Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Natural variation in GmRVE4d facilitates soybean latitudinal adaptation by regulating GmPRR5a-dependent flowering and maturity

Baiquan Sun

sunbaiquan@caas.cn

National Nanfan Research Institute, Chinese Academy of Agricultural Sciences

Tingting Wu

Ministry of Agriculture Key Laboratory of Soybean Biology, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing

Bingjun Jiang

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Xuegang Sun

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Hongchang Jia

Heihe Branch of Heilongjiang Academy of Agricultural Sciences

Mahmoud Naser

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Maoxiang Yang

Fujian Agriculture and Forestry University

Xiangyu Yao

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Yanhui Sun

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Qimeng Li

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Peiguo Wang

National Nanfan Research Institute, Chinese Academy of Agricultural Sciences

Shibo Sun

National Nanfan Research Institute, Chinese Academy of Agricultural Sciences

Chao Qin

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Shan Yuan

Ministry of Agriculture Key Laboratory of Soybean Biology, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing

Kunhui He

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Wensheng Hou

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Cunxiang Wu

Ministry of Agriculture Key Laboratory of Soybean Biology, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing

Huihui Li

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences https://orcid.org/0000-0002-9117-5011

Shi Sun

Ministry of Agriculture Key Laboratory of Soybean Biology, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing

Tianfu Han

MOA Key Laboratory of Soybean Biology (Beijing), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences

Liwei Jiang

National Nanfan Research Institute, Chinese Academy of Agricultural Sciences

---

Article

Keywords: soybean (Glycine max (L.) Merr.), genome-wide association study, deep learning, GmRVE4d, circadian clock, maturity

Posted Date: March 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9015910/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on September 8th, 2026. See the published version at https://doi.org/10.1038/s41467-026-77316-3.

Natural variation in *GmRVE4d* facilitates soybean latitudinal adaptation by regulating *GmPRR5a*-dependent flowering and maturity

Baiquan Sun[a,b†], Tingting Wu[a†], Bingjun Jiang[a†], Xuegang Sun[a], Hongchang Jia[c], Mahmoud Naser[a], Maoxiang Yang[a], Xiangyu Yao[a], Yanhui Sun[a], Qimeng Li[a], Peiguo Wang[a,b], Liwei Jiang[a,b], Shibo Sun[a,b], Chao Qin[a], Shan Yuan[a], Kunhui He[a,b], Wensheng Hou[a], Cunxiang Wu[a], Huihui Li[a,b*], Shi Sun[a,*] & Tianfu Han[a,b*]

[a] The State Key Laboratory of Crop Gene Resources and Breeding, Ministry of Agriculture and Rural Affairs Key Laboratory of Soybean Biology (Beijing), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[b] National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya 572024, Hainan, China

[c] Heihe Branch of Heilongjiang Academy of Agricultural Sciences, Heihe 164300, Heilongjiang, China

[†] **These authors contributed equally to this work.**

[*] **Correspondence:**

E-mail addresses: hantianfu@caas.cn (T. Han); sunshi@caas.cn (S. Sun); lihuihui@caas.cn (H. Li)

Abstract

Timely flowering and maturity are crucial for plant reproduction and environmental adaptation. As a core molecular system for plants to perceive environmental cues, the circadian clock integrates cyclic photothermal signals with intrinsic developmental programs, exerting a pivotal regulatory role in these processes. Soybean is a short-day thermophilic crop sensitive to photoperiod and temperature, but the molecular mechanism of the circadian clock in soybean photothermal adaptation remains largely elusive. Herein, we performed genome-wide association study (GWAS) combined with deep learning approach with maturity variations in soybean, and first identified *GmRVE4d*, a homologue of the circadian clock gene *REVEILLE*, which significantly associated with maturity in different photothermal conditions. Genetic analysis revealed that *GmRVE4d* underwent domestication-related artificial selection, and its favorable allele $GmRVE4d^{H3}$ was characterized. Further functional analyses demonstrate that GmRVE4d directly binds to the promoter of *GmPRR5a*, the soybean ortholog of *Arabidopsis* circadian clock central component *PSEUDO-RESPONSE REGULATOR5* (*PRR5*), and represses its transcriptional expression, thereby delaying soybean flowering and maturity. Thus, *GmRVE4d* is a novel negative regulator of soybean flowering and maturity that functions by directly repressing *GmPRR5a* expression, and is a promising molecular target for improving soybean photothermal adaptation and expanding cultivation latitudes through molecular breeding.

Keywords: soybean (*Glycine max* (L.) Merr.); genome-wide association study; deep learning; *GmRVE4d*; circadian clock; maturity

Introduction

Flowering and maturity are the key traits for ecological adaptation in plants, jointly regulated by endogenous gene networks and photothermal environments (Li *et al*., 2022). As core components of the photothermal molecular regulatory pathway, circadian clock genes play pivotal roles in mediating photothermal responses and shaping genetic diversity across plant species (Bu *et al*., 2021; Xu *et al*., 2022).

Soybean (*Glycine max* (L.) Merr.) is originally domesticated in the temperate

regions of China between latitudes 32°N and 40°N, has subsequently spread globally and is now cultivated from 53°N to 35°S (Zhang *et al*., 2020; Zhu *et al*., 2025; Jiang *et al*., 2026). It is a typical short-day crop sensitive to photothermal cues. A majority of varieties switch from vegetative growth to reproductive growth after photoperiod is shortened to a certain threshold (Garner and Allard, 1920; Yang *et al*., 2019). Therefore, photothermal sensitivity is a key trait influencing the flowering time, maturity, and ecological adaptability of soybean (Lin *et al*., 2021; Fang *et al*., 2024). In the photoperiod flowering pathway, plants detect light changes through phytochromes (phyA-phyE), cryptochromes (cry1, cry2), phototropins (phot1 and phot2), ZEITLUPE (ZTLs) and UV-B receptors (Kottke *et al*., 2018; Gendron and Staiger, 2023). These signals are transmitted to the circadian clock via the ELF3, TIC1 and FHY3 for integration and conversion into new signals (Su *et al*., 2025). Ultimately, flowering time is regulated via the *GI-CO-FT* (*GIGANTEA - CONSTANS - FLOWERING LOCUS T*) pathway (Cao *et al*., 2015).

The circadian clock, is an endogenous molecular oscillator that regulates the circadian rhythm of plants with a period of approximately 24 hours, endowing plants with adaptive advantages (Li *et al*., 2025). The circadian clock of *Arabidopsis* is a complex network of interlocking feedback loops involving both repressor and activator transcription factors (McClung, 2019). Morning-phase genes include *CIRCADIAN CLOCK-ASSOCIATED 1* (*CCA1*), *LATE ELONGATED HYPOCOTYL* (*LHY*), *REVEILLEs* (*RVE4, RVE6*, *RVE8*), *NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED* (*LNK1, LNK2*) and *PSEUDO-RESPONSE REGULATORs* (*PRR9, PRR7*). Evening-phase genes include *PRR5*, *PRR3*, *PRR1* (*TIMING OF CAB EXPRESSION 1*, *TOC1*), *ZTL*, *GIGANTEA* (*GI*), and evening complex (EC) formed by *EARLY FLOWERING* (*ELF3, ELF4*) and *LUX ARRHYTHMO/PHYTOCLOCK1* (*LUX/LC1*) (Xu *et al*., 2022; Liu *et al*., 2026). RVEs are key components of the plant circadian clock system and belong to a small gene family encoding MYB domain-containing transcription factors (Rawat *et al*., 2011). The MYB protein family

is a group of functionally diverse transcriptional activators in plants, characterized by a highly conserved N-terminal domain consisting of 1-4 tandem, incompletely R motifs (Rosinski and Atchley, 1998). This family holds potential as transcription factors for plant breeding and improvement (Zhang *et al*., 2025). In *Arabidopsis*, *RVE8* regulates the circadian rhythm antagonistically to *LHY*/*CCA1* and promotes the expression of *PRR5* and *TOC1* as well as other evening genes, including *LUX*, *ELF4*, and *GI*, by directly binding to the evening element (EE) in their promoters (Sorkin *et al*., 2022). It is also implicated in controlling flowering time (Rawat *et al*., 2011; Hsu *et al*., 2013). Previous studies have shown that *RVE4*, *6* and *8* exhibit partially redundant functions in accelerating the circadian clock pace (Hsu *et al*., 2013; Scandola *et al*., 2022). Interestingly, flowering time is slightly delayed in *rve4 rve6 rve8* triple mutants, a phenotype opposite to that of *rve8* single mutant (Rawat *et al*., 2011; Gray *et al*., 2017).

In recent years, at least thirteen circadian clock genes in soybean have been identified and functionally characterized, including *GmELF3*/*J* (Yue *et al*., 2016; Lu *et al*., 2017), *Gp1*/*qFT12-1*/*Tof12*/*GmPRR37*/*GmPRR3b* (Li *et al*., 2019; Li *et al*., 2020; Lu *et al*., 2020; Wang *et al*., 2020a), *Tof11*/*Gp11*/*GmPRR3a* (Lu *et al*., 2020), *Tof16*/*GmLHY1a*, *GmLHY1b*/*E11*, *GmLHY1c* and *GmLHY1d* (Dong *et al*., 2021), *LUX1* and *LUX2* (Bu *et al*., 2021; Fang *et al*., 2021), *E2* (Wang *et al*., 2022), *GmTOC1b* (Zhang *et al*., 2023), *GmEID1* (Qin *et al*., 2023) and *Dgt1*/*GmRVE1* (Cheng *et al*., 2026). These genes are all involved in the photoperiod-dependent regulation of soybean growth. Recent studies have uncovered that under long-day (LD) conditions, *E2* interacts with the blue-light receptor Flavin-binding, Kelch repeat, F box 1 (FKF1), inducing the degradation of GmELF3/J. The EC also suppresses *E2* expression by binding to the *E2* promoter, eventually forming a soybean photoperiod-dependent regulatory loop (Zhao *et al*., 2024). In addition, Tof11 and Tof12 directly suppress *GmRVE1* expression, which represses flowering dependent on the *E1* (Cheng *et al*., 2026). However, the circadian clock in soybean exhibits a

complex regulatory mechanism, and the underlying molecular regulatory network still remains to be further elucidated. For instance, it remains unclear to date what the function of *GmRVE4s* is and what its downstream regulatory mechanism is.

In this study, a circadian clock gene, *GmRVE4d*, was identified as the candidate gene for a major maturity QTL through combining genome-wide association study (GWAS) and a deep learning (DL) approach, DeepGOPlus. *GmRVE4d* encodes a homologous protein of *Arabidopsis* RVE4 and RVE8. Through analysis of CRISPR-Cas9 knockout mutants, phenotyping of transformants, and investigation of its regulatory mechanism, we found that *GmRVE4d* regulates flowering time, maturity and soybean growth period structure by inhibiting the transcriptional expression of *GmPRR5a* and its haplotype $GmRVE4d^{H3}$ exerts a more potent regulatory effect than $GmRVE4d^{H1}$. Furthermore, the nucleotide polymorphisms in the promoter region of *GmRVE4d* have undergone strong artificial selection during domestication. These results suggest that *GmRVE4d* is not only an inhibitor of flowering and maturity, primarily acting during the soybean vegetative growth stage, but also may be involved in the formation of environmental adaptation and the domestication process from wild soybeans (*G. soja*) to cultivated soybeans.

Results

Phenotypic analysis of maturity in multiple environments

The maturity of 257 accessions, representing various maturity groups worldwide (Supplementary Data 1), was evaluated in the field in Beijing (2020 and 2021) (BJ2020 and BJ2021) and Xinxiang, China (2021) (XX2021). The results revealed significant differences in maturity phenotypes between locations, but no significant differences between years at the same location (Fig. 1a). Additionally, the maturity phenotypes of different accessions differed significantly ($P < 0.05$) across all tested environments (Supplementary Fig. 1). Furthermore, maturity was significantly delayed ($P < 0.05$) with the advancement of maturity groups, demonstrating that

soybean accessions from different maturity groups exhibit abundant natural variation in maturity-related traits.

Resequencing and genomic variation of soybean accessions from different maturity groups

Whole-genome resequencing was performed on a panel of 257 accessions on the Illumina NovaSeq platform, yielding an average clean read depth of 9× per accession. The clean reads were aligned to the soybean reference genome (*Glycine max* Wm82.a2.v1), achieving an average mapping coverage of 94% (Supplementary Data 2). Following stringent quality control and filtering, a total of 2,757,937 high-quality single-nucleotide polymorphisms (SNPs) distributed across the 20 soybean chromosomes, along with 315,478 insertions/deletions (InDels), were identified (Supplementary Fig. 2).

Genetic relationships among the 257 accessions were investigated using multiple analytical approaches, including principal component analysis (PCA), geographical distribution analysis, phylogenetic analysis, and kinship assessment. These analyses consistently revealed three distinct groups: Northeast China, Huanghuaihai, and Southern China accessions, reflecting the major soybean-growing regions in China (Fig. 1b and Supplementary Fig. 3, 4).

Linkage disequilibrium (LD) decay patterns in the population were further characterized. LD, quantified by the squared correlation coefficient (r^2) between SNP alleles, decreased as the physical distance between SNPs was increased. The LD decay distance was defined as the physical distance at which r^2 declines to the critical value of 0.4, was approximately 50 kb (Supplementary Fig. 5).

Identification of maturity-related QTLs by GWAS

Using phenotypic data collected from BJ2020, BJ2021 and XX2021, a genome-wide association study (GWAS) was conducted for the maturity traits. We identified a consistently significant QTL region ($-\log_{10} P \geq 6$) on Chr. 19 across all three environments, indicating that this genomic region plays a robust and critical role in

controlling soybean maturity (Fig. 1c, Supplementary Fig. 6 and Supplementary Data 3). Other significant associations detected on different chromosomes likely correspond to known QTLs (*EMF1*, *E1*, *FR1*, *E2*, *TFL3*, *SOC1* and *MYB33*) and novel QTLs, but these are not explored further here.

Linkage disequilibrium (LD) block structure analysis revealed that Chr 19:43,721,425, Chr 19:43,687,363 and Chr 19:43,684,936 are located within the same genetic locus ($r^2 \geq 0.4$) and span an 84.1-kb interval (43.672 - 43.756 Mb) within the candidate genomic region (Supplementary Fig. 7). This interval contains 9 annotated genes according to the soybean reference genome (*Glycine max* Wm82.a2.v1) (Supplementary Data 4).

Identification of the new locus controlling maturity

We first identified the candidate genes using DeepGOPlus (Yang *et al*., 2023). Gene Ontology (GO) term predictions by DeepGOPlus revealed that only *Glyma.19G178000* is involved in the photoperiod response and flowering regulation processes (Fig. 1d and Supplementary Data 5). Then, we performed quantitative real-time PCR (qRT-PCR) to analyze the transcriptional profiles of candidate genes in the Jack cultivar. Under both short-day (SD, 12 h: 12 h, light: dark) and long-day (LD, 16 h: 8 h, light: dark) conditions, only *Glyma.19G178000* showed a circadian rhythmic expression pattern, indicating that *Glyma.19G178000* expression is regulated by the circadian clock (Fig. 1e-f and Supplementary Fig. 8). Further analysis indicated that *Glyma.19G178000* mRNA was highly expressed in trifoliate leaves, followed by flowers and true leaves (Supplementary Fig. 9), which are key tissues for perceiving environmental light signals in soybean. The GmRVE4d-GFP fusion protein was localized to the nucleus (Supplementary Fig. 10), suggesting that GmRVE4d may participate in the regulation of the circadian rhythm as a transcription factor in soybean. Moreover, *Glyma.19G178000* is homologous to *Arabidopsis RVE4* and *RVE8*, contains an R motif encoded by 50 amino acids, and belongs to the 1R-MYB subfamily (Supplementary Fig. 11), which was thus designated as

GmRVE4d. We hypothesized that *GmRVE4d* might play a role in regulating soybean flowering and maturity and thus focused on *GmRVE4d* as a high-confidence candidate gene for the new maturity locus.

Natural variations and haplotype analysis of *GmRVE4d*

To validate the function of *GmRVE4d*, we investigated its nucleotide polymorphisms (SNPs/InDels) and identified 10 SNPs and 1 InDel in its coding sequence (CDS). Among these polymorphisms, a 2-bp (TG) deletion at position +1707 bp resulted in a frameshift mutation (Fig. 2a).

Notably, the maturity phenotypes of accessions carrying $GmRVE4d^{H3}$ exhibit significantly delayed maturity compared with those carrying $GmRVE4d^{H1}$ across all three environments ($P < 0.01$) (Fig. 2d), possibly due to an amino acid substitution at position 77 in the conserved domain (Ser to Asn) (Fig. 2b and Supplementary Fig. 12).

Geographical distribution analysis revealed that $GmRVE4d^{H1}$ was predominant in the northern region (NR), while $GmRVE4d^{H3}$ was more prevalent in the southern region (SR) (Fig. 2c); the other haplotypes with low sample sizes were not further investigated. Additionally, to explore the domestication dynamics of *GmRVE4d*, we compared the nucleotide diversity (F_{ST} and π) in the 2.7-kb promoter region of *GmRVE4d* across 257 accessions. The results from F_{ST} and π showed that the nucleotide diversity of cultivated soybeans was significantly lower than that of wild soybeans, suggesting that a strong selective sweep has occurred in the chromosomal region harboring *GmRVE4d* during the domestication process (Fig. 2e, f). These findings collectively indicated that *GmRVE4d* is likely the causal gene underlying soybean maturity variation.

***GmRVE4d* acts as a flowering and maturity inhibitor**

To further verify the role of *GmRVE4d* in controlling flowering time and maturity, we employed CRISPR-Cas9 technology to create *Gmrve4d* mutants in the Jack background. From these transformations, we identified three homozygous *Gmrve4d*

mutants for further phenotypic analysis (Supplementary Fig. 13).

Under SD conditions, the time from emergence to flowering of three independent mutant lines was significantly advanced by 2-3 days compared with the Jack ($P < 0.05$), and the maturity was significantly advanced by 9-10 days ($P < 0.05$) (Supplementary Fig. 14). Under LD conditions, the flowering time of three independent mutant lines was significantly advanced by 4-7 days ($P < 0.05$), and the maturity was significantly advanced by 11 days ($P < 0.05$) (Supplementary Fig. 14).

Under natural long-day (NLD) conditions (June to October) in Beijing (40° 23' N, 116° 57' E) in 2023, the flowering time of three independent mutant lines was significantly advanced by 5 days ($P < 0.05$) (Fig. 3a, b), and the maturity was significantly advanced by 5-7 days ($P < 0.05$) (Fig. 3a, d). Similarly, under NLD conditions in Xinxiang (35° 10' N, 113° 48' E) in 2023, the flowering time of three independent mutant lines was significantly advanced by 2-3 days ($P < 0.05$) (Fig. 3f), and the maturity was significantly advanced by 4-5 days, respectively ($P < 0.05$) (Fig. 3h). Consistent phenotypic results were observed in the 2025 field trial (Fig. 3c, e, g, i).

To further confirm the role of *GmRVE4d* in regulating flowering time and maturity, we generated a plant expression construct containing the CDS of $GmRVE4d^{H3}$ driven by the constitutive *CaMV35S* promoter, and introduced it into the soybean cultivar Jack via Agrobacterium-mediated transformation. qRT-PCR analysis revealed a significant increase in *GmRVE4d* transcription levels in T_3 homozygous transgenic plants (*GmRVE4d-OE* lines), showing transcript levels 77-, 30-, and 80-fold higher than those in the Jack plants, respectively (Supplementary Fig. 15). Furthermore, Western blot analysis confirmed the presence of GmRVE4d-GFP fusion proteins in the *GmRVE4d-OE* lines (Supplementary Fig. 16).

Under SD conditions, the flowering times of *GmRVE4d-OE* lines were significantly delayed by 2-3 days compared to the Jack ($P < 0.05$) , and the maturity was significantly delayed by 4-6 days ($P < 0.05$) (Supplementary Fig. 17). Under LD

conditions, the flowering time of *GmRVE4d-OE* lines was significantly delayed by 8-9 days compared to Jack plants ($P < 0.05$), and the maturity was significantly delayed by 10-13 days ($P < 0.05$) (Supplementary Fig. 17).

Under natural long-day (NLD) conditions (June to October) in Beijing in 2023, the flowering time of *GmRVE4d-OE* lines was significantly delayed by 5-9 days ($P < 0.05$) (Fig. 3a, b), and the maturity was significantly delayed by 7 days ($P < 0.05$) (Fig. 3a, d). Similarly, under NLD conditions in Xinxiang in 2023, the flowering time of *GmRVE4d-OE* lines was significantly delayed by 10-14 days ($P < 0.05$) (Fig. 3f), and the maturity was significantly delayed by 11-14 days, respectively ($P < 0.05$) (Fig. 3h). Consistent phenotypic results were observed in the 2025 field trial (Fig. 3c, e, g, i). The above results demonstrate that *GmRVE4d* functions as a photoperiod-independent flowering and maturity regulator in soybean. However, the phenotypic effect of *Gmrve4d* mutant lines on flowering time and maturity was weaker than that of *GmRVE4d-OE* lines. This difference may be related to the fact that soybeans require a certain time to complete biomass accumulation during the vegetative stage, and the loss-of-function mutation of *GmRVE4d* only partially relieves the transcriptional repression of *GmPRR5a*.

Notably, the growth period structure (Reproductive/Vegetative, R/V) ratio of Jack, *GmRVE4d-OE*, and *Gmrve4d* mutant lines under different photoperiod conditions was the highest in the *Gmrve4d* mutant lines, followed by the Jack and *GmRVE4d-OE* (Supplementary Fig. 18). These results indicate that GmRVE4d predominantly modulates the vegetative growth stage in soybean, thereby regulating maturity by affecting flowering time.

GmRVE4d directly inhibits the transcription of *GmPRR5a*

To gain further insight into how *GmRVE4d* delays flowering and maturity, we performed RNA-seq to compare the genome-wide gene expression profiles in the Jack, *GmRVE4d-OE,* and *Gmrve4d* mutant lines (Supplementary Data 6). A Venn diagram revealed 209 commonly differentially expressed genes in the *GmRVE4d-OE* and

Gmrve4d mutant lines compared with the Jack (Supplementary Fig. 19). Subsequently, DNA Affinity Purification Sequencing (DAP-seq) was performed to identify the direct regulatory targets of GmRVE4d at the whole-genome scale (Supplementary Fig. 20). By integrating the RNA-seq and DAP-seq results, 56 candidate genes were identified (Fig. 4a), among which *GmPRR5a* is the soybean ortholog of *PSEUDO-RESPONSE REGULATOR5* (*PRR5*), a central component of the core circadian clock genes from *Arabidopsis*. MEME analysis of GmRVE4d binding peaks revealed a significantly enriched motif with a 10-bp conserved core sequence (Fig. 4b, c). Expression pattern analysis further revealed that *GmPRR5a* displayed a circadian expression pattern opposite to that of *GmRVE4d* under both SD and LD conditions (Fig. 1e-f and Supplementary Fig. 21). Therefore, by combining the results of RNA-seq and DAP-seq, *GmPRR5a* was selected as a candidate target gene.

To determine whether GmRVE4d binds to the *GmPRR5a* promoter, we performed a Dual-Luciferase reporter assay, which further confirmed that GmRVE4d could directly bind to the *GmPRR5a* promoter to inhibit reporter gene expression (Fig. 4d-f). We also conducted electrophoretic mobility shift assays (EMSA) to analyze the in vitro binding of GmRVE4d to the corresponding synthetic oligonucleotide probes (5′-GATATTTTTT-3′) containing the conserved sequences in the promoter regions of *GmPRR5a* (Fig. 4g).

We performed a chromatin immunoprecipitation (ChIP)-qPCR experiment using transgenic *35Spro:GmRVE4d-GFP* plants. Trifoliate leaves from transgenic plants grown under SD photoperiod were harvested and subjected to ChIP-qPCR with an immunoglobulin G (IgG) antibody. GmRVE4d enrichment was detected in the *GmPRR5a* promoter (Fig. 4h), further supporting that GmRVE4d might bind the promoter of *GmPRR5a*.

In terms of phenotype effects, $GmRVE4d^{H3}$ exerted a significantly stronger effect than $GmRVE4d^{H1}$. To further confirm our hypothesis, we performed Yeast One-Hybrid (Y1H) assays. The results indicate that GmRVE4d can directly bind to

the *GmPRR5a* promoter in yeast, and *GmRVE4d*^{H3} similarly exerted a stronger binding effect (Fig. 4i).

In conclusion, *GmRVE4d* directly binds to the promoter region of its downstream target gene *GmPRR5a* to inhibit its transcriptional expression, and natural variation has led to the enhanced functional activity of *GmRVE4d*^{H3}. This renders soybean more sensitive to photothermal cues, resulting in late flowering and maturity, and ultimately improves the adaptive ability of soybean in low-latitude regions.

A novel *GmRVE4-GmPRR5a-GmFT2a/GmFT5a* regulatory cascade modulates flowering time and maturity in soybean

To further explore whether *GmPRR5a* regulates soybean flowering time and maturity, we used CRISPR-Cas9 technology to generate *Gmprr5a* mutants in the Jack background. From these transformations, we identified two homozygous *Gmprr5a* mutants for further phenotypic analysis (Supplementary Fig. 22).

Under SD conditions, the time from emergence to flowering of two independent mutant lines was significantly delayed by 1-2 days compared with the Jack ($P < 0.05$) (Fig. 5a, b), and the maturity was significantly delayed by 8 days ($P < 0.05$) (Fig. 5a, e). Under LD conditions, the flowering time of the two independent mutant lines was significantly delayed by 5-6 days ($P < 0.05$) (Fig. 5c), and the maturity was significantly delayed by 8-9 days ($P < 0.05$) (Fig. 5f).

Under natural short-day (NSD) conditions (August to October) in Sanya (18°14' N, 109°13' E) in 2025, the flowering time of two independent mutants lines was significantly delayed by 2 days ($P < 0.05$) (Fig. 5d), and the maturity was significantly delayed by 4 days ($P < 0.05$) (Fig. 5g).

Expression analysis further revealed that compared with the Jack, the transcript levels of *GmFT2a* and *GmFT5a* in *gmprr5a* mutant plants were significantly downregulated ($P < 0.05$) (Supplementary Fig. 23). This indicates that the *GmRVE4d-GmPRR5a* pathway also regulates soybean flowering and maturity by modulating *GmFT2a/GmFT5a* expression. However, the specific mechanism by

which *GmPRR5a* regulates *GmFT2a/GmFT5a* still remains to be further investigated.

Discussion

GmRVE4d is a major gene affecting maturity and adaptability of soybean

Flowering time and maturity are crucial traits for soybean ecological adaptation, and are tightly regulated by photoperiod-temperature interactions (Mao *et al*., 2017). Photoperiod determines the progression through developmental stages, while temperature influences the rate of development (growth) (Han, 2007). In *Arabidopsis*, *RVE4* and *RVE8* play key roles in responding to different photoperiods and temperature stresses (Gray *et al*., 2017; Kidokoro *et al*., 2023). *GmRVE4d*^{H3} is a domestication-selected gene that confers adaptive advantages to soybean in low-latitude regions for adapting to photothermal changes, whereas *GmRVE4d*^{H1} lacks such adaptive capacity in these regions. These findings imply that *GmRVE4d* is a major gene controlling soybean maturity and adaptability.

In soybean, the majority of circadian clock genes function as negative regulators of flowering and maturity, including *E2*, *GmPRR3a*, *GmPRR3b*, and *GmLHY1a*, while a few (*GmELF3* and *GmLUX)* act as positive regulators (Wu *et al*., 2023). These genes exhibit distinct expression patterns under different photoperiod conditions, which contributes to their distinct regulatory effects. In this study, we identified *GmRVE4d* as a novel negative regulator of flowering and maturity in soybean, while its downstream target gene *GmPRR5a* acts as a positive regulator. Moreover, our findings reveal that *GmRVE4d*, much like *GmPRR3b* and *GmELF3*, exerts a strong negative regulatory influence on soybean flowering and maturity processes, which further enriches our understanding of the complex gene regulatory network controlling these crucial developmental stages in soybean (Yue *et al*., 2016; Wang *et al*., 2020a; Lu *et al*., 2020).

Previous research has highlighted maturity as a fundamental domestication trait (Lu *et al*., 2022), with the gradual shift in the core growth period structure from wild

to cultivated soybeans driven by artificial selection for higher yield (Han *et al*., 1998; Zhu *et al*., 2012). Furthermore, the latitudinal expansion of soybean is associated with the selection of new alleles of core circadian clock genes (Gerhardt and Mehta, 2025). Therefore, GmRVE4d not only functions as a regulator of flowering and maturity-primarily during the vegetative growth stage-but its haplotypic variation also serves as the genetic basis for soybean environmental adaptability and the domestication process from wild soybeans.

GmRVE4d provides a new molecular target for improving ecological adaptability of soybean

Improving soybean adaptability across different latitudes and expanding its cultivation area is a crucial strategy to meet the increasing consumer demand for soybean products. Previous studies have found that the wide distribution of soybean results from rich natural variations and different combinations of genes and QTLs regulating flowering time and maturity (Wang *et al*., 2020a). Although several major genes associated with maturity traits have been cloned and identified (Wu *et al*., 2023), there is still an urgent need for a large number of photoperiod-insensitive super-early maturity varieties in high-latitude regions. Therefore, uncovering new major genes that regulate soybean maturity could unlock its potential for future genetic improvements, thereby enhancing its ecological adaptability.

In recent years, researchers have proposed a modern breeding strategy called "Potalaization", which allows breeding of widely adapted soybean varieties through the use of multiple molecular tools in existing elite widely adapted varieties (Qi *et al*., 2021). We found that the CRISPR-Cas9-mediated targeted mutagenesis of *GmRVE4d* in the main soybean cultivar Jack resulted in *Gmrve4d* mutants that displayed early maturity by approximately 7 days under NLD conditions (Fig. 3d). Based on this finding, CRISPR-Cas9 technology could be utilized to create *Gmrve4d* mutants in Jack or other elite soybean varieties. This advancement allows these varieties to be introduced for cultivation in higher-latitude regions or planted in spring in multiple

cropping systems. In addition, transgenic overexpression of *GmRVE4d* significantly delayed flowering and maturity when sown in summer in Beijing and Xinxiang (Fig. 3). Thus, *GmRVE4d* could be used to breed cultivars with a longer growth period and delayed maturity in temperate regions.

Molecular mechanism underlying *GmRVE4d*-mediated regulation of flowering and maturity in soybean

This study demonstrated that GmRVE4d directly binds to the promoter of *GmPRR5a* and inhibits its transcriptional expression, as shown by the Y1H, EMSA, ChIP-qPCR, and Dual-Luciferase reporter assays. In *Arabidopsis*, members of the Groucho/Tup1 corepressor family, TOPLESS/TOPLESS-RELATED (TPL/TPR), interact with PRR5 protein at the *CCA1* and *LHY* promoters to repress transcription and alter circadian period (Wang *et al.*, 2013).

Furthermore, previous studies have demonstrated that *GmLCLs*/*GmLHYs* mutations can result in late flowering in soybean (Wang *et al*., 2020b; Dong *et al*., 2021). Therefore, GmPRR5a may activate its downstream target genes *GmLCLs/GmLHYs*, which may have distinct functions in soybean compared to their orthologs in *Arabidopsis*. Notably, LNK1 is an essential transcriptional coactivator for the transcriptional expression of the circadian clock gene *PRR5*, through recruitment to the *PRR5* promoters via interaction with bona fide DNA binding proteins such as RVE4 and RVE8 (Xie *et al*., 2014; Gray *et al*., 2017). Additionally, AlphaFold3-based structural predictions (Abramson *et al*., 2024) suggested that GmRVE4d may exert its regulatory function by recruiting GmLNK1/GmLNK2 to form a protein complex (Supplementary Fig. 24). However, whether GmRVE4d physically interacts with GmLNK1 and GmLNK2 proteins remains to be further validated by molecular experiments like co-immunoprecipitation (Co-IP) or bimolecular fluorescence complementation (BiFC), and whether *GmPRR5a* regulates the expression of downstream genes via E2/EC and thereby affects the flowering and maturity processes of soybean remains to be further clarified (Zhao *et al*., 2024).

Therefore, further studies of *GmRVE4d* and *GmPRR5a* will help clarify the molecular mechanism of the circadian clock regulatory network and the photothermal regulation of flowering and maturity in soybean.

Taken together, these results support a regulatory model (Fig. 6) that integrates photothermal signal inputs, *GmRVE4d* haplotypic variation, and downstream cascades: Photothermal signals trigger the expression of *GmRVE4d* alleles, including $GmRVE4d^{H3}$ (the favorable variant from low-latitude accessions) and $GmRVE4d^{H1}$ (the favorable variant from high-latitude accessions). Both alleles physically bind to the promoter of *GmPRR5a* to repress its transcription, with $GmRVE4d^{H3}$ exerting a stronger inhibitory effect. Reduced *GmPRR5a* expression consequently downregulates the transcription of *GmFT2a/GmFT5a* (core flowering-promoting genes), ultimately driving late flowering and maturity. In contrast, loss-of-function mutation of GmRVE4d relieves the transcriptional repression of *GmPRR5a*, thereby activating the expression of *GmFT2a/GmFT5a* and accelerating flowering and maturity in soybean.

Methods

Plant materials, growth conditions and phenotyping

A panel of 257 accessions was selected from 14 consecutive maturity groups (MG 0000 – X), including 10 wild soybeans (*G. soja*), 33 landraces, and 214 cultivated soybeans. The soybean seeds were provided by the Ministry of Agriculture and Rural Affairs Key Laboratory of Soybean Biology (Beijing), ICS, CAAS. The 257-accession panel was grown under natural daylength conditions in Beijing (40° 23' N, 116° 57' E) in 2020 and 2021 and in Xinxiang (35° 10' N, 113° 48' E) in 2021, China, to evaluate the maturity phenotype. To accurately evaluate the flowering and maturity phenotypes, Jack, *GmRVE4d* transformants, and CRISPR-Cas9 knockout mutants were grown under natural field conditions from in Beijing and Xinxiang from June to October, in Sanya (18°14' N, 109°13' E) from August to October, as well as under controlled growth chamber conditions: short-day (SD) conditions (light: dark, 12 h: 12 h, 26°C) and long-day (LD) conditions (light: dark, 16 h: 8 h, 26°C). Each

location-year combination was regarded as an independent experimental environment, and these eight environments were named BJ2020, BJ2021, BJ2023, BJ2025, XX2021, XX2023, XX2025 and SY2025. Flowering time was recorded at the R1 stage (days from emergence to the first one open flower at any nodes on the main stem). Maturity was recorded at the R7 stage (days from emergence to the stage at which one normal pod on the main stem has reached its mature pod color) (Fehr and Carviness 1977). The reproductive/vegetative (R/V) growth period ratio was used as a key indicator to characterize the maturity structure (Sun *et al*., 1990).

Primers and accession numbers

All primers used in this study are listed in Supplementary Data 7. Gene sequences can be obtained from the Phytozome database (*Glycine max* Wm82.a2.v1) under the following accession numbers: *GmRVE4d* (*Glyma.19G178000*), *GmPRR5a* (*Glyma.19G260400*), *GmFT2a* (*Glyma.16G150700*) and *GmFT5a* (Glyma.16G044100).

DNA extraction, sequencing, and SNP calling

Genomic DNA was extracted from each sample using the Plant Genomic DNA Kit (TIANGEN). Genome sequencing libraries were constructed using the DNA Library Prep Kit (Vazyme) and subsequently sequenced on the Illumina NovaSeq platform. Raw reads were quality-filtered and trimmed using Trimmomatic with the following parameters: ILLUMINACLIP:adaptor.sequence:2:30:10, TRAILING:3, LEADING:3, SLIDINGWINDOW:4:10, MINLEN:20 (Bolger *et al*., 2014). Clean reads were then aligned to the soybean reference genome (*Glycine max* Wm82.a2.v1) using BWA-MEM with the default parameters (Li and Durbin 2009). SNP calling and annotation were performed with the SAMtools (Li *et al*., 2009), BCFtools (Li 2011), and SnpEff (Cingolani *et al*., 2012).

Subsequently, only high-quality SNPs were retained with VCFtools (Danecek *et al*., 2011) based on the following filtering criteria: missing rate ≤ 0.1 (max-missing = 0.1), quality score (QUAL) ≥ 50.0, read depth (DP) ≥ 5.0, quality by depth (QD) ≥ 5.0, mapping quality (MQ) ≥ 30, genome coverage ≥ 90%, max/min-alleles = 2, and minor allele frequency (MAF) ≥ 0.01. Finally, PLINK (Purcell *et al*., 2007) and TASSEL (Bradbury *et al*., 2007) were used to format the sequencing data and filter out InDels for downstream genetic analyses.

Phylogenetic structure and linkage disequilibrium analysis

High-quality SNPs were used to construct a Neighbor-Joining (NJ) tree with SNPhylo software (Lee *et al*., 2014), the resulting phylogenetic tree was visualized using the online interactive tool iTOL (https://itol.embl.de). Principal component analysis (PCA) was performed using the 'Princomp' function in R (Price *et al*., 2006) and visualized with the 'Scatterplot3d' package (Ligges and Mächler 2003). Genome-wide SNP distribution density was analyzed with the 'CMplot' package in R (https://github.com/YinLiLin/R-CMplot). The kinship matrix was computed using the 'ComplexHeatmap' package in R and visualized as a heat map (Gu *et al*., 2016). Linkage disequilibrium (LD) calculations were conducted using PopLDdecay software. LD decay was estimated based on the squared correlation coefficient r^2 between pairwise SNPs, with values averaged in 1-kb sliding windows across a maximum distance of 1 Mb (Zhang *et al*., 2019). The LD decay threshold was defined as the physical distance at which r^2 declined to half of its maximum value (Huang *et al*., 2010).

GWAS for maturity

Genome-wide association study (GWAS) for maturity traits was performed using 2,757,937 high-quality SNPs in GAPIT v3 (http://zzlab.net/GAPIT) with the Mixed Linear Model (MLM). This model incorporated the first three principal components as fixed effects to correct for population structure, and accounted for individual relatedness among the 257 soybean accessions. The significant association threshold was set as $1/n$ (where n represents the total number of SNPs). Manhattan plots and quantile-quantile (QQ) plots were generated using the 'qqman' package in R (Turner, 2014).

Identification of candidate genes using a deep learning (DL) method

A deep learning tool, DeepGOPlus (Yang *et al*., 2023), was used to capture protein molecular features from amino acid sequences with a prediction threshold of 0.3. The model consists of an input layer, a convolutional layer, a pooling layer, and a fully connected layer. The input data is passed through a set of CNN layers with different filter sizes, each CNN layer containing 512 filters to learn a particular sequence of a

particular size. Subsequently, the MaxPooling layer extracts the maximum feature values from the representations acquired by each filter. Through this configuration, we obtained an output containing 8192 (16 × 512) convolution filters as sequence features. Finally, a sigmoid activation function combined with a fully connected layer is used for the final prediction and classification.

Haplotype and genetic diversity analysis

Haplotype analysis of SNPs and InDels within the exons of *GmRVE4d* was performed using Python scripts. The geographical distribution pattern of haplotypes was analyzed using PopART v1.7 software. Nucleotide diversity (π) and population differentiation (F_{ST}) values among wild soybeans, landraces, and cultivated soybean populations were calculated using VCFtools.

Sample collection, RNA isolation and qRT-PCR analysis

To investigate the tissue-specific expression pattern of *GmRVE4d,* samples were collected from various tissues (root, hypocotyl, cotyledon, stem, true leaf, trifoliolate leaf, growing point, flower, pod, and seed) at key developmental stages. For analysis of the rhythmic expression pattern of *GmRVE4d*, Jack plants were grown under SD and LD conditions in growth chambers for three weeks. Fully expanded trifoliolate leaves were collected every 4 hours over a 48-hour period under both photoperiod conditions. To compare the dynamic transcriptional levels of *GmRVE4d* among the Jack, *GmRVE4d* transformants, and CRISPR-Cas9 knockout mutants, fully expanded trifoliolate leaves were harvested from 3-week-old plants grown under SD conditions. Total RNA was extracted from all samples using the RNAprep Pure Plant Kit (TIANGEN Biotechnology, Beijing, China). Complementary DNA (cDNA) synthesis and genomic DNA (gDNA) removal were performed with the FastKing RT Kit with genomic DNA Eraser (TIANGEN Biotechnology, Beijing, China). Quantitative real-time PCR (qRT-PCR) was conducted on an ABI 7600 real-time PCR system using 384-well optical plates and SYBR® Premix Ex Taq™ II Kit (Takara Biotechnology, Shiga, Japan). *GmACTIN* was used as the internal reference gene to normalize the relative expression levels of target genes. All assays included three biological replicates with three technical replicates per biological replicate.

Plasmid construction and plant transformation

549 The coding sequence (CDS) of the candidate gene *GmRVE4d* was amplified from
550 Jack cultivar using KOD-One DNA Polymerase (TOYOBO Life Science, Osaka,
551 Japan) with primers listed in Supplementary data 6. PCR products were separated by
552 1.0% agarose gel electrophoresis and purified using the V-ELUTE Gel Mini
553 Purification Kit (ZOMANBIO, Beijing, China).

554 The purified PCR product was ligated into the pGEM-T Easy vector (Promega
555 Corporation, Madison, WI, USA) and verified by Sanger sequencing. Positive clones
556 with confirmed inserts were then digested with the restriction endonuclease *Xba* I, and
557 the target fragment was subcloned into the *PTF101.1-GFP-GmRVE4d* expression
558 vector. Target sequence adapters for *GmRVE4d* and *GmPRR5a* were designed using
559 the CRISPR-P v2.0 web tool (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR)
560 (Liu *et al*., 2017).

561 Based on their genomic locations and GC contents, one optimal target sequence
562 per gene was selected and integrated into a single guide DNA (sgDNA) expression
563 cassette, which was then cloned into the CRISPR-Cas9 binary vector. All
564 recombinant plasmids were individually transformed into competent cells of
565 *Agrobacterium tumefaciens* strain EHA105 and subsequently used to transform Jack
566 via the cotyledon-node method (Chen *et al*., 2018).

567 **Subcellular localization**

568 For the subcellular localization assay, the recombinant plasmid
569 *PTF101.1-GFP-GmRVE4d* was transformed into Agrobacterium tumefaciens strain
570 EHA105 and then transiently expressed in the leaves of 6-week-old *Nicotiana*
571 *benthamiana*, as described by Yang *et al*. (2000). Fluorescence images were acquired
572 48 hours post-agroinfiltration using a ZEISS LSM 700 laser scanning confocal
573 microscope (Carl Zeiss AG, Oberkochen, Germany) with excitation/emission
574 wavelengths of 488 nm/509 nm for GFP.

575 **Western blot**

576 To verify the expression of the target protein in transgenic soybean plants, total

proteins were extracted from Jack and *PTF101.1-GFP-GmRVE4d* overexpression transgenic lines using RIPA lysis buffer (radioimmunoprecipitation assay buffer) supplemented with 1 mM PMSF (phenylmethanesulfonyl fluoride, a protease inhibitor). Protein concentrations were determined using the Bradford assay, and immunoblot analyses were performed using HSP90 antibodies.

RNA-seq and data analysis

To compare the transcriptomic profiles of Jack, *GmRVE4d* overexpression transformants, and *Gmrve4d* CRISPR-Cas9 knockout mutants, soybean seedlings were grown in growth chambers under SD conditions for 13 days. On the 13th day under SD conditions, at the end of the dark period, fully expanded trifoliolate leaves were harvested and immediately frozen in liquid nitrogen. Three biological replicates were prepared for each sample. Total RNA extraction, cDNA library construction, Illumina NovaSeq sequencing, and raw data processing were performed as previously described. Differentially expressed genes were visualized using the 'VennDiagram' package in R.

DAP-Seq profiling and analysis

DAP-seq experiments were performed following a published protocol (O'Malley *et al*., 2016). Briefly, the GmRVE4d-HaloTag fusion protein was incubated with the constructed DNA library at 25℃ for 1 h, following which the GmRVE4d fusion protein was immobilized using Halo Tag beads and removal of unbound DNA fragments by washing. The GmRVE4d-binding DNA fragments were recovered and subjected to PCR analysis. An Illumina NovaSeq 6,000 instrument was used for sequencing, and reads were mapped to the Wm82a2.v1, using BWA-MEM. MACS2 callpeak and Homer software were used to merge the peaks of the two biological duplicates with $Q < 0.05$. Motif discovery was performed using the MEME-ChIP software. The bound peaks were annotated using ChIPseeker software (Yu *et al*., 2015).

Yeast one-hybrid

To verify the direct binding of GmRVE4d to the *GmPRR5a* promoter, the CDS of *GmRVE4d* was cloned into the *pB42AD* activation domain vector, while a 2-kb upstream promoter sequence of *GmPRR5a* was inserted into the *pLacZ2u* reporter vector. Primers used for fragment amplification are listed in Supplementary Data 7. Recombinant plasmids were cotransformed into yeast strain EGY48 competent cells using the EX-Yeast Transformation Kit (ZOMANBIO, Beijing, China). Transformed yeast cells were first cultured on SD/-Trp/-Ura selective medium at 30℃ for 3 days. Positive transformants were then transferred to SD/Raf/Gal/-Trp/-Ura/X-Gal indicator medium, and the binding activity of GmRVE4d to the *GmPRR5a* promoter was evaluated by monitoring the development of blue yeast colonies.

Electrophoretic mobility shift assay

The coding sequence of *GmRVE4d* was amplified and cloned into the pET-28a (+) expression vector. The restriction enzyme sites *NdeI* and *XhoI* were selected, and a HIS tag was added to the N-terminus of the fusion vector. EMSA analyses were performed using the LightShift Chemiluminescent EMSA Kit (Beyotime Biotechnology, Shanghai, China). Probe sequences are shown in Supplementary Data 7.

Chromatin immunoprecipitation analysis

Soybean samples were collected from the third open trifoliate leaves of *35Spro:GmRVE4d-GFP* plants. The ChIP experiment was performed as previously described (Gendrel *et al*., 2005). Samples were ground in liquid nitrogen and ChIP Lysis Buffer was added, then formaldehyde was added. Nuclei were isolated and sonicated as previously described (Gendrel *et al*., 2005). The soluble chromatin was immunoprecipitated by antibody against the GFP epitope. The immunoprecipitated DNA was recovered and analyzed by qRT-PCR in triplicate. Relative fold enrichment was calculated by normalizing the value of input DNA. The primers used for amplification are listed in Supplementary Data 7.

Dual-Luciferase Reporter Assay

To construct the effector vector, the CDS of *GmRVE4d* was cloned into the *pGreenII 62-SK* vector. The empty *pGreenII 62-SK* vector was used as a negative control for the effector. For reporter plasmid construction, a 2-kb upstream promoter sequence of *GmPRR5a* was amplified and cloned into the *pGreenII 0800-LUC* vector.

These plasmids were cotransformed into competent cells of *Agrobacterium tumefaciens* strain GV3101 (pSoup-P19). Firefly luciferase (LUC) and Renilla luciferase (REN) activities were measured using the Dual Glo Luciferase Reporter Gene Assay Kit (YEASEN Biotechnology, Shanghai, China). The relative activity of the GmPRR5a promoter was determined as the ratio of LUC to REN, with three biological replicates performed for each assay. The activities of firefly luciferase (LUC) and Renilla luciferase (REN) were quantified using the Dual Glo Luciferase Reporter Gene Assay Kit (YEASEN Biotechnology, Shanghai, China). The relative activity of the *GmPRR5a* promoter was determined by calculating the LUC to REN ratio from three biological replicates.

Statistical analyses

For phenotypic evaluation of soybean accessions, 12 individual plants were analyzed per accession. For qRT-PCR-based expression analyses, samples were prepared by pooling tissues from three individual plants per biological replicate, and three technical replicates were performed for each qRT-PCR reaction. Statistical analysis of mean values was performed using one-way analysis of variance (ANOVA) for multiple groups or two-tailed Student's t-tests for pairwise comparisons, as appropriate. All statistical analyses were conducted using SPSS software (v.23, IBM Corporation, Armonk, NY, USA), with $P < 0.05$ considered statistically significant. Details of statistical tests and significance levels are indicated in the relevant figure legends.

Acknowledgements

This work was supported by the National Key R&D Program of China (2023YFD1201300), the Hainan Provincial Sanya Yazhou Bay Science and

Technology Innovation Joint Project (ZDYF2025GXJS150) and the Postdoctoral Research Project Funding of Hainan Province (BH424090).

Conflict of interest

The authors declare that they have no conflicts of interest.

Data availability statement

All whole-genome sequencing data generated in this study are available from NCBI under accession number **PRJNA1114037**.

Author contributions

B.S. designed and performed the experiments, and drafted the manuscript. T.W. collected the natural population. B.J. provided the sequencing data. H.J., X.S., M.Y., X.Y., P.W., L.J. and Y.S. performed the phenotypic characterization. S.Y., M.N., C.Q., K.H., S.S., W.H. and C.W. critically revised the manuscript. T.H., S.S. and L.H. conceived and designed the research, and critically revised the manuscript.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O. *et al*. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature*, **630**, 493-500.

Bolger, A.M., Lohse, B. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114-2120.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T M., Ramdoss, Y. and Buckler, E.S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. **23**, 2633-2635.

Bu, T., Lu, S., Wang, K., Dong, L., Li, S., Xie, Q., Xu, X. *et al*. (2021). A critical role of the soybean evening complex in the control of photoperiod sensitivity and adaptation. *Proc*. *Natl. Acad*. *Sci*. *USA*. **118**, e2010241118.

Cao, D., Li, Y., Lu, S., Wang, J., Nan, H., Li, X., Shi, D. *et al*. (2015). GmCOL1a and GmCOL1b function as flowering repressors in soybean under long-day conditions. *Plant Cell Physiol.* **56**, 2409-2422.

689 Chen, L., Cai, Y., Liu, X., Yao, W., Guo, C., Sun, S., Wu, C. *et al*. (2018).
690 Improvement of soybean *Agrobacterium*-mediated transformation efficiency by
691 adding glutamine and asparagine into the culture media. *Int*. *J*. *Mol*. *Sci*. **19**,
692 3039.
693 Cheng, Q., Yuan, J., Huang, Y., Ye, W., Gao, C., Li, T., Zhong, W. *et al*. (2026).
694 Stepwise selection of *Tof11/12* and *GmRVE1* facilitates soybean adaptation to
695 high-latitude regions. *Cell Rep*. **45**, 117049.
696 Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M. and Lu, X.
697 (2012). Using *Drosophila melanogaster* as a model for genotoxic chemical
698 mutational studies with a new program, SnpSift. *Front*. *Genet*. **3**, 35.
699 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A.,
700 Handsaker, R.E. *et al*. (2011). The variant call format and VCFtools.
701 *Bioinformatics*. **27**, 2156-2158.
702 Dong, L., Fang, C., Cheng, Q., Su, T., Kou, K., Kong, L., Zhang, C. *et al*. (2021).
703 Genetic basis and adaptation trajectory of soybean from its temperate origin to
704 tropics. *Nat*. *Commun*. **12**, 5445.
705 Fang, X., Han, Y., Liu, M., Jiang. J., Li, X., Lian, Q., Xie, X. *et al*. (2021).
706 Modulation of evening complex activity enables north-to-south adaptation of
707 soybean. *Sci*. *China Life Sci*. **64**, 179-195.
708 Fang, C., Sun, Z., Li, S., Su, T., Wang, L., Dong, L., Li, H. *et al*. (2024).
709 Subfunctionalisation and self-repression of duplicated *E1* homologues finetunes
710 soybean flowering and adaptation. *Nat. Commun*. **15**, 6184.
711 Fehr, W.R. and Caviness, C.E. (1977). Stages of Soybean Development. Special
712 Report 80. Ames, Iowa: Iowa State University. 1-12.
713 Garner, W.W. and Allard, H.A. (1920). Effect of the relative length of day and night
714 and other factors of the environment on growth and reproduction in plants. *J*.
715 *Agric*. *Res*. **18**, 553-606.
716 Gerhardt, A. and Mehta, D. (2025). Engineering the plant circadian clock for
717 latitudinal adaptation as a strategy to secure agricultural productivity on a
718 changing planet. *New Phytol*. **248**, 68-75.
719 Gendrel, A.V., Lippman, Z., Martienssen, R., and Colot, V. (2005). Profiling histone
720 modification patterns in plants using genomic tiling microarrays. *Nat. Methods.* **2**,
721 213-218.

722 Gendron, J.M., and Staiger, D. (2023). New horizons in plant photoperiodism. *Annu.*
723 *Rev. Plant Biol*. **74**, 481-509.
724 Gray, J.A., Shalit-Kaneh, A., Chu, D.N., Hsu, P.Y. and Harmer, S.L. (2017). The
725 REVEILLE clock genes inhibit growth of juvenile and adult plants by control of
726 cell size. *Plant Physiol*. **173**, 2308-2322.
727 Gu, Z., Eils, R. and Schlesner, M. (2016). Complex heatmaps reveal patterns and
728 correlations in multidimensional genomic data. *Bioinformatics*. **32**, 2847-2849.
729 Han, T., Gai, J., Chen, F. and Qiu, J. (1998). Photoperiodic response and agronomic
730 characters of soybean varieties with different maturity structures. *Acta*
731 *Agronomica Sinica*. **24**, 550-557.
732 Han, T., (2007). Photoperiodism in soybean. In: Wang J, Guo Q (eds) Contemporary
733 soybean research in China. Jindu, Beijing, pp. 211-220.
734 Hsu, P.Y., Devisetty, U.K. and Harmer, S.L. (2013). Accurate timekeeping is
735 controlled by a cycling activator in *Arabidopsis*. *eLife*. **2**, e00473.
736 Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C. *et al*. (2010).
737 Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat.*
738 *Genet*. **42**, 961-969.
739 Jiang, X., Zhang, M., Yuan, X., Wang, L., Jiao, W., Mao, J., Ye, W. *et al*. (2026).
740 Population epigenomics reveals epigenetic drivers of replicated evolution and
741 missing heritability in soybean. *Mol. Plant*.
742 https://doi.org/10.1016/j.molp.2026.02.013
743 Kidokoro, S., Konoura, I., Soma, F., Suzuki, T., Miyakawa, T., Tanokura, M.,
744 Shinozaki, K., *et al*. (2023). Clock-regulated coactivators selectively control
745 gene expression in response to different temperature stress conditions in
746 *Arabidopsis*. *Proc*. *Natl. Acad*. *Sci*. *USA*. **120**, e2216183120.
747 Kottke, T., Xie, A., Larsen, D.S., and Hoff, W.D. (2018). Photoreceptors take charge:
748 Emerging principles for light sensing. *Annu. Rev. Biophys.* **47**, 291-313.
749 Lee, T.H., Guo,H., Wang, X., Kim, C. and Paterson, A.H. (2014). SNPhylo: a
750 pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*.
751 **15**, 162.
752 Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with
753 Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. *et al*. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**, 2078-2079.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. **27**, 2987-2993.

Li, Y., Dong, Y., Wu, H., Hu, B., Zhai, H., Yang, J. and Xia, Z. (2019). 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, C., Li, Y., Li, Y., Lu, H., Hong, H., Yu, T., Li, H. *et al*. (2020). A domestication-associated gene *GmPRR3b* regulates circadian clock and flowering time in soybean. *Mol*. *Plant*. **13**, 745-759.

Li, X., Liang, T., and Liu, H. (2022). How plants coordinate their development in response to light and temperature signals. *Plant Cell*. **34**, 955-966.

Li, J., Yang, M.K., Zeng, J., Chen, L., and Huang, W. (2025). Transcriptional activation and repression in the plant circadian clock: Revisiting core oscillator feedback loops and output pathways. *Plant Commun*. **6**, 101415.

Ligges, U. and Mächler, M. (2003). Scatterplot3d-An R package for visualizing multivariate data. *J. stat. softw*. **8**, 1-20.

Lin, X., Liu, B., Weller, J.L., Abe, J. and Kong, F. (2021). Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *J. Integr*. *Plant Biol*. **63**, 981-994.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K. and Chen, L. (2017). CRISPR-P 2.0: An improved CRISPR-Cas9 tool for genome editing in plants. *Mol*. *Plant*. **10**, 530-532.

Liu, M., Liu, Y., Gao, S., Liang, L., Xu, W., Huang, Z. *et al*. (2026). Temporal orchestration of PRRs fine-tunes circadian pacing and anticipates environmental cues. *Cell Rep*. **45**, 116787.

Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., Fang, C. *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., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., Chen, L. *et al*. (2020). Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat*. *Genet*. **52**, 428-436.

Lu, S., Fang, C., Abe, J., Kong, F. and Liu, B. (2022). Current overview on the genetic basis of key genes involved in soybean domestication. *aBIOTECH*. **3**, 126-139.

Mao, T., Li, J., Wen, Z., Wu, T., Wu, C., Sun, S., Jiang, B., *et al*. (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.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*. **165**, 1280-1292.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat*. *Genet*. **38**, 904-909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J. *et al*. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am*. *J*. *Hum*. *Genet*. **81**, 559-575.

Qi, X., Jiang, B., Wu, T., Sun, S., Wang, C., Song, W., Wu, C. *et al*. (2021). Genomic dissection of widely planted soybean cultivars leads to a new breeding strategy of crops in the post-genomic era. *Crop J*. **9**, 1079-1087.

Qin, C., Li, H., Zhang, S., Lin, X., Jia, Z., Zhao, F., Wei, X. (2023). GmEID1 modulates light signaling through the Evening Complex to control flowering time and yield in soybean. *Proc*. *Natl. Acad*. *Sci*. *USA*. **120**, e2212468120.

Rawat, R., Takahashi, N., Hsu, P.Y., Jones, M.A., Schwartz, J., Salemi, M.R., Phinney, B.S. *et al*. (2011). REVEILLE8 and PSEUDO-REPONSE REGULATOR5 form a negative feedback loop within the Arabidopsis circadian clock. *PLoS Genet*. **7**, e1001350.

Rosinski, J.A. and Atchley, W.R. (1998). Molecular evolution of the myb family of transcription factors: Evidence for polyphyletic origin. *J*. *Mol*. *Evol*. **46**, 74-83.

Scandola, S., Mehta, D., Li, Q., Rodriguez Gallo, M. C., Castillo, B., & Uhrig, R. G. (2022). Multi-omic analysis shows *REVEILLE* clock genes are involved in

carbohydrate metabolism and proteasome function. *Plant Physiol.* **190**, 1005-1023.

Sorkin, M.L., Tzeng, S.C., Romanowski, A., Kahle, N., Bindbeutel, R., Hiltbrunner, A., Yanovsky, M.J. *et al*. (2022). COR27/28 regulate the evening transcriptional activity of the RVE8-LNK1/2 circadian complex. *bioRxiv*. https://doi.org/10.1101/2022.05.16.492168.

Su, C., Qin, Y., Yu, Y., Wang, Y., He, Y., Davis, S. J., and Wang, L. (2025). A circadian morning complex modulates far-red light signaling in *Arabidopsis*. *Nat. Commun*. **16**, 9398.

Sun, Z., Tian, P. and Wang, J. (1990). Preliminary analysis of the maturity structure of soybean varieties in Northeast China. *Soybean Sci*. **9**, 198-205.

Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *biorxiv*. https://doi.org/10.1101/005165

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*. **2**, 761-766.

Wang, L., Sun, S., Wu, T., Liu, L., Sun, X., Cai, Y., Li, J. *et al*. (2020a). Natural variation and CRISPR/Cas9-mediated mutation in *GmPRR37* affect photoperiodic flowering and contribute to regional adaptation of soybean. *Plant Biotechnol*. *J*. **18**, 1869-1881.

Wang, Y., Yuan, L., Su, T., Wang, Q., Gao, Y., Zhang, S., Jia, Q. *et al*. (2020b). Light-and temperature-entrainable circadian clock in soybean development. *Plant Cell Environ*. **43**, 637-648.

Wang, L., Li, H., He, M., Dong, L., Huang, Z., Chen, L., Nan, H., *et al*. (2022) *GIGANTEA* orthologs, *E2* members, redundantly determine photoperiodic flowering and yield in soybean. *J*. *Integr*. *Plant Biol*. **65**, 188-202.

Wu, T., Lu, S., Cai, Y., Xu, X., Zhang, L., Chen, F., Jiang, B. *et al*. (2023). Molecular breeding for improvement of photothermal adaptability in soybean. *Mol*. *Breed*. **43**, 60.

Xu, X., Yuan, L., Yang, X., Zhang, X., Wang, L. and Xie, Q. (2022). Circadian clock in plants: Linking timing to fitness. *J*. *Integr*. *Plant Biol*. **64**, 792-811.

Yang, Y., Li, R. and Qi, M. (2000). *In vivo* analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J*. **22**, 543-551.

Yang, W., Wu, T., Zhang, X., Song, W., Xu, C., Sun, S., Hou, W. *et al*. (2019). Critical photoperiod measurement of soybean varieties in different maturity groups. *Crop Sci.* **59**, 2055-2061.

Yang, M., Chen, S., Huang, Z., Gao, S., Yu, T., Du, T., Zhang, H., Li, X., Liu, C. M., Chen, S., and Li, H. (2023). Deep learning-enabled discovery and characterization of *HKT* genes in *Spartina alterniflora*. *Plant J.* **116**, 69-705.

Yu, G., Wang, L.G., He, Q.Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. **31**, 2382-2383.

Yue, Y., Liu, N., Jiang, B., Li, M., Wang, H., Jiang, Z., Pan, H. *et al*. (2016). A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol*. *Plant* **10**, 656-658.

Zhang, C., Dong, S., Xu, J., He, W. and Yang, T. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. **35**, 1786-1788.

Zhang, L., Liu, W., Tsegaw, M., Xu, X., Qi, Y., Sapey, E., Liu, L. *et al*. (2020). Principles and practices of the photo-thermal adaptability improvement in soybean. *J. Integr Agrc*. **19**, 295-310.

Zhang, Y., Du, H., Zhao, T., Liao, C., Feng, T., Qin, J., Liu, B.. (2023). *GmTOC1b* negatively regulates resistance to *Soybean mosaic virus*. *Crop J*. **11**, 1762-1773.

Zhang, D., Zhou, H., Zhang, Y., Zhao, Y., Zhang, Y., Feng, X., & Lin, H. (2025). Diverse roles of MYB transcription factors in plants. *J. Integr. Plant Biol.* **67**, 539-562.

Zhao, X., Li, H., Wang L., Wang J., Huang Z., Du H., Li Y. *et al*. (2024). A critical suppression feedback loop determines soybean photoperiod sensitivity. *Dev*. *Cell*. **59**, 1-14.

Zhu, B., Sun, S., Han, T. and Wu, C. (2012). Comparison of maturity and its structure traits between wild and cultivated soybeans in China. *Soybean Sci*. **31**, 894-898.

Zhu, Z., Wang, Y., Liu, S., Wang, S., Li, J., Fang, C., Liu, Y. *et al*. (2025). Genomic atlas of 8,105 accessions reveals stepwise domestication, global dissemination, and improvement trajectories in soybean. *Cell*. **188**, 6519-6535.

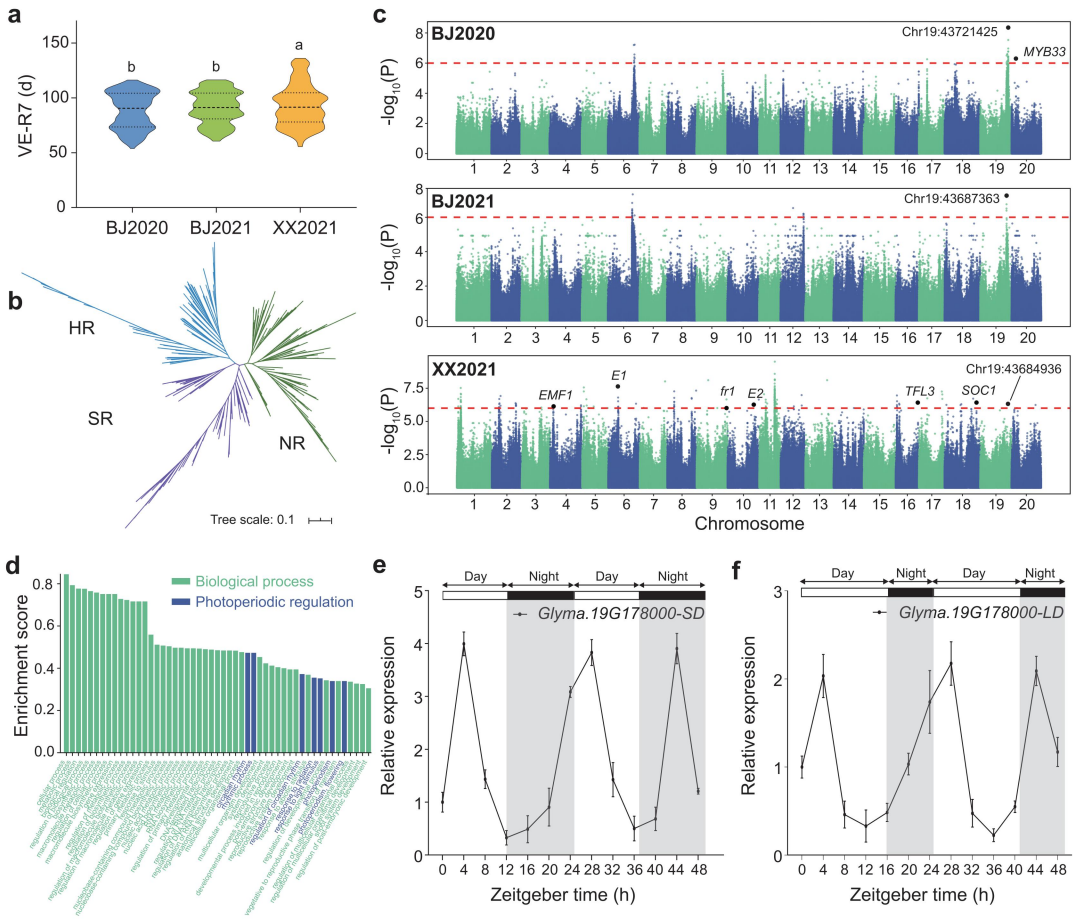
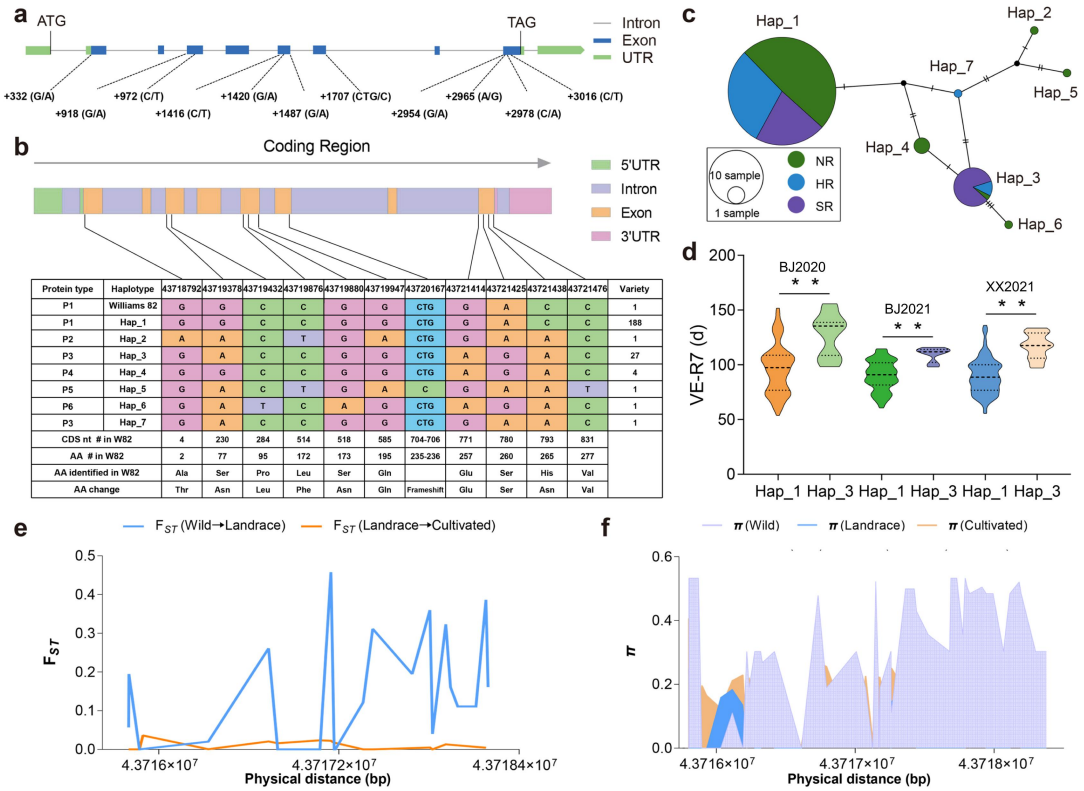


Fig. 1: Identification of *GmRVE4d* locus in soybean. a Effect of different environments on soybean maturity (VE-R7). **b** Population phylogenetic tree based on genome-wide SNP analysis. NR represents soybeans from the Northern Spring Planting Region. HR represents soybeans from the Huang-Huai-Hai Summer Planting Region. SR represents soybeans from the Southern Multiple Cropping Region. **c** GWAS results for soybean maturity across different environments. BJ2020 refers to Beijing (40°23' N, 116°57' E) in 2020. BJ2021 refers to Beijing in 2021. XX2021 refers to Xinxiang (35°10' N, 113°48' E) in 2021. The dashed horizontal lines represent the significance threshold determined by the Bonferroni correction. **d** Gene ontology (GO) enrichment scores for biological process (BP) in the candidate genomic region linked to soybean maturity, calculated using the deep learning-based DeepGOPlus model. *GmRVE4d* transcription levels under short-day (SD, 12 h: 12 h, light: dark) **(e)** and long-day conditions (LD, 16 h: 8 h, light: dark) **(f)**. Data are presented as mean ± SD of three independent biological replicates.



Protein type	Haplotype	43718792	43719378	43719432	43719876	43719880	43719947	43720167	43721414	43721425	43721438	43721476	Variety
P1	Williams 82	G	G	C	C	G	G	CTG	G	A	C	C	1
P1	Hap_1	G	G	C	C	G	G	CTG	G	A	C	C	188
P2	Hap_2	A	A	C	T	G	A	CTG	G	A	A	C	1
P3	Hap_3	G	A	C	C	G	G	CTG	A	G	A	C	27
P4	Hap_4	G	G	C	C	G	G	CTG	A	G	A	C	4
P5	Hap_5	G	A	C	T	G	A	C	G	A	A	T	1
P6	Hap_6	G	A	T	C	A	G	CTG	A	G	A	C	1
P3	Hap_7	G	A	C	C	G	G	CTG	G	A	A	C	1
CDS nt # in W82		4	230	284	514	518	585	704-706	771	780	793	831	
AA # in W82		2	77	95	172	173	195	235-236	257	260	265	277	
AA identified in W82		Ala	Ser	Pro	Leu	Ser	Gln		Glu	Ser	His	Val	
AA change		Thr	Asn	Leu	Phe	Asn	Gln	Frameshift	Glu	Ser	Asn	Val	

Fig. 2: Cloning of *GmRVE4d* locus in soybean. a Gene structure and coding region variations of *GmRVE4d*. **b** Haplotypes detected in the coding region of *GmRVE4d*. **c** Geographical distribution of different *GmRVE4d* haplotypes. NR represents soybeans from the Northern Spring Planting Region. HR represents soybeans from the Huang-Huai-Hai Summer Planting Region. SR represents soybeans from the Southern Multiple Cropping Region. **d** Comparison of maturity between different haplotypes of *GmRVE4d* across multiple environments. VE, vegetative stage of emergence. R7, physiological maturity. d, day. BJ2020 refers to Beijing (40°23' N, 116°57' E) in 2020. BJ2021 refers to Beijing in 2021. XX2021 refers to Xinxiang (35°10' N, 113°48' E) in 2021. ** indicates significant differences at $P < 0.01$. *F*ST **(e)** and π **(f)** values in wild soybeans (*G. soja*), landraces and cultivated soybeans across the 2.7-kb promoter region.

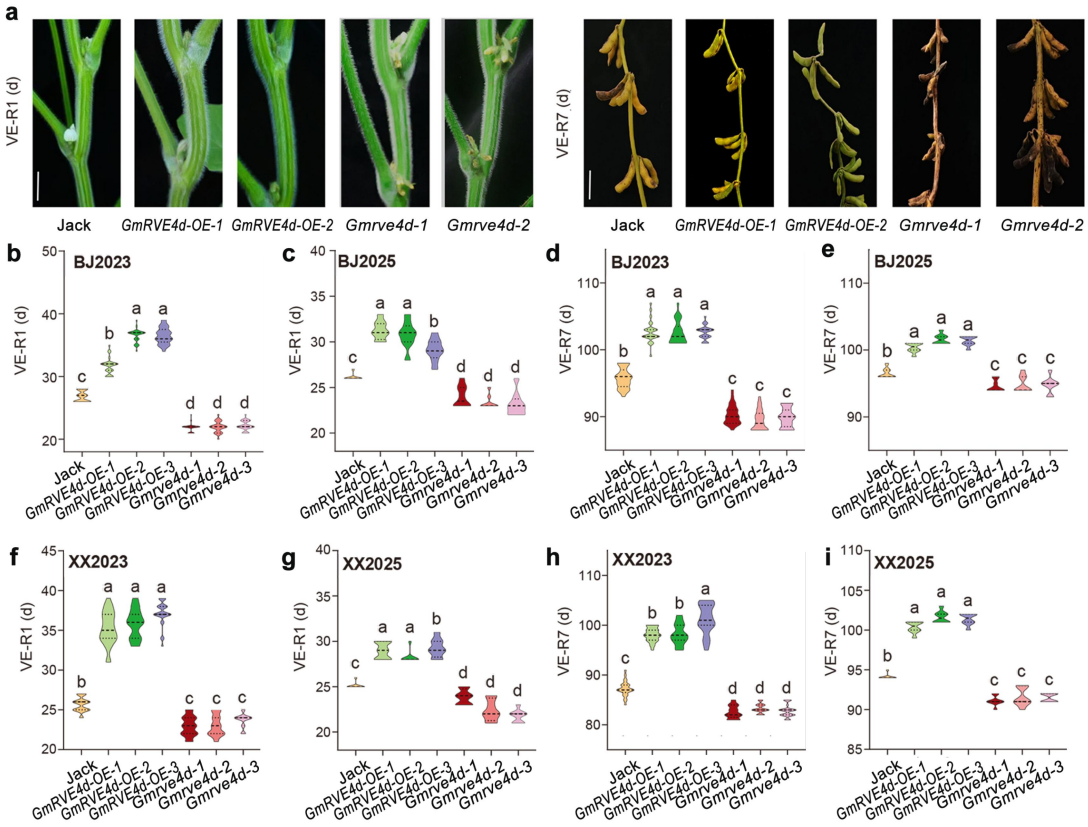


Fig. 3: Functional characterization of *GmRVE4d* in controlling flowering time and maturity. **a** Plant phenotypes related to flowering time and maturity under natural long-day (NLD) conditions. Flowering time (days from VE to R1) across Beijing (40°23' N, 116°57' E) in 2023 **(b)** and 2025 **(c)**, Xinxiang (35°10' N, 113°48' E) in 2023 **(f)** and 2025 **(g)**. VE, vegetative stage of emergence. R1, beginning bloom. Maturity (days from VE to R7) across Beijing in 2023 **(d)** and 2025 **(e)**, Xinxiang in 2023 **(h)** and 2025 **(i)**. R7, physiological maturity. n = 12 plants. d, day. Flowering stage scale bars: 3 cm. Maturity stage scale bars: 5 cm. Different lowercase letters indicate significant differences at $P < 0.05$.

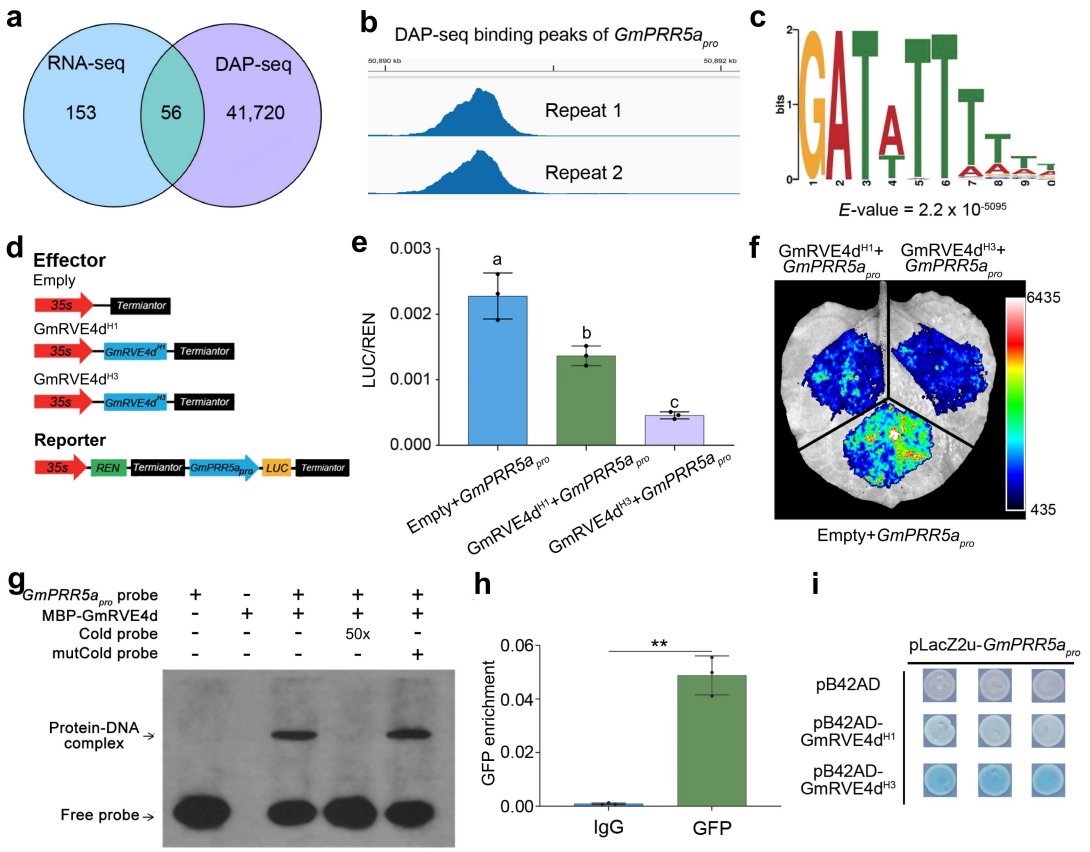


Fig. 4: Identification of downstream target genes of GmRVE4d. **a** Venn diagram of overlapping genes in two datasets: RNA-seq and DAP-seq. **b** Localization of GmRVE4d-enriched binding peaks in the *GmPRR5a* promoter region. **c** Putative GmRVE4d-binding motif identified in promoter regions. *P*-value was calculated using MEME. **d** Schematic of constructs used in the Dual-Luciferase reporter assay. **e, f** Relative reporter activity (LUC/REN) in tobacco leaves. Different lowercase letters indicate significant differences at $P < 0.05$. **g** Yeast one-hybrid assay showing direct binding of GmRVE4d to the *GmPRR5a* promoter. Control group: pB42AD empty vector. **h** Electrophoretic mobility shift assay (EMSA) of MBP-GmRVE4d binding to the Gm*PRR5a* promoter. MBP-GmRVE4d fusion proteins were purified from *Escherichia coli*. Biotin-labeled oligonucleotides (GATATTTTTT) were used as probes. The probe sequence of Gm*PRR5a* was determined based on the location of the binding peaks in DAP-seq results. An excess of non-labeled probe or mutated probe was used as a control. **i** ChIP-qPCR assays validating GmRVE4d binding to *GmPRR5a*. ** indicates significance at $P < 0.01$. Data are means ± SD of three independent biological repeats.

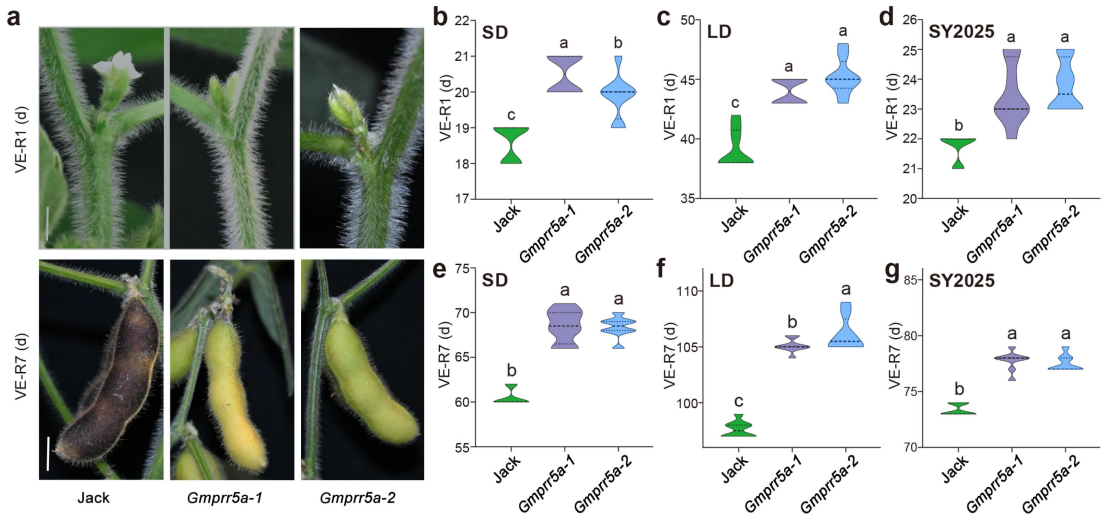


Fig. 5: Functional characterization of *GmPRR5a* in controlling flowering time and maturity. **a** Plant phenotypes related to flowering time and maturity under short-day (SD, 12 h: 12 h, light: dark) conditions. Flowering stage scale bars: 0.5 cm. Maturity stage scale bars: 1 cm. Flowering time (days from VE to R1) across SD **(b)**, LD (16 h: 8 h, light: dark) **(c)**, and Sanya (18°14' N, 109°13' E) in 2025 **(d)**. VE, vegetative stage of emergence. R1, beginning bloom. Maturity (days from VE to R7) across SD **(e)**, LD **(f)**, and Sanya in 2025 **(g)**. R7, physiological maturity. $n = 12$ plants. d, day. Different lowercase letters indicate significant differences at $P < 0.05$.

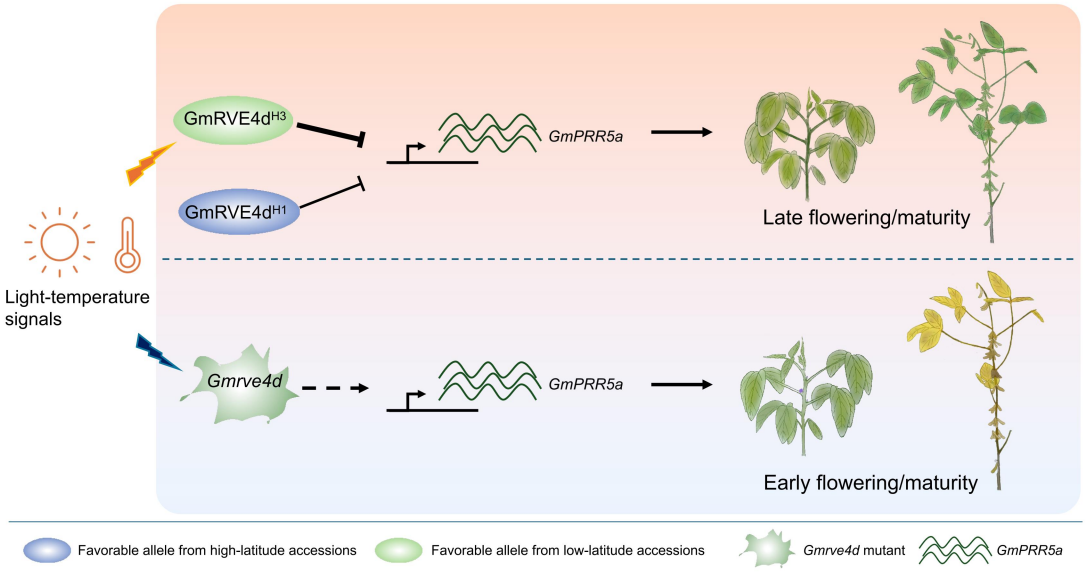
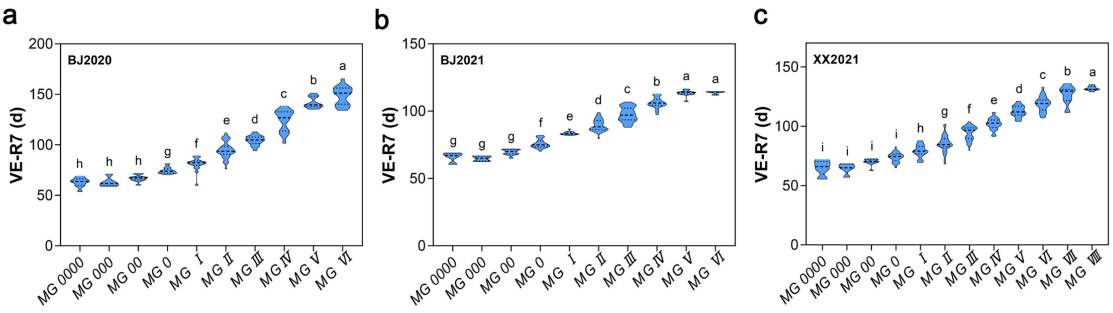
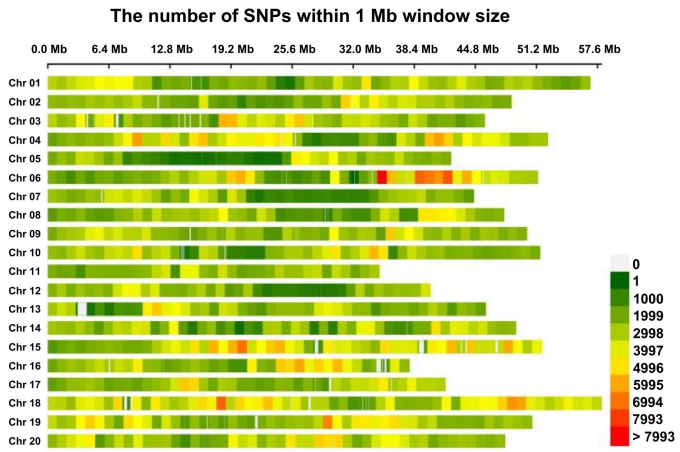


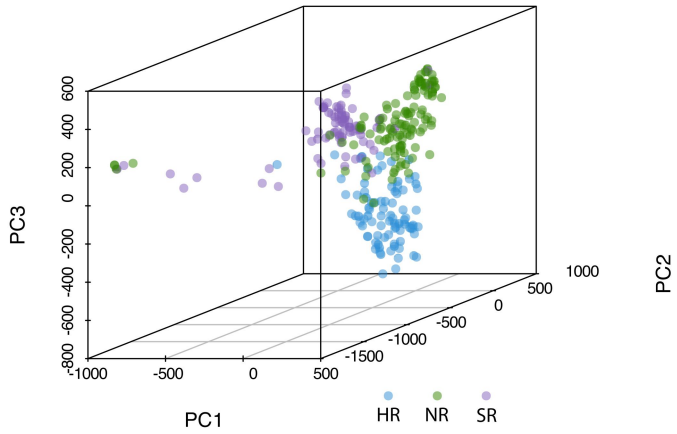
Fig. 6: Model summarizing the regulatory mechanism of *GmRVE4d*. Photothermal signals trigger the expression of *GmRVE4d* alleles, including *GmRVE4d*^{H3} (the favorable variant from low-latitude accessions) and *GmRVE4d*^{H1} (the favorable variant from high-latitude accessions). Both alleles physically bind to the promoter of *GmPRR5a* to repress its transcription, with GmRVE4d^{H3} exerting a stronger inhibitory effect. Reduced *GmPRR5a* expression consequently downregulates the transcription of *GmFT2a/GmFT5a* (core flowering-promoting genes), ultimately driving late flowering and maturity. In contrast, loss-of-function mutation of GmRVE4d relieves the transcriptional repression of *GmPRR5a*, thereby activating the expression of *GmFT2a/GmFT5a* and accelerating flowering and maturity in soybean.



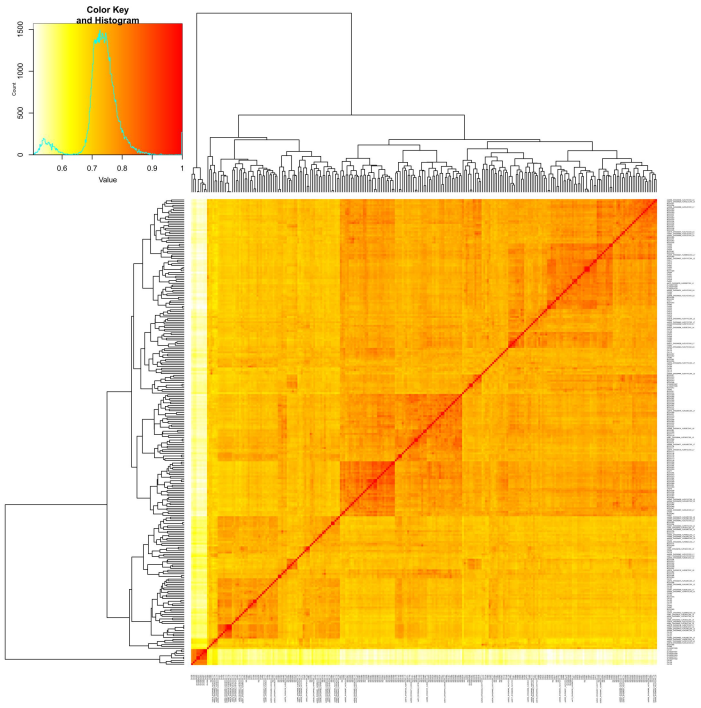
Supplementary Fig. 1 Distribution of maturity for the 257-accession panel over three environments. BJ2020 refers to Beijing in 2020. BJ2021 refers to Beijing in 2021. XX2021 refers to Xinxiang in 2021. MG, maturity group. R7, physiological maturity. d, day. Different lowercase letters indicate significant differences at $P < 0.05$.



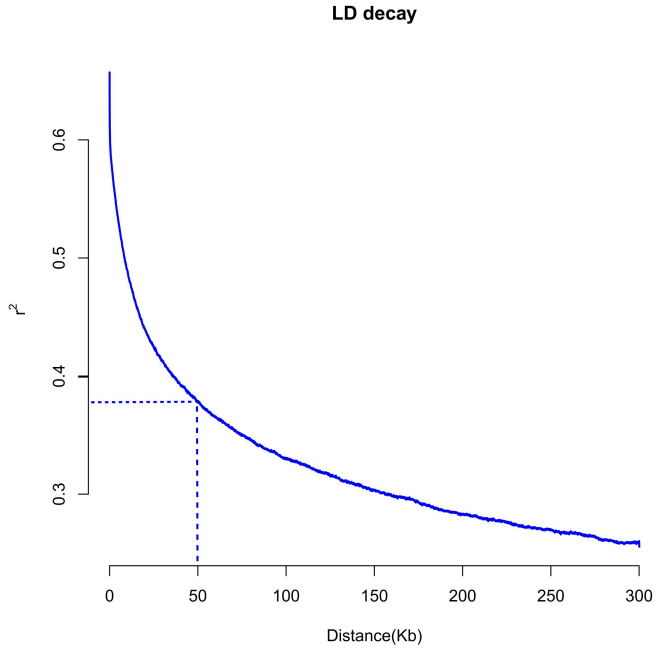
Supplementary Fig. 2 SNP distribution density across the soybean genome.



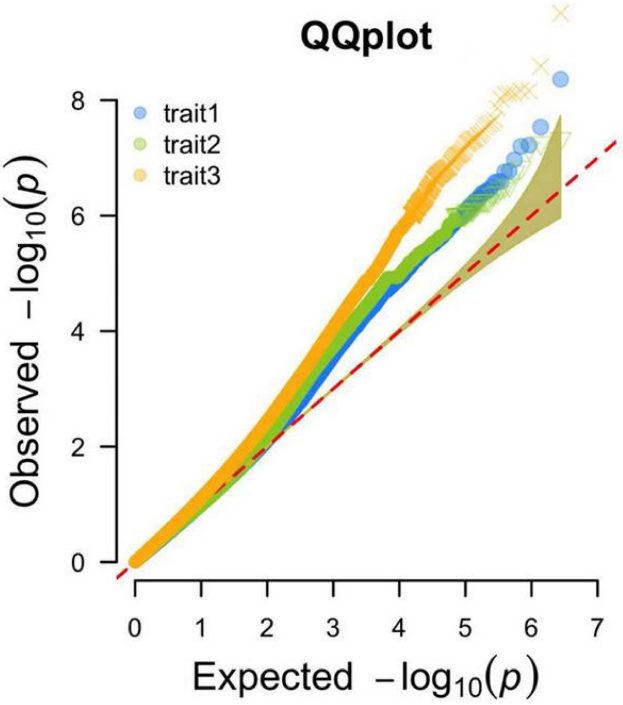
Supplementary Fig. 3 3D Plot of principal component analysis (PCA) of 257 accessions. NR represents soybeans from the Northern Spring Planting Region. HR represents soybeans from the Huang-Huai-Hai Summer Planting Region. SR represents soybeans from the Southern Multiple Cropping Region.



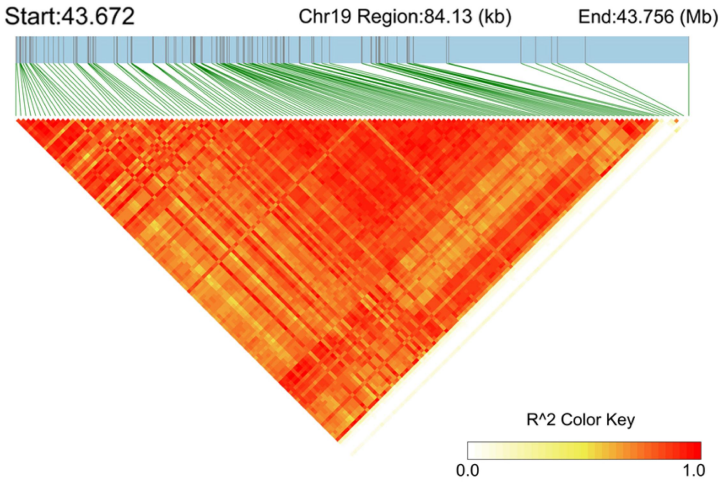
Supplementary Fig. 4 3D Plot of principal component analysis (PCA) of 257 accessions.



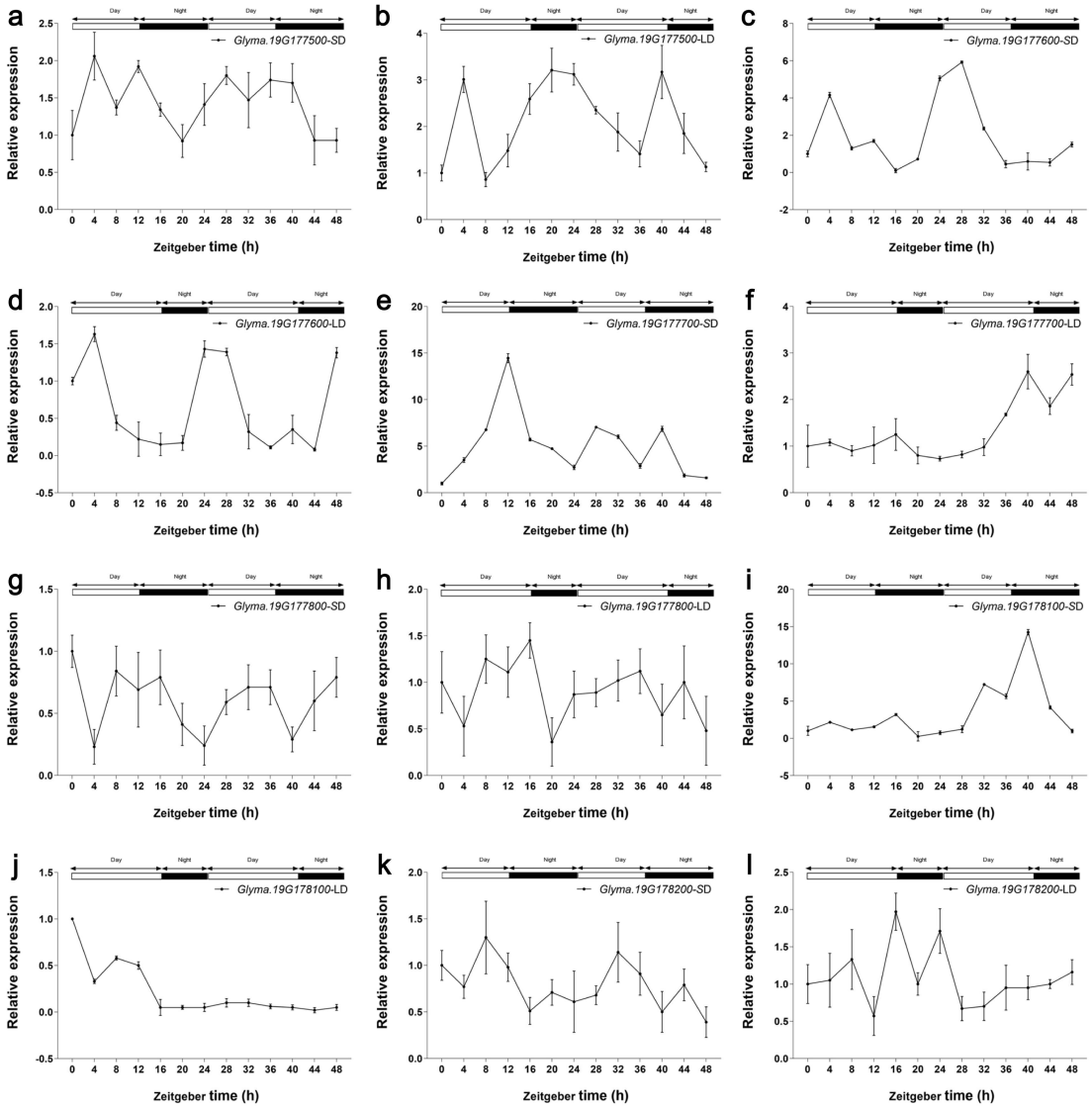
Supplementary Fig. 5 Genome-wide linkage disequilibrium (LD) decay plot.



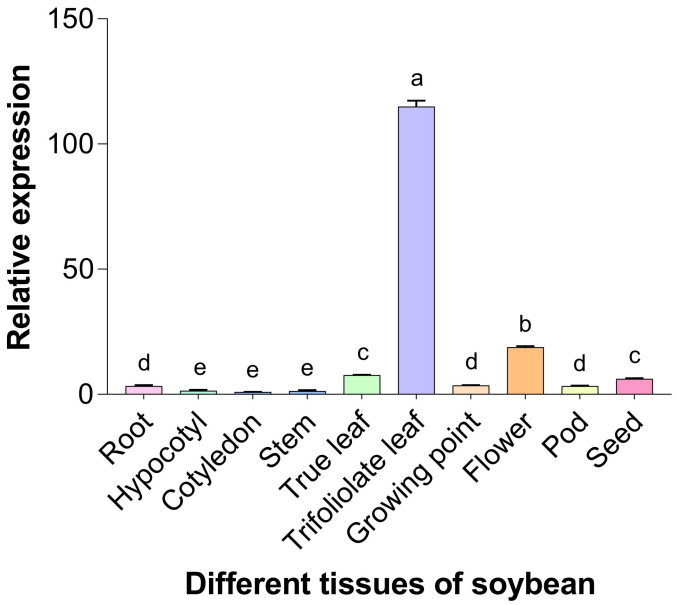
Supplementary Fig. 6 Quantile-quantile (QQ) plot for soybean maturity across different environments. Trait 1 refers to Beijing (40°23' N, 116°57' E) in 2020. Trait 2 refers to Beijing in 2021. Trait 3 refers to Xinxiang (35°10' N, 113°48' E) in 2021.



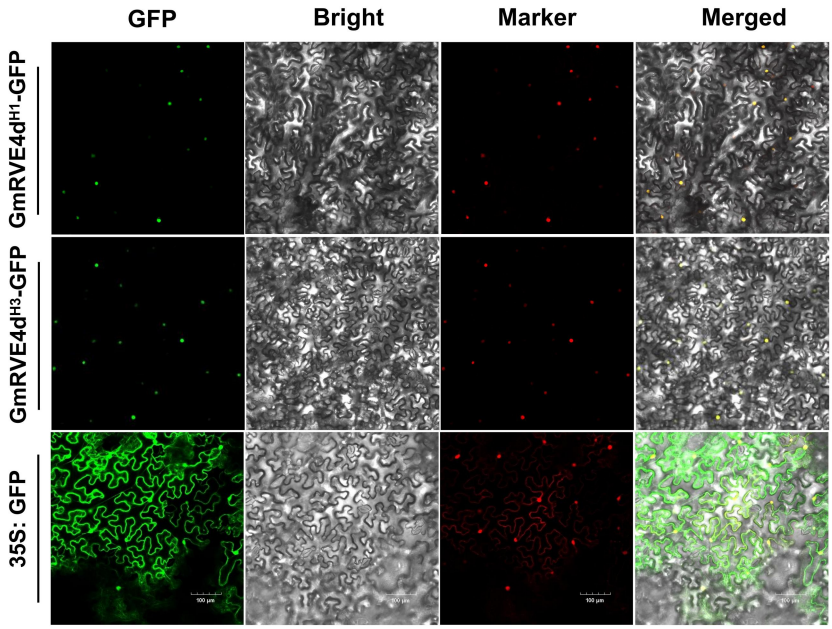
Supplementary Fig. 7 Linkage disequilibrium (LD) heatmap of the *GmRVE4d* locus.



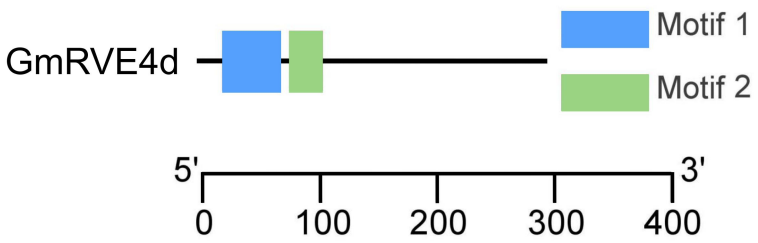
Supplementary Fig. 8 Diurnal transcriptional levels of candidate genes under short-day and long-day conditions. Data are presented as mean ± SD of three independent biological repeats.



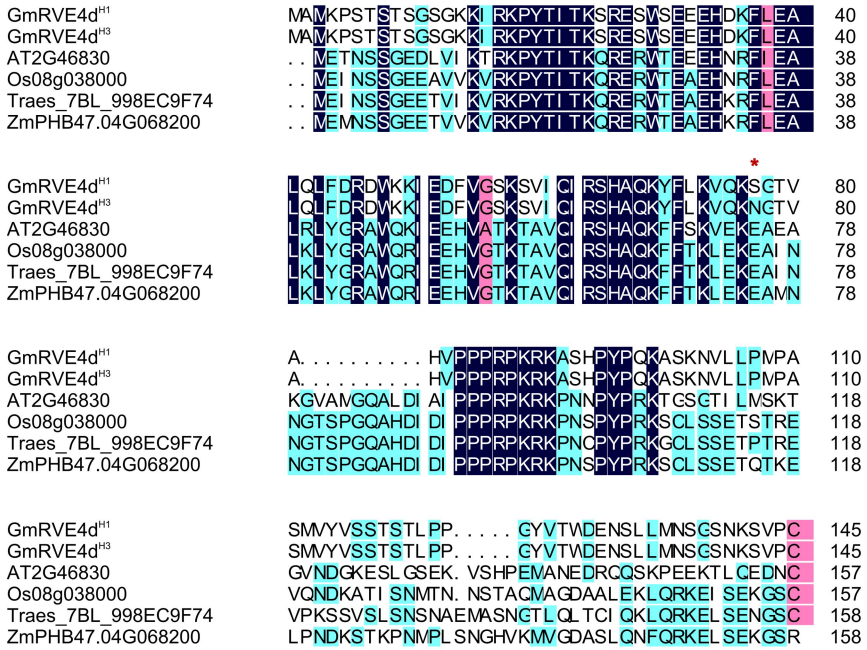
Supplementary Fig. 9 Transcriptional levels of *GmRVE4d* in different organs. Different lowercase letters indicate significant differences at $P < 0.05$.



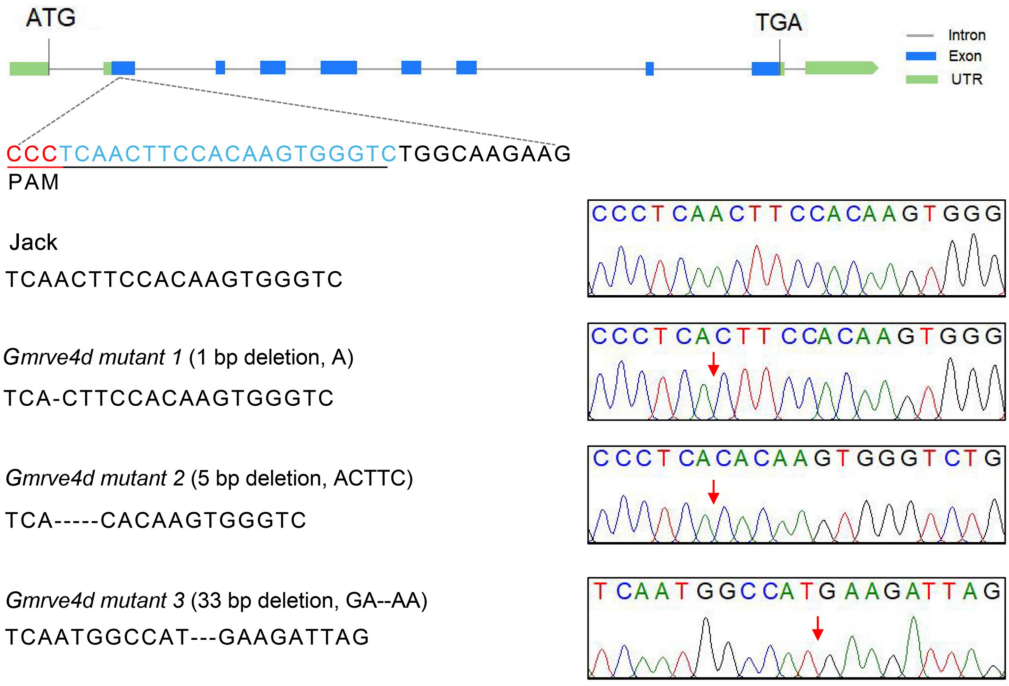
Supplementary Fig. 10 Subcellular localization of GmRVE4d-GFP. Motif 1 corresponds to the MYB repeat (R). Each recombinant expression construct of the indicated GmRVE4d haplotype, fused with GFP, was ectopically expressed in tobacco. Marker, mCherry fused with a nuclear localization signal (NLS) as the nucleus marker; Merged, merged signals of GFP and mCherry. GFP alone was used as a negative control. Scale bar: 100 μm.



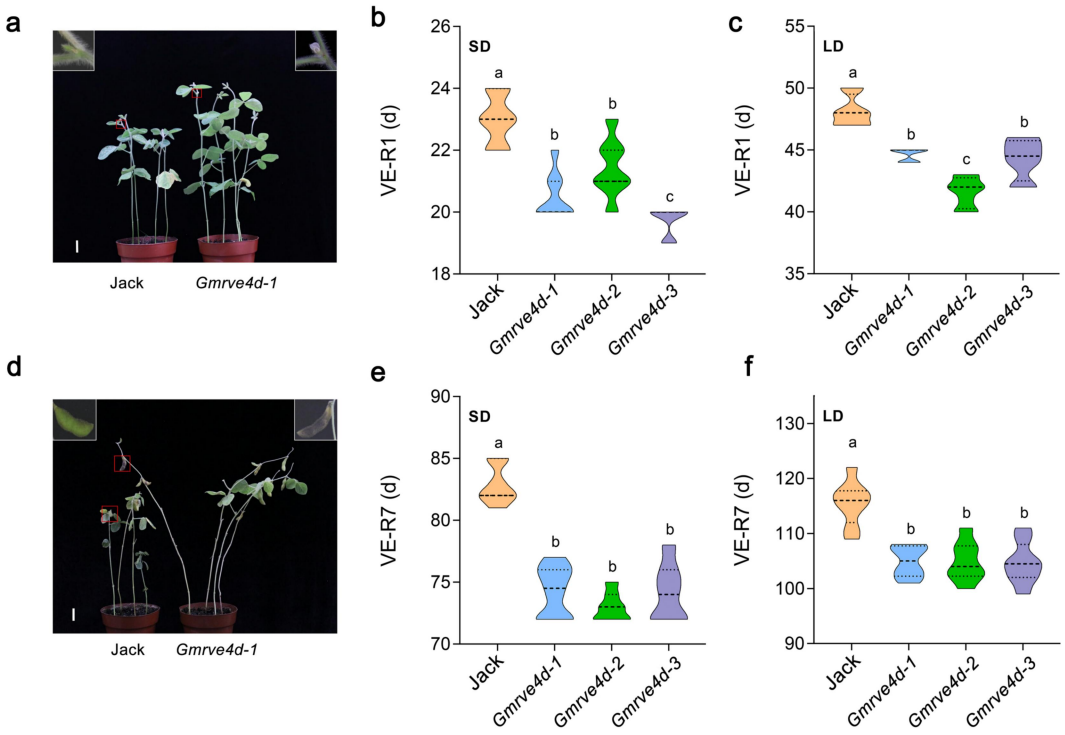
Supplementary Fig. 11 Motif analysis of GmRVE4d. Motif 1 corresponds to the MYB repeat (R).



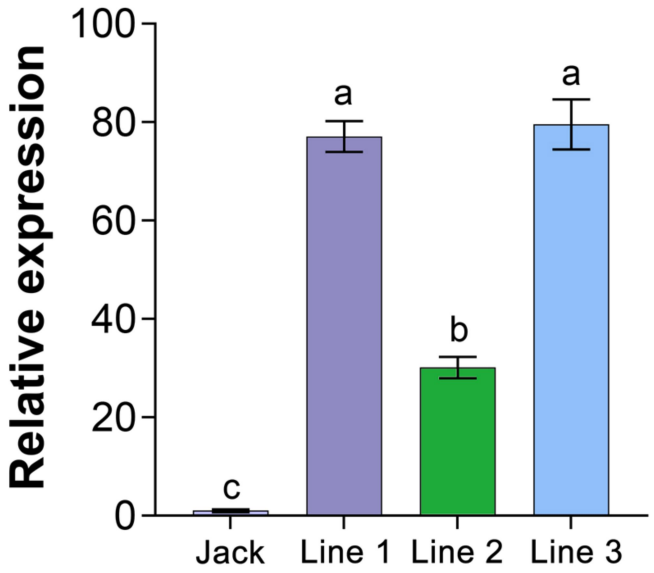
Supplementary Fig. 12 Alignment of amino acid sequences of GmRVE4d haplotypes and its homologs. The amino acid sequences were aligned by DNAMAN with manually adjustments. The red asterisks indicate the variations of the amino acid of GmRVE4d haplotypes.



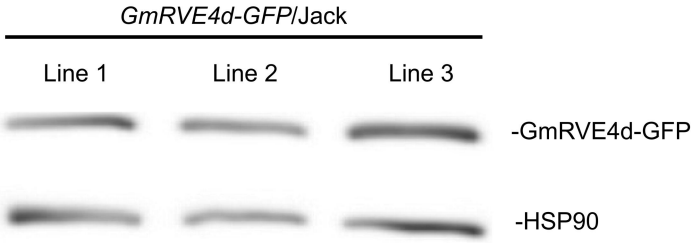
Supplementary Fig. 13 Homozygous targeted mutagenesis of *GmRVE4d* mediated by CRISPR/Cas9 in *Gmrve4d* mutant plants. Red arrows indicate the mutation sites.



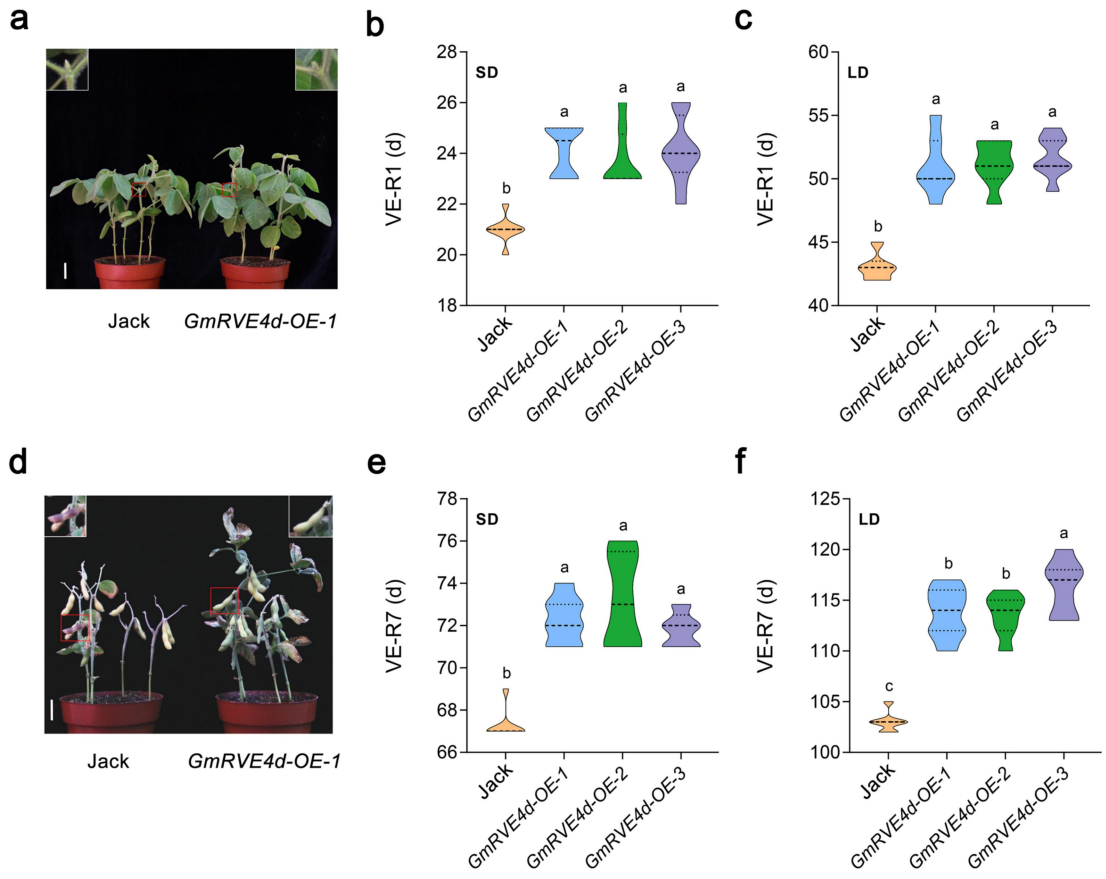
Supplementary Fig. 14 Phenotypic identification of flowering time and maturity in *Gmrve4d mutant* under short-day and long-day conditions. a Plant phenotypes related to flowering time under short-day (SD, 12 h: 12 h, light: dark) conditions. Flowering time (days from VE to R1) across SD **(b)** and LD (16 h: 8 h, light: dark) conditions **(c)**. VE, vegetative stage of emergence. R1, beginning bloom. **d** Plant phenotypes related to maturity under SD conditions. Maturity (days from VE to R7) across SD **(e)** and LD **(f)**. R7, physiological maturity. n = 12 plants. d, day. Flowering stage scale bars: 3 cm. Maturity stage scale bars: 5 cm. Different lowercase letters indicate significant differences at $P < 0.05$.



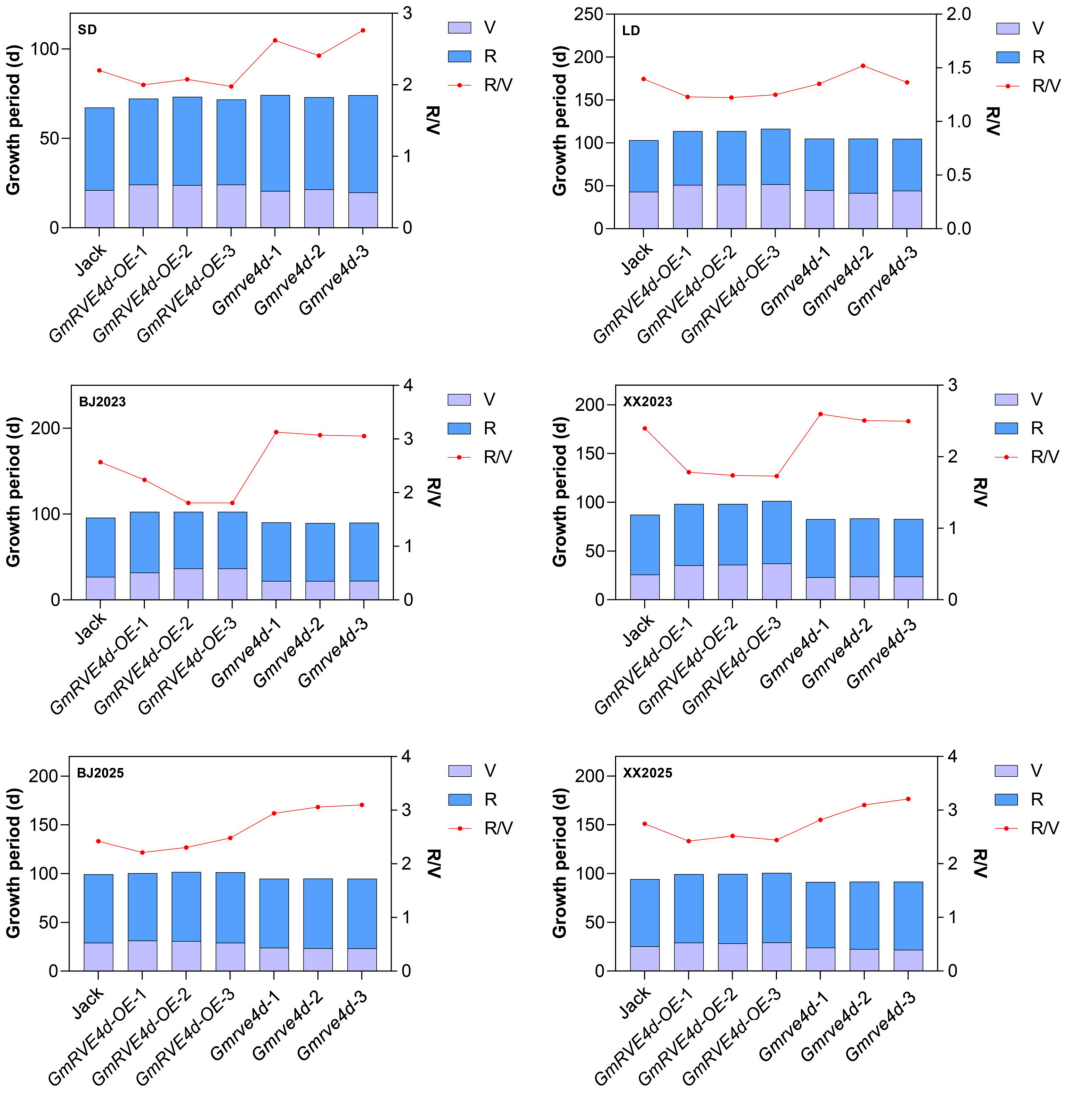
Supplementary Fig. 15 qRT-PCR validation of *GmRVE4d* expression in *GmRVE4d-OE* lines. Data are shown as mean ± SD of three independent biological replicates. Different lowercase letters indicate significant differences at $P < 0.05$.



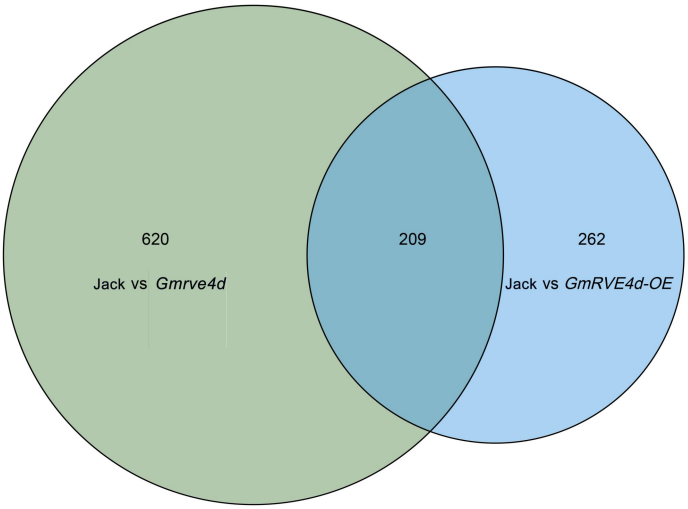
Supplementary Fig. 16 Western blot analysis of GmRVE4d protein levels. Membranes were re-probed with HSP90 antibody as a loading control.



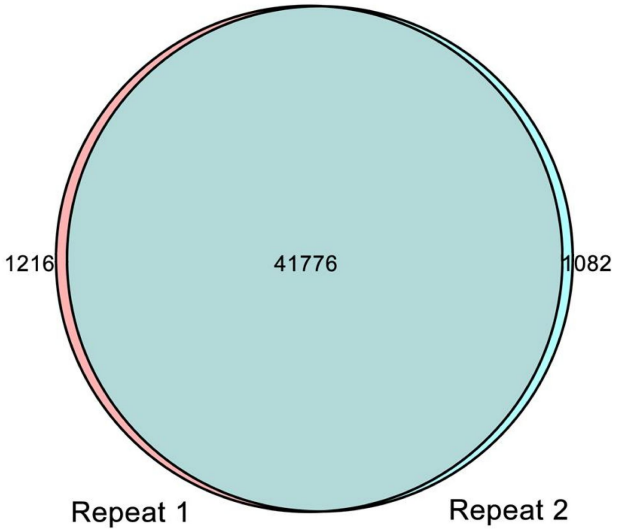
Supplementary Fig. 17 Phenotypic identification of flowering time and maturity in *GmRVE4d-OE* lines under short-day and long-day conditions. a Plant phenotypes related to flowering time under short-day (SD, 12 h: 12 h, light: dark) conditions. Flowering time (days from VE to R1) across SD **(b)** and LD (16 h: 8 h, light: dark) conditions **(c)**. VE, vegetative stage of emergence. R1, beginning bloom. **d** Plant phenotypes related to maturity under SD conditions. Maturity (days from VE to R7) across SD **(e)** and LD **(f)**. R7, physiological maturity. $n = 12$ plants. d, day. Flowering stage scale bars: 3 cm. Maturity stage scale bars: 5 cm. Different lowercase letters indicate significant differences at $P < 0.05$.



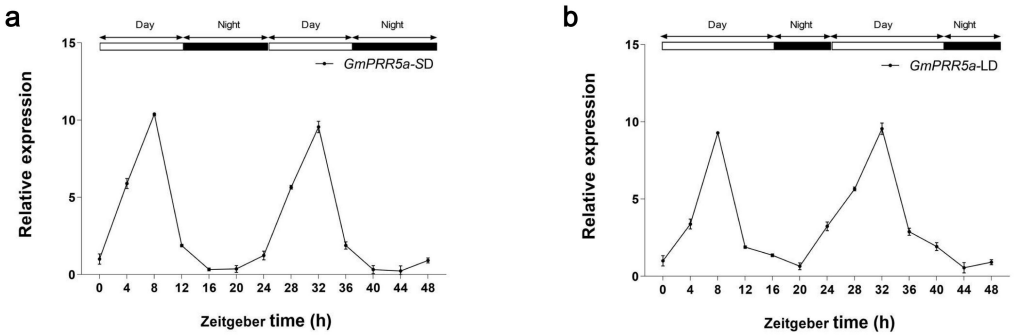
Supplementary Fig. 18 Regulatory effect of *GmRVE4d* on soybean growth period structure under diverse environmental conditions. V indicates the vegetative growth period. R indicates the reproductive growth period. R/V indicates the ratio of the growth period structure. SD (short-day, 12 h: 12 h, light: dark). LD (long-day, 16 h: 8 h, light: dark). BJ2023 and BJ2025 refer to Beijing (40°23' N, 116°57' E) in 2023 and 2025, respectively. XX2023 and XX2025 refer to Xinxiang (35°10' N, 113°48' E) in 2023 and 2025, respectively.



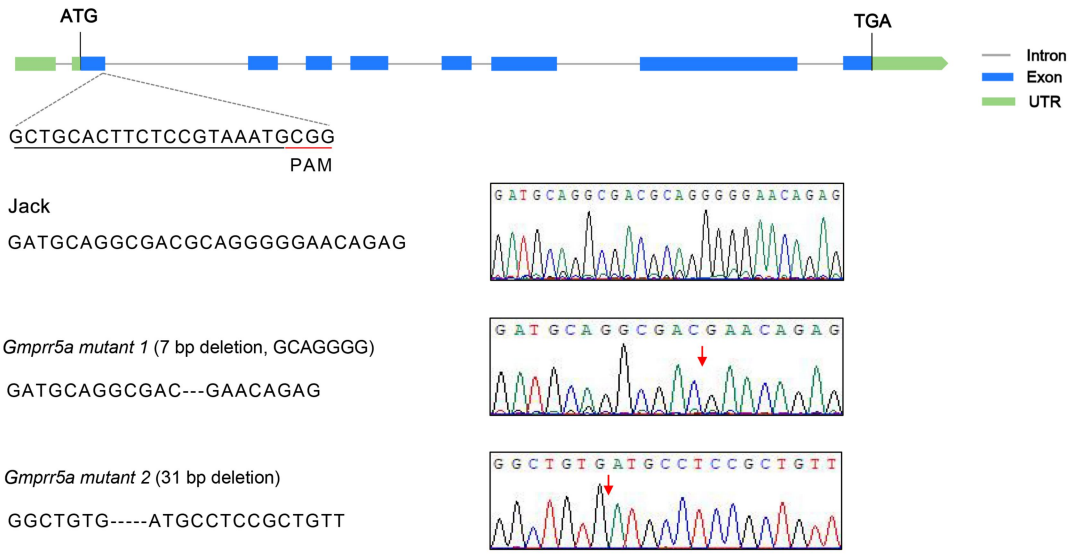
Supplementary Fig. 19 Venn diagram of differentially expressed genes (DEGs) identified in comparisons between *GmRVE4d-OE* and wild-type, and between *Gmrve4d* mutant and wild-type.



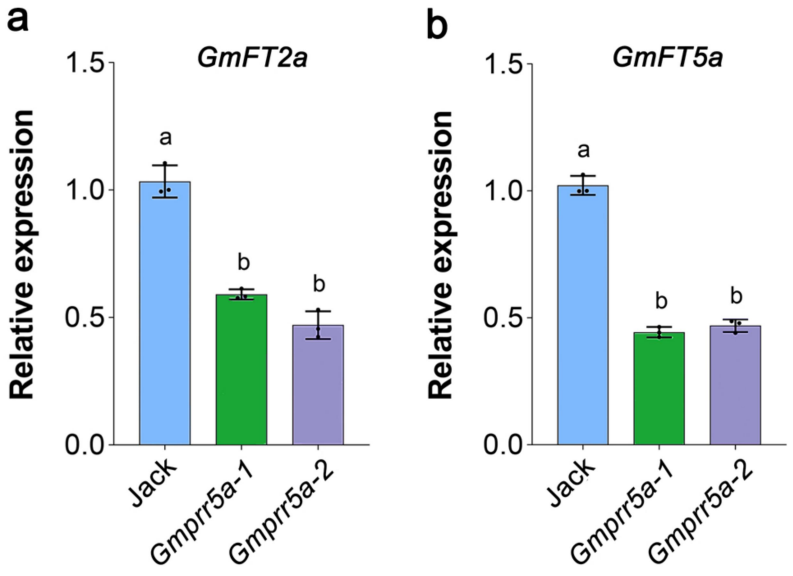
Supplementary Fig. 20 Genome-wide distribution of GmRVE4d binding sites identified by DAP-seq. Repeat 1 and Repeat 2 indicate two biological replicates.



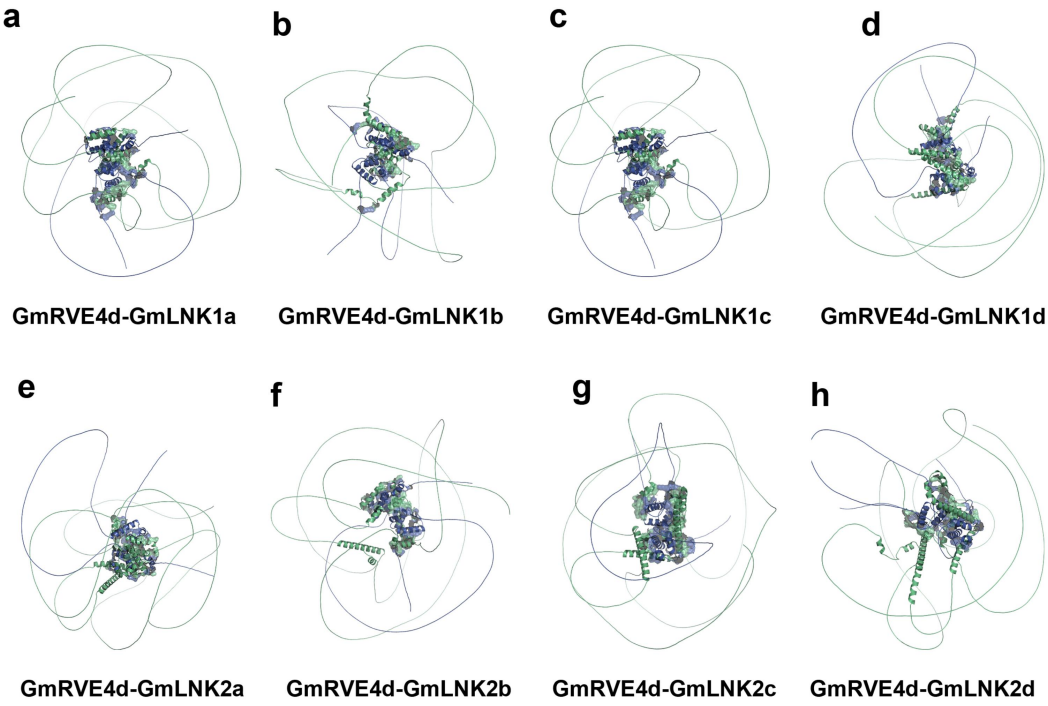
Supplementary Fig. 21 Relative expression levels of *GmPRR5a* under short-day and long-day conditions. a SD, 12 h: 12 h, light: dark. **b** LD, 16 h: 8 h, light: dark. Data are shown as mean ± SD of three independent biological repeats.



Supplementary Fig. 22 Homozygous targeted mutagenesis of *GmPRR5a* mediated by CRISPR/Cas9 in *Gmrve4d* mutant plants. Red arrows indicate the mutation sites.



Supplementary Fig. 23 Relative expression patterns of *GmFT2a/5a*, key regulatory genes controlling soybean flowering. Data are shown as mean ± SD of three independent biological repeats. Different lowercase letters indicate significant differences at $P < 0.05$.



Supplementary Fig. 24 AlphaFold3-predicted 3D structural models of interaction complexes between GmRVE4d and GmLNK1/2 family proteins. Blue/green tones distinguish proteins, secondary elements: α-helices, β-sheets, dashed regions indicate flexible protein segments.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData7.xlsx
- SupplementaryData3.docx
- SupplementaryData5.xlsx
- SupplementaryData4.docx
- SupplementaryData1.docx
- SupplementaryData6.docx
- SupplementaryData2.docx