Genetic regulations of the oil and protein contents in soybean seeds and strategies for improvement

Ailin Liu[a], Sau-Shan Cheng[a], Wai-Shing Yung[a], Man-Wah Li[a,b], and Hon-Ming Lam[a,b,*]

[a]Centre for Soybean Research of the State Key Laboratory of Agrobiotechnology and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China
[b]Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
*Corresponding author: e-mail address: honming@cuhk.edu.hk

Contents

Abstract

Soybean [*Glycine max* (L.) Merr.] is an important legume crop that provides high-quality vegetable protein and oil. In general, oil makes up around 19% of the dry weight of a soybean seed, with the five most abundant fatty acids being palmitic, stearic, oleic, linoleic, and linolenic acids. Storage proteins make up roughly 40% of the dry seed weight,

Advances in Botanical Research
ISSN 0065-2296
https://doi.org/10.1016/bs.abr.2022.03.002

mainly consisting of β-conglycinin (7S globulin) and glycinin (11S globulin). The content and composition of soybean seeds are valuable traits controlled by a complex genetic background and various growth conditions. Environmental stresses, including drought, temperature, and salinity, could severely reduce seed productivity and alter the chemical compositions. Understanding the genetic regulations of oil and protein biosynthesis and how they are affected by abiotic stresses is imperative to enhancing the seed quality. This chapter reviews how oil and protein contents and compositions are affected by environmental conditions, summarizes the genes involved in protein and oil biosynthesis/metabolism, emphasizes the genetic mechanisms of oil/protein accumulation, and offers strategies for improving soybean production both quantitatively and qualitatively.

1. Introduction

Protein and oil are the two principal seed constituents that make soybean (*Glycine max*) an important crop. Due to population growth and strong demand for oil and feed protein, the soybean acreage has increased globally. Soybean plantation areas in the United States was around 87.6 million acres (USDA, 2022) in 2021. Between 2000 and 2020, soybean production more than doubled, from 161 to nearly 353 million tons, and soybean oil production increased from 25.6 to 59.9 million tons (FAOSTAT, 2020). The United States Department of Agriculture (USDA) food composition data show that soybean seeds contain 36.49% protein, 19.94% oil, of which more than half are unsaturated fatty acids (11.2%), 30.16% carbohydrates, and 9.3% fiber (FoodData [USDA], 2022). The protein concentration in soybean ranges from 341 to 568 g/kg of total seed weight, and oil content ranges from 83 to 279 g/kg (Wilson, 2004). Increasing the oil accumulation and protein content in soybean seeds has been a major objective of soybean breeding programs for decades (Van & McHale, 2017). However, the negative correlations in soybean seeds between protein content and oil content as well as yield render it challenging to increase soybean protein content while maintaining the desired seed oil level and seed yield (Patil et al., 2017; Rincker et al., 2014).

The wide range in protein and oil concentrations is not only genetically controlled but is also affected by multiple conditions like agronomic practices, soil conditions, disease pressure and environmental factors including climate change, increasing average temperature and decreasing water availability (Alsajri et al., 2020; Miransari, 2016a). Genetic mapping studies utilizing biparental populations have been carried out for the identification of quantitative trait loci (QTLs) for soybean seed composition. Nowadays,

over 700 QTLs controlling soybean seed oil and protein contents have been identified using accessions with different genetic backgrounds, and 57 QTLs have been confirmed (Van & McHale, 2017). The biosynthesis of oil has been well studied, and the fatty acid (FA) composition is found to be highly regulated via the spatial separation of biosynthetic steps within different subcellular compartments.

A better understanding of the biosynthetic genes and regulatory mechanisms of soybean protein and oil biosynthesis could facilitate the breeding of varieties with high protein or oil content via genetic engineering approaches. Thus, this chapter describes the soybean seed compositions of protein and oil, how seed contents are altered under different environmental conditions, and summarizes the biosynthetic genes and genetic regulations determining protein and FA compositions. Also, we highlight the strategies for improving protein and oil contents in soybean from multiple approaches.

2. Oil and protein contents in soybean seeds

2.1 Oil content and fatty acid composition of soybean seeds

As a dominant oilseed, the oil content in dry soybean seeds averages around 19% and varies from 6.5% to 28.7% depending on the soybean varieties and growth conditions (Greenberg & Hartung, 1998). The yield of soybean oil accounted for nearly 60% of the world's oilseed production in 2019/2020 (SoyStats, 2022). Soybean oil mainly consists of high amounts of fatty acids (FAs), tocopherols and triacylglycerols (TAGs) (Jokić et al., 2013). The fiver major FAs found in soybean seeds are palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3), with concentrations accounting for 10, 4, 18, 55, and 13%, respectively (Clemente & Cahoon, 2009; Lee, Bilyeu, & Shannon, 2007).

The soybean oil quality is determined by the FA composition. Modification of the ratios between saturated FAs (C16:0 and C18:0), oleic acid, and linolenic acid is an important goal in improving the oil quality. Nowadays, soybeans are bred to have higher amounts of monounsaturated FAs and lower amounts of polyunsaturated FAs (PUFA). Unsaturated FAs play an important role in immune system regulation, blood clotting, cholesterol metabolism, and the structure of membrane phospholipids in the brain and retina (Djuricic & Calder, 2021). Foods high in saturated FAs are associated with high cholesterol and high risk for heart disease. Lowering the content of saturated FAs makes soybean oil more acceptable for consumers

and food manufacturers (Wilson, 2004). Moreover, PUFAs such as linoleic (C18:2) and linolenic acids (C18:3) are susceptible to oxidation due to a high number of double bonds, which reduce the oxidative stability and shelf life of the final product. The common way to improve the stability of soybean oil is to increase the saturation by hydrogenation. Yet, there is also a major concern about the trans-fats produced during the process. Thus, increasing the oleic acid (C18:1) content is more favorable for both health and product stability concerns (Yao et al., 2020). It is found that the linolenic acid concentration is the most stable across environments. Soybean lines with stable oleic acid and linolenic acid across environments could thus be used as a source of germplasms with a desirable FA composition (Oliva et al., 2006). Studies on soybean seed fatty acid profiles of various germplasms under different growth conditions have been conducted (Abdelghany et al., 2020; Anwar, Kamal, Nadeem, & Shabir, 2016; Ciabotti et al., 2019; Do Wee, Hashiguchi, Anai, Suzuki, & Akashi, 2017; La et al., 2019; Song et al., 2016). Breeding programs that can make use of this information to alter the oil composition is a priority in improving soybean oil for both food and industrial uses.

2.2 Protein content and composition of soybean seeds

Soybean is grown for its production of seeds with a rich protein content and substantial nutritional value for human and animal consumption worldwide. On average, the protein content accounts for around 40% of the dry weight of soybean seeds, consisting mainly of storage proteins (Krishnan, Natarajan, Mahmoud, & Nelson, 2007; Medic, Atkinson, & Hurburgh, 2014). Glycinin (11S globulin) and conglycinin (7S globulin) are two major components, constituting around 70% of the total storage proteins (Liu et al., 2007; Medic et al., 2014; Paek, Imsande, Shoemaker, & Shibles, 1997; Panthee et al., 2004). Soybean seeds also contain a considerable amount of Bowman-Birk inhibitors (BBI) and Kunitz trypsin inhibitors (KTI). During food processing, inactivation by cleavage or aggregation of these inhibitors is needed due to their inhibitory effects on protein digestion by trypsin and chymotrypsin (Gillman, Kim, & Krishnan, 2015). In spite of the adverse effects on the nutritional value, these protease inhibitors were reported to possess anticarcinogenic activities (Kennedy, 1998; Kobayashi, 2013).

Soybean seeds also comprise about 1% lectin, which is a class of glycoproteins that specifically bind to carbohydrates (Sharon & Lis, 2002). Similar to the protease inhibitors, lectins have been shown to adversely impact nutrient uptake in animals (Friedman & Brandon, 2001). Although lectins cause damaging effects when bound to the intestinal epithelium, they have also been reported to have potential anticarcinogenic activity and therapeutic uses (de Mejia, Bradford, & Hasler, 2003; George, Bhide, Thengane, Hosseini, & Manjaya, 2008; Kik et al., 1989). Besides, the immunodominant allergen, P34, constitutes around 1% of the soybean seed proteins. P34 was identified as a 34-kDa protein localized in the protein storage vacuoles (Kalinski, Melroy, Dwivedi, & Herman, 1992). Around 2% of the soybean seed proteome is contributed by oleosins, which play an essential role in maintaining the stability of oleosomes and thus affect lipid storage (Schmidt et al., 2011). The presence of active ureases also reduces the nutritional value of soybean meal especially when mixed with other food sources in animal feed (Real-Guerra, Staniscuaski, & Carlini, 2013).

The soybean seed protein content is tightly regulated, as numerous early attempts have failed to utilize the soybean system for the significant accumulation of recombinant proteins, such as vaccine antigens, milk protein and growth hormone (Herman, 2014; Lau & Sun, 2009; Oakes, Bost, & Piller, 2009; Philip, Darnowski, Maughan, & Vodkin, 2001; Russell et al., 2005). Also, the innate program in soybean to rebalance the seed protein composition was functional even the synthesis of major storage proteins was disrupted. In transgenic plants with the α and α′ subunits of conglycinin genes silenced, glycinin, proglycinin and a precursor of P34 were significantly accumulated in seeds, resulting in a similar seed protein content to the wild type. The accumulated proteins were also stored in protein bodies in addition to the normal storage vacuoles (Kinney, Jung, & Herman, 2001). Moreover, knocking down both conglycinin and glycinin synthesis did not lead to a significant change in the total protein content. A detailed investigation in the protein composition identified that the lack of conglycinin and glycinin was compensated by the accumulation of other storage proteins such as lectin, P34 and KTI, while increases in the abundance of other proteins may also have participated in the rebalancing. Further analysis also indicated that the amino acid profile of the mutant seeds was similar to that of the wild type (Schmidt et al., 2011).

3. Alterations in oil and protein contents and compositions under diverse environmental conditions

Crop production is hampered by environmental stresses such as drought, flooding, salinity and heat. Stresses adversely affect plant physiology and metabolism, inhibiting plant growth and eventually leading to deterioration in seed quality and yield. Indeed, environment has been reported as the major factor influencing protein concentration in soybean (Assefa et al., 2019). It is well documented that a negative correlation exists between oil and protein contents, and when oil concentration decreases, the protein content increases (Bandillo et al., 2015; Chung et al., 2003; Cober, Voldeng, & H., 2000; Wilcox, 1998). These environmental factors are also likely the cause of variations between seed protein and oil concentrations from different geographic locations. Numerous studies have reported the impacts of different stress factors on the oil and protein production of soybean seeds and they are summarized in this section.

3.1 Changes in oil and protein contents due to drought stress

Drought stress is a severe abiotic stress factor limiting soybean production worldwide. It is reported that soybean yield is most sensitive to drought stress during the pod filling stage of development (Rodrigues et al., 2012). The drought effects at the pod filling stage were first assessed in the soybean cultivars "Gnome" and "Hodgson 78." Both protein and oil contents were affected. Drought stress increased seed protein content by around 10% and the protein contents in both cultivars increased linearly with the increase in drought stress intensity (Dornbos & Mullen, 1992). On the other hand, the oil content of "Gnome" decreased up to 12.4% mainly due to a reduction in oleic acid content (Dornbos & Mullen, 1992).

Lipids are a major form of carbon and energy storage in the seed for germination. Drought causes oil content reduction owing to the inhibition on the production of digestible carbohydrates such as glucose, fructose, and sucrose. Drought also affects the FA composition due to reduced unloading of sugars from the stem to developing seeds (Bellaloui, Hu, Mengistu, Kassem, & Abel, 2013). A study on the effects of drought treatment on the slow-wilting genotypes, NTCPR94-5157 and N04-9646, showed that seed protein, oleic acid and sugar contents increased whereas seed oil,

linoleic and linolenic acids decreased, and concluded that the accumulation of oleic acid could be an adaptive mechanism to maintain cell turgor under drought stress (Bellaloui, Gillen, et al., 2013). Seeds of the soybean cultivar, Hutcheson, grown under drought conditions (−90 to −100 kPa soil water potential) had higher total oil and oleic acid concentrations and lower linoleic and linolenic concentrations compared to irrigated soybeans (−15 to −20 kPa soil water potential) (Bellaloui, Hu, et al., 2013). A study on the irrigation effect on seven soybean germplasms found that it was not significant on the PUFA concentration but it changes the composition depending on the genotypes (Lee, Oliva, Sleper, & Shannon, 2008).

The effect of drought stress on the reduction of seed oil is more significant than on the seed protein content (Miransari, 2016a). The explanation was that drought stress could trigger the remobilization of nitrogen sources from the leaf (Turner, Davies, Plummer, & Siddique, 2005). However, most of the later studies reported a negative impact of drought stress on seed protein content (Rotundo & Westgate, 2009). It was found that seeds from the cultivars M7, M9 and Hobit stressed by drought at the seed filling stage contained significantly lower protein contents than the control, while the same stress applied at earlier stages such as the flowering and podding stages had no significant effects on seed protein content (Maleki et al., 2013). In line with this finding, Nakagawa et al. (2018) reported that the seed protein content was >50% lower in drought-stressed soybean cultivar "Fukuyutaka" compared to that in control. A similar reduction was reported in another study on "Sinara" and "Sigalia," although the drought treatment was applied at the flowering stage (Anda, Simon, Soos, da Silva, & Menyhart, 2021). The seed protein contents of "Asgrow AG5332" and "Progeny 5333RY" were also reported to correlate positively with the degree of drought stress in terms of soil moisture (Wijewardana, Reddy, & Bellaloui, 2019).

In a study investigating the different effects drought had on the components of seed storage proteins between a drought-tolerant (BOSA) and a drought-sensitive (L 121) genotype under drought stress (Blanusa, Stikic, Vucelic-Radovic, Barac, & Velickovic, 2000), it was found that drought affected the stage at which the accumulation of the β subunit of conglycinin took place in each genotype, although the differences were not observed at consistent stages in two experimental years (Blanusa et al., 2000). The alterations in the accumulation of the β subunit of conglycinin were also reported in another study on the cultivated soybean Harosoy, in which the accumulation was observed at a later developmental stage when under drought stress (Samarah, Mullen, Cianzio, & Scott, 2006). Also, an earlier accumulation of

heat-stable dehydrin proteins was observed under greenhouse conditions but not in the field (Samarah et al., 2006).

3.2 Changes in oil and protein contents under salinity stress

Seed yield, oil, FA composition and protein production are all affected by salinity (Miransari, 2016b). Under salt stress, the percentage of oil in soybean seeds declined while the total oil content per seed increased during the seed filling stage (Ghassemi-Golezani & Farhangi-Abriz, 2018). Irrigation with 6000 mg/L saline solution on the soybean cultivar Giza 111 led to a significant decrease in seed yield (the product of pod weight and seed number/plant) and oil content compared to the control plants irrigated with tap water (Sadak et al., 2020). It was found that different salt concentrations might have opposite effects on the seed protein content (Phang, Shao, & Lam, 2008; Wan, Shao, Chen, & Yan, 2002).

A life-long exposure to mild salinity at 4 dS/m had no significant effect on the proportion of protein in the seeds of cultivar M7, while higher salinity conditions at 7 and 10 dS/m reduced the proportion of protein in seeds (Farhangi-Abriz & Ghassemi-Golezani, 2016). The seedlings grown under all the above salinity conditions had lower overall yield and total protein production per plant. Although most of the amino acids had higher concentrations per gram dry weight, those of leucine and valine were reduced under higher salinity (Farhangi-Abriz & Ghassemi-Golezani, 2016). Alterations in the metabolic pathways of branched-chain amino acids were reported as one of the key features of salt stress adaptation in soybean and therefore may affect the resultant accumulation of these amino acids in seeds (Liu et al., 2019; Yung et al., 2022).

3.3 Effects of heat stress on oil and protein contents in soybean seeds

Heat stress harms almost all the stages of soybean development from germination to maturity, and results in a substantial decrease in crop yield. Heat stress inhibits photosynthetic activities and hastens senescence prematurely, which decreases the synthesis and distribution of assimilates to seeds (McDonald & Paulsen, 1997). Increasing temperature can lead to a reduced number of cells per cotyledon, later onset of seed fill and maturity, and lower seed weight and size (Tacarindua, Shiraiwa, Homma, Kumagai, & Sameshima, 2012). The soybean oil content is determined by both soybean varieties and temperature conditions. A study on the temperature effect on

20 soybean cultivars across 60 locations found that seed oil concentration increased with temperature (14–28°C) (Piper & Boote, 1999). In an earlier study, a slight increase in seed protein concentration was observed in "Hodgson 78" grown at 27, 29, 33, and 35°C (Dornbos & Mullen, 1992). At 29°C the percentage of protein increased by 4% and oil content decreased by 2.6%, compared to those at 35°C. High temperature decreased the proportion of unsaturated acid including linoleic and linolenic acids in seeds but increased oleic acid concentration (Dornbos & Mullen, 1992). A study on different soybean varieties including Hutcheson, DT97-4290, AG 4403 and AG 4903 showed that growing at higher temperatures (36/28°C, day/night) resulted in higher total oil and oleic acid concentrations and lower linoleic and linolenic concentrations than those grown at normal temperatures (25/20°C, day/night) (Bellaloui, Reddy, & Mengistu, 2015).

The effect of elevated temperatures on the protein concentration in soybean seeds was variable among cultivars (Ortiz, De Smet, Sozzani, & Locke, 2022; Piper & Boote, 1999). Generally, an elevated temperature led to an increase in protein concentration and a decrease in total protein content in seeds (Rotundo & Westgate, 2009). Similar to drought stress, heat stress applied from the seed filling stage onwards led to a lower percentage of protein in the seeds of cultivar "Fukuyutaka" (Nakagawa et al., 2020). A severe heat regime of 42°C/26°C (day/night) cycles starting from the flower initiation stage was reported to impair the accumulation of glycinin and conglycinin in both heat-tolerant (DS25-1) and sensitive cultivars (DT97-4290) (Chebrolu et al., 2016). When these cultivars were exposed to moderate heat stress (36°C/24°C day/night cycles), the levels of proline and leucine in seeds were higher, while the levels of aspartate were lower, than those grown in the optimal condition (Chebrolu et al., 2016). Similarly, a lower accumulation of the β subunit of conglycinin, lipoxygenase1, sucrose binding proteins, P34 and Bowman-Birk inhibitor was observed in the seeds of "DS25-1" and "DT97-4290" exposed to the same severe heat stress condition in a later study (Krishnan, Kim, Oehrle, Smith, & Gillman, 2020). As expected, the seeds from both cultivars contained higher levels of heat shock proteins HSP70 and HSP17.6 under severe heat stress (Krishnan et al., 2020). As shown in the transmission electron microscopy images, the cotyledon cells were plasmolyzed and the protein storage vacuoles were destroyed in the sensitive cultivar (DT97-4290) under severe heat, whereas the ultrastructure in the tolerant cultivar (DS25-1) was much less affected, although the plant storage vacuoles were only partially filled (Krishnan et al., 2020).

3.4 Other abiotic stresses and factors

The seed protein and oil contents of soybean are also impacted by cold and flooding stresses. Under low temperature conditions, various studies have shown a decrease in the oil content in soybean seeds (Alsajri et al., 2020; Khan et al., 2011; Rotundo, Miller-Garvin, & Naeve, 2016). However, cold stress was found to have no significant effect on the seed protein content, although the exposure to long-term cold stress (12°C/6°C day/night cycles for 9 days) after seed sowing resulted in higher seed yields in 12 out of 16 selected cultivars (Staniak, Stepien-Warda, Czopek, Kocira, & Baca, 2021). Similarly, flooding stress was reported to cause significant changes in other seed components such as fatty acids and isoflavones but not in the protein content (VanToai et al., 2012). In addition, phosphate deficiency was found to reduce the protein yield per plant while an enhanced carbon dioxide concentration at 800 $\mu mol\,mol^{-1}$ exerted an opposite effect (Singh, Barnaby, Reddy, & Sicher, 2016).

4. Genetic strategies for improving oil and protein contents

4.1 Genetic mapping and identification of genes associated with oil and protein contents

Seed oil content is a quantitative trait that is governed by multiple genes. In the soybean genome, there are more than 1100 annotated acyl lipid genes (Schmutz et al., 2010), about double that in the model plant Arabidopsis. This is probably due to the recent whole-genome duplication of the soybean genome and could potentially explain the high oil content in the soybean seed. Thus far, QTL mapping and genome-wide association study (GWAS) have identified a large number of genomic regions controlling the soybean seed oil content (Cao et al., 2017; Di et al., 2022; Eskandari, Cober, & Rajcan, 2013; Huang et al., 2020; Hwang et al., 2014; Karikari et al., 2019; Li, Xu, Yang, & Zhao, 2019; Liu et al., 2020; Priolli, Campos, Stabellini, Pinheiro, & Vello, 2015; Silva et al., 2021; Zhang et al., 2018; Zhao et al., 2019). One of the main goals of soybean breeding is to increase the seed oil content. It has been found that a large portion of a selected region in the soybean genome is occupied by overlapping QTL regions related to oil content (Zhou et al., 2015), some of which contain known biosynthetic genes for fatty acids.

Soybean seed protein and oil contents often share overlapping QTL regions. As previously reviewed, improvement in the soybean protein

content is challenging as soybean oil content and yield are negatively correlated with its protein content (Bandillo et al., 2015; Chaudhary et al., 2015; Pandurangan et al., 2012; Patil et al., 2017). Since the selection for a higher and better oil content is a higher priority than for the protein content, any improvements in the protein content often takes a backseat in breeding programs. Analyses on the breeding efforts over the past 60 years revealed that efforts were biased toward selecting for higher oil content rather than protein content in the hope of improving nutritional values (Mahmoud et al., 2006). Wild soybean or its recombinant inbred lines serve as a good reservoir of the genetic variations for different traits. Based on restriction fragment length polymorphism (RFLP) analyses, *G. max* contains alleles that are associated with higher oil contents than in *G. soja*, and vice versa for the alleles associated with protein contents (Diers, Keim, Fehr, & Shoemaker, 1992). Wild soybean often contains specific QTLs that are associated with higher protein contents and higher heterogeneity of the proteins (Natarajan, Xu, Bae, Caperna, & Garrett, 2006; Sebolt, Shoemaker, & Diers, 2000). The wild soybean usually contains a higher 11S:7S ratio than the cultivated germplasms, leading to the suggestion that some regulatory genes or elements in the cultivars might have been altered during the domestication of soybean (Kwanyuen, Pantalone, Burton, & Wilson, 1997). These qualities of the wild soybean enable it to be a good genetic resource for improved protein levels.

Undoubtedly, cultivated soybean might also possess certain desirable traits after years of breeding and selection. It is reported that the wild soybean genotype contains more abundant allergenic and antinutritional proteins such as the subunits of β-conglycinin, glycinin, and KTI than the cultivated genotypes (Natarajan et al., 2006). Some *G. max* accessions such as PI603570A and PI567476 were reported to have a lower level of the allergen P34 (Bilyeu, Ren, Nguyen, Herman, & Sleper, 2009; Joseph, Hymowitz, Schmidt, & Herman, 2006; Koo et al., 2011). Some of the cultivars carry four amino acid substitutions in the P34 sequence, and some only express a truncated P34 protein (Bilyeu et al., 2009; Joseph et al., 2006; Koo et al., 2011). Some of the cultivated germplasms might show high protein contents upon selection in some geographical regions. PI407788A and "Enrei" contain the (CC+) *GmSWEET39* allele at chromosome 15 for higher protein content in South Korea and Japan respectively (Brzostowski & Diers, 2017; Shimomura et al., 2015; Zhang et al., 2020). Other successful breeding of the high-protein elites, such as R95-1705, R05-1415, and R05-1772, were reported to obtain stable high-protein-content germplasms by crossing (Chen, Ishibashi, Dombek,

& Rupe, 2011; Chen, Sneller, Ishibashi, & Cornelious, 2008). Thus, the breeding program for developing protein-rich soybean should include these different lines as the parents.

4.2 Gene regulations on the content, nutrition and stability of soybean oil

4.2.1 The FA biosynthesis pathway in soybean seeds

The biosynthesis of oils in the seed has been well described (Bates, Stymne, & Ohlrogge, 2013). In brief, FAs are assembled in the plastid. Acetyl-CoA is first converted to malonyl-CoA by acetyl-CoA carboxylase (ACCase). The acyl chain is elongated on the acyl carrier protein (ACP) by fatty acid synthase (FAS) by adding two carbon atoms every time until 16 carbon acyl-ACP (C16:0) is produced. The 16:0-ACP may be converted further to 18:0-ACP by ketoacyl-ACP synthase II (KASII) and may subsequently be desaturated by stearoyl-ACP desaturase (SAD). Desaturation of the fatty acids may also take place through the action of fatty acid desaturase (FAD) to form polyunsaturated FAs (PUFAs). The free fatty acids (C16:0, C18:0, C18:1, etc.) are then released from ACP by acyl-ACP thioesterase (FAT) and exported to the endoplasmic reticulum. There, in the de novo TAG synthesis pathway (Kennedy Pathway), free fatty acids will be incorporated into glycerol-3-phosphate (G3P) by acyl-CoA:G3P acyltransferase (GPAT) to form lyso-phosphatidic acid (LPA). Further additions of FA to LPA by acyl-CoA:LPA acyltransferase (LPAAT) will yield phosphatidic acid (PA), which will be converted to diacylglycerol (DAG) by PA phosphatase (PAP). At the final step, acyl-CoA:DAG acyltransferase (DGAT) catalyzes the formation of triacylglycerol (TAG) from DAG through the addition of the last FA. Alternatively, TAG can also be produced through the phosphatidylcholine (PC) pathway, although this is not the major source of TAG in soybean seeds. The TAG in the seed is mostly stored in oleosomes (oil bodies), which are wrapped in a single layer of membrane compositing of phospholipids and proteins such as oleosins (Nikiforidis, 2019).

4.2.2 Regulatory genes that can improve the FA composition of soybean seeds

There are two major goals in improving the soybean seed oil content. The first is to increase the total oil content of the seed, ideally without sacrificing the overall yield. The second is to "improve" the oil quality by increasing the beneficial fatty acids while reducing those detrimental ones. Thus far, a

number of genes, including not only the biosynthetic ones, but also a number of transcription factors from soybean have been characterized.

The most common genetic strategy to reduce the PUFA content in seeds is to impair the function of fatty acid desaturase (FAD) to create a high-oleic acid low-linoleic acid soybean. In the past few decades, different molecular techniques have been used to manipulate *FAD* genes in soybean, including induced mutagenesis (Anai et al., 2008), RNA interference (RNAi) (Yang et al., 2018), transcription activator-like effector nucleases (TALENs) genome editing (Demorest et al., 2016; Haun et al., 2014), CRISPR-cas9-mediated genome editing (al Amin et al., 2019; Wu et al., 2020). FAD2 and FAD3 desaturate oleic acid and linoleic acid, respectively. The mutation of either copy of *FAD2* genes (*GmFAD2-1A* and *GmFAD2-1B*) can increase the seed oleic acid content to 50% of the total and reduce linoleic acid down to 30% (Anai et al., 2008). The *Gmfad2-1A Gmfad2-1B* double mutant can further increase oleic acid to 80% and reduce linoleic acid to less than 5% (Pham, Shannon, & Bilyeu, 2012). There are three *FAD3* genes in the soybean genome, namely *GmFAD3A*, *GmFAD3B*, and *GmFAD3C* (Bilyeu, Palavalli, Sleper, & Beuselinck, 2005). The *Gmfad3A Gmfad3C* double mutant has a 3% linolenic acid content (Bilyeu et al., 2005), while the triple mutant has the linolenic acid content as low as 1% (Pham et al., 2012). Stacking the mutations of *fad2* and *fad3* together generated lines with high oleic acid, low linoleic acid and low linolenic acid (Demorest et al., 2016; Pham et al., 2012). A high-oleic low-linolenic soybean produced by TALEN genome editing has been deemed non-regulated by USDA in 2020, making it the first commercially available genome-edited crop.

Acyl-ACP thioesterase (FAT) is responsible for the release of saturated FAs from ACPs. TILLING-by-Sequencing+ identified a number of soybean mutants with amino acid substitutions in different *FAT* genes (Zhou et al., 2021). These mutations cause a reduction in saturated FAs and an increase in oleic acid (Zhou et al., 2021). The loss of function of either *GmFAT1a* or *GmFAT1b* can reduce ~50% palmitic and ~30% stearic acid contents without altering the unsaturated FA content (Ma, Sun, Whelan, & Shou, 2021). Although the *Gmfat1a Gmfat1b* double mutant has significantly reduced saturated FA contents in the leaf, it also suffers from growth retardation and male sterility, which is not practical for crop production (Ma et al., 2021).

Through genetic mapping, a mutation disrupting the donor splice site of *GmKASIIIA* (*Glyma09g41380*) was found to be associated with reduced palmitic acid content (Cardinal et al., 2014). Given that KASIII participates

in the reaction between acetyl-CoA and malonyl-ACP, the reason why the mutation of *GmKASIIIA* should lead to a reduction in palmitic acid is a mystery. However, both the mutation of *KASIIIA* (Cardinal et al., 2014) and the ectopic expression of *KASIIIA* caused a reduction in the total oil content (Dehesh, Tai, Edwards, Byrne, & Jaworski, 2001). This suggests that the ratio of KASIII to malonyl-ACP may be crucial for efficient FA biosynthesis.

In some food applications, oils with a higher saturated FA content are needed. To fulfill this demand, soybean lines with high stearic acid are also being generated. A germplasm with as high as 35% stearic acid has been found (Pantalone, Wilson, Novitzky, & Burton, 2002). Yet, the cause for the high stearic acid content remains unknown. Through mutant screening, a mutation in a delta-9-stearoyl-acyl-carrier protein desaturase (SACPD-C)-encoding gene was found to increase the stearic acid content in soybean seeds (Carrero-Colon, Abshire, Sweeney, Gaskin, & Hudson, 2014). SPACD-C is responsible for converting stearic acid to oleic acid. Thus, impairing the function of SPACD-C can increase stearic acid at the expense of oleic acid (Carrero-Colon et al., 2014). This is also true in high-oleic acid lines. Stacking the *sacpd* mutation with the *fad2-1a fad2-1b* double mutation caused a 10% increase in stearic acid but also a 10% decrease in oleic acid in the triple mutant compared to the *fad2-1a fad2-1b* double mutant (Bilyeu et al., 2018). An increase in stearic acid content can also be achieved by the ectopic expression of the stearoyl-ACP thioesterase-encoding gene (Park, Graef, Xu, Tenopir, & Clemente, 2014). The overexpression of a codon-optimized stearoyl-ACP thioesterase from mangosteen in soybean led to a significant enhancement in the stearic acid content in the seed from 3% in the wild type to around 10% in the transgenic lines (Park et al., 2014). Nevertheless, a slight reduction in other FAs was also observed in these transgenic lines (Park et al., 2014).

4.2.3 Transcription regulation of oil accumulation in soybean seeds

Oil biosynthesis in the soybean seed is governed by a complex transcription regulation network. WRINKLED1 (WRI) is a transcription factor in the APETALA2/ethylene responsive element-binding protein (AP2/EREBP) family and is responsible for the biosynthesis of storage compounds in Arabidopsis. Its homologs in soybean appear to play a central role in oil accumulation in the seed. There are two WRI1-encoding genes in the soybean genome, *GmWRI1a* (*Glyma15g34770/Glyma.15G221600*) and *GmWRI1b* (*Glyma08g24420/Glyma.08G227700*). *GmWRI1a* is highly expressed in maturing soybean seeds (Chen et al., 2018). Overexpression

of *GmWRI1a* with a seed-specific promoter resulted in significantly higher total oil and fatty acid contents in the seed, accompanied by a bigger seed size and higher total seed weight per plant and per hundred seeds (Chen et al., 2018). The most significant changes in fatty acid contents were the significant increases in oleic acid (18:1) and linoleic acid (C18:2). Results from gene expression study and electrophoretic mobility shift assay (EMSA) showed that GmWRI1a is able to act on the AW-box on the promoter of genes encoding ketoacyl-ACP synthase (KASII and KASIII), enoyl-ACP reductase, stearoyl-ACP desaturase (SAD), acyl-ACP thioesterase, and long-chain acyl-CoA synthetase (Chen et al., 2018).

Another seed-specific DREB-type AP2 transcription factor, GmDREBL, also governs the oil accumulation in soybean seeds (Zhang et al., 2016). The expression of *GmDREBL* is positively regulated by GmABI3 and GmABI5 (Zhang et al., 2016). Overexpression of *GmDREBL* in soybean hairy roots increased the expression of *GmWRI1a*, but not that of *GmWRI1b* (Chen et al., 2018; Zhang et al., 2016). The ectopic expression of *GmDREBL* in Arabidopsis enhanced the expressions of *KAS1*, *FATA*, *LPD1* and *SAD* (*At5g16240*) through the activation of *WRI1* (Zhang et al., 2016). The activation of these genes resulted in larger seeds and higher FA contents (C18:1, C18:2, C18:3, and C20:1) in the seed (Zhang et al., 2016). Evidence also showed that *GmDREBL* may be subjected to selection during domestication, as cultivated soybeans, which have high oil contents, tend to have a higher expression of *GmDREBL* (Zhang et al., 2016).

A nuclear transcription factor Y subunit A-encoding gene (*GmNYFA*, *Glyma02g47380*) was differentially expressed between the developing seed of cultivated soybeans and wild soybeans (Lu et al., 2016). This gene is also located within a QTL for seed oil content (Lu et al., 2016). Similar to *GmDREBL*, the overexpression of *GmNYFA* in Arabidopsis also enhanced the expressions of *WRI1*, *KAS1*, *LPD1*, *FATA*, *SAD*, and a gene encoding thioesterase, resulting in an elevation in seed FAs (C16:0, C18:1, C18:2, C18:3, and C20:1) and oil content (Lu et al., 2016). It is probable that GmNYFA acts directly on the CCAAT box on the *WRI1* promoter to boost the expression of *WRI1* and indirectly increase the expression of the other oil biosynthetic genes (Lu et al., 2016). As evidence of a potential target of soybean domestication, the promoter of *GmNYFA* in cultivated soybeans and some wild soybeans harbor a 1500-bp insertion. This insertion is associated with the higher expression of *GmNYFA* and oil content in cultivated soybeans (Lu et al., 2016).

The expression of *GmWRI1a* was found to be regulated by GmLEC2a, a B-box-containing transcription factor expressed mainly in the flower, pod and seed (Manan et al., 2017). The ectopic overexpression of *GmLEC2a* can enhance the total FA content in transgenic Arabidopsis seeds and soybean hairy roots (Manan et al., 2017). Comparisons between the transcriptomes of the wild type and transgenic hairy roots expressing *GmLEC2a* revealed that GmLEC2a is a positive regulator for a group of transcription factors involved in TAG biosynthesis, including *GmWRI1a*, *GmWRI1b*, *GmDof11*, *GmFUS3*, *GmABI3* and *GmLEC1*. The expressions of TAG biosynthetic genes, including *KAS*, *GPAT*, *DGAT*, *PGAT*, *PAP* and so on, were also regulated by the overexpression of *GmLEC2a*. In contrast, sucrose synthase genes were largely suppressed to shut off the biosynthesis of starch in order to redirect the materials for lipid biosynthesis (Manan et al., 2017).

GmZF351 encodes a tandem CCCH zinc finger protein in soybean, functioning as a transcription activator that governs oil accumulation in seed (Li et al., 2017). Ectopic overexpression of *GmZF351* can increase FA and TAG accumulation in the seed of both transgenic Arabidopsis and soybean (Li et al., 2017). GmZF351 can directly activate the expression of *WRI1*, *Biotin Carboxyl Carrier Protein 2* (*BCCP2*), *KASIII*, *DGAT1*, *OLE2* in soybean (Li et al., 2017). Through the activation of *WRI1*, lipid biosynthetic genes downstream of *WRI1* are also regulated by GmZF351 (Li et al., 2017). In general, the expression of *GmZF351* is higher in cultivated soybean than in wild soybean and it is positively correlated with the lipid content in seeds (Li et al., 2017). An analysis of the promoter of *GmZF351* identified four major haplotypes, one for cultivated soybean and three for wild soybean (Li et al., 2017). This suggests that domestication may have selected a haplotype of *GmZF351* with a higher expression that can in turn increase seed oil content.

Two soybean DNA binding with one finger (Dof) transcription factors, GmDof4 and GmDof11, were also found to be involved in lipid accumulation (Wang et al., 2007). The expression of *GmDof4* is predominantly in flower and pod while that of *GmDof11* is constitutively expressed in leaf, flower and pod (Wang et al., 2007). As mentioned in a previous paragraph, *GmDof11*, among other transcription factors, is also a target of GmLEC2a (Manan et al., 2017). The ectopic expressions of *GmDof4* and *GmDof11* can elevate the total FA contents in Arabidopsis (Wang et al., 2007), *Brassica napus* L. (Sun et al., 2018) and *Chlorella ellipsoidea* (Zhang et al., 2014). Since the gene expression study was done in a heterologous

transgenic system, it may not truly reflect the functions of GmDof4 and GmDof11 in soybean (Sun et al., 2018; Wang et al., 2007; Zhang et al., 2014). However, the available evidence suggests that GmDof4 and GmDof11 may target different sets of genes related to TAG biosynthesis and accumulation. While only GmDof4 was shown to interact with *accD* that encodes the β-subunit of ACCase in Arabidopsis (Wang et al., 2007), the ectopic expressions of both *GmDof4* and *GmDof11* could enhance the expression of *accD* in *B. napus* L. (Sun et al., 2018). Furthermore, GmDof11 was shown to positively regulate a long-chain-acyl CoA synthetase-encoding gene in Arabidopsis while GmDof4 was shown to act on *FAB2*, *FAD3* and *FAD8* in *B. napus* L. (Sun et al., 2018; Wang et al., 2007).

GmbZIP123 is expressed highly in multiple organs, especially root and developing seed (Song, Li, et al., 2013). The translation of the Arabidopsis homologs of *GmbZIP123* is regulated by sucrose due to the presence of upstream open reading frames (uORFs) in their transcripts (Song, Li, et al., 2013). Two uORFs are also found in the transcript of *GmbZIP123*, implying that the translation of *GmbZIP123* may also be regulated by sucrose. The ectopic overexpression of *GmbZIP123* in Arabidopsis significantly increased the FA content in seeds without altering the 1000-seed weight (Song, Li, et al., 2013). Unlike *GmWRI1a* which altered only oleic acid and linoleic acid contents, the overexpression of *GmbZIP123* altered all major FA contents (C16:0, C18:1, C18:2, C18:3 and C20:1) (Song, Li, et al., 2013). RNA-seq analyses showed that instead of acting directly on FA biosynthetic genes, GmbZIP123 enhanced the expression of sucrose transporter-encoding genes including *SUC1*, *SUC5*, *cell wall invertase 1* (*cwINV1*), *cwINV3*, and *cwINV6* in transgenic Arabidopsis (Song, Li, et al., 2013). Consistent with the expression study, the glucose, fructose, and sucrose levels were all higher in the developing seeds of the transgenic line than those in the wild-type plant (Song, Li, et al., 2013). These results suggested that GmbZIP123 can enhance FA accumulation by allocating carbon resources to the developing seeds.

GmMYB73 is highly expressed in the flower, root, and early developing seed in soybean (Liu et al., 2014). The ectopic overexpression of *GmMYB73* in Arabidopsis caused multiple phenotypic changes including the repression of trichome formation, increase in seed size, and enhanced oil accumulation (Liu et al., 2014). The overexpression can also improve the FA accumulation in *Lotus japonica* seed and soybean hairy roots (Liu et al., 2014), suggesting the function of GmMYB73 is conserved in these species. The overexpression of *GmMYB73* functionally resembled its Arabidopsis homolog,

CAPRICE-like proteins (CPC and TRY), by derepressing the expression of *phospholipase D alpha 1* (*PLDα1*) and other genes through the repression of *GLABRA 2* (*GL2*) by forming a repressor complex with GL3 and Enhance of GLABRA (EGL3) in Arabidopsis (Liu et al., 2014).

4.2.4 Altering soybean oil accumulation by modifying biosynthetic genes

DGAT is responsible for the acylation of the sn-3 position of DAG to form TAG. DGAT is considered to be the rate-limiting enzyme in TAG biosynthesis and the expression of *DGAT* is correlated with TAG accumulation during seed development (Settlage, Kwanyuen, & Wilson, 1998). There are three major DGATs, namely DGAT1, DGAT2, and DGAT3, each with different affinities and activities toward different acyl-CoAs. At least three GmDGAT1-, five GmDGAT2-, and two GmDGAT3-encoding genes have been identified in the soybean genome (Xu et al., 2021; Zhao et al., 2019). Ectopic expressions of the three *GmDGAT1s* in Arabidopsis led to a significant increase in oil content (Xu et al., 2021; Zhao, Bi, et al., 2019; Zhao, Chang, et al., 2019), while the expressions of *GmDGAT2* and *GmDGAT3* mainly altered the FA composition, with less prominent effects on total oil content (Xu et al., 2021; Zhao, Bi, et al., 2019; Zhao, Chang, et al., 2019). This makes *DGAT1* an attractive target for genetic manipulation to increase soybean oil production. In a later experiment, a 3% increase in total FA was observed in transgenic soybean seeds overexpressing *GmDGAT1B* (*Glyma17g06120*), accompanied by a significant increase in oleic acid but a decrease in linoleic acid (Xu et al., 2021). Although the manipulation of soybean *DGAT* yielded seeds with an improved oil content, it also came with penalties in the protein content. However, a study has demonstrated that the expression of *DGAT2A* from a soil fungus, *Umbelopsis ramanniana*, can significantly increase soybean seed oil content with a minimal trade-off in the protein content (Lardizabal et al., 2008). Furthermore, expressing a modified GmDGAT with high oleic acid affinity (based on the analysis of the mutant library of DGAT from *Corylus americana*, which is a species with high oleic acid and high oil content) added an extra 3% of oil to the transgenic soybean seed (Roesler et al., 2016). This implies protein engineering using base-editing technology to modify the endogenous *GmDGAT* genes may be a possible way to improve soybean oil content in the future.

G3P is the essential building block of TAG. Therefore improving the G3P supply could also increase TAG accumulation. The overexpression

of a plastidic NAD^+-dependent glycerol-3-phosphate dehydrogenase gene, *GmGPDHp1*, produced transgenic soybean seeds with a higher G3P content (Zhao et al., 2021). The increase in TAG also brought about a 20% increase in seed oil content, but potentially at the expense of soluble sugars in mature seeds (Zhao et al., 2021). The transgenic seeds have lower free FAs but higher PC, PE, DAG, and TAG, suggesting that the overexpression of *GmGPDHp1* has enhanced the incorporation of FAs with G3P.

Impairing the catabolism of TAG is also an effective way to improve soybean seed oil content. TAG lipases hydrolyze TAG into DAG and a carboxylate. Four TAG lipase-encoding genes, namely *Sugar Dependent1-1* (*GmSDP1-1*), *GmSDP1-2*, *GmSDP1-3*, and *GmSDP1-4*, are identified in soybean (Kanai, Yamada, Hayashi, Mano, & Nishimura, 2019). Knocking down the expressions of all four *GmSDPs* by RNAi improved the FA content in the seed and the overall oil yield per plant (Kanai et al., 2019). Moreover, without altering the seed number per plant, the seed yield was also improved due to the increase in seed weight (Kanai et al., 2019).

4.2.5 Improving oil accumulation by altering the sink and the source

Apart from directly improving the biosynthesis of oil and FAs, enhancing the sink can also improve the oil content in soybean seed. A GWAS of 219 diverse soybean accessions identified an oleosin protein-encoding gene (*GmOLEO1*) that affects oil content in soybean seeds, which is a potential target of artificial selection (Zhang et al., 2019). Oleosin is an important protein constituent of the oil body membrane. The constitutive overexpression of *GmOLEO1* in soybean led to a 10% increase in oil contents in the seed, which was mainly attributable to the increase in linoleic acid (C18:2) and linolenic acid (C18:3) (Zhang et al., 2019). Although the seed size was reduced in the transgenic soybean, the overall seed yield in total seed weight per plant showed a significant increase (Zhang et al., 2019). The higher *GmOLEO1* expression in the transgenic line facilitated the formation of smaller but more numerous oil bodies in the seed (Zhang et al., 2019), which may be the reason why more oil could be packed. At the same time, the enlargement in the sink also provided a positive feedback to increase the expressions of oil biosynthesis-related genes and other oleosin-encoding genes (Zhang et al., 2019). Tajima's D analysis demonstrated that the promoter, but not the coding region, of *GmOLEO1* was under positive selection (Zhang et al., 2019).

The biosynthesis of FAs in the seed largely relies on the carbon source from photosynthetic tissues. The efficiency of sucrose reallocation can thus determine the rate of oil accumulation in the seed. Through population studies, two Sugar Will Eventually be Exported Transporter-encoding genes (*GmSWEET39/GmSWEET10a/Glyma.15G049200* and *GmSWEET10b/Glyma.08G183500*) were identified to be involved in the domestication-related enhancement of oil accumulation in soybean seed (Miao et al., 2020; Wang et al., 2020; Zhang et al., 2020). They are mainly expressed in the thin-walled parenchyma of the seed coat, and the resulting proteins are localized on the plasma membrane (Miao et al., 2020; Wang et al., 2020; Zhang et al., 2020). Both GmSWEET10a and GmSWEET10b are able to transport sucrose and hexoses in a heterologous cell culture system and Xenopus oocytes (Wang et al., 2020). Swapping promoters and coding sequences of different alleles of *GmSWEET10a* in transgenic analyses demonstrated that both the gene expression level and the protein sequence of GmSWEET10a contributed to the oil accumulation (Wang et al., 2020). The overexpression of *GmSWEET10a* and *GmSWEET10b* led to higher oil content, larger seed size, but lower protein content in the transgenic soybean seeds (Wang et al., 2020). Consistent with the results from overexpression, knockout mutants of these two genes had lower oil content, smaller seed size, and higher protein content (Wang et al., 2020). Together with the fact that knocking out both genes caused a significant reduction in sucrose and hexoses in the developing embryo but not in the seed coat (Wang et al., 2020), it is suggested that GmSWEET10a and GmSWEET10b function to upload sugar into the embryo from the seed coat.

4.3 Gene regulation and strategy for improving protein contents in soybean seeds

4.3.1 Regulatory genes involved in soybean protein contents

There are over 200 protein-related quantitative trait loci (QTLs) in the soybean genome as recorded by Soybase (Grant, Nelson, Cannon, & Shoemaker, 2009). Unlike the seed oil content, there is no clear story on how soybean seed protein content is regulated. QTLs controlling soybean proteins are scattered among all the soybean chromosomes except chromosome 16 (Huang et al., 2019). Previous studies showed that the region mapped to the soybean linkage group I (LG-I) on chromosome 20 showed a strong correlation with seed protein content (Bandillo et al., 2015; Bolon et al., 2010; Diers et al., 1992; Pandurangan et al., 2012; Sebolt et al., 2000). The seed protein-associated QTLs in this region on chromosome 20 were

further narrowed down to an 8.4-Mb region (Bandillo et al., 2015), a 3-Mb region (Hwang et al., 2014) and a 1-Mb region (Vaughn, Nelson, Song, Cregan, & Li, 2014), respectively. Yet these regions were still too large for the selection of protein content-controlling genes for characterization. Moreover, the soybean protein content is thought to be regulated by multiple genes, while the phenotype is highly subjected to environmental influences such as temperature, latitude, and humidity (Bolon et al., 2010; Huang et al., 2019; Qin et al., 2014). Moreover, the soybean genome duplication event has led to the rearrangement of homologous chromosomes and resulted in the duplication of different QTL regions, eventually diluting the efforts of studying protein content regulation (Lestari, Van, Lee, Kang, & Lee, 2013). These factors increase the difficulty in studying protein content regulating mechanisms. Recent researches have coupled QTL analyses with microarrays or RNA-sequencing to screen for potential gene candidates that regulate soybean seed protein contents (Bolon et al., 2010; Wang et al., 2021). This approach was able to help resolve single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) between the high- and low-protein genotypes (Hwang et al., 2014; Wang et al., 2021). Furthermore, QTL analyses related to the soybean seed content could then focus more on the non-shared gene contents within the duplicated genomic regions to consolidate research efforts (Lestari et al., 2013). Although a relationship between the protein content and some genes was proposed, not many of them were validated for their corresponding gene functions.

4.3.2 Rebalancing storage proteins in soybean seeds by the molecular manipulation of globulin genes

There are two principal components of soybean seed storage proteins as categorized by their sedimentation coefficients, the 7S β-conglycinin, and the 11S glycinin. These proteins constitute about 70% of the total storage proteins in the seed (Liu et al., 2007). The 7S β-conglycinin is a trimer composed of three subunits: α′, α, and β subunits (Koshiyama, 1968). The 11S glycinin is a hexamer encoded by the *Gy1*, *Gy2*, and *Gy3* (group I) and *Gy4* and *Gy5* (group II) glycinin-encoding genes (Fischer & Goldberg, 1982; Nielsen et al., 1989). The β-conglycinin and glycinin-encoding genes consist of multiple gene members, and each of them is regulated transcriptionally, translationally, or post-translationally after flowering and during embryogenesis (Dickinson, Hussein, & Nielsen, 1989; Harada, Barker, & Goldberg, 1989). The methionine and cysteine concentrations showed a strong correlation with the 11S:7S globulin ratio, while threonine and tyrosine

showed a moderate correlation (Kwanyuen et al., 1997). The 11S glycinin contains more cysteine, methionine, threonine, and tyrosine. It is believed that a higher 11S:7S ratio could lead to beneficial changes in the nutritional quality of soybean meal.

Soybean protein accumulation is subjected to delicate adjustments during seed filling to maintain a relatively stable total protein content in the seed. Repression of the storage proteins would result in other proteins being accumulated as compensation (Kwanyuen et al., 1997; Ogawa, Tayama, Kitamura, & Kaizuma, 1989; Schmidt et al., 2011; Yang, Yu, Zheng, & James, 2016). The low-7S β-conglycinin soybean varieties, Keburi and Mo-shi-dou Gong 503, were found not to express the α′ subunit of β-conglycinin and, as a result, had a lowered level of the α and β subunits of β-conglycinin, without showing any abnormality in the agronomic traits (Ogawa et al., 1989). The co-suppression of β-conglycinin α′ and α subunits resulted in similar seed storage protein rebalancing as observed in the low-7S β-conglycinin soybean varieties, Keburi and Mo-shi-dou Gong 503 (Kinney et al., 2001; Ogawa et al., 1989). As shown in the electron microscopy-immunogold assays, P34 and 11S glycinin concentrations were boosted in the protein bodies in these two soybean varieties (Kinney et al., 2001). It is believed that the accumulation of P34 and 11S glycinin in the endoplasmic reticulum (ER)–derived protein bodies was enhanced to compensate for the loss of seed storage proteins (Kinney et al., 2001). Concomitant soybean protein rebalancing was also observed in the transgenic soybean with glycinin and β-conglycinin-encoding genes suppressed by RNA interference (RNAi) (Schmidt et al., 2011). The concentrations of P34 and KTI surged in the soybean seed to compensate for the missing storage proteins, eventually attaining a comparable seed protein content with that of untransformed lines (Schmidt et al., 2011). However, the introgression of exogenous green fluorescence protein (GFP) into the co-repressed background did not alter the proteome of the soybean seed, suggesting the presence of an intrinsic selection mechanism for specific compensating proteins (Schmidt et al., 2011). The soybean seed globulin rebalancing was also reported in studies conducted with cultivars expressing different levels of a subunit of the 11S glycinin, 11SA4 (Yang et al., 2016). It was demonstrated that the 11SA4 deficiency in the soybean seed would induce a compensatory accumulation of 7S globulins and result in minimal changes in the total protein content (Yang et al., 2016).

The compensatory nature of soybean seed storage proteins suggests an approach for improving the nutritional value of soybean seeds, such as

higher cysteine and methionine contents (Kwanyuen et al., 1997). The regulatory pathway for glycinin and β-conglycinin rebalancing and compensation might provide good study models for understanding protein loading to seeds for storage in soybean (Kinney et al., 2001). Undoubtedly, manipulating the ratio of the 11S and 7S globulins could affect the nutritional value of the soybean seed. It was reported that the ratio of globulin members might affect the texture and processing method required in food processing, such as the ability to form aggregates and precipitates, as 11S and 7S globulins have different association and dissociation properties upon thermal treatments (Saio & Watanabe, 1978; Utsumi, Damodaran, & Kinsella, 1984). Hence the resulting changes in physical properties should be considered in making soybean seed protein content improvements (Saio & Watanabe, 1978).

4.3.3 Improving the nutritional value of soybean seed by increasing protein content and changing amino acid composition

A lot of effort has been made on manipulating the original soybean seed composition for better nutritional values, but manipulating proteins other than glycinin and β-conglycinin might be an instant solution before we can achieve a better understanding of the protein content regulation mechanisms in soybean seeds. The ectopic expression of *β-zein*, a methionine-rich storage protein from maize, could increase the soybean seed methionine level but the impact was relatively small (Guo et al., 2020). GmSWEET39 is a sucrose transporter that facilitates sugar transport and the supply of intermediates for protein and oil metabolism in soybean seeds (Zhai, Liu, Xu, & Shanklin, 2017). It was reported that a 2-bp (CC) deletion in *GmSWEET39* encoded a truncated transporter that resulted in a high-oil low-protein content in the seed (Zhang et al., 2020). The high level of asparagine synthetase (AS) in soybean leaves was positively correlated with high seed protein concentration among different cultivars with varying degrees of AS expression (Wan, Shao, Shan, Zeng, & Lam, 2006). Furthermore, the asparagine synthetase (AS) and asparaginase (ASPG) levels were positively correlated with the protein contents from a study of 73 recombinant inbred lines (Pandurangan et al., 2012). It is suggested that the transcript and protein (enzyme) levels, and enzyme activities in the seed coat might explain the differences in protein contents among these lines (Pandurangan et al., 2012). There are not many well-known transcriptional and translational controls on soybean seed protein composition. One of them is the regulation by C4 of nuclear factor Y (NF-YC4) and Qua-Quine Starch (QQS). *Qua-Quine Starch* (*QQS*) is an orphan gene from *Arabidopsis thaliana*, and

the ectopic expression of *QQS* modulated protein compositions in the leaf and seed by affecting the carbon and nitrogen partitioning in soybean (Li et al., 2015; Li & Wurtele, 2015). QQS also interacts with the conserved subunit C4 of nuclear factor Y (NF-YC4) to mediate the transcription of a variety of CCAAT box-containing genes (Li et al., 2015). The ectopic expression of *QQS* could lead to an overall increase in protein content by around 20% with a reduction in oil content by around 10% (Li & Wurtele, 2015). Other than the ectopic expression of *QQS* in soybean, overexpression of its interacting partner *GmNF-YC4-2* (*Glyma.04g196200*) also resulted in a higher protein level in the soybean seed along with increased yield (O'conner et al., 2021). Besides, *GmNF-YC4-2*-overexpressing lines showed increased resistance to biotic stress such as bacterial, viral, and fungal challenges (O'conner et al., 2021). As mentioned in a previous section on oil accumulation regulation in the soybean seed by transcription factors, GmLEC2 also acts as a master regulator of soybean seed protein content. The ectopic expression of *GmLEC2a* in Arabidopsis resulted in an increase in protein, while the *Atlec2* loss-of-function mutant seeds contained approximately 40% less protein than wild type seeds (Manan et al., 2017).

Altering the amino acid content in the soybean seed might be a feasible way to improve its nitrogen-based nutritional value (Ishimoto et al., 2010; Pandurangan et al., 2012; Wan et al., 2006). As already mentioned, selective breeding in the past 60 years has minimally affected the overall amino acid composition of cultivated soybean seeds, while those from the wild soybean *G. soja* contain a significantly higher level of various amino acids (Mahmoud et al., 2006). Several studies utilized feedback-insensitive enzymes in amino acid metabolism for boosting the respective amino acid contents in seeds (Falco et al., 1995; Ishimoto et al., 2010; Kita et al., 2010; Song, Hou, et al., 2013; Yu et al., 2018). The ectopic expression of a rice tryptophan-feedback-resistant subunit of anthranilate synthase (OASA1D) increased the tryptophan level in the seed, using either a 35S promoter or a seed-specific promoter (Ishimoto et al., 2010; Kita et al., 2010). OASA1D is a regulatory enzyme in the tryptophan (Trp) biosynthesis pathway. The overexpression of *OASA1D* bypassed the negative tryptophan feedback mechanism, which led to the accumulation and deposition of tryptophan in the soybean seed without detrimental effects on its growth and development (Ishimoto et al., 2010). The ectopic expression of *OASA1D* in both low- and high-free amino acid breeding lines of soybean also resulted in a similar boost in the level of tryptophan in the seed (Kita et al., 2010). The ectopic expression of an Arabidopsis methionine-feedback-resistant form of

cystathionine γ-synthase (CGS) in soybean using either a seed-specific or a constitutive promoter was shown to boost the methionine storage in seed (Song, Hou, et al., 2013; Yu et al., 2018). Cystathionine γ-synthase (CGS) is the first committed enzyme bridging the aspartate family biosynthesis pathway with methionine metabolism. Similar to the bypass of the negative tryptophan feedback mechanism, the transgenic soybean seeds had an elevated methionine content without affecting their agronomic performance (Song, Hou, et al., 2013; Yu et al., 2018). Such uses of feedback-insensitive enzymes in amino acid metabolism also worked on soybean seed lysine content. The ectopic expression of lysine-feedback-insensitive *dihydrodipicolinate synthase* (*DHDPS*) from *E. coli* and *aspartokinase* (*AK*) from *Corynebacterium* using a seed-specific promoter resulted in a five-fold increase in the seed lysine level (Falco et al., 1995).

5. Conclusions

Seed protein and oil contents are two valuable traits determined by genotypes and planting conditions. To produce soybean seeds with higher protein/oil contents, efforts should be made to screen the wild soybeans and germplasms from various geographical locations and under different stress environments for tolerant genotypes and incorporate them into breeding programs. It is also vital to identify all the genes, regulators, proteins, and metabolites involved in the protein and oil metabolism pathways to decipher the molecular mechanisms that determine seed composition. Understanding the development, growth, seed productivity, seed quality and sink-source relations in the soybean seed will give us better insights into the physiological responses and underlying genetics of stress tolerance, with the ultimate goal of enhancing soybean seed yield and quality.

Acknowledgments

This work was supported by Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), Lo Kwee-Seong Biomedical Research Fund and Guangdong Provincial Department of Science and Technology 2020 Key Areas Research and Development Programs: Breeding for High Yield and High Quality New Soybean Cultivars for Tropics and Subtropics (2020B020220008). J.Y. Chu copy-edited the text. Any opinions, findings, conclusions or recommendations expressed in this publication do not reflect the views of the Government of Hong Kong Special Administrative Region or the Innovation and Technology Commission.

References

Abdelghany, A. M., Zhang, S., Azam, M., Shaibu, A. S., Feng, Y., Qi, J., et al. (2020). Natural variation in fatty acid composition of diverse world soybean germplasms grown in China. *Agronomy*, *10*, 24.

al Amin, N., Ahmad, N., Wu, N., Pu, X. M., Ma, T., Du, Y. Y., et al. (2019). CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max*. L). *BMC Biotechnology*, *19*.

Alsajri, F. A., Wijewardana, C., Irby, J. T., Bellaloui, N., Krutz, L. J., Golden, B., et al. (2020). Developing functional relationships between temperature and soybean yield and seed quality. *Agronomy Journal*, *112*, 194–204.

Anai, T., Yamada, T., Hideshima, R., Kinoshita, T., Rahman, S. M., & Takagi, Y. (2008). Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by GmFAD2-1a. *Breeding Science*, *58*, 447–452.

Anda, A., Simon, B., Soos, G., da Silva, J. A. T., & Menyhart, L. (2021). Water stress modifies canopy light environment and qualitative and quantitative yield components in two soybean varieties. *Irrigation Science*, *39*, 549–566.

Anwar, F., Kamal, G. M., Nadeem, F., & Shabir, G. (2016). Variations of quality characteristics among oils of different soybean varieties. *Journal of King Saud University*, *28*, 332–338.

Assefa, Y., et al. (2019). Assessing variation in US soybean seed composition (protein and oil). *Frontiers in Plant Science*, *10*, 298.

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., et al. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome*, *8*.

Bates, P. D., Stymne, S., & Ohlrogge, J. (2013). Biochemical pathways in seed oil synthesis. *Current Opinion in Plant Biology*, *16*, 358–364.

Bellaloui, N., Gillen, A. M., Mengistu, A., Kebede, H., Fisher, D. K., Smith, J. R., et al. (2013). Responses of nitrogen metabolism and seed nutrition to drought stress in soybean genotypes differing in slow-wilting phenotype. *Frontiers in Plant Science*, *4*, 498.

Bellaloui, N., Hu, Y., Mengistu, A., Kassem, M. A., & Abel, C. A. (2013). Effects of foliar boron application on seed composition, cell wall boron, and seed δ15N and δ13C isotopes in water-stressed soybean plants. *Frontiers in Plant Science*, *4*, 270.

Bellaloui, N., Reddy, K. N., & Mengistu, A. (2015). Drought and heat stress effects on soybean fatty acid composition and oil stability. In *Processing and impact on active components in food* (pp. 377–384). Elsevier.

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2005). Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. *Crop Science*, *45*, 1830–1836.

Bilyeu, K., Ren, C., Nguyen, H. T., Herman, E., & Sleper, D. A. (2009). Association of a four-Basepair insertion in the P34 gene with the low-allergen trait in soybean. *Plant Genome*, *2*, 141–148.

Bilyeu, K., Skrabisova, M., Allen, D., Rajcan, I., Palmquist, D. E., Gillen, A., et al. (2018). The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. *Journal of the American Oil Chemists' Society*, *95*, 39–49.

Blanusa, T., Stikic, R., Vucelic-Radovic, B., Barac, M., & Velickovic, D. (2000). Dynamics of seed protein biosynthesis in two soybean genotypes differing in drought susceptibility. *Biologia Plantarum*, *43*, 55–59.

Bolon, Y. T., et al. (2010). Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biology*, *10*.

Brzostowski, L., & Diers, B. (2017). Agronomic evaluation of a high protein allele from PI407788A on chromosome 15 across two soybean backgrounds. *Crop Science*, *57*(6), 2972–2978. https://doi.org/10.2135/cropsci2017.02.0083.

Cao, Y. C., Li, S. G., Wang, Z. L., Chang, F. G., Kong, J. J., Gai, J. Y., et al. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Frontiers in Plant Science*, *8*.

Cardinal, A. J., et al. (2014). Mapping the low palmitate fap1 mutation and validation of its effects in soybean oil and agronomic traits in three soybean populations. *Theoretical and Applied Genetics*, *127*, 97–111.

Carrero-Colon, M., Abshire, N., Sweeney, D., Gaskin, E., & Hudson, K. (2014). Mutations in SACPD-C result in a range of elevated stearic acid concentration in soybean seed. *PLoS One*, *9*.

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyodan, B., et al. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Frontiers in Plant Science*, *6*, 1–16.

Chebrolu, K. K., Fritschi, F. B., Ye, S. Q., Krishnan, H. B., Smith, J. R., & Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed metabolome. *Metabolomics*, *12*.

Chen, P., Ishibashi, T., Dombek, D. G., & Rupe, J. C. (2011). Registration of R05-1415 and R05-1772 high-protein soybean germplasm lines. *Journal of Plant Registrations*, *5*, 410–413.

Chen, P., Sneller, C. H., Ishibashi, T., & Cornelious, B. (2008). Registration of high-protein soybean germplasm line R95-1705. *Journal of Plant Registrations*, *2*, 58–59.

Chen, L., Zheng, Y. H., Dong, Z. M., Meng, F. F., Sun, X. M., Fan, X. H., et al. (2018). Soybean (*Glycine max*) WRINKLED1 transcription factor, GmWRI1a, positively regulates seed oil accumulation. *Molecular Genetics and Genomics*, *293*, 401–415.

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., et al. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Cell Biology and Molecular Genetics*, *43*, 1053–1067.

Ciabotti, S., Juhász, A. C. P., Mandarino, J. M. G., Costa, L. L., Corrêa, A. D., Simão, A. A., et al. (2019). Chemical composition and lipoxygenase activity of soybean (*Glycine max* L. Merrill.) genotypes, specific for human consumption, with different tegument colours. *Brazilian Journal of Food Technology*, *22*.

Clemente, T. E., & Cahoon, E. B. (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiology*, *151*, 1030–1040.

Cober, E. R., Voldeng, D., & H. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Science*, *40*, 39–42.

de Mejia, E. G., Bradford, T., & Hasler, C. (2003). The anticarcinogenic potential of soybean lectin and lunasin. *Nutrition Reviews*, *61*, 239–246.

Dehesh, K., Tai, H., Edwards, P., Byrne, J., & Jaworski, J. G. (2001). Overexpression of 3-ketoacyl-acyl-carrier protein synthase IIIs in plants reduces the rate of lipid synthesis. *Plant Physiology*, *125*, 1103–1114.

Demorest, Z. L., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biology*, *16*.

Di, Q., Piersanti, A., Zhang, Q., Miceli, C., Li, H., & Liu, X. Y. (2022). Genome-wide association study identifies candidate genes related to the linoleic acid content in soybean seeds. *International Journal of Molecular Sciences*, *23*.

Dickinson, C. D., Hussein, E. H., & Nielsen, N. C. (1989). Role of posttranslational cleavage in glycinin assembly. *Plant Cell*, *1*, 459–469.

Diers, B. W., Keim, P., Fehr, W. R., & Shoemaker, R. C. (1992). RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, *83*, 608–612.

Djuricic, I., & Calder, P. C. (2021). Beneficial outcomes of omega-6 and omega-3 polyunsaturated fatty acids on human health: An update for 2021. *Nutrients*, *13*, 2421.

Do Wee, C., Hashiguchi, M., Anai, T., Suzuki, A., & Akashi, R. (2017). Fatty acid composition and distribution in wild soybean (*Glycine soja*) seeds collected in Japan. *Asian Journal of Plant Sciences*, *16*, 52–64.

Dornbos, D. L., & Mullen, R. E. (1992). Soybean seed protein and oil contents and fatty-acid composition adjustments by drought and temperature. *Journal of the American Oil Chemists' Society*, *69*, 228–231.

Eskandari, M., Cober, E. R., & Rajcan, I. (2013). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theoretical and Applied Genetics*, *126*, 483–495.

Falco, S. C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R. T., et al. (1995). Transgenic canola and soybean seeds with increased lysine. *Biotechnology (N. Y.)*, *13*, 577–582.

FAOSTAT. (2020). *FAOSTAT*.

Farhangi-Abriz, S., & Ghassemi-Golezani, K. (2016). Improving amino acid composition of soybean under salt stress by salicylic acid and jasmonic acid. *Journal of Applied Botany and Food Quality*, *89*, 243–248.

Fischer, R. L., & Goldberg, R. B. (1982). Structure and flanking regions of soybean seed protein genes. *Cell*, *29*, 651–660.

FoodData (USDA). (2022). *FoodData*.

Friedman, M., & Brandon, D. L. (2001). Nutritional and health benefits of soy proteins. *Journal of Agricultural and Food Chemistry*, *49*, 1069–1086.

George, M. A., Bhide, S. V., Thengane, R. J., Hosseini, G. H., & Manjaya, J. G. (2008). Identification of low lectin mutants in soybean. *Plant Breeding*, *127*, 150–153.

Ghassemi-Golezani, K., & Farhangi-Abriz, S. (2018). Changes in oil accumulation and fatty acid composition of soybean seeds under salt stress in response to salicylic acid and jasmonic acid. *Russian Journal of Plant Physiology*, *65*, 229–236.

Gillman, J. D., Kim, W. S., & Krishnan, H. B. (2015). Identification of a new soybean Kunitz trypsin inhibitor mutation and its effect on Bowman-Birk protease inhibitor content in soybean seed. *Journal of Agricultural and Food Chemistry*, *63*, 1352–1359.

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2009). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research*, *38*, 843–846.

Greenberg, P., & Hartung, H. N. (1998). *The whole soy cookbook: 175 Delicious, nutritious, easy-to-prepare recipes featuring tofu, tempeh, and various forms of nature's healthiest bean*. CA: Three Rivers Press.

Guo, C., Liu, X., Chen, L., Cai, Y., Yao, W., Yuan, S., et al. (2020). Elevated methionine content in soybean seed by overexpressing maize β-zein protein. *Oil Crop Science*, *5*, 11–16.

Harada, J. J., Barker, S. J., & Goldberg, R. B. (1989). Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*, *1*, 415–425.

Haun, W., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnology Journal*, *12*, 934–940.

Herman, E. M. (2014). Soybean seed proteome rebalancing. *Frontiers in Plant Science*, *5*.

Huang, J. H., Ma, Q. B., Cai, Z. D., Xia, Q. J., Li, S. X., Jia, J., 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*, 6448–6460.

Huang, S., et al. (2019). Identification of soybean genes related to soybean seed protein content based on quantitative trait loci collinearity analysis. *Journal of Agricultural and Food Chemistry*, *67*, 258–274.

Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics*, *15*, 1–12.

Ishimoto, M., et al. (2010). Evaluation of amino acid content and nutritional quality of transgenic soybean seeds with high-level tryptophan accumulation. *Molecular Breeding*, *25*, 313–326.

Jokić, S., SudaR, R., Svilović, S., Vidović, S., Bilić, M., Velić, D., et al. (2013). Fatty acid composition of oil obtained from soybeans by extraction with supercritical carbon dioxide. *Czech Journal of Food Sciences*, *31*, 116–125.

Joseph, L. M., Hymowitz, T., Schmidt, M. A., & Herman, E. M. (2006). Evaluation of glycine germplasm for nulls of the immunodominant allergen P34/Gly m Bd 30k. *Crop Science*, *46*, 1755–1763.

Kalinski, A., Melroy, D. L., Dwivedi, R. S., & Herman, E. M. (1992). A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation. *The Journal of Biological Chemistry*, *267*, 12068–12076.

Kanai, M., Yamada, T., Hayashi, M., Mano, S., & Nishimura, M. (2019). Soybean (*Glycine max* L.) triacylglycerol lipase GmSDP1 regulates the quality and quantity of seed oil. *Scientific Reports*, *9*.

Karikari, B., Li, S. G., Bhat, J. A., Cao, Y. C., Kong, J. J., Yang, J. Y., et al. (2019). Genome-wide detection of major and epistatic effect QTLs for seed protein and oil content in soybean under multiple environments using high-density bin map. *International Journal of Molecular Sciences*, *20*.

Kennedy, A. R. (1998). The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. *The American Journal of Clinical Nutrition*, *68*, 1406s–1412s.

Khan, A. Z., Shah, P., Khan, H., Nigar, S., Perveen, S., Shah, M. K., et al. (2011). Seed quality and vigor of soybean cultivars as influenced by canopy temperature. *Pakistan Journal of Botany*, *43*, 642–648.

Kik, M. J. L., Rojer, J. M., Mouwen, J. M. V. M., Koninkx, J. F. J. G., Vandijk, J. E., & Vanderhage, M. H. (1989). The interaction between plant-lectins and the small intestinal epithelium—A primary cause of intestinal disturbance. *The Veterinary Quarterly*, *11*, 108–115.

Kinney, A. J., Jung, R., & Herman, E. M. (2001). Cosuppression of the α subunits of β-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum-derived protein bodies. *Plant Cell*, *13*, 1165–1178.

Kita, Y., Nakamoto, Y., Takahashi, M., Kitamura, K., Wakasa, K., & Ishimoto, M. (2010). Manipulation of amino acid composition in soybean seeds by the combination of deregulated tryptophan biosynthesis and storage protein deficiency. *Plant Cell Reports*, *29*, 87–95.

Kobayashi, H. (2013). Prevention of cancer and inflammation by soybean protease inhibitors. *Frontiers in Bioscience (Elite Edition)*, *5*, 966–973.

Koo, S. C., Bae, D. W., Seo, J. S., Park, K. M., Choi, M. S., Kim, S. H., et al. (2011). Proteomic analysis of seed storage proteins in low allergenic soybean accession. *Journal of Applied Biological Chemistry*, *54*, 332–339.

Koshiyama, I. (1968). Chemical and physical properties of a 7S protein in soybean globulins. *Cereal Chemistry*, *45*, 394–404.

Krishnan, H. B., Kim, W. S., Oehrle, N. W., Smith, J. R., & Gillman, J. D. (2020). Effect of heat stress on seed protein composition and ultrastructure of protein storage vacuoles in the cotyledonary parenchyma cells of soybean genotypes that are either tolerant or sensitive to elevated temperatures. *International Journal of Molecular Sciences*, *21*.

Krishnan, H. B., Natarajan, S. S., Mahmoud, A. A., & Nelson, R. L. (2007). Identification of glycinin and beta-conglycinin subunits that contribute to the increased protein content of high-protein soybean lines. *Journal of Agricultural and Food Chemistry*, *55*, 1839–1845.

Kwanyuen, P., Pantalone, V. R., Burton, J. W., & Wilson, R. F. (1997). A new approach to genetic alteration of soybean protein composition and quality. *JAOCS, Journal of the American Oil Chemists' Society*, *74*, 983–987.

La, T., Large, E., Taliercio, E., Song, Q., Gillman, J. D., Xu, D., et al. (2019). Characterization of select wild soybean accessions in the USDA germplasm collection for seed composition and agronomic traits. *Crop Science*, *59*, 233–251.

Lardizabal, K., Effertz, R., Levering, C., Mai, J., Pedroso, M. C., Jury, T., et al. (2008). Expression of Umbelopsis ramanniana DGAT2A in seed increases oil in soybean. *Plant Physiology*, *148*, 89–96.

Lau, O. S., & Sun, S. S. M. (2009). Plant seeds as bioreactors for recombinant protein production. *Biotechnology Advances*, *27*, 1015–1022.

Lee, J.-D., Bilyeu, K. D., & Shannon, J. G. (2007). Genetics and breeding for modified fatty acid profile in soybean seed oil. *Journal of Crop Science and Biotechnology*, *10*, 201–210.

Lee, J. D., Oliva, M. L., Sleper, D. A., & Shannon, J. G. (2008). Irrigation has little effect on unsaturated fatty acid content in soya bean seed oil within genotypes differing in fatty acid profile. *Journal of Agronomy and Crop Science*, *194*, 320–324.

Lestari, P., Van, K., Lee, J., Kang, Y. J., & Lee, S. H. (2013). Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. *Frontiers in Plant Science*, *4*, 1–8.

Li, L., & Wurtele, E. S. (2015). The QQS orphan gene of Arabidopsis modulates carbon and nitrogen allocation in soybean. *Plant Biotechnology Journal*, *13*, 177–187.

Li, S., Xu, H., Yang, J., & Zhao, T. (2019). Dissecting the genetic architecture of seed protein and oil content in soybean from the Yangtze and Huaihe river valleys using multi-locus genome- wide association studies. *International Journal of Molecular Sciences*, *20*.

Li, L., et al. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 14734–14739.

Li, Q. T., et al. (2017). Selection for a zinc-finger protein contributes to seed oil increase during soybean domestication. *Plant Physiology*, *173*, 2208–2224.

Liu, X. Y., Qin, D., Piersanti, A., Zhang, Q., Miceli, C., & Wang, P. W. (2020). Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. *BMC Plant Biology*, *20*.

Liu, C., Wang, H., Cui, Z., He, X., Wang, X., Zeng, X., et al. (2007). Optimization of extraction and isolation for 11S and 7S globulins of soybean seed storage protein. *Food Chemistry*, *102*, 1310–1316.

Liu, A. L., Xiao, Z. X., Li, M. W., Wong, F. L., Yung, W. S., Ku, Y. S., et al. (2019). Transcriptomic reprogramming in soybean seedlings under salt stress. *Plant, Cell & Environment*, *42*, 98–114.

Liu, Y. F., et al. (2014). Soybean GmMYB73 promotes lipid accumulation in transgenic plants. *BMC Plant Biology*, *14*.

Lu, X., Li, Q. T., Xiong, Q., Li, W., Bi, Y. D., Lai, Y. C., et al. (2016). The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *The Plant Journal*, *86*, 530–544.

Ma, J., Sun, S., Whelan, J., & Shou, H. X. (2021). CRISPR/Cas9-mediated knockout of GmFATB1 significantly reduced the amount of saturated fatty acids in soybean seeds. *International Journal of Molecular Sciences*, *22*.

Mahmoud, A. A., Natarajan, S. S., Bennett, J. O., Mawhinney, T. P., Wiebold, W. J., & Krishnan, H. B. (2006). Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. *Journal of Agricultural and Food Chemistry*, *54*, 3916–3922.

Maleki, A., Naderi, A., Naseri, R., Fathi, A., Bahamin, S., & Maleki, R. (2013). Physiological performance of soybean cultivars under drought stress. *Bulletin of Environment, Pharmacology and Life Sciences*, *2*, 38–44.

Manan, S., Ahmad, M. Z., Zhang, G., Chen, B., Haq, B. U., Yang, J., et al. (2017). Soybean LEC2 regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development. *Frontiers in Plant Science*, *8*, 1–16.

McDonald, G. K., & Paulsen, G. M. (1997). High temperature effects on photosynthesis and water relations of grain legumes. *Plant and Soil*, *196*, 47–58.

Medic, J., Atkinson, C., & Hurburgh, C. R. (2014). Current knowledge in soybean composition. *Journal of the American Oil Chemists' Society*, *91*, 363–384.

Miao, L., Yang, S. N., Zhang, K., He, J. B., Wu, C. H., Ren, Y. H., et al. (2020). Natural variation and selection in GmSWEET39 affect soybean seed oil content. *The New Phytologist*, *225*, 1651–1666.

Miransari, M. (2016a). *Environmental stresses in soybean production: Soybean production. Vol. 2.* Academic Press.

Miransari, M. (2016b). Soybean, protein, and oil production under stress. In *Environmental stresses in soybean production* (pp. 157–176). Elsevier.

Nakagawa, A. C. S., Ario, N., Tomita, Y., Tanaka, S., Murayama, N., Mizuta, C., et al. (2020). High temperature during soybean seed development differentially alters lipid and protein metabolism. *Plant Production Science*, *23*, 504–512.

Nakagawa, A. C. S., Itoyama, H., Ariyoshi, Y., Ario, N., Tomita, Y., Kondo, Y., et al. (2018). Drought stress during soybean seed filling affects storage compounds through regulation of lipid and protein metabolism. *Acta Physiologiae Plantarum*, *40*.

Natarajan, S., Xu, C., Bae, H., Caperna, T. J., & Garrett, W. M. (2006). Proteomic analysis of allergen and antinutritional proteins in wild and cultivated soybean seeds. *Journal of Plant Biochemistry and Biotechnology*, *15*, 103–108.

Nielsen, N. C., Dickinson, C. D., Cho, T. J., Thanh, V. H., Scallon, B. J., Fischer, R. L., et al. (1989). Characterization of the glycinin gene family in soybean. *Plant Cell*, *1*, 313–328.

Nikiforidis, C. V. (2019). Structure and functions of oleosomes (oil bodies). *Advances in Colloid and Interface Science*, *274*.

O'conner, S., Zheng, W., Qi, M., Kandel, Y., Fuller, R., Whitham, S. A., et al. (2021). Gmnf-yc4-2 increases protein, exhibits broad disease resistance and expedites maturity in soybean. *International Journal of Molecular Sciences*, *22*.

Oakes, J. L., Bost, K. L., & Piller, K. J. (2009). Stability of a soybean seed-derived vaccine antigen following long-term storage, processing and transport in the absence of a cold chain. *Journal of the Science of Food and Agriculture*, *89*, 2191–2199.

Ogawa, T., Tayama, E., Kitamura, K., & Kaizuma, N. (1989). Genetic improvement of seed storage proteins using three variant alleles of 7S globulin subunits in soybean (*Glycine max* L.). *Japanese Journal of Breeding*, *39*, 137–147.

Oliva, M. L., Shannon, J. G., Sleper, D. A., Ellersieck, M. R., Cardinal, A. J., Paris, R. L., et al. (2006). Stability of fatty acid profile in soybean genotypes with modified seed oil composition. *Crop Science*, *46*, 2069–2075.

Ortiz, A. C., De Smet, I., Sozzani, R., & Locke, A. M. (2022). Field-grown soybean shows genotypic variation in physiological and seed composition responses to heat stress during seed development. *Environmental and Experimental Botany*, *195*, 104768.

Paek, N. C., Imsande, J., Shoemaker, R. C., & Shibles, R. (1997). Nutritional control of soybean seed storage protein. *Crop Science*, *37*, 498–503.

Pandurangan, S., Pajak, A., Molnar, S. J., Cober, E. R., Dhaubhadel, S., Hernández-Sebasti, C., et al. (2012). Relationship between asparagine metabolism and protein concentration in soybean seed. *Journal of Experimental Botany*, *63*, 3173–3184.

Pantalone, V. R., Wilson, R. F., Novitzky, W. P., & Burton, J. W. (2002). Genetic regulation of elevated stearic acid concentration in soybean oil. *Journal of the American Oil Chemists' Society*, *79*, 549–553.

Panthee, D. R., Kwanyuen, P., Sams, C. E., West, D. R., Saxton, A. M., & Pantalone, V. R. (2004). Quantitative trait loci for beta-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. *Journal of the American Oil Chemists' Society*, *81*, 1005–1012.

Park, H., Graef, G., Xu, Y. X., Tenopir, P., & Clemente, T. E. (2014). Stacking of a stearoyl-ACP thioesterase with a dual-silenced palmitoyl-ACP thioesterase and Delta 12 fatty acid desaturase in transgenic soybean. *Plant Biotechnology Journal*, *12*, 1035–1043.

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., et al. (2017). Molecular mapping and genomics of soybean seed protein: A review and perspective for the future. *Theoretical and Applied Genetics*, *130*, 1975–1991.

Pham, A. T., Shannon, J. G., & Bilyeu, K. D. (2012). Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theoretical and Applied Genetics*, *125*, 503–515.

Phang, T. H., Shao, G. H., & Lam, H. M. (2008). Salt tolerance in soybean. *Journal of Integrative Plant Biology*, *50*, 1196–1212.

Philip, R., Darnowski, D. W., Maughan, P. J., & Vodkin, L. O. (2001). Processing and localization of bovine beta-casein expressed in transgenic soybean seeds under control of a soybean lectin expression cassette. *Plant Science*, *161*, 323–335.

Piper, E. L., & Boote, K. J. (1999). Temperature and cultivar effects on soybean seed oil and protein concentrations. *Journal of the American Oil Chemists' Society*, *76*, 1233–1241.

Priolli, R. H. G., Campos, J. B., Stabellini, N. S., Pinheiro, J. B., & Vello, N. A. (2015). Association mapping of oil content and fatty acid components in soybean. *Euphytica*, *203*, 83–96.

Qin, P., Song, W., Yang, X., Sun, S., Zhou, X., Yang, R., et al. (2014). Regional distribution of protein and oil compositions of soybean cultivars in China. *Crop Science*, *54*, 1139–1146.

Real-Guerra, R., Staniscuaski, F., & Carlini, C. R. (2013). Soybean urease: Over a hundred years of knowledge. In *A comprehensive survey of international soybean research—Genetics, physiology, agronomy and nitrogen relationships* (pp. 317–339).

Rincker, K., Nelson, R., Specht, J., Sleper, D., Cary, T., Cianzio, S. R., et al. (2014). Genetic improvement of US soybean in maturity groups II, III, and IV. *Crop Science*, *54*, 1419–1432.

Rodrigues, F. A., Marcolino-Gomes, J., de Carvalho, J. F. C., do Nascimento, L. C., Neumaier, N., Farias, J. R. B., et al. (2012). Subtractive libraries for prospecting differentially expressed genes in the soybean under water deficit. *Genetics and Molecular Biology*, *35*, 304–314.

Roesler, K., Shen, B., Bermudez, E., Li, C. J., Hunt, J., Damude, H. G., et al. (2016). An improved variant of soybean type 1 diacylglycerol acyltransferase increases the oil content and decreases the soluble carbohydrate content of soybeans. *Plant Physiology*, *171*, 878–893.

Rotundo, J. L., Miller-Garvin, J. E., & Naeve, S. L. (2016). Regional and temporal variation in soybean seed protein and oil across the United States. *Crop Science*, *56*, 797–808.

Rotundo, J. L., & Westgate, M. E. (2009). Meta-analysis of environmental effects on soybean seed composition. *Field Crops Research*, *110*, 147–156.

Russell, D. A., Spatola, L. A., Dian, T., Paradkar, V. M., Dufield, D. R., Carroll, J. A., et al. (2005). Host limits to accurate human growth hormone production in multiple plant systems. *Biotechnology and Bioengineering*, *89*, 775–782.

Sadak, M. S., El-Hameid, A., Asmaa, R., Zaki, F. S. A., Dawood, M. G., & El-Awadi, M. E. (2020). Physiological and biochemical responses of soybean (*Glycine max* L.) to cysteine application under sea salt stress. *Bulletin of the National Research Centre*, *44*, 1–10.

Saio, K., & Watanabe, T. (1978). Differences in functional properties of 7S and 11S soybean proteins. *Journal of Texture Studies*, *9*, 135–157.

Samarah, N. H., Mullen, R. E., Cianzio, S. R., & Scott, P. (2006). Dehydrin-like proteins in soybean seeds in response to drought stress during seed filling. *Crop Science*, *46*, 2141–2150.

Schmidt, M. A., Barbazuk, W. B., Sandford, M., May, G., Song, Z., Zhou, W., et al. (2011). Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiology*, *156*, 330–345.

Schmutz, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*, 178–183.

Sebolt, A. M., Shoemaker, R. C., & Diers, B. W. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science*, *40*, 1438–1444.

Settlage, S. B., Kwanyuen, P., & Wilson, R. F. (1998). Relation between diacylglycerol acyltransferase activity and oil concentration in soybean. *Journal of the American Oil Chemists' Society*, *75*, 775–781.

Sharon, N., & Lis, H. (2002). How proteins bind carbohydrates: Lessons from legume lectins. *Journal of Agricultural and Food Chemistry*, *50*, 6586–6591.

Shimomura, M., Kanamori, H., Komatsu, S., Namiki, N., Mukai, Y., Kurita, K., et al. (2015). The Glycine max cv. Enrei genome for improvement of Japanese soybean cultivars. *International Journal of Genomics*, *2015*. https://doi.org/10.1155/2015/358127.

Silva, E., Perez Da Graça, J., Porto, C., Martin Do Prado, R., Nunes, E., Corrêa Marcelino-Guimarães, F., et al. (2021). Untargeted metabolomics analysis by UHPLC-MS/MS of soybean plant in a compatible response to Phakopsora pachyrhizi infection. *Metabolites*, *11*.

Singh, S. K., Barnaby, J. Y., Reddy, V. R., & Sicher, R. C. (2016). Varying response of the concentration and yield of soybean seed mineral elements, carbohydrates, organic acids, amino acids, protein, and oil to phosphorus starvation and CO2 enrichment. *Frontiers in Plant Science*, *7*.

Song, S. K., Hou, W. S., Godo, I., Wu, C. X., Yu, Y., Matityahu, I., et al. (2013). Soybean seeds expressing feedback-insensitive cystathionine -synthase exhibit a higher content of methionine. *Journal of Experimental Botany*, *64*, 1917–1926.

Song, Q. X., Li, Q. T., Liu, Y. F., Zhang, F. X., Ma, B., Zhang, W. K., et al. (2013). Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic Arabidopsis plants. *Journal of Experimental Botany*, *64*, 4329–4341.

Song, W., Yang, R., Wu, T., Wu, C., Sun, S., Zhang, S., et al. (2016). Analyzing the effects of climate factors on soybean protein, oil contents, and composition by extensive and high-density sampling in China. *Journal of Agricultural and Food Chemistry*, *64*, 4121–4130.

SoyStats. (2022). *SoyStats* (p. 2021). St. Louis, Missouri: American Soybean Association (ASA).

Staniak, M., Stepien-Warda, A., Czopek, K., Kocira, A., & Baca, E. (2021). Seeds quality and quantity of soybean [*Glycine max* (L.) Merr.] cultivars in response to cold stress. *Agronomy-Basel*, *11*.

Sun, Q., Xue, J., Lin, L., Liu, D., Wu, J., Jiang, J., et al. (2018). Overexpression of soybean transcription factors GmDof4 and GmDof11 significantly increase the oleic acid content in seed of *Brassica napus* L. *Agronomy*, *8*, 222.

Tacarindua, C. R. P., Shiraiwa, T., Homma, K., Kumagai, E., & Sameshima, R. (2012). The response of soybean seed growth characteristics to increased temperature under near-field conditions in a temperature gradient chamber. *Field Crops Research*, *131*, 26–31.

Turner, N., Davies, S. L., Plummer, J., & Siddique, K. (2005). Seed filling in grain legumes under water deficits, with emphasis on chickpeas. *Advances in Agronomy*, *87*, 211–250.

USDA [N.A.S.S.]. (2022). *USDA, National Agricultural Statistics Service*.

Utsumi, S., Damodaran, S., & Kinsella, J. E. (1984). Heat-induced interactions between soybean proteins: Preferential association of 11S basic subunits and β subunits of 7S. *Journal of Agricultural and Food Chemistry*, *32*, 1406–1412.

Van, K., & McHale, L. K. (2017). Meta-analyses of QTLs associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] seed. *International Journal of Molecular Sciences*, *18*, 1180.

VanToai, T. T., Lee, J. D., Goulart, P. F. P., Shannon, J. G., Alves, J. D., Nguyen, H. T., et al. (2012). Soybean (*Glycine max* L. Merr.) seed composition response to soil flooding stress. *Journal of Food, Agriculture and Environment*, *10*, 795–801.

Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., & Li, Z. (2014). The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes, Genomes, Genetics*, *4*, 2283–2294.

Wan, C., Shao, G., Chen, Y., & Yan, S. (2002). Relationship between salt tolerance and chemical quality of soybean under salt stress. *Chinese Journal of Oil Crop Sciences*, *24*, 67–72.

Wan, T. F., Shao, G. H., Shan, X. C., Zeng, N. Y., & Lam, H. M. (2006). Correlation between AS1 gene expression and seed protein contents in different soybean (*Glycine max* [L.] Merr.) cultivars. *Plant Biology*, *8*, 271–276.

Wang, J., Mao, L., Zeng, Z., Yu, X., Lian, J., Feng, J., et al. (2021). Genetic mapping high protein content QTL from soybean 'Nanxiadou 25' and candidate gene analysis. *BMC Plant Biology*, *21*, 1–13.

Wang, H. W., Zhang, B., Hao, Y. J., Huang, J., Tian, A. G., Liao, Y., et al. (2007). The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic Arabidopsis plants. *The Plant Journal*, *52*, 716–729.

Wang, S. D., 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*, 1776–1786.

Wijewardana, C., Reddy, K. R., & Bellaloui, N. (2019). Soybean seed physiology, quality, and chemical composition under soil moisture stress. *Food Chemistry*, *278*, 92–100.

Wilcox, J. R. (1998). Increasing seed protein in soybean with eight cycles of recurrent selection. *Crop Science*, *38*, 1536–1540.

Wilson, R. F. (2004). Seed composition. In *Vol. 16. Soybeans Improv. Prod. uses* (pp. 621–677).

Wu, N., Lu, Q., Wang, P. W., Zhang, Q., Zhang, J., Qu, J., et al. (2020). Construction and analysis of GmFAD2-1A and GmFAD2-2A soybean fatty acid desaturase mutants based on CRISPR/Cas9 technology. *International Journal of Molecular Sciences*, *21*.

Xu, Y., Yan, F., Liu, Y. J., Wang, Y., Gao, H., Zhao, S. H., et al. (2021). Quantitative proteomic and lipidomics analyses of high oil content GmDGAT1-2 transgenic soybean illustrate the regulatory mechanism of lipoxygenase and oleosin. *Plant Cell Reports*, *40*, 2303–2323.

Yang, J., Xing, G. J., Niu, L., He, H. L., Guo, D. Q., Du, Q., et al. (2018). Improved oil quality in transgenic soybean seeds by RNAi-mediated knockdown of GmFAD2-1B. *Transgenic Research*, *27*, 155–166.

Yang, A., Yu, X., Zheng, A., & James, A. T. (2016). Rebalance between 7S and 11S globulins in soybean seeds of differing protein content and 11SA4. *Food Chemistry*, *210*, 148–155.

Yao, Y., You, Q., Duan, G., Ren, J., Chu, S., Zhao, J., et al. (2020). Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. *BMC Plant Biology*, *20*, 1–13.

Yu, Y., Sheng Hou, W., Hacham, Y., Sun, S., Xiang Wu, C., Matityahu, I., et al. (2018). Constitutive expression of feedback-insensitive cystathionine γ-synthase increases methionine levels in soybean leaves and seeds. *Journal of Integrative Agriculture*, *17*, 54–62.

Yung, W. S., Wang, Q. W., Huang, M. K., Wong, F. L., Liu, A. L., Ng, M. S., et al. (2022). Priming-induced alterations in histone modifications modulate transcriptional responses in soybean under salt stress. *The Plant Journal*.

Zhai, Z., Liu, H., Xu, C., & Shanklin, J. (2017). Sugar potentiation of fatty acid and triacylglycerol accumulation. *Plant Physiology*, *175*, 696–707.

Zhang, H., Goettel, W., Song, Q., Jiang, H., Hu, Z., Wang, M. L., et al. (2020). Dual use and selection of GmSWEET39 for oil and protein improvement in soybean. *PLoS Genetics*, *16*.

Zhang, Y. Q., Lu, X., Zhao, F. Y., Li, Q. T., Niu, S. L., Wei, W., et al. (2016). Soybean GmDREBL increases lipid content in seeds of transgenic Arabidopsis. *Scientific Reports*, *6*.

Zhang, J., et al. (2014). Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of *Chlorella ellipsoidea*. *Biotechnology for Biofuels*, *7*, 128.

Zhang, D. J., et al. (2018). Elevation of soybean seed oil content through selection for seed coat shininess. *Nature Plants*, *4*, 30–35.

Zhang, D., et al. (2019). Artificial selection on GmOLEO1 contributes to the increase in seed oil during soybean domestication. *PLoS Genetics*, *15*.

Zhao, J. Z., Bi, R. R., Li, S. X., Zhou, D., Bai, Y., Jing, G. Q., et al. (2019). Genome-wide analysis and functional characterization of Acyl-CoA:diacylglycerol acyltransferase from soybean identify GmDGAT1A and 1B roles in oil synthesis in Arabidopsis seeds. *Journal of Plant Physiology*, *242*.

Zhao, X., Chang, H., Feng, L., Jing, Y., Teng, W. L., Qiu, L. J., et al. (2019). Genome-wide association mapping and candidate gene analysis for saturated fatty acid content in soybean seed. *Plant Breeding*, *138*, 588–598.

Zhao, Y., et al. (2021). Enhanced production of seed oil with improved fatty acid composition by overexpressing NAD(+)-dependent glycerol-3-phosphate dehydrogenase in soybean. *Journal of Integrative Plant Biology*, *63*, 1036–1053.

Zhou, Z., Lakhssassi, N., Knizia, D., Cullen, M. A., El Baz, A., Embaby, M. G., et al. (2021). Genome-wide identification and analysis of soybean acyl-ACP thioesterase gene family reveals the role of GmFAT to improve fatty acid composition in soybean seed. *Theoretical and Applied Genetics*, *134*, 3611–3623.

Zhou, Z. K., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, *33*, 408–U125.