BMC Genomics

RESEARCH

Open Access

Transcriptome and WGCNA reveals the potential genetic basis of photoperiod-sensitive male sterility in soybean

Yuhua Yang[1,2*], Suqin He[2,3], Lihong Xu[1,3], Minggui Wang[1,3], Shuichun Chen[1,3], Zhiyuan Bai[1], Tingting Yang[3], Bo Zhao[2,3], Lixiang Wang[2,3], Haiping Zhang[1], Jiangjiang Zhang[4*] and Ruijun Zhang[1*]

Abstract

Background Soybean (*Glycine max* (L.) Merr.) is a crucial crop due to its high plant protein and oil content. Previous studies have shown that soybeans exhibit significant heterosis in terms of yield and protein content However, the practical application of soybean heterosis remains difficult, as the molecular mechanisms underlying photoperiod-sensitive genic male sterile (PGMS) is still unclear.

Results This study characterized the PGMS line 88-428BY, which is sterile under short-day (SD) conditions and fertile under long-day (LD) conditions. To elucidate the genetic basis for this trait, we collected anthers, from 88-428BY under SD and LD conditions at three developmental stages, resulting in the identification of differentially expressed genes (DEGs) (2333, 2727 and 7282 DEGs, respectively) using Illumina transcriptome analysis. Using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses, we fund that among the DEGs, enriched genes were associated with photoperiod stress, light stimulus, oxidation-reduction processes, multicellular organism development and protein phosphorylation. Additionally, weighted correlation network analysis identified four modules (blue, brown, red, and yellow) that were significantly correlated with PGMS, revealing co-expressed hub genes with potential regulatory roles. Functional annotation of 224 DEGs with|KME| >0.9 across the four modules in seven databases highlighted their involvement in light stimulus, oxidation-reduction processes, multicellular organism development, and protein phosphorylation, suggesting their importance in soybean PGMS. By integrating fertility-related genes previously identified by other studies with the DEGs from our analysis, we identified eight candidate genes associated with the photosensitive sterility in soybeans.

Conclusions This study enhances the understanding of PGMS in soybean and establishes the genetic basis for a two-line hybrid seed production system in soybean.

Keywords Soybean, Photoperiod-sensitive genic male sterility, WGCNA, Photoperiod stress, Hybrid breeding

*Correspondence:
Yuhua Yang
yangyuhua@sxau.edu.cn
Jiangjiang Zhang
zhangjiangjiang6@163.com
Ruijun Zhang
zrj013835@126.com
Full list of author information is available at the end of the article

Introduction

Soybean is considered one of the most economically important leguminous crops due to its high plant protein and oil contents in the seed [1, 2]. Despite significant advancements in soybean breeding in recent decades, increasing yield remains challenging [1, 3]. Studies have shown that hybrid cultivars, which exhibit a 15–50% yield increase compared to inbred cultivars, can harness the heterosis present in soybean [4]. Heterosis, a phenomenon successfully utilized in maize, rice, sorghum, rapeseed and sunflower, is recognized as an effective method to increase crop yields [5–7]. However, the application of heterosis in soybean has been limited by factors such as natural outcrossing and a narrow genetic background in male-sterile lines and restorers [8, 9].

Selecting a male sterility (MS) system is a crucial step in hybrid seed production, as it forms the basis for crop heterosis and the identification of sterile genes [10–14]. There are three types of male sterility (MS): genetic or nuclear male sterility (GMS/NMS), cytoplasmic male sterility (CMS), and cytoplasmic genetic male sterility (CGMS) [15]. Approximately 50 cm and 30 N/GMS lines/mutants have been identified in soybean [7, 10, 12–14, 16–19]. GMS systems include structural MS, partially MS, and photosensitive MS, and CMS is primarily utilized in three-line breeding systems, with each line only being viable in specific soybean planting regions. Matched lines are selected from a large number of germplasm resources, especially an abundance of restorers lines for CMS [19, 20]. Inefficient maintenance of pure MS lines has hindered the application of GMS in soybean hybrid seed production. Among GMS, environment-sensitive genic male sterility (EGMS) has shown promise in two-line hybrids compared to three-line hybrids developed from the CMS system [20]. EGMS lines exhibit male sterility induced by specific environmental factors, such as temperature and day length [7, 19]. Understanding its genetic basis and effectively utilizing these genetic tools can accelerate two-line hybrid seed production heterosis in soybean.

In soybean, the first identified EGMS line was MS3, which exhibits male sterility influenced by photoperiod. A recent study revealed that *MS3* encodes a homeodomain-finger protein, with natural variations in *ms3*, leading to pre-termination and loss of the PHD-finger domain [20]. Another EGMS line, 88-428BY, also displayed male sterility induced by photoperiod [21]. To determine the molecular mechanisms and gene networks underlying male sterility in 88-428BY, RNA sequencing (RNA-seq) analysis was conducted to examine the genome-wide expression patterns of relevant genes at three stages of anther development [22] under short-day (SD) and long-day (LD) conditions. Differentially expressed genes (DEGs) in response to photoperiod and light stimulus, oxidation-reduction processes, multicellular organism development and protein phosphorylation were identified. DEGs involved in the chlorophyll metabolization as well as toxicological, forming interactive networks, were also detected using weighted gene co-expression network analysis (WGCNA). The findings of this study enhance our understanding of photoperiod-sensitive genic male sterility in soybean.

Materials and methods

Plant materials and different day-length treatments

Soybean seeds of 88-428BY were provided by Baoguo Wei [21]. Seeds of 88-428BY were cultivated in an experimental field in Taiyuan, and designated as 88-428BY (LD: 14.5–15.0 h light). SD (12 h light/12 h dark) treatments were performed in Taiyuan using a black shading cloth. The 88-428BY seedlings were cultured under SD treatment for 20 days, following which the shade cloth were removed. The SD experiments were also conducted in a paddy field at Sanya, between 2005 and 2022, with an average light duration of 11 h, 5 min, and 30 s during November and December, to validate the PGMS phenotype of 88-428BY.

Phenotype characterization

Plants were photographed using a Nikon D7100 digital camera (Japan). To analyze pollen fertility, anthers were collected from three flower buds (1 day before flowering) per plant and stored in 70% alcohol. Pollen grains and mature pollen were dyed using a 1% potassium iodide solution (I_2-KI) and observed under an OLYMPUS CX33 microscope (Japan) to assess the fertility of mature pollen.

Sampling and RNA extraction

Anthers at development stage 4 (S4), stage 8 (S8) and stage 12 (S12) were sampled in soybean 88-428BY (LD) and 88-428BY (SD) and immediately flash-frozen in liquid nitrogen before being stored at −80℃ for further analysis. Ten individual plants were sampled at each stage. Total RNA was extracted using an RNA isolation kit (Tiangen, China), following the manufacturer's instructions. The RNA quality was evaluated using a Bioanalyzer 2100 (Aligent, USA).

RNA sequencing and data analysis

Illumina Novaseq™ 6000 (LC-BIO, Hangzhou, Zhejiang, China) was utilized for paired-end sequencing of approximately 10 μg of total RNA, obtained from three biological replicates following a standard protocol. A total of 18

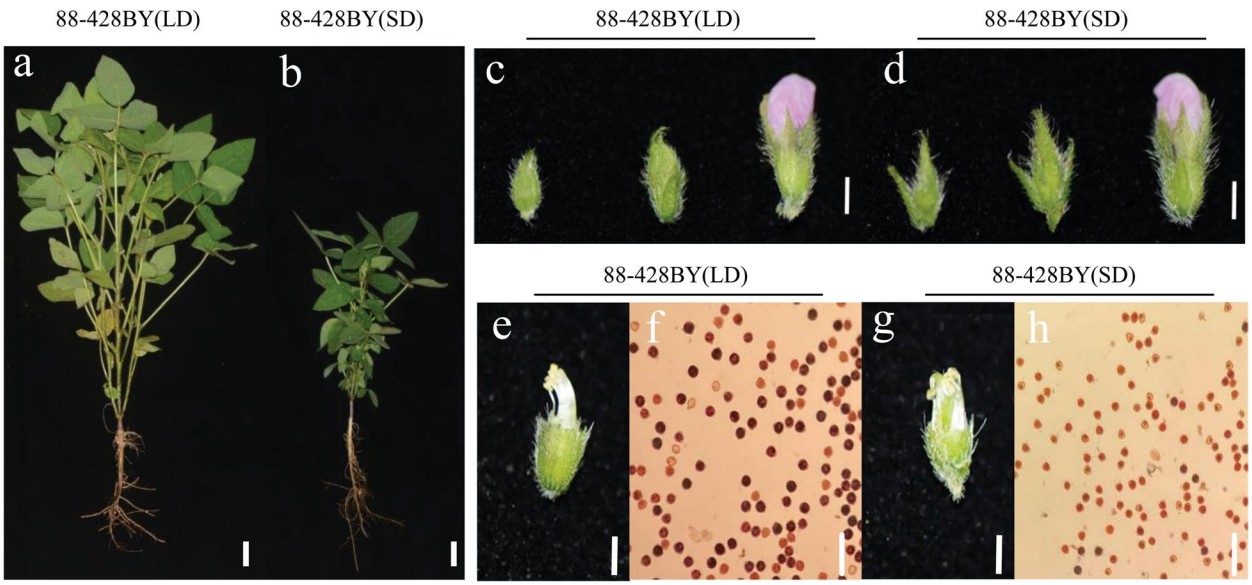

Fig. 1 Comparison of the phenotypes of 88-428BY under long day (LD) and short day (SD) conditions. **a**, **b** Whole plants at the flowering stage. **c**, **d** Bud phenotype at three anther development stages: 4, 8 and 12. **e**, **g** Buds with the petals and sepals removed. **f**, **h** Pollen grains stained with I 2 -KI. (Scale bar: a,b = 5 cm; c,d, e and g = 2 cm; f, h = 100 μm.)

cDNA libraries (comprising three biological replicates from three anther stages under both LD and SD conditions) were generated, yielding 150 bp paired-end reads with an average of 45 million raw data points per sample. Sequence quality assessment was conducted using a standard protocol [23]. Raw reads were filtered using Cutadapt to obtain high quality clean reads [24]. The clean reads were subsequently aligned to the reference genome (*Glyma.Wm82. a2. v1*) using HISAT2.2.4 [25]. The fragments per kilobase of exon model per million mapped reads (FPKM) value was calculated to quantify the gene expression levels [26]. Genes meeting stringent criteria ($|\log2foldchange| \geq 2$, $p < 0.05$) were identified as DEGs using edgeR software [27]. These DEGs were then anlyzed to identify their gene ontology (GO) terms and kyoto encyclopedia of genes and genomes (KEGG) pathways, revealing their enriched biological processes and metabolic pathways.

Weighted gene co-expression network analysis and candidate genes related to GMS

This study utilized the WGCNA package in R software (version 4.1.2) to construct gene co-expression networks [28]. The core DEGs were grouped into 16 modules using the WGCNA package in R, and the PGMS for each module was calculated. To determine the correlation between the module and PGMS, an association coefficient was calculated following the method outlined by Chen et al. [29]. A total of 28 genes from plants with known functions related to environment-sensitive genic, male sterility were identified from the NCBI website (https://www.ncbi.nlm.nih.gov/).

Validation of gene expression by qRT-PCR

Total RNA was extracted and followed by reverse transcription using the PrimeScriptTM^{RT} reagent kit

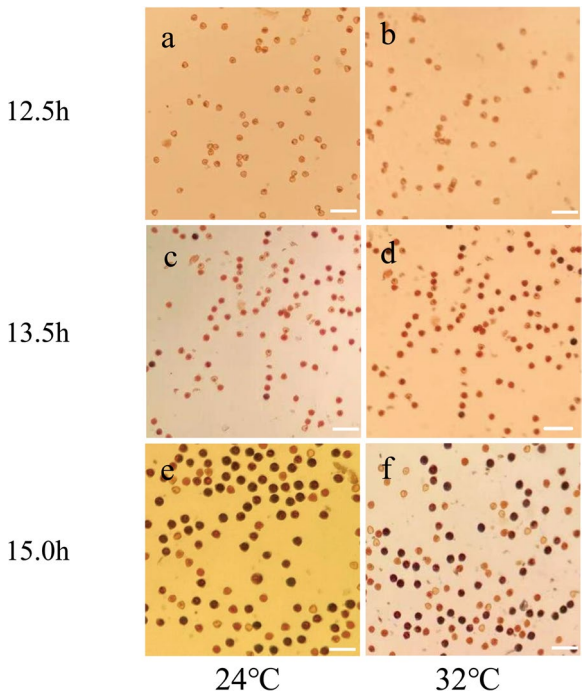

Fig. 2 I 2 -KI stained pollen grains of 88-428BY grown under different temperature and photoperiods in a growth room. (Scale bars: 100)

Table 1 Statistics of RNA-Seq results and corresponding quality control information of all samples

Sample	Raw Data		Valid Data		Valid Ratio (%)	Q20 (%)	Q30 (%)	GC Content (%)
	Read	Base	Read	Base				
LD_S4_1	44239928	6.64G	44125702	6.62G	99.74	99.97	98.84	42.50
LD_S4_2	50323360	7.55G	50192228	7.53G	99.74	99.98	98.74	42.50
LD_S4_3	47598132	7.14G	47487406	7.12G	99.77	99.98	98.79	42.50
LD_S8_1	47551512	7.13G	47444788	7.12G	99.78	99.97	98.75	42.50
LD_S8_2	44448652	6.67G	44347726	6.65G	99.77	99.98	98.80	42.50
LD_S8_3	50010854	7.50G	49863504	7.48G	99.71	99.98	98.73	41.50
LD_S12_1	43088996	6.46G	42968090	6.45G	99.72	99.98	98.34	43.00
LD_S12_2	47232370	7.08G	47087790	7.06G	99.69	99.99	98.32	42.00
LD_S12_3	49408874	7.41G	49218102	7.38G	99.61	99.99	98.87	43.00
SD_S4_1	46887548	7.03G	46714178	7.01G	99.63	99.98	98.64	42.50
SD_S4_2	47755714	7.16G	47620782	7.14G	99.72	99.98	98.63	43.00
SD_S4_3	49495808	7.42G	49373988	7.41G	99.75	99.98	98.55	43.00
SD_S8_1	48645194	7.30G	48521380	7.28G	99.75	99.98	98.74	43.00
SD_S8_2	42964364	6.44G	42857948	6.43G	99.75	99.98	98.47	43.00
SD_S8_3	45529538	6.83G	45382184	6.81G	99.68	99.98	98.61	43.00
SD_S12_1	41777986	6.27G	41667446	6.25G	99.74	99.99	98.74	43.00
SD_S12_2	42207288	6.33G	42050514	6.31G	99.63	99.98	98.65	42.50
SD_S12_3	43920464	6.59G	43767292	6.57G	99.65	99.98	98.64	42.50

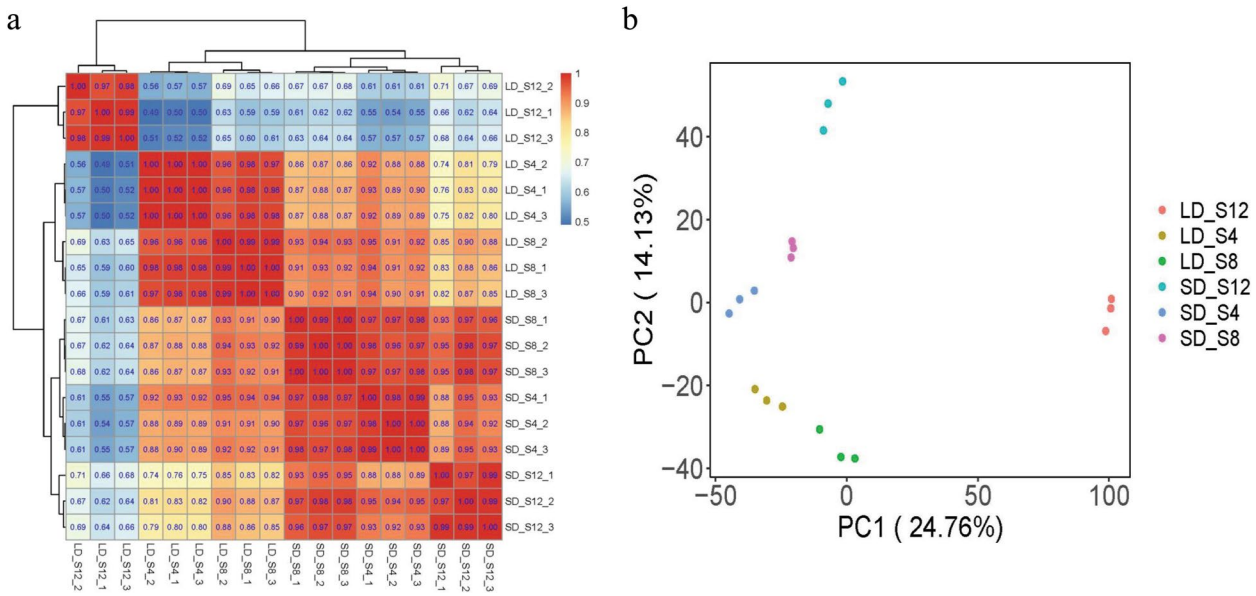

Fig. 3 **a** Pearson's correlation co-efficient analysis (PCC) and (**b**) principal component analysis (PCA)

(TaKaRa, Japan). The resulting first-strand cDNA was then amplified in accordance with the instructions provided for RealMasterMix (SYBR Green, TIANGEN). The relative expression levels of 10 DEGs across tissues and treatments were determined using the $2^{-\Delta\Delta Ct}$ method [30]. *GmActin11*(accession number: LOC100781831)was selected as the reference gene to assess the template load [31]. The experiment was conducted with three dependent biological replicates. The primers were designed using Primer 3 software (Table S1).

Results

Phenotype of photoperiod-sensitive genic male sterility line 88-428BY

When exposed to a photoperiod of 14.5–15.0 h (LD), soybean 88-428BY exhibited normal plant height and pollen grain fertility (Fig. 1a, c, e and f). In contrast, under a 12.0–13.0 h (SD) condition, the plants exhibited dwarfism, with less smaller flowers and complete male sterility (Fig. 1b, d, g and h). The consistent observations from 1998 to 2022 indicate that the photoperiod plays a pivotal role in regulating male sterility in 88-428BY.

Experiments with varying day lengths (12.5 h, 13.5 h, 15.0 h) and temperatures (24 °C and 32 °C) were conducted to determine whether the fertility phenotype of 88-428BY was specifically caused by photoperiod. As shown in Fig. 2, pollen from 88-428BY plants grown under 12.5 and 13.5 h photoperiods exhibited complete male sterility at both 24 and 32 °C, whereas plants grown under 15.0 /9.0 h photoperiod produced fertile pollen (Fig. 2e, f). These results confirm that the fertility of 88-428BY is predominantly controlled by photoperiod rather than temperature, aligning with previous research findings [21].

Analysis of the global transcriptome between samples

To investigate the molecular mechanism of male sterility in 88-428BY, transcriptome analysis was conducted at three stages of anther development under SD and LD conditions. Flower buds of 88-428BY at stages 4, 8, 12 were labeled as LD-S4, LD-S8, LD-S12, SD-S4, SD-S8, and SD-S12, depending on the photoperiod treatment. Illumina Novaseq™ 6000 sequencing generated approximately 830 million reads, with an average of 46 million reads per sample, a Q30 base ratio of over 98%, and GC content of about 42% (Table 1). Nearly 96% of valid data were mapped to the reference genome (Glyma.Wm82.a2. v1), with approximately 83% of these reads located at unique chromosomal positions (Table S2).

Principal component analysis (PCA) was conducted using the FPKM values, and samples were separated into 6 distinct groups along with their replicates (Fig. 3a). Samples from LD_S4, SD_S4, and SD_S8 clustered on the negative side of the x-axis, while LD_S12 was positioned on the positive side of the x-axis. On the other side, samples from SD_S12 and SD_S8 were located on the positive side of the y-axis, whereas LD_S4 and SD_S8 were on the negative side of the y-axis. Except for LD_S4 and SD_S4, which were on the same side of the lactated axis, samples from other stages under LD or SD conditions were distributed across different axes. PCA, further revealed that a significant proportion of the variation (38.89%), was explained by two principal components (PC1: 24.76%, PC2: 14.13%) (Fig. 3b).

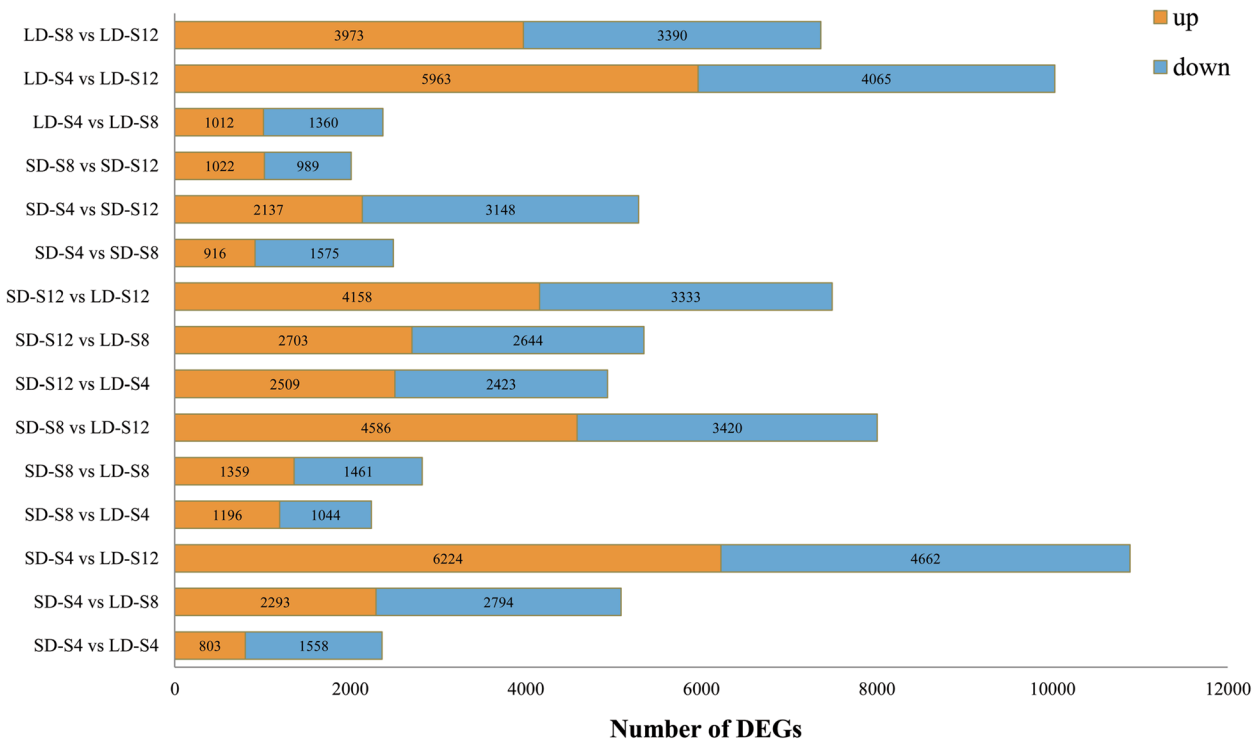

Fig. 4 Up- and down-regulated DEGs between LD and SD at three anther developmental stages

The median expression of all genes in the samples studied, was relatively stable, with minor variations, except for a decline in LD-S12 and SD-S12 (Fig. S1). Analysis of whole-genome FPKMs showed that the sample from the S4 and S8 stages had higher values compared to those from the S12 stage (Fig. S2). The Pearson's correlation coefficient (PCC) was used to confirm the association with RNA-seq data, revealing a significant positive correlation (≥ 0.97) among the 3 replicates of each sample. There was a relatively high correlation between the other samples (≥ 0.74), except for LD-S12. PCA showed that samples from the S12 stage under two different conditions clustered together but with a greater distance compared to the other samples. These results were consistent with PCC analysis, suggesting that the 88-428BY genotype exhibited high gene expression similarity during the S4 and S8 stages under LD/SD treatments, which may be linked to anther development.

Comparative analysis of differentially expressed genes during anther development under different photoperiod conditions

To identify DEGs across various treatments and anther developmental stages, a comparative analysis was conducted under LD and SD conditions at three distinct stages. The number of DEGs identified varied from 2,011 (comparing SD-S8 to SD-S12) to 10,028 (comparing LD-S4 to LD-S12) (Fig. 4). Interestingly, DEGs, were more abundant at stage 12 than at stages 4 and 8 under both LD and SD conditions. GO analysis was utilized to categorize the DEGs into biological processes, cellular components, and molecular functions (Fig. 5a, Table S3-S5). The comparison of treatments revealed that DEGs were associated with various GO terms, such as transcription regulation process, oxidation-reduction process, multicellular organism development, response to light stimulus, and protein phosphorylation. KEGG

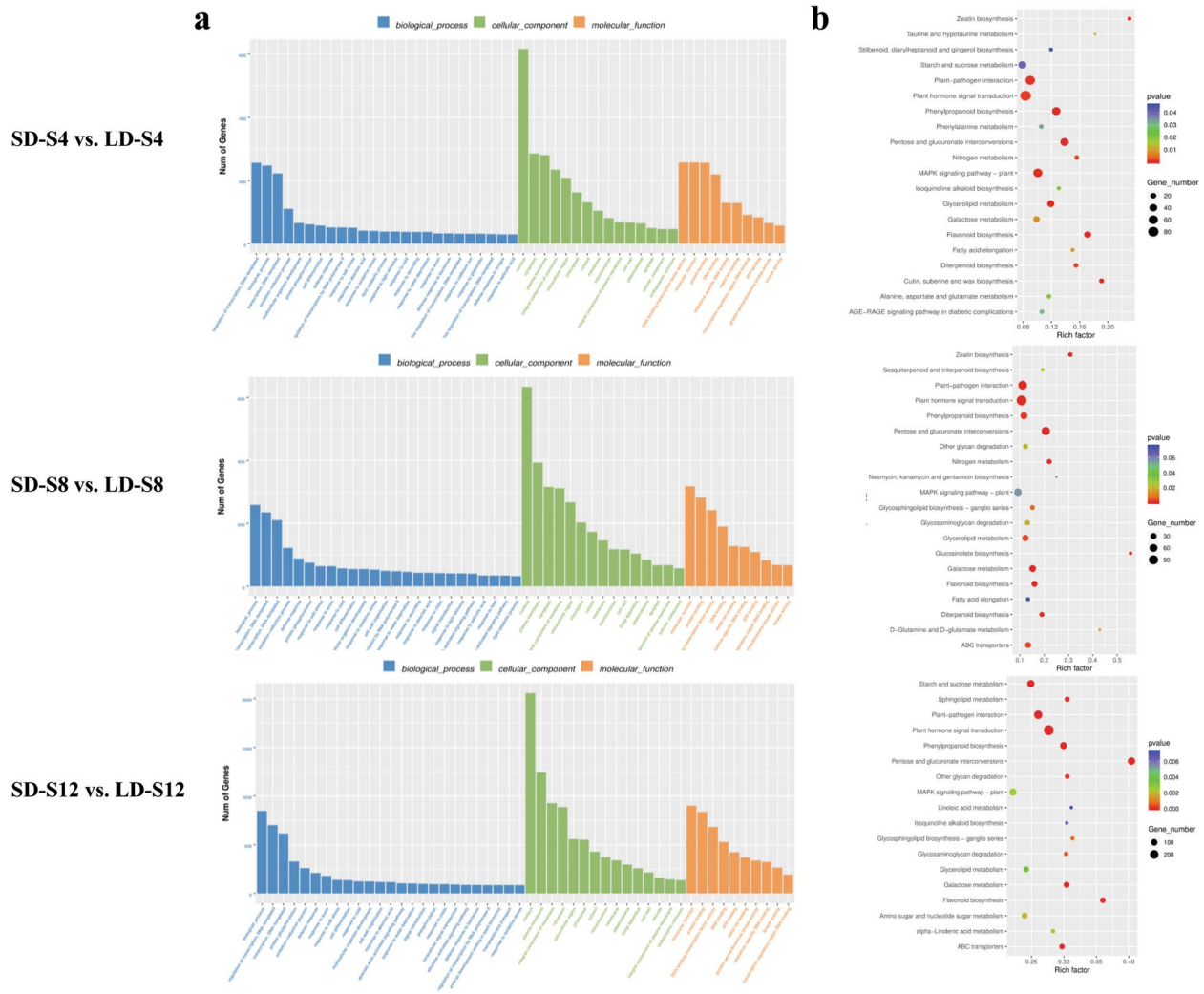

Fig. 5 **a** Gene Ontology and (**b**) Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of all DEGs

enrichment analysis shed light on the pathways enriched with DEGs. The DEGs were found to be enriched in common pathways across the three different anther development stages between 88-428BY under SD and LD conditions. These pathways included plant hormone signal transduction (KO04075), plant-pathogen interaction (KO04626), pentose and glucuronate interconversions (KO00040), MAPK signaling pathway (KO04016), phenylpropanoid biosynthesis (KO00940), galactose metabolism (KO00052), starch and sucrose metabolism (KO00500), glycerolipid metabolism (KO00561), flavonoid biosynthesis (KO00941), and ABC transporters (KO02010) (Fig. 5b).

Co-expression network analysis reveals candidate genes regulating photosensitive sterility in soybean

To identify highly associated genes in 88-428BY, we utilized WGCNA with a soft threshold power of 20 to construct a gene co-expression network (Fig. S3). A total of 17 co-expression modules were generated and visualized in a clustergram with distinct colors (Fig. 6a and Fig. S4). The network heatmap, which exhibited topological associations between the identified modules, shows lighter blue colors representing low association and darker red colors indicating strong association. By examining the correlation of eigengenes, we identified three clusters with strong interactions: modules brown and green, blue and red, and black and yellow (Fig. 6b).

The brown, blue, red, and yellow modules showed significant correlations with the PGMS phenotype ($r = 0.58$, 0.64, 0.72 and 0.87) in soybean (Fig. 7), indicating that genes within these modules have a positive regulatory effect on PGMS. A total of 244 genes were identified as hub genes in these four modules, based on the criterion of |KME|>0.9. The brown, blue, red, and yellow modules were composed 84, 97, 12, and 33 DEGs, respectively (Table S6). Functional annotation of 224 DEGs (|KME|>0.9) in the four modules was performed using seven databases (Fig. 8). Several DEGs were associated with light stimulus, oxidation-reduction processes, multicellular organism development, and protein phosphorylation, suggesting their importance in the photosensitive sterility process in soybean.

Among the identified genes, *Glyma.08G250800* and *Glyma.10G240900* encode MADS-box transcription factors. Previous research has highlighted the significant roles of MADS-box transcription factors, such as *GmNMH7* and *GmMADS28*, in reproductive development and plant fertility [32, 33]. The *Glyma.19G219000* encodes a MYB2 transcription factor that regulates plant mortality and senescence by inhibiting cytokinin production, leading to less apical meristem under environmental stress and a shortened vegetative stage [34, 35].

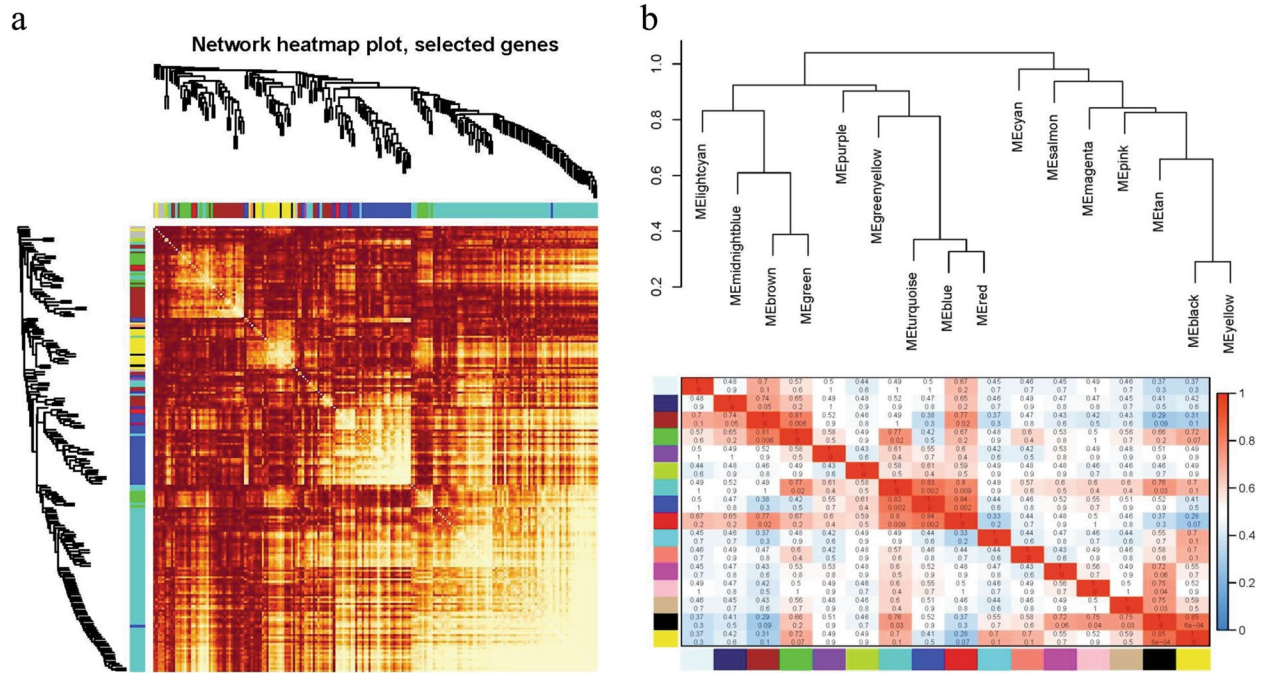

Fig. 6 WGCNA of high-variance DEGs in 88-428BY at different anther developmental stages. **a** Cluster dendrogram and network heatmap plot. **b** Module-sample associations based on Pearson correlations. The color key indicates r 2 values from – 1 to 1 in a gradient from blue to red

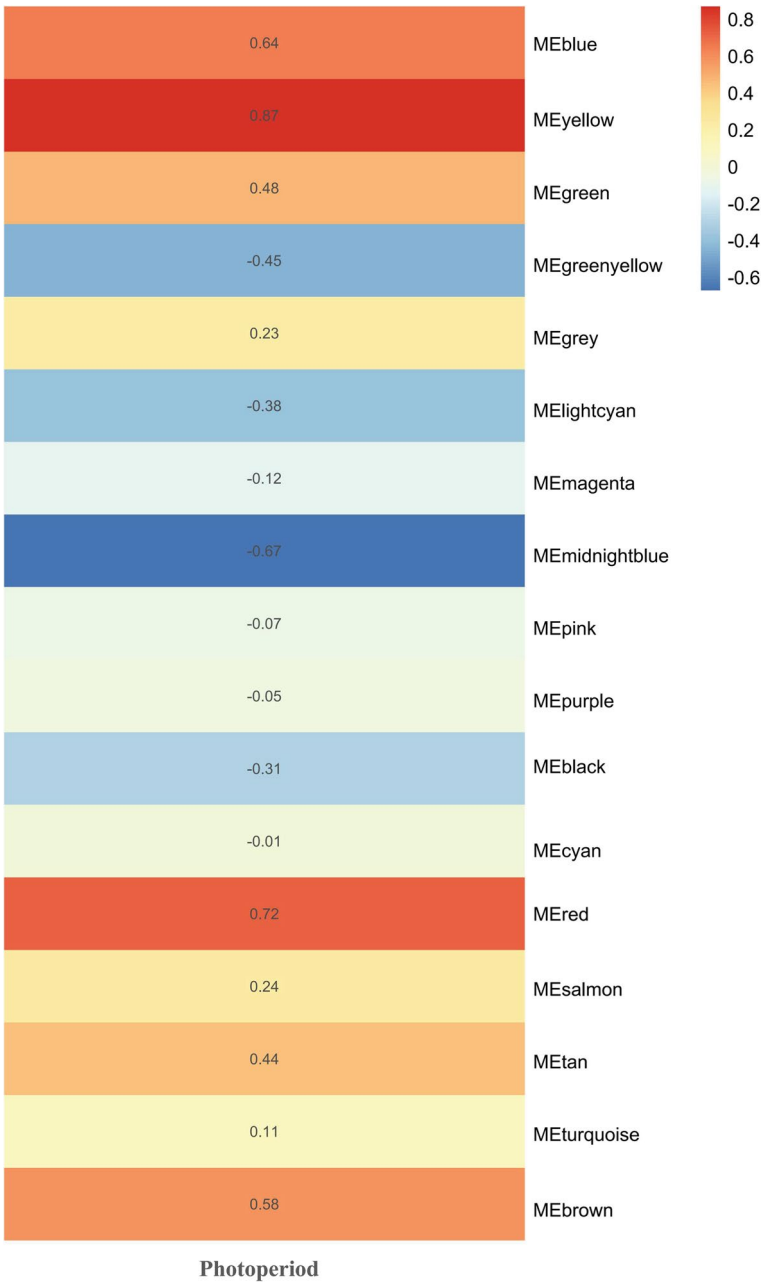

Fig. 7 Module-trait relationships plots. The numbers represent the Pearson's correlation coefficients, with positive correlations in red and negative correlations in blue

Furthermore, 10 hub genes were selected to validate the reliability of the DEGs identified from RNA-seq data. qRT-PCR analysis was conducted to assess gene expression levels at the different anther development stages under two treatments. Figure 9 illustrates that these 10 genes exhibited a consistent trend with the result of RNA-seq. Thus, our study confirms the relevance of the identified hub genes in the soybean response to day-length treatment.

Selection of photoperiod-sensitive genic male sterile candidate genes

A total of 28 cloned genes associated with male sterility in various crops have been identified in the previous studies (Table S7) [12, 18, 20, 36–53]. Among these, 20 genes were found to be expressed in the transcriptome (Table S8). The expression levels of 5 genes exhibited significant differences between LD and SD conditions (Fig. 10), namely *cas*, *CalS5*, *res3*, *ms1*, and *ms6*, which

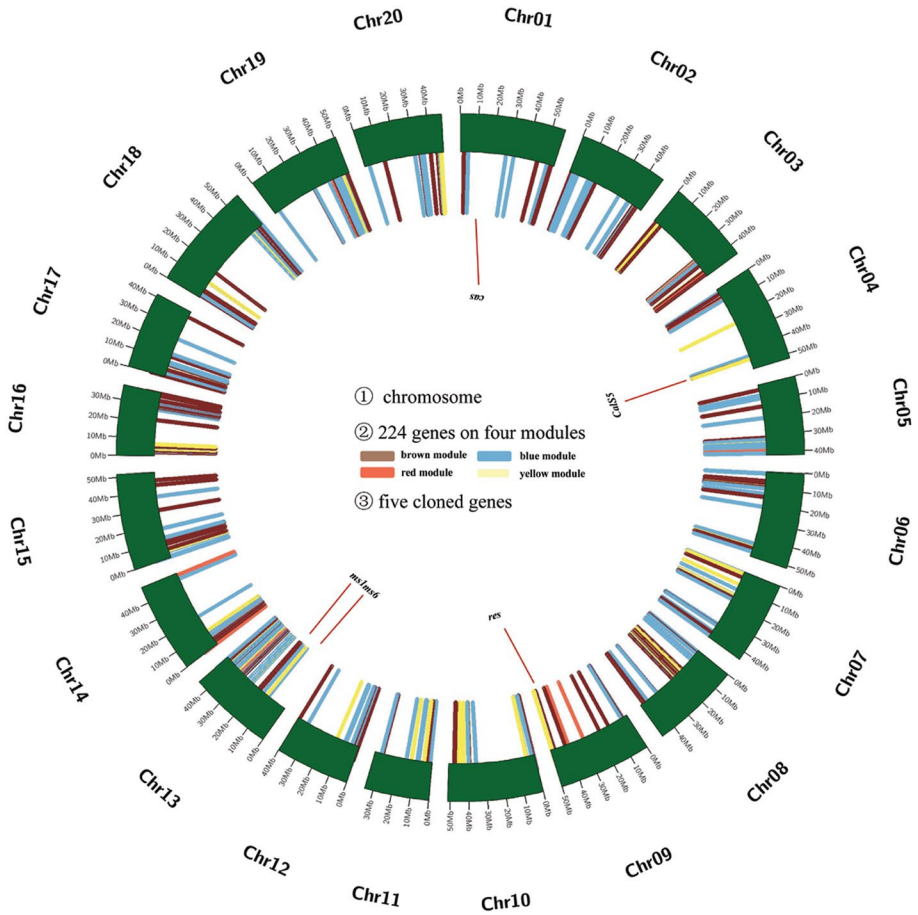

Fig. 8 Integration of candidate genes for GMS on the soybean reference genome. The Circos program was used to draw different features in concentric circles, including:① 20 soybean chromosomes, ② 224 genes in the four modules, different colors represent different modules, and ③ 5 cloned genes

validated the reliability of the transcriptome sequencing data in this study. These genes showed significant differences in the transcriptome of the soybean line 88-428BY, suggesting their involvement in the regulation of male sterility in soybeans. The observed variations in gene expression levels may influence the genetic control of male fertility through diverse molecular mechanisms. These findings imply a possible link between the expression of these genes and the development of male sterility in the specific soybean line under study. Investigating the functions of these genes and their regulatory mechanisms may shed light on the genetic basis of male sterility and its manipulation in soybean breeding programs.

Discussion

Circadian rhythms and light signals play a crucial role as a key environmental factors in plant growth and development. Plants sense day-length signals to regulate their physiological and developmental responses [54–56]. Soybean was the first plant studied for its photoperiod response. It is a classic short day-length plant that is sensitive to day-length, and that is unable to flower when exposed to sunlight beyond a certain threshold. This can lead to photoperiod stress, affecting the flowering, maturity and other biological processes of soybean [57–59]. Despite this, there are limited reports on natural variation in soybean sensitivity to photoperiods leading to nuclear sterility. The soybean *MS3* gene has been successfully identified and cloned, playing a key role in controlling photoperiod-sensitive genic male sterility [20]. In this study, high-throughput RNA-seq and WGCNA were applied to identify pathways and candidate genes associated with photoperiod sensitivity in another soybean genic male sterile line 88-428BY [21]. A total of 224 core DEGs were identified in pathways responsive to photoperiod stress (Table S6, Fig. 8). GO and KEGG enrichment analysis of these core DEGs revealed involvement in biological processes and pathways related to transcription factors, protein binding, MAPK signaling, starch and sucrose metabolism pathways (Figs. 5 and 6).

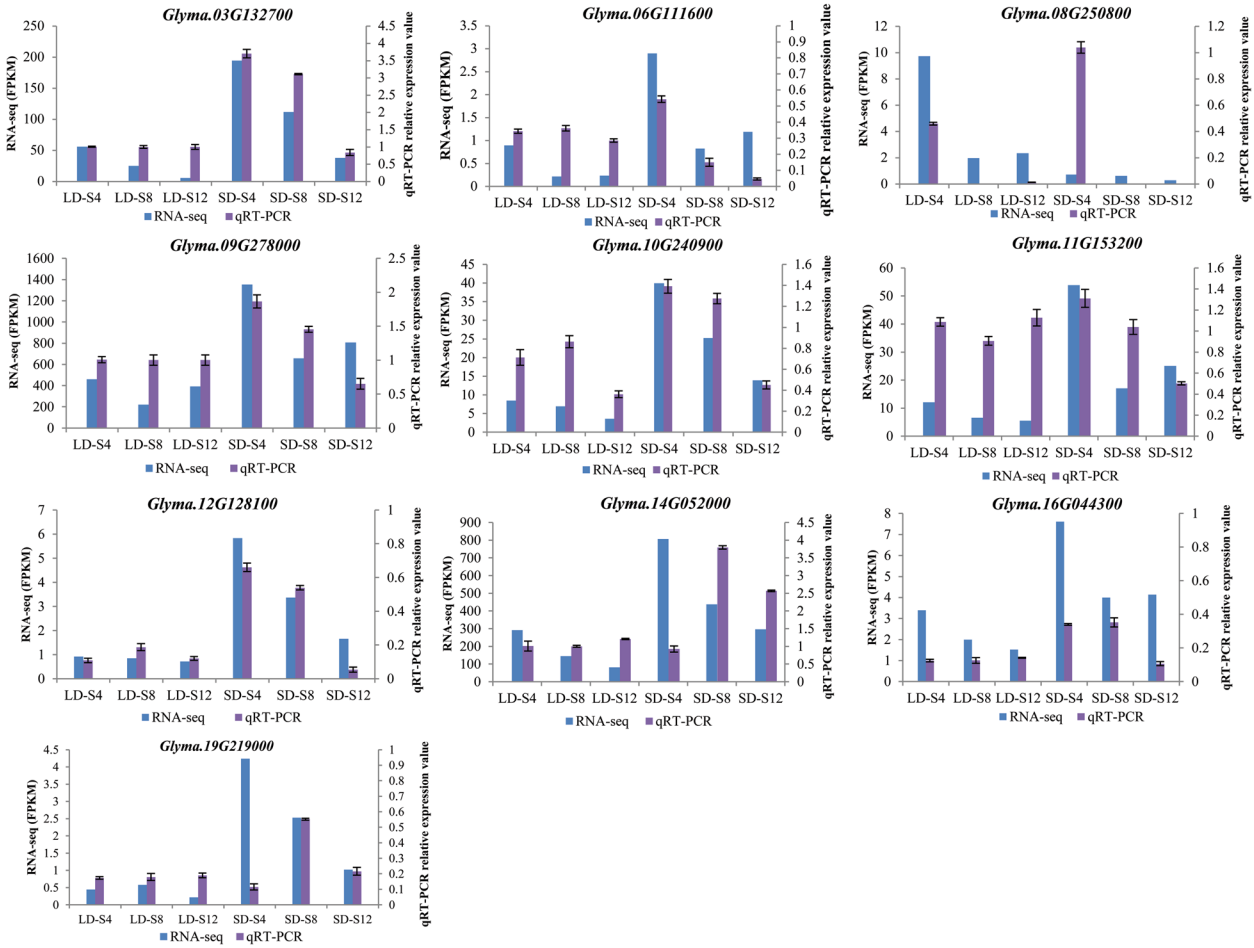

Fig. 9 Validation of 10 DEGs by qRT-PCR. The FPKM values of RNA-seq were compared with the relative expression level determined by qRT-PCR. The qRT-PCR results represent the mean ± standard error of the three replicates

The MYB and MADS-box family members were found to play important roles in various plant biological processes including plant growth, development, and abiotic stresses response [60–63]. Specifically, MYB proteins, including AtMYB2, AID1, and OSMYOXIB, have been highlighted for their roles in anther development and plant growth under photoperiod stress [64–66]. The MYB transcription factor *Glyma.19G219000* showed significant differential expression during soybean anther development under photoperiod stress. MADS-box family members have been implicated in regulating flowering and maintaining floral organ identity, with studies in rice demonstrating that mutations in these genes, lead to male sterility. GmMADS28 has been identified as a causative factor for male sterility in soybean [61–63]. This study identified expression variation in the MADS-box transcription factor gene, *Glyma.10G240900*, under different day-length conditions.

CSA shared homology with the *Glyma.01G049600* located on chromosome 01, showing a similarity of 71.5%. *Glyma.01G049600* demonstrated higher expression levels under LD conditions than under SD conditions, with its expression decreasing as anther development progressed. This expression pattern suggests that *Glyma.01G049600* is involved in anther development, potentially playing a role in the early stages of this process. The decline in expression during anther maturation indicates that its function is most crucial during the initial phases of development, after which its significance diminishes as the anther matures. This expression profile is indicative of genes, initiating or regulating developmental pathways in anthers. Previous research has revealed that *CSA* encodes an R2R3 MYB-type transcription factor and is vital in directing sugar allocation from source tissues, where photosynthesis takes place, to the developing anthers, which act as the sink tissues. Mutations in this gene may lead to male sterility in rice plants under short-day (SD) conditions; however, male fertility can be restored under long-day (LD) conditions [43].

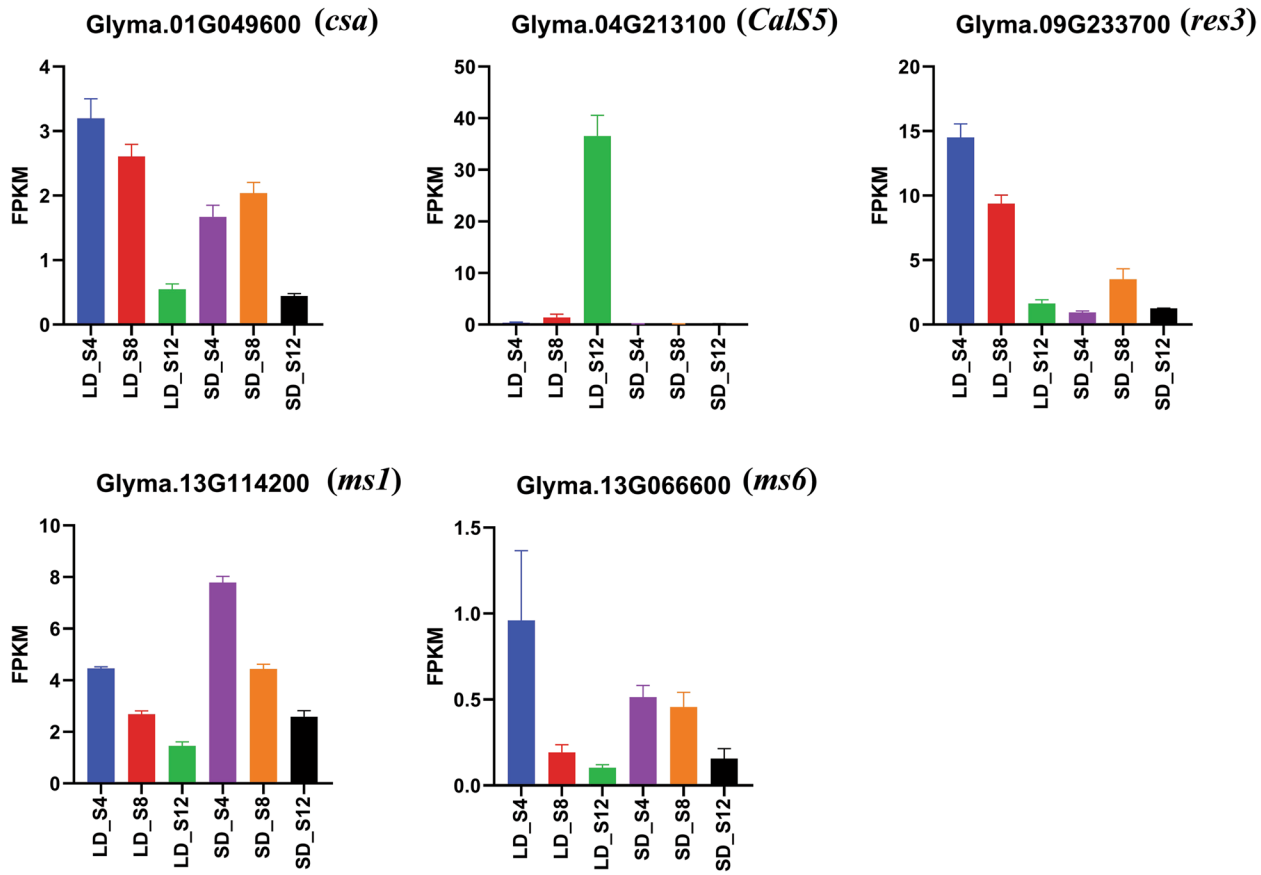

Fig. 10 FPKM values of five reported genes in RNA-seq

The *CalS5* gene shared 76.3% similarity with *Glyma.04G213100* located on soybean chromosome 04. *Glyma.04G213100* exhibited differential expression between LD and SD conditions, with peak expression levels observed in the LD-S12 stage and relatively low or no expression in other developmental stages. This expression pattern indicates that the gene's function is primarily active during the anther maturation phase. The hypothesis that its function may be associated with pollen viability is supported by its peak expression, which coincided with a crucial period in anther development when pollen maturity was determined. The function of the *CalS5* gene is directly linked to male fertility in *Arabidopsis thaliana*, as the *cals5-2* mutant exhibits male sterility under normal conditions, but it can regain fertility under permissive conditions, such as low temperature or SD photoperiod [42]. The *res3* gene has been shown to restore fertility in various known P/TGMS lines with pollen wall formation defects [37], and it is homologous to the soybean gene *Glyma.09G233700*, sharing a 49.4% similarity. *Glyma.09G233700* exhibited significantly lower expression levels in SD conditions compared to LD conditions, and its expression gradually decreases as anther development progresses. This expression pattern suggests that *Glyma.09G233700* plays a crucial role in anther development, particularly during the early stages of anther growth under LD conditions.

MS1 (*Glyma.13G114200*) and *MS6* (*Glyma.13G066600*) were identified, and their functions wrer linked to male sterility in soybean. *MS1* expression was notably higher under SD than under LD conditions, and its expression gradually decreased as the anthers developed. Previous research generated *ms1* mutant through CRISPR/Cas-mediated gene editing, exhibiting a complete male-sterile phenotype [18]. This highlights the crucial role of *MS1* in soybean male fertility. The expression pattern of *MS6* mirrored that of *MS1*, with *MS6* exhibiting higher expression levels under LD conditions than under SD conditions. Previous studies have indicated that *MS6*, which was highly expressed in anthers, encodes an R2R3 MYB transcription factor known as *GmTDF1-1* [10]. This transcription factor is homologous to Tapetal Development and Function 1 (*TDF1*), a key factor in anther development identified in *Arabidopsis* and rice. The homology and role of *MS6* suggest its involvement in regulatory pathways crucial for anther and pollen development in soybean.

Overall, these results offer valuable insights into the cloning of the mutated allele of PGMS in 88-428BY and its potential application in enhancing soybean heterosis.

Conclusions

This study identified soybean 88-428BY, which exchibited the PGMS trait. RNA-seq, WGCNA and qRT-PCR were employed to uncover the genes associated with PGMS in 88-428BY under photoperiod stress. Eight candidate genes demonstrated a strong correlation with PGMS in soybean, enhancing our understanding of its genetic foundation and paving the way for future functional validation studies. This research establishes a genetic basis for developing new photoperiod-sensitive nuclear sterile materials in soybean, further utilization of these candidate genes in soybean breeding would provide important gene materials for hybrid soybean varieties cultivation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11314-5.

Supplementary Material 1. Fig. S1. Boxplot for FPKM values of all samples.

Supplementary Material 2. Fig. S2. Kernel plot of the overall expression density.

Supplementary Material 3. Fig. S3. Analysis of network topology for different soft–thresholding powers of 88-428BY in different samples. The x-axis represents the soft-thresholding power. The y-axis represents the scale-free topology model fit index (a) and the mean connectivity (b).

Supplementary Material 4. Fig. S4. Dendrogram of co-expression modules identified by WGCNA.

Supplementary Material 5. Table S1. Makers for qRT-PCR and their related information.

Supplementary Material 6. Table S2.Reference genome comparison read statistics for all samples.

Supplementary Material 7. Table S3. Detailed subcategory GO enrichment analysis SD-S4 vs. LD-S4.

Supplementary Material 8. Table S4. Detailed subcategory GO enrichment analysis SD-S8 vs. LD-S8.

Supplementary Material 9. Table S5. Detailed subcategory GO enrichment analysis SD-S12 vs. LD-S12.

Supplementary Material 10. Table S6. A total of 244 genes were discovered and subsequently defined as hub genes in these four modules.

Supplementary Material 11. Table S7. The 28 cloned male sterility-related genes collected from crops in the reported studies.

Supplementary Material 12. Table S8. Expression of cloned genes in the transcriptome.

Acknowledgements
The authors would like to acknowledge Professor Baoguo Wei for providing the soybean resource 88-428BY used in this study.

Authors' contributions
J.Z. and Y.Y. conceived and designed research; S.H., L.X., M.W., Z.B. and B.W. performed the sample collection; Y.Y., J.Z., T.Y., B.Z., H.Z., S.C. and R.Z. collected and analyzed the data; L.W. and Y.Y. wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding
This work was supported by the Major Science and Technology Special Program of Shanxi Province (No.202201140601025), the Natural Science Foundation of Shanxi Province (No.202303021211085), the Science and Technology Innovation Project of Colleges and Universities in Shanxi Province (No. 2021L091), the National Agricultural Science and Technology Major Projects (No.NK202308010403), the China Postdoctoral Science Foundation, Young Scholars of Shanxi Province (No.20210302124152), the Central Guide Local Science and Technology Development Special Project Fund (No. YDZJSX2022C020), and the Modern Agro-industry Technology Research System, Science and Technology Innovation Young Talent Team of Shanxi Province (No.202204051001020).

Data availability
The original transcriptome sequence data in this publication have been submitted to the National Center for Biotechnology Information under accession number CRA017761.

Declarations

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals. The collection materials of the plants, complies the relevant institutional, national, and international guidelines and legislation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Center of Agricultural Genetic Resources Research, Shanxi Agricultural University, Taiyuan, Shanxi 030031, China. [2]Houji laboratory in Shanxi Province, Shanxi Agricultural University, Taiyuan, Shanxi 030031, China. [3]College of Agricultural, Shanxi Agricultural University, Taigu 030801, China. [4]Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205, China.

Received: 17 July 2024 Accepted: 31 January 2025
Published online: 11 February 2025

References

1. Du H, Fang C, Li Y, Kong F, Liu B. Understandings and future challenges in soybean functional genomics and molecular breeding. J Integr Plant Biol. 2023;65(2):468–95.
2. Zhang M, Liu S, Wang Z, Yuan Y, Zhang Z, Liang Q, Yang X, Duan Z, Liu Y, Kong F, et al. Progress in soybean functional genomics over the past decade. Plant Biotechnol J. 2022;20(2):256–82.
3. Li MW, Wang Z, Jiang B, Kaga A, Wong FL, Zhang G, Han T, Chung G, Nguyen H, Lam HM. Impacts of genomic research on soybean improvement in East Asia. Theor Appl Genet. 2020;133(5):1655–78.
4. Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. Science. 2010;327(5967):818–22.
5. Rice heterosis. Quantitatively characterized and optimized hybrid breeding. Nat Genet. 2023;55(10):1619–20.
6. Wang B, Hou M, Shi J, Ku L, Song W, Li C, Ning Q, Li X, Li C, Zhao B, et al. Author correction: De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. Nat Genet. 2023;55(2):355.
7. Chen L, Liu YG. Male sterility and fertility restoration in crops. Annu Rev Plant Biol. 2014;65:579–606.
8. Ding X, Guo J, Lv M, Wang H, Sheng Y, Liu Y, Gai J, Yang S. The miR156b-GmSPL2b module mediates male fertility regulation of cytoplasmic male sterility-based restorer line under high-temperature stress in soybean. Plant Biotechnol J. 2023;21(8):1542–59.

9. Wang T, He T, Ding X, Zhang Q, Yang L, Nie Z, Zhao T, Gai J, Yang S. Confirmation of GmPPR576 as a fertility restorer gene of cytoplasmic male sterility in soybean. J Exp Bot. 2021;72(22):7729–42.
10. Yu J, Zhao G, Li W, Zhang Y, Wang P, Fu A, Zhao L, Zhang C, Xu M. A single nucleotide polymorphism in an R2R3 MYB transcription factor gene triggers the male sterility in soybean ms6 (Ames1). Theor Appl Genet. 2021;134(11):3661–74.
11. Yang Y, Speth BD, Boonyoo N, Baumert E, Atkinson TR, Palmer RG, Sandhu D. Molecular mapping of three male-sterile, female-fertile mutants and generation of a comprehensive map of all known male sterility genes in soybean. Genome. 2014;57(3):155–60.
12. Thu SW, Rai KM, Sandhu D, Rajangam A, Balasubramanian VK, Palmer RG, Mendu V. Mutation in a PHD-finger protein MS4 causes male sterility in soybean. BMC Plant Biol. 2019;19(1):378.
13. Nie Z, Zhao T, Liu M, Dai J, He T, Lyu D, Zhao J, Yang S, Gai J. Molecular mapping of a novel male-sterile gene ms(NJ) in soybean [*Glycine max* (L.) Merr]. Plant Reprod. 2019;32(4):371–80.
14. Zhao Q, Tong Y, Yang C, Yang Y, Zhang M. Identification and mapping of a new soybean male-sterile gene, mst-M. Front Plant Sci. 2019;10: 94.
15. Gautam R, Shukla P, Kirti PB. Male sterility in plants: an overview of advancements from natural CMS to genetically manipulated systems for hybrid seed production. Theor Appl Genet. 2023;136(9):195.
16. Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY. Progress in research and development on hybrid rice: a super-domesticate in China. Ann Bot-London. 2007;100(5):959–66.
17. Fang X, Sun X, Yang X, Li Q, Lin C, Xu J, Gong W, Wang Y, Liu L, Zhao L, et al. MS1 is essential for male fertility by regulating the microsporocyte cell plate expansion in soybean. Sci China Life Sci. 2021;64(9):1533–45.
18. Nadeem M, Chen A, Hong H, Li D, Li J, Zhao D, Wang W, Wang X, Qiu L. GmMs1 encodes a kinesin-like protein essential for male fertility in soybean (*Glycine max* L). J Integr Plant Biol. 2021;63(6):1054–64.
19. Fang X, Sun Y, Li J, Li M, Zhang C. Male sterility and hybrid breeding in soybean. Mol Breed. 2023;43(6):47.
20. Hou J, Fan W, Ma R, Li B, Yuan Z, Huang W, Wu Y, Hu Q, Lin C, Zhao X, et al. MALE STERILITY 3 encodes a plant homeodomain-finger protein for male fertility in soybean. J Integr Plant Biol. 2022;64(5):1076–86.
21. Wei B. Preliminary report on the discovery of soybean light temperature sensitive male sterile lines. Crop Germplasm Resources (Chinese). 1991;03:12.
22. Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu Y-C, Lee PY, Truong MT, Beals TP, Goldberg RB. Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex Plant Reprod. 1999;11(6):297–322.
23. Walsh CA, Bolger JC, Byrne C, Cocchiglia S, Hao Y, Fagan A, Qin L, Cahalin A, McCartan D, McIlroy M, et al. Global gene repression by the steroid receptor coactivator SRC-1 promotes oncogenesis. Cancer Res. 2014;74(9):2533–44.
24. Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017;24(11):1138–43.
25. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.
26. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40.
27. Smoot M, Ono K, Ideker T, Maere S. PiNGO: a Cytoscape plugin to find candidate genes in biological. Networks Bioinf. 2011;27(7):1030–1.
28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9: 559.
29. Chen X, Mei Q, Liang W, Sun J, Wang X, Zhou J, Wang J, Zhou Y, Zheng B, Yang Y, et al. Gene Mapping, Genome-Wide Transcriptome Analysis, and WGCNA reveals the molecular mechanism for triggering programmed cell death in Rice Mutant pir1. Plants. 2020;9(11):1607.
30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25(4):402–8.
31. Sharmin RA, Bhuiyan MR, Lv W, Yu Z, Zhao TJ. RNA-Seq based transcriptomic analysis revealed genes associated with seed-flooding tolerance in wild soybean (Glycine soja Sieb. & Zucc). Environ Exp Bot. 2020;171:103906.
32. Wu C, Ma Q, Yam KM, Cheung MY, Xu Y, Han T, Lam HM, Chong K. In situ expression of the GmNMH7 gene is photoperiod-dependent in a unique soybean (*Glycine max* [L.] Merr.) Flowering reversion system. Planta. 2006;223(4):725–35.
33. Huang F, Xu G, Chi Y, Liu H, Xue Q, Zhao T, Gai J, Yu D. A soybean MADS-box protein modulates floral organ numbers, petal identity and sterility. BMC Plant Biol. 2014;14: 89.
34. Jia T, Zhang K, Li F, Huang Y, Fan M, Huang T. The AtMYB2 inhibits the formation of axillary meristem in Arabidopsis by repressing RAX1 gene under environmental stresses. Plant Cell Rep. 2020;39(12):1755–65.
35. Guo Y, Gan S. AtMYB2 regulates whole plant senescence by inhibiting cytokinin-mediated branching at late stages of development in Arabidopsis. Plant Physiol. 2011;156(3):1612–9.
36. Shi QS, Lou Y, Shen SY, Wang SH, Zhou L, Wang JJ, Liu XL, Xiong SX, Han Y, Zhou HS, et al. A cellular mechanism underlying the restoration of thermo/photoperiod-sensitive genic male sterility. Mol Plant. 2021;14(12):2104–14.
37. Wang KQ, Yu YH, Jia XL, Zhou SD, Zhang F, Zhao X, Zhai MY, Gong Y, Lu JY, Guo Y, et al. Delayed callose degradation restores the fertility of multiple P/TGMS lines in Arabidopsis. J Integr Plant Biol. 2022;64(3):717–30.
38. Wang X, Wang J, Liu Z, Yang X, Chen X, Zhang L, Song X. The R2R3 MYB gene TaMYB305 positively regulates anther and pollen development in thermo-sensitive male-sterility wheat with Aegilops Kotschyi cytoplasm. Planta. 2024;259(3):64.
39. Murai K, Takumi S, Koga H, Ogihara Y. Pistillody, homeotic transformation of stamens into pistil-like structures, caused by nuclear-cytoplasm interaction in wheat. Plant J. 2002;29(2):169–81.
40. Shi C, Zou W, Zhu Y, Zhang J, Teng C, Wei H, He H, He W, Liu X, Zhang B, et al. mRNA cleavage by 21-nucleotide phasiRNAs determines temperature-sensitive male sterility in rice. Plant Physiol. 2024;194(4):2354–71.
41. Shi ZH, Zhang C, Xu XF, Zhu J, Zhou Q, Ma LJ, Niu J, Yang ZN. Overexpression of AtTTP affects ARF17 expression and leads to male sterility in Arabidopsis. PLoS ONE. 2015;10(3): e0117317.
42. Zhang C, Xu T, Ren MY, Zhu J, Shi QS, Zhang YF, Qi YW, Huang MJ, Song L, Xu P, et al. Slow development restores the fertility of photoperiod-sensitive male-sterile plant lines. Plant Physiol. 2020;184(2):923–32.
43. Zhang H, Xu C, He Y, Zong J, Yang X, Si H, Sun Z, Hu J, Liang W, Zhang D. Mutation in CSA creates a new photoperiod-sensitive genic male sterile line applicable for hybrid rice seed production. P Natl Acad Sci USA. 2013;110(1):76–81.
44. Teng C, Zhang H, Hammond R, Huang K, Meyers BC, Walbot V. Dicer-like 5 deficiency confers temperature-sensitive male sterility in maize. Nat Commun. 2020;11(1):2912.
45. Fernández-Gómez J, Talle B, Wilson ZA. Increased expression of the MALE STERILITY1 transcription factor gene results in temperature-sensitive male sterility in barley. J Exp Bot. 2020;71(20):6328–39.
46. Xue Z, Xu X, Zhou Y, Wang X, Zhang Y, Liu D, Zhao B, Duan L, Qi X. Deficiency of a triterpene pathway results in humidity-sensitive genic male sterility in rice. Nat Commun. 2018;9(1):604.
47. Han Y, Jiang SZ, Zhong X, Chen X, Ma CK, Yang YM, Mao YC, Zhou SD, Zhou L, Zhang YF, et al. Low temperature compensates for defective tapetum initiation to restore the fertility of the novel TGMS line ostms15. Plant Biotechnol J. 2023;21(8):1659–70.
48. Zhou L, Mao YC, Yang YM, Wang JJ, Zhong X, Han Y, Zhang YF, Shi QS, Huang XH, Meyers BC, et al. Temperature and light reverse the fertility of rice P/TGMS line ostms19 via reactive oxygen species homeostasis. Plant Biotechnol J. 2024;22(7):2020–32.
49. Fan Y, Yang J, Mathioni SM, Yu J, Shen J, Yang X, Wang L, Zhang Q, Cai Z, Xu C, et al. PMS1T, producing phased small-interfering RNAs, regulates photoperiod-sensitive male sterility in rice. P Natl Acad Sci USA. 2016;113(52):15144–9.
50. Barman HN, Sheng Z, Fiaz S, Zhong M, Wu Y, Cai Y, Wang W, Jiao G, Tang S, Wei X, et al. Generation of a new thermo-sensitive genic male sterile rice line by targeted mutagenesis of TMS5 gene through CRISPR/Cas9 system. BMC Plant Biol. 2019;19(1):109.
51. Zhang YF, Li YL, Zhong X, Wang JJ, Zhou L, Han Y, Li DD, Wang N, Huang XH, Zhu J, et al. Mutation of glucose-methanol-choline oxidoreductase leads to thermosensitive genic male sterility in rice and Arabidopsis. Plant Biotechnol J. 2022;20(10):2023–35.
52. Wu S, Yu Z, Wang F, Li W, Yang Q, Ye C, Sun Y, Jin D, Zhao J, Wang B. Identification and characterization of a novel adenine phosphoribosyltransferase gene (ZmAPT2) from maize (Zea mays L). DNA Seq. 2008;19(3):357–65.

53. Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. P Natl Acad Sci USA. 2012;109(7):2654–9.
54. Harmer SL. The circadian system in higher plants. Annu Rev Plant Biol. 2009;60:357–77.
55. López-Juez E, Dillon E, Magyar Z, Khan S, Hazeldine S, de Jager SM, Murray JA, Beemster GT, Bögre L, Shanahan H. Distinct light-initiated gene expression and cell cycle programs in the shoot apex and cotyledons of Arabidopsis. Plant Cell. 2008;20(4):947–68.
56. Izawa T, Mihara M, Suzuki Y, Gupta M, Itoh H, Nagano AJ, Motoyama R, Sawada Y, Yano M, Hirai MY, et al. Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. Plant Cell. 2011;23(5):1741–55.
57. Zheng S, Hu H, Ren H, Yang Z, Qiu Q, Qi W, Liu X, Chen X, Cui X, Li S, et al. The Arabidopsis H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. Nat Commun. 2019;10(1):1303.
58. Bandillo NB, Anderson JE, Kantar MB, Stupar RM, Specht JE, Graef GL, Lorenz AJ. Dissecting the genetic basis of local adaptation in soybean. Sci Rep-UK. 2017;7(1):17195.
59. Chen L, Cai Y, Qu M, Wang L, Sun H, Jiang B, Wu T, Liu L, Sun S, Wu C, et al. Soybean adaption to high-latitude regions is associated with natural variations of GmFT2b, an ortholog of FLOWERING LOCUS T. Plant Cell Environ. 2020;43(4):934–44.
60. Schilling S, Pan S, Kennedy A, Melzer R. MADS-box genes and crop domestication: the jack of all traits. J Exp Bot. 2018;69(7):1447–69.
61. Theissen G. Development of floral organ identity: stories from the MADS House. Curr Opin Plant Biol. 2001;4(1):75–85.
62. Tsaftaris A, Pasentsis K, Makris A, Darzentas N, Polidoros A, Kalivas A, Argiriou A. The study of the E-class SEPALLATA3-like MADS-box genes in wild-type and mutant flowers of cultivated saffron crocus (*Crocus sativus* L.) and its putative progenitors. J Plant Physiol. 2011;168(14):1675–84.
63. Immink RG, Gadella TW Jr, Ferrario S, Busscher M, Angenent GC. Analysis of MADS box protein-protein interactions in living plant cells. P Natl Acad Sci USA. 2002;99(4):2416–21.
64. Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K. Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell. 1997;9(10):1859–68.
65. Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K. An Arabidopsis myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell. 1993;5(11):1529–39.
66. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, et al. The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol. 2006;60(1):107–24.

Publisher's Note