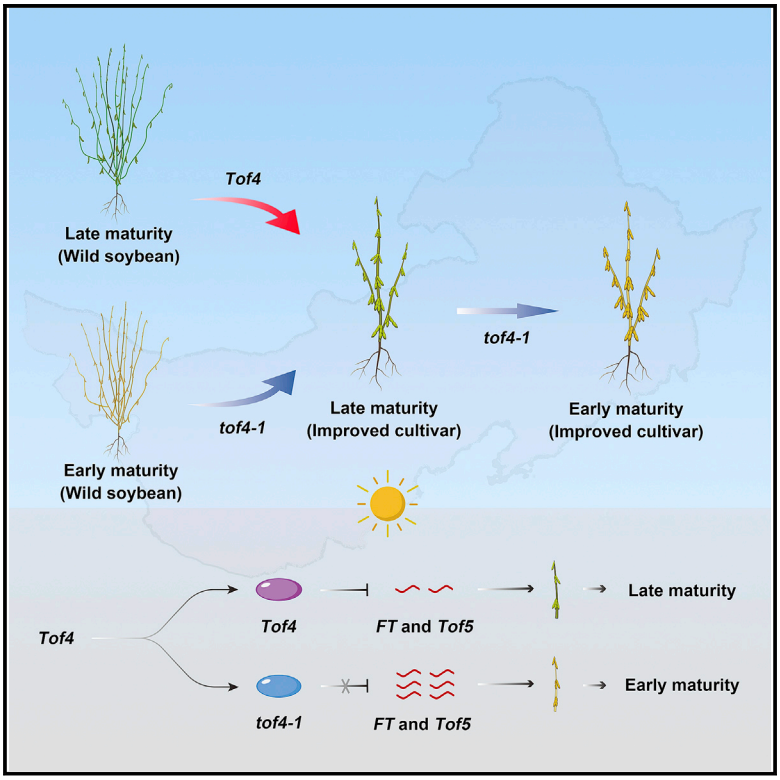
Article

Current Biology

The genetic basis of high-latitude adaptation in wild soybean

Graphical abstract



Highlights

- A new locus *Tof4* controlling flowering time in wild soybean was identified by GWAS
- Mutant alleles of *Tof4* facilitate the adaptation of wild soybean to high latitudes
- Tof4 directly binds to the promoters of *FT2a* and *FT5a* to control maturity in soybean
- *Tof4* and *Tof5*^{H2} form the genetic basis of wild soybean adaptation to high latitudes

Authors

Lidong Dong, Shichen Li, Lingshuang Wang, ..., Chao Fang, Fanjiang Kong, Baohui Liu

Correspondence

chengqun0118@gzhu.edu.cn (Q.C.), fangchao@gzhu.edu.cn (C.F.), kongfj@gzhu.edu.cn (F.K.), liubh@gzhu.edu.cn (B.L.)

In brief

Identification of genes or alleles from wild soybean to introgress into modern elite cultivars will improve soybean productivity. Dong et al. show that *Tof4* has undergone natural selection in wild soybean and facilitated adaptation to high latitudes. This study provides a new approach for precision breeding of early-maturity soybean varieties.

Dong et al., 2023, Current Biology *33*, 252–262
January 23, 2023
https://doi.org/10.1016/j.cub.2022.11.061

CellPress

CellPress
OPEN ACCESS

Current Biology

Article

The genetic basis of high-latitude adaptation in wild soybean

Lidong Dong,[1,6] Shichen Li,[1,6] Lingshuang Wang,[1,6] Tong Su,[1,6] Chunbao Zhang,[4] Yingdong Bi,[5] Yongcai Lai,[5] Lingping Kong,[1] Fan Wang,[1] Xinxin Pei,[1] Haiyang Li,[1] Zhihong Hou,[1] Haiping Du,[1] Hao Du,[1] Tai Li,[1] Qun Cheng,[1,*] Chao Fang,[1,*] Fanjiang Kong,[1,2,7,*] and Baohui Liu[1,3,*]

[1]Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou 510006, China
[2]College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China
[3]The Innovative Academy of Seed Design, Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, China
[4]Soybean Research Institute, National Engineering Research Center for Soybean, Jilin Academy of Agricultural Sciences, Changchun 130033, China
[5]Institute of Crops Tillage and Cultivation, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China
[6]These authors contributed equally
[7]Lead contact
*Correspondence: chengqun0118@gzhu.edu.cn (Q. C.), fangchao@gzhu.edu.cn (C. F.), kongfj@gzhu.edu.cn (F. K.), liubh@gzhu.edu.cn (B. L.)
https://doi.org/10.1016/j.cub.2022.11.061

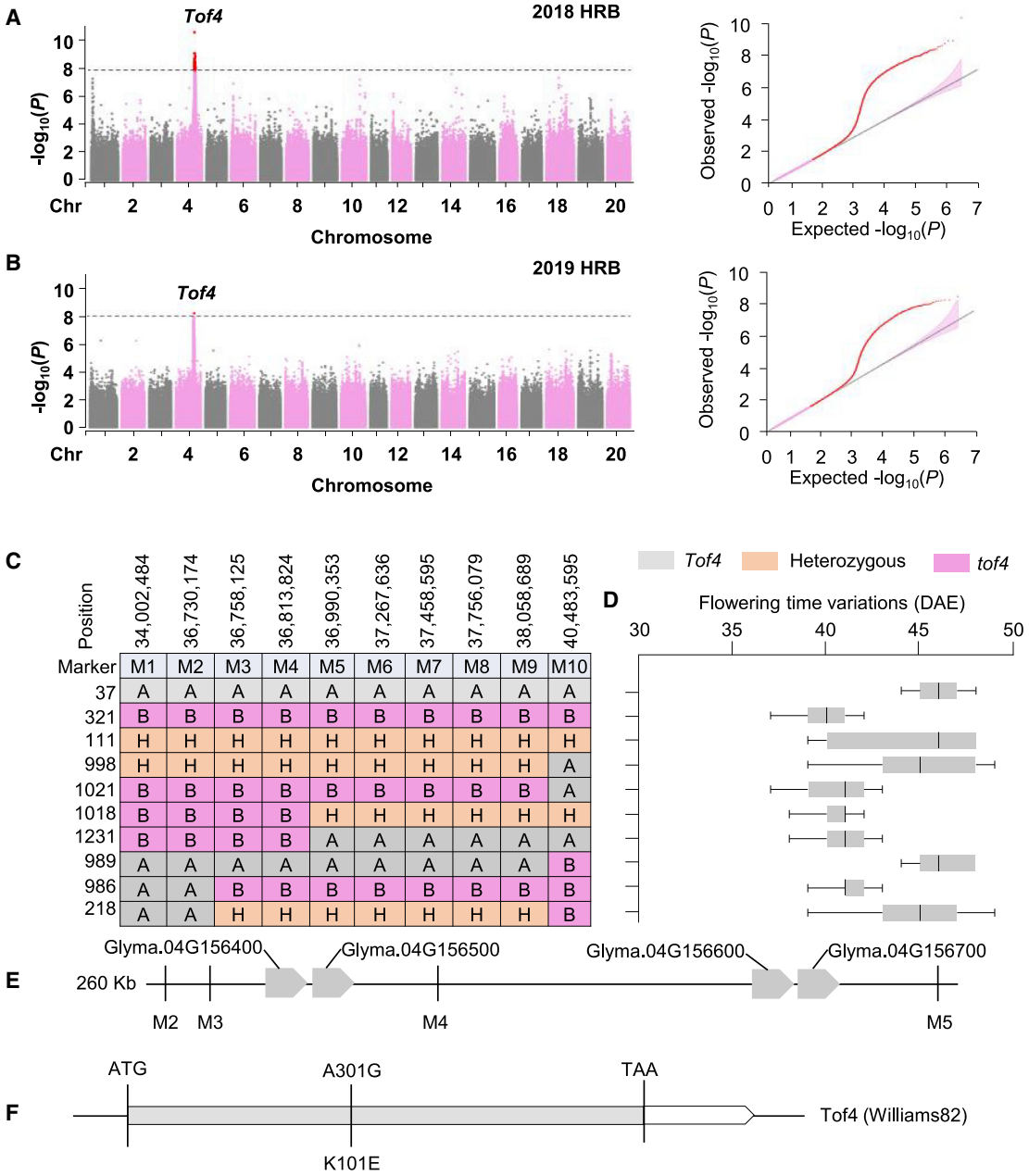
SUMMARY

In many plants, flowering time is influenced by daylength as an adaptive response. In soybean (*Glycine max*) cultivars, however, photoperiodic flowering reduces crop yield and quality in high-latitude regions. Understanding the genetic basis of wild soybean (*Glycine soja*) adaptation to high latitudes could aid breeding of improved cultivars. Here, we identify the *Tof4* (*Time of flowering 4*) locus, which encodes by an E1-like protein, E1La, that represses flowering and enhances adaptation to high latitudes in wild soybean. Moreover, we found that Tof4 physically associates with the promoters of two important *FLOWERING LOCUS T* (*FT2a* and *FT5a*) and with *Tof5* to inhibit their transcription under long photoperiods. The effect of *Tof4* on flowering and maturity is mediated by *FT2a* and *FT5a* proteins. Intriguingly, *Tof4* and the key flowering repressor *E1* independently but additively regulate flowering time, maturity, and grain yield in soybean. We determined that weak alleles of *Tof4* have undergone natural selection, facilitating adaptation to high latitudes in wild soybean. Notably, over 71.5% of wild soybean accessions harbor the mutated alleles of *Tof4* or a previously reported gain-of-function allele $Tof5^{H2}$, suggesting that these two loci are the genetic basis of wild soybean adaptation to high latitudes. Almost no cultivated soybean carries the mutated *tof4* allele. Introgression of the *tof4-1* and $Tof5^{H2}$ alleles into modern soybean or editing *E1* family genes thus represents promising avenues to obtain early-maturity soybean, thereby improving productivity in high latitudes.

INTRODUCTION

Soybean (*Glycine max* [L.] Merr.) is an important edible oil and protein feed crop.[4] In 2020, 100.31 million tons of soybean were imported into China, while the domestic soybean production was only 19.6 million tons. The Heilongjiang province accounts for 56.11% of China's total soybean production, mainly in the north, for example, in the cities of Heihe, Qiqihar, and Jiagedaqi at latitudes of over 48° N (National Bureau of Statistics, NBSC). Notably, soybean production has also experienced rapid growth in Russia, with the Amur region of the Far Eastern Federal District (48°51′–57°04′ N) providing nearly 30% of Russia's total production.[5] Despite being major soybean-producing and variety-breeding areas, these northern regions in both China and Russia usually have only short frost-free periods. Thus, early-maturity or super-early-maturity soybean varieties are needed to ensure natural maturity before early frost in these regions. However, few early-flowering and -maturity genes have been identified, limiting the soybean varieties suitable for planting in these high-latitude regions with short seasons.

Cultivated soybean was domesticated from its wild relative (*Glycine soja* [Sieb. and Zucc.]) in central China ~5,000 years ago.[6,7] It is a facultative short-day (SD) plant and is extremely sensitive to photoperiod.[7] Cultivated soybean gained the ability to flower under long-day (LD) conditions during the growing season at higher latitudes, mainly through the reduction of photoperiod sensitivity.[8] In recent decades, studies focused on the molecular pathways of soybean flowering and adaptation to high latitude have determined that a legume-specific *E1* gene-mediated photoperiodic network plays a crucial role in soybean.[9–12] The PHYTOCHROME A (PHYA) photoreceptors E3 (PHYA3) and E4 (PHYA2) receive light signals and convert them into oscillatory signals through the circadian clock, which directly or indirectly regulate the expression of *E1* and in turn release its transcriptional suppression of downstream genes, such as *FLOWERING LOCUS T 2a* (*FT2a*), *FT5a*, and *Time of*



G

Tof4 MSLQSQIPSKFKIIFISSNMSNPSDEKEQCQKKRKSTICEASNFKTSRRRFFSNKNEEDMNKGVSTTLKLYDD 73
tof4 MSLQSQIPSKFKIIFISSNMSNPSDEKEQCQKKRKSTICEASNFKTSRRRFFSNKNEEDMNKGVSTTLKLYDD 73

Tof4 PWKIKKTLTDSDLGILSRLSLAADLVKKQILPMLGADHARAAETEEGTPVRVWDMDTKSMHQLVLKRWSSSK 145
tof4 PWKIKKTLTDSDLGILSRLSLAADLVKEQILPMLGADHARAAETEEGTPVRVWDMDTKSMHQLVLKRWSSSK 145

Tof4 SYVLIGKWNQDFVRRRDLKKGDEIGFHWDPYNCVFNFCVLKRAMPEN* 192
tof4 SYVLIGKWNQDFVRRRDLKKGDEIGFHWDPYNCVFNFCVLKRAMPEN* 192

Figure 1. Identification of the *Tof4* gene by map-based cloning

(A and B) GWAS of flowering time (R1 stage) using data from 295 wild soybean accessions grown over the 2018 (A) and 2019 (B) field seasons in Harbin, China. (C) Delimitation of the *Tof4* locus to a 260-kb region of chromosome 4 in a segregating heterozygous inbred family (n = 2,210) from a cross between Williams82 (W82) and Dongnong50 (DN50). A, homozygous for the *Tof4* allele from W82; B, homozygous for the *tof4* allele from DN50; H, heterozygous.

(legend continued on next page)

flowering 5 (*Tof5*; a homolog of *FRUITFULL*), ultimately controlling the flowering time and maturity under LD conditions.[11–13] In addition, loss-of-function alleles of *E2*, *E1-like-b* (*E1Lb*), *Tof11*, and *Tof12* as well as gain-of-function alleles of *Tof5* facilitate early flowering and maturity and improve soybean adaptation to high latitude.[2,13–15] Allelic variation in *E1–E4* genes explains 62%–66% of the variation in soybean flowering, indicating that allelic combinations of *E1–E4* genes play major roles in determining soybean flowering time and thus adaptation zone.[16] The allelic combination $e1^{as}$/*e2*/*e3*/*E4* ($e1^{as}$ is a weak mutant allele of *E1*) plays a dominant role in southern Northeast China, and only cultivars carrying mutations at all four *E1–E4* loci are distributed in high-latitude (above 47° N) cold regions.[17,18] However, wild soybean is distributed across a broad geographical range (24°–53° N, 97°–143° E), which includes the northern Heilongjiang province in China and Far Eastern Federal District in Russia at latitudes over 48° N.[19,20] Therefore, we reasoned that elucidating the genetic basis of wild soybean's adaptation to high-latitude regions with short seasons could point us toward useful genetic resources that could be reintroduced into cultivated soybean via breeding.

In this study, we identified a novel locus, *Tof4*, that controls flowering time and latitude adaptation in wild soybean under LD conditions. Using a genome-wide association study (GWAS) and positional cloning of quantitative trait loci (QTLs), we show that *Tof4* accounts for a major portion of the natural variation in flowering and adaptation to high latitude in wild soybean. *Tof4* encodes a homolog of the legume-specific E1 protein (E1La) that represses flowering and maturity by directly regulating the transcription of *FT2a*, *FT5a*, and *Tof5* in soybean. Furthermore, we determined that weak alleles of *Tof4* have undergone natural selection, facilitating adaptation to high latitudes in wild soybean. Finally, we establish that *Tof4* inhibits flowering under LD conditions, independently of *E1*, and plays a major role in the control of flowering in soybean. By modulating combinations of *E1* homolog mutations or different *Tof4* and *E1* alleles, it should be possible to breed early-maturity and super-early-maturity soybean varieties suitable for high-latitude regions with short seasons. Our findings thus uncover the major genetic basis for the adaptation of wild soybean to high latitudes and provide a new approach for precise breeding to improve soybean crop adaptation to such regions.

RESULTS

Identification of the *Tof4* locus

To identify the QTLs that control flowering time in wild soybean, we first evaluated the flowering time of 295 wild soybean accessions, including 200 wild soybean accessions that were described in a previous study[1,3] and 95 wild soybean accessions newly collected and sequenced in this study (Data S1), grown under natural LD conditions in Harbin (45.75′ N). We conducted a GWAS using 4,893,018 single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) $\geq$ 0.05 and identified one consistent significantly associated locus on chromosome 4 (hereafter referred to as *Time of flowering 4* [*Tof4*]) controlling flowering time under natural LD conditions in both 2018 and 2019 (Figures 1A–1D). Consistent with the GWAS results, we previously identified a major QTL controlling flowering time and maturity on chromosome 4, using a biparental recombinant inbred line population derived from a cross between Dongnong50 (DN50, a cultivar introgressed from wild soybean) and Williams 82 (W82).[21] We concluded that *Tof4* denotes a major QTL regulating flowering time in wild soybean under LD conditions.

Tof4 encodes an E1La protein and represses soybean flowering

We generated a large (n = 2,210) inbred F_8 population from DN50 × W82 for fine mapping of the *Tof4* locus by recurrent selection for heterozygosity at *Tof4* from the F_2 to F_7 generations. Analysis of this population located *Tof4* within a 260-kilobase (kb) region harboring four genes (Figure 1E; Table S1), based on the W82 reference genome.[22] We cloned and sequenced all four of the predicted genes coding regions in the two parents; of these, only the sequence of the *E1La* gene (*Glyma.04G156400*) differed between the two parents (W82 and DN50). The early flowering parent DN50 harbored one SNP predicted to convert a lysine into a glutamate, in which the lysine is conserved in leguminous species (Figures 1F, 1G, and S1), indicating that this SNP may impair the functions of E1La. The presence of this mutation in DN50 suggested that the *E1La* gene was a strong candidate for the *Tof4* locus.

The examination of tissue-specific expression showed that *E1la* was highly expressed in leaves and almost not expressed in other tissues (Figure S2A). To examine the specific effect of the *Tof4* locus, we generated two F_8 near-isogenic lines (NILs) carrying either the functional *Tof4* allele (NIL-*Tof4*) or mutated *tof4* allele (NIL-*tof4-1*) and evaluated the flowering time of both NILs under natural LD conditions in Changchun, China. Compared with NIL-*Tof4*, NIL-*tof4-1* showed significantly earlier flowering time and maturity (Figures 2A–2C), along with decreased plant height, node number, and pod number (Figures S2B–S2D), as well as reduced grain yield (Figure 2D). These results support the notion that the *tof4-1* allele promotes flowering and maturity in soybean under LD conditions. Like the function of E1, E1La also acts as a flowering repressor. The loss of function of *E1* and *E1La* might have contributed to the adaptation of soybean to high latitude.

To validate whether *E1La* is associated with flowering time in soybean, we generated loss-of-function mutants of *E1La* (named $tof4^{CR}$) in the W82 background, using CRISPR-Cas9-mediated

(D) Segregation of flowering time is shown in boxplot format, where the interquartile region, median, and range are represented by the box, the bold vertical line, and the horizontal line, respectively (n = 20 plants). Each boxplot corresponds to the segregant on the same row in (C). For phenotypic investigations, the days from emergence to the first flowering (DAE), corresponding to the R1 stage, were scored.
(E) The 260-kb genomic region between markers M2 and M5 contains four predicted genes in the reference genome *Glycine max* Wm82.a2.v1.
(F) Allelic variation in the *Tof4* candidate gene *Glyma04G156400* in W82 and DN50. The top line is base change, and the bottom line is amino acid change.
(G) Sequence comparison of W82 and DN50. The red indicates the position of the single amino acid variation in DN50. An asterisk indicates the termination of translation. See also Figure S1, Table S1, and Data S1.

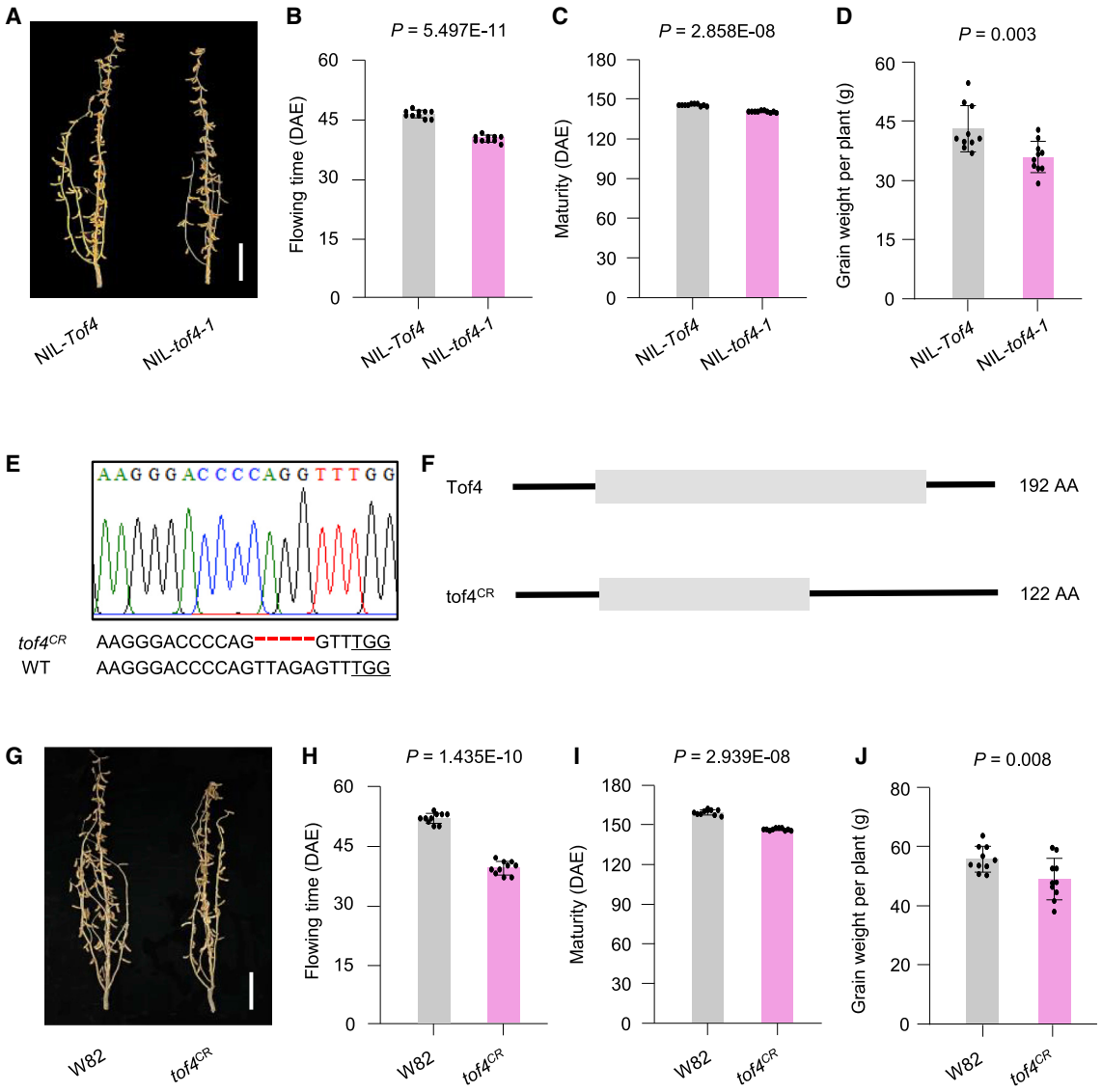


Figure 2. *Tof4* negatively regulates flowering time in soybean under long-day conditions

(A) Phenotypes of NIL-*Tof4* and NIL-*tof4-1* under natural long-day (LD) conditions in Changchun. Scale bars, 20 cm.
(B) Flowering time. DAE, days after emergence.
(C) Time to maturity.
(D) Grain yield per plant.
(E) Generation of *Tof4* mutants by CRISPR-Cas9 in W82 ($e1^{as}E3E4$, $e1^{as}$ is a weak mutant allele of *E1*) background.
(F) Schematic figure of Tof4 protein in $tof4^{CR}$ mutants.
(G) Phenotypes of $Tof4^{CR}$ and W82 under natural LD conditions in Changchun. Scale bars, 20 cm.
(H) Flowering time.
(I) Time to maturity.
(J) Grain yield per plant.
All data are shown as mean ± SD (n = 10 plants). One-tailed Student's t test was used to generate the p values. See also Figure S2.

gene editing (Figure 2E). The mutation led to a truncation of the E1La protein (Figure 2F). The $tof4^{CR}$ plants showed significantly earlier flowering time and maturity, altered yield-related traits, and reduced overall grain yield relative to wild-type W82 (Figures 2G–2J and S2E–S2G). We also genetically transformed W82 to overexpress the coding sequence of *Tof4* and characterized six independent T_3 homozygous overexpression (*Tof4-OE*) transgenic lines (Figure S2H). The *Tof4-OE* transgenic lines showed significantly later flowering and maturity than wild-type W82 and thus an improved grain yield under natural LD conditions in Shijiazhuang, China (Figures S2I–S2L). These observations establish *E1La* as the candidate for the *Tof4* locus and indicate that the natural mutations in DN50 cause *Tof4* flowering.

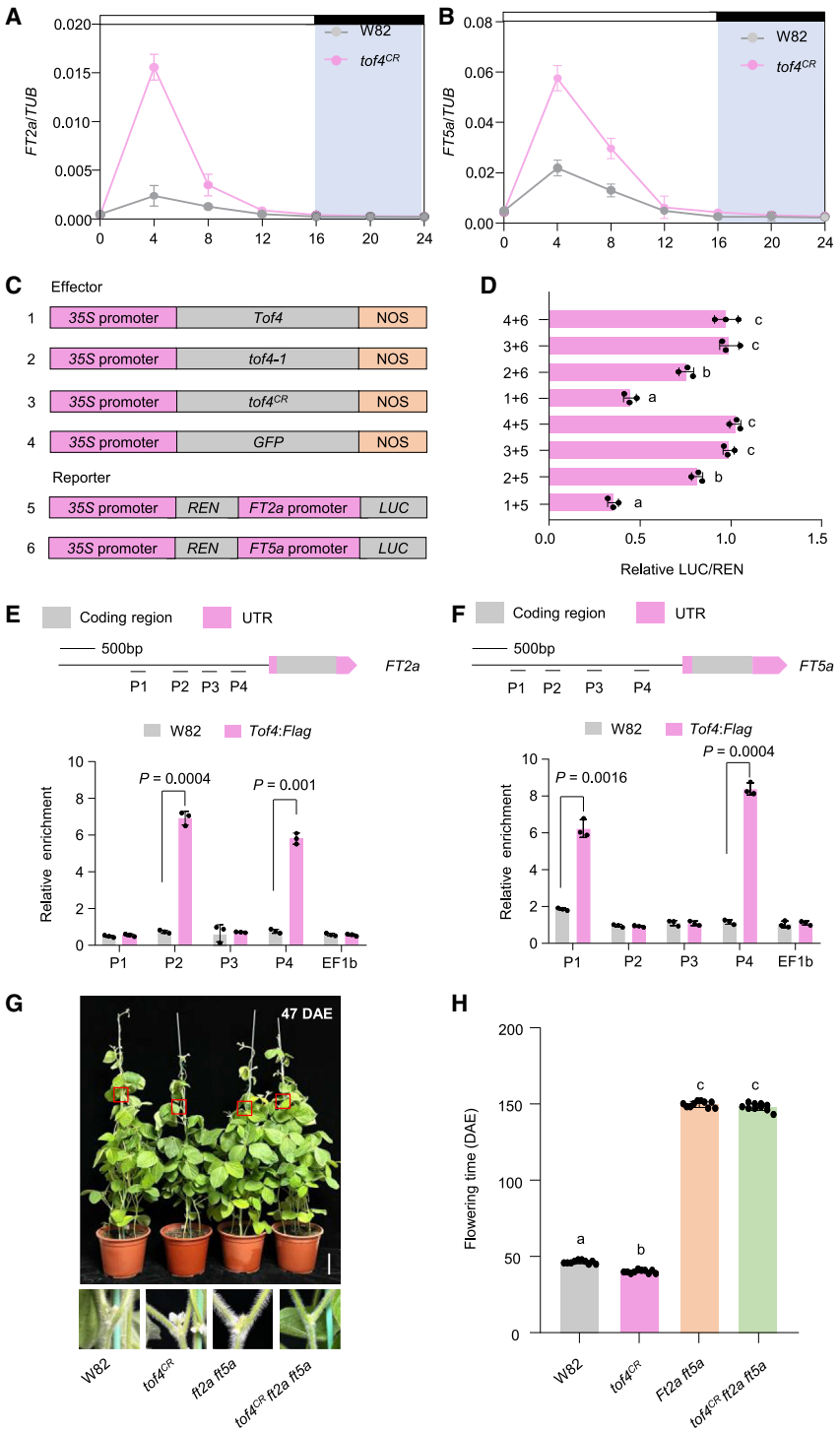


Figure 3. Tof4 directly represses the expression of *FT2a* and *FT5a*

(A and B) Diurnal variation in transcript levels of *FT2a* (A) and *FT5a* (B) in Williams82 and *tof4*^{CR} plants under LD conditions (16 h light/8 h dark) in growth chamber. All data are given as mean ± SD (n = 3 repeats).

(C) Constructs used for the transient transfection assay.

(D) Luciferase activity under the control of the *FT2a* or *FT5a* promoter, with the results from three independent replicates shown; the value for each replicate is represented by a dot. The combinations in (D) link to the numbers in (C). A Student's t test was used to generate the p values.

(E) ChIP-qPCR results demonstrating the direct binding of Tof4 to the *FT2a* promoter. *EF1b* as negative control.

(F) ChIP-qPCR results demonstrating the direct binding of Tof4 to the *FT5a* promoter; the value for each independent replicate is represented by a dot. A Student's t test was used to generate the p values.

(G) Phenotypes of difference allelic combinations at *Tof4* and *FT2a FT5a* at 47 days after emergence under LD conditions in growth chamber. Red box, magnified view. Flower buds were seen more clearly in the small photos. Scale bars, 10 cm.

(H) Flowering time of different allelic combinations at *Tof4* and *FT2a FT5a* under LD conditions (16 h light/8 h dark) in growth chamber.

All data are shown as mean ± SD (n = 10 plants); the value of each plant is represented by a dot. Different lowercase letters above the histogram bars in (H) denote significant differences across the two panels ($p > 0.05$). One-way ANOVA was used to generate the p values. See also Figure S3.

Tof4 directly represses the expression of *FT2a*, *FT5a*, and *Tof5*

Two florigen gene homologs, *FT2a* and *FT5a*, coordinately integrate photoperiod responses to control soybean flowering and adaptation.[23,24] To gain more insight into the molecular mechanism underlying the regulation of flowering and maturity by *Tof4*, we investigated the expression of these genes in *tof4*^{CR} and NILs of *Tof4* under LD conditions (16 h light/8 h dark, 16L/8D) in growth chamber. *FT2a* and *FT5a* were expressed at higher levels in the *tof4*^{CR} mutant and NIL-*tof4-1*, compared with those in wild-type W82 and NIL-*Tof4*, respectively (Figures 3A, 3B, S3A, and S3B). These results suggest that *Tof4* acts upstream of *FT2a* and *FT5a* in the soybean flowering pathway. To further explore the molecular nature of the relationship between *Tof4* and *FT2a*/*FT5a*, we performed a dual-luciferase transient expression and chromatin immunoprecipitation (ChIP)-qPCR assay. Tof4 proteins could directly associate with the promoters of *FT2a* and *FT5a* (Figures 3E and 3F) and significantly inhibited the transcription of *FT2a* and *FT5a* (Figures 3C and 3D). To further investigate the genetic relationship between *Tof4* and *FT*, we developed a *tof4*^{CR} *ft2a ft5a* triple mutant in the W82 background by crossing the *tof4*^{CR} mutant and a *ft2a ft5a* double mutant[24] and performed a phenotypic analysis under LD conditions in growth chamber. The *tof4*^{CR} mutant flowered significantly earlier than wild-type W82; however, there was no difference in flowering time between the *ft2a ft5a* double mutant and the *tof4*^{CR} *ft2a ft5a* triple mutant, suggesting that *Tof4* genetically depends on

FT2a and *FT5a* to control flowering in soybean (Figures 3G and 3H). Collectively, these results indicate that Tof4 directly represses *FT* transcription (*FT2a* and *FT5a*) to control flowering and maturity in soybean.

Our previous report confirmed that E1 inhibited the transcription of *Tof5* by associating with its promoters.[1] To determine whether Tof4 also regulates *Tof5* expression, we performed reverse-transcriptase quantitative PCR (RT-qPCR) assays to determine the transcription levels of *Tof5* in the $tof4^{CR}$ mutant and in NILs of *Tof4*. *Tof5* was expressed more highly in the $tof4^{CR}$ mutant and NIL-*tof4-1*, compared with that in wild-type W82 and NIL-*Tof4*, respectively (Figures S3C and S3D). Consistent with the finding that E1 binds to the promoter of *Tof5*, we also determined that Tof4 could bind to the promoter of *Tof5* near the adjacent fragment sites P2 and P3 by ChIP-qPCR assay (Figure S3E), and significantly repressed the transcription of *Tof5* (Figures S3F and S3G). These results indicate that Tof4 can directly bind to the promoter of *Tof5* and inhibit its expression. Therefore, Tof4 proteins directly associate with the promotors of flowering enhancers *FT* and *FUL* to suppress their transcription, thus delaying soybean flowering and maturity.

Loss-of-function alleles of *Tof4* have the potential for breeding early-maturing soybean varieties

To explore the genetic relationship between *Tof4*/*E1La* and *E1*, we developed a $tof4^{CR}$ $e1^{CR}$ double mutant by crossing the $tof4^{CR}$ mutant and the $e1^{CR}$ mutant that were described in a previous study.[1] We subjected the mutants to phenotypic evaluation in Harbin (45.75′ N) and Changchun (43.88′ N) under natural LD conditions. In Changchun, the $tof4^{CR}$ $e1^{CR}$ double mutant flowered earlier than the $e1^{CR}$ mutant and wild-type W82 and had altered yield-related traits (Figures 4A–4D and S4D–S4F) However, in Harbin, wild-type W82 and the $e1^{CR}$ mutant failed to mature, and it was difficult to harvest their seeds naturally until the end of the growing season, whereas the $tof4^{CR}$ $e1^{CR}$ double mutant showed earlier flowering and maturity, and high-quality seeds were harvested (Figures 4E–4H and S4G–S4I). The soybean genome contains two *E1* homologs (*E1La* and *E1Lb*), and their amino acid sequences maintain high similarity.[25] We generated a $tof4^{CR}$ $e1^{CR}$ $e1lb^{CR}$ triple mutant using CRISPR-Cas9 technology and performed phenotypic analysis (Figures S4A–S4C). The $tof4^{CR}$ $e1^{CR}$ $e1lb^{CR}$ triple mutant showed super-early flowering and maturity under natural LD conditions in two high-latitude locations, Changchun and Harbin (Figures 4A–4H and S4D–S4I). We tested the expression levels of *FT2a*, *FT5a*, and *Tof5* in these mutants and determined that the transcription of *FT2a*, *FT5a*, and *Tof5* correlated with flowering, maturity, and yield under LD conditions (Figures 4I, 4J, and S4J). Together, these results suggest that loss-of-function alleles of *Tof4* promote soybean adaptation to high latitudes and that *E1La* and *E1* independently but additively regulate flowering time, maturity, and grain yield in soybean.

Geographical distribution of *Tof4* alleles in wild soybean

To explore the evolutionary origin of different *Tof4* alleles, we examined variations of the *Tof4* coding sequence in 2,387 resequenced accessions, including 2,292 previously described accessions[13,2,1,3] and 95 wild soybean accessions used in this study. We identified five unique high-confidence haplotypes for *Tof4* (Figure S5A; Data S2). Median-joining network analysis indicated that $Tof4^{H3}$ originated from $Tof4^{H2}$ (*tof4-1*) and $Tof4^{H5}$ originated from $Tof4^{H4}$ (Figure S5B). To further examine the functional significance of the different alleles of *Tof4*, we performed a dual-luciferase transient expression assay in *Arabidopsis* (*Arabidopsis thaliana*) protoplasts. The results demonstrate that the $Tof4^{H2}$ (*tof4-1*) and $Tof4^{H3}$ (named *tof4-2*) alleles partially impaired the inhibitory ability of *Tof4* in the expression of *FT5a*, but the $Tof4^{H4}$ and $Tof4^{H5}$ alleles had no effects (Figures S5C and S5D). We generated a second null functional *tof4* mutant (named $tof4^{CR}$*-D*) in the DN50 (*tof4-1* allele) background (Figures S5E and S5F). Phenotypic analysis showed that the $tof4^{CR}$*-D* mutant flowered significantly earlier than wild-type DN50 under LD conditions in growth chamber (Figures 5A and 5B). We also performed a population genetic association analysis of flowering time in the 295 wild soybean accessions, and the results indicated that wild soybean carrying the *tof4-1* allele showed earlier flowering than wild accessions carrying the *Tof4* allele (Figures 5C and 5D), suggesting that *tof4-1* is a weak allele.

To further investigate the evolutionary history of *Tof4*, we analyzed the percentages of different alleles in accessions of cultivated soybean, landraces, and wild soybean in our panel of 2,387 resequenced accessions. Weak alleles of *Tof4* (*tof4-1* and *tof4-2*) were present in 32.9% of wild soybean accessions; however, only 0.35% (3/857) of cultivated soybean accessions contained the *tof4-1* allele, indicating that the mutated *tof4* alleles observed in wild soybean are rarely present in cultivated soybean (Figure 5E). However, three cultivated soybean accessions that contained the mutated *tof4-1* allele may be derived from natural or artificial introgressions of wild soybean (Figure 5E).

We examined the geographical distributions of different *Tof4* alleles in 441 wild soybean accessions from China. The frequency of *tof4* alleles was 61.6%, with those of *tof4-1* and *tof4-2* being 58.8% and 2.8%, respectively, in the northeast regions of China (NE), whereas *tof4* alleles were absent in the wild soybean accessions from medium- or low-latitude regions (Figure 5F). Nucleotide diversity analysis showed that mutated *tof4* alleles exhibited lower nucleotide diversity than *Tof4* allele in the NE (Figure 5G). In addition, wild soybean accessions carrying *tof4* alleles were found at a higher mean latitude than accessions carrying *Tof4* (Figure 5H). These results indicate that the *tof4-1* allele was the major genetic variant in wild soybean to adapt to high latitudes. Our previous research showed that gain-of-function alleles of *Tof5* ($Tof5^{H1}$ and $Tof5^{H2}$) and loss-of-function alleles of *E3* also contributed to the adaptation of wild soybean to high-latitude regions.[1] We therefore examined the geographical distribution of different *Tof5* and *E3* alleles in the 441 wild accessions. The frequencies of the $Tof5^{H2}$ and $Tof5^{H1}$ alleles were 15% and 18%, respectively, but that of the *e3* allele was only 4% in NE; the $Tof5^{H1}$ and *e3* alleles also arose in other regions of China (Figure 5F). Notably, more than 71.5% of accessions harbored the mutated alleles of *Tof4* or gain-of-function alleles of *Tof5* (Data S3), suggesting that these genes are the major genetic basis of adaptation to high latitude in wild soybean.

Based on our data, we propose a model describing how the natural selection of *Tof4* and *Tof5* facilitated the adaptation of wild soybean to high latitudes (Figure 5I). Tof4 represses *FT* and *Tof5* expression, and Tof4 physically associates with the

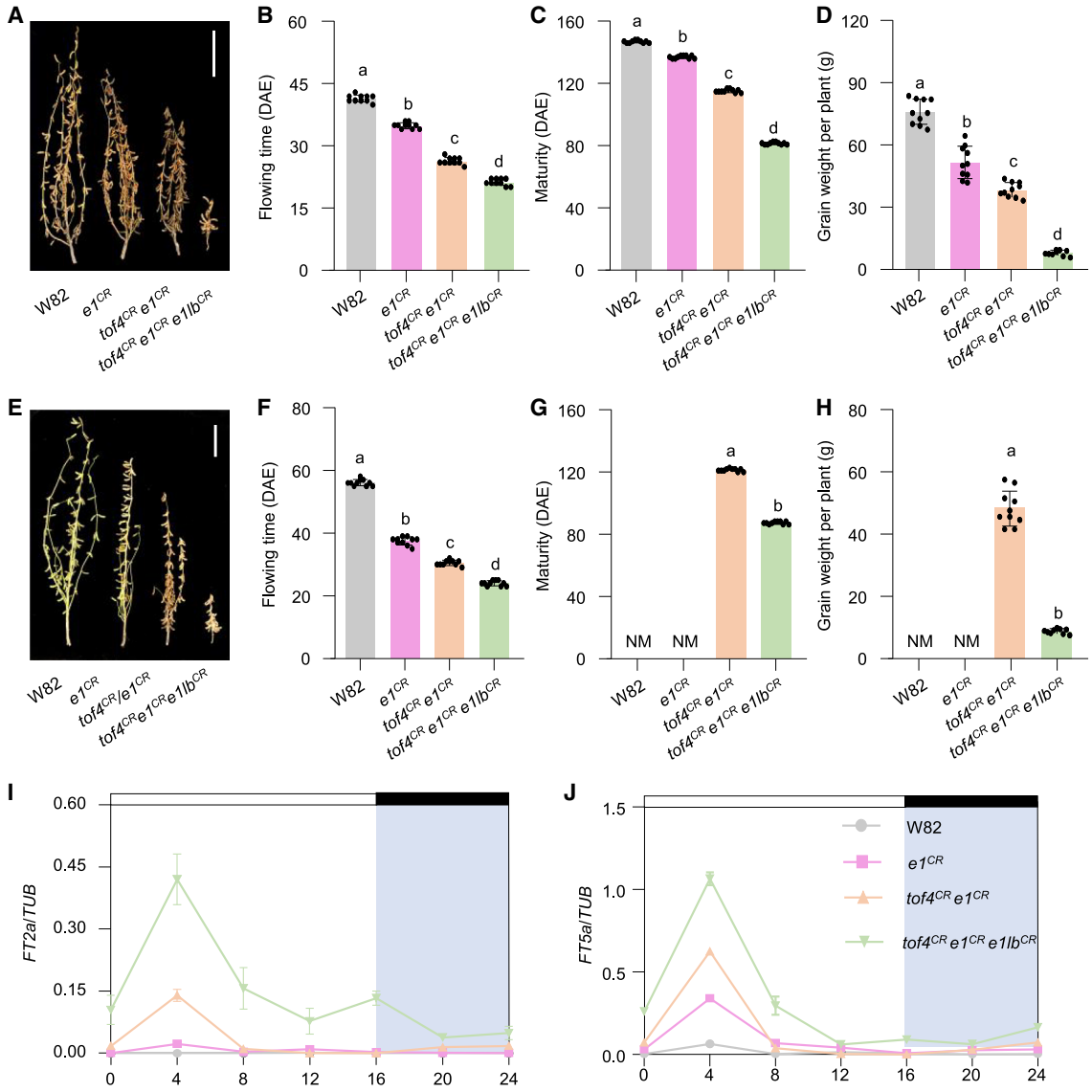


Figure 4. Tof4 and E1 additively control flowering time and latitude adaptation

(A) Phenotypes of Williams82 (W82), *e1*^{CR}, *tof4*^{C} *e1*^{CR}, and *tof4*^{CR} *e1*^{CR} *e1lb*^{CR} plants under natural LD conditions in Changchun. Scale bars, 20 cm.

(B) Flowering time.

(C) Time to maturity.

(D) Grain yield per plant.

(E) Phenotypes of W82, *e1*^{CR}, *tof4*^{CR} *e1*^{CR}, and *tof4*^{CR} *e1*^{CR} *e1lb*^{CR} plants under natural LD conditions in Harbin. Scale bars, 20 cm.

(F) Flowering time.

(G) Time to maturity.

(H) Grain yield per plant.

All data are shown as mean ± SD (n = 10 plants); the value of each plant is represented by a dot. Different lowercase letters above the histogram bars in (B)–(D) and (F)–(H) denote significant differences across the two panels ($p > 0.05$). One-way ANOVA was used to generate the p values.

(I) Diurnal variation in transcript levels of *FT2a* in W82, *e1*^{CR}, *tof4*^{CR} *e1*^{CR}, and *tof4*^{CR} *e1*^{CR} *e1lb*^{CR} plants under LD conditions (16 h light/8 h dark) in growth chamber.

(J) Diurnal variation in transcript levels of *FT5a* in W82, *e1*^{CR}, *tof4*^{CR} *e1*^{CR}, and *tof4*^{CR} *e1*^{CR} *e1lb*^{CR} plants under LD conditions (16 h light/8 h dark) in growth chamber.

All data are given as mean ± SD (n = 3 repeats). See also Figure S4.

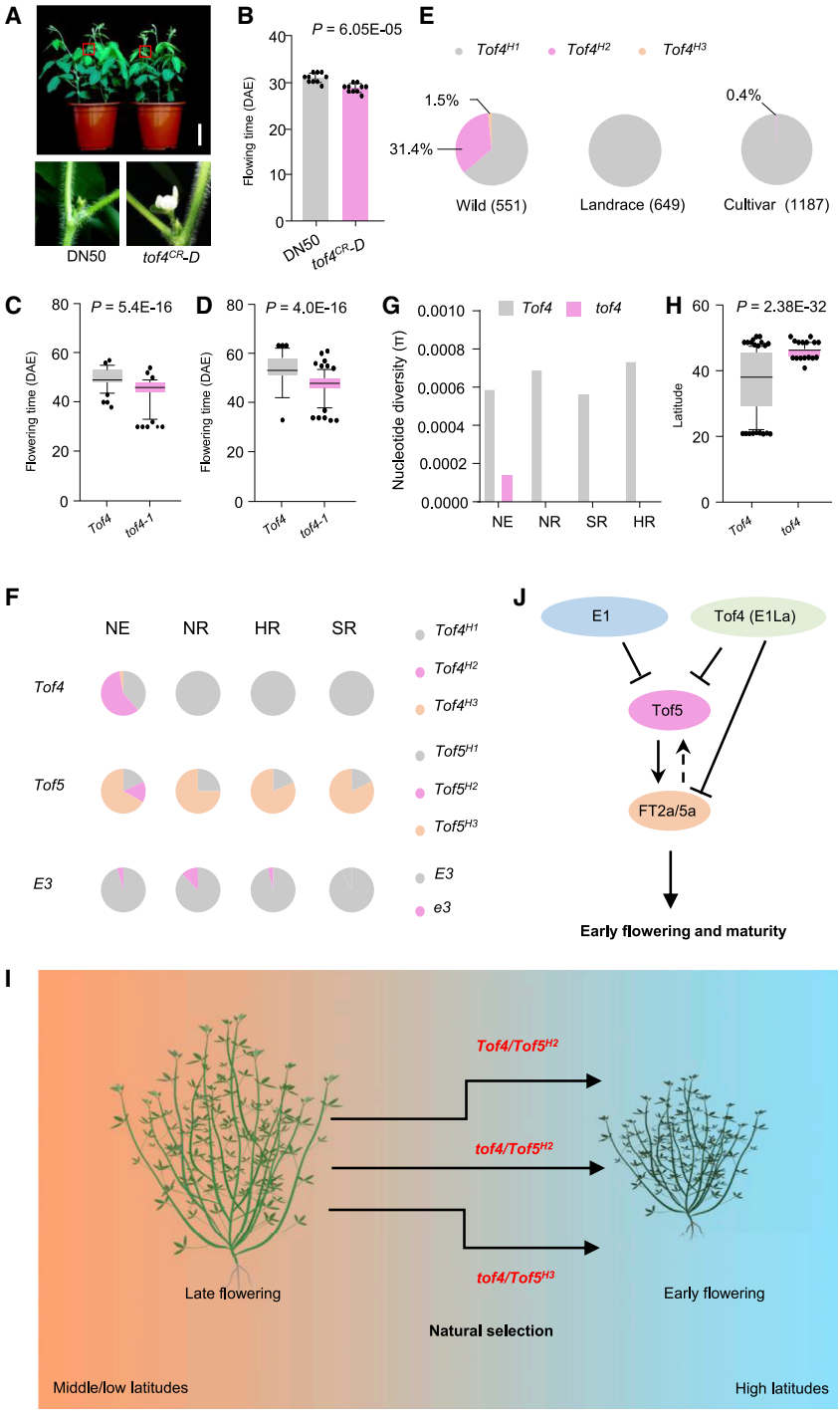


Figure 5. Geographical distribution of genetic diversity of *Tof4*

(A) Phenotypes of Dongnong50 (DN50) and *tof4*^{CR}-*D* plants under LD conditions (16 h light/8 h dark) in growth chamber. Red box, magnified view. Scale bars, 10 cm. In *tof4*^{CR}-D, D indicates that it is CRISPR in DN50.

(B) Flowering time of DN50 and *tof4*^{CR}-*D* plants under LD conditions (16 h light/8 h dark) in growth chamber.

(C) Flowering time variations in 379 wild soybean accessions carrying *Tof4* and *tof4-1* alleles under natural LD conditions in 2018 in Harbin.

(D) Flowering time variations in 379 wild soybean accessions carrying *Tof4* and *tof4-1* alleles under natural LD conditions in 2019 in Harbin. For box-and-whisker plots, the central line, box, and whiskers indicate the median, interquartile range (IQR), and 1.5 times the IQR, respectively. One-tailed Student's t test was used to generate the p values.

(E) Proportions of *Tof4* alleles and their co-occurrence with each of the three germplasm groups.

(F) Latitudinal distribution of all wild Chinese soybean accessions possessing different alleles at *Tof4*. NE, northeast region of China; NR, north region of China; HR, Huanghuai region of China; SR, south region of China.

(G) Nucleotide diversity analysis of the region surrounding *Tof4* in different region of China.

(H) Latitudinal distribution of all wild Chinese soybean accession possessing different alleles at *Tof4*. For box-and-whisker plots, the central line, box, and whiskers indicate the median, IQR, and 1.5 times the IQR, respectively. A Student's t test was used to generate the p values.

(I) A proposed model illustrating natural selection of *Tof4* and *Tof5*.

(J) Molecular regulatory mechanism of *Tof4* in soybean flowering and adaptation. See also Figure S5 and Data S2 and S3.

promoters of the two key soybean *FT* homologs and *Tof5* to promote their expression, ultimately resulting in the promotion of flowering and earlier maturity (Figure 5J).

DISCUSSION

Determining the genetic basis of wild soybean adaptation to high latitudes will help us identify valuable genes or alleles from wild soybean that can be introgressed into modern elite cultivars to improve soybean productivity. In the current study, we identified the novel locus *Tof4*, which encodes the E1La protein, and determined that mutated *tof4* alleles were targeted by natural selection and contributed to earlier flowering and maturity and geographical adaptation to high-latitude regions in wild soybean. We confirmed that *Tof4* depends genetically on two key soybean *FT* homologs (*FT2a* and *FT5a*) and that Tof4 directly binds to the promoters of *FT* and *Tof5* to repress their transcriptions. *Tof4* inhibits flowering and maturity independently of *E1* and plays a major role in the control of flowering in soybean under LD conditions. The weak alleles of *Tof4* that were identified in this study provide valuable genetic resources for molecular breeding to improve soybean adaptation and productivity at high latitudes.

Crop domestication is the process during which wild species are converted into crop plants through artificial selection.[26] Due to artificial selection and population/genetic bottlenecks,

modern crops have much lower genetic diversity than their wild relatives.[27–29] This reduced variation may have led to the loss of some genes or alleles that are important for adapting to different environments. Natural variations from wild ancestors can be introduced into modern crops via breeding to improve plant adaptation and yield. For instance, the *Upright plant architecture2* (*UPA2*) allele from teosinte was lost during maize domestication, and re-introducing the wild *UPA2* allele into modern maize hybrids facilitated high-density planting and increased yields.[30] Our previous report established that the $Tof5^{H2}$ allele leads to earlier flowering and maturity than the $Tof5^{H3}$ allele in wild soybean, representing a good genetic resource for improving cultivated soybean adaptation to high latitudes.[1] Here, we show that mutated *Tof4* alleles from wild soybean promoted earlier flowering and maturity and facilitated wild soybean adaptation to high-latitude short-season regions. The weak alleles of *Tof4* mainly occurred in wild soybean accessions from the NE regions of China but rarely (3/1,187) occurred in cultivated soybeans. Because cultivated soybean was domesticated from its wild relative *Glycine soja* in temperate regions (32° N and 40° N) of China ~5,000 years ago,[6,7] the wild soybean from the NE regions of China has not undergone the process of domestication; therefore, mutated *Tof4* alleles have not been used in modern breeding of cultivated soybean. Plants with loss-of-function alleles of both *E1* and *Tof4* exhibited very early maturity and matured normally in high-latitude regions with short frost-free periods. RNAi suppression of *E1* and its paralogs resulted in a near-complete loss of photoperiod sensitivity and was sufficient to convert an extremely late-flowering maturity group (MG) VIII cultivar to MG 000.[31] Here, we also determined that the triple mutant of *E1* family genes confers super-early maturity and adaptation to high-latitude short-season regions. Therefore, introducing mutated *tof4* alleles from wild soybean into modern cultivated soybean can improve cultivated soybean adaptation and could be used to breed earlier-maturity and high-yield soybean.

The adaptation of soybean to high latitudes was facilitated by the presence of natural loss-of-function alleles of six flowering suppressors—*E1*, *E2* (homolog of *GIGANTEA*), *E3*, *E4*, and two *PSEUDO-RESPONSE-REGULATOR3* homologs, *Tof11* and *Tof12*—and the gain-of-function allele of flowering activator *Tof5*.[9,11,14,2,1,32] Although these genes controlling flowering and maturity have been identified in various cultivated soybean accessions, the genetic basis of wild soybean adaptation to high latitudes is still unclear. Our previous study report that *Tof5* contributes to the wild soybean adaptation to high-latitude regions. The frequency of the $Tof5^{H2}$ allele in the northeast regions of China was only 18.8% in the 257 wild soybean accessions from China. In addition, natural variation in *E1La* may be present in a high frequency in wild soybean accessions adapted to high latitudes, and it has likely been used only rarely in North American cultivar development.[33] However, the genetic basis of wild soybean (*G. soja*) adaptation to high-latitude cold regions remains largely unclear. In this study, we collected 379 wild soybeans to perform a population genetic association analysis of flowering time and observed that more than 61.6% of wild soybean accessions harbor mutations of *tof4* alleles and about 15% of wild soybean accessions harbor $Tof5^{H2}$. Moreover, more than 71.5% of wild soybean accessions in high-latitude regions harbor *tof4* mutant alleles or *Tof5* gain-of-function alleles. Together, these results suggest that mutated alleles of *Tof4* and gain-of-function $Tof5^{H2}$ alleles are the major genetic basis of wild soybean adaptation to high latitudes. The identification of novel genes involved in the geographical adaptation of wild soybean will improve our understanding of flowering mechanisms in wild soybean and facilitate the improvement of cultivated soybean.

In the LD model plant *Arabidopsis*, CONSTANS (CO), a transcriptional activator of *FT*,[34–36] plays a key role in the regulation of photoperiodic flowering.[37–39] Transcriptional and post-transcriptional up-regulation of *CO* results in accumulation of the CO protein in the late afternoon under LD conditions, which in turn activates *FT* expression. The SD model plant rice *CO* ortholog *HEADING DATE1* (*HD1*) acts as an activator of *HD3A* expression under SD conditions but as a suppressor under LD conditions.[40–42] Unlike *Arabidopsis* and rice, there is a *PHYTOCHROME A*-mediated transcriptional induction of legume-specific *E1* by light that plays a critical role in photoperiodic induction of flowering in soybean.[11,43] The maturity gene *E1*, which has the largest effect on flowering in soybean and is a B3-related transcription factor, suppresses the expression of *FT2a* and *FT5a*.[9,16,44,45] *E1* has two homologs, *E1La* and *E1Lb*, which share high nucleotide and amino acid identities and exhibit similar expression patterns to *E1*.[25] Virus-induced silencing of *E1La* and *E1Lb* up-regulates the expression of *FT2a* and *FT5a* and leads to early flowering.[25] A loss-of-function allele of *E1Lb* identified in far-eastern Russian soybean cultivars promotes flowering by up-regulating the expression of soybean *FT2a* and *FT5a* under LD conditions.[15] Despite these effects, we do not know how E1La regulates the expression of *FT2a* and *FT5a* to control soybean flowering under LD conditions. Here, we generated stable genetic soybean plants and sufficient biochemical evidence to conclude that Tof4 (E1La) directly binds to the promoters of *FT2a* and *FT5a* and represses their expression, and *Tof4* depends genetically on *FT2a* and *FT5a* to control soybean flowering under LD conditions. In addition, we demonstrate that Tof4 also directly regulates the transcription level of *Tof5*. Therefore, the roles of *Tof4* in controlling flowering are integrated into the classical *E3/E4-E1-FT* pathway. *E1* and its homologs (*E1La* and *E1Lb*) are the critical regulators in photoperiodic flowering in wild and cultivated soybean. The loss-of-function alleles of *E1* family genes thus can be used to develop early-maturity and high-yield soybean cultivars for high-latitude regions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant materials and growth conditions
- METHOD DETAILS
  - Resequencing, mapping, and variation calling
  - Genome-wide association analysis

- DNA isolation and map-based cloning
- RNA extraction and RT-qPCR
- Plasmid construction and plant transformation
- Transient expression assay
- Immunoblot analysis
- ChIP assay

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.11.061.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant nos. 31930083 to B.L., 32090064 to F.K., and 32090065 to L.D.), the Major Program of Guangdong Basic and Applied Research (grant no. 2019B030302006 to F.K. and B.L.), and the Science and Technology Planning Project of Guangzhou (grant nos. 202102010388 to L.D. and 202102010389 to Q.C.).

AUTHOR CONTRIBUTIONS

F.K. and L.D. designed and interpreted the results. F.K., L.D., and B.L. supervised the experiments and coordinated the projects. L.D., S.L., Q.C., L.W., T.S., C.Z., L.K., Z.H., H.L., Haiping Du, T.L., and C.F. performed the experiments. Y.B. and Y.L. provided soybean accessions. L.D., Q.C., F.W., X.P., Hao Du, B.L., and C.F. performed the data analysis. F.K. and L.D. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 20, 2022
Revised: November 1, 2022
Accepted: November 24, 2022
Published: December 19, 2022

REFERENCES

1. Dong, L., Cheng, Q., Fang, C., Kong, L., Yang, H., Hou, Z., Li, Y., Nan, H., Zhang, Y., Chen, Q., et al. (2022). Parallel selection of distinct *Tof5* alleles drove the adaptation of cultivated and wild soybean to high latitudes. Mol. Plant *15*, 308–321.
2. Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., Chen, L., Su, T., Nan, H., Zhang, D., et al. (2020). Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. Nat. Genet. *52*, 428–436.
3. Kou, K., Yang, H., Li, H., Fang, C., Chen, L., Yue, L., Nan, H., Kong, L., Li, X., Wang, F., et al. (2022). A functionally divergent SOC1 homolog improves soybean yield and latitudinal adaptation. Curr. Biol. *32*. 1728.e6–1742.e6.
4. Graham, P.H., and Vance, C.P. (2003). Legumes: importance and constraints to greater use. Plant Physiol. *131*, 872–877.
5. Sinegovskii, M., Yuan, S., Sinegovskaya, V.T., and Han, T. (2018). Current status of the soybean industry and research in the Russian Federation. Soybean Sci. *37*, 1–7.
6. Hymowitz, T. (1970). On the domestication of the soybean. Econ. Bot. *24*, 408–421.
7. Carter, T.E., Nelson, R., Sneller, C.H., and Cui, Z. (2004). Ch. 8. Genetic diversity in soybean. In Soybeans: Improvement, Production and Uses, Third Edition, R.M. Shibbles, J.E. Harper, R.F. Wilson, and R.C. Shoemaker, eds. (American Society of Agronomy).
8. Watanabe, S., Harada, K., and Abe, J. (2012). Genetic and molecular bases of photoperiod responses of flowering in soybean. Breed. Sci. *61*, 531–543.
9. Xia, Z., Watanabe, S., Yamada, T., Tsubokura, Y., Nakashima, H., Zhai, H., Anai, T., Sato, S., Yamazaki, T., Lü, S., et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. USA *109*, E2155–E2164.
10. Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A., and Kong, F. (2017). Molecular mechanisms of flowering under long days and stem growth habit in soybean. J. Exp. Bot. *68*, 1873–1884.
11. 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.
12. Lin, X., Fang, C., Liu, B., and Kong, F. (2021). Natural variation and artificial selection of photoperiodic flowering genes and their applications in crop adaptation. Abiotech *2*, 156–169.
13. Dong, L., Fang, C., Cheng, Q., Su, T., Kou, K., Kong, L., Zhang, C., Li, H., Hou, Z., Zhang, Y., et al. (2021). Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. Nat. Commun. *12*, 5445.
14. Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K., and Harada, K. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. Genetics *188*, 395–407.
15. Zhu, J., Takeshima, R., Harigai, K., Xu, M., Kong, F., Liu, B., Kanazawa, A., Yamada, T., and Abe, J. (2018). Loss of function of the E1-Like-b gene associates with early flowering under long-day conditions in soybean. Front. Plant Sci. *9*, 1867.
16. Tsubokura, Y., Watanabe, S., Xia, Z., Kanamori, H., Yamagata, H., Kaga, A., Katayose, Y., Abe, J., Ishimoto, M., and Harada, K. (2014). Natural variation in the genes responsible for maturity loci *E1*, *E2*, *E3* and *E4* in soybean. Ann. Bot. *113*, 429–441.
17. Liu, L., Song, W., Wang, L., Sun, X., Qi, Y., Wu, T., Sun, S., Jiang, B., Wu, C., Hou, W., et al. (2020). Allele combinations of maturity genes *E1-E4* affect adaptation of soybean to diverse geographic regions and farming systems in China. PLoS One *15*, e0235397.
18. Zhang, X., Wu, T., Wen, H., Song, W., Xu, C., Han, T., Sun, S., and Wu, C. (2021). Allelic variation of soybean maturity genes *E1–E4* in the Huang-Huai-Hai river valley and the northwest China. Agriculture *11*, 478.
19. Qiu, L.J., Xing, L.L., Guo, Y., Wang, J., Jackson, S.A., and Chang, R.Z. (2013). A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol. Biol. *83*, 41–50.
20. Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., et al. (2014). De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat. Biotechnol. *32*, 1045–1052.
21. Kong, L., Lu, S., Wang, Y., Fang, C., Wang, F., Nan, H., Su, T., Li, S., Zhang, F., Li, X., et al. (2018). Quantitative trait locus mapping of flowering time and maturity in soybean using next-generation sequencing-based analysis. Front. Plant Sci. *9*, 995.
22. Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J.X., Mitros, T., Nelson, W., Hyten, D.L., Song, Q.J., Thelen, J.J., Cheng, J.L., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature *463*, 178–183.
23. Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B.M., Watanabe, S., Yamada, T., Tabata, S., Kanazawa, A., Harada, K., and Abe, J. (2010). Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. Plant Physiol. *154*, 1220–1231.
24. Li, X., Fang, C., Yang, Y., Lv, T., Su, T., Chen, L., Nan, H., Li, S., Zhao, X., Lu, S., et al. (2021). Overcoming the genetic compensation response of soybean florigens to improve adaptation and yield at low latitudes. Curr. Biol. *31*, 3. 755.e4–3767.e4.

25. Xu, M., Yamagishi, N., Zhao, C., Takeshima, R., Kasai, M., Watanabe, S., Kanazawa, A., Yoshikawa, N., Liu, B., Yamada, T., and Abe, J. (2015). The soybean-specific maturity gene *E1* Family of floral repressors controls night-break responses through down-regulation of *FLOWERING LOCUS T* orthologs. Plant Physiol. *168*, 1735–1746.

26. Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., Yu, Y., Fu, F., Wang, Y., Qian, F., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. Nat. Commun. *8*, 15323.

27. Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. USA *103*, 16666–16671.

28. Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., Li, M.W., He, W., Qin, N., Wang, B., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat. Genet. *42*, 1053–1059.

29. Fernie, A.R., and Yan, J. (2019). De novo domestication: an alternative route toward new crops for the future. Mol. Plant *12*, 615–631.

30. Tian, J., Wang, C., Xia, J., Wu, L., Xu, G., Wu, W., Li, D., Qin, W., Han, X., Chen, Q., et al. (2019). Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. Science *365*, 658–664.

31. Liu, L., Gao, L., Zhang, L., Cai, Y., Song, W., Chen, L., and Han, T. (2022). Co-silencing *E1* and its homologs in an extremely late-maturing soybean cultivar confers super-early maturity and adaptation to high-latitude short-season regions. J. Integr. Agr. *21*, 326–335.

32. Lin, X., Dong, L., Tang, Y., Li, H., Cheng, Q., Li, H., Zhang, T., Ma, L., Xiang, H., Chen, L., et al. (2022). Novel and multifaceted regulations of photoperiodic flowering by phytochrome A in soybean. Proc. Natl. Acad. Sci. USA *119*. e2208708119.

33. Dietz, N., Combs-Giroir, R., Cooper, G., Stacey, M., Miranda, C., and Bilyeu, K. (2021). Geographic distribution of the *E1* family of genes and their effects on reproductive timing in soybean. BMC Plant Biol. *21*, 441.

34. Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science *288*, 1613–1616.

35. Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. Nature *410*, 1116–1120.

36. Yanovsky, M.J., and Kay, S.A. (2002). Molecular basis of seasonal time measurement in Arabidopsis. Nature *419*, 308–312.

37. Kobayashi, Y., and Weigel, D. (2007). Move on up, it's time for change—mobile signals controlling photoperiod-dependent flowering. Genes Dev. *21*, 2371–2384.

38. Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. Nat. Rev. Genet. *13*, 627–639.

39. Song, Y.H., Ito, S., and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. Trends Plant Sci. *18*, 575–583.

40. Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., and Sasaki, T. (2000). *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. Plant Cell *12*, 2473–2484.

41. Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature *422*, 719–722.

42. Ishikawa, R., Aoki, M., Kurotani, K., Yokoi, S., Shinomura, T., Takano, M., and Shimamoto, K. (2011). Phytochrome B regulates *Heading date 1* (*Hd1*)-mediated expression of rice florigen *Hd3a* and critical day length in rice. Mol. Genet. Genomics *285*, 461–470.

43. Bu, T., Lu, S., Wang, K., Dong, L., Li, S., Xie, Q., Xu, X., Cheng, Q., Chen, L., Fang, C., 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.

44. McBlain, B.A., Hesketh, J.D., and Bernard, R.L. (1987). Genetic effects on reproductive phenology in soybean isolines differing in maturity genes. Can. J. Plant Sci. *67*, 105–115.

45. Upadhyay, A.P., Ellis, R.H., Summerfield, R.J., Roberts, E.H., and Qi, A. (1994). Characterization of photothermal flowering responses in maturity isolines of soyabean [*Glycine max* (L.) Merrill] cv. Clark. Ann. Bot. *74*, 87–96.

46. Fehr, W.R., and Cavines, C.E. (1977). Stages of Soybean Development Special Report (Iowa State University).

47. Lu, S., Zhao, X., Hu, Y., Liu, S., Nan, H., Li, X., Fang, C., Cao, D., Shi, X., Kong, L., et al. (2017). Natural variation at the soybean *J* locus improves adaptation to the tropics and enhances yield. Nat. Genet. *49*, 773–779.

48. Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., Xing, S., Du, J., Ma, S., and Tian, Z. (2018). De novo assembly of a Chinese soybean genome. Sci. China Life Sci. *61*, 871–884.

49. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

50. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

51. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

52. Cingolani, P., Platts, A., Wang, le L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly *6*, 80–92.

53. Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. Genome Biol. *18*, 161.

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

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-FLAG mouse monoclonal antibody	Abcam	Cat# ab125243; RRID: AB_11001232
Bacterial and virus strains		
Agrobacterium tumefaciens (strain GV3101)	Bu et al.[43]	N/A
Agrobacterium tumefaciens (strain EHA101)	Lu et al.[2]	N/A
E. coli (strain DH5α)	Widely distributed	N/A
Critical commercial assays		
NuClean Plant Genomic DNA Kit	CWBIO	Cat# CW0531M
Ultrapure RNA Kit	CWBIO	Cat# CW0581M
Reverse Transcriptase Kit (M-MLV)	Takara	Cat#ZR102
LightCycler® 480 SYBR Green I Master	Roche	Cat# 04887352001
Dual-Luciferase Reporter Assay System	Promega	Cat# E1910
Deposited data		
Details of genome resequencing data	This paper	BioProject: PRJNA859249
Details of genome resequencing data	Dong et al.[1]	BioProject: PRJNA743225
Details of genome resequencing data	Kou et al.[3]	BioProject: PRJNA776405
Details of genome resequencing data	Lu et al.[2]	BioProject: PRJNA394629; PRJCA000205; PRJCA001691;
Experimental models: Organisms/strains		
Diverse soybean accessions	China, Russia, and the United States	See Data S2
Soybean: *Williams82*	Widely distributed	N/A
Soybean: $tof4^{CR}$, $tof4^{CR}e1^{CR}$, $tof4^{CR}e1^{CR}e1lb^{CR}$, $tof4^{CR}$-D	This paper	N/A
Soybean: *Tof4-OE*	This paper	N/A
Soybean: *ft2a ft5a*	Li et al.[24]	N/A
Soybean: $e1^{CR}$	Dong et al.[1]	N/A
Oligonucleotides		
Primer sequences used for vector construction, sequencing and expression analysis	This paper	See Table S2
Recombinant DNA		
pTF101	Lu et al.[2]	N/A
pGreenII 0800-LUC	Bu et al.[43]	N/A
pRT107-35S$_{pro}$:SOC1b	This paper	N/A
proFT2a-LUC-REN	This paper	N/A
proFT5a-LUC-REN	This paper	N/A
proTof5-LUC-REN	This paper	N/A
proTof4-Tof4-Flag	This paper	N/A
pro35S-Tof4-Flag	This paper	N/A
pro35S-tof4-1-Flag	This paper	N/A
pro35S-tof4-2-Flag	This paper	N/A
pro35S-Tof4-H4-Flag	This paper	N/A
pro35S-Tof4-H5-Flag	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
BWA version 0.7.17-r1188	Lu et al.[2]	http://bio-bwa.sourceforge.net/
Picard version V1.109	Lu et al.[2]	http://broadinstitute.github.io/picard/
Genome Analysis Toolkit version V3.2-2	Lu et al.[2]	https://gatk.broadinstitute.org/hc/en-us
GATK and SAMtools version 1.9	Lu et al.[2]	http://samtools.sourceforge.net/
SnpEff version 1.9.6	Lu et al.[2]	https://pcingola.github.io/SnpEff/
PHYLIP version 3.68	Lu et al.[2]	https://evolution.gs.washington.edu/phylip/
EIGENSOFT version 6.1.1	Lu et al.[2]	https://github.com/DReichLab/EIG
PLINK version 0.72	Lu et al.[2]	http://pngu.mgh.harvard.edu/purcell/plink/
EMMAX beta version	Lu et al.[2]	https://genome.sph.umich.edu/wiki/EMMAX
VCFtools version 0.1.16	Lu et al.[2]	https://vcftools.github.io/man_latest.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Fanjiang Kong (kongfj@gzhu.edu.cn).

Materials availability

The genetic materials that support the findings of this study are available from the corresponding authors upon request.

Data availability

The sequencing data of the 95 wild accessions used in this study have been deposited into the NCBI database: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA859249. The previously reported sequence data of 372 accessions used in this study have been deposited into the NCBI database: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA743225.[1] The previously reported sequence data of 1,295 accessions were deposited into the NCBI database: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA394629 and the GSA database of the BIG Data Center: https://ngdc.cncb.ac.cn/search/?dbId=gsa&q=PRJCA000205 and https://ngdc.cncb.ac.cn/search/?dbId=gsa&q=PRJCA001691.[2] The previously reported sequence data of 349 accessions were deposited into the NCBI database: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA776405.[3] Accession numbers are listed in the key resources table.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant materials and growth conditions

The 295-accession panel was grown during the cultivation season (May to October) in 2018 and 2019 in Harbin (45.75′ N, 126°63′ E). Flowering time was investigated in 2018 and 2019. For map-based cloning, one heterozygous inbred progeny population of 2,210 individuals segregating at the *Tof4* locus was subsequently developed. NILs for the *Tof4* locus were selected from F_8 progeny of this same cross using molecular markers for *Tof4*. The heterozygous inbred progenies, NILs, and CRISPR–Cas9 knockout mutants used for phenotyping were grown under natural LD conditions in the field from 2018 to 2021 in Harbin or under natural LD conditions in 2021 at the Experimental Station of the Jilin Academy of Agricultural Sciences (43.88′ N, 125°35′ E). For natural LD conditions, plants were sown in the beginning of May, spaced 0.15 m apart in rows 5 m long with 0.7 m between rows, and harvested in September or October. Plants used for expression analysis and ChIP assays were grown under LD conditions (16 h light/8 h dark) in a plant growth cabinet (Conviron Adaptis A1000) with a light intensity of 500 μmol m^{-2} s^{-1}. For phenotypic investigations, the days from emergence to the first flowering, corresponding to the R1 stage,[46] were scored. The days from emergence to when pods attained a mature color, corresponding to the R8 stage,[46] were also scored. Yield-related traits were recorded at the R8 stage.[47]

METHOD DETAILS

Resequencing, mapping, and variation calling

The libraries for each accession in the 95-accession panel were constructed following the manufacturer's instructions (Illumina Inc., San Diego, CA, USA). The DNA-seq libraries were sequenced on the HiSeq X Ten system (150-bp paired-end reads). Resequencing reads of the 95 accessions sequenced in this study and the 2,292 previously sequenced accessions, as well as mapping and SNP

calling, were performed as described previously.[13,2,1,3] In brief, all quality-controlled reads were aligned to the cultivated soybean W82 genome (Wm82.a2.v1) using BWA 0.6.1-r104 software with default parameters.[22,48] DNA variants, including SNPs and InDels, were called using GATK (ver. 3.1.1) and SAMtools (ver. 0.1.19) software independently,[49–51] the common sites identified by both methods were retained for further analysis. SNPs with missing data minor allele frequency (MAF) < 1% were filtered, and InDels with a maximum length of 10 bp were included. Annotations of SNPs and InDels were performed based on the W82 genome using snpEff (ver. 3.1) software.[52] SNPs in coding sequences were classified as synonymous SNPs or nonsynonymous SNPs. InDels in exons were further categorized according to whether they led to a frameshift.

Genome-wide association analysis

Principal components as determined with TASSEL5 were used as population structure (Q) matrices. Kinship (K) matrix was calculated using centered IBS method in TASSEL5. To identify loci influencing flowering time in the 295-accession panel, we conducted a GWAS through the MLM implemented in Efficient Mixed-Model Association eXpedited (EMMAX) software based on the SNPs with a MAF > 0.05 and using the flowering time in Harbin in 2018 and 2019 for each accession.[53]

DNA isolation and map-based cloning

Genomic DNA was extracted from fresh trifoliate leaves of 2-week-old seedlings with a SurePlant DNA kit (CWBIO) and used to amplify InDel (Insertion-Deletion) markers. The primer sequences used to amplify the markers for mapping are listed in Table S2. For fine mapping, InDel and simple sequence repeat markers were developed in the regions of *Tof4* based on the resequencing data of the two parents, W82 and DN50. Ten recombinants were identified in the fine-mapping population of *Tof4* using seven markers. The flowering time of the progeny of these recombinants was evaluated to delimit the genomic interval containing *Tof4*.

RNA extraction and RT-qPCR

Leaves were sampled from W82 and *tof4*^{CR} mutant, and NILs of *Tof4* soybean plants at 20 DAE under long-day (16 h light/8 h dark) conditions for expression analysis of *FT2a*, *FT5a* and *Tof5*.Total RNA was extracted using an Ultrapure RNA kit (CWBIO), and the RNA was reverse transcribed using a Super Script First-strand cDNA Synthesis System (Takara, Dalian, China). Quantitative PCR (qPCR) was performed using SYBR Green Real-Time PCR Master Mix (Roche). Three independent RNA samples were prepared for biological replicates. The soybean *Tubulin* (Glyma.05G157300) gene was used as the internal reference. All qPCR primers are listed in Table S2.

Plasmid construction and plant transformation

The coding sequence of *Tof4* was amplified from cDNA of W82 using the primer set Tof4-3Flag-F/R and inserted into the *Xba* I and *Mlu* I sites of the *pTF101-3×Flag* vector (containing the *BAR* gene for glufosinate resistance) to generate the *pro35S*$_{pro}$*:Tof4-3×Flag* construct. To obtain soybean *tof4*^{CR} and *tof4*^{CR} *e1*^{CR} *e1Ib*^{CR} triple mutants, one 20-bp sequences of *Tof4* were selected as target site for Cas9 cleavage and cloned into the pYLCRISPR–Cas9–p35S-BS vector as previously described.[1] The recombinant vector was introduced into *Agrobacterium tumefaciens* strain EHA101 and used to transform W82 via A. *tumefaciens*–mediated transformation.[54] Primers used for vector construction are listed in Table S2.

Transient expression assay

The *FT2a*$_{pro}$*:LUC* and *FT5a*$_{pro}$*:LUC* reporter constructs were described previously.[1] The different alleles of *Tof4* (*tof4–1* and *tof4–2*, *Tof4–H4*, *Tof4–H5*) were introduced into the *pTF101–3×Flag* vector to generate the constructs *35S*$_{pro}$*:Tof4–3×Flag*, *35S*$_{pro}$*:tof4–1–3×Flag*, *35S*$_{pro}$*:tof4–2–3×Flag*, *35S*$_{pro}$: *Tof4–H4–3×Flag*, and *35S*$_{pro}$: *Tof4–H5–3×Flag*. The *FT2a*$_{pro}$*:LUC* and *FT5a*$_{pro}$*:LUC* reporter construct was used as the reporter, and various *Tof4* constructs were used as the effectors in the Arabidopsis protoplast transient expression system to test whether Tof4 suppresses the transcription of *FT2a* or *FT5a*. Arabidopsis mesophyll protoplasts were prepared, transfected and cultured as previously described.[1] The LUC and REN activities were measured under the manufacturers' instructions (Promega). The LUC/REN ratio was presented with three biological replicates, and the primers used are listed in Table S2.

Immunoblot analysis

To analyze protein expression in the transgenic plants, total proteins were extracted from W82 and the *Tof4*$_{pro}$*:Tof4–3×Flag* transgenic lines with protein extraction buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 2 mM PMSF and 1× Protease Inhibitor Cocktail). Samples were centrifuged at 16,000g for 5 min at 4 °Cand the supernatants were electrophoresed on a sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) mini-gel to separate the proteins.

For immunoblotting the separated proteins were transferred onto a polyvinylidene fluoride membrane (PVDF) by the Trans-Blot Turbo Transfer System (Bio-rad). Blots were probed with anti-Flag antibody (Sigma, A8692) antibody conjugated to horseradish peroxidase (HRP). The blots were developed with a 1:1 mix of SuperSignal West Femto Maximum Sensitivity and SuperSignal West Dura Extended Duration Substrates and signals were detected on Amersham Imager600 System (GE). Anti-Flag (HRP) antibody was used at 1:5,000-fold dilutions.

ChIP assay

Leaf samples were collected from 20-day-old seedlings at Zeitgeber time 4 under LD conditions from W82 and *Tof4*$_{pro}$*:Tof4–3×Flag* transgenic lines. The samples were fixed in 1% formaldehyde on ice for 15 min under a vacuum. Nuclei were isolated from the samples and sonicated to generate DNA fragments with an average size of ~250–500 bp. The solubilized chromatin was immunoprecipitated with anti-Flag M2 magnetic beads (Sigma, M8823). The coimmunoprecipitated DNA was recovered and analyzed by qPCR in triplicate. Relative fold enrichment was calculated by normalizing the amount of a target DNA fragment against that of a genomic fragment of the reference gene *Tubulin* (Glyma.05G157300) and then against the respective input DNA samples. The enrichment of the *EF1b* genomic fragment was used as a negative control. Three biological replicates were performed in ChIP-qPCR. The primers used for amplification are listed in Table S2.

QUANTIFICATION AND STATISTICAL ANALYSIS

In this study, all values were presented as mean ± s.d. and numbers (*n*) of samples or replicates are indicated in figure legends. Data were analyzed with GraphPad Prism 8 (ver. 8.0.1). Significance levels of differences were calculated by one-tailed, two-sample Student's *t*-tests or one-way ANOVA with GraphPad Prism 8 (ver. 8.0.1). To evaluate the phenotypes of the 329 accessions, at least 10 individual plants of each accession were analyzed.