CellPress
Partner Journal

Molecular Plant
Research article

The domestication-associated *L1* gene encodes a eucomic acid synthase pleiotropically modulating pod pigmentation and shattering in soybean

Xiangguang Lyu[1,5], Ying-hui Li[1,2,5], Yanfei Li[1,2,4], Delin Li[1,2], Chao Han[3], Huilong Hong[1,2], Yu Tian[1,2], Lida Han[3], Bin Liu[1,*] and Li-juan Qiu[1,2,*]

[1]The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, P.R. China

[2]Key Laboratory of Grain Crop Genetic Resources Evaluation and Utilization, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, P.R. China

[3]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, P.R. China

[4]Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People''s Republic of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, P.R. China

[5]These authors contributed equally to this article.

***Correspondence: Bin Liu (liubin05@caas.cn), Li-juan Qiu (qiulijuan@caas.cn)**

https://doi.org/10.1016/j.molp.2023.06.003

ABSTRACT

Pod coloration is a domestication-related trait in soybean, with modern cultivars typically displaying brown or tan pods, while their wild relative, *Glycine soja*, possesses black pods. However, the factors regulating this color variation remain unknown. In this study, we cloned and characterized *L1*, the classical locus responsible for black pods in soybean. By using map-based cloning and genetic analyses, we identified the causal gene of *L1* and revealed that it encodes a hydroxymethylglutaryl-coenzyme A (CoA) lyase-like (HMGL-like) domain protein. Biochemical assays showed that L1 functions as a eucomic acid synthase and facilitates the synthesis of eucomic acid and piscidic acid, both of which contribute to coloration of pods and seed coats in soybean. Interestingly, we found that *L1* plants are more prone to pod shattering under light exposure than *l1* null mutants because dark pigmentation increases photothermal efficiency. Hence, pleiotropic effects of *L1* on pod color and shattering, as well as seed pigmentation, likely contributed to the preference for *l1* alleles during soybean domestication and improvement. Collectively, our study provides new insights into the mechanism of pod coloration and identifies a new target for future *de novo* domestication of legume crops.

Key words: soybean, domestication, pod pigmentation, pod shattering, pleiotropic effects

Lyu X., Li Y.-h., Li Y., Li D., Han C., Hong H., Tian Y., Han L., Liu B., and Qiu L.-j. (2023). The domestication-associated *L1* gene encodes a eucomic acid synthase pleiotropically modulating pod pigmentation and shattering in soybean. Mol. Plant. **16**, 1178–1191.

INTRODUCTION

Soybean (*Glycine max*) is one of the primary sources of plant-based protein and oil for human consumption, animal feed, and industrial processing (Sá et al., 2020). It is believed that soybean was domesticated from annual, wild *G. soja* in East Asia 5000–9000 years ago (Zhou et al., 2015; Sedivy et al., 2017). Humans have selectively modified the physical and genetic characteristics of *G. soja* throughout the complex evolutionary process of soybean domestication and genetic improvement to cultivate desirable traits (Doebley et al., 2006). Therefore, a deeper understanding of the morphological and genetic alterations from wild *G. soja* to domesticated soybean provides an excellent window into soybean domestication as well as valuable information for improving soybean yield and quality (Lu et al., 2020; Wang et al., 2020).

The process of soybean domestication has involved the selection of a wide range of traits, including coloration of the mature pods and seed coats (Gillman et al., 2011; Gao et al., 2021). Wild *G. soja* pods are typically black, whereas domesticated soybean pods tend to be brown or tan (Woodworth, 1923). In addition, it

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

 © 2023 The Author.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

has been reported that black pods are often accompanied by smudged seeds. This may result from the pigment in black pods permeating the parenchymatous cells of the seed coat, leading to a commercially unattractive appearance (Owen, 1927, 1928; Williams, 1952). Classically, genetic research has shown that the color of soybean pods is controlled by the epistatic interaction of two loci, *L1* and *L2*. The *L1* allele results in black pods, while the *l1* alleles cause brown or tan pods in the *L2* or *l2* background, respectively (Woodworth, 1923; Bernard, 1967; Shoemaker and Specht, 1995). However, the genes responsible for pod color remain unknown.

Genes responsible for pigmentation often demonstrate close linkage or pleiotropic effects on key traits (Cieslak et al., 2011; Paauw et al., 2019). In soybean, pod color is possibly associated with pod shattering, a critical trait in the process of domestication (Dong et al., 2014; Funatsuki et al., 2014; Sedivy et al., 2017). In a natural environment, pod shattering is essential in seed dispersal, providing sufficient space for the offspring to grow and increasing their chances of survival (Parker et al., 2021). However, uncontrolled pod dehiscence before harvest can result in devastating yield losses in agriculture, particularly in tropical regions (Tiwari and Bhatia, 1995; Tukamuhabwa et al., 2002). Hence, during soybean domestication, humans tend to select plants that are resistant to pod shattering. Pod shattering is a complex trait influenced by several inherent factors, including the chemical composition, morphology, and anatomical structures of the pods, as well as the architecture of the plant (Bara et al., 2013; Dong et al., 2014). Furthermore, environmental factors such as high temperatures, rapid temperature fluctuations, low humidity, and light-intensity levels have a strong impact on pod shattering (Krisnawati and Adie, 2019; Liu et al., 2022). Therefore, pod shattering in soybean depends on the complicated interactions between genes and the environment, making it challenging to detect and identify genes related to pod shattering that are influenced by environmental factors under field conditions.

In this study, we utilized map-based cloning to discover the *L1* gene, which encodes a Fabaceae-specific hydroxymethylglutaryl-coenzyme A (CoA) lyase-like (HMGL) domain protein that is necessary for the biosynthesis of eucomic acid and piscidic acid in soybean pods. Furthermore, our findings show that the *L1* gene has pleiotropic effects on the coloration of pods and seed coats, as well as pod shattering under light conditions, compared with *l1* null mutants, suggesting that *L1*-mediated dark pod pigmentation intensifies the impact of photothermal effects on pod shattering.

RESULTS

Map-based cloning of the *L1* candidate gene

To understand the molecular mechanisms underlying the change from black to non-black pods, we performed a cross between two near-isogenic lines (NILs), R171-BLk with black pods and R171-LBn with light-brown pods. These NILs were derived from recombinant inbred line (RIL) R171 that was created by crossing ZP03-5373 and ZhongHuang 13 (Supplemental Figure 1). Genetic analysis of the F_2 population revealed a 3:1 segregation of black-podded and light-brown-podded plants ($\chi^2(1:3) = 0.05 < \chi^2\ 0.05 = 3.84$) (Figure 1A), suggesting that dark pigmentation of the pod is controlled by a single dominant gene. Previous studies have shown a correlation between black pods and smudged seed coats in soybeans (Owen, 1928). Consistent with these findings, we observed smudged seed coats in R171-BLk but not in R171-LBn (Supplemental Figure 1). Additionally, the smudged seed coats cosegregated with black pods in the F_2 and $F_{2:4}$ populations (Figure 1A), suggesting that the *L1* locus contributes to both black pod pigmentation and the smudged appearance of seed coats.

To identify the causal gene responsible for black pod coloration in soybeans, we performed bulked segregant RNA sequencing (BSR-seq) using the leaves of 30 black-podded and 30 light-brown-podded plants from the F_2 population. After stringent quality control and data filtration, we identified a total of 149 157 high-quality SNPs. Employing an allele frequency difference >0.8 and a linkage probability >0.05 as filtration criteria, we found 229 SNP markers that displayed high probability in association with black pod coloration (Supplemental Table S1). Of these, 185 SNP markers (80.8%) were localized to a 1.7-Mb genomic interval spanning the 36.5–38.2 Mb region of chromosome 19, which overlaps with the *L1* locus (Figure 1B) (Bandillo et al., 2017; Chang et al., 2021).

To further narrow down the location of the *L1* locus, we conducted recombinant screening of the $F_{2:4}$ population ($n = 2662$) and identified a 16.2-kb region between markers M5 and M8 that contains only a part of a putative gene, *Glyma.19G120400* (Wm82.a2.v1) (Figure 1C). We sequenced *Glyma.19G120400* in two parental lines and identified a C-to-T variant at the 91st nucleotide in the first exon, leading to an amino acid substitution from arginine to cysteine (p.R31C) (Figure 1C). Real-time reverse transcription–PCR (qRT–PCR) analysis showed that *Glyma.19G120400* was predominantly expressed in pods and that its expression level increased gradually as the pods developed and matured (Figure 1D). However, expression of this gene was not significantly different between the NILs (R171-BLk and R171-LBn) (Supplemental Figure 2A), suggesting that the C91T variant may affect gene function at the post-transcriptional level. These findings indicate that *Glyma.19G120400* is a strong candidate gene for the *L1* locus.

Genetic confirmation of the *L1* gene responsible for black pod coloration

To test the role of this candidate gene, we utilized CRISPR–Cas9 technology to knock out *Glyma.19G120400* in the R171-BLk background. We obtained multiple independent mutant lines, including *l1-cr1* with a single-base insertion and *l1-cr2* with a 68-base insertion in the first exon. Phenotypic analysis revealed that dysfunction of the *Glyma.19G120400* gene led to a change in pod color from black to light brown (Figure 2A and 2B), indicating that *Glyma.19G120400* is the causal gene for the *L1* locus. Hereafter, we refer to *Glyma.19G120400* in R171-BLk as *L1* and its loss-of-function variant in R171-LBn as *l1-1*. Interestingly, knockout of the *L1* gene also eliminated the smudged seed coat phenotype (Figure 2A), supporting the idea that the *L1* gene affects both pod and seed coat color.

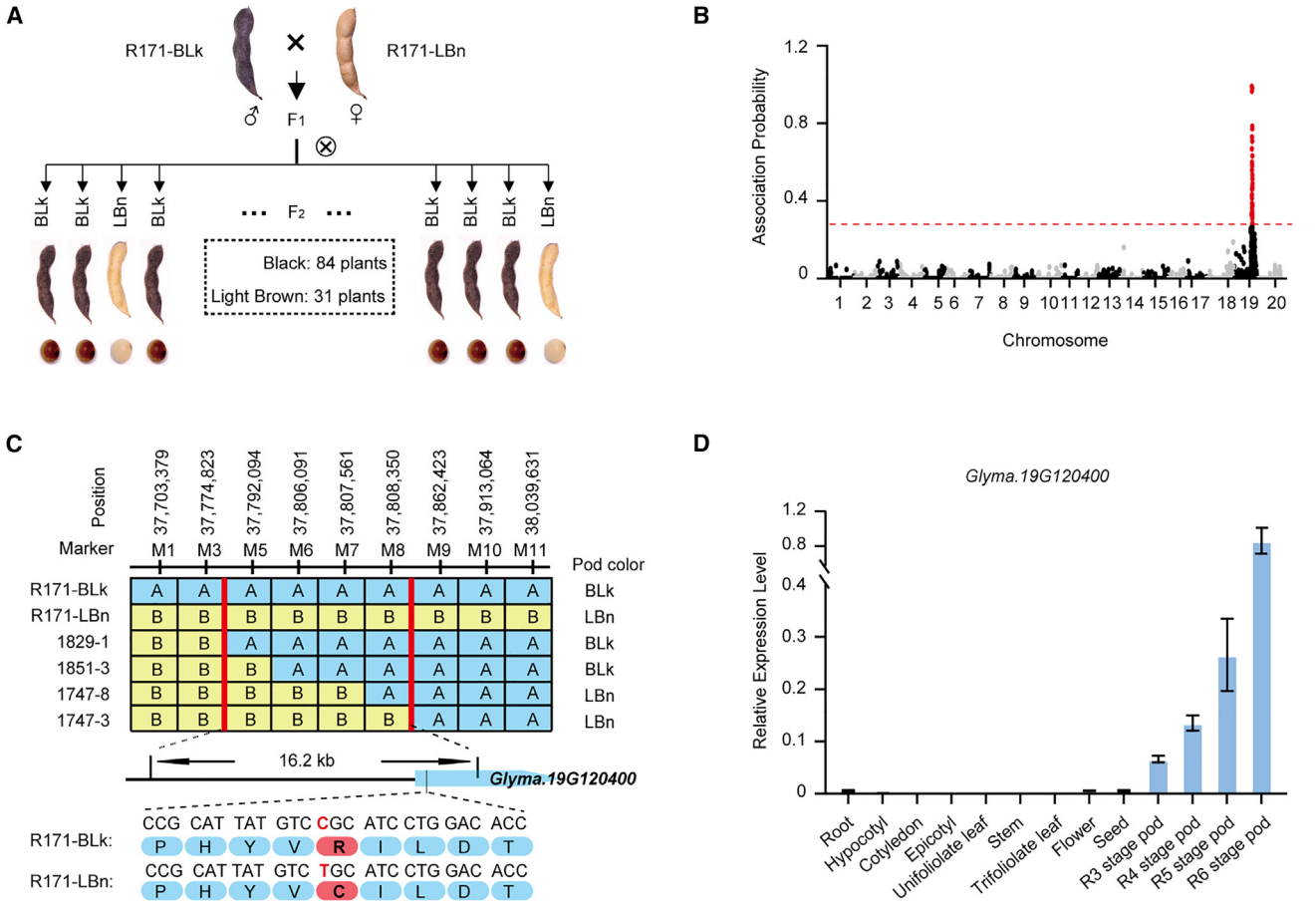

Figure 1. Identification of the *L1* gene by map-based cloning.

(A) Segregation of soybean pod and seed coloration in the population derived from a cross between the NILs R171-BLk (black pods) and R171-LBn (light-brown pods).

(B) Bulked segregant RNA sequencing (BSR-seq) analysis of the two extreme pools containing leaves from 30 black and 30 light-brown podded plants. The dashed line indicates the threshold association probability value (0.25) for the putative *L1* locus.

(C) The M5 and M8 markers delimit the *L1* locus to a 16.2-kb region on chromosome 19, which contains a nucleotide variant in the *Glyma.19g120400* gene between R171-BLk and R171-LBn. The letter A denotes *L1* homozygous, and the letter B denotes *l1* homozygous.

(D) The expression pattern of *Glyma.19g120400* in R171-BLk was determined by qPCR. Data are presented as the mean ± SD of three biological replicates, with *GmActin* as the internal control.

To further examine the effects of the *L1* gene, we amplified the *L1* coding DNA sequence (CDS) from R171-BLk and transformed TL1, a cultivar with light-colored pods, with the *L1* CDS under the control of the *CaMV 35S* promoter. Phenotype analysis showed that two independent transgenic overexpression lines, *L1-OE1* and *L1-OE2*, produced black pods instead of the light-brown pods observed in wild-type TL1 plants (Figure 2C and 2D). This further confirmed that *Glyma.19G120400* is the *L1* gene.

L1 encodes an HMGL-like domain protein

Upon analysis of the protein sequence, we discovered that the L1 protein features a highly conserved N-terminal HMGL-like domain accompanied by a C-terminal LeuA regulatory domain (Supplemental Figure 2B) (Chen and Frantom, 2019). The HMGL-like domain proteins are categorized as part of the DRE-TIM metallolyase superfamily, which facilitate cleavage or condensation reactions of C–C bonds and can be further divided into four distinct subgroups: Claisen-condensation-like (CC-like), carboxylase-like, lyase-like, and aldolase-like (Conte and Frantom, 2020).

We generated a phylogenetic tree containing 203 members by aligning L1 homologous proteins from several Archaeplastida species (Figure 2E). The analysis revealed that the L1 protein is closely related to α-isopropylmalate synthase (IPMS) and methylthioalkylmalate synthase (MAM), which belong to the CC-like subgroup (Felnagle et al., 2012). Within this subgroup, enzymes catalyze a condensation reaction between acetyl-CoA and various α-keto acids, resulting in the production of a 2-malate derivative and CoA (Conte and Frantom, 2020). Notably, this subgroup exhibits significant functional diversity, with at least six enzymes showing distinct substrate specificities to date. Our analysis revealed that the 203 taxa in the phylogenetic tree were predominantly classified into three subgroups, each potentially possessing different substrate specificities (Figure 2E). For instance, IPMS1/2 is an essential enzyme in leucine biosynthesis, as it catalyzes the formation of 2-isopropylmalate from

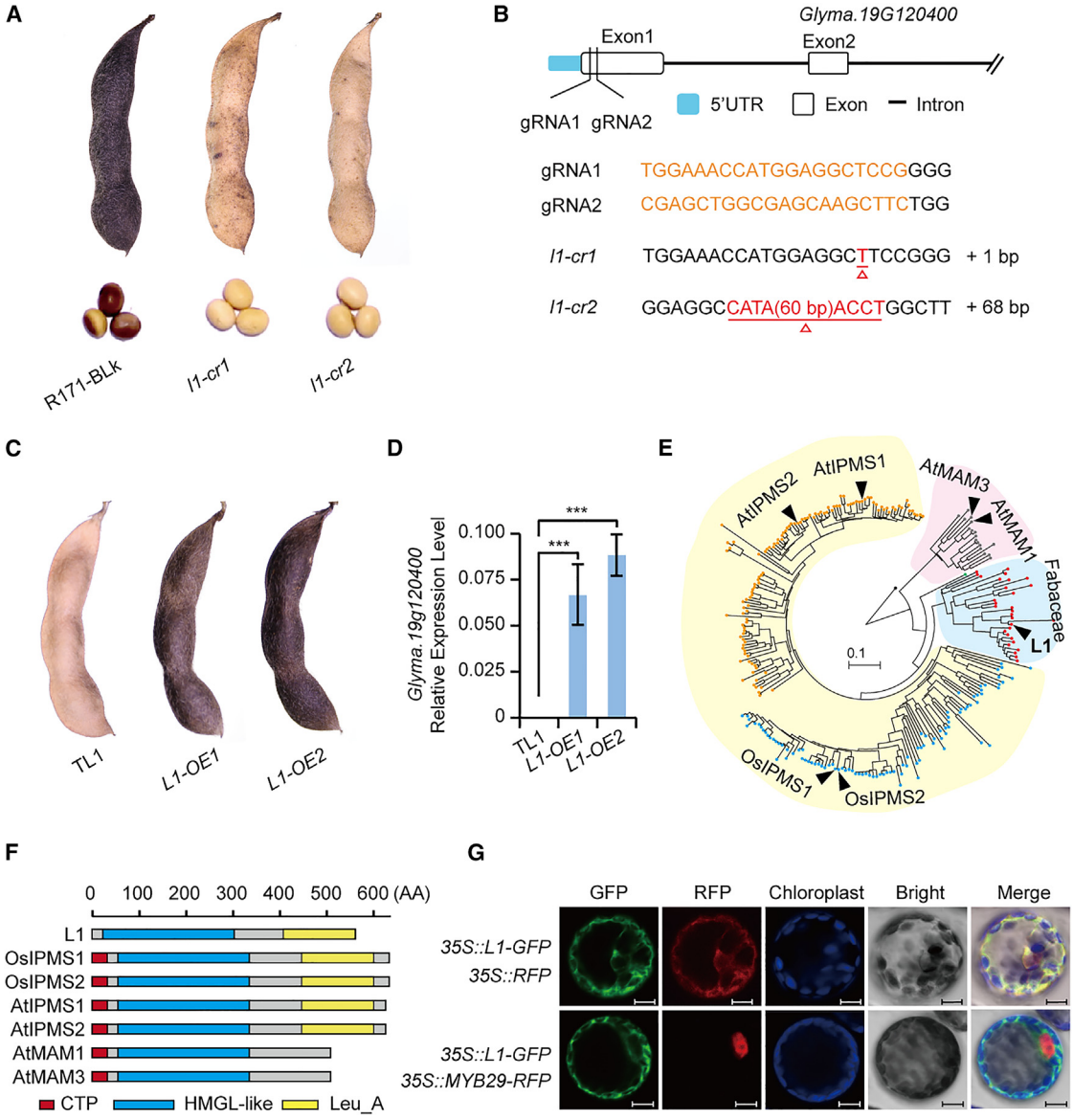

Figure 2. Characterization of the *L1* gene.

(A) The pod and seed phenotype of the wild-type plant (R171-BLk) and CRISPR–Cas9-engineered *Glyma.19g120400* mutants (*l1-cr1* and *l1-cr2*).

(B) Schematic diagram showing the genomic structure of the *Glyma.19g120400* gene. The orange letters indicate two, single guide RNAs targeting the first exon of the candidate gene *L1*. The red letters above the lines indicate the positions of the nucleotide insertions in the CRISPR–Cas9-mediated *l1* mutants. Light-blue boxes represent 5′ untranslated regions (UTRs); white boxes denote exons; black lines are introns.

(C) Pod color of the wild-type plant (TL1) and two transgenic *Glyma.19g120400.1*-overexpression lines.

(D) Expression levels of the *L1* gene in the leaves of the indicated plants. Data are presented as the mean ± SD of three biological replicates, with *GmActin* as the internal control. Statistically significant differences were determined by a two-tailed Student's *t*-test (***$p < 0.001$).

(E) Neighbor-joining phylogenetic tree of L1 and its homologs from Archaeplastida. The scale bar represents branch length.

(F) The protein domain of L1 and its homologs from rice and *Arabidopsis*. The red boxes represent chloroplast transit peptide (CTP); the blue rectangles denote the HMGL-like domain; the yellow rectangles indicate the LeuA dimerization domain.

(G) Subcellular localization of L1-GFP fusion proteins in *Nicotiana benthamiana* protoplasts. Red fluorescent protein (RFP) alone was used as a cytoplasmic marker, and MYB29-RFP was used as a nuclear marker. Scale bars represent 10 μm.

2-ketoisovalerate and acetyl-CoA (De Kraker et al., 2007; He et al., 2019). Additionally, MAM1/3 contributes to glucosinolate formation by condensing ω-methylthio-2-oxoalkanoic acids with acetyl-CoA (Knill et al., 2008).

Despite sharing 59%–67% sequence identity with its *Arabidopsis* and rice homologs, the L1 protein forms a separate subgroup that comprises members solely derived from Fabaceae species (Figure 2E). In contrast to its homologs, the L1 protein lacks chloroplast transit peptides and is localized in the cytoplasm (Figure 2F and 2G; Supplemental Figure 2C) (Field et al., 2004; Textor et al., 2007). These observations suggest that the L1 protein evolved differently, possibly resulting in distinct substrate specificities for the condensation reaction in soybean.

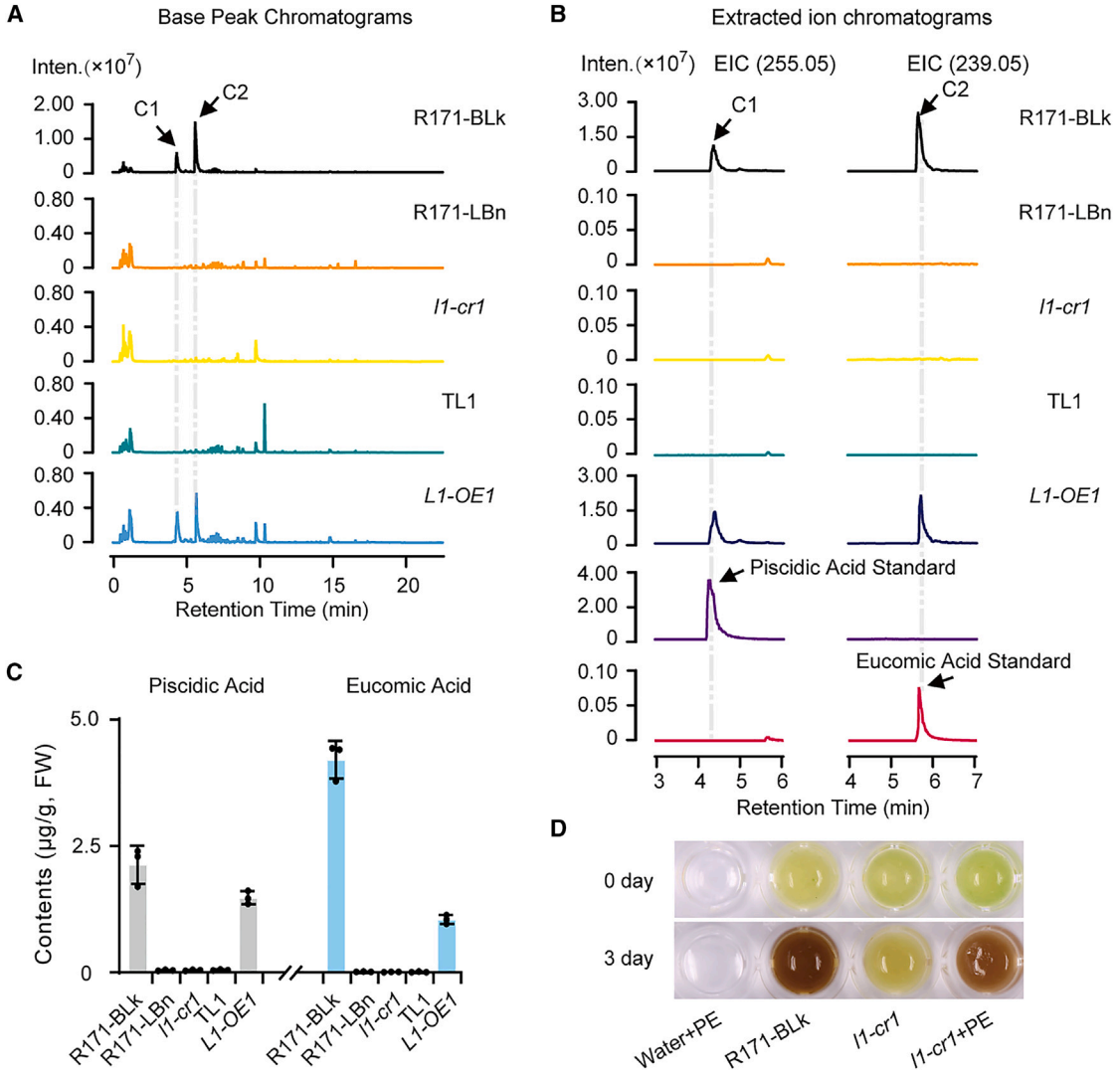

Figure 3. Analysis of metabolites in soybean pod.

(A) LC–MS analysis of extracts derived from immature pods of two parent lines (R171-BLk and R171-LBn), *l1* mutant (*l1-cr1*), wild-type plant (TL1), and *L1*-overexpression transgenic line (*L1-OE1*).

(B) Comparative LC–MS extracted ion chromatograms (EICs) of R171-BLk, R171-LBn, *l1-cr1*, and *L1-OE1* extracts with piscidic acid and eucomic acid standards.

(C) The content of piscidic acid and eucomic acid detected in pod extracts from two parent lines (R171-BLk and R171-LBn), *l1* mutant (*l1-cr1*), wild type (TL1), and *L1*-overexpression transgenic line (*L1-OE1*).

(D) Pigmentations are derived from piscidic acid and eucomic acid. PE stands for piscidic acid and eucomic acid.

Piscidic acid and eucomic acid are responsible for pod pigmentation

To investigate the compounds responsible for diverse pod colors, we utilized an ultra-performance liquid chromatography quadrupole time-of-flight mass spectrometry (UPLC-Q-TOF-MS) system to analyze and compare the metabolites in immature pods with different alleles. Our metabolomic profiling revealed the presence of two major components (C1 and C2) in R171-BLk and *L1-OE1* pod extracts, which were absent in R171-LBn, *l1-cr1*, or TL1 pod extracts (Figure 3A). Through fragmentation analysis, we deduced C1 to be piscidic acid with a mass-to-charge ratio (*m/z*) of 255.0534 ($[M-H]^-$), while C2 was identified as likely eucomic acid with an *m/z* ratio of 239.0586 ($[M-H]^-$) (Supplemental Figure 3A–3D). For further validation, we compared the column retention times and MS data of these two compounds to commercially available standards, which unequivocally confirmed them as piscidic acid and eucomic acid (Figure 3B; Supplemental Figure 3E and 3F).

Piscidic acid and eucomic acid were first isolated from the bark of the Jamaica dogwood tree and the bulbs of *Eucomis punctata*, respectively (Bridge et al., 1948; Heller and Tamm, 1974). We found that these compounds were abundant in black pods but absent from tan or brown pods (Figure 3C). These compounds are phenolic acids with antioxidant properties. Certain phenolic acids and their by-products can undergo polymerization, resulting in the formation of pigmented

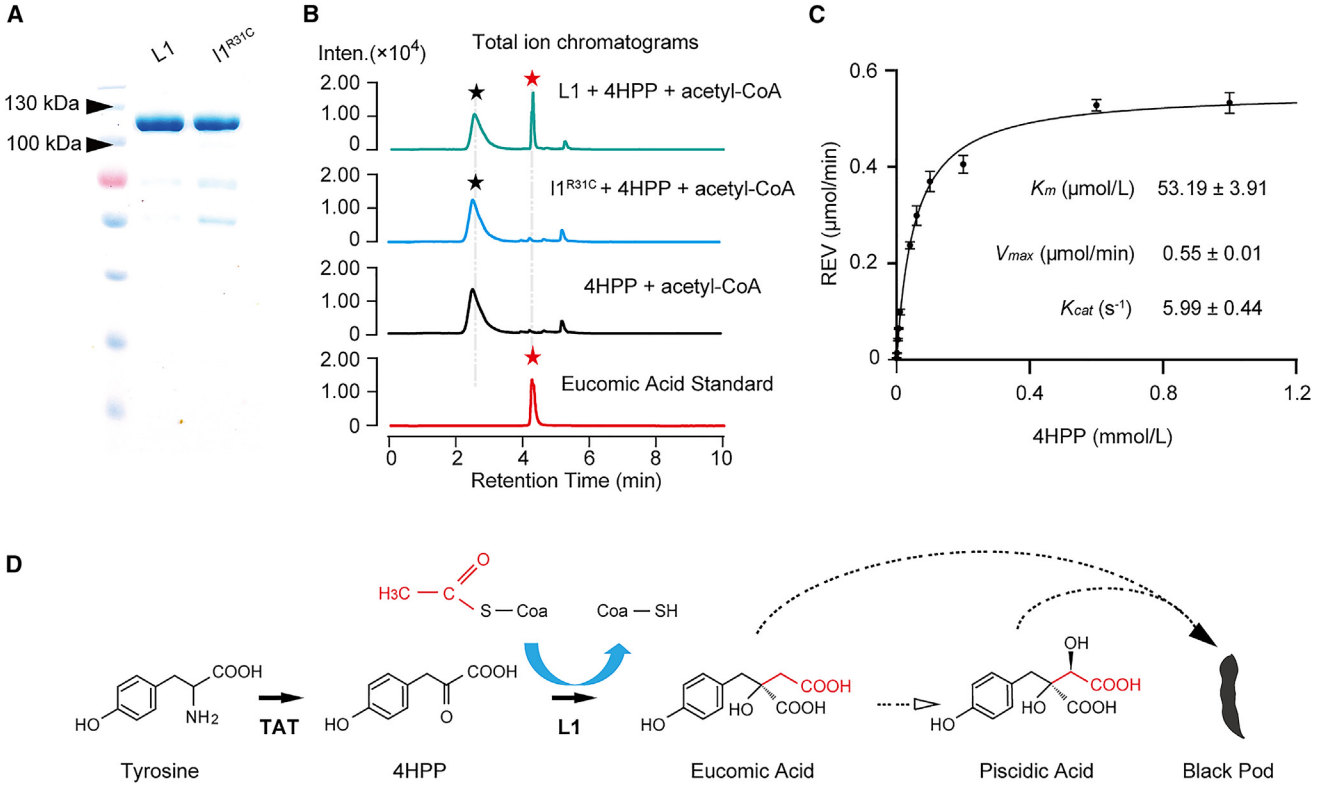

Figure 4. L1 is required for eucomic and piscidic acid biosynthesis.

(A) Polyacrylamide gel electrophoresis showing the purified His-tagged L1 and I1^{R31C} proteins.

(B) LC–MS analysis of the products obtained from the enzymatic reactions of wild-type L1 and I^{R31C} using acetyl-CoA and 4-hydroxyphenylpyruvic acid (4-HPP) as substrates.

(C) The L1 protein follows Michaelis–Menten kinetics. L1 activity was measured in Tris–HCl buffer (pH 8.0) containing 1.5 nmol L1 protein, 5 mM acetyl-CoA, and different concentrations of 4-HPP. Data are presented as the mean ± SD of three biological replicates.

(D) Proposed pathway for the biosynthesis of black pigments in soybean pods. Enzymes used for pathway reconstitution are shown in bold. TAT represents tyrosine aminotransferase. The enzyme responsible for the oxidation of eucomic acid to piscidic acid is unknown.

compounds such as lignin and tannin. These acids can also act as potential agents for protein crosslinking, resulting in phenol–protein interactions and corresponding color changes (Rohn et al., 2006; Prodpran et al., 2012; Insaward et al., 2015). It is likely that the dark color of R171-BLk and *L1-OE1* pods is due to the abundance of piscidic acid and eucomic acid. Consistent with this hypothesis, we observed that the light-green extraction solution obtained from the immature R171-BLk pods became progressively darker over 3 days, while the extract solution from *I1-cr1* pods retained its light-green color (Figure 3D).

To investigate the effect of piscidic acid and eucomic acid on pod color, we supplemented these compounds in the *I1-cr1* pod extraction solution at a concentration equivalent to that in the R171-BLk pod extraction solution, while water was used as a negative control. The results demonstrate that adding piscidic and eucomic acids led to progressive darkening of the *I1-cr1* pod extraction solution over a period of 3 days, reminiscent of the black pod extraction solution (Figure 3D). Conversely, no color change was observed when these compounds were added to water (Figure 3D). These results indicate that eucomic acid and piscidic acid are crucial in the development of black pigmentation in mature soybean pods.

L1 functions as a eucomic acid synthase

Considering that eucomic acid is a 2-malate derivative, we postulated that the L1 protein might be capable of synthesizing this compound via a Claisen condensation reaction involving 4-hydroxyphenylpyruvate (4-HPP), a 2-oxo-monocarboxylic acid intermediate in the tyrosine metabolic pathway, and acetyl-CoA. To test this hypothesis, we conducted *in vitro* assays to assess the activity of the L1 protein. In brief, we cloned the CDS of the *L1* and *I1-1* genotypes into the pCold-TF vector, transformed the constructs into *Escherichia coli* BL21, and purified the recombinant proteins using nickel–nitrilotriacetic acid (Ni-NTA) agarose (Figure 4A).

The purified recombinant L1 protein exhibited effective eucomic acid synthase activity in an assay mixture containing 4-HPP and acetyl-CoA, as expected (Figure 4B). In contrast, the I1^{R31C} protein was inactive, indicating that the amino acid changes (p.R31C) disrupted the function of the L1 protein (Figure 4B). Furthermore, biochemical characterization of L1 revealed that its K_m for 4-HPP was 53.19 ± 3.91 μM, and its catalytic efficiency (k_{cat}/K_m) was 13.9 ± 1.8 mM^{-1} s^{-1} (Figure 4C; Supplemental Figure 4A and 4B). These values are similar to the catalytic efficiencies reported for 4-hydroxyphenylpyruvate dioxygenase and 4-hydroxyphenylpyruvate reductase, which are responsible

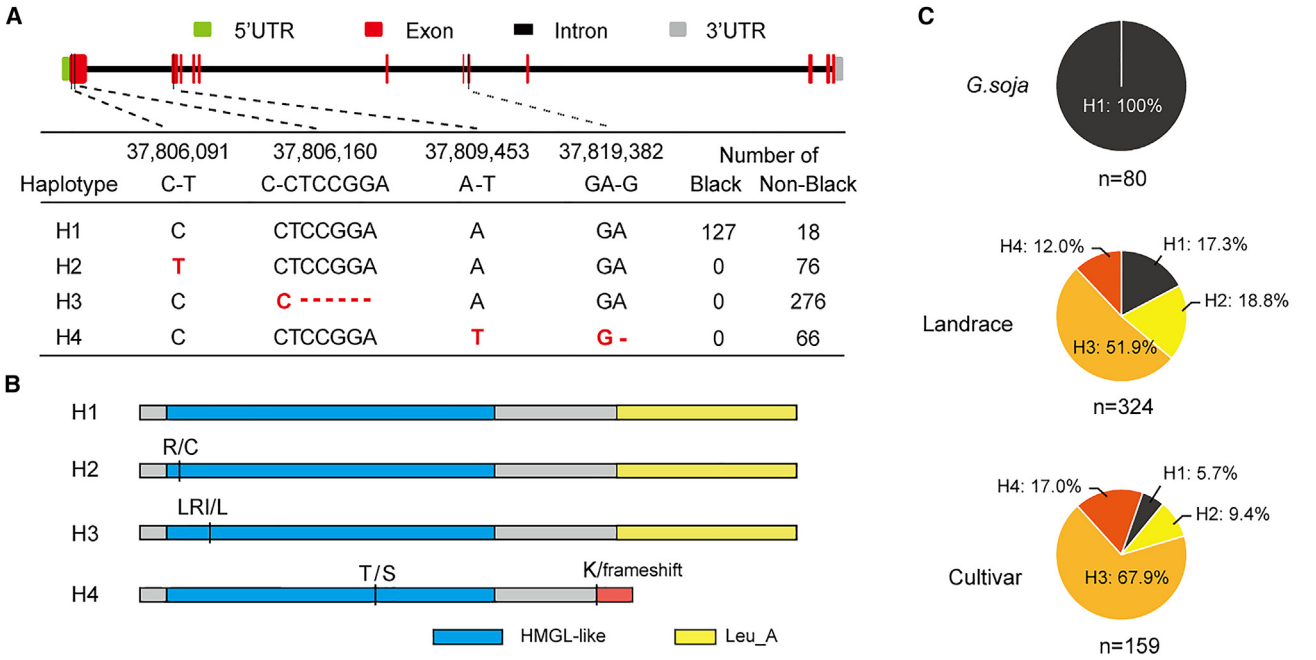

Figure 5. Haplotypes of the *L1* gene and their relation to soybean domestication.

(A) Structure and haplotypes of the *L1* gene. In the top panel, the linear gene structure displays the UTRs, CDS, and introns. The vertical solid lines represent the SNP loci across 563 soybean varieties. The dotted lines link the SNPs to their physical positions and nucleotides in the bottom panel. The functional SNPs are highlighted in red.

(B) The structures of proteins are encoded by four *L1* haplotypes. The amino acid abbreviations before and after slashes represent the reference (L1 in R171-BLk) and the alternative amino acids, respectively. Frameshift refers to a shift in the reading frame during protein translation.

(C) From top to bottom, the pie charts are for *G. soja*, landraces, and improved cultivars. The different colored portions in each pie chart represent the percentages of plants with different haplotypes.

for the oxidation and reduction of 4-HPP, respectively (Häusler et al., 1991; Wang et al., 2017). Piscidic acid is notably difficult to generate directly from α-keto acid and acetyl-CoA condensation. Given the high structural similarity between piscidic and eucomic acids (Heller and Tamm, 1974), we argue that piscidic acid might be derived from eucomic acid by an unknown mechanism (Figure 4D). Taken together, our data demonstrate that L1 is essential for the biosynthesis of piscidic and eucomic acids, both of which contribute to the black pigmentation of soybean pods.

Artificial selection of *I1* alleles during domestication

Next, we investigated natural variation of the *L1* gene in soybean by analyzing its CDS from 1407 sequenced accessions (Li et al., 2022). Despite synonymous mutations, we found seven polymorphisms, including five SNPs that led to amino acid substitutions (p.R31C, p.G40D, p.T196S, p.V276A, and p.N418D) and two InDels that resulted in either amino acid deletion or protein truncation (p.L54delinsLRI and p.K403fs). A total of four common variants, with a minor allele frequency (MAF) over 3%, were identified, which formed four haplotypes (H1 to H4) among 563 of 1407 accessions without missing data on these four variants (Figure 5A).

Haplotypes H1 and H2 respectively corresponded to the *L1* and *l1-1* alleles. Variants p.L54delinsLRI and p.T196S are present in haplotypes H3 and H4, respectively. Haplotype H4 also contains another InDel (p.K403fs), which results in CDS frameshifts (Figure 5A). These SNPs and InDels in haplotypes H2, H3, and H4 affect highly conserved residues on the N-terminal HMGL-like domain, which is essential for L1 function (Figure 5B). After analyzing the relationship between the haplotypes and phenotypes of 563 accessions, we observed that all 127 accessions featuring black pods possess the H1 haplotype. On the other hand, 89.7% (418 out of 466) of the accessions with non-black pods harbored the H2, H3, or H4 haplotypes (Figure 5A). Additionally, after including three rare variants that fell below 3% MAF, we discovered that all six accessions with a T-to-C transition at the 827th nucleotide in the fourth exon carried non-black pods (Supplemental Table 2). In conclusion, these five variants of *Glyma.19G120400* account for 97.9% of the phenotypic variation in 563 soybean accessions. These results demonstrate a strong association between *L1* alleles (H1 haplotype) and black pods and between *l1* alleles (five variants) and non-black pods.

Our analysis of 1407 accessions revealed that all wild *G. soja* had black pods, while approximately 14.7% of landraces and only 3.3% of the improved cultivars exhibited black pods (Supplemental Figure 5A). To further understand the evolutionary history of the *L1* gene, we analyzed the four haplotypes in 593 soybean accessions. Our results showed that all wild *G. soja* possessed the H1 haplotype. However, the proportion of H1 was reduced to 17.3% (56 of 324) in landraces and further decreased to 5.7% (9 of 159) in cultivars (Figure 5C). Furthermore, we investigated the selection on the *L1* gene at the sequence level by utilizing F_{ST} and nucleotide diversity (π). Our findings showed that sweep signals within the *L1* gene surpass the top 5%

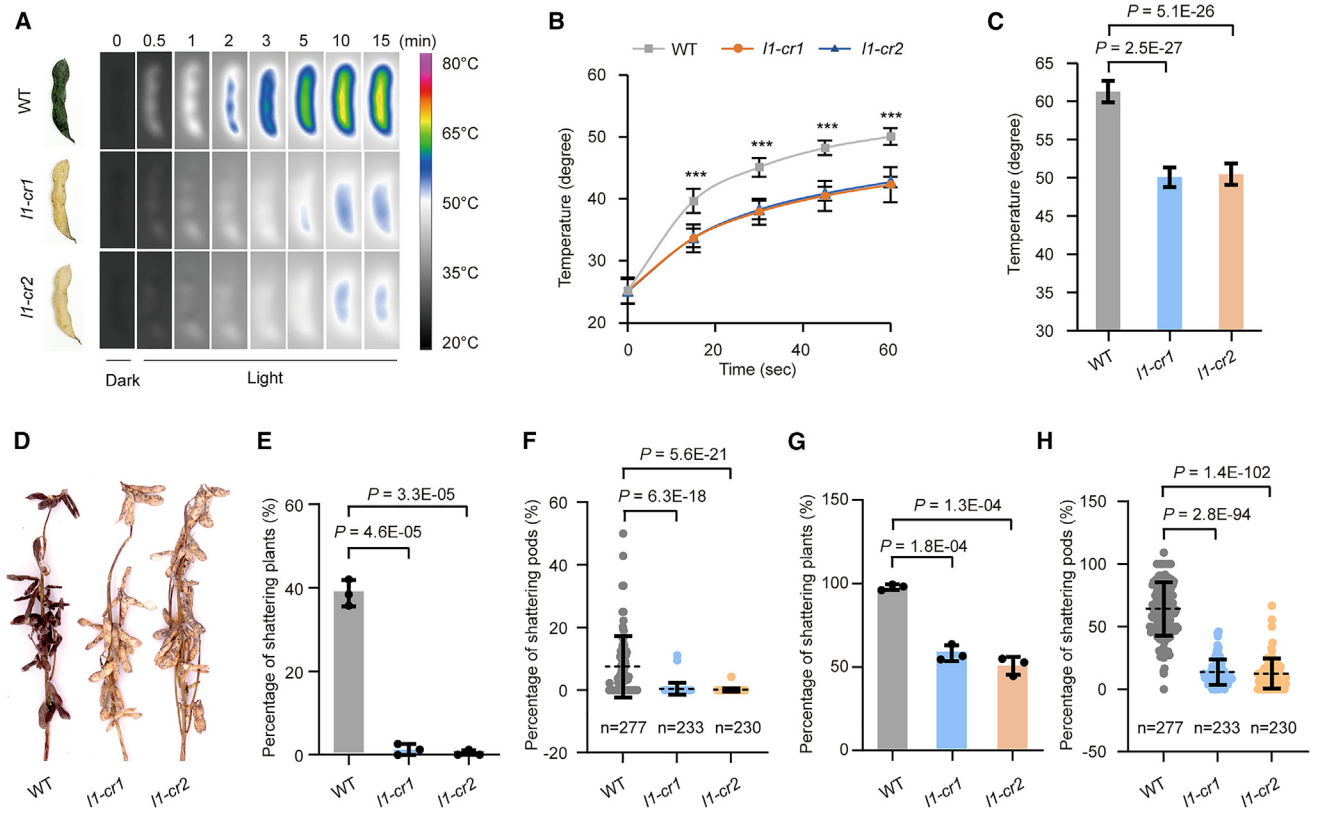

Figure 6. *L1*-mediated dark pod pigmentation intensifies photothermal-induced pod shattering.

(A) Infrared image of wild-type pods (R171-BLk) and two mutant pods (*l1-cr1* and *l1-cr2*) under simulated sunlight conditions.

(B) Temperature measurements of wild-type pod (R171-BLk) and two mutant pods (*l1-cr1* and *l1-cr2*) under simulated sunlight conditions for the first 60 s.

(C) Temperature of the wild-type pod (R171-BLk) and two mutant pods (*l1-cr1* and *l1-cr2*) at stabilized stages under simulated sunlight conditions.

(D) Representative image of a wild-type plant (R171-BLk) and two *L1* mutant lines (*l1-cr1* and *l1-cr2*) under field conditions after fully maturing for 3 weeks.

(E and F) Percentage of plants with shattered pods per plot and percentages of shattered pods per plant for wild type (R171-BLk) and *L1* mutants (*l1-cr1* and *l1-cr2*) at the full maturity stage. Data are presented as the mean ± SD, and *n* represents biological replicates.

(G and H) Percentage of plants with shattered pods per plot and percentages of shattered pods per plant for wild type (R171-BLk) and *L1* mutants (*l1-cr1* and *l1-cr2*) after 3 weeks at full maturity. Data are presented as the mean ± SD, and *n* represents biological replicates.

threshold observed throughout the whole genome sequence. These signals were consistently observed in both the comparison of *G. soja* versus landrace and landrace versus improved cultivar (Supplemental Figure 5B and 5C). These results indicate that *l1* alleles have undergone strong artificial selection during the domestication and genetic improvement of soybean.

L1 intensifies the impact of photothermal effects on pod shattering

Objects can absorb light and convert it to thermal energy through a process commonly known as the photothermal effect (Terazima et al., 2004). This effect is particularly pronounced for dark-colored objects. However, little research has explored the photothermal effect on soybean pods of different colors. To address this, we compared the temperatures of black pods (R171-BLk) with those of light-brown pods from *L1* gene-edited lines (*l1-cr1* and *l1-cr2*) under simulated sunlight (photosynthetically active radiation, approximately 1000 μmol $m^{-2} \cdot s^{-1}$). Our analysis of infrared images revealed that during the first 30 s of light exposure, black pods exhibited a significantly more substantial temperature change than light-brown pods (Figure 6A and 6B). After 1 min, the temperature of the black pods rose from 25°C to 50°C, while that of the light-brown pods only increased by 17°C (Figure 6B). Moreover, after 5 min, the temperature of the black soybean pods stabilized at 61°C, constituting an 11°C difference from the light-brown pods (Figure 6C). These results indicate that black-colored soybean pods undergo a rapid and intense temperature increase with short-term light exposure and that their temperature exceeds that of non-black pods when exposed to prolonged light.

Temperature is a critical factor influencing pod shattering in soybean plants (Tsuchiya, 1987). This observation raises an intriguing question regarding the contribution of the *L1* gene to pod shattering under light conditions. Thus, we compared the pod-shattering tendency of R171-BLk with that of two *L1* gene-edited lines under simulated sunlight conditions. Our results indicate that black pods of R171-BLk were more susceptible to shattering than those of the *l1-cr1* and *l1-cr2* lines (Supplemental Video 1). These results suggest an association between *L1*-mediated pod pigmentation and photothermal-induced pod shattering in soybeans.

To investigate the impact of *L1* on soybean pod shattering in agricultural practice, we conducted field experiments in Hainan

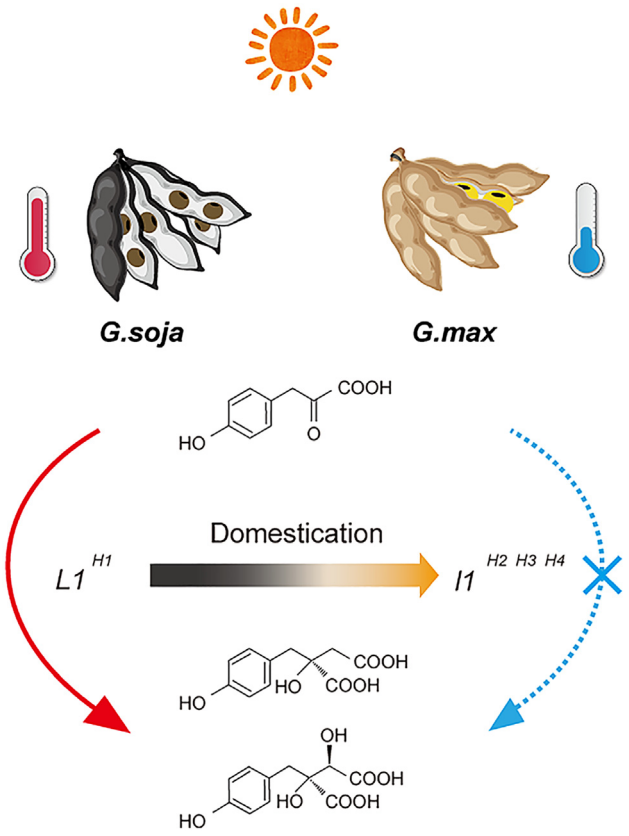

Figure 7. A proposed model illustrating how the domestication-associated *L1* gene modulates pod color and shattering in soybean.

L1 functions as a eucomic acid synthase, facilitating the synthesis of eucomic acid and piscidic acid, and leading to black pods in *G. soja*. The dark coloration of the pods enhances the photothermal conversion efficiency, promoting the shattering of pods and contributing to the survival of *G. soja*. The loss-of-function alleles $I1^{H2}$, $I1^{H3}$, and $I1^{H4}$ with light pod color and resistance to pod shattering were selected during soybean domestication and improvement.

Province, China. We measured the pod-shattering rate at full plant maturity (when all pods had reached their mature color) and 3 weeks later (Figure 6D). Field tests demonstrated that only a small proportion of mutant plants (*l1-cr1*, 1.31%; *l1-cr2*, 0.41%) produced shattering pods at the full maturity stage, compared to over 38.73% of the R171-BLk line (Figure 6E). Consistently, the percentages of shattered pods per plant for mutants (*l1-cr1*, 0.39%; *l1-cr2*, 0.05%) were significantly lower than for the R171-BLk line (6.44%) (Figure 6F).

Delayed harvesting can exacerbate pod shattering in soybean (Zhang and Bellaloui, 2012). After 3 weeks of maturity, approximately 50% of the mutant lines had at least one shattering pod, yet this rate was still much lower than that of the R171-BLk line (96.24%) (Figure 6G). The light-brown pods exhibited significant resistance to shattering (*l1-cr1*, 13.61%; *l1-cr2*, 11.52%), whereas the ratio of pod shattering in the black pods was as high as 62.98% (Figure 6H). These results demonstrate that precise genomic editing of *L1* can notably enhance pod-shattering resistance in the field by reducing pod pigmentation in soybean. Therefore, it is likely that the preference for pod-shattering resistance has contributed to the artificial selection of *I1* alleles in cultivated soybean varieties (Figure 7).

DISCUSSION

Plant coloration, or pigmentation, is an age-old subject in botanical sciences (Willmer et al., 2009; Zhu et al., 2010; Samanta et al., 2011; Aviezer and Lev-Yadun, 2015; Wang et al., 2018; Nelson and Whitehead, 2021; Roy et al., 2022). In 1865, Mendel's pioneering experiments on pea characteristics, specifically the colors of cotyledons, flowers, and pods, unveiled the foundational principles of genetics and marked the beginning of modern genetics. However, the mechanism behind pod coloration in legumes, including soybeans, remains elusive to this day (Shirasawa et al., 2021). Despite identifying that the *L1* locus governs black pod coloration in soybeans over a century ago, the gene accountable for this trait continues to be a subject of ongoing debate. He et al. (2015) proposed *Glyma19g27460* (equivalent to *Glyma.19G101700* in Wm82.a2.v1), which encodes an MYB DNA-binding domain protein, as the most likely candidate for *L1*. However, other researchers have located the *L1* locus in a region situated approximately 2.7 Mb downstream of *Glyma19g27460* (Bandillo et al., 2017; Chang et al., 2021). In this study, we employed a map-based cloning strategy and transgenic plants to identify and verify that *Glyma.19g120400* corresponds to the *L1* gene.

We discovered that the *L1* gene encodes a cytoplasmic protein with an HMGL-like domain, which is a member of the CC-like subgroup of the DRE-TIM metallolyase superfamily. This subgroup of enzymes is responsible for catalyzing condensation reactions between various α-keto acids and acetyl-CoA and is involved in the biosynthesis of branched-chain amino acids or secondary metabolites in plants (Kumar et al., 2016). We found that L1 functions as a eucomic acid synthase, catalyzing the condensation reaction between 4-HPP and acetyl-CoA to produce eucomic acid and possibly piscidic acid. Notably, L1 belongs to a distinct subgroup consisting only of members derived from Fabaceae species in the phylogenetic tree, which is consistent with the observation that eucomic and piscidic acids are abundant in Fabaceae species, including pigeonpea, faba bean, and *Lotus japonicus* (Ae et al., 1990; Okada et al., 2009; Valente et al., 2018). We further demonstrated that the presence of eucomic and piscidic acids is necessary for dark pigmentation in pods. These compounds are thought to be involved in the process of oxidation or polymerization of other compounds, but the precise mechanism remains unclear and requires further investigation (Owen, 1928).

Our research indicates that *L1* alleles underwent strong artificial selection during the process of soybean domestication and improvement. This is consistent with a previous study indicating that *L1* was in a selective sweep region (with a Tajima's *D* score of >2) (Zhou et al., 2016). However, the selection pressure against the black pod trait in soybean remains unclear. It is possible that the black pod trait potentially conflicts with human cultural preferences or is associated with other characteristics that negatively affect agricultural production. For instance, the green seed coat was eliminated during soybean domestication because it was positively linked with uneven emergence times of seedlings (Wang et al., 2018). Indeed, during the process of

crop domestication, pigmentation traits that lacked agronomic value were frequently subjected to negative selection, leading to their elimination from the cultivated varieties (Paauw et al., 2019).

Black pods are usually accompanied by smudged seeds due to the presence of "oxidation pigments" that are different from anthocyanins (Owen, 1928; Williams, 1952; Bernard, 1967). Interestingly, the quantitative trait loci responsible for the total antioxidant content of seeds precisely overlap with the *L1* and *L2* loci (Qi et al., 2014; Li et al., 2016). Our study provides evidence of the pleiotropic effects of the *L1* locus, which regulates pigmentation, involving oxidation or polymerization, in black pods and smudged appearance in the seed coat. Seed coloration typically acts as an adaptive feature that provides protective camouflage, thus helping seeds evade predators (Porter, 2013; Aviezer and Lev-Yadun, 2015). However, modern agricultural practices consider smudged seeds undesirable because they lead to reduced soybean seed quality and lower market prices (Sinclair, 1992). According to the United States Standards for Soybeans, the No. 1 yellow soybean must contain no more than 1% soybeans of other colors (SBOC), with seeds classified as "mixed" if SBOC exceeds 10%. The unsightly appearance of soybean seeds could therefore have been a contributing factor in the selection of *l1* alleles during soybean domestication and improvement, motivated by aesthetic or cultural reasons (Doebley et al., 2006). This, in turn, could indirectly influence the artificial selection of pod color in soybean.

Moreover, we showed that black pods exhibit a more prominent photothermal effect, resulting in a rapid and significant increase in temperature even with brief exposure to intense light. High temperatures and rapid thermal fluctuations can lead to alterations in the structure and moisture content of soybean pods, making the pods more susceptible to shattering (Tukamuhabwa et al., 2002). Indeed, Prine et al. (1964) have reported that soybean pods on the south side of east–west rows, which receive the most sunlight, experience the highest rate of pod shattering. Additionally, in tropical regions, where sunlight is particularly strong, soybeans are more vulnerable to pod shattering (Fasusi et al., 2022). These agricultural practices emphasize the significant impact of the photothermal effect on soybean pod shattering. Our studies demonstrated that *L1*-mediated dark pigmentation intensifies photothermal-induced pod shattering in soybean. The *L1* plants were more susceptible to pod shattering under light conditions than *l1* null mutants. This phenomenon highlights the importance of selecting light-colored pods during soybean domestication and agricultural practices to minimize yield losses caused by pod shattering (Figure 7). Therefore, our findings suggest that targeting *L1* could potentially provide a solution to this issue.

In summary, we have identified L1 as a crucial synthase of eucomic acid, thereby enhancing our knowledge of the enzymatic activity of the HMGL-like domain family. Our findings also shed light on the multifaceted regulatory role of *L1* in the pigmentation of soybean pods and seeds and its indirect facilitation of pod shattering. Consequently, *L1* can be considered a potential target for the *de novo* domestication of *G. soja* to strengthen resistance against pod shattering and enhance seed quality.

METHODS

Plant materials and growth conditions

R171-BLk and R171-LBn are two NILs with different pod colors segregated in an $F_{10:11}$ RIL, R171, by crossing Zhonghuang13 (ZH13) and Zhongpin03-5373 (ZP03). CRISPR–Cas9-engineered mutants were generated on the background of R171-BLk. Transgenic overexpression lines were generated on the background of the elite soybean cultivar Tianlong1 (TL1), which was provided by the Oil Crops Research Institute, Chinese Academy of Agricultural Sciences. Plants were cultivated at the experimental station in Beijing (40.1°N, 116.7°E) and Sanya City (18.3°N, 109.6°E).

BSR-seq

Leaves of 30 black-podded and 30 light-brown-podded plants from the F_2 population were sampled to isolate RNA. Total RNA was quantified with a bioanalyzer and subjected to RNA sequencing using an Illumina HiSeq2500 platform with a 125-bp paired-end sequencing strategy. The sequenced reads were first scanned for trimming of low-quality bases. Bases with PHRED quality values <15 (i.e., error rates < ~3%) were removed, and the remaining reads were filtered with a minimum length of 60 bp. The trimmed reads were then aligned to the Wm82 reference genome (a2.v1) with GSNAP (Wu and Nacu, 2010). The uniquely aligned reads were used for SNP calling while ignoring the first 3 and last 3 bp of each read and only used sites with a PHRED score ≥20. The biallelic SNPs having alleles different from those of the reference genome were identified for those two poolings using an in-house script (Liu et al., 2012). Raw SNPs were filtered with at least five reads in each pooled sample. The 45 684 retained SNPs were used for BSR-seq mapping, with an empirical Bayesian approach. SNPs with an association probability of >0.05 were defined as associated with the trait.

Map-based cloning of *L1* locus

We used a classical map-based cloning strategy to identify the *L1* locus. First, we crossed the two NILs, R171-BLk (*L1*) and R171-LBn (*L1*), in 2013. F_1 plants were grown at the Beijing Experimental Station in 2014, and F_2 populations were derived from self-fertilization. Subsequently, an $F_{2:4}$ population of 2662 individuals was used to screen recombinants with different breakpoints at the *L1* locus using nine InDel and SNP markers in 2015 at the Shunyi Experimental Station of the Institute of Crop Sciences of the Chinese Academy of Agricultural Sciences (40.1°N, 116.7°E).

Primers, accession numbers, and vector construction

Supplemental Table 3 lists the primers used in this study. The *L1* gene sequence was obtained from the Phytozome database (Wm82.a2.v1) with accession number *Glyma.19G120400*. The CDS of the dominant *L1* allele was amplified from the R171-BLk cDNA and cloned into the binary vector p0641-FLAG under the control of the *CaMV 35S* promoter. Subsequently, the *35S*::*L1-FLAG* construct was introduced into *Agrobacterium* strain EHA105 and transformed into cultivar TL1 using the *Agrobacterium*-cotyledonary node transformation system.

The CRISPR–Cas9 knockout construct for *L1* was designed to create a mutant in the first exon using two guide RNAs (gRNAs). The gRNAs were designed using the CRISPRdirect web tool (http://crispr.dbcls.jp/) (Naito et al., 2015). At least four different gRNAs were selected to construct the CRISPR–Cas9 expression vectors. According to the soybean hairy root system, each candidate gRNA was evaluated for editing efficiency, and two efficient gRNAs were selected. The soybean cotyledonary explants were then infected with the EHA105 strain containing CRISPR–Cas9 constructs. A pair of gene-specific primers (*L1-CR*, Supplemental Table 3) was designed to amplify DNA fragments comprising the target sites in the transgenic T1 plants and their progenies.

Sample collection, RNA extraction, and qRT–PCR

We collected various tissues of R171-BLk plants at 16:00 under field conditions, including roots, hypocotyls, cotyledons, epicotyls, unifoliolate

leaves, trifoliolate leaves, stems, flowers, and pods. Total RNA was extracted from the samples using TRIzol reagent. First-strand cDNA was synthesized with 1–3 μg total RNA in a 20-μl volume using an oligo(dT)$_{18}$ primer. Gene transcription levels were determined by qRT–PCR using the gene-specific primer on the qTOWER3G platform (Analytik Jena) according to the manufacturer's instructions. In brief, 2 μl of 10-fold diluted cDNA was used as a template and amplified using ChamQ Universal SYBR qPCR Master Mix (Vazyme) in a volume of 20 μl. Cycling conditions included 30 s of predenaturation at 95°C, followed by a 40-cycle program (95°C for 10 s, 60°C for 30 s, per cycle). The expression level of soybean actin (*Glyma.18G290800*) was measured as an endogenous control (Lyu et al., 2021). For each set of experiments, the transcription levels of genes were calculated using the $2^{-\Delta Ct}$ method, and at least three biological replicates were used.

Soybean transformation

The *Agrobacterium*-cotyledonary node transformation method was used to develop the CRISPR–Cas9-mediated *l1* mutants and the *L1* overexpression lines. Seeds of TL1 or R171-BLk were placed in a vacuum desiccator containing chlorine gas for 16–18 h and then placed in sterile water overnight. The seed coat was carefully removed, and two cotyledonary explants were prepared by making a horizontal incision through the hypocotyl and a vertical incision between the cotyledons. After a 30-min immersion in *Agrobacterium* EHA105 inoculum containing the constructs, the explants were transferred to a cocultivation plate for 3 days at 25°C in darkness. The explants were then cultured on shoot initiation medium at 25°C under photoperiodic conditions (12-h light/12-h dark) and subcultured on fresh medium every 2 weeks. After the shoot initiation phase, the elongated shoots were moved to the rooting medium.

Subcellular localization

To verify the subcellular localization of the L1 protein, the CDS of the *L1* gene was cloned into the pTF101 vector with green fluorescent protein (GFP) at the C terminus. We introduced the *35S::L1-GFP* construct into *Agrobacterium* strain EHA105 and infiltrated it into young *Nicotiana benthamiana* leaves. After incubation under light for 36 h, leaves were cut into strips and incubated in enzyme solutions (1% Cellulase R-10, 0.5% Macerozyme R-10, 0.5 M mannitol, 50 mM 2-(*N-morpholino)ethanesulfonic acid* [pH 5.8]) to isolate protoplasts. Protoplasts were then examined using a Zeiss LSM780 confocal laser scanning microscope and processed using ZEN 2009 Light Edition software.

Protein domain and phylogenetic analysis

The homologs of L1 were found in Phytozome using BLAST. The amino acid sequences of homologs were downloaded and then aligned with COBALT (NCBI) tools. A phylogenetic tree was constructed using the neighbor-joining method with 1000 bootstrap replicates. Conserved domains were obtained from the HMMER web server (https://www.ebi.ac.uk/Tools/hmmer/).

Metabolite extraction and analysis

Pods of two parental lines, R171-BLk and R171-LBn, as well as *l1-cr1*, TL1, and *L1-OE1*, were harvested and ground to a fine powder. Subsequently, each powdered sample was subjected to extraction with solvent (water/methanol/acetonitrile = 1:2:2) in a rotator at 4°C overnight. After centrifugation at 15 000 *g* for 20 min, the supernatant was filtered using ultrafiltration with a 3-kDa molecular weight cutoff and subjected to UPLC-QTOF-MS analysis (Thermo Vanquish F-AB SCIEX 6600) equipped with an Acquity UPLC HSS T3 column (1.8 μm, 2.1 mm internal diameter [ID] × 100 mm, Waters) eluting at a flow rate of 0.4 ml/min. Mobile phase A was water/formic acid (1000:1, v/v), and mobile phase B was acetonitrile/formic acid (1000:1, v/v). The UPLC separations were 25 min/sample using the following scheme: (1) 0–1 min, 5% B; (2) 5 min, 25% B; (3) 15–18 min, 100% B; and (4) 18.1–25 min, 5% B. All the changes were linear, and the flow rate was set at 0.4 ml/min. Targeted metabolomics analysis of LC–MS data was performed using SCIEX OS 2.0.1 software.

Purification of the expressed His-tag protein

CDSs of the genes (*L1* and *l1-1*) were amplified from the R171-BLk and R171-LBn cDNA and cloned into the pCold-TF vector with a His tag. The constructs were then induced in *E. coli* BL21 (DE3) cells. A single colony was inoculated into 10 ml of Luria-Bertani (LB) liquid medium and grown overnight at 37°C to produce a seed culture. An aliquot of 3 ml of the seed culture was transferred to 300 ml of LB liquid medium and shaken at 37°C for approximately 4 h until the OD_{600} reached 0.6. The culture was then cooled and induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside at 16°C for 24 h. Cells were harvested by centrifugation at 8000 *g* for 5 min at 4°C and then resuspended in lysis buffer (50 mM Tris–HCl [pH 8.0], 500 mM NaCl, 0.25 mM dithiothreitol, 1 mM phenylmethylsulfonyl fluoride, and 0.2% lysozyme) on ice for 30 min. Bacterial cells were lysed using a sonicator and centrifuged at 12 000 *g* at 4°C for 20 min. The supernatant was collected using a sterile 0.25-μm filter. A volume of 1 ml of 50% Ni-NTA agarose (Qiagen), rinsed three times with lysis buffer, was added to the supernatant and placed in a rotator for 8 h at 4°C to allow the His-tag protein to bind to the resin. The lysate–Ni-NTA mixture was collected by centrifugation at 700 *g* for 2 min at 4°C and washed more than six times with wash buffer (50 mM Tris–HCl [pH 8.0], 500 mM NaCl, 20 mM imidazole). His-tag protein was then eluted with 3 ml of elution buffer (50 mM Tris–HCl [pH 8.0], 500 mM NaCl, 500 mM imidazole). Subsequently, the eluent was filtered through an Amicon Ultra 0.5-ml centrifugal filter with 10 kDa molecular weight and desalted in 4 ml of a 50 mM Tris buffer (pH 8.0) containing 1 mM $MgCl_2$ and 10% glycerol.

Enzyme assay

L1 and l1^{R31C} enzyme assays were performed according to the methods of De Kraker et al. (2007) with minor modifications. In brief, a 150-μl reaction mixture contained 1 nm purified recombinant enzyme, 500 μM acetyl-CoA, 1 mM 4-HPP, 4 mM $MgCl_2$, and 100 mM Tris–HCl (pH 8.0). Incubation was performed at 37°C for 10 min and terminated by the addition of 50 μl of ethanol. After centrifugation at 15 000 *g* for 20 min, 5 μl of the supernatant was subjected to triple-quadrupole LC–MS (Agilent 1290-6460) analysis equipped with an Acquity UPLC HSS T3 column (1.8 μm, 2.1 mm ID × 100 mm, Waters) eluting at a flow rate of 0.4 ml/min. Mobile phase A was water/formic acid (1000:1, v/v), and mobile phase B was acetonitrile/formic acid (1000:1, v/v). The UPLC separations were 12 min/sample using the following scheme: (1) 0–1 min, 5% B; (2) 5 min, 25% B; (3) 7–9 min, 100% B; and (4) 9.1–12 min, 5% B. All the changes were linear, and the flow rate was set at 0.4 ml/min. For the determination of K_m and V_{max}, different concentrations of 4-HPP ranging from 0.01 μM to 1 mM were used in the presence of 500 μM acetyl-CoA, 4 mM $MgCl_2$, 100 mM Tris–HCl (pH 8.0), and equal amounts of purified recombinant proteins. K_m and V_{max} were determined by nonlinear fitting to the Michaelis–Menten equation using GraphPad Prism v8.0.0. Triplicate assays were performed for all data points.

Gene diversity analysis

To identify potential selective signals of *Glyma.19G120400* during soybean domestication and improvement breeding, we used a sliding-window approach (10-kb windows sliding in 5-kb steps) to quantify the levels of nucleotide diversity (π) and genetic differentiation (F_{ST}) between *G. soja*, landrace, and improved cultivars using VCFtools (v0.1.14) (Danecek et al., 2011). The windows with the top 5% of F_{ST}, π ratio (π_{pop1}/π_{pop2}) were considered as candidate sweeps.

DATA AVAILABILITY

All data supporting the findings of this study are available in the paper, the extended data, and the supplementary information. The sequencing data used in this study have been deposited in the Genome Sequence Archive

(GSA) database in the BIG Data Center (https://bigd.big.ac.cn/gsa/index.jsp) under accession number GSA: PRJCA010353.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the National Key Research and Development Program of China (2021YFD1201601, 2021YFF1001201), the earmarked fund for CARS (CARS-04-PS01), the Agricultural Science and Technology Innovation Program (ASTIP) of the Chinese Academy of Agricultural Sciences, the National Natural Science Foundation of China (grant no. 32201734), and the China Postdoctoral Science Foundation (grant no. 2021M693465).

AUTHOR CONTRIBUTIONS

L.-j.Q. and B.L. conceived the project. Y.-h.L. and B.L. designed and planned the experiments. X.L., Y.-h.L., Y.L., D.L., C.H., H.H., Y.T., and L.H. performed the experiments and collected and analyzed the data. L.-j.Q., B.L., Y.-h.L., and X.L. wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank Prof. Patrick S. Schnable (Iowa State University, Ames) for fruitful discussions and suggestions on the BSR-seq. The authors declare no competing interests.

Received: March 21, 2023
Revised: May 23, 2023
Accepted: June 13, 2023
Published: June 19, 2023

REFERENCES

Ae, N., Arihara, J., Okada, K., Yoshihara, T., and Johansen, C. (1990). Phosphorus uptake by pigeon pea and its role in cropping systems of the Indian subcontinent. Science **248**:477–480.

Aviezer, I., and Lev-Yadun, S. (2015). Pod and seed defensive coloration (camouflage and mimicry) in the genus Pisum. Isr. J. Plant Sci. **62**:39–51.

Bandillo, N.B., Lorenz, A.J., Graef, G.L., Jarquin, D., Hyten, D.L., Nelson, R.L., and Specht, J.E. (2017). Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. Plant Genome **10**.

Bara, N., Khare, D., and Shrivastava, A.N. (2013). Studies on the factors affecting pod shattering in soybean. Indian J. Genet. Plant Breed. **73**:270–277.

Bernard, R.L. (1967). The inheritance of pod color in soybeans. J. Hered. **58**:165–168.

Bridge, W., Coleman, F., and Robertson, A. (1948). Constituents of "Cortex Piscidiæ erythrinæ." Part I. The structure of piscidic acid. J. Chem. Soc. **174**:257–260.

Chang, F., Lv, W., Lv, P., Xiao, Y., Yan, W., Chen, S., Zheng, L., Xie, P., Wang, L., Karikari, B., et al. (2021). Exploring genetic architecture for pod-related traits in soybean using image-based phenotyping. Mol. Breed. **41**:28.

Chen, W., and Frantom, P.A. (2019). Distinct mechanisms of substrate selectivity in the DRE-TIM metallolyase superfamily: A role for the LeuA dimer regulatory domain. Arch. Biochem. Biophys. **664**:1–8.

Cieslak, M., Reissmann, M., Hofreiter, M., and Ludwig, A. (2011). Colours of domestication. Biol. Rev. **86**:885–899.

Conte, J.V., and Frantom, P.A. (2020). Biochemical characterization of 2-phosphinomethylmalate synthase from Streptomyces hygroscopicus: A member of the DRE-TIM metallolyase superfamily. Arch. Biochem. Biophys. **691**, 108489.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158.

De Kraker, J.-W., Luck, K., Textor, S., Tokuhisa, J.G., and Gershenzon, J. (2007). Two *Arabidopsis* genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. Plant Physiol. **143**:970–986.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127**:1309–1321.

Dong, Y., Yang, X., Liu, J., Wang, B.H., Liu, B.L., and Wang, Y.Z. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. Nat. Commun. **5**:3352.

Fasusi, S.A., Kim, J.-M., and Kang, S. (2022). Current status of soybean production in nigeria: constraint and prospect. Korean J. Int. Agric. **34**:149–156.

Felnagle, E.A., Chaubey, A., Noey, E.L., Houk, K.N., and Liao, J.C. (2012). Engineering synthetic recursive pathways to generate non-natural small molecules. Nat. Chem. Biol. **8**:518–526.

Field, B., Cardon, G., Traka, M., Botterman, J., Vancanneyt, G., and Mithen, R. (2004). Glucosinolate and amino acid biosynthesis in *Arabidopsis*. Plant Physiol. **135**:828–839.

Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., Komatsu, K., Katayama, T., Sayama, T., Ishimoto, M., and Fujino, K. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. Proc. Natl. Acad. Sci. USA **111**:17797–17802.

Gao, R., Han, T., Xun, H., Zeng, X., Li, P., Li, Y., Wang, Y., Shao, Y., Cheng, X., Feng, X., et al. (2021). MYB transcription factors GmMYBA2 and GmMYBR function in a feedback loop to control pigmentation of seed coat in soybean. J. Exp. Bot. **72**:4401–4418.

Gillman, J.D., Tetlow, A., Lee, J.-D., Shannon, J.G., and Bilyeu, K. (2011). Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. BMC Plant Biol. **11**:155.

Häusler, E., Petersen, M., and Alfermann, A.W. (1991). Hydroxyphenylpyruvate reductase from cell suspension cultures of coleus Blumei Benth. Z. Naturforsch., C: J. Biosci. **46**:371–376.

He, Q., Yang, H., Xiang, S., Tian, D., Wang, W., Zhao, T., and Gai, J. (2015). Fine mapping of the genetic locus L1 conferring black pods using a chromosome segment substitution line population of soybean. Plant Breed. **134**:437–445.

He, Y., Cheng, J., He, Y., Yang, B., Cheng, Y., Yang, C., Zhang, H., and Wang, Z. (2019). Influence of isopropylmalate synthase OsIPMS1 on seed vigour associated with amino acid and energy metabolism in rice. Plant Biotechnol. J. **17**:322–337.

Heller, W., and Tamm, C. (1974). Isolierung, konstitution und synthese der (r)-(−)-eucominsäure. Helv. Chim. Acta **57**:1766–1784.

Insaward, A., Duangmal, K., and Mahawanich, T. (2015). Mechanical, Optical, and Barrier Properties of Soy Protein Film As Affected by Phenolic Acid Addition. J. Agric. Food Chem. **63**:9421–9426.

Knill, T., Schuster, J., Reichelt, M., Gershenzon, J., and Binder, S. (2008). *Arabidopsis* branched-chain aminotransferase 3 functions in both amino acid and glucosinolate biosynthesis. Plant Physiol. **146**:1028–1039.

Krisnawati, A., and Adie, M.M. (2019). Genetic variability of soybean (Glycine max L. Merrill) genotypes for pod shattering resistance. IOP Conf. Ser. Earth Environ. Sci. **293**:012003.

Kumar, G., Johnson, J.L., and Frantom, P.A. (2016). Improving Functional Annotation in the DRE-TIM Metallolyase Superfamily through Identification of Active Site Fingerprints. Biochemistry **55**:1863–1872.

Li, M.-W., Muñoz, N.B., Wong, C.-F., Wong, F.-L., Wong, K.-S., Wong, J.W.-H., Qi, X., Li, K.-P., Ng, M.-S., and Lam, H.-M. (2016). QTLs Regulating the Contents of Antioxidants, Phenolics, and Flavonoids in Soybean Seeds Share a Common Genomic Region. Front. Plant Sci. **7**:854.

Li, Y.H., Qin, C., Wang, L., Jiao, C., Hong, H., Tian, Y., Li, Y., Xing, G., Wang, J., Gu, Y., et al. (2022). Genome-wide signatures of the geographic expansion and breeding of soybean. Sci. China Life Sci. **66**:350–365.

Liu, J., Zhang, Y., Jiang, Y., Sun, H., Duan, R., Qu, J., Yao, D., Liu, S., and Guan, S. (2022). Formation mechanism and occurrence law of pod shattering in soybean: A Review. Phyton **91**:1327–1340.

Liu, S., Yeh, C.-T., Tang, H.M., Nettleton, D., and Schnable, P.S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One **7**, e36406.

Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., Chen, L., Su, T., Nan, H., Zhang, D., et al. (2020). Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. Nat. Genet. **52**:428–436.

Lyu, X., Cheng, Q., Qin, C., Li, Y., Xu, X., Ji, R., Mu, R., Li, H., Zhao, T., Liu, J., et al. (2021). GmCRY1s modulate gibberellin metabolism to regulate soybean shade avoidance in response to reduced blue light. Mol. Plant **14**:298–314.

Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. Bioinformatics **31**:1120–1123.

Nelson, A.S., and Whitehead, S.R. (2021). Fruit secondary metabolites shape seed dispersal effectiveness. Trends Ecol. Evol. **36**:1113–1123.

Okada, M., Park, S., Koshizawa, T., and Ueda, M. (2009). (R)-Eucomic acid, a leaf-opening factor of the model organism, Lotus japonicus. Tetrahedron **65**:2136–2141.

Owen, F. (1927). Hereditary and environmental factors that produce mottling in soy beans. J. Agric. Res. **34**:559–587.

Owen, F.V. (1928). Inheritance studies in soybeans. III. Seed-coat color and summary of all other mendelian characters thus far reported. Genetics **13**:50–79.

Paauw, M., Koes, R., and Quattrocchio, F.M. (2019). Alteration of flavonoid pigmentation patterns during domestication of food crops. J. Exp. Bot. **70**:3719–3735.

Parker, T.A., Lo, S., and Gepts, P. (2021). Pod shattering in grain legumes: emerging genetic and environment-related patterns. Plant Cell **33**:179–199.

Porter, S.S. (2013). Adaptive divergence in seed color camouflage in contrasting soil environments. New Phytol. **197**:1311–1320.

Prine, G.M., West, S.H., and Hinson, K. (1964). Shattering, moisture content and seed temperature of soybeans as influenced by row direction. Agron. J. **56**:594–595.

Prodpran, T., Benjakul, S., and Phatcharat, S. (2012). Effect of phenolic compounds on protein cross-linking and properties of film from fish myofibrillar protein. Int. J. Biol. Macromol. **51**:774–782.

Qi, X., Li, M.-W., Xie, M., Liu, X., Ni, M., Shao, G., Song, C., Kay-Yuen Yim, A., Tao, Y., Wong, F.-L., et al. (2014). Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing. Nat. Commun. **5**:4340.

Rohn, S., Petzke, K.J., Rawel, H.M., and Kroll, J. (2006). Reactions of chlorogenic acid and quercetin with a soy protein isolate – Influence on the in vivo food protein quality in rats. Mol. Nutr. Food Res. **50**:696–704.

Roy, R., Moreno, N., Brockman, S.A., Kostanecki, A., Zambre, A., Holl, C., Solhaug, E.M., Minami, A., Snell-Rood, E.C., Hampton, M., et al. (2022). Convergent evolution of a blood-red nectar pigment in vertebrate-pollinated flowers. Proc. Natl. Acad. Sci. USA **119**, e2114420119.

Sá, A.G.A., Moreno, Y.M.F., and Carciofi, B.A.M. (2020). Plant proteins as high-quality nutritional source for human diet. Trends Food Sci. Technol. **97**:170–184.

Samanta, A., Das, G., and Das, S.K. (2011). Roles of flavonoids in plants. Carbon **100**:12–35.

Sedivy, E.J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. New Phytol. **214**:539–553.

Shirasawa, K., Sasaki, K., Hirakawa, H., and Isobe, S. (2021). Genomic region associated with pod color variation in pea (Pisum sativum). G3 **11**:jkab081.

Shoemaker, R.C., and Specht, J.E. (1995). Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. **35**:436–446.

Sinclair, J. (1992). Discoloration of soybean seeds–an indicator of quality. Plant Dis. **76**:1087.

Terazima, M., Hirota, N., Braslavsky, S.E., Mandelis, A., Bialkowski, S.E., Diebold, G.J., Miller, R.J.D., Fournier, D., Palmer, R.A., and Tam, A. (2004). Quantities, terminology, and symbols in photothermal and related spectroscopies (IUPAC Recommendations 2004). Pure Appl. Chem. **76**:1083–1118.

Textor, S., De Kraker, J.-W., Hause, B., Gershenzon, J., and Tokuhisa, J.G. (2007). MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in *Arabidopsis*. Plant Physiol. **144**:60–71.

Tiwari, S., and Bhatia, V. (1995). Characters of pod anatomy associated with resistance to pod-shattering in soybean. Ann. Bot. **76**:483–485.

Tsuchiya, T. (1987). Physiological and genetic analysis of pod shattering in soybean. JARQ **21**:166–175.

Tukamuhabwa, P., Dashiell, K.E., Rubaihayo, P., and Nabasirye, M. (2002). Determination of field yield loss and effect of environment on pod shattering in soybean. Afri. Crop Sci. J. **10**:203–209.

Valente, I.M., Maia, M.R.G., Malushi, N., Oliveira, H.M., Papa, L., Rodrigues, J.A., Fonseca, A.J.M., and Cabrita, A.R.J. (2018). Profiling of phenolic compounds and antioxidant properties of European varieties and cultivars of Vicia faba L. pods. Phytochemistry **152**:223–229.

Wang, G.Q., Chen, J.F., Yi, B., Tan, H.X., Zhang, L., and Chen, W.S. (2017). HPPR encodes the hydroxyphenylpyruvate reductase required for the biosynthesis of hydrophilic phenolic acids in Salvia miltiorrhiza. Chin. J. Nat. Med. **15**:917–927.

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., Yang, R., Zhang, M., Liu, S., Lu, S., et al. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. Nat. Genet. **50**:1435–1441.

Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y.C., Liu, Z., Frommer, W.B., Ma, J.F., Chen, L.Q., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. Natl. Sci. Rev. **7**:1776–1786.

Williams, L.F. (1952). The Inheritance of Certain Black and Brown Pigments in the Soybean. Genetics **37**:208–215.

Willmer, P., Stanley, D.A., Steijven, K., Matthews, I.M., and Nuttman, C.V. (2009). Bidirectional Flower Color and Shape Changes Allow a Second Opportunity for Pollination. Curr. Biol. **19**:919–923.

Woodworth, C.M. (1923). Inheritance of growth habit, pod color, and flower color in soybeans. Agron. J. **15**:481–495.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics **26**:873–881.

Zhang, L., and Bellaloui, N. (2012). Effects of planting and maturity dates on shattering patterns under early soybean production system. Am. J. Plant Sci. **03**:119–124.

Zhou, L., Luo, L., Zuo, J.F., Yang, L., Zhang, L., Guang, X., Niu, Y., Jian, J., Geng, Q.C., Liang, L., et al. (2016). Identification and validation of candidate genes associated with domesticated and improved traits in soybean. Plant Genome **9**. https://doi.org/10.3835/plantgenome2015.09.0090.

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., Yu, Y., Shu, L., Zhao, Y., Ma, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. **33**:408–414.

Zhu, C., Bai, C., Sanahuja, G., Yuan, D., Farré, G., Naqvi, S., Shi, L., Capell, T., and Christou, P. (2010). The regulation of carotenoid pigmentation in flowers. Arch. Biochem. Biophys. **504**:132–141.