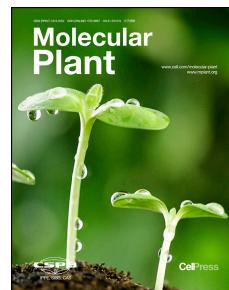


Journal Pre-proof

A Domestication-Associated Gene *GmPRR3b* Regulates Circadian Clock and Flowering Time in Soybean

Cong Li, Ying-hui Li, Yanfei Li, Hongfeng Lu, Huilong Hong, Tian Yu, Hongyu Li, Tao Zhao, Xiaowei Zhou, Jun Liu, Xianan Zhou, Scott A. Jackson, Bin Liu, Li-juan Qiu



PII: S1674-2052(20)30030-7
DOI: <https://doi.org/10.1016/j.molp.2020.01.014>
Reference: MOLP 888

To appear in: MOLECULAR PLANT
Accepted Date: 29 January 2020

Please cite this article as: **Li C., Li Y.-h., Li Y., Lu H., Hong H., Yu T., Li H., Zhao T., Zhou X., Liu J., Zhou X., Jackson S.A., Liu B., and Qiu L.-j.** (2020). A Domestication-Associated Gene *GmPRR3b* Regulates Circadian Clock and Flowering Time in Soybean. Mol. Plant. doi: <https://doi.org/10.1016/j.molp.2020.01.014>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

All studies published in MOLECULAR PLANT are embargoed until 3PM ET of the day they are published as corrected proofs on-line. Studies cannot be publicized as accepted manuscripts or uncorrected proofs.

A Domestication-Associated Gene *GmPRR3b* Regulates Circadian Clock and Flowering Time in Soybean

Cong Li^{1,5}, Ying-hui Li^{1,5}, Yanfei Li^{1,5}, Hongfeng Lu², Hui long Hong¹, Tian Yu¹, Hongyu Li¹, Tao Zhao¹, Xiaowei Zhou², Jun Liu¹, Xinan Zhou³, Scott A Jackson⁴, Bin Liu^{1*}, Li-juan Qiu^{1*}

¹The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, P.R. China

²Novogene Bioinformatics Institute, Beijing, P.R. China.

³Key Laboratory of Oil Crop Biology (MOA), Oil Crops Research Institute of Chinese Academy of Agriculture Sciences, Wuhan, China

⁴Center for Applied Genetic Technologies, University of Georgia, Athens, Georgia, USA.

⁵These authors contributed equally to this article.

*Correspondence: Li-juan Qiu (qiulijuan@caas.cn), Bin Liu (liubin05@caas.cn)

SHORT SUMMARY

Natural variation in flowering time is critical for geographic adaption of soybean. A circadian clock gene *GmPRR3b* was identified as a major flowering time QTL through genome-wide association study (GWAS). Its haplotype H6 associated with early flowering and vigorous growth traits has been extensively utilized during domestication and breeding of modern cultivars.

ABSTRACT

Improved soybean cultivars have been adapted to grow at a wide range of latitudes, enabling expansion of cultivation worldwide. However, the genetic basis for this broad adaptation is still not clear. Here we report the identification of *GmPRR3b* as a major flowering time regulatory gene that has been selected during domestication and genetic improvement for geographic expansion. Through genome-wide association study (GWAS) of a diverse landrace panel of soybean (comprised of 279 accessions), we identified 16 candidate quantitative loci associated with flowering time and maturity time. The strongest signal resides on the known flowering gene *E2*, verify the effectiveness of our approach. We detected strong association signals of a genomic region containing *GmPRR3b* with both flowering and maturity time. Haplotype analysis revealed that *GmPRR3b*^{H6} is the major form that has been utilized during recent breeding of modern cultivars. mRNA profiling analysis showed that *GmPRR3b*^{H6} displays a rhythmic and photoperiod-dependent expression that is preferentially induced under long day conditions. Overexpression of *GmPRR3b*^{H6} conferred increased main stem node number and high yield phenotypes, while knockout of *GmPRR3b*^{H6} by CRISPR-Cas9 technology retarded growth and delayed floral transition. *GmPRR3b*^{H6}

appears to act as a transcriptional repressor of multiple predicted circadian clock genes, including *GmCCA1a* which directly upregulates *J/GmELF3a* to modulate flowering time. Taken together, the causal SNP (Chr12:5520945) likely endows *GmPRR3b^{H6}* a moderate but appropriate activity, which is associated with early flowering and vigorous growth traits preferentially selected for broad adaption in landraces and improved cultivars.

Key words: flowering time, soybean landrace, domestication, genome-wide association study, *GmPRR3b*, circadian clock

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) is a legume crop providing plant oil and protein for humans and livestock. Archaeological and evolutionary studies have shown that soybean was domesticated from its progenitor (*Glycine soja* Sieb. & Zucc.) approximately 5,000 years ago in East Asia (Lee et al., 2011). A suite of domestication-related traits, including flowering time, were selected in traditional landraces and the breeding of improved cultivars (Wang et al., 2018a). Wild soybean is a typical short-day plant that flowers early when the day length is shorter than a certain threshold. Day length-insensitive flowering has been an essential trait for the cultivation of soybean and its geographical spread to lower and/or higher latitudes (Sedivy et al., 2017).

Previous population genetics analyses identified a number of genes as putative targets of selection for changes in flowering time during domestication, including at least ten flowering loci, designated *E1* to *E9* and *J* (Lam et al., 2010; Li et al., 2019c; Li et al., 2014; Possart et al., 2014; Wu et al., 2017; Zhang et al., 2017). Among them, *E1* encodes a legume-specific flowering repressor which inhibits flowering under long day conditions; *E2* encodes a homolog of *Arabidopsis* GIGANTEA (GI); both *E3* and *E4* encode homologs of PHYTOCHROME A (PhyA); *E9* encodes a homolog of FLOWERING LOCUS T (FT); and *J* encodes a homolog of EARLY FLOWERING 3 (ELF3), a key component of the evening complex (Cober et al., 2010; Gould et al., 2013; Kong et al., 2014; Lu et al., 2017). However, genetic differentiation analyses of these genes between *G. soja* and landraces suggest that only *E2* was associated with domestication (Wang et al., 2016). The causal alleles of the other genes likely occurred after the separation of *G. max* and *G. soja* and enabled local adaptation to new climatic regions, e.g. *E1* and *E4* for higher-latitude regions (Jiang et al., 2014; Tsubokura et al., 2013), and *J* and *E9* (*GmFT2a*) for lower-latitude tropical regions (Lu et al., 2017; Zhao et al., 2016).

Plants rely on an endogenous circadian clock system to coordinate their growth, development, defense, and metabolism to daily and seasonal environmental changes. Previous studies in the model plant *Arabidopsis thaliana* showed that the circadian clock is a self-sustaining timekeeping mechanism consisting of multiple interlocking feedback loops

(Hsu and Harmer, 2014). The morning loop comprises morning-expressed *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *LATE ELONGATION HYPOCOTYL* (*LHY*), *PSEUDORESPONSE REGULATOR 9* (*PRR9*) and *PRR7*. The evening loop comprises evening-expressed *TIMING OF CAB EXPRESSION 1* (*TOC1*)/*PRR1*, *EARLY FLOWERING 3* (*ELF3*), *ELF4* and *LUX ARRHYTHMO* (*LUX*). These two loops are connected by the central oscillator loop consisting of *CCA1*, *LHY*, and *TOC1* (Sanchez and Kay, 2016). Most of these components in *Arabidopsis* are transcriptional repressors, underlying a ‘repressor’ model for the oscillating machine (McClung, 2019).

PRRs (designated as APRRs in *Arabidopsis* and including *APRR9*, *APRR7*, *APRR5*, *APRR3*, and *APRR1/TOC1*) are key components of the plant circadian clock. PRRs belong to a protein family containing a pseudo receiver (PR) domain at the N-terminal and a CCT domain (named after the presence in CONSTANS (CO), CO-like and TOC1) at the C-terminal. The PR domain is essential for protein-protein interactions. Most PRR proteins contain an EAR-like motif that recruits TOPLESS (TPL) and histone deacetylases to repress target gene expression by epigenetic modifications (Wang et al., 2013; Yang et al., 2018). PRRs regulate many physiological processes in plants. *Arabidopsis* mutants in the *PRR* gene family show a wide range of defects in the circadian clock, photoperiodic flowering, growth, and stress responses (Hayama et al., 2017; Li et al., 2019a). In parallel to the model plant *Arabidopsis thaliana*, numerous studies have identified distinct natural variation in *PRR* homologs in various crops that were selected for adaptation to different regions of cultivation (Klein et al., 2015; Koo et al., 2013; Liu et al., 2015; Murphy et al., 2011). Here, we report *GmPRR3b* as an essential flowering quantitative trait locus (QTL), which has been selected during soybean domestication and facilitated the widespread adaptation of soybean.

RESULTS

Identification of flowering time QTLs by GWAS in soybean landraces

In our previous GWAS and genome-scan study using 1,938 Chinese soybean landraces, we found that two flowering time-associated variants on chromosome 12, Chr12:5470311 and Chr12:5914898, exhibit strong selection signatures during domestication (Li et al., 2019e). The genomic regions covering these two domestication variants are still very large (92.9 kb and 73.5 kb, respectively), due to the limited density of SNPs inferred from the tGBS® approach (Ott et al., 2017). To identify new regulatory genes of flowering time in soybean, we selected a genetically representative panel of 279 diverse accessions from 1,938 landraces (Supplemental Table 1) for whole-genome re-sequencing analysis (with an average depth of 6.03× per genome and covering 95.3% of the soybean reference genome *Glycine max* Wm82.a2.v1 (<https://phytozome.jgi.doe.gov/>). These landraces were collected from sites

across China, ranging from 19.4 to 51.4 °N latitude, 85.6 to 132.1 °E longitude. We obtained a total of 1,770 Gb of clean sequence. In addition, we downloaded re-sequencing data for 106 soybean accessions (58 *G. soja* accessions and 48 improved cultivars) from the NCBI Short Read Archive (accession number: SRA045129) published previously (Zhou et al., 2015b). After calibrating SNP calling quality using all 385 genomes and discarding SNPs with missing rates > 0.4 and a minor allele frequency ≤ 0.01, a dataset of 8,624,466 high-quality SNPs was recovered. Among these, 73.5% (6,341,742 SNPs) were observed for soybean landraces.

Using these 6,341,742 SNPs and phenotype data collected in Beijing (40.1 °N) and Wuhan (30.5 °N) for 279 soybean landraces, GWAS analyses were performed for flowering time and maturity time using Compressed Mixed Linear Model (cMLM). We detected 16 candidate QTL regions for flowering time (using $r^2 \geq 0.8$) with association signals exceeding a significant threshold ($-\log_{10} P \geq 7$) in both test sites (Beijing and Wuhan) (Supplemental Figure 1 and Supplemental Table 2). Notably, the strongest association was represented by the peak SNP Chr10:45310798, which is known to be the causal variant of flowering gene *E2* (Wang et al., 2016). Both the second strongest signal (Chr12:5483364) in Beijing and the top signal in Wuhan (Chr12:5520945) for flowering time are located on Chromosome 12 (Figure 1A and Supplemental Figure 1). GWAS analyses for maturity time in the Wuhan site also identified a strong signal (Chr12:5520945) on chromosome 12 ($P = 3.9E - 09$) (Figure 1B and Supplemental Figure 2). These results are consistent with our earlier report that the region between Chr12:5470311 and Chr12:5914898 on chromosome 12 contains important flowering time regulatory gene (Li et al., 2019e) and further narrow down the region to a likely region between Chr12:5483364 and Chr12:5520945.

Linkage disequilibrium block structure analysis indicated that Chr12:5483364 and Chr12:5520945 belonged to the same genetic locus ($r^2 \geq 0.8$) and covered a 59.4-kb candidate genomic region (from 5,483,364 bp to 5,542,737 bp) (Supplemental Table 2). This region contains three annotated genes, *Glyma.12G073800*, *Glyma.12G073900* and *Glyma.12G07400*. Of these, *Glyma.12G073900* is an ortholog of *Arabidopsis PSEUDO-RESPONSE REGULATORS3 (AtPRR3)* and thus referred to as *GmPRR3b* (Li et al., 2019b; Para et al., 2007). Further analysis of the linkage disequilibrium block surrounding Chr12:5520945 using *HAPLOVIEW* (Barrett et al., 2005) narrowed 59.4-kb down to a 24.3-kb genomic region (5,505,719 bp to 5,530,042 bp) (Figure 1C) containing one annotated gene, *GmPRR3b*. We observed no variants resulting in amino acid changes in *Glyma.12G073800*, and one non-synonymous SNP in *Glyma.12G07400* which exhibited no significant association with flowering time. These observations indicated that *GmPRR3b* was the most likely causal gene for flowering time.

The association signal Chr12:5520945 was located in the stop codon of *GmPRR3b* in the

Williams 82 reference genome. Flowering and maturity time, when stratified by genotypes at Chr12:5520945 (T/C) exhibited a significant difference among soybean landraces (Supplemental Figure 3). The average flowering time of landraces carrying the Chr12:5520945_TT genotype was significantly earlier ($P < 0.01$) than that of landraces carrying the alternate genotype by 12.3 days in the Beijing field test site and 6.5 days in the Wuhan test site (Supplemental Figure 3). These observations reaffirmed that Chr12:5520945 as a causal variant for the observed flowering time change.

GmPRR3b haplotype 6 was selected during domestication and utilized by modern breeders

To understand the evolutionary process for *GmPRR3b* during domestication, we investigated the selection signal in the genomic region harboring *GmPRR3b* based on pairwise analyses of F_{ST} and the ratio of nucleotide diversity (θ_π) between three evolutionary populations: 58 wild accessions, 279 landraces; and 48 improved cultivars (Supplemental Figure 4). The results from F_{ST} and the ratio of θ_π demonstrated that *GmPRR3b* is located in a selective sweep region showing evidence of selection during both soybean domestication and subsequent genetic improvement.

Six variants which resulted in amino acids changed were detected in the coding region of *GmPRR3b* across our panel of 385 soybean accessions, including four non-synonymous variants and two variants (Chr12:5509061 and Chr12:5520945) which caused stop-codon changes and subsequently resulted in the elongation or truncation of the peptide (Figure 1D and 1E). Thus, a total of eight *GmPRR3b* haplotypes (hereafter referred to as H1 to H8) were detected (Figure 1D). H1 carrying Chr12:5509061 and H6 carrying Chr12:5520945 encoded two truncations of the peptide, compared to other haplotypes (Figure 1E). H6 exhibited the highest frequency (63.4%), followed by H4 (12.2%). Whereas H1-H3 occurred at relatively lower frequencies, ranging from 1% to 1.3% (Figure 2A), which were distinguished from the other haplotypes by Chr12:5509061, Chr12:5519426, and Chr12:5520716 alleles in that they were specific to the *G. soja* population. The no. of haplotypes decreased from eight in *G. soja* to five in landraces then to one in improved cultivars (Figure 2A), suggesting that *GmPRR3b* was strongly selected by human during domestication and genetic improvement.

To evaluate the phenotypic effects of the different haplotypes, we analyzed differences in flowering time and maturity time in landraces planted in two test sites, Beijing and Wuhan. The landraces carrying H6-H8 exhibited significantly earlier flowering and maturity in Wuhan and significantly earlier flowering in Beijing than landraces carrying H4 and H5 (Figure 1F and Supplemental Figure 5). Two single nucleotide substitutions, one of which led to a single amino acid change (Chr12:5509317) and the other (Chr12:5520945) to a 99 amino acids change, distinguished early flowering H6-H8 from late flowering H4 and H5 (Figure 1F

and Supplemental Figure 5). This indicates that both Chr12:5509317 and Chr12:5520945 are likely causal genetic variants and that changes from late flowering to early flowering are associated with the independent polymorphisms, Chr12:5509317 and Chr12:5520945. The observations that the frequencies of H6, H7 and H8 were largely increased from 3.4%, 1.7% and 1.7% in *G. soja* to 69.4%, 7.5% and 10.8% in landraces respectively (Figure 2A), signified that early flowering alleles in these two causal genetic variants (Chr12:5509317, and Chr12:5520945 for) were derived from standing variations in the wild progenitor *G. soja*. Median-joining network analysis indicates that H6/H7 and H8 are closely related to H4 and H5, respectively (Figure 2B).

The geographical distribution of the *GmPRR3b* haplotypes was surveyed in the 383 soybean accessions collected from East Asia, the native area of soybean (Figure 2C). Interestingly, 13 of 14 *G. soja* accessions carrying H1, H2, or H3 were collected from high latitude areas in northeast China and Russia. The *G. soja* accessions carrying H4 and H5 were collected all over East Asia, whereas the landraces carrying H4 and H5 were mainly from south China, suggesting these two late-flowering haplotypes were selected in south China for longer maturity which confers higher yields in the short-day conditions (Figure 2C). Only four *G. soja* accessions carried H6, H7 or H8, one from Japan (PI366121), two from South Korea (PI407246 and PI407275) and one from China (PI549046). After domestication, the *G. max* accessions carrying H6 were distributed all over China from low to high latitudes, whereas those carrying H7 and H8 were mainly distributed in central China. Only H6 was utilized by modern breeders during soybean genetic improvement; all 48 improved cultivars studied here carried only H6 (Figure 2A). Therefore, we focused on the functional mechanism of the *GmPRR3b* haplotype 6 (hereafter referred to as *GmPRR3b^{H6}*).

Transcript profiles and subcellular localization of GmPRR3b^{H6}

To interrogate the function and underlying molecular mechanism of *GmPRR3b^{H6}* in flowering time regulation, we performed quantitative RT-PCR (qRT-PCR) to check the transcript profiles of *GmPRR3b^{H6}* in cultivar Williams 82. The results showed that its mRNA was highly expressed in unifoliolate and trifoliolate leaves, which are the tissues important for the perception of ambient light signals including photoperiod (Figure 3A). In addition, *GmPRR3b^{H6}* mRNA exhibited a rhythmic expression pattern peaking at the middle of the day under diurnal and consequent free-running conditions, implying *GmPRR3b* is associated with circadian clock regulation (Figure 3B). Its overall expression level was markedly induced under long day conditions as compared to short day photoperiods, reminiscent of *E1*, which acts as the key flowering repressor in soybean (Lu et al., 2017; Xia et al., 2012).

Next, we analyzed the subcellular localization of GmPRR3b. The SNP Chr12:5520945 results in a premature termination codon (CAA to TAA) in *GmPRR3b^{H6}*, which leads to a

truncated protein absent the C-terminal CCT domain, in comparison to $GmPRR3b^{H4}$ (Figure 3C). Two putative nuclear localization signals (NLS1 and NLS2) were identified in the full-length GmPRR3b protein by the LOCALIZER program (Sperschneider et al., 2017). NLS2 is located in the CCT domain, suggesting that the CCT domain may be required for the nuclear localization of GmPRR3b. To test this possibility, we made H4-YFP and H6-YFP ($GmPRR3b^{H4}$ and $GmPRR3b^{H6}$ fused with the yellow fluorescent protein, respectively) constructs and transformed them into soybean mesophyll protoplasts (Figure 3D). Confocal microscopy showed that nearly 95% of the protoplasts that expressing H4-YFP and 75% of protoplasts that expressing H6-YFP had fluorescent signals restricted to the nucleus (Figure 3E), demonstrating that absence of the CCT domain reduces but does not abolish the nuclear localization of $GmPRR3b^{H6}$. To further test if NLS1 is required for the nuclear localization, we made an $H6^m$ -YFP construct by site-direct mutation of NLS1 (LGEV to PADF) (Figure 3C). Mutation of NLS1 led to a proportion of protoplasts with nuclear fluorescent signals decreased to less than 20% (Figure 3E), suggesting that NLS1 and NLS2 are functionally additive for the nuclear localization of GmPRR3b, and the absence of NLS2 in $GmPRR3b^{H6}$ may reduce but not abolish its biological activity.

Overexpression of $GmPRR3b^{H6}$ dampens the circadian clock

To address whether $GmPRR3b^{H6}$ is biologically functional, we made the $35S::YFP-H6$ and $35S::H6-YFP$ overexpression constructs and obtained multiple transgenic lines which were verified by immunoblotting probed with the anti-GFP antibody (Supplemental Figure 6A and B, two representative lines were shown). Phenotypic analysis showed that both the $YFP-H6$ and $H6-YFP$ lines flowered significantly later than wild-type (WT, cultivar Tianlong 1 harboring $GmPRR3b^{H6}$) plants under natural conditions (ND, summer in Beijing), as well as under long day (LD) and short day (SD) conditions in growth chambers (Supplemental Figure 6C), suggesting that overexpression of $GmPRR3b^{H6}$ may delay flowering via a photoperiod-independent pathway. In addition, the transgenic lines grown in the field were different from the WT controls in several aspects including elongated flowering period, increased node and grain number, and enhanced grain yield per plant (Supplemental Figure 7), again indicating why there may have been preferential selection of $GmPRR3b^{H6}$ during the breeding of elite cultivars. As the $YFP-H6$ and $H6-YFP$ overexpression lines showed similar phenotypes, only the $YFP-H6$ line was used in subsequent experiments.

To gain insights into how overexpression of $GmPRR3b^{H6}$ conferred the later flowering phenotype, we performed a time-course expression analysis of several well-documented flowering or circadian clock-associated genes in soybean under LD conditions following a free-running regimen of continuous light. The qRT-PCR results indicated that the transcriptional levels of $GmFT2a$ and $GmFT5a$, both encoding florigens in soybean (Nan et

al., 2014), were decreased in the *YFP-H6* line as compared with the WT. Consistent with this, the mRNA level of *E1*, a key repressor of flowering in soybean (Jiang et al., 2014), was upregulated in the *YFP-H6* line (Supplemental Figure 8). Interestingly, the mRNA levels of multiple circadian clock genes including *GmCCA1s*, *GmPRR3a*, *GmPRR5a*, *GmELF3s*, *GmELF4a*, *GmLUX1* and *GmLUX2* decreased to varying degrees in the *YFP-H6* line (Figure 4A and B; Supplemental Figure 9-11), suggesting that *GmPRR3b^{H6}* may act as a general transcriptional repressor in the central circadian oscillator machine. Consistent with this notion, overexpression of *YFP-H6* dramatically dampedened the rhythmic expression amplitude of *GmCAB1*(Figure 4C), which serves as an output marker of the plant circadian clock, reinforcing the notion that *GmPRR3b^{H6}* is a key component of the circadian oscillator and exerts pleiotropic effects on agronomic traits in soybean.

GmPRR3b^{H6}* is a transcriptional repressor of *GmCCA1a

Given that *GmCCA1a* transcription was downregulated in the *YFP-H6* line, together with the fact that *GmPRR3b^{H6}* contains an EAR motif, we suspected that *GmPRR3b^{H6}* acts as a direct repressor of *GmCCA1a*. To test this hypothesis, we checked the DNA sequence of *GmCCA1a* and found several G-box, TBS and TGTG elements located in its promoter region (Gendron et al., 2012) (Figure 4D). Chromatin immunoprecipitation (ChIP) experiments were performed using the *YFP-H6* line and GFP-trap beads. The result showed that *GmPRR3b^{H6}* robustly bound to the 1, 2 and 3 sites which contain the G-box, TBS and TGTG elements, respectively, supporting that *GmCCA1a* is a direct target gene negatively regulated by *GmPRR3b^{H6}* (Figure 4E).

GmCCA1a* directly promotes the transcription of *J/ELF3a

Studies in *Arabidopsis* previously showed that CCA1 negatively regulates the transcription of evening complex genes including *ELF3*, *ELF4* and *LUX* (Lu et al., 2012; Nagel et al., 2015). However, we found that both *GmCCA1a* and *J/GmELF3a* were downregulated in the *YFP-H6* line as compared with WT (Figure 4A and 4B), suggesting that *GmCCA1a* might positively regulate the transcription of *J/GmELF3a* in soybean. To test this possibility, we established a Root Induced Callus Expression (RICE) system to analyze the effect of *GmCCA1a* on the transcription of *J/GmELF3a*. Briefly, gene expression construct harbored in *Agrobacterium tumefaciens* was transformed into the soybean hair roots, which were further induced to generate uniform callus through tissue culture. Thus, the regulating network can be investigated by analyzing the transcript profiles of putative downstream genes in the transformed callus cultivated under specific regimens.

To test the feasibility of the RICE system, we first checked the rhythmic expression of *GmPRR3b* and *GmCCA1a* genes in WT callus using qRT-PCR. Both genes displayed comparable rhythmic expression patterns in callus (Figure 5A and 5B) and in trifoliolate leaves

(Figure 3B and 4B). Then we transformed the *H6-YFP* construct into the soybean hairy root and generated *H6-YFP* overexpression callus. Transcriptional assays demonstrated that the expression of *GmCCA1a* was significantly downregulated in the *H6-YFP* overexpression callus as compared with that in WT callus (Figure 5B). Hence, similar results of *H6-YFP* repressing *GmCCA1a* transcription in both callus and trifoliolate leaves validated the feasibility of RICE system.

We then generated the *GmCCA1a-YFP* overexpression callus (Figure 5C) and confirmed that the *GmCCA1a* indeed could upregulate the transcription levels of *J/GmELF3a* as well as other putative evening complex genes in the RICE system (Figure 5D and Supplemental Figure 12). To further test if *GmCCA1a* directly regulates *J/GmELF3a* expression, we performed ChIP experiments using the *GmCCA1a-YFP* overexpression callus and GFP-trap beads. The result detected robust binding signals of *GmCCA1a-YFP* to the promoter region of *J* and other evening complex genes (Figure 5E and 5F; Supplemental Figure 13). Furthermore, we examined whether *GmCCA1a* might act as a transcriptional activator of *J* using a dual-luciferase assay in mesophyll protoplasts according to a previously reported procedure (Zhang et al., 2014). Briefly, the potential transcriptional activity of *GmCCA1a* was tested by its effect on the *firefly luciferase* (*LUC*) reporter gene driven by the promoter of *J/GmELF3a*, using the *Renilla reniformis luciferase* (*REN*) driven by the standard 35S promoter as the internal control (Figure 5G). The result showed that *GmCCA1a-YFP* protein significantly increased the expression of the reporter gene as compared with the YFP control protein (Figure 5H), supporting the inference that *GmCCA1a* acts as a transcriptional activator of *J/GmELF3a* in soybean. Taken together, these results indicate that overexpression of *GmPRR3b^{H6}* may cause late flowering phenotypes by repressing *GmCCA1a* and then *J/GmELF3a* to upregulate the expression of flowering repressor *E1*.

GmPRR3b^{H6}* has a moderate activity to repress *GmCCA1a* expression in comparison to *GmPRR3b^{H4}

Median-joining network analysis showed that *GmPRR3b* haplotype H6 is most likely derived from H4 (Figure 2B). To get an insight into how H6 is functionally different from H4, we tested the transcriptional levels of *GmPRR3b* in soybean accessions harboring H4 or H6 under both LD and SD conditions. Twenty accessions of H6 and ten accessions of H4 were selected for qRT-PCR analysis (Supplemental Table 4). The results showed that the expression levels of *GmPRR3b* tended to be lower in the H6 accessions than in the H4 accessions at three time points under both LD (ZT4, 8 and 12) and SD (ZT0, 4 and 8) conditions (Figure 6A and 6B and Supplemental Figure 14). In sharp contrast, the expression levels of *GmCCA1a* in the majority H6 accessions were higher than those in the H4 accessions. These observations are consistent with our previous observation that *GmPRR3b* and *GmCCA1a* reciprocally repress

the transcription of each other (Figure 4B and 5A).

To test if the causal SNP (Chr12:5520945) in the H6 accessions affects the *GmPRR3b* activity, we ectopically expressed *GmPRR3b^{H4}* and *GmPRR3b^{H6}* in *Arabidopsis*. The results showed that *GmPRR3b^{H4}* was more effective than *GmPRR3b^{H6}* to promote hypocotyl elongation and repress flowering in *Arabidopsis* (Supplemental Figure 15 and 16). Then we further compared their activity using the RICE system in soybean. We generated dozens of independent transgenic callus lines expressing comparable levels of *YFP-H4* or *YFP-H6* (Figure 6C, left panel). The results showed that the expression levels of *GmCCA1a* were more effectively repressed in the *YFP-H4* callus than in the *YFP-H6* callus at all three time points (ZT4, 8 and 12) (Figure 6C, right panel). Correlation analysis between *GmPRR3b* and *GmCCA1a* mRNA levels further consolidates that *GmPRR3b^{H6}* is less effective than *GmPRR3b^{H4}* in repressing *GmCCA1a* transcription (Figure 6D). Taken together, *GmPRR3b^{H6}* harbors a moderate activity in comparison to *GmPRR3b^{H4}*, which may be associated with favorite traits for domestication and breeding improvement.

Knockout of *GmPRR3b^{H6}* retards growth and delayed floral transition in soybean

To further characterize the function of *GmPRR3b^{H6}* in soybean cultivar Tianlong 1, we generated the *Gmprr3b* loss-of-function mutants using the CRISPR-Cas9 technology (Figure 7A). DNA sequencing identified two putative null mutants, *Gmprr3b-1* and *Gmprr3b-2*, which harbor a 4-bp deletion and a 5-bp deletion with a 1-bp substitution, respectively, in the first exon of *GmPRR3b^{H6}*, causing frameshifts and pre-termination of protein translation (Figure 7A). Morphological analysis showed that the *Gmprr3b* mutants and the *GmPRR3b^{H6}* overexpression lines displayed opposite phenotypes in several aspects including plant height, node and grain number, and yield per plant (Figure 7B and 7C). The *Gmprr3b* null mutants produced less main stem nodes and grains in the mature stage, suggesting that *GmPRR3b^{H6}* is essential for vigorous growth and high yield in soybean.

Unexpectedly, the *Gmprr3b* mutant plants displayed a slightly later flowering phenotype under ND and LD conditions (Figure 7D). To interrogate why the absence of *GmPRR3b^{H6}* delays flowering, we tested the transcription profiles of flowering-associated genes under both LD and SD conditions (Supplemental Figure 17 and 18). The results showed that the expression levels of *GmCCA1a*, *J* and *GmFT2a* (flowering activators) were upregulated, while *E1* (flowering repressor) was down-regulated in the *Gmprr3b-1* mutant in comparison to WT (Figure 7E and Supplemental Figure 17 and 18), demonstrating that the abnormal flowering time of the *Gmprr3b* mutants is not mediated by the *J-E1-GmFT2a* pathway (Figure 7G) (Lu et al., 2017). Alternatively, the delayed flowering of the *Gmprr3b* mutants may be a consequence of retarded growth since that the main stem nodes of mutants emerged significantly slower than the wild type, especially under LD conditions (Figure 7F and

Supplemental Figure 19).

DISCUSSION

GmPRR3b is a domestication associated gene selected for early flowering Domestication is a process involving human selection for morphological and physiological traits that allowed for cultivation (Gaut et al., 2018). Selection is usually accompanied by loss of allelic diversity during domestication and genetic improvement (Zhou et al., 2015b). In soybean, a reduction of 16.2-25% of genetic diversity was detected from the landraces to improved cultivars by whole genome resequencing (Li et al., 2013; Zhou et al., 2015b); thus, early farmer-derived landraces have a reservoir of genetic diversity. In view of this, landraces have been used to explore the genetics of domestication of agronomic traits in many species (Navarro et al., 2017). Flowering time is one of few rare traits, that was expansively selected for during domestication, geographical expansion and genetic improvement, and obvious differences are observed between domesticated crops and their progenitors, among crops from different geographic regions or even among genotypes from similar geographic areas in the case of soybean (Li et al., 2014; Wu et al., 2017).

In this study, a set of soybean landraces from the general predicted domestication area in China was used to identify QTL underlying flowering time using a GWAS approach. We detected one QTL that mapped to a cloned flowering time gene *E2* (Fang et al., 2017; Mao et al., 2017), and a location-stable flowering gene *GmPRR3b*. Both genes, *E2* and *GmPRR3b*, showed evidence of selection during domestication (Wang et al., 2016). Whilst under selection during domestication and subsequent genetic improvement, the frequency of the early flowering allele in two causal variants (Chr12:5509317 and Chr12:5520945) at *GmPRR3b* were increased and likely fixed in improved cultivars. Therefore, we were unable to detect *GmPRR3b* in populations that consisted of only improved cultivars, even when a few landraces were included (Fang et al., 2017; Mao et al., 2017), as they were filtered out in GWAS analysis due to low frequency. Therefore, focusing on landrace populations allowed us to identify new genetic variants or genes controlling agronomic traits.

We found eight haplotypes encompassing this gene in our population consisting of wild soybean (*G. soja*), landraces and improved cultivars. Of the haplotypes, H1 to H3 are found in only *G. soja*, H4 to H8 are in both *G. soja* and landraces. H6, H7 and H8 are phylogenetically associated with H4 and H5, respectively (Figure 2B). Accessions carrying H6, H7 and H8 are geographic distributed in a higher latitude than the ones harboring H4 and H5 (Figure 2C), consistent with the observation that H6, H7 and H8 are associated with early flowering and maturation traits as compared with H4 and H5, suggesting that *GmPRR3b* is not only the target of artificial selection during domestication and genetic improvement, but also associated with soybean expansion.

GmPRR3b^{H6} may have a pleiotropic role favored for domestication and genetic improvement Intriguingly, although H6 was also associated with early flowering and maturation, only the landraces possessing H6, but not H7 or H8, were expanded throughout China. Moreover, H6 was exclusively selected in the improved cultivars (Figure 2A), suggesting that H6 might have other ‘benefits’ for domestication and genetic improvement. Congruent with this, the overexpression and knockout of *GmPRR3b^{H6}* displayed multiple opposite phenotypes in plant height, main stem node number, grain number, and grain yield per plant (Figure 7), demonstrating a pleiotropic role of *GmPRR3b* in regulating growth and development. In addition, knockout of *GmPRR3b^{H6}* in cultivar Tianlong1 delayed flowering and retarded growth under LD conditions (Figure 7D and F; Supplemental Figure 19), implying a possibility that H6 has advantages over other haplotypes in simultaneously accelerating flowering and ensuring high yield.

Possible modes of GmPRR3b^{H6} action Recent studies examining the genetics of flowering time in soybean determined that the chromosomal region containing *GmPRR3b* was associated with the flowering time (Li et al., 2019d; Li et al., 2019e; Zhou et al., 2015a). However, the functional mechanism of this locus, including *GmPRR3b*, is still unclear. In this study, we found that GmPRR3b is a direct repressor of the flowering enhancer *GmCCA1a* (Wang et al., 2019). Moreover, GmCCA1a acts as a direct activator of *J* which targets and represses the key legume-specific flowering repressor *E1*. Hence, our data established a pathway linking the central circadian clock to flowering time regulation in soybean (Figure 7G). We deduced that the full-length GmPRR3b ancestral form (with CCT domain, such as GmPRR3b^{H4}) acts to delay flowering in non-inductive LD conditions. The gain of a stop codon in the truncated form GmPRR3b^{H6}, leading to the loss of the CCT domain, reduces protein function and causes early flowering trait. Whereas complete knockout of *GmPRR3b^{H6}* reduces the growth rate of node in the vegetative stage (Figure 7F and Supplemental Figure 19), which may prolong the vegetative stage for preparation of floral transition. Taken together, these possibilities warrant further investigation to illustrate the mechanism of how *GmPRR3b^{H6}* does confer better agronomic traits than other haplotypes.

METHODS

Plant materials and growth conditions

A panel of 279 landraces was selected from a Chinese primary core collection to capture as much of the representatives of diversity of the collection of soybean landraces present in the Chinese National Soybean GeneBank (CNSGB) as possible (Qiu et al., 2013). These landraces were planted in two experimental fields, Wuhan city in Hubei province (30.5 °N,

114.3 °E), and Beijing city (40.1 °N, 116.7 °E) with two replicates in 2015. Flowering time was scored in Beijing and Wuhan test sites, maturity time and plant height were scored only in Wuhan site according to Qiu et al.(Qiu et al., 2006) since only a small part of landraces (127 landraces, 45.5% of 279 landraces) were mature properly in the Beijing site. To confirm the flowering dates and periods, wild-type (Cultivar Tianlong 1) and transgenic soybean were grown under natural condition from May to October in a field in Beijing (40.1 °N, 116.7 °E), LD conditions (16 h light/ 8 h dark, 27°C), or SD conditions (8 h light/ 16 h dark, 27°C) in a controlled growth chamber. To investigate a series of agronomic traits of mature soybean under natural condition, whole plants were harvested and dried for two weeks to measure the following agronomic traits, including node number, plant height, grain number per plant, and yield per plant.

DNA extraction, sequencing, and SNP calling

Genomic DNA was extracted from each sample using the CTAB method and prepared in libraries for sequencing using a TruSeq Nano DNA HT sample preparation kit (Illumina, USA) following the manufacturer's recommendations. Briefly, the genomic DNA samples were fragmented by sonication to a size of ~350 bp, then end-polished, A-tailed, and ligated with the full-length adapters for Illumina sequencing with further PCR amplification. Index codes were added to facilitate the differentiation of sequences from each sample. The PCR products were purified (AMPure XP bead system; Beckman Coulter, USA) and the libraries were analyzed for their size distribution using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and quantified using RT-PCR. The libraries were sequenced using the Illumina HiSeq X platform (Illumina). To ensure reliable reads without artificial bias, the low-quality paired reads ($\geq 10\%$ unidentified nucleotides (N); > 10 nucleotides aligned to the adaptor, allowing $\leq 10\%$ mismatches; $> 50\%$ bases with a Phred quality less than 5) were removed. The remaining high-quality paired-end reads were mapped to the *Glycine max* Wm82.a2.v1 reference genome using Burrows-Wheeler Aligner software with the command ‘mem -t 4 -k 32 -M’ (Li and Durbin, 2009). To reduce mismatches generated by the PCR amplification before sequencing, the duplicated reads were removed with the help of SAMtools (v0.1.19)(Li et al., 2009). After alignment, SNP calling on a population scale was performed with the Genome Analysis Toolkit (GATK, version v3.8) with the HaplotypeCaller method. Subsequently, to exclude SNP-calling errors caused by incorrect mapping, only high-quality SNPs (SNP in which heterozygosity according with Hardy–Weinberg expectation, missing ratio of samples within population $\leq 10\%$ and minor allele frequency (MAF) ≥ 0.01) were retained for subsequent analyses (Wang et al., 2018b).

GWAS analysis

Association analysis for flowering time and maturity time with 6,341,742 SNPs were

performed using Compressed Mixed Linear Model (cMLM) in the GAPIT program to account for sample structure in 279 landraces (Lipka et al., 2012). The first four principal components were included as fixed effects. With the number of SNPs analyzed ($n = 6,341,742$), the threshold for significance was estimated to be approximately $P = 1 \times 10^{-7}$ (that is, 1/6,341,742) by the Bonferroni correction method. The pair-wise r^2 was calculated between significant SNPs and its two sides 1Mb-context SNPs, the LD block region of QTL was defined by its farthest 1Mb-context SNPs with a $r^2 \geq 0.8$ in each side.

Selection signal detection

Nucleotide diversity θ_π was calculated using VariScan (v2.0.3) and fixation index (F_{ST}) was calculated using VCFtools (v0.1.14) (Danecek et al., 2011). These population statistics were analyzed using the sliding-window approach (20-kb windows with 10-kb increments).

Primers and accession numbers

All primers used in this study were listed in Supplemental Table 3. Gene Sequences were downloaded from the *Glycine max* Wm82.a2.v1 (Soybean) database (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax). The accession numbers are *GmPRR3b* (*Glyma.I2G073900*), *GmPRR3a* (*Glyma.U034500*), *GmPRR1a* (*Glyma.04G166300*), *GmPRR1d* (*Glyma.17G102200*), *GmPRR5a* (*Glyma.04G228300*), *GmPRR5d* (*Glyma.07G049400*), *GmPRR7a* (*Glyma.10G048100*), *GmCCA1a* (*Glyma.07G048500*), *GmCCA1b* (*Glyma.16G017400*), *GmCCA1c* (*Glyma.03G261800*), *GmCCA1d* (*Glyma.19G260900*), *J/GmELF3a* (*Glyma.04G050200*), *GmELF3b* (*Glyma.14G091900*), *GmELF3c* (*Glyma.17G231600*), *GmELF4a* (*Glyma.11G229700*), *GmLUX1* (*Glyma.12G060200*), *GmLUX2* (*Glyma.03G123400*), *E1* (*Glyma.06G207800*), *GmFT2a* (*Glyma.16G150700*), *GmFT5a* (*Glyma.16G044100*), *GmFT4* (*Glyma.08G363100*), *GmCAB1* (*Glyma.05G128000*) and *GmActin* (*Glyma.18G290800*). Raw sequencing reads are PRJNA552939 deposited in NCBI and PRJCA001583 deposited in BioProject.

Plasmid construction and generation of transgenic plants

To construct *H6-YFP* and *YFP-H6* plant transformation plasmids, coding DNA sequence (CDS) of *GmPRR3b^{H6}* (1,878 bp) was amplified from the cDNA of young leaves of Williams 82 using primer Attb-H6-F and Attb-H6-R and inserted into the overexpression vector pEarleyGate 101 or pEarleyGate 104 by the Gateway system following the manufacturer's instructions (Invitrogen). To generate the CRISPR/Cas9-engineered *Gmprr3b* mutant, gRNAs were designed using CRISPR-P website (<http://crispr.hzau.edu.cn>). The 19-bp DNA fragment coding the gRNA was inserted into the pCas9-AtU6-sgRNA plasmid at the *Xba*I site. The CRISPR/Cas9 expression vector was constructed with Cas9 CDS driven by *Arabidopsis RPS5A* promoter and the customized gRNA transcribed by *Arabidopsis U6* promoter (Tsutsui

and Higashiyama, 2017). Above expression plasmids were individually introduced into *Agrobacterium tumefaciens* strain EHA105 via electroporation and then transformed into wild-type soybean (Tianlong 1, *E1e2E3E4* and *GmPRR3b^{H6}* background) by the cotyledon-node method (Paz et al., 2006).

Sample Collection, RNA isolation and qRT-PCR analysis

To study the tissue-specific expression pattern of *GmPRR3b^{H6}* in various tissues, Williams 82 plants were grown under LD conditions for three weeks. The root, hypocotyl, cotyledon, unifoliolate leaves, trifoliolate leaves, stem, and shoot apical samples were collected 4 h after the light was turned on. To analyze the rhythmic expression pattern of *GmPRR3b^{H6}*, Williams 82 plants were grown under LD or SD conditions for three weeks. The second fully expanded trifoliolate leaves were collected every 4 h during a 24 h LD or SD photoperiod, and then a 24 h continue light period (free-running condition). To compare dynamic transcription levels of indicated genes in WT and *YFP-H6* transgenic plants, the second fully expanded trifoliolate leaves of 3-week-old plants grown under LD conditions were collected every 4 h during a 24 h LD photoperiod and then a 24 h free-running period. To compare gene transcriptional levels in WT and *Gmprr3b* mutant plants, the second fully expanded trifoliolate leaves were collected at ZT4 and ZT12 from 3-week-old plants under LD conditions. To analyze the *GmPRR3b* expression level in different accessions carrying H4 or H6, the second fully expanded trifoliolate leaves were collected at ZT4, ZT8, ZT12 under LD conditions or at ZT0, ZT4, ZT8 under SD conditions from 3-week-old plants. All above samples were used to extract total RNA using TRIzol reagent (Invitrogen). For qRT-PCR, the first complementary DNA and cDNA were synthesized from DNase-treated total RNA (3 µg, reaction total volume 20 ul) using a reverse transcription kit (TRAN). qRT-PCR was performed in 96-well optical plates using a SYBR Green RT-PCR kit (Takara) and a Roche Light Cycler 480. The mRNA level of *GmActin* was used as the internal control. Three independent biological replicates were analyzed, and three replicate reactions were used for each sample.

Subcellular Localization in soybean mesophyll protoplasts

To investigate the subcellular localization of indicated proteins, the coding sequences of *GmPRR3b^{H4}*, *GmPRR3b^{H6}*, *GmPRR3b^{H6m}* were inserted into PA7-YFP vector at the *BamHI* and *SmaI* sites using the In-fusion system (Clontech), to generate the *H4-YFP*, *H6-YFP*, *H6^m-YFP* transient expression constructs driven by 35S promoter. PA7-YFP empty vector was used as control. Plasmids were transformed into soybean mesophyll protoplasts using our previous method (Xiong et al., 2019). The subcellular localization images were captured under the Zeiss LSM780 confocal laser scanning microscope and processed using ZEN 2009 Light Edition software.

ChIP qRT-PCR Assay

ChIP assay was performed following the protocol described previously with some modifications (Zhang et al., 2016). 4 g leaves of the 4-week-old plant or hairy root callus of indicated genotypes grown under LD conditions were collected 8 h after the light was turned on. Tissues were incubated twice in 30.8 mL formaldehyde buffer (0.4 M Suc, 10 mM Tris-HCl (pH 8.0), 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM EDTA, and 1% (v/v) formaldehyde) and vacuumed for 15 min each time. Then 1.6 mL of 2 M Glycine buffer was added to stop cross-linked reaction for 5 min under vacuum. The cross-linked material was repeatedly rinsed with ddH₂O to remove formaldehyde buffer, ground in liquid nitrogen, resuspended in 15 mL Honda buffer (0.44 M Suc, 1.25% Ficoll, 2.5% Dextran T40, 20 mM HEPES KOH (pH 7.4), 10 mM MgCl₂, 0.5% Triton X-100, 5 mM dithiothreitol, 1 mM PMSF, and 1 tablet/50 mL of protease inhibitor cocktail), filtered with two layers of Miracloth and centrifuged for 15 min at 2000 g, 4 °C. Sediment was resuspended with 500 uL nuclei lysis buffer (50 mM Tris-HCl (pH 8.8), 10 mM EDTA, 1% SDS, 1 mM PMSF, and 1 tablet/50 mL of protease inhibitor cocktail), subjected to ultrasound (pulse on 15 s, pulse off 60 s, 5-6 times), centrifuged for 10 min at 13000 rpm, 4°C. 200 uL supernatant solutions were taken and incubated with GFP-trap_A agarose beads overnight at 4°C, centrifuged (1 min, 1500 g, 4°C). The precipitated DNA was recovered with water and used as templates for qRT-PCR. The enrichment folds of three biological replicates were calculated as the ratio between the transgenic sample and wild-type control sample using *GmActin* as the negative control.

Dual-Luciferase Reporter Assay

To construct the reporter plasmid, a 3,671 bp sequence corresponding to the *J* promoter was amplified and inserted into *pGreen0800-LUC* vector. To construct the effector plasmid, the full-length *GmCCA1a* cDNA was amplified and inserted into the *pA7-YFP* vector. The reporter and effector plasmids were transiently co-transformed into *Arabidopsis* mesophyll protoplasts. The firefly luciferase (LUC) and Renilla luciferase (REN) activities were quantified using the Dual-Luciferase Reporter Assay System (Promega, United States). The relative activity of *J* promoter was calculated as LUC to REN ratio of three biological replicates.

RICE system to investigate gene expression

The *GmCCA1a-YFP* and *H6-YFP* plasmids were introduced into *Agrobacterium tumefaciens* strain K599 which further infected the young seedlings of Tianlong 1 at the hypocotyl region to induce transgenic hairy roots according to the previously reported method (Kereszt et al., 2007). The hairy roots induced by Empty K599 was used as the wild-type control. The callus induction medium (2.22 g / L Murashige & Skoog Basal Medium with Vitamins, 0.59 g / L MES monohydrate, 30 g / L sucrose, 1 mg / L 2, 4-D, 0.1 mg / L 6-BA, 0.1 g / L Timentin)

was prepared as previously described with minor modifications (Leshina and Bulko, 2014; Wang et al., 2001). The transgenic roots were grown on the callus induction medium for two weeks under LD conditions. Those transgenic callus lines confirmed by qRT-PCR were transferred to fresh callus induction medium for subculture. Three independent hairy root callus lines of each indicated genotype were grown under indicated conditions and used for transcriptional analysis or the ChIP assay of target genes. To compare the transcriptional repression activity of GmPRR3b^{H4} and GmPRR3b^{H6} on *GmCCA1a*, at least ten independent *YFP-H4* or *YFP-H6* hairy root callus lines were generated for qRT-PCR analysis.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was partially supported by the National Key Research and Development Plan (2016YFD0101005, 2016YFD0100201 and 2016YFD0100304), the National Natural Science Foundation of China (31871705, 31422041), the Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences, and the Central Public-Interest Scientific Institution Basal Research Fund (Y2016JC13).

AUTHOR CONTRIBUTIONS

B.L., Y.H.L., and L.J.Q. designed the research. C.L., Y.H.L., Y.F.L., T.Z., J.L. and H.Y.L. performed the experiments. Y.F.L., Y.T., H.L.H. collected the phenotypic data. Y.H.L., H.F.L., Y.F.L., X.W.Z. and C.L. analyzed data. B.L., Y.H.L., S.A.J. and L.J.Q. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Kenneth M. Olsen for the critical reading of this manuscript and valuable suggestions. No conflict of interest declared.

REFERENCES

- Barrett, J.C., Fry, B., Maller, J., and Daly, M.J.** (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263-265.
- Cober, E.R., Molnar, S.J., Charette, M., and Voldeng, H.D.** (2010). A new locus for early maturity in soybean. *Crop Sci.* **50**: 524-527.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E.,**

- Lunter, G., Marth, G.T., Sherry, S.T., et al.** (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.
- Fang, C., Ma, Y., Wu, S., Z., L., Wang, Z., Yang, R., Hu, G., Zhou, Z., and Yu, H.** (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**: 161.
- Gaut, B.S., Seymour, D.K., Liu, Q., and Zhou, Y.** (2018). Demography and its effects on genomic variation in crop domestication. *Nat. Plants* **4**: 512-520.
- Gendron, J.M., Pruneda-Paz, J.L., Doherty, C.J., Gross, A.M., Kang, S.E., and Kay, S.A.** (2012). *Arabidopsis* circadian clock protein, TOC1, is a DNA-binding transcription factor. *Proc. Natl. Acad. Sci. USA* **109**: 3167-3172.
- Gould, P.D., Ugarte, N., Domijan, M., Costa, M., Foreman, J., Macgregor, D., Rose, K., Griffiths, J., Millar, A.J., Finkenstadt, B., et al.** (2013). Network balance via *CRY* signalling controls the *Arabidopsis* circadian clock over ambient temperatures. *Mol. Syst. Biol.* **9**: 650.
- Hayama, R., Sarid-Krebs, L., Richter, R., Fernandez, V., Jang, S., and Coupland, G.** (2017). PSEUDO RESPONSE REGULATORs stabilize CONSTANS protein to promote flowering in response to day length. *EMBO J.* **36**: 904-918.
- Hsu, P.Y., and Harmer, S.L.** (2014). Wheels within wheels: the plant circadian system. *Trends Plant Sci.* **19**: 240-249.
- Jiang, B., Nan, H., Gao, Y., Tang, L., Yue, Y., Lu, S., Ma, L., Cao, D., Sun, S., Wang, J., et al.** (2014). Allelic combinations of soybean maturity loci *E1*, *E2*, *E3* and *E4* result in diversity of maturity and adaptation to different latitudes. *PLoS One* **9**: e106042.
- Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C.D., Nontachaiyapoom, S., Kinkema, M., and Gresshoff, P.M.** (2007). Agrobacterium rhizogenes-mediated transformation of soybean to study root biology. *Nat. Protoc.* **2**: 948-952.
- Klein, R.R., Miller, F.R., Dugas, D.V., Brown, P.J., Burrell, A.M., and Klein, P.E.** (2015). Allelic variants in the *PRR37* gene and the human-mediated dispersal and diversification of sorghum. *Theor. Appl. Genet.* **128**: 1669-1683.
- Kong, F., Nan, H., Cao, D., Li, Y., Wu, F., Wang, J., Lu, S., Yuan, X., Cober, E.R., and Abe, J.** (2014). A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Sci.* **54**: 2529-2535.
- Koo, B.H., Yoo, S.C., Park, J.W., Kwon, C.T., Lee, B.D., An, G., Zhang, Z., Li, J., Li, Z., and Paek, N.C.** (2013). Natural variation in *OsPRR37* regulates heading date and contributes to rice cultivation at a wide range of latitudes. *Mol. Plant* **6**: 1877-1888.
- Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.W., He, W., Qin, N., Wang, B., et al.** (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* **42**: 1053-1059.
- Lee, G.A., Crawford, G.W., Liu, L., Sasaki, Y., and Chen, X.** (2011). Archaeological soybean (*Glycine max*) in East Asia: does size matter? *PLoS One* **6**: e26720.
- Leshina, L.G., and Bulko, O.V.** (2014). Plants regeneration from genetically transformed root and callus cultures of periwinkle *Vinca minor* L. and foxglove purple *Digitalis purpurea* L. *Tsitol Genet.* **48**: 36-42.
- Li, B., Wang, Y., Zhang, Y., Tian, W., Chong, K., Jang, J.C., and Wang, L.** (2019a). *PRR5*, *7* and *9* positively modulate *TOR* signaling-mediated root cell proliferation by repressing *TANDEM ZINC FINGER 1* in *Arabidopsis*. *Nucleic Acids Res.* **47**: 5001-5015.

- Li, H., and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S.** (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Li, M.W., Liu, W., Lam, H.M., and Gendron, J.M.** (2019b). Characterization of two growth period QTLs reveals modification of PRR3 genes during soybean domestication. *Plant Cell Physiol.* **60**: 407-420.
- Li, X., Zhang, X., Zhu, L., Bu, Y., Wang, X., Zhang, X., Zhou, Y., Wang, X., Guo, N., Qiu, L., et al.** (2019c). Genome-wide association study of four yield-related traits at the R6 stage in soybean. *BMC Genet.* **20**: 39.
- Li, Y., Dong, Y., Wu, H., Hu, B., Zhai, H., Yang, J., and Xia, Z.** (2019d). Positional cloning of the flowering time QTL *qFT12-1* reveals the link between the clock related *PRR* homolog with photoperiodic response in soybeans. *Front. Plant Sci.* **10**: 1303.
- Li, Y.H., Li, D., Jiao, Y.Q., Schnable, J.C., Li, Y.F., Li, H.H., Chen, H.Z., Hong, H.L., Zhang, T., Liu, B., et al.** (2019e). Identification of loci controlling adaptation in Chinese soya bean landraces via a combination of conventional and bioclimatic GWAS. *Plant Biotechnol J.* <https://doi.org/10.1111/pbi.13206>.
- Li, Y.H., Zhao, S.C., Ma, J.X., Li, D., Yan, L., Li, J., Qi, X.T., Guo, X.S., Zhang, L., He, W.M., et al.** (2013). Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. *BMC Genomics* **14**: 579.
- Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., et al.** (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* **32**: 1045-1052.
- Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., Gore, M.A., Buckler, E.S., and Zhang, Z.** (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**: 2397-2399.
- Liu, C., Song, G., Zhou, Y., Qu, X., Guo, Z., Liu, Z., Jiang, D., and Yang, D.** (2015). *OsPRR37* and *Ghd7* are the major genes for general combining ability of *DTH*, *PH* and *SPP* in rice. *Sci. Rep.* **5**: 12803.
- Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., Fang, C., Cao, D., Shi, X., Kong, L., et al.** (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* **49**: 773-779.
- Lu, S.X., Webb, C.J., Knowles, S.M., Kim, S.H., Wang, Z., and Tobin, E.M.** (2012). CCA1 and ELF3 interact in the control of hypocotyl length and flowering time in *Arabidopsis*. *Plant Physiol.* **158**: 1079-1088.
- Mao, T., Li, J., Wen, Z., Wu, T., Wu, C., Sun, S., and Wang, D.** (2017). Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. *BMC Genomics* **18**: 415.
- McClung, C.R.** (2019). The plant circadian oscillator. *Biology* **8**: 14. <https://doi.org/10.3390/biology8010014>.
- Murphy, R.L., Klein, R.R., Morishige, D.T., Brady, J.A., Rooney, W.L., Miller, F.R., Dugas, D.V., Klein, P.E., and Mullet, J.E.** (2011). Coincident light and clock regulation of *pseudoresponse regulator protein 37* (*PRR37*) controls photoperiodic flowering in sorghum. *Proc. Natl. Acad. Sci.*

- USA **108**: 16469-16474.
- Nagel, D.H., Doherty, C.J., Pruneda-Paz, J.L., Schmitz, R.J., Ecker, J.R., and Kay, S.A.** (2015). Genome-wide identification of *CCA1* targets uncovers an expanded clock network in *Arabidopsis*. Proc. Natl. Acad. Sci. USA **112**: E4802-4810.
- Nan, H., Cao, D., Zhang, D., Li, Y., Lu, S., Tang, L., Yuan, X., Liu, B., and Kong, F.** (2014). *GmFT2a* and *GmFT5a* redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. PLoS One **9**: e97669.
- Navarro, J., Willcox, M., Burgueño, J., Romay, C., Swarts, K., Trachsel, S., and Ortega, A.** (2017). A study of allelic diversity underlying flowering-time adaptation in maize landraces. Nat. Genet. **49**: 476-480.
- Ott, A., Liu, S., Schnable, J.C., Yeh, C., Wang, K.S., and Schnable, P.S.** (2017). tGBS(R) genotyping-by-sequencing enables reliable genotyping of heterozygous loci. Nucleic Acids Res. **45**: e178.
- Para, A., Farre, E., Imaizumi, T., Pruneda-Paz, J., Harmon, F., and Kay, S.** (2007). *PRR3* is a vascular regulator of *TOC1* stability in the *Arabidopsis* circadian clock. Plant Cell **19**: 3462-3473.
- Paz, M.M., Martinez, J.C., Kalvig, A.B., Fonger, T.M., and Wang, K.** (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. Plant Cell Rep. **25**: 206-213.
- Possart, A., Fleck, C., and Hiltbrunner, A.** (2014). Shedding (far-red) light on phytochrome mechanisms and responses in land plants. Plant Sci. **217-218**: 36-46.
- Qiu, L.J., Chang, R.Z., Liu, Z.X., Guan, R.X., and Li, Y.H.** (2006). Descriptors and data standard for soybean (*Glycine* spp.) (Beijing: Chinese Agricultural Press).
- Qiu, L.J., Xing, L.L., Guo, Y., Wang, J., Jackson, S.A., and Chang, R.Z.** (2013). A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol. Biol. **83**: 41-50.
- Sanchez, S.E., and Kay, S.A.** (2016). The plant circadian clock: from a simple timekeeper to a complex developmental manager. Cold Spring Harb. Perspect. Biol. **8**.
- Sedivy, E.J., Wu, F., and Hanzawa, Y.** (2017). Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol. **214**: 539-553.
- Sperschneider, J., Catanzariti, A.M., DeBoer, K., Petre, B., Gardiner, D.M., Singh, K.B., Dodds, P.N., and Taylor, J.M.** (2017). LOCALIZER: subcellular localization prediction of both plant and effector proteins in the Plant Cell. Sci. Rep. **7**: 44598.
- Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., Katayose, Y., Abe, J., Ishimoto, M., and Harada, K.** (2013). Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. Ann. Bot. **113**: 429-441.
- Tsutsui, H., and Higashiyama, T.** (2017). pKAMA-ITACHI vectors for highly efficient CRISPR/Cas9-mediated gene knockout in *Arabidopsis thaliana*. Plant Cell Physiol. **58**: 46-56.
- Wang, L., Kim, J., and Somers, D.E.** (2013). Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. Proc. Natl. Acad. Sci. USA **110**: 761-766.
- Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., Yang, R., Zhang, M., Liu, S., Lu, S., et al.** (2018a). Parallel selection on a dormancy gene during domestication of crops from multiple families. Nat. Genet. **50**: 1435-1441.
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R.,**

- Zhang, F., et al.** (2018b). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43-49.
- Wang, Y., Gu, Y., Gao, H., Qiu, L., Chang, R., Chen, S., and He, C.** (2016). Molecular and geographic evolutionary support for the essential role of GIGANTEAa in soybean domestication of flowering time. *BMC Evol. Biol.* **16**: 79.
- Wang, Y., Yuan, L., Su, T., Wang, Q., Gao, Y., Zhang, S., Jia, Q., Yu, G., Fu, Y., Cheng, Q., et al.** (2019). Light- and temperature-entrainable circadian clock in soybean development. *Plant Cell Environ.* <https://doi.org/10.1111/pce.13678>.
- Wang, Y.M., Jiang, B.W., Da, L., and Jing, F.J.** (2001). Regeneration of plants from callus cultures of roots induced by Agrobacterium rhizogenes on *Alhagi pseudoalhagi*. *Cell Res.* **279-284**: 279-284.
- Wu, F., Sedivy, E.J., Price, W.B., Haider, W., and Hanzawa, Y.** (2017). Evolutionary trajectories of duplicated *FT* homologues and their roles in soybean domestication. *Plant J.* **90**: 941-953.
- Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lu, S., et al.** (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. USA* **109**: E2155-2164.
- Xiong, L., Li, C., Li, H., Lyu, X., Zhao, T., Liu, J., Zuo, Z., and Liu, B.** (2019). A transient expression system in soybean mesophyll protoplasts reveals the formation of cytoplasmic GmCRY1 photobody-like structures. *Sci. China Life Sci.* **62**: 1070-1077.
- Yang, P., Wang, J., Huang, F.Y., Yang, S., and Wu, K.** (2018). The plant circadian clock and chromatin modifications. *Genes* **9**: 561.
- Zhai, H., Lu, S., Liang, S., Wu, H., Zhang, X., Liu, B., Kong, F., Yuan, X., Li, J., and Xia, Z.** (2014). *GmFT4*, a homolog of *FLOWERING LOCUS T*, is positively regulated by *E1* and functions as a flowering repressor in soybean. *PloS One* **9**: e89030.
- Zhang, C., Liu, J., Zhao, T., Gomez, A., Li, C., Yu, C., Li, H., Lin, J., Yang, Y., Liu, B., et al.** (2016). A drought-inducible transcription factor delays reproductive timing in rice. *Plant Physiol.* **171**: 334-343.
- Zhang, S.R., Wang, H., Wang, Z., Ren, Y., Niu, L., Liu, J., and Liu, B.** (2017). Photoperiodism dynamics during the domestication and improvement of soybean. *Sci. China Life Sci.* **60**: 1416-1427.
- Zhang, Z., Ji, R., Li, H., Zhao, T., Liu, J., Lin, C., and Liu, B.** (2014). *CONSTANS-LIKE 7 (COL7)* is involved in *phytochrome B (phyB)*-mediated light-quality regulation of auxin homeostasis. *Mol. Plant* **7**: 1429-1440.
- Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., Kanazawa, A., Liu, B., Kong, F., Yamada, T., et al.** (2016). A recessive allele for delayed flowering at the soybean maturity locus *E9* is a leaky allele of *FT2a*, a *FLOWERING LOCUS T* ortholog. *BMC Plant biol.* **16**: 20.
- Zhou, L., Wang, S.B., Jian, J., Geng, Q.C., Wen, J., Song, Q., Wu, Z., Li, G.J., Liu, Y.Q., Dunwell, J.M., et al.** (2015a). Identification of domestication-related loci associated with flowering time and seed size in soybean with the RAD-seq genotyping method. *Sci. Rep.* **5**: 9350.
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., and Ma, Y.** (2015b). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**: 408-414.

Figure 1. Identification of *GmPRR3b* as a domestication gene in regulating flowering time and maturation time by genome-wide association study.

Figure 2. The evolutionary and geographical distribution characterization of *GmPRR3b* haplotypes.

Figure 3. NLS1 and NLS2 are essential for the nuclear localization of GmPRR3b.

Figure 4. *GmPRR3b^{H6}* directly represses *GmCCA1a* expression.

Figure 5. *GmCCA1a* directly upregulates *J* expression.

Figure 6. *GmPRR3b^{H4}* is more effective than *GmPRR3b^{H6}* to repress *GmCCA1a* transcription.

Figure 7. CRISPR/Cas9-engineered mutations in *GmPRR3b^{H6}* confer reduced node number and late flowering phenotypes.

Supplemental Figure 1. Genome-wide association studies for flowering time in soybean landraces.

Supplemental Figure 2. Genome-wide association studies for maturity time.

Supplemental Figure 3. Boxplots for flowering time in soybean landraces based on the different genotypes of Chr12:5520945 in Beijing and Wuhan test sites.

Supplemental Figure 4. The ratio of θ_π (gene diversity) between pairwise of three evolutionary populations (*G. soja*, landrace and improved cultivars) of genomic regions harboring *GmPRR3b*.

Supplemental Figure 5. Violin plots of maturity time for haplotypes of *GmPRR3b* in Beijing and Wuhan test sites.

Supplemental Figure 6. Overexpression of *GmPRR3b^{H6}* confers late flowering phenotype.

Supplemental Figure 7. Statistical analysis of the agronomic traits of *H6-YFP* and *YFP-H6* overexpression plants.

Supplemental Figure 8. Dynamic transcriptional levels of each indicated flowering associated gene in WT and *YFP-H6* plant under LD and free running conditions.

Supplemental Figure 9. Dynamic transcriptional level of *GmCCA1* genes in WT and *YFP-H6* plants under LD and free running conditions.

Supplemental Figure 10. Dynamic transcriptional levels of *GmPRR* genes in WT and *YFP-H6* plants under LD and free running conditions.

Supplemental Figure 11. Dynamic transcriptional levels of putative evening complex genes in WT and *YFP-H6* plants under LD and free running conditions.

Supplemental Figure 12. Dynamic transcriptional levels of putative evening complex genes and *E1* gene in WT and *GmCCA1a-YFP* hairy root calluses cell under LD condition.

Supplemental Figure 13. ChIP-qPCR analysis of the binding of *GmCCA1a* to *GmLUX1* and *GmELF4a* promoters using the *GmCCA1a-YFP* hair root callus.

Supplemental Figure 14. Comparison of *GmPRR3b* and *GmCCA1a* expression levels in different accessions carrying haplotype H4 or H6.

Supplemental Figure 15. The hypocotyl phenotypes of *GmPRR3b^{H4}* or *GmPRR3b^{H6}* overexpression lines in *Arabidopsis*.

Supplemental Figure 16. The flowering time phenotype of *GmPRR3b^{H4}* or *GmPRR3b^{H6}* overexpression lines in *Arabidopsis*.

Supplemental Figure 17. Dynamic transcriptional levels of indicated genes in WT and *Gmprr3b-1* mutant under LD condition.

Supplemental Figure 18. Dynamic transcriptional levels of indicated genes in WT, *YFP-H6* and

Gmprr3b-1 mutant under SD conditions.

Supplemental Figure 19. Statistical analysis of plant height and main stem node number in WT and *Gmprr3b-1* mutant grown under LD or SD conditions for 28 or 35 days.

Supplemental Figure 20. GmCCA1a directly interacting with *GmFT4* promoter to activate its expression.

Supplemental Table 1. 385 soybean accessions used for GWAS.

Supplemental Table 2. 18 SNPs associated with flowering time in soybean landraces measured at Beijing and Wuhan test sites.

Supplemental Table 3. Primers for PCR and qRT-PCR.

Supplemental Table 4. Flowering time in H4 and H6 accessions at Beijing or Wuhan test sites.

Figure Legends

Figure 1. Identification of *GmPRR3b* as a domestication gene in regulating flowering time and maturation time by genome-wide association study.

- (A and B) Local Manhattan plots for flowering time and maturity time in Wuhan environment. Dashed red lines indicate the candidate region for the peak. Arrowhead indicates the peak association signal1.
- (C) Heatmap surrounding the candidate region for the peak in Wuhan environment. Red line indicates the location of Chr12:5520945. Sky blue triangle indicates the LD block surrounding Chr12:5520945.
- (D) The genome structure and haplotypes of the *GmPRR3b* gene. The linear gene structure in the top panel displays the UTRs (black rectangles), CDS regions (teal rectangles), and introns (horizontal solid black lines). The vertical solid lines represent the SNP loci in 385 soybean samples. The dot lines link the SNPs to their physical positions and nucleotides in the bottom panel. The SNP stressed by red are the functional ones.
- (E) The protein structures of eight haplotypes. The amino acid abbreviation before and after slash represent the reference and the alternative one, respectively. Stop-gain indicates the premature translation of termination.
- (F) Violin plots for the flowering time of indicated haplotype groups in Beijing and Wuhan environments. The letters above the bars indicate the significant differences ($P<0.05$) as determined by one-way ANOVA analysis.

Figure 2. The evolutionary and geographical distribution characterization of *GmPRR3b* haplotypes.

- (A) The pies from top to bottom were wild soybean, landraces and improved cultivars, respectively. The different colored portions in each pie represent the percentage of different haplotypes.
- (B) Median-joining network of eight *GmPRR3b* haplotypes, each represented by a numbered circle fill in different color.
- (C) The geographical distribution of 383 soybean accessions carrying different haplotypes. Accession details are provided in supplementary Table 1.

Figure 3. NLS1 and NLS2 are essential for the nuclear localization of GmPRR3b.

- (A) Transcriptional analysis of *GmPRR3b^{H6}* in different tissues of soybean cultivar Williams 82 by quantitative reverse transcription PCR (qRT-PCR). The relative expression levels (REL) are shown as means \pm s.d. ($n = 3$). *GmActin* was used as an internal control and the REL of root sample was arbitrarily set to 1.
- (B) Transcriptional analysis of *GmPRR3b^{H6}* in a time course manner under diurnal and free running conditions. The REL of the 0 h sample under LD conditions was arbitrarily set to 1.
- (C) Diagram of protein structures of indicated proteins. PR, pseudo receiver domain; EAR, ERF associated amphiphilic repression motif; CCT, CONSTANS, CO-like, and TOC1 (CCT) domain;

NLS, nuclear localization signal. H6^m was generated by mutating the NLS1 amino acid residues LEGV to PADF at position 451-454 within the GmPRR3b^{H6} protein.

(D) Subcellular localization of H4-YFP (GmPRR3b^{H4}-YFP), H6-YFP (GmPRR3b^{H6}-YFP), and H6^m-YFP (GmPRR3b^{H6m}-YFP) proteins in soybean mesophyll protoplasts. PA7-YFP (YFP alone) was used as control. Scale bar = 2 μm.

(E) The percentage of the protoplasts harboring YFP fluorescence signal in nucleus or speckles. Mean values ± s.d. (n = 3) are shown.

Figure 4. GmPRR3b^{H6} directly represses *GmCCA1a* expression.

(A-C) The dynamic transcriptional level of each indicated gene in *YFP-H6* line and WT in a time course manner under LD and free running conditions. Mean values ± s.d. (n = 3) are shown. The REL of 0 h WT sample was arbitrarily set to 1.

(D) A diagram of *GmCCA1a* promoter for the ChIP assay. Nucleotide position is coordinate relative to the start codon (+1). The black, red or blue dots indicate the G-box (CACGTG, -1483 and -1540), TBS (GGCCCA, -1428 and -1419) or TGTG motif (-1067) respectively. The bars represent the amplicons containing the indicated motif.

(E) ChIP-qPCR assay of YFP-H6 binding to the *GmCCA1a* promoter using *YFP-H6* line. Enrichment fold was quantified by normalization of the immunoprecipitation signal with the corresponding input signal. Mean values ± s.d. (n = 3) are shown. *GmActin* was used as negative control. The values in *YFP-H6* line were compared with that in WT by Student's *t*-tests (** P < 0.01).

Figure 5. GmCCA1a directly upregulates *J* expression.

(A-D) The dynamic transcriptional levels of indicated genes in WT, *H6-YFP* or *GmCCA1a-YFP* hairy root calluses in a time course manner under LD conditions. Mean values ± s.d. (n = 3) are shown. The REL of WT sample at 0 h was arbitrarily set to 1.

(E) A diagram representing the *J* promoter for the ChIP assay. Nucleotide position is coordinate relative to the start codon (+1). The black dot indicates CBS-motif (AAAAATCT, -3607, -3359 and -2514); the diamond dot indicated EE-motif (AAATATCT, -1040 and -1022); the white dot indicates the recognition motif (TAACTG, -1388, -1298 and -749) of plant MYB proteins. The bars represent the amplicons containing the indicated motif.

(F) ChIP-qPCR assay of *GmCCA1a-YFP* binding to the *J* promoter using the *GmCCA1a-YFP* hair root callus. Enrichment fold was quantified by normalization of the immunoprecipitation signal with the corresponding input signal. Mean values ± s.d. (n = 3) are shown. *GmActin* was used as negative control.

(G) Constructs of effector and reporter used for the dual-luciferase assay.

(H) Comparison of relative reporter activities in the presence of indicated effectors in *Arabidopsis* protoplasts. Relative expression Units (REU) were calculated by normalizing the LUC activity to REN activity. Mean values ± s.d. (n = 3) are shown. All above comparisons were performed by Student's *t*-tests (** P < 0.01).

Figure 6. *GmPRR3b^{H4}* is more effective than *GmPRR3b^{H6}* to repress *GmCCA1a* transcription.

(A and B) Scatter plots showing the correlation between the expression levels of *GmPRR3b* and *GmCCA1a* genes in different accessions carrying haplotype H4 or H6 under LD (**A**) or SD (**B**) conditions. Details of accessions were provided in Supplementary Table 4.

(C) The transcriptional levels of *GmPRR3b* and *GmCCA1a* in the indicated root hair calluses under LD conditions. The numbers below the x axis indicate zeitgeber times (ZTs) of the day. Violin plots were created using 10 independent calluses of each genotype. All above comparisons were performed by Student's *t*-tests (** P < 0.01).

(D) Correlation analysis between the expression levels of *GmPRR3b* and *GmCCA1a* genes in the indicated calluses grown under LD conditions at ZT4, ZT8 or ZT12 respectively. Three independent lines with different expression level of *GmPRR3b* were used to make the scatter plots.

Figure 7. CRISPR/Cas9-engineered mutations in *GmPRR3b^{H6}* confer reduced node number and late flowering phenotypes.

(A) Two single-guide RNAs (sgRNA, red arrows) were designed to target the first exon of *GmPRR3b^{H6}*. The mutant sequences of two representative homozygous mutants (*GmPRR3b-1* and *GmPRR3b-2*) at T₂ generation are shown. The target sites of gRNA are highlighted in red letters with the protospacer-adjacent motif (PAM) in bold. The black letter and dashed lines within the target sites denote nucleotide insertion and deletion respectively.

(B) Representative images of indicated lines at maturation stage grown in Beijing field in 2019. Scale bar = 10 cm.

(C) Statistical analysis of the agronomic traits of indicated lines as in (B). Data are shown as means ± s.d. (n ≥ 15). The letters above the bars indicate the significant differences (P<0.05) as determined by one-way ANOVA analysis.

(D) Comparison of flowering time between *Gmprr3b* mutants and WT grown under different photoperiods by Student's *t*-tests (** P < 0.01). Data are means ± s.d. (n = 20).

(E) The transcriptional levels of indicated genes in the *Gmprr3b-1* mutant and WT grown under LD conditions. The numbers below the x axis indicate zeitgeber times (ZTs) of the day. *GmActin* was used as an internal control. The REL of WT sample at ZT4 was arbitrarily set to 1. Data are shown as means ± s.d. (n = 3).

(F) The representative images of the *Gmprr3b* mutants and WT grown under LD conditions for 24 days. The white arrows denote the position of nodes. Scale bar = 10 cm.

(G) A proposed model depicts the roles of *GmPRR3b* in flowering time and growth regulations in soybean. The solid arrows indicate activation of transcription, dotted lines denote relationship that are not established to be direct, and solid lines ending with a dash denote repression of transcription.

