Seed Biology

REVIEW

Open Access

https://doi.org/10.48130/seedbio-0026-0004

Seed Biology **2026**, 5: e008

Advances in understanding domestication-related genes for critical seed traits in soybean

Baoyin Chen[#], Congcong Wang[#], Yongbin Zhuang, Xiaoming Li, Jinfei Zhang and Dajian Zhang[*]

National Key Laboratory of Wheat Improvement, College of Agronomy, Shandong Agricultural University, Tai'an 271018, China

[#] Authors contributed equally: Baoyin Chen, Congcong Wang

[*] Correspondence: dajianzhang@sdau.edu.cn (Zhang D)

Abstract

Soybeans are a vital seed crop, supplying a substantial amount of protein and oil globally, for both human and animal consumption. Modern soybean (*Glycine max* L. Merr.) varieties have been developed through long-term domestication and genetic introgression of wild ancestors (*Glycine soja* Sieb. and Zucc.). This process involved significant alterations in seed traits, including seed morphology, size, and quality. In recent years, quantitative trait loci (QTLs), and key genes associated with seed domestication traits have been identified via analytical methods such as QTL mapping, genome-wide association studies, whole-genome resequencing, and -omics applications. This review summarizes recent research on domestication-related genes that regulate key seed traits, including size, oil content, and protein content. Additionally, we discuss the application of new technologies, such as -omics technologies, to advance research involving domestication-related genes. This work will help us better understand soybean domestication in terms of seed traits and should expedite the development of elite soybean varieties to meet the growing demands for soybean production, and ensure global food security.

Citation: Chen B, Wang C, Zhuang Y, Li X, Zhang J, et al. 2026. Advances in understanding domestication-related genes for critical seed traits in soybean. *Seed Biology* 5: e008 https://doi.org/10.48130/seedbio-0026-0004

Introduction

Domestication represents the earliest form of plant breeding and is a pivotal technological milestone in human history. During domestication, humans cultivated and improved wild plants, promoting reliable harvests that met their needs. Since the rise of agricultural civilizations around 10,000 years ago, numerous plant species have been domesticated from their wild ancestor, including legumes, rice, wheat, and corn—crops that now sustain billions of people globally. Traditionally, the process of domestication was extremely slow, often taking over a millennium, or even several millennia, to fix a single trait in a founder population[1]. This slow pace was largely due to the lack of effective tools (such as CRISPR-Cas9, high-throughput phenotyping, marker-assisted selection, and genomic selection), and the limited seed dispersal of available germplasm by ancient farmers[2]. However, current rapid advances in biotechnology have made modern breeding much more efficient, often requiring only a few decades to fix an undesirable trait in a population. To expedite the *de novo* domestication of wild plants and improve existing crops, it is essential for plant geneticists and breeders to understand the genetic principles underlying specific domestication traits, and to trace the crucial molecular changes that have occurred during the domestication process.

Seeds are crucial not only as a primary source of nutrition, but also for propagating food crops[3]. With the global population steadily increasing, the demand for food crops is expected to rise considerably. However, seed production faces multiple challenges due to environmental changes, including global temperature rise, soil salinization, desertification, and extreme weather events, all of which threaten food security. These changes, many of which are unpredictable, are likely to dynamically alter the breeding objectives of crop breeders, and impact the domestication of new seed crops.

Cultivated soybean (*Glycine max* L. Merr.) is one of the most economically significant leguminous seed crops, providing a rich source of protein, essential amino acids, oil, and metabolizable energy. The seeds of the cultivated soybean typically contain approximately 40% protein, 20% oil, 35% carbohydrates, and 5% ash, in addition to other important components such as amino acids, macronutrients, micronutrients, and sugars[4,5]. Dietary protein and vegetable oil provide the two main economic benefits of soybean, contributing up to 69% and 30% of the world's food and animal feed, respectively (www.usda.gov). Cultivated soybean was domesticated from its wild ancestor (*G. soja* Sieb. & Zucc.) in China approximately 6,000–9,000 years ago[6]. During soybean domestication, various traits were selected for, including the loss of seed shattering, reduced seed oil content, and increased seed size. Accordingly, *G. soja* has been used to study the nature of domesticated seed traits in *G. max*[7–10].

Since the release of the genomes for the soybean cultivar Williams 82[11], and the wild soybean variety IT182932[12], research into the evolutionary history of domesticated soybeans has advanced rapidly. Numerous nucleotide sequence differences between wild and cultivated soybeans have been identified[12,13]. In recent years, advances in resequencing and next-generation sequencing have significantly enhanced soybean population genetics. More than 500 selective regions related to domestication have been identified[7,14–16]. These recent domestication-selective strategies now provide a valuable resource for characterizing new genes involved in soybean domestication.

To further identify and characterize domestication-related genes, a genome-wide association study (GWAS) analysis was employed to examine the genetic basis of seed traits underlying domestication in soybean, traits such as seed size/weight, and oil and protein contents[5,17–20]. Recently, multi-omics analysis has identified candidate domestication genes associated with desirable soybean seed traits. For instance, the integration of GWAS, expression QTL (eQTL) information, and transcriptome-wide association studies (TWAS), has identified 22 candidate causal genes responsible for preferred seed traits. These include *GmRWOS1*, which regulates seed weight and oil content[21]. This review highlights recent research progress in understanding the domestication of soybean seed traits such as

size, oil content, and protein content, and it focuses on the evolution of these traits during the domestication process (Fig. 1,Table 1). Additionally, we discuss the applications of new technologies in basic domestication research, along with the challenges and prospects for future soybean studies and breeding.

Seed size and weight

Regulation of seed size

Seed size is a crucial agronomic trait for domesticated crops. Increasing seed size is a key objective for yield improvement[22]. Therefore, enhancing yield by increasing seed size is a primary breeding objective for plant breeders. *G. soja* has small seeds compared to *G. max*[23], and consequently seed size is a key indicator of domestication and can be characterized by three main features: length, width, and thickness[24]. Compared to wild seeds, cultivated soybean seeds are longer, wider, and thicker, with a strong positive correlation between seed size and seed weight. This pleiotropic trait is controlled by numerous genes and environmental factors, making it a crucial selection factor during the domestication process[25]. In recent years, over 400 QTLs associated with soybean seed size have been mapped, and they are found across nearly all 20 of the soybean chromosomes[26–28]. Most of these QTLs are recorded in SoyBase (www.soybase.org).

Overall, cultivated soybean seeds are rounder than wild soybean seeds. This morphological trait is primarily determined by thickness, assessed by the diameter perpendicular to the hilum. Through GWAS and map-based cloning, the *Seed Thickness 1* (*ST1*) locus on chromosome 8 was identified; it encodes a UDP-D-glucuronate 4-epimerase that influences seed thickness by catalyzing pectin biosynthesis[29]. This locus is also located within a soybean oil content QTL, situated in a ~160-kb 'selective sweep' region shaped by soybean domestication[29–31]. Overexpression of *ST1* enhances oil accumulation in seeds by indirectly affecting glycolytic levels, thereby influencing key metabolites in fatty acid (FA) metabolic pathways[29]. Analysis of sequence variation at the *ST1* locus among 1,209 accessions revealed that *ST1* has undergone selection during soybean domestication[29].

Through a GWAS of 1,853 soybean accessions, a natural allelic variation at *GmST05* (*Seed Thickness on Chromosome 5*) was identified as predominantly controlling seed thickness and size in soybean[17]. *GmST05* encodes a protein in the phosphatidylethanolamine-binding protein (PEBP) family that is homologous to *MOTHER OF TFL1 AND FT* (*MFT*) in Arabidopsis (*Arabidopsis thaliana*)[17,32]. Two haplotypes of *GmST05* showed significant differences at the transcriptional level, with haplotype I exhibiting a significantly greater seed thickness than haplotype II[17]. The proportion of $GmST05^{HapI}$ to $GmST05^{HapII}$ increased progressively from wild soybean, to landraces, to cultivars, suggesting that *GmST05* underwent artificial selection and subsequent local breeding during soybean domestication[17]. Transgenic experiments have demonstrated that *GmST05* positively regulated seed size and oil content, while negatively regulating protein content, possibly by regulating the expression of *GmSWEET10a* (also known as *GmSWEET39*), which is involved in the transport of sucrose[17].

In addition to seed thickness, seed morphology is determined by seed length and width. Approximately 52 QTLs are related to seed length, and 32 QTLs are related to seed width (http://soybase.ncgr.org). Recently, a major QTL, *GmSW17* (*Seed Width 17*), that determines soybean seed width/weight in the natural population, was reported on chromosome 17[20]. Transgenic experiments have demonstrated that *GmSW17* positively regulated seed size. *GmSW17* encodes a ubiquitin-specific protease orthologous to UBP22 of the ubiquitin-specific protease family. It forms a deubiquitinase module with GmSGF11 and GmENY2 that modulates H2B ubiquitination levels, thereby negatively regulating the expression of *GmDP-E2F-1*. Although population analysis showed that *GmSW17* underwent artificial selection during soybean domestication, its functional alleles have not been fixed in modern breeding. As no other key genes controlling seed length or seed width have been identified to date, further investigation is needed to identify genes determining these two traits, and to pinpoint precisely how seed morphology is regulated in soybean.

Regulation of seed weight

Seed weight is a crucial trait influencing yield, and it is typically measured by the weight of 100 seeds. Over 300 QTLs associated

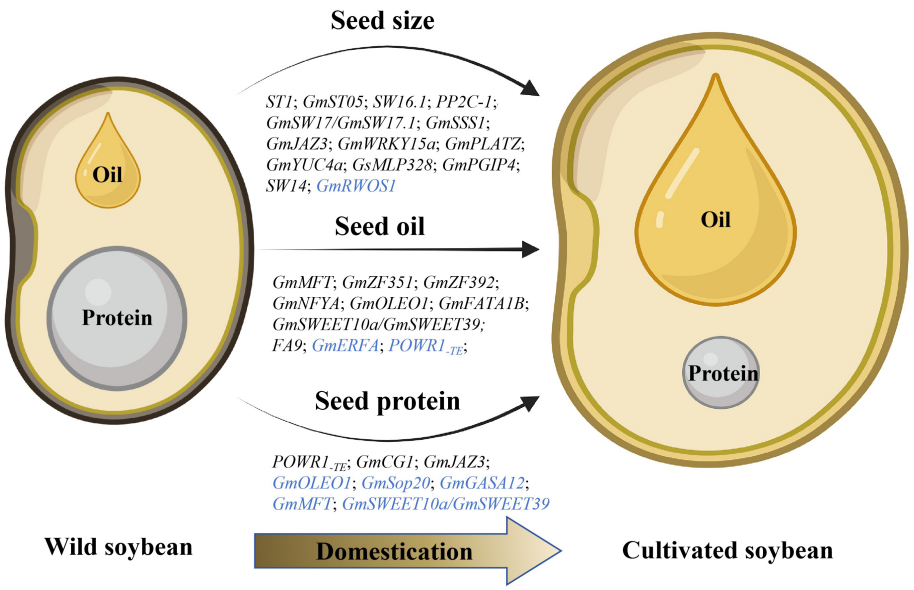

Fig. 1 Domestication genes related to critical seed traits in soybean. Black and blue letters represent positive and negative regulatory genes, respectively.

Table 1. Published domestication genes related to seed traits in soybean.

Trait	Name	Locus	Conserved domain or function	Mutant type	Ref.
Seed size	*ST1*	*Glyma.08G109100*	UDP-D-glucuronate 4-epimerase	SNP in CDS	[29]
	GmST05/GmMFT	*SoyZH13_05G229200*	Mother of TFL1 and FT (MFT)	Indel and SNP in promoter and SNP in CDS	[17,50]
Seed weight	*SW16.1*	*Glyma.16G198300*	LIM domain-containing transcription factor	Indel in CDS	[35]
	SW14	*Glyma.14G010000*	Nuclear factor Y subunit	SNP in CDS	[33]
	PP2C-1	*Glyma17g33690*	Putative phosphatase 2C protein	Indel and SNPin CDS	[36]
	GmSW17/GmSW17.1	*SoyZH13_17G105400/ Glyma.17G109100*	Ubiquitin-specific protease	SNP in CDS	[20,37]
	GmSSS1	*Glyma.19G196000*	Putative O-GlcNAc transferase	SNP in CDS	[40]
	GmJAZ3	*Glyma.09G123600*	JASMONATE-ZIM DOMAIN (JAZ)	Indel and SNPin promoter	[44]
	GmWRKY15a	*Glyma05g20710*	WRKY transcription factor	Indel in 5' UTR	[45]
	GmPLATZ	*Glyma.13G147800*	AT-rich sequence and zinc-binding (PLATZ) family protein	Indel and SNP in promoter	[46]
	GmRWOS1	*Glyma.12G064800*	K-stimulated pyrophosphate-energized sodium pump protein	Indel and SNP in CDS	[21]
	GsMLP328	*Glysoja.20G052861*	Soybean major latex protein (MLP)	SNP in CDS	[42]
	GmPGIP4	*Glyma.08G079200*	Polygalacturonase-inhibiting proteins (PGIP)	SNP in CDS	[42]
	GmYUC4a	*Glyma.14G124100*	YUCCA (YUC) flavin monooxygenases	Indel in 3' UTR and SNP in downstream region	[10]
Seed oil content and component	*GmST05/GmMFT*	*SoyZH13_05G229200*	Mother of TFL1 and FT (MFT)	Indel and SNP in promoter and SNP in CDS	[17,50]
	GmZF351	*Glyma06g44440*	Tandem zinc-finger CCCH-domain protein	Indel and SNP in promoter	[51]
	GmZF392	*Glyma.12G205700*	Tandem zinc-finger CCCH-domain protein	SNP in promoter	[54]
	GmSWEET39/ GmSWEET10a	*Glyma.15G049200*	Sugars will eventually be exported transporter	Indel in CDS	[18,19,56,57]
	GmNFYA	*Glyma02g47380*	Nuclear factor Y subunit	Indel in promoter	[54]
	POWR1	*Glyma.20G085100*	CCT domain	Indel in CDS	[5]
	GmOLEO1	*Glyma.20G196600*	Putative oleosin protein	Indel and SNP in promoter	[59]
	GmFATA1B	*Glyma.08G349200*	Acyl-ACP thioesterases (FAT)	Indel in CDS	[67]
	FA9	*Glyma.09G250400*	SEIPIN protein	SNP in CDS	[68]
	GmERFA	*Glyma.02G016100*	AP2/ERF-type transcription factor	Indel and SNP in promoter	[55]
Tocopherol content	*GmMQT/GmST05*	*Glyma.NDD2.05G269100*	Mother of TFL1 and FT (MFT)	SNP in promoter	[79]
	GmZF351	*Glyma06g44440*	Tandem zinc-finger CCCH-domain protein	SNP in promoter	[76]
Seed protein content	*POWR1*	*Glyma.20G085100*	CCT domain	Indel in CDS	[5]
	GmSWEET10a/ GmSWEET39	*Glyma.15G049200*	Sugars will eventually be exported transporter	Indel in CDS	[18,19,56,57]
	GmST05/ GmMFT	*SoyZH13_05G229200*	Mother of TFL1 and FT (MFT)	Indel and SNP in promoter and SNP in CDS	[17,50]
	GmOLEO1	*Glyma.20G196600*	Putative oleosin protein	Indel in promoter	[59]
	GmJAZ3	*Glyma.09G123600*	JASMONATE-ZIM DOMAIN (JAZ)	SNP in promoter	[44]
	GmSop20	*Glyma.20G005900*	C2H2-type zinc-finger transcription factor	Indel in CDS	[84]
	GmCG-1	*Glyma.10G246300*	α' subunit of β-conglycinin	SNP in CDS	[88]
	GmGASA12	*Glyma.08G291900*	GA-regulated family protein	SNP in CDS and promoter	[90]

with soybean seed weight have been identified, scattered across its 20 chromosomes (www.soybase.org). On chromosome 14, *Nuclear factor Y Subunit A* (*NFYA*) (*Glyma.14G010000*), which encodes Seed Weight 14 (SW14), a positive regulator of seed weight, was identified through GWAS[33]. SW14 physically interacts with soybean Leafy Cotyledon 1 (GmLEC1), a key regulator of seed development, and subsequently inhibits the formation and transactivating activity of the GmLEC1/GmNF-YC2/GmbZIP67 trimeric complex. A haplotype analysis found that the natural allele, *SW14*^{H3}, has undergone artificial selection during soybean domestication. A QTL for 100-seed weight, *SW16.1*, was identified on chromosome 16 using mapping populations derived from chromosome segment substitution lines of wild soybean[34,35]. *SW16.1* encodes a transcription factor containing a LIM domain, which contains two zinc fingers, and is involved in protein interactions. It controls 100-seed weight by regulating the transcription of *Metallothionein 4* (*MT4*), which encodes a positive regulator of seed weight[35]. Polymorphism analysis reveals that the frequencies of superior *SW16.1* alleles have increased in cultivated soybean compared to wild soybean[35]. On chromosome 17, a QTL for 100-seed weight, *phosphatase 2C-1* (*PP2C-1*), was mapped using a genetically stable population of RILs derived from a cross between wild soybean ZYD7, and cultivated soybean HN44[36]. PP2C-1 promotes cell size in association with GmBZR1, and the *PP2C* variant ZYD7 from wild soybean contributes to increased seed weight, suggesting a potential role in determining seed weight during soybean domestication[36].

GmSW17.1 (also named *GmSW17*) is another locus on chromosome 17 that is associated with 100-seed weight. It was identified using GWAS and joint linkage mapping[37]. Knockout of *GmSW17.1* significantly reduced 100-seed weight in 'Williams 82'. *GmSW17.1* encodes a ubiquitin-specific protease that regulates 100-seed weight by interacting with a deubiquitinating module component, GmSGF11,

in plant cell nuclei[37,38]. Two natural allelic variants of *GmSW17.1* (*GmSW17.1*^{T} and *GmSW17.1*^{C}) result in significantly different 100-seed weights, with *GmSW17.1*^{T} being more frequent in cultivated soybean, aligning with the trend of increasing 100-seed weight during domestication[37].

Association mapping showed that *qHSW18-1* (*Hundred-Seed Weight 18-1*) on chromosome 18 was associated with 100-seed weight, with *Glyma.18G242400* emerging as the most promising candidate gene for regulating 100-seed weight in soybean[39]. Haplotype analysis indicated that a superior haplotype of *qHSW18-1* underwent intense selection during domestication and subsequent cultivar development[39]. On chromosome 19, *Soybean Seed Size 1* (*SSS1*), which regulates seed weight, was cloned using MutMap analysis using an F_2 segregating population, derived from crossing 'Zhong-pin 661' (Zp661) with an *sss1* mutant (following EMS mutagenesis of Zp661)[40,41]. Haplotype evolution and molecular function analysis showed that an amino acid variant (Glu-to-Gln) at position 182 in one of its haplotypes, is characteristic of elite alleles selected during soybean domestication[40].

A recent GWAS study by Yang et al.[42] identified *GsMLP328* (*Major Latex Protein 328*), a gene also associated with 100-seed weight, localized on chromosome 20, using a population of 236 wild soybean accessions[42]. *GsMLP328* encodes an MLP that interacts with Polygalacturonase-Inhibiting Protein 4 (GmPGIP4), to coordinately regulate seed weight and associated traits, including seed dimensions, protein, and oil content. Two superior haplotypes, Hap_2 and Hap_3, exhibited significantly higher 100-seed weights and seed weights per plant compared to Hap_1. Hap_3 showed a predominant distribution in Asia, especially the Huang-Huai-Hai region of southeast China. However, this haplotype may be in the nascent stage of selection, as its proportion increased only slightly in landrace and cultivated accessions, relative to wild accessions. Moreover, the *GmPGIP4* is located in a selection region, and the distribution ratio of its elite haplotype (Hap_2) has been continuously expanding from wild to landrace and cultivated accessions[42].

Transcriptome analysis of differentially expressed genes (DEGs) between wild and cultivated soybeans, combined with gene coexpression network analysis (WGCNA), offers a valuable perspective for exploring domestication genes. Through this approach, *GA20OX* and *NFYA* (*Glyma02g47380*), which are associated with seed weight and oil content, were found to be significantly more highly expressed in cultivated soybeans compared to their wild counterparts[43]. Utilizing gene coexpression network analysis, *JASMONATE-ZIM DOMAIN 3* (*GmJAZ3*) was identified as a key regulator of soybean seed weight and size that promotes cell proliferation[44]. Haplotype analysis of the *JAZ3* promoters revealed that Hap3 of *JAZ3* underwent selection and fixation during domestication[44]. Using DEGs from RNA-seq data and then mapping these DEGs with QTLs associated with seed size, a candidate regulatory gene, *GmWRKY15a*, was identified. *GmWRKY15a* in *G. max* is significantly more highly expressed than *GsWRKY15a* in *G. soja*, and its expression level correlates with seed weight[45]. *GmPLATZ*, which encodes a Plant AT-rich sequence and zinc-binding (PLATZ) family protein, was identified through the analysis of seed transcriptomes from 45 cultivated soybean varieties and developing seeds at seven stages[46]. As a PLATZ-type transcription factor, GmPLATZ can directly bind to the promoters of cyclin genes and *GmGA20OX* to regulate their expression and promote cell proliferation[46]. Hap3 of *GmPLATZ* is an elite allele with elevated promoter activity that is associated with increased seed weight[46].

De novo genome assemblies of nine legume species and pangenome analysis of the genetic variations that impacted seed weight revealed convergent selection in hundreds of genes during legume evolution[10]. Among these, a *YUCCA* family gene, *GmYUCCA4a,* located on soybean chromosome 14, was identified as a potential target of convergent selection for increased seed weight in soybean, pigeon pea [*Cajanus cajan* (L.) Millspaugh], chickpea (*Cicer arietinum* L.), and pea (*Pisum sativum* L.). The Hap1 haplotype of *GmYUCCA4a* exhibited lower seed weights than Hap2 or Hap3, with a frequency of 94.5% in wild soybeans that declined to 4.3% in landraces, and 0.4% in cultivars. Through integrating GWAS, eQTL analysis, and TWAS, a candidate causal gene for regulating seed weight and oil content, *Regulator of Weight and Oil of Seed 1* (*GmRWOS1*), was identified[21]. *GmRWOS1* encodes a K^+-stimulated pyrophosphate-energized sodium pump that regulates seed weight and oil content. Allelic variation analysis revealed that *GmRWOS1* was strongly selected during soybean domestication[21].

Seed oil

Regulation of oil content

Due to the high global demand for edible vegetable oil, increasing seed oil content is a crucial breeding goal for soybeans. Soybean breeding has more than doubled the oil content in soybean seeds, with cultivated soybean seeds typically containing about 18%–22% oil, whereas wild soybean seeds have only about 8%–10%[47]. The oil content in soybean seeds is a complex trait regulated by multiple genes. More than 300 QTLs related to oil content have been mapped across all 20 soybean chromosomes (https://soybase.org)[48].

qOil-5-1, a stable QTL associated with soybean oil content, was identified at the end of chromosome 5 using RILs from the cultivars Huachun 2 and Wayao[49]. A combination of fine mapping and GWAS confirmed that *GmMFT* (also known as *GmST05*), is the causal gene for this QTL[50]. GmMFT belongs to the PEBP family, and the loss of function of GmMFT causes downregulation of FA biosynthesis, and decreased expression of *SWEET* genes[50]. Haplotype analysis has revealed that the high seed oil content haplotype of *GmMFT* was selected during modern soybean breeding[50].

By comparing the transcriptomes of developing seeds from cultivated and wild soybeans, the domestication gene *GmZF351*, which is associated with seed oil accumulation, has been identified on chromosome 6. *GmZF351* is located within the known QTL *Seed Oil Plus Protein-2*[43,51]. This gene encodes a zinc-finger protein that positively regulates lipid biosynthesis by enhancing the activity of WRINKLED1 (WRI1)[51]. WRI1 is a transcription factor in the APETALA2/ethylene-responsive element-binding protein family that is considered to be a master regulator of seed oil accumulation. It interacts with multiple oil-regulated proteins and enhances the expression of various genes involved in glycolysis and FA biosynthesis[52,53].

Additionally, another tandem Cys_3-His (CCCH) zinc-finger protein GmZF392 functions as a positive regulator of lipid accumulation by physically interacting with GmZF351[54]. Domestication analysis of the promoter sequence of *GmZF392* indicates that it has undergone selection from wild to cultivated soybeans[54]. Both *GmZF392* and *GmZF351* are upregulated by GmNFYA, a transcription factor associated with oil content accumulation that is expressed at significantly higher levels in cultivated soybeans compared to wild ones[43,54]. An AP2/ERF-type transcription factor (GmERFA) can inhibit the transcriptional activity of *GmNFYA* via a physical interaction[55]. There are three haplotypes of the *GmERFA* promoter (Hap1-3), of which Hap3 is the major haplotype in cultivated soybeans and is associated with a lower *GmERFA* expression level, higher oil content, and lower

protein content. This indicates that *GmERFA* may have undergone selection during domestication[55].

Association analysis showed that the sugar transporter gene *GmSWEET10a* is closely linked to seed oil and protein QTLs[18,19,56,57]. *GmSWEET10a* is highly expressed in soybean seeds[18]. Variations in its promoter and coding regions results in differing oil contents in the seeds of RILs, and the superior alleles have been selectively transferred from *G. soja* to *G. max* by breeding[18,19,57]. Additionally, *GmSWEET10b*, which is functionally redundant with *GmSWEET10a*, is also undergoing selection in current breeding programs[57,58].

Important QTLs affecting oil content in soybean seeds have also been mapped on chromosome 20 in a region including *POWR1* (*protein, oil, weight, regulator 1*), which encodes a CONSTANS, CO-like, and TOC1 (CCT)-domain protein, playing a major role in oil and protein production[5]. The favorable allele has a transposable element (TE) insertion that truncates the protein's CCT domain, significantly increasing seed oil content, weight, and yield[5]. *POWR1* is considered to be a domestication gene, and the TE insertion in *POWR1* underwent artificial selection during soybean domestication, resulting in higher seed oil content and weight in cultivated soybeans compared to wild-type soybeans[5]. Through GWAS analysis, an environmentally stable QTL, *GqOil20*, that is associated with oil content was also identified on chromosome 20. Molecular assays have demonstrated that this QTL contains *GmOLEO1*, which encodes an oleosin that contributes to oil accumulation in soybean seeds by stabilizing oil bodies within cells[59]. Further analysis revealed that the promoter region of *GmOLEO1* is located within artificial selection sites, leading to its higher expression in cultivated soybeans compared to wild soybeans, thereby increasing oil accumulation in seeds of cultivated soybeans[59].

Improvement of oil composition

In recent years, an increasing number of studies have focused on the relationship between domestication and seed oil components. Soybean oil primarily consists of five types of fatty acids (FAs): approximately 10% palmitic acid (16:0), 4% stearic acid (18:0), 18% oleic acid (18:1, ω-9), 55% linoleic acid (18:2, ω-6), and 13% α-linolenic acid (18:3, ω-3), with 14% being saturated and 86% unsaturated FAs[48]. Unsaturated FAs are known to have beneficial effects on human health. One important unsaturated FA in determining soybean oil quality is ω-3. On one hand, higher levels can reduce oxidative stability[48]. On the other hand, the World Health Organization and numerous studies recommend that reducing the ω-6/ω-3 ratio to less than 4:1 in modern diets is beneficial to human health[60]. Notably, oil from cultivated soybeans generally has a ω-6 to ω-3 ratio of 6:1 to 7:1[61]. In contrast, wild soybeans have nearly double the ω-3 concentration, and a much lower ω-6/ω-3 ratio of 4:1 compared to cultivated soybeans[61]. Additionally, wild soybeans possess different QTLs controlling ω-3 levels than those found in cultivated soybeans[62]. This suggests that the higher ω-6 to ω-3 FA ratio in cultivated soybeans was selected during domestication. The high ω-3 concentration trait in wild soybeans is governed by a set of FA desaturase (FAD) alleles involved in ω-3 biosynthesis, including *FAD3*, *FAD7*, and *FAD8*[48,62–64]. However, none of these genes were preferentially selected during domestication, despite exhibiting a significantly lower missense mutation frequency than the genome-wide average in domesticated populations[65].

In the soybean genome, over 228 QTLs associated with FA metabolism have been identified across all 20 chromosomes (www.soybase.org)[30,66]. Recently, a few causal genes within these QTL loci have also been identified. One such gene, *GmFATA1B*, which encodes an acyl-ACP thioesterase, was pinpointed as a causal gene in the seed oil-associated locus *qOil-8-1*[49,67]. Heterologous overexpression of *GmFATA1B* in Arabidopsis led to an increase in oleic acid, and its derivative, linoleic acid, of approximately 27% and 53%, respectively[67]. Polymorphism spectrum analysis revealed that *GmFATA1B* has undergone strong purifying selection, resulting in a very low frequency of deleterious alleles at this locus[67]. Another gene, *GmSEIPIN1A*, located within a meta-QTL region for soybean oil content on chromosome 9, known as *Fatty Acid 9* (*FA9*), was identified as a significant contributor to seed FA content[68]. Loss of FA9 function decreases linoleic acid and oil contents, while significantly increasing oleic acid content, indicating that FA9 is a key regulator of the seed FA profile. Haplotype 2 of FA9 is absent in wild soybeans, but present in 13% of landraces and 26% of cultivars, suggesting it may have been selected during post-domestication improvement of soybeans[68].

Improvement of tocopherol content

Tocopherols, commonly known as vitamin E, play a crucial role in human health due to their unique antioxidant properties. These compounds are composed of four types of fat-soluble molecules: α-, β-, γ-, and δ-tocopherol[69]. Among these, α-tocopherol is the primary contributor to the antioxidant activity of tocopherols, and it offers significant benefits to human cells[70]. Tocopherols reduce lipid peroxidation, thereby enhancing the storage stability of soybean oil[71]. As lipophilic antioxidants, tocopherols are present in all plant species, with soybean seed oil containing a higher tocopherol content than most other oilseed crops[72]. In soybean oil, γ-tocopherol is the most abundant (60%–66% of total tocopherol), followed by δ-tocopherol (24%–29%), α-tocopherol (4%–10%), and β-tocopherol (less than 3%)[73]. Due to their importance in oil quality, understanding the genetic basis of tocopherol content has become a significant focus of recent research.

QTLs related to tocopherol content have been identified on chromosomes 5, 9, 11, and 12 using methods such as mGWAS and BSA-Seq[74–77]. On chromosome 9, the promoter of *γ-TMT3*, which encodes a tocopherol methyltransferase from a high α-tocopherol cultivar (Keszthelyi Aproszemu Sarga), had higher activity than the promoter from a low α-tocopherol cultivar (Ichihime), likely due to single-nucleotide polymorphisms (SNPs) in *cis*-regulatory elements[77]. A molecular genetic study of wild soybean accessions with high α-tocopherol levels suggested that novel *γ-TMT3* promoter haplotypes could enhance the genetic diversity of α-tocopherol biosynthesis in soybeans[78]. Analysis of a soybean population of over 800 accessions showed that the tocopherol content increased during domestication, with a strong positive correlation observed between tocopherol and oil content[76].

Further analysis revealed that the FA regulatory transcription factor GmZF351 activated the expression of several tocopherol pathway genes, thereby increasing both FA and tocopherol contents in soybean seeds[76]. On chromosome 5, SV (structural variation)-GWAS analysis identified *GmMQT* (*Multiple Quality Traits*, also known as *GmST05*) as affecting important seed traits, including tocopherol and oil content[79]. Twelve structural variations in *GmMQT* in the 547 accessions examined defined two haplotypes, Hap-Ref and Hap-Alt. Hap-Ref exhibited significantly higher oil content (19.05%) than did Hap-Alt (17.11%), with similar trends for γ-tocopherol[79]. These findings demonstrate that tocopherol content was domesticated along with the domestication of fatty acids.

Seed protein

Regulation of protein content

In addition to being a vital oil source, soybeans are a significant source of high-quality protein for both human and animal consumption, contributing over 25% of the global protein supply[80]. Soybean protein is particularly nutritious as it contains all the essential amino acids required by humans. Seeds of cultivated soybean varieties contain 35%–40% protein, making them among the most abundant plant-based protein sources. In contrast, wild soybean seeds typically have about 50% protein[81]. During domestication, an increase in seed oil content was accompanied by a decrease in seed protein content due to a negative genetic correlation[48]. This suggests that the artificial selection of soybeans throughout domestication primarily targeted oil content and yield[82]. In modern soybean breeding, both seed oil and protein content are key objectives. Therefore, understanding the genetic mechanisms that determine soybean oil and protein content, and any potential tradeoffs, is crucial for breeding efforts.

To date, 241 QTLs related to protein content in soybean have been identified, with 16 confirmed (www.soybase.org). Two major QTLs for protein and oil content were discovered on chromosomes 15 and 20 using RFLP markers in an F_2 population derived from a cross between *G. max* (A81-356022), and *G. soja* (PI 468916)[83]. *POWR1*, another key domestication-related gene associated with soybean protein content, is a pleiotropic regulator that influences both seed weight and oil content[5]. Besides *POWR1*, several domestication genes involved in soybean seed development have been identified. These genes have pleiotropic effects on seed size, oil, and protein content, often negatively regulating protein content. Notable among them are *GmSWEET10a*/*GmSWEET39*[18,68], *GmMFT*[50], *GmOLEO1*[59], and *GmJAZ3*[44]. A recent study integrating GWAS/TWAS has identified *GmSop20* (*Glyma.20G005900*) on chromosome 20 as a key regulator of the oil-to-protein ratio[84]. The domestication-selected allele $GmSop20^C$ has undergone intense artificial selection, as revealed by genetic diversity analysis, and is prevalent in cultivars across northern China and the USA[84]. Functional analysis showed that GmSop20 directly activates the expression of *GmSWEET10a*. This synergizes two artificially selected loci within a unified regulatory network, enhancing sugar distribution from the seed coat to the embryo, and thereby increasing oil accumulation[84]. Although the precise metabolic connections between oil and storage protein synthesis remain unclear, the carbon flux in developing soybean seeds is mainly divided between protein and oil[69,84,85]. Understanding any shifts in the carbon flux favoring oil or protein will help improve soybean varieties.

Improvement of protein quality

To enhance the quality of soybean seeds, it is crucial not only to increase protein content, but also to improve the efficiency of soybean protein utilization, and optimize seed protein quality. The primary storage proteins in soybeans are 7S and 11S globulins, which constitute 70%–80% of the total seed protein, and greatly influence soybean quality[86]. The 11S globulin is abundant in sulfur-containing amino acids essential for human nutrition, and the 7S globulin is a major allergen often responsible for allergies in humans and animals[87]. Consequently, strategies to enhance soybean seed quality focus on reducing 7S globulin levels and increasing 11S globulin levels. *GmCG-1* encodes the α' subunit of β-conglycinin. Suppressing expression of GmCG-1 and its paralogues GmCG-2 and GmCG-3 decreases β-conglycinin content while also increasing the 11S/7S ratio, total protein content, and sulfur-containing amino acid content, thereby enhancing the nutritional profile of soybean seeds. Population evolution analysis indicates that GmCG-1, GmCG-2, and GmCG-3 were selected during soybean domestication, and the frequency of the superior Hap1 haplotype of GmCG-1 has gradually declined[88].

A recent study has demonstrated that *High Seed Storage Protein* (*HSSP1*), a B3 domain protein increases seed storage protein content by directly binding to the *cis*-acting element of GmCG1, upregulating its expression[89]. Selecting for larger seed size and higher oil content in soybeans may result in slightly less desirable seed protein traits. Another study demonstrated that *Glycine max gibberellic acid-stimulated Arabidopsis 12* (*GmGASA12*), which encodes a gibberellin-regulated protein, cooperatively regulates the biosynthesis of β-conglycinin and glycinin by interacting with GmCG6[90]. Evolutionary analyses have shown that elite *GmGASA12* haplotypes were strongly favored during domestication, with 94% of cultivars possessing beneficial alleles[90].

Challenges and perspectives

Because several domesticated traits of soybean seeds are influenced by multiple genes[16], comparative genomics analysis related to the seed's morphological index is a promising approach for identifying domestication genes[10]. With the advancement of sequencing technology and the growing availability of population data, numerous QTLs controlling seed traits in soybeans have been identified within the domestication region of the soybean genome[16]. However, only a few key genes have been isolated and functionally validated.

Integrating phenomics and multi-omics to uncover domestication genes in soybean seeds

A primary barrier to gene isolation and characterization is that most seed traits are significantly influenced by the growing environment. This makes it challenging to collect reproducible seed phenotypic data from various geographic areas and time periods, hindering the fine mapping of target traits. However, recent advances in seed phenomics, such as high-throughput phenotyping analysis systems, have greatly enhanced the accuracy and reliability of seed trait measurements. The high-throughput of these systems also facilitates the analysis of more crosses and replicates, in more diverse environments[91,92]. Various high-throughput image analysis and spectroscopic tools have been developed for seed phenology[93]. These systems also provide precise and detailed seed phenotypic information for individual lines and large mapping populations, potentially identifying correlations between genomic variation and phenotypic information in seed traits. Notably, the development of multispectral imaging, which offers a nondestructive physical technique, is a valuable tool for evaluating seed quality traits such as vigor, moisture, and germination status[94–96]. These advances are likely to reveal additional potential domestication genes and identify domesticated seed traits.

Genomic approaches have provided valuable insights into the domestication of seed traits in soybeans, and the advent of -omics technologies have transformed traditional methods. These technologies integrate information from various biological levels, including transcriptomics, proteomics, metabolomics, epigenomics, and even single-cell and tissue spatial transcriptomics[21,97]. The multidimensional and dynamic nature of seed -omics approaches

has not only accelerated the identification of domesticated genes, but also opened up new opportunities for crop improvement. For instance, three-dimensional genome architecture (Hi-C) epigenomic technology, when combined with GWAS and long-read sequencing, has been used to construct a high-quality chromosome-scale reference genome for oilseed Camellia (*Camellia oleifera*)[98,99]. This research demonstrated that the artificial selection of elite alleles involved in oil biosynthesis significantly contributed to the domestication of oilseed Camellia[98].

Furthermore, a spatially resolved single-cell atlas of gene expression and chromatin accessibility in developing soybean seeds has led to the identification of several genes specifically expressed in the endosperm or embryo, including *SWEET10a*, a domesticated gene that regulates soybean seed size and oil content[100]. In the future, combining -omics technologies to deepen our understanding of seed trait domestication throughout seed development will become increasingly valuable. These innovative -omics technologies have significantly enhanced our comprehension of genotype–phenotype associations during seed domestication, and have provided information that is essential for plant breeding, particularly in modern breeding efforts.

Bottlenecks in exploring key domesticated genes of soybean

There are several bottlenecks in exploring key domesticated genes in soybeans during current research. First, most domesticated traits are controlled by multiple genes, which form different pathways, networks, or modules, that collectively regulate one or more traits. For example, GmZF392 physically interacts with GmZF351 to synergistically promote the expression of downstream genes. Furthermore, both GmZF392 and GmZF351 are upregulated by GmNFYA[54]. This complexity limits our ability to identify the central genes controlling specific domesticated traits. Second, many pleiotropic genes have been confirmed to regulate multiple traits during soybean domestication. For instance, *GmST05* can regulate seed oil and protein content by affecting *GmSWEET10a* expression, in addition to influencing seed thickness[17,50]. This complexity complicates phenotypic screening and results in difficult trade-offs regarding which traits are beneficial for breeding objectives during the breeding process. Third, some traits exhibit negative coupling that cannot be resolved through traditional breeding approaches during domestication, such as the relationship between soybean protein and oil content. Breeders and early farmers likely shifted the balance of pleiotropic gene expression through selection pressures to mitigate this negative correlation.

Moreover, due to directional selection at specific genomic regions underlying agronomically important traits, the genome-wide genetic diversity of domesticated crops has been reduced[101]. Approximately half of the genetic diversity of soybean has been lost during domestication, resulting in a genetic bottleneck in current soybean breeding[7]. Meanwhile, deleterious mutations have accumulated during the domestication process, increasing the costs associated with soybean domestication[102]. Therefore, exploring genetic resources from wild soybean varieties, and utilizing modern biotechnologies such as genome editing and mutagenesis techniques to create more desirable allelic variations, could help overcome this genetic bottleneck and reduce the costs of future soybean breeding.

Progress and applications of gene editing technology in soybean improvement

In recent years, gene editing technology has rapidly advanced in the field of plant cultivation. The development of CRISPR/Cas9 technology has significantly expanded its range of applications. For instance, the introduction of the Flanking Nicks Prime Editor (FLICK-PE) system facilitates precise genome modification in soybeans, enhancing editing efficiency by over 21.1%[103]. Moreover, efficient genetic transformation methods have gradually been established. Recently, the Cut-Dip-Budding (CDB) gene delivery system, along with developmental regulators (DRs) such as *WUSCHEL2* (*WUS2*) and the gene encoding isopentenyltransferase (IPT), has significantly accelerated soybean tissue culture, reducing the transformation process from months to weeks[104,105]. These advances not only have the potential to accelerate the functional analysis and breeding improvement of soybean domestication genes, but have also been demonstrated to be effective strategies for uncoupling the negative correlation between protein and oil content. For example, mutation in nodule inception gene *Rhizobially Induced Cle1a/2a* (*RIC1a/2a*), achieved through a multiplexed CRISPR–Cas9 mutagenesis system, lead to a significant increase in seed protein content without a concomitant decrease in oil content[106]. The soybean invertase inhibitor *GmCIF1*, which has been shown to reduce protein content without affecting oil content, may serve as a viable strategy for dissociating the negative correlation between protein and oil content when knocked out using CRISPR-Cas9 technology[107]. Furthermore, gene editing combined with AlphaFold prediction can create novel genotypes that do not exist in natural populations, overcoming the limitations of natural variation and addressing genetic bottlenecks[108].

In conclusion, with advancements in agricultural technologies and cropping systems, the focus of breeding efforts has evolved. Breeding objectives now must not only encompass improved yield and seed quality, but also align with current agricultural practices, such as selecting traits suitable for machine harvesting. Consequently, new technologies are swiftly transforming the domestication patterns of crop seeds. In the next stage of seed domestication, there is a pressing need to develop more diverse food systems through genomic editing-based breeding, or introgression breeding with multiple breeding populations. Therefore, we have summarized the key domestication genes regulating major seed traits in soybean (Fig. 1, Table 1). These genes can serve as elite targets for overcoming the bottleneck of low yield and quality in soybean by integrating them into modern breeding.

Author contributions

The authors confirm their contributions to the paper as follows: study conception and design: Chen B, Zhang D; data collection: Chen B, Wang C; draft manuscript preparation: Chen B, Wang C, Zhuang Y, Li X, Zhang J, Zhang D; funding: Chen B, Zhuang Y, Li X, Zhang D. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.

Acknowledgments

This work was supported by the National key R&D program of China (2024YFF1000502), the National Natural Science Foundation of China (32322062, 32441057), the Biological Breeding-National Science and Technology Major Project (2024ZD0407802), and the Natural Science Foundation of Shandong Province (ZR2023JQ009).

Dates

Received 5 November 2025; Revised 21 December 2025; Accepted 12 January 2026; Published online 27 March 2026

References

[1] Fuller DQ, Qin L, Zheng Y, Zhao Z, Chen X, et al. 2009. The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* 323:1607–1610

[2] Anand A, Subramanian M, Kar D. 2023. Breeding techniques to dispense higher genetic gains. *Frontiers in Plant Science* 13:1076094

[3] Huang X, Zhao P, Peng X, Sun MX. 2023. Seed development in Arabidopsis: what we have learnt in the past 30 years. *Seed Biology* 2:6

[4] Liu K. 1997. Chemistry and nutritional value of soybean components. In *Soybeans*. Boston, MA: Springer. pp. 25–113 doi: 10.1007/978-1-4615-1763-4_2

[5] Goettel W, Zhang H, Li Y, Qiao Z, Jiang H, et al. 2022. *POWR1* is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nature Communications* 13:3051

[6] Sedivy EJ, Wu F, Hanzawa Y. 2017. Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytologist* 214:539–553

[7] Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, et al. 2015. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* 33:408–414

[8] Zhuang Y, Wang X, Li X, Hu J, Fan L, et al. 2022. Phylogenomics of the genus *Glycine* sheds light on polyploid evolution and life-strategy transition. *Nature Plants* 8:233–244

[9] Zhuang Y, Li X, Hu J, Xu R, Zhang D. 2022. Expanding the gene pool for soybean improvement with its wild relatives. *aBIOTECH* 3:115–125

[10] Wang L, Jiang X, Jiao W, Mao J, Ye W, et al. 2025. Pangenome analysis provides insights into legume evolution and breeding. *Nature Genetics* 57:2052–2061

[11] Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183

[12] Kim MY, Lee S, Van K, Kim TH, Jeong SC, et al. 2010. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proceedings of the National Academy of Sciences of the United States of America* 107:22032–22037

[13] Li YH, Zhou G, Ma J, Jiang W, Jin LG, et al. 2014. *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology* 32:1045–1052

[14] Kim MS, Lozano R, Kim JH, Bae DN, Kim ST, et al. 2021. The patterns of deleterious mutations during the domestication of soybean. *Nature Communications* 12:97

[15] Han Y, Zhao X, Liu D, Li Y, Lightfoot DA, et al. 2016. Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytologist* 209:871–884

[16] Zuo JF, Ikram M, Liu JY, Han CY, Niu Y, et al. 2022. Domestication and improvement genes reveal the differences of seed size- and oil-related traits in soybean domestication and improvement. *Computational and Structural Biotechnology Journal* 20:2951–2964

[17] Duan Z, Zhang M, Zhang Z, Liang S, Fan L, et al. 2022. Natural allelic variation of *GmST05* controlling seed size and quality in soybean. *Plant Biotechnology Journal* 20:1807–1818

[18] Miao L, Yang S, Zhang K, He J, Wu C, et al. 2020. Natural variation and selection in *GmSWEET39* affect soybean seed oil content. *New Phytologist* 225:1651–1666

[19] Zhang H, Goettel W, Song Q, Jiang H, Hu Z, et al. 2020. Selection of *GmSWEET39* for oil and protein improvement in soybean. *PLoS Genetics* 16:e1009114

[20] Liang S, Duan Z, He X, Yang X, Yuan Y, Liang Q, et al. 2024. Natural variation in *GmSW17* controls seed size in soybean. *Nature Communications* 15:7417

[21] Yuan X, Jiang X, Zhang M, Wang L, Jiao W, et al. 2024. Integrative omics analysis elucidates the genetic basis underlying seed weight and oil content in soybean. *The Plant Cell* 36:2160–2175

[22] Xu G, Zhang X. 2023. Mechanisms controlling seed size by early endosperm development. *Seed Biology* 2:1

[23] Chen Y, Nelson RL. 2004. Genetic variation and relationships among cultivated, wild, and semiwild soybean. *Crop Science* 44:316–325

[24] Xing Y, Zhang Q. 2010. Genetic and molecular bases of rice yield. *Annual Review of Plant Biology* 61:421–442

[25] Cui B, Chen L, Yang Y, Liao H. 2020. Genetic analysis and map-based delimitation of a major locus qSS3 for seed size in soybean. *Plant Breeding* 139:1145–1157

[26] Xie FT, Niu Y, Zhang J, Bu SH, Zhang HZ, et al. 2014. Fine mapping of quantitative trait loci for seed size traits in soybean. *Molecular Breeding* 34:2165–2178

[27] Xu Y, Li HN, Li GJ, Wang X, Cheng LG, et al. 2011. Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics* 122:581–594

[28] Kumawat G, Xu D. 2021. A major and stable quantitative trait locus *qSS2* for seed size and shape traits in a soybean RIL population. *Frontiers in Genetics* 12:646102

[29] Li J, Zhang Y, Ma R, Huang W, Hou J, et al. 2022. Identification of *ST1* reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnology Journal* 20(6):1110–1121

[30] Zhang J, Wang X, Lu Y, Bhusal SJ, Song Q, et al. 2018. Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Molecular Plant* 11(3):460–472

[31] Yan L, Di R, Wu C, Liu Q, Wei Y, et al. 2019. Haplotype analysis of a major and stable QTL underlying soybean (*Glycine max*) seed oil content reveals footprint of artificial selection. *Molecular Breeding* 39:57

[32] Wang Z, Zhou Z, Liu Y, Liu T, Li Q, et al. 2015. Functional evolution of phosphatidylethanolamine binding proteins in soybean and *Arabidopsis*. *The Plant Cell* 27(2):323–336

[33] Zhang C, Li W, Tan C, Huang M, Wu H, et al. 2025. Natural allelic variation in *SW14* determines seed weight and quality in soybean. *Nature Communications* 16(1):8070

[34] Wang W, He Q, Yang H, Xiang S, Xing G, et al. 2014. Identification of QTL/segments related to seed-quality traits in *G. soja* using chromosome segment substitution lines. *Plant Genetic Resources* 12(S1):S65–S69

[35] Chen X, Liu C, Guo P, Hao X, Pan Y, et al. 2023. Differential *SW16.1* allelic effects and genetic backgrounds contributed to increased seed weight after soybean domestication. *Journal of Integrative Plant Biology* 65(7):1734–1752

[36] Lu X, Xiong Q, Cheng T, Li QT, Liu XL, et al. 2017. A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Molecular Plant* 10(5):670–684

[37] Zhang H, Yang L, Guo S, Tian Y, Yang C, et al. 2024. A natural allelic variant of *GmSW17.1* confers high 100-seed weight in soybean. *The Crop Journal* 12(6):1709–1717

[38] Pfab A, Bruckmann A, Nazet J, Merkl R, Grasser KD. 2018. The adaptor protein ENY2 is a component of the deubiquitination module of the *Arabidopsis* SAGA transcriptional co-activator complex but not of the TREX-2 complex. *Journal of Molecular Biology* 430(10):1479–1494

[39] Zhang Y, Yang X, Bhat JA, Zhang Y, Bu M, et al. 2024. Identification of superior haplotypes and candidate gene for seed size-related traits in soybean (*Glycine max* L.). *Molecular Breeding* 45(1):3

[40] Zhu W, Yang C, Yong B, Wang Y, Li B, et al. 2022. An enhancing effect attributed to a nonsynonymous mutation in *SOYBEAN SEED SIZE 1*, a *SPINDLY*-like gene, is exploited in soybean domestication and improvement. *New Phytologist* 236(4):1375–1392

[41] Li Z, Jiang L, Ma Y, Wei Z, Hong H, et al. 2017. Development and utilization of a new chemically - induced soybean library with a high mutation density. *Journal of Integrative Plant Biology* 59(1):60–74

[42] Yang Z, Lu S, Li W, Wang Z, Hu D, et al. 2025. A major latex protein, GsMLP328, modulates seed traits in soybean. *Journal of Integrative Plant Biology* 67(11):2790–2792

[43] Lu X, Li QT, Xiong Q, Li W, Bi YD, et al. 2016. The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *The Plant Journal* 86(6):530−544

[44] Hu Y, Liu Y, Tao JJ, Lu L, Jiang ZH, et al. 2023. GmJAZ3 interacts with GmRR18a and GmMYC2a to regulate seed traits in soybean. *Journal of Integrative Plant Biology* 65:1983−2000

[45] Gu Y, Li W, Jiang H, Wang Y, Gao H, et al. 2017. Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *Journal of Experimental Botany* 68(11):2717−2729

[46] Hu Y, Liu Y, Lu L, Tao JJ, Cheng T, et al. 2023. Global analysis of seed transcriptomes reveals a novel PLATZ regulator for seed size and weight control in soybean. *New Phytologist* 240(6):2436−2454

[47] Patil G, Vuong TD, Kale S, Valliyodan B, Deshmukh R, et al. 2018. Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal* 16(11):1939−1953

[48] Yao Y, You Q, Duan G, Ren J, Chu S, et al. 2020. Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. *BMC Plant Biology* 20(1):51

[49] Huang J, Ma Q, Cai Z, Xia Q, Li S, et al. 2020. Identification and mapping of stable QTLs for seed oil and protein content in soybean [*Glycine max* (L.) Merr.]. *Journal of Agricultural and Food Chemistry* 68(23):6448−6460

[50] Cai Z, Xian P, Cheng Y, Zhong Y, Yang Y, et al. 2023. MOTHER-OF-FT-AND-TFL1 regulates the seed oil and protein content in soybean. *New Phytologist* 239(3):905−919

[51] Li QT, Lu X, Song QX, Chen HW, Wei W, et al. 2017. Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. *Plant Physiology* 173(4):2208−2224

[52] Andre C, Froehlich JE, Moll MR, Benning C. 2007. A heteromeric plastidic pyruvate kinase complex involved in seed oil biosynthesis in *Arabidopsis*. *The Plant Cell* 19(6):2006−2022

[53] Wei W, Wang LF, Tao JJ, Zhang WK, Chen SY, et al. 2025. The comprehensive regulatory network in seed oil biosynthesis. *Journal of Integrative Plant Biology* 67(3):649−668

[54] Lu L, Wei W, Li QT, Bian XH, Lu X, et al. 2021. A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytologist* 231(2):661−678

[55] Liu Y, Hu Y, Wei JJ, Jiang ZH, Han JQ, et al. 2025. Transcription factor GmERFA interacts with GmNFYA and acts as a negative regulator of seed fatty acid accumulation in soybean. *Plant Biotechnology Journal* 23(12):5917−5933

[56] Yang H, Wang W, He Q, Xiang S, Tian D, et al. 2019. Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. *Theoretical and Applied Genetics* 132:2793−2807

[57] Wang S, Liu S, Wang J, Yokosho K, Zhou B, et al. 2020. Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication. *National Science Review* 7(11):1776−1786

[58] Sun J, Li W, Wei X, Shou H, Tran LP, et al. 2025. Mechanistic roles of *GmSWEET10a/b* and *GmSUT1* in the oil–protein balance in soybean mature seeds at transcriptional and metabolic levels. *The Plant Journal* 123(4):e70435

[59] Zhang D, Zhang H, Hu Z, Chu S, Yu K, et al. 2019. Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. *PLoS Genetics* 15:e1008267

[60] Mariamenatu AH, Abdu EM. 2021. Overconsumption of omega-6 polyunsaturated fatty acids (PUFAs) versus deficiency of omega-3 PUFAs in modern-day diets: the disturbing factor for their "balanced antagonistic metabolic functions" in the human body. *Journal of Lipids* 2021(1):8848161

[61] Asekova S, Chae JH, Ha BK, Dhakal KH, Chung G, et al. 2014. Stability of elevated α-linolenic acid derived from wild soybean (*Glycine soja* Sieb. & Zucc.) across environments. *Euphytica* 195:409−418

[62] Ha BK, Kim HJ, Velusamy V, Vuong TD, Nguyen HT, et al. 2014. Identification of quantitative trait loci controlling linolenic acid concentration in PI483463 (*Glycine soja*). *Theoretical and Applied Genetics* 127(7):1501−1512

[63] Pantalone VR, Rebetzke GJ, Burton JW, Wilson RF. 1997. Genetic regulation of linolenic acid concentration in wild soybean *Glycine soja* accessions. *Journal of the American Oil Chemists' Society* 74:159−163

[64] Gishini MFS, Kachroo P, Hildebrand D. 2025. Fatty acid desaturase 3-mediated α-linolenic acid biosynthesis in plants. *Plant Physiology* 197(2):kiaf012

[65] Derbyshire MC, Marsh J, Tirnaz S, Nguyen HT, Batley J, et al. 2023. Diversity of fatty acid biosynthesis genes across the soybean pangenome. *The Plant Genome* 16(2):e20334

[66] Li B, Fan S, Yu F, Chen Y, Zhang S, et al. 2017. High-resolution mapping of QTL for fatty acid composition in soybean using specific-locus amplified fragment sequencing. *Theoretical and Applied Genetics* 130(7):1467−1479

[67] Cai Z, Xian P, Cheng Y, Yang Y, Zhang Y, et al. 2023. Natural variation of *GmFATA1B* regulates seed oil content and composition in soybean. *Journal of Integrative Plant Biology* 65(10):2368−2379

[68] Qi Z, Guo C, Li H, Qiu H, Li H, et al. 2024. Natural variation in *Fatty Acid 9* is a determinant of fatty acid and protein content. *Plant Biotechnology Journal* 22(3):759−773

[69] Boschin G, Arnoldi A. 2011. Legumes are valuable sources of tocopherols. *Food Chemistry* 127(3):1199−1203

[70] Azzi A. 2007. Molecular mechanism of α-tocopherol action. *Free Radical Biology and Medicine* 43(1):16−21

[71] Kamal-Eldin, A. 2006. Effect of fatty acids and tocopherols on the oxidative stability of vegetable oils. *European Journal of Lipid Science and Technology* 108(12):1051−1061

[72] Méjean M, Brunelle A, Touboul D. 2015. Quantification of tocopherols and tocotrienols in soybean oil by supercritical-fluid chromatography coupled to high-resolution mass spectrometry. *Analytical and Bioanalytical Chemistry* 407(17):5133−5142

[73] Carrera CS, Seguin P. 2016. Factors affecting tocopherol concentrations in soybean seeds. *Journal of Agricultural and Food Chemistry* 64(50):9465−9474

[74] Park C, Dwiyanti MS, Nagano AJ, Liu B, Yamada T, et al. 2019. Identification of quantitative trait loci for increased α-tocopherol biosynthesis in wild soybean using a high-density genetic map. *BMC Plant Biology* 19(1):510

[75] Ghosh S, Zhang S, Azam M, Agyenim-Boateng KG, Qi J, et al. 2022. Identification of genomic loci and candidate genes related to seed tocopherol content in soybean. *Plants* 11(13):1703

[76] Chu D, Zhang Z, Hu Y, Fang C, Xu X, et al. 2023. Genome-wide scan for oil quality reveals a coregulation mechanism of tocopherols and fatty acids in soybean seeds. *Plant Communications* 4(5):100598

[77] Dwiyanti MS, Yamada T, Sato M, Abe J, Kitamura K. 2011. Genetic variation of γ-tocopherol methyltransferase gene contributes to elevated α-tocopherol content in soybean seeds. *BMC Plant Biology* 11(1):152

[78] Dwiyanti MS, Maruyama S, Hirono M, Sato M, Park E, et al. 2016. Natural diversity of seed α-tocopherol ratio in wild soybean (*Glycine soja*) germplasm collection. *Breeding Science* 66(4):653−657

[79] Zhang C, Shao Z, Kong Y, Du H, Li W, et al. 2024. High-quality genome of a modern soybean cultivar and resequencing of 547 accessions provide insights into the role of structural variation. *Nature Genetics* 56:2247−2258

[80] Lu S, Fang C, Abe J, Kong F, Liu B. 2022. Current overview on the genetic basis of key genes involved in soybean domestication. *aBIOTECH* 3(2):126−139

[81] Kim WJ, Kang BH, Kang S, Shin S, Chowdhury S, et al. 2023. A genome-wide association study of protein, oil, and amino acid content in wild soybean (*Glycine soja*). *Plants* 12(8):1665

[82] Brummer EC, Graef GL, Orf J, Wilcox JR, Shoemaker RC. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Science* 37(2):370−378

[83] Diers BW, Keim P, Fehr WR, Shoemaker RC. 1992. RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics* 83:608−612

[84] Zheng H, Feng X, Wang L, Shao W, Guo S, et al. 2025. *GmSop20* functions as a key coordinator of the oil-to-protein ratio in soybean seeds. *Advanced Science* 12(38):e05181

[85] Clemente TE, Cahoon EB. 2009. Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiology* 151(3):1030−1040

[86] Krishnan HB. 2000. Biochemistry and molecular biology of soybean seed storage proteins. *Journal of New Seeds* 2(3):1−25

[87] Krishnan HB, Kim WS, Jang S, Kerley MS. 2009. All three subunits of soybean β-conglycinin are potential food allergens. *Journal of Agricultural and Food Chemistry* 57(3):938−943

[88] Yang R, Ma Y, Yang Z, Pu Y, Liu M, et al. 2024. Knockdown of β-conglycinin α' and α subunits alters seed protein composition and improves salt tolerance in soybean. *The Plant Journal* 120(4):1488−1507

[89] Tian H, Yin Y, Li X, Zhang Z, Feng S, et al. 2025. Identification of *HSSP1* as a regulator of soybean protein content through QTL analysis and Soy-SPCC network. *Plant Biotechnology Journal* 23(7):2673−2688

[90] Yang Y, Zhang L, Zuo H, Yang Y, Hu D, et al. 2025. GmGASA12 coordinates hormonal dynamics to enhance soybean water-soluble protein accumulation and seed size. *Journal of Integrative Plant Biology* 67(9):2401−2415

[91] Padhy AK, Singh A, Chaurasia S, Parida SK, Tripathi K, et al. 2025. Key determinants of seed size for enhancing genetic gain in legumes. *Plant, Cell & Environment* 00:Eary view

[92] Yu LA, Sussman H, Khmelnitsky O, Rahmati Ishka M, Srinivasan A, et al. 2024. Development of a mobile, high-throughput, and low-cost image-based plant growth phenotyping system. *Plant Physiology* 196(2):810−829

[93] Dwivedi SL, Spillane C, Lopez F, Ayele BT, Ortiz R. 2021. First the seed: Genomic advances in seed science for improved crop productivity and food security. *Crop Science* 61(3):1501−1526

[94] Shi R, Zhang H, Wang C, Zhou Y, Kang K, et al. 2025. Data fusion-driven hyperspectral imaging for non-destructive detection of single maize seed vigor. *Measurement* 253:117416

[95] Xue H, Xu X, Yang Y, Hu D, Niu G. 2024. Rapid and non-destructive prediction of moisture content in maize seeds using hyperspectral imaging. *Sensors* 24(6):1855

[96] Jiang W, Wang J, Lin R, Chen R, Chen W, et al. 2024. Machine learning-based non-destructive terahertz detection of seed quality in peanut. *Food Chemistry: X* 23:101675

[97] Chen Z, Wei Y, Hou J, Huang J, Zhu X, et al. 2024. Transcriptional atlas for embryo development in soybean. *Seed Biology* 3:e022

[98] Lin P, Wang K, Wang Y, Hu Z, Yan C, et al. 2022. The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. *Genome Biology* 23(1):14

[99] Šimková H, Câmara AS, Mascher M. 2024. Hi-C techniques: from genome assemblies to transcription regulation. *Journal of Experimental Botany* 75(17):5357−5365

[100] Zhang X, Luo Z, Marand AP, Yan H, Jang H, et al. 2025. A spatially resolved multi-omic single-cell atlas of soybean development. *Cell* 188(2):550−567.e19

[101] Zhang M, Kong D, Wang H. 2023. Genomic landscape of maize domestication and breeding improvement. *Seed Biology* 2:9

[102] Lam HM, Xu X, Liu X, Chen W, Yang G, et al. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* 42(12):1053−1059

[103] Bai M, Zhang J, Lin W, Zhou Y, Jiang M, et al. 2026. A flanking-nicks prime editor (FLICK-PE) system to boost prime editing in dicots. *Nature Communication* 17:337

[104] Cao X, Xie H, Wang Z, Guo R, Jing F, et al. 2026. An efficient tissue-culture-free soybean genetic transformation technology using the extremely simple cut-dip-budding strategy. *The Innovation* 7(3):101221

[105] Lin W, Li C, Li M, Guan Y. 2025. Emerging nucleases in crop genome editing: towards intellectual property independence and technical flexibility. *Seed Biology* 4:e008

[106] Zhong X, Wang J, Shi X, Bai M, Yuan C, et al. 2024. Genetically optimizing soybean nodulation improves yield and protein content. Nature. *Plants* 10(5):736−742

[107] Tang X, Su T, Han M, Wei L, Wang W, et al. 2017. Suppression of extracellular invertase inhibitor gene expression improves seed weight in soybean (*Glycine max*). *Journal of Experimental Botany* 68(3):469−482

[108] Wang J, Zhang L, Wang S, Wang X, Li S, et al. 2025. AlphaFold-guided bespoke gene editing enhances field-grown soybean oil contents. *Advanced Science* 12(23):e2500290