ORIGINAL ARTICLE

Fine-mapping of a major QTL controlling plant height by BSA-seq and transcriptome sequencing in cotton

Chao Li[1,2] · Longyu Huang[1,3,4] · Yiwen Huang[1] · Meng Kuang[1] · Yuzhen Wu[1] · Zhiying Ma[2] · Xiaoqiong Fu[1]

Abstract

Key message *GhSOT (GH_D05G3950)* plays a negative role in regulating plant height development by modulating the GA signaling.

Abstract Plant height is an important indicator affecting mechanical harvesting for cotton. Therefore, understanding the genes associated with the plant height is crucial for cotton breeding and production. In this study, we used bulk segregant analysis sequencing to identify a new quantitative trait locu (QTL) called *qPH5.1*, which is linked to plant height. Local QTL mapping using seven kompetitive allele-specific PCR (KASP) markers and linkage analysis successfully narrowed down *qPH5.1* to ~ 0.34 Mb region harbored five candidate genes. Subsequently, RNA sequencing (RNA-seq) analysis and examination of expression patterns revealed that *GhSOT* exhibited the highest likelihood of being the candidate gene responsible for the plant height at this locus. Seven SNP site variations were identified in the *GhSOT* promoter between the two parents, and Luciferase experiments confirmed that the promoter of *GhSOT* from *cz3* enhances downstream gene expression more effectively. Additionally, suppression of *GhSOT* in *cz3* resulted in the restoration of plant height, further emphasizing the functional significance of this gene. Application of exogenous gibberellin acid (GA) significantly restored plant height in *cz3*, as demonstrated by RNA-seq analysis and exogenous hormone treatment, which revealed alterations in genes associated with GA signaling pathways. These results reveal *GhSOT* is a key gene controlling plant height, which may affect plant height by regulating GA signaling.

Introduction

Cotton is the most important source of natural textile fiber in the world. Upland cotton (*Gossypium hirsutum* L.) has high lint yield, accounting for more than 95% of world cotton production (Li et al. 2016; Hu et al. 2019). Given the evolving trade dynamics in the cotton industry and the push for mechanization, the cultivation of high-yield and high-quality cotton varieties becomes paramount. The plant architecture plays a vital role in determining cotton yield and the effectiveness of mechanical harvesting (Dong et al. 2018). Plant architecture includes various agronomic traits, such as plant height, number of fruit branches, distribution of fruit branch, as well as bolls and leaves.

Plant height is a crucial crop trait that determines overall architecture impacting light interception, photosynthesis, and harvesting in various crop species (Teng et al. 2013). Breeders typically prefer plants that approach an optimal height value, as excessively short plants tend to exhibit lower yields compared to semi-dwarf varieties. The identification and characterization of plant height-related

---

Chao Li and Longyu Huang have contributed equally to this work.

✉ Zhiying Ma
  mzhy@hebau.edu.cn

✉ Xiaoqiong Fu
  m13663726262@163.com

1 State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China

2 State Key Laboratory of North China Crop Improvement and Regulation, Key Laboratory for Crop Germplasm Resources of Hebei, Hebei Agricultural University, Baoding 071000, China

3 National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya 572024, China

4 Hainan Seed Industry Laboratory, Sanya 572025, China

genes are increasing, and several genes have been successfully cloned. Notably, recent research has demonstrated that the down-regulation of *GhPAS1* plays an important role in modulating architectural traits by regulating plant height and fruit branch angle (Wu et al. 2021). The over-expression of *GhDREB1B* has been found to result in a notable decrease in plant height, branch length, and branch angle. Regulating the expression of *GhDREB1B* offers a potential means to optimize plant architecture for dense planting (Ji et al. 2021). Despite the insights provided by previous studies, the genetic and molecular mechanisms underlying cotton plant height is not clear.

Presently, the regulation of cotton plant height is achieved through chemical means, thereby incurring supplementary expenses and engendering the risk of chemical contamination. Alternatively, given the escalating demand for mechanical harvesting in China, the cultivation of varieties with ideal plant height emerges as the most dependable and viable approach. However, plant height inherits as a quantitative trait governed by multiple genes and is readily influenced by environmental factors. Numerous QTLs have been reported for cotton plant height through the utilization of diverse genetic populations. For instance, Wu et al. (2022) employed QTL mapping in a recombinant inbred line (RIL) population across nine environments and three best linear unbiased predictions (BLUPs) to identify 60 plant height QTLs. Ma et al. (2019) utilized SLAF-seq technology to identify a stable QTL, qPH-Dt1-1, and a gene, *GhPIN3*, that regulates cotton plant height in an interspecific *G.hirsutum × G.barbadense* backcross inbred line (BIL) population. Su et al. (2018) identified eight SNPs associated with plant height, and observed that the silenced plants (*Gh_D03G0922*) exhibited increased plant height. Additionally, qRT-PCR analysis revealed that the expression of *Gh_D03G0922* was upregulated in the apical buds and young leaves of short and compact cotton varieties. Additionally, Ye et al. (2023) conducted genome-wide association studies in three field conditions using unmanned aerial vehicles and identified two genetic loci *GhUBP15* and *GhCUL1* associated with plant height. However, the identification of candidate genes associated with plant height within these QTL regions is still limited.

To provide additional information for breeding, the genetic basis of plant height must be further investigated in cotton. In this study, we mapped the plant height QTL of mutant cotton genotype *cz3* precisely by BSA-seq and RNA-seq to locate the candidate gene region on the chromosome D05. *GhSOT* may plays an important role in regulating plant height development. This work will provide a foundation to verify the plant height function of candidate genes and ultimately realize the use of ideal plant height resource.

Materials and methods

## Plant materials and field experiments

A dwarf cotton mutant named *cz3* was identified from EMS-mutagenized seeds of cotton germplasm resources. *cz3* and CIR12 were used as female and male parents, respectively, to generate an F_2 population of 600 individuals. *cz3* ($n = 48$), CIR12 ($n = 48$), and F_2 ($n = 600$) individuals were planted in the field at Anyang, China in 2021. Each row of field plots was 6 m long and 80 cm wide, with the seeds placed 20 cm apart within the row. The sowing depth was about 8 cm. Plants from the center of each plot were selected for plant height evaluation.

DNA extraction and construction of BSA library

Genomic DNA was extracted from parental and F_2 plants leaves with the cetyltrimethylammonium bromide (CTAB) method and followed by purification through RNase A treatment. Two segregating pools, namely the dwarf and tall bulk, were created by combining equal amounts of DNA from 30 dwarf lines and 30 tall lines. The quality and concentration of the DNA were assessed using a Beckman Coulter DU800 spectrophotometer, and necessary adjustments were made to achieve a final DNA concentration of 100 ng mL^{-1}. Approximately 5 μg of DNA from the two segregating pools and two parental lines were utilized for the construction of sequencing libraries.

RNA-seq

Total RNA was extracted from stem tissues of *cz3* and CIR12 during the four-leaf development period, and RNA libraries were prepared using Illumina 2500 platform sequencing technology. The sequencing was performed by Beijing Genomics Institute (BGI) in Shenzhen, China. For the RNA-seq experiment, three biological replicates were performed to ensure the reliability of the results. The TM-1 (AD1) genome NAU-NBI assembly v1.1 was used as the reference genome for mapping and analysis (Nanjing Agricultural University-Novogene Bioinformatics Technology). Gene expression levels were calculated and normalized into fragments per kilobase per million mapped fragments (FPKM) values using the stringtie software (Zhao et al. 2021). To identify differentially expressed genes (DEGs), the deseq2 package was used with the following parameters: a fold change of more than two and an adjusted P-value less than 0.01. Gene ontology enrichment was analyzed by David (https://david.ncifcrf.gov). KEGG pathway enrichment was
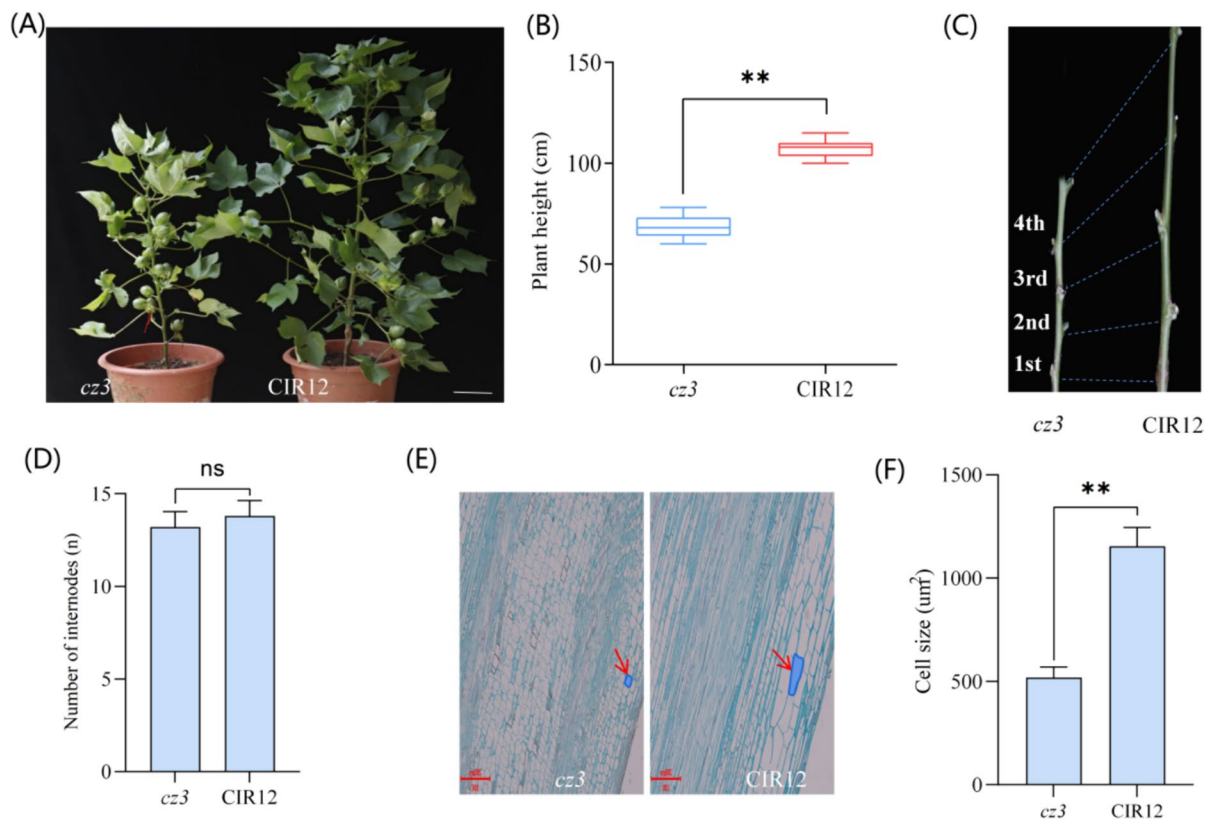
Fig. 1 Plant height and longitudinal sections of the stem at the mature stage. **A**, **B** Plant height of *cz3* and CIR12 at the mature stage. Scale bar = 10 cm. **C** The spacing between four fruit branches on the stem of *cz3* and CIR12. **D** Internode numbers on the stem of *cz3* and CIR12, 'ns' indicates no significant difference. **E** Hand-cut longi-

tudinal sections of the stem in *cz3* and CIR12, respectively. Scale bar = 200 μm. **F** The internode cell size of *cz3* and CIR12. Six slides of each sample were observed, and more than 100 cells were measured for each slide. **P < 0.01, Student's *t* test was used to calculate the *P* value

examined via the Novogene platform (https://magic.novogene.com).

LUC imaging assay

The coding sequences of pro*GhSOT*$_{cz3}$ and pro*GhSOT*$_{CIR12}$ were PCR-amplified using specific primer sets listed in Table S1. pro*GhSOT*$_{cz3}$ and pro*GhSOT*$_{CIR12}$ were cloned in-frame with the p0800MINI vector. These constructs were transformed into *A. tumefaciens* strain GV3101 and subsequently infiltrated into *N. benthamiana* leaves. After infiltration, the plants were maintained under a 16 h light/8 h dark cycle for 2 days. Firefly LUC and Renilla luciferase (REN) activities were quantified using a dual-luciferase reporter assay system (Promega, Madison, WI, USA).

Gene cloning and plant transformation

For virus-induced gene silencing (VIGS) experiment, about 100–200 bp DNA fragments of candidate gene were each into pTRV2 vectors that was transformed into *A. tumefaciens*

strains GV3101 to generate pTRV::candidate gene in *cz3*. The GV3101 cells with each of the above plasmids were mixed with an equal volume of the cells harboring pTRV1 and infiltrated into 10-day-old cotton seedling cotyledons as previously described (Chen et al. 2021; Burch-Smith et al. 2004; Pang et al. 2013). The VIGS experiments were repeated three times, and each time with more than 10 plants. TRV::PDS was used as the positive control and the empty vector pTRV2 was used as the negative control. All the plants grew in the greenhouse that were facilitated with 28 °C/22 °C day/night temperature, 16 h/8 h light/dark photoperiod.

Scanning electron microscopy (SEM)

SEM was employed to examine the stems of both *cz3* and CIR12 cotton plants, adapting the method outlined with slight modifications to suit the specific requirements of the samples being analyzed (Challa et al. 2021).
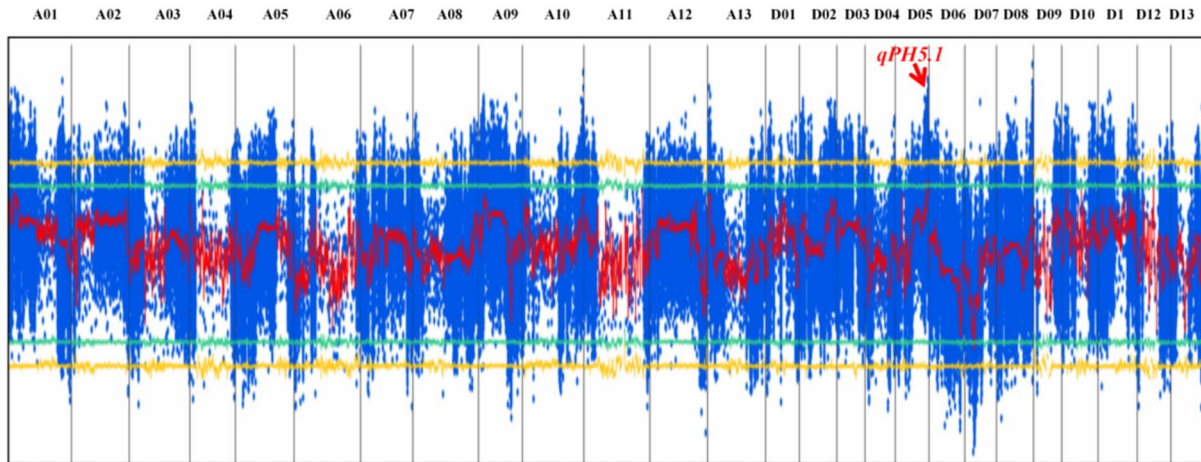
Fig. 2 QTL-seq approach for mapping genomic regions controlling cotton plant height. The colored dots represent the Δ(SNP index) value of every SNP locus. The red lines show the Δ(SNP index) value of the fitting results, the green dotted line shows the association threshold value (0.95) and the yellow dotted line shows the association threshold value (0.99)

Subcellular localization of *GhSOT*

The full-length coding sequence (CDS) of *GhSOT,* excluding the stop codon, was amplified from CIR12 using specific primers (Table S1) and cloned into the pCAMBIA2300-GFP vector to construct a 35S::*GhSOT*-GFP fusion protein for transient expression. Along with the control plasmid (35S::GFP), the recombinant plasmid was transferred into *A. tumefaciens* strain GV3101. The subcellular localization of *GhSOT* in living cells was analyzed in the leaves of 4-week-old *N. benthamiana* plants infiltrated with *A. tumefaciens* cells carrying the *GhSOT* construct. After 48 h, the epidermal cells of the infiltrated tobacco leaves were examined under an OLYMPUS FV1200 confocal microscope (Olympus, Tokyo, Japan) to determine the localization of the *GhSOT*-GFP fusion protein. The primers used for subcellular localization are listed in Table S1.

Real-time quantitative RT-PCR

Total RNA was extracted from cotton samples using a plant total RNA extraction kit (Tiangen Biotech, Beijing, China) following the instructions provided with the kit. Subsequently, first-strand cDNA was synthesized using the PrimeScript Rt reagent Kit with gDND Eraser reagent (Takara Biotech, Beijing, China) for the purpose of analyzing the expression levels of relevant genes through qRT-PCR. The qRT-PCR was conducted using the MonAmp in review SYBR Green qPCR Mix Kit (Monad Biotech, Suzhou, China). The relative expression level was calculated according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). In order to standardize the gene expression, an internal control *GhUBQ-7* was used.

Statistical analysis

Differences between parent and progeny plant height were detected using SPSS18.0 (IBM, Armonk, NY, USA). Data represent means ± standard deviations. Graphs were drawn using edgeR (http://www.r-project.org/) and Origin 2018 (OriginLab, Northampton, MA, USA).

Results and analysis

Phenotypic identification and genetic analysis of plant height

cz3 displayed a dwarf plant phenotype with height of ~ 65 cm, significantly lower than CIR12 (Fig. 1A, B). However, *cz3* and CIR12 have the same number of internodes, indicating that the observed plant height difference between the two lines is caused by differences in the length of the internodes, rather than the number of internodes (Fig. 1C, D). Cytological observation revealed that internode cell sizes in CIR12 were significantly larger than those of *cz3* ($P < 0.01$) (Fig. 1E, F), indicating that the longer internodes in CIR12 relative to *cz3* are caused by increased cell sizes, not increased cell numbers.

BSA-seq analysis

In order to facilitate phenotypic scoring, we opted to utilize plant height as the quantitative trait for the initial identification of the plant height locus. To achieve this, a pooled sequencing strategy was employed, involving two parent-bulk samples, *cz3* (P1-pool) and CIR12 (P2-pool). From these samples, we selected 30 extreme-dwarf plants
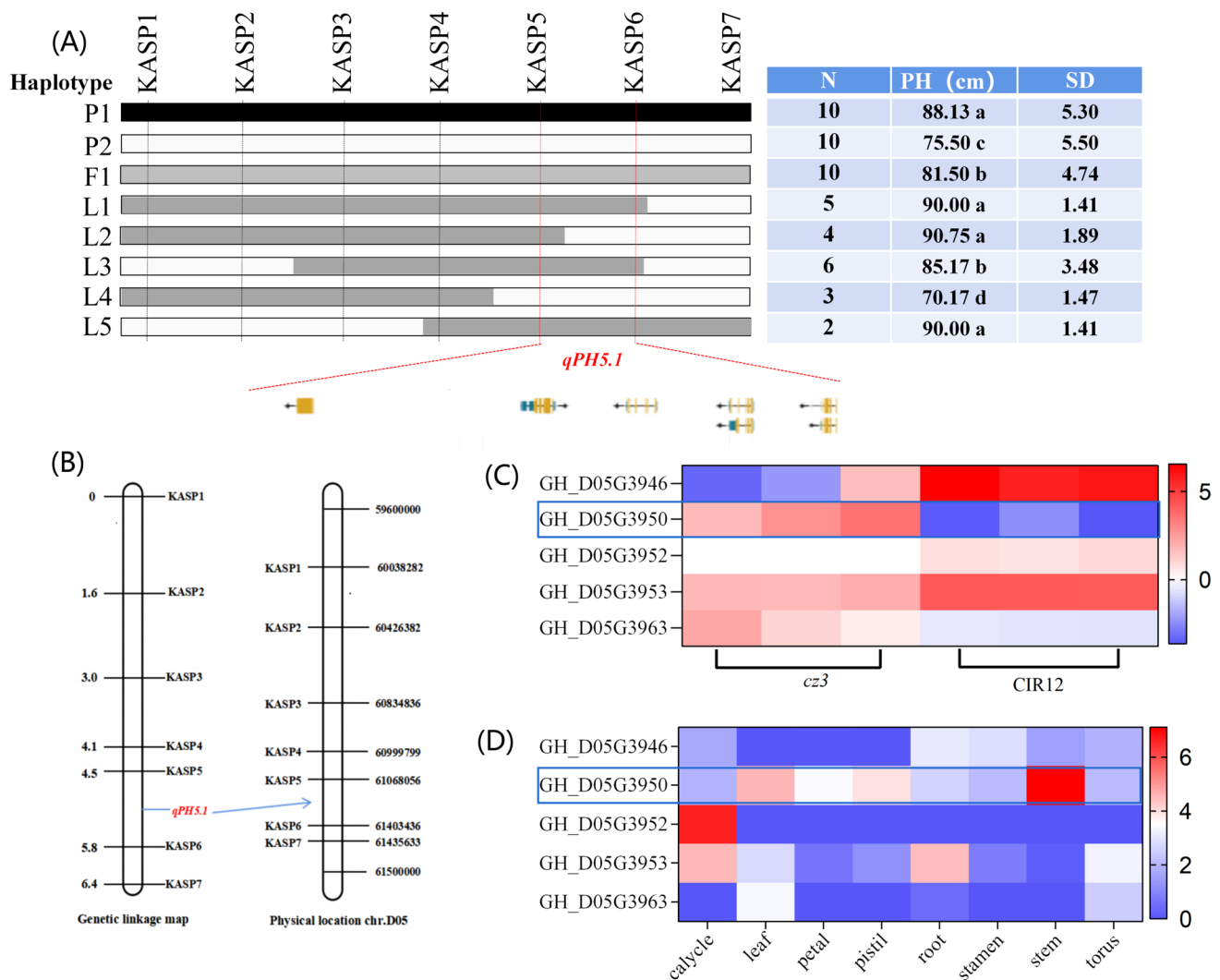
Fig. 3 Fine-mapping and analysis of candidate genes. **A** Fine-mapping of the *qPH5.1* locus. Genotyping of recombinant plants from 1800 F$_2$ individuals. Annotation of the ~0.34 Mb region. Letters represent significant differences at the 0.05 level based on Tukey's test. Yellow boxes indicate exons, and black lines indicate introns. Arrow indicates the orientation of the gene. **B** Mapping interval for *qPH5.1* locus and its corresponding physical map of chromosome D05 on cotton. **C** Heat map of gene expressions in the candidate gene region between *cz3* and CIR12 from RNA-Seq, with expression level presented as log$_2$-transformed mean values from three biological replicates. The color spectrum from blue to red represents low to high expression levels, respectively. **D** Expression patterns of the five candidate genes in *cz3* and CIR12. Data were obtained from the website (http://cotton.zju.edu.cn/10.rnasearch.html)

(D-pool) and 30 extreme-tall plants (T-pool) based on F$_2$ plant height data, which were then pooled into two bulks (Fig. S1). The HAU versions of TM-1 genomes were chosen as the references. Subsequently, an average SNP-index was calculated using a sliding window approach (with a 1 Mb interval and 1 kb step width) for the two extreme bulks. The mapping ratio of the clean reads is between 97.07 and 97.76%, while the sequencing depths varied from 12.30-fold to 20.92-fold for *cz3*, CIR12, D-pool, and the T-pool, respectively. The Delta SNP-index was calculated and graphed across the entire genome using the Delta SNP-indexes for D-pool and T-pool. Regions with SNP-index values exceeding the threshold were identified as candidate QTL regions. Under a 95% level of confidence, only one genomic region (58.1–62.1 Mb) on chromosome D05 qualified presence a major-effect QTL controlling plant height, which was designated as *qPH5.1* (Fig. 2).

Fine-mapping of *qPH5.1* locus

In order to conduct a more detailed mapping of *qPH5.1*, a larger population of F$_2$ individuals consisting of 1200 plants was developed in the year 2022. A total of 1800 F$_2$ plants, including 600 plants for bulk construction, were genotyped using KASP1 and KASP7 markers, which were developed based on BSA data (Table S2, Fig. S3). Additionally, all
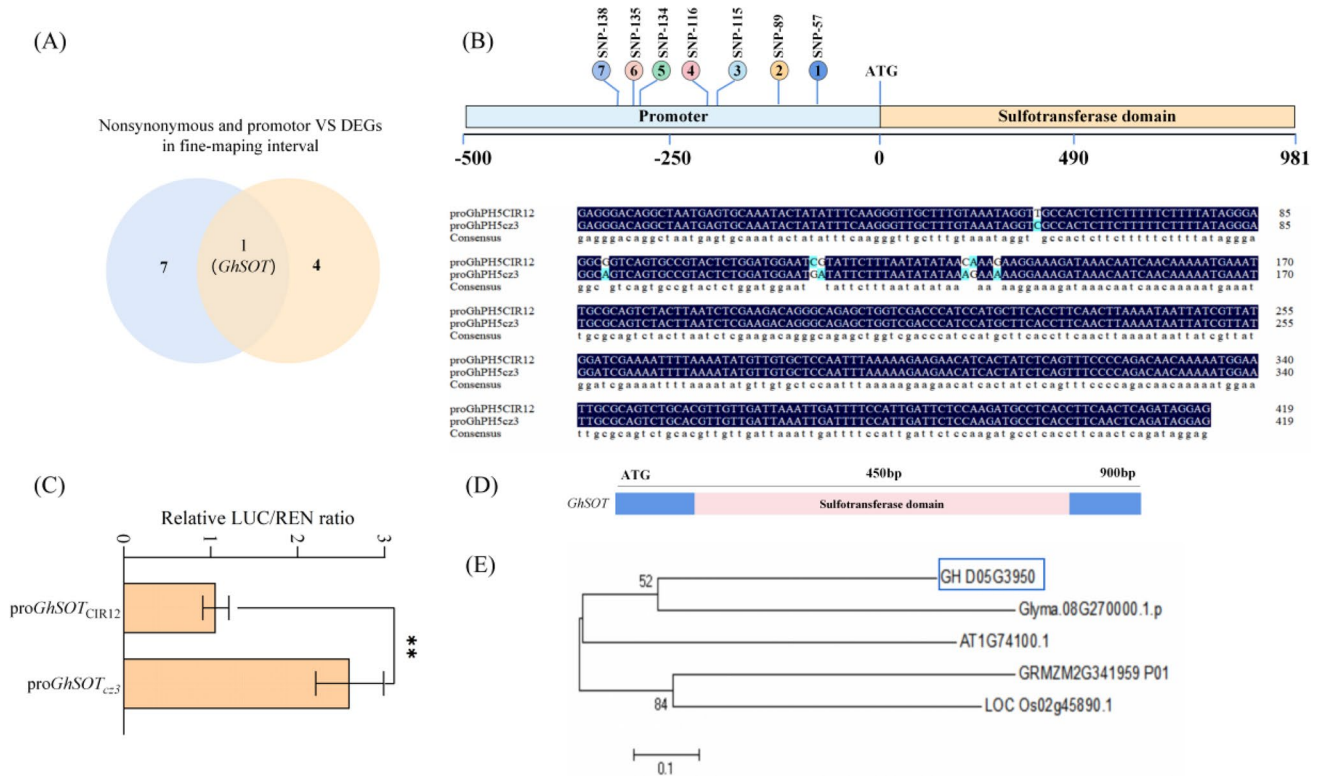
Fig. 4 Promoter analysis of the candidate genes. **A** Venn diagram showing the overlap between DEGs and nonsynonymous genes in fine-mapping interval. **B** Seven SNP variants of the *GhSOT* promoter region in CIR12 and *cz3*. **C** LUC experiments of *GhSOT* promoter in CIR12 and *cz3*, demonstrating that the promoter of *GhSOT* from *cz3* exhibited a greater enhancement of downstream gene expression. ****P < 0.01**, Student's *t* test was used to calculate the *P* value. **D** Gene structure of *GhSOT*. **E** Evolutionary analysis of *GhSOT* (*GH_D05G3950*)

progeny plants were genotyped using 7 markers (KASP1-KASP7), resulting in the identification of 20 recombinant plants. Subsequently, we assessed the performance of these plants for plant height in the field. Statistical analysis revealed a significant decrease in plant height for L3 and L4 in comparison to CIR12 (*P* < 0.01), while no significant differences were observed between contrasting genotypes in L1, L2, and L5 families. These findings provide evidence that the candidate gene is located within a physical interval of approximately ~ 0.34 Mb, which is flanked by markers KASP5 (61,068,056 bp) and KASP6 (61,403,436 bp) (Fig. 3A), this mapping interval is consistent with the *qPH5.1* (Fig. 3B). The BSA data of QTL-seq and the findings from fine-mapping reveal a total of 4838 SNP sites, with 4402 being intergenic, 91 exonic, 30 synonymous, and 60 non-synonymous mutations. Among these, 22 SNP sites are located on the *qPH5.1* interval, with 14 in exon, one in the interregion, and seven on the promoter. Out of these 14 exonic SNP sites, eight are synonymous, six are nonsynonymous. Further investigation will involve a comprehensive analysis of the genes associated with these SNP sites.

Identification of expressed genes in the candidate interval

To further investigate candidate genes, we conducted an analysis of the transcriptome data from stem tissues in *cz3* and CIR12. A total of 968 genes showed statistically significant expression differences between *cz3* and CIR12 (Table S3). Compared to CIR12, two genes were up-regulated and three genes were down-regulated of the stable QTL in *cz3* (Fig. 3C). Additionally, examination of the expression patterns of these five genes revealed that *GhSOT* (GH_D05G3950) is predominantly expressed in the stem, as reported in the CottonFGD database (http://cotton.zju.edu.cn/10.rnasearch.html) (Fig. 3D). Subsequent analysis identified seven SNP variants in the promoter region of *GhSOT* among the five DEGs, supporting this finding through sequencing of *GhSOT* from CIR12 and *cz3* (Fig. 4A, B). LUC experiments provided additional support, demonstrating that the promoter of *GhSOT* from *cz3* exhibited a greater enhancement of downstream gene expression (Fig. 4C). The alteration in the promoter sequence may influence the expression of the downstream gene, potentially contributing to the observed change in plant height. The amino acid
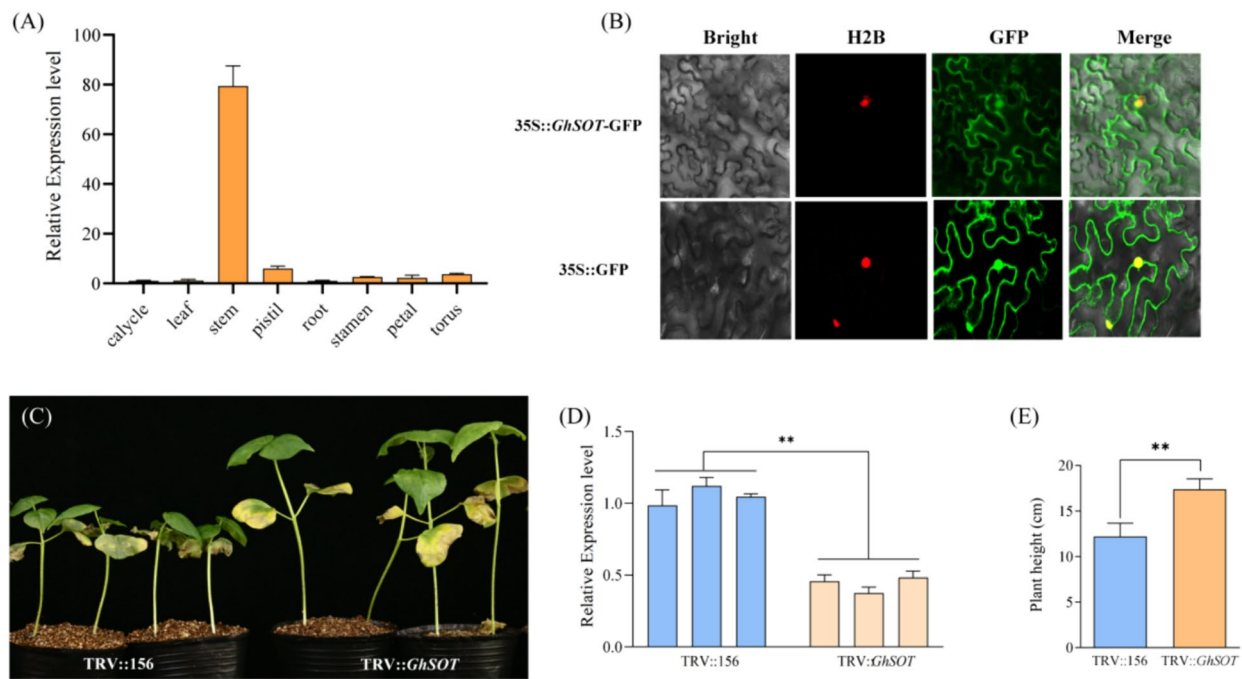
Fig. 5 Functional validation of *GhSOT*. **A** Expression analysis of *GhSOT* gene in different tissues. *GhHistone3* was used as an internal reference gene for data normalization. Data are presented as mean ± SD for three independent experiments. **B** Subcellular localization of *GhSOT* protein. Fusion of *GhSOT* with GFP protein was transiently expressed in *N. benthamiana* leaves, with free GFP serving as the control. H2B was employed as a nuclear maker. The visualization was conducted using confocal microscopy. **C** Phenotypes of blank control and *GhSOT* gene silenced in *cz3*. TRV::156 (empty, negative vector control), TRV::*GhSOT* (silenced cotton plants). **D** Quantification of *GhSOT* expression in blank control and VIGS in *cz3*. *GhHistone3* was used as an internal reference gene for data normalization. **E** Plant height statistics of TRV::*GhSOT* and TRV::156 of *cz3*. Data are presented as mean ± SD for three independent experiments. **P < 0.01, Student's *t* test was used to calculate the *P* value

sequences of *GhSOT* were subjected to BLAST searches against public databases for the identification of homologous sequences. The protein sequences of *GhSOT* and its four most similar homologous genes were used for phylogenetic analysis. The results showed that *GhSOT* shares 45% amino acid sequence identity with the soybean and 44% identity with the *Arabidopsis* SOT16 proteins (Fig. 4D, E). These results suggest that *GhSOT* may be the candidate gene for plant height.

VIGS of *GhSOT* brings an increase in plant height

To better understand *GhSOT* function, we conducted quantitative real-time RT-PCR analysis to examine its relative expression patterns in various tissues. Our analysis indicated that *GhSOT* was mainly expressed in the stem, with significantly lower expression in the leaf, bract, sepal, torus, and fiber tissue (Fig. 5A). In addition, to explore the subcellular localization of *GhSOT* in cell, a coding sequence construct was introduced into the epidermal cells of *N. benthamiana* leaves. The construct was fused with GFP at the N-terminal of *GhSOT* The results demonstrated that the *GhSOT*-GFP

fusion protein was localized in the nucleus and plasma membrane (Fig. 5B). To elucidate the role of *GhSOT* in cotton, VIGS experiments were conducted on *cz3* to suppress the expression of *GhSOT* genes. Results showed that silencing *GhSOT* in *cz3* plants led to a 41.0% increase in plant height compared to control plants, suggesting that down-regulation of *GhSOT* partially restored plant height (Fig. 5C–E). Conversely, VIGS of CIR12 did not result in a significant change in plant height (Fig. S2), indicating that the expression level of *GhSOT* plays an important role in regulating plant development.

GhSOT may regulate plant height based on GA signaling

We employed the RNA-seq to investigate the pathways associated with stem development in a mutant organism, and identified the top 20 pathways that exhibited enrichment in DEGs. Notably, the Glycerolipid metabolism pathway displayed significant enrichment, with down-regulation observed in the related genes (*GPAT3*, *GPAT5*, *EXL3*, etc.) (Fig. 6A). Previous studies have indicated that *GPAT1* plays a role in regulating plant height by influencing cell length
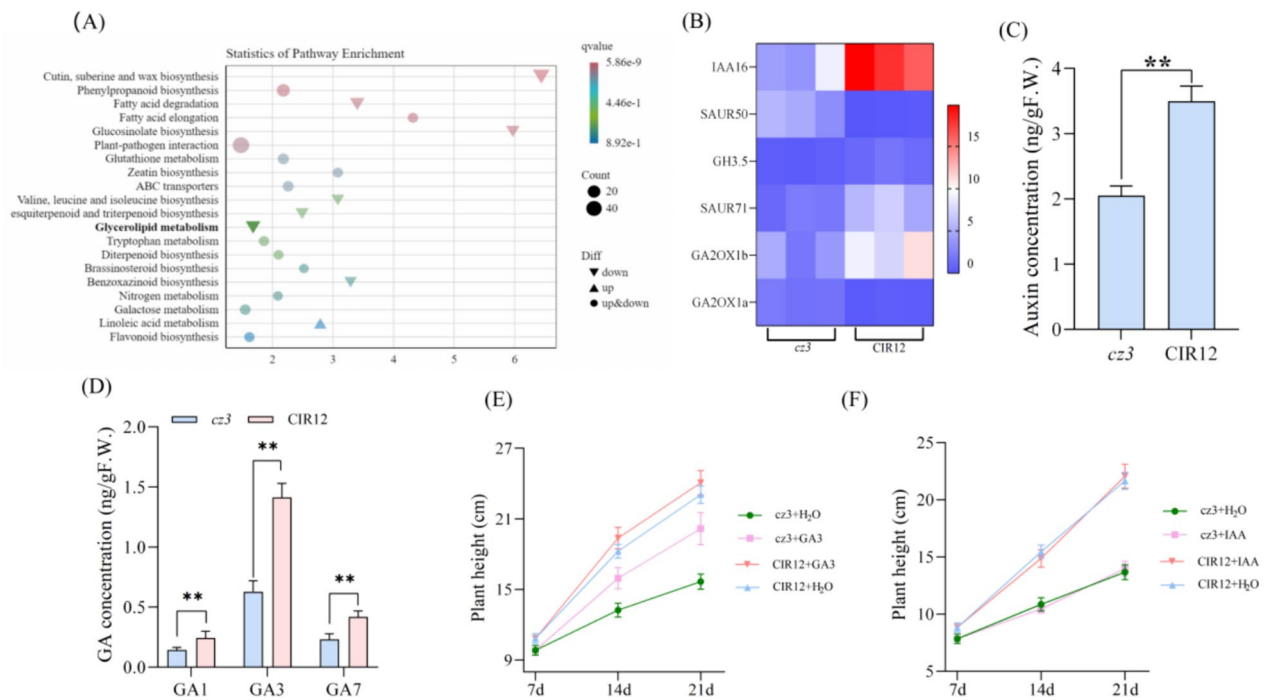
Fig. 6 Analysis of regulatory pathways. **A** KEGG pathway enrichment analysis of DEGs between *cz3* and CIR12. The top 20 KEGG pathways with the most abundant DEGs were presented. The y-axis and x-axis represent the pathway name and rich factor, respectively. The size of the circle dot indicates the gene count in each pathway. **B** RNA-seq is used to evaluate expression level of putative downstream genes involved in Aux-related and GA signaling between *cz3* and CIR12, with expression level presented as $\log_2$-transformed mean values from three biological replicates. The color spectrum from blue to red represents low to high expression levels. **C**, **D** IAA and GA quantification from the main stem between first and second fruit branches of *cz3* and CIR12. Data show mean $\pm$ SD ($n=3$), **$P<0.01$, Student's t test was used to calculate the P value. **E**, **F** Plant height in *cz3* and CIR12 plants after treatment with GA$_3$, IAA. Plant height is shown as the means $\pm$ SD ($n>30$)

(Bai et al. 2021). Additionally, four DEGs have been identified as key players in Auxin-related signaling pathways, including an Aux/IAA member, an Aux-responsive GH3 member, and two SAUR genes (Fig. 6B). Quantitative analysis of IAA and GA levels in the main stem between the first and second fruit branches revealed significantly higher levels in CIR12 compared to *cz3* ($P<0.01$) (Fig. 6C, D). Further, we started spraying IAA, GA$_3$ and H$_2$O on the cotton plants planted in the laboratory from the seventh day of sowing, with a concentration of 100 mg L^{-1} every two days. It is obvious that the plant height increment of *cz3* is higher than CIR12 compared to the control after spraying GA$_3$ (Fig. 6E). However, there was no significant change in plant height increment after spraying IAA with a concentration of 100 mg L^{-1} every two days (Fig. 6F). These findings indicate that the regulation of cotton plant height by *GhSOT* potentially relies on GA signaling.

Discussion

Cotton plant architecture breeding programs have become an effective method for improving yields (Song and Zhang 2009). The plant height is a crucial aspect of its overall structure and has implications for various agricultural factors such as mechanized harvesting, harvest index, and final yield performance. Plant height is a quantitative trait and many plant height QTLs in cotton have been reported that they are distributed over multiple linkage groups (Ye et al. 2023; Liu et al. 2020; Huang et al. 2017; Su et al. 2018; Wen et al. 2019; Zhang et al. 2019). These QTL fine-mapping and map-based cloning are supported by high-density genetic map and a series of near-isogenic lines. And the BSA-seq technology has proven to be successful for rapidly identifying the chromosome region harboring the genes/QTLs of interest in numerous crop species (Takagi et al. 2015; Lei et al. 2020; Xin et al. 2020). The BSA-seq technology can reduce the genotyping cost of large sets of individuals by choosing only extreme individuals for genotyping (Elshire et al. 2011). The present study mapped a major locus,

qPH5.1, which is located at a different position from all of the previous studies in cotton.

Within the mapping interval, it was confirmed that *GhSOT* belonging to the sulfotransferase (*SOT/ST*) family regulates the development of plant height. Analysis of the *GhSOT* promoter revealed seven SNP site variations between the two parental genes. LUC experiments indicated that the promoter from *cz3* effectively enhances the expression of the downstream gene *GhSOT*. Previous research has demonstrated the significant involvement of *SOT/ST* family genes in various plant processes (Komori et al. 2009; Igarashi et al. 2012; Zhou et al. 2010). In this study, RNA-seq analysis was conducted to investigate the role of *GhSOT* in regulating plant height development, with a suggesting potential modulation by the GA signaling pathway. Previous research has extensively documented the role of GA in the regulation of plant height (Wang and Wang 2022). The development of semi-dwarf varieties with increased yields during the 'Green revolution' is attributed to significant modifications in GA biosynthesis and signaling pathways (Evenson and Gollin 2003). Improving plant height by altering the GA pathway has proven to be an effective way (Tester and Langridge 2010; Wang et al. 2017). Additionally, there is evidence of crosstalk between GA and other phytohormones, such as Aux, brassinosteroids (BRs), and jasmonates (JAs), in the regulation of plant height (Ferrero-Serrano et al. 2019). Auxin response factor (ARF) and Aux/IAA family members play a direct role in up-regulating the expression of GA catabolic genes *GA20ox* and *GA2ox*, thereby coordinating GA metabolism during the process of organ elongation (Alamin et al. 2018). Therefore, GA signaling plays a crucial role in controlling plant height.

In this study, a *SOT/ST* family gene *GhSOT* was detected via BSA-seq and transcriptome sequencing. We have found seven SNP variants in the gene promoter between CIR12 and *cz3*. RT-PCR analysis indicated that *GhSOT* was mainly expressed in the stem. And its function was verified by VIGS, knock-down *GhSOT* expression partially restored the height of *cz3* plants. These results reveal *GhSOT* is a key gene controlling plant height. Transcriptomics data analysis has demonstrated that the expression levels of genes associated with GA biosynthesis and signaling, as well as those involved in cell wall organization and biogenesis, ultimately led to the elongation of cell length and consequently an increase in plant height. We found the plant height increment of *cz3* is higher than CIR12 compared to the control after application of GA3. There was no significant change in plant height increment after spraying IAA (Fig. 6E, F). We speculated the regulation of cotton plant height by *GhSOT* potentially relies on GA signaling. However, this inference needs further experiments to be authenticated.

Conclusion

In the present research, we have mapped a new QTL, *qPH5.1*, associated with plant height on chromosome D05 using BSA-seq and RNA-seq methods with ~ 0.34 Mb region. This region contains five genes, including *GhSOT* which is the candidate gene for plant height. These findings will boost further research on the genetic mechanisms underlying plant height.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Alamin M, Zeng DD, Sultana MH et al (2018) Rice *SDSFL1* plays a critical role in the regulation of plant structure through the control of different phytohormones and altered cell structure. J Plant Physiol 231:110–123

Bai Y, Shen Y, Zhang ZQ et al (2021) A *GPAT1* mutation in *Arabidopsis* enhances plant height but impairs seed oil biosynthesis. Int J Mol Sci 22:785

Burch-Smith TM, Anderson JC, Martin GB et al (2004) Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J 39:734–746

Challa KR, Rath M, Sharma AN et al (2021) Active suppression of leaflet emergence as a mechanism of simple leaf development. Nat Plants 7:1264–1275

Chen B, Zhang Y, Yang J et al (2021) The G-protein α subunit *GhGPA* positively regulates *Gossypium hirsutum* resistance to *Verticillium dahliae* via induction of SA and JA signaling pathways and ROS accumulation. Crop J 9:823–833

Dong HZ, Zhang YJ, Zhang DM et al (2018) New grouped harvesting-based population structures of cotton. Sci Agric Sin 51:4615–4624

Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Evenson RE, Gollin D (2003) Assessing the impact of the green revolution, 1960 to 2000. Science 300:758–762

Ferrero-Serrano Á, Cantos C, Assmann SM (2019) The role of dwarfing traits in historical and modern agriculture with a focus on rice. Cold Spring Harb Perspect Biol 11:a034645

Hu Y, Chen JD, Fang L et al (2019) *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet 51:739–748

Huang C, Nie XH, Shen C et al (2017) Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. Plant Biotechnol J 15:1374–1386

Igarashi D, Tsuda K, Katagiri F (2012) The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. Plant J 71:194–204

Ji GX, Liang CZ, Cai YF et al (2021) A copy number variant at the HPDA-D12 locus confers compact plant architecture in cotton. New Phytol 229:2091–2103

Komori R, Amano Y, Ogawa-Ohnishi M et al (2009) Identification of tyrosylprotein sulfotransferase in *Arabidopsis*. Proc Natl Acad Sci U S A 106:15067–15072

Lei L, Zheng HL, Bi YL et al (2020) Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.) using QTL-Seq and RNA-Seq. Rice 13:55

Li XM, Jin X, Wang HT et al (2016) Structure, evolution, and comparative genomics of tetraploid cotton based on a high-density genetic linkage map. DNA Res 23:283–293

Liu RX, Xiao XH, Gong JW et al (2020) QTL mapping for plant height and fruit branch number based on RIL population of upland cotton. J Cotton Res 3:9

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402–408

Ma JJ, Pei WF, Ma QF et al (2019) QTL analysis and candidate gene identification for plant height in cotton based on an interspecific backcross inbred line population of *Gossypium hirsutum* × *Gossypium barbadense*. Theor Appl Genet 132:2663–2676

Pang JH, Zhu Y, Li Q et al (2013) Development of Agrobacterium-mediated virus-induced gene silencing and performance evaluation of four marker genes in *Gossypium barbadense*. PLoS ONE 8:e73211

Song XL, Zhang TZ (2009) Quantitative trait loci controlling plant architectural traits in cotton. Plant Sci 177:317–323

Su JJ, Li LB, Zhang C et al (2018) Genome-wide association study identified genetic variations and candidate genes for plant architecture component traits in Chinese upland cotton. Theor Appl Genet 131:1299–1314

Takagi H, Tamiru M, Abe A et al (2015) MutMap accelerates breeding of a salt-tolerant rice cultivar. Nat Biotechnol 33:445–449

Teng F, Zhai LH, Liu RX et al (2013) *ZmGA3ox2*, a candidate gene for a major QTL, *qPH3.1*, for plant height in maize. Plant J 73:405–416

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science 327:818–822

Wang SS, Wang YJ (2022) Harnessing hormone gibberellin knowledge for plant height regulation. Plant Cell Rep 41:1945–1953

Wang YJ, Zhao J, Lu WJ et al (2017) Gibberellin in plant height control: old player, new story. Plant Cell Rep 36:391–398

Wen TW, Dai BS, Wang T et al (2019) Genetic variations in plant architecture traits in cotton (*Gossypium hirsutum*) revealed by a genomewide association study. Crop J 7:209–216

Wu HH, Ren ZY, Zheng L et al (2021) The bHLH transcription factor *GhPAS1* mediates BR signaling to regulate plant development and architecture in cotton. Crop J 9:1049–1059

Wu J, Mao LL, Tao JC et al (2022) Dynamic quantitative trait loci mapping for plant height in recombinant inbred line population of upland cotton. Front Plant Sci 13:914140

Xin F, Zhu T, Wei SW et al (2020) QTL mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. Sci Rep 10:25

Ye LL, Wang PL, Zhang M et al (2023) UAV-based time-series phenotyping reveals the genetic basis of plant height in upland cotton. Plant J 115:937–951

Zhang Z, Liu AY, Huang Z et al (2019) Quantitative trait locus mapping for plant height and branch number in an upland cotton recombinant inbred line with an SNP-based high-density genetic map. Euphytica 215:110

Zhao YD, Li MC, Konaté MM et al (2021) TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. J Transl Med 19:269

Zhou WK, Wei LR, Xu J et al (2010) *Arabidopsis* tyrosylprotein sulfotransferase acts in the auxin/plethora pathway in regulating postembryonic maintenance of the root stem cell niche. Plant Cell 22:3692–3709