Detecting QTL and Candidate Genes for Plant Height in Soybean via Linkage Analysis and GWAS

Jiajing Wang†, Bo Hu†, Yuliang Jing, Xiping Hu, Yue Guo, Jiankun Chen, Yuxi Liu, Jianhui Hao, Wen-Xia Li* and Hailong Ning*

OPEN ACCESS

Edited by:
Yuan-Ming Zhang, Huazhong Agricultural University, China

Reviewed by:
Weiren Wu, Fujian Agriculture and Forestry University, China
Caixiang Wang, Gansu Agricultural University, China
Xhuan Li, Agricultural University of Hebei, China
Chao Fang, Michigan State University, United States

*Correspondence:
Wen-Xia Li
liwenxianeau@126.com
Hailong Ning
ninghailongneau@126.com

†These authors share first authorship

Specialty section:
This article was submitted to Plant Breeding, a section of the journal Frontiers in Plant Science

Received: 28 October 2021
Accepted: 20 December 2021
Published: 21 January 2022

Citation:
Wang J, Hu B, Jing Y, Hu X, Guo Y, Chen J, Liu Y, Hao J, Li W-X and Ning H (2022) Detecting QTL and Candidate Genes for Plant Height in Soybean via Linkage Analysis and GWAS. Front. Plant Sci. 12:803820. doi: 10.3389/fpls.2021.803820

Soybean is an important global crop for edible protein and oil, and plant height is a main breeding goal which is closely related to its plant shape and yield. In this research, a high-density genetic linkage map was constructed by 1996 SNP-bin markers on the basis of a recombinant inbred line population derived from Dongnong L13 × Henong 60. A total of 33 QTL related to plant height were identified, of which five were repeatedly detected in multiple environments. In addition, a 455-germplasm population with 63,306 SNP markers was used for multi-locus association analysis. A total of 62 plant height QTN were detected, of which 26 were detected repeatedly under multiple methods. Two candidate genes, Glyma.02G133000 and Glyma.05G240600, involving in plant height were predicted by pathway analysis in the regions identified by multiple environments and backgrounds, and validated by qRT-PCR. These results enriched the soybean plant height regulatory network and contributed to molecular selection-assisted breeding.

Keywords: soybean, plant height, QTL, QTN, candidate genes

INTRODUCTION

Soybean [Glycine max (L.) Merr.] is one of the most important crops in the world as a major source of protein and oil (Feng et al., 2021). Plant height (PH) as the main trait of soybean plant shape is related to soybean yield (Assefa et al., 2019). Low plants result in lower yields, while too high plants are prone to yield reduction due to lodging. Plant height also affects yield through other traits such as number of pods per plant and number of nodes in the main stem (Chang et al., 2018; Li M. W. et al., 2020). PH was a complex quantitative trait which was controlled by multiple genes and influenced by environmental conditions (Lee et al., 1996).

With the objective to breeding efficiently, QTL mapping for PH were conducted by linkage and genome-wide association analysis (GWAS) analysis. Based on the bi-parent derived populations and linkage analysis (Xu et al., 2017), 238 QTLs associated with plant height had been listed on all 20 chromosomes. In these researches, most of the linkage maps were constructed by low-throughput molecular markers such as restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), and simple sequence repeat (SSR) markers, which led to low marker density, large genomic region intervals for QTL localization. It was difficult to predict candidate
genes and design marker-assisted selection for PH. With the continuous development of molecular markers, high-throughput and high-density single nucleotide polymorphism markers (SNPs) were used as major markers for linkage analysis for mapping QTL (Adewusi et al., 2017; Gomez-Casati et al., 2017; Zhang et al., 2019; Karikari et al., 2020; Tian et al., 2020; Salari et al., 2021; Silva et al., 2021). In order to construct effective linkage intervals to identify QTL, SNP bin maker technology were gradually used in construction of linkage map. Cao et al. (2019) constructed two linkage maps by 3,958 and 2,600 SNP bin markers and detected 39 PH QTLs on chromosome 5, 6, 7, 9, 10, 12, 15, 16, 18, and 20, and the phenotypic variation explained (PVE) ranged from 1.11 to 18.99% based on a recombinant inbred line population of soybean. The second method for detecting QTL was GWAS, which has been extensively studied through recombinant inbred lines and germplasm populations of soybean (Lü et al., 2016; Qi et al., 2020; Song et al., 2020). With the objective of overcoming the shortage of false positives (Sonah et al., 2015), combinations of linkage and association analysis were gradually used in detecting genome regions controlling quantitative traits (Zhang et al., 2019; Song et al., 2020; Li X. et al., 2021). However, few studies combining both methods have been conducted on PH of soybean.

Based on the results of linkage and GWAS, some genes controlling PH formation were gradually mined, such as GA20ox, GA3ox (Fernandez et al., 2009), and GmGW1 (Han et al., 2021) control flowering time and PH in soybean. GmTFL1b determines PH and growth habit, which a candidate gene for Dt1 (Liu et al., 2020). In this research, QTL/QTN localization of soybean plant height was performed via linkage analysis of a recombinant inbred lines and GWAS of a 455-germplasm population. In the region of QTL/QTN, candidate genes related to PH formation were predicted and initially validated by qRT-PCR. The objective of this research was to lay foundation for probing genetic basis and molecular assistant selection of PH.

**MATERIALS AND METHODS**

**Plant Populations**

Two soybean varieties with large differences in PH, Dongnong L13 obtained from a cross between Heinong 40 and Jiujiao 5640 and Henong 60 obtained from a cross between Beifeng 11 and Hobbit, were used as parents to make cross in 2008 in Harbin, Heilongjiang Province (E 126.63°, N 45.75°). F1 was planted in Yacheng City, Hainan Province (E 109.00°, N 17.50°) in the winter of the same year. After five consecutive generations from 2010 to 2014 by planting in Harbin and Yacheng alternatively, 139 recombinant inbred lines were obtained and a population formed and were used to conduct linkage analysis. Furthermore, a 455-germplasm population, including 4 local soybean varieties, 387 domestic varieties and 44 foreign varieties, was used for GWAS. The variety name was described and published earlier by Li X. et al. (2020).

**Field Trials and Phenotypic Measurement**

RIL6013 were planted in eight environments at three locations: Harbin (E 126.63°, N 45.75°), Keshan (E 125.64°, N 48.25°) and Shuangcheng (E 126.92°, N 45.75°). About 455 germplasm resources were planted in Harbin and Shuangyashan (E 131.16°, N 46.64°) in 2018 and 2019, respectively. The detail information for each plant environment was summarized in

| Env. | P1 | P2 | Minimum (cm) | Maximum (cm) | Range (cm) | Mean ± STD (cm) | Skew | Kurt | CV (%) |
|------|----|----|-------------|-------------|------------|----------------|------|------|--------|
| E1   | –  | –  | 77.00       | 144.00      | 67.00      | 117.52 ± 15.40 | –0.63 | –0.22 | 13.10  |
| E2   | –  | –  | 67.40       | 151.40      | 84.00      | 120.49 ± 19.02 | –0.75 | –0.28 | 15.79  |
| E3   | 134.00 | 63.80 | 72.00      | 157.00      | 85.00      | 120.96 ± 19.31 | –0.48 | –0.54 | 15.96  |
| E4   | 132.00 | 63.00 | 65.00      | 155.00      | 90.00      | 120.42 ± 16.35 | –0.83 | 0.96  | 13.58  |
| E5   | 139.67 | 75.67 | 65.33      | 149.00      | 83.67      | 123.58 ± 17.97 | –0.80 | –0.17 | 14.54  |
| E6   | 135.00 | 72.00 | 60.00      | 163.33      | 103.33     | 133.23 ± 19.47 | –0.93 | 0.68  | 14.61  |
| E7   | 144.67 | 77.50 | 62.00      | 162.67      | 100.67     | 130.80 ± 19.01 | –0.80 | 0.63  | 14.53  |
| E8   | 137.50 | 77.00 | 58.00      | 152.00      | 94.00      | 122.04 ± 19.85 | –1.20 | 1.23  | 16.27  |
| E9   | –  | –  | 37.00       | 131.00      | 94.00      | 86.82 ± 15.65 | –0.19 | 0.54  | 18.03  |
| E10  | –  | –  | 38.33       | 176.67      | 138.34     | 95.83 ± 17.44 | 0.48  | 2.40  | 18.20  |
| E11  | –  | –  | 40.00       | 168.67      | 128.67     | 97.35 ± 19.77 | –0.10 | 0.08  | 20.31  |
| E12  | –  | –  | 47.00       | 158.00      | 111.00     | 93.13 ± 16.50 | 0.12  | 0.22  | 17.71  |

*aParents: P1, female cultivar "Dongnong L13"; P2, male cultivar "Henong 60". bRange, difference between maximum and minimum value. cMean ± standard deviation of the observed values. dCoefficient of variation.*

Frontiers in Plant Science | www.frontiersin.org 2 January 2022 | Volume 12 | Article 803820
Supplementary Table 1. The field experiments were conducted in a randomized block in replication design (RBRD). RBRD is a randomized incomplete block design with three replicates used in each environment. With the objective to eliminate the difference of blot among large amounts of lines in a replicate (block), the replicates were divided into multiple sub-blocks which contain about 15 lines. Three ridges were contained in one blot, and the ridges were 3 m in length and 0.67 m in width. The seeds were

![FIGURE 1](https://example.com/figure1.png)

**FIGURE 1** | Frequency histograms of plant height in RIL 6013 in eight environments.
TABLE 2 | Joint ANOVA of PH of RIL6013 population and 455 germplasm resources in multiple environment and heritability.

| Population     | Source                | DF  | SS        | MS        | F       | Pr > F      | σ²  |
|----------------|-----------------------|-----|-----------|-----------|---------|-------------|-----|
| RIL6013        | Replication           | 2   | 208,033   | 104,012   | 0.36    | 0.6949      |     |
|                | Environment           | 7   | 93,199,500| 13,314,212| 48.6    | <0.0001     |     |
|                | Block                 | 9   | 239,34,861| 26,59,434 | 9.31    | <0.0001     |     |
|                | Replication × Block   | 18  | 874,039,2 | 485,583   | 1.7     | 0.033       |     |
|                | Block (Genotype)      | 129 | 3,511,27,234| 2,721,923 | 9.53    | <0.0001     | 74.412 |
|                | Replication × Environment| 14 | 310,7,823 | 221,994   | 0.78    | 0.6954      |     |
|                | Environment × Block   | 63  | 679,99,444| 10,79,362 | 3.78    | <0.0001     |     |
|                | Replication × Environment × Block | 126 | 386,42,481| 306,691  | 1.07    | 0.2779      |     |
|                | Environment × Block (Genotype) | 896 | 849,663,510| 948,291  | 3.32    | <0.0001     | 220.863 |
|                | Error                 | 2050| 585,738,432| 285,733  |         | 285.733     |     |
|                | Total                 | 3314| 201,927,931|          |         |             |     |
| Germplasm      | Replication           | 2   | 1,353,231 | 676,612   | 2.46    | 0.0852      | 65% |
|                | Environment           | 3   | 1,011,53,852| 337,17,952| 122.83  | <0.0001     |     |
|                | Block                 | 30  | 197,653,374| 658,845  | 24      | <0.0001     |     |
|                | Replication × Block   | 60  | 172,56,132 | 287,630  | 1.05    | 0.3769      |     |
|                | Block (Genotype)      | 424 | 91,488,0,761| 215,7,744| 7.86    | <0.0001     | 139.413 |
|                | Replication × Environment | 6  | 381,323   | 63,533   | 0.23    | 0.9665      |     |
|                | Environment × Block   | 86  | 11,277,3,772| 131,1,321| 4.78    | <0.0001     |     |
|                | Replication × Environment × Block | 172 | 459,13,361| 266,942  | 0.97    | 0.5865      |     |
|                | Environment × Block (Genotype) | 1178 | 753,903,492| 639,993  | 2.33    | <0.0001     | 125.211 |
|                | Error                 | 3204| 879,537,283| 274,514  |         | 274.512     |     |
|                | Total                 | 5165| 307,162,641|          |         |             |     |

h² 65%

h² 72%

FIGURE 2 | The high-density bin map of RIL6013.
sowed in single row on the ridges with the plant space set 0.07. The whole experiments were managed as the same as local field production. Ten plants were randomly selected from each plot to determine PH at the maturity stage. The average value of the 10 plants was used as the observation value of the plot, and finally the average value of the three blots was used for QTL and QTN mapping.

Statistical Analysis of Phenotype Data

Frequency distribution histograms were drawn from the phenotypic data of PH in each environment and descriptive statistics were performed. Analysis of variance (ANOVA) and estimation of generalized heritability were also performed as Li X. et al. (2021). The statistical model for the multi-environment ANOVA for RBRD was showed as follows:

\[
x_{ijk} = \mu + E_i + R_j + B_k + R_{Bjk} + B_k(G_i) + ER_{ej} + EB_{ek} + ERR_{eijk} + EB_{G}(G_i) + \varepsilon_{eijk}
\]

where \( \mu \) is the grand average, \( G_i \) is the \( i \)th genotype effect, \( E_i \) is the \( i \)th environment effect, \( R_j \) is the \( j \)th replication effect, \( B_k \) is the \( k \)th block in \( j \)th replicate, \( RB_{jk} \) is the interaction effect between \( j \)th replication and \( k \)th block, \( B_k(G_i) \) is \( i \)th genotype in \( k \)th block, \( ER_{ej} \) is interaction between \( e \)th environment and \( j \)th replication, \( EB_{ek} \) is interaction between \( e \)th environment and \( k \)th block, \( ERR_{eijk} \) is interaction effect among \( e \)th environment and \( j \)th replication and \( k \)th block, \( EB_{G}(G_i) \) is \( i \)th genotype under interaction of \( e \)th environment and \( k \)th block, and \( \varepsilon_{eijk} \) is the error effect following \( N(0, \sigma^2) \). The broad-sense heritability in multiple environments was showed as following:

\[
h^2 = \frac{\sigma^2_{B(G)}}{\sigma^2_{B(G)} + \frac{\sigma^2_{ER(G)}}{e} + \sigma^2_{\varepsilon}}
\]

where \( h^2 \) is the broad sense generalized heritability of average in over multiple environments, \( \sigma^2_{B(G)} \) is the variance of genotype under block, \( \sigma^2_{ER(G)} \) is the variance of genotype under environment \( \times \) block interaction, \( \sigma^2_{\varepsilon} \) is the error variance, \( e \) is the number of environments, and \( r \) is the number of repetitions in each environment. Significance of each factor was tested by the general linear model method and variance were estimated using mixed method implemented by SAS 9.2 (SAS Institute, Cary, NC, United States).

SNP Genotyping

DNA samples extracted by CTAB method from RIL6013 were genotyped for SNPs using a soybean SNP660K microarray at Beijing Boao Biotechnology Co., Ltd. A total of 54,836 SNPs were screened on 20 chromosomes. A total of 63,306 SNPs were screened on 20 chromosomes using a soybean SNP180K microarray for SNP genotyping of 455 DNA samples from germplasm resources at Beidahuang Kenfeng Seed Co., Ltd. The obtained SNP markers were screened according to the following criteria: minimum allele frequency for markers (MAF > 5%) and maximum deletion rate <10% for each SNP (Belamkar et al., 2016).

Bin Maker Map Construction and QTL Localization

Here the SNP data from RIL6013 was used to identify possible crossovers via python 2.7snpbinner, and the minimum distance between crossovers is 0.2% of the chromosome length. The aggregated breakpoints generated from the crossover points were then used to create representative bins for the entire population (minimum distance of 30 kb per bin). The obtained bin markers were used to construct a high-density genetic linkage map of SNPBins using the map function (Linkage map construction) in the software QTL IcimappingV 4.1 (Wang, 2009).

QTL IcimappingV 4.1 (Wang, 2009) software was used to locate additive QTL using two mapping methods: interval mapping (IM-ADD) and inclusive composite interval mapping (ICIM-ADD). The scan step was set to 1.00 cM and the LOD threshold was set to 2.50. The PIN value of the ICIM-ADD method was set to 0.001. The QTL were named using the method of McCouch (1997).

Genome-Wide Association Analysis

The population structure and LD of the germplasm resource population were described and published earlier by Li X. et al. (2020). The germplasm resource population consisted of two subpopulations containing 132 (29.01%) and 323 (70.99%) lineages, respectively (K = 2). And the physical distance of LD decay was estimated as the position where \( r^2 \) dropped
to half of its maximum value, the LD decay distance was estimated to be 86 kb.

Genome-wide association analysis was performed using the mrMLM.GUI package (Zhang et al., 2020), and the six methods (mrMLM (Wang et al., 2016), FASTmrMLM (Tamba and Zhang, 2018), FASTmrEMMA (Wen et al., 2019), pLARmEB (Zhang et al., 2017), ISIS EM-BLASSO (Tamba et al., 2017), and pKWmEB (Ren et al., 2018) were used to detect significant QTN. In the first stage, the critical $p$-value parameter was set to 0.005 for all methods except FASTmrEMMA, and the critical LOD value for significant QTN was set to 3 in the final stage. The kinship matrix used in the analysis was also calculated by the software itself.

**Candidate Gene Prediction**

Genomic regions repeatedly identified in multiple environments or two populations were used to predict genes involving in PH formation. Specifically, the genome region of QTL interval localized in multiple environments with a genomic region less than 300 kb and the LD decay distance of 86 kb of the QTN localized within the QTL genomic region were selected, and the genes were searched for by the Phytozome website. The genes expressed in the stems were then screened. Finally, candidate genes related with PH were identified by combining annotation information of genes, pathway analysis in the Kyoto Encyclopedia of Genes and Genomes (KEGG) and previous studies.

**Candidate Gene Validation**

Two parents (Dongnong L13 and Henong 60), two varieties (HN400 and HN451) with lower PH and two varieties (HN369 and HN477) with higher PH, were selected in the RIL6013 population based on the PH phenotype data. The qRT-PCR was used to study the relative expression of candidate genes in these six varieties. These varieties were planted in Harbin in the same environment as E1. Stems were sampled at 10-day intervals starting from the R1 period when elongation is the fastest. The third node down from the top of the main stem was taken and replicated three times per plant. Total RNA was extracted using the OminiPlant RNA Kit (Dnase I) (CWBIO, Jiangsu, China). Two microgram of total RNA was extracted using the EasyScript® One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen Biotech, Beijing, China). The first strand cDNA was synthesized from 2 µg of total RNA using the EasyScript® One-Step gDNA Removal kit.

**TABLE 4** | Five QTL detected in multiple environments.

| QTL   | Env.     | Chr. | Marker Interval | LOD$^a$ | PVE (%)$^b$ | ADD$^c$ | Physical Region (Mb) | Method       |
|-------|----------|------|-----------------|---------|-------------|---------|-----------------------|--------------|
| qPH-1-1 | E3/E5/E8/E8 | 1 | 60c01056-60c01052 | 3.34/4.28/3.62/3.62 | 1.67/3.19/2.68/11.72 | 16.70/41.16/31.10/31.10 | 48.83–50.55 | IM/IM/IM/ICIM |
| qPH-2-1 | E3/E4/E8 | 2 | 60c02058-60c02059 | 2.87/3.09/2.96 | 1.60/6.46/3.53 | 15.53/18.93/27.09 | 13.66–13.98 | IM/IM/IM |
| qPH-3-5 | E5/E6/E8/E8 | 3 | 60c03076-60c03080 | 2.63/2.73/3.51/2.73 | 4.72/3.92/2.19/3.92 | 21.31/25.74/32.81/25.74 | 42.10–42.72 | IM/IM/IM/ICIM |
| qPH-12-1 | E2/E3/E2 | 12 | 60c12024-60c12032 | 2.68/2.90/2.68 | 5.60/1.47/5.80 | 15.77/15.54/15.77 | 11.79–14.97 | IM/IM/IM/ICIM |
| qPH-18-3 | E5/E8 | 18 | 60c18061-60c18058 | 4.37/3.48 | 3.25/0.86 | 24.58/40.77 | 49.04–52.21 | IM/IM |

$^a$LOD, logarithm of odds. $^b$PVE, phenotypic variation explained by QTL. $^c$ADD, additive effect.

---

2https://phytozome.jgi.doe.gov

3https://www.kegg.jp
and cDNA Synthesis SuperMix kit (TransGen Biotech, Beijing, China). Twenty microliter reaction volume was determined for qRT-PCR using the SYBR® Green doping method from Roche Light Cycle™ containing the following components: 10 µL SYBR® Green Realtime PCR Master Mix (TOYOBO, Japan), 0.8 µL of each primer (10 µM), 6.4 µL of distilled water and 2 µL of diluted cDNA. The whole reaction was run under the following conditions: pre-denaturation 95°C for 30 s; PCR40 cycles, 95°C for 5 s, 60°C for 20 s, 72°C for 15 s; solubility curve analysis 95°C 10 s, 65°C for 60 s, and 97°C for 1 s. All PCR reactions were repeated three times. Data were processed using the $2^{-\Delta\Delta Ct}$ method using FBOX as the internal reference gene (Bansal et al., 2015), the primers used are shown in Supplementary Table 2.

**Molecular Marker Identification**

With the objective to verify the effect of gene and develop markers for molecular assistant selection, the markers with polymorphism in the 100k bp interval of the genes were evaluated for the association with plant height in the 455-germplasm population. The significant differences of averages between allelic genotypes were determined by analysis of variance, and the probability to determine the significance was set 0.05.

**RESULTS**

**Phenotypic Variation Analysis**

Phenotypic data collected from 139 lines of RIL6013 in eight environments were analyzed. 455 germplasm resource populations in four environments were analyzed early by Wang et al. (2021). The results of descriptive statistics (Table 1) showed that the absolute values of kurtosis and skewness were less than 1 in all the eight environments of RIL6013 except E8, which was close to 1. It showed that PH distributed normally (Figure 1). The range of PH in RIL6013 contained those of parents, which indicated a transgressive segregation in the two populations. The coefficient of variation ranged from 13.10 to 22.00% for the RIL6013 population and from 18.03 to 20.31% for the 455-germplasm population, which suggested that a wide range of
variation in plant height in two populations and a different genetic basis in different environments.

The results of ANOVA (Table 2) showed that there were highly significant differences in environment, genotype, and genotype × environment interaction effect, which indicated that PH was influenced not only by genotype and environment but also by genotype by environment interaction effect. Higher broad sense heritability (65 and 72%) was found in RIL6013 and 455 germplasm resource populations, respectively, which indicated that the variation of soybean plant height mainly come from genetic effect.

**Bin Map and QTL Localization for RIL6013**

A high-density SNP bin genetic linkage map covered all 20 chromosomes containing 1996 bin markers, and the total length of the map was 2874.72 cm. The number of SNP bin markers per chromosome ranged from 59 to 158, and the length of each linkage group ranged from 82.37 to 238.98 cm. The average number of markers per linkage group was 99.8, and the average distance between markers was 1.48 cm (Figure 2 and Table 3).

A total of 33 QTLs associated with plant height were localized in the RIL6013 population on 12 chromosomes of soybean using two methods IM and ICIM based on bin mapping (Figure 3 and Supplementary Table 3). The number of QTL localized on each chromosome ranged from one (Chr02, Chr04, Chr12, and Chr13) to six (Chr14), with phenotypic contributions ratio ranging from 0.55 to 13.64%. 2, 16, 6, 9, 1, 1, and 6 QTLs were localized in E2–E8, respectively. A total of three QTL (qPH-1-1, qPH-6-2, and qPH-18-4) showed phenotypic contributions ratio more than 10% and can be considered as the main effective QTL for plant height.

A total of five QTLs were localized in multiple environments (Table 4), and the additive effects were all positive, indicating that the parent Dongnong L13 could increase plant height via these QTL. qPH-2-1 was localized on chr02 in E3, E4, and E8 environments with LOD values of 2.87–3.09 and phenotypic contributions ratio of 1.60–6.46%. The genomic region of qPH-2-1 was shorter than 320 kb, which is suitable for searching candidate genes.

**Multi-Locus GWAS for Germplasm**

A total of 62 QTN were detected on 18 chromosomes (except for chr11 and chr20) using six multilocus methods within the mrMLM package: mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, ISIS EM-BLASSO, and pKWmEB, respectively. LOD values ranged from 3.02 to 10.45, and the ratio of phenotypic variation explained by QTN ranged from 1.12 to 13.12%. Six methods detected 20, 10, 3, 18, 25, and 29 QTN, respectively, while 15, 23, 13, and 11 QTN were detected within E1, E2, E3, and E4, respectively.

Of all QTN, 26 detected by multiple methods were located on chromosomes 1, 2, 3, 4, 6, 7, 9, 10, 13, 14, 15, 16, and 18, respectively, with LOD values ranging from 3.04 to 10.45. The proportion of phenotypic variation explained by QTN ranged from 1.12 to 6.62%. The detected QTN effects (positive or negative) were consistent between methods (Wang et al., 2021).

**Co-detected Regions by Linkage Analysis and Association Analysis**

The regions detected by GWAS were compared with those of the linkage analysis. The results showed that two QTN loci fell within the genomic region where the two QTLs identified in the RIL6013 population (Figure 4). Among them, AX-90484715 was located within the interval of qPH-5-4 and AX-90349538 was located within the interval of qPH-14-2. Candidate genes were searched within 43 kb flanking these two QTN loci based on LD distance.
Candidate Gene Prediction
Based on the above results, candidate genes were selected to search within 13.66-13.98 Mb on Chr02, 41.55-41.63 Mb on Chr05 and 4.01-4.09 Mb on chr14. A total of 50 candidate genes were searched, of which 46 genes were expressed in the stems. The pathway analysis on 46 genes showed that a total of 18 genes (39.1%) had annotations (Supplementary Table 4).

Candidate Gene Validation
The relative expression of the three candidate genes in the two parents, HN400, HN451, HN369, and HN477, were characterized by applying qRT-PCR. The plant height of the six varieties continued to grow from R1 to day 30, with highly significant differences in plant height from day 10 after R1 (Supplementary Table 5 and Figure 5A). Relative expression amount (REA) of Glyma.02G132200 did not differ significantly among varieties at the whole stages, which indicated Glyma.02G132200 was not directly related to PH. It could be related to PH (Figure 5B). The expression of Glyma.05G240600 in the six cultivars continued to increase from R1 to day 30. REA of Glyma.02G133000, REA of the six varieties increased continuously from R1 to day 20 and started to decrease from day 20 to day 30. REA from day 10 to day 30 of HN369, HN477 and Dongnong L13 was larger significant than that of HN400, HN451 and Henong 60 (Figure 5B). The expression of Glyma.02G133000.1, Glyma.05G240600.1 and Glyma.02G133000 (Table 6), which indicated that the two genes controlling plant height. Among these markers, AX-90483488, AX-90490846, and AX-90515514 were detected in three environments, while the rest five markers were detected in only one environment. These eight markers could be used improve plant height commonly or specifically.

DISCUSSION
Improving the Accuracy of QTL Analysis and GWAS by Multi-Environment Experiments and Sufficient SNP Markers
The small amount of RFLP, AFLP, and SSR markers used in previous studies made it difficult to ensure the accuracy of linkage analysis (Singh et al., 2016; Bhat et al., 2020), and most of the previously localized QTL were analyzed in a single environment, which is prone to false positive results (Fang et al., 2020). QTL detected repeatedly in multiple environments are more authentic than those detected in a single environment (Fulton et al., 1997).

Comparison With Previous Results of Localized QTL
Here, the five QTLs located by RIL6013 repeatedly in multiple environments and the two QTNs identified by a combination of linkage and association analysis were compared with 238 QTLs associated with PH located by previous researches in the soybase.

### TABLE 6 | SNPs markers associated with plant height near Glyma.02G133000 and Glyma.05G240600.1.

| Genes        | SNP probe       | Physical Position | Genotype | 18H³ | 19H | 19S |
|--------------|-----------------|-------------------|----------|------|-----|-----|
| Glyma.02G133000.1 | AX-90394810     | 13746369          | C/T      | 9.11 | 0.0027 |     |
| Glyma.05G240600.1 | AX-90351367     | 13765868          | A/G      | 4.95 | 0.0267 |     |
| Glyma.02G133000.1 | AX-90460406     | 13766378          | A/G      |      |      | 4.12 | 0.043 |
| Glyma.02G133000.1 | AX-90483488     | 13787209          | G/C      | 14.5 | 0.0002 | 10.19 | 0.0015 |
| Glyma.05G240600.1 | AX-90515514     | 41489472          | C/T      | 6.04 | 0.0144 | 7.1  | 0.008 |

³18H represent the 455-germplasm population was planted in Harbin in 2018, 19H represent Harbin in 2019, 19S represent Shuangyashan in 2019.
and the specific regulation of plant height by these two candidate genes needs to be investigated in follow-up studies.

SUMMARY

Here, five multi-environmental QTL and 26 multi-method QTN were detected by linkage analysis and association analysis, respectively, and two candidate genes associated with plant height were identified by pathway analysis and qRT-PCR validation. These results lay the foundation for marker-assisted selection.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

W-XL and HN conceived and designed the experiments. JW, BH, YJ, XH, YG, JC, YL, and JH performed the field experiments. JW and HN analyzed and interpreted the results. JW, BH, and HN drafted the manuscript. W-XL provided the laboratory conditions. All authors contributed to the manuscript revision.

FUNDING

This research was supported by the National Natural Science Foundation of China (32072015) to W-XL and Hundred-thousand and Million Project of Heilongjiang Province for Engineering and Technology Science/Soybean Breeding Technology Innovation and New Cultivar Breeding (2019ZX16B01) to HN.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.803820/full#supplementary-material

REFERENCES

Adewusi, O., Odiiyi, A., and Akinyele, B. (2017). Identification of genomic region governing yield related characters in soybean, glycin max (L.) merrill using SNP Markers. J. Adv. Biol. Biotechnol. 2017, 1–9. doi: 10.9734/JABB/2017/36290

Assefa, T., Otyama, P. I., Brown, A. V., Kalberer, S. R., Kulkarni, R. S., and Cannon, S. B. (2016). Genomics-assisted characterization of a breeding collection of Apios americana, an edible tuberous legume. Sci. Rep. 6, 1–17. doi: 10.1038/srep34908

Bhat, J. A., Deshmukh, R., Zhao, T., Patil, G., Deokar, A., Shinde, S., et al. (2020). Harnessing high-throughput phenotyping and genotyping for enhanced drought tolerance in crop plants. J. Biotechnol. 2020:10. doi: 10.1016/j.jbiotec.2020.11.010

Cao, Y., Li, S., Chen, G., Wang, Y., Bhat, J. A., Karikari, B., et al. (2019). Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants 8:373. doi: 10.3390/plants8100373

Chang, F., Guo, C., Sun, F., Zhang, J., Wang, Z., Kong, J., 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. doi: 10.3389/fpls.2018.01184
Chen, L., Nan, H., Kong, L., Yue, L., Yang, H., Zhao, Q., et al. (2020). Soybean AP1 homologs control flowering time and plant height. J. Integr. Plant Biol. 62, 1806–1879. doi: 10.1111/jipb.129888
Dai, C., Lee, Y., Lee, I.-C., Nam, H. G., and Kwak, J. M. (2018). Calmodulin 1 regulates senescence and ABA response in Arabidopsis. Front. Plant Sci. 9:803. doi: 10.3389/fpls.2018.00803
Fang, Y., Liu, S., Dong, Q., Zhang, K., Tian, Z., Li, X., et al. (2020). Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. Front. Plant Sci. 11:99. doi: 10.3389/fpls.2020.00009
Feng, Y.-Y., He, J., Jin, Y., and Li, F.-M. (2021). High phosphorus acquisition and allocation strategy is associated with soybean seed yield under water-and P-limited conditions. Agronomy 11:574. doi: 10.3390/agronomy11030574
Fernandez, M. G. S., Becraft, P. W., Yin, Y., and Lübberstedt, T. (2009). From TIR-NB-LEU repeats to TIR-NB-LRRs: exploration of a major disease resistance gene family in Arabidopsis. Theor. Appl. Genet. 118, 677–689. doi: 10.1007/s00122-008-0987-2
Kim, K.-S., Diers, B., Hyten, D., Mian, M. R., Shannon, J., and Nelson, R. (2012). Transcriptome analysis of an advanced backcross of Lycopersicon peruvianum to L. esculentum and comparisons with QTLs found in other wild species. Front. Plant Sci. 3:246. doi: 10.3389/fpls.2012.00246
Kim, S. A., Punshon, T., Lanzirotti, A., Li, L., Alonso, J. M., Ecker, J. R., et al. (2006). Genetic mapping and confirmation of quantitative trait loci for flowering time, branch initiation height, stem diameter, and flowering time in an alien introgression derived Brassica napus DH population. Front. Plant Sci. 9:390. doi: 10.3389/fpls.2018.00390
Lee, S., Bailey, M., Mian, M., Shipe, E., Ashley, D., Parrott, W., et al. (1996). Linkage and association mapping-by-sequencing and linkage analysis. Theor. Appl. Genet. 92, 516–523. doi: 10.1007/BF00224553
Lee, S., Jun, T., Michel, A. P., and Mian, M. R. (2015). SNP markers linked to QTLDI conditioning plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. 92, 516–523. doi: 10.1007/BF00224553
Li, A., Huang, G., and Ma, C. (2020). Detection of QTL and QTN and candidate genes for oil content in soybean using a combination of four-way-RIL and genomewide populations. Crop J. 8, 802–811. doi: 10.1016/cj.2020.07.004
Li, Z. F., Guo, Y., Ou, L., Hong, H., Wang, J., Liu, Z. X., et al. (2018). Identification of the dwarf gene GmDW1 in soybean (Glycine max L.) by combining mapping-by-sequencing and linkage analysis. Theor. Appl. Genet. 131, 1001–1016. doi: 10.1007/s00122-017-3044-8
Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010). The soybean stem growth habit gene D1 is an ortholog of Arabidopsis TERMINAL FLOWER1. Plant Physiol. 153, 198–210. doi: 10.1104/pp.109.150607
Li, H.-Y., Li, H.-W., Fan, R., Li, H.-Y., Yin, J.-Y., Zhang, J.-L., et al. (2016). Genome-wide association study of dynamic developmental plant height in soybean. Can. J. Plant Sci. 97, 308–315. doi: 10.1139/cjps-2016-0152
McCouch, S. (1997). Report on QTL nomenclature. Rice Genet. News. 11, 14–13.
Pathan, S., Vuong, T., Clark, K., Lee, J., Shannon, J., Roberts, C., et al. (2013). Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 53, 765–774. doi: 10.2135/cropsci2013.03.0153
Qi, Z., Song, J., Zhang, K., Liu, S., Tian, X., Wang, Y., et al. (2020). Identification of QTNs controlling 100-seed weight in soybean using multilocus genome-wide association studies. Front. Genet. 11:689. doi: 10.3389/fgene.2020.00689
Ren, W.-L., Wen, Y.-J., Dunwell, J. M., and Zhang, Y.-M. (2018). pKwMeBiE: integration of Kruskal–Walls test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. Heredity 120, 208–218. doi: 10.1038/s41437-017-0007-4
Salarí, M. W., Ongom, P. O., Thapa, R., Nguyen, H. T., Vuong, T. D., and Rainey, K. M. (2021). Mapping QTTL controlling soybean seed sucrose and oligosaccharides in a single family of soybean nested association mapping (SoyNAM) population. Plant Breed. 140, 110–122. doi: 10.1111/pbr.12883
Shen, Y., Xiang, Y., Xu, E., Ge, X., and Li, Z. (2018). Major co-localized QTL for plant height, branch initiation height, stem diameter, and flowering time in an alien introgression derived Brassica napus DH population. Front. Plant Sci. 9:390. doi: 10.3389/fpls.2018.00390
Song, J., Sun, X., Zhang, K., Liu, S., Wang, J., Yang, C., et al. (2020). Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. Mol. Breed. 40, 1–14. doi: 10.1007/s11032-020-01404-w
Sun, D., Li, W., Zhang, Z., Chen, Q., Ning, H., Qiu, L., et al. (2006). Quantitative trait loci analysis for the developmental behavior of soybean (Glycininmax L. Merr.). Theor. Appl. Genet. 112, 665–673. doi: 10.1007/s00122-005-0169-y
Tamba, C. L., Ni, Y.-L., and Zhang, Y.-M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput. Biol. 13:e1005357. doi: 10.1371/journal.pcbi.1005357
Tamba, C. L., and Zhang, Y.-M. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. BioRxiv 2018:341784. doi: 10.1101/341784
Tian, X., Zhang, K., Liu, S., Sun, X., Li, X., Song, J., et al. (2020). Quantitative trait loci analysis of protein and oil content in response to planting density in chickpea (Cicer arietinum L.). Plant Biotechnol. J. 14, 2110–2119. doi: 10.1111/pbi.13567
Wang, J. (2009). Inclusive composite interval mapping of quantitative trait genes. Acta Agronomica Sinica 35, 239–245. doi: 10.1007/s11032-007-0663-5
Wang, J., Hu, B., Huang, S., Hu, X., Siyal, M., Yang, C., et al. (2021). SNP-bin linkage analysis and genome-wide association study of plant height in soybean. *Crop. Pasture. Sci.*

Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., et al. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep19444

Wen, Y.-J., Zhang, Y.-W., Zhang, J., Feng, J.-Y., Dunwell, J. M., and Zhang, Y.-M. (2019). An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F2. *Brief. Bioinform.* 20, 1913–1924. doi: 10.1093/bib/bby058

Xu, P., Guo, Q., Meng, S., Zhang, X., Xu, Z., Guo, W., et al. (2021). Genome-wide association analysis reveals genetic variations and candidate genes associated with salt tolerance related traits in *Gossypium hirsutum*. *BMC Genomics* 22:66236. doi: 10.21203/rs.3.rs-66236/v4

Xu, Y., Li, P., Yang, Z., and Xu, C. (2017). Genetic mapping of quantitative trait loci in crops. *Crop J.* 5, 175–184. doi: 10.1016/j.cj.2016.06.003

Yao, D., Liu, Z., Zhang, J., Liu, S., Qu, J., Guan, S., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14, 6101–6109. doi: 10.4238/2015.June.8.8

Yue, L., Li, X., Fang, C., Shen, L., Yang, H., Yang, J., et al. (2021). FT5a interferes with the Dt1-AP1 feedback loop to control flowering time and shoot determinacy in soybean. *J. Integr. Plant Biol.* 63, 1004–1020. doi: 10.1111/jipb.13070

Zhang, J., Feng, J.-Y., Ni, Y., Wen, Y., Niu, Y., Tamba, C., et al. (2017). pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity* 118, 517–524. doi: 10.1038/hdy.2017.8

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:1–11. doi: 10.1186/s12864-015-1441-4

Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., et al. (2019). A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* 20:5915. doi: 10.3390/ijms20235915

Zhang, Y.-W., Tamba, C. L., Wen, Y.-J., Li, P., Ren, W.-L., Ni, Y.-L., et al. (2020). mrMLM v4. 0.2: an r platform for multi-locus genome-wide association studies. *Genom. Proteom. Bioinf.* 18, 481–487. doi: 10.1016/j.gpb.2020.06.006

Zhao, X., Cao, D., Huang, Z., Wang, J., Lu, S., Xu, Y., et al. (2015). Dual functions of GmTOE4a in the regulation of photoperiod-mediated flowering and plant morphology in soybean. *Plant Mol. Biol.* 88, 343–355. doi: 10.1007/s11103-015-0322-1

**Conflict of Interest:** XH is employed by Key Laboratory of Crop Biotechnology Breeding of the Ministry of Agriculture, Beidahuang Kenfeng Seed Co., Ltd., Harbin, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

**Copyright © 2022 Wang, Hu, Jing, Hu, Guo, Chen, Liu, Hao, Li and Ning. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**