

Article

GmSWEET46 Regulates Seed Oil and Protein Content in Soybean

Dezhi Han ^{1,†}, Huiyi Su ^{2,†}, Qiuzhen Lai ³, Wei Li ¹, Wencheng Lu ^{1,*} and Tianxiao Lv ^{3,*}

¹ Heihe Branch of Heilongjiang Academy of Agricultural Sciences, No. 345 Huanchengxi Road, Heihe 152052, China; handezhi2008@163.com (D.H.); liweizjyzc@163.com (W.L.)

² College of Agriculture, South China Agricultural University, No. 483 Wushan Road, Guangzhou 510642, China; hknmat@163.com

³ Guangdong Provincial Key Laboratory of Plant Adaptation and Molecular Design, Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou University, Guangzhou Higher Education Mega Center, Guangzhou 510006, China; 2112314047@e.gzhu.edu.cn

* Correspondence: 13845674288@163.com (W.L.); ljl_vv@gzhu.edu.cn (T.L.)

† These authors contributed equally to this work.

Abstract

Seed oil and protein contents are critical agronomic traits that determine soybean quality. However, the key loci and corresponding genes controlling these quality traits remain to be elucidated. Here, we performed bulked segregant analysis by sequencing (BSA-seq) using an F4 population derived from a cross between the cultivars HeiNong 35 (HN35) and Dengke 3 (DK3). A major soybean oil and protein quantitative trait locus (QTL) designated as *q-OP18* was identified on chromosome 18, and the sugar transporter gene *GmSWEET46* was further cloned. Haplotype analysis revealed that a single-nucleotide polymorphism (SNP) in the sixth exon of *GmSWEET46* results in an amino acid change between HN35 and DK3 and is associated with seed oil and protein content, suggesting its important role in determining seed quality in soybean. *GmSWEET46* is expressed during the early stages of seed and pod development and localizes to the plasma membrane, indicating its potential function as a sugar transporter. Further studies demonstrated that *GmSWEET46* can regulate seed protein content, oil content, and seed size in Arabidopsis and soybean. Collectively, this study provides a novel locus and gene for regulating soybean seed traits and offers valuable resources for the breeding of high-quality and high-yielding soybean cultivars.

Keywords: soybean; BSA-seq; seed quality; SWEET



Academic Editor: Steven R. Larson

Received: 9 August 2025

Revised: 4 September 2025

Accepted: 9 September 2025

Published: 16 September 2025

Citation: Han, D.; Su, H.; Lai, Q.; Li, W.; Lu, W.; Lv, T. GmSWEET46 Regulates Seed Oil and Protein Content in Soybean. *Agronomy* **2025**, *15*, 2198. <https://doi.org/10.3390/agronomy15092198>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Soybean (*Glycine max* L. Merrill) is widely cultivated in many countries and serves as an important source of edible oil and protein for both humans and domestic animals [1]. The recognition of soybean's value as a significant source of oil and protein dates back to the early 20th century. Since 1987, soybean yield has nearly doubled, making it one of the most sustainable crops to meet the rapidly increasing global demand for plant-based oil and protein [2]. However, it is noted that both seed protein content and oil content are quantitatively inherited and are negatively correlated with each other in soybean [3,4]. This inverse relationship presents a significant challenge for increasing soybean protein content while maintaining the desired levels of seed oil and yield [5].

Sucrose is produced in photosynthetically active leaves (sources) and then transported to nonphotosynthetic tissues (sinks) with sucrose carriers and efflux transporters, known as SWEETs (Sugars Will Eventually be Exported Transporters) [6]. Since sucrose acts as the main source of carbon energy delivered via the phloem to developing seeds [7], both sucrose and SWEETs play pivotal roles in seed development in many species. For example, mutations in *AtSWEET11/12/15* in *Arabidopsis* impair sucrose delivery to the embryo, causing severe seed defects [8]. In rice, knockout of *OsSWEET11/15* leads to a complete loss of endosperm development [9]. In soybean, simultaneous knockout of *GmSWEET15a/15b* results in a high rate of seed abortion [10]. Recent studies have demonstrated that *GmSWEET10a* and *GmSWEET10b*, which have undergone stepwise selection during soybean domestication, play a crucial role in regulating seed size and quality by facilitating sugar allocation from the seed coat to the embryo [11]. To date, only *GmSWEET10a/b* have been implicated in controlling the seed oil and protein content in soybean [4,11,12]. Despite the identification of 52 putative *GmSWEET* genes in soybean [13], the specific functions of these genes in seed traits, including *GmSWEET46*, remain largely uncharacterized.

Although a large number of quantitative trait loci (QTLs) that regulate seed oil and protein content have been identified across 20 chromosomes in soybean [14], the underlying genes for these QTLs have rarely been isolated due to the complex structure of the soybean genome [15]. Here, we identified a seed oil and protein content associated locus, termed *q-OP18*, on chromosome 18 using an F4 population derived from a cross between the cultivars HeiNong 35 (HN35) and Dengke 3 (DK3) via bulked segregant analysis by sequencing (BSA-seq). Among the genes in the *q-OP18* locus, the sugar transporter gene *GmSWEET46* was considered as the candidate for regulating seed quality. Further analysis revealed that *GmSWEET46* can control seed protein content, oil content, and seed size in both *Arabidopsis* and soybean. Our findings provide insights into the role of *SWEET* genes in seed traits and offer a valuable genetic resource for improving quality in soybean.

2. Materials and Methods

2.1. Plant Materials and Phenotypic Identification

A population was derived by single-seed descent from F2 offspring of a cross between cultivar Hei Nong 35 (high protein content) and cultivar Deng ke 3 (high oil content). The F2:4 generations with 238 families along with parents were grown in Hailun, China (126°14' E, 46°58' N), during the spring of 2022 and 2023. The wild-type Williams 82 (Wm82) and the EMS-induced mutant line M499 were grown in Yazhou, Hainan, China (109.17° E, 18.37° N), during the autumn of 2024. The M499 mutant was obtained from the Wm82 EMS mutant library developed by Nanjing Agricultural University. Both the population and the EMS-induced mutant lines were arranged in a randomized complete block design. Each plot consisted of a single row measuring 1 m in length, with row spacing of 0.5 m and plant spacing of 0.1 m.

Protein and crude fat contents were determined using a near-infrared spectroscopy system NIRS DS2500 (FOSS, Denmark). Ten plants were randomly selected from each of the parental and F2:4 populations, NILs, Wm82, and EMS-induced mutant lines for the measurement of protein and oil content. Each plant was tested in three technical replicates to ensure data reliability. Statistical analysis was performed using SPSS 14 software, with key parameters including mean, maximum, minimum, skewness, kurtosis, and coefficient of variation (CV) calculated for all tested lines.

2.2. DNA Bulk Samples and BSA Sequencing

Based on the protein and oil content of F2 population individuals, 25 plants with the highest oil contents and 25 plants with the lowest oil content were selected to construct the

high-oil pool (HO) and low-oil pool (LO) groups, respectively. For each parent, 10 plants were randomly selected to construct the parent pool. Subsequent DNA extraction, library preparation, and sequencing were completed by Biomarker Technologies Corporation in Beijing. Genomic DNA was extracted from young tissues using the CTAB method, and its DNA concentration and purity were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). A sequencing library with an insert size of 350 bp was constructed and sequenced on an Illumina/BGI platform to generate 150 bp paired-end reads. Raw reads were processed through the BMK-Cloud bioinformatics platform www.biocloud.net (accessed on 12 January 2024) to obtain high-quality clean data, which were subsequently aligned to the reference genome (Glycine max Wm82.a4.v1) using BWA-MEM2. Variant calling was performed with GATK, and high-quality SNPs and INDELs were obtained after filtering. Allele frequency differences between pools were analyzed by calculating Euclidean distance (ED) and Δ (SNP-index). Association thresholds were defined using the DISTANCE/SNPNUM and LOESS methods, and genomic regions exceeding the threshold were identified as candidate intervals.

2.3. SNP-Index Analysis

Homozygous SNPs between parental lines and high-quality SNPs (with a minimum read depth of 10 and an SNP base quality ≥ 100 in the bulks) were selected for SNP-index analysis. The SNP-index was calculated at each SNP position in both the high-oil (HO) and low-oil (LO) bulks using the alternative allele from DK as the reference. The SNP-index was assigned a value of 0 or 1 when all short reads contained genomic fragments derived from HN35 or DK3, respectively. The Δ (SNP-index) was derived by subtracting the SNP-index of the LO bulk from that of the HO bulk. A high Δ (SNP-index) value at a given locus indicates that the allele was highly frequent in the HO bulk and scarce in the LO bulk. Additionally, a Fisher’s exact test was performed at each SNP locus to assess the significance of allele frequency differences between the two bulks.

2.4. DNA, RNA Extraction, and RT-qPCR

Soybean genomic DNA was extracted using a NuClean Plant Genomic DNA Kit (CW0531M, CwBio, Beijing, China). RNA was extracted from three replicates of different tissue including roots, stems, leaves, flowers, cotyledons, and the developing seeds (10 DAP, 25 DAP, 35 DAP, and 45 DAP) using a Plant RNA Extraction Kit (LS1040, Promega Shanghai, China). All the tissues and seeds were grown in a growth chamber under controlled conditions. cDNA was synthesized with a Promega reverse transcription kit (A3500, Promega). RT-qPCR was performed with a KAPA SYBR FAST qPCR Kit (KK4601, Sigma-Aldrich, St. Louis, MO, USA) using a Roche LightCycler 480. GmActin was used as an internal control. Primers used are listed in Table 1.

Table 1. Primers for fragment amplification, plasmid construction, and qRT-PCR.

	Forward Primer (5'-3')	Reverse Primer (5'-3')
Primers for gene cloning and constructs		
SWEET46	ATGGTTATCAGTCACCATAC	GACTAGGCAGTCTTGGCCCT
SWEET46-GFP	AAAAAGCTTGATATCGAATTCATGGT TATCAGTCACCATAC	AGCGAATTATCTAGAACTAGTGACTAGG CAGTCTTGGCCCT
SWEET46-flag	GGTCCCTACGTAGTCACGTGATGGTT ATCAGTCACCATAC	GAACCTCCGGACGTCACGTGGACTAGG CAGTCTTGGCCCT
Primers for qPCR		
SWEET46-qRT Actin	CCCTTCGTGTCCAAGTCCTC CGGTGGTTCTATCTTGGCATC	GGCACTCAATGTGAGGGTGA GTCTTTCGCTTCAATAACCCT

2.5. Subcellular Localization

To determine the subcellular localization of GmSWEET46, its coding sequence was cloned into the PstI site of the pGreen-35S:GFP vector to generate an in-frame fusion with GFP. The primers used for this construction are listed in Table 1. The resulting plasmid was transiently expressed in *Arabidopsis thaliana* mesophyll protoplasts. Fluorescence signals were observed and imaged using a confocal laser scanning microscope (TCS SP5; Leica, Wetzlar, Hessen, Germany).

2.6. Multiple Alignment Analysis

The gene sequences of SWEET46 from HN35 and DK3 were sequenced and analyzed. The candidate gene was subjected to PCR by using TransStart® FastPfu DNA Polymerase and sent to TSINGKE Biological Technology Co., Ltd. (Guangzhou, China) for sequencing after agarose gel electrophoresis. DNA homology alignment used Invitrogen Vector NTI 11.5.1 software. Homologous SWEET46 protein sequences were downloaded from Phytozome <https://phytozome-next.jgi.doe.gov> (accessed on 4 May 2024). Amino acid sequences were aligned using the ClustalW program with manual adjustments. The transmembrane (TM) domains and cytoplasmic C-terminal tail of GmSWEET46, predicted to be essential for disaccharide transport, were identified using TMHMM-2.0. Additionally, the 3D protein structures of HN35 and DK3 GmSWEET46 were analyzed using an on-line structural prediction tool <https://golgi.sandbox.google.com/about> (accessed on 2 June 2025).

2.7. Plasmid Construction and Plant Transformation

The coding sequence of GmSWEET46 was amplified from W82 cDNA and cloned into the pCAMBIA1311 vector to generate the overexpression construct. The primers used are listed in Table 1. Transgenic plants were obtained using the floral dip method, and positive transformants were selected on MS medium containing hygromycin. Transgenic hairy roots were produced following a previously established method [16].

3. Results

3.1. Phenotypic Characterization of Parental and F4 Population

To explore the genetic regulation of seed oil and protein content in soybean, a segregating F4 population was generated by crossing the cultivars Heinong 35 (HN35) and Dengke 3 (DK3). Comparative analyses revealed that HN35 exhibited significant increases in seed protein content but a decrease in oil content compared to DK3 in 2022 and 2023 (Figure 1a,b). We also grew the F4 population to investigate the distribution of seed oil and protein content in Hailun (126°14' E, 46°58' N) over the course of two years. The protein content ranged from 36.0% to 45.0% in 2022 and from 35.0% to 45.0% in 2023, while the oil content ranged from 16.8% to 21.9% in 2022 and from 16.5% to 22.0% in 2023 (Figure 1d–g). These results demonstrated continuous variation and an approximately normal distribution, suggesting that seed oil and protein content are complex traits under the control of multiple genes. Notably, the oil and protein content consistently exhibited a stable negative correlation across both years (Figure 1f). Therefore, we chose the extreme pools using both oil and protein content as criteria to identify the key genetic factors that contribute to seed oil and protein content in soybean.

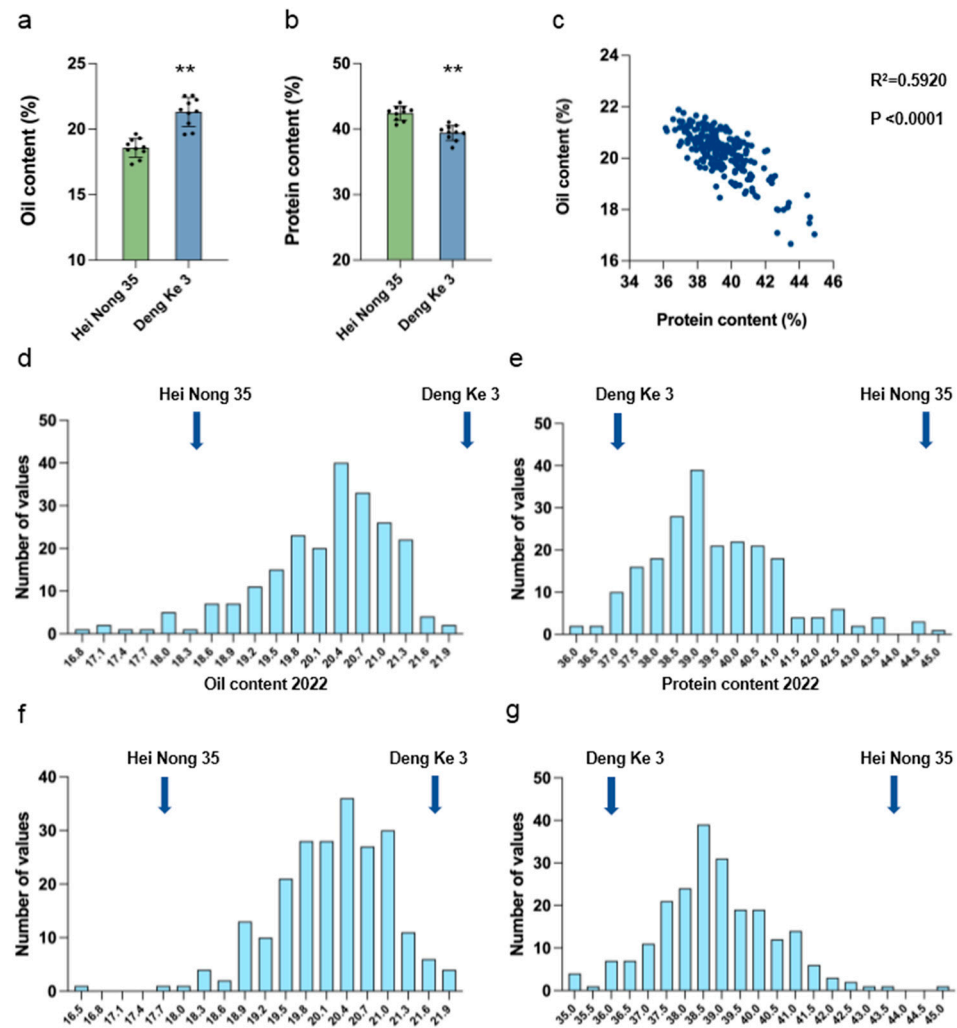


Figure 1. Frequency distribution of oil and protein content in the F4 population along with two parents. (a,b) Comparison of seed oil and protein content between the two parental lines in Suihua field during the 2021 growing season. Values represent means $\pm$ SD ($n = 10$, one plot indicates one plant). Asterisks indicate significant differences between the two parents (** $p < 0.01$, Student's t -test). (c) Seed protein and oil content of the 238-accession panel and the correlation analysis between protein and oil content. Pearson correlation analysis was used to generate the R^2 and the p value. (d–g) Histograms depict the distribution of oil and protein content across 238 F4 plants from the Hailun location in the 2022 and 2023 growing seasons.

3.2. BSA-seq for Seed Oil and Protein Content

The bulked-segregant analysis by sequencing (BSA-seq) was employed to characterize the loci associated with seed oil and protein content. Genomic DNA was extracted from two extreme pools, each comprising 25 F4 plants: one pool with extremely high protein and low oil content and another pool with extremely low protein and high oil content. These phenotypes were confirmed over the years 2022 and 2023. Additionally, genomic DNA was also extracted from the two parental lines for the BSA-seq analysis. A total of 102 Gbp of high-quality, clean sequencing data was obtained. The sequencing depths were $20\times$ for the parental lines and $30\times$ for the extreme phenotypic pools, respectively. The high-quality reads were aligned to the Wm82.a4 reference genome sequence. After filtering, a total of 2,100,647 high-quality single-nucleotide polymorphisms (SNPs) and 474,073 high-quality insertions/deletions (InDels) were identified and used for the subsequent analysis.

To identify the candidate QTLs controlling the seed oil and protein content, we employed the $\Delta(\text{SNP}/\text{InDel-index})$ algorithm, a powerful approach for detecting genomic

regions associated with quantitative traits. By integrating the results from SNPs and InDels, only one genomic interval on chromosome 18 exceeded the threshold (90% confidence interval) and was termed *q-OP18* (Figure 2a). The size of the *q-OP18* locus ranged from 57,985,698 to 58,281,145, with a physical distance of 0.3 Mb. An assessment of the Wm82.a4 reference sequence revealed that the *q-OP18* locus contains 17 annotated genes, including 13 with non-synonymous mutations, 1 with stop gain mutations, and 1 with frameshift mutations between the parental lines (Table 2). The *q-OP18* locus is considered to be involved in regulating seed oil and protein content in soybean.

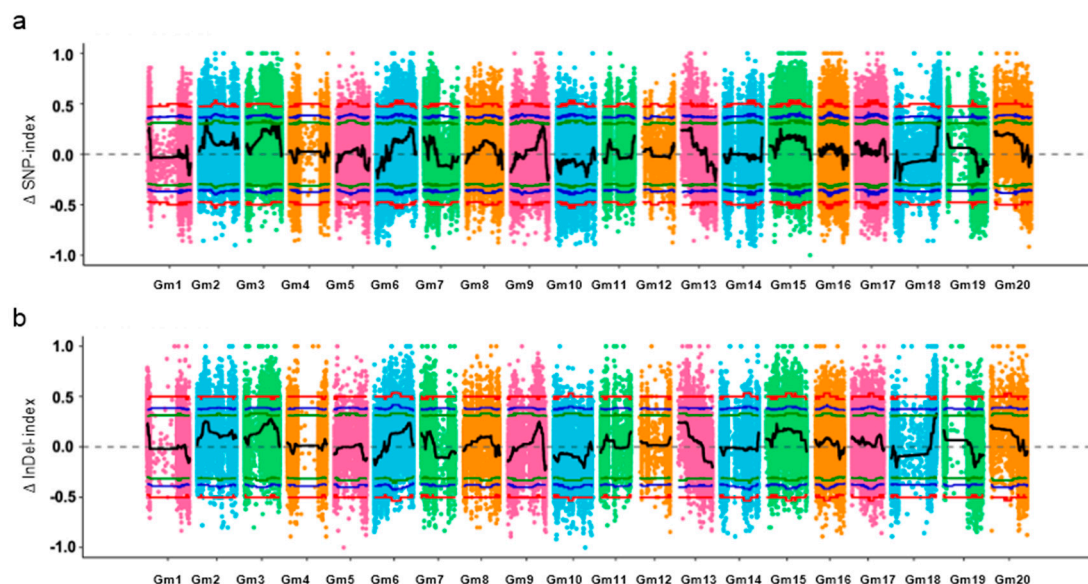


Figure 2. Distribution of SNP-index and InDel-index association values across chromosomes. (a,b) The horizontal coordinate is the chromosome name, the colored dots represent the index value of each SNP or InDel locus, the black line indicates the fitted index value, and the green dashed line represents the significance association threshold 90%; the higher the index value, the better the phenotypic association effect of the locus.

Table 2. Candidate gene predictions for the *q-OP18* locus within the final merged SNP-index and indel-index region.

Gene_ID(W82a4)	Alt	Effect	Position	Annotation
Glyma.18G299500	G > T	Non-synonymous	58017132	EPSIN3
Glyma.18G299600	T > C	Non-synonymous	58021695	Phosphoenolpyruvate carboxylase-like protein
Glyma.18G299800	G > A	Non-synonymous	58029268	Solute carrier family 35, member F1/2 (SLC35F1_2)
Glyma.18G299900	G > A	Non-synonymous	58035105	Probable methyltransferase PMT19
Glyma.18G300200	A > G	Non-synonymous	58072431	Callose synthase 3
Glyma.18G300300	C > T	Non-synonymous	58093863	U-box domain-containing protein 15
Glyma.18G300400	C > T	Non-synonymous	58100916	Protein NRT1/PTR FAMILY 7.1
Glyma.18G300450	A > T	Non-synonymous	58105800	—
Glyma.18G300700	A > C	Non-synonymous	58124213	Signal peptide peptidase-like 5

Table 2. Cont.

Gene_ID(W82a4)	Alt	Effect	Position	Annotation
Glyma.18G300800	A > T	Non-synonymous	58129157	Pollen ole e 1 allergen and extensin family protein
Glyma.18G301200	T > A	Non-synonymous	58172713	Bidirectional sugar transporter sweet15
Glyma.18G301300	G > T	Non-synonymous	58183227	—
Glyma.18G301600	A > AG	Frame shift	58211303	Uncharacterized protein At2g33490
Glyma.18G301950	G > A	Stop gain	58227083	Replication factor a 1, RFA1
Glyma.18G302200	G > T	Non-synonymous	58277925	Histone-lysine N-methyltransferase ASHR3

3.3. Identification of *GmSWEET46* in the *q-OP18* Locus

To screen out candidate genes that influence oil and protein content in soybean seeds, we analyzed all candidate genes with non-synonymous mutations in the *q-OP18* locus (Table 2). Among these, the gene *Glyma.18G301200* encodes a protein that is an ortholog of the *Arabidopsis* AtSWEET15 sugar transporter (AT5G13170), which has been shown to play a role in lipid accumulation in seeds [17]. In a previous analysis of the SWEET gene family, this gene was designated as *GmSWEET46* [13], so we adopt this designation in the present study. Haplotype analysis showed that one SNP in the sixth exon results in an amino acid change of the *GmSWEET46* protein from Glutamic acid (Glu) in HN35 to Valine (Val) in DK3 (Figure 3a,b and Figure S1). Like other SWEET proteins, *GmSWEET46* also have seven predicted transmembrane (TM) domains and a cytoplasmic C-terminal tail (Figure 3c). The amino acid change of the *GmSWEET46* protein caused an obvious change in the predicted 3D protein structure between the haplotypes of HN35 and DK3 (Figure 3d). Previous studies have highlighted the important roles that SWEET proteins play in facilitating sugar translocation to seeds, thereby affecting seed-related traits in plants [11,18]. Given this evidence, we propose that *GmSWEET46* is a candidate gene in the *q-OP18* locus for controlling the seed oil and protein content in soybean.

To elucidate the role of *GmSWEET46* in soybean seed development, we examined its expression pattern using multi-omics database reverse transcription quantitative PCR (RT-qPCR) across different soybean tissues. The results showed that *GmSWEET46* was highly expressed in the early stage of pod development and slightly expressed in the early stages of seed development (Figure 4a,b). This expression profile suggests that *GmSWEET46* may be involved in the regulation of seed development in soybean. To further explore the subcellular localization of the *GmSWEET46* protein, we generated a fusion construct by fusing the coding sequence of *GmSWEET46* to *GFP* and transiently expressed it in leaves of *Nicotiana benthamiana*. As shown in Figure 4c, the *GmSWEET46*-GFP fusion protein was localized to the plasma membrane, as evidenced by its co-localization with AtSCAMP-mCherry, a well-established marker for plasma membrane proteins [16]. This finding is consistent with the subcellular localization patterns for other soybean SWEET proteins [10,12].

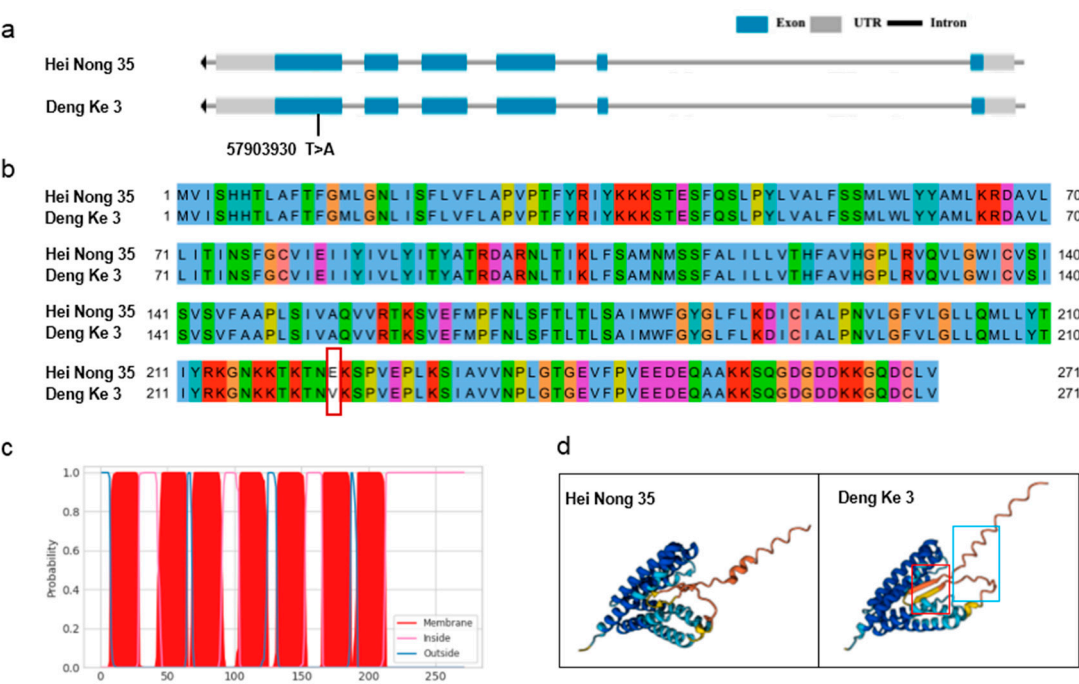


Figure 3. Sequence and analysis of *GmSWEET46* alleles. (a) Allelic variation in the *GmSWEET46* between Hei Nong 35 and Deng Ke 3. (b) Peptide sequence alignment of *GmSWEET46*: rectangular windows indicate different *GmSWEET46* peptides. (c) *GmSWEET46* is conserved at the amino acids essential for disaccharide transport. (d) Comparison of the 3D protein structure between HN35 and DK3 *GmSWEET46* proteins. The difference in the structure at the C-terminal tails is highlighted with rectangles in blue and red.

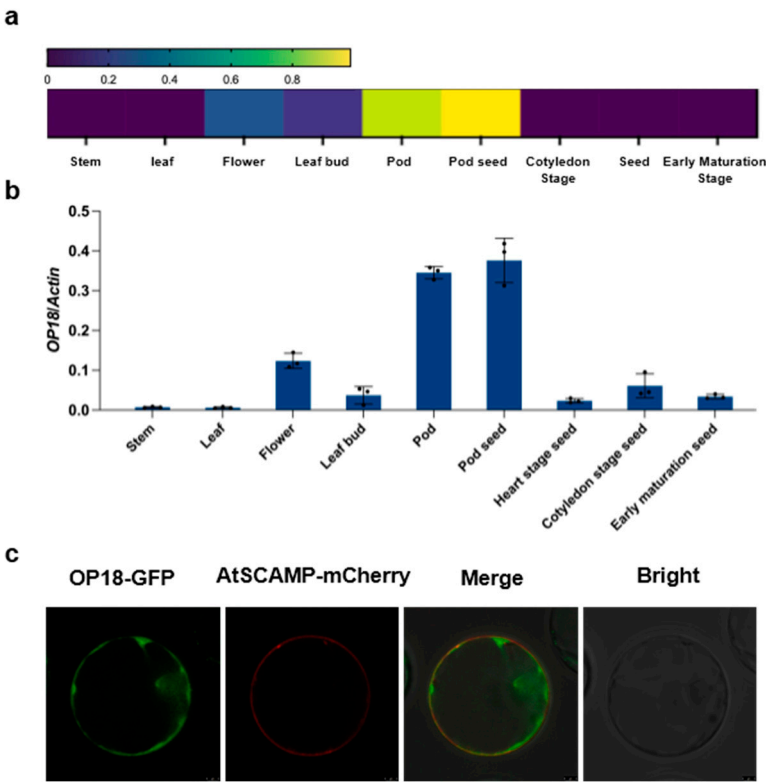


Figure 4. Expression pattern of the candidate gene *GmSWEET46*. (a) The TPM (Transcripts Per Million) value was extracted from the SoyMD multi-omics database (<https://yanglab.hzau.edu.cn/SoyMD/#/>, accessed on 4 May 2024). (b) The expression pattern *GmSWEET46* in Williams82 under

LDs. *Actin* was used as an internal control. Data are means $\pm$ SEM ($n = 3$); the value of each technical replicate is represented by a dot. (c) Subcellular localization of the GmSWEET46 protein in soybean protoplasts. GmSWEET46-GFP is the GmSWEET46-GFP fusion protein' localization; AtSCAMP-mCherry is the plasma membrane marker; and Merge is the merge of GFP and mCherry. Scale bars, 5 μ m.

3.4. GmSWEET46 Regulates Seed Oil and Protein Content

In order to investigate the function of *GmSWEET46* in regulating seed oil and protein content, the full-length *GmSWEET46* cDNA was introduced into the *Arabidopsis* ecotype Col-0 background under the control of the strong and constitutive CaMV 35S promoter. We obtained three independent transgenic lines (OE-1, OE-2, and OE-3) with significantly elevated *GmSWEET46* expression levels for further analyses (Figure S2). Compared with the Col-0 wild type, the *GmSWEET46* overexpressing lines exhibited significantly higher oil content and lower protein content in mature seeds (Figure 5a,b). In addition, overexpression of *GmSWEET46* also led to an increase in seed size compared to Col-0 (Figure 5b). These results indicate that *GmSWEET46* can regulate seed weight and quality in *Arabidopsis*.

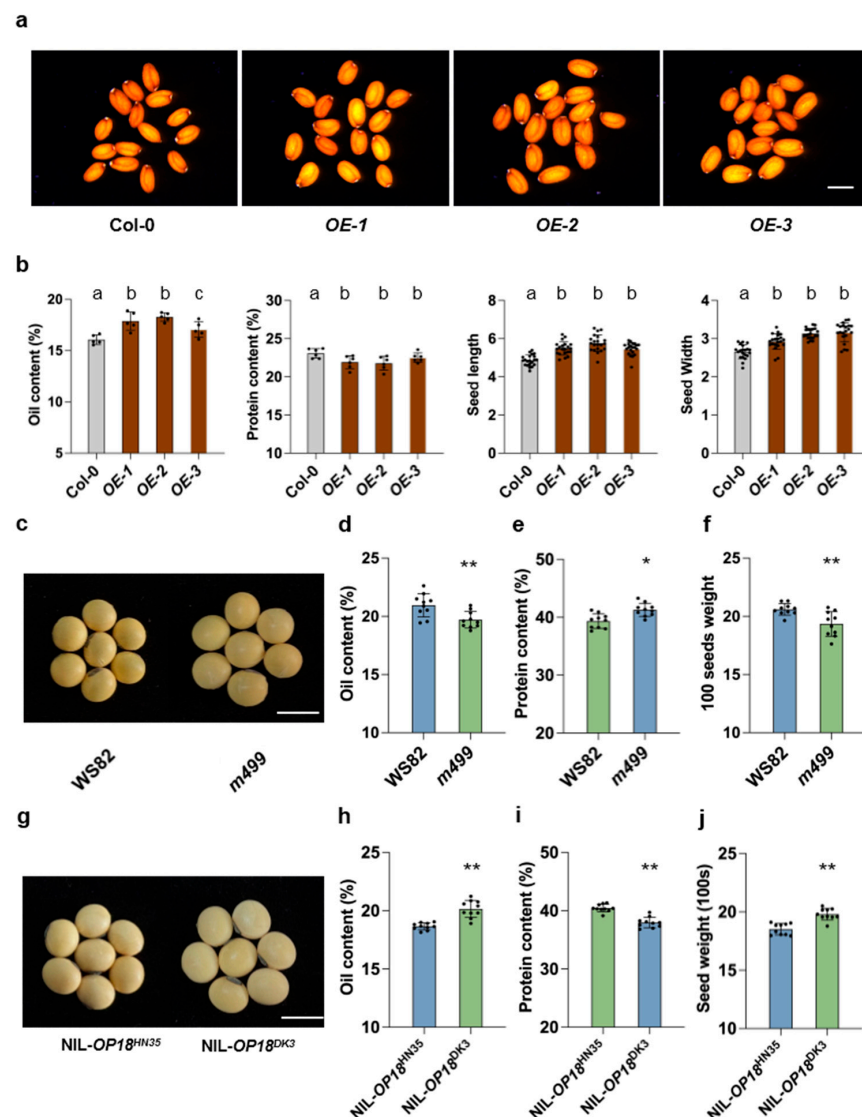


Figure 5. Functional characterization of *GmSWEET46* in controlling oil content and seed weight. (a,b) Verification of *GmSWEET46*-overexpressing *Arabidopsis* lines. Values represent means $\pm$ SD ($n = 5$, one plot indicates one plant). Different lowercase letters indicate statistically significant differences (one-way

ANOVA, $p < 0.01$). (c–f) Seed phenotypes of the near-isogenic lines (NILs) possessing homozygous NIL-*GmSWEET46*^{HN35} and NIL-*GmSWEET46*^{DK3} in Suihua field. Scale bar, 1 cm. b–e Statistical analysis of the oil content (d), protein content (e), and 100-seed weight (f) of NILs. Values represent means $\pm$ SD ($n = 10$, one plot indicates one plant). (g–j) Seed phenotypes of Williams 82 (W82) and the *GmSWEET46* EMS mutant line m499 in Sanya field. Scale bar, 1 cm. (h–j) Statistical analysis of the oil content (h), protein content (i), and 100-seed weight (j) of W82 and the EMS mutant lines. Values represent means $\pm$ SD ($n = 10$, one plot indicates one plant). Asterisks indicate significant differences between the NILs and EMS mutant lines (* $p < 0.05$, ** $p < 0.01$, Student's *t*-test).

To preliminarily test the function of *GmSWEET46* related to oil and protein content in soybeans, we measured the oil and protein content in EMS mutant line m499, which exhibited a non-synonymous mutation in *GmSWEET46*. Compared to the wild-type Wm82, m499 had a higher protein content but a lower oil content and seed weight (Figure 5c–f). We next investigated the phenotypic effects of different *GmSWEET46* haplotypes on seed traits in soybean. To this end, we developed a set of near-isogenic lines (NILs) containing the *GmSWEET46* locus by crossing the HN35 and DK3 parents, which harbor the *GmSWEET46*^{HN35} haplotype and the *GmSWEET46*^{DK3} haplotype, respectively. Phenotypic analysis revealed that NIL-*GmSWEET46*^{DK3} exhibited significantly higher 100-seed weight and oil content compared to NIL-*GmSWEET46*^{HN35} (Figure 5g–i). Conversely, the protein content in NIL-*GmSWEET46*^{DK3} was lower than that in NIL-*GmSWEET46*^{HN35} (Figure 5j). Taken together, these results suggest that *GmSWEET46* plays an important role in determining oil content, protein content, and seed weight in soybean.

4. Discussion

Soybean is a globally important commercial crop that serves as a primary source of edible oil and protein for both human consumption and livestock feed. Seed protein content, oil content, and seed size, quantitative traits regulated by multiple genetic loci, are key determinants of soybean quality and yield [18]. Elucidating the key genes for these seed traits is critical for advancing soybean breeding programs. With rising global demand for soybean products, there is an urgent need to develop high-quality/yield varieties through the integration of genomic technologies with conventional breeding strategies [19]. However, the genetic mechanisms governing natural variation in seed composition traits remain poorly understood. In the present study, we identified a major locus associated with seed quality traits and cloned the corresponding gene, *GmSWEET46*. Functional characterization demonstrated that *GmSWEET46* modulates seed protein content, oil content, and seed size in both Arabidopsis and soybean EMS mutants. However, the detailed functions of *GmSWEET46* in soybean seeds still need to be further investigated.

The SWEET protein family has been extensively studied for its role in sugar transport from source to sink tissues, with developing seeds representing a major sink organ. Numerous studies have demonstrated that SWEET genes play crucial players in plant seed development [8,9]. In soybean, 52 SWEET genes have been identified. Among them, *GmSWEET10a* and *GmSWEET10b* have been reported to control seed traits by regulating sugar allocation from the seed coat to embryo, while *GmSWEET15* was shown to mediate sucrose transport from the endosperm to the early embryo [10,11]. Additionally, the expression levels of several SWEET genes are modulated by *GmMFT*, a member of the PEBP family, which in turn influences the regulation of seed weight and quality [20]. In potato, a SWEET protein physically interacts with an FT homolog (another PEBP member) to inhibit apoplastic sucrose leakage and promote symplastic sucrose assimilation [21]. Consistent with these findings, our results also demonstrated that the *GmSWEET46* protein localizes to the plasma membrane, indicating its potential function as a sugar transporter in soybean. However, the underlying molecular mechanisms by which *GmSWEET46* regulates seed

traits still need to be further elucidated; for instance, whether *GmSWEET46* is regulated by or interacts with other factors to influence seed traits requires further investigation.

The natural variation within the soybean genome plays a crucial role in shaping its phenotypic traits [22]. This genetic diversity provides the foundation for the adaptation and improvement of soybean varieties to meet diverse agricultural demands [23]. For instance, natural variation in *GmSWEET39* has been shown to influence oil and protein contents as well as seed size in soybean, and *GmSWEET39* has been selected to increase seed oil content during soybean domestication and improvement [4,11,12]. In this study, we identified one SNP in the sixth exon of *GmSWEET46* that results in an amino acid change between the varieties HN35 and DK3. Further analysis revealed that this natural variation in *GmSWEET46* affects seed oil content, protein content, and seed size, highlighting its potential as a valuable genetic resource for molecular breeding efforts in soybean. However, whether this variation in *GmSWEET46* was selected during domestication remains an open question, and how it impacts the protein's function also requires further investigation. Additionally, the potential of *GmSWEET46* for application in breeding high-quality soybean cultivars requires further evaluation. Pyramiding this favorable allele with other beneficial alleles, such as *GmSWEET39* [12] and *POWR1* [5], presents a promising strategy for overcoming the typical trade-off between protein and oil content in soybean seeds.

5. Conclusions

This study employed BSA-seq on an F_4 population from soybean lines HN35 and DK3 to identify a major QTL, *q-OP18*, governing seed oil and protein content. Further analyses revealed that the causal gene underlying this QTL is *GmSWEET46*, which encodes a sugar transporter protein. A key SNP in its sixth exon causes an amino acid change between parental lines, which significantly correlates with oil and protein content. Functional analyses showed *GmSWEET46* is highly expressed during early seed development, localizes to the plasma membrane, and modulates seed composition and size, as validated in transgenic *Arabidopsis* and soybean EMS mutants. These findings deepen understanding of *SWEET* genes in soybean seed traits and provide a key genetic resource for quality improvement. Notably, *GmSWEET46* and its SNP can be developed as molecular markers for soybean MAS breeding.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/agronomy15092198/s1>, Figure S1. Sequence alignment of *GmSWEET46*. Rectangular windows indicate different *GmSWEET46* alleles. Figure S2. *GmSWEET46* expression levels in transgenic *Arabidopsis* plants. Col-0 was used as a negative control. *GmActin* was used as an internal control. Values represent means $\pm$ SD ($n = 3$).

Author Contributions: T.L. and W.L. (Wencheng Lu) planned and coordinated this research. T.L. designed the experimental details. D.H., H.S., Q.L., and W.L. (Wei Li) performed the wet-lab experiments and the bioinformatics analyses. T.L., D.H., and H.S. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Key Research and Development Plan of Heilongjiang Province (2024ZXDxB34), the National Natural Science Foundation of China (32201800), and the program of the Natural Science Foundation of Guangdong province (2023A1515011645). This work was also supported by the Heilongjiang Province agricultural science and technology innovation leap project (CX25JC01).

Data Availability Statement: The data presented in the study are available in the “Bioproject” under accession number “PRJNA896173”.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Leamy, L.J.; Zhang, H.; Li, C.; Chen, C.Y.; Song, B.H. A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genom.* **2017**, *18*, 18. [\[CrossRef\]](#)
2. Godfray, H.C.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food security: The challenge of feeding 9 billion people. *Science* **2010**, *327*, 812–818. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Lee, S.; Van, K.; Sung, M.; Nelson, R.; LaMantia, J.; McHale, L.K.; Mian, M.A.R. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *TAG. Theor. Appl. Genet.* **2019**, *132*, 1639–1659. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Zhang, H.; Goettel, W.; Song, Q.; Jiang, H.; Hu, Z.; Wang, M.L.; An, Y.C. Selection of GmSWEET39 for oil and protein improvement in soybean. *PLoS Genet.* **2020**, *16*, e1009114. [\[CrossRef\]](#)
5. Goettel, W.; Zhang, H.; Li, Y.; Qiao, Z.; Jiang, H.; Hou, D.; Song, Q.; Pantalone, V.R.; Song, B.H.; Yu, D.; et al. POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nat. Commun.* **2022**, *13*, 3051. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Baker, R.F.; Leach, K.A.; Braun, D.M. SWEET as sugar: New sucrose effluxers in plants. *Mol. Plant* **2012**, *5*, 766–768. [\[CrossRef\]](#)
7. Ruan, Y.L. Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. *Annu. Rev. Plant Biol.* **2014**, *65*, 33–67. [\[CrossRef\]](#)
8. Chen, L.Q.; Lin, I.W.; Qu, X.Q.; Sosso, D.; McFarlane, H.E.; Londono, A.; Samuels, A.L.; Frommer, W.B. A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the Arabidopsis embryo. *Plant Cell* **2015**, *27*, 607–619. [\[CrossRef\]](#)
9. Yang, J.; Luo, D.; Yang, B.; Frommer, W.B.; Eom, J.S. SWEET11 and 15 as key players in seed filling in rice. *New Phytol.* **2018**, *218*, 604–615. [\[CrossRef\]](#)
10. Wang, S.; Yokosho, K.; Guo, R.; Whelan, J.; Ruan, Y.L.; Ma, J.F.; Shou, H. The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo. *Plant Physiol.* **2019**, *180*, 2133–2141. [\[CrossRef\]](#)
11. 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. Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *Natl. Sci. Rev.* **2020**, *7*, 1776–1786. [\[CrossRef\]](#)
12. Miao, L.; Yang, S.; Zhang, K.; He, J.; Wu, C.; Ren, Y.; Gai, J.; Li, Y. Natural variation and selection in GmSWEET39 affect soybean seed oil content. *New Phytol.* **2020**, *225*, 1651–1666.
13. Patil, G.; Valliyodan, B.; Deshmukh, R.; Prince, S.; Nicander, B.; Zhao, M.; Sonah, H.; Song, L.; Lin, L.; Chaudhary, J.; et al. Soybean (*Glycine max*) SWEET gene family: Insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genom.* **2015**, *16*, 520. [\[CrossRef\]](#)
14. Van, K.; McHale, L.K. Meta-Analyses of QTLs associated with oil and protein contents and compositions in soybean [*Glycine max* (L.) Merr.] seed. *Int. J. Mol. Sci.* **2017**, *18*, 1180.
15. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome sequence of the palaeopolyploid soybean. *Nature* **2010**, *463*, 178–183. [\[CrossRef\]](#)
16. Li, X.; Chen, Z.; Li, H.; Yue, L.; Tan, C.; Liu, H.; Hu, Y.; Yang, Y.; Yao, X.; Kong, L.; et al. Dt1 inhibits SWEET-mediated sucrose transport to regulate photoperiod-dependent seed weight in soybean. *Mol. Plant* **2024**, *17*, 496–508.
17. Chen, L.Q.; Cheung, L.S.; Feng, L.; Tanner, W.; Frommer, W.B. Transport of sugars. *Annu. Rev. Biochem.* **2015**, *84*, 865–894. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Bheemanahalli, R.; Poudel, S.; Alsajri, F.A.; Reddy, K.R. Phenotyping of southern United States soybean cultivars for potential seed weight and seed quality compositions. *Agronomy* **2022**, *12*, 839. [\[CrossRef\]](#)
19. Yuan, X.; Jiang, X.; Zhang, M.; Wang, L.; Jiao, W.; Chen, H.; Mao, J.; Ye, W.; Song, Q. Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. *Plant Cell* **2024**, *36*, 2160–2175. [\[CrossRef\]](#)
20. 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*, 905–919.
21. Abelenda, J.A.; Bergonzi, S.; Oortwijn, M.; Sonnewald, S.; Du, M.; Visser, R.G.F.; Sonnewald, U.; Bachem, C.W.B. Source–sink regulation is mediated by interaction of an FT homolog with a SWEET protein in potato. *Curr. Biol.* **2019**, *29*, 1178–1186. [\[CrossRef\]](#)
22. Valliyodan, B.; Dan, Q.; Patil, G.; Zeng, P.; Huang, J.; Dai, L.; Chen, C.; Li, Y.; Joshi, T.; Song, L.; et al. Landscape of genomic diversity and trait discovery in soybean. *Sci. Rep.* **2016**, *6*, 23598. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Li, J.; Li, Y.; Agyenim-Boateng, K.G.; Shaibu, A.S.; Liu, Y.; Feng, Y.; Qi, J.; Li, B.; Zhang, S.; Sun, J. Natural variation of domestication-related genes contributed to latitudinal expansion and adaptation in soybean. *BMC Plant Biol.* **2024**, *24*, 651. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.