Natural variation of the $Dt2$ promoter controls plant height and node number in semi-determinant soybean

Kun Kou · Tong Su · Yanping Wang · Hui Yang · Hao Du · Milan He · Tai Li · Lixin Ma · Chunmei Liao · Cen Yang · Wenqian Shi · Linnan Chen · Yongli Li · Bize Yang · Lingping Kong · Shichen Li · Lingshuang Wang · Xiaohui Zhao · Sijia Lu · Baohui Liu · Fanjiang Kong · Chao Fang

Received: 25 February 2021 / Accepted: 23 May 2021 / Published online: 13 June 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Soybean (Glycine max (L.) Merrill) is an important legume crop worldwide. Plant height (PH) is a quantitative trait that is closely related to node number (NN) and internode length (IL) on the main stem, which together affect soybean yield. To identify candidate genes controlling these three traits in soybean, we examined a recombinant inbred line (RIL) population derived from a cross between two soybean varieties with semi-determinate stems ($Dt1Dt1Dt2Dt2$), JKK378 and HXW. A quantitative trait locus (QTL) named $qPH18$ was identified that simultaneously controls PH, NN, and IL; this region harbors the semi-determinant gene $Dt2$. Sequencing of the $Dt2$ promoter from JKK378 identified three polymorphisms relative to HXW, including two single nucleotide polymorphism (SNPs) and an 18-bp insertion/deletion polymorphism (Indel). $Dt2$ expression was lower in the $qPH18^{JKK378}$ group than in the $qPH18^{HXW}$ group, whereas the expression level of the downstream gene $Dt1$ showed the opposite tendency. A transient transfection assay confirmed that $Dt2$ promoter activity is lower in JKK378 compared to HXW. We propose that the polymorphisms in the dominant $Dt2$ promoter underlie the differences in $Dt2$ expression and its downstream gene $Dt1$ in the two parents, thereby affecting PH, NN, IL, and grain weight per plant without altering stem growth habit. Compared to the $PH18^{HXW}$ allele, the $qPH18^{JKK378}$ allele suppresses $Dt2$ expression, which releases the inhibition of $Dt1$ expression, thus enhancing NN and grain yield. Our findings shed light on the mechanism underlying NN and PH in soybean and provide a molecular marker to facilitate breeding.

Keywords Soybean · QTL · Genotyping-by-sequencing (GBS) · Plant height · Node number · $Dt2$

Y. Wang
Mudanjiang Branch of Heilongjiang Academy of Agricultural Sciences, Mudanjiang, China

H. Yang · H. Du · M. He · T. Li · L. Ma · C. Liao · C. Yang · W. Shi · L. Chen · Y. Li · B. Yang · L. Kong · X. Zhao · S. Lu · B. Liu · F. Kong · C. Fang
Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou, China

e-mail: fanchao@gzhu.edu.cn
Introduction

Soybean (Glycine max (L.) Merrill) is an economically important crop that represents an excellent source of protein and edible oil. Generating soybean lines with high yield and quality has long since been an important goal of breeders. Soybean yield is a complex quantitative trait that is affected by various traits, including plant height (PH), node number (NN), and internode length (IL). A significant positive correlation has been detected between PH and NN and between PH and IL (Liu et al. 2011; Allen et al. 2018; Chang et al. 2018; Assefa et al. 2019). Taller soybean plants have greater NN and IL, thereby producing more pods. IL and NN affect the strength of the main stem and hence the yield per plant.

PH and NN on the main stem are mainly determined by the stem growth habit of the plant, which is controlled by two major loci: Dt1 and Dt2 (Bernard 1971; Heatherly and Smith 2004). In the Dt1 genetic background, Dt2 and dt2 genotypes produce semi-determinate and indeterminate phenotypes, respectively. In the dt1 genetic background, the phenotype is determinate, as the dt1 allele has an epistatic effect on the Dt2 locus (Liu et al. 2010; Ping et al. 2014). Studies of near isogenic lines of Dt1 showed that Dt1 function is regulated by photoperiod, as the effect of Dt1 on regulating stem growth habit is not obvious under short-day conditions, but differences in NN, PH, and pod number per plant are observed under long-day conditions (Bernard 1972; Liu et al. 2010). The Dt2 locus harbors FUL3b, encoding a MADS box transcription factor; several variations in the first intron of this gene are believed to be responsible for the differences in plants with the Dt2 vs. dt2 genotype (Ping et al. 2014). Dt2 is expressed at its highest levels in the stem apices of plants at the V2 stage: at this stage, the first trifoliate leaflets at node 2 are fully expanded, while the second trifoliate leaflets at node 3 have not yet unfolded (Ping et al. 2014). Furthermore, the Dt2 directly binds to the Dt1 promoter to inhibit its transcription, leading to the development of a semi-determinate habit (Liu et al. 2016). Phytochromes E3 and E4 induce Dt1 transcription indirectly, thus promoting NN and pod number per plant (Xu et al. 2013). FT5a might control PH and NN by inducing AP1 transcription, thereby inhibiting Dt1 expression, thus controlling yield per plant (Takeshima et al. 2019). AP1 inhibits the expression of Dt1 to control NN, PH, and pod number per plant (Chen et al. 2020).

Identifying the QTLs/genes determining NN in soybean is of great importance for breeding plants with increased NN and yield with the appropriate PH for lodging tolerance, like the semi-dwarf trait in rice (Oryza sativa) and wheat (Triticum aestivum) (Peng et al. 1999; Ashikari et al. 2002; Monna et al. 2002; Hedden 2003). In the past decades, most soybean cultivars grown at high latitudes have been semi-determinate. Such cultivars may produce more pods per plant than indeterminate cultivars (Setiyono et al. 2007). Semi-determinate soybean plants maintain stem elongation following the photoperiod-induced initiation of the floral transition, which allows for more overlap between vegetative growth and reproductive development. This trait helps these plants better adapt to a shorter growing season.

In the current study, we used a recombinant inbred line (RIL) population derived from a cross between two semi-determinate (Dt1Dt1Dt2Dt2) varieties, JKK378 and HXW, to identify QTLs (in genes other than Dt1 and Dt2) for PH, NN, and IL under long-day conditions. Intriguingly, we identified a QTL related to PH and NN whose candidate causal gene is Dt2. qRT-PCR and transient transfection assays confirmed that polymorphisms in the promoter of Dt2 between plants of the qPH18JKK378 and qPH18HXW groups caused the differences in Dt2 and Dt1 expression and external performance (PH, NN, IL, and grain weight per plant). The results of this study provide basic resources and lay the foundation for further exploring the mechanism underlying yield in soybean.

Materials and methods

Plant materials and field trials

The JH population of 267 F6:8 soybean RILs was obtained by single seed descent (SSD) from a cross between JKK378 (e1-as/e2/E3/E4/Dt1/Dt2) and HXW (e1-as/e2/E3/E4/Dt1/Dt2). The JH population was grown in Harbin (45°45′N, 126°38′E), Heilongjiang Province, China, in 2018 and 2019. The population and parents were planted in 2-m-long rows, each containing 15 plants. PH, NN, and IL were measured and calculated after harvesting. NN and PH were measured from the cotyledonary node to the top of the apex at maturity. IL = PH / NN of the main stem per plant.
Statistical analysis of phenotypic data in the population

Correlation analysis was performed to test the significance of the differences in PH, NN, and IL between parents using SPSS 18.0 (SPSS Inc., Chicago, IL, USA).

DNA isolation

Fresh, fully developed trifoliate leaves from the parents and the RIL individuals were collected, frozen in liquid nitrogen, and stored at −80 °C. Total genomic DNA was extracted from the samples using a Plant Genomic DNA Kit (CWBIO, Beijing, China). The integrity and quality of the extracted DNA were appraised by 1% agarose gel electrophoresis. The DNA concentration of each sample was measured using a Qubit® 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and a NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA).

Genotyping by high-throughput sequencing

Approximately 1.5 µg of DNA was prepared from each parent for whole-genome resequencing to generate sequencing libraries as previously described (Cheng et al. 2015). Two parental libraries were sequenced on the Illumina Hi-Seq 2500 platform (Illumina, Inc., San Diego, CA, USA), and 150 bp paired end reads with an insert size of approximately 300 bp were obtained. Genotypes were identified by low coverage sequencing of the JH population based on the parental polymorphic loci as a reference. The genomic DNA was incubated at 37 °C with the restriction endonuclease MseI (New England Biolabs, NEB, Ipswich, MA, USA), T4 DNA ligase (NEB), ATP (NEB), and the MseI Y-adapter N containing a barcode. The restriction-ligation reactions were heat-inactivated at 65 °C and digested by adding the restriction enzymes MseI+NlaIII at 37 °C. The restriction-ligation samples were purified with Agencourt AMPure XP (Beckman Coulter, Brea, CA, USA). PCR was conducted using diluted restriction-ligation DNA samples, dNTPs, Phusion Master Mix (NEB) universal primers, and index primers. The PCR products were purified using Agencourt AMPure XP (Beckman Coulter, Brea, CA, USA), pooled, and separated by 2% agarose gel electrophoresis. Fragments ranging in size from 375 to 400 bp (including indicators and adapters) were separated using a Gel Extraction kit (Qiagen), purified using Agencourt AMPure XP (Beckman, Irvine, CA, USA), and diluted for sequencing. Paired-end sequencing (each end was 150 bp long) was performed on the Illumina Hi-Seq 2500 system (Illumina, Inc., San Diego, CA, USA) following the manufacturer’s recommendations.

Sequence data grouping and SNP identification

Raw data (raw reads) were processed via a series of quality control (QC) procedures using in-house C programs to ensure that the reads were reliable without artificial bias. The QC standards included removal of the following: (1) reads with >50% bases and Phred quality <5; (2) reads with ≥10% unidentified nucleotides (N); (3) reads containing MseI and/or NlaIII cut-site remnant sequences; (4) reads with >10 nt aligned to the adapter, which allowed for ≤10% mismatches; and low-quality raw data (which were deleted first). The clean reads from each sample were aligned against the reference genome using Burrows-Wheeler Aligner (settings: mem-t4-M-R) (BWA v0.7.10) (Li 2009), and the alignment files were converted to BAM format with SAMtools software (v1.7.6) (Li et al. 2009). Variant calling of all samples was conducted using GATK (v3.8.0) (Wang et al. 2010). Single nucleotide polymorphism (SNPs) were filtered using a Perl script and annotated using ANNOVAR (v20170716) based on the GFF files of the reference genome (Wang et al. 2010). Markers with >30% missing genotype data, segregation distortion (P<0.01), or abnormal bases were screened out during map construction.

High-density genetic map construction

Chi-square (χ²) tests were carried out to detect segregation distortion for all SNPs. After genotyping, the markers were distributed into 20 linkage groups (or chromosomes) based on their physical positions, which were calculated (in cM) using QTL IciMapping software (www.isbreeding.net) (Meng et al. 2015).

QTL analysis for PH, NN, and IL using the ICIM method

QTLs for NN, PH, and IL were detected using QTL IciMapping4.2 software with inclusive composite interval mapping (ICIM) to identify QTLs with
the parameters PIN (probability in stepwise regression) = 0.001 and step = 1.0 cM. Nonsignificant QTLs with logarithm-of-odds (LOD) scores between 2.5 and the permutation test LOD threshold were also included in subsequent analysis and were considered to be suggestive QTLs (Kong et al. 2018a, b). To eliminate the influence of environment, we calculated the correlation coefficients of NN and PH from the RILs in 2018 and 2019. The PH trait showed strong correlation over the 2 years, as did NN (Supplementary Table 1). Therefore, we also used the average data over the 2 years for PH and NN, as well as IL, to detect QTLs.

qRT-PCR analysis of Dt2 and Dt1

RNA was extracted from the stem apical meristems (SAMS) of plants at the V2 stage using TRIzol Reagent (Ultrapure RNA Kit, CWBIO, China). The RNA was reverse transcribed into cDNA using a kit (Transcript®, China). The cDNA was subjected to qPCR on the LightCycle 480 system (Roche, Roche Diagnostics, Rotkreuz, Switzerland) using TBGreen II qPCR Mix (Vazyme Biotech, Nanjing, China). Using GmTubulin as internal reference, the expression levels of DT2 and Dt1 were determined using the following formula (Song et al. 2013): Relative Expression = 2\(^{\Delta C_t} = 2^{C_t(\text{GmTubulin})−C_t(\text{target genes})}\). Dt2-specific primers for qRT-PCR were designed with Primer Premier 5.0. The primers are listed in Supplementary Table 11.

Construction of the transient transformation vector and transient transformation of Arabidopsis thaliana protoplasts

Two different promoter sequences were cloned with specific primers using sequences from JKK378 and HXW as templates. The purified and recovered promoter sequence fragments were cloned into the pGreen-0800 vector using Exnase® II (ClonExpress® II One Step Cloning Kit, Vazyme, Nanjing, China). The vector included Renilla luciferase (REN) driven by the 35S promoter and luciferase (LUC) driven by the target promoter. The recombined vector was transformed into E. coli DH5α cells. Following sequence confirmation of positive plasmids, the plasmids were transformed into Arabidopsis protoplasts via PEG-mediated transformation (Lu et al. 2020). Healthy, fully expanded leaves of wild-type Arabidopsis Columbia were used to prepare protoplasts. Fluorescence detection was performed 18 h after conversion using a microplate reader (Biotek Synergy H1, USA). Promoter activity was measured based on the ratio of LUC activity divided by REN activity.

Results

Statistical analysis of phenotypic data

The parents of the RIL population have the same genotypes (e1-as/e2/E3/E4/Dt1/Dt2) in the major flowering time genes E1–E4 and the stem growth habit genes Dt1 and Dt2. JKK378 and HXW showed significant differences in PH and NN, but their IL did not greatly differ from each other in 2018 or 2019 (Table 1 and Fig. 1). However, a continuous distribution and transgressive segregation were observed for all three traits in the RIL population in 2018 and 2019 (Table 1 and Supplementary Fig. 1). The absolute value of skewness and the kurtosis was <1 for all three traits in both years (Table 1). These results indicate that PH, NN, and IL are quantitatively controlled by multiple loci, and they imply the existence of respective loci in the two parents controlling PH, NN, and IL. The heritability (\(h^2\)) values of PH, NN, and IL were 90.62%, 80.98%, and 58.49%, respectively (Table 1). The PH and NN data were strongly correlated over the 2 years, as was IL (Supplementary Table 1). Therefore, we also used the average data for PH, NN, and IL to detect QTLs.

Analysis of sequencing data and construction of genetic linkage maps

JKK378 and HXW were sequenced at an average depth of 15.61\(\times\) and 13.42\(\times\) and generated 14,817,826,500 and 12,740,326,800 bases, respectively. Of the total reads, 92.93% and 92.85% were of high quality, with an average Q30 ratio and GC content of 35% and 36%, respectively. The coverage of reads that mapped to the reference genome was 91.60% for both parental cultivars (Supplementary Table 2). We identified a total of 647,829 SNPs. Ultimately, 4031 SNP markers were used to construct a linkage map covering 2909.01 cM, with an average marker interval of 0.72 cM (Supplementary Fig. 2 and Supplementary Table 3).
QTL mapping for PH, NN, and IL

The thresholds of the LOD scores from the ICIM method used to evaluate the statistical significance of the QTL effect are shown in Supplementary Table 4. Common loci located in adjacent regions with the same effect in both years were considered to be consistent QTLs.

We detected a QTL associated with the PH trait, qPH18 (QTL for PH on chromosome 18), in both years (2018 and 2019). As expected, this QTL was also detected from the average data for both years. This QTL explained 7.16%, 10.90%, and 8.17% of the observed PVE with maximum LOD scores of 5.68, 9.92, and 8.99 and an additive effect of 0.51, 0.67, and 0.50, respectively. The QTL qIL18 for the IL trait explained 2.90%, 11.30%, and 13.12% of the observed PVE, with a maximum LOD score of 9.05, 9.46, and 9.47, respectively. The physical region harboring the QTL for the IL trait on chromosome 18 overlapped with those of the NN and PH traits. We conclude that the QTLs named qPH18, qNN18, and qIL18, which were detected on chromosome 18, are the same QTL and that this QTL simultaneously regulates PH, NN, and IL and is the major QTL in the JH RIL population (Table 2 and Supplementary Fig. 3).

The qPH18 allele, which was derived from JKK378, results in more nodes and greater PH and IL. Another QTL for PH was detected at a similar position as the QTL for IL on chromosome 4 named qPH4.2. qPH4.2 was detected in Harbin in 2018 and 2019 and in the average data for the two growing seasons (Supplementary Table 5 and Supplementary Fig. 3). qPH4.2 is located between markers

---

Table 1: Results of statistical analysis of the parents and the entire JH population from 2018 to 2019

| Trait | Environment | Parents | RIL |
|-------|-------------|---------|-----|
|       |             | JKK378  | HXW | Range | P-value | Min | Max | Range | Mean | CV% | Skewness | Kurtosis | h² | |
| PH    | 2018        | 115.10  | 76.00 | 39.1  | 2.57E-04 | 72.50 | 198.80 | 126.30 | 139.43 | 19%  | 0.41  | −0.52  | 90.62% |
|       | 2019        | 112.55  | 66.30 | 46.25 | 3.79E-09 | 62.20 | 178.70 | 116.50 | 123.02 | 31%  | −0.01 | −0.03  | |
|       | Average PH  | 112.50  | 71.50 | 41.00 | 2.07E-07 | 73.70 | 186.20 | 112.51 | 132.08 | 17%  | 0.28  | −0.42  | |
| NN    | 2018        | 24.00   | 16.40 | 7.60  | 2.67E-04 | 17.33 | 27.00  | 9.67  | 20.94 | 28%  | 0.50  | 0.94   | 80.98% |
|       | 2019        | 23.75   | 14.20 | 9.55  | 1.11E-08 | 15.80 | 29.00  | 13.2  | 20.94 | 28%  | 0.50  | 0.94   | |
|       | Average NN  | 23.87   | 15.30 | 8.57  | 5.56E-07 | 17.53 | 26.30  | 8.77  | 21.54 | 9%   | 0.17  | 0.26   | |
| IL    | 2018        | 4.82    | 4.66  | 0.16  | 6.80E-01 | 4.02  | 9.97   | 5.95  | 6.35  | 20%  | 0.51  | −0.24  | 58.49% |
|       | 2019        | 4.63    | 4.67  | 0.04  | 6.90E-01 | 3.49  | 8.38   | 4.89  | 5.88  | 30%  | 0.42  | −0.04  | |
|       | Average IL  | 4.72    | 4.66  | 0.06  | 7.20E-01 | 3.76  | 8.89   | 5.13  | 6.14  | 16%  | 0.55  | 0.08   | |

aPlant height is the distance of the main stem (in cm) from the cotyledonary node to the top of the apex at maturity
bAverage plant height in two environments (2018 and 2019 in Harbin)
cAverage node number in two environments (2018 and 2019 in Harbin)
dAverage internode length in two environments (2018 and 2019 in Harbin)
eMinimum value of the population
fMaximum value of the population
gThe difference between the minimum and maximum values
hAverage data from the entire population
iCoefficient of variance in percentage type
jBroad sense heritability
Fig. 1 Phenotypes of the parents of the JH population. JKK378 is shown on the left and HXW is shown on the right. a Fully mature JKK378 and HXW plants. b, e Plant height of the parents in 2018 (b) and 2019 (e) in Harbin. c, f Node number of the parents in 2018 (c) and 2019 (f) in Harbin. d, g Internode length of the parents in 2018 (d) and 2019 (g) in Harbin. Bar = 20 cm.

Table 2 Detailed information about the major QTLs for PH, NN, and IL on chromosome 18 in the JH population detected by ICIM

| QTL name | Environment | Chr. | Left marker | Right marker | Physical length | Max LODa | PVE (%)b | Addc |
|-----------|-------------|------|-------------|--------------|----------------|----------|---------|-------|
| qPH18     | 2018        | 18   | 18-MAR-61   | 18-MAR-73    | 53,950,157–56,334,200 | 16.95    | 21.69   | 10.87 |
| qPH18     | 2019        | 18   | 18-MAR-64   | 18-MAR-65    | 54,577,297–55,740,447 | 8.92     | 12.56   | 6.56  |
| qPH18     | Average     | 18   | 18-MAR-64   | 18-MAR-65    | 54,577,297–55,740,447 | 27.32    | 29.80   | 12.44 |
| qNN18     | 2018        | 18   | 18-MAR-67   | 18-MAR-68    | 55,740,294–55,826,575 | 5.68     | 7.16    | 0.51  |
| qNN18     | 2019        | 18   | 18-MAR-61   | 18-MAR-73    | 53,950,157–56,334,200 | 9.92     | 10.90   | 0.67  |
| qNN18     | Average     | 18   | 18-MAR-67   | 18-MAR-68    | 55,740,294–55,826,575 | 8.99     | 8.17    | 0.50  |
| qIL18     | 2018        | 18   | 18-MAR-72   | 18-MAR-73    | 55,826,366–56,334,200 | 9.05     | 2.94    | 0.37  |
| qIL18     | 2019        | 18   | 18-MAR-61   | 18-MAR-73    | 53,950,157–56,334,200 | 9.46     | 11.30   | 0.34  |
| qIL18     | Average     | 18   | 18-MAR-72   | 18-MAR-73    | 55,826,366–56,334,200 | 9.47     | 13.12   | 0.33  |

a Chromosome
b Maximum logarithm-of-odds (LOD) scores
c Percentage of phenotypic variance explained by the QTL
d Additive effects contributed by QTLs. Positive values represent alleles from HXW, and negative values represent alleles from JKK378
e LOD scores ≥ the threshold
4-MAR-21 and 4-MAR-22 at a physical position of 2,452,312–2,443,173 bp. This QTL explained 4.54%, 4.93%, and 2.56% of the observed PVE, with a maximum LOD score of 4.16, 4.08, and 3.28 and an additive effect of 4.95, 4.12, and 3.65, respectively. The minor QTL qPH9 is located between markers 9-MAR-27 and 9-MAR-28 on chromosome 9 at a similar physical position (4,970,193–6,972,683 bp) to the QTL for NN. This QTL explained 2.8% of the observed PVE, with a maximum LOD score of 2.5 and an additive effect of 3.91 (Supplementary Table 5 and Supplementary Fig. 3).

Candidate gene prediction

In summary, we detected one major QTL at the end of the short arm of chromosome 18 related to PH, NN, and IL. We combined the intervals of these interrelated QTLs to obtain a larger QTL interval and used it to mine candidate genes. To identify candidate genes, we combined the genes in the larger QTL regions with interparental gene variation and performed Gene Ontology (GO) annotation and functional annotation of orthologs from Arabidopsis.

We identified 295 genes in the major qPH18 interval 53,950,157–56,334,200 bp on chromosome 18 according to Gmax_275_Wm82.a2.v1 (Schmutz et al. 2010). Based on SNP identification and Indel calling of the resequencing results of the parents, 247 SNPs resulting in amino acid changes or frameshift mutations were identified in the exons of 54 genes (Supplementary Table 6). Among the 295 genes, 255 genes could be functionally annotated using WEGO2.0 (Supplementary Table 7). Of these, 161, 62, and 223 genes were functionally annotated to the biological processes, cellular components, and molecular function categories, respectively (Supplementary Fig. 4). Of these genes, seven were related to transcription factor activity, including Dt2 (Glyma.18G273600), and were therefore considered to be major candidate genes. In addition, when we compared the 295 genes, two genes, Glyma.18G275200 and Glyma.18G282000, were associated with plant growth and development based on the annotation information and had mutations in the coding sequence (CDS) region (Supplementary Table 8). Previous studies have shown that homologs of these genes in Arabidopsis or rice influence plant architecture (Kasajima et al. 2006; Nobutoshi et al. 2007; Ivanova et al. 2014; Zhuang et al. 2010). Therefore, these genes were classified as the most likely candidate genes (Supplementary Table 8).

Expression analysis of Dt2 of qPH18 for PH, NN, and IL

Based on our genotyping-by-sequencing results, the RILs were divided into two subgroups depending on whether they possessed homozygous alleles of JKK378 or HXW at qPH18. Analysis of variance revealed significant differences in PH, NN, IL, and grain weight per plant in 2018 and 2019; members of the subgroup harboring the qPH18 JKK378 allele showed increased PH, NN, and IL, ultimately leading to increased yield (Fig. 2). The list of candidate genes of qPH18 included Dt2. Even though both parents are semi-determinate, possessing Dt2, we still considered Dt2 to be a likely candidate gene due to its key role in regulating NN, PH, and growth habit (Ping et al. 2014). Thus, we performed genomic sequencing and compared the 3-kb promoter sequences between JKK378 and HXW. Three polymorphisms were detected in the Dt2 promoter, including two SNPs and one Indel. The polymorphisms SNP_-2570 (T-A), SNP_-2420 (T-C), and Indel_-2376 (TAT A TA TAT A TA TAT A) led to changes in the cis-acting elements POLLEN1LELAT52-motif (sense strand), a CAAT-box (sense strand), and a TATA-box (sense strand) based on NEW PLACE analysis. These variations in the promoters of the two dominant Dt2 alleles from JKK378 and HXW might lead to a difference in the level of Dt2 gene expression.

To test this possibility, we examined the expression level of Dt2 in the SAMs of plants at the V2 stage, including the two parents and two subgroups divided based on the polymorphisms in the Dt2 promoter. Dt2 was expressed at significantly lower levels in JKK378 than in HXW. In addition, several random samples from the subgroup harboring the qPH18 JKK378 allele showed significantly lower levels of Dt2 expression than HXW and members of the other subgroup (Fig. 3a). There were no significant differences in the expression levels of Glyma.18G275200 or Glyma.18G282000 in the SAMs of the parents at the V2 stage (Supplementary Fig. 5), which helped confirm the identity of the candidate gene. In addition, we compared the expression levels of Dt1 between the two subgroups. As expected, Dt1 expression was
inhibited in the subgroup harboring the qPH18\textsubscript{JKK378} allele (Fig. 3b). These results support the model in which \textit{Dt2} inhibits \textit{Dt1} to regulate NN, IL, and PH.

To validate the functional differences in the \textit{Dt2} promoters, we performed a transient transfection assay to determine whether polymorphisms in the \textit{Dt2} promoter cause a difference in promoter activity. The two different promoter sequences were cloned and introduced into the \textit{pGreen0800-LUC} vector to generate reporter constructs (Fig. 3c). Subsequently, plasmids containing the reporter were transformed into \textit{Arabidopsis} protoplasts. After 18 h, the LUC and REN activities of the two reporter constructs were detected in a microplate assay. Two-way ANOVA ($P=2.61E-18$) revealed a significant difference in the LUC/REN ratios of protoplasts harboring the different promoters. LUC activity associated with the \textit{Dt2} promoter was significantly lower in JKK378 than in HXW (Fig. 3d), suggesting that the sequence variations in the promoters lead to the strong functional differences in \textit{Dt2} activity. Taken together, we conclude that \textit{Dt2} is the candidate gene of qPH18, although fine mapping is still required to confirm this.

In addition, we developed a dCAPs molecular marker focused on SNP\_\textsubscript{-2420} (T-C) (Supplementary Fig. 6). The products obtained by two rounds of PCR using specific primers for HXW, followed by digestion with the restriction enzyme \textit{MluI}, could be distinguished by polyacrylamide gel electrophoresis. After digestion, the fragment from HXW was 172 bp long, while that from JKK378 was 197 bp long. These results demonstrate that the dCAPs marker could be used...
to genotype the two different promoters between JKK378 and HXW, which has broad applications for molecular breeding.

Discussion

A genetic map constructed using a RIL population represents an excellent resource for identifying QTLs underlying a trait of interest. With the emergence of new molecular markers, high-throughput sequencing has become an ideal solution for obtaining SNP markers and for constructing high-density linkage maps for QTL mapping (Zhou et al. 2016; Kong et al. 2018a, b; Fang et al. 2019). In the present study, we constructed a high-density genetic map to facilitate QTL analysis. Based on this map, we detected several QTLs associated with PH, NN, and IL in soybean and identified candidate genes in these loci using multiple methods.

PH is an important yield-related agronomic trait in soybean. PH and NN as well as PH and IL are positively correlated (Liu et al. 2011; Allen et al. 2018; Chang et al. 2018; Assefa et al. 2019). Therefore, the same QTL may affect multiple traits due to its pleiotropic effects. In the present study, we identified a major QTL that affects PH, NN, and IL simultaneously. This QTL, named $qPH18$, is located on the terminus of chromosome 18 in the physical region of 53,950,157–56,334,200 bp (Table 2 and Supplementary Fig. 3). This interval has been identified multiple times and is also associated with several other traits, such as pod maturity, lodging, seed yield, and seed oil content (Supplementary Table 9) (Kabelka et al. 2004; Sun et al. 2006; Guzman et al. 2007; Qi et al. 2011; Kim et al. 2012). However, most QTLs only affect PH and NN or PH and IL, such as $qPH9$, which is associated with PH and NN, and $qPH4.2$, which affects PH and IL (Supplementary Fig. 3 and Supplementary Table 5). These findings indicate that PH is not affected by either NN or IL alone but is instead affected by the joint effects of both traits, which makes the genetic dissection of the PH trait
more complicated. In addition, the environmentally specific QTL qPH9 was detected only in 2018. We examined meteorological data such as effective accumulated temperature, daylength, and rainfall during the developmental stage examined in 2018 and 2019 and established that effective accumulated temperature and rainfall were significantly different during the flowering period (June and July) in 2018 vs. 2019 (Supplementary Table 10). These differences might help explain why this QTL was only detected in a single year.

To identify candidate genes, we selected two homologous genes related to PH, namely Glyma.18G275200 and Glyma.18G282000. Glyma.18G275200 encodes an auxin transport protein. The homologous gene in Arabidopsis is AT3G02260, whose mutants are characterized by organ elongation defects (dwarfism), altered root structure, decreased apical dominance, defective light responses, and abnormal auxin transport (Kasajima et al. 2006; Nobutoshi et al. 2007; Ivanova et al. 2014). Glyma.18G282000 encodes an ARF GTPase-activating protein; mutants of Os02g0198300, its homologous gene in rice, show abnormal auxin influx, vesicle trafficking, and root development (Zhuang et al. 2010). Homologous genes with the potential to influence the plant architecture of soybean were considered to be possible candidate genes.

A series of QTLs/loci associated with PH have already been reported. To date, two important genes, Dt1 and Dt2, are known to affect PH by regulating stem growth habit (Liu et al. 2010; Ping et al. 2014). Dt2 protein directly interacts with the Dt1 promoter to repress its transcription, resulting in a semi-determinate growth habit affecting PH. In the current study, we mapped the major QTL qPH18 onto chromosome 18; Dt2 is located within this region, as are some known Arabidopsis PH homologs and transcription factor genes (Supplementary Table 8). The phenotypic values of PH, NN, IL, and grain weight per plant showed obvious differences in two subgroups harboring the alleles of qPH18 from different parents, suggesting that qPH18 ultimately increases yield by affecting PH, NN, and IL (Fig. 2). This QTL deserves further study and could be used in subsequent breeding.

Even though both parents have a semi-determinate stem growth habit, indicating that they share the dominant Dt2 allele, we analyzed the expression levels of Dt2 in both subgroups at the V2 stage because Dt2 is an important regulator of NN, PH, and growth habit. The expression level of Dt2 indeed differed between JKK378 and HXW (Fig. 3a). A transient transfection assay also showed that the promoter activities of this gene in JKK378 and HXW were different, which is consistent with its gene expression levels (Fig. 3d). These results suggest that Dt2 is the candidate gene of qPH18. The dominant Dt2 allele does not show a single level of functional strength; there are at least two dominant Dt2 alleles with different intensities in nature. Classical genetic analyses showed that the functional strength (Dt1 or dt1) and expression level of Dt1 (regulated by Dt2) determine NN, PH, and stem growth habit in soybean. We propose that a certain threshold of the functional strength and expression level of Dt1 exists, which corresponds to the semi-determinate growth habit (Dt1Dt1-Dt2Dt2). Between this threshold and dt1dt1 (determinate growth habit), various levels of Dt1 expression can occur due to the various expression levels of Dt2, which ultimately appear as different NN and PH values. These insights could facilitate molecular breeding for NN and PH using dominant Dt2 with different levels of promoter activity.

Fine mapping of the above QTLs should help elucidate the molecular mechanisms underlying the regulation of PH, NN, and IL. Furthermore, our study provides soybean materials and a molecular marker to facilitate breeding of soybean with improved yield.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-021-01235-y.

Acknowledgements We would like to acknowledge Mrs. Yafeng Liu for phenotyping and managing the field.

Author contribution CF, FK, and BL designed the experiments. KK, TS, YW, HY, HD, MH, TL, LM, CL, CY, WS, LC, YL, and BY carried out the experiments. KK, LK, SL, LW, XZ, SL, BL, FK, and CF analyzed the data. KK, FK, BL, and CF wrote the paper.

Funding This work was supported by National Natural Science Foundation of China (31930083) and EUCLEG Horizon 2020 of European Union (No.727312). This work was also funded by the National Key Research and Development Program (Nos. 2017YFE0111000), National Natural Science Foundation of China (31701445, 31930083, 31725021, 31801384), the Major Program of Guangdong Basic and Applied Research (2019B030302006), and Open Project...
Foundation of National Key Laboratory for Crop Genetics and Germplasm Enhancement (ZKW201901).

Data availability Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

Allen LH, Zhang L, Boote KJ, Hauser BA (2018) Elevated temperature intensity, timing, and duration of exposure affect soybean internode elongation, mainstem node number, and pod number per plant. Crop J 6:148–161. https://doi.org/10.1016/j.cj.2017.10.005

Ashikari M, Sasaki A, Ueguchi-Tanaka M et al (2002) Loss-of-function of a rice gibberellin biosynthetic gene, GA20 oxidase (GA20ox2), Led to the Rice “Green Revolution.” Breed Sci 52:143–150

Assefa T, Otyama PI, Brown AV et al (2019) Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. BMC Genomics 20:527

Bernard RL (1971) Two major genes for time of flowering and maturity in soybeans. Crop Sci 11:242–244. https://doi.org/10.2135/cropsci1971.0011183X001100020022x

Bernard RL (1972) Two genes affecting stem termination in soybeans. Crop Sci 12:235–239

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. https://doi.org/10.3389/fpls.2018.01184

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

Cheng W, Liu F, Li M et al (2015) Variation detection based on next-generation sequencing of type Chinese 1 strains of Toxoplasma gondii with different virulence from China. BMC Genomics 16:888. https://doi.org/10.1186/s12864-015-2106-z

Fang C, Chiu L, Nan H (2019) Rapid identification of consistent novel QTLs underlying long-juvenile trait in soybean by multiple genetic populations and genotyping-by-sequencing. Mol Breed 39:80

Guzman PS, Diers BW, Neece DJ et al (2007) QTL associated with yield in three backcross-derived populations of soybean. Crop Sci 47:111. https://doi.org/10.2135/cropsci2006.01.0003

Heatherly L, Smith J (2004) Effect of soybean stem growth habit on height and node number after beginning bloom in the midsouthern USA. Crop Sci 44:1855–1858. https://doi.org/10.2135/cropsci2004.1855

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

Ivanova A, Law SR, Narsai R et al (2014) A functional antagonistic relationship between auxin and mitochondrial retrograde signaling regulates alternative oxidase A expression in arabidopsis. Plant Physiol 165:1233

Kabelka EA, Diers BW, Fehr WR et al (2004) Putative alleles for increased yield from soybean plant introductions. Crop Sci 44:784–791. https://doi.org/10.2135/cropsci2004.0784

Kasajima I, Okhama-Ohtsu N et al (2006) The BIG gene is involved in regulation of sulfur deficiency-responsive genes in Arabidopsis thaliana. Physiol Plant 129(2):351–363

Kim KS, Diers BW, Hyten DL et al (2012) Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. Theor Appl Genet 125:1353–1369. https://doi.org/10.1007/s00122-012-1944-1

Kong L, Lu S, Wang Y et al (2018a) Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. Front Plant Sci 9:1–20. https://doi.org/10.3389/fpls.2018.00995

Kong L, Sijia L, Yanping W et al (2018) Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. Frontiers in Plant Science 9:995

Li H (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics (Oxford, England) 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu B, Watanabe S, Uchiyama T et al (2010) Innovation of a regulatory PRR genes controlling flowering and maturity in soybean. J Crop Sci Biotechnol 14:65–70

Liu W, Kim MY, Van K et al (2011) QTL identification of yield-related traits and their association with flowering and maturity in soybean. J Crop Sci Biotechnol 14:65–70

Liu Y, Zhang D, Ping J et al (2016) The BIG gene is involved in regulation of sulfur deficiency-responsive genes in Arabidopsis thaliana. Physiol Plant 129(2):351–363

Liu Y, Zhang D, Ping J et al (2016) The BIG gene is involved in regulation of sulfur deficiency-responsive genes in Arabidopsis thaliana. Physiol Plant 129(2):351–363

Monna L, Noriyuki K, Rika Y et al (2002) Positional cloning of rice semidwarfing gene, sd-1: rice “green revolution
gene" encodes a mutant enzyme involved in gibberellin synthesis. DNA Res 9(1):11–17
Nobutoshi Y, Mitsuhiro S, Hidehiro F et al (2007) CRM1/BIG-mediated auxin action regulates arabidopsis inflorescence development. Plant Cell Physiol 48:1275–1290
Peng J, Richards DE, Hartley NM et al (1999) “Green revolution” genes encode mutant gibberellin response modulators. Nature 400:256–261
Ping J, Liu Y, Sun L et al (2014) Dt2 Is a gain-of-function MADS-domain factor gene that specifies Semideterminacy in Soybean. Plant Cell 26:2831–2842
Qi ZM, Wu Q, Han X et al (2011) Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. Euphytica 179:499–514. https://doi.org/10.1007/s10681-011-0386-1
Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. https://doi.org/10.1038/nature08670
Setiyono T, Weiss A, Specht J et al (2007) Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. Field Crop Res 100:257–271. https://doi.org/10.1016/j.fcr.2006.07.011
Song QX, Li QT et al (2013) Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic Arabidopsis plants. J Exp Bot 64(14):4329–4341
Sun D, Li W, Zhang Z et al (2006) Quantitative trait loci analysis for the developmental behavior of Soybean (Glycine max L. Merr.). Theor Appl Genet 112:665–673. https://doi.org/10.1007/s00122-005-0169-y
Takeshima R, Nan H, Harigai K, Dong L (2019) Functional divergence between soybean flowering Locus T orthologues, FT2a and FT5a, in post-flowering stem growth. J Exp Bot 67:5247–5258
Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:164–164. https://doi.org/10.1093/nar/gkq603
Xu M, Xu Z, Liu B et al (2013) Genetic variation in four maturity genes affects photoperiod insensitivity and PHYA-regulated post-flowering responses of soybean. BMC Plant Biol 13:91
Zhou Z, Zhang C, Zhou Y et al (2016) Genetic dissection of maize plant architecture with an ultra-high density bin map based on recombinant inbred lines. BMC Genomics 17:1–15. https://doi.org/10.1186/s12864-016-2555-z
Zhuang X, Jiang J, Li J et al (2010) Over-expression of OsA-GAP, an ARF-GAP, interferes with auxin influx, vesicle trafficking and root development. Plant J 48:581–591

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.