID	Source
112	C117	2016D1092 HZDD2001	Moldova	272	C032	2016D0199 HZDD1960	America
113	C120	2016D2475 HZDD1700	Russia	273	C233	2016D0758 HZDD1668	Czech Republic
114	C123	2016D1974 HZDD828	Moldova	274	C305	Heinong52	China
115	C140	2016D4426 HZDD4449	Ukraine	275	C313	2016D2828 HZDD2231	Serbia
116	C143	2016D1190 HZDD4036	America	276	C316	2016D1432 HZDD155	America
117	C150	2016D1038 HZDD768	Romania	277	C319	2016D3698 HZDD3466	Japan
118	C152	2016D0184 HZDD1805	Romania	278	C322	2016D1231 HZDD1267	Russia
119	C154	2016D1025 HZDD565	Germany	279	C006	2016D2932 HZDD2361	America
120	C155	2016D1164 HZDD3476	Algeria	280	C008	2016D1079 HZDD1370	China
121	C156	2016D1207 HZDD4300	Sweden	281	C012	2016D1406 HZDD118	Japan
122	C161	2016D1162 HNDD3469	Japan	282	C016	2016D1052 HZDD919	Russia
123	C162	2016D1160 HNDD3464	Sweden	283	C018	2016D0196 HZDD1023	Russia
124	C164	2016D1048 HZDD881	Poland	284	C040	2016D2370 HZDD1529	Russia
125	C165	2016D1091 HZDD1822	Czech Republic	285	C044	2016D2138 HZDD1113	China
126	C166	2016D1167 HZDD3481	Algeria	286	C050	2016D1794 HZDD619	Russia
127	C168	2016D1978 HZDD834	Moldova	287	C056	2016D1135 HZDD2927	China
128	C169	2016D3262 HZDD2828	Serbia	288	C060	2016D1118 HZDD2806	Russia
129	C170	2016D0807 HZDD3381	China	289	C064	2016D1018 HZDD477	Australia
130	C171	2016D1013 HZDD382	Germany	290	C078	2016D1480 HZDD229	Russia
131	C172	2016D1161 HZDD3465	Sweden	291	C086	2016D2536 HZDD1799	Romania
132	C178	2016D3263 HZDD2829	Serbia	292	C094	2016D0940 HZDD2711	Belarus
133	C179	Keshan1	China	293	C250	2016D2759 HZDD2128	Russia
134	C185	2016D1173 HZDD3500	Algeria	294	C258	P768 HZDD1365	America
135	C189	2016D2469 HZDD1693	America	295	C276	2016D1399 HZDD108	America
136	C191	2016D2686 HZDD2025	Czech Republic	296	C302	2016D1333 HZDD39	Russia
137	C192	2016D2687 HZDD2026	Czech Republic	297	C304	Dongnong51	China
138	C194	2016D2890 HZDD2310	France	298	C312	2016D2676 HZDD2013	Ukraine
139	C195	2016D3157 HZDD2702	Moldova	299	C314	2016D0654 HZDD3733	America
140	C198	2016D3965 HZDD3834	America	300	C318	2016D2864 HZDD2282	Moldova
141	C199	2016D4379 HZDD4398	Canada	301	C321	2016D4302 HZDD4312	Canada
142	C200	2016D0524 HZDD3652	Germany	302	C002	2016D1070 HZDD1101	China
143	C205	2016D2883 HZDD2302	France	303	C004	2016D1548 HZDD314	Italy
144	C214	Henong76	China	304	C010	2016D4096 HZDD4004	Moldova
145	C220	Mudou10	China	305	C022	2016D3069 HZDD2583	Algeria
146	C229	2016D2782 HZDD2156	America	306	C024	2016D1986 HZDD845	Ukraine
147	C231	2016D4432 HZDD4456	Romania	307	C042	2016D2226 HZDD1292	Moldova
148	C246	2016D3713 HZDD3502	Serbia	308	C068	2016D1059 HZDD998	China
149	C264	2016D4474 HZDD4498	Romania	309	C124	2016D3619 HZDD3346	China
150	C277	2016D1446 HZDD195	America	310	C306	Kenfeng16	China
151	C285	2016D2978 HZDD2432	Ukraine	311	C307	Longdou4	China
152	C320	2016D4080 HZDD3979	America	312	C020	2016D3170 HZDD2720	Hungary
153	C111	2016D1042 HZDD821	Czech Republic	313	C026	2016D0255 HZDD1119	China
154	C114	2016D1559 HZDD333	Moldova	314	C028	2016D3700 HZDD3468	Japan
155	C184	2016D0536 HZDD4155	Australia	315	C058	2016D1119 HZDD2807	Russia
156	C197	2016D3712 HZDD3498	Algeria	316	C070	2016D0187 HZDD2502	Russia
157	C122	2016D2561 HZDD1838	Canada	317	C128	2016D2334 HZDD1477	China
158	C134	2016D1891 HZDD731	France	318	C132	2016D1369 HZDD74	Moldova
159	C146	Jinyuan55	China	319	C323	2016D0026 HZSLD1120	Russia
160	C180	Longxiaolidou2	China	320	C324	2016D1245 HZDD2889	China

Table A2. Significance analysis of soybean main stem node number, bottom pod height, and branch number in six environments.

1. Significance Analysis of Soybean Main Stem Node Number in 6 Environments.						
Environment (I)	**Environment (J)**	**Mean Difference (I-J)**	**Standard Error**	**Significance**	**95% Confidence Interval**	
					Lower Limit	**Upper Limit**
E1	E2	−0.68419 *	0.21	0.001	−1.10	−0.27
	E3	−1.06531 *	0.21	0.000	−1.48	−0.65
	E4	0.87787 *	0.21	0.000	0.47	1.29
	E5	0.11	0.21	0.602	−0.30	0.52
	E6	0.07	0.21	0.737	−0.34	0.48
E2	E1	0.68419 *	0.21	0.001	0.27	1.10
	E3	−0.38	0.21	0.069	−0.79	0.03
	E4	1.56206 *	0.21	0.000	1.15	1.97
	E5	0.79350 *	0.21	0.000	0.38	1.20
	E6	0.75466 *	0.21	0.000	0.34	1.17
E3	E1	1.06531 *	0.21	0.000	0.65	1.48
	E2	0.38	0.21	0.069	−0.03	0.79
	E4	1.94319 *	0.21	0.000	1.53	2.35
	E5	1.17463 *	0.21	0.000	0.76	1.59
	E6	1.13578 *	0.21	0.000	0.72	1.55
E4	E1	−0.87787 *	0.21	0.000	−1.29	−0.47
	E2	−1.56206 *	0.21	0.000	−1.97	−1.15
	E3	−1.94319 *	0.21	0.000	−2.35	−1.53
	E5	−0.76856 *	0.21	0.000	−1.18	−0.36
	E6	−0.80741 *	0.21	0.000	−1.22	−0.40
E5	E1	−0.11	0.21	0.602	−0.52	0.30
	E2	−0.79350 *	0.21	0.000	−1.20	−0.38
	E3	−1.17463 *	0.21	0.000	−1.59	−0.76
	E4	0.76856 *	0.21	0.000	0.36	1.18
	E6	−0.04	0.21	0.853	−0.45	0.37
E6	E1	−0.07	0.21	0.737	−0.48	0.34
	E2	−0.75466 *	0.21	0.000	−1.17	−0.34
	E3	−1.13578 *	0.21	0.000	−1.55	−0.72
	E4	0.80741 *	0.21	0.000	0.40	1.22
	E5	0.04	0.21	0.853	−0.37	0.45
2. Significance Analysis of Soybean Pod Height in 6 Environments.						
Environment (I)	**Environment (J)**	**Mean Difference (I-J)**	**Standard Error**	**Significance**	**95% Confidence Interval**	
					Lower Limit	**Upper Limit**
E1	E2	4.15147 *	0.48	0.00	3.21	5.09
	E3	1.18269 *	0.48	0.01	0.24	2.12
	E4	5.08622 *	0.48	0.00	4.15	6.03
	E5	−0.94	0.48	0.05	−1.88	0.00
	E6	−3.65003 *	0.48	0.00	−4.59	−2.71
E2	E1	−4.15147 *	0.48	0.00	−5.09	−3.21
	E3	−2.96878 *	0.48	0.00	−3.91	−2.03
	E4	0.93	0.48	0.05	−0.01	1.87
	E5	−5.09037 *	0.48	0.00	−6.03	−4.15
	E6	−7.80150 *	0.48	0.00	−8.74	−6.86
E3	E1	−1.18269 *	0.48	0.01	−2.12	−0.24
	E2	2.96878 *	0.48	0.00	2.03	3.91
	E4	3.90353 *	0.48	0.00	2.96	4.84
	E5	−2.12159 *	0.48	0.00	−3.06	−1.18
	E6	−4.83272 *	0.48	0.00	−5.77	−3.89

Table A2. *Cont.*

2. Significance Analysis of Soybean Pod Height in 6 Environments.						
Environment (I)	**Environment (J)**	**Mean Difference (I-J)**	**Standard Error**	**Significance**	**95% Confidence Interval**	
					Lower Limit	**Upper Limit**
E4	E1	−5.08622 *	0.48	0.00	−6.03	−4.15
	E2	−0.93	0.48	0.05	−1.87	0.01
	E3	−3.90353 *	0.48	0.00	−4.84	−2.96
	E5	−6.02512 *	0.48	0.00	−6.97	−5.09
	E6	−8.73625 *	0.48	0.00	−9.68	−7.80
E5	E1	0.94	0.48	0.05	0.00	1.88
	E2	5.09037 *	0.48	0.00	4.15	6.03
	E3	2.12159 *	0.48	0.00	1.18	3.06
	E4	6.02512 *	0.48	0.00	5.09	6.97
	E6	−2.71112 *	0.48	0.00	−3.65	−1.77
E6	E1	3.65003 *	0.48	0.00	2.71	4.59
	E2	7.80150 *	0.48	0.00	6.86	8.74
	E3	4.83272 *	0.48	0.00	3.89	5.77
	E4	8.73625 *	0.48	0.00	7.80	9.68
	E5	2.71112 *	0.48	0.00	1.77	3.65
3. Significance Analysis of Soybean Branching Number in 6 Environments.						
Environment (I)	**Environment (J)**	**Mean Difference (I-J)**	**Standard Error**	**Significance**	**95% Confidence Interval**	
					Lower Limit	**Upper Limit**
E1	E2	−1.04931 *	0.12	0.00	−1.29	−0.81
	E3	−0.32916 *	0.12	0.01	−0.57	−0.09
	E4	−1.06469 *	0.12	0.00	−1.31	−0.82
	E5	−0.31488 *	0.12	0.01	−0.56	−0.07
	E6	−0.44878 *	0.12	0.00	−0.69	−0.21
E2	E1	1.04931 *	0.12	0.00	0.81	1.29
	E3	0.72016 *	0.12	0.00	0.48	0.96
	E4	−0.02	0.12	0.90	−0.26	0.23
	E5	0.73444 *	0.12	0.00	0.49	0.98
	E6	0.60053 *	0.12	0.00	0.36	0.84
E3	E1	0.32916 *	0.12	0.01	0.09	0.57
	E2	−0.72016 *	0.12	0.00	−0.96	−0.48
	E4	−0.73553 *	0.12	0.00	−0.98	−0.49
	E5	0.01	0.12	0.91	−0.23	0.26
	E6	−0.12	0.12	0.33	−0.36	0.12
E4	E1	1.06469 *	0.12	0.00	0.82	1.31
	E2	0.02	0.12	0.90	−0.23	0.26
	E3	0.73553 *	0.12	0.00	0.49	0.98
	E5	0.74981 *	0.12	0.00	0.51	0.99
	E6	0.61591 *	0.12	0.00	0.38	0.86
E5	E1	0.31488 *	0.12	0.01	0.07	0.56
	E2	−0.73444 *	0.12	0.00	−0.98	−0.49
	E3	−0.01	0.12	0.91	−0.26	0.23
	E4	−0.74981 *	0.12	0.00	−0.99	−0.51
	E6	−0.13	0.12	0.28	−0.37	0.11
E6	E1	0.44878 *	0.12	0.00	0.21	0.69
	E2	−0.60053 *	0.12	0.00	−0.84	−0.36
	E3	0.12	0.12	0.33	−0.12	0.36
	E4	−0.61591 *	0.12	0.00	−0.86	−0.38
	E5	0.13	0.12	0.28	−0.11	0.37

* Significant difference at $p < 0.05$.

Table A3. SNP loci located by MLM analysis method for the number of main stem nodes.

SNP	CHROM	POS	Effect	SE	Env.
Chr19.44532590	Chr19	44532590	−1.591641594	0.269548872	E1
Chr19.44544551	Chr19	44544551	−1.646482973	0.276852835	E1
Chr19.44553045	Chr19	44553045	−1.639898908	0.276182318	E1
Chr19.44565523	Chr19	44565523	−1.588502512	0.270752574	E1
Chr19.44576961	Chr19	44576961	−1.722565811	0.269548355	E1, E2
Chr19.44577127	Chr19	44577127	−1.655939418	0.275091589	E1
Chr19.44579861	Chr19	44579861	−1.683883501	0.274307136	E1
Chr19.44579904	Chr19	44579904	−1.68785785	0.277215324	E1
Chr19.44649344	Chr19	44649344	−1.545748592	0.264223584	E1
Chr19.44698321	Chr19	44698321	−1.597329448	0.268285306	E1
Chr19.44568904	Chr19	44568904	−1.12873311	0.191099787	E2

E1 was planted in Harbin in 2018, and E2 was planted in Harbin in 2019.

Table A4. SNP loci located at the bottom pod height using MLM analysis method.

SNP	CHROM	POS	Effect	SE	Env.
Chr08.218444	Chr08	218444	−5.701518389	0.956528827	E3
Chr08.44018897	Chr08	44018897	13.3229696	2.097661364	E5
Chr08.44157110	Chr08	44157110	11.45072168	1.932836277	E5
Chr08.44289397	Chr08	44289397	13.32141264	2.082459225	E5

E3 was planted in Qing'an in 2018, E5 was planted in Xiangyang in 2021.

Table A5. SNP loci located by branching number under MLM analysis method.

SNP	CHROM	POS	Effect	SE	Env.
Chr08.21754137	Chr08	21754137	4.599298967	0.760812051	E3
Chr08.21754146	Chr08	21754146	6.544646908	0.923782272	E3
Chr14.6866752	Chr14	6866752	−5.599884266	0.936501031	E3

E3 was planted in Qing'an in 2018.

Table A6. Detailed information on candidate genes in the genome region of the main stem node number in the KEGG database.

Gene Stable ID	GO Term Accession	GO Term Name
Chr.19G033600	GO:0016020	membrane
Chr.19G033700	GO:0032259	methylation
Chr.19G033800	GO:0033356	UDP-L-arabinose metabolic process
Chr.19G033900	GO:0016491	oxidoreductase activity
Chr.19G034000	GO:0016491	oxidoreductase activity
Chr.19G034200	GO:0047134	protein-disulfide reductase (NAD(P)) activity
Chr.19G034300	GO:0070042	rRNA (uridine-N3-)-methyltransferase activity
Chr.19G034000	GO:0016491	oxidoreductase activity
Chr.19G034200	GO:0047134	protein-disulfide reductase (NAD(P)) activity
Chr.19G034300	GO:0070042	rRNA (uridine-N3-)-methyltransferase activity
Chr.19G034500	GO:0003700	DNA-binding transcription factor activity
Chr.19G034600	GO:0000978	RNA polymerase II cis-regulatory region sequence-specific DNA binding
Chr.19G185700	GO:0050790	regulation of catalytic activity
Chr.19G185800	GO:0009941	chloroplast envelope
Chr.19G185900	GO:0009570	chloroplast stroma
Chr.19G186000	GO:0046166	glyceraldehyde-3-phosphate biosynthetic process
Chr.19G186100	GO:0006833	water transport

Table A6. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.19G186200	GO:0003676	nucleic acid binding
Chr.19G186300	GO:0012505	endomembrane system
Chr.19G186400	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kinase activity
Chr.19G186600	GO:0005515	protein binding
Chr.19G186700	GO:0071555	cell wall organization
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G186100	GO:0006833	water transport
Chr.19G186200	GO:0003676	nucleic acid binding
Chr.19G186300	GO:0005794	Golgi apparatus
Chr.19G186300	GO:0012505	endomembrane system
Chr.19G186400	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kinase activity
Chr.19G186600	GO:0005515	protein binding
Chr.19G186700	GO:0071555	cell wall organization
Chr.19G187000	GO:0080043	quercetin 3-O-glucosyltransferase activity
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G188100	GO:0016491	oxidoreductase activity
Chr.19G186200	GO:0003676	nucleic acid binding
Chr.19G186300	GO:0012505	endomembrane system
Chr.19G186400	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kinase activity
Chr.19G186600	GO:0005515	protein binding
Chr.19G186700	GO:0071555	cell wall organization
Chr.19G187000	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187100	GO:0080043	quercetin 3-O-glucosyltransferase activity
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0016758	hexosyltransferase activity
Chr.19G187500	GO:0080043	quercetin 3-O-glucosyltransferase activity
Chr.19G187600	GO:0016758	hexosyltransferase activity
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G188100	GO:0016491	oxidoreductase activity
Chr.19G188300	GO:0015031	protein transport
Chr.19G186400	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kinase activity
Chr.19G186600	GO:0005515	protein binding
Chr.19G186700	GO:0071555	cell wall organization
Chr.19G187000	GO:0080043	quercetin 3-O-glucosyltransferase activity
Chr.19G187100	GO:0080043	quercetin 3-O-glucosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity

Table A6. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G188100	GO:0046872	metal ion binding
Chr.19G188400	GO:0071944	cell periphery
Chr.19G188500	GO:0034641	cellular nitrogen compound metabolic process
Chr.19G186300	GO:0012505	endomembrane system
Chr.19G186400	GO:0016308	1-phosphatidylinositol-4-phosphate 5-kinase activity
Chr.19G186600	GO:0005515	protein binding
Chr.19G186700	GO:0071555	cell wall organization
Chr.19G187000	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187800	GO:0016021	integral component of membrane
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0046872	metal ion binding
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G188100	GO:0016491	oxidoreductase activity
Chr.19G188300	GO:0046907	intracellular transport
Chr.19G188400	GO:0071944	cell periphery
Chr.19G188500	GO:0034641	cellular nitrogen compound metabolic process
Chr.19G187000	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187700	GO:0016758	hexosyltransferase activity
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0046872	metal ion binding
Chr.19G188100	GO:0046872	metal ion binding
Chr.19G188300	GO:0046907	intracellular transport
Chr.19G188400	GO:0071944	cell periphery
Chr.19G188500	GO:0034641	cellular nitrogen compound metabolic process
Chr.19G188600	GO:0006508	proteolysis
Chr.19G188700	GO:0005737	cytoplasm
Chr.19G189000	GO:0009658	chloroplast organization
Chr.19G189100	GO:0000209	protein polyubiquitination
Chr.19G187000	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187100	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187500	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187800	GO:0052689	carboxylic ester hydrolase activity
Chr.19G187900	GO:0003682	chromatin binding
Chr.19G188100	GO:0016491	oxidoreductase activity
Chr.19G188300	GO:0015031	protein transport
Chr.19G188400	GO:0071944	cell periphery
Chr.19G188500	GO:0005737	cytoplasm
Chr.19G188600	GO:0016021	integral component of membrane
Chr.19G188700	GO:0005737	cytoplasm
Chr.19G188800	GO:0005737	cytoplasm
Chr.19G188900	GO:0006468	protein phosphorylation
Chr.19G189000	GO:0009658	chloroplast organization
Chr.19G189100	GO:0000209	protein polyubiquitination

Table A6. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.19G189200	GO:0005515	protein binding
Chr.19G189300	GO:0005794	Golgi apparatus
Chr.19G188300	GO:0046907	intracellular transport
Chr.19G188400	GO:0071944	cell periphery
Chr.19G188500	GO:0034641	cellular nitrogen compound metabolic process
Chr.19G188600	GO:0006508	proteolysis
Chr.19G188700	GO:0005737	cytoplasm
Chr.19G188900	GO:0006468	protein phosphorylation
Chr.19G189000	GO:0009658	chloroplast organization
Chr.19G189100	GO:0000209	protein polyubiquitination
Chr.19G189200	GO:0005515	protein binding
Chr.19G189300	GO:0005794	Golgi apparatus
Chr.19G189500	GO:0016020	membrane
Chr.19G189600	GO:0002181	cytoplasmic translation
Chr.19G189700	GO:0006357	regulation of transcription by RNA polymerase II
Chr.19G189900	GO:0050155	ornithine(lysine) transaminase activity
Chr.19G190200	GO:0016020	membrane
Chr.19G190300	GO:0045927	positive regulation of growth
Chr.19G187700	GO:0008194	UDP-glycosyltransferase activity
Chr.19G190500	GO:0006470	protein dephosphorylation
Chr.19G190600	GO:0016757	glycosyltransferase activity
Chr.19G033500	GO:0010215	cellulose microfibril organization
Chr.19G190100	GO:0019752	carboxylic acid metabolic process
Chr.19G189800	GO:0071902	positive regulation of protein serine/threonine kinase activity
Chr.19G187000	GO:0008194	UDP-glycosyltransferase activity
Chr.19G188800	GO:0005681	spliceosomal complex
Chr.19G187600	GO:0008194	UDP-glycosyltransferase activity
Chr.19G187400	GO:0008194	UDP-glycosyltransferase activity

Table A7. Detailed information on candidate genes for the basal pod height genome region in the KEGG database.

Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G320500	GO:0005515	protein binding
Chr.08G320700	GO:0006351	transcription, DNA-templated
Chr.08G320900	GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
Chr.08G321100	GO:0008233	peptidase activity
Chr.08G321200	GO:0008233	peptidase activity
Chr.08G321300	GO:0016020	membrane
Chr.08G321400	GO:0006508	proteolysis
Chr.08G321500	GO:0004190	aspartic-type endopeptidase activity
Chr.08G321600	GO:0016020	membrane
Chr.08G321700	GO:0016020	membrane
Chr.08G321800	GO:0005515	protein binding
Chr.08G321900	GO:0005634	nucleus
Chr.08G322000	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322200	GO:0000911	cytokinesis by cell plate formation
Chr.08G322300	GO:0032259	methylation
Chr.08G322000	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322300	GO:0032259	methylation
Chr.08G322400	GO:0008289	lipid binding
Chr.08G322500	GO:0016020	membrane

Table A7. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G322600	GO:0015031	protein transport
Chr.08G322700	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322900	GO:1902600	proton transmembrane transport
Chr.08G323000	GO:0071806	protein transmembrane transport
Chr.08G323100	GO:0006629	lipid metabolic process
Chr.08G323200	GO:0006952	defense response
Chr.08G323300	GO:0004359	glutaminase activity
Chr.08G323400	GO:0006629	lipid metabolic process
Chr.08G323500	GO:0003678	DNA helicase activity
Chr.08G323600	GO:0006629	lipid metabolic process
Chr.08G323700	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322000	GO:0000911	cytokinesis by cell plate formation
Chr.08G322200	GO:0032259	methylation
Chr.08G322300	GO:0008289	lipid binding
Chr.08G322400	GO:0016020	membrane
Chr.08G322500	GO:0015031	protein transport
Chr.08G322600	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322700	GO:0048285	organelle fission
Chr.08G322800	GO:1902600	proton transmembrane transport
Chr.08G322900	GO:0071806	protein transmembrane transport
Chr.08G323000	GO:0006629	lipid metabolic process
Chr.08G323100	GO:0006952	defense response
Chr.08G323200	GO:0005737	cytoplasm
Chr.08G323300	GO:0006629	lipid metabolic process
Chr.08G323400	GO:0003678	DNA helicase activity
Chr.08G323500	GO:0006629	lipid metabolic process
Chr.08G323600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322700	GO:0048285	organelle fission
Chr.08G322800	GO:1902600	proton transmembrane transport
Chr.08G322900	GO:0071806	protein transmembrane transport
Chr.08G323000	GO:0006629	lipid metabolic process
Chr.08G323100	GO:0006952	defense response
Chr.08G323200	GO:0004359	glutaminase activity
Chr.08G323300	GO:0006629	lipid metabolic process
Chr.08G323400	GO:0003678	DNA helicase activity
Chr.08G323500	GO:0006629	lipid metabolic process
Chr.08G323600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity

Table A7. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:0006629	lipid metabolic process
Chr.08G325600	GO:0016020	membrane
Chr.08G323600	GO:0016020	membrane
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G325600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0005504	fatty acid binding
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G322200	GO:0000911	cytokinesis by cell plate formation
Chr.08G321000	GO:0003677	DNA binding
Chr.08G322800	GO:0048285	organelle fission
Chr.08G323800	GO:0005634	nucleus

Table A8. Detailed information on candidate genes for the basal branch number genome region in the KEGG database.

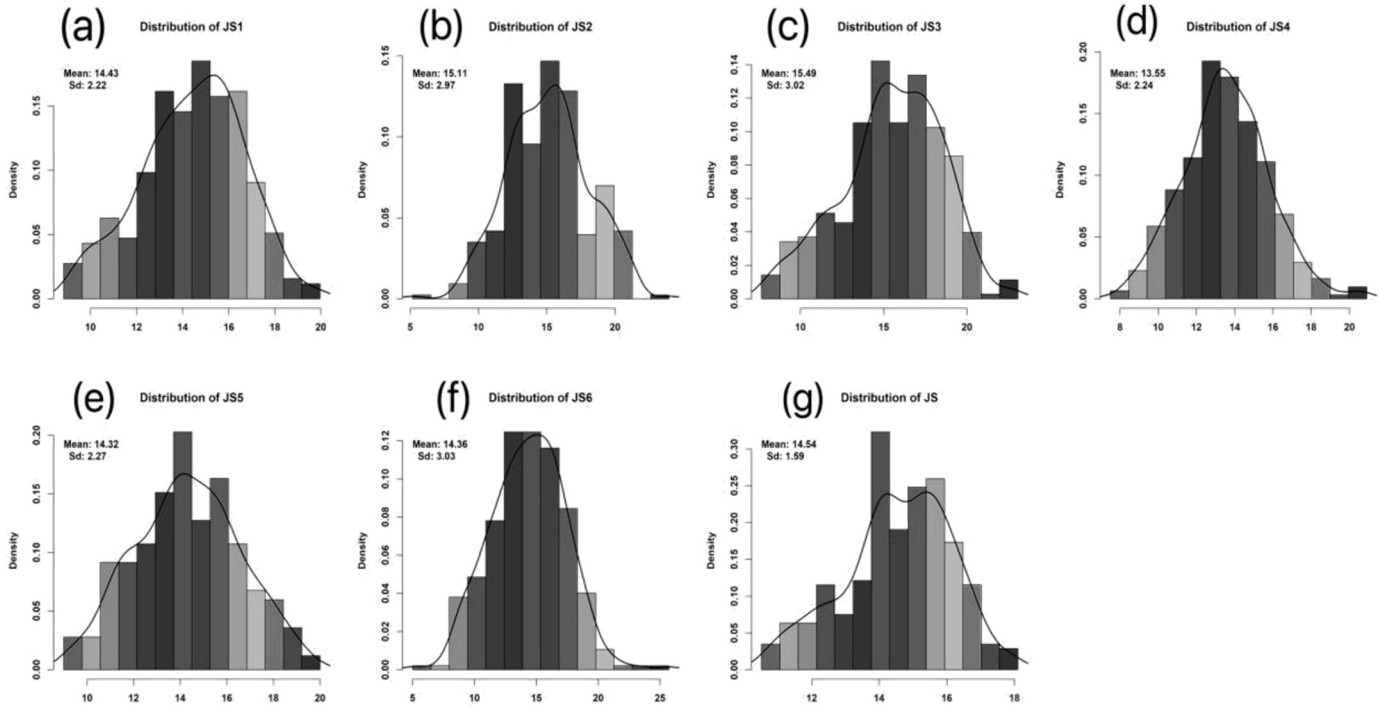
Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G320500	GO:0005515	protein binding
Chr.08G320700	GO:0006351	transcription, DNA-templated
Chr.08G320900	GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
Chr.08G321100	GO:0008233	peptidase activity
Chr.08G321200	GO:0008233	peptidase activity
Chr.08G321300	GO:0016020	membrane
Chr.08G321400	GO:0006508	proteolysis
Chr.08G321500	GO:0004190	aspartic-type endopeptidase activity
Chr.08G321600	GO:0016020	membrane
Chr.08G321700	GO:0016020	membrane
Chr.08G321800	GO:0005515	protein binding
Chr.08G321900	GO:0005634	nucleus
Chr.08G322000	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322200	GO:0000911	cytokinesis by cell plate formation
Chr.08G322300	GO:0032259	methylation
Chr.08G322000	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322300	GO:0032259	methylation
Chr.08G322400	GO:0008289	lipid binding
Chr.08G322500	GO:0016020	membrane
Chr.08G322600	GO:0015031	protein transport
Chr.08G322700	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322900	GO:1902600	proton transmembrane transport
Chr.08G323000	GO:0071806	protein transmembrane transport
Chr.08G323100	GO:0006629	lipid metabolic process
Chr.08G323200	GO:0006952	defense response
Chr.08G323300	GO:0004359	glutaminase activity
Chr.08G323400	GO:0006629	lipid metabolic process
Chr.08G323500	GO:0003678	DNA helicase activity
Chr.08G323600	GO:0006629	lipid metabolic process
Chr.08G323700	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0000774	adenyl-nucleotide exchange factor activity
Chr.08G322000	GO:0000911	cytokinesis by cell plate formation
Chr.08G322200	GO:0032259	methylation
Chr.08G322300	GO:0008289	lipid binding
Chr.08G322400	GO:0016020	membrane
Chr.08G322500	GO:0015031	protein transport
Chr.08G322600	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322700	GO:0048285	organelle fission
Chr.08G322800	GO:1902600	proton transmembrane transport
Chr.08G322900	GO:0071806	protein transmembrane transport
Chr.08G323000	GO:0006629	lipid metabolic process
Chr.08G323100	GO:0006952	defense response
Chr.08G323200	GO:0005737	cytoplasm
Chr.08G323300	GO:0006629	lipid metabolic process
Chr.08G323400	GO:0003678	DNA helicase activity
Chr.08G323500	GO:0006629	lipid metabolic process

Table A8. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G323600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0016251	RNA polymerase II general transcription initiation factor activity
Chr.08G322700	GO:0048285	organelle fission
Chr.08G322800	GO:1902600	proton transmembrane transport
Chr.08G322900	GO:0071806	protein transmembrane transport
Chr.08G323000	GO:0006629	lipid metabolic process
Chr.08G323100	GO:0006952	defense response
Chr.08G323200	GO:0004359	glutaminase activity
Chr.08G323300	GO:0006629	lipid metabolic process
Chr.08G323400	GO:0003678	DNA helicase activity
Chr.08G323500	GO:0006629	lipid metabolic process
Chr.08G323600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:0006629	lipid metabolic process
Chr.08G325600	GO:0016020	membrane
Chr.08G323600	GO:0016020	membrane
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G325600	GO:0016020	membrane
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0009627	systemic acquired resistance
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity

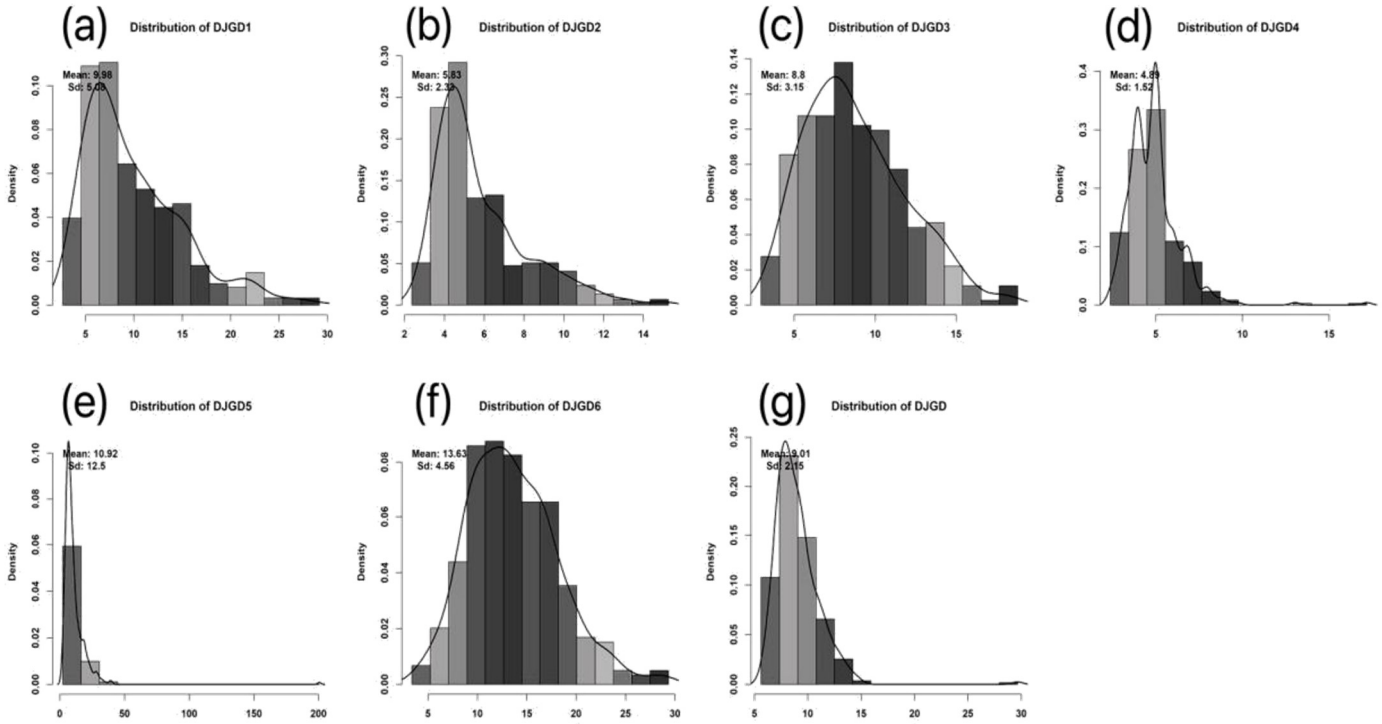
Table A8. *Cont.*

Gene Stable ID	GO Term Accession	GO Term Name
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G323800	GO:0005634	nucleus
Chr.08G324000	GO:0005504	fatty acid binding
Chr.08G324100	GO:0098656	anion transmembrane transport
Chr.08G324200	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324300	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324400	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324500	GO:0008889	glycerophosphodiester phosphodiesterase activity
Chr.08G324600	GO:0016763	pentosyltransferase activity
Chr.08G324800	GO:0016763	pentosyltransferase activity
Chr.08G324900	GO:0022626	cytosolic ribosome
Chr.08G325000	GO:0009691	cytokinin biosynthetic process
Chr.08G325100	GO:1990939	ATP-dependent microtubule motor activity
Chr.08G325200	GO:0000166	nucleotide binding
Chr.08G325400	GO:0003723	RNA binding
Chr.08G325500	GO:0035556	intracellular signal transduction
Chr.08G322200	GO:0000911	cytokinesis by cell plate formation
Chr.08G321000	GO:0003677	DNA binding
Chr.08G322800	GO:0048285	organelle fission
Chr.08G323800	GO:0005634	nucleus

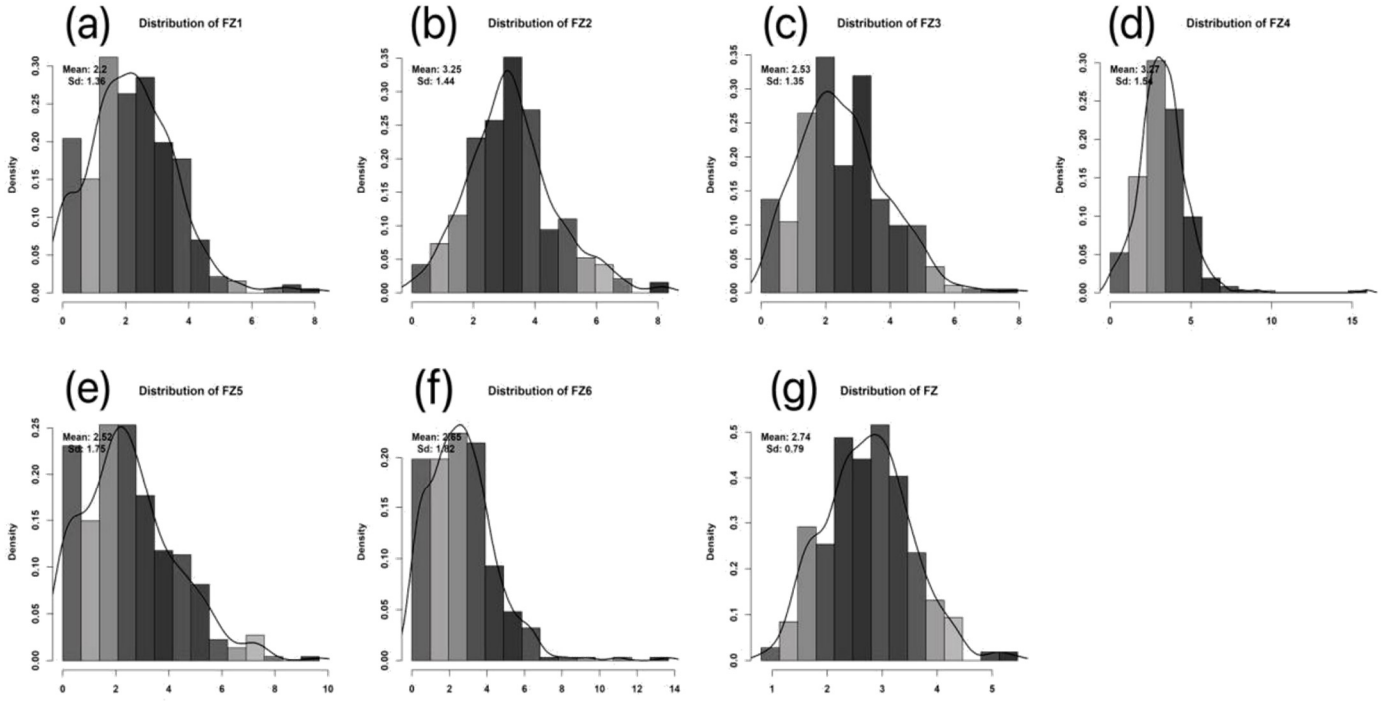


(**A**) Histogram of frequency distribution of soybean main stem nodes

Figure A1. *Cont.*



(**B**) Histogram of frequency distribution of soybean pod height



(**C**) Histogram of frequency distribution of soybean branching number

Figure A1. Frequency distribution of three yield-related traits across six environments and their best linear unbiased predictions (BLUP). Figures (**a**–**c**) are histograms of branch numbers for planting populations in Harbin, Mudanjiang, and Qing'an. Figure (**d**) is a histogram of the number of branches in the Yilan planting population in 2019. Figures (**e**,**f**) show the histogram distribution of branch numbers in Xiangyang and Minzhu planting communities in 2021. Figure (**g**) is a histogram of the number of branches after BLUP processing.

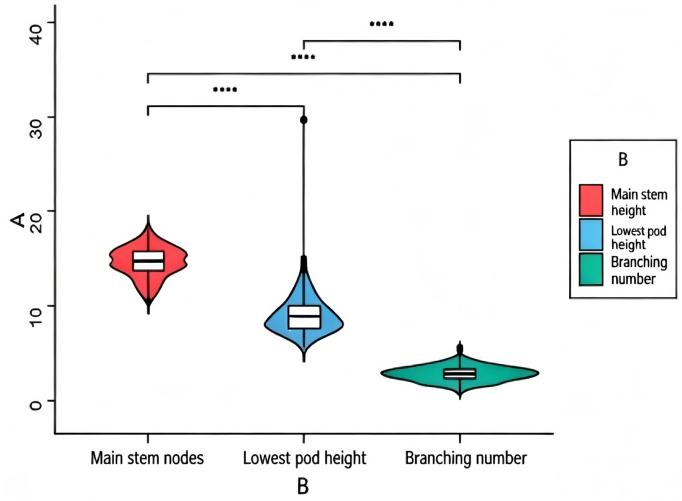


Figure A2. Correlation analysis between main traits of soybean natural population. Pearson correlation analysis among three key yield-related traits in the soybean natural population. Colors represent different traits: red for main stem node number, blue for lowest pod height, and green for branch number. A: y-axis scale (0–40); B: trait names. ●●●● $p < 0.001$.

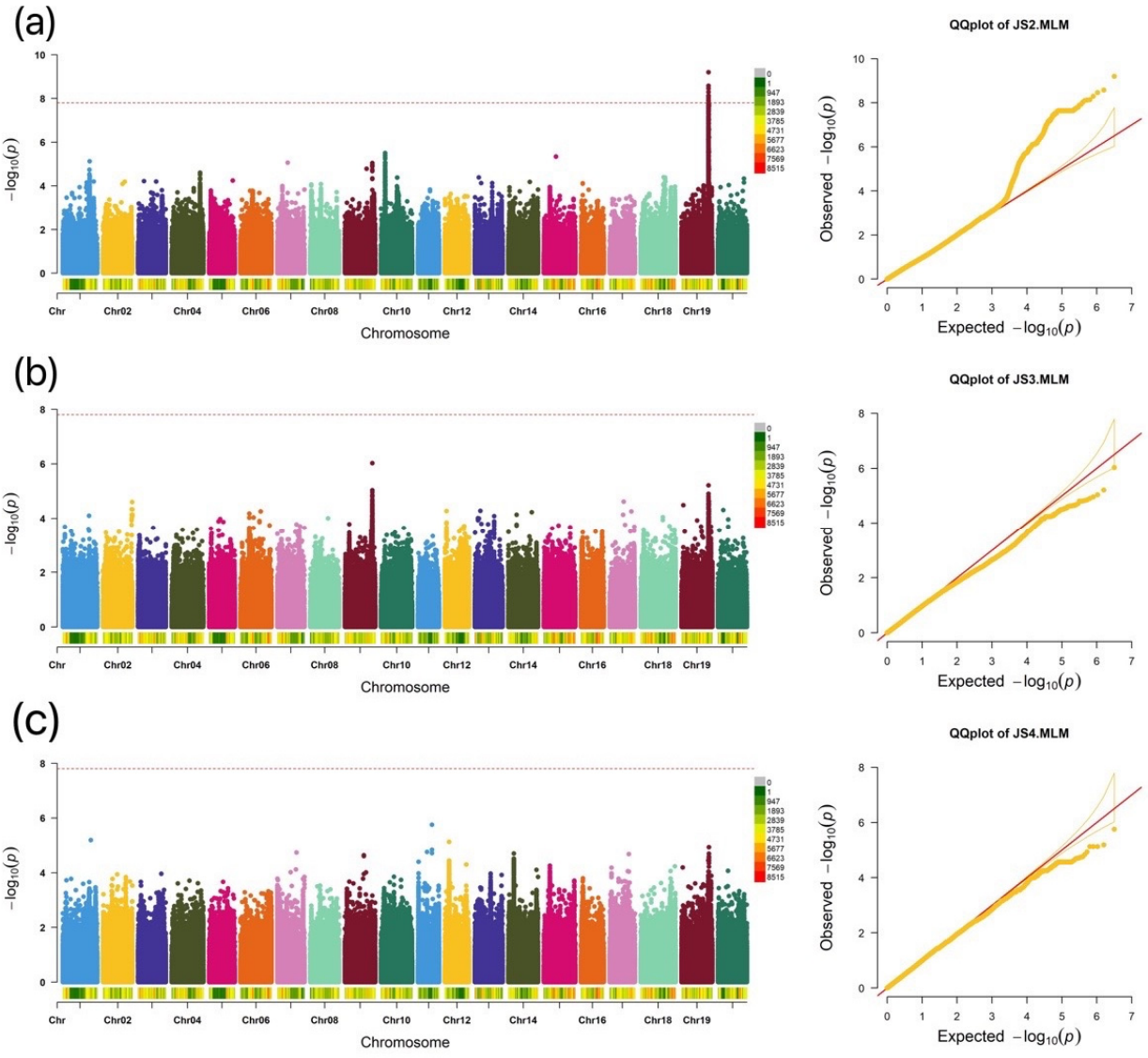


Figure A3. *Cont.*

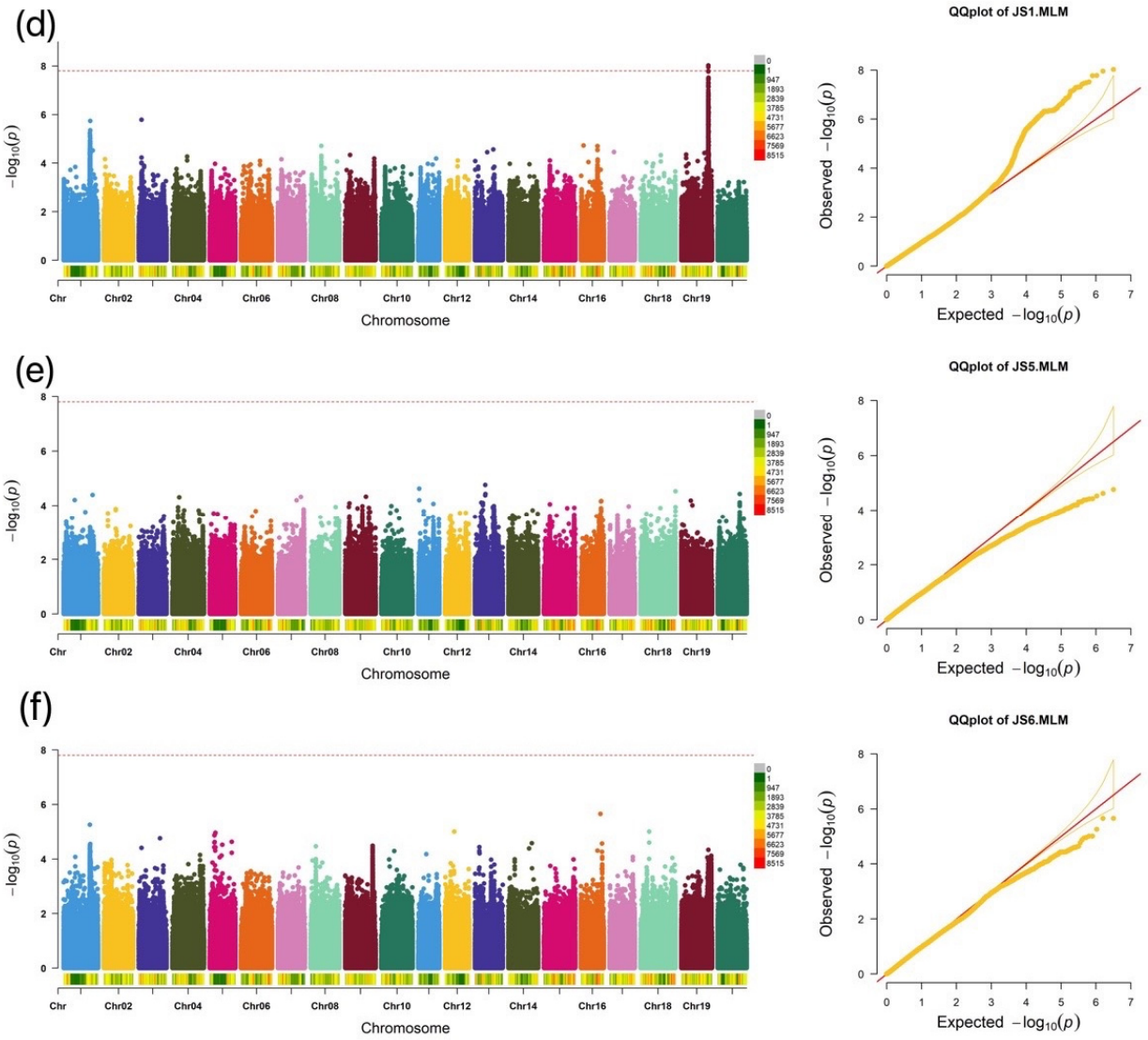


Figure A3. Manhattan and Q-Q plots of the correlation results of the number of main stem nodes at six locations. Figures (**a**–**c**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in the 2018 planting populations in Harbin, Mudanjiang, and Qing'an using the MLM method. Figure (**d**) shows the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in the Yilan planting population in 2019 using the MLM method. Figures (**e**,**f**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in Xiangyang and Minzhu planting populations in 2021 using MLM. Color coding and significance threshold are the same as in Figure 1.

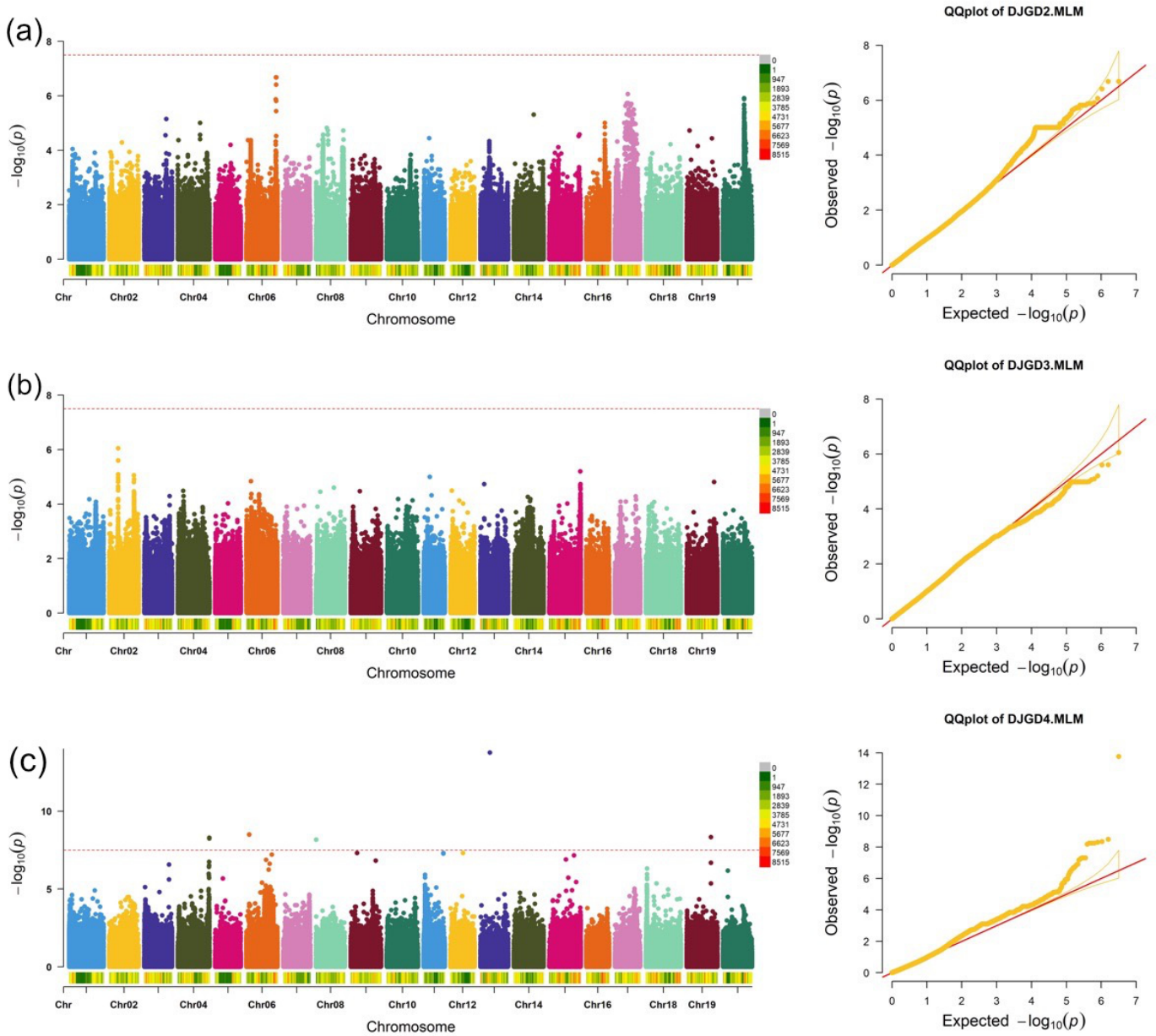


Figure A4. *Cont.*

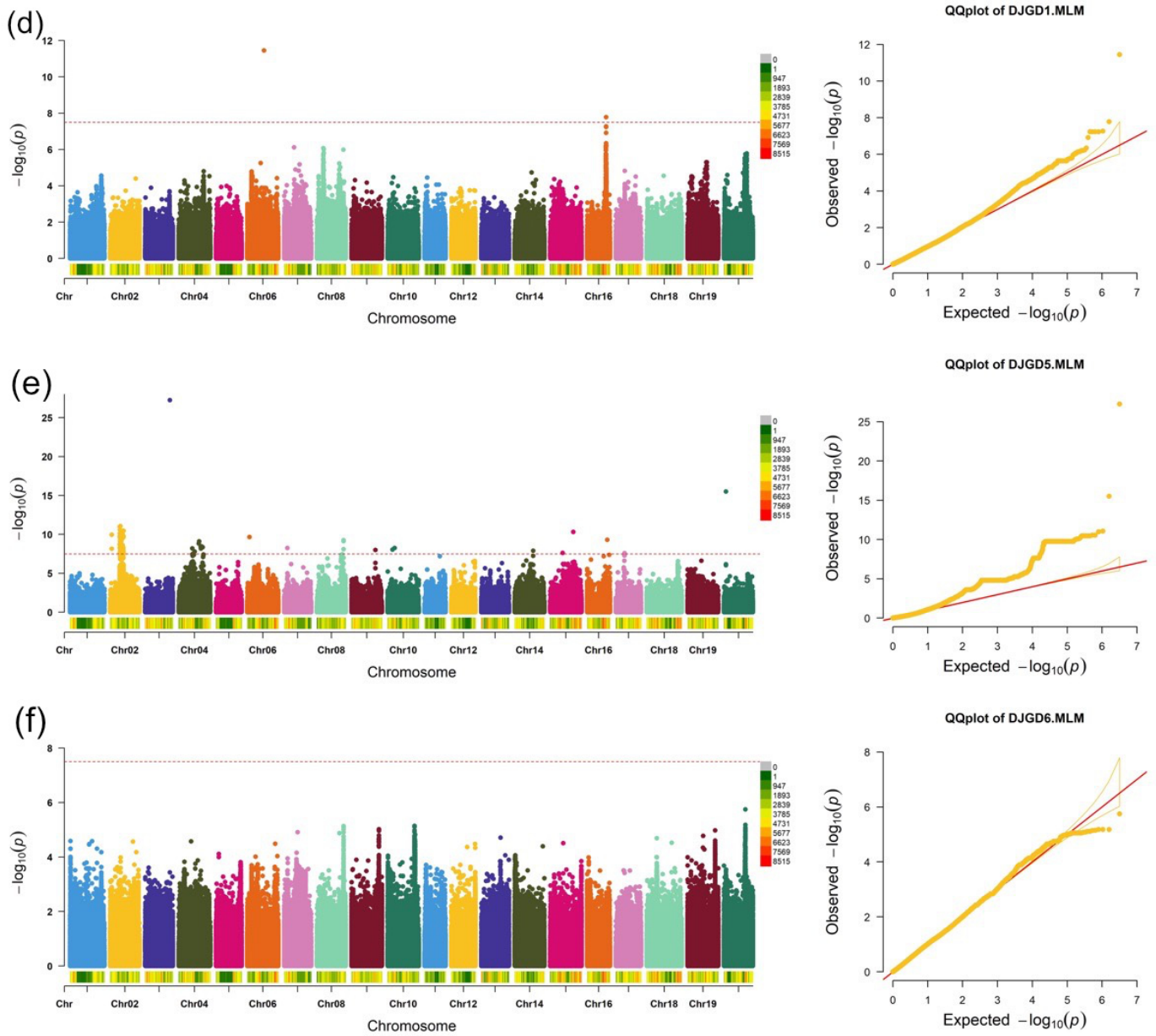


Figure A4. Manhattan and Q-Q plots of the correlation results of bottom pod height at six locations. Figures (**a**–**c**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in the 2018 planting populations in Harbin, Mudanjiang, and Qing'an using the MLM method. Figure (**d**) shows the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in the Yilan planting population in 2019 using the MLM method. Figures (**e**,**f**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of the main stem node number trait in Xiangyang and Minzhu planting populations in 2021 using MLM. Color coding and significance threshold are the same as in Figure 1.

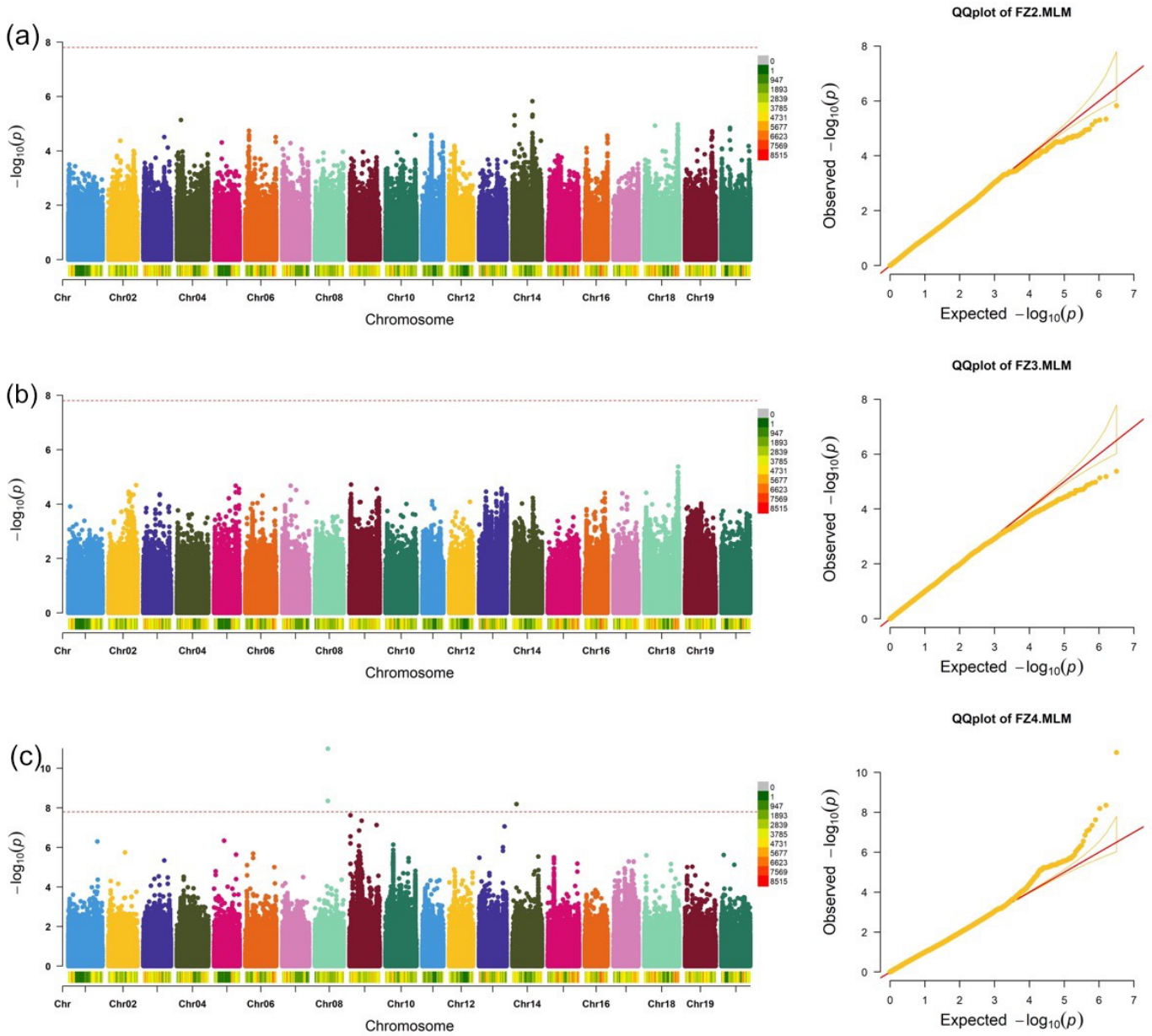


Figure A5. *Cont.*

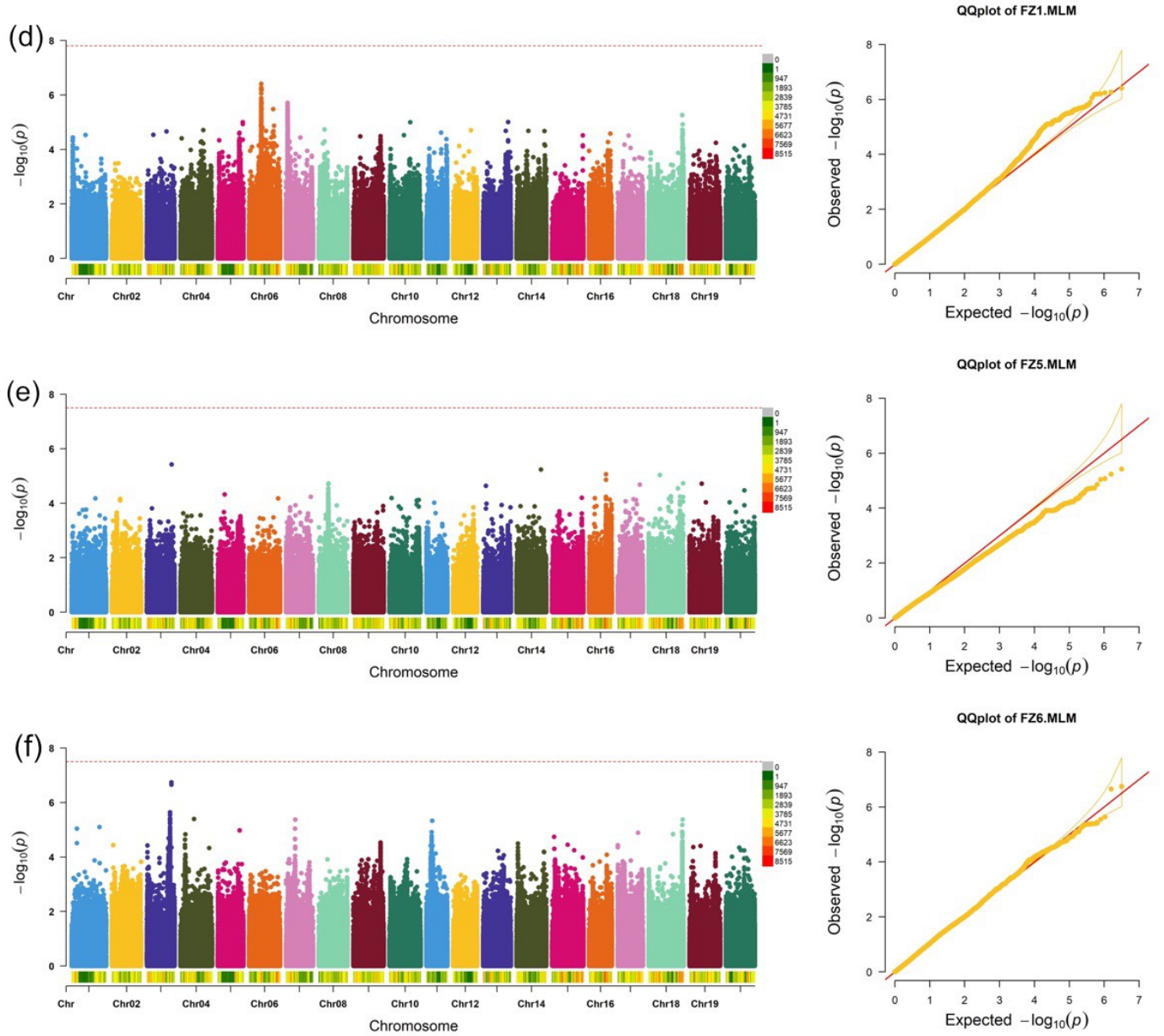


Figure A5. Manhattan and Q-Q plots of the correlation results of branch numbers at six locations. Figures (**a**–**c**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of branch number traits in the 2018 planting populations in Harbin, Mudanjiang, and Qing'an using MLM. Figure (**d**) shows the Manhattan and Q-Q plots obtained from the genome-wide association analysis of branch number traits in the Yilan planting population in 2019 using the MLM method. Figures (**e**,**f**) are the Manhattan and Q-Q plots obtained from the genome-wide association analysis of branch number traits in Xiangyang and Minzhu planting populations in 2021 using MLM.Color coding and significance threshold are the same as in Figure 1.

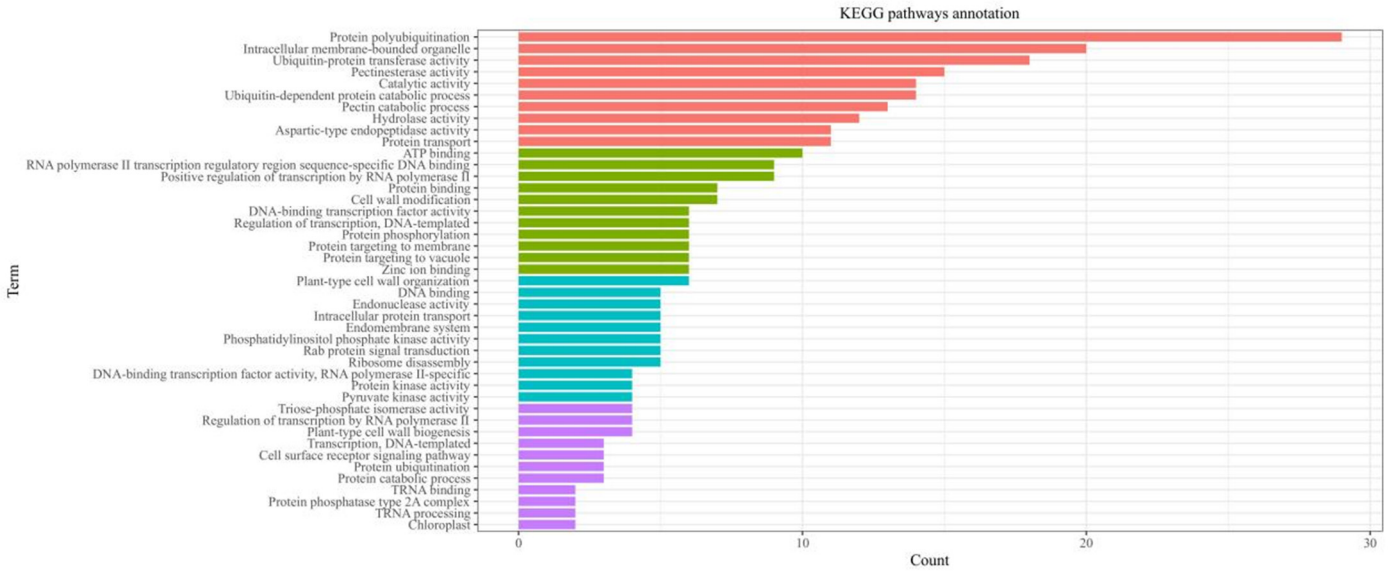


Figure A6. Summary of KEGG analysis results on the number of main stem nodes. Colors represent different GO categories: red for Molecular Function (MF), green for Biological Process (BP), blue for Cellular Component (CC), and purple for other functional terms. The x-axis shows the count of candidate genes enriched in each term, and the y-axis lists the corresponding functional terms. Only the top enriched terms are presented.

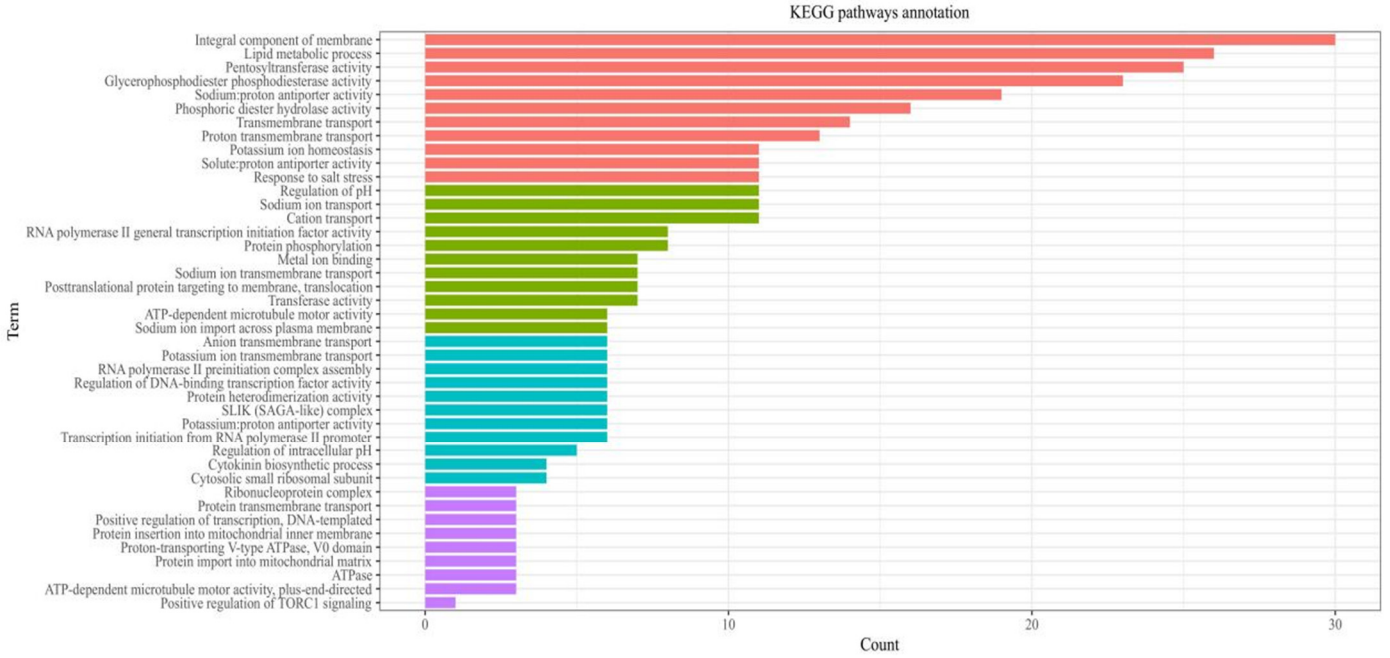


Figure A7. Summary of KEGG analysis results for bottom pod height. Colors represent different GO categories: red for Molecular Function (MF), green for Biological Process (BP), blue for Cellular Component (CC), and purple for other functional terms. The x-axis shows the count of candidate genes enriched in each term, and the y-axis lists the corresponding functional terms. Only the top enriched terms are presented.

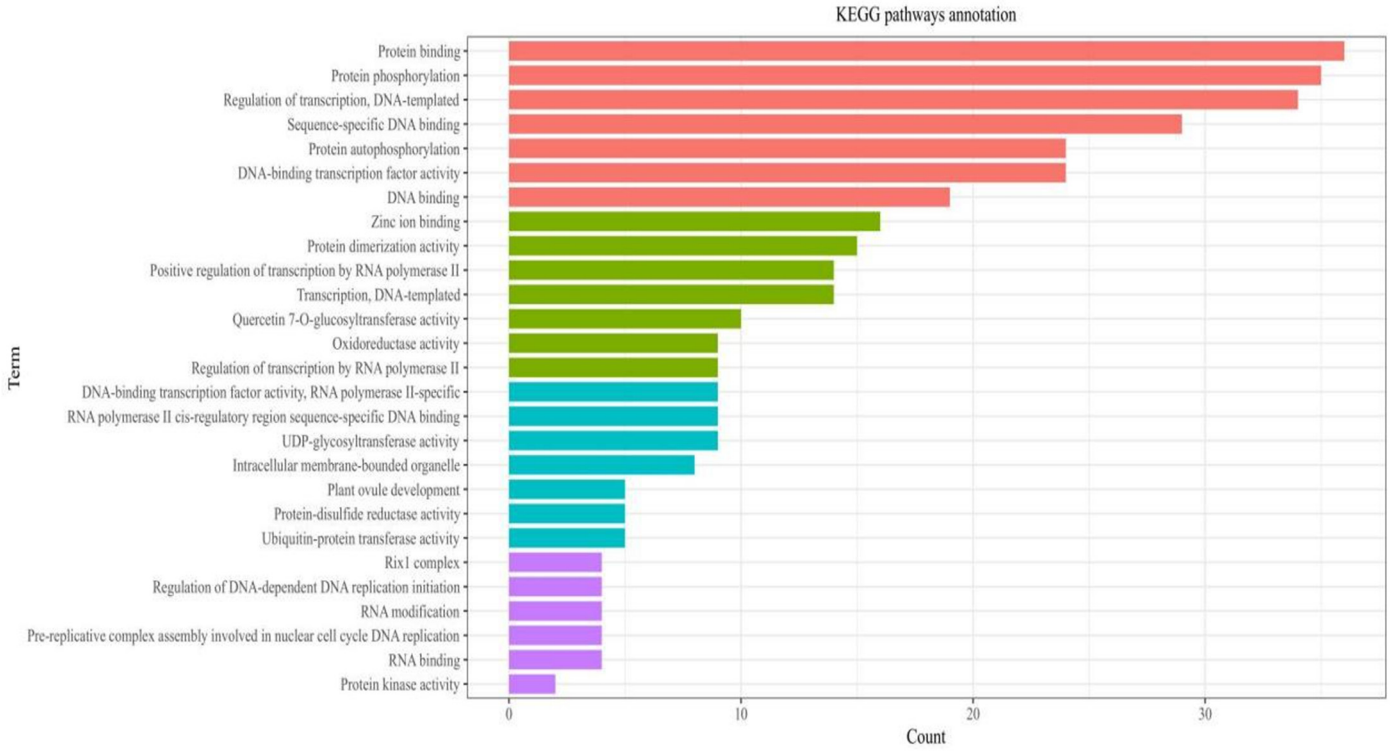


Figure A8. Summary of the analysis results of branch number KEGG. Colors represent different GO categories: red for Molecular Function (MF), green for Biological Process (BP), blue for Cellular Component (CC), and purple for other functional terms. The x-axis shows the count of candidate genes enriched in each term, and the y-axis lists the corresponding functional terms. Only the top enriched terms are presented.

References

1. Lai, M.C.; Lai, Z.Y.; Jhan, L.H.; Lai, Y.S.; Kao, C.F. Prioritization and Evaluation of Flooding Tolerance Genes in Soybean *Glycine max* (L.) Merr. *Front. Genet.* **2021**, *11*, 612131. [CrossRef]
2. Hu, D.; Zhang, H.; Du, Q.; Hu, Z.; Yang, Z.; Li, X.; Wang, J.; Huang, F.; Yu, D.; Wang, H.; et al. Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (*Glycine soja* Sieb. and Zucc.). *Planta* **2020**, *251*, 39. [CrossRef] [PubMed]
3. Li, X.; Wang, P.; Zhang, K.; Liu, S.; Qi, Z.; Fang, Y.; Wang, Y.; Tian, X.; Song, J.; Wang, J.; et al. Fine mapping QTL and mining genes for protein content in soybean by the combination of linkage and association analysis. *Theor. Appl. Genet.* **2021**, *134*, 1095–1122. [CrossRef] [PubMed]
4. Xia, L.W.; Wang, P.; Zhang, H.; Su, X.; Yu, T.; Li, H.; Hu, Y.; Li, C.; Mahfishan, S.; Rameez, R.V.; et al. QTL for Main Stem Node Number and Its Response to Plant Densities in 144 Soybean FW-RILs. *Front. Plant Sci.* **2021**, *12*, 666796. [CrossRef] [PubMed]
5. Zhang, X.; Ding, W.; Xue, D.; Li, X.; Zhou, Y.; Shen, J.; Feng, J.; Guo, N.; Qiu, L.; Xing, H.; et al. Genome-wide association studies of plant architecture-related traits and 100-seed weight in soybean landraces. *BMC Genom. Data* **2021**, *22*, 10. [CrossRef]
6. Zhang, C.; Li, W.; Tan, C.; Huang, M.; Wu, H.; Liu, S.; Liu, H.; Li, X.; Miao, Y.; Liu, B.; et al. Natural allelic variation in *SW14* determines seed weight and quality in soybean. *Nat. Commun.* **2025**, *16*, 8070. [CrossRef]
7. Li, C.; Li, Y.H.; Li, Y.; Lu, H.; Hong, H.; Tian, Y.; Li, H.; Zhao, T.; Zhou, X.; Liu, J.; et al. A Domestication-Associated Gene *GmPRR3b* Regulates the Circadian Clock and Flowering Time in Soybean. *Mol. Plant* **2020**, *13*, 745–759. [CrossRef]
8. Li, X.; Xue, H.; Zhang, K.; Li, W.; Fang, Y.; Qi, Z.; Wang, Y.; Tian, X.; Song, J.; Li, W.; et al. Mapping QTLs for protein and oil content in soybean by removing the influence of related traits in a four-way recombinant inbred line population. *J. Agric. Sci.* **2019**, *157*, 659–675. [CrossRef]
9. Fang, C.; Ma, Y.; Wu, S.; Liu, Z.; Wang, Z.; Yang, R.; Hu, G.; Zhou, Z.; Yu, H.; Zhang, M.; et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **2017**, *18*, 161. [CrossRef]
10. Zhu, C.; Gore, M.; Buckler, E.S.; Yu, J. Status and prospects of association mapping in plants. *Plant Genome* **2008**, *1*, 5–13. [CrossRef]
11. Qu, S.; Hu, S.; Song, G.; Sun, H.; Liu, F.; Zhang, M.; Zhan, Y.; Li, Y.; Teng, W.; Li, H.; et al. Genome-wide association study and molecular marker development for resistance to soybean cyst nematode in soybean. *Plant Genome* **2025**, *18*, e70083. [CrossRef] [PubMed]

12. Sung, M.; Van, K.; Lee, S.; Nelson, R.; LaMantia, J.; Taliercio, E.; McHale, L.K.; Mian, M.A.R. Identification of SNP markers associated with soybean fatty acids contents by genome-wide association analyses. *Mol. Breed.* **2021**, *41*, 24. [CrossRef]
13. Zhang, T.; Wu, T.; Wang, L.; Jiang, B.; Zhen, C.; Yuan, S.; Hou, W.; Wu, C.; Han, T.; Sun, S. A combined linkage and GWAS analysis identified QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* **2019**, *20*, 5915. [CrossRef]
14. Zhang, S.; Hao, D.; Zhang, S.; Zhang, D.; Wang, H.; Du, H.; Kan, G.; Yu, D. Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Mol. Genet. Genom.* **2021**, *296*, 91–102. [CrossRef]
15. Hong, H.; Li, M.; Chen, Y.; Wang, H.; Wang, J.; Guo, B.; Gao, H.; Ren, H.; Yuan, M.; Han, Y.; et al. Genome-wide association studies for soybean epicotyl length in two environments using 3VmrMLM. *Front. Plant Sci.* **2022**, *13*, 1033120. [CrossRef]
16. Wang, J.; Hu, B.; Huang, S.; Hu, X.; Siyal, M.; Yang, C.; Zhao, H.; Yang, T.; Li, H.; Hou, Y.; et al. SNP-bin linkage analysis and genome-wide association study of plant height in soybean. *Crop Pasture Sci.* **2020**, *71*, 222–237. [CrossRef]
17. Dhanapal, A.P.; Ray, J.D.; Singh, S.K.; Hoyos-Villegas, V.; Smith, J.R.; Purcell, L.C.; King, C.A.; Cregan, P.B.; Song, Q.; Fritschi, F.B. Genome-wide association study (GWAS) of carbon isotope ratio (δ^{13}C) in diverse soybean [*Glycine max* (L.) Merr.] genotypes. *Theor. Appl. Genet.* **2015**, *128*, 73–91. [CrossRef]
18. Rani, R.; Raza, G.; Ashfaq, H.; Rizwan, M.; Razzaq, M.K.; Waheed, M.Q.; Shimelis, H.; Babar, A.D.; Arif, M. Genome-wide association study of soybean (*Glycine max* [L.] Merr.) germplasm for dissecting the quantitative trait nucleotides and candidate genes underlying yield-related traits. *Front. Plant Sci.* **2023**, *14*, 1229495. [CrossRef] [PubMed]
19. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
20. Valliyodan, B.; Cannon, S.B.; Bayer, P.E.; Shu, S.; Brown, A.V.; Ren, L.; Jenkins, J.; Chung, C.Y.-L.; Chan, T.-F.; Daum, C.S.; et al. Construction and comparison of three reference-quality genome assemblies for soybean. *Plant J.* **2019**, *100*, 1066–1082. [CrossRef] [PubMed]
21. Zhao, X.; Bao, D.; Wang, W.; Zhang, C.; Jing, Y.; Jiang, H.; Qiu, L.; Li, W.; Han, Y. Loci and candidate gene identification for soybean resistance to *Phytophthora* root rot race 1 in combination with association and linkage mapping. *Mol. Breed.* **2020**, *40*, 94. [CrossRef]
22. Waly, E.F.; Zied, A.A.A.; Mourad, A.H.K.; Abdelghany, M. Identification of high-yielding and stable Egyptian soybean genotypes for breeding across varied environments. *BMC Plant Biol.* **2026**, *26*, 168. [CrossRef] [PubMed]
23. Cai, L.; Guo, Z.; Ding, J.; Gai, Z.; Liu, J.; Meng, Q.; Yang, X.; Zhang, N.; Wang, Q. Genome-Wide Identification and Exogenous Hormone and Stress Response Expression Analysis of the GARP Gene Family in Soybean (*Glycine max*). *Agriculture* **2024**, *14*, 2109. [CrossRef]
24. Zhao, D.; Zheng, H.; Li, J.; Wan, M.; Shu, K.; Wang, W.; Hu, X.; Hu, Y.; Qiu, L.; Wang, X. Natural Variation in the Promoter of *GmSPL9d* Affects Branch Number in Soybean. *Int. J. Mol. Sci.* **2024**, *25*, 5991. [CrossRef]
25. Borgohain, T.; Suma, R.; Muttappagol, M.; Saikia, B.; Keithellakpam, A.; Laskar, A.; Hiremath, S.S.; Basu, U.; Velmurugan, N.; Palakolanu, S.R.; et al. Precision breeding in a changing climate: Unlocking resilience through omics and gene editing. *Funct. Integr. Genom.* **2026**, *26*, 26. [CrossRef] [PubMed]
26. Dhanapal, P.A.; Li, Y.M.; Kim, H.A.; Fritschi, F.B. Genome-Wide Association Study of Topsoil Root System Architecture in Field-Grown Soybean [*Glycine max* (L.) Merr.]. *Front. Plant Sci.* **2021**, *11*, 590179. [CrossRef]
27. Maldonado dos Santos, J.V.; Ferreira, E.G.C.; Passianotto, A.L.D.L.; Brumer, B.B.; Santos, A.B.D.; Soares, R.M.; Torkamaneh, D.; Arias, C.A.A.; Belzile, F.; Abdelnoor, R.V.; et al. Association mapping of a locus that confers southern stem canker resistance in soybean and SNP marker development. *BMC Genom.* **2019**, *20*, 798. [CrossRef]
28. Du, W.; Ning, L.; Liu, Y.; Zhang, S.; Yang, Y.; Wang, Q.; Chao, S.; Yang, H.; Huang, F.; Cheng, H.; et al. Identification of loci and candidate gene *GmSPX-RING1* responsible for phosphorus efficiency in soybean via genome-wide association analysis. *BMC Genom.* **2020**, *21*, 725. [CrossRef]
29. Zha, B.; Zhang, C.; Yuan, R.; Zhao, K.; Sun, J.; Liu, X.; Wang, X.; Zhang, F.; Zhang, B.; Lamlom, S.F.; et al. Integrative QTL mapping and candidate gene analysis for main stem node number in soybean. *BMC Plant Biol.* **2025**, *25*, 422. [CrossRef]
30. Itoh, Y.; Nagano, H.; Sato, Y. Analysis of genetic property of phenotypic plasticity evolution in asian cultivated and wild rice with special reference to year environments in one of the northern-limit regions of rice cultivation. *Cereal Res. Commun.* **2025**, *54*, 75–96. [CrossRef]
31. Crevelari, J.A.; Souza, Y.P.D.; Santos, J.S.; Ambrósio, M.; Gonçalves, V.M.L.; Pereira, M.G. Adaptability and stability of corn hybrids for silage via genotype and genotype × environment interaction biplot. *Agron. J.* **2023**, *115*, 687–697. [CrossRef]
32. Jin, H.; Yang, X.; Zhao, H.; Song, X.; Tsvetkov, Y.D.; Wu, Y.; Gao, Q.; Zhang, R.; Zhang, J. Genetic analysis of protein content and oil content in soybean by genome-wide association study. *Front. Plant Sci.* **2023**, *14*, 1182771. [CrossRef]
33. Liu, X.; Li, C.; Cao, J.; Zhang, X.; Wang, C.; He, J.; Xing, G.; Wang, W.; Zhao, J.; Gai, J. Growth period QTL-allele constitution of global soybeans and its differential evolution changes in geographic adaptation versus maturity group extension. *Plant J.* **2021**, *108*, 1624–1643. [CrossRef]
34. Lin, Y.; Dang, X.; Hu, X.; Li, Y.; Liu, S. Stability and Adaptability of Sugar Beet Examined Based on AMMI and BLUP–GGE Biplot Analyses. *Sugar Tech.* **2025**, *27*, 821–831. [CrossRef]

35. Abdelkhalek, M.S.; Abdelrahman, M.; Mazal, M.T.; Kimiko, I.; Elshenawy, M.; Aamer, S.; Elshenawy, M.; Aamer, S.; Hassan, A.A.; Attia, K.A.; et al. Integrating in vitro breeding, BLUP prediction, and marker analysis to enhance rice yield, quality, and blast resistance. *Front. Plant Sci.* **2025**, *16*, 1588427. [CrossRef] [PubMed]
36. Fu, W.; Li, S.; Dai, X.; Wu, H.; Yin, T. Growth performance and stability analysis of poplar clones based on BLUP and GGE biplot under two different geographic sites in China. *Ind. Crops Prod.* **2025**, *230*, 121039. [CrossRef]
37. Makaju, S.O.; Chhetri, H.B.; Abeyratne, C.R.; Pavicic, M.; Poudel, H.; Irfan, J.A.; Giabardo, A.; Devos, K.M.; Jacobson, D.; Missaoui, A.M. Drought adaptation index (DAI) based on BLUP as a selection approach for drought-resilient switchgrass germplasm. *Front. Genet.* **2025**, *16*, 1626083. [PubMed]
38. Fredriksen, H.B.; Eiselt, K.U.; Good, P. Exploring Global Temperature Oscillations Using a Generalized Linear Box Model. *Geophys. Res. Lett.* **2025**, *52*, e2024GL113994. [CrossRef]
39. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [CrossRef]
40. Pilario, K.E.; Wu, Z. Fast mixed kernel canonical variate analysis for learning-based nonlinear model predictive control. *Chem. Eng. Res. Des.* **2025**, *219*, 19–33. [CrossRef]
41. Arkwazee, H.A.; Wallace, L.T.; Hart, J.P.; Griffiths, P.D.; Myers, J.R. Genome-wide association study (GWAS) of white mold resistance in snap bean. *Genes* **2022**, *13*, 2297. [CrossRef] [PubMed]
42. Pawan, K.; Somveer, N.; Singh, S.R.; Neeraj, B.; Varsha, S.; Chandra, M.D.; Sagar; Choudhary, R.R. Identification of Novel Marker–Trait Associations for Lint Yield Contributing Traits in Upland Cotton (*Gossypium hirsutum* L.) Using SSRs. *Front. Plant Sci.* **2021**, *12*, 653270. [CrossRef] [PubMed]
43. Xiong, C.; Lu, R.; Wolk, D.; Shaw, L.M.; Gleason, C.E.; Johnson, S.C.; Agboola, F.; Schindler, S.E.; Morris, J.C.; Luo, J. A General Linear Mixed Effect Model to Infer Biomarker Correlations by Bridging Retrospectively Measured Data Across Multiple Studies. *Stat. Med.* **2025**, *44*, e70200. [CrossRef]
44. Dai, K.; Wang, X.; Liu, H.; Qiao, P.; Wang, J.; Shi, W.; Guo, J.; Diao, X. Efficient identification of QTL for agronomic traits in foxtail millet (*Setaria italica*) using RTM- and MLM-GWAS. *Theor. Appl. Genet.* **2024**, *137*, 18. [CrossRef]
45. Bermann, M.; Munera, A.A.; Legarra, A.; Aguilar, I.; Misztal, I.; Lourenco, D. Monte Carlo approximation of the logarithm of the determinant of large matrices with applications for linear mixed models in quantitative genetics. *Genet. Sel. Evol.* **2025**, *57*, 44. [CrossRef]
46. Wang, F.; Nan, H.; Chen, L.; Fang, C.; Zhang, H.; Su, T.; Li, S.; Cheng, Q.; Dong, L.; Liu, B.; et al. A new dominant locus, *E11*, controls early flowering time and maturity in soybean. *Mol. Breed.* **2016**, *36*, 13. [CrossRef]
47. Zhang, J.; Song, Q.; Cregan, P.B.; Nelson, R.L.; Wang, X.; Wu, J.; Jiang, G.-L. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genom.* **2015**, *16*, 217. [CrossRef] [PubMed]
48. Fliege, C.; Ward, R.A.; Vogel, P.; Nguyen, H.; Quach, T.; Guo, M.; Viana, J.P.G.; dos Santos, L.B.; Specht, J.E.; Clemente, T.E.; et al. Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *Plant J.* **2022**, *110*, 114–128. [CrossRef]
49. Guo, Z.; Magwire, M.M.; Basten, C.J.; Xu, Z.; Wang, D. Evaluation of the utility of GWAS summary data for fine-mapping and cross-population prediction in maize. *Plant Cell* **2023**, *35*, 2505–2522.