Theoretical and Applied Genetics (2025) 138:84
https://doi.org/10.1007/s00122-025-04875-2

ORIGINAL ARTICLE

A genome-wide association study prioritizes *VRN1-2* as a candidate gene associated with plant height in soybean

Le Wang[1] · Hong Xue[2] · Zhenbin Hu[3] · Yang Li[4] · Tuya Siqin[5] · Hengyou Zhang[1]

Received: 18 December 2024 / Accepted: 6 March 2025

Abstract

Plant height is an important architectural trait that affects crop growth, yield, and stress resistance. Tremendous efforts have been dedicated to revealing the genetic basis or regulatory mechanism; however, the underlying molecular mechanism remains largely unknown primarily due to the lack of controlling genes. In this study, we conducted a single-nucleotide resolution genome-wide association study (GWAS) of plant height using a diverse soybean panel collected worldwide with 6.7 million genome-wide variants (SNPs and Indels). The GWAS of plant height identified three QTLs on chromosomes 10, 18, and 19, of which the one on chromosome 19 precisely co-localized with *Dt1*, known as a major stem growth habit-controlling gene. Other loci without reported genes for plant height were regarded to be new. A close investigation within QTL intervals proposed nine genes that were likely involved in the regulation of plant height according to the expression specificity in developing shoot tip meristems. *VRN1-2* underlies the significant QTL on chromosome 10 was prioritized as the most promising candidate gene. *VRN1-2* shows higher expression in Williams 82 with indeterminate growth habit than Dongnong50 with semi-determinate growth habit across vegetative (V2, V3) and reproductive (R1) growth stages. *VRN1-2* carries non-synonymous variants in the coding region that were significantly associated with plant height variation. The GT allele conferring short plant height was likely subjected to artificial selection during domestication. These results provide a source of new loci and genes for further elaborating the regulatory mechanism of plant height and the key variants would facilitate soybean molecular breeding.

Introduction

Soybean (*Glycine max* L. Merr.) is a major plant-based protein provider since its seed is rich in protein, essential amino acids, and edible oil. It has become the most cultivated oilseed crop that provides nearly half of the oil and protein consumption worldwide (Anderson 2019). It is also a major product to feed animals and poultry because of the high protein concentration (Patil et al. 2017). It is estimated that plant-based protein production needs to be doubled to meet the increasing demands for plant meals by 2050 (Godfray et al. 2010). As a primary source of plant-based protein, it is urgently needed to continuously enhance soybean production due to the high demand for more plant meals and the continuous growth of the world population. The development of soybean cultivars with desirable traits would further increase the production of soybeans,

Breeding semi-dwarf crops is found to be critical for yield improvement, also known as the "Green Revolution," particularly for rice and wheat where a significant increase in planting density because of reduced plant height greatly

Communicated by Istvan Rajcan.

✉ Hengyou Zhang
zhanghengyou@iga.ac.cn

[1] State Key Laboratory of Black Soils Conservation and Utilization, Key Laboratory of Soybean Molecular Design and Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, China

[2] Keshan Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161100, China

[3] Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705, USA

[4] Institute of Operations Research and Information Engineering, Beijing University of Technology, Beijing 100124, China

[5] Heilongjiang Institute of Atomic Energy, Harbin 150086, China

Published online: 27 March 2025

enhanced grain yield. The major genes that are responsible for plant height greatly contribute to the Green Revolution, such as *reduced height* (*Rht-B1b*) in wheat and the *semi-dwarf 1* (*sd1*) in rice (Peng et al. 1999; Sasaki et al. 2002). *sd1* and *Rht-B1b* encode GA20 oxidase and DELLA, respectively, which are key components in the gibberellin biosynthesis pathway, and the mutation alleles reduced gibberellin abundance or response by which reduced plant height. Other participants in the GA pathway regulated plant height were also identified in plant species (Fu et al. 2016; Teng et al. 2013; Zhang et al. 2008; Xue et al. 2022; Ford et al. 2018), strongly suggesting that the GA pathway functioned importantly in regulating plant height. Compared with the significant increase in cereals' yield due to reduced height, it lags largely behind in soybean (Liu et al. 2020a), despite an estimated 14-fold increase in soybean production from 1961 to 2023 (nass.usda.gov). It is challenging to simply apply the aforementioned semi-dwarf controlling genes in soybeans because of the dramatic differences in plant height factors between soybeans and cereals. Whether those mentioned genes are involved in soybean plant height control remains to be elaborated.

Plant height is one of essential plant architecture traits in soybean that is tightly associated with the number of pods per plant, the number of nodes in the main stem, and the plant density, which are all critical to yield (Chang et al. 2018; Li et al. 2020a). Therefore, plant height is a key component of ideal plant architecture contributing to higher yield likely through improved resistance to lodging. Like semi-dwarf cultivars in wheat and rice with a great improvement in yield (Hedden 2003; Peng et al. 1999), semi-dwarf soybean cultivars were also developed to increase yield, such as Hobbit87, Apex, and Charleston with increased yield than the controls (Cooper et al. 1995, 2003). Short plants also greatly contributed to the high-density cultivation of soybean, particularly in the high latitude areas with short growth periods, such as Northeast China or Canada. Therefore, the plant height for soybean is of great importance while its molecular mechanism has yet to be fully exploited for breeding high-yielding cultivars.

Plant height in soybean is a multi-gene controlled trait. Thus far, over 200 quantitative trait loci (QTLs) across 20 chromosomes were identified to be associated with plant height according to SoyBase (https://soybase.org). The majority of studies investigate plant height using bi-parent segregating population (Zhang et al. 2018; Liu et al. 2022; Cao et al. 2019; Lee et al. 2015; Li et al. 2020b; Wang et al. 2022b), some QTLs were specific to a certain population, and the intervals were quite large which is challenging to reveal the underlying genes in a short time. On the other hand, whether the variation as revealed in linkage mapping is representative of the soybean population remains to be further determined. As an alternative to linkage mapping in discovering genes, a genome-wide association study (GWAS) was demonstrated to be a helpful approach in investigating the genetic architecture of quantitative traits as it takes nearly all variations that occur in the evolutionary process within a population into account and thereby the variation as revealed by GWAS is usually representative of the population. With GWAS, some QTLs associated with plant height were identified at the population scale and candidate genes including genes known to control flowering such as *E2* were proposed (Wang et al. 2022b; Yang et al. 2021a; Yu et al. 2023; Zhang et al. 2021), however, these studies use different sets of panels consisting 133–455 diverse individuals and only a few of the QTLs were overlapped (Chang et al. 2018), suggesting that QTLs identified were largely dependent on the population composition, scale, and likely environment (Yang et al. 2021a). However, rare of the QTLs was cloned. Therefore, a large representative population with a saturated resolution of genome-wide variants might be helpful in gene discovery for plant height.

Indeterminate growth 1 (*Dt1*) is identified to be a key gene affecting plant height through and yielding its regulation of determinate growth traits, it is a homolog of *Arabidopsis Terminal Flower* 1 (*TFL1*), a key regulator of flowering time and the development of the inflorescence meristem in *Arabidopsis* (Liu et al. 2020a; Tian et al. 2010; Hanano and Goto 2011). Other genes involved in GA biosynthesis or flowering time were also identified to affect soybean plant height, such as *PH24* (Zhang et al. 2015), *GmTOE4a* (Zhao et al. 2015), *E2* (Watanabe et al. 2011), *GmAP1* (Chen et al. 2020), *GmGAMYB* (Yang et al. 2021b), and *GmIAA27* (Su et al. 2022), suggesting that the GA pathway and/or flowering time is also likely a role in controlling plant height. Despite progress, the majority of the genes have yet to be cloned or discovered, thus the underlying mechanism is still largely unknown. It is needed to continue the investigation of plant height to further understand the regulatory mechanism.

Here, we report a high-resolution GWAS for plant height using a genome-sequenced diverse panel that is more diverse and enhanced mapping resolution than those in previous studies (Yu et al. 2023; Wang et al. 2022a; Yang et al. 2021a; Zhang et al. 2021; Han et al. 2021). This collection of USDA germplasm has successfully identified two major QTLs controlling high protein in soybean (Goettel et al. 2022; Zhang et al. 2020), whereas has not been used in investigating plant height. We precisely detected the previously identified *Dt1* locus and further determined several QTLs that have yet to be revealed for plant height. The high resolution permitted the identification of candidates that were likely involved in the regulation of plant height. The results reported herein provide a comprehensive genetic architecture of plant height in soybean that is likely to be of interest to researchers interested in elaborating the underlying mechanism with the

candidate genes or leveraging them to develop high-yield soybean varieties.

Material and methods

Plant materials

The soybean accessions used in this study came from the USDA soybean germplasm collection (Song et al. 2015). The collection consists of 1008 soybean accessions, including modern soybean varieties, landrace (*G. max*), and wild soybean (*G. soja*). The accessions with both plant height phenotypic values and the availability of genome-sequencing data were obtained from GRIN (USDA-ARS Germplasm Resources Information Network (GRIN) (ars-grin.gov)) for association study. The phenotype data were originally obtained from field conditions that were performed in multiple locations in the USA across several years. Details for the phenotyping of plant height can be found in the USDA-GRIN database. The phenotypic data was collected in two repeats, and the average value of two repeats for each accession was used for GWAS. The data was unadjusted. Additional information for the accessions including locations, country of origin, and accession numbers could be found in the GRIN database (Table S1).

Genome-sequencing data analysis

The soybean accessions with values of plant height were used for searching the availability of genome resequencing data at the NCBI SRA database, and the sequencing data were mainly from previously published literature (Zhou et al. 2015; Valliyodan et al. 2021; Liu et al. 2020b; Fang et al. 2017). Raw sequencing data were aligned to the soybean reference genome of *Glycine max Wm82.a2.v1* (https://phytozome-next.jgi.doe.gov/) with BWA using the default parameters (Li and Durbin 2009). After the removal of duplicated reads with the Picard package (http://broadinstitute.github.io/picard), the Genome Analysis Toolkit (GATK, version 3.70) was used for SNP and indel discovery and genotyping across all the accessions according to GATK Best Practices recommendations (McKenna et al. 2010).

Genome-wide association analysis

GWAS for the plant height was carried out using the TASSEL 5.0 pipeline command line (Bradbury et al. 2007). In the analysis, SNPs with minor allele frequency (MAF) ≥ 0.01 and missing rate < 0.1 in the population were used in the GWAS. A principal component analysis (PCA) of the population was carried out using the TASSEL with a reduced amount of genome-wide SNPs (20 k between adjacent SNPs) using VCFtools (Danecek et al. 2011), and the first five PCs were used as the covariance to control the population structure. The Bonferroni threshold was used with $\alpha = 0.05$ (0.05/6895529 = 7.25e-09), which is equivalent to a -log10P score of 8.15. An SNP that is above the score is considered a significant SNP, and those SNPs that were physically close within 5 Mb were regarded as one QTL (Zhang et al. 2019). The GWAS result was visualized using a Manhattan plot using the R package *qqman,* and the threshold was highlighted.

Candidate gene search

Based on a previously identified linkage disequilibrium (LD) decay distance at an average of 200 kb, any gene models that fall within the 200 kb region on either side of the significant SNPs were collected as the source of candidate genes. The annotation file for the well-annotated genome *Glycine max Wm82.a2.v1* was downloaded from the database Phytozome 12 (https://phytozome-next.jgi.doe.gov/). Sequence variation analysis for the promising candidate genes was also investigated using the panel for GWAS, and only those carrying unambiguous variations in the regions of interest were presented. Raw sequencing reads for 28 tissues of developing soybean plants (Fang et al. 2017) were downloaded from the NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) and re-analyzed using RNA-Seq analysis pipeline comprising read aligner STAR followed by differential expression analysis with EdgeR (Dobin et al. 2013; Robinson et al. 2010). The expression abundance for each gene across the 28 tissues was used to investigate the expression pattern for the candidate genes.

Quantitative real-time PCR

Total RNA was isolated from soybean roots, stems, leaves, and shoot apical meristem using TRIzol reagent (Invitrogen, China) according to the manufacturer's protocol, and qPCR analysis was performed on a LightCycler 96 instrument (Roche, Switzerland) with a real-time PCR kit (Vazyme, China). The housekeeping gene of soybean *GmTUB4* (GenBank accession no. NM_001252709) was used as the internal control (Table S2). The reaction conditions consisted of an initial 5 min pre-incubation at 94 °C, 40 cycles at 94 °C for 30 s, 59 °C for 30 s, and 72 °C for 40 s, followed by a melting-curve analysis from 55 °C to 100 °C with a final cooling for 10 min at 72 °C. The relative transcript abundance of each target gene was calculated using the $2^{-\Delta\Delta CT}$ method. Three biological replicates were tested for each line in each analysis.

Phylogenetic analysis

The homologous protein sequences were retrieved from the NCBI database using BLASTp with the full-length VRN1-2 protein as query. Phylogenetic reconstruction was conducted in MEGA11 through the neighbor-joining method, and the reliability of the tree branches was assessed using the bootstrap method with 1000 replicates.

Results

The soybean collection showed a wide variation in plant height

A total of 1008 accessions with both the values of plant height and genome resequencing data were collected in this study to make the association panel. The panel comprises all of the three subspecies of soybean, including 315 cultivars, 635 landraces, and 58 wild soybeans. A close investigation of the panel revealed that the collection comprised diverse germplasm from a total of 23 countries worldwide, with 63.10% (636) of which from China, followed by 14.09% (142) from the Korean Peninsula, 11.51% (116) from the USA, 4.96% (50) from Japan, and less than 5% in Russia (21) and Canada (10). The remaining accessions were from other 16 countries each with 1–4 accessions. Therefore, the panel is expected to contain a high level of diversity in genome and phenotypes including plant height.

Phenotypic investigation reveals a wide range of variation in plant height in the collection, ranging from 33 to 235 cm, with a mean of 91.8 cm (Fig. 1A, B). Significant differences in plant height among the three subspecies were also observed, which is expected and also revealed in other studies (Lu et al. 2022; Zhou et al. 2015). In the panel, the average plant height for cultivars is 89.2 cm, which is significantly lower ($p = 0.01$) than that of landraces (94.1 cm), consistent with a breeding goal for shorter varieties. It is unexpected that, despite large variation (SD = 26.3 cm), the average value for wild soybean is 79.5 cm. The analyses indicate that the association panel is a diverse collection with a large variation in plant height, which is appropriate for conducting GWAS analysis.

High-resolution GWAS identified significant QTLs

Analysis of the resequencing data reveals a total of 6,895,529 variations (MAF ≥ 0.01 and missing rate < 0.1); this allows the examination of the variation-trait association at the resolution of 1 SNP/140 bp. With such high-resolution variation, the association mapping identified a total of 106 significant SNPs with a threshold of 8.15. The significant SNPs were identified on three chromosomes including 10, 18, and 19 (Fig. 2). The significant SNPs include one SNP on chromosome 10, 14 SNPs on chromosome 18, and 50 SNPs on chromosome 19. The QTLs explain 29.76–30.64% of the phenotypic variance for plant height.

A search of literature or previously identified plant height-associated QTLs in SoyBase revealed co-localized QTLs (Table 1). The major associated QTL detected on chromosome 19 contains a cluster of SNPs in the 89.5-kb interval of chr19: 45,132,143- 45,221,712, with the lead SNP at chr19:45,158,536. The lead SNP is 24.8 kb away from previously identified *Dt1* (*Glyma.19G194300*) which is located at Chr19: 45,183,356-45,185,175 (Tian et al. 2010; Liu et al. 2010). This QTL was also co-localized with previously identified QTLs, plant height 13–8, plant height 4–2, and plant height 4–4 (Lee et al. 1996). The significant region on chromosome 18 was found to overlap with plant height-associated QTLs, such as *qPH-007* (Yao et al. 2015; Kabelka et al. 2004; Sun et al. 2006), while the candidate genes remain unknown. No known plant height-associated QTLs were detected overlapping with the significant SNPs

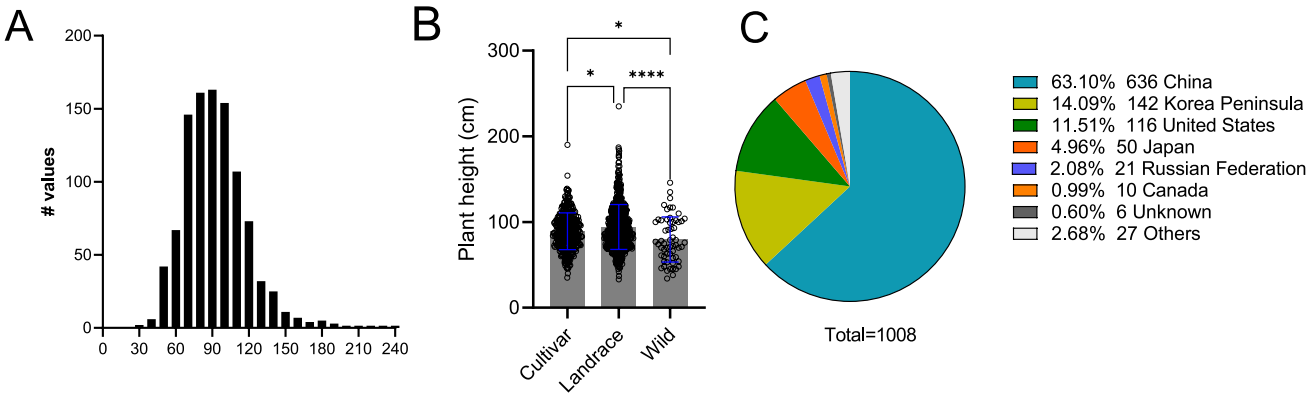

Fig. 1 Statistics of the plant height in the population. **A** Distribution of plant height in the population. **B** Comparison of the plant height among cultivar, landrace, and wild soybeans. **C** Percentage of the origin of the accessions in the panel

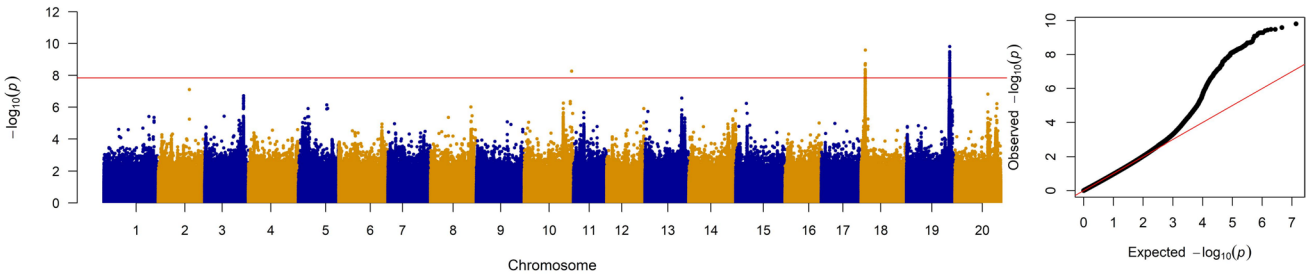

Fig. 2 Manhattan plots of the GWAS of plant height in soybean

Table 1 Information for the significant SNPs identified in GWAS

Significant SNPs	Chr	Position	*P*. value	MAF (%)	R^2 (%)	Adjusted *P*-values	Co-localized plant height QTLs
Chr10_49993502	10	49,993,502	3.11E-12	19	30.64	1.16E-06	
Chr18_4351572	18	4,351,572	2.56E-10	13	29.76	0.000112	*mqPlant height-007*
Chr19_45158536	19	45,158,536	1.56E-10	32	29.86	2.56E-05	*Plant height 13–8, Plant height 4–2, Plant height 4–4*

on chromosome 10 (lead SNP at 50.0 Mb), which were regarded as a new QTL.

Candidate genes associated with plant height

The prominent clustered SNPs at the chr19 QTL span an 86-kb genomic region (45,132,143–45,221,712), and it co-harbors the *Dt1* gene (Chr19:45,183,356–45,185,175) with a demonstrated role in plant growth habit that is tightly associated with plant height (Liu et al. 2010; Tian et al. 2010), demonstrating the robustness of the GWAS study in determining plant height associations and candidate genes. This result urges us to predict candidate genes associated with plant height in other QTLs. In total, within the flanking genomic region centered with each of the significant SNPs, a total of 135 predicted gene models were identified within the region according to the annotation of the *Wm82.a2.v1* reference genome. *Dt1*, a major QTL gene controlling soybean stem growth habit (Tian et al. 2010; Liu et al. 2010), is among the 51 gene models within the QTL region on chromosome 19. Thus it is regarded to be the best candidate gene at this locus. Other than *Dt1*, we also observed several genes that likely play roles associated with plant height, thereby we proposed in this study (Table 2).

We next examined the candidacy of the gene models associated with plant height primarily based on literature, annotations, and expression patterns in 28 developing tissues in soybean. Among the genes (Table S2), three genes (*Glyma.10G272300*, *Glyma.10G273000*, *Glyma.10G276100*) with the Arabidopsis orthologs involved in gibberellic acid (GA) biosynthesis and signaling were nominated the promising candidate for the identified QTL on chr10 since the demonstrated role of GA pathway in the regulation of plant height (Wang and Wang 2022). For example, *Glyma.10G273000* is an ortholog of *AtMYB62*, and overexpression of *AtMYB62* resulted in GA deficiency symptoms such as delayed bolting and senescence, and smaller plants, which could be partially reversed by exogenous GA application (Devaiah et al. 2009); *Glyma.10G276100* is an ortholog of PAN that is involved in GI involved flowering pathway and connected to the shoot apical meristem (SAM) regulatory network (Maier et al. 2011). In addition, three tandem AP2/B3-like transcriptional factor family proteins (*Glyma.10G281100* (*VRN1-1*), *Glyma.10G281200* (*VRN1-2*), *Glyma.10G281300* (*VRN1-3*), spanning chr10: 50,240,905–50,252,837), also known as *REDUCED VERNALIZATION RESPONSE 1* (*VRN1*) that are involved in flowering, growth habit, spike determinacy in several plant species (Konopatskaia et al. 2016; Li et al. 2019; Lu et al. 2015). These genes were located within the QTL region on chromosome 10 (chr10: 49,566,600- 50,700,770). Other genes with annotations were associated with shoot development, auxin responsiveness, and associated light signaling such as *Glyma.05G064600* (*VAP27-4*) and *Glyma.10G286700* (*CIP8*) (DeCook et al. 2006; Nawkar et al. 2017).

Expression patterns of selected candidate genes

In addition to the aforementioned genes with functions involved in plant height, we also examined whether many of the genes expressing in shoot meristem, which play important roles in plant height determination. To do this, we analyzed genes' expression in 28 developing tissues including

Table 2 Selected candidate genes and the annotation

Lead SNPs	Chr	SNP position	Candidate gene	Gene position	Ara ortholog	TAIR10_symbol	TAIR10_name
Chr10_49993502	10	49,993,502	Glyma.10G272300	Chr10:49,452,028-49453400	AT1G68360.1		C2H2 and C2HC zinc fingers superfamily protein
			Glyma.10G273000	Chr10:49,546,836-49,548,462	AT1G68320.1	AtMYB62	myb domain protein 62
			Glyma.10G276100	Chr10:49,857,795-49,863,059	AT1G68640.1	PAN	bZIP transcription factor family protein
			Glyma.10G278200	Chr10:50,022,810-50029454	AT5G05010.1	GIS3 (GLABROUS INFLORESCENCE STEMS 3)	clathrin adaptor complexes medium subunit family protein
			Glyma.10G281100	Chr10:50,240,905-50243956	AT3G18990.1	REM39,VRN1	AP2/B3-like transcriptional factor family protein
			Glyma.10G281200	Chr10:50,246,998-50249746	AT3G18990.1	REM39,VRN1	AP2/B3-like transcriptional factor family protein
			Glyma.10G281300	Chr10:50,251,156-50252837	AT3G18990.1	REM39,VRN1	AP2/B3-like transcriptional factor family protein
			Glyma.10G286700	Chr10:50,629,654-50631072	AT5G64920.1	CIP8	COP1-interacting protein 8
Chr18_4351572	18	4,351,572	Unknown				
Chr19_45158536	19	45,158,536	Glyma.19G194300	Chr19:45,183,356-45,185,175	AT5G03840.1	TFL-1,TFL1	PEBP (phosphatidylethanolamine-binding protein) family protein

roots, cotyledon, shoot meristem, and different stages of leaves, leaf buds, flowers, seeds, pod seeds, and pods (Fang et al. 2017). We revealed that 40 of 135 genes (11.6%) were expressed in shoot meristem at different abundance levels (Fig. 3, Table S2). Genes with expression in specific tissues deserve more attention. For example, *Glyma.10G278200* shows relatively stronger expression in early flower and shoot meristem compared with other tested tissues, suggesting the possible specific roles associated with plant height.

We were particularly interested in the three *VRN1* homologs (*VRN1-1*, *VRN1-2*, *VRN1-3*,) on chromosome 10 and evaluated the relative expressions in different tissues (roots, stems, leaves, and shoot apical meristem (SAM)) across multiple stages (V1, V2, V3, R1) using qRT-PCR (Fig. 4). The result indicates that, of the three *VRN1*-like genes, VRN1-2 was dramatically higher than the other two in the testing tissues (roots, stem, leaves, and SAM) in both vegetable growth stage 1 (V1) and reproductivity growth stage 1 (R1). In contrast, *VRN1-2* and *VRN1-3* were barely expressed in the testing tissues in both V1 and R1 (Fig. 4A). We next examined the expression of *VRN1-2* in stem and SAM tissues in two representative soybean cultivars Willimas 82 (W82, indeterminate growth habit) and Dongnong50 (DN50, semi-determinate growth habit) showing different determination growth habits and plant heights in Harbin, China, where Willimas 82 was 97.85 ± 7.24 cm, in average, which is 36.32% higher than DN50 (71.78 ± 2.44 cm) (Fig. 4B,), The comparison showed that the expression of *VRN1-2* was significantly higher in both stem and SAM tissues in W82 than DN50, in both V2, V3 and R1 stages. We further investigated the amino acids for VRN1-2, and it contains three B3 domains (Fig. S1). In addition to the expression patterns, we also construct a phylogenetic tree using full-length protein sequences of VRN1-2 and its homologues from diverse species (Fig. S2), it appears that the three soybean VRN proteins (VRN1-1, VRN1-2, VRN1-3) were spited into two clusters, while VRN1-2 is phylogenetically closer to VRN-like proteins from *Vigna unguiculata, Phaseolus acutifolius, and P. vulgaris* than the other two VRNs, suggesting a functional difference between VRN1-2 from VRN1-2 and VRN1-3. These results suggest that, among the three *VRN1* homologs in soybean, *VRN1-2* might play a role associated with plant growth and development, and it might contribute to the difference in plant height between W82 and DN50.

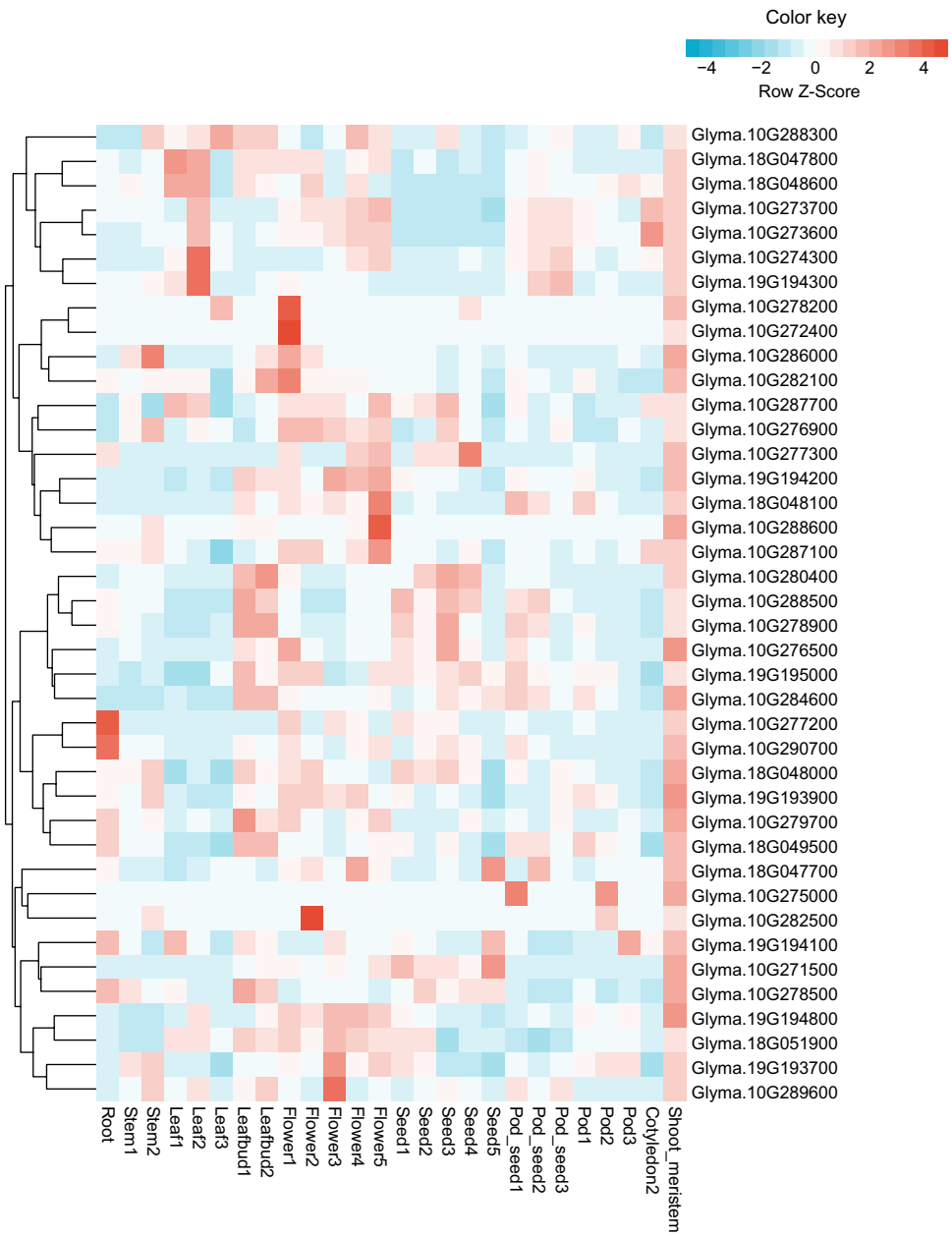

Fig. 3 A heatmap illustrating the relative expression of candidate genes

VRN1-2 haplotype association with plant height variation

To evaluate if *VRN1-2* contains variation that is associated with plant height, we analyzed the sequence variation in *VRN1-2* in 806 germplasms, a subset of the association panel. The sequence analysis reveals two major variants (G/A at 50,248,210, T/C at 50,248,523) in the CDS region (exon1 and exon2) (Fig. 5A). The two variants were non-synonymous which results in amino acid changes from R to K for G/A variation and F to S for T/C variation. To evaluate the genetic effects of the haplotypes, we also include the key variation of C/T (K/R) in *Dt1* (Tian et al. 2010). The combination analysis defined four haplotypes (Hap1-Hap4). The phenotypic comparison showed that soybean lines carrying AC (Hap1) were significantly higher by 3.72cm, on average, than those (Hap2) carrying GT in *Dt1*-C background (Fig. 5B); in contrast, there is no significant difference in plant height for those lines (Hap3 and Hap4) in *Dt1*-T background. Further, we calculated the allelic frequency in cultivated soybeans and wild soybeans. The results show a dramatic difference in the allelic frequency between wild and cultivated soybeans (Fig. 5C). For example, the wild soybean population contains approximately 22% GT of *VRN1-2*, whereas the frequency is much higher in cultivated soybeans (71%), suggesting that GT was likely subjected to artificial selection during domestication, which is rational since the GT-C haplotype (Hap2) is statistically, on average,

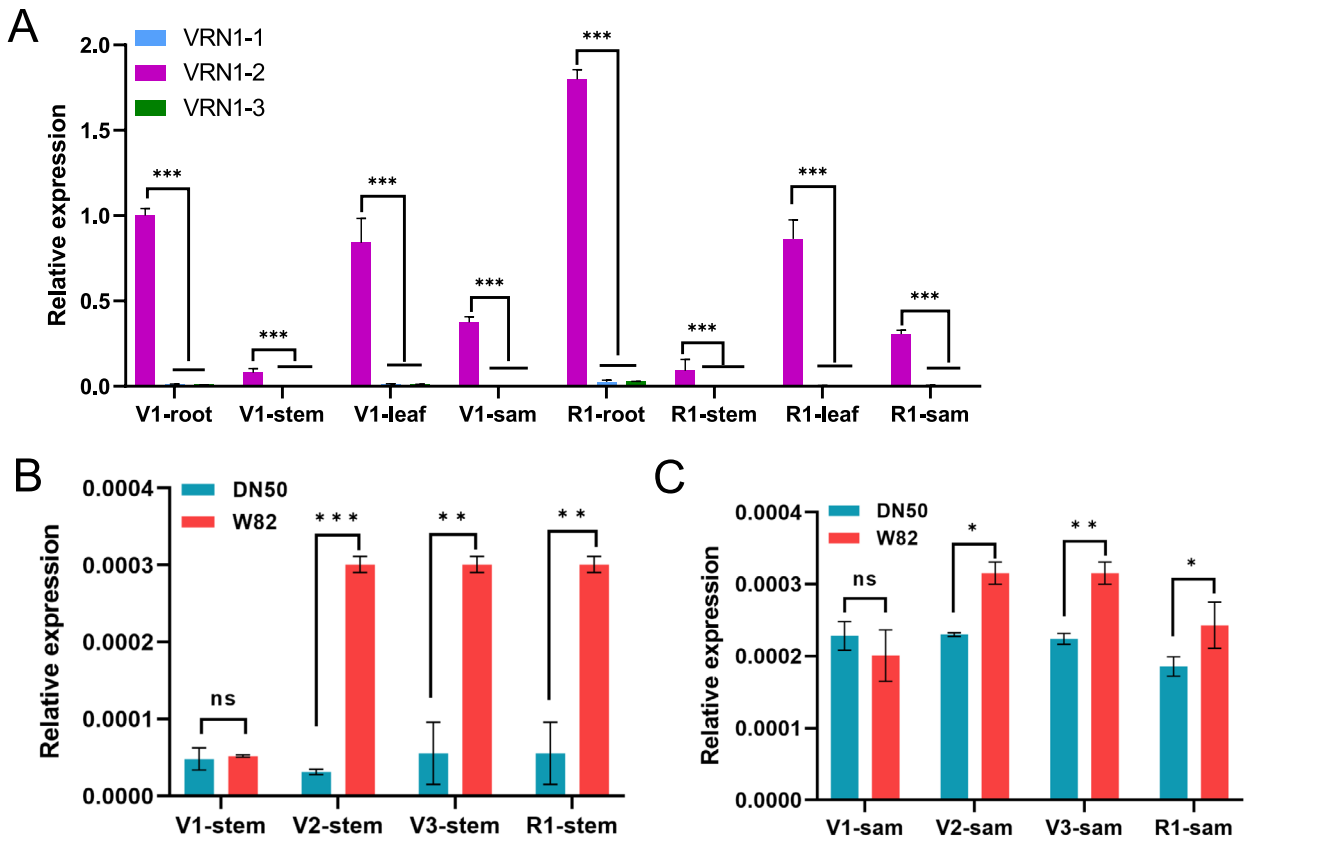

Fig. 4 Expression patterns and comparison of three *VRN1* genes in different tissues and genotypes. **A** Expression levels of three *VRN1* genes (*VRN1-1*, *VRN1-2*, and *VRN1-3*) in soybean roots, stems, leaves, and SAM tissues of W82. **B** The expression of *VRN1-2* in stems between DN50 and W82 during V1, V2, V3, and R1 stages. **C** The expression of *VRN1-2* in SAM tissues between DN50 and W82 during V1, V2, V3, and R1 stages

lower than other three haplotypes (Hap1, Hap3, and Hap4), which is rational because preferential selection for GT might contribute to reduced plant height, which is one of the goals in high-yield breeding.

Discussion

Plant height has been an important trait that has been investigated in various crop species for decades. It has been demonstrated that a reduction in plant height may allow logging resistance and high-density planting to achieve production increase. As such, short plant height without penalty of seed production per plant would be an ideal architecture trait in soybean. However, the molecular mechanism of plant height in soybean is largely unknown, mainly due to the lack of information on height-controlling genes. There are a number of QTLs associated with plant height identified in soybean, while uncovering the underlying genes remains challenging because of large intervals of many QTLs or low-resolution genome-wide variation (Han et al. 2021; Lee et al. 2015; Li et al. 2020b; Wang et al. 2022a; Yu et al. 2023; Zhang et al. 2021). The purpose of this study was to identify SNPs and candidate genes that are associated with plant height using a larger, more diverse soybean population with higher resolution of genome-wide variations. We are specifically interested in whether we might discover QTLs that have yet to be identified previously using either linkage mapping or GWAS. The variation used in this study includes both SNPs and Indels, and the amount of the variants permits the survey of variation-trait associations at single-nucleotide resolution. By using the high-resolution GWAS, we were able to identify associations for the plant height, one of which precisely co-localized with previously identified *Dt1* with demonstrated role regulating growth habit trait (Liu et al. 2010; Tian et al. 2010), whereas others appeared to be novel. *Dt1* (Chr19:45,183,356-45,185,175) resided within the QTL region (Chr19: 45,132,143-45,221,712) and is 24.8 kb away from the lead SNP (Chr19_45158536). This result demonstrates the robustness of our GWAS analysis in uncovering QTLs, and it urges us to predict genes for other QTLs that have yet to be cloned.

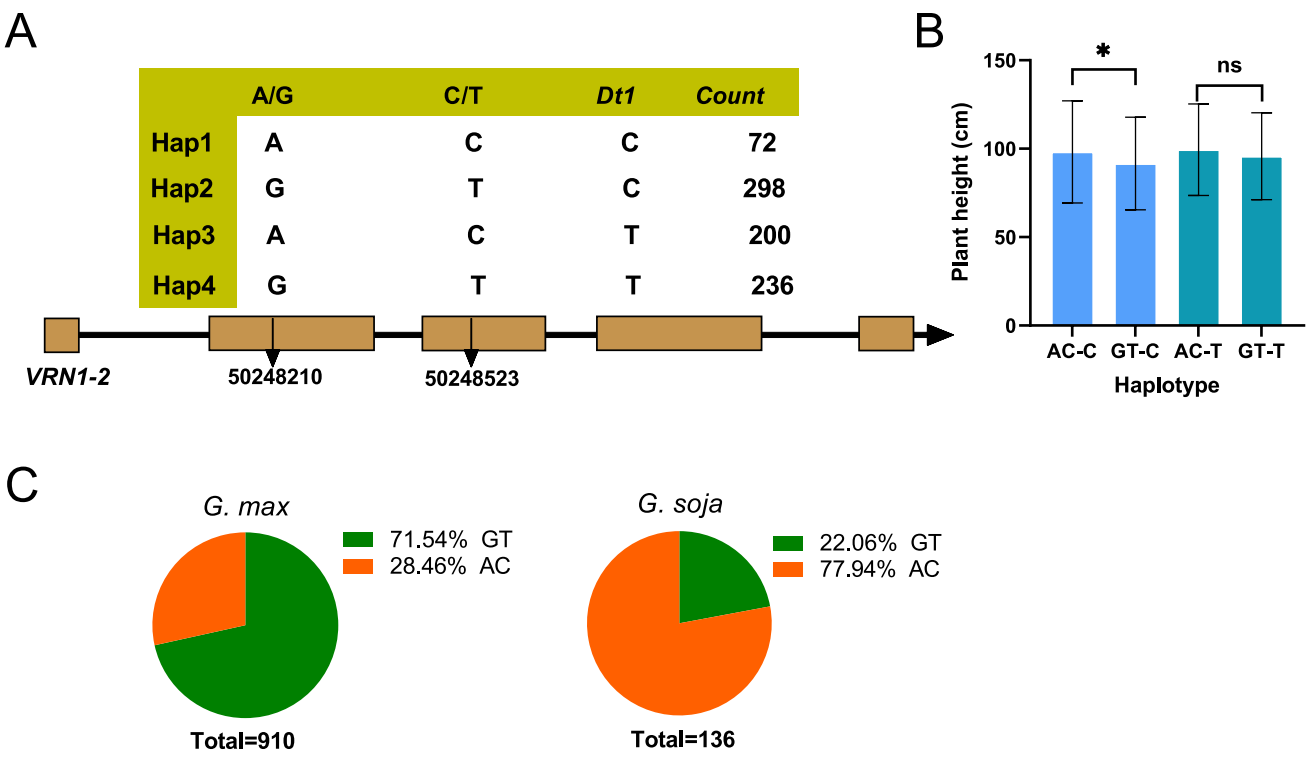

Fig. 5 Genetic diversity analysis of VRN1-2 in soybean population. **A** Haplotypic analysis of *VRN1-2* in the panel. **B** Phenotypic comparison between haplotypes. **C** Allelic frequency of *VRN1-2* in two subspecies *G. max* and *G. soja*

VRN1 was identified for its role in Arabidopsis vernalization, which is also an essential physiological process for many crops to flower (Lu et al. 2015; Konopatskaia et al. 2016; Deng et al. 2015; Levy et al. 2002), for example, overexpression of *VRN1* resulted in early flowering (Levy et al. 2002; Lu et al. 2015). It is also demonstrated that *VRN1* has much broader roles. For example, it was identified to regulate spikelet development and spike determinacy in wheat, suggesting that it is likely involved in the regulation of plant height in other crop species which needs further experimental determination. It might also be involved in cold tolerance and hormone metabolism based on the detection of its targeting sequences (Deng et al. 2015). In soybean, a *VRN1* homolog on chromosome 11 was recently identified to be involved in the regulation of flowering when expressed in *Arabidopsis thaliana* (Lu et al. 2015), suggesting that it might be a conserved role for *VRN1* in the regulation of flowering between dicot and monocot plant species (Levy et al. 2002; Konopatskaia et al. 2016). Further, *VRN1* in wheat has been demonstrated to regulate spring growth habit (Konopatskaia et al. 2016). Here, we identified tandem *VRN1* genes in soybean genome and the locus resides on the QTL associated with plant height on chromosome 10. The expression and sequence phenotype correlation analysis demonstrated that *VRN1-2* was likely associated with soybean plant height. It is likely that *VRN1-2* functioning in the regulation of plant height is tightly associated with Dt1-C since the observation of plant height difference between Hap1 and Hap2, but not between Hap3 and Hap4. The Hap2 is likely subjected to artificial selection mainly because of the dramatic increase in the frequency for GT of *VRN1-2* in *G. max* than *G. soja*. Continuous use of this locus might be useful for soybean breeding for short soybean plants that allow compact planting in high latitudes.

A potential concern of our analysis is that the phenotypic values were collected from the USDA Soybean germplasm collection, and the values were obtained in multiple fields across different years in past decades. Recent studies demonstrated that there is a minor difference in identifying major QTLs associated with complex traits using GRIN-derived unadjusted and adjusted phenotypic data (Bandillo et al. 2015; Goettel et al. 2022; Zhang et al. 2020). Precise identification of *Dt1* in this study is supportive that the plant height values in this panel were appropriate for uncovering associated SNPs and revealing candidate genes, which would be a useful resource, in addition to previous findings, for further investigation and continued genetic improvement.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-025-04875-2.

Acknowledgements This study was supported by National Key R&D Program of China (2024YFE0112800), the Natural Science

Foundation of Heilongjiang Province of China (JQ2022C005), Hainan Seed Industry Laboratory and China National Seed Group (project of B23YQ1503), the National Natural Science Foundation of China (32272176), the Innovation Team Project of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences (2022CXTD03), and Natural Science Foundation of Heilongjiang Academy of Sciences (YZ2022YZN02).

Author contribution statement Hengyou Zhang conceived and designed the study; Le Wang and Hong Xue performed the experiments and analyzed the data; Zhenbin Hu, Yang Li and Tuya Siqin assisted the data preparation and analysis; Le Wang and Hengyou Zhang wrote the manuscript. Hengyou Zhang reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding This study was supported by National Key R&D Program of China (2024YFE0112800), the Natural Science Foundation of Heilongjiang Province of China (JQ2022C005), Hainan Seed Industry Laboratory and China National Seed Group (pro ject of B23YQ1503), the National Natural Science Foundation of China (32272176), the Innovation Team Project of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences (2022CXTD03), and Natural Science Foundation of Heilongjiang Academy of Sciences (YZ2022YZN02).

Declarations

Conflict of interest The authors declare no competing interests.

References

Anderson E (2019) Soybean [*Glycine max* (L.) Merr.] breeding: history, improvement, production and future opportunities. In: Al-Khayri JM, Jain SM, Johnson DV (eds) Advances in Plant Breeding Strategies: Legumes, vol 7, Springer

Bandillo N, Jarquin D, Song Q, Nelson R, Cregan P, Specht J, Lorenz A (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. The Plant Genome. https://doi.org/10.3835/plantgenome2015.04.0024

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635

Cao YC, Li SG, Chen GL et al (2019) Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants-Basel 8(10):373

Chang F, Guo C, Sun F et al (2018) Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. Front Plant Sci 9:1184

Chen L, Nan H, Kong L et al (2020) Soybean AP1 homologs control flowering time and plant height. J Integr Plant Biol 62(12):1868–1879

Cooper RL, Martin RJ, Martin SK (1995) Registration of 'Charleston' soybean. Crop Sci 35(2):593

Cooper RL, Mendiola T, Martin SK (2003) Registration of 'apex' soybean. Crop Sci 43(4):1563–1564

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158

DeCook R, Lall S, Nettleton D, Howell SH (2006) Genetic regulation of gene expression during shoot development in Arabidopsis. Genetics 172(2):1155–1164

Deng WW, Casao MC, Wang PH, Sato K, Hayes PM, Finnegan EJ, Trevaskis B (2015) Direct links between the vernalization response and other key traits of cereal crops. Nat Commun 6:5882

Devaiah BN, Madhuvanthi R, Karthikeyan AS, Raghothama KG (2009) Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in Arabidopsis. Mol Plant 2(1):43–58

Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29(1):15–21

Fang C, Ma Y, Wu S et al (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol 18(1):161

Ford BA, Foo E, Sharwood R et al (2018) Rht18 semidwarfism in wheat is due to increased GA 2-oxidaseA9 expression and reduced GA content. Plant Physiol 177(1):168–180

Fu J, Ren F, Lu X et al (2016) A tandem array of ent-kaurene synthases in maize with roles in gibberellin and more specialized metabolism. Plant Physiol 170(2):742–751

Godfray HCJ, Beddington JR, Crute IR et al (2010) Food security: the challenge of feeding 9 billion people. Science 327(5967):812–818

Goettel W, Zhang HY, Li Y et al (2022) POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. Nat Commun 13(1):3051

Han X, Xu ZR, Zhou L, Han CY, Zhang YM (2021) Identification of QTNs and their candidate genes for flowering time and plant height in soybean using multi-locus genome-wide association studies. Mol Breed 41(6):39

Hanano S, Goto K (2011) Arabidopsis TERMINAL FLOWER1 Is involved in the regulation of flowering time and inflorescence development through transcriptional repression. Plant Cell 23(9):3172–3184

Hedden P (2003) The genes of the green revolution. Trends Genet 19(1):5–9

Kabelka EA, Diers BW, Fehr WR et al (2004) Putative alleles for increased yield from soybean plant introductions. Crop Sci 44(3):784–791

Konopatskaia I, Vavilova V, Kondratenko EY, Blinov A, Goncharov NP (2016) VRN1 genes variability in tetraploid wheat species with a spring growth habit. BMC Plant Biol 16(Suppl 3):244

Lee SH, Bailey MA, Mian MA et al (1996) Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor Appl Genet 92(5):516–523

Lee S, Jun TH, Michel AP, Mian MAR (2015) SNP markers linked to QTL conditioning plant height, lodging, and maturity in soybean. Euphytica 203(3):521–532

Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C (2002) Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. Science 297(5579):243–246

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760

Li CX, Lin HQ, Chen A, Lau M, Jernstedt J, Dubcovsky J (2019) Wheat VRN1, FUL2 and FUL3 play critical and redundant roles in spikelet development and spike determinacy. Development 146(14):175398

Li MW, Wang Z, Jiang B et al (2020a) Impacts of genomic research on soybean improvement in East Asia. Theor Appl Genet 133(5):1655–1678

Li RC, Jiang HW, Zhang ZG et al (2020b) Combined linkage mapping and bsa to identify QTL and candidate genes for plant height and the number of nodes on the main stem in soybean. Int J Mol Sci 21(1):42

Liu BH, Watanabe S, Uchiyama T et al (2010) The soybean stem growth habit gene *Dt1* is an ortholog of Arabidopsis *TERMINAL FLOWER1*. Plant Physiol 153(1):198–210

Liu SL, Zhang M, Feng F, Tian ZX (2020a) Toward a "green revolution" for soybean. Mol Plant 13(5):688–697
Liu Y, Du H, Li P et al (2020b) Pan-genome of wild and cultivated soybeans. Cell 182(1):162–176
Liu C, Tian Y, Liu ZX et al (2022) Identification and characterization of long-InDels through whole genome resequencing to facilitate fine-mapping of a QTL for plant height in soybean (Glycine max L. Merr.). J Integr Agric 21(7):1903–1912
Lu J, Suo H, Yi R, Ma Q, Nian H (2015) Glyma11g13220, a homolog of the vernalization pathway gene VERNALIZATION 1 from soybean [Glycine max (L.) Merr.], promotes flowering in Arabidopsis thaliana. BMC Plant Biol 15:232
Lu Y, Zhang J, Guo X, Chen J, Chang R, Guan R, Qiu L (2022) Identification of genomic regions associated with vine growth and plant height of soybean. Int J Mol Sci 23(10):5823
Maier AT, Stehling-Sun S, Offenburger SL, Lohmann JU (2011) The bZIP transcription factor PERIANTHIA: a multifunctional hub for meristem control. Front Plant Sci 2:79
McKenna A, Hanna M, Banks E et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20(9):1297–1303
Nawkar GM, Kang CH, Maibam P et al (2017) HY5, a positive regulator of light signaling, negatively controls the unfolded protein response in Arabidopsis. Proc Natl Acad Sci USA 114(8):2084–2089
Patil G, Mian R, Vuong T et al (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theor Appl Genet 130(10):1975–1991
Peng J, Richards DE, Hartley NM et al (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400(6741):256–261
Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140
Sasaki A, Ashikari M, Ueguchi-Tanaka M et al (2002) Green revolution: a mutant gibberellin-synthesis gene in rice - new insight into the rice variant that helped to avert famine over thirty years ago. Nature 416(6882):701–702
Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2015) Fingerprinting soybean germplasm and its utility in genomic research. G3 (Bethesda) 5(10):1999–2006
Su B, Wu H, Guo Y, Gao H, Wei Z, Zhao Y, Qiu L (2022) GmIAA27 encodes an AUX/IAA protein involved in dwarfing and multibranching in soybean. Int J Mol Sci 23(15):8643
Sun DH, Li WB, Zhang ZC, Chen QS, Ning HL, Qiu LJ, Sun GL (2006) Quantitative trait loci analysis for the developmental behavior of Soybean (Glycine max L. Merr.). Theor Appl Genet 112(4):665–673
Teng F, Zhai L, Liu R et al (2013) ZmGA3ox2, a candidate gene for a major QTL, qPH31, for plant height in maize. Plant J 73(3):405–416
Tian Z, Wang X, Lee R et al (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci U S A 107(19):8563–8568
Valliyodan B, Brown AV, Wang JX et al (2021) Genetic variation among 481 diverse soybean accessions, inferred from genomic re-sequencing. Sci Data 8(1):50
Wang SS, Wang YJ (2022) Harnessing hormone gibberellin knowledge for plant height regulation. Plant Cell Rep 41(10):1945–1953
Wang JJ, Hu B, Huang SS et al (2022a) SNP-bin linkage analysis and genome-wide association study of plant height in soybean. Crop Pasture Sci 73(3):222–237
Wang JJ, Hu B, Jing YL et al (2022b) Detecting QTL and candidate genes for plant height in soybean via linkage analysis and GWAS. Front Plant Sci 12:803820
Watanabe S, Xia ZJ, Hideshima R et al (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the gene is involved in soybean maturity and flowering. Genetics 188(2):395-U260
Xue Y, Zhang Y, Shan J et al (2022) Growth repressor GmRAV binds to the GmGA3ox promoter to negatively regulate plant height development in soybean. Int J Mol Sci 23(3):1721
Yang Q, Lin GM, Lv HY, Wang CH, Yang YQ, Liao H (2021a) Environmental and genetic regulation of plant height in soybean. BMC Plant Biol 21(1):63
Yang X, Li X, Shan J et al (2021b) Overexpression of *GmGAMYB* accelerates the transition to flowering and increases plant height in soybean. Front Plant Sci 12:667242
Yao D, Liu ZZ, Zhang J et al (2015) Analysis of quantitative trait loci for main plant traits in soybean. Genet Mol Res 14(2):6101–6109
Yu H, Bhat JA, Li CD, Zhao BF, Guo T, Feng XZ (2023) Genome-wide survey identified superior and rare haplotypes for plant height in the north-eastern soybean germplasm of China. Mol Breed 43(4):22
Zhang Y, Zhu Y, Peng Y et al (2008) Gibberellin homeostasis and plant height control by EUI and a role for gibberellin in root gravity responses in rice. Cell Res 18(3):412–421
Zhang J, Song Q, Cregan PB, Nelson RL, Wang X, Wu J, Jiang GL (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (Glycine max) germplasm. BMC Genom 16(1):217
Zhang XL, Wang WB, Guo N, Zhang YY, Bu YP, Zhao JM, Xing H (2018) Combining QTL-seq and linkage mapping to fine map a wild soybean allele characteristic of greater plant height. BMC Genom 19(1):226
Zhang D, Zhang HY, Hu ZB et al (2019) Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. Plos Genet 15(7):e1008267
Zhang HY, Goettel W, Song QJ, Jiang H, Hu ZB, Wang ML, An YQC (2020) Dual use and selection of GmSWEET39 for oil and protein improvement in soybean. Plos Genet 16(11):e1009114
Zhang XL, Ding WT, Xue D et al (2021) Genome-wide association studies of plant architecture-related traits and 100-seed weight in soybean landraces. BMC Genom Data 22(1):10
Zhao X, Cao D, Huang Z et al (2015) Dual functions of GmTOE4a in the regulation of photoperiod-mediated flowering and plant morphology in soybean. Plant Mol Biol 88(4–5):343–355
Zhou Z, Jiang Y, Wang Z et al (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol 33(4):408–414