BMC Plant Biology

RESEARCH

Open Access

Loss of phytochromobilin synthase activity leads to larger seeds with higher protein content in soybean

Xin Su[1,2†], Hao-Rang Wang[3†], Yong Zhang[4], Hui-Long Hong[2], Xu-hong Sun[4], Lei Wang[4], Ji-Ling Song[4], Meng-Ping Yang[4], Xing-Yong Yang[4], Ying-Peng Han[1*] and Li-juan Qiu[2*]

Abstract

Seed weight is an important agronomic trait that is related to seed size and determines yield in soybean (*Glycine max*). We previously identified a spontaneous soybean mutant with light green leaves called *ygl2*. Here, we cloned *YGL2*, which encodes a phytochromobilin (PΦB) synthase involved in synthesizing the chromophore of the photoreceptor phytochrome. The lesion in *ygl2* is a 10-bp deletion, causing a frameshift mutation and a premature stop codon that truncates the encoded protein. In contrast to the wild type, *ygl2* lacks PΦB synthase activity and function. This appears to promote cell expansion, thus increasing seed weight. Surprisingly, the *ygl2* mutant also exhibits excellent traits including early maturity and high protein content. Moreover, under the condition of dense planting (3 cm), the yield of YGL2 mutant was significantly increased. Mutants harboring *ygl2* mutations that we generated via gene editing had enlarged seeds with high protein content. Moreover, the expression levels of the photoperiod sensitive genes (*E1*, *FT2a*, *FT5a*) were lower in the *ygl2* mutant than in the wild type. Mutating the *YGL2* gene resulted in increased biliverdin content and decreased heme content. We determined that Lhcb4, a chlorophyll *a/b* binding protein in photosystem II, interacts with YGL2 but not with the mutant version of the protein. We thus identified a mutation in a PΦB synthase gene that enhances seed weight in soybean, providing a promising breeding target for this important crop.

Keywords GmYGL2, QTL, Seed size, Seed weight, Soybean

Introduction

Soybean (*Glycine max*) is an economically important crop that provides edible oil, protein, and other nutrients for humans and feed for animals worldwide [60, 61, 67]. Soybean seeds have a balanced composition of many essential amino acids, with protein representing 37–48% of seed weight and fat representing another 16.3–25%, making soybean a high-quality protein source. Large seeds have typically been selected during crop domestication [14, 60, 61]. Larger seeds accumulate plentiful nutrients that confer better stress tolerance and/or better seed germination rates and early seedling development compared to smaller seeds [41]. But if the seeds are too large, they also usually have problems in germination and

†Xin Su and Hao-Rang Wang contributed equally to this work.

*Correspondence:
Ying-Peng Han
hyp234286@aliyun.com
Li-juan Qiu
qiulijuan@caas.cn
[1] Key Laboratory of Soybean Biology in Chinese Ministry of Education (Northeastern Key Laboratory of Soybean Biology and Genetics and Breeding in Chinese Ministry of Agriculture), Northeast Agricultural University, Harbin 150030, China
[2] State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, People's Republic of China
[3] Jiangsu Xuhuai Regional Institute of Agricultural Sciences, Xuzhou 221131, China
[4] Keshan Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161606, China

development. Thus, seed size and weight are important traits for adaptation and crop domestication.

Seed size and weight are precisely regulated by internal genetic information and environmental signals. The ubiquitin–proteasome pathway, G-protein signaling, mitogen-activated protein kinase (MAPK) signaling, and transcriptional regulators control seed development via various physiological pathways [30]. Several factors that regulate seed size and weight have been identified in soybean. For instance, the phytohormones auxin and ethylene play diverse roles in plant growth and influence seed size and weight [34]. Similarly, the gibberellin biosynthesis enzyme *GIBBERELLIN 3-OXIDASE 1 (GmGA3OX1)* promotes hundred-seed weight in soybean by increasing cell size [20]. *PROTEIN PHOSPHATASE 2C-1 (PP2C-1)* interacts with and dephosphorylates *BRASSINAZOLE-RESISTANT 1 (GmBZR1)* to activate brassinosteroid signaling and regulate seed size and weight [39]. Signaling from the jasmonic acid and cytokinin pathways is integrated through *JASMONATE-ZIM-DOMAIN PROTEIN 3 (GmJAZ3)* to positively regulate hundred-grain weight in soybean [15, 22]. *GmGA2ox* and *GmGA3ox1* regulate seed weight by catalyzing the biosynthesis of bioactive gibberellin [20]. The *SUGARS WILL EVENTUALLY BE EXPORTED TRANSPORTER* (*SWEET*) gene family and *CELL WALL/VACUOLAR INHIBITOR OF FRUCTOSIDASE 1* (*GmCIF1*) influence seed weight because they encode proteins that regulate sugar transport and metabolism [60, 61]. In addition, *Glycine max Soybean Seed Size 1* (*GmSSS1*,*Hundred-seed weight* (*HSW*,*Protein, Oil, Weight, Regulator 1* (*POWR1*,*Seed Thickness 1* (*ST1*,*GmST05* (also named *MOTHER OF FT AND TFL1* [*GmMFT*]); and *FATTY ACID DESATURASE*s (*GmFAD*s) control seed-related traits in soybean [6, 9, 12, 16, 31, 53, 64, 76].

Plant perception of neighbor proximity or canopy shading is primarily mediated by the red/far-red light phytochromes [33]. Photoactive phytochromes consist of a PHY apoprotein covalently attached to a linear tetrapyrrole chromophore (phytochromobilin, 3E-PΦB) [56]. PΦB synthase, a ferredoxin-dependent biliverdin (BV) reductase encoded by *LONG HYPOCOTYL 2* (*HY2*) in Arabidopsis (*Arabidopsis thaliana*), plays an important role in phytochrome biosynthesis [25, 28]. Mutants defective in phytochrome function display changes in leaf color and have been extensively studied in the context of light perception. For example, the Arabidopsis *hy1* and *hy2* mutants [27, 42] and the tomato (*Solanum lycopersicum*) *yellow-green-2* (*yg-2*) mutant [8, 13] exhibit an early flowering phenotype due to blocked phytochrome biosynthesis. Agronomic traits such as plant height, grain number, and seed set are significantly reduced in the rice (*Oryza sativa*) mutants *pale yellow revertible 1* (*pyr1)* and *yellow-green leaf 98* (*ygl98*) relative to their wild-type parents [10, 55], and so on.

Compared to model plants such as Arabidopsis and rice, few studies have focused on soybean mutants with leaf color phenotypes. Of the 28 such soybean mutants reported to date, only 16 mutations have been mapped to genomic regions [25, 35, 36, 46] . Only the *YELLOW LEAF 1* (*YL1*), *YL2*, *YELLOW 11* (*Y11*), *Mg-chelatase subunit ChlI1a*, *TRANSLOCON AT THE INNER ENVELOPE MEMBRANE OF CHLOROPLASTS 110* (*Tic110*), *psbP*, and *YELLOW GREEN LEAF (YGL)* genes have been cloned [7, 35, 36, 59, 63, 65]. Of these, the gene mutated in the mutant line *NJ9903-5* encodes a transporter localized to the chloroplast membrane whose loss of function has negative effects on plant height, main stem node number, grain number per plant, and single plant pod number but has little effect on hundred-seed weight, protein content, or oil content [26]. The role of PΦB synthase in seed weight has not yet been reported.

In this study, we mapped a QTL controlling hundred-seed weight using the spontaneous *ygl2* mutant, which shows yellow-green leaves in the field. We identified a 10-bp deletion leading to a frameshift mutation in a gene encoding PΦB synthase. While the mutant protein lost its original enzymatic role, it also acquired a new function in promoting seed size that has been selected during soybean domestication and genetic improvement. These findings illustrate a new evolutionary and developmental mechanism whereby seed weight and size can be controlled by the phytochrome signaling pathway, revealing new genetic variation and mechanisms during soybean domestication and evolution and providing new ideas for crop breeding.

Materials and methods

Plant materials and growth conditions

The original soybean *ygl2* mutant was identified in the field. In 2017, this plant was crossed to wild-type soybean variety *YGL2* to construct a mapping population. All F1 plants showed a normal green leaf phenotype, whereas the leaf color trait segregated in the F2 population. Of the 567 F2 plants examined, 412 had normal green leaves and 155 had yellow-green leaves. A Chi-squared test (Table S1) showed that the segregation ratio between plants with normal green vs. yellow-green leaves matched a segregation ratio of 3:1 ($P > 0.05$), indicating that the yellow-green leaf trait of *ygl2* is controlled by a single recessive nuclear gene. In the same year (October), F1 seeds were planted in Hainan, and an F2 population was obtained; these plants were propagated continuously from 2017 to 2022 at the Chinese Academy of Agricultural Sciences, China.

Morphological and physiological analysis

Flowering time was measured at the R1 stage, which is defined as the time from seedling emergence to the opening of the first flower [2]. Plant height and internode length were recorded at maturation (R8 stage), while grain protein content and hundred-seed weight were measured after harvest. Agronomic traits were studied at the mature stage using conventional methods. The contents of chlorophyll, other pigments, heme, and biliverdin IXa were determined in leaves using targeted metabolomics (LC-QQQ-MS, conducted by Allwegene Company). Starch, sucrose, fructose, and glucose contents were measured using a Solybo reagent kit. Chloroplasts in the leaves of 15-day-old seedlings were examined by transmission electron microscopy (Hitachi, H-600, Tokyo, Japan), and samples were prepared as described previously [77]. Semi-thin cross Sects. (1 μm) were obtained from the seeds of 138-day-old YGL2 and *ygl2* plants at the R8 stage for cell observation. The sections were stained with 1% toluidine blue. The inner surfaces of cotyledons from mature dry seeds were observed under a scanning electron microscope (SEM), and the cells in the middle region were observed and analyzed. Three seeds (one cotyledon per seed) were analyzed for each genotype. The cell number and cell area from each 600 × SEM image (3.26E-2 mm^2) were measured, and the total number of cells per cotyledon area was calculated. Cell and cotyledon areas were measured using AxioVision Rel. 4.7 imaging software (Carl Zeiss AG, Oberkochen, Germany). Statistical significance was assessed by Student's *t*-test or Duncan's test.

Map-based cloning of *YGL2*

Based on the initial mapping results [62], individual F2 plants heterozygous across chromosome 2 were selected and allowed to self. A set of 1440 F3 plants was previously obtained for fine-mapping of the *YGL2* locus, which was mapped to a region between markers RM02104 and RM02107 on chromosome 2, with nine annotated genes in the interval. In the current study, genes within this interval were screened for single nucleotide polymorphisms (SNPs) and insertion/deletion (InDel) polymorphisms to identify the candidate gene, revealing the presence of nonsynonymous mutations in the *YGL2* (Glyma.02g304700) gene.

Prediction of protein structure

The 3D structures of wild-type YGL2 and mutant ygl2 proteins were predicted using SWISS-model (https://swissmodel.expasy.org/).

Vector construction for CRISPR/Cas9-mediated editing and transformation

The clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated nuclease 9 (Cas9) system was used to knock out *YGL2*. The sequence and other information about the *YGL2* gene was downloaded from the Phytozome website (www.phytozome.net/). The single-guide RNA (sgRNA) was designed using the web tool CRISPR-P (http://cbi.hzau.edu.cn/crispr/). One sgRNA that targets the second exon of *YGL2* was used in this study. Two DNA oligonucleotides were synthesized by Sangon Biotech (Beijing, China) and cloned into the CRISPR/Cas9 vector. The resulting construct was transformed into Agrobacterium (*Agrobacterium tumefaciens*) strain EHA105 and used for soybean transformation [3]. The primers used in this study are listed in Table S2.

Subcellular localization of ygl2

The vector 35S-pCEP01-GFP was selected for cloning, and KnpI and XbaI restriction sites were used for cloning. Primers were designed to amplify the full-length coding sequence of *ygl2*, with the forward primer containing a KnpI restriction site before the start codon (ATG). To detect the subcellular localization of ygl2 protein, the open reading frame of *ygl2* without the stop codon was cloned into pSuper1300 containing the *GFP* reporter gene driven by the CaMV35S to generate the *35S::ygl2-GFP* construct. The CaMV 35S::GFP empty vector was used as a control. The constructs were transformed into Agrobacterium strain EHA105. The Agrobacterium cells were injected into the epidermal cells on the abaxial surfaces of *Nicotiana benthamiana* leaves. The GFP signals were observed 48 h after injection under a confocal laser-scanning microscope (Olympus, Tokyo, Japan). The excitation laser wavelength for GFP was 488 nm. The sequences of the 35S-pCEP01-GFP-*ygl2*-F and 35S-pCEP01-GFP-*ygl2*-R primers are listed in Table S3. The fluorescent signals from GFP were observed under a confocal laser-scanning microscope (Nikon, A1, Tokyo, Japan) at 48 h after infiltration.

Yeast two-hybrid assay

The yeast two-hybrid assay was performed using the Yeastmaker Yeast Transformation System 2 (Takara) according to the manufacturer's instructions. The full-length coding sequences of *GmYGL2*, *Gmygl2*, and *GmLchb4* were amplified using wild-type YGL2 cDNA as a template and cloned into the pGBKT7 and pGADT7 vectors. Different combinations of pGADT7 and pGBKT7 vectors carrying the target genes were co-transfected into yeast strain AH109. The yeast cells were grown on Leu and Trp drop-out plates and transferred to

X-α-Gal and Leu, Trp, His, Ade drop-out plates at 28 °C. The primers used for vector construction are listed in Table S4.

Bimolecular fluorescence complementation (BiFC) assay

The full-length coding sequences of *GmYGL2* and *Gmygl2* were individually cloned into the pCAMBIA2300-nYFP vector; the full-length coding sequence of *GmLhcb4* was cloned into the pCAMBIA2300-cYFP vector. These vectors were transfected into Agrobacterium strain EHA105 and transiently expressed in *N. benthamiana* leaves. Fluorescent signals were observed under a Zeiss LSM 780 confocal laser-scanning microscope. Primers used for vector construction are listed in Table S5.

Luciferase complementation imaging (LCI) assays

To test the interaction between GmYGL2 and GmLhcb4, LCI assays were performed using the leaves of 4-week-old *N. benthamiana* plants. The full-length coding sequences of *GmYGL2* and *Gmygl2* were cloned into the pCAMBIA1300-NLUC vector, and the full-length coding sequence of *GmLhcb4* was cloned into the pCAMBIA1300-CLUC vector between the KpnI and XbaI restriction sites [74]. The recombinant plasmids were introduced into Agrobacterium strain EHA105 and co-infiltrated into *N. benthamiana* leaves as described previously [74]. LUC activity was analyzed by chemiluminescence imaging (4600SF,Tanon, China at 48 h after infiltration. The primers used for the LUC assay are listed in Table S6.

RNA extraction and reverse-transcription quantitative PCR (RT-qPCR)

GmYGL2 was previously shown to be expressed in all soybean tissues [62]. To analyze the expression patterns of flowering- and photoperiod-related genes in wild-type YGL2 and *ygl2* plants, seedlings were exposed to continuous dark conditions (DD), continuous light conditions (LL), and long-day (LD, 16 h light/8 h dark) and short-day (SD, 12 h light/12 h dark) conditions starting 20 days after emergence at 25 °C. Fully expanded trilobed leaves were collected every 4 h starting on day 15. The leaves were immediately frozen in liquid nitrogen after harvest. RNA was extracted from the leaves using TRIzol reagent (Ambion) and from other organs using CTAB buffer as previously described [39]. A Maxima H Minus First Strand cDNA Synthesis Kit with dsDNase (Thermo Fisher) and SYBR Green PCR Master Mix (Takara) were used for reverse transcription and the detection of gene expression, respectively, according to the manufacturers' instructions. The primers used in this study are listed in Table S7.

RNA-seq analysis

Total RNA was extracted from three independent biological samples of leaves from YGL2 and *ygl2* soybean plants at the V3 stage. Seeds were harvested from YGL2 and *ygl2* plants with three biological repeats and quickly frozen in liquid nitrogen. Total RNA was isolated from the samples using the cetyltrimethylammonium bromide (CTAB) method as previously described [39] and subjected to RNA sequencing by BioMarker (Beijing, China). HISAT2 (v2.1.0) was used to map the clean data to the soybean reference genome (https://phytozome-next.jgi.doe.gov/), and SAMtools (v1.9) was used to sort and convert the resulting files [24, 29]. FPKM (Fragments per kilobase of transcript sequence per million mapped reads) values were calculated to measure gene expression levels [57]. DESeq2 was used to identify differentially expressed genes (DEGs) [37]. The changes in gene expression were analyzed by principal component analysis (PCA) using the stats package in R. Gene Ontology (GO) annotation was performed using SoyBase (https://soybase.org). The R package clusterProfiler (version 4.0.5) was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

Results

Phenotypic analysis of *ygl2*

We previously identified the *ygl2* mutant as a recessive mutant that arose spontaneously in the field [60, 61]. Phenotypic analyses at multiple locations and years revealed that the hundred-seed weight, seed length, and seed width were significantly larger for *ygl2* plants than for the wild-type parental line YGL2 (Yellow Green Leaf 2) (Fig. 1A–D). Analysis of transverse sections showed that *ygl2* seeds were larger than the wild type due to longitudinal cell elongation (Fig. 1E). As previously noted, *ygl2* produced yellow-green leaves throughout its growth period compared to the dark green leaves observed in the wild type (Fig. 1F). We investigated the changes in chloroplast ultrastructure in *ygl2* vs. the wild type using transmission electron microscopy. The *ygl2* mutant presented an abnormal chloroplast structure, with fewer and thinner grana than the wild type; the mutant grana contained fewer layers and had an irregular size and shape (Fig. 1G).

Mapping the locus contributing to hundred-seed weight

Our earlier mapping efforts [60, 61] narrowed down the target interval of *ygl2* to a 56.1-kb region on chromosome 2 containing nine annotated genes (Fig. 2A). We looked for nonsynonymous mutations in the coding regions of all nine genes in this region using resequencing data from the wild type and *ygl2* (Table S1). Notably, we detected a 10-bp deletion in the coding region

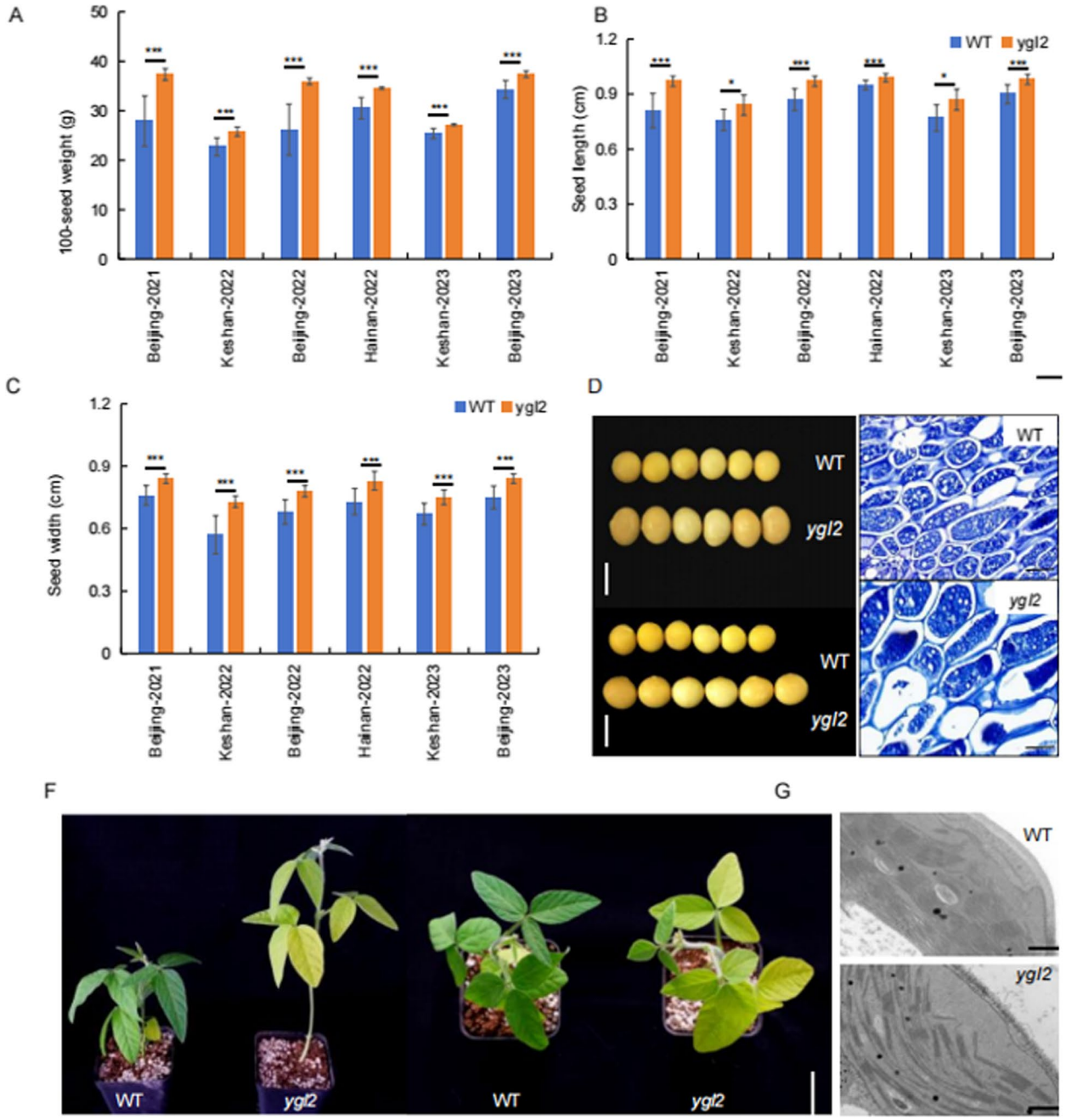


Fig. 1 Characterization of the growth of wild-type (WT) and *ygl2* plants. **A** Hundred-seed weight of WT and *ygl2* plants grown in Beijing, Keshan, or Hainan in 2021–2023 (Beijing) or 2022–2023 (Keshan, Hainan). **B**, **C** Seed length (**B**) and seed width (**C**) of WT and *ygl2* plants grown in Beijing, Keshan, or Hainan in 2021–2023 (Beijing) or 2022–2023 (Keshan, Hainan)(SD, $n = 5$). **D** Seed morphology of WT and *ygl2* plants. Scale bars, 7 mm. **E** Transverse sections of WT and *ygl2* seeds at the R6 stage. **F** Representative photographs of 15-day-old WT and *ygl2* plants. **G** Electron micrographs of chloroplast ultrastructure in 15-day-old WT and *ygl2* plants, Scale bar = 1 cm

of Glyma.02G304700 in *ygl2* compared to the wild type, introducing a frameshift mutation leading to early termination of translation (Fig. 2B–D).

To confirm the identity of *YGL2*, we generated transgenic soybean lines in the Jack cultivar background targeting the *YGL2* gene for genome editing via CRISPR/Cas9. We generated 30 independent positive transgenic

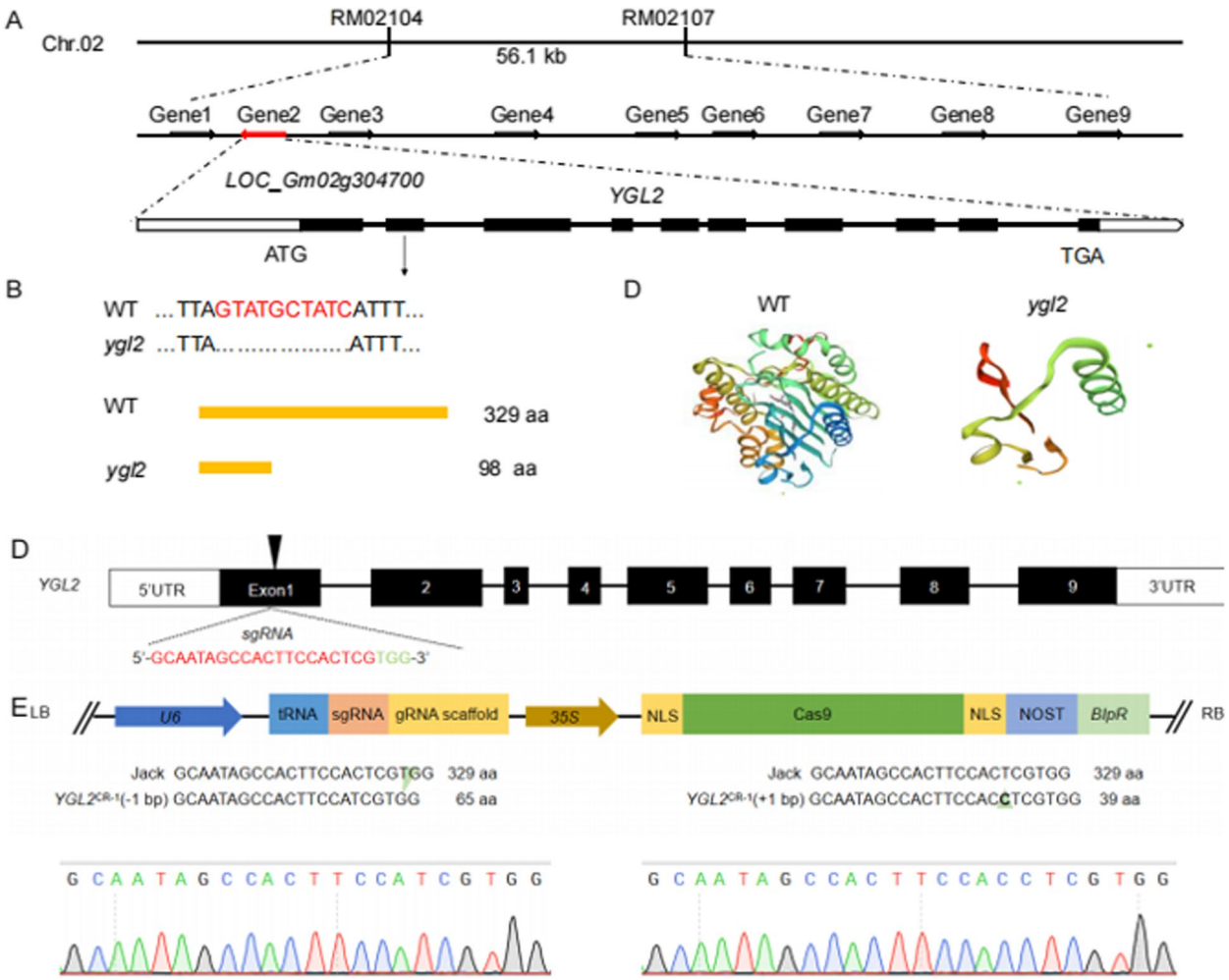


Fig. 2 Cloning of *YGL2* and characterization of seed weight. **A** Fine-mapping of *YGL2*. Top, fine-mapping the interval of *ygl2* showing the nine predicted genes in the interval. Bottom, diagram showing the structure of the *YGL2* locus. Black boxes represent exons. ATG and TGA indicate the start and stop codons, respectively. **B** The *ygl2* mutant harbors a 10-bp deletion in Glyma.02g304700. The red letters represent the 10-bp deletion in the second exon of *YGL2*. **C** The frameshift mutation in *ygl2* results in the early termination of protein translation, resulting in a protein of only 98 rather than 329 amino acids; yellow represents amino acids. **D** Predicted three-dimensional structure of YGL2 in the wild type (left) and the *ygl2* mutant (right). **(E)** Diagram showing the location of the sgRNA targeting *YGL2*. **(F)** Genotyping results of two gene-edited mutants in *YGL2*. The green inverted triangles indicate the 1-bp deletion in $YGL2^{CR-1}$ and the 1-bp insertion in $YGL2^{CR-2}$

individuals in the Jack background in the T0 generation, from which 26 transgenic lines exhibited phenotypes consistent with *ygl2* up to the T4 generation. We chose two independent transgenic lines for characterization. Sequencing of the genomic region targeted by the single guide RNA (sgRNA) used for editing revealed a 1-bp deletion and a 1-bp insertion in the two mutant lines relative to the wild type (Fig. 2E). Both mutations led to a frameshift that produced premature termination of translation, resulting in non-functional proteins consisting of only 65 amino acids (aa) or 39 aa compared to 329 aa for wild-type YGL2 (Fig. 2F).

The two independent gene-edited *ygl2* mutants had significantly larger grains compared to the wild type (Fig. 3) and significantly higher hundred-seed weight (Fig. 4). In addition, the protein content in seeds of these edited mutants was higher than that of the wild type (Fig. 4). The mutants produced yellow-green leaves, as did the original spontaneous *ygl2* mutant (Fig. 4). We conclude that Glyma.02g304700 is the gene mutated in the *ygl2* mutant and that its loss-of-function alleles increase hundred-seed weight in soybean.

ygl2 lacks phytochromobilin synthase activity

YGL2 (Glyma.02g304700) is annotated as a phytochromobilin (PΦB) synthase and is associated with the biosynthesis of the chromophore for the photoreceptor phytochrome (Table S8). In the *ygl2* mutant, Glyma.02g304700 was predicted to yield a truncated protein of 98 aa (Fig. 2C) with a much simpler structure than the wild type (Fig. 2D). We tested the subcellular localization of the mutant ygl2 protein by cloning the full-length Glyma.02g304700 coding sequence derived from the *ygl2* mutant in-frame and upstream of the sequence encoding green fluorescent protein (GFP). We expressed the resulting construct in the leaves of *N. benthamiana* plants via Agrobacterium-mediated infiltration. This truncated protein still localized to the chloroplast (Fig. 5).

We also measured chlorophyll (Ch), carotenoid, heme, and biliverdin contents in wild type and *ygl2* plants. The levels of total Chl, as well as Chl *a* and Chl *b*, were

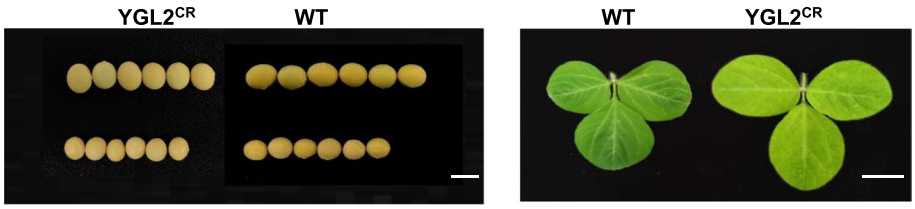


Fig. 3 Seed morphology of wild-type soybean and two independent transgenic lines. **A** Length and width of seeds from the transgenic lines (top) and wild type (bottom). Scale bars, 1 cm (**B**) Leaf color of the wild type (left) and one transgenic line (right). Scale bars, 5 cm

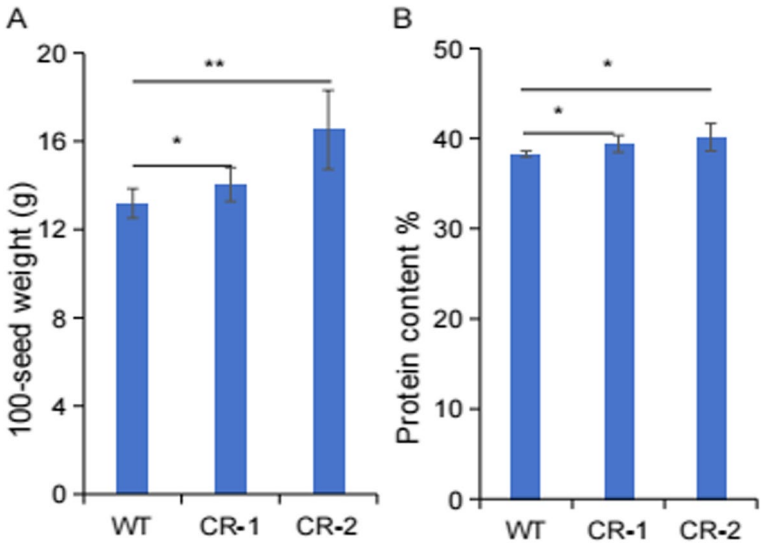


Fig. 4 Hundred-seed weight and protein content in soybean CRISPR mutants. **A** Analysis of 100-seed weight in the wild type and two CRISPR homozygous mutants (CR-1 and CR-2). **B** Analysis of protein content in the wild type and two CRISPR homozygous mutants (CR-1 and CR-2)

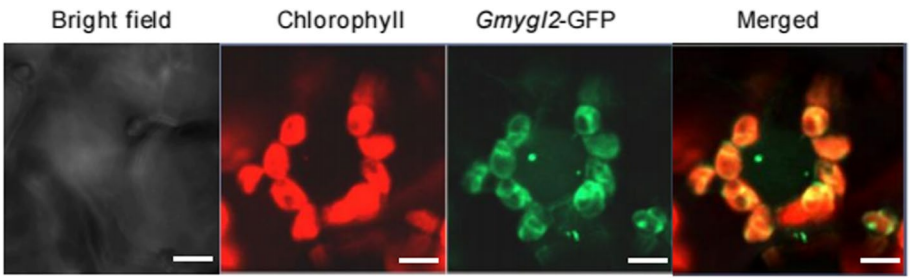


Fig. 5 Fluorescence microscopy of *35 s-pCEP01-GFP-ygl2*. Green signals represent GFP fluorescence, and red signals represent chlorophyll autofluorescence. Scale bar = 30 μm

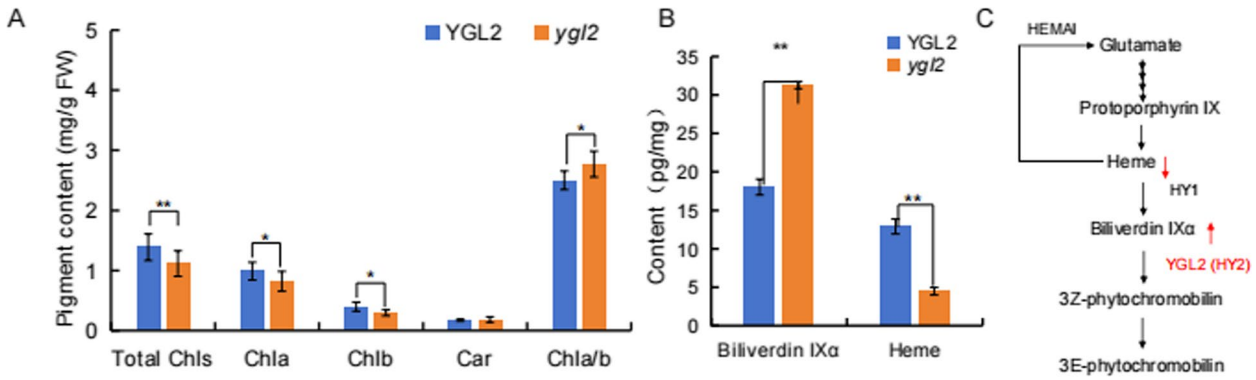


Fig. 6 Pigment contents in wild-type YGL2 and *ygl2* plants. **A** Chlorophyll contents, carotenoid contents, and related indices in wild-type YGL2 and *ygl2* plants. **B** Heme and biliverdin contents in wild-type YGL2 and *ygl2* plants. **C** Simplified diagram of the tetrapyrrole biosynthetic pathway. All samples were collected from plants at the V3 leaf stage. Data are shown as means ± standard error (SE) from three replicates. Statistical significance was determined using Student's *t*-tests: **, $P < 0.01$

significantly lower in the *ygl2* mutant compared to the wild type, along with a higher Chl *a*/Chl *b* ratio in the mutant (Fig. 6A). These results are consistent with the leaf color phenotype observed in the mutant. The carotenoid contents were similar between the wild type and *ygl2*. In agreement with its role as a PΦB synthase, we detected lower levels of biliverdin IXa, the substrate of PΦB synthase, and lower levels of heme in the mutant compared to the wild type (Fig. 6B).

The mutation of *YGL2* positively affects the accumulation of seed storage compounds

Plant storage tissues store assimilates or other substances, while sink tissues use these substances to support their growth. The grain is the most important sink tissue in soybean and other seed plants, providing the nutrients necessary for early seedling growth and development [4]. To explore the effect of the *ygl2* mutation on the sink–storage balance, we evaluated the levels of photosynthetic metabolites in leaves and seeds of wild type and *ygl2* plants. We observed a significant increase in glucose contents in *ygl2* seeds compared to the wild type, with a concomitant decrease in glucose levels in *ygl2* leaves. We also noticed a decrease in fructose levels in *ygl2* leaves, but this drop was not accompanied by an increase in fructose levels in *ygl2* seeds. There was no significant difference in sucrose content between the two genotypes in either tissue. Importantly, compared to the wild type, the starch content in *ygl2* increased by 52.23% in leaves and 37.91% in seeds (Fig. 7A, B).

In addition to changes in photosynthetic metabolites, flowering time was also significantly affected by the mutation of *YGL2* (Fig. 7C). We investigated the flowering times of wild-type YGL2 and *ygl2* plants under different photoperiod conditions (long day [LD] conditions: 16 h light/8 h dark; and short day [SD] conditions: 12 h light/12 h dark). Under LD conditions, the flowering time of the *ygl2* mutant was 37 ± 0.84 days, while that of the wild type was 45 ± 0.84 days, thus flowering 8 days later than the mutant (Fig. 7C). When grown under SD conditions, both wild type and *ygl2* plants flowered at ~ 30 days (Fig. 7C).

In addition to flowering time, the *ygl2* mutant showed increased plant height and higher protein content in seeds vs. the wild type and all of the mutants had a precocious phenotype(LD conditions) (Fig. 7D, E, F). Also, we found that the 100-seed weight of the *ygl2* mutant increased, but the yield per plant decreased (Compared with YGL2) (Figure S1A). Therefore, we conducted field yield measurement experiments under three density conditions of 3/4/5 cm, and the research results showed that, under the density condition of 3 cm plant spacing, the yield of the *ygl2* mutant plot significantly increased(Figure S1B), indicating that the *ygl2* mutant can increase yield through dense planting, thereby compensating for the yield reduction caused by high 100-seed weight. Under the density condition of 3 cm, the yield of the *ygl2* mutant plot significantly increased, thereby compensating for the yield reduction caused by high 100-seed-weight.

Loss of photoperiod sensitivity in the *ygl2* mutant

We performed transcriptome deep sequencing (RNA-seq) of wild-type YGL2 and *ygl2* plants at the V3 stage. This analysis yielded 165 upregulated differentially expressed genes (DEGs) and 279 downregulated DEGs in *ygl2* relative to wild type (Table S9). We conducted KEGG enrichment analysis of the DEGs (upregulated DEGs and downregulated DEGs), finding significant enrichment for genes associated with photosynthesis. All DEGs involved

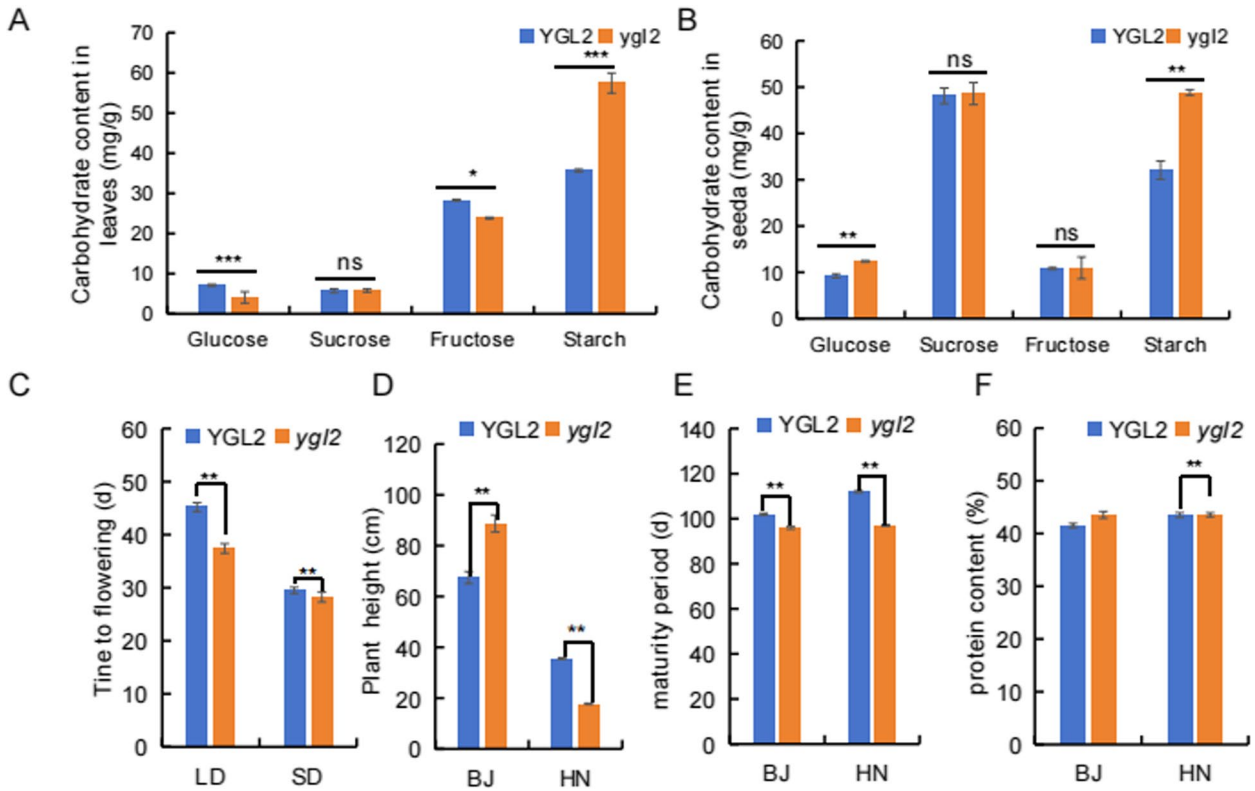


Fig. 7 Economically important characteristics of YGL2 and *ygl2* plants. **A**, **B** Glucose, starch, soluble sugar, sucrose, fructose, and starch contents in leaves (**A**) and seeds (**B**) of wild-type YGL2 and *ygl2* plants. (**C**) Flowering time (**D**), plant height (**E**), maturity period (**F**), and protein content of wild-type YGL2 and *ygl2* plants. BJ-LD, HN-SD. All WT and *ygl2* plants were grown in Beijing 2022–2023

in photosynthesis-related pathways were downregulated in *ygl2*, including genes encoding light harvesting chlorophyll *a/b* binding proteins (LHCs) and photosystem II proteins, among others (Fig. 8A). *Lhcb4*, encoding a component of photosystem II (PSII), was downregulated in the mutant (Fig. 8B).

We asked whether the expression of *YGL2* and *ygl2* are regulated by the circadian clock. Accordingly, we grew plants under LD conditions for 20 days before transferring them to continuous light (LL) or continuous dark (DD) conditions. We then collected samples every 4 h over 48 h. In the wild type, *YGL2* expression was initially higher in LL than in DD but gradually decreased in LL (Fig. 8C), Whereas no obvious pattern was observed in the mutants (Fig. 8D). Suggesting that *ygl2* mutant expression may not be regulated by the circadian clock. In addition, we observed no clear circadian rhythms in *YGL2* transcript levels. Moreover, we tested the expression of the photoperiod-sensitive genes *E1*, *FLOWERING LOCUS T 2a* (*GmFT2a*), and *GmFT5a* in wild type and *ygl2* plants grown under LD or SD conditions. Their expression levels were lower in the mutant than the wild type under both photoperiods (Fig. 8E–J).

ygl2 does not bind to Lhcb4

To better understand the underlying molecular mechanism of YGL2, we conducted a yeast two-hybrid screen to look for proteins that interact with YGL2. Among the potential interactors, we identified the PSII-localized protein Lhcb4. LHCs are membrane proteins that bind to pigment molecules in plant photosystems. In addition to capturing and transmitting light energy, LHCs are also widely involved in regulating the distribution of excitation energy between PSI and PSII, maintaining thylakoid membrane structure, photoprotection, and responses to various environmental conditions. We confirmed the interaction between YGL2 and Lhcb4 in a targeted yeast two-hybrid assay (Fig. 9A), a bimolecular fluorescence complementation (BiFC) assay (Fig. 9B), and a luciferase (LUC) complementation imaging assay (Fig. 9C). Importantly, the truncated YGL2 protein produced by the *ygl2* mutant was no longer able to interact with Lhcb4 in any of these assays.

YGL2 allele was domesticate during soybean improvement

We used resequencing data from 2533 soybean accessions to analyze the *YGL2* sequence and its haplotypes.

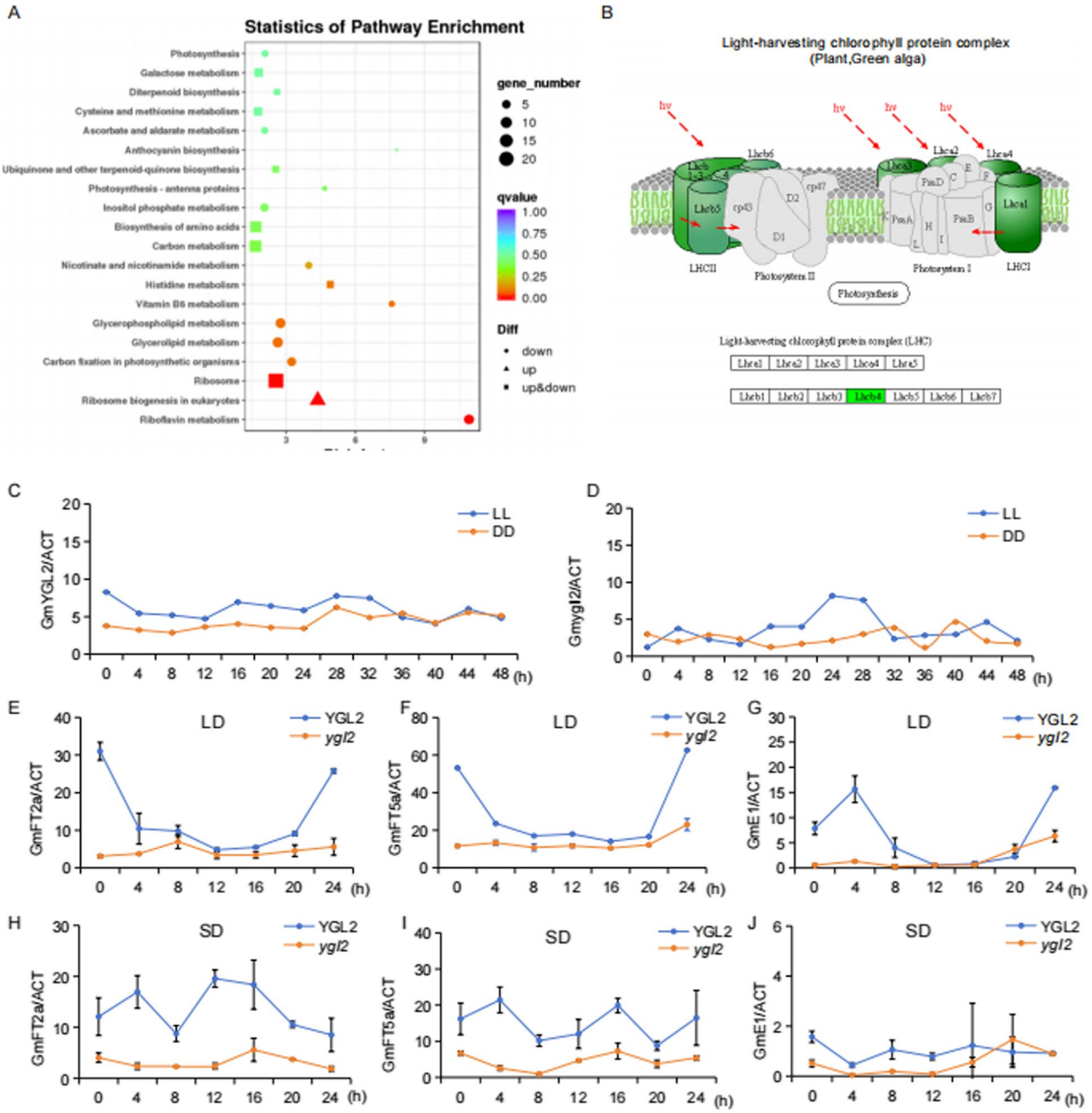


Fig. 8 Gene expression analysis in *ygl2* under different light treatment conditions. **A, B** Gene Ontology (**A**) and Kyoto Encyclopedia of Genes and Genomes (**B**) pathway enrichment analyses of the differentially expressed genes (DEGs) between wild-type YGL2 and the *ygl2* mutant. Lhcb4 highlighted in green represents downregulation. **C** Relative *YGL2* transcript levels in wild-type YGL2 in continuous dark and continuous light conditions. **D** Relative *YGL2* transcript levels in the *ygl2* mutant in continuous dark and continuous light conditions. **E**, **H** Relative *GmFT2a* transcript levels over 24 h in LD (**E**) or SD (**H**) conditions in wild-type YGL2 and *ygl2*. (**F**, **I**) Relative *GmFT5a* transcript levels over 24 h in LD (**F**) or SD (**I**) conditions. (**G, J**) Relative *E1* transcript levels over 24 h in LD (**G**) or SD (**J**) conditions

We detected single nucleotide polymorphisms (SNPs) at four sites, representing a total of five haplotypes. Among these, accessions carrying Hap3 (including the soybean reference cultivar Williams 82 [Wm82]) and Hap2 showed substantially different seed weight (Fig. 10C). A sequence significantly associated with hundred-seed weight was identified in Hap2 (GG/AA, 47,977,820) (Fig. 10C). We counted the number of wild accessions, landraces, and cultivars that harbor Hap2 or Hap3 (Fig. 10B). Hap2 was selected during the evolution of wild accessions to landraces, while Hap3 was selected during the evolution of landraces to cultivars.

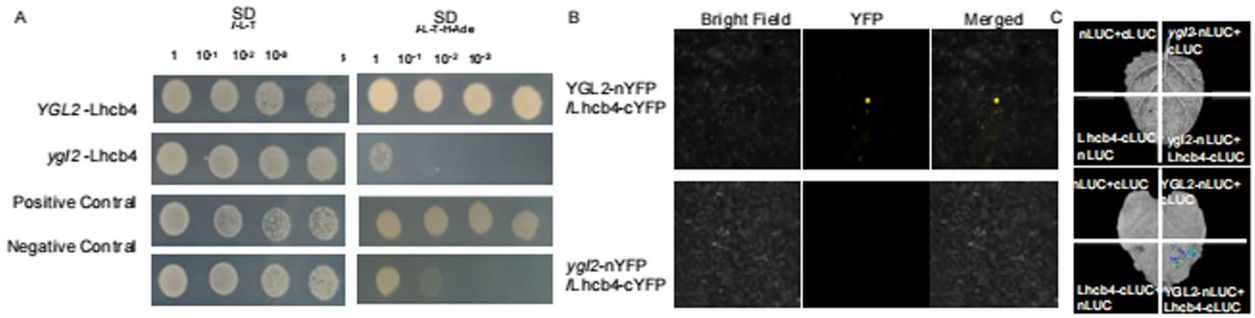
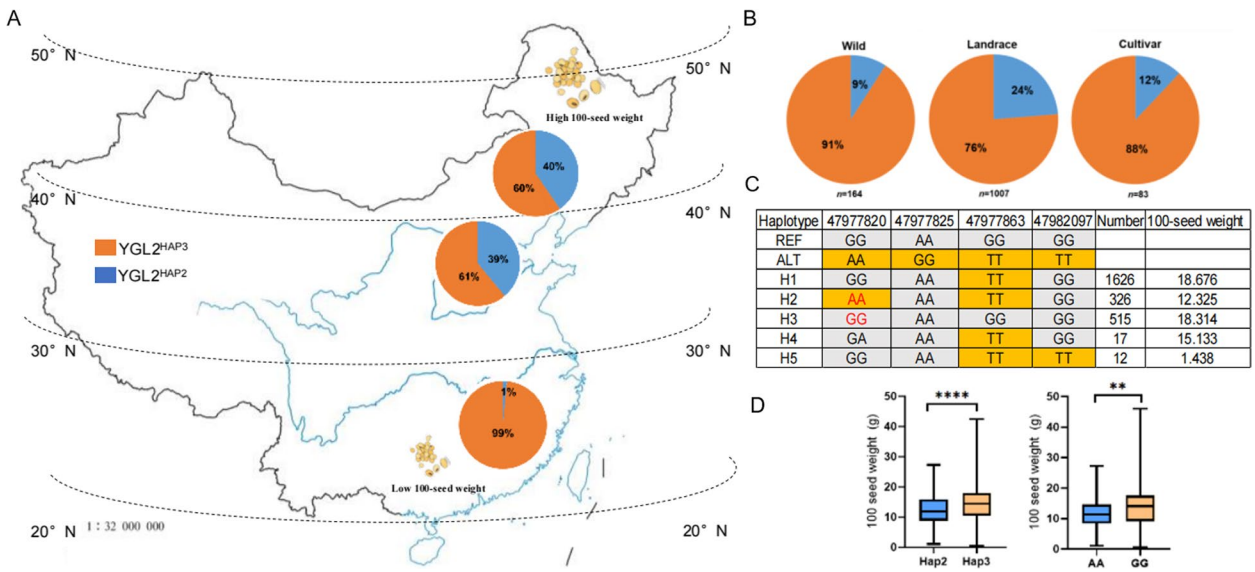


Fig. 9 Testing the interaction between GmYGL2 or Gmygl2 and GmLhcb4. **A** Yeast two-hybrid assay showing the interaction between GmYGL2 or Gmygl2 and GmLhcb4.–LWHA + X-α-Gal indicates X-α-Gal and Leu, Trp, His, Ade drop-out plates, and–LW indicates Leu and Trp drop-out plates. **B** Bimolecular fluorescence complementation assay of the GmYGL2 or Gmygl2 and GmLhcb4 interaction in *Nicotiana benthamiana* leaves. Scale bar = 20 μm. **C** LUC signals revealing the interaction between GmYGL2 or Gmygl2 and GmLhcb4. A leaf transiently co-expressing Gmygl2 and GmLhcb4 is shown at the top; a leaf transiently co-expressing GmYGL2 and GmLhcb4 is shown at the bottom



Haplotype	47977820	47977825	47977863	47982097	Number	100-seed weight
REF	GG	AA	GG	GG		
ALT	AA	GG	TT	TT		
H1	GG	AA	TT	GG	1626	18.676
H2	AA	AA	TT	GG	326	12.325
H3	GG	AA	GG	GG	515	18.314
H4	GA	AA	TT	GG	17	15.133
H5	GG	AA	TT	TT	12	1.438

Fig. 10 Geographical distribution of *YGL2* haplotypes. **A** Geographical distribution of $YGL2^{H2}$ and $YGL2^{H3}$ haplotypes in China. **B** Distribution of the $YGL2^{H2}$ and $YGL2^{H3}$ haplotypes among wild soybean accessions, landraces, and cultivars. **C** Haplotype analysis of *YGL2*. **D** Hundred-seed weight of soybean accessions carrying $YGL2^{H2}$ or $YGL2.^{H3}$ AA and GG, including wild accessions, landraces, and cultivars (There were no significant differences between the remaining haplotypes)

The 100-seed weight of accessions with Hap2 was 33.3% higher than that of Hap3 (Fig. 10C). Moreover, the hundred-seed weight of accessions carrying Hap1 or Hap3 was higher than that of the other three haplotypes, but the hundred-grain weight of lines harboring Hap1 was not significantly different from that of Hap3 (Fig. 10C). Hap 2 and Hap 3 show significant differences in the haplotypes of 100 seed weights (Fig. 10D), indicating that the *YGL2* allele was domesticate during soybean improvement.

Discussion

Loss of phytochromobilin synthase function affects seed weight and size

Soybean seed size is a major domestication trait associated with high yield. However, the genetic basis behind variation in seed size in soybean is largely unknown. In this study, we cloned the gene that is impaired in the recessive mutant *ygl2*, which encodes a truncated PΦB synthase due to a 10-bp deletion in the gene. The substrate for PΦB synthase, biliverdin IXa, accumulates in

the *ygl2* mutant, indicating that this truncated enzyme has no enzymatic activity. Although YGL2 and Lhcb4 proteins were both localized to chloroplasts, this study found that their colocalization occurred in the nucleus, which may be due to the presence of nuclear localization signal peptides in Lhcb4 protein, which was originally masked after the interaction of the two proteins to form a complex, resulting in the transport of the complex to the nucleus, while still relying on the chloroplast localization signal when it exists alone. Jin et al. [23] found that CDNI is a dual-localized protein that can localize to both the nucleus and chloroplasts. It contains chloroplast signal peptide (cTP) and nuclear localization signal sequence (NLS), which are similar to the results of this study. Therefore, the next step in this study will be to perform a knockout experiment on the Lhcb4 gene, knock out the nuclear localization signal sequence of Lhcb4, and observe whether the subcellular localization of the two proteins changes.

PΦB synthase is a ferredoxin-dependent biliverdin reductase and a key enzyme in the biosynthesis of the chromophore that confers phytochrome its ability to perceive changes in environmental light conditions to regulate plant growth and development [25]. Mutants of PΦB synthase genes have been identified in many species, the first of which was Arabidopsis *long hypocotyl 2* (*hy2*), which exhibits long hypocotyls when grown in red or white light due to a defect in phytochrome function [27]. Subsequently, mutants in orthologs of Arabidopsis *HY2* have been described in many crops, including tomato, maize (*Zea mays*), rice, and cucumber (*Cucumis sativus*) [8, 21, 50]. Zhu et al. [75] and Zhang et al. [73] found that *HY2a* deletion mutants would lead to the obstruction of photochromin synthesis and reduce photoperiod sensitivity, which was consistent with the results of this study. These mutants are all characterized by the absence of biosynthesis of the chromophore of phytochrome, leading to partial or complete photoperiod insensitivity and a loss of light-mediated regulation of growth and development. Indeed, we observed that flowering time in the *ygl2* mutant is largely photoperiod insensitive (SD), in contrast to the strong photoperiod sensitivity displayed by the wild type. Surprisingly, the loss of YGL2 function also resulted in larger and heavier seeds. This was confirmed in lines with editing of *YGL2* (Fig. 3), which produced even shorter versions of the YGL2 protein than that in the original spontaneous *ygl2* mutant. We established that the larger seeds of *ygl2* are due to enhanced cell elongation (Fig. 1).

Induced loss-of-function mutations dramatically alter plant physiology and morphology, with loss-of-function mutations being either neutral or deleterious. Likewise, naturally occurring mutations can have diverse effects on plant development and adaptation [66]. Natural loss-of-function mutations, particularly those that induce frameshifts, can also lead to the functional divergence of duplicated genes [43, 47] or alter the biochemical properties of the proteins they encode [1]. These evolutionary events might contribute to morphological and physiological variation. Mutations that modify the functions of genes from a duplicated pair have occurred during soybean domestication. For example, a polymorphism in a CC dinucleotide pair in the *GmSWEET39* coding sequence affects seed oil and protein contents [38, 40, 69, 70]. On the contrary, the results of the current study, the loss-of-function of GmSSSS1 via gene editing led to a decrease in the 100-seed weight of soybean [17, 76]. Variation at *GmSSS1* might thus represent a rare case of "gain by loss" of a neomorphic mutation during soybean domestication to increase seed weight. Here we demonstrated that the *ygl2* mutant is defective in PΦB synthase activity but that the mutant protein somehow acquired a new role in controlling seed size. In the next step, we will take a deep dive into whether mutants have an impact on shade, and the results will have a positive impact on the production application of mutants.

Using the pleiotropy of *YGL2* to cultivate high-protein soybeans in high-latitude regions

Soybean seeds contain high-quality protein; consuming 25 g of soy protein per day could reduce the risk of heart disease (1999, Food and Drug Administration, FDA). Therefore, high-protein soybeans are attractive to both soybean growers and end-users. Meal produced from soybean seeds after pressing and extracting oil contains high-quality protein. The United Nations Food and Agriculture Organization [32] called soybean meal "the most important preferred source of high-quality plant protein for animal feed". Globally, nearly 98% of soybean meal is used as animal feed, indicating that soybean protein is highly valued.

Soybean protein content is greatly affected by the environment, with more rain and higher temperatures associated with higher protein content [19]. Protein content also shows clear geographical distribution, with lower protein content at higher latitudes [68]. The major genes related to protein content that have been cloned to date are *GmSWEET39* [69, 70] and Glyma.20g085100 [51]. The role of *GmSWEET39* was revealed in a segregating population derived from a cross between Wm82 and wild soybean (*Glycine soja*) accession PI479752 [69, 70]. Variation in *Glyma.20g085100* was shown to be related to protein content, with plants harboring the allele associated with high seed protein content maturing 0–5 days earlier than plants with low seed protein content [5, 45, 51], with no reported effects on flowering time. Enrei

(a variety cultivated in Japan) is late flowering and late maturing, but its protein content is relatively high [52]. In the current study, the *ygl2* mutant had a high seed protein content and matured 4–8 days earlier than the wild type (Fig. 7).

Any phenotypic effects mediated by the ripening period may show genotype-specific responses to the environment in different regions. In the current study, the high-protein-content *ygl2* mutant flowered significantly earlier than wild type when grown at both high latitudes (Beijing) and low latitudes (Hainan). We suggest that the PΦB synthase encoded by Glyma.02g304700 somehow regulates protein content, but further studies are needed to investigate the nature of this relationship, and we will continue to further study the regional yields of mutants under different photoperiod conditions. Changes in protein content often have pleiotropic effects. The *ygl2* mutant showed significantly increased protein content (by 4.65%) compared to the wild type, with the hundred-seed weight increasing by 26.8%, suggesting that *YGL2* has pleiotropic effects and may provide additional beneficial traits for farmers.

YGL2 is an excellent resource for increasing hundred-seed weight and protein content at high latitudes

The seeds of cultivated crops are generally larger than the seeds of their wild ancestors, underscoring the parallel artificial selection trajectories these crops have followed [11]. To date, although some QTLs controlling seed size and seed weight have been identified in soybean, few such genes have been cloned [73]. The *GmPP2C-1* allele from wild soybean line ZYD7 significantly increased hundred-seed weight in soybean [39]. After introducing the superior *GmST1* allele into soybean cultivar KeFeng 1, the hundred-seed weight increased significantly, but the seed length/thickness ratio decreased significantly [31]. *GmST05* positively regulates soybean grain size and increases hundred-seed weight [12]. According to The Annals of Chinese Soybean Cultivars, soybean seeds weighing 25.0–29.9 g per 200 grains are considered to be large. The Japanese soybean variety 'Danish' produces the largest seeds worldwide, with its hundred-seed weight surpassing 30 g; the Korean soybean cultivars 'Huayan', 'Huacheng', and 'Xiliang' also have hundred-seed weights of over 30 g. Due to different tillage systems and different meteorological conditions, the sowing period of soybean in different regions varies greatly in northern China and is relatively short [72].

Wang et al. [58] analyzed the ecological distribution of soybean varieties in China and found that the hundred-seed weight varies widely across different regions. The hundred-seed weight was approximately 18–22 g in northeast China but only 13–16 g in the arid saline-alkali region of northeast China [58]. Zhang et al. examined the cultivation of soybean varieties in China and found that the hundred-seed weight of soybean varieties in northeast China was 17–23 g, while that in the Huang-Huai-Hai region was approximately 14–22 g [71]. In the current study, the hundred-seed weight of the *ygl2* mutant was as high as 34 g and did not change with latitude (Fig. 10C), making the mutant genotype suitable for the directional cultivation of new soybean varieties with large grains and the appropriate characteristics for different regions, especially high-latitude regions (Fig. 10C, D).

The source-sink theory was proposed by Mason and Maskell in 1928 and gradually developed into the guiding theory of physiological research on crop yield. Source tissues with net photosynthate output capacity (such as mature leaves) produce and supply photosynthates, which are consumed and stored in sink tissues. Plants undergoing growth and development are the main reservoirs for the accumulation of photosynthetic products. Crop yield involves the coordination of the structures and functions of sink and storage tissues, with large storage tissues being at the core of high yields [54]. The coordinated source–sink relationship was shown to promote photosynthesis in leaves, improve the efficiency of assimilate transport, and improve yield in cucumber [73], regulate the accumulation of the photosynthetic product sucrose in sugarcane (*Saccharum officinarum*) [49], balance carbon distribution in sunflower (*Helianthus annuus*) [44], improve stress tolerance in crops [48], and regulate the composition of the soil microbiome [18].

In the current study, the starch contents of leaves and seeds of the *ygl2* mutant increased significantly compared to the wild type (Fig. 5A, B). The storage and transport of carbohydrates in soybean seeds are mainly driven by the degradation of starch in leaves. In *ygl2*, the starch contents in leaves and seeds significantly increased compared to the wild type, but there was no significant difference in sucrose content, indicating that starch gradually accumulates in *ygl2* leaves and that sugar is transported to seeds and converted to starch. When the seeds mature, some of the sugar in the stems and leaves forms starch, while the rest is ultimately used to produce protein. Therefore, the protein content in *ygl2* seeds also significantly increased (Fig. 7F). We also suggest that the starch content is maintained at a relatively stable level to maintain the source–sink balance. Sucrose is the main photosynthate used for long-distance transport, with different concentrations at either end of the source-sink resulting in a change in the flow from source tissue to sink tissue. In the current study, the sucrose contents in the leaves and seeds of the *ygl2* mutant were comparable to the levels measured in the wild type (Fig. 7A and Fig. 7B). The mutation of *YGL2* resulted in significant

increases in soluble sugar and protein contents, indicating that the high carbohydrate level is beneficial for the accumulation of seed storage materials, ultimately promoting grain size. Therefore, this mutation could serve as a genetic resource for soybean breeding for large grains and high protein content. In the next step, we will deeply analyze the adaptability mechanism of 100-seed weight genes in different ecological environments, and cultivate high-yielding varieties with strong adaptability and stable yield.

In conclusion, we cloned a key gene that controls the weight of soybean grains. PΦB is an important plant pigment that is commonly involved in light signaling and plant development processes. Loss of function of this gene may affect grain development and protein accumulation in soybean. Therefore, this trait can be used to improve soybean varieties to show excellent traits in both grain weight and protein content. Furthermore, we will further study how YGL2 gene affects grain weight and protein content of soybean by regulating optical signaling and protein synthesis pathways, which is of great significance to meet the global demand for high-protein soybeans.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06298-z.

Supplementary Material 1.

Authors' contributions

X.S. and L.J.Q. designed the experiments. X.S. wrote the initial draft of the manuscript, which was revised by L.J.Q.. H.R.W. analyzed the data. X.S. and H.R.W. performed most of the experiments. Y.P.H., R.H.W., Y.Z., H.L.H., J.L.S., L.W., X.H.S., M.P.Y., and X.Y.Y. provided the wild and cultivated soybeans and contributed to data analysis and material preparation. All authors have read and approved of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32172005), Heilongjiang Academy of Agricultural Sciences (2021YYYF035) and Collaborative Innovation Center for Modern Crop Production co-sponsored by Province and Ministry.

Data availability

The data reported in this paper have been deposited in the GenBase [1] in National Genomics Data Center [3], Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation, under accession number C_AA099487.1 that is publicly accessible at https://ngdc.cncb.ac.cn/genbase.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 December 2024 Accepted: 25 February 2025
Published online: 28 May 2025

References

1. Bartonek L, Braun D, Zagrovic B. Frameshifting preserves key physicochemical properties of proteins. Proc Natl Acad Sci USA. 2020;117:5907–12.
2. Bernard R. Two major genes for time of flowering and maturity in soybeans. Crop Sci. 1971;(11):242.
3. Bian XH, Li W, Niu CF, Wei W, Hu Y, Han JQ, Lu X, Tao JJ, Jin M, Qin H. A class B heat shock factor selected for during soybean domestication contributes to salt tolerance by promoting flavonoid biosynthesis. New Phytol. 2020;225:268–83.
4. Board JE, Harvilleh BG. Late planted soybean yield response to reproductive source sink stress. Crop Sci. 1998;38:763–71.
5. Brzostowski LF, Pruski TI, Specht JE, Diers BW. Impact of seed protein alleles from three soybean sources on seed composition and agronomic traits. Theor Appl Genet. 2017;130:2315–26.
6. Cai Z, Xian P, Cheng Y, Zhong Y, Yang Y, Zhou Q, Lian T, Ma Q, Nian H, Ge L. MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. New Phytol. 2023;239:905–19.
7. Campbell BW, Mani D, Curtin SJ, Slattery RA, Michno J, Ort DR, Schaus PJ, Palmer RG, Orf JH, Stupar RM. Identical substitutions in magnesium chelatase paralogs result in chlorophyll-deficient soybean mutants. G3: Genes. Genomes, Genetics. 2014;5:123–31.
8. Chen X, Chen Z, Zhao H, Zhao Y, Cheng B, Xiang Y. Genomewide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. PLoS ONE. 2014;9:e87156.
9. Chen L, Nan H, Kong L, Yue L, Yang H, Zhao Q, Fang C, Li H, Cheng Q, Lu S. Soybean AP1 homologs control flowering time and plant height. J Integr Plant Biol. 2020;62:1868–79.
10. Cheng X, Ren DY, Ma J, Zhu XY, Sang XC, Ling YH, Zhao FM, He GH. Identification and Gene Mapping of Leaf Pale Yellow-Revertible Mutant pyr1 in Rice. Acta Agron Sin. 2013;39:992–8.
11. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006;127:1309–21.
12. Duan Z, Zhang M, Zhang Z, Liang S, Fan L, Yang X, Yuan Y, Pan Y, Zhou G, Liu S, Tian Z. Natural allelic variation of GmST05 controlling seed size and quality in soybean. Plant Biotechnol J. 2017;20(9):1807–1818.
13. Elzinga JA, Atlan A, Biere A, Gigord L, Weis AE, Bernasconi G. Time after time: flowering phenology and biotic interactions. Trends Ecol Evol. 2007;22:432–9.
14. Gaut BS, Seymour DK, Liu Q, Zhou Y. Demography and its effects on genomic variation in crop domestication. Nature Plants. 2018;4:512–20.
15. Ge L, Yu J, Wang H, Luth D, Bai G, Wang K, Chen R. Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. Proc Nat Acad Sci USA. 2016;113:12414–9.
16. Goettel W, Zhang H, Li Y, Qiao Z, Jiang H, Hou D, Song Q, Pantalone VR, Song BH, Yu D-Q, Charles An, Y.Q. POWR1 is a domestication gene leiotropically regulating seed quality and yield in soybean. Nat Commun. 2022;13:3051.
17. Gu Y, Li W, Jiang H, Wang Y, Gao H, Liu M, Chen Q, Lai Y, He C. Differential expression of a WRKY gene between wild and cultivated soybeans correlates to seed size. J Exp Bot. 2017;68:2717–29.
18. Hall JP, Wood AJ, Harrison E, Brockhurst MA. Source-sink plasmid transfer dynamics maintain gene mobility in soil bacterial communities. Proc Natl Acad Sci USA. 2016;29:8260–5.
19. Han X, Qi HD, Qi ZM, Chen QS. Analysis of soybean protein content for environment stability based on GGE biplot. J Agricult Sci. 2018;8:801–8.
20. Hu D, Li X, Yang Z, Liu S, Hao D, Chao M, Zhang J, Yang H, Su X, Jiang M, Lu S, Zhang D, Wang L, Kan G, Wang H, Cheng H, Jiao W, Huang F, Tian ZX, Yu DD. Down regulation of a gibberellin 3 beta-hydroxylase enhances photosynthesis and increases seed yield in soybean. New Phytol. 2022;235:502–17.

21. Hu L, Liu P, Jin Z, Sun J, Weng Y, Chen P, Du S, Wei A, Li Y. A mutation in CsHY2 encoding a phytochromobilin (PΦB) synthase leads to an elongated hypocotyl 1 (*elh1*) phenotype in cucumber (*Cucumis sativus L.*). Theor Appl Genet. 2021;134:2639–52.
22. Hu Y, Liu Y, Tao JJ, Lu L, Jiang ZH, Wei JJ, Wu CM, Yin CC, Li W, Bi YD, Lai YC, Wei W, Zhang WK, Chen SY, Zhang JS. GmJAZ3 interacts with GmRR18a and GmMYC2a to regulate seed traits in soybean. J Integr Plant Biol. 2023;65:1983–2000.
23. Jin HL, Duan S, Zhang P, Yang Z, Zeng Y, Chen Z, Hong L, MLi, Luo L, Chang Z, Hu J, Wang HB. Dual roles for CND1 in maintenance of nuclear and chloroplast genome stability in plants. Cell Rep. 2023;42(3):112268.
24. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15.
25. Kohchi T, Mukougawa K, Frankenberg N, Masuda M, Yokota A, Lagarias JC. The Arabidopsis HY2 gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. Plant Cell. 2001;13:425–36.
26. Kong KK, Xu MG, Wang YQ, Kong JJ, Amin GM, Zhao TJ. Gene mapping and character performance of A Yellow-green Leaf mutant NJ9903-5 in Soybean. Soybean Sci. 2017;36:494–501.
27. Koornneef M, Rolff E, Spruit CJP. Genetic control of light-inhibited hypocotyl elongation in *Arabidopsis thaliana* (L.). Heynh Zeitschrift für Pflanzenphysiologie. 1980;100:147–60.
28. Legris M, Ince YC, Fankhauser C. Molecular mechanisms underlying phytochrome-controlled morphogenesis in plants. Nat Commun. 2019;10:5219.
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth, G, Abecasis G, Durbin R and Genome Project Data Processing. The sequence alignment/map format and SAM tools. Bioinformatics. 2009;25:2078–2079.
30. Li N, Xu R, Li Y. Molecular networks of seed size control in plants. Annu Rev Plant Biol. 2019;70:435–63.
31. Li J, Zhang Y, Ma R, Huang W, Hou J, Fang C, Wang L, Yuan Z, Sun Q, Dong X. Identification of ST1 reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. Plant Biotechnol J. 2022;20:1110–21.
32. Lilian F, B., and Brian, W., Diers. Agronomic Evaluation of a High Protein Allele from PI407788A on Chromosome 15 across Two Soybean Backgrounds. Crop Sci. 2017;57:2972–8.
33. Liu Y, Jafari F, Wang H. Integration of light and hormone signaling pathways in the regulation of plant shade avoidance syndrome. aBIOTECH. 2021;2:131–45.
34. Liu SY, Li Y, Wei J. Research progress on the regulatory mechanisms of plant seed development. Agric Technol. 2022a;42(18):37–40.
35. Liu M, Wang Y, Nie Z, Gai J, Bhat JA, Kong J, Zhao T. Double mutation of two homologous genes YL1 and YL2 results in a leaf yellowing phenotype in soybean [*Glycine max (L.) Merr*]. Plant Molecular Bio. 2020;103:527–43.
36. Liu S, Zhang M, Feng F, Tian ZX. Toward a 'Green Revolution' for soybean. Mol Plant. 2020;13:688–97.
37. Love MI, Huber W, Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
38. Lu S, Dong L, Fang C, Liu S, Cheng Q, Kong L, Chen L, Su T, Nan H, Zhang D, Zhang L, Wang Z, Yang Y, Yu D, Liu X, Yang Q, Lin X, Tang Y, Zhao X, Yang X, Tian C, Xie Q, Li X, Yuan X, Tian Z, Liu B, Weller JL, Kong FJ. Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat Genet. 2020;52:428–36.
39. Lu X, Xiong Q, Cheng T, Li QT, Liu XL, Bi YD, Li W, Zhang WK, Ma B, Lai YC, Du WG, Man WQ, Chen SY, Zhang JS. A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. Mol Plant. 2017;10:670–84.
40. Miao L, Yang SN, Zhang K, He JB, Wu CH, Ren YH, Gai JY, Li Y. Natural variation and selection in *GmSWEET39* affect soybean seed oil content. New Phytol. 2020;225:1651–66.
41. Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Westoby M. A brief history of seed size. Science. 2005;307:576–80.
42. Muramoto T, Kohchi T, Yokota A, Hwang I, Goodman HM. The *Arabidopsis* photomorphogenic mutant hy1 is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase. Plant Cell. 1999;11:335–48.
43. Nguyen CX, Paddock KJ, Zhang Z, Stacey MG. GmKIX8-1 regulates organ size in soybean and is the causative gene for the major seed weight QTLqSw17-1. New Phytol. 2020;229:920–34.
44. Pincovici S, Cochavi A, Kamieli A, Ephrth J, Rachmilrvitch S. Source-sink relations of sunflower plants as affected by a parasite modifies carbon allocations and leaf traits. Plant Sci. 2018;271:100–7.
45. Prenger EM, Yates J, Rouf Mian MA, Buckley B, Boerma HR, Li Z. Introgression of a high protein allele into an elite soybean cultivar results in a high-protein near-isogenic line with yield parity. Crop Sci. 2019;59:2498–508.
46. Qin DD, Li MF, Xu FC. Agronomic traits of barley and its regulatory genes in ygl. J Wheat Crop. 2019;39:1–6.
47. Raes J, Van D, Peer Y. Functional divergence of proteins through frameshift mutations. Trends Genet. 2005;21:428–31.
48. Rorigues J, Inze D, Nelissen H, Saibo NJM. Source-Sink Regulation in Crops under Water Deficit. Trends Plant Sci. 2019;7:652–63.
49. Saez JV, Mariotti JA, Vega CRC. Source-sink relationships during early crop development influence earliness of sugar accumulation in sugarcane. J Exp Bot. 2019;19:5157–71.
50. Sawers RJ, Linley PJ, Gutierrez-Marcos JF, Delli-Bovi T, Farmer PR, Kohchi T, Terry MJ, Brutnell TP. The Elm1 (ZmHy2) gene of maize encodes a phytochromobilin synthase. Plant Physiol. 2004;136:2771–81.
51. Sebolt AM, Shoemaker RC, Diers BW. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Sci. 2000;40:1438–1444.
52. Shimomura M, Kanamori H, Komatsu S, Namiki N, Mukai Y, Kurita K, Kamatsuki K, Ikawa H, Yano R, Ishimoto M, Kaga A, Katayose Y. The *Glycine max* cv. Enrei genome for improvement of Japanese soybean cultivars. Int J Genomics. 2015;15:1–12.
53. Singh AK, Fu DQ, El-Habbak M, Navarre D, Ghabrial S, Kachroo A. Silencing genes encoding omega-3 fatty acid desaturase alters seed size and accumulation of Bean pod mottle virus in soybean. Mol Plant Microbe Interact. 2011;24:506–15.
54. Sun F, Wang K. Research on the economic relationship between farming and animal husbandry in the composite eco-economic system of the agro-pastoral zone. Res Agric Modernization. 2007;(2):221–224.
55. Sun XQ, Wang B, Xiao YH, Wan CM, Deng XJ, Wang PR. Genetic analysis and fine mapping of gene ygl98 for yellow-green leaf of rice. Acta Agron Sin. 2011;37:991–7.
56. Terry MJ. Phytochrome chromophore-deficient mutants. Plant Cell Enviro. 1997;20:740–745.
57. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–515.
58. Wang JL, Chen Y. Ecological distribution of soybean variety resources in China. J Northeast Agric Univ. 1983;(2):1–8.
59. Wang M, Li WZ, Fang C, Xu F, Liu YC, Wang Z, Yang R, Zhang M, Liu SL, Lu SJ, Lin T, Tang JY, Wang YQ, Wang HR, Lin H, Zhu BG, Chen MS, Kong FJ, Liu BH, Zeng DL, Jackson SA, Chu CC, Tian ZX. Parallel selection on a dormancy gene during domestication of crops from multiple families. Nat Genetics. 2018;50:1435–41.
60. Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu YC, Liu Z, Frommer WB, Ma JF, Guan CLQ, Y.F., Shou, H.X., and Tian, Z.X. Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. Natl Sci Rev. 2020;7:1776–86.
61. Wang SD, Liu SL, Wang J, Yokosho K, Zhou B, Yu YC, Liu Z, Frommer WB, Ma JF, Chen LQ, Guan YF, Shou HX, Tian ZX. Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. Natl Sci Rev. 2020;7:1776–86.
62. Wang HR, Zang Y, Yu CM, Dong QZ, Li WW, Hu KF, Zhang MM, Xue H, Yang MP, Song JL, Wang L, Yang XY, Qiu LJ. Fine mapping of yellow-green leaf gene HY2 in soybean (Glycine max L.). Acta Agro Sinica. 2022;48:791–800.
63. Watanabe S, Xia Z, Hideshima R, Tsubokura Y, Sato S, Yamanaka N, Takahashi R, Anai T, Tabata D, Kitamura K, Harada K. A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genet. 2011;188:395–407.
64. Wei SM, Yong B, Jiang BW, An ZH, Wang Y, Li BB, Yang C, Zhu WW, Chen QS, He CY. A loss-of-function mutant allele of a glycosyl hydrolase gene has been co-opted for seed weight control during soybean domestication. J Integr Plant Biol. 2023. https://doi.org/10.1111/jipb.13559.

65. Xia Y, Li Z, Wang J, Li Y, Ren Y, Du J, Song Q, Ma S, Song Y, Zhao H, Yang Z, Zhang G, Niu N. Isolation and identification of a TaTDR-like wheat gene encoding a bHLH domain protein, which negatively regulates chlorophyll biosynthesis in Arabidopsis. Int J of Mol Sci. 2020;21:629–42.
66. Xu YC, Guo YL. Less is more, natural loss-of-function mutation is a strategy for adaptation. Plant Commun. 2020;1:100–3.
67. Yang HY, Wang WB, He QY, Xiang SH, Gai JY. Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. Theor Appl Genet. 2019;132:2793–807.
68. Zeng XM. The startegic planning of soybean industry development in China. Chin J Agricult Res Region Plan. 2017;38:89–97.
69. Zhang HY, Goettel W, Song QJ, Jiang H, Hu ZB, Wang ML, An YC. Selection of GmSWEET39 for oil and protein improvement in soybean. PLoS Genet. 2020;16:1–18.
70. Zhang B, Li C, Li Y, Yu H. Mobile TERMINAL FLOWER1 determines seed size in *Arabidopsis*. Nat Plants. 2020;6:1146–57.
71. Zhang B, Qiu LJ, Chang RZ. Research progress on genetic diversity and core collection of soybean cultivars. Crop J. 2003;(3):46–49.
72. Zhang Q, Yao YB, Li YH. Research progress and prospects on monitoring, forecasting and disaster reduction technologies of meteorological drought in Northwest China. Adv Earth Sci. 2015;30(2):196–213.
73. Zhang Z, Yang S, Wang Q, Yu H, Zhao B, Wu T, Tang K, Ma J, Yang X, Feng X. Soybean GmHY2a encodes a phytochromobilin synthase that regulates internode length and flowering time. J Exp Bot. 2022;73(19):6646–62.
74. Zhou Z, Bi G, Zhou JM. Luciferase complementation assay for protein-protein interactions in plants. Curr Protoc Plant Biol. 2018;3:42–50.
75. Zhu X, Wang H, Li Y, Rao D, Wang F, Gao Y, Zhong W, Zhao Y, Wu S, Chen X, Qiu H, Zhang W, Xia Z. A Novel 10-Base Pair Deletion in the First Exon of *GmHY2a* Promotes Hypocotyl Elongation, Induces Early Maturation, and Impairs Photosynthetic Performance in Soybean. Int J Mol Sci. 2024;25(12):6483.
76. Zhu WW, Yang C, Yong B, Wang Y, Li BB, Gu YZ, Wei SM, An ZH, Sun WK, Qiu LJ, He CY. An enhancing effect attributed to a nonsynonymous mutation in SOYBEAN SEED SIZE 1, a SPINDLY-like gene, is exploited in soybean domestication and improvement. New Phytol. 2022;236:1375–92.
77. Zou T, Xiao Q, Li W, Luo T, Yuan G, He Z, Liu M, Li Q, Xu P, Zhu J, Liang YY, Deng QM, Wang SQ, Zheng AP, Wang LX, Li P, Li SC. OsLAP6/OsPKS1, an orthologue of *Arabidopsis* PKSA/LAP6, is critical for proper pollen exine formation. Rice. 2017;10:53–9.

Publisher's Note