nature portfolio

Peer Review File

Natural variation in histone demethylase GmLDL2 confers flowering time divergence for geographical expansion in soybean

Corresponding Author: Professor Yingxiang Wang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This article employs a forward approach to pinpoint a gene, LDL2, that influences the flowering time and latitudinal adaptability of soybeans. This gene possesses demethylase activity towards H3K4me1 and H3K4me2, and through the application of ChIP and RNA-seq techniques, it has identified a downstream gene of LDL2, namely FERONIA. LDL2 also exhibits two haplotypes, each adapted to high-latitude and low-latitude regions, respectively. The LDL2-FERONIA signaling pathway for regulating flowering that the article has identified represents a relatively novel pathway in soybeans,, and the article's line of thought is clear. I believe this article deserves to be published, but I do have a few minor questions.

As you mentioned in the discussion section, FER is involved in many signaling pathways in Arabidopsis. However, in this study, the LDL2 mutation leads to the upregulation of FER transcription and results only in an early flowering phenotype. I would like to know if the overexpression or knocking out FER in soybean plants also leads solely to an early flowering phenotype, or if there are other phenotypes observed. Additionally, could it be that FER has several paralogs in soybean that play roles in regulating other aspects of plant development?

One mutant of FERONIA leads to premature termination, while the other two cause non-synonymous mutations in the malectin domain and kinase domain, respectively. Did these two mutations both result in changes to particularly important amino acids within the domains? The phenotypes of these three lines appear to be identical.

Typo:
On page 2, line 43, it should not be "anef1", right? It should be "ef1"?

Reviewer #2

(Remarks to the Author)
In this study, Liu et al. identified GmLDL2 that control flowering time through map-based cloning using an early flowering mutant (ef1). They found that GmLDL2 has demethylase activity towards H3K4me1 and H3K4me2 and directly inhibits the expression of GmFER. They verified the function of GmLDL2 and GmFER via transgenic analyses. They also found that natural variation of GmLDL2 may be related to the soybean adaptation. Although the function of these two genes in controlling flowering time have been disclosed in Arabidopsis, this study provides more information of the natural variations and its possible role in regulating flowering time and affecting adaptation, which may facilitate the breeding of varieties adapted to different latitudes.
Overall, the results are logically. To make the story more complete and solid, some additional evidences are needed.
1. To confirm the function of GmLDL2, a complementary test using the mutant as receptor is better to provided.
2. The authors claimed that the two haplotypes of GmLDL2 may show varied function. The authors used different accessions to test the differences in flowering time. Because the variations of many genes control flowering time in soybean, and GmLDL2 is not a major QTL. How the authors eliminate the effect of natural variations of other genes? All these accessions have the same genetic backgrounds?
To make the conclusion solid, more evidences are needed. For example, the use of pair of isogenic line, or to do the complementary test using the mutant.

3. Similarly, in the proposed model, the authors claimed that the two haplotypes of GmLDL2 have different effect on GmFER. This also need more evidence, to test the changes of GmFER in histone modification and transcription.
4. Regarding GmFER is the solo target of GmLDL2, I can not understand well. Since GmLDL2 displays histone demethylase activity towards H3K4 mono- and di-ethylation globally in the genome, why it only target a solo gene? Does that mean GmFER is the gene that have the histone demethylase activity? I do not think so.
Minor comments:
1. Line 268, 'further supporting the idea that GmLDL2 may only affect H3K4 methylation at specific loci', "idea" better to be "hypothesis";
2. The authors transformed the construction of GmFER fused to 3×Flag tags driven by the CaMV 35S promoter (35Spro:GmFER-Flag) into wild type (Col) Arabidopsis plant. This experiment is not necessary, as the gene function has been clearly revealed in Arabidopsis.
3. Line 369: 'These results suggested that GmLDL2 is likely under strong natural selection during domestication', is it "natural selection" or "artificial selection"?
4. The discussion has too much redundancy with the results, much of them can be compressed.

Reviewer #3

(Remarks to the Author)
In this manuscript, the authors isolated an early flowering mutant (ef1) in soybean, and revealed the putative causal gene for this mutant was the lysine-specific demethylase like 2 (GmLDL2). They further demonstrated that the encoding protein of GmLDL2 had a demethylase activity towards H3K4me1 and H3K4me2. Moreover, GmLDL2 negatively affected the expression of GmFER, and the function of GmFER in flowering time was revealed. They also performed natural variation analysis, and revealed that GmLDL2H1 and GmLDL2H2 haplotypes displayed distinct demethylase activity, and the two haplotypes respectively facilitated the adaptation of soybeans to low- and high-latitude regions. These findings provide new data and useful information for the understanding of flower time control and regional adaptability of soybeans. However, several major concerns are arisen during my review, and the manuscript has to be significantly modified.

Major concerns:
No complementary experiment was conducted for the conclusion that "GmLDL2 is the causal gene for the ef1 mutant". This is a serious issue.

The authors claimed that they established "the GmLDL2-GmFER module in regulating soybean flowering". However, the evidence is not sufficient for this at all. The author indeed generated overexpression GmFER in Arabidopsis and soybean lines, and analyzed three independent soybean mutant lines of GmFER: NJAU0800, NJAU1719, and NJAU1817 to support the role of GmFER in flowering time. However, the phenotypic variation in flowering time in these three so-called independent mutant lines was not fully supported by the detected mutations in GmFER. They should cross reciprocally them to observe the phenotype of F1. Most importantly, the genetic interaction between GmLDL2 and GmFER should be one of the key evidence to propose the GmLDL2-GmFER module. No such evidence is provided, although GmLDL2 regulates target gene GmFER transcription via modulating H3K4 methylation. The genetic interaction should be investigated for the main conclusion in this work. In addition, I would like to suggest that the over-expression of Arabidopsis work may be presented in Supplementary data, while the work on three independent soybean mutant lines of GmFER could be provided as Figure in main text.

In the natural population analysis, the provided evidence is insufficient to make any solid statements.
Line 354-357, the authors claimed that "Extremely high Tajima's D values suggested that the landrace population may shrink as a result of balancing selection. These results suggest that GmLDL2 was selected during soybean domestication, leading to different alleles rapid accumulation in landraces and cultivars." Why did the authors make the conclusion that "GmLDL2 was selected during soybean domestication" based on these limited information? Please clarify the statement from the variation of Tajima's D and nucleotide diversity (π). Do these analyses and results really support this statement?
Line 364-366, the authors claimed that "The frequency of GmLDL2H1 allele was present in 87.1% of wild soybeans, whereas GmLDL2H2 was only 7.9%, indicating that GmLDL2H1 probably was the original haplotype in soybean ancestors." I do not think that the frequency of haplotypes is directly linked to their ancestral haplotypes. Please provide direct evidences for this statement on the ancestral state.
In Fig. 7c, do the accessions in the pie chart include wild soybeans? I would like to suggest to separately indicate wild soybeans, landraces, and cultivated soybeans. Moreover, why are the accessions from Huanghuai region of China and Northern region of China mixed together for this analysis? The geographic category of soybeans may refer to a previous work (Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010) Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol. 188: 242–253).

Minor concerns:
Line 131-135 and Fig. 1d, significance test should be performed to support the related statement.

In the transcriptomic analysis, the leaf samples at ONLY 55 DAE with two biological replicates were used. Why the authors harvested the material at this specific stage? Whether the single developmental stage is sufficient for such work? Whether two biological replicates are sufficient for such work? What kind of leaf material and from which position was harvested? These information should be clarified in the manuscript.

In view of that GmLDL2 did not affect the change of global level of H3K4me1/2 in vivo, the subtitle that “GmLDL2 displays histone demethylase activity towards H3K4 mono- and di-methylation” should be modified.

In Supplementary Fig. 3a, which period of leaves were used for expression pattern analysis?

Reviewer #4

(Remarks to the Author)
This study identified GmLDL2, a histone H3K4me1/2 demethylase as a crucial flowering time regulator in soybean by a forward genetic method. Moreover, the effect of Gmldl2 mutation on gene expression and histone methylation were determined by omics data, revealing GmFER as a critical target gene of GmLDL2. Finally, the study indicated that the GmLDL2H1 and GmLDL2H2 haploids are likely involved in the adaptation of soybeans to low and high latitude regions, respectively. These findings identified a previously unknown flowering time-associated mutant in soybean, and revealed the molecular mechanism by which how it affects flowering time and the adaptation of soybeans to different latitudes. These results will greatly contribute to understanding of the chromatin regulation of flowering time in soybean. While parts of the results in the manuscript are convincing, some others need to be validated by additional experiments to improve the manuscript.
1. While the early flowering phenotype of the ef1 mutant was well described, it is also necessary to describe whether the mutation also affect other phenotypes, especially those yield-related traits.

2. In line 137-139, 87 plants showing the flowering time comparable to the wild type and 32 plants exhibiting early flowering phenotype were identified. Because the flowering time shown in Figure 1e is consecutive, the method to divide the flowering time into wild type and mutant classes needs to be described.

3. The number of H3K4me1/2/3 ChIP-seq peaks identified in the ef1 mutant and the HX3 control was listed in the text (line255-259). I found that the number of H3K4me3 peaks in the ef1 mutant is much higher than in the HX3 control. The number of H3K4me1 and H3K4me2 peaks between two replicates showed a high level of variations. If the number is not meaningful, I suggest either remove the statement from the text or provide a plausible explanation for the variations.

4. Figure S10 showed the correlation between the two replicates of each sample. However, the analysis is not sufficient for determining the correlation between different samples. A combined PCA analysis for the correlation of all samples with two replicates needs to be performed. Even the correlation between different samples was comparable to the correlation between the replications of each sample, this analysis will provide more complete information.

5. In the ef1 mutant, the regions showing increased levels of H3K4me1 and H3K4me2 were identified (line 286-288). Although the ef1 mutant was supposed to increase H3K4me1 and H3K4me2, the regions with decreased levels of H3K4me1 and H3K4me2 should also be indicated. Further, both genes with increased and decreased levels of H3K4me1/2 should be compared with the GmLDL2 ChIP-seq data. This may reveal a higher degree of overlap between GmLDL2-bound genes and the genes showing increased levels of H3K4me1/2 than between GmLDL2-bound genes and the genes showing decreased levels of H3K4me1/2. This analysis is important for supporting the inference that GmLDL2 is involved in H3K4me1/2 demethylation.

6. The H3K4me1 and H3K4me2 ChIP-seq signals were thought to increase at GmFER gene. However, the H3K4me1 signal and the GmLDL2 signal were outside of the GmFER gene, while the H3K4me2 signal was within the GmFER gene. If GmLDL2 is not located within the GmFER gene, how it mediates H3K4me2 demethylation within the gene.

7. Moreover, at the GmFER gene, GmLDL2 ChIP-seq signal was comparable between the wild-type HX3 and the ef1 mutant, even though the highlighted region showed a little different. I am not sure whether the GmLDL2 ChIP-seq is in a high quality. Were the GmLDL2 ChIP-seq peaks repeatedly identified in two replicates?

8. GmFER-OE lines showed early flowering, while Gmfer mutant lines exhibited delayed flowering, supporting the notion that GmLDL2 is involved in regulating flowering time by affecting the expression of GmFER. However, this is not the direct evidence for supporting the inference. I wonder whether the authors tried to generate the double mutant of Gmldl2 and Gmfer by genetic crossing or CRISPR-Cas9. If GmFER functions at the downstream step of GmLDL2, the early flowering phenotype of Gmldl2 mutant is supposed to be rescued by the Gmfer mutation in the double mutant.

9. The histone demethylase activity of GmLDL2H1 and GmLDL2H2 was assessed by luminescence assays in tobacco leaves, indicating GmLDL2H1 showed a higher level of histone demethylase activity than GmLDL2H2. I feel that the in vitro histone demethylase activity assay using bacterially expressed GmLDL2 proteins (as shown in Figure 3a) is a more strong evidence for supporting the conclusion.

10. It is interesting that the geographical distributions of GmLDL2H1 and GmLDL2H2 alleles differ, and contribute to the flowering time difference between them. The soybean accessions harboring GmLDL2H1 allele showed later flowering than the varieties containing GmLDL2H2 allele. I suggest to perform a complementation test for the ef1 mutant using the

GmLDL2H1 and GmLDL2H2 transgenes. This experiment will provide a direct evidence for supporting the role of GmLDL2 H1 and H2 differentiation in adaptation.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors had addressed all my concerns and I have no further comments.

Reviewer #2

(Remarks to the Author)
Thank the authors for their efforts to address my questions.
Most of my questions are addressed, only GmLDL2 overexpression which was driven by 35S promoter in the ef1 mutant background in fact may not be the truly complementary test. However, it may prove the function of LDL2. In the future, a true complementary test may be better, as well as the functional difference test of the two hypolotypes.
I do not have further questions.

Reviewer #3

(Remarks to the Author)
I have carefully reviewed the revised version and the authors' responses, and found that the authors did their best to address the comments and concerns raised by the Reviewers. We appreciate their efforts, particularly the additional work, and the revised manuscript has been significantly improved compared to the previous version. However, several minor issues should be further clarified.
(1) In Fig. 2a and Fig. 6g, the growth conditions (SD or LD) should be stated in the corresponding legends.
(2) In Supplementary Fig. 19a, I cannot see a higher nucleotide diversity (π) in wild soybeans compared to landraces and cultivars. However, the nucleotide diversity of wild soybeans was lower than that of landraces and cultivars in the GmLDL2 region as the authors indicated by an arrow. Overall, I cannot see a signal for artificial selection for this gene. Moreover, previous work showed that domesticated soybeans retain 66% (π) and 49% (θ) of the nucleotide diversity of wild soybeans after the domestication bottleneck (Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA. 2006; 103: 16666-71.). Therefore, I am not really convinced by the presented data and results to suggest the artificial selection on GmLDL2. Please clarify the related statements.
(3) The frequency of haplotype may largely indicate its adaptability. The presented result in Supplementary Fig. 20b did not support the identification of ancestral haplotypes. A rooted tree with the outgroup haplotype may be necessary, which may provide evidence for identifying ancestral haplotypes.
(4) The authors claimed that the GmLDL2H1 and GmLDL2H2 did not show obvious geographical distribution trend in the wild and landraces accessions (Supplementary Fig. 21). I do not think that the authors accurately described the results. What I see is that the frequencies of GmLDL2H1 and GmLDL2H2 exhibited distinct geographical patterns between the wild soybeans (Supplementary Fig. 21a) and landraces (Supplementary Fig. 21b). Clarification or proper interpretation of this point is required.
(5) The results indicated that soybean accessions carrying GmLDL2H1 showed significantly later flowering time, increased plant height, and grain weight per plant compared to those carrying GmLDL2H2 under natural short-day conditions (Fig. 7h-j). Apparently, early flowering of accessions carrying GmLDL2H2 significantly decreased the soybean yield. How to understand the artificial selection on GmLDL2H2? What is the advantage of GmLDL2H2?

Reviewer #4

(Remarks to the Author)
Most of my concerns have been well addressed in the revised manuscript. Two additional results presented in the response-to-reviewer file should be incorporated into the revised manuscript. This is crucial for offering a comprehensive understanding of the function of GmLDL2.

1. Response Table 1, which shows the effect of ef1 on yield-related traits, should be presented in the Supplemental Information. This will offer a more comprehensive understanding of the biological function of GmLDL2.

2. Although the results presented in Response Figure 6 are unexpected and cannot be explained at present, they will significantly contribute to understanding the multifaceted role of GmLDL2. This is of great importance to other researchers in the field. Thus, I strongly recommend adding these results to the revised manuscript.

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
I have evaluated the revision and gone through the authors' response letter. I think that the authors have made the greatest efforts to address the concerns of the Reviewers. The manuscript has been improved. I have no further comments.

Point-by-point response to reviewers' comments
(Questions in black and responses in blue)

Dear Reviewers:

Thank you very much for taking the time to review the manuscript. We appreciate the time and effort you dedicated to providing feedback on our manuscript and are grateful for the insightful comments that will help us further improve our paper. We are pleased to resubmit our revised manuscript for your consideration. Below, please find our point-by-point response (questions in black and responses in blue).

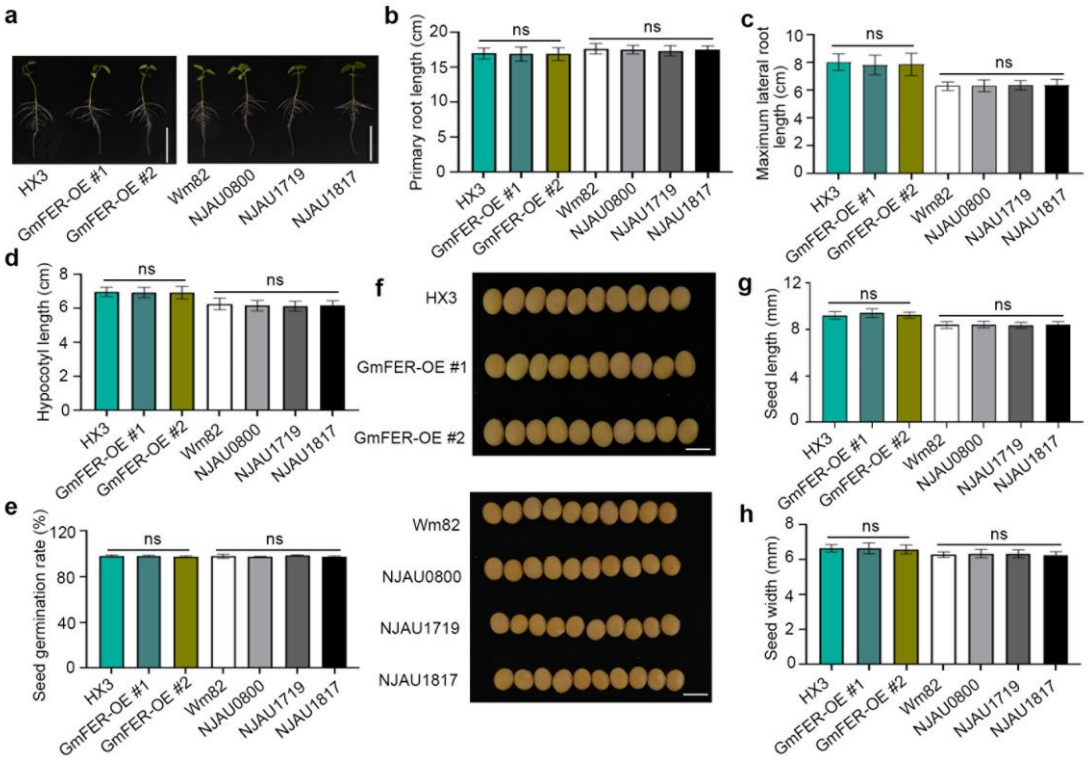
Reviewer #1 comments:

This article employs a forward approach to pinpoint a gene, LDL2, that influences the flowering time and latitudinal adaptability of soybeans. This gene possesses demethylase activity towards H3K4me1 and H3K4me2, and through the application of ChIP and RNA-seq techniques, it has identified a downstream gene of LDL2, namely FERONIA. LDL2 also exhibits two haplotypes, each adapted to high-latitude and low-latitude regions, respectively. The LDL2-FERONIA signaling pathway for regulating flowering that the article has identified represents a relatively novel pathway in soybeans, and the article's line of thought is clear. I believe this article deserves to be published, but I do have a few minor questions.

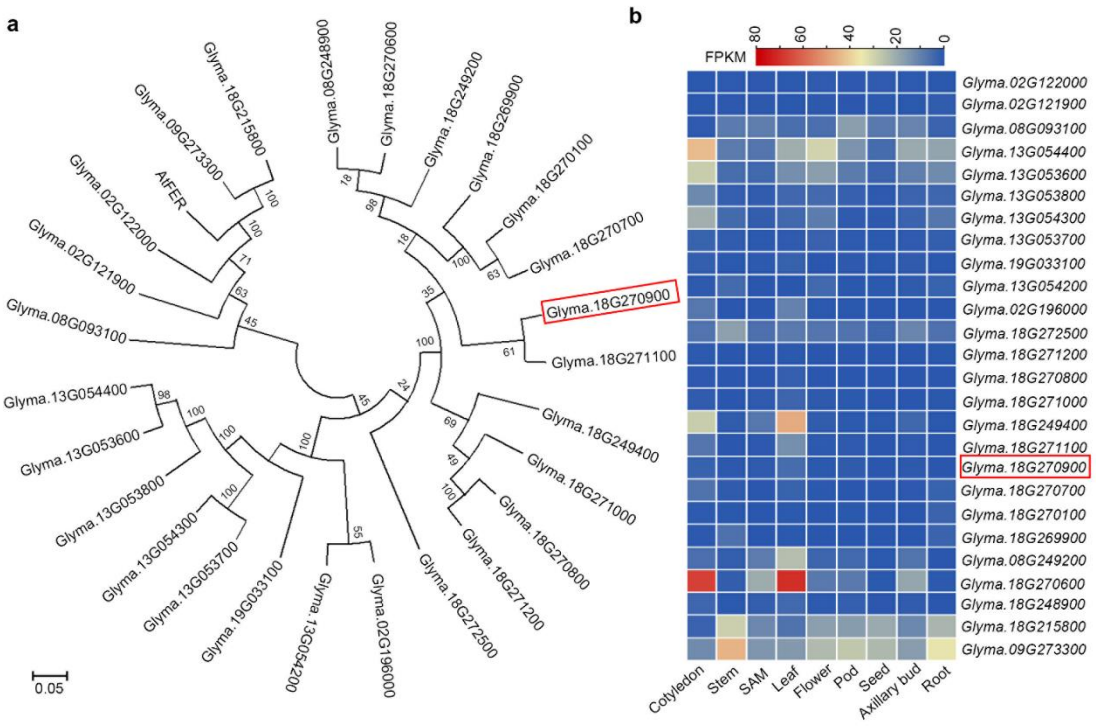
Response: We appreciate the reviewer's positive assessment of our work.

Question 1: As you mentioned in the discussion section, FER is involved in many signaling pathways in Arabidopsis. However, in this study, the LDL2 mutation leads to the upregulation of FER transcription and results only in an early flowering phenotype. I would like to know if the overexpression or knocking out FER in soybean plants also leads solely to an early flowering phenotype, or if there are other phenotypes observed. Additionally, could it be that FER has several paralogs in soybean that play roles in regulating other aspects of plant development?

Response: Thank you for this suggestion! We systematically characterized phenotypic traits in *GmFER*-overexpression and mutant lines, including root architecture (primary root length and maximal lateral root length), hypocotyl length, and seed development (size and germination rate). No statistically significant differences were observed between transgenic/mutant lines and wild type control (Response Fig. 1, see below). Phylogenetic analysis revealed that *GmFER* has 26 paralogs in soybean, which exhibit various expression patterns. For example, *Glyma.18G270600* displayed preferential expressions in cotyledons and leaves, whereas *Glyma.09G273300* showed predominant expression in stems and roots (Response Fig. 2, see below). The divergent expression profiles suggest functional diversification among *GmFER* paralogs, potentially regulating different aspects of plant development. This may explain why *GmFER* overexpression or knockout specifically alters flowering time without causing additional developmental defects in soybean.



Response Fig. 1 The other traits of GmFER-OE transgenic plants and *Gmfer* mutant lines. **a** Phenotype of 7-day-old wild type, GmFER-OE, and *Gmfer* mutant plants. Scale bars, 10 cm. **b-d** Statistics of primary root length, maximal lateral root length, and hypocotyl length of wild type, GmFER-OE, and *Gmfer* mutant plants. Error bars indicate standard deviation (n=10). Statistical analyses were performed by Student's *t*-test (ns: no significance). **e** Seed germination rate of wild type, GmFER-OE, and *Gmfer* mutant plants. The mean values are from three independent biological replicates. Statistical analyses were performed by Student's *t*-test (ns: no significance). **f** The seed size of wild type, GmFER-OE, and *Gmfer* mutant plants. Scale bars, 1 cm. **g** and **h** Seed length and width of wild type, GmFER-OE, and *Gmfer* mutant plants. Error bars indicate standard deviation (n=20). Statistical analyses were performed by Student's *t*-test (ns: no significance).

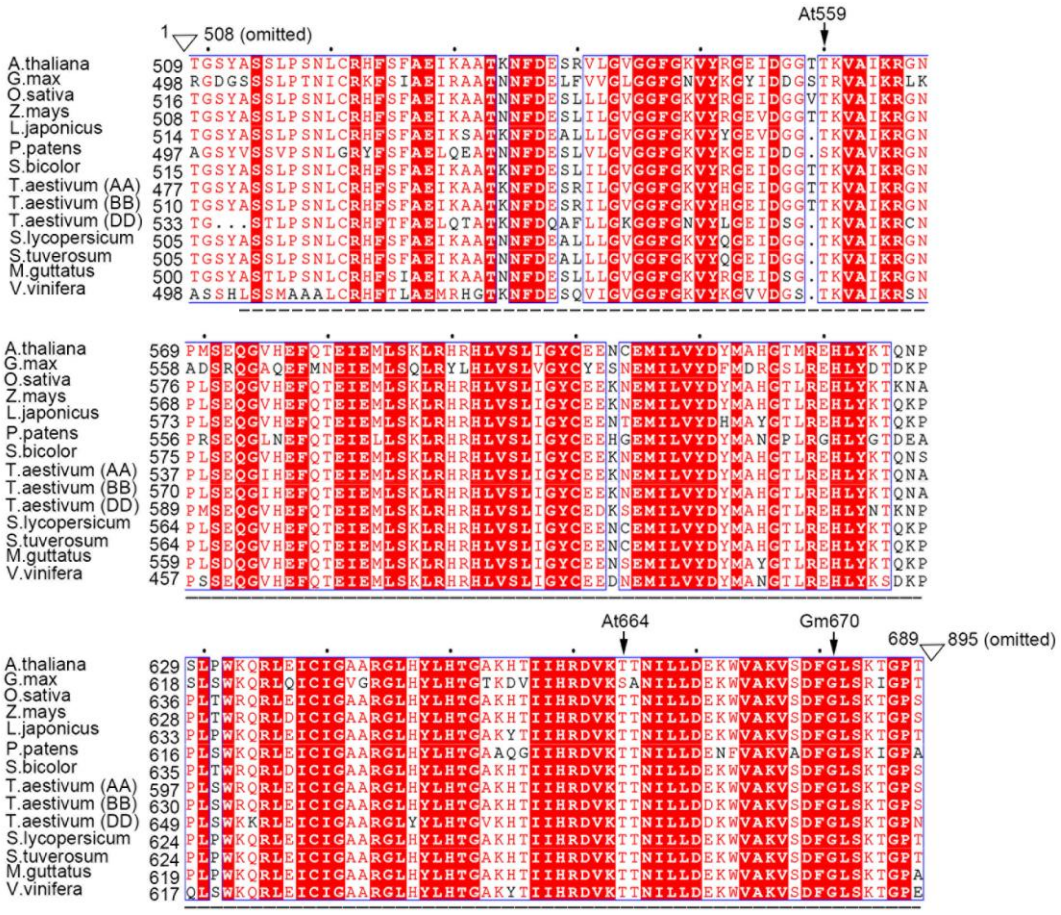


Response Fig. 2 Evolutionary and expression analysis of GmFER paralogs in soybean. **a** Phylogenetic tree of FER in *Arabidopsis* and soybean. The phylogenetic tree was constructed by the neighbor-joining method using MEGA 11.0. The bootstrap values were 1,000 replications for major branches. GmFER is highlighted by a red rectangle. **b** Heat map of expression profiles of *GmFER-like* genes in different tissues. The expression data was based on SoyOmics database. *GmFER* is highlighted by a red rectangle.

Question 2: One mutant of FERONIA leads to premature termination, while the other two cause non-synonymous mutations in the malectin domain and kinase domain, respectively. Did these two mutations both result in changes to particularly important amino acids within the domains?

Response: A previous study identified T559 and T664 as essential catalytic residues for AtFER kinase activity[1]. Our sequence alignment revealed that the two *GmFER* mutations (L404M and G670R) occur at non-homologous positions distinct from these critical sites. Notably, the glycine residue at position 670 in the kinase domain shows high evolutionary conservation among FER homologs (Response Fig. 3, see below), suggesting its potential functional importance, potentially in maintaining structural integrity or modulating catalytic activity.

The L404M mutation, located in the malectin domain, substitutes a high-hydrophobicity leucine (Kyte-Doolittle index: 3.8) with a moderately hydrophobicity methionine (index: 1.9). This substantial reduction in local hydrophobicity (Δindex = -1.9) may perturb the malectin domain's interaction interface[2,3]. Given that the malectin domain is involved in ligand binding, we hypothesize that this mutation may influence FER's affinity for secreted peptides such as RALF1 (Rapid Alkalinization Factor 1)[4] and PCP-B (POLLEN COAT PROTEIN B-class)[5], thereby modulating signal perception.



Response Fig. 3 Amino acid sequence alignment of FER homolog.
Numbers indicate amino acid residues, with similar residues highlighted in red. The kinase domain is marked by a dashed line. The triangle denotes the spacer region of AtFER.

Question 3: On page 2, line 43, it should not be "an *ef1*", right? It should be "*ef1*"?

Response: Yes, it should be "*ef1*". Thank you for your reminder.

Reviewer #2 comments:

In this study, Liu et al. identified GmLDL2 that control flowering time through map-based cloning using an early flowering mutant (ef1). They found that GmLDL2 has demethylase activity towards H3K4me1 and H3K4me2 and directly inhibits the expression of GmFER. They verified the function of GmLDL2 and GmFER via transgenic analyses. They also found that natural variation of GmLDL2 may be related to the soybean adaptation. Although the function of these two genes in controlling flowering time have been disclosed in Arabidopsis, this study provides more information of the natural variations and its possible role in regulating flowering time and affecting adaptation, which may facilitate the breeding of varieties adapted to different latitudes.

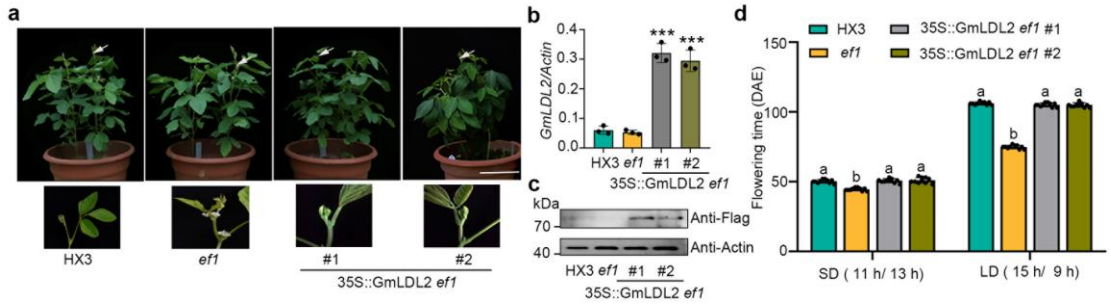
Overall, the results are logically. To make the story more complete and solid, some additional evidences are needed.

Response: We appreciate the reviewer's positive evaluation on our paper and constructive comments as below.

Question 1: To confirm the function of GmLDL2, a complementary test using the mutant as

receptor is better to provided.

Response: We agree with the reviewer that a complementary experiment would strengthen our findings. During the review and revision of the manuscript, we obtained several independent *GmLDL2* overexpression transgenic lines in the *ef1* mutant background. Based on RT-qPCR and western blotting data, we selected two representative transgenic lines exhibiting significantly delayed flowering compared to the *ef1* mutant (revised Fig. 2a, b, see Response Fig. 4 below), providing additional evidence that *GmLDL2* is the causal gene responsible for the *ef1* mutant phenotype. We have added this information in lines 173-179 in the revised manuscript.



Response Fig. 4 GmLDL2 delays flowering time in soybean.

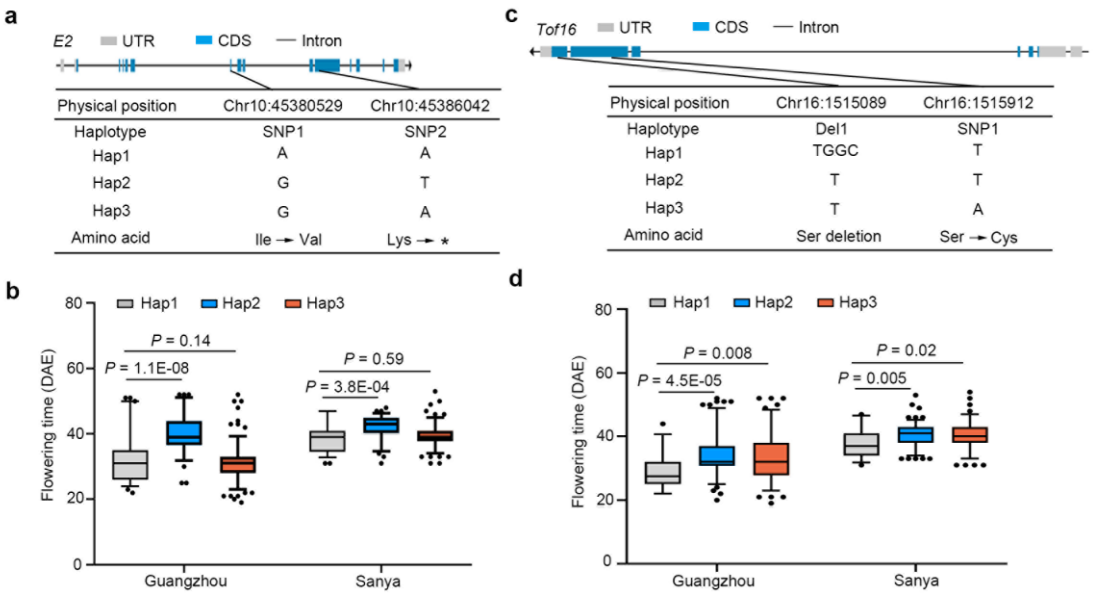
a The phenotype of whole plants and flowers of HX3, *ef1*, and two *GmLDL2* overexpressed transgenic lines. The two *GmLDL2* complementary lines showed later flowering compared to *ef1*. Scale bars, 10 cm. **b** Relative expression level of *GmLDL2* in HX3, *ef1*, and two *GmLDL2* complementary lines. *GmActin* was used as a reference control. Error bars indicate standard deviation (n=3), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$). **c** Validation of GmLDL2 expression in GmLDL2-OE transgenic plants by western blot with anti-Flag antibody. Actin serves as the loading control. **d** The flowering time of HX3, *ef1*, and two *GmLDL2* overexpressed transgenic lines under SD (11 h/13 h) and LD (15 h/9 h) conditions. Error bars indicate standard deviation (n=10), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$).

Question 2: The authors claimed that the two haplotypes of GmLDL2 may show varied function. The authors used different accessions to test the differences in flowering time. Because the variations of many genes control flowering time in soybean, and GmLDL2 is not a major QTL. How the authors eliminate the effect of natural variations of other genes? All these accessions have the same genetic backgrounds? To make the conclusion solid, more evidences are needed. For example, the use of pair of isogenic line, or to do the complementary test using the mutant.

Response: We apologize for not completely accounting for the potential effects of other genes. To address this, we systematically analyzed natural variations in known soybean flowering-time genes, including *E1*, *E2*, *E3*, *E4*, *J*, *LUXs*, and *Tof16*, across our experimental accessions. The results revealed that no coding sequence (CDS) variations were detected in *E3*, *E4*, *LUXs*, *Tof18*, *GmFT2a,* and *GmFT5a*. Then, we excluded the soybean accessions carrying known functional alleles, including the *e1-as* allele (E1 transaction variant causing early flowering)[6], the *HT1m* mutation (J splice variant conferring later flowering)[7], *Tof11-1* to *Tof11-10* (loss-of-function alleles of *Tof11*), and *Tof12-1* to *Tof12-4* (loss-of-function alleles of *Tof12*)[8]. Next, we characterized

coding sequence polymorphisms in *E2* and *Tof16*. Two exonic nonsynonymous SNPs defined three *E2* haplotypes, while three *Tof16* haplotypes were classified by one exonic nonsynonymous SNP and an in-frame deletion. To evaluate the phenotypic effects of the three haplotypes of *E2* and *Tof16*, we compared flowering time differences among different haplotypes. The results showed that haplotype 2 of *E2* flowered significantly later than the other two haplotypes, while both haplotype 2 and haplotype 3 of *Tof16* flowered later than haplotype 1 (revised Supplementary Fig. 24, see also below). Consequently, we also excluded accessions with $E2^{H2}$, $Tof16^{H2}$, or $Tof16^{H3}$. Finally, we selected 64 soybean accessions with nearly identical genetic backgrounds to assess the functions of $GmLDL2^{H1}$ and $GmLDL2^{H2}$. The results indicated that soybean accessions carrying $GmLDL2^{H1}$ showed significantly later flowering time, increased plant height, and grain weight per plant compared to those carrying $GmLDL2^{H2}$ under natural short-day conditions (revised Fig. 7h-j, see also below). We have added this information in lines 412-432 in the revised manuscript.

We agree with the reviewer that using near-isogenic lines or conducting complementary tests would indeed be an excellent approach to validating the function of $GmLDL2^{H1}$ and $GmLDL2^{H2}$. However, at this moment, we have not yet developed near-isogenic lines of different haplotypes of *GmLDL2* in our recombinant inbred line population. Additionally, the genetic transformation of soybean is also time-consuming. Nevertheless, through the multi-step genetic filtering as described above, we believe that the observed phenotypic differences of our experimental accessions are attributable to *GmLDL2* rather than background genetic variation. We hope the reviewer can be satisfied with this explanation.



Supplementary Fig. 24 Haplotypes and their flowering time of *E2* and *Tof16*.
a Haplotypes of *E2* in the CDS region. **b** The flowering time of soybean accessions carrying $E2^{H1}$, $E2^{H2}$, and $E2^{H3}$ under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. **c** Haplotypes of *Tof16* in the CDS region. **d** The flowering time of soybean accessions carrying $Tof16^{H1}$, $Tof16^{H2}$, and $Tof16^{H3}$ under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. In all the box plots, the center line indicates the median, the edges of the box represent the first and third quartiles, and the whiskers extend to the smallest and

largest data points within 1.5 interquartile ranges from the edges. Statistical significance is determined using a two-sided *t*-test.

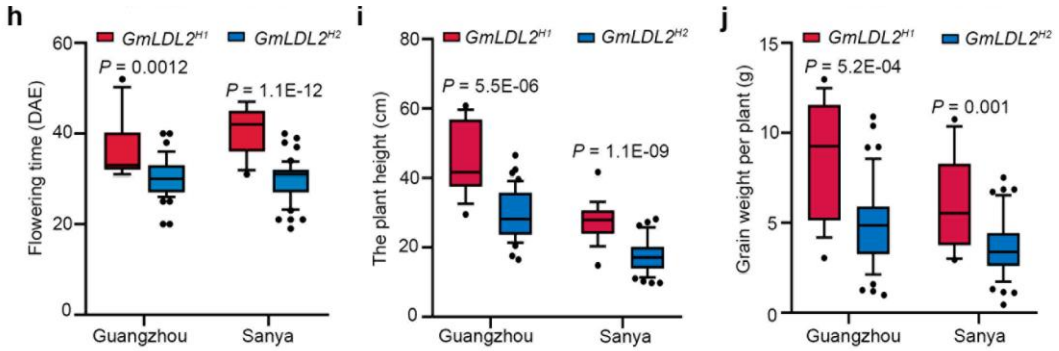
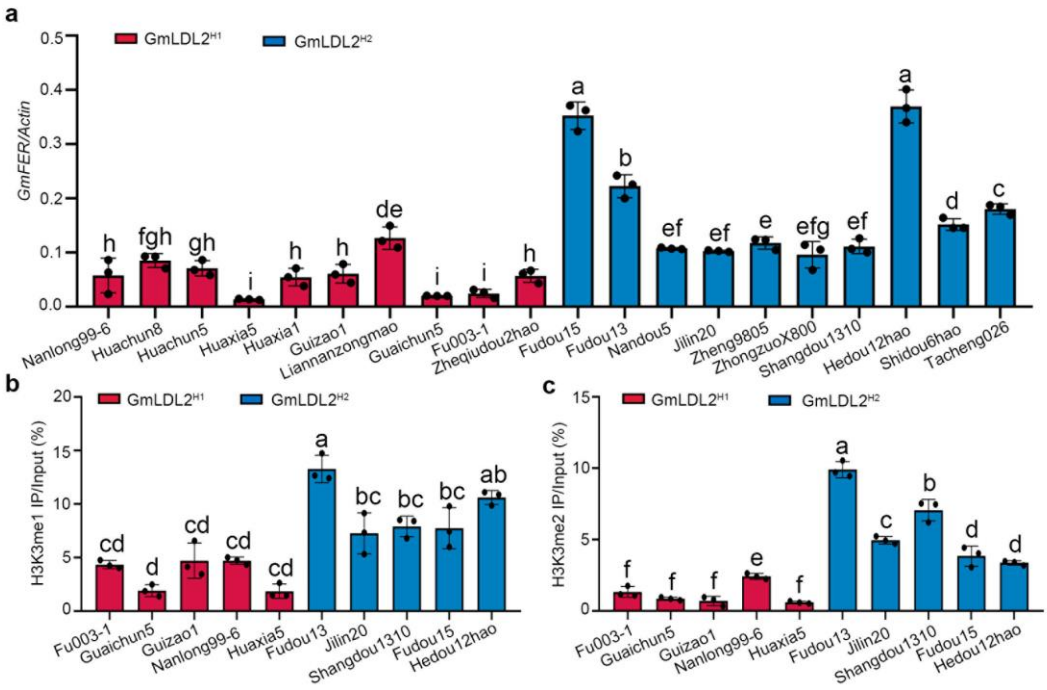


Fig. 7 GmLDL2 contributes to the latitudinal adaptation of soybean.

h-i Statistics of flowering time, plant height, and grain weight of 18 soybean accessions carrying *GmLDL2*^{H1} and 46 varieties carrying *GmLDL2*^{H2} under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. In all the box plots, the center line indicates the median, the edges of the box represent the first and third quartiles, and the whiskers extend to the smallest and largest data points within 1.5 interquartile ranges from the edges. Statistical significance is determined using a two-sided *t*-test.

Question 3: Similarly, in the proposed model, the authors claimed that the two haplotypes of GmLDL2 have different effect on GmFER. This also need more evidence, to test the changes of GmFER in histone modification and transcription.

Response: Based on the reviewer's suggestion, we examined the expression levels by RT-qPCR and H3K4me1/2 enrichment by ChIP-qPCR of *GmFER* in soybean accessions with different haplotypes of *GmLDL2*. The results showed that varieties carrying *GmLDL2*^{H2} exhibited significantly higher H3K4me1/2 enrichment and transcription levels of *GmFER* compared to those carrying *GmLDL2*^{H1} (revised Supplementary Fig. 25, see also below), providing robust evidence to support the conclusion that the two haplotypes of *GmLDL2* have different effects on *GmFER*. We have added this information in lines 433-440 in the revised manuscript.



Supplementary Fig. 25 The expression and H3K4me1/2 levels of *GmFER* in *GmLDL2^{H1}* and *GmLDL2^{H2}* accessions.

a qRT-PCR analysis of *GmFER* expression level in soybean accessions carrying *GmLDL2^{H1}* and soybean accessions carrying *GmLDL2^{H2}* at ZT4 in SD (11 h/ 13 h) conditions. *GmActin* was used as an internal control. Error bars indicate standard deviation (n=3), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$). **b** and **c** ChIP-qPCR analysis of H3K4me1 and H3K4me2 methylation status at the *GmFER* locus. The mean values are from three independent biological replicates, and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$).

Question 4: Regarding GmFER is the solo target of GmLDL2, I can not understand well. Since GmLDL2 displays histone demethylase activity towards H3K4 mono- and di-ethylation globally in the genome, why it only target a solo gene? Does that mean GmFER is the gene that have the histone demethylase activity? I do not think so.

Response: We apologize for the lack of clarity in our original description. We agree with the reviewer that *GmFER* is probably not the solo target of GmLDL2 across the genome. In this paper, we would like to provide the following explanation. (1) As mentioned in the text, due to functional redundancy with other family members, GmLDL2 does not appear to influence the global levels of H3K4me1/2 across the genome (Fig. 5a, Supplementary Fig. 6, revised manuscript). Instead, we speculate that GmLDL2 may only modulate the H3K4me1/2 methylation levels at specific loci. (2) Considering the complexity of epigenetic regulation, a single gene is often subject to multiple layers of modifications, such as DNA methylation, histone modifications, and chromatin remodeling, acting in combination to regulate its expression. GmLDL2 modulates the H3K4me1/2 methylation level of target genes, whereas the ultimate expression level of these genes depends on the integrated effects of all participating epigenetic modifications. (3) To identify the downstream

targets of GmLDL2 in regulating soybean flowering, the samples of RNA-seq and ChIP-seq were harvested under standard growth conditions. It should be noted that GmLDL2, as an epigenetic regulator, its expression is induced by various environmental stresses including temperature, low Pi, and salinity[9]. We hypothesize that the GmLDL2-targeted genes are most likely involved in those stress responses, yet their expression remains unaltered under standard growth conditions. Furthermore, we performed RNA-seq analysis on two-week-old leaves of both the *ef1* mutant and wild type plants under low phosphate and low temperature conditions. Compared to normal growth conditions, a significantly higher number of differentially expressed genes (DEGs) was observed in *ef1* under low temperature stress, with 4,901 genes upregulated and 5,611 downregulated, whereas the wild type showed 3,859 upregulated and 2,849 downregulated DEGs. Similarly, under low phosphate conditions, *ef1* exhibited 1,103 upregulated and 2,655 downregulated DEGs, in contrast to the wild type with only 637 upregulated and 141 downregulated genes. These results further support our hypothesis that *ef1* undergoes more extensive transcriptional reprogramming in response to environmental stress. Due to these data obviously beyond the scope of this study, so we did not include these data in the revised manuscript.

As a result, compared to the wild type, we observed differential expression of only a limited number of genes in the *ef1* mutant under standard growth conditions, with 440 genes upregulated and 372 downregulated (Fig. 4a, revised manuscript). Through integrated analysis of RNA-seq and ChIP-seq data, we found that the only significantly upregulated gene, *GmFER*, overlapped within the regions showing both extremely strong increases in H3K4me1/2 signals and anti-GmLDL2 occupancy, making it the most likely candidate target of GmLDL2 in regulating soybean flowering. This is also supported by the observation that the *ef1* mutant only shows a flowering time phenotype under normal conditions. We hope this clarification addresses the reviewer's concerns.

Question 5: Line 268, 'further supporting the idea that GmLDL2 may only affect H3K4 methylation at specific loci', "idea" better to be "hypothesis".
Response: Thanks for your suggestion, we have replaced "idea" with "hypothesis" in the revised manuscript (lines 269-270, revised manuscript).

Question 6: The authors transformed the construction of GmFER fused to 3×Flag tags driven by the CaMV 35S promoter (35Spro:GmFER-Flag) into wild type (Col) Arabidopsis plant. This experiment is not necessary, as the gene function has been clearly revealed in Arabidopsis.
Response: Thank you! As suggested by Reviewer #3, we have presented this work in Supplementary Fig. 16 (revised manuscript). If the reviewer does not accept it, we will remove it in the next version.

Question 7: Line 369: 'These results suggested that GmLDL2 is likely under strong natural selection during domestication', is it "natural selection" or "artificial selection"?
Response: We are sorry for this. Yes, you are right, it is artificial selection. We have replaced "natural selection" with "artificial selection" in the revised manuscript (line 387, revised manuscript).

Question 8: The discussion has too much redundancy with the results, much of them can be compressed.

Response: According to your advice, we have made some adjustments and improved the discussion section in our revised manuscript (lines 500-509, 512-516).

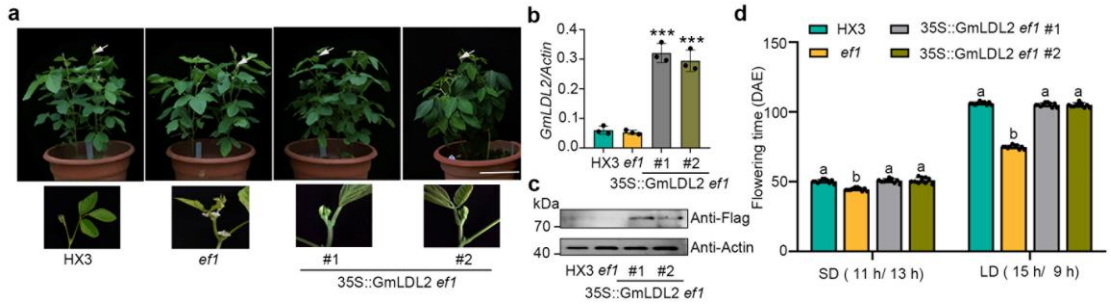
Reviewer #3 comments:

In this manuscript, the authors isolated an early flowering mutant (ef1) in soybean, and revealed the putative causal gene for this mutant was the lysine-specific demethylase like 2 (GmLDL2). They further demonstrated that the encoding protein of GmLDL2 had a demethylase activity towards H3K4me1 and H3K4me2. Moreover, GmLDL2 negatively affected the expression of GmFER, and the function of GmFER in flowering time was revealed. They also performed natural variation analysis, and revealed that GmLDL2^{H1} and GmLDL2^{H2} haplotypes displayed distinct demethylase activity, and the two haplotypes respectively facilitated the adaptation of soybeans to low- and high-latitude regions. These findings provide new data and useful information for the understanding of flower time control and regional adaptability of soybeans. However, several major concerns are arisen during my review, and the manuscript has to be significantly modified.

Response: Thank you so much for your encouragement and valuable comments, which are very helpful for us to improve the work.

Question 1: No complementary experiment was conducted for the conclusion that "GmLDL2 is the causal gene for the ef1 mutant". This is a serious issue.

Response: We fully acknowledge the reviewer's suggestion regarding complementary validation, which was also raised by Reviewer #2. During the review and revision processes, we obtained several independent *GmLDL2* overexpression transgenic lines in the *ef1* mutant background. Based on RT-qPCR and western blotting data, we selected two representative transgenic lines exhibiting significantly delayed flowering, similar to wild type control (revised Fig. 2a, b, see Response Fig. 4 below), providing additional evidence that *GmLDL2* is the causal gene for the *ef1* mutant phenotype. We have added this information in lines 173-179 in the revised manuscript.



Response Fig. 4 GmLDL2 delays flowering time in soybean.

a The phenotype of whole plants and flowers of HX3, *ef1*, and two *GmLDL2* overexpressed transgenic lines. The two *GmLDL2* complementary lines showed later flowering compared to *ef1*. Scale bars, 10 cm. **b** Relative expression level of *GmLDL2* in HX3, *ef1*, and two *GmLDL2* complementary lines. *GmActin* was used as a reference control. Error bars indicate standard deviation (n=3), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$). **c** Validation of GmLDL2 expression in GmLDL2-OE transgenic plants by western blot

with anti-Flag antibody. Actin serves as the loading control. **d** The flowering time of HX3, *ef1*, and two *GmLDL2* overexpressed transgenic lines under SD (11 h/13 h) and LD (15 h/9 h) conditions. Error bars indicate standard deviation (n=10), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$).

Question 2: The authors claimed that they established "the GmLDL2-GmFER module in regulating soybean flowering". However, the evidence is not sufficient for this at all. The author indeed generated overexpression GmFER in Arabidopsis and soybean lines, and analyzed three independent soybean mutant lines of GmFER: NJAU0800, NJAU1719, and NJAU1817 to support the role of GmFER in flowering time. However, the phenotypic variation in flowering time in these three so-called independent mutant lines was not fully supported by the detected mutations in GmFER. They should cross reciprocally them to observe the phenotype of F_1. Most importantly, the genetic interaction between GmLDL2 and GmFER should be one of the key evidence to propose the GmLDL2-GmFER module. No such evidence is provided, although GmLDL2 regulates target gene GmFER transcription via modulating H3K4 methylation. The genetic interaction should be investigated for the main conclusion in this work.

Response: We fully agree with the reviewer that elucidating the genetic relationship between *GmLDL2* and *GmFER* is essential for proposing the GmLDL2-GmFER module in regulating soybean flowering. To address this, we generated the *Gmldl2/fer* double mutant in the *Gmldl2* mutant background using CRISPR/Cas9 technology. Compared to the *Gmldl2* single mutant, the *Gmldl2/fer* double mutant plants displayed a significantly delayed flowering time, even later than the wild type control (revised Fig. 6e-h, see also below). The phenotype of *Gmldl2/fer* double mutant closely resembled that of the *Gmfer* single mutant (revised Supplementary Fig. 17), providing strong evidence to support the conclusion that *GmLDL2* regulates soybean flowering time via modulating the target gene *GmFER*. We have added this information in lines 335-350 in the revised manuscript.

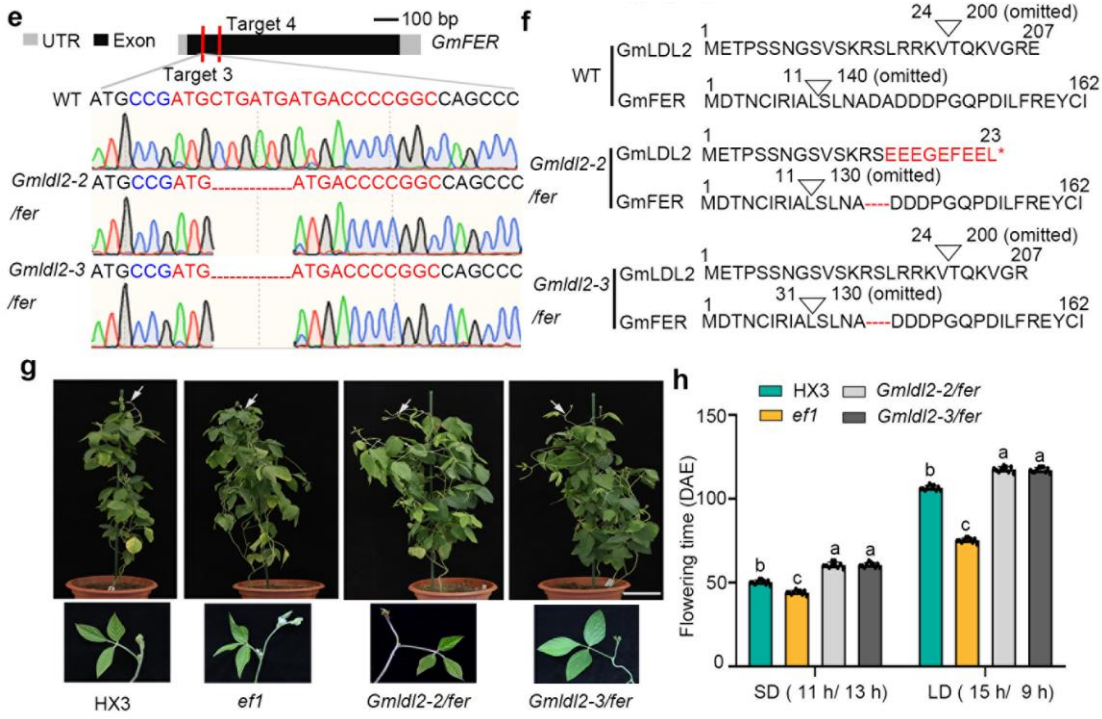


Fig. 6 GmLDL2 controls flowering in a *GmFER*-dependent manner.
e The design of sgRNAs for CRISPR/Cas9 to knock out *GmFER* in *Gmldl2* mutant background. Top: sgRNAs targeting CDS are labeled in red. Bottom: Results of Sanger sequencing of target sites in T_1 generations in HX3 and *Gmldl2/fer* double mutants. sgRNA target and PAM sequences were shown in red and blue, respectively. **f** Predicted amino acid sequences of GmLDL2 and GmFER in HX3 and *Gmldl2/fer* double mutants. Missense amino acids resulting from the frameshift caused by CRISPR/Cas9-induced mutations are shown in red. The asterisk indicates the termination of translation resulting from a stop codon. **g** The phenotypes of whole plants and flowers of HX3, *ef1*, *Gmldl2-2/fer,* and *Gmldl2-3/fer*. All *Gmldl2/fer* double mutants showed later flowering compared to *ef1*. Scale bars, 10 cm. **h** The flowering time of all varieties examined under SD (11 h/13 h) and LD (15 h/9 h) conditions. Error bars indicate standard deviation (n=10), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$).

Question 3: In addition, I would like to suggest that the over-expression of Arabidopsis work may be presented in Supplementary data, while the work on three independent soybean mutant lines of GmFER could be provided as Figure in main text.

Response: Thank you for your suggestion. We have presented the over-expression work of Arabidopsis in Supplementary Fig. 16. Additionally, given that we have obtained the *Gmldl2/fer* double mutant, the work of three independent soybean mutant lines is still presented in the Supplementary data (revised Supplementary Fig. 17).

Question 4: Line 354-357, the authors claimed that "Extremely high Tajima's D values suggested that the landrace population may shrink as a result of balancing selection. These results suggest that GmLDL2 was selected during soybean domestication, leading to different alleles rapid accumulation in landraces and cultivars." Why did the authors make the conclusion that "GmLDL2 was selected during soybean domestication" based on these limited information? Please clarify the statement from the variation of Tajima's D and nucleotide diversity (π). Do these analyses and results really support this statement?

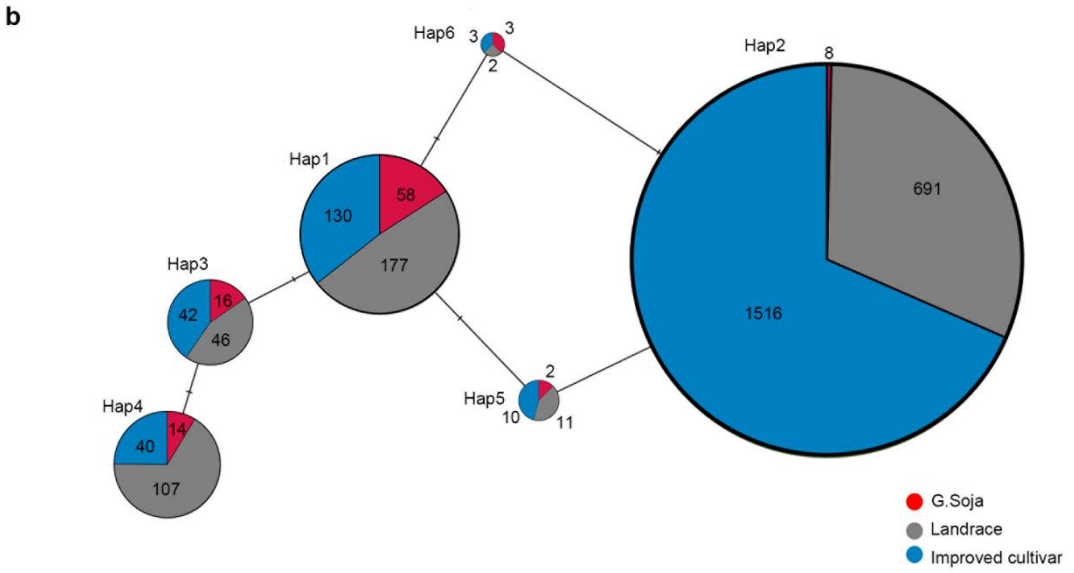
Response: We apologize for this unclear description. In this manuscript, we have estimated key population genetic parameters, including nucleotide diversity (π), Tajima's *D*, and the fixation index (F_{ST}), across the soybean genome. A detailed view of the 3 Mb genomic region surrounding *GmLDL2* is presented in Fig. 7a and Supplemental Fig. 19 in the revised manuscript. The results indicated that *GmLDL2* is located in a region with a high F_{ST} value (0.34) between wild and landrace (lines 360-361, revised manuscript), higher π value in the wild soybean than cultivar (lines 363-364, revised manuscript), and a high Tajima's *D* (4.68) in landraces (lines 366-367, revised manuscript). As originally described[10,11], F_{ST} values from 0.05 to 0.15 were taken to indicate moderate differentiation between populations, from 0.15 to 0.25 is high differentiation, and greater than 0.25 is very high differentiation. The high F_{ST} value (0.34) between wild and landrace populations in the region involved *GmLDL2*, indicating strong genetic differentiation, which is often a hallmark of local adaptation or directional selection. A higher nucleotide diversity (π) in wild soybeans compared to landraces and cultivars in this region, which suggests a reduction in diversity (a pattern commonly associated with selective sweeps or artificial selection). Moreover, Tajima's *D* value above 2 indicates an excess of common alleles suggestive of balancing selection[12]. In our analysis, the high Tajima's *D* value (4.68) in the landrace population may reflect balancing selection, population structure, or historical bottlenecks. In this context, the elevated Tajima's *D* likely reflects the maintenance of multiple alleles in landraces, possibly due

to selection for diverse stress-response traits during the early stages of domestication. Together, these results support our conclusion that *GmLDL2* underwent selection during soybean domestication.

Numerous genes have been identified as targets of selection during domestication through the analysis of these population genetic parameters, such as *OsSD6* and *OsCTB5* in rice[13,14], *ZmTB1* and *ZmTSH4* in maize[15,16], and *GmSW17*, *GmFKF1*, *Tof5, Tof11, Tof12,* and *Tof18* in soybean[8,17–20]. In more detail, the study of *GmSW17* revealed a higher π value in the wild soybean than landraces and improved cultivars, indicating an artificial selection of *GmSW17* during soybean domestication[17]. In light of these precedents, we believe that our analysis can support the corresponding results.

Question 5: Line 364-366, the authors claimed that "The frequency of GmLDL2H1 allele was present in 87.1% of wild soybeans, whereas GmLDL2H2 was only 7.9%, indicating that GmLDL2H1 probably was the original haplotype in soybean ancestors." I do not think that the frequency of haplotypes is directly linked to their ancestral haplotypes. Please provide direct evidences for this statement on the ancestral state.

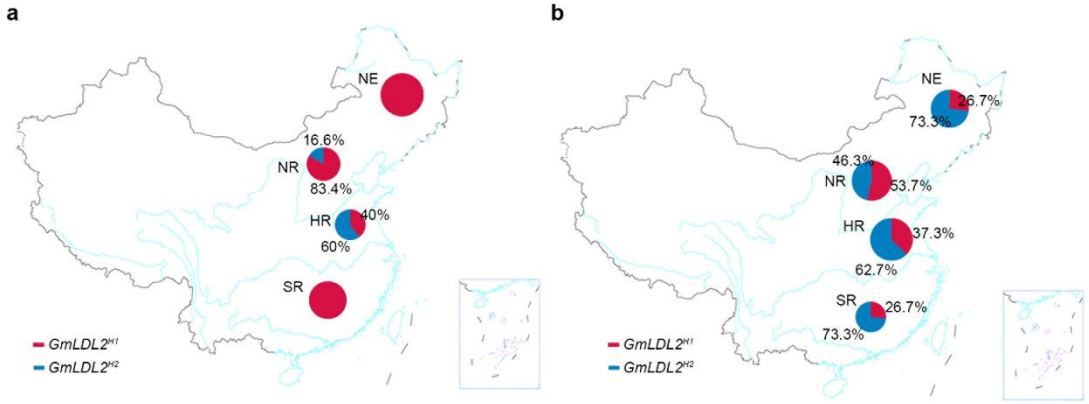
Response: We thank the reviewer for this insightful comment. We agree that haplotype frequency alone is insufficient to determine the ancestral haplotype. We further inferred the evolutionary relationships among *GmLDL2* haplotypes by constructing a minimum spanning tree (MST) using POPART software[21], based on four SNP variations in the exon regions of *GmLDL2* (revised Supplementary Fig. 20b, see also below). In the MST, the *GmLDL2*^{H1} is centrally located and connected to other haplotypes with the fewest mutational steps, suggesting it may represent the ancestral haplotype. Moreover, *GmLDL2*^{H1} is also the most frequent haplotype in wild soybeans, which is consistent with previous findings that ancestral haplotypes tend to be both frequent in wild populations and centrally located in the haplotype network[22]. We have added this information in lines 381-384 in the revised manuscript.



Supplementary Fig. 20 Haplotypes and minimum spanning tree (MST) analysis of *GmLDL2*. **b** Haplotype origins of *GmLDL2*. Red represented the wild soybeans, grey represented the landraces, and blue represented the improved cultivars.

Question 6: In Fig. 7c, do the accessions in the pie chart include wild soybeans? I would like to suggest to separately indicate wild soybeans, landraces, and cultivated soybeans. Moreover, why are the accessions from Huanghuai region of China and Northern region of China mixed together for this analysis? The geographic category of soybeans may refer to a previous work (Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010) Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol. 188: 242–253).

Response: Yes, which includes wild soybeans. As you suggested, we have separated wild soybeans, landraces, and cultivated soybeans. Furthermore, based on the previous work[23], we classified the geographic distribution of soybean accessions into four regions: Northeast, Northern, Huang-Huai, and Southern regions of China. The results indicated that in the wild soybeans and landrace, the frequencies of *GmLDL2*^{H1} and *GmLDL2*^{H2} did not exhibit distinct geographical patterns (revised Supplementary Fig. 21, see also below). In contrast, among improved cultivars, the frequency of *GmLDL2*^{H1} increased from the Northeast to the Southern region, while the frequency of *GmLDL2*^{H2} displayed an opposite trend, decreasing from the Northeast to the Southern region (revised Fig. 7c, see also below). We have added this information in lines 390-397 and the reference in lines 723-724 in the revised manuscript.



Supplementary Fig. 21 The geographical distributions of the wild (a) and landrace (b) soybean accessions harboring *GmLDL2*^{H1} and *GmLDL2*^{H2} across China.

NE, Northeastern region of China; NR, Northern region of China; HR, Huanghuai region of China; SR, Southern region of China.

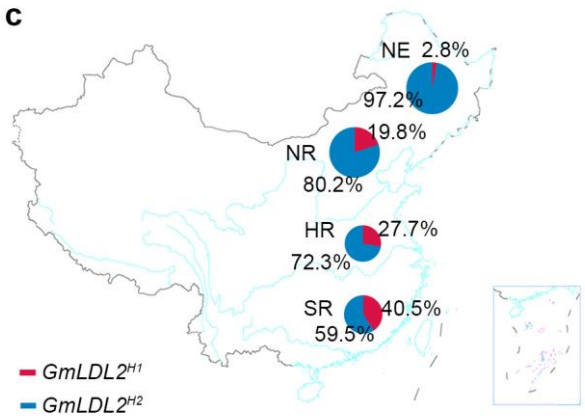


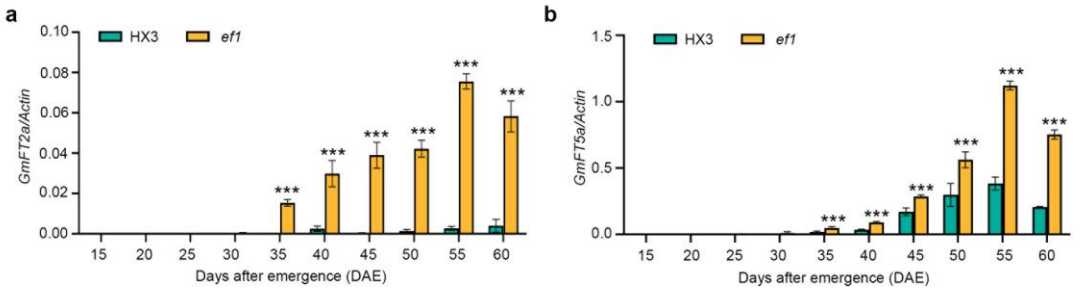
Fig. 7 GmLDL2 contributes the latitudinal adaptation of soybean.
c The geographical distributions of cultivar accessions harboring *GmLDL2*^{H1} and *GmLDL2*^{H2} across China. NE, Northeastern region of China; NR, Northern region of China; HR, Huanghuai region of China; SR, Southern region of China.

Question 7: Line 131-135 and Fig. 1d, significance test should be performed to support the related statement.
Response: Thanks for your suggestion. We have presented the significance test in the revised Fig. 1d.

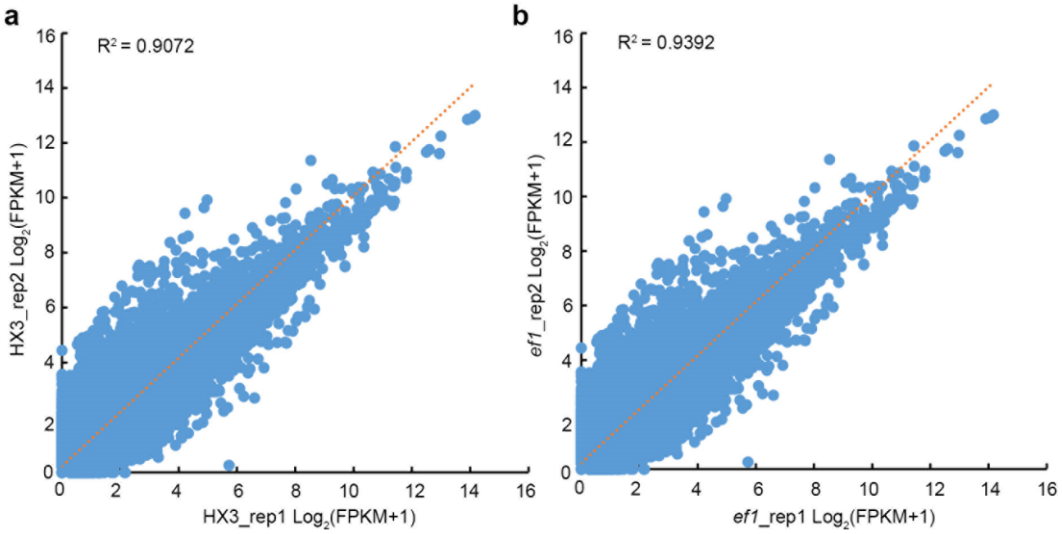
Question 8: In the transcriptomic analysis, the leaf samples at ONLY 55 DAE with two biological replicates were used. Why the authors harvested the material at this specific stage? Whether the single developmental stage is sufficient for such work? Whether two biological replicates are sufficient for such work? What kind of leaf material and from which position was harvested? These information should be clarified in the manuscript.
Response: Given that the *ef1* mutant exhibits an early flowering phenotype, we compared the expression levels of *GmFT2a* and *GmFT5a*, key flowering regulators in soybean, between *ef1* and wild type. The results showed that the expression levels of *GmFT2a* and *GmFT5a* peak at the 55-day stage; meanwhile, their expression levels in the *ef1* mutant are significantly higher than in the wild type (see Response Fig. 5 below). Based on these results, we selected leaf samples from this stage for RNA-seq analysis.

In recent studies on soybean flowering, RNA-seq samples have predominantly been collected from a single developmental stage[20,24–27]. For example, in the studies on *GmLUXs*, leaf samples for RNA-seq were harvested at 15 DAE under SD conditions[24] or 52 DAE under field conditions[25] with two biological replicates. Similarly, in the study of *GmSOC1*, samples were collected at 30 DAE under SD conditions[20]. Moreover, our RNA-seq data show a high correlation coefficient (R^2), indicating good biological reproducibility (Supplementary Fig. 7, see also below). Collectively, we consider that sampling at a single development stage with two biological replicates is sufficient for our RNA-seq analysis.

The samples for our RNA-seq are the fully expanded trifoliolate leaves at the last node, and we have added this information in the revised manuscript (lines 588-590). Thanks for your suggestion.



Response Fig. 5 The expression levels of *GmFT2a* (a) and *GmFT5a* (b) during different development stages in the *ef1* mutant and the wild type.



Supplementary Fig. 7 The correlation analysis of RNA-seq replicates.
a The Correlation Coefficient r-squared of HX3 two replicates calculated by gene expression levels. **b** The Correlation Coefficient r-squared of *ef1* two replicates calculated by gene expression levels.

Question 9: In view of that GmLDL2 did not affect the change of global level of H3K4me1/2 in vivo, the subtitle that "GmLDL2 displays histone demethylase activity towards H3K4 mono- and di-methylation" should be modified.

Response: As per your suggestion, we have modified the subtitle to "The histone demethylase activity assay of GmLDL2" (line 195, revised manuscript). Actually, the results in this section only demonstrated that *GmLDL2* encodes a histone demethylase and exhibits enzymatic activity towards H3K4me1/2 using different experimental systems assays. As said in the text, due to the potential functional redundancy with other members, GmLDL2 does not affect the global level of H3K4me1/2. In contrast, when we transiently overexpressed GmLDL2 in tobacco leaves, we can clearly observe the reduction of H3K4me1/2 methylation level in the presence of GmLDL2-expressed nuclei, suggesting that GmLDL2 has histone demethylase activity *in vivo*. Furthermore, the *in vitro* histone demethylase activity assay using bacterially expressed GmLDL2 proteins confirmed that GmLDL2 exhibits demethylation activity toward H3K4me1/2. Taken together, we concluded that GmLDL2 has a demethylase activity towards H3K4me1/2.

Question 10: In Supplementary Fig. 3a, which period of leaves were used for expression pattern analysis?

Response: We are sorry for omitting this information. Leaves, shoot apical meristem (SAM), stem, and root were collected at 20 DAE. Flowers were harvested at the R1 stage, while pods and nodules were collected at the R4 stage, which has been added to the revised manuscript (lines 577-579).

Reviewer #4 comments:

This study identified GmLDL2, a histone H3K4me1/2 demethylase as a crucial flowering time regulator in soybean by a forward genetic method. Moreover, the effect of Gmldl2 mutation on gene expression and histone methylation were determined by omics data, revealing GmFER as a critical target gene of GmLDL2. Finally, the study indicated that the GmLDL2H1 and GmLDL2H2 haploids are likely involved in the adaptation of soybeans to low and high latitude regions, respectively. These findings identified a previously unknown flowering time-associated mutant in soybean, and revealed the molecular mechanism by which how it affects flowering time and the adaptation of soybeans to different latitudes. These results will greatly contribute to understanding of the chromatin regulation of flowering time in soybean. While parts of the results in the manuscript are convincing, some others need to be validated by additional experiments to improve the manuscript.

Response: We appreciate the reviewer's positive assessment of our manuscript and have considered their comments below.

Question 1: While the early flowering phenotype of the ef1 mutant was well described, it is also necessary to describe whether the mutation also affect other phenotypes, especially those yield-related traits.

Response: We appreciate these suggestions. We have measured several yield-related traits of the *ef1* mutant, including plant height, node number, valid branch number, pod number, one hundred seed weight, and grain weight per plant. Compared with the wild type, the mutant exhibits a slight reduction in pod number and grain weight per plant (Response Table 1, see below).

Response Table 1 The yield-related traits of wild type and *ef1* mutant.

Traits	Wild type	*ef1* mutant
Plant height	63.43±3.15	62.39 ±2.04
Nod number	21.67±0.72	20.22±0.52
Valid branch number	8.33±0.27	8.16±0.2
Pod number	102.33±7.62	98.33 ±6.63*
One hundred seed weight	14.68 ±0.62	14.82 ±0.41
Grain weight per plant	130.33±5.52	113.67±8.10*

Question 2: In line 137-139, 87 plants showing the flowering time comparable to the wild type and 32 plants exhibiting early flowering phenotype were identified. Because the flowering time shown in Figure 1e is consecutive, the method to divide the flowering time into wild type and mutant classes needs to be described.

Response: To determine whether F_2 plants (particularly those with flowering times intermediate between wild type and mutant) belong to the wild type or the mutant category, we also rechecked the flowering time of their $F_{2:3}$ progeny. If all $F_{2:3}$ plants flowered earlier than the wild type, the

corresponding F_2 line was classified as mutant, like line 4 and 36. Conversely, if the flowering time of $F_{2:3}$ progeny was variable or similar to that of the wild type, the F_2 line was classified as wild type, like line 41 and 53 (Response Table 2, see below). We have incorporated this information into the revised manuscript as per your suggestion (lines 552-556).

Response Table 2 The phenotype of partial $F_{2:3}$ individuals.

F_2 Line	Flowering time (d)	$F_{2:3}$ number	Flowering time of $F_{2:3}$(d)	The category of F_2
HX3	52±2		52±1	
ef1	41±1		40±2	
1	47	5	41±1	mutant
4	46	18	39±2	mutant
11	46	16	38±2	mutant
17	47	15	40±1	mutant
36	47	25	38±2	mutant
5	49	18	54±2	WT
34	48	35	48±6	WT
37	48	32	53±2	WT
40	48	39	49±8	WT
41	49	24	50±8	WT
47	48	19	53±3	WT
48	48	18	54±3	WT
46	49	9	50±1	WT
29	49	12	51±2	WT
58	50	22	50±7	WT
63	48	19	51±8	WT

Question 3: The number of H3K4me1/2/3 ChIP-seq peaks identified in the ef1 mutant and the HX3 control was listed in the text (line255-259). I found that the number of H3K4me3 peaks in the ef1 mutant is much higher than in the HX3 control. The number of H3K4me1 and H3K4me2 peaks between two replicates showed a high level of variations. If the number is not meaningful, I suggest either remove the statement from the text or provide a plausible explanation for the variations.

Response: Thank you. We have removed this information from the main text in the revised manuscript. The peak numbers in our results are unnecessarily required for detecting differential genomic regions for H3K4me1/2/3 signals between wild type and mutation plants (see Methods for data processing in the manuscript). However, they are meaningful for assessing data quality. We acknowledge the observed differences in H3K4me3 peak numbers between the *ef1* mutant and HX3 control, as well as the variability in H3K4me1 and H3K4me2 peaks between replicates. The discrepancy of peak numbers could arise from biological differences, technical biases, batch effects, antibody efficiency, data quality, and MACS2 peak calling parameter choices etc. For ChIP-seq data sets, a low signal-to-noise ratio (S/N ratio) can reduce peak detection accuracy (more false peaks)

and make it difficult to distinguish true binding sites from sequencing artifacts. ENCODE (Encyclopedia of DNA Elements) Consortium emphasizes the Fraction of Reads in Peaks (FRiP) score as a crucial quality control metric for ChIP-seq experiments. The FRiP score represents the proportion of mapped reads that fall within the identified peak regions, serving as an indicator of signal enrichment relative to background noise. It measures the proportion of mapped reads that fall within called peaks, reflecting signal strength relative to background noise. In the ENCODE ChIP-seq dataset, approximately 80% of FRiP scores exceed 1%, making 1% a commonly used soft threshold for evaluating ChIP-seq experiment quality, while scores below this suggest poor IP efficiency or excessive background noise. As per ENCODE guidelines, FRiP% values around 5% or higher generally reflect successful enrichment[28]. Therefore, we calculated the FRiP score to assess the quality of the ChIP-seq libraries, as shown below (Response Table 3). All libraries exhibit high data quality, except for H3K4me1, which appears to have a slightly lower FRiP score. However, this is due to H3K4me1 is a broad-peak histone mark. Broad histone modifications tend to have more diffuse binding across large genomic regions, leading to a lower proportion of reads falling within called peaks. When using MACS2 with the parameters `--broad --broad-cutoff 0.1` for histone modifications, we observed a higher FRiP score exceeding 5%, indicating strong enrichment and data quality. Moreover, we analyzed the ratio of overlapping peaks and found that 72.90% of peaks in *ef1*_H3K4me3_rep2 overlapped with those in *ef1*_H3K4me3_rep1. Although the two samples showed a large difference in peak numbers, the high overlap indicates strong replication consistency. Specifically, for the H3K4me1 mark, which initially showed a slightly low peak overlap ratio, but we observed more than 50% peak overlap when using the `--broad` parameter.

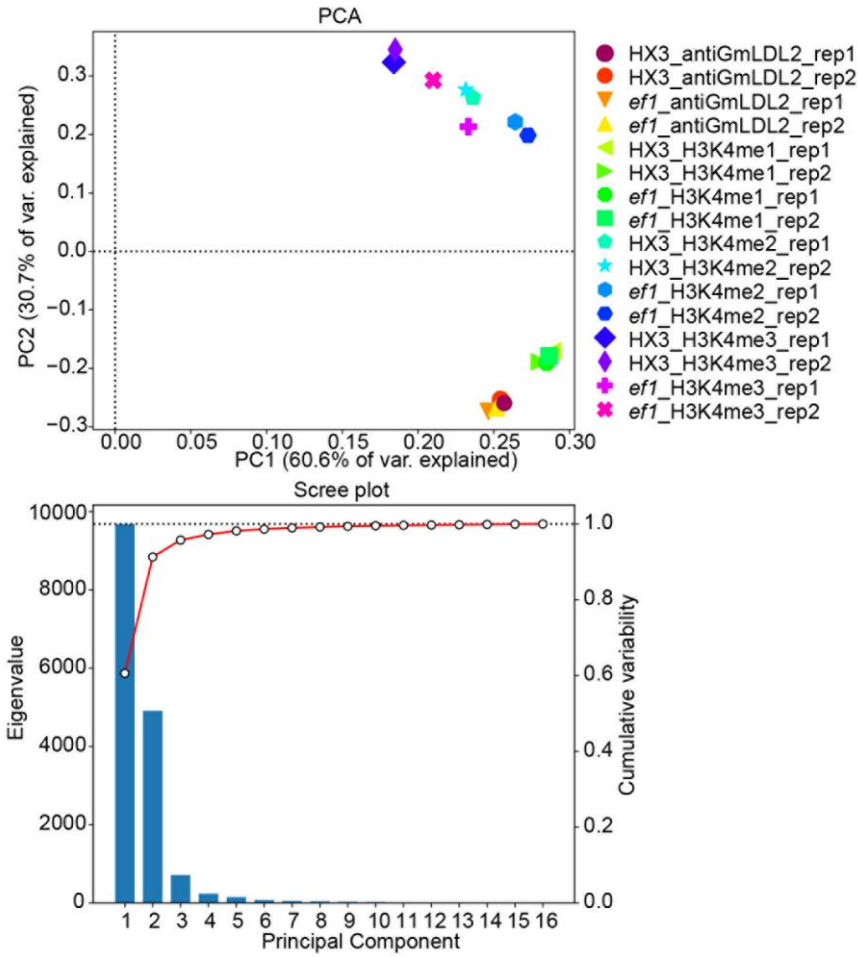
Response Table 3 ChIP-seq data sets and quality assessment of libraries.

Sample	Peak number (narrow)	Overlap ratio	> 5-fold enrichment	ratio (> 5-fold enrichment)	FRiP (narrow)	peak number (broad)	FRiP (broad)
WT_LDL2_rep1	39,366	59.28%	10,226	25.98%	12.76%		
WT_LDL2_rep2	13,769		5,464	39.68%	5.93%		
*ef1*_LDL2_rep1	31,146	51.34%	13,462	43.22%	14.34%		
*ef1*_LDL2_rep2	24,358		8,078	33.16%	10.08%		
WT_H3K4me1_rep1	8,580	30.69%	2,057	23.97%	2.35%	33,099	8.86%
WT_H3K4me1_rep2	5,119		1,876	36.65%	1.80%	22,666	6.78%
*ef1*_H3K4me1_rep1	9,968	25.87%	2,492	25.00%	2.73%	37,595	10.22%
*ef1*_H3K4me1_rep2	8,287		1,966	23.72%	2.16%	21,329	6.01%
WT_H3K4me2_rep1	28,070	61.88%	4,997	17.80%	11.43%		
WT_H3K4me2_rep2	21,514		3,375	15.69%	8.07%		

*ef1*_ H3K4me2_rep1	24,514	53.03%	3,338	13.62%	8.65%
*ef1*_ H3K4me2_rep2	19,100		2,732	14.30%	6.64%
WT_ H3K4me3_rep1	20,695	52.18%	10,541	50.94%	10.87%
WT_ H3K4me3_rep2	24,180		8,605	35.59%	10.65%
*ef1*_ H3K4me3_rep1	61,694	72.90%	25,318	41.04%	24.19%
*ef1*_ H3K4me3_rep2	33,466		14,238	42.54%	14.25%

Question 4: Figure S10 showed the correlation between the two replicates of each sample. However, the analysis is not sufficient for determining the correlation between different samples. A combined PCA analysis for the correlation of all samples with two replicates needs to be performed. Even the correlation between different samples was comparable to the correlation between the replications of each sample, this analysis will provide more complete information.

Response: Thanks for your valuable suggestions. By segmenting the whole genome into 500 bp windows, we compute the average scores for each file within every genomic region. Then, we conducted a combined PCA analysis incorporating all samples with two replicates to better assess their correlation (revised Supplementary Fig. 12, see also below). The first two principal components (PC1: 60.6%, PC2: 30.7%) explain 91.3% of the total variance. Distinct separation between different histone modification states (H3K4me1, H3K4me2, H3K4me3) and anti-GmLDL2. Samples suggest clear biological differences. HX3 and *ef1* samples exhibit distinct clustering, indicating that the *ef1* mutation may significantly impact chromatin modifications. We have added this information in lines 247-250 and lines 256-258 in the revised manuscript.



Supplementary Fig. 12 Principal Component Analysis (PCA) of ChIP-seq samples and scree plot.

Different samples, including histone modifications (H3K4me1, H3K4me2, H3K4me3) and anti-GmLDL2 from both HX3 and *ef1* mutants, are represented by distinct colors and shapes. The scree plot displays the eigenvalues associated with each principal component. The first two principal components capture the majority of the variance, as indicated by the steep drop in eigenvalues after PC2. The cumulative variance curve (red line) approaches saturation after PC2, demonstrating that higher-order components contribute minimally to the dataset's overall variance.

Question 5: In the ef1 mutant, the regions showing increased levels of H3K4me1 and H3K4me2 were identified (line 286-288). Although the ef1 mutant was supposed to increase H3K4me1 and H3K4me2, the regions with decreased levels of H3K4me1 and H3K4me2 should also be indicated. Further, both genes with increased and decreased levels of H3K4me1/2 should be compared with the GmLDL2 ChIP-seq data. This may reveal a higher degree of overlap between GmLDL2-bound genes and the genes showing increased levels of H3K4me1/2 than between GmLDL2-bound genes and the genes showing decreased levels of H3K4me1/2. This analysis is important for supporting the inference that GmLDL2 is involved in H3K4me1/2 demethylation.

Response: We thank the reviewer for this constructive suggestion. As recommended, we analyzed both hyper- and hypo-methylated regions of H3K4me1 and H3K4me2 in the *ef1* mutant, and identified the associated genes that overlapped with GmLDL2-bound genes. We further compared

both gene sets (with increased and decreased H3K4me1/2 levels) to our GmLDL2 ChIP-seq data. Interestingly, we found that the proportion of GmLDL2-bound genes was similar between the hyper- and hypo-methylated groups (see figure below). While this result is not entirely expected for a histone demethylase mutant, we believe this reflects the complexity of epigenetic regulation.

There are several possible explanations: (1) Not all GmLDL2-bound sites are necessarily catalytic targets; some may serve as docking platforms for other chromatin-modifying complexes or function in structural chromatin organization. (2) In the absence of GmLDL2, feedback or compensatory regulation by other histone-modifying enzymes may lead to hypo-methylation at certain loci. Particularly, demethylation of H3K4 by LSD requires the presence of a long stretch of H3 tail sequence and is strongly influenced by other modifications of H3 tails, sensing the presence of additional epigenetic marks on the histone substrate[29–33]. (3) H3K4me1/2 levels are dynamic and context-dependent, and GmLDL2 activity may vary depending on chromatin state, developmental stage, or environmental conditions. Similar concerns were also raised by Reviewer #2. To further address this issue, we performed RNA-seq analysis on two-week-old leaves of both the *ef1* mutant and wild type plants under low phosphate and low temperature conditions. Compared to normal growth conditions, a significantly higher number of differentially expressed genes (DEGs) was observed in *ef1* under low temperature stress, with 4,901 genes upregulated and 5,611 downregulated, whereas the wild type showed 3,859 upregulated and 2,849 downregulated DEGs. Similarly, under low phosphate conditions, *ef1* exhibited 1,103 upregulated and 2,655 downregulated DEGs, in contrast to the wild type with only 637 upregulated and 141 downregulated genes. These results further support our hypothesis that H3K4me1/2 levels are dynamically regulated in response to environmental stress. Due to these data obviously beyond the scope of this paper, so we did not include them in the revised manuscript.

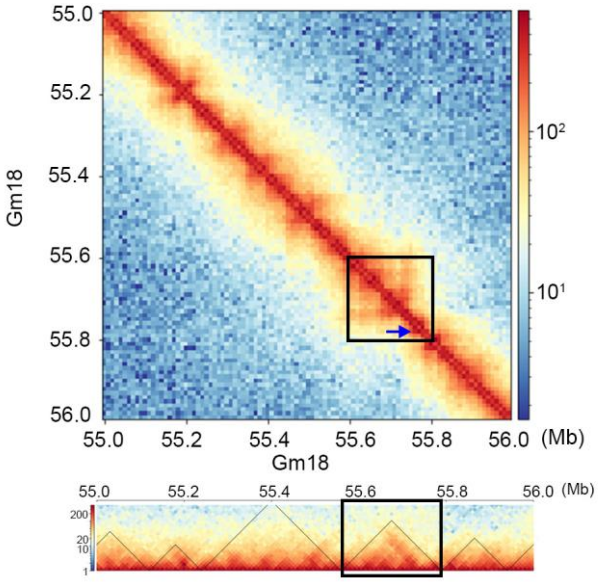
Overall, these findings highlight the multifaceted role of GmLDL2 in regulating gene expression, and support its involvement in chromatin remodeling rather than acting as a simple on-off switch for histone demethylation.

Response Fig. 6 Hyper- and hypo-H3K4 methylated genes in *ef1* mutant overlap with GmLDL2 binding genes.

Question 6: The H3K4me1 and H3K4me2 ChIP-seq signals were thought to increase at GmFER gene. However, the H3K4me1 signal and the GmLDL2 signal were outside of the GmFER gene, while the H3K4me2 signal was within the GmFER gene. If GmLDL2 is not located within the GmFER gene, how it mediates H3K4me2 demethylation within the gene.

Response: We thank the reviewer for raising this important point. To address this, we analyzed available Hi-C data[34] and found that the genomic region encompassing the *GmFER* gene

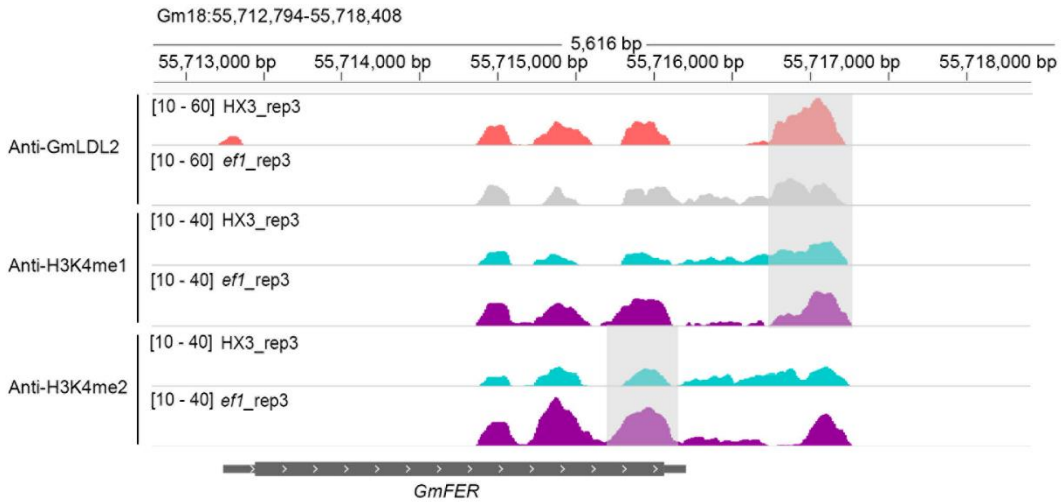
(Gm18:55,713,244..55,716,218) resides within a well-defined topologically associating domain (TAD) (see figure below). TADs are known to facilitate physical interactions among regulatory elements and genes located within the same domain. The *GmFER* gene and the GmLDL2 binding site both reside within this TAD, a region with enriched local chromatin interactions, which facilitates long-range regulatory communication. Therefore, it is plausible that GmLDL2, although binding downstream of *GmFER*, may still influence histone modifications within the gene body through chromatin configuration. We hope the reviewer will be satisfied with this explanation.



Response Fig. 7 Hi-C contact map and TAD structure around the *GmFER* locus showing the chromatin interaction matrix from 55.0 to 56.0 Mb on chromosome 18.
Black contours indicated the TADs and boundaries.

Question 7: Moreover, at the GmFER gene, GmLDL2 ChIP-seq signal was comparable between the wild-type HX3 and the ef1 mutant, even though the highlighted region showed a little different. I am not sure whether the GmLDL2 ChIP-seq is in a high quality. Were the GmLDL2 ChIP-seq peaks repeatedly identified in two replicates?

Response: Yes, the GmLDL2 ChIP-seq peak was consistently identified in two independent replicates. In the revised manuscript, we have included both individual replicates of anti-GmLDL2 ChIP-seq data (revised Fig. 5d), instead of the merged data presented in the original version. Quality control analysis revealed robust FRiP scores (>5%) for the GmLDL2 ChIP-seq datasets. RiP% values around 5% or higher generally reflect successful enrichment. Additional comprehensive quality metrics, including principal component analysis of all ChIP-seq samples were detailed in Response to Question 3 (Response Table 2, see above) and Question 4 (Supplementary Fig. 12, see above). To address this point and reinforce our conclusions, we have supplemented the study with an additional biological replicate for the key H3K4me1/2 and anti-GmLDL2 ChIP-seq assays for both the *ef1* mutant and wild type. The result still revealed a strong H3K4me1/2 and anti-GmLDL2 ChIP-seq signals around the 3'terminus and in the downstream region of *GmFER* (Response Fig. 8, see also below).



Response Fig. 8 ChIP-seq profiles of GmLDL2 binding and H3K4me1/2 histone modifications at the *GmFER* locus.

Question 8: GmFER-OE lines showed early flowering, while Gmfer mutant lines exhibited delayed flowering, supporting the notion that GmLDL2 is involved in regulating flowering time by affecting the expression of GmFER. However, this is not the direct evidence for supporting the inference. I wonder whether the authors tried to generate the double mutant of Gmldl2 and Gmfer by genetic crossing or CRISPR-Cas9. If GmFER functions at the downstream step of GmLDL2, the early flowering phenotype of Gmldl2 mutant is supposed to be rescued by the Gmfer mutation in the double mutant.

Response: We greatly appreciate this valuable suggestion to enhance the clarity and accessibility of our manuscript, which was also emphasized by Reviewer #3. To expedite genetic analysis, we used the *Gmldl2* mutant as the background to obtain the *Gmldl2/fer* double mutant. The double mutant exhibited a pronounced delay in flowering time compared to the *Gmldl2* single mutant, surpassing even the wild type control (revised Fig. 6e-h, see also below). Notably, its phenotype closely mirrored that of the *Gmfer* single mutants (later flowering, revised Supplementary Fig. 17). These results provide compelling evidence supporting the conclusion that *GmFER* acts downstream of *GmLDL2* in regulating soybean flowering. We have added this information in lines 335-350 in the revised manuscript.

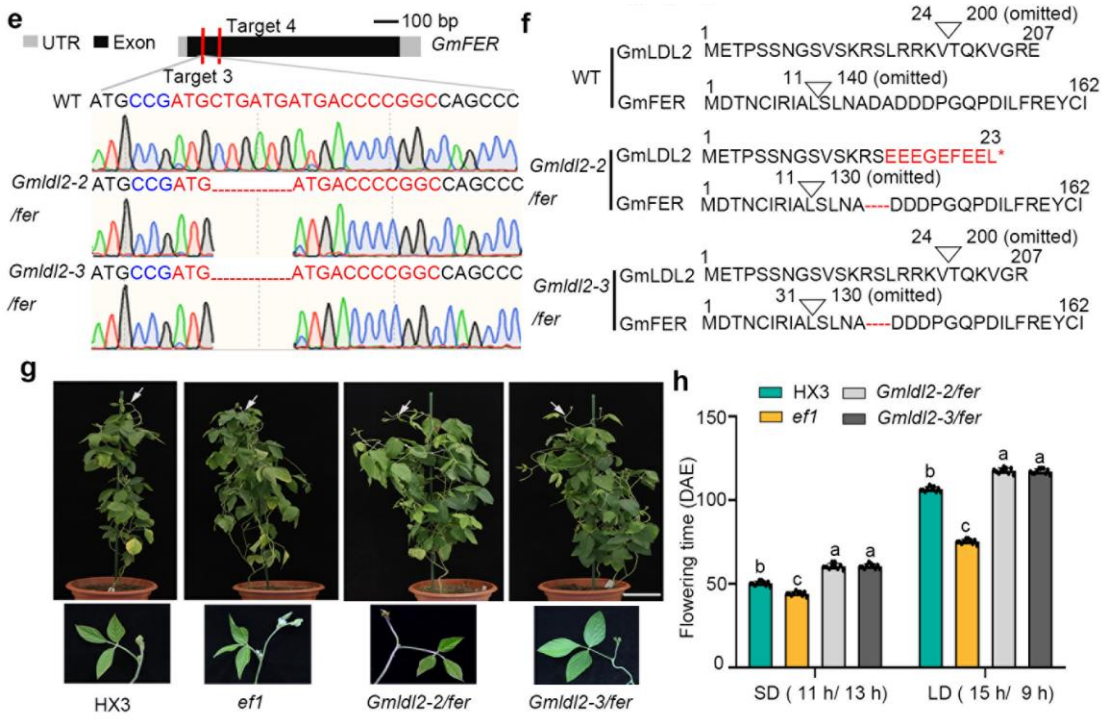


Fig. 6 GmLDL2 controls flowering in a *GmFER*-dependent manner.

e The design of sgRNAs for CRISPR/Cas9 to knock out *GmFER* in *Gmldl2* mutant background. Top: sgRNAs targeting CDS are labeled in red. Bottom: Results of Sanger sequencing of target sites in T_1 generations in HX3 and *Gmldl2/fer* double mutants. sgRNA target and PAM sequences were shown in red and blue, respectively. **f** Predicted amino acid sequences of GmLDL2 and GmFER in HX3 and *Gmldl2/fer* double mutants. Missense amino acids resulting from the frameshift caused by CRISPR/Cas9-induced mutations are shown in red. The asterisk indicates the termination of translation resulting from a stop codon. **g** The phenotypes of whole plants and flowers of HX3, *ef1*, *Gmldl2-2/fer,* and *Gmldl2-3/fer*. All *Gmldl2/fer* double mutants showed later flowering compared to *ef1*. Scale bars, 10 cm. **h** The flowering time of all varieties examined under SD (11 h/13 h) and LD (15 h/9 h) conditions. Error bars indicate standard deviation (n=10), and different lowercase letters indicate significant differences between the two panels, as determined by one-way analysis of variance (ANOVA) followed by Tukey's post hoc test ($P < 0.05$).

Question 9: The histone demethylase activity of GmLDL2H1 and GmLDL2H2 was assessed by luminescence assays in tobacco leaves, indicating GmLDL2H1 showed a higher level of histone demethylase activity than GmLDL2H2. I feel that the in vitro histone demethylase activity assay using bacterially expressed GmLDL2 proteins (as shown in Figure 3a) is a more strong evidence for supporting the conclusion.

Response: As recommended by the reviewer, we performed an *in vitro* histone demethylase activity assay using affinity-purified recombinant $GmLDL2^{H1}$ and $GmLDL2^{H2}$ protein. The results demonstrated that the $GmLDL2^{H1}$ exhibits significantly higher demethylase activity compared to $GmLDL2^{H2}$, which is consistent with the findings from the luminescence assay (revised Fig. 7f, g, see also below). We have added this information in lines 433-440 in the revised manuscript.

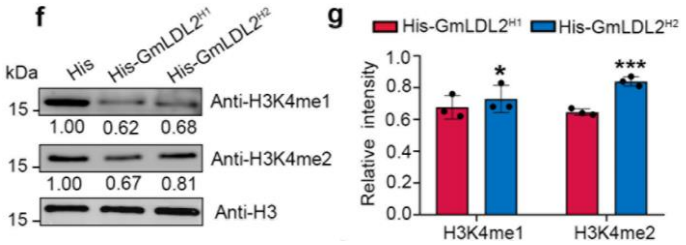
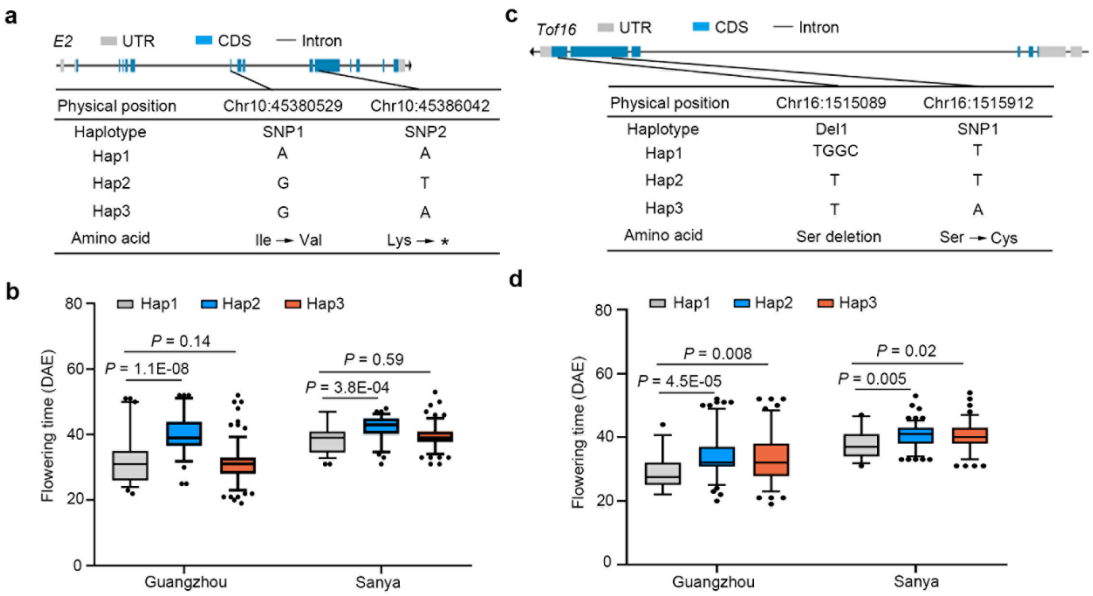


Fig. 7 GmLDL2 contributes the latitudinal adaptation of soybean.

f The affinity-purified recombinant His-GmLDL2^{H1} and His-GmLDL2^{H2} from *E.coli* were incubated with calf thymus histone for H3K4me1/2 demethylase activity assay, and the histone methylation status was determined using methylation specific antibodies. His was used as negative control and anti-H3 was used as a loading control. **g** Quantification analysis of (**f**) by ImageJ. The mean values are from three independent biological replicates. Statistical analyses were performed by Student's t-test (ns: no significance, *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$).

Question 10: It is interesting that the geographical distributions of GmLDL2H1 and GmLDL2H2 alleles differ, and contribute to the flowering time difference between them. The soybean accessions harboring GmLDL2H1 allele showed later flowering than the varieties containing GmLDL2H2 allele. I suggest to perform a complementation test for the ef1 mutant using the GmLDL2H1 and GmLDL2H2 transgenes. This experiment will provide a direct evidence for supporting the role of GmLDL2 H1 and H2 differentiation in adaptation.

Response: This is a great question, which is also noted by Reviewer #2. Since soybean transformation is very time-consuming, we have not had the complementation test of GmLDL2^{H1} and GmLDL2^{H2} at this moment. Nevertheless, as suggested by Reviewer #2, to exclude the other genes required for flowering time control in the analyzed accessions, we systematically analyzed natural variations in known soybean flowering-time genes, including *E1*, *E2*, *E3*, *E4*, *J*, *LUXs*, and *Tof16*, across our experimental accessions. The results revealed that no coding sequence (CDS) variations were detected in *E3*, *E4*, *LUXs*, *Tof18*, *GmFT2a,* and *GmFT5a*. Then, we excluded the soybean accessions carrying known functional alleles, including the *e1-as* allele (E1 transaction variant causing early flowering)[6], the *HT1m* mutation (J splice variant conferring later flowering)[7], *Tof11-1* to *Tof11-10* (loss-of-function alleles of *Tof11*), and *Tof12-1* to *Tof12-4* (loss-of-function alleles of *Tof12*)[8]. Next, we characterized coding sequence polymorphisms in *E2* and *Tof16*. Two exonic nonsynonymous SNPs defined three *E2* haplotypes, while three *Tof16* haplotypes were classified by one exonic nonsynonymous SNP and an in-frame deletion. To evaluate the phenotypic effects of the three haplotypes of *E2* and *Tof16*, we compared flowering time differences among different haplotypes. The results showed that haplotype 2 of *E2* flowered significantly later than the other two haplotypes, while both haplotype 2 and haplotype 3 of *Tof16* flowered later than haplotype 1 (Supplementary Fig. 24, see also below). Consequently, we also excluded accessions with *E2*^{H2}, *Tof16*^{H2}, or *Tof16*^{H3}. Finally, 64 soybean accessions with nearly identical genetic backgrounds were selected to assess the functions of *GmLDL2*^{H1} and *GmLDL2*^{H2}. Under natural short-day conditions, accessions carrying *GmLDL2*^{H1} exhibited significantly delayed flowering time, increased plant height, and greater grain weight per plant compared to those with *GmLDL2*^{H2} (revised Fig. 7h-j, see also below). We hope the reviewer will be satisfied with this explanation. This information has been added in lines 412-432 in the revised manuscript.



Supplementary Fig. 24 Haplotypes and their flowering time of *E2* and *Tof16*.

a Haplotypes of *E2* in the CDS region. **b** The flowering time of soybean accessions carrying $E2^{H1}$, $E2^{H2}$, and $E2^{H3}$ under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. **c** Haplotypes of *Tof16* in the CDS region. **d** The flowering time of soybean accessions carrying $Tof16^{H1}$, $Tof16^{H2}$, and $Tof16^{H3}$ under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. In all the box plots, the center line indicates the median, the edges of the box represent the first and third quartiles, and the whiskers extend to the smallest and largest data points within 1.5 interquartile ranges from the edges. Statistical significance is determined using a two-sided *t*-test.

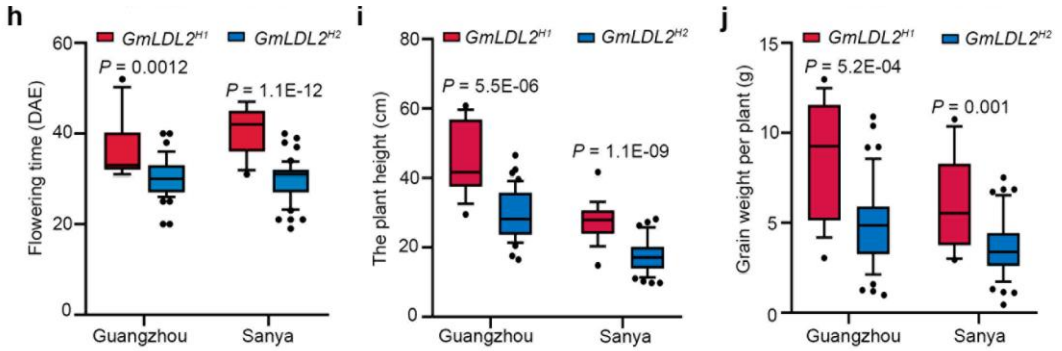


Fig. 7 GmLDL2 contributes to the latitudinal adaptation of soybean.

h-i Statistics of flowering time, plant height, and grain weight of 18 soybean accessions carrying $GmLDL2^{H1}$ and 46 varieties carrying $GmLDL2^{H2}$ under natural short-day conditions in Guangzhou and Sanya. At least 6 plants were tested per variety. Significant differences among haplotypes based on the two-tailed Student's *t*-test. In all the box plots, the center line indicates the median, the edges of the box represent the first and third quartiles, and the whiskers extend to the smallest and largest data points within 1.5 interquartile ranges from the edges. Statistical significance is determined using a two-sided *t*-test.

Reference

1. Choi, J.-H., Kim, J.-W. & Oh, M.-H. Identification of feronia-interacting proteins in *Arabidopsis thaliana*. *Genes Genomics* **44**, 1477–1485 (2022).
2. Cedano, J. A., Querol, E. & Mozo-Villarías, A. How hydrophobicity shapes the architecture of protein assemblies. *Eur. Phys. J. E* **46**, 62 (2023).
3. Teng, S., Srivastava, A. K., Schwartz, C. E., Alexov, E. & Wang, L. Structural assessment of the effects of amino acid substitutions on protein stability and protein protein interaction. *Int. J. Comput. Biol. Drug Des.* **3**, 334 (2010).
4. Xiao, Y. *et al.* Mechanisms of RALF peptide perception by a heterotypic receptor complex. *Nature* **572**, 270–274 (2019).
5. Liu, C. *et al.* Pollen PCP-B peptides unlock a stigma peptide–receptor kinase gating mechanism for pollination. *Science* **372**, 171–175 (2021).
6. 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.* **109**, (2012).
7. Yue, Y. *et al.* A single nucleotide deletion in *J* encoding GmELF3 confers long juvenility and is associated with adaption of tropic soybean. *Mol. Plant* **10**, 656–658 (2017).
8. Lu, S. *et al.* Stepwise selection on homeologous *PRR* genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* **52**, 428–436 (2020).
9. Liu, M. *et al.* Functional characterization of the lysine-specific histone demethylases family in soybean. *Plants* **11**, 1398 (2022).
10 Hall, S. J. G. Genetic differentiation among livestock breeds—values for *Fst*. *Animals* **12**, 1115 (2022).
11. Wright, S. The interpretation of population structure by f-statistics with special regaro to systems of mating. *Evolution* **19**, 395–420 (1965).
12. Eckshtain-Levi, N., Weisberg, A. J. & Vinatzer, B. A. The population genetic test Tajima's *D* identifies genes encoding pathogen-associated molecular patterns and other virulence-related genes in *Ralstonia solanacearum*. *Mol. Plant Pathol.* **19**, 2187–2192 (2018).
13. Xu, F. *et al.* Antagonistic control of seed dormancy in rice by two bHLH transcription factors. *Nat. Genet.* **54**, 1972–1982 (2022).
14. Guo, H. *et al.* Natural variation of *CTB5* confers cold adaptation in plateau *japonica* rice. *Nat. Commun.* **16**, 1032 (2025).
15. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. The limits of selection during maize domestication. *Nature* **398**, 236–239 (1999).
16. Dong, Z. *et al.* A regulatory network controlling developmental boundaries and meristem fates contributed to maize domestication. *Nat. Genet.* **56**, 2528–2537 (2024).
17. Liang, S. *et al.* Natural variation in *GmSW17* controls seed size in soybean. *Nat. Commun.* **15**, 7417 (2024).
18. Li, H. *et al.* Natural variation of *FKF1* controls flowering and adaptation during soybean domestication and improvement. *New Phytol.* **238**, 1671–1684 (2023).
19. Dong, L. *et al.* Parallel selection of distinct *Tof5* alleles drove the adaptation of cultivated and wild soybean to high latitudes. *Mol. Plant* **15**, 308–321 (2022).
20. Kou, K. *et al.* A functionally divergent *SOC1* homolog improves soybean yield and latitudinal adaptation. *Curr. Biol.* **32**, 1728-1742.e6 (2022).
21. Leigh, J. W. & Bryant, D. POPART: full-feature software for haplotype network construction.

Methods Ecol. Evol. **6**, 1110–1116 (2015).
22. Rendón-Anaya, M. *et al.* Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* **18**, 60 (2017).
23. Li, Y. *et al.* Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol.* **188**, 242–253 (2010).
24. Bu, T. *et al.* A critical role of the soybean evening complex in the control of photoperiod sensitivity and adaptation. *Proc. Natl. Acad. Sci.* **118**, e2010241118 (2021).
25. Fang, X. *et al.* Modulation of evening complex activity enables north-to-south adaptation of soybean. *Sci. China Life Sci.* **64**, 179–195 (2021).
26. Wang, Y. *et al. GmRAV* confers ecological adaptation through photoperiod control of flowering time and maturity in soybean. *Plant Physiol.* **187**, 361–377 (2021).
27. Zhai, H. *et al. GmMDE* genes bridge the maturity gene *E1* and florigens in photoperiodic regulation of flowering in soybean. *Plant Physiol.* **189**, 1021–1036 (2022).
28. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
29. Chen, Y. *et al.* Crystal structure of human histone lysine-specific demethylase 1 (LSD1). *Proc. Natl. Acad. Sci.* **103**, 13956–13961 (2006).
30. Chen, F. *et al.* Structural insight into substrate recognition by histone demethylase LSD2/KDM1b. *Cell Res.* **23**, 306–309 (2013).
31. Kim, S.-A., Zhu, J., Yennawar, N., Eek, P. & Tan, S. Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. *Mol. Cell* **78**, 903-914.e4 (2020).
32. Zhang, Q. *et al.* Structure-function analysis reveals a novel mechanism for regulation of histone demethylase LSD2/AOF1/KDM1b. *Cell Res.* **23**, 225–241 (2013).
33. Karytinos, A. *et al.* A novel mammalian flavin-dependent histone demethylase. *J. Biol. Chem.* **284**, 17775–17782 (2009).
34. Wang, L. *et al.* Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. *Plant Cell* **33**, 1430–1446 (2021).

Point-by-point response to reviewers' comments
(Questions in black and responses in blue)

Reviewer #1 (Remarks to the Author):
The authors had addressed all my concerns and I have no further comments.
Response: Thank you so much for your encouragement and valuable comments.

Reviewer #2 (Remarks to the Author):
Thank the authors for their efforts to address my questions.
Most of my questions are addressed, only GmLDL2 overexpression which was driven by 35S promoter in the ef1 mutant background in fact may not be the truly complementary test. However, it may prove the function of LDL2. In the future, a true complementary test may be better, as well as the functional difference test of the two hypotypes.
I do not have further questions.
Response: We are grateful for the reviewer's satisfaction with our response. I do agree with that further studies using native promoter of LDL2 and functional test of two hypotypes are deserved.

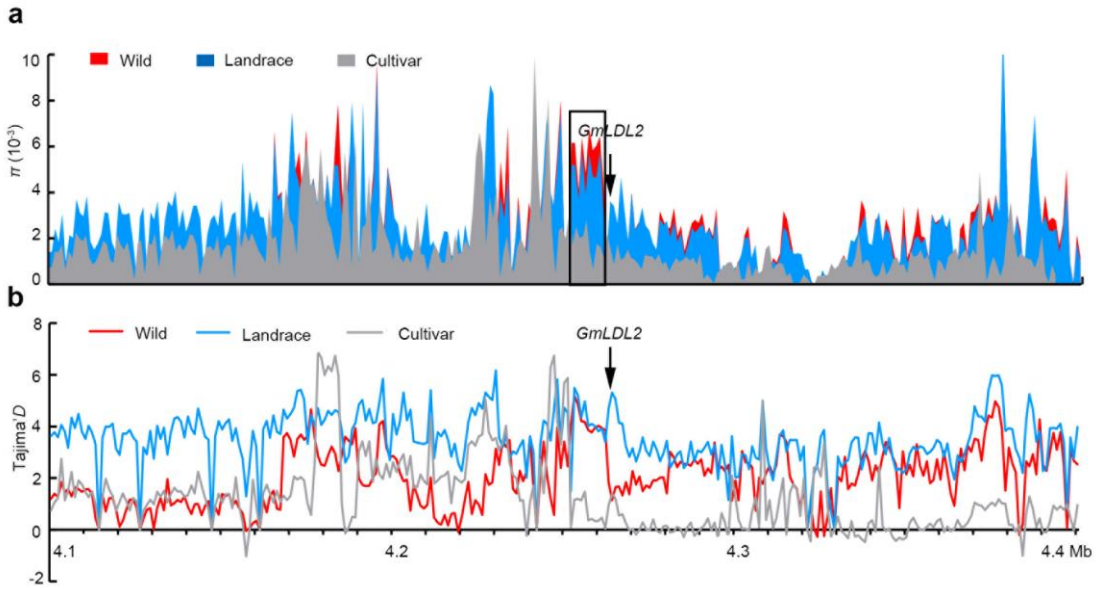
Reviewer #3 (Remarks to the Author):
I have carefully reviewed the revised version and the authors' responses, and found that the authors did their best to address the comments and concerns raised by the Reviewers. We appreciate their efforts, particularly the additional work, and the revised manuscript has been significantly improved compared to the previous version. However, several minor issues should be further clarified.
Response: We appreciate the reviewer's satisfaction with our response.

Question 1: In Fig. 2a and Fig. 6g, the growth conditions (SD or LD) should be stated in the corresponding legends.
Response: Thank you for pointing this out. We have added the growth conditions in the corresponding legends (Line 1016 and Line 1117-1118 of the revised manuscript).

Question 2: In Supplementary Fig. 19a, I cannot see a higher nucleotide diversity (π) in wild soybeans compared to landraces and cultivars. However, the nucleotide diversity of wild soybeans was lower than that of landraces and cultivars in the GmLDL2 region as the authors indicated by an arrow. Overall, I cannot see a signal for artificial selection for this gene. Moreover, previous work showed that domesticated soybeans retain 66% (π) and 49% (θ) of the nucleotide diversity of wild soybeans after the domestication bottleneck (Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, et al. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA. 2006; 103: 16666-71.). Therefore, I am not really convinced by the presented data and results to suggest the artificial selection on GmLDL2. Please clarify the related statements.
Response: We appreciate the reviewer's thoughtful comments. We apologize for this inappropriate description and inaccurate labeling in the original manuscript and Supplementary Fig. 19a.

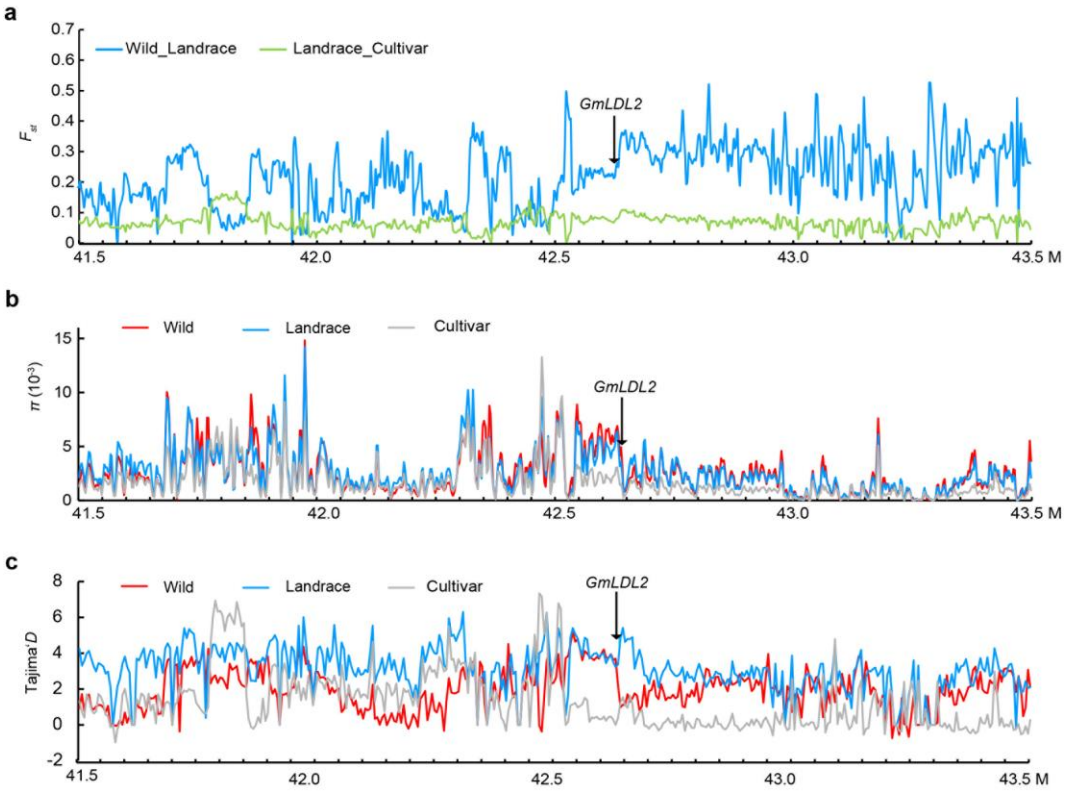
The *GmLDL2* gene (*SoyZH13_06G229900*) is located at the region of Chr06:42631396-42636682 in the ZH13 reference genome. According to the description in method sections of the manuscript: "VCFtools was used to calculate the pairwise genomic differentiation values for wild, landrace, and improved cultivar soybean populations per 10 kb sliding window, including the

fixation index (*Fst*) of population differentiation, nucleotide diversity (π) and Tajima's D". Due to the gene located around the boundary of the highlight boxed region (Response Fig 1: Supplementary Fig. 19 in the manuscript, see also below), which showed nucleotide diversity (π) differs significantly among the three groups. Wild soybeans (red) show the highest π values, indicating greater genetic diversity. Landraces (blue) exhibit intermediate diversity, while cultivars (gray) display markedly reduced π, suggesting a strong selective sweep or genetic bottleneck during domestication or improvement in this region.



Response Fig. 1 The Nucleotide diversity (*π*) and Tajimai's *D* value across the *GmLDL2* locus. **a** Nucleotide diversity (*π*) in wild soybeans, landraces, and improved cultivars across the 3 Mb genomic regions surrounding *GmLDL2*. **b** Tajimai's *D* value in wild soybeans, landraces, and improved cultivars across the 3 Mb genomic regions surrounding *GmLDL2*.

Therefore, we reduced the sliding window size from 10 kb to 5 kb and set the additional parameter --window-pi-step 2500 to increase resolution. This adjustment enabled a finer-scale analysis of nucleotide diversity (π), improving our ability to detect localized selection signals. In the window containing *GmLDL2*, the nucleotide diversity (π) was 0.00330208 in wild soybeans, 0.002325 in landraces, and 0.000992 in cultivars (revised Supplementary Fig. 20a, see Response Fig. 2b below). The corresponding F_{ST} values were 0.262575 (wild vs. cultivars) and 0.0808876 (landraces vs. cultivars) (revised Fig. 7a, see Response Fig. 2a below), and Tajima's *D* values were 2.24323, 3.36964, and 0.378166 for wild soybeans, landraces, and cultivars, respectively (revised Supplementary Fig. 20b, see Response Fig. 2c below). We have revised the relevant description in lines 359-365 (revised manuscript).

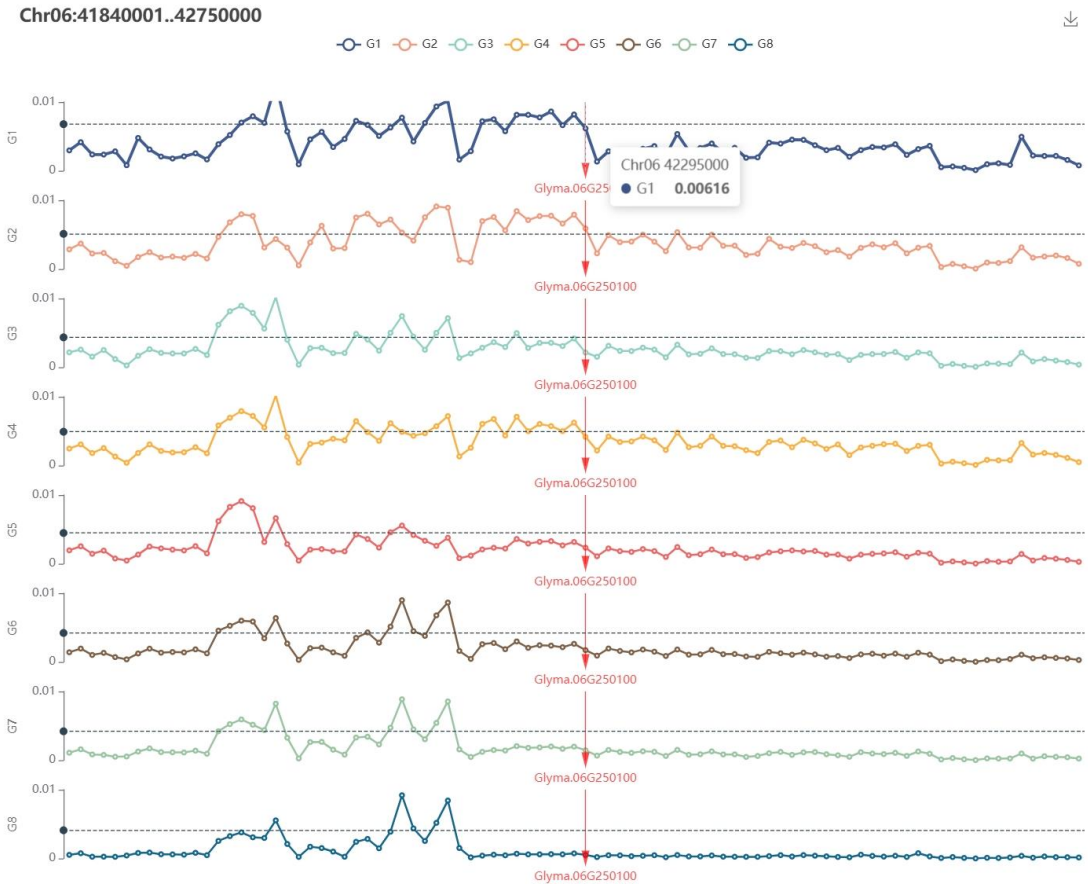


Response Fig. 2 The *Fst,* Nucleotide diversity (π) and Tajimai's *D* value across the *GmLDL2* locus.

a The nucleotide variation indicated by *Fst* in wild soybeans, landraces, and improved cultivars across the 2 Mb genomic regions surrounding *GmLDL2*. **b** Nucleotide diversity (π) in wild soybeans, landraces, and improved cultivars across the 2 Mb genomic regions surrounding *GmLDL2*. **c** Tajimai's *D* value in wild soybeans, landraces, and improved cultivars across the 2 Mb genomic regions surrounding *GmLDL2*.

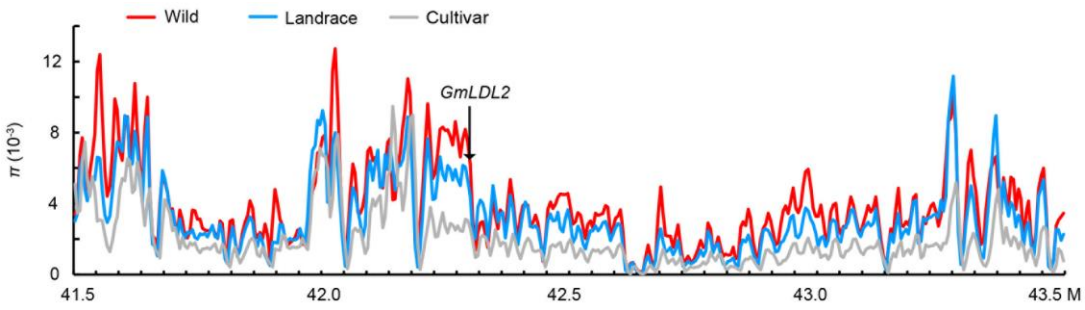
In addition, we consider that 103 wild accessions used in this study for calculating π are primarily from northeastern Asia and may represent a relatively narrow geographic and genetic background. In contrast, 1048 landraces and 1747 cultivars were collected from a broader range of geographic origins. This difference in population structure may have led to an underestimation of π in wild soybeans, while slightly elevating π in landraces and cultivars. Hence, we search for another larger SoyMD database (https://yanglab.hzau.edu.cn/SoyMD), comprising 4,414 soybean accessions, of which 588 are wild soybeans. The reference genome of this database is Wm82.a2.v1, and the *GmLDL2* gene (*Glyma.06G250100*) is located at the region of Chr06:42299582-42304722 in the Wm82 reference genome. Obviously, we detected the very high π in the G1 wild population (0.00616) (Response Fig. 3 was generated using the SoyMD website, see also below). Furthermore, we downloaded the VCF file for all 4,414 accessions from the SoyMD database. The accessions were then classified into three subpopulations: wild, landrace, and cultivar. We recalculated nucleotide diversity (π) based on this database from SoyMD database and observed values of 0.006172 for wild soybeans, 0.004374 for landraces, and 0.002592 for cultivars (Response Fig. 4, see below). Therefore, the marked reduction in nucleotide diversity in landrace and cultivars supports the hypothesis that *GmLDL2* may have undergone artificial selection. We hope the

reviewer can be satisfied with this explanation.



Response Fig. 3 Nucleotide diversity (π) across eight subpopulations (G1–G8) in the genomic region surrounding the *GmLDL2* gene, based on data from SoyMD database.

G1 (N = 588): Wild soybean accessions. G2 (N = 439): Landrace accessions from Yellow River, Huai River region and Hai River (Huang-Huai-Hai) basins of China. G3 (N = 379): Cultivars from Huang-Huai-Hai region of China. G4 (N = 686): Landraces and cultivars from southern China. G5 (N = 251): Landraces from Japan and Korea. G6 (N = 687): Landraces and cultivars from America and some cultivars from the Huang-Huai-Hai and northern China regions. G7 (N = 539): Landraces of northern China. G8 (N = 842): Cultivars of northern China.

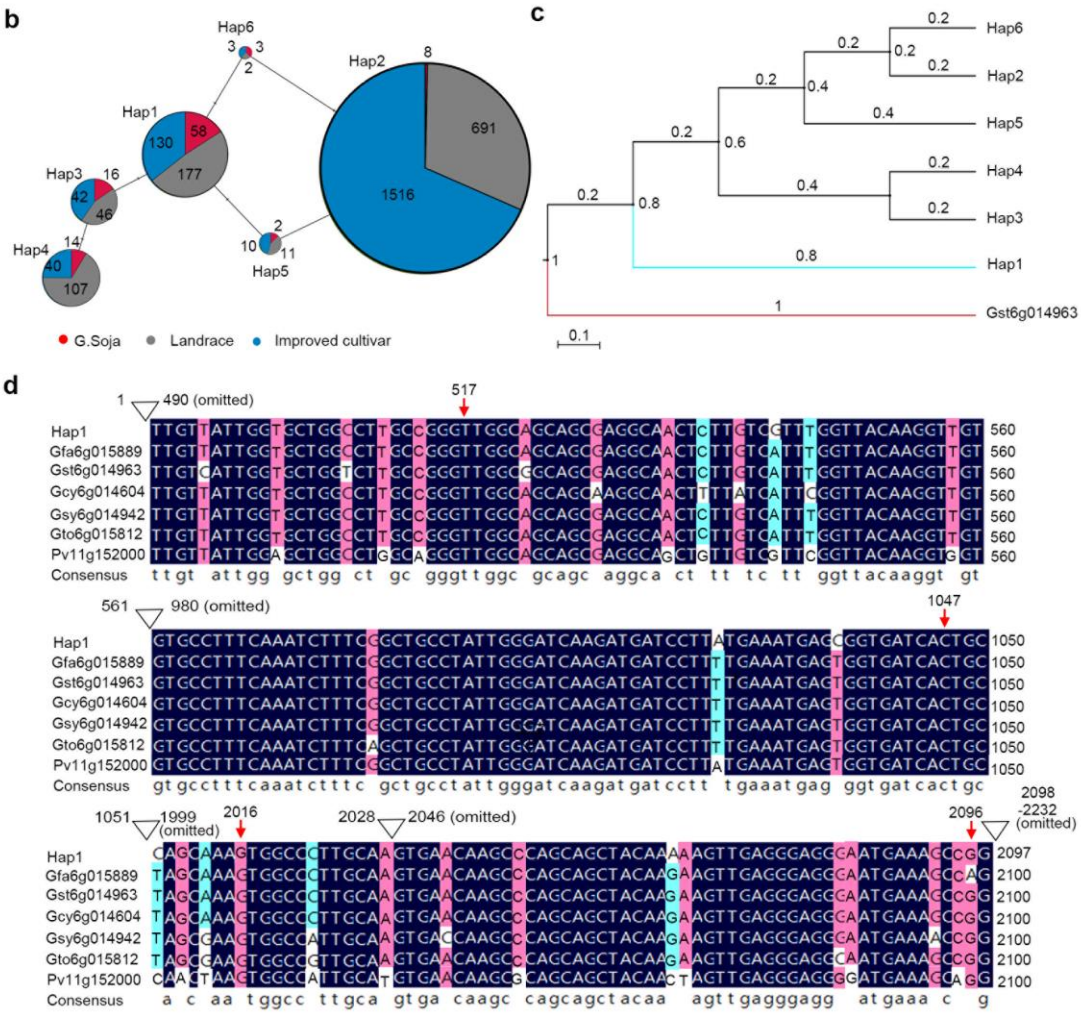


Response Fig. 4 The nucleotide diversity (π) value across the *GmLDL2* locus.

The nucleotide variation indicated by *Fst* in wild soybeans, landraces, and improved cultivars across the 2 Mb genomic regions surrounding *GmLDL2* based on data from the 4414 panel of 588 wild soybeans, 1253 landrace accessions, and 2573 improved cultivars.

Question 3: The frequency of haplotype may largely indicate its adaptability. The presented result in Supplementary Fig. 20b did not support the identification of ancestral haplotypes. A rooted tree with the outgroup haplotype may be necessary, which may provide evidence for identifying ancestral haplotypes.

Response: We agree that the current results based on the construction of a minimum spanning tree (MST) do not provide sufficient evidence to confidently identify ancestral haplotypes. As per your suggestion, we used the *Glycine stenophita*, a wild perennial species in the subgenus *Glycine*, as the outgroup to construct a rooted tree. The results indicate that Hap1 is more closely related to this species, suggesting that Hap1 may represent the ancestral haplotype of *GmLDL2* (revised Supplementary Fig. 21c, see also below). We also aligned the CDS of the *LDL2-like* gene from other wild perennial *Glycine* species, including *G.falcata*, *G.stenophita*, *G.cyrtoloba*, *G.syndetika*, and *G.tomentella* D3, with *Phaseolus vulgaris* as the outgroup. The results showed that the base T at position 517, C (1047), G (2016), and G (2096) were highly conserved among *LDL2-like* homologs (revised Supplementary Fig. 21d, see also below). This conservation pattern further supports the hypothesis that Hap1 represents the ancestral haplotype of *GmLDL2*. We hope the reviewer can be satisfied with this explanation, and thanks again for the constructive comment. We have added this information in lines 382-384 in the revised manuscript.



Supplementary Fig. 21 Haplotypes and their origins of *GmLDL2*.

a Haplotypes (H1-H6) of GmLDL2 in the CDS region. The numbers indicate position relative to ATG (bp). **b** Haplotype origins of GmLDL2. Red represented the wild soybeans, grey represented the landraces, and blue represented the improved cultivars. **c** Phylogenetic relationships of the different haplotypes of *GmLDL2* and *LDL2-like* genes from the species of *Glycine stenophita*. The phylogenetic tree was constructed by the neighbor-joining method using MEGA 11.0, with genetic distances calculated via the maximum composite likelihood method. The bootstrap values were 1,000 replications for major branches. **d** The CDS sequence alignment of *GmLDL2* homolog. Gfa: *G.falcata*, Gst: *G.stenophita*, Gcy: *G.cyrtoloba*, Gsy: *G.syndetika*, Gto: *G.tomentella* D3, Pv: *Phaseolus vulgaris*. The triangle denotes the spacer region of *GmLDL2*.

Question 4: The authors claimed that the GmLDL2H1 and GmLDL2H2 did not show obvious geographical distribution trend in the wild and landraces accessions (Supplementary Fig. 21). I do not think that the authors accurately described the results. What I see is that the frequencies of GmLDL2H1 and GmLDL2H2 exhibited distinct geographical patterns between the wild soybeans (Supplementary Fig. 21a) and landraces (Supplementary Fig. 21b). Clarification or proper interpretation of this point is required.

Response: We apologize for this unclear description. As mentioned in the text, unlike in the cultivar subset, the frequencies of $GmLDL2^{H1}$ and $GmLDL2^{H2}$ do not show a distinct geographical distribution pattern in wild or landrace accessions. And as you noted, these two alleles exhibit contrasting patterns between wild and landrace accessions. In the wild soybean accessions, the frequency of $GmLDL2^{H1}$ increased from 40% in the Huang-Huai region to 100% in the Northeast and Southern regions. Conversely, in landrace varieties, it decreased from 37.3% to 26.7% across these regions. We have revised the relevant description in lines 391-395 (revised manuscript).

Question 5: The results indicated that soybean accessions carrying GmLDL2H1 showed significantly later flowering time, increased plant height, and grain weight per plant compared to those carrying GmLDL2H2 under natural short-day conditions (Fig. 7h-j). Apparently, early flowering of accessions carrying GmLDL2H2 significantly decreased the soybean yield. How to understand the artificial selection on GmLDL2H2? What is the advantage of GmLDL2H2?

Response: This is a good question. Since *GmLDL2* encodes a histone demethylase that negatively regulates soybean flowering, and enzymatic assays demonstrated that $GmLDL2^{H2}$ exhibits significantly lower demethylase activity than $GmLDL2^{H1}$ (Fig. 7d-g), we conclude that $GmLDL2^{H2}$ is a hypomorphic allele (partial loss-of-function variant). Phenotypic analysis showed that *Gmldl2* knockout mutants flower significantly earlier than wild type under both short-day and long-day conditions (Fig. 2e,f). We thus speculate that the soybean accessions carrying $GmLDL2^{H2}$ may flower earlier than those carrying $GmLDL2^{H1}$ under natural long-day conditions, mirroring the phenotype observed under natural short-day conditions. This earlier flowering phenotype could facilitate timely maturation before frost onset, potentially maximizing yield in the high-latitude regions. Consequently, integrating the $GmLDL2^{H2}$ allele into elite varieties from low-latitude regions enhances their adaptation to the high-latitude areas, which deserves study in the future. We hope the reviewer can be satisfied with this explanation.

Reviewer #4 (Remarks to the Author):

Most of my concerns have been well addressed in the revised manuscript. Two additional results presented in the response-to-reviewer file should be incorporated into the revised manuscript. This is crucial for offering a comprehensive understanding of the function of GmLDL2.

Response: Thanks for your positive assessment of our revised manuscript.

Question 1: Response Table 1, which shows the effect of ef1 on yield-related traits, should be presented in the Supplemental Information. This will offer a more comprehensive understanding of the biological function of GmLDL2.

Response: Thanks for your suggestion. We have added this information in lines 120-124 and included Response Table 1 as Supplementary Table 1 (revised manuscript).

Question 2: Although the results presented in Response Figure 6 are unexpected and cannot be explained at present, they will significantly contribute to understanding the multifaceted role of GmLDL2. This is of great importance to other researchers in the field. Thus, I strongly recommend adding these results to the revised manuscript.

Response: We appreciate this suggestion. As recommended, we have added these results in lines 264-269 and included Response Figure 6 as Supplementary Fig.15 in the revised manuscript.

Point-by-point response to reviewers' comments
(Questions in black and responses in blue)

Reviewer #3 (Remarks to the Author):

I have evaluated the revision and gone through the authors' response letter. I think that the authors have made the greatest efforts to address the concerns of the Reviewers. The manuscript has been improved. I have no further comments.

Response: Thank you for your review. We deeply appreciate your time and effort in helping us improve the quality of this article.