Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics

Lidong Dong1,7, Chao Fang1,7, Qun Cheng1,7, Tong Su2,7, Kun Kou2,7, Lingping Kong1, Chuanbo Zhang3, Haiyang Li1, Zhihong Hou1, Yuhang Zhang1, Liyu Chen1, Lin Yue1, Lingshuang Wang2, Kai Wang4, Yongli Li1, Zhuoran Gan1, Xiaohui Yuan5, James L. Weller6, Sijia Lu1, Fanjiang Kong1,2✉ & Baohui Liu1,2✉

Soybean (Glycine max) serves as a major source of protein and edible oils worldwide. The genetic and genomic bases of the adaptation of soybean to tropical regions remain largely unclear. Here, we identify the novel locus Time of Flowering 16 (Tof16), which confers delay flowering and improve yield at low latitudes and determines that it harbors the soybean homolog of LATE ELONGATED HYPOCOTYL (LHY). Tof16 and the previously identified J locus genetically additively but independently control yield under short-day conditions. More than 80% accessions in low latitude harbor the mutations of tof16 and j, which suggests that loss of functions of Tof16 and J are the major genetic basis of soybean adaptation into tropics. We suggest that maturity and yield traits can be quantitatively improved by modulating the genetic complexity of various alleles of the LHY homologs, J and E1. Our findings uncover the adaptation trajectory of soybean from its temperate origin to the tropics.

https://doi.org/10.1038/s41467-021-25800-3
soybean (Glycine max [L.] Merr.) is one of the most economically important leguminous crops, as it provides more than one-quarter of the world’s protein for human and animal consumption. In 2020/2021, global soybean production was 362 million tons; soybean production in low latitude regions (Brazil, Argentina, and India) comprised 53.45 percent of the total worldwide production (United States Department of Agriculture, 2020). Therefore, improving soybean productivity in the tropics will have a great effect on fulfilling the demands of the growing worldwide population.

Cultivated soybean was domesticated from its wild progenitor (Glycine soja Sieb. & Zucc.) in temperate regions of China ~5000 years ago. This facultative short-day (SD) plant is extremely sensitive to photoperiod. This high sensitivity to photoperiod dramatically impedes the improvement of soybean productivity in the tropics. For instance, cultivars that are adapted to temperate regions flower early and produce extremely low grain yields when grown in low-latitude regions. The long-juvenile (LJ) trait, which delays flowering and enhances grain yields under SD conditions, was first introduced into soybean cultivars in Brazil in the 1970s, which expanded soybean cultivation into the tropics. Only two loci, J and E6, have been reported to control the LJ trait; both contain the same ortholog of Arabidopsis thaliana EARLY FLOWERING 3 (E1)

In the current study, we identify a novel locus that controls flowering time and yield in soybean under SD conditions. This locus, Tof16, harbors a LATE ELONGATED HYPOCOTYL (LHY) gene, as revealed by whole-genome resequencing and genome-wide association studies (GWAS) and by positional cloning of quantitative trait loci (QTL). We confirm that four LHY homologs redundantly control flowering time and yield. We also demonstrate that Tof16 and J additively but independently control soybean flowering. Both Tof16 and J directly bind to the promoter of E1, which encodes a core flowering suppressor, to repress its transcription. Interestingly, weak mutant alleles for both tof16 and J are initially selected, and additional null alleles are subsequently acquired in addition to these weak alleles and further selected, suggesting that stepwise selection of natural mutations of both genes occur during the adaptation of soybean from temperate regions to the tropics. By modulating the combinations of mutations of LHY homologs or combinations of different alleles of Tof16, J, and E1, maturity and yield are altered in a quantitative manner. Our findings uncover the genetic and genomic basis for the adaptation of soybean to low-latitude regions and provides a new approach for precise breeding for improve soybean productivity in the tropics.

Identification of the Tof16 locus. Using a linear mixed model for GWAS of the panel of 329 accessions, we identified one consistent significantly associated locus on chromosome 16 (P < 1.13 × 10^{-8}) controlling flowering time under natural SD conditions in both 2018 and 2019 in Guangzhou. (Fig. 1a, b). This GWAS peak is consistent with the SD flowering QTL on chromosome 16 that we previously identified using two F2 segregation populations, PI591429 × PI628930 and PI240664 × BR12116. These analyses across multiple biparental and natural populations indicate that variation at this locus (hereafter referred to as Time of Flowering 16, Tof16) is widespread and substantially contributes to the control of flowering time under SD conditions.

Tof16 encodes LHY1a. We generated a large (n = 2418) inbred F6 population from the PI591429 × PI628930 cross for fine-mapping of the Tof16 locus by recurrent selection for heterozygosity at Tof16 from the F2 to F6 generations. Analysis of this population located Tof16 within a 120-kilobase (kb) region harboring 12 genes based on the Williams 82 reference genome. Using two F2 segregation populations, PI591429 and PI628930 (Supplementary Fig. 3b). The late flowering parent PI628930 harbored two SNPs predicted to cause a gain of a stop codon, resulting in premature termination of translation after 159 amino acids in the 750-amino-acid LHY1a protein (Supplementary Figs. 3b and 4). We also sequenced the coding region of LHY1a in PI240664 and BR121, finding that in PI240664, this gene harbored one SNP predicted to convert a serine into a cysteine, which is a conserved site in leguminous plants (Supplementary Figs. 4 and 5). The presence of different mutations in these two lines suggests that the LHY1a gene was a strong candidate for the Tof16 locus.

To validate whether LHY1a is the causative gene of the Tof16 locus, we generated loss-of-function mutations of LHY1a (named tof16CR) in the Harosoy background using CRISPR/Cas9-mediated gene editing and evaluated the phenotypes of the mutants vs. wild-type Harosoy (Supplementary Fig. 6). The tof16CR plants showed significantly delayed flowering time and maturity, altered yield-related traits, and improved overall grain yield relative to Harosoy (Fig. 1c–f, Supplementary Fig. 7a–d). These results confirm the notion that LHY1a is the causative gene of the Tof16 locus and that two natural mutations arose in PI628930 and PI240664 (hereafter referred to as tof16-1 and tof16-2, respectively).

To examine the specific effects of the Tof16 locus, we also compared the phenotypes of two F2, near-isogenic lines (NILs) carrying either the functional Tof16 allele (NIL-Tof16) or the non-functional tof16-1 alleles (NIL-tof16-1 or NIL-tof16-2) in different genetic backgrounds. NIL-tof16-1 and NIL-tof16-2 showed significantly delayed flowering time and maturity (Supplementary Fig. 8a–c, e–g), along with increased plant height, node number, pod number, branch number (Supplementary Fig. 7e–l), and grain yield compared to the functional NILs (Supplementary Fig. 8d, h), confirming the notion that the tof16-1 and tof16-2 alleles delay flowering and greatly enhance grain yields in soybean under SD conditions. Therefore, like J, Tof16 also functions as a flowering enhancer; the loss of function of both genes might have contributed to the adaptation of soybean to the tropics.
Fig. 1 Identification of Tof16. a-b GWAS scan for flowering time (R1 stage) using data from the 329-accession panel grown over the 2018 (a) and 2019 (b) field seasons in Guangzhou, China. c Phenotypes of tof16 mutation and wild-type Harosoy under SD (12 h light/12 h dark) conditions. Scale bar, 10 cm. d Flowering time. e Time to maturity. f Grain yield per plant of tof16CR mutation and Harosoy. All data were given as mean ± s.e.m. (n = 10 plants), the value of each plant was represented by a dot. One-tailed Student’s t-test was used to generate the P values. g Diurnal variation in transcript levels of in wild-type Harosoy and tof6CR mutant under SD. All data were given as mean ± s.e.m. (n = 5 plants). h Location of the AATATC-motif in E1 gene promoter. i ChIP-qPCR results demonstrated that Tof16 directly bound to the promoter of E1 gene. Values are mean ± s.e.m. (n = 3 biologically independent samples), the value of each replication was represented by a dot. Source data underlying Fig. 1d-g and i are provided as a Source Data file.
Tof16 is genetically dependent on the legume-specific flowering repressor E1. E1 plays a central role in photoperiodic flowering by repressing the expression of two key FT homologs, FT2a and FT5a. There are three major natural alleles of E1 in soybean: E1, e1as, and e1fl. e1fl is a weak mutant allele and e1fl is a null functional allele. To examine the genetic interaction of Tof16 and E1, we developed a NIL set for E1/Tof16, E1/tof16CR, e1as/Tof16, e1as/tof16CR, e1fl/Tof16, and e1fl/tof16CR in the Harosoy background and performed phenotypic analysis. The Tof16/e1fl allele delayed flowering and maturity in both the E1 and e1as genetic backgrounds, but the effect was weaker in the e1as background. By contrast, the effect of Tof16 on flowering was completely eliminated in the e1fl null functional background (Supplementary Fig. 9a–c), implying that the full effect of Tof16 on flowering depends on E1.

We evaluated the effect of Tof16 on the transcriptional regulation of E1, FT2a, and FT5a under SD (12 h light/12 h dark) conditions using tof16CR and Harosoy or independent NIL pairs for each locus. As expected, E1 was expressed at higher levels in tof16CR (Fig. 1g), and FT2a and FT5a were expressed at lower levels in this line compared to Harosoy (Supplementary Fig. 10a, b). A similar result was obtained for the NILs. These results indicate that functional alleles of Tof16 repress E1 expression and increase FT2a and FT5a expression relative to the mutant alleles (Supplementary Figs. 10c, d and 11).

To further explore the molecular nature of the relationship between Tof16 and E1, we determined whether Tof16 directly binds to the promoter of E1 in vivo. We generated transfectants overexpressing Tof16-6HA in the Williams 82 background and subjected them to chromatin immunoprecipitation (ChIP)-qPCR assays (Supplementary Fig. 12). Tof16 directly associated with the E1 promoter regions that contained AATATC motif (a part of the EE motif, Fig. 1h, i). These results confirm our previous finding that Tof16 protein could bind to AATATC motif in the E1 promoter in vitro. Taken together, these results indicate that Tof16 enhances early flowering and maturity by direct associating with the E1 promoter to suppress E1 expression, thus releasing FT2a and FT5a transcription.

Four LHY homologs redundantly control soybean flowering and grain yield. The soybean genome contains four LHY/CCA1 homologs (LHY1a, LHY1b, LHY2a, and LHY2b) and its amino acid sequences were high homology (Supplementary Fig. 5). Since we demonstrated that Tof16/LHY1a controls flowering time, maturity, and grain yield in soybean, we asked whether the other LHY family members also control these soybean traits and whether these homologs could potentially be used for agricultural applications. We crossed the lhy1a/lhy1b/lhy2a/lhy2b quadruple mutants with wild-type Harosoy and obtained all 15 homozygous mutational combinations of LHY (Fig. 2a). We examined the phenotypic differences of the mutants grown in fields in Guangzhou under natural SD conditions. Among the single mutants, lhy1a and lhy1b, but not lhy2a or lhy2b, showed significantly delayed flowering time and maturity and improved overall grain yield relative to Harosoy (Fig. 2a–e, Supplementary Fig. 13a–d). All of the multiple mutants except the lhy2c/2d double mutants showed significantly delayed flowering time and increased grain yield in a quantitative manner compared to wild-type Harosoy (Fig. 2a–e, Supplementary Fig. 13a–d). Strikingly, the lhy1a/lhy2b triple mutants exhibited the best architecture and highest grain yield but shorter flowering time and earlier maturity compared to the lhy1a/lhy2b quadruple mutants under natural SD conditions (Fig. 2a–e, Supplementary Fig. 13a–d). Furthermore, all mutants possessing the lhy1a (tof16) mutation had higher grain yields than those without this mutation, indicating that lhy1a plays a crucial role in controlling flowering time, maturity, and grain yield under SD conditions (Fig. 2a–e, Supplementary Fig. 13a–d).

To gain further insight into how LHY homologs control flowering time and grain yields under long-day (LD) conditions, we also evaluated the mutants in the field in Changchun under natural LD conditions. All mutants except lhy2a, lhy2b, and lhy2a/2b showed delayed flowering time and maturity compared to the wild type (Supplementary Fig. 14a–i). Due to the larger effect of LHY1a on flowering time, multiple mutants carrying the lhy1a mutation failed to mature, and it was difficult to harvest the seeds naturally until the end of the growing season (Supplementary Fig. 14a–i). However, unlike their performance under SD conditions, the lhy1a and lhy1b mutants exhibited the best architecture and improved grain productivity under natural LD conditions. These results suggest that LHY homologs have redundant but divergent functions in controlling flowering time, maturity, and grain yield in soybean under both SD and LD conditions. We further examined the expression of four LHY genes under LD and SD conditions and found that the expression of four LHY genes were no significant difference under LD and SD conditions (Supplementary Fig. 15). We also tested the expression levels of E1 in lhy multiple mutants, the result showed that the amount of E1 correlated with flowering and maturity under LD and SD conditions, but not yield (Supplementary Figs. 16–19). Therefore, manipulating the combinations of these alleles and modulating the genetic complexity of the LHY homologos could help create the appropriate genotypes to maximize the adaptation and yield potential of soybean at different latitudes.

Tof16 and J additively control flowering time and grain yield. The adaptation of soybean to low latitude or tropical regions largely depends on the natural loss of function of the flowering enhancer J. Therefore, it is critical to explore the genetic relationship between Tof16 and J. We tested the reciprocal transcriptional regulation between Tof16 and J using mutants or NIL sets of tof16 or j. No mutual transcriptional regulation was observed between Tof16 and J (Supplementary Fig. 20). To further explore the genetic interaction of Tof16 and J, we developed two NIL sets for the four different homozygous allelic combinations at two loci from a cross between tof16CR/J and NIL-j/Nil or a cross between tof16CR/e1as and NIL-j/e1as in the Harosoy background and subjected them to phenotypic evaluation. The presence of a recessive allele of either Tof16 or J delayed flowering and maturity and enhanced grain yield in both the E1 and e1as genetic backgrounds. However, the double recessive mutant tof16 j showed significantly later flowering, maturity, and higher grain yield than either the j or tof16 single mutant in both the E1 and e1as genetic backgrounds. For the double recessive mutant, both flowering and grain yield were signficantly lower than those of E1 and j.

Consistent with this genetic effect, E1 transcript levels were highest in NIL-tof16CR/j, followed by NIL-Tof16/j or NIL-tof16CR/J, and were lowest in NIL-Tof16/J. As expected, FT2a and FT5a showed the opposite expression pattern (Fig. 3e–g). These results, together with our previous findings, suggest that the positive regulators of flowering Tof16 and J both promote flowering, which depends on the function of E1. However, Tof16 and J independently but additively regulate flowering time, maturity, and grain yield in soybean.

Based on these findings, we proposed two possible methods for the quantitative improvement of flowering time and grain yield in soybean in tropical regions (Fig. 3h, i): (1) Due to the genetically redundant and divergent effects of LHY homologs, genotypes with a quantitative series of flowering time, maturity, and yield
traits could be created by combining various CRISPR/Cas9-generated mutants (Fig. 3h). (2) Due to the genetic effects of Tof16, J, and E1, their various natural or artificial alleles could be combined, allowing another set of genotypes conditioning quantitative traits to be produced (Fig. 3i). How these genotypes could be selected or utilized remains to be explored and depends on the photoperiod or latitudinal environment.

**Stepwise selection of Tof16 and J during soybean adaptation.**

Our findings indicate that the positive regulators of flowering Tof16 and J play essential roles in the adaptation of soybean to SDs and yield development. To gain insight into the evolutionary trajectory of soybean adaptation from its temperate origin to the tropics, we examined the genomic variations in the Tof16 and J coding sequences in 1624 resequenced soybean accessions, including 1295 previously described accessions and the 329-accession panel used in the current study. We identified 15 unique high-confidence haplotypes in Tof16. In addition to haplotypes H10 (tof16-1) and H11 (tof16-2), we identified two novel loss-of-function haplotypes: H1 (named tof16-3) and H8 (named tof16-4) (Supplementary Fig. 22a, Supplemental Data 2). Analysis of haplotype origin indicated that the tof16-2 allele (H11, Tof16-SNPA1276T) first occurred in wild soybeans originating in the Huanghui region and were subsequently domesticated into landraces in areas where soybean domestication occurred (Fig. 4a, Supplementary Fig. 22b). Following domestication and during the adaptation of soybean to low latitudes, H11 (tof16-2) was intensively selected in the accessions that adapted to low-latitude regions, suggesting that this haplotype plays critical roles in soybean adaptation to the tropics (Fig. 4a). Interestingly, H10 (tof16-1) and H8 (tof16-4) originated from H11 (tof16-2), suggesting that loss-of-function alleles of tof16-2, tof16-1, and tof16-4 were under stepwise selection during adaptation to low latitudes. However, H8 (tof16-3) originated from H1 and only occurred in India and Nepal, while tof16-1 and tof16-4 mainly occurred in Brazil, indicating that unique loss-of-function alleles of tof16 were selected in different regions to enhance the adaptation of soybean (Fig. 4a, Supplementary Fig. 22b).

To further validate the functional significance of tof16-1, tof16-2, tof16-3, and tof16-4, we performed a transient transfection

---

**Fig. 2 Redundancy among four LHY genes regulates soybean flowering time and yield under SD (12 h light/12 h dark) conditions.** a Phenotypes of lhy mutants. Scale bar, 10 cm. b Flowering time. c Time to maturity. d Total pods per plant height. e Grain yield per plant. All data were given as mean ± s.e.m. (n = 10 plants), the value of each plant was represented by a dot. The presence of the same lowercase letter above the histogram bars in (b–e), denotes nonsignificant differences across the two panels (P > 0.05). One-way ANOVA was used to generate the P values. Source data underlying Fig. 2b–e are provided as a Source Data file.
assay in Arabidopsis thaliana protoplasts. tof16-2 partial impaired the ability of Tof16 to repress the expression of E1, but tof16-1, tof16-4, and tof16-3 completely impaired this ability (Supplementary Fig. 23). These results imply that partial loss of function of tof16-2 (standing variations from soybean in the central area of origin) was first selected during soybean adaptation to the tropics but was not sufficient for full adaptation. In this genetic background, null alleles of tof16-1 and tof16-4 occurred and were further selected for better adaptation to the tropics. Population genetic association analysis of flowering time in the 329-accession...
Flowering time variations of eight allelic combinations of Tof16 and J in Guangzhou 2018 (j-1, j-3, j-6 and j-10) were shown in Supplementary Fig. 25, indicating that H21 (j-SNPG920T) arose during the adaptation of soybean to low latitudes. The presence of the same lowercase letter above the histogram bars in (b, c, e, f, h, i) denotes nonsignificant differences across the two panels (P > 0.05). One-way ANOVA was used to generate the P values. All data were given as mean ± s.e.m. The value of each plant was represented by a dot. Source data underlying Fig. 4b, c, e, f, h, and i are provided as a Source Data file.

Fig. 4 Geographical distribution of genetic diversity of Tof16 and J. a Loss-of-function alleles of Tof16 frequency is highly correlated with low latitude regions. Data from 1624 diversity panels. b–f Flowering time (DAE) variations in 329 accessions possess Tof16 and tof16 alleles in Guangzhou 2018 (b), and Guangzhou 2019 (c). Proportions of j alleles in upper pie chart. Dark gray represents J alleles and light gray represents j alleles. d Loss-of-function alleles of J frequency is highly correlated with low latitude regions. Data from 1624 diversity panels. j-SNPG920T represents j-11 allele including j-SNPG920T, β20bm represents SNP920-based mutations including j-1, j-3, j-6 and j-10 mutational alleles. j-others represents j mutations other than j-SNPG920T. e–f Flowering time (DAE) variations in 329 accessions possess J and j alleles in Guangzhou 2018 (e), and Guangzhou 2019 (f). Proportions of tof16 alleles in upper pie chart. Dark gray represents tof16 alleles and light gray represents tof16 alleles. g Loss-of-function Tof16 and J improve soybean adaptation to low latitudes. Data from 1624 diversity panels. h–i Flowering time variations of eight allelic combinations of Tof16 and J in Guangzhou 2018 (h), and Guangzhou 2019 (i).
groups, J, j^20, j^20bm (SNP920-based mutations, including the j-1, j-3, j-6, and j-10 mutant alleles), and j^other (mutations other than j^20bm) and evaluated their associations with flowering in the 329-accession panel. J^20bm flowered the latest, followed by j^other, j^20, and J in both 2018 and 2019 (Fig. 4e, f). Taken together, our data suggest that like the selection at Tof16, the selection of j^20bm and j^20 also arose in a stepwise manner.

Selection of natural mutations of Tof16 and J allowed soybean to move into the tropics. We firstly investigated that the distribution of the various natural alleles of E1, Tof16, and J in tropical soybean accessions. We found that all of the tropical accessions harbor the dominant E1 allele (Supplementary Data 3), and the frequency of the tof16-2 allele (26.75%) and j-11 allele (47.42%) were highly variable in tropical accessions (Supplementary Tables 2 and 3). To further explore the contributions of mutations of Tof16 and J to the adaptation of soybean to low latitudes, we grouped eight allelic combinations (Tof16/J, Tof16/j^20, Tof16/j^other, tof16-2/J, tof16-2/j^20, tof16-2/j^other, tof16-1,j^20, tof16-3/j^20) and examined the geographic distributions of the eight Tof16/j allelic combinations in the 1624 accessions covering all latitudes. Accessions carrying loss-function-of Tof16 alleles were enriched in Brazil, but accessions carrying loss-function-of J alleles were enriched in Southern China, Southeast Asia, and Brazil (Fig. 4g), suggesting that the selection of mutations of tof16 and j might have occurred independently. Interestingly, by extracting the 165 accessions from low latitude, we found that more than 80% accessions harbor the loss of functions of Tof16 and J suggest the two genes are the major genetic forces to drive soybean adaptation from temperate into tropics (Supplementary Data 4, Supplementary Table 4). By contrast, most accessions adapted to high-latitude regions such as Northern China, the United States, and Canada, where early flowering is required, carried the full functional alleles of Tof16 and J (Fig. 4a, d, g), suggesting that the mutations of these two genes helped soybean move from its temperate origins into the tropics.

Finally, to further explore the functional significance of tof16 and j, we examined their association with flowering time in the 329-accession panel in two environments in Guangzhou (2018 and 2019). Accessions carrying two recessive alleles flowered significantly later than accessions carrying single recessive alleles tof16 or j, and accessions carrying both functional alleles (Tof16/J) flowered earlier than the other three genotypic groups in all environments (Fig. 4h, i). These results further confirm the genetic additive effects of Tof16 and J revealed from their interactions in NILs (Fig. 3b–d). Therefore, the selection of natural mutations of both genes has played substantial roles in expanding soybean cultivation from its temperate origins into the tropics and has facilitated the improvement of soybean adaptation and yield in the tropics.

Discussion
Soybean production in low-latitude regions already accounts for approximately half of the world’s total production. Understanding the genetic and genomic basis of soybean adaptation to low latitudes will greatly facilitate the improvement of soybean productivity. In the current study, we identified the novel locus Tof16, which harbors the LHY1a gene, and determined that it contributes to geographical adaptation to low latitude regions and improves soybean yields. Consistent with our previous study, we confirmed that Tof16 depends genetically on the legume-specific flowering repressor E1 and that Tof16 directly binds the promoter of E1 to repress its transcription, which in turn releases its transcriptional suppression of two key soybean FT homologs (Fig. 1g, h and Supplementary Figs. 10 and 11).

The adaptation of soybean to the tropics also depends on selection for natural mutations of J, a homolog of the circadian clock gene ELF3. J also directly binds to the promoter of E1 to repress its transcription. A recent report indicates that J associates with two LUX homologs to form the evening complex, which binds to the E1 promoter through the LUX MYB binding domain. The two LUX homologs redundantly control soybean flowering, as the single mutants show no phenotypes but the lux1 lux2 double mutant exhibits the same extremely late flowering under both SD and LD conditions, suggesting that the evening complex plays key roles in photoperiodic flowering and photo-period sensitivity in soybean.

Two other circadian clock genes, Tof11 and Tof12, which are homologs of PRR3, play essential roles in soybean domestication and in cultivars adapted to high latitudes. Both Tof11 and Tof12 control flowering via E1, but they do this indirectly by inducing E1 transcription by binding to the promoters of LHY homologs. Collectively, these findings strongly suggest that the legume-specific transcription factor E1 plays critical roles in photoperiodic flowering, thus controlling adaptation and yield development in soybean. These findings also indicate that circadian clock genes act upstream of E1 to regulate its transcription. Therefore, dissecting the functions and elucidating the genetic networks of circadian clock genes would increase our understanding of photoperiodic flowering and adaptation in soybean.

Decoding the genomic basis of the natural variation of flowering genes is central to understanding crop adaptation and yield improvement. Loss-of-function alleles of J were successfully introduced into soybean cultivars in central-western Brazil, which subsequently enabled the expansion of soybean production to regions below 15° latitude and even as far as the equator. Our analysis of the selection of natural variations of Tof16 and J sheds light on how soybean adapted from its temperate mid-latitude origins to low latitudes. Both tof16 and j (tof16-2 and j-11) are relatively weak alleles encoding proteins with amino acid substitutions that occurred in wild soybean from the Huanghuai region in China, the center of origin of soybean (Fig. 4a, d, Supplementary Figs. 22–25). During domestication, these two alleles were retained and passed onto the landraces in this region (Fig. 5 and Supplementary Data 2). When the landraces disseminated into low-latitude regions, these two alleles were first selected due to their partial functional impairments, which delay flowering at low latitudes (Fig. 5). However, this delayed flowering was not sufficient to meet the requirements for complete adaptation and yield development. Consequently, more complete loss-of-function mutations occurred on top of the two earlier weak alleles and were intensively selected and incorporated into breeding programs to further improve plant adaptation and yield in the tropics (Fig. 4a–f).

The interesting stepwise selection process of soybean adaptation to the tropics was not identified previously due to the lack of available whole-genome sequences of large groups of accessions. Stepwise selection of flowering-time genes was also identified for Tof12 and Tof11 homologs during the adaptation of soybean to high latitudes. This led to the gradual loss of photoperiod sensitivity to facilitate growth during the short growing season and to synchronize the harvest. Stepwise selection of different cis-regulatory variants in an FT homolog also occurred at different evolutionary times for local adaptation in maize. A recent study also indicated that stepwise selection of a CONSTANS LIKE gene and a Phytochrome A gene facilitated the adaptation of common bean to high latitudes. These findings suggest that the gradual selection of different genes or different alleles of one gene is a common strategy during crop evolution and adaptation.
Genome-editing techniques have been developed to introduce precise, predictable genome modifications into plants in order to obtain the desired traits. These techniques are giving rise to precision breeding techniques that are defining the next-generation of plant breeding24–26. CRISPR/Cas9-mediated genome editing has been widely used to create novel allelic variation in plants27,28. For example, new alleles developed by gene editing of the MADS-box genes J2, EJ2, and LIN in tomato have allowed in fluorescence architecture to be optimized, leading to improved yields29. In recent years, several cases of multiplex gene editing have drastically accelerated gene stacking for important traits. In a recent example, 13 rice pyrabactin resistance 1 (PYR1)/PYR1-like (PYL) family genes were edited by CRISPR/Cas9. Characterization of the combinatorial mutants suggested that pyl1/4/6 exhibited the best growth and improved grain productivity in natural paddy field conditions30. Here, we generated all combinatorial null mutants of four LHY homologs by CRISPR/Cas9 and found that lhy1a/1b/2b exhibited the most improved grain yield under natural LD conditions (Supplementary Fig. 14). Perhaps combining natural variation alleles of the four LHY homologs could also improve soybean productivity (Fig. 3h).

### Methods

**Resequencing, mapping, and variation calling.** The libraries for each accession of 329 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 329 accessions sequenced in this study and the 1295 previously sequenced accessions mapping and SNP calling were performed as described previously14. In brief, paired-end resequencing reads were mapped to the Williams 82 genome (Wm82.a2.v1) with BWA 0.6.1-r104 software with default parameters31. The SNPs and indels were called with GATK (ver. 3.1.1)32,33 and SAMtools (ver. 0.1.19) software34, independently; then, 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 SNP and INDEL were performed based on the Williams 82 genome using snpEff (ver. 3.1) software35. 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.

**Population genetic analysis.** To conduct phylogenetic analysis, SNPs with MAF < 0.05 were filtered out for all soybean accessions. The remaining SNPs were used to construct a neighbor-joining tree with MEGA v6.06 software and were visualized with the online tool iTOL (https://itol.embl.de). PCA was performed with this SNP set with the smartpca program in the EIGENSOFT v.5.0.1 package36,37.

**Linkage disequilibrium analysis.** Linkage disequilibrium was calculated for each subpopulation with SNPs with MAF > 0.05. To perform the linkage disequilibrium calculation, plink v1.9 software was applied with the parameters (–ld-window-r2 0–ld-window 99999–ld-window-kb 1000). Linkage disequilibrium decay was calculated based on $r^2$ between two SNPs and averaged in 1-kb windows with a maximum distance of 1 Mb (ref. 38).
GWAS for flowering time. We used 4,435,213 high-quality SNPs (MAF > 0.05) to perform GWAS for flowering time in the 329 accessions. Association analyses were performed by MLM implemented in efficient mixed-model association expedited (EMMAX) software20. Kinship was derived from all of these SNPs. The threshold for GWAS was determined by Bonferroni correction (that is, corrected $P = 0.05/n$, in which $n$ is the number of independent SNPs across the genome). The significantly associated regions were manually verified from the aligned resequencing reads against the Williams genome with SAMTools v0.1.18 (ref. 34).

Soybean accessions, growth conditions, and phenotyping. The 329-accession panel was grown during the cultivation season (July to December) in 2018 and 2019 at the experimental station of the Guangzhou University in Guangzhou (23°16′N, 113°23′E). Flowering time was investigated in 2018 and 2019. For map-based cloning, one heterozygous inbred progeny population of 2418 individuals segregating at Tof16 locus was subsequently developed. NILs for the Tof16 locus were selected from F7 progeny of this same cross using molecular markers for Tof16. The heterozygous inbred progenies, NILs, and CRISPR-Cas9 knockout mutants used for phenotyping were grown under natural SD conditions in the field (day length 13 h light/11 h dark) from 2018 to 2020 at the Experimental Station of the Guangzhou University or under natural LD conditions (day length 15 h light/9 h dark) in 2020 at the Experimental Station of the Jilin Academy of Agricultural Sciences. For natural SD conditions, plants were sown in the beginning of July, spaced 0.15 m apart in rows 5 m long, with 0.7 m between rows, and harvested in November or December of each year. 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 SD conditions (day length 12 h light/12 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 stage39 were scored. The days from emergence to pods attained mature color, corresponding to the R8 stage39 were scored. Yield-related traits were recorded at the R8 stage43.

DNA isolation and map-based cloning. Genomic DNA was extracted from fresh trifoliate leaves of 2-week-old seedlings with a SurePlant DNA kit (CWBOI) and used to amplify indel markers. The primer sequences used to amplify the markers for mapping are listed in Supplementary Table 9. For fine mapping, dCAPS markers were developed in the regions of Tof16 based on the resequencing data of the two parents, PI91429 and PI628930. Seven recombinants were identified in the fine-mapping population of Tof16 using seven markers. The flowering time of the progeny of these recombinants was evaluated to delimit the genomic interval containing Tof16.

RNA extraction and quantitative RT-PCR. Total RNA was extracted using a Ultrapure RNA kit (CWBOI) and the RNA was reversely transcribed using a Super Script First-strand cDNA Synthesis System (Takara, Dalian, China). Quantitative reverse-transcription PCR (qRT-PCR) 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 Supplementary Table 5.

Plasmid construction and plant transformation. The coding sequence of Tof16 was amplified with the primer set Tof16-6HA-F/R and inserted into the XbaI and MluI I sites of the pTF101-6HA vector (containing the Bar gene for glufosinate resistance) to generate the pro35S-Tof16-6HA construct. The recombinant vector was introduced into Agrobacterium tumefaciens strain EHA101 and used to transform Williams 82 via Agrobacterium tumefaciens-mediated transformation40,41.

Transient expression assay. To generate the proEF1-LUC reporter construct, the 3147-bp promoter sequence of EF1 was amplified from Williams 82 and introduced into the pGreenII 0800-LUC vector42. The different alleles of Tof16 (Tof16, tof16-1, tof16-2, tof16-3, tof16-4) were introduced into the pTF101-6HA vector to generate the constructs pro35S-Tof16-6HA, pro35S-Tof16-1-6HA, pro35S-Tof16-2-6HA, pro35S-Tof16-3-6HA, and pro35S-Tof16-4-6HA. The different alleles of J1 (J-H1, J-H28, j-11) were introduced into the pTF101-6HA vector to generate the constructs pro35S-J1-H1-6HA, pro35S-J1-H28-6HA, and pro35S-j11-6HA. The proEF1-LUC construct was used as the reporter and various Tof16 and J1 constructs were used as the effectors in the Arabidopsis protoplast transient expression system to test whether Tof16 and J suppress the transcription of EF1.

Immunoblot analysis. To analyze protein expression in the transgenic plants, total proteins were extracted from Williams 82 and the pro35S-Tof16-6HA transgenic lines in protein extraction buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100) and protein inhibitor cocktail and used for immunoblot analysis. Immunoblot analysis was performed as described previously32. In brief, total proteins were separated by SDS-PAGE. After electrophoresis, the proteins were transferred to polyvinylidene difluoride membranes (Millipore) and probed using antibodies anti-HA antibody (ab18181, 1:5000 dilution). The anti-HA antibody (ab18181) was obtained from Abcam.

ChIP assay. Leaf samples were collected from 20-day-old seedlings at Zeitgeber time 0 under SD conditions from Williams 82 and pro35S-Tof16-6HA 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 Protein G PLUS agarose (16-201, Millipore) with anti-HA antibody (ab18181). The immunoprecipitated DNA was recovered and analyzed by qRT-PCR in triplicate. Data normalized with input transcripts are means from three biological repeats. The enrichment of the EF1$^\alpha$ genomic fragment was used as a negative control. The primers used for amplification are listed in Supplementary Table 5.

Statistical analyses. In this study, all values were presented as mean ± s.e.m. 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). For 329 accessions phenotypic evaluation, at least 10 individual plants were analyzed.

Characterization of diversity panel. Resequencing of the newly assembled 329-accession diversity panel (including 110 wild soybeans, 45 landraces, and 174 improved cultivars) generated a total of 21.8 billion paired-end reads of 100 bp in length (6.5 Tb of sequences), with an average coverage depth of more than 10x for each line. After mapping to the reference genome of soybean cultivar Williams 82 (W82)32, 31,925,948 SNPs and 3,266,897 indels (shorter than or equal to 6 bp) were identified (Supplementary Table 6). In addition, the depth of sequencing data allowed us to identify a total of 704 tandem duplications, 7160 segmental deletions, 3781 segmental insertions, and 497 inversions (Supplementary Table 6).

Genome-wide association analyses. To identify loci influencing flowering time in the 329-accession panel, we conducted a GWAS through the MLM implemented in Efficient Mixed-Model Association Expeditied (EMMAX) software30 based on the SNPs with a MAF > 0.05 and using the flowering time at GuangZhou in 2018 and 2019 for each accession.

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

Data availability. The data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The sequencing data (329 accessions) used in this study have been deposited into the NCBI database under accession number PRJNA278982. The previously reported sequence data14 were deposited into the NCBI database under accession number PRJNA394629 and the GSA database in BIG Data Center under accession number PRJCA000205 and PRJCA001691. Source data are provided with this paper.

Received: 4 June 2021; Accepted: 31 August 2021; Published online: 14 September 2021

References
1. Graham, P. H. & Vance, C. P. Legumes: importance and constraints to greater use. Plant Physiol. 131, 872–877 (2003).
2. Hymowitz, T. On the domestication of the soybean. Econ. Bot. 24, 408–421 (1970).
3. Carter, T. E. et al. Ch. 8 (American Society of Agronomy, 2004).
4. Li, Y. et al. Genetic structure and diversity of cultivated soybean (Glycine max (L.) Merr.) landraces in China. Theor. Appl. Genet. 117, 857–871 (2008).
5. Watanabe, S., Harada, K. & Abe, J. Genetic and molecular bases of photoperiod responses of flowering in soybean. Breed. Sci. 61, 531–543 (2012).
6. Neumaier, N. & James, A. T. Exploiting the long-juvenile trait to improve adaptation of soybeans to the tropics. Food Legume News, 8, 12–14 (1993).
7. Spehar, C. R. Impact of strategic genes in soybean on agricultural development in the Brazilian tropical savannah. Field Crops Res. 41, 141–146 (1993).
8. Desterro, D., Carpentieri-Pipolo, V., Kühli, R. A. S. & Almeida, L. A. Photoperiodism and genetic control of the long juvenile period in soybean: a review. Crop Breed. Appl. Biotechnol. 1, 72–92 (2001).
9. Hartwig, E. E. & Kühli, R. A. S. Identification and utilization of a delayed flowering character in soybeans for short-day conditions. Field Crops Res 2, 145–151 (1979).
10. Bonato, E. R. & Vello, N. A. E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genet. Mol. Biol.* **22**, 2 (1999).

11. Ray, J. D., Hinson, K., Mankono, J. & Malo, M. F. Genetic control of a long-juvenile trait in soybean. *Crop Sci.* **35**, 1001–1006 (1995).

12. Lu, S. et al. Natural variation at the soybean locus improves adaptation to the tropics and enhances yield. *Nat. Genet.* **49**, 773–779 (2017).

13. Fang, C. et al. A recent retrotransposon insertion of *E6* locus facilitates soybean adaptation into low latitude. *J. Integr. Plant Biol.* https://doi.org/10.1111/jipb.13034 (2020).

14. Lu, S. et al. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* **52**, 428–436 (2020).

15. Zhou, Z. et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **33**, 408–414 (2015).

16. Fang, C. et al. Rapid identification of consistent novel QTLs underlying long-juvenile trait in soybean by multiple genetic populations and genotyping-by-sequencing. *Mol. Breed.* **39**, 80 (2020).

17. Schmutz, J. et al. Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 1278–1283 (2010).

18. Xia, Z. et al. 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 (2012).

19. Cheng, Q. et al. CRISPR/Cas9-mediated targeted mutagenesis of GmLHY genes alters plant height and internode length in soybean. *BMC Plant Biol.* **19**, 562 (2020).

20. Fang, C. et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**, 161 (2017).

21. Bu, T. 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).

22. Guo, L. et al. Stepwise cis-regulatory changes in *ZCN8* contribute to maize flowering-time adaptation. *Curr. Biol.* **28**, 3005–3015 (2018).

23. González, A. M. et al. Ancient relaxation of an obligate short-day requirement in common bean through loss of a domestication gene. *Genet. Mol. Biol.* **201**, 169 (2018).

24. Mou, D. et al. Genome editing with CRISPR-Cas9 for genome engineering and precision plant breeding in agriculture. *Curr. Opin. Biotechnol.* **6058**, 667 (2018).

25. Gao, C. X. Genome engineering for crop improvement and future agriculture. *Cell* **184**, 1621–1635 (2021).

26. Chen, K. L., Wang, Y. P., Zhang, R., Zhang, H. W. & Gao, C. X. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* **70**, 667–697 (2019).

27. Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N. J. & Nekrasov, V. Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **6063**, 2078–2081 (2018).

28. Soyk, S. et al. Bypassing negative epistasis on yield in tomato imposed by a domestication gene. *Curr. Biol.* **28**, 1155.e12 (2017).

29. Miao, C. B. et al. Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity. *Proc. Natl Acad. Sci. USA* **115**, 6058–6063 (2018).

30. Shen, Y. et al. De novo assembly of a Chinese soybean genome. *Sci. China Life Sci.* **61**, 871–884 (2018).

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

32. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

33. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

34. Cingolani, P. et al. 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 (2012).

35. Felsenstein, J. *PHYLIP-Phylogeny Inference Package* (version 3.2). *Cladistics* **5**, 164–166 (1989).

36. 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).

37. Fehr, W. R. & Caviness, C. E. Stages of Soybean Development Special Report (Iowa State Univ., 1977).

38. Nan, H. et al. GmFT2a and GmFT5a redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. *PLoS ONE* **9**, e97669 (2014).

39. Li, X. et al. Overcoming the genetic compensation response of soybeanflorigens to improve adaptation and yield at low latitudes. *Curr. Biol.* https://doi.org/10.1016/j.cub.2021.06.037 (2021).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32090064 and 31725021 to F.K., 31901568 to Q.C., 32022062 to S.L., 32015058 to L.D., 32001530 to K.W., 32010503 to C.F.), and also supported by the Major Program of Guangdong Basic and Applied Research (Grant no. 2019B03032006 to F.K. and B.L.).

Author contributions

F.K. designed and interpreted the results. F.K. and B.L. supervised the experiments and coordinated the projects. L.D., C.F., Q.C., K.K., T.S., L.K., C.Z., H.L., Z.H., Y.Z., L.C., L.Y., L.W., K.Y., Z.G. and S.L. performed the experiments. L.D., C.F., Q.C., H.L., X.Y., J.L.W., S.L. and F.K. performed the data analysis. F.K., J.L.W. and B.L. were the corresponding authors. F.K. and B.L. wrote the manuscript.

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/s41467-021-25800-3.

**Correspondence** and requests for materials should be addressed to James L. Weller, Sijia Lu, Fanjiang Kong or Baohui Liu.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contributions to the peer review of this work.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2021