nature communications

Article | Article in Press

https://doi.org/10.1038/s41467-026-77316-3

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

Received: 2 March 2026

Accepted: 21 August 2026

Published online: 08 September 2026

Check for updates

Baiquan Sun[1,2,4], Tingting Wu[1,4], Bingjun Jiang[1,4], Xuegang Sun[1], Hongchang Jia[3], Mahmoud Naser[1], Maoxiang Yang[1], Xiangyu Yao[1], Yanhui Sun[1], Qimeng Li[1], Peiguo Wang[1,2], Liwei Jiang[1,2], Shibo Sun[1,2], Chao Qin[1], Shan Yuan[1], Kunhui He[1,2], Wensheng Hou[1], Cunxiang Wu[1], Huihui Li[1,2] ✉, Shi Sun[1] ✉ & Tianfu Han[1,2] ✉

Timely flowering and maturity are crucial for plant reproduction and environmental adaptation. Light–dark cycle-associated regulatory networks integrate photoperiodic cues with intrinsic developmental programs and play pivotal roles in flowering, maturity and environmental adaptation. Soybean is a short-day crop with strong photoperiod responsiveness; however, the molecular mechanisms by which these networks mediate soybean adaptation to diverse environments remain largely elusive. Here, we combine genome-wide association analysis with deep learning-based assessment of soybean maturity variation and identify *GmRVE4d*, a homolog of the *Arabidopsis* clock-associated *REVEILLE*, as a maturity-associated locus across both field environments. Genetic analysis reveals evidence of domestication-related artificial selection at *GmRVE4d* and identifies the late-maturing haplotype *GmRVE4d*^{H3}. Functional analyses further demonstrate that the GmRVE4d protein directly binds to the promoter of *GmPRR5a*, a soybean homolog of the *Arabidopsis* clock-associated component *PSEUDO-RESPONSE REGULATOR 5*, and represses its transcription, thereby delaying soybean flowering and maturity. Taken together, our findings suggest that GmRVE4d is a negative regulator of soybean flowering and maturity that functions by directly repressing *GmPRR5a* expression and represents a promising molecular target for expanding cultivation latitudes through molecular breeding.

Flowering time and maturity are the key traits for ecological adaptation in plants, jointly regulated by endogenous genetic networks and environmental cues[1]. As core components of molecular pathways that perceive and integrate environmental signals, circadian clock-associated genes coordinate external cues with intrinsic developmental programs and play pivotal roles in shaping adaptive genetic diversity across plant species[2,3].

Soybean (*Glycine max* (L.) Merr.) was originally domesticated in the temperate regions of China and has since spread globally; it is now cultivated from 53°N to 35°S[4–6]. Soybean is a typical short-day crop that

A full list of affiliations appears at the end of the paper. ✉e-mail: lihuihui@caas.cn; sunshi@caas.cn; hantianfu@caas.cn

is highly sensitive to photoperiod and temperature. A majority of soybean varieties switch from vegetative to reproductive growth after the photoperiod is shortened below a certain threshold[7,8], indicating that photoperiod is a primary environmental signal governing key developmental transitions such as floral induction. By contrast, temperature largely modulates developmental rate, affecting how rapidly plants progress through vegetative and reproductive phases under a given photoperiod[9]. Therefore, identifying circadian clock-associated regulators of flowering and maturity is essential for understanding the genetic basis of soybean adaptation and for improving maturity-related traits through molecular breeding[10,11].

In soybean, flowering time and maturity are controlled by a conserved but legume-specialized regulatory network centered on the classic maturity loci *E1–E4* and the long-juvenile gene *J*, which determine photoperiod sensitivity and regional adaptation[12,13]. *E1* acts as a soybean-specific flowering repressor by suppressing the florigen genes *GmFT2a* and *GmFT5a*, whereas *E2/GmGIa*, *E3/GmPhyA3*, *E4/GmPhyA2*, and *J/GmELF3* link photoperiod perception and circadian regulation with downstream flowering responses. These pathways ultimately regulate the florigen genes *GmFT2a* and *GmFT5a*, which function as key downstream outputs of the photoperiodic flowering network and promote floral transition under inductive conditions[14,15].

The circadian clock is an endogenous molecular oscillator with an approximately 24-h period that coordinates plant growth and development with daily environmental cycles, thereby conferring adaptive advantages[16]. In *Arabidopsis*, the circadian clock comprises a complex network of interlocking feedback loops involving both transcriptional repressors and activators[17]. Morning-phase genes include *CIRCADIAN CLOCK-ASSOCIATED 1* (*CCA1*), *LATE ELONGATED HYPOCOTYL* (*LHY*), *REVEILLE* genes such as *RVE4, RVE6, RVE8*, *NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED* genes, such as *LNK1* and *LNK2*, and *PSEUDO-RESPONSE REGULATOR* genes including *PRR9* and *PRR7*. Evening-phase genes include *PRR5*, *PRR3*, *PRR1*/*TIMING OF CAB EXPRESSION 1* (*TOC1*), *ZEITLUPE* (*ZTL*), *GIGANTEA* (*GI*), and the evening complex (EC), which is formed by *EARLY FLOWERING 3* (*ELF3*), *ELF4* and *LUX ARRHYTHMO/PHYTOCLOCK 1* (*LUX/PCL1*)[3,18].

REVEILLE genes encode MYB domain-containing transcription factors and are important components of the plant circadian clock system[19]. The MYB protein family comprises functionally diverse transcriptional regulators characterized by highly conserved N-terminal MYB domains containing one to four imperfect tandem repeats, and members of this family have substantial potential for crop improvement through transcriptional regulation[20,21]. In *Arabidopsis*, RVE8 regulates circadian rhythm antagonistically to *LHY* and *CCA1* and promotes the expression of *PRR5*, *TOC1* and other evening-phased genes, including *LUX*, *ELF4* and *GI*, by directly binding to the evening element (EE; AAAATATCT) in their promoters[22]. The *RVE8* gene is also implicated in the control of flowering time[19,23]. RVE4 and RVE8 are closely related members of the *Arabidopsis* RVE family, sharing conserved protein domains and overlapping biological functions. Previous studies have shown that *RVE4*, *RVE6* and *RVE8* function partially redundantly to accelerate the circadian clock pace[23,24]. Intriguingly, flowering time is slightly delayed in *rve4 rve6 rve8* triple mutants, a phenotype opposite to that of the *rve8* single mutant[19,25].

In recent years, multiple soybean genes homologous to known circadian clock components or associated with clock-regulated pathways have been identified. Several genes have been implicated in circadian or clock-associated rhythmic regulation, including *GmLHY1a*/*Tof16*, *GmLHY1b*/*E11*, *GmLHY1c*, *GmLHY1d*[26], *LUX1*, *LUX2*[2,27], *J*[12,28], *GmTOC1b*[29] and *E2*[30]. These genes are homologous to conserved clock or EC components and have been linked to circadian regulation and photoperiodic flowering in soybean. By contrast, other loci, such as *Gp1*/*qFT12-1*/*Tof12*/*GmPRR37*/*GmPRR3b*[31–34], *Tof11*/*Gp11*/*GmPRR3a*[33], *GmEID1*[35] and *Dgt1*/*GmRVE1*[36], are homologous to, or associated with, clock-related genes but have so far mainly been characterized as regulators of flowering time, maturity, plant architecture or photoperiodic adaptation. Recent studies have shown that, under long-day conditions, E2 interacts with the blue-light receptor Flavin-binding, Kelch repeat, F-box 1 (FKF1), thereby inducing the degradation of J. The EC also suppresses *E2* expression by binding to the *E2* promoter, eventually forming a soybean photoperiod-dependent regulatory loop[37]. In addition, Tof11 and Tof12 directly suppress *GmRVE1* expression, and *GmRVE1* represses flowering in an *E1*-dependent manner[36]. However, the soybean circadian clock appears to have a complex regulatory architecture, and its underlying molecular network remains incompletely understood. In particular, the functions of soybean *GmRVE4* genes and their downstream regulatory mechanisms remain largely unknown.

In this study, we identified *GmRVE4d*, a soybean homolog of the *Arabidopsis REVEILLE* gene, as the candidate gene for a major maturity quantitative trait locus by combining genome-wide association study (GWAS) with a deep learning (DL)-based approach, DeepGOPlus. *GmRVE4d* encodes a protein homologous to *Arabidopsis* RVE4 and RVE8. Through analysis of CRISPR-Cas9 knockout mutants, phenotyping of transgenic materials and investigation of its regulatory mechanism, we found that *Gmrve4d* mutants flowered and matured earlier under the tested conditions, whereas enhanced *GmRVE4d* activity delayed flowering and maturity. Mechanistically, we showed that *GmRVE4d* regulates flowering time and maturity by inhibiting the transcriptional expression of *GmPRR5a*. Moreover, the *GmRVE4d*^{H3} haplotype exerted a stronger regulatory effect than *GmRVE4d*^{H1}. We further demonstrated that the *GmRVE4d*-containing haplotype block underwent strong artificial selection during soybean domestication. Collectively, these results suggest that *GmRVE4d* is a negative regulator of soybean flowering and maturity and may have contributed to the domestication process from wild soybean (*G. soja*) to cultivated soybean.

Results

Phenotypic analysis of maturity in both environments

The maturity of 257 accessions representing diverse maturity groups (MGs) worldwide was evaluated under field conditions in Beijing, China, in 2020 and 2021 (BJ2020 and BJ2021) (Supplementary Data 1). The results revealed no significant differences between the two years at the Beijing location (Fig. 1a). Significant phenotypic variation in maturity was detected among accessions across all tested environments ($p < 0.05$; Supplementary Fig. 1). Moreover, maturity was significantly delayed with increasing MG ($p < 0.05$), indicating substantial natural variation in maturity-related traits among soybean accessions from different MGs.

Resequencing and genomic variation of soybean accessions from different maturity groups

Whole-genome resequencing was performed for the 257 soybean accessions using the Illumina NovaSeq 6000 platform, generating an average clean read depth of 9× per accession. Clean reads were aligned to the soybean reference genome assembly *Glycine max* Wm82.a2.v1, with an average mapping coverage of 94% (Supplementary Data 2). After stringent quality control and filtering, we identified 2,757,937 high-quality single-nucleotide polymorphisms (SNPs) distributed across the 20 soybean chromosomes, together with 315,478 insertions/deletions (InDels) (Supplementary Fig. 2).

Genetic relationships among the 257 accessions were assessed using multiple approaches, including principal component analysis (Supplementary Fig. 3), population genetic structure analysis (Supplementary Fig. 4) and phylogenetic analysis (Fig. 1b). These analyses revealed broad clustering patterns that generally corresponded to the geographic origins of Northern Spring Planting Region, the Huang–Huai–Hai Summer Planting Region, and the Southern Multiple

https://doi.org/10.1038/s41467-026-77316-3

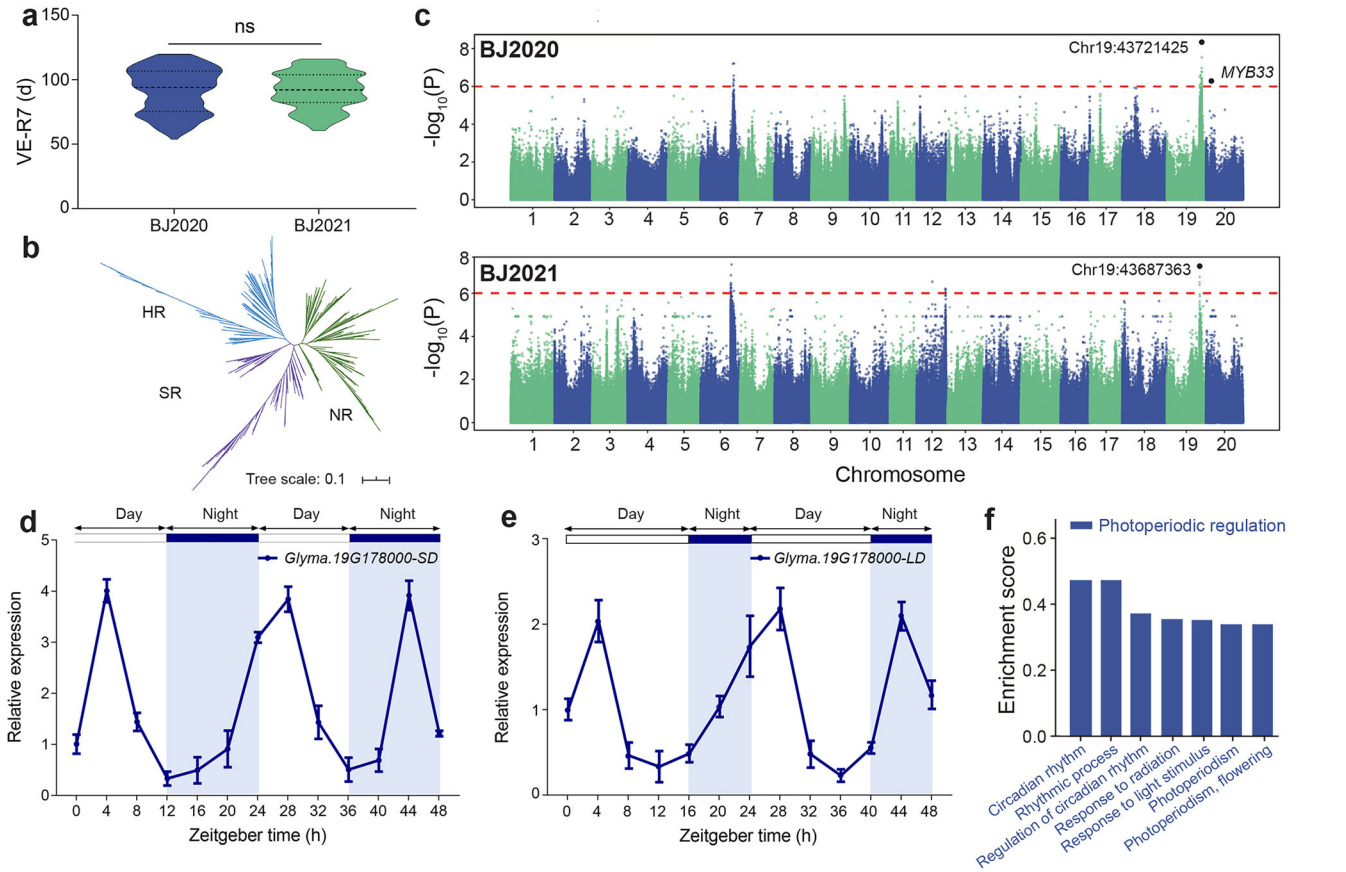


Fig. 1 | Genome-wide association analysis identifies *GmRVE4d* as a candidate locus associated with soybean maturity. a Phenotypic variation in soybean maturity duration across two environments, measured as days from emergence to physiological maturity. VE, emergence; R7, physiological maturity; VE–R7, days from VE to R7. ns indicates no significant difference among environments, as determined by one-way ANOVA followed by Tukey's test, $p \geq 0.05$; $n = 257$ accessions per environment. **b** Phylogenetic relationships of the soybean diversity panel inferred from genome-wide SNPs. NR, accessions from the Northern Spring Planting Region; HR, accessions from the Huang–Huai–Hai Summer Planting Region; SR, accessions from the Southern Multiple Cropping Region. **c** Genome-wide association study results for VE–R7 across two environments. BJ2020, Beijing (40°23′N, 116°57′E), China, in 2020; BJ2021, Beijing, China, in 2021. Marker–trait associations were evaluated using a mixed linear model with two-sided tests. Dashed horizontal lines indicate the Bonferroni-corrected significance threshold adjusted for multiple comparisons. Candidate loci and previously reported flowering- or maturity-related genes are annotated. Diurnal expression profiles of *Glyma.19G178000* under short-day conditions, SD, 12 h light/12 h dark (**d**), and long-day conditions, LD, 16 h light/8 h dark (**e**). Data are shown as mean ± SD from three independent biological replicates. **f** Gene Ontology biological process terms associated with *Glyma.19G178000*, predicted using the deep learning-based DeepGOPlus model. Selected terms relevant to photoperiodic regulation are shown.

Cropping Region in China. However, the ADMIXTURE analysis indicated substantial admixture among accessions from these regions, suggesting that geographic origin is not equivalent to genetic ancestry or a genetically discrete population boundary.

We further characterized linkage disequilibrium in this population. Linkage disequilibrium, quantified as the squared correlation coefficient (r^2) between SNP alleles, decreased with increasing physical distance. The linkage disequilibrium decay distance, defined as the physical distance at which r^2 decreased to half of its maximum value, was approximately 50 kb (Supplementary Fig. 5).

Identification of maturity-related QTLs by GWAS

Using maturity phenotypes collected from BJ2020 and BJ2021, we performed GWAS separately for each environment. A threshold of $-\log_{10} p \geq 6$, corresponding to $1/n$, where n represents the total number of SNPs, was used to identify significant marker–trait associations. To identify loci consistently detected in both environments, we further compared the physical positions of association peaks across both environments. Although association signals were detected on several chromosomes, including chromosome 6, these peaks were not consistently observed at similar genomic positions across both environments. By contrast, a locus on chromosome 19 showed a reproducible association peak at a consistent genomic position in both environments and was therefore selected for further analysis (Fig. 1c, Supplementary Fig. 6 and Supplementary Data 3). Other association signals detected on different chromosomes may correspond to previously reported maturity-related loci (*MYB33*) or may represent novel loci. However, these loci were not investigated further in this study because association peaks at the corresponding genomic positions were not consistently detected across environments, unlike the chromosome 19 locus.

To define the candidate region on chromosome 19, we examined the local linkage disequilibrium structure around the lead-associated SNPs based on the estimated genome-wide linkage disequilibrium decay distance of approximately 50 kb. Linkage disequilibrium block analysis showed that the associated SNPs at Chr19:43,721,425 and Chr19:43,687,363 were located within the same linkage disequilibrium block. This strongly linked block spanned an 84.1-kb interval from 43.672 to 43.756 Mb within the candidate genomic region (Supplementary Fig. 7). According to the soybean reference genome assembly *Glycine max* Wm82.a2.v1, this interval contains nine annotated genes (Supplementary Data 4).

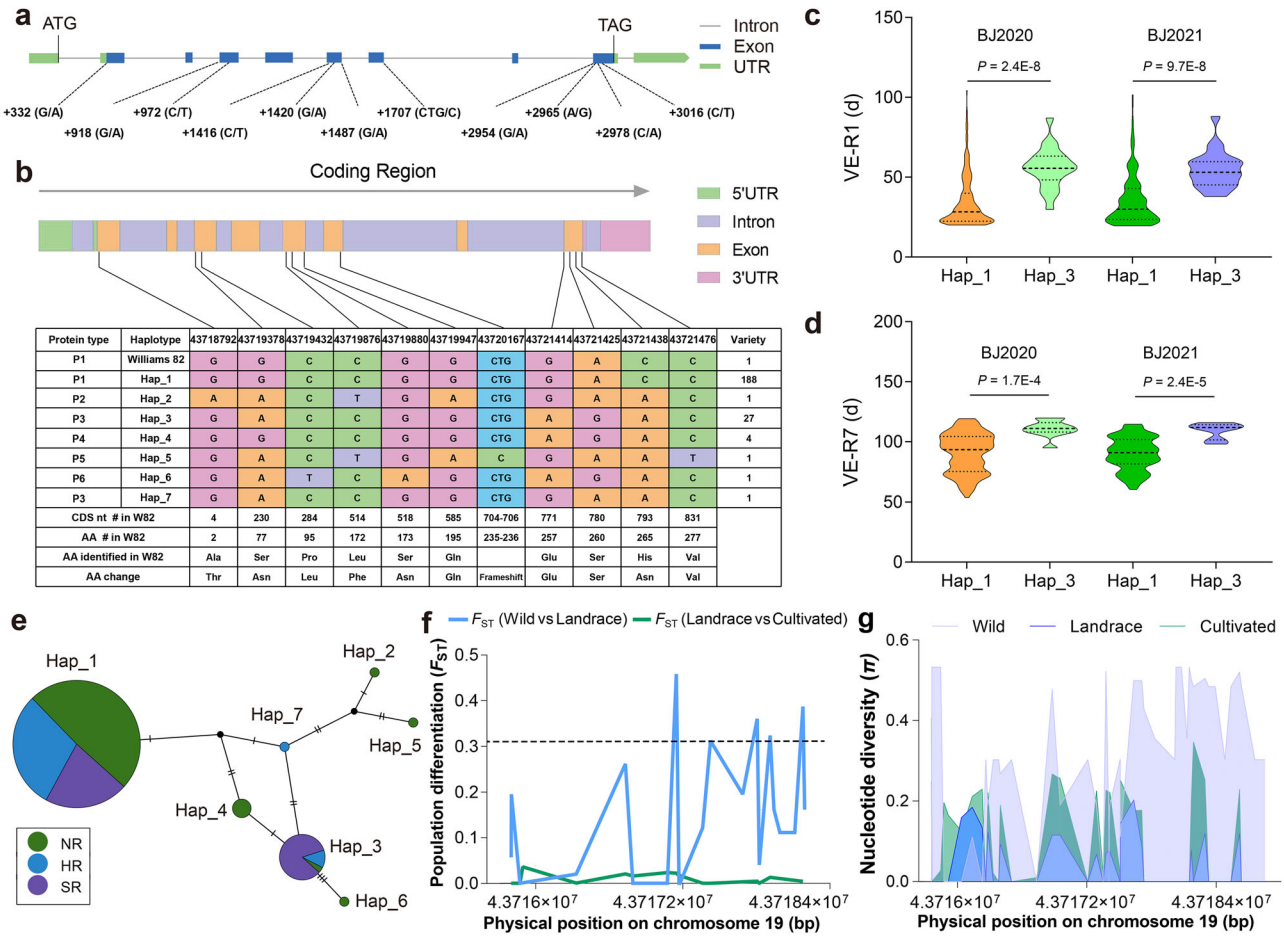
https://doi.org/10.1038/s41467-026-77316-3

Identification of a candidate maturity-associated locus

To prioritize candidate genes within the chromosome 19 maturity-associated locus, we first performed quantitative real-time PCR (qRT–PCR) to analyze the transcriptional profiles of genes in the candidate interval in the soybean cultivar Jack. Under both short-day (SD; 12 h light/12 h dark) and long-day (LD; 16 h light/8 h dark) conditions, only *Glyma.19G177500*, *Glyma.19G177600*, *Glyma.19G178000* and *Glyma.19G178200* showed time-dependent expression changes, suggesting that their expression was regulated by the light-dark cycle (Fig. 1d, e and Supplementary Fig. 8). We then predicted their functional annotations using DeepGOPlus. Among these four genes, *Glyma.19G178000* and *Glyma.19G178200* were predicted to be associated with photoperiodic regulation (Fig. 1f and Supplementary Data 5).

We further surveyed natural variation in the coding sequences (CDSs) of these two candidate genes and used the identified variants to define haplotypes. Nonsynonymous variants leading to amino acid substitutions were detected in *Glyma.19G178000*, but not in *Glyma.19G178200* (Fig. 2a, b and Supplementary Fig. 9). Accessions carrying *Glyma.19G178000*^{H3} showed significantly delayed flowering and maturity compared with those carrying *Glyma.19G178000*^{H1} across both environments ($p < 0.01$) (Fig. 2c, d), possibly owing to a Ser-to-Asn amino acid substitution at position 77 within the conserved domain (Fig. 2b and Supplementary Fig. 10). Geographical distribution analysis further revealed that *Glyma.19G178000*^{H1} was predominant in the Northern Spring Planting Region, whereas *Glyma.19G178000*^{H3} was more prevalent in the Southern Multiple Cropping Region (Fig. 2e); other haplotypes with small sample sizes ($n < 5$) were not further analyzed.

To determine whether *Glyma.19G178000* was associated with selection during soybean domestication, we performed a genome-wide cross-population composite likelihood ratio (XP-CLR) scan comparing wild soybeans and landraces. The genomic region harboring *Glyma.19G178000* showed evidence of selection during domestication (Supplementary Fig. 11). We next examined the local haplotype structure across the 257 soybean accessions. Within the 500-kb region flanking *Glyma.19G178000*, the gene was located



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 | Molecular characterization and haplotype analysis of the *GmRVE4d* locus in soybean. a Gene structure and coding region variants of *GmRVE4d* in the different haplotypes. Exons, introns and untranslated regions are indicated. ATG and TAG denote the translation start and stop codons, respectively. **b** Haplotypes detected in the coding region of *GmRVE4d* and their predicted amino acid changes. W82 denotes Williams 82, the soybean reference cultivar; CDS, coding sequence; nt, nucleotide; AA, amino acid; #, nucleotide or amino acid position. Comparison of flowering time, measured as days from emergence to first bloom, VE–R1 (**c**), and maturity duration, measured as days from emergence to physiological maturity, VE–R7 (**d**), between the major haplotypes *GmRVE4d*^{H1} and *GmRVE4d*^{H3} across two environments. VE, emergence; R1, first bloom; R7, physiological maturity. BJ2020, Beijing, China, in 2020; BJ2021, Beijing, China, in 2021. Statistical significance was determined using a two-sided Student's *t* test. **e** Haplotype network of *GmRVE4d* based on coding-region variants. Circle size is proportional to the number of accessions carrying each haplotype. Colors indicate soybean accessions from different ecological regions: NR, Northern Spring Planting Region; HR, Huang–Huai–Hai Summer Planting Region; SR, Southern Multiple Cropping Region. Short bars on branches indicate mutational steps between haplotypes. **f** Population differentiation, measured by F_{ST}, across the 2.7-kb promoter region of *GmRVE4d* in wild soybeans, landraces and cultivated soybeans. Blue and green lines indicate F_{ST} between wild soybeans and landraces, and between landraces and cultivated soybeans, respectively. The dashed line indicates the threshold of $F_{ST} = 0.31$. **g** Nucleotide diversity, π, across the 2.7-kb promoter region of *GmRVE4d* in wild soybeans, landraces and cultivated soybeans.

within an approximately 200-kb haplotype block characterized by strong linkage disequilibrium (Supplementary Fig. 12). We therefore conducted population genetic analyses across the entire linkage disequilibrium block using Tajima's *D*, population differentiation (F_{ST}) and nucleotide diversity (π), and further examined per-site F_{ST} and π in the 2.7-kb promoter region of *Glyma.19G178000*. These analyses revealed elevated differentiation between wild and domesticated soybeans, together with reduced nucleotide diversity in landraces and cultivated soybeans relative to wild soybean (Fig. 2f, g and Supplementary Figs. 13–15). These results supported the occurrence of domestication-associated selection at the genomic region harboring *Glyma.19G178000*.

Phylogenetic and domain analyses showed that *Glyma.19G178000* is homologous to *Arabidopsis RVE4* and encodes a 1R-MYB transcription factor containing a 50-amino-acid R motif (Supplementary Fig. 16). Accordingly, we named this gene *GmRVE4d*. Collectively, the expression profile, predicted function, haplotype-phenotype association, geographic distribution and population-genetic signatures described above supported the identification of *GmRVE4d* as a strong candidate gene at the chromosome 19 maturity-associated locus.

Tissue-specific expression analysis and subcellular localization of *GmRVE4d*

To determine the expression pattern of *GmRVE4d*, we performed qRT–PCR to examine *GmRVE4d* transcript levels in different soybean tissues. *GmRVE4d* was highly expressed in trifoliolate leaves, followed by flowers and true leaves (Supplementary Fig. 17), consistent with a potential role in tissues involved in perceiving environmental photoperiod signals in soybean. The GmRVE4d-GFP fusion protein was predominantly localized to the nucleus (Supplementary Fig. 18), consistent with its predicted function as a transcription factor that regulates downstream gene expression.

GmRVE4d acts as an inhibitor of flowering and maturity

To further verify the role of *GmRVE4d* in controlling flowering time and maturity, we used CRISPR-Cas9 to generate *Gmrve4d* mutants in the Jack background, which carries the *GmRVE4d*^{H1} haplotype. Three independent homozygous *Gmrve4d* mutant lines carrying 1-bp, 5-bp, and 33-bp deletions, respectively, were obtained for phenotypic analysis (Supplementary Fig. 19).

Under SD conditions, the time from emergence to flowering was significantly advanced by 2–3 days in the three independent mutant lines compared with Jack wild type (WT) ($p < 0.05$), and maturity was significantly advanced by 9–10 days ($p < 0.05$; Supplementary Fig. 20). Under LD conditions, flowering time was significantly advanced by 4–7 days in the three independent mutant lines ($p < 0.05$), and maturity was significantly advanced by 11 days ($p < 0.05$; Supplementary Fig. 20).

Under natural long-day (NLD) conditions from June to October in Beijing, China (40°23′N, 116°57′E), in 2023, flowering time was significantly advanced by 5 days in the three independent mutant lines ($p < 0.05$; Fig. 3a, b), and maturity was significantly advanced by 5–6 days ($p < 0.05$; Fig. 3a, d). Similarly, under NLD conditions in Xinxiang, Henan Province, China (35°10′N, 113°48′E), in 2023, flowering time was significantly advanced by 2–3 days in the three independent mutant lines ($p < 0.05$; Fig. 3f), and maturity was significantly advanced by 4–5 days ($p < 0.05$; 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 significantly increased *GmRVE4d* transcription levels in T_3 homozygous transgenic plants, hereafter referred to as *GmRVE4d*-OE lines, with transcript levels 77-, 30- and 80-fold higher than those in Jack plants, respectively (Supplementary Fig. 21). Western blot analysis further confirmed the presence of GmRVE4d-GFP fusion proteins in the *GmRVE4d*-OE lines (Supplementary Fig. 22).

Under SD conditions, flowering time was significantly delayed by 3 days in the *GmRVE4d*-OE lines compared with Jack ($p < 0.05$), and maturity was significantly delayed by 5–6 days ($p < 0.05$; Supplementary Fig. 23). Under LD conditions, flowering time was significantly delayed by 8–9 days in the *GmRVE4d*-OE lines compared with Jack plants ($p < 0.05$), and maturity was significantly delayed by 11–13 days ($p < 0.05$; Supplementary Fig. 23).

Under NLD conditions in Beijing in 2023, flowering time was significantly delayed by 5–10 days in the *GmRVE4d*-OE lines ($p < 0.05$; Fig. 3a, b), and maturity was significantly delayed by 7 days ($p < 0.05$; Fig. 3a, d). Similarly, under NLD conditions in Xinxiang in 2023, flowering time was significantly delayed by 10–11 days in the *GmRVE4d*-OE lines ($p < 0.05$; Fig. 3f), and maturity was significantly delayed by 11–14 days ($p < 0.05$; Fig. 3h). Consistent phenotypic results were observed in the 2025 field trial (Fig. 3c, e, g, i). These results demonstrated that *GmRVE4d* delayed flowering and maturity under both controlled and natural photoperiodic conditions.

Notably, the growth period structure, defined here as the ratio of reproductive to vegetative growth duration (R/V), was highest in the *Gmrve4d* mutant lines under each photoperiod condition, followed by Jack and *GmRVE4d*-OE lines (Supplementary Fig. 24). These results suggested that elevated GmRVE4d activity prolonged pre-flowering development. However, because 35S-driven expression bypasses native temporal and environmental regulation, detailed phase-specific functions of endogenous *GmRVE4d* will require further analysis using native-promoter materials.

GmRVE4d directly inhibits the transcription of *GmPRR5a*

To gain further insight into how *GmRVE4d* delays flowering and maturity, we performed RNA-seq to compare genome-wide gene expression profiles among Jack, *GmRVE4d*-OE, and *Gmrve4d* mutant lines (Supplementary Data 6). A Venn diagram identified 209 differentially expressed genes shared between the comparisons of the *GmRVE4d-OE* and *Gmrve4d* mutant lines with Jack (Supplementary Fig. 25). We then performed DNA affinity purification sequencing (DAP-seq) to identify direct regulatory targets of GmRVE4d at the whole-genome scale (Supplementary Fig. 26). By integrating the RNA-seq and DAP-seq results, we identified 56 candidate target genes (Fig. 4a). Among these genes, *GmPRR5a* is a soybean homolog of *Arabidopsis PSEUDO-RESPONSE REGULATOR 5* (*PRR5*), which encodes a central component of the circadian clock (Supplementary Fig. 27). MEME analysis of GmRVE4d-binding peaks revealed a significantly enriched motif corresponding to the canonical EE, with a conserved 9-bp core sequence 5′-AGATATTTT-3′ that is the reverse complement of the widely accepted EE consensus sequence (Fig. 4b, c). Expression pattern analysis further showed that *GmPRR5a* displayed a time-dependent expression pattern opposite to that of *GmRVE4d* under both SD and LD conditions, suggesting that their expression is regulated by the light–dark cycle (Fig. 1d, e and Supplementary Fig. 28). Therefore, on the basis of the integrated RNA-seq and DAP-seq analyses, *GmPRR5a* was selected as a candidate direct target gene of GmRVE4d.

To determine whether GmRVE4d binds to the *GmPRR5a* promoter, we performed a dual-luciferase reporter assay. These assays showed that GmRVE4d inhibited reporter gene expression driven by the *GmPRR5a* promoter, supporting a direct repressive effect on *GmPRR5a* transcription (Fig. 4d-f). We also conducted electrophoretic mobility shift assays (EMSAs) to examine the in vitro binding of GmRVE4d to synthetic oligonucleotide probes containing the

https://doi.org/10.1038/s41467-026-77316-3

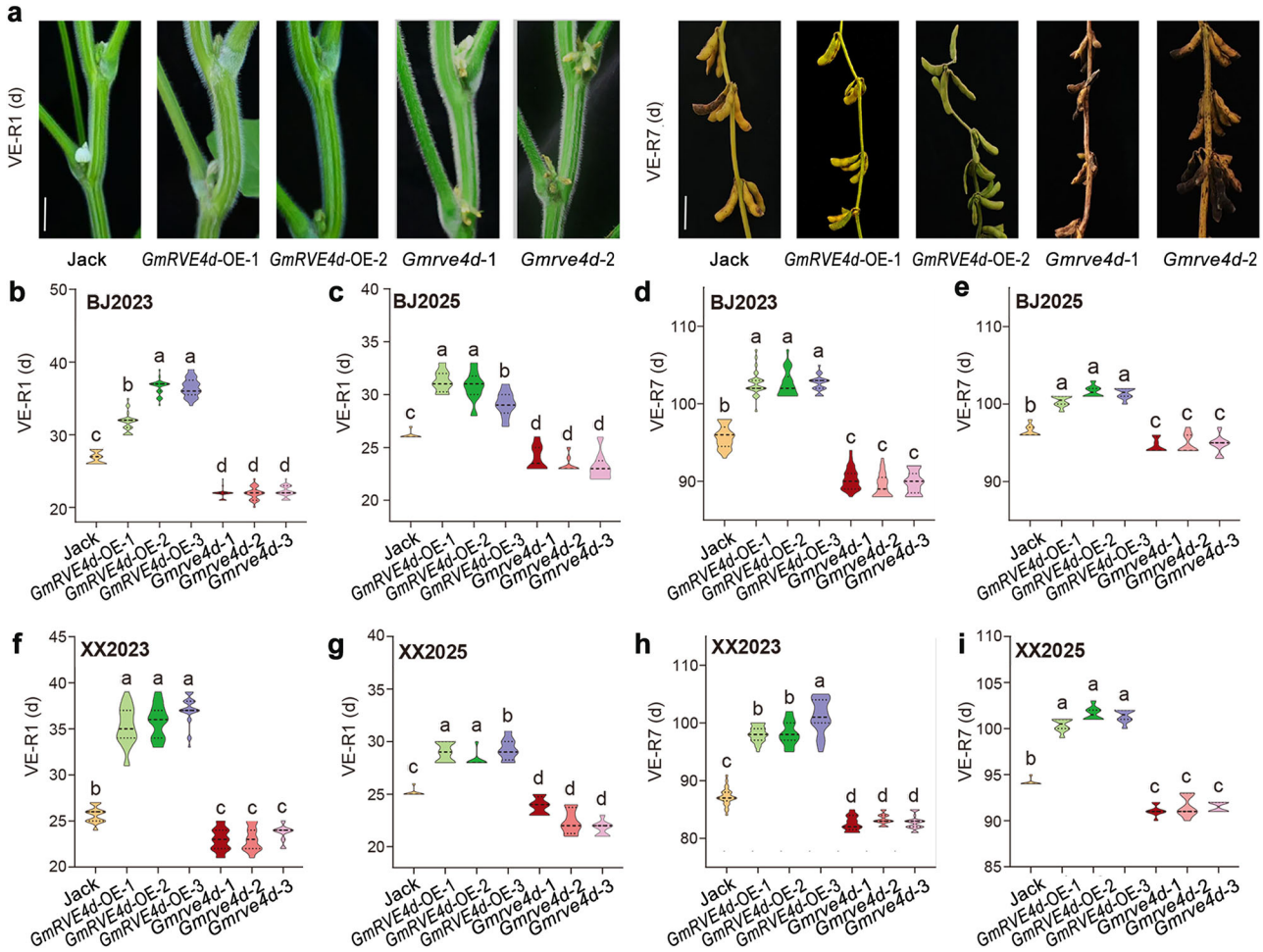


Fig. 3 | Functional characterization of *GmRVE4d* in regulating soybean flowering time and maturity. a Plant phenotypes related to flowering time and maturity under natural long-day (NLD) conditions. Flowering time (days from VE to R1) across Beijing in 2023 (**b**) and 2025 (**c**), Xinxiang in 2023 (**f**) and 2025 (**g**). VE, 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. $n = 25$ plants (2023) and 8 plants (2025). Flowering stage scale bars: 3 cm. Maturity stage scale bars: 5 cm. Different lowercase letters indicate significant differences among genotypes, as determined by one-way ANOVA followed by Tukey's test, $p < 0.05$.

conserved sequence 5′-AGATATTTT-3′ from the *GmPRR5a* promoter region (Fig. 4g).

We further performed chromatin immunoprecipitation followed by qPCR (ChIP-qPCR) using transgenic 35Spro:*GmRVE4d*-GFP plants. Trifoliolate leaves from transgenic plants grown under SD photoperiod were harvested for ChIP-qPCR using an anti-GFP antibody, with immunoglobulin G (IgG) as a negative control. GmRVE4d enrichment was detected at the *GmPRR5a* promoter (Fig. 4h), further supporting the binding of GmRVE4d to the *GmPRR5a* promoter in vivo.

Phenotypic analyses indicated that *GmRVE4d*^{H3} exerted a stronger effect than *GmRVE4d*^{H1}. To test whether this difference was associated with promoter-binding activity, we performed yeast one-hybrid (Y1H) assays. The results showed that GmRVE4d directly bound to the *GmPRR5a* promoter in yeast, and *GmRVE4d*^{H3} showed stronger binding activity than *GmRVE4d*^{H1} (Fig. 4i).

Collectively, these results demonstrated that GmRVE4d directly bound to the promoter of its downstream target gene *GmPRR5a* and repressed its transcription. Natural variation in *GmRVE4d* enhanced the activity of the GmRVE4d^{H3} protein, which was associated with delayed flowering and maturity. These findings suggested that the *GmRVE4d*^{H3} haplotype might have contributed to soybean adaptation to lower-latitude regions where delayed flowering and maturity could be agronomically favorable.

The *GmRVE4d-GmPRR5a-GmFT2a/GmFT5a* regulatory cascade modulates flowering time and maturity in soybean

To further determine whether *GmPRR5a* regulates soybean flowering time and maturity, we used CRISPR-Cas9 to generate *Gmprr5a* mutants in the Jack background. Two independent homozygous *Gmprr5a* mutant lines were identified and used for phenotypic analysis (Supplementary Fig. 29).

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

Under natural short-day (NSD) conditions from August to October in Sanya, China (18°14′N, 109°13′E), in 2025, flowering time was significantly delayed by 2 days in the two independent mutant lines ($p < 0.05$; Fig. 5d), and maturity was significantly delayed by 4 days ($p < 0.05$; Fig. 5g). These results indicated that *GmPRR5a* promoted flowering and maturity in soybean, consistent with the delayed-flowering and delayed-maturity phenotypes observed when *GmPRR5a* function was disrupted.

Expression analysis further revealed that the transcript levels of the florigen-encoding genes *GmFT2a* and *GmFT5a* were significantly

https://doi.org/10.1038/s41467-026-77316-3

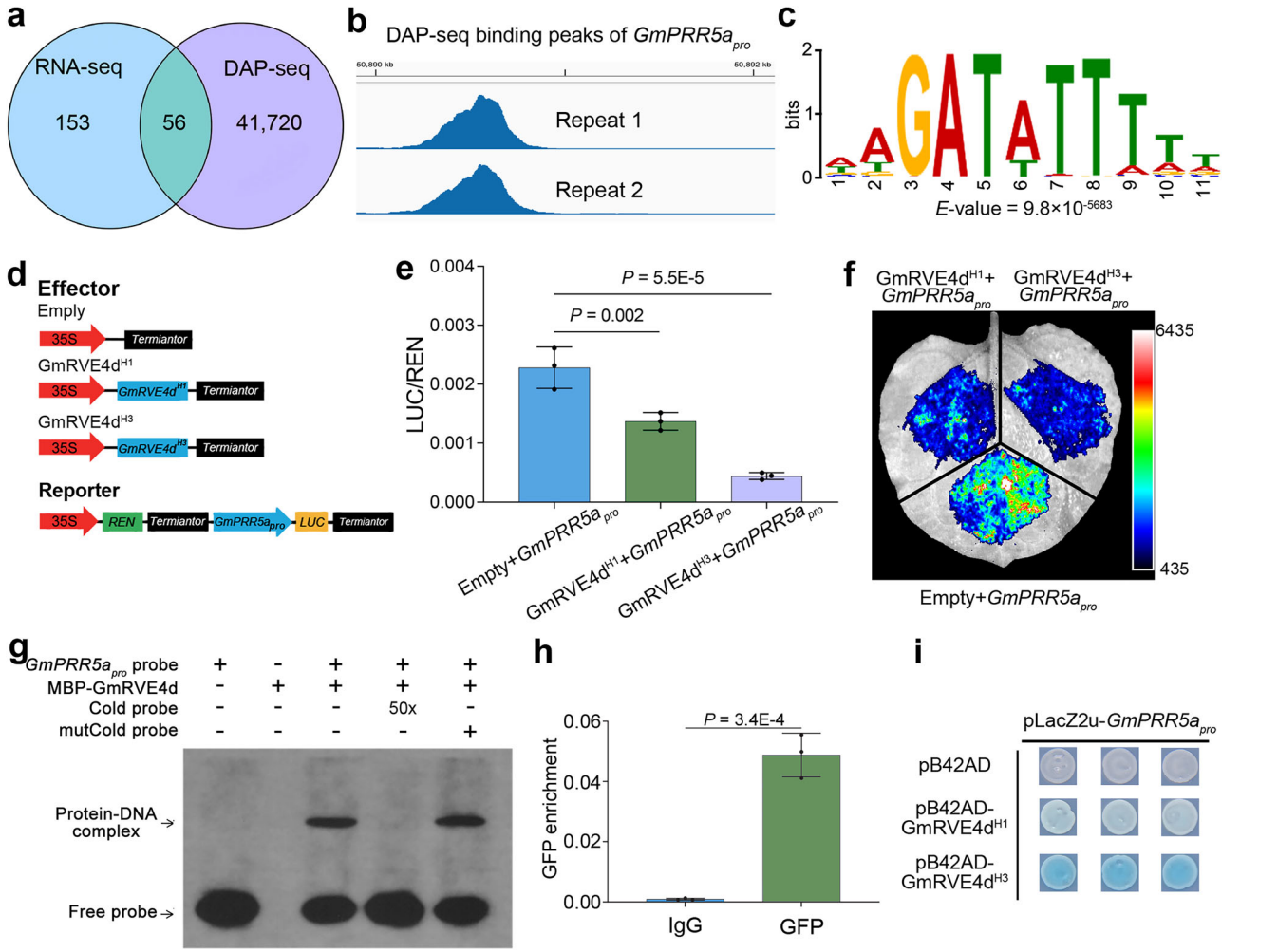


Fig. 4 | Identification of downstream target genes of GmRVE4d. a Venn diagram showing the overlap between differentially expressed genes identified by RNA-seq and candidate target genes identified by DAP-seq. **b** DAP-seq binding peaks showing enrichment of GmRVE4d in the promoter region of *GmPRR5a*. **c** Putative GmRVE4d-binding motif identified from DAP-seq peak-associated sequences using MEME. **d** Schematic representation of the effector and reporter constructs used in the dual-luciferase reporter assay. The effector constructs contained the empty vector, *GmRVE4d*^{H1} or *GmRVE4d*^{H3} driven by the CaMV 35S promoter, and the reporter construct contained the *GmPRR5a* promoter driving LUC and the CaMV 35S promoter driving REN. **e** Relative LUC/REN activity showing that *GmRVE4d*^{H1} or *GmRVE4d*^{H3} repress the transcriptional activity of the *GmPRR5a* promoter in tobacco leaves. **f** Representative bioluminescence images of tobacco leaves transiently expressing the indicated effector and reporter constructs. **g** EMSA showing the direct binding of MBP–GmRVE4d to the *GmPRR5a* promoter probe. Biotin-labeled oligonucleotides containing the AGATATTTT motif were used as probes. The probe sequence of Gm*PRR5a* was determined based on the location of the binding peaks in DAP-seq results. Excess unlabeled probe was used as a competitor, and mutated unlabeled probe was used as a negative control. The experiment was independently repeated three times with similar results. **h** ChIP–qPCR assay validating the in vivo enrichment of GmRVE4d at the *GmPRR5a* promoter region. **i** Yeast one-hybrid assay showing the direct binding of GmRVE4d to the *GmPRR5a* promoter. The pB42AD empty vector was used as the negative control. Data are means ± SD of three independent biological replicates. Statistical significance was determined using one-way ANOVA followed by a multiple-comparisons test. Exact adjusted *p* values are shown in the figure. The MEME *E*-value for the identified motif is 9.8×10^{-5683}.

reduced in *Gmprr5a* mutant plants compared with Jack ($p < 0.05$; Supplementary Fig. 30), consistent with the delayed-flowering phenotype of the mutants. Given that *Arabidopsis* PRR proteins regulate flowering partly through the CDF–CO–FT module, we further examined the expression of soybean *CDF* and *CO*-like homologs. The expression levels of *GmCDF1* and *GmCDF2* were significantly decreased in the *Gmprr5a* mutants, whereas those of *GmCOL1a* and *GmCOL1b* were increased ($p < 0.05$; Supplementary Fig. 30). These results indicated that *GmPRR5a* affected the expression of soybean *CDF* and *CO*-like homologs. However, the reduced expression of soybean *GmCDF1/2* and increased expression of *GmCOL1a/b* in the mutants, together with the downregulation of *GmFT2a/GmFT5a*, suggested that regulation of *GmFT2a/GmFT5a* by *GmPRR5a* could not be fully explained by a simple conserved PRR–CDF–CO–FT cascade as described in *Arabidopsis*. Thus, *GmPRR5a* may regulate soybean flowering through a more complex or soybean-specific mechanism that ultimately affects *GmFT2a* and *GmFT5a* expression.

Discussion

GmRVE4d is a major gene affecting soybean maturity and latitudinal adaptability

Flowering time and maturity are crucial traits for soybean ecological adaptation and are tightly regulated by photoperiod[38]. In *Arabidopsis*, RVE4 and RVE8 play important roles in responses to photoperiod stress[25,39]. Our results suggest that *GmRVE4d* is associated with domestication-related selection signals and that the *GmRVE4d*^{H3} haplotype contributes to soybean adaptation in low-latitude regions, possibly through altered responses to photoperiod-related environmental conditions, whereas *GmRVE4d*^{H1} is more prevalent in high-

https://doi.org/10.1038/s41467-026-77316-3

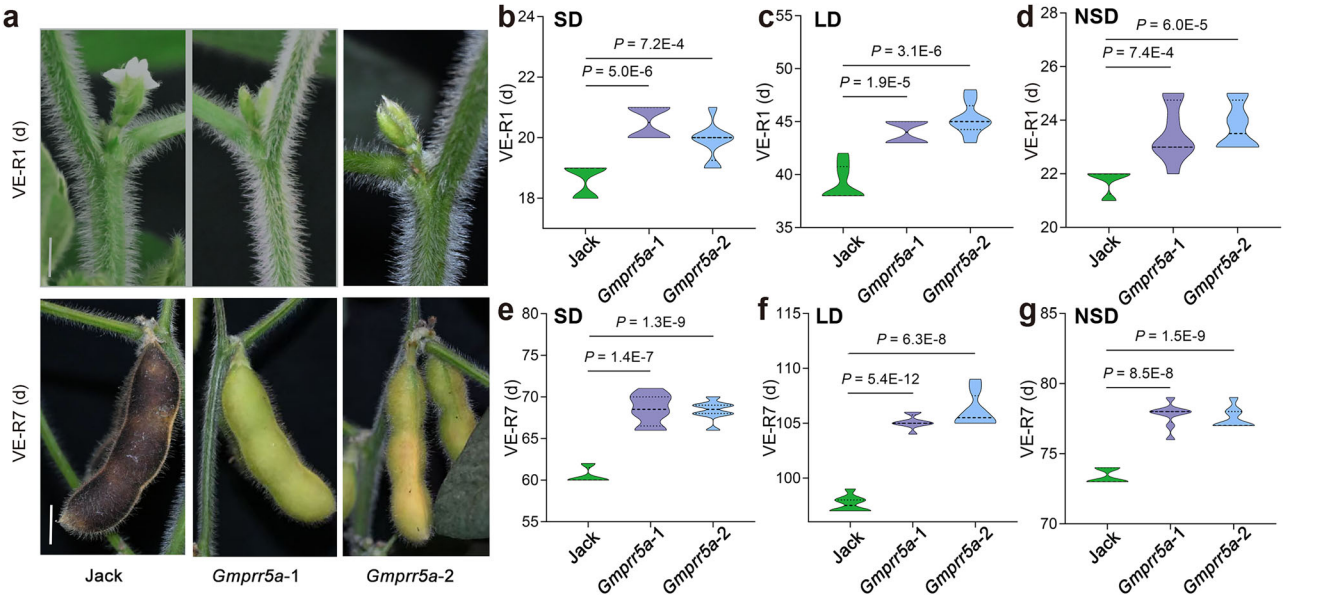


Fig. 5 | Functional characterization of *GmPRR5a* in regulating flowering time and maturity. a Plant phenotypes of the wild-type cultivar Jack and the *Gmprr5a* mutants (*Gmprr5a*-1 and *Gmprr5a*-2) at the flowering and maturity stages under short-day conditions. At the flowering stage, Jack had opened flowers, whereas *Gmprr5a*-1 and *Gmprr5a*-2 showed delayed floral development. At the maturity stage, Jack showed mature dark-brown pods, whereas the *Gmprr5a* mutants retained yellow-green immature pods. Scale bars, 0.5 cm for the flowering stage and 1 cm for the maturity stage. Flowering time, measured as days from vegetative emergence to beginning bloom (VE–R1), of Jack, *Gmprr5a*-1 and *Gmprr5a*-2 under short-day conditions (SD; 12 h light/12 h dark) (**b**), long-day conditions (LD; 16 h light/8 h dark) (**c**) and natural short-day conditions (NSD) in Sanya, Hainan Province, China (18°14′N, 109°13′E), in 2025 (**d**). Maturity, measured as days from vegetative emergence to physiological maturity (VE–R7), of Jack, *Gmprr5a*-1 and *Gmprr5a*-2 under SD (**e**), LD (**f**) and NSD (**g**) conditions. VE, vegetative emergence; R1, beginning bloom; R7, physiological maturity. $n = 8$ plants. One-way ANOVA followed by Tukey's multiple-comparison test was used to assess differences among genotypes. Exact adjusted p values are shown in the figure.

latitude regions. These findings indicate that *GmRVE4d* is an important gene controlling soybean maturity and adaptability.

In soybean, many circadian clock-related genes function as negative regulators of flowering and maturity, including *E2*, *GmPRR3a*, *GmPRR3b* and *GmLHY1a*, whereas several others, such as *GmELF3* and *GmLUX*, act as positive regulators[13]. These genes exhibit distinct expression patterns under different photoperiod conditions, which contribute to their diverse regulatory effects. In this study, we identified *GmRVE4d* as a negative regulator of flowering and maturity in soybean, whereas its downstream target gene *GmPRR5a* acts as a positive regulator. Moreover, our findings reveal that *GmRVE4d*, similar to *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[28,33,34].

Previous research has highlighted maturity as a fundamental domestication trait[40], with gradual changes in growth-period structure from wild to cultivated soybeans driven by artificial selection for higher yield and broader adaptation[41,42]. Furthermore, the latitudinal expansion of soybean is associated with selection on alleles of core circadian clock-associated genes[43]. Consistent with this notion, the genomic region harboring *GmRVE4d* showed elevated population differentiation and reduced nucleotide diversity between wild and cultivated soybeans, together with distinct haplotype distributions across geographic regions. We note, however, that the number of wild soybean accessions included in this study was limited; therefore, the nucleotide-diversity comparison between wild and cultivated groups should be interpreted with caution. Nevertheless, when considered together with evidence from XP-CLR, F_{ST}, Tajima's D, local haplotype structure, and haplotype-phenotype association analyses, these results collectively support *GmRVE4d* as a major maturity-associated gene contributing to soybean latitudinal adaptation.

GmRVE4d is a candidate molecular target for improving soybean ecological adaptability

Improving soybean adaptability across different latitudes and expanding its cultivation area is a crucial strategy to meet the increasing demand for soybean products. Previous studies have shown that the wide geographical distribution of soybean is associated with abundant natural variation and diverse combinations of genes and QTLs regulating flowering time and maturity[34]. Although several major genes associated with maturity traits have been cloned and characterized[13], there remains 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 provide valuable targets for future genetic improvements and enhance soybean ecological adaptability.

In recent years, a modern breeding strategy termed "Potalaization" has been proposed to generate widely adapted soybean varieties by combining multiple molecular tools in existing elite and broadly adapted genetic backgrounds[44]. In this study, CRISPR-Cas9-mediated targeted mutagenesis of *GmRVE4d* in the soybean cultivar Jack generated *Gmrve4d* mutants that matured approximately 5–6 days earlier under NLD conditions in Beijing and 4–5 days earlier under NLD conditions in Xinxiang in the 2023 field trial (Fig. 3d, h). Consistent early-maturity phenotypes were also observed in the 2025 field trial (Fig. 3e, i). These results suggested that CRISPR-Cas9 technology could be used to create *Gmrve4d* mutants in Jack or other elite soybean varieties, enabling their potential introduction into higher-latitude regions or use in spring sowing within multiple-cropping systems. In addition, transgenic overexpression of *GmRVE4d* significantly delayed flowering and maturity when plants were grown under NLD conditions in Beijing and Xinxiang (Fig. 3). Thus, *GmRVE4d* could also be used to breed cultivars with a longer growth period and delayed maturity in temperate or lower-latitude regions where extended vegetative growth is agronomically favorable.

https://doi.org/10.1038/s41467-026-77316-3

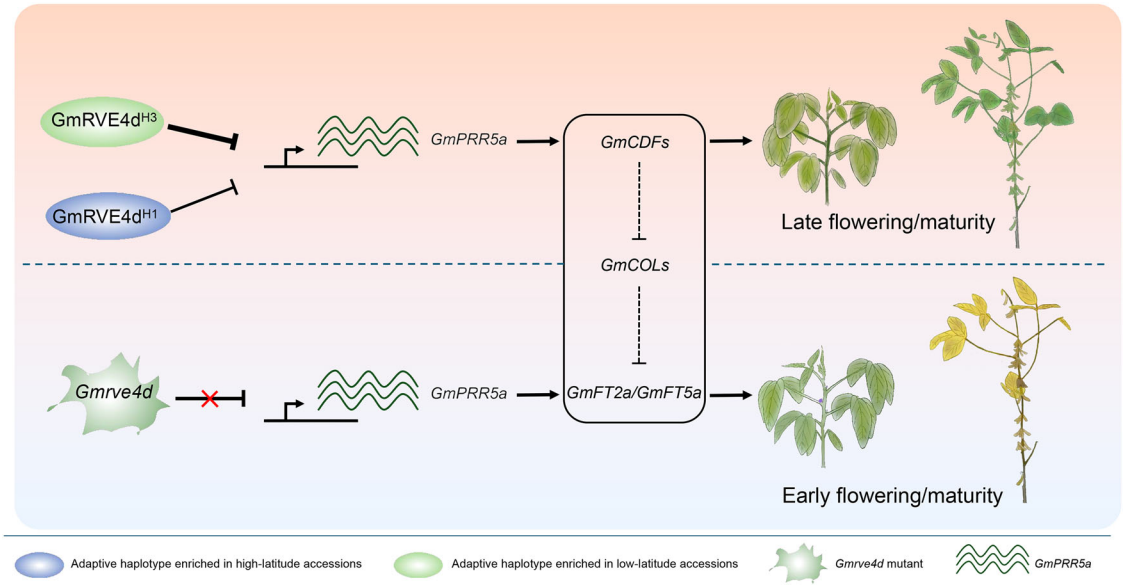


Fig. 6 | Proposed model for *GmRVE4d*-mediated regulation of soybean flowering and maturity. Adaptive haplotypes of *GmRVE4d* contribute to soybean flowering and maturity regulation by repressing *GmPRR5a* expression. The low-latitude-enriched haplotype *GmRVE4d*^{H3} and the high-latitude-enriched haplotype *GmRVE4d*^{H1} directly regulate the *GmPRR5a* promoter, with their regulatory effects being associated with changes in a downstream photoperiod-related module involving *GmCDFs*, *GmCOLs* and *GmFT2a/GmFT5a*. In plants carrying functional *GmRVE4d* haplotypes, repression of *GmPRR5a* is associated with reduced *GmFT2a/GmFT5a* expression and delayed flowering and maturity. In contrast, mutation of *GmRVE4d* releases *GmPRR5a* expression and is associated with increased *GmFT2a/GmFT5a* expression, resulting in early flowering and early maturity. *GmCDFs* and *GmCOLs* are indicated as gene families. Arrows indicate positive regulation or transcriptional activation, whereas T-bars indicate negative regulation or transcriptional repression. Solid lines indicate direct regulatory relationships supported by experimental evidence in this study, and dashed lines indicate putative or expression-associated links that remain to be further validated.

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 supported by 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[45]. These findings suggested that PRR proteins and their associated regulatory modules could have central roles in circadian transcriptional regulation. In soybean, our results supported a model in which GmRVE4d acts upstream of *GmPRR5a* and delays flowering and maturity by repressing *GmPRR5a* transcription.

Previous studies have shown that mutations in soybean *GmLCL/GmLHY* genes can result in late flowering[26,46]. Therefore, *GmPRR5a* may regulate downstream clock-associated genes, including *GmLCL/GmLHY* homologs, which may have functions distinct from those of their *Arabidopsis* orthologues. Notably, LNK1 is an essential transcriptional coactivator for the expression of the circadian clock gene *PRR5* in *Arabidopsis*; it is recruited to the *PRR5* promoter through interaction with bona fide DNA-binding proteins such as RVE4 and RVE8[25,47]. In this study, AlphaFold3-based structural predictions suggested that GmRVE4d may exert its regulatory function by recruiting GmLNK1 and GmLNK2 to form a protein complex (Supplementary Fig. 31)[48]. However, whether GmRVE4d physically interacts with GmLNK1 or GmLNK2 remains to be experimentally validated using molecular assays such as co-immunoprecipitation (Co-IP) or bimolecular fluorescence complementation (BiFC). In addition, whether *GmPRR5a* regulates downstream genes through the *E2*/EC pathway and thereby affects soybean flowering and maturity remains to be further clarified[37]. Further studies of *GmRVE4d* and *GmPRR5a* will help elucidate the molecular mechanisms underlying the soybean light–dark cycle-associated regulatory network and its contribution to photoperiod-dependent regulation of flowering and maturity.

Taken together, our results supported a regulatory model in which GmRVE4d-mediated repression of *GmPRR5a* contributes to soybean flowering and maturity regulation under different photoperiod conditions and field environments (Fig. 6). The GmRVE4d^{H3} protein, encoded by the low-latitude-enriched *GmRVE4d*^{H3} haplotype, and the GmRVE4d^{H1} protein, encoded by the high-latitude-enriched GmRVE4d^{H1} haplotype, both bind to the *GmPRR5a* promoter and repress its transcription, with GmRVE4d^{H3} showing stronger repressive activity. Because *GmPRR5a* promotes *GmFT2a* and *GmFT5a* expression and accelerates flowering and maturity, stronger repression of *GmPRR5a* by GmRVE4d^{H3} is associated with delayed flowering and maturity. We further incorporated a putative photoperiod-related module into the model, in which *GmPRR5a* may promote *GmCDF* expression or activity, GmCDFs may repress *GmCOL* components, and FT-repressive GmCOLs may negatively affect *GmFT2a* and *GmFT5a* expression. This provides a possible route by which the *GmRVE4d–GmPRR5a* module may be connected to florigen transcript output. Nevertheless, because the *GmPRR5a–GmCDF–GmCOL–GmFT2a/GmFT5a* connections were not directly tested by molecular interaction, DNA-binding or genetic epistasis assays in this study, we depicted these relationships as putative expression-associated links. Overall, our findings identified the *GmRVE4d–GmPRR5a* module as a clock-associated regulatory node that links natural haplotype variation with soybean flowering and maturity adaptation, providing a potential target for fine-tuning maturity in region-specific molecular breeding.

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, 33 landraces and 214 cultivated soybeans. 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, China (40°23′N, 116°57′E), in 2020 and 2021, to evaluate maturity phenotypes.

To evaluate the flowering and maturity phenotypes of Jack, *GmRVE4d* transgenic lines, *Gmrve4d* mutants and *Gmprr5a* mutants, plants were grown under natural field conditions in Beijing and Xinxiang, Henan Province, China (35°10′N, 113°48′E) from June to October, and in Sanya, Hainan Province, China (18°14′N, 109°13′E), from August to October. Plants were also grown under controlled growth chamber conditions: SD and LD. Each location–year combination was regarded as an independent field environment, and the field environments were named BJ2020, BJ2021, BJ2023, BJ2025, XX2023, XX2025, and SY2025.

Flowering time was recorded at the R1 stage, defined as the number of days from emergence to the first one open flower at any node on the main stem. Maturity was recorded at the R7 stage, defined as the number of days from emergence to the stage at which one normal pod on the main stem had reached its mature pod color[49]. The reproductive/vegetative growth period ratio (R/V) was used as an indicator of maturity structure[50].

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 based on the soybean reference genome *Glycine max* Wm82.a2.v1 under the following accession numbers: *GmRVE4d* (*Glyma.19G178000*), *GmPRR5a* (*Glyma.19G260400*), *GmCDF1* (*Glyma.13G062500*), *GmCDF2* (*Glyma.05G025900*), *GmCOL1a* (*Glyma.04G062400*), *GmCOL1b* (*Glyma.06G063000*), *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, Beijing, China) according to the manufacturer's instructions. DNA concentration and purity were assessed using a NanoDrop spectrophotometer and Qubit fluorometer, and DNA integrity was examined by agarose gel electrophoresis. Only high-quality DNA samples were used for library construction. Genome sequencing libraries were constructed using the DNA Library Prep Kit (Vazyme, Nanjing, China) following the manufacturer's protocol and sequenced on the Illumina NovaSeq 6000 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[51]. Clean reads were then aligned to the soybean reference genome *Glycine max* Wm82.a2.v1 using BWA v0.7.19 with default parameters[52]. SNP calling and annotation were performed using SAMtools[53], BCFtools,[54] and SnpEff[55].

SNP quality control was performed using VCFtools[56]. Only high-quality biallelic SNPs were retained for downstream analyses. SNPs with a missing rate >10%, QUAL < 50.0, QD < 5.0, MQ < 30.0 or minor allele frequency (MAF) < 0.01 were removed. InDels and non-biallelic variants were excluded. After quality control and MAF filtering, 2,757,937 high-quality SNPs were retained for genome-wide association analysis. To generate an approximately independent marker set for population structure analyses, linkage disequilibrium pruning was performed using PLINK with the parameter --indep-pairwise 50 10 0.2, corresponding to a sliding window of 50 SNPs, a step size of 10 SNPs and a linkage disequilibrium threshold of $r^2 = 0.2$[57]. The linkage disequilibrium-pruned SNP set was used for principal component analysis, ADMIXTURE analysis and phylogenetic analysis.

Population structure and linkage disequilibrium analysis

Population structure was inferred using ADMIXTURE based on the MAF-filtered SNP dataset. The number of ancestral populations, *K*, was tested from 2 to 15 with cross-validation, and the optimal *K* value was determined according to the lowest cross-validation error. High-quality SNPs were used to construct a neighbor-joining tree with SNPhylo software[58], and the resulting phylogenetic tree was visualized using the online interactive tool iTOL (https://itol.embl.de). Principal component analysis was performed using the princomp function in R[59] and visualized with the Scatterplot3d package[60]. Genome-wide SNP distribution density was analyzed with the CMplot package in R.

Genome-wide linkage disequilibrium decay was estimated using PopLDdecay[61]. Linkage disequilibrium decay was calculated separately for wild soybeans, landraces, cultivated soybeans and the total population using SNPs with MAF ≥ 0.01. Pairwise linkage disequilibrium was measured as the squared correlation coefficient r^2 between SNP pairs within a maximum physical distance of 500 kb. Linkage disequilibrium decay curves were plotted using the Plot_MultiPop.pl script implemented in PopLDdecay. The linkage disequilibrium decay distance was defined as the physical distance at which r^2 declined to half of its maximum value[62].

GWAS for maturity traits

GWAS for maturity traits was performed using GAPIT v3 with a mixed linear model (MLM) (http://zzlab.net/GAPIT)[63]. The full set of 2,757,937 high-quality SNPs was used for association testing to retain genome-wide marker information. To control for population structure and relatedness among the 257 soybean accessions, the first three principal components were included as fixed-effect covariates, and a genomic relationship matrix calculated in GAPIT was incorporated as a random effect. Manhattan plots and quantile-quantile plots were generated using the qqman package in R[64]. Model fit was evaluated using quantile-quantile plots and genomic inflation factors. The significant association threshold was set as $1/n$, where n represents the total number of SNPs.

For the phenotypic data used in GWAS, two biological replicates were evaluated per accession per trial, with five plants per replicate. Individual plant values were averaged to obtain accession-level means, which were used for association analysis.

XP-CLR, nucleotide diversity, Tajima's *D*, and population differentiation analyses

A genome-wide XP-CLR analysis was performed between wild soybeans and landraces[65]. XP-CLR scores were calculated using 100-kb sliding windows with a 50-kb step size, an LD threshold of 0.95, and a maximum of 500 SNPs per window. Genomic intervals with XP-CLR scores in the top 5% of the genome-wide distribution were defined as candidate regions under selection. The 200-kb haplotype block encompassing the candidate gene on chromosome 19 (Chr19: 43,681-43,881 kb; reference genome Wm82.a2.v1) was extracted from the quality-filtered VCF file using VCFtools. Tajima's *D* was calculated separately for wild soybeans, landraces and cultivated soybeans using non-overlapping 10-kb windows. π was estimated for each group using a 10-kb window and a 5-kb step size. F_{ST} was calculated using the Weir and Cockerham method implemented in VCFtools. Pairwise F_{ST} was estimated between wild soybeans and landraces and between landraces and cultivated soybeans using 10-kb sliding windows with a 5-kb step size.

Identification of candidate genes using a DL method

A DL tool DeepGOPlus[66,67], was used to predict functions 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. Input protein sequences were passed through a set of convolutional neural network layers with different filter sizes, each containing 512 filters to learn sequence features of a specific size. A max-pooling layer was then used to extract the maximum feature values from the representations acquired by each filter. This configuration generated an output containing 8192 convolutional filters as sequence features. Finally, a sigmoid activation function combined with a fully connected layer was used for prediction and classification.

Haplotype and genetic diversity analysis

Haplotype analysis of SNPs and InDels within the exons of *GmRVE4d* was performed using vcf2phylip v2.0[68]. The geographical distribution patterns of haplotypes were analyzed using PopART v1.7. π and F_{ST} among wild soybean, 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, including root, hypocotyl, cotyledon, stem, true leaf, trifoliolate leaf, growing point, flower, pod, and seed at key developmental stages. To analyze the diurnal expression pattern of *GmRVE4d* and candidate genes, Jack plants were grown under SD and LD conditions in growth chambers for three weeks. Fully expanded trifoliolate leaves were collected every 4 h over a 48-h period under both photoperiod conditions. To compare transcript levels of *GmRVE4d* among Jack, *GmRVE4d*-OE lines and *Gmrve4d* mutants, fully expanded trifoliolate leaves were harvested from 3-week-old plants grown under SD conditions. For each biological replicate, tissues from three individual plants were pooled, immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

Total RNA was extracted from all samples using the RNAprep Pure Plant Kit (TIANGEN, Beijing, China) according to the manufacturer's instructions. RNA concentration and purity were determined using a NanoDrop spectrophotometer, and RNA integrity was assessed by agarose gel electrophoresis. Complementary DNA synthesis and genomic DNA removal were performed with the FastKing RT Kit with genomic DNA Eraser (TIANGEN, Beijing, China) using 1 μg of total RNA per reaction according to the manufacturer's protocol.

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, Shiga, Japan). Each reaction was performed in a final volume of 10 μL containing 5 μL SYBR Premix, 1.2 μL diluted cDNA, 0.4 μM forward and reverse primers, and nuclease-free water. The amplification program was 95 °C for 30 s, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. A melting curve analysis was performed to confirm amplification specificity. *GmActin* was used as the internal reference gene to normalize the relative expression levels of target genes. Relative expression levels were calculated using the 2^−ΔΔCt method. All assays included three biological replicates, with three technical replicates per biological replicate.

Plasmid construction and plant transformation

For overexpression analysis, the CDS of *GmRVE4d*^{H3} was amplified from a soybean accession carrying *GmRVE4d*^{H3} using KOD-One DNA Polymerase (TOYOBO Life Science, Osaka, Japan), with primers listed in Supplementary data 7. PCR products were separated by 1.0% agarose gel electrophoresis and purified using the V-ELUTE Gel Mini Purification Kit (ZOMANBIO, Beijing, China). The purified PCR product was ligated into the pGEM-T Easy vector (Promega Corporation, Madison, WI, USA) and verified by Sanger sequencing. Positive clones with confirmed inserts were then digested with the restriction endonuclease XbaI, and the target fragment was subcloned into the PTF101.1-*GmRVE4d*-GFP expression vector driven by the CaMV 35S promoter.

Target sequence adapters for *GmRVE4d* and *GmPRR5a* were designed using the CRISPR-P v2.0 web tool (http://crispr.hzau.edu.cn/cgi-bin/CRISPR2/CRISPR)[69]. Based on their genomic locations and GC contents, one optimal target sequence per gene was selected and integrated into a single guide DNA expression cassette, which was then cloned into the CRISPR-Cas9 binary vector. All recombinant plasmids were individually transformed into competent cells of *Agrobacterium tumefaciens* strain EHA105 and subsequently introduced into the soybean cultivar Jack using the cotyledon-node method[70].

Subcellular localization

For the subcellular localization assay, the recombinant plasmid PTF101.1-*GmRVE4d*-GFP was transformed into *Agrobacterium tumefaciens* strain EHA105 and transiently expressed in the leaves of 6-week-old *Nicotiana benthamiana* plants, as described previously[71]. Agrobacterium cells were resuspended in infiltration buffer containing 10 mM $MgCl_2$, 10 mM MES, pH 5.6, and 150 μM acetosyringone to a final OD_{600} of 0.6. The empty GFP vector was used as a control. Fluorescence images were acquired 48 h after agroinfiltration using a ZEISS LSM 700 laser scanning confocal microscope (Carl Zeiss, Oberkochen, Germany) with excitation and emission wavelengths of 488 nm and 509 nm, respectively, for GFP. The assay was repeated independently three times.

Western blot

Total proteins were extracted from young soybean leaves by grinding approximately 100 mg of fresh tissue in liquid nitrogen and homogenizing the powder in 500 μL of ice-cold extraction buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 0.1% NP-40, 3 mM DTT and 1 mM PMSF. Lysates were incubated on ice for 30 min and centrifuged at 12,000 × *g* for 15 min at 4 °C. Protein concentrations in the supernatants were determined using the Bradford assay.

For immunoblot analysis, 30 μg of total protein per sample was mixed with 5× protein loading buffer (ZOMANBIO, ZS306), boiled at 95 °C for 10 min, separated by 10% SDS-PAGE and transferred onto 0.45 μm nitrocellulose membranes. Membranes were blocked in TBST buffer containing 5% (w/v) non-fat milk for 1 h at room temperature. TBST contained 20 mM Tris-HCl, pH 7.5, 150 mM NaCl and 0.1% Tween-20. Membranes were then incubated overnight at 4 °C with anti-GFP antibody (TransGen Biotech, HT801; 1:2000). After washing three times with TBST for 10 min each, membranes were incubated with HRP-conjugated anti-mouse secondary antibody (TransGen Biotech, HS201; 1:5000) for 1 h at room temperature. Signals were detected using an ECL substrate (Tanon, 180-506) and a Tanon 5200 chemiluminescence imaging system.

After GFP detection, membranes were stripped with 0.2 M glycine-HCl, pH 2.5, for 1 h, re-blocked in TBST containing 5% non-fat milk for 1 h, and reprobed with anti-β-Actin antibody (TransGen Biotech, HC201; 1:5000) for 1 h at room temperature as the loading control. After washing three times with TBST for 10 min each, membranes were incubated with the same HRP-conjugated anti-mouse secondary antibody as described above for 1 h at room temperature. Protein molecular masses were estimated using the Solarbio Rainbow 180 Prestained Protein Marker. The immunoblot assay was independently repeated three times with similar results.

RNA-seq and data analysis

To compare the transcriptomic profiles among Jack, *GmRVE4d*-OE lines, and *Gmrve4d* CRISPR-Cas9 knockout mutants, soybean seedlings were grown in growth chambers under SD conditions for 13 days. On day 13 under SD conditions, fully expanded trifoliolate leaves were harvested at the end of the dark period and immediately frozen in liquid nitrogen. Three biological replicates were prepared for each genotype. For each biological replicate, trifoliolate leaves from three individual plants were pooled.

Total RNA was extracted using the RNAprep Pure Plant Kit (TIANGEN, Beijing, China) according to the manufacturer's instructions. RNA-seq libraries were prepared using an RNA-seq library preparation kit (Vazyme, Nanjing, China) following the manufacturer's protocol and sequenced on the Illumina NovaSeq 6000 platform. Raw sequencing reads were first assessed using FastQC v0.11.9, and adapter sequences and low-quality bases were removed using Trimmomatic v0.39. The resulting clean reads were aligned to the soybean reference genome *Glycine max* Wm82.a2.v1 using HISAT2 v2.2.1 with default parameters. Gene-level read counts were generated using

featureCounts v2.0.1 based on the corresponding Wm82.a2.v1 genome annotation. Differential expression analysis was performed in R using DESeq2 v1.34.0. Raw read counts were normalized using the median-of-ratios method implemented in DESeq2, and differential expression was assessed using the Wald test. *p* values were adjusted for multiple testing using the Benjamini-Hochberg procedure. Genes with a false discovery rate (FDR) < 0.05 and an absolute $\log_2$-transformed fold change ≥1 were defined as differentially expressed genes. Overlaps among differentially expressed gene sets were visualized using the VennDiagram package in R.

DAP-Seq profiling and analysis

DAP-seq experiments were performed following a published protocol[72]. Briefly, the coding sequence of *GmRVE4d* was fused in-frame with the HaloTag sequence, and the GmRVE4d-HaloTag fusion protein was expressed using an in vitro cell-free protein expression kit according to the manufacturer's instructions. After incubation at 25 °C for 2 h, the expression of the target fusion protein was verified by Western blotting.

Genomic DNA was extracted from leaves of Jack cultivar and used for DAP-seq library construction. DAP-seq libraries were prepared using the MICH TLX DNA-Seq Kit (MICH, Baoding, China) according to the manufacturer's instructions and the published DAP-seq protocol.

The GmRVE4d-HaloTag fusion protein was incubated with the constructed DNA library at 25 °C for 1 h. The GmRVE4d fusion protein was then immobilized using HaloTag beads (Vazyme, Nanjing, China), and unbound DNA fragments were removed by washing. GmRVE4d-bound DNA fragments were recovered and subjected to PCR amplification. Sequencing was performed using the Illumina NovaSeq 6000 platform. Reads were mapped to the soybean reference genome Wm82a2.v1 using BWA v0.7.19. Peaks were identified using MACS2 callpeak with a *q*-value cutoff of 0.05, and reproducible peaks from two biological replicates were merged using HOMER. Motif discovery was performed using MEME-ChIP, and bound peaks were annotated using ChIPseeker[73].

Yeast one-hybrid assay

To verify the direct binding of GmRVE4d to the *GmPRR5a* promoter, the CDSs of $GmRVE4d^{H1}$ and $GmRVE4d^{H3}$ were cloned separately into the pB42AD activation domain vector, whereas 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 competent cells of yeast strain EGY48 using the EX-Yeast Transformation Kit (ZOMANBIO, Beijing, China) according to the manufacturer's instructions. Transformed yeast cells were first cultured on SD/−Trp/−Ura selective medium at 30 °C 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. The empty pB42AD vector cotransformed with the *GmPRR5a* promoter reporter was used as the negative control. Each assay was repeated independently three times.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were performed using a chemiluminescent EMSA kit (Beyotime, Shanghai, China) according to the manufacturer's instructions. The MBP-GmRVE4d fusion protein was expressed in *Escherichia coli* BL21 (DE3) cells and purified by amylose resin affinity chromatography. Briefly, bacterial cultures were induced with 0.5 mM IPTG at 20 °C overnight when the OD_{600} reached 0.5–0.6. Cells were harvested and lysed by sonication in ice-cold column buffer containing 20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA and 1 mM DTT, pH 7.4, supplemented with PMSF, lysozyme and protease inhibitors. After centrifugation, the soluble fraction was loaded onto an amylose resin column, washed with column buffer, and eluted with column buffer containing 10 mM maltose according to the manufacturer's instructions. The purified MBP-GmRVE4d protein was verified by SDS-PAGE, aliquoted and stored at −80 °C until use. Biotin-labeled and unlabeled probes were synthesized by Sangon Biotech, and probe sequences are listed in Supplementary Data 7. Purified MBP-GmRVE4d protein was incubated with the biotin-labeled probe containing the putative binding region. For competition assays, a 50-fold excess of unlabeled wild-type probe was added before the labeled probe. An unlabeled mutated probe was used as a negative competitor to test binding specificity. The DNA-protein complexes were separated on a 4% native polyacrylamide gel in 0.5× TBE buffer at 4 °C and transferred to a nylon membrane. After UV crosslinking, signals were detected using streptavidin-HRP-mediated chemiluminescence.

Chromatin immunoprecipitation analysis

Soybean samples were collected from the third fully expanded trifoliolate leaves of 35Spro:*GmRVE4d*-GFP plants grown under SD conditions. Chromatin immunoprecipitation followed by quantitative PCR was performed as described previously[74]. Samples were immediately fixed in 1% formaldehyde on ice for 15 min under vacuum. Nuclei were isolated, and chromatin was sonicated to shear DNA to an average size of 250–500 bp. Soluble chromatin was immunoprecipitated with an antibody against the GFP epitope (Proteintech, 50430-2-AP), with IgG used as a negative control. The immunoprecipitated DNA was recovered and analyzed by qPCR with three technical replicates. ChIP−qPCR signals were first normalized to input DNA, and relative fold enrichment was calculated as the enrichment of anti-GFP ChIP over the IgG control. Three independent biological replicates were performed, each with three technical replicates for qPCR. Primers used for amplification are listed in Supplementary Data 7.

Dual-luciferase reporter assay

To construct the effector vector, the CDSs of $GmRVE4d^{H1}$ and $GmRVE4d^{H3}$ were cloned separately into the pGreenII 62-SK vector. The empty pGreenII 62-SK vector was used as the negative effector control. 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 carrying pSoup-P19. Agrobacterium cultures were resuspended in infiltration buffer containing 10 mM $MgCl_2$, 10 mM MES, pH 5.6, and 150 μM acetosyringone to a final OD_{600} of 0.6. Effector and reporter strains were mixed at a ratio of 1:1 and infiltrated into leaves of 6-week-old *Nicotiana benthamiana* plants. Leaf samples were collected 48 h after infiltration.

Firefly luciferase (LUC) and Renilla luciferase (REN) activities were quantified using the Dual-Glo Luciferase Reporter Gene Assay Kit (YEASEN Biotechnology, Shanghai, China) according to the manufacturer's instructions using a GloMax 20/20 luminometer. The relative activity of the *GmPRR5a* promoter was determined as the LUC/REN ratio. Three biological replicates were performed for each assay.

Statistical analysis

For phenotypic evaluation of the 257-accession panel used for GWAS, two biological replicates were evaluated for each accession in each environment, with five plants per replicate. For phenotypic evaluation of transgenic and mutant materials under different photoperiod conditions, 25 individual plants per genotype were analyzed in 2023, and 8 individual plants per genotype were analyzed in 2025. 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-group comparisons or two-

tailed Student's *t* tests for pairwise comparisons, as appropriate. All statistical analyses were conducted using SPSS software v.23 (IBM, Armonk, NY, USA), with $p < 0.05$ considered statistically significant. Details of statistical tests and significance levels are indicated in the relevant figure legends. Exact sample sizes are provided in the figure legends.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data generated in this study, including whole-genome sequencing, RNA-seq, and DAP-seq data, have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1114037. The *Glycine max* Wm82.a2.v1 reference genome was obtained from *Phytozome* (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1). All other data supporting the findings of this study are available within the article and its supplementary information files. Source data are provided with this paper.

References

1. Li, X., Liang, T. & Liu, H. How plants coordinate their development in response to light and temperature signals. *Plant Cell* **34**, 955–966 (2022).
2. Bu, T. et al. A critical role of the soybean evening complex in the control of photoperiod sensitivity and adaptation. *Proc. Natl. Acad. Sci. USA.* **118**, e2010241118 (2021).
3. Xu, X. et al. Circadian clock in plants: linking timing to fitness. *J. Integr. Plant Biol.* **64**, 792–811 (2022).
4. Zhang, L. et al. Principles and practices of the photo-thermal adaptability improvement in soybean. *J. Integr. Agric.* **19**, 295–310 (2020).
5. Zhu, Z. et al. Genomic atlas of 8,105 accessions reveals stepwise domestication, global dissemination, and improvement trajectories in soybean. *Cell* **188**, 6519–6535 (2025).
6. Jiang, X. et al. Population epigenomics reveals epigenetic drivers of replicated evolution and missing heritability in soybean. *Mol. Plant.* **19**, 1100–1116 (2026).
7. Garner, W. W. & Allard, H. A. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J. Agric. Res.* **18**, 553–606 (1920).
8. Yang, W. et al. Critical photoperiod measurement of soybean varieties in different maturity groups. *Crop Sci.* **59**, 2055–2061 (2019).
9. Wu, T. et al. Analysis of the independent- and interactive-photothermal effects on soybean flowering. *J. Integr. Agric.* **14**, 622–632 (2015).
10. Lin, X., Liu, B., Weller, J. L., Abe, J. & Kong, F. Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *J. Integr. Plant Biol.* **63**, 981–994 (2021).
11. Fang, C. et al. Subfunctionalisation and self-repression of duplicated *E1* homologues finetunes soybean flowering and adaptation. *Nat. Commun.* **15**, 6184 (2024).
12. Lu, S. et al. Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* **49**, 773–779 (2017).
13. Wu, T. et al. Molecular breeding for improvement of photothermal adaptability in soybean. *Mol. Breed.* **43**, 60 (2023).
14. Cao, D. et al. GmCOL1a and GmCOL1b function as flowering repressors in soybean under long-day conditions. *Plant Cell Physiol.* **56**, 2409–2422 (2015).
15. Su, C. et al. A circadian morning complex modulates far-red light signaling in *Arabidopsis*. *Nat. Commun.* **16**, 9398 (2025).
16. Li, J., Yang, M. K., Zeng, J., Chen, L. & Huang, W. Transcriptional activation and repression in the plant circadian clock: revisiting core oscillator feedback loops and output pathways. *Plant Commun.* **6**, 101415 (2025).
17. McClung, C. R. The plant circadian oscillator. *Biology* **8**, 14 (2019).
18. Liu, M. et al. Temporal orchestration of PRRs fine-tunes circadian pacing and anticipates environmental cues. *Cell Rep* **45**, 116787 (2026).
19. Rawat, R. et al. REVEILLE8 and pseudo-reponse regulator5 form a negative feedback loop within the Arabidopsis circadian clock. *PLoS Genet* **7**, e1001350 (2011).
20. Rosinski, J. A. & Atchley, W. R. Molecular evolution of the myb family of transcription factors: evidence for polyphyletic origin. *J. Mol. Evol.* **46**, 74–83 (1998).
21. Zhang, D. et al. Diverse roles of MYB transcription factors in plants. *J. Integr. Plant Biol.* **67**, 539–562 (2025).
22. Sorkin, M. L. et al. COR27/28 regulate the evening transcriptional activity of the RVE8-LNK1/2 circadian complex. Preprint at: https://doi.org/10.1101/2022.05.16.492168 (2022).
23. Hsu, P. Y., Devisetty, U. K. & Harmer, S. L. Accurate timekeeping is controlled by a cycling activator in *Arabidopsis*. *eLife* **2**, e00473 (2013).
24. Scandola, S. et al. Multi-omic analysis shows *REVEILLE* clock genes are involved in carbohydrate metabolism and proteasome function. *Plant Physiol* **190**, 1005–1023 (2022).
25. Gray, J. A., Shalit-Kaneh, A., Chu, D. N., Hsu, P. Y. & Harmer, S. L. The REVEILLE clock genes inhibit growth of juvenile and adult plants by control of cell size. *Plant Physiol* **173**, 2308–2322 (2017).
26. Dong, L. et al. Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* **12**, 5445 (2021).
27. Fang, X. et al. Modulation of evening complex activity enables north-to-south adaptation of soybean. *Sci. China Life Sci.* **64**, 179–195 (2021).
28. Yue, Y. et al. A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant* **10**, 656–658 (2017).
29. Zhang, Y. et al. *GmTOC1b* negatively regulates resistance to *Soybean mosaic virus*. *Crop J.* **11**, 1762–1773 (2023).
30. Wang, L. et al. *GIGANTEA* orthologs, *E2* members, redundantly determine photoperiodic flowering and yield in soybean. *J. Integr. Plant Biol.* **65**, 188–202 (2023).
31. Li, Y. et al. 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 (2019).
32. Li, C. et al. A domestication-associated gene *GmPRR3b* regulates circadian clock and flowering time in soybean. *Mol. Plant.* **13**, 745–759 (2020).
33. Lu, S. et al. Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* **52**, 428–436 (2020).
34. Wang, L. et al. 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 (2020).
35. Qin, C. et al. GmEID1 modulates light signaling through the Evening Complex to control flowering time and yield in soybean. *Proc. Natl. Acad. Sci. USA.* **120**, e2212468120 (2023).
36. Cheng, Q. et al. Stepwise selection of *Tof11/12* and *GmRVE1* facilitates soybean adaptation to high-latitude regions. *Cell Rep.* **45**, 117049 (2026).
37. Zhao, X. et al. A critical suppression feedback loop determines soybean photoperiod sensitivity. *Dev. Cell.* **59**, 1–14 (2024).
38. Mao, T. et al. Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. *BMC Genom.* **18**, 415 (2017).
39. Kidokoro, S. et al. Clock-regulated coactivators selectively control gene expression in response to different temperature stress

conditions in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA.* **120**, e2216183120 (2023).
40. Lu, S., Fang, C., Abe, J., Kong, F. & Liu, B. Current overview on the genetic basis of key genes involved in soybean domestication. *aBIOTECH* **3**, 126–139 (2022).
41. Han, T., Gai, J., Chen, F. & Qiu, J. Photoperiodic response and agronomic characters of soybean varieties with different maturity structures. *Acta Agronom. Sin.* **24**, 550–557 (1998).
42. Zhu, B., Sun, S., Han, T. & Wu, C. Comparison of maturity and its structure traits between wild and cultivated soybeans in China. *Soybean Sci* **31**, 894–898 (2012).
43. Gerhardt, A. & Mehta, D. Engineering the plant circadian clock for latitudinal adaptation as a strategy to secure agricultural productivity on a changing planet. *New Phytol.* **248**, 68–75 (2025).
44. Qi, X. et al. 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 (2021).
45. Wang, L., Kim, J. & Somers, D. E. Transcriptional corepressor TOPLESS complexes with pseudoresponse regulator proteins and histone deacetylases to regulate circadian transcription. *Proc. Natl. Acad. Sci. USA.* **110**, 761–766 (2013).
46. Wang, Y. et al. Light-and temperature-entrainable circadian clock in soybean development. *Plant Cell Environ.* **43**, 637–648 (2020).
47. Xie, Q. et al. LNK1 and LNK2 are transcriptional coactivators in the Arabidopsis circadian oscillator. *Plant Cell* **26**, 2843–2857 (2014).
48. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold3. *Nature* **630**, 493–500 (2024).
49. Fehr, W. R. & Caviness, C. E. *Stages of Soybean Development. Special Report 80*. 1–12 (Iowa State University, 1977).
50. Sun, Z., Tian, P. & Wang, J. Preliminary analysis of the maturity structure of soybean varieties in Northeast China. *Soybean Sci* **9**, 198–205 (1990).
51. Bolger, A. M., Lohse, B. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
55. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
56. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
57. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
58. Lee, T. H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genom.* **15**, 162 (2014).
59. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
60. Ligges, U. & Mächler, M. scatterplot3d: an R package for visualizing multivariate data. *J. Stat. Softw.* **8**, 1–20 (2003).
61. Zhang, C., Dong, S., Xu, J., He, W. & Yang, T. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
62. Huang, X. et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961–969 (2010).
63. Wang, J. & Zhang, Z. GAPIT Version 3: boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinform.* **19**, 629–640 (2021).
64. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. Preprint at: https://doi.org/10.1101/005165 (2014).
65. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
66. Kulmanov, M. & Hoehndorf, R. DeepGOPlus: improved protein function prediction from sequence. *Bioinformatics* **36**, 422–429 (2020).
67. Yang, M. et al. Deep learning-enabled discovery and characterization of *HKT* genes in *Spartina alterniflora*. *Plant J.* **116**, 690–705 (2023).
68. Ortiz, E. M. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. Preprint at: https://doi.org/10.5281/zenodo.2540861 (2019).
69. Liu, H. et al. CRISPR-P 2.0: an improved CRISPR-Cas9 tool for genome editing in plants. *Mol. Plant.* **10**, 530–532 (2017).
70. Chen, L. et al. Improvement of soybean *Agrobacterium*-mediated transformation efficiency by adding glutamine and asparagine into the culture media. *Int. J. Mol. Sci.* **19**, 3039 (2018).
71. Yang, Y., Li, R. & Qi, M. In vivo analysis of plant promoters and transcription factors by agroinfiltration of tobacco leaves. *Plant J* **22**, 543–551 (2000).
72. O'Malley, R. C. et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016).
73. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
74. Gendrel, A. V., Lippman, Z., Martienssen, R. & Colot, V. Profiling histone modification patterns in plants using genomic tiling microarrays. *Nat. Methods.* **2**, 213–218 (2005).

Author contributions

B.S. designed and performed the experiments and drafted the manuscript. T.W. collected natural populations. 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. Q.L., S.Y., M.N., C.Q., K.H., S.B.S., W.H. and C.W. critically revised the manuscript. T.H., S.S. and H.L. conceived and designed the study and critically revised the manuscript.

Funding

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).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-77316-3.

Correspondence and requests for materials should be addressed to Huihui Li, Shi Sun or Tianfu Han.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

[1]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, China. [2]National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya, China. [3]Heihe Branch of Heilongjiang Academy of Agricultural Sciences, Heihe, China. [4]These authors contributed equally: Baiquan Sun, Tingting Wu, Bingjun Jiang. ✉e-mail: lihuihui@caas.cn; sunshi@caas.cn; hantianfu@caas.cn