We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Food Grade Soybean Breeding, Current Status and Future Directions

Souframanien Jegadeesan and Kangfu Yu

Abstract

Soybeans possess average 20% oil and 40% protein content and are a major source of protein and fatty acids in human and animal nutrition. Soybean cultivars are classified as commodity type, which are used for edible or industrial oil and animal feed, and food-type, which are used for human consumption in fermented foods and non-fermented foods. Major breeding targets for food grade soybeans are high protein and sucrose content. Developing cultivars with desired seed size and appearance depends on the type of soyfood for which the soybeans are destined. Seed with high protein content (>45%), low oil content, high sucrose, and low oligosaccharide content are suitable for making soymilk and tofu. For soyfood such as natto, soybean seed with a high content of carbohydrates are preferred. Since, molecular markers linked to the target food traits have been developed, transfer of the food grade traits among soybean varieties is possible through marker-assisted selection (MAS) to track the target gene/QTLs. Introgression of wild soybean alleles through genomics assisted breeding (e.g., GWAS, haplotype blocks, NIL, etc.), high-throughput phenotyping, mutagenesis and genome engineering/editing would improve protein without yield drag, pleiotropic effects, and background/allelic effects in breeding food grade soybean.

Keywords: soybean, soyfood, selection, mar-assisted selection

1. Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most important crops in the world. Soybean seeds are distinguished from other legume grains especially by their high protein content. It is a major source of vegetable protein and oil in human food industries because of its nutritive and health benefits [1]. Owing to its health benefits, soy food has been part of the Asian diet for several centuries [2]. Soyfood has nutritional qualities that reduce human blood serum cholesterol levels and lower the risk of cardiovascular diseases [3]. Soy food is a natural source of isoflavones (daidzein, genistein, and glycitein) and a good source of calcium [4].

Soybeans typically possess protein and oil contents of approximately 40 and 20%, respectively. This composition gives the possibility for a broad variety of applications such as feed, biodiesel, edible oils, and other food products. Commercially, soybean can be categorized as (i) commodity type, mainly used for oil and animal feed, and (ii) food type soybean, mainly used for human consumption. One important application of soybean is its use in animal feed. Approximately 85%

of the soybean is crushed into oil and meal. The greatest part of the meal is used as protein source in animal feed, Approximately 6% of the produced soybean is directly used for food purposes [5]. When soybean is exclusively used for consumption by the farm animals, nutritionally, it should have at least 36% protein to meet part of their daily requirement for protein. The food-type soybean, on the other hand, contains protein higher than feed-grade soybean; it usually ranges between 40 and 45%. It helps in making better quality soy-based food products. Usually, food-grade soybeans are used for production of soy milk, tofu or soy paneer, soy sprouts, and other soy-based food items.

Globally soybean meal accounts for 63% of all the protein sources in animal feed. Next to the soybean meal, soybean oil obtained after crushing is refined and used as cooking oils and salad oils, and they can also find their use in mayonnaise and butter substitute. Soybean oil is refined and lecithin is obtained as a main by-product. Lecithins are found in many products such as in chocolate, margarine, and emulsifying agents [6]. Furthermore, the unraveling of the health and nutritional benefits of soybean contributed to increased interest for soybeans destined for food production, the so-called food-grade soybeans. Hence, the breeding of cultivars with enhanced food-grade traits is gaining great importance [7]. Soy foods available in the market can be divided in traditional and non-traditional soy foods. Traditional soy foods includes edamame (green vegetable soybeans), soy milk (a drink produced by soaking and grinding soybeans, boiling the mixture, and filtering out remaining particulates), tofu (soybean curd), tempeh (made of whole cooked soybeans), miso (fermented soup-based paste), soy sauce, okara, natto (fermented whole soybeans), and soy sprouts. They originated from Asia where soybean has been grown for centuries before its introduction to the rest of the world. Non-traditional foods are, for instance, soy yoghurt, soy cheese, pudding, snacks, etc. [6]. Soy food available in the market is classified into two groups based on the soybean seed size. Soy foods made of large seeds (>20 g/100 seeds) include tofu, edamame, miso, and soy milk. Soy foods made of small seeds (<12 g/100 seeds) include natto, soy sauce, tempeh, and bean sprouts.

Soybean seed needs to satisfy specific physical and chemical requirements for soy food production. In addition to seed size, visual appearance such as uniformity of seed size and shape with light-colored hilum and yellow seed coat without physical damage such as mottling, splits, shriveling, purple stain, and insect damage are the main consideration of food-grade soybeans [8]. It is reported that the seed size uniformity affected water absorption and the quality of the final soy product [9]. Shrunken or discolored seeds were undesirable due to the consumer requirements [10]. Stone seeds that do not absorb water during soaking cause serious problems for food processing as it affects the texture and consistency of the soy products particularly for fermented soy food such as natto [11]. Seeds with harder texture have higher calcium content and absorb less water [12]. Seed's hardness could be estimated by the seed swell ratio that was related to seed weight or water volume change before and after soaking [13]. High water absorption was required to obtain soft steamed seeds [14]. Thus, soybean seeds with rapid and high water uptake are preferred by processors in order to provide more products per unit of time [9]. However, seed composition requirements vary according to the type of soy food. Soybean seeds with high protein content (>45%), low oil content, high sucrose content, and low oligosaccharides (raffinose and stachyose) content are suitable for making tofu. For soy food such as natto made through short fermentation process, soybean seed with a high carbohydrate content are preferred for the purpose of getting a quick conversion to simple sugars [10].

2. Anti-nutritional factors

Like other plants, soybean too synthesizes a range of secondary metabolites for their adaptation and self-protection, called anti-nutritional factor (ANF). From nutritional point of view, the ANF is considered as harmful and toxic, as it interferes with normal growth, reproduction, and health. Therefore, soybean, in general, and food-grade soybeans, in particular, are expected to be free from anti-nutritional factors or allergens that may cause harm to the consumers. Soybean contains several anti-nutritional factors among which trypsin inhibitor (TI) and phytase are the most important ones. Kunitz trypsin inhibitor (KTI), which constitutes more than 80% of the total TI, is primarily responsible for improper digestibility of soybean, if consumed unprocessed. Although KTI is heat labile, yet heat inactivation process is neither fully effective nor economic. Genetic elimination of KTI is the most effective way of making soybean free from KTI [15].

2.1 Phytase

Soybean seeds contain phosphorus in the form of inositol hexaphosphate commonly known as phytate [16]. Besides sequestering inorganic phosphate, phytate may also chelate divalent cations such as Fe, thereby decreasing their availability. It also contributes toward water pollution by eutrophication as the phosphate-rich waste discharges into water bodies. So, reduction of seed phytase can enhance mineral and protein bioavailability in soybean. Microarray-based gene expression profiling of phytic acid biosynthesis pathway indicated stepwise regulation of eight genes, viz. myo-inositol-3-phosphatesynthaes (MIPS), inositol phosphate kinase (IPK1–4), etc. Gene silencing constructs were used to silence *GmIKP1* and *GmIPK2* through seed-specific vicilin/conglycinin promoters [17]. Similarly, mutation breeding through gamma irradiation has also been attempted to develop plants with reduced or zero phytase.

3. Cadmium content

Vast areas of agricultural soils are contaminated with Cadmium (Cd) through the use of super phosphate fertilizers, sewage sludge, and inputs from the mining and smelting industries [18]. Cadmium (Cd) is a highly toxic element for human beings because of its extremely long biological half-life. Soybeans grown in cadmium-contaminated soil take up cadmium by roots and translocate into aerial organs, where it affects photosynthesis and consequently root and shoot growth. Many soybean cultivars can accumulate high Cd concentration in seed when grown on Cd-polluted soil [19, 20]. Consumption of food containing excessive Cd leads to a risk of chronic toxicity. In humans, it can damage kidneys, causing a loss of calcium and associated osteoporosis [21]. To reduce the health risk, it is desirable to limit the concentration of Cd in crops used for human consumption. Due to growing concern about safety of foods and human health, the Codex Alimentarius Commission of Food and Agriculture Organization/World Health Organization (FAO/WHO) has proposed an upper limit of 0.2 mg kg^{-1} for Cd concentration in soybean grain [22]. However, a large-scale survey of agricultural products revealed that the Cd concentration of 16.7% of soybean seeds exceeded the international allowable limit of 0.2 mg kg^{-1}, which is much higher than that of other upland crops [23]. Cultivars with reduced uptake of Cd are needed for

human consumption. Cd uptake depends both on the Cd concentration in the soil and on the characteristics of the specific cultivars. Breeding cultivar with reduced Cd is an attractive method for changing the element profile of crops as the benefit will persist in the seed that can reduce the requirement for other management practices [24].

4. Nutritional factors

4.1 Soybean protein

Soybean contains about 40% protein and is noteworthy as it is the most complete vegetable protein [25]. Concretely, with exception of sulfur-containing amino acids such as methionine, the amino acid pattern of soybean resembles the pattern derived from high-quality animal protein sources [25]. In fact, soybean protein can even enhance the nutritional quality of other vegetable protein. Protein sources that are deficient in some amino acids can be complemented by soybean. Soybean is rich in lysine, tryptophan, threonine, isoleucine, and valine and therefore complements well with cereal grains that are deficient in those amino acids [26]. By ultracentrifugation studies, four different fractions have been revealed, with approximate Svedberg coefficients of 2S, 7S, 11S, and 15S [6]. The 2S fraction contains from 8 to 22% of the extractable soybean protein. It consists of several enzymes, including the trypsin inhibitors, Bowman-Birk and Kunitz inhibitors [6]. Trypsin inhibitors inhibit the protein-cleavage effect of proteases (such as trypsin) affecting the digestibility and leading to growth depression in animals. Therefore, soybean meal needs first to be heated in order to inactivate the trypsin inhibitors. However, trypsin inhibitors have been found to be powerful anti-carcinogenic agents in humans and therefore they can be considered as functional components of soybeans [27].

More than 70% of the soybean seed storage protein is composed of 7S β-conglycinin and 11S glycinin. The 7S fraction makes up 35% of the extractable soybean protein. The quantity and the quality of the protein in the seed are the major biochemical components influencing the quality of tofu and other soy food products [28]. The mean glycinin to β-conglycinin protein ratio is known to influence the protein quality of soybeans, and greatly affects the functional properties of food products made from soybeans [29, 30]. Glycinin and β-conglycinin also differ in amino acid composition, with glycinin being higher in sulfur (S), containing amino acids that account for 3–4.5% of the total amino acid residues [31]. G1, G2, and G7 glycinin subunits contain a higher amount of methionine (6–7 per subunit) compared to G3, G4, and G5 glycinin subunits, which contain 5, 2, and 4 methionine residues per subunit, respectively [31]. By comparison, β-conglycinin is devoid of methionine [32, 33] and β-conglycinin contains a major allergen in its subunit [34]. Increased glycinin content in soybean protein is an important trait for increasing the concentration of the S-containing amino acids [35]. Because glycinin and β-conglycinin have a great impact on the nutritional value and quality of soybean products, these two storage proteins have been extensively studied and targeted for genetic manipulation in breeding programs. Soybean mutant genotypes differing in seed storage glycinin and β-conglycinin subunit composition were developed and tested for their effects on tofu quality [30]. It was shown that group IIb (A_3) glycinin played the major role in contributing to tofu firmness with any coagulant, while the group IIa (A_4) subunit could have a negative effect on tofu quality. Yu et al. [36] reported that soybean cultivars with 7S α′ and 11 S a4 nulls always make firm tofu than the check cultivar Harovinton. The hardness of gels from glycinin

decreased in the order of group IIa, IIb, and I [37, 38]. Protein subunit composition also affects the quality and stability of soymilk [39].

Other soybean seed proteins include lipoxygenase and lectins. The lipoxygenase enzyme constitutes about 1–2% of the soybean protein. The lipoxygenase enzyme generates a grassy-beany flavor when it oxidizes fats and is not preferred by consumers in some countries. It is possible to avoid the oxidation of the fats by heat inactivation of the lipoxygenase enzyme; however, this is cost-ineffective and leads to insolubilization of proteins. Therefore, the genetic elimination of the lipoxygenase is preferred in order to reduce the beany flavor. Genotypic variation and the influence of growing environment on lipoxygenase accumulation in soybean seed are well documented in the literature [2, 26]. Lipoxygenase 1, 2, and 3 null germplasm lines were developed and showed that the grassy-beany flavor was eliminated [40]. Triple-null soybeans can be used for edible soy products, such as soymilk and tofu [40]. Similarly, saponins and isoflavones may also be the cause of undesirable taste in soy products although this is not well documented yet. The breeding of cultivars with low isoflavones and saponins is possible [2]. The 11S fraction comprises 31–52% of the extractable soybean proteins [6]. The 11S fraction is responsible for the gelling character of tofu, and hence, the proportion of this fraction compared to 7S plays an important role in tofu firmness [2]. The 15 S fraction comprises about 5% of the total extractable protein. It is only poorly characterized and is thought to be composed of polymers of the other soybean proteins [6].

4.2 Carbohydrates

Dry soybeans contain on average 35% of carbohydrates, which can be divided into soluble and insoluble carbohydrates [27]. Soybean seeds possess 15–20 different soluble carbohydrates that makes up approximately 15–25% of dry weight [41]. Sucrose, raffinose, and stachyose are the most relevant soluble carbohydrates for breeding of food-grade soybean. Sucrose in dry soybean seeds is found in contents of typically 5.5% [27]. Sucrose is important for improving taste in soybean-based products. The oligosaccharides raffinose and stachyose typically constitute about 0.9 and 3.5% of dry soybean seeds, respectively [27]. The seed coat of soybeans contains a major part of insoluble carbohydrates such as cellulose, hemicellulose, pectin, and a trace amount of starch [27]. Consumers, especially in countries where fermented and vegetable soybean are not in vogue, may be skeptical toward the use of soy products because of flatulence and poor digestibility. These effects are caused by oligosaccharides, stachyose and raffinose. Humans and monogastric animals do not possess the enzyme called α-galactosidase necessary for hydrolyzing the linkages present in these oligosaccharides, so they cannot be digested when consumed. Intact oligosaccharides reach the lower intestine and undergo anaerobic fermentation by bacteria with gas expulsion (H_2, CO_2, and traces of CH_4), causing the flatus effect and sometimes diarrhea and abdominal pain. Although raffinose and stachyose can be reduced to an extent by soaking or boiling, genetic reduction is one of the prime plant breeding objectives.

4.3 Soybean oil

The major components of crude soybean oil are triglycerides. After refinement of the oil, soybean oil is composed of 99% of triglycerides. Triglycerides are neutral lipids composed of one glycerol linking three fatty acids [27]. The saturated fatty acids in soybean oil are palmitic acid (16:0) and stearic acid (18:0), with average concentrations of about 11 and 4% (relative to the oil), respectively, and they are useful in making low trans-fat margarines. Soybean oil contains an average of

22% monounsaturated fatty acid, oleic acid (18:1). Monounsaturated fatty acids are healthy and have good oil stability [42]. Soybean oil possesses the two polyunsaturated fatty acids: linoleic acid (18:2), an omega-6 fatty acid, and linolenic acid (18:3), an omega-3 fatty acid [26]. They can be found in average concentrations of 53 and 8% of the oil, for linoleic and linolenic acid, respectively. Low (reduced) linolenic soybeans have half the linolenic acid level of standard soybeans, which reduces the need for hydrogenation, a process used in converting vegetable oils to margarine that results in the production of unhealthy trans fatty acids.

Soybean crude oil is also shown to consist of phospholipids, unsaponifiable material, free fatty acids, and metals. Unsaponifiable material consists of tocopherols, phytosterols, and hydrocarbons [27]. Tocopherols and phytosterols are considered as functional components. Soybean oil provides an additional benefit due to presence of enriched amounts of α-tocopherol or natural vitamin E. Oils containing low contents of linolenic acid (18:3) have been shown to contain high amount of α-tocopherol and results in lowered amount of ϒ-tocopherol [5].

4.4 Vitamins and minerals

Soybeans contain water-soluble and oil-soluble vitamins. The water-soluble vitamins such as vitamin B1 (thiamin), vitamin B2 (riboflavin), vitamin B5 (pantothenic acid), and vitamin B6 (niacin) and the oil-soluble vitamins vitamin A and vitamin E (tocopherols) are present in soybean. Vitamin A mainly exists in the form of β-carotene in immature and germinated seeds, whereas it is present in negligible amount in mature seeds [27]. Most of the minerals are found in the meal fraction rather than in the soybean oil fraction. Dry soybean seeds contain on an average concentration ranging from 0.2 to 2.1% major minerals such as potassium, which is present in the highest concentration followed by phosphorus, magnesium, sulfur, calcium, chloride, and sodium [27]. Minor minerals found in soybeans include silicon, iron, zinc, manganese, copper, molybdenum, fluorine, chromium, selenium, cobalt, cadmium, lead, arsenic, mercury, and iodine [27].

4.5 Functional components

Functional components of soybeans include isoflavones, saponins, lecithin, trypsin inhibitors, lectins, oligosaccharides, tocopherols, and phytosterols [27]. Presence of such biological ingredient creates interest to consider soybean food products as functional foods, i.e., foods that contain biological components that deliver special health benefits, e.g., anticancer, hypocholesteromic, and antioxidative effects to the consumer [26]. Isoflavones are phytoestrogens and are known to have positive health effects such as the reduction of the risks for coronary heart disease, osteoporosis, certain types of cancer, and the moderation of postmenopausal symptoms in women [43]. Soybean possesses 0.1–0.4% of isoflavones on a dry weight basis; hence, soybean possesses the highest amount of isoflavones compared to all other crops [27]. The isoflavone concentration varies considerably depending upon the genotype and environmental conditions. It is thought that isoflavones are mainly responsible for most of the health benefits from soybean-based foods. Therefore, they gained more and more attention from the scientific world [27], and research on breeding for enhanced isoflavone content is increasing. Refined soybean oil possesses about 1000–2000 mg/kg. Tocopherol exists in four isomers, three of them being α-, ϒ-, and δ-isomers that are present in soybean oil. α-tocopherol (natural vitamin E) in soybean is the leading commercial source of this vitamin. Tocopherols protect the polyunsaturated fatty acids from oxidation; hence, they are antioxidants and used in pharmaceutical applications [42].

5. Soyfoods

5.1 Tofu

Soybeans with large seed size and high protein levels are primarily used for soymilk and tofu production. Other traditional food from soybean includes tempeh, miso, soy sauce, okara, soynuts, soy milk, yoghurt, meat, and cheese alternatives. Tofu is perhaps the most widely consumed soy food in the world. Tofu is naturally processed and it retains a good amount of nutrients and phytochemicals such as the isoflavones [5]. Tofu typically contains 7.8% protein and 4.2% lipid on a wet basis [5]. It has a relatively low carbohydrate and fiber content, making it easier to digest. There are two main types of tofu: silken, or soft tofu and hard tofu. They are made by soaking whole soybeans and grinding them into a slurry with water. The slurry is cooked to form soymilk and a coagulant is added. The most commonly used coagulants are magnesium chloride, calcium sulfate, or glucono-D-lactone; the coagulants can be used purely or in combinations to achieve different flavor or textural characteristics. Heating is also usually applied in order to facilitate the coagulation. The result of the coagulation is that after a few minutes, the soymilk begins to curdle and large white clouds of tofu curd are formed. The water in the curds are then removed and placing the tofu curd in cloth-lined forming boxes where pressure is applied from the top results in the formation of hard tofu. Silken tofu in comparison to hard tofu is not pressed and is often coagulated in the container in which it is to be sold. [2].

5.2 Soymilk

The popularity of soymilk has expanded from Asia to the U.S. and Europe since the 1980s. Traditionally, it is made from whole beans in the same way as the first few steps of tofu manufacture. This soy milk contains nutrients, saponins, isoflavones, and other soluble components of the soybean from which the soy milk is made. Some manufacturers add isoflavones back into the soy milk in order to make health claims about the product. Additionally, soymilks are also fortified with vitamins and minerals, such as β-carotene and calcium or docosahexaenoic acid (DHA), an omega-3 fatty acid [2]. However, beverage-quality soy milks available in the market are usually prepared from soy protein isolate, to which sugars, fats, and carbohydrates are added to improve flavor and generate a nutritional profile similar to that of cow's milk [2].

5.3 Vegetable soybeans (edamame, mukimame)

Vegetable soybean consists of the whole soybean picked at the R6-R7 stage and seeds are bigger and sweeter. At this stage, the soybean has a firm texture, contains a high level of sucrose, chlorophyll, and is at its peak of green maturity. The harvested pod can be left entire or be shucked into individual beans. After being blanched and frozen, the soybean can be sold as "edamame," referring to the entire pod, or "mukimame," referring to individual beans [2]. Nutritionally, it is highly rich in protein (11–16%), monounsaturated fatty acid, vitamin C, fiber, iron, zinc, calcium, phosphorous, folate, magnesium, potassium, tocopherol, and anticancer isoflavones [44]. It also has a pleasant flavor and soft texture and is easier to cook. Cooked vegetable soybean has the highest net protein utilization value (NPU: ratio of amino acid converted to protein) among all soy products. Vegetable soybean also has 60% more calcium and twice the phosphorus and potassium levels of green peas, which is India's most commonly consumed fresh legume (https://www.gov.uk/government/case-studies/dfid). The vegetable soybean, in general,

carries a flavor, called "beany flavor" or "grassy flavor." Genotypes with high levels of sucrose, aspartic acid, glutamic acid, and alanine are found to have acceptable taste [44]. Biochemical analysis has established that production of "beany flavor" in soybean or soy-based products is primarily due to the lipoxygenase or the oxidative rancidity of unsaturated fatty acids [45]. Plant lipids are sequentially degraded into volatile and nonvolatile compounds by a series of enzymes via the lipoxygenase pathway, which catalyzes the hydroperoxidation of polyunsaturated fatty acids to form the aldehyde and alcohols that are responsible for the grassy-beany flavor [46]. *Organic food-grade* soybeans are produced using cultivation practices that do not use synthetic compounds. In the U.S., growers producing and selling soybeans that are labeled "organic" must be certified by a USDA-approved state or private agency. The top selling organically produced soy products in the US are tofu and soymilk. Other specialty soybeans include varieties with low saturated fat, high isoflavone, high sucrose, high oleic acid, high stearate, or high protein. Large-seeded soybeans with thin seed coat and a clear hilum are preferred for the soynut market, while small- to medium-sized seeds are preferred for sprouts.

5.4 Breeding targets for food grade soybeans

Breeding for food-grade soybeans with unique seed composition has focused on a specific nutritional trait of the soybean seed. Examples of such varieties are given according to the fraction from which the targeted trait origins. Food-grade soybean that targets a specific trait such as varieties high in total protein content, high in β-conglycinin, low in lipoxygenase, high in specific amino acids such as lysine, methionine, and threonine, and low in allergenic proteins [13]. High-protein soybeans (>43%) are used for tofu, soymilk, soy sauce, beverages, baked goods, pudding, cheese, and meat analogs. The breeding of food-grade soybeans can be classified into three major categories: the breeding of large-seeded soybeans, the breeding of small-seeded soybeans, and the breeding of soybean with unique seed composition [13].

5.5 Large-seeded soybeans

By targeting specific traits, soybean breeders try to develop soybeans with good yield and quality [5]. Large-seeded soybeans are bred for tofu, soymilk, miso, edamame, and soynuts [13]. An important factor for the breeding of tofu soybeans is the tofu yield, which is defined as the weight of fresh tofu produced from a unit of harvested soybean. Seed size and seed appearance are also of importance for tofu soybeans. Tofu soybeans are larger than 20 g/100-seeds [13]. It is possible to produce good quality tofu with dark hilum beans but this requires prior dehulling of the beans and careful soymilk filtration [5]. In order to avoid these additional processing steps, soybeans with a yellow cotyledon, yellow seed coat, and clear hilum are preferred. Moreover, a thin but strong seed coat that is free from cracking and discoloration is desirable [13]. Soybean seeds with high protein content exceeding 45% on dry matter basis and improved ratio of 11S/7S is desirable for tofu soybeans as this enhances tofu yield and gelling characteristics, respectively [5]. A high protein/oil ratio provides a higher tofu yield and firmer texture; therefore, low oil content is preferred. Moreover, tofu soybeans should have high water uptake, a low calcium content, and a high germination rate. The carbohydrate content and composition influence the taste of tofu and soymilk [13]. High total sugar content (above 8% on dry matter basis) [5], high sucrose, low raffinose, and low stachyose are highly desirable for tofu and soymilk [13]. Examples of tofu and soymilk varieties: Black Kato, Toyopro, Grande, Proto (from Minnesota), Vinton-81, HP 204, IA1007, IA1008 (from Iowa), and Harovinton [13, 47].

5.5.1 Vegetable soybean

Vegetable soybean varieties should meet certain requirements such as sweeter seeds with thin seed coat and large seed size (>30 g/100-seeds dry weight) [13]. As the pods are eaten directly, genotypes with sparse gray pubescence with green and thin seed coat are preferred [13]. Moreover, edamame cultivars should possess as less as possible of one-seeded pods as they require greater effort to shell by consumers. Those cultivars with genetically "stay green" and delayed yellowing toward maturity make it possible for growers to have extended harvest period closer to maturity. Vegetable-type soybean should possess important nutritional traits such as high content of sugar (sucrose and maltose) and free amino acids to impart sweet and delicious taste. Sucrose is primarily responsible for the sweetness of vegetable soybeans, where sucrose content is preferably higher than 10% on dry matter basis. Certain free amino acids, such as glutamic acids, are major contributors to the taste of vegetable soybeans [13].

5.6 Small-seeded soybeans

5.6.1 Natto

Natto beans are small-seeded soybeans typically used for the fermented soy foods popular in Japan. For natto, small to ultrasmall soybeans (smaller than 9 g/100-seeds [2] of maximum of 5.5 mm diameter are preferred for better fermentation. The seeds have preferably a near spherical shape as this reduces the ratio of the tough seed coat to softer cotyledon [2]. Also, clear hilum and thin seed coat are desirable traits for natto soybeans. Natto soybeans are nutritionally characterized by a high carbohydrate content [13]. A high content of soluble sugars (>10%) on a dry weight basis results in a softer natto product, an important requirement for natto [5]. The composition of sugars is important for the effectiveness of fermentation [13]. To obtain a steady and controlled fermentation, low sucrose content with high stachyose and raffinose content is favored [5]. Soybean with moderately high protein content is desirable in order to provide amino acids for fermentation. Oil content must be low, i.e., less than 18% of the dry matter as it enhances water absorption [13]. For a softer natto product, seeds must additionally possess high water absorption capacity during soaking, which is the first step of natto manufacturing. Breeders use standard small-seeded lines, such as the cultivar Vance (known for having a medium ability for water uptake), to compare selected lines for water absorption capacity [5]. Soybeans with medium seed size (10–12 g/100-seeds) and a high germination rate are preferred for bean sprouts. High-protein, high-isoflavone, high-sugar, and lipoxygenase-free soybeans are desirable for soybean sprouts [13].

It is reported that the Asian small-seeded lines exhibited high diversity indices than the U.S. lines for seed hardness, calcium content, and stone seed rate. In addition, the average genetic diversity of the U.S. small-seeded soybeans (1.48) was lower than that of Asian small-seeded soybeans (1.57), suggesting narrower genetic base in the U.S. lines. Seed uniformity, hardness, protein, and calcium content appeared to be relatively high in diversity index for both the U.S. and the Asian large-seeded lines. The U.S. small-seeded soybeans were desirable for natto production because of their softer texture with higher water absorption capacity and lower stone seed ratio. However, the Asian large-seeded soybeans had a lower stone seed ratio and a higher water absorption capacity. Therefore, using the Asian large-seeded genotypes may potentially improve seed quality for tofu and soymilk [48]. Therefore, the Asian soybean gene pool may serve as valuable genetic source for increasing protein content of the U.S. food-grade soybeans.

6. Breeding for protein content

Availability of genetic variability for soybean food-grade traits offers scope to improve through breeding. Breeding cultivated soybean varieties with high protein or high oil are an extremely important and promising objective. High protein and low oil content add nutritional value to soy foods. Germplasms that cover a wide range in protein content (33.1–55.9%) and oil content (13.6–23.6%) are available for breeders to modify the seed/oil ratio in the breeding program. The negative correlation between protein and oil facilitates the development of high protein and low oil lines. High protein content is generally associated with low yield, which makes the development of lines that combine high protein and high yield difficult. However, high yield is mostly achieved by selection for moderately high protein content (43–45%) [13]. Seed protein and oil content are two valuable quality traits controlled by multiple genes in soybean. The phenotypic range of protein content of soybean has been reported to be 34.1–56.8% of seed dry mass, and oil content ranged from 8.3 to 27.9% [49], suggesting that there is great potential for genetic improvement of soybean seed protein and oil content. The negative correlation between oil and protein content makes improvement of both traits simultaneously a challenging task using conventional breeding [50]. Therefore, the identification of molecular markers associated with quantitative trait loci (QTLs) controlling protein and oil content is a prerequisite for breaking the negative correlations between both traits [51].

In the SoyBase database, 241 QTLs for protein content and 315 QTLs for oil content were reported and found to be distributed over 20 soybean chromosomes [52]. A majority of these QTLs were mapped by linkage mapping based on bi-parental populations and limited by the relatively small phenotypic variation and by the fact that only two alleles per locus can be studied simultaneously. The broad chromosome regions of QTLs make it difficult to identify putative candidate genes of interest [53]. With the advancement of genetic map construction, the availability of a well-annotated reference genome, resources for association mapping, and whole-genome resequencing (WGRS) data, a large number of QTLs for seed protein content have been identified (**Table 1**).

Several genome-wide association studies [50, 66, 68] and QTL analysis [53, 56] have shown similar QTL genomic loci (e.g., Chrs20, 15, and 5) for protein and oil indicating negative pleiotropic effect or linkage (larger LD). The QTL on Chr20 was

Trait	Major QTL (Chr.)	References
Seed protein	20, 15, 18	Akond et al. [54]; Diers et al. [55]; Nichols et al. [56]; Brummer et al. [57]; Lee et al. [58]; Sebolt et al. [59]; Chung et al. [60]; Fasoula et al. [61]; Panthee et al. [62]; Lu et al. [63]; Pathan et al. [53]; Wang et al. [64]; Phansak et al. [65]; Hwang et al. [50]; Vaughn et al. [66]; Sonah et al. [67]; Bandillo et al. [68]; Valliyodan et al. [69]; Zhou et al. [70]; Tajuddin et al. [71]
	5, 6, 8	Hyten et al. [72]; Pathan et al. [53], Sonah et al. [67]; Bandillo et al. [68]
	10, 14	Wang et al. [64]; Phansak et al. [65]
	13, 03, 17, 12, 11	Zhou et al. [70]
Sucrose and oligosaccharides content	5, 7, 8, 11, 13, 18, 16, 19, 20	Maughan et al. [73]; Kim et al. [74]; Kim et al. [75]; Alkond et al. [76], Cicek [77]; Clevinger [78]; Maroof et al. [79]; Feng et al. [80]; Zeng et al. [81]; Wang et al. [82]

Table 1.
Major QTLs for seed protein, sucrose and oligosaccharide content reported in soybean.

most likely in the genomic region of 29.8–31.6 Mbp that was supported by integrating GWAS, transcriptome, and QTL mapping analysis (**Table 1**) [68]. It was observed that the gene order was conserved and 18 identified genes were tandemly duplicated on Chr10 and showed similar gene ontology [83]. Three putative candidate genes were identified on Chr20 and suggested that these non-duplicated genes might be related to protein content [68]. Similarly, Chr15 QTL (38.1–39.7 Mbp) showed an inversely duplicated genomic block on Chr8. The QTL on Chr15 comprises 18 putative genes, 13 of which were duplicated with similar gene function. Syntenic analysis provided a basis for divergence of QTL regions that took place during recent genome duplication and suggested the retention or loss of several genes that might be responsible for oil content and protein in soybean. In addition to pleiotropic effects of protein on oil and yield, variation in seed protein concentration significantly affects seed size, crop growth, and development [84]. High-protein genotypes showed lower leaf area and harvest index when compared with high-yielding genotypes. While high-protein small seed showed higher leaf area at the beginning of seed fill, more canopy biomass production, and low levels of assimilate per seed [84]. Therefore, breaking the undesirable genetic linkage between protein, oil, and yield related loci through repetitive recombination and random mating is necessary.

7. Breeding for 11S/7S ratio

Consumers have preference for firmer tofu texture that partly depends upon the protein composition. The genotypic variation in this trait is partly due to the ratio of 11S-to-7S protein fraction in the seed. The 11S fraction generally possesses greater gelling potential than 7S; hence, high 11S-to-7S ratio is desirable as it results into harder than those with low ratio. The 11S-to-7S ratio is reported to range from 0.3 to 4.9. However, genotypes with same 11S-to-7S ratio do not always result in the same firmness because of different 11S subunit composition. In general, a high 11S-to-7S ratio as well as suitable 11S composition is of importance for good tofu firmness.

The selection and manipulation of specific subunit composition will play a major role in the development of improved protein quality. Molecular markers linked to the various subunit of glycinin and β-conglycinin have been reported previously. PCR-based markers were reported for the identification of β-conglycinin genes [85, 86]. An RFLP marker associated with the *Scg-1* (suppressor of β-conglycinin) gene was developed by using the α-subunit gene as probe [87]. SNPs in the β subunit genes were used to map the *Scg-1* gene, and the chromosomal region associated with β-conglycinin deficiency was located on linkage group I of the soybean genetic map [86]. Hayashi et al. [88] reported AFLP markers linked to the recessive allele, cgdef, controlling the mutant line lacking 7S globulin subunits (α, α′, β). Markers linked to the glycinin genes were reported. RFLP markers were identified for both *Gy*4 and *Gy*5 and mapped in linkage group O and F on the public soybean linkage map [31, 89]. *Gy*1, *Gy*2, and g*y*6 are linked in tandem to one another on linkage group N, while *Gy*3 and *Gy*7 are linked to one another on linkage group L [90]. KASP-SNP markers linked to 7S α′ and 11S A_1, A_3, and A_4 subunits have been reported [91]. Three SSR markers (Satt461, Satt292, and Satt156) were found to be associated with glycinin QTLs that were distributed on linkage group D_2, I, and L, whereas two β-conglycinin QTL-associated SSRs (Satt461 and Satt249) were distributed on LG D_2 and J [35].

Functional markers (FMs) have advantages over the linked markers, because their polymorphic sites have been derived from the genes involved in phenotypic trait variation [92]. Glycinin genes have high degree of conservation within the subgenus *Soja*, but there are more variations within subgenus *Glycine* [93]. Despite the high degree of similarity among the subunits in Group I and Group II, gene

primer pairs specific to *Gy1*, *Gy2*, and *Gy5* were designed [93–95]. PCR primers were designed for identification of the *Gy4* null allele and demonstrated selection of soybean without the A_4 peptide can be done by null allele specific primers which overcome the drawbacks of SDS-PAGE gel-based selection for A_4 peptide [95].

8. Breeding for amino acid composition

Besides breeding for increased protein content, protein composition is important for its nutritional value. Based on solubility properties, globulins and albumins are two major components of dicot seed storage protein, and soybean primarily belongs to the globulin (~70%) family [96]. The soybean globulins (glycinin and β-conglycinin) are relatively low in sulfur-containing amino acids methionine (Met) and cysteine (Cys) as well as threonine (Thr) and lysine (Lys) [97]. Increasing the soybean storage protein content of seed along with improving the ratio of glycinin to β-conglycinin is of great potential for food grade soybean improvement [98, 99]. Therefore, besides increased protein content, enhancing sulfur containing amino acids (Met, Thr, Cys, and Lys) would improve the nutritional value. More than 70% of the essential amino acid enriched meal is used in the feed industry [97, 100]. Although soybean cultivars with improved protein content have been successfully developed, only a few studies have been conducted to identify genomic regions controlling amino acid composition. The difficulty in breeding for improved amino acids could be due to lack of genetic variability, lack of high throughput, and cost-effective phenotyping platform to screen a large number of samples for amino acids. Panthee et al. [99] identified QTL for essential amino acids in a F_6-derived recombinant inbred population. In another study, a major QTL for essential amino acids and crude protein was identified on Chr20 [97]. Moreover, negative correlations of crude protein with Lys and Thr and a positive correlation between Thr with Lys were also observed [97]. Among the essential amino acids, Met, Lys, and Thr are synthesized from a common precursor aspartate; thus, they are strongly correlated. Krishnan et al. [101] introgressed leginsulin (Cys-rich protein) and a high protein trait from an Asian soybean germplasm, PI 427138, into North American experimental line (LD00–3309). While they were successful in introgressing leginsulin and improving protein content, the overall concentration of sulfur-containing amino acids was not changed compared to parental lines.

Seed protein content and composition are dependent on the genetic background of an elite parent that plays an important role in the expression of a newly introgressed allele because of complex epistatic interactions [102]. It has been found that most of the QTLs affecting seed protein and yield and yield-related components were detectable only in one of the parental genetic backgrounds (GBs) in introgression lines of reciprocal crosses [103]. The high protein allele within a different genetic background resulted into reduced Thr and Lys content [103]. The high protein allele from Danbaekkong on Chr20 has been demonstrated to increase seed protein content in several maturity groups (III–VIII) in various genetic backgrounds with little drag on seed yield [104]. On the other hand, yield drag was observed for the protein QTL alleles on Chr20 from other sources, including wild *G. soja* [56, 60]. Hence, it is not feasible to select only the major crude protein QTL on Chr20 to improve protein quality. Improvement of protein and amino acid profiles has been limited by the narrow genetic base and genome complexity of soybean. Mutation breeding can be used to enhance the genetic variability. Mutagenized populations (physical, chemical, transposon tagging or transformation-induced mutagens) have been useful in crop improvement [105]. In soybean, mutations for seed traits, including oleic acid [106], oil [105, 107], stearic acid [108], and lipoxygenase [109] were identified using induced mutation.

9. Genomics-assisted breeding (GAB)

The integration of genomic tools and breeding practices are the core components of genomics-assisted breeding (GAB) for developing improved cultivars for any given trait. Near-isogenic lines (NILs) can be developed for major QTL (e.g., protein QTL on Chr20) by backcross breeding. Using NILs, the effect of a QTL and the phenotype it produces (i.e., protein or amino acid content) can be estimated precisely without the confounding effects of differences in genetic backgrounds. Additionally, developing NILs in a range of maturity groups is desirable to study the effect of environment and maturity on seed protein content. Marker-assisted backcrossing selection approach was utilized to produce a NIL-(cgy-2–NIL)-containing mutant cgy-2 allele, responsible for the absence of allergenic α-subunit of β-conglycinin [110]. It is also possible to incorporate multiple genes/QTL into elite lines in a cyclic forward crossing scheme and employing marker-assisted recurrent selection (MARS) as an effective approach [111, 112]. Recurrent selection was effectively utilized for increased gain yield, protein, oil, and oleic acid content [111, 113, 114]. Furthermore, the next-generation sequencing (NGS) data can be used effectively for genomic selection (GS) to identify desirable parents and progenies. Jarquin et al. [115] assessed the genomic and phenotypic data of over 9000 accessions and developed genomic predication models to evaluate the genetic value for protein, oil, and yield traits. Similarly, genomics-assisted haplotype analysis is a promising approach if the information of a major QTL is available and that can be applied to select desirable haplotype blocks for parental selection and crossing by design [116].

In order to widen the genetic base, it may be necessary to utilize wild species accessions as introgression libraries as well as developing interspecific populations. On the other hand, elite cultivars and landraces can be used to develop mapping populations, and training populations [114]. Wild soybean (*G. soja*) serves as a unique resource to study regulation of protein and amino acid biosynthesis, because the seed concentration of these components is higher in *G. soja* compared with *G. max.* Utilization of *G. soja* in breeding program is hampered due to linkage drag on favorable agronomic characteristics [113]. However, this issue could be resolved by advanced backcross QTL-based breeding, which was utilized for introgressing alleles from wild tomato to cultivated type for yield improvement [117], or through mutation breeding approaches.

10. Breeding for carbohydrate content in soybean seeds

10.1 Sucrose content

Breeders aim to increase the sucrose content in soybean seeds which contribute to the sweet taste of soy foods, especially for tofu, soy milk, and edamame. The sucrose content in soybeans ranges from 1.5 to 10.2%, and germplasm with even higher content, 13.6%, has been identified [13]. Varieties that target a specific component of the carbohydrate fraction are varieties high in sucrose content and varieties low in oligosaccharides [13]. Compared to conventional soybeans, high-sucrose soybeans contain 40% more sucrose but 90% less stachyose and raffinose. High-sucrose soybeans are used to produce tofu, soymilk, beverages, baked goods, puddings, cheese, and meat analogs [13]. The genotypic correlation between sucrose and 100-seed weight is positive and significant, as well as the genotypic correlation of 1000-seed weight with protein. Moreover, the heritability for 1000-seed weight is high. Hence, the breeding program selection on 100-seed weight would result in a good response on relative protein and sucrose content.

10.2 Oligosaccharides content

Stachyose and raffinose are not readily digestible and cause flatulence when soy foods are consumed. Therefore, breeders aim to develop soybean seeds with reduced oligosaccharide content. Stachyose and raffinose content among soybean germplasm range from 1.4 to 6.7%, and 0.1 to 2.1%, respectively. Breeding lines with less than 1% stachyose and raffinose have been developed [13]. Soybean germplasm "V99–5089" was developed with high sucrose, low raffinose, and low stachyose content to use as a parent in food-grade soybean breeding programs [118]. The genetic variability of seed sugars has significant allelic difference in the genes controlling the biosynthetic enzymes. QTL mapping of soluble sugars in soybean seed were reported and of which 28 were for seed sucrose (**Table 1**). These 28 QTLs were mapped on LGs A1 and E; 3 QTLs on A2, I, and F, and 3 QTLs on L, M, and B1 [73], two QTLs on L, D1b, 7 QTLs on L [74], and B2, D1B, E, H, J [75]. The genomic regions associated with sucrose, raffinose, and stachyose were identified in segregating $F_{2–10}$ RILs [74].

Alkond et al. [76] reported 14 significant QTLs associated with sucrose and oligosaccharides that were mapped on 8 different linkage groups (LGs) and chromosomes (Chr). Seven QTL were identified for raffinose content on LGs D1a (Chr1), N (Chr3), C2 (Chr6), K (Chr9), B2 (Chr14), and J (Chr16). Four QTL for stachyose content were identified on LG D1a (Chr1), C2 (Chr6), H (Chr12), and B2 (Chr14) [76]. Three QTL for seed sucrose content were identified on LGs N (Chr3), K (Chr9), and E (Chr15). The region of Chr15 (LG E) that has been reported to be associated with sucrose was detected by others [73, 75, 77, 78], but the position of the QTL was different [76]. The two of the regions underlying seed sucrose QTLs identified on LG N (Chr3) and K (Chr9) are additions to the loci previously reported on LGs D1b (Chr2), A1 (Chr5), M (Chr7), A2 (Chr8), B1 (Chr11), H (Chr12), F (Chr13), G (Chr18), J (Chr16), L (Chr19), and I (Chr20) [73, 75, 77, 78]. The selection for beneficial alleles of these QTLs could facilitate breeding strategies to develop soybean lines with higher concentrations of sucrose and lower levels of raffinose and stachyose.

11. Breeding for functional components

11.1 Lipoxygenase

Normal soybean seeds contain three lipoxygenase isozymes that are responsible for the grassy beany flavor and bitter taste of soy food. Research is being conducted for the genetic elimination of lipoxygenase from soybean seeds to reduce undesirable flavors in soy food products. Soybean seed lipoxygenase exists in three isozymic forms, namely lipoxygenase-1, −2, and −3 controlled by single dominant genes, viz. *Lx1*, *Lx2*, and *Lx3*, respectively. Their recessive forms, i.e., *lx1*, *lx2*, and *lx3* cause the loss in activity of corresponding isozyme [119]. Several combinations of lipoxygenase null mutants have already been developed: 0-, 00-, and 000-genotypes with one, two, and three of the isozymes eliminated respectively, In the 000-genotype, absence of the grassy and beany flavor was observed, as there was no detectable level of the lipoxygenase proteins in mature soybean seeds. The presence or absence of three lipoxygenase isozymes is determined by gel electrophoresis and spectrophotometer or by immunological or colorimetric methods [13]. Of the three lipoxygenases, *Lx2* locus has been mapped on *chr13*, which corresponds to linkage group F, and has been reported tightly linked with *Lx1* locus [120]. *Lx3* gene has been reported to be present on *chr15* and is inherited independent of *Lx1* and *Lx2.* SSR marker Satt656 tightly linked with Lox2 [121] has been deployed in the development of Lox-2 free soybean genotypes NRC109 and NRC110 in India [121].

Based on *Lx3* mutant gene sequence, SNP (Lox3PM1) and STS marker (Lox3–3′) were developed for identification of Lox*3* null individual [122].

11.2 Isoflavones

Soybean cultivars with good isoflavone content are desirable as it contributes health benefits. High-isoflavone soybeans contain more than 0.4% isoflavones compared to levels of 0.15–0.25% for traditional soybean varieties [13]. Isoflavone content is influenced by genetic factors and environmental factors such as temperature and irrigation during seed maturation [13]. For instance, the total isoflavone content of soybean seeds appears to be negatively related to growth temperature [5]. Understanding the genetic regulation of this pathway may be necessary for obtaining cultivars with good isoflavone levels. Interest has been put in the phenylpropanoid synthetic pathway which is catalyzed in its first step by isoflavone synthase (IFS). Two genes for IFS have been identified in soybean. Furthermore, negative correlation has been found between total isoflavone content and linolenic acid (18:3) concentration. Other data suggest negative correlation between isoflavone content and protein content [5]. QTLs affecting isoflavones were identified using recombinant inbred line population and found five QTLs contributed to the concentration of isoflavones, having single or multiple additive effects on isoflavone component traits [123]. Similarly, six QTLs were identified using the linkage map constructed with specific length amplified fragment sequencing, of which one major QTL (qIF20-2) contributed to a majority of isoflavone components across various environments and explained a high amount of phenotypic variance (8.7–35.3%) [124]. Akond et al. [125] identified QTL controlling isoflavone content in a set of recombinant inbred line (RIL) populations of soybean derived from "MD96–5722" by "Spencer" cultivars. Wide variations were found for seed concentrations of daidzein, glycitein, genistein, and total isoflavones among RIL populations. Three QTLs were identified on three different linkage groups (LG). One QTL that controlled daidzein content was identified on LG A1 (Chr 5) and two QTLs that underlay glycitein content were identified on LG K (Chr 9) and LG B2 (Chr 14). Identified QTLs could be used to develop soybean with preferable isoflavone concentrations in the seeds through MAS.

11.3 Seed oil concentration

Increasing the seed oil concentration has been a breeding goal for centuries. The ancestor of the domesticated soybean used to have small, hard, black seeds with low oil content, high protein content, and low yield. It is known that an increase in oil content is positively correlated with yield and negatively correlated with protein content. Selection for yield, agronomic characteristics and seed quality, large yellow seeds with typical averages of 20% oil and 40% protein were obtained. However, soybean is appreciated for its high protein meal and versatile vegetable oils; therefore, breeders mostly prefer to obtain modest gains in oil and yield without substantial loss in protein concentration [42]. Breeding for oil quality such as with reduced saturated fatty acids are prime focus as it is responsible for elevating cholesterol. The saturated fatty acids present in soybean oil are palmitic acid, 16:0, and stearic acid, 18:0. Especially, palmitic acid is a health concern as it is correlated to cardiovascular disease. It has been suggested that saturated fatty acids should be kept below 7–10% on a daily basis [42]. Soybean oil contains the monounsaturated fatty acid, oleic acid, 18:1. The oxidative stability of the soybean oil is enhanced by increasing three times higher the concentration of monounsaturated fatty acid such as oleic (18:1) than the normal content which is about 22%. Therefore, breeders target a concentration of 18:1 of about 65–75% of total lipid in soybean. By the means

of genetic engineering, 18:1 levels of about 80% total lipid have been achieved [42]. In general, soybean varieties with unique fatty acid composition such as high oleic acid content, high stearic acid content, low linolenic acid content, or low palmitic acid content are preferred [13].

Assessment of agronomic traits has been used to evaluate phenotypic diversity in 20,570 Chinese soybean accessions and it was reported that seed coat color had the highest diversity index among the qualitative traits [126]. Plant's height had the most variation among quantitative traits, and followed by seed size, protein content, growth period, and oil content. The seed size of those accessions ranged from smaller than 2 to as large as 46 g/100-seeds. The protein content ranged from 30 to 53%; and oil content ranged from 10 to 25%. The variances of seed size, protein content, and oil content of the U.S. cultivars were lower than the Chinese cultivars [127]. The Southern U.S. soybeans were more variable in oil and protein contents and less variable in seed size than the Northern U.S. soybeans. The food-grade soybean breeding aims to increase the nutritional content and quality of protein and oil [128]. Greater genetic diversity of protein content, seed hardness, calcium content, and seed size uniformity than other quality traits in both small and large-seeded genotypes were evaluated [128]. The U.S. soybean genotypes with small seed were more diverse and exhibited higher swell ratio and oil content but lower stone seed ratio and protein content than the Asian accessions [128]. Among the large-seeded accessions, the U.S. genotypes had higher stone seed ratio and oil content but lower swell ratio and protein content, and were less diverse than the Asian genotypes [128]. The characterization of diverse food grade soybeans will facilitate parent selection in specialty soybean breeding [1].

12. Breeding for reduced trypsin inhibitor

Soybean germplasm PI542044, also known as Kunitz soybean, contains the null allele of KTI, i.e., kti that encodes a truncated protein and it was developed in a backcross program involving Williams 82 and PI157440 [129]. Introgression of kti is complicated by a number of factors viz., (i) kti being recessive in inheritance, each conventional backcross generation would be requiring selfing followed by estimation of KTI content in the seeds so as to identify a target plant. However, three recessive null alleles, viz. Kunitz trypsin inhibitor, soybean agglutinin, and P34 allergen null were stacked in the background of "Williams 82" and were termed as "Triple Null" [130]. Three SSR markers, viz. Satt228, Satt409, and Satt429 have been reported to be closely linked (0–10 cM) with the null allele of Kunitz trypsin inhibitor [131]. These SSR markers was also validated in the mapping population generated using Indian soybean genotypes as the recipient parent (*TiTi*) and PI542044 (*titi*) as the donor for the null allele [132]. Further, a gene-specific marker has also been designed from the null allele of KTI from genotype PI157440 [15] and has been deployed in the selection of plants carrying the null allele of KTI derived from PI542044 [121]. The null allele of KTI from PI542044 was introgressed into the cultivar "JS97–52" (recurrent parent) through marker-assisted backcrossing using the SSR marker Satt228, tightly linked with a trypsin inhibitor *Ti* locus. An introgressed line JS97–52 with reduced trypsin inhibitor (68.8–83.5%) content was developed [133].

13. Breeding for reduced cadmium content in soybean

Based on the importance of soybean as a staple food crop, the development of low Cd soybean cultivars should be a priority. The genetic variability for Cd accumulation within a species provides an opportunity to select soybean genotypes with low Cd

concentration. In soybean grain, Cd concentration was found to be controlled by a single gene, with low Cd dominant in the crosses studied [134]. Lines with the low Cd trait had restricted root-to-shoot translocation, which limited the Cd accumulation in the grain. Genetic variability in soybean [19, 135] has been reported. An understanding of genetics and heritability of the Cd accumulation is essential in designing the breeding strategy to incorporate gene(s) controlling low Cd accumulation in modern cultivars. However, identifying low Cd phenotypes by analysis of the grain is challenging due to the high cost of analysis [136]. Developing inexpensive methods would assist in transferring the low Cd accumulation traits with other desirable traits.

14. Developing markers for marker-assisted selection of low Cd accumulation

Marker-assisted selection (MAS) could be an alternative to phenotypic selection. In soybean, DNA markers linked to low Cd accumulation were identified using RIL population (*F*6:8) derived from the cross AC Hime (high Cd accumulation in seeds) and Westag-97 (low Cd accumulation in seeds). The distribution of Cd concentration of 166 RILs ranged from 0.067 to 0.898 mg kg^{-1}, with a mean of 0.268 ± 0.013 mg kg^{-1} [134]. Using the RIL population, seven simple sequence repeat (SSR) markers, SatK138, SatK139, SatK140 (0.5 cM), SatK147, SacK149, SaatK150, and SattK152 (0.3 cM), were reported to be linked to *Cda1* in soybean seed. It was also reported that all the linked markers were mapped to the same linkage group (LG) K. SSR markers closely linked to *Cda*1 in soybean seeds have the potential to be used for MAS to develop low Cd-accumulating cultivars in a breeding program [134]. In a similar mapping approach, Benitez et al. [137] identified a major QTL cd1 on chromosome 9 (LG-K) across years and generations which accounted for 82, 57, and 75% of the genetic variation. Near-isogenic lines (NILs) were used to confirm the effect of the QTL and the peak of the QTL that was located in the vicinity of two SSR markers, Gm09:4770663 and Gm09:4790483. Both the studies revealed a major QTL for seed *Cd* content, *Cda*1 at a similar genomic location, suggesting that *cd*1 and *Cda1* may be identical. Candidate genes related to heavy metal transport or homeostasis were located in the vicinity of the identified QTL (Cda1). Protein kinase, putative adagio-like protein, and plasma membrane H^+-ATPase were found in the QTL vicinity. The presence of protein kinase and plasma membrane H^+-ATPase genes near the tightly linked SSR markers suggests that the regulation of this enzyme may play a vital role in Cd stress [134]. This was later supported by a major QTL-controlling Cd concentration (*cd*1) identified in soybean [137]. The gene was designated as *GmHMA1.* In *GmHMA1a*, one base substitution from G to A at nucleotide position 2095 resulted in a loss of function of the ATPase and was found to be associated with Cd uptake [137]. The SSR markers linked to the *Cda*1 and*Cd*1gene(s)/or QTLs and the SNP marker in the P1B-ATPase metal ion transporter gene in soybean can be utilized in MAS for developing soybean cultivars with low Cd content.

15. Future directions

Breeding for soybean seed composition traits is a complicated process; fortunately, ample genomic resources and tools are now available to soybean breeders/researchers for dissection of seed composition traits. The combination of conventional breeding strategy and genomic approaches will help to identify genomic loci, haplotypes, and FMs in breeding for improvement of seed composition traits. For improvement of protein, the major protein QTL, which was repeatedly mapped

on Chr20, Chr15, and Chr18, may facilitate breeders to select parental lines and consider them for crossing schemes or introgression into locally adapted superior yielding cultivars through genomics-assisted breeding and MAS. Issues related to protein increase without yield drag, pleiotropic effects, and background/allelic effects could be addressed via screening diverse germplasm, considering wild soybean alleles for introgression, undertaking genomics-assisted breeding, precise high-throughput phenotyping, mutational breeding, and genome editing through Crisp/Cas. Integrating these aspects will extend our current genetic and genomic portfolio far beyond that of traditional breeding. Finally, when a cultivar with improved food-grade characteristics is developed, a further step is the evaluation of the quality of the product obtained from this cultivar. This is important as the success of a food-grade soybean cultivar is determined by the preferences of the consumers.

Author details

Souframanien Jegadeesan[1] and Kangfu Yu[2]*

1 Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Mumbai, Maharashtra, India

2 Harrow Research and Development Centre, Agriculture and Agri-Food Canada, Ontario, Canada

*Address all correspondence to: kangfu.yu@canada.ca

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] Zhang B, Chen P, Florez-Palacios SL, Shi A, Hou A, Ishibashi T. Seed quality attributes of food-grade soybeans from the U.S. and Asia. Euphytica. 2010;**173**:387-396

[2] Cui ZAT, James S, Miyazaki RF, Wilson TE. Carter. Breeding of specialty soybeans for traditional and new soyfoods. In: Liu K, editor. Soybeans as a Functional Food. Champaign: AOCS Press; 2004. pp. 264-322

[3] Omoni AO, Aluko RE. Soybean foods and their benefits: Potential mechanisms of action. Nutrition Reviews. 2005;**63**:272-283

[4] US Soyfoods Directory. 2006. Available from: http://www.soyfoods.com/soyfoodsdescriptions/descriptions.html

[5] Soyatech. Soyfood Facts. Available from: http://www.soyatech.com/soyfoods_facts.htm [Cited: 20 January 2012]

[6] L'Hocine L, Boye JI. Allergenicity of soybean: New developments in identification of allergenic proteins, cross-reactivities and hypo-allergenization technologies. Critical Reviews in Food Science and Nutrition. 2007;**47**:127-143

[7] Jaureguy LM, Chen P, Scaboo AM. Heritability and correlations among food-grade traits in soybean. Plant Breeding. 2011;**130**:647-652

[8] Graef GL, Specht JE. Fitting the Niche Food Grade Soybean Production: A New Opportunity for Nebraska Soybean Producers. Lincoln: Nebraska Department of Agriculture; 1989. pp. 18-27

[9] Poysa V, Cober E, Reid J, Pietrzak L, Mullin J, Voldeng H. Improving the Quality of Food Grade Soybeans. Bulletin National Research Council (NRC), Plant Biology Institute (PBI). 2002

[10] Bachman MS, Tamulonis JP, Nickell CD, Bent AF. Molecular markers linked to brown stem rot resistance genes, Rbs1 and Rbs2, in soybean. Crop Science. 2001;**41**:527-535

[11] Mullin WJ, Xu W. Study of soybean seed coat components and their relationship to water absorption. Journal of Agricultural and Food Chemistry. 2001;**49**:5331-5335

[12] Saio K. Soybeans resistant to water absorption. Cereal Foods World. 1976;**21**:168-173

[13] Chen P. Developing high quality identity-preserved soybean for the specialty soyfood market. In: Production and Marketing of Identity-Preserved Soybean. MO, USA: American Soybean Association; 2004. pp. 23-31

[14] Taira H. Quality of soybeans for processed foods in Japan. Japan Agricultural Research Quarterly. 1990;**24**:224-230

[15] Moraes R, Soares CB, Colombo LR, Salla MFS, Barros JGA, Piovesan ND, et al. Assisted selection by specific DNA markers for genetic elimination of the kunitz trypsin inhibitor and lectin in soybean seeds. Euphytica. 2006;**149**:221-226

[16] Raboy V. Seed phosphorous and the development of low-phytate crop. In: Inositol Phosphates: Linking Agriculture and the Environment Turner BL, Richardson AE, Mullany EJ.). Wallingford, U.K: CAB International; 2007. pp. 111-132.

[17] Talukdar A, Shivkumar M. Genetic improvement of food-grade soybean in India: Current status and future prospects. The Indian Journal of Genetics. 2016;**76**:626-630

[18] McGrath SP, Zhao FJ, Lombi E. Plant and rhizosphere processes involved

in phyto remediation of metal-contaminated soils. Plant and Soil. 2001;**232**:207-214

[19] Arao T, Ae N, Sugiyama M, Takahash M. Genotypic differences in cadmium uptake and distribution in soybeans. Plant and Soil. 2003;**251**:247-253

[20] Sugiyama M, Ae N, Hajika M. Developing of a simple method for screening soybean seedling cadmium accumulation to select soybean genotypes with low seed cadmium. Plant and Soil. 2011;**341**:413-422

[21] Kazantzis G. Cadmium, osteoporosis and calcium metabolism. Biometals. 2004;**17**:493-498

[22] Codex Alimentarius Commission. Report of the 33rd Session of the Codex Committee on Food Additives and Contaminants. Hague, The Netherlands: Codex Alimentarius Commission; 2001. p. 285

[23] Ministry of Agriculture, Forestry, and Fisheries of Japan. Investigation of Cd Concentration of Staple Crops: The Outline of the Results. Tokyo: Ministry of Agriculture, Forestry, and Fisheries of Japan; 2002

[24] Cupit M, Larsson O, de Meeus C, Eduljee GH, Hutton M. Assessment and management of risks arising from exposure to cadmium in fertilisers: II. Science of the Total Environment. 2002;**291**:189-206

[25] Endres JG. Protein quality and human nutrition. In: Endres JG, editor. Soy Protein Products: Characteristics, Nutritional Aspects, and Utilization. Revised and Expanded Edition. Champaign, Illinois, USA: AOCS; 2001. pp. 10-19

[26] Kumar V, Rani A, Chauhan GS. Nutritional value of soybean. In: Singh G, editor. The Soybean: Botany, Production and Uses. Wallingford, U.K: CABI; 2010. pp. 375-403

[27] Liu K. Soybeans as a powerhouse of nutrients and phytochemicals. In: Liu K, editor. Soybeans as Functional Foods and Ingredients. New York: AOCS Publishing; 2004

[28] Poysa V, Woodrow L. Stability of soybean seed composition and its effect on soymilk and tofu yield and quality. Food Research International. 2002;**35**:337-345

[29] Cai T, Chang KC. Processing effect on soybean storage proteins and their relationship with tofu quality. Journal of Agricultural and Food Chemistry. 1999;**47**:720-727

[30] Poysa V, Woodrow L, Yu K. Effect of soy protein subunit composition on tofu quality. Food Research International. 2006;**39**:309-317

[31] Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, et al. Characterization of the glycinin gene family in soybean. The Plant Cell. 1989;**1**:313-328

[32] Maruyama N, Adachi M, Takahashi K, Yagasaki K, Kohna M, Takenaka Y, et al. Crystal structures of recombinant and native soybean β -conglycinin β homotrimers. European Journal of Biochemistry. 2001;**268**:3595-3604

[33] Thanh VH, Shibasaki KK. Major proteins of soybean seeds. A straight forward fractionation and their characterization. Journal of Agricultural and Food Chemistry. 1976;**24**:1117-1121

[34] Ogawa T, Bando N, Tsuji H, Nishikawa K, Kitamura K. 1995. Alpha-subunit of beta-conglycinin, an allergenic protein recognized by IgE antibodies of soybean-sensitive patients with atopic dermatitis. Bioscience,

Biotechnology, and Biochemistry. 59:831-833.

[35] Panthee DR, Kwanyuen P, Sams CE, West DR, Saxton AM, Pantalone VR. Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. Journal of the American Oil Chemists' Society. 2004;**81**:1005-1012

[36] Yu K, Woodrow L, Shi C, Anderson D, Poysa V. Registration of 7S β-conglycinin α′ and 11S glycinin A4 null food-grade soybean germplasm, HS-162. Canadian Journal of Plant Science. 2017;**97**:536-538

[37] Yagasaki K, Kousaka F, Kitamura K. Potential improvement of soymilk gelation properties using soybean with modified protein subunit composition. Breeding Science. 2000;**50**:101-107

[38] Tezuka SL, Taira H, Igarashi Y, Yagasaki K, Ono T. Properties of tofus and soy milks prepared from soybeans having different subunits of glycinin. Journal of Agricultural and Food Chemistry. 2000;**48**:1111-1117

[39] Nik AM, Tosh SM, Woodrow L, Poysa V, Corredig M. Effect of soy protein subunit composition and processing conditions on stability and particle size distribution of soymilk. LWT- Food Science and Technology. 2009;**42**:1245-1252

[40] Yu K, Woodrow L, Poysa V. Registration of lipoxygenase free food grade soybean Germplasm, HS-151. Canadian Journal of Plant Science. 2016;**96**:148-150

[41] Obendorf RL, Kosina SM. Soluble carbohydrates in soybean. In: Tzi-Bun N, editor. Soybean-Biochemistry, Chemistry and Physiology. UK: Intechopen; 2011. pp. 201-228

[42] Cober ER, Cianzion SR, Pantalone VR, Rajcan I. In: Vollmann J, Rajcan I, editors. Soybean, in Oil Crops. Handbook of Plant Breeding. Vol. 4. New York; Springer; 2009. pp. 57-90

[43] Sathyapalan T, Aye M, Rigby A, Thatcher NJ, Dargham SR, Kilpatrick ES, et al. Soy isoflavones improve cardiovascular disease risk markers in women during the early menopause. Nutrition, Metabolism, and Cardiovascular Diseases. 2018;**28**:691-697

[44] Esler I. Prospects for Vegetable Soybean in India and Its Market Acceptance. Report. Taiwan: AVRDC—The World Vegetable Centre; 2011. pp. 1-26

[45] Lee JY, Min S, Choe EO, Min DB. Formation of volatile compounds in soy flour by singlet oxygen oxidation during storage under light. Journal of Food Science. 2003;**68**:1933-1937

[46] Iassonova DR, Johnson LA, Hammond EG, Beattie SE. Evidence of anenzymatic source of off flavors in lipoxygenase-null soybeans. Journal of the American Oil Chemists' Society. 2009;**86**:59-64

[47] Buzzell RI, Anderson TR, Hamill AS, Welacky TW. Harovinton soybean. Canadian Journal of Plant Science. 1991;**71**:525-526

[48] Griffis G, Wiedermann L. Marketing Food-Quality Soybeans in Japan. 3rd ed. St. Louis: American Soybean Association; 1990

[49] Wilson RF. Seed composition. In: Boerma HR, Specht JE, editors. Soybeans: Improvement, Production, and Uses. 3rd ed. Madison: American Society of Agronomy, Inc./Crop Science Society of America, Inc./Soil Science Society of America, Inc.; 2004, 2004. p. 621

[50] Hwang E, Song Q, Jia G, Specht J, Hyten D, Costa J. A genome-wide association study of seed protein

and oil content in soybean. BMC Genomics. 2014;**15**:1

[51] Eskandari M, Cober ER, Rajcan I. Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. Theoretical and Applied Genetics. 2013;**126**:1677-1687

[52] SoyBase. Available from: https://soybase.org/

[53] Pathan SM, Vuong T, Clark K, Lee JD, Shannon JG, Roberts CA, et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Science. 2013;**53**:765-774

[54] Akond M, Ragin B, Bazzelle R, Kantartzi SK, Meksem K. Quantitative trait loci associated with moisture, protein, and oil content in soybean [*Glycine max* (L.) Merr.]. Journal of Agricultural Science. 2012;**4**:16-25

[55] Diers B, Keim P, Fehr W, Shoemaker R. RFLP analysis of soybean seed protein and oil content. Theoretical and Applied Genetics. 1992;**83**:608-612

[56] Nichols D, Glover K, Carlson S, Specht J, Diers B. Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Science. 2006;**46**:834-839

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

[58] Lee SH, Bailey MA, Mian MAR, Carter TE, Shipe ER, Ashley DA, et al. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theoretical and Applied Genetics. 1996;**93**:649-657

[59] Sebolt A, Shoemaker R, Diers B. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science. 2000;**40**:1438-1444

[60] Chung J, Babka H, Graef G, Staswick P, Lee D, Cregan P, et al. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science. 2003;**43**:1053-1067

[61] Fasoula VA, Harris DK, Boerma HR. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Science. 2004;**44**(4):1218

[62] Panthee DR, Pantalone VR, West DR, Saxton AM, Sams CE. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Science. 2005;**45**:2015

[63] Lu W, Wen Z, Li H, Yuan D, Li J, Zhang H, et al. Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. Theoretical and Applied Genetics. 2013;**126**:425-433

[64] Wang J, Chen P, Wang D, Shannon G, Zeng A, Orazaly M, et al. Identification and mapping of stable QTL for protein content in soybean seeds. Molecular Breeding. 2015;**35**:1-10

[65] Phansak P, Soonsuwon W, Hyten DL, Song Q, Cregan PB, Graef GL, et al. Multi-population selective genotyping to identify soybean (*Glycine max* (L.) Merr.) seed protein and oil QTLs. G3: Genes, Genomes, Genetics. 2016;**6**:1635-1648

[66] Vaughn JN, Nelson RL, Song Q, Cregan PB, Li Z. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3: Genes, Genomes, Genetics. 2014;**4**:2283-2294

[67] Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F. Identification of loci governing eight agronomic traits using a GBSGWAS approach and validation by QTL mapping in soybean. Plant Biotechnology Journal. 2015;**3**:211-221

[68] Bandillo N, Jarquin D, Song QJ, Nelson R, Cregan P, Specht J, et al. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome. 2015;**8**:1-13

[69] Valliyodan B, Qiu D, Patil G, Zeng P, Huang J, Dai L, et al. Landscape of genomic diversity and trait discovery in soybean. Scientific Reports. 2016;**6**:23598

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

[71] Tajuddin T, Watanabe S, Yamanaka N, Harada K. Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. Breeding Science. 2003;**53**:133-140

[72] Hyten DL, Pantalone VR, Sams CE, Saxton AM, Landau-Ellis D, Stefaniak TR, et al. Seed quality QTL in a prominent soybean population. Theoretical and Applied Genetics. 2004;**109**:552-561

[73] Maughan PJ, Maroof MAS, Buss GR. Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max* L.). Molecular Breeding. 2000;**6**:105-111

[74] Kim HK, Kang ST, Cho JH, Choung MG, Suh DY. Quantitative trait loci associated with Oligo saccharide and sucrose contents in soybean (*Glycine max* L.). Journal of Plant Biology. 2005;**48**:106-112

[75] Kim HK, Kang ST, Oh DW. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. Journal of Plant Research. 2006;**119**:533-538

[76] Akond M, Liu S, Kantartzi SK, Meksem K, Bellaloui N, David, et al. Quantitative trait loci underlying seed sugars content in "MD96-5722" by "Spencer" recombinant inbredline population of soybean. Food and Nutrition Sciences. 2015;**6**:964-973

[77] Cicek M. Genetic marker analysis of three major carbohydrates in soybean seeds [PhD dissertation]. Blacksburg: Virginia Polytechnic Institute & State University; 2001

[78] Clevinger EM. Mapping quantitative trait loci for soybean quality traits from two different sources [PhD dissertation]. Blacksburg: Virginia Polytechnic Institute and State University; 2006

[79] Maroof MAS, Buss GR. Low phytic, low stachyose, high sucrose soybean lines. Patent Application Number: 20080199591; US20080199591 A1; US 8003856 B2; 2008

[80] Feng C, Morsy M, Giannoccaro E, Zhang B, Chen P. Soybean seed sugar content and quantitative trait loci mapping. In: Li CJ et al., editors. Plant Nutrition for Food Security, Human Health, and Environmental Protection. Beijing: Tsinghua University Press; 2005. p. 438

[81] Zeng A, Chen P, Shi A, Wang D, Zhang B, Orazaly M, et al. Identification of quantitative trait loci for sucrose content in soybean seed. Crop Science. 2014;**54**:554-564

[82] Wang Y, Chen P, Zhang B. Quantitative trait loci analysis of soluble sugar contents in soybean. Plant Breeding. 2014;**133**:493-498

[83] Patil G, Chaudhary J, Vuong TD, Jenkins B, Qiu D, Kadam S, et al. Development of SNP genotyping assays for seed composition traits in soybean. International Journal of Plant Genomics. 2017:1-12. DOI: 10.1155/2017/6572969

[84] Poeta F, Borrás L, Rotundo JL. Variation in seed protein concentration and seed size affects soybean crop growth and development. Crop Science. 2016;**6**:3196-3208

[85] Ishikawa G, Takada Y, Nakamura T. A PCR based method to test for the presence or absence of β-conglycinin α´- and α-subunits in soybean. Molecular Breeding. 2006;**17**:365-374

[86] Tsubokura Y, Hajika M, Harada K. Molecular marker associated with β –conglycinin deficiency in soybean. Breeding Science. 2006;**56**:113-117

[87] Teraishi M, Takahashi M, Hajika M, Matsunaga R, Uematsu Y, Ishimoto M. Suppression of soybean β-conglycinin genes by a dominant gene, Scg-1. Theoretical and Applied Genetics. 2001;**103**:1266-1272

[88] Hayashi M, Nishioka M, Kitamura K, Harda K. Identification of AFLP markers tightly linked to the gene for deficiency of the 7S globulin in soybean seed and characterization of abnormal phenotypes involved in the mutation. Breeding Science. 2000;**50**:123-129

[89] Diers BW, Beilinson V, Nielsen NC, Shoemaker RC. Genetic mapping of the Gy4 and Gy5 glycinin genes in soybean and the analysis of a variant of Gy4. Theoretical and Applied Genetics. 1994;**89**:297-304

[90] Beilinson V, Chen Z, Shoemaker RC, Fischer RL, Goldberg RB, Neilsen NC. Genomic organization of glycinin genes in soybean. Theoretical and Applied Genetics. 2002;**104**:1132-1140

[91] Boehm JD Jr, Nguyen V, Tashiro RM, Anderson D, Shi C, Wu X, et al. Genetic mapping and validation of the 7S α' and 11S A-type storage protein subunits in soybean [*Glycine max* (L.) Merr.]. Theoretical and Applied Genetics (TAG). 2018;**131**(3):659-671

[92] Andersen JR, Lubberstedt T. Functional markers in plants. Trends in Plant Science. 2003;**8**:554-560

[93] Wang C, Wu X, Jia F, Zhang J, Chen S. Genetic variations of glycinin genes among cultivated and wild type soybean species. Progress in Natural Science. 2008;**18**:33-41

[94] Mahmoud AA, Natarajan SS, Bennett JO. Effect of six decades of selective breeding on soybean protein composition and quality: A biochemical and molecular analysis. Journal of Agricultural and Food Chemistry. 2006;**54**:3916-3922

[95] Souframanien J, Yu K, Lorna Woodrow L, Wang Y, Shi C, Poysa V. Molecular analysis of glycinin genes in soybean mutants for development of gene specific markers. Theoretical and Applied Genetics. 2012;**124**:365-372

[96] Mandal S, Mandal R. Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Current Science. 2000;**79**:576-589

[97] Warrington C, Abdel-Haleem H, Hyten D, Cregan P, Orf J, Killam A, et al. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. Theoretical and Applied Genetics. 2015;**128**:839-850

[98] Ma Y, Kan G, Zhang X, Wang Y, Zhang W, Du H, et al. Quantitative trait loci (QTL) mapping for glycinin and beta-conglycinin contents in soybean (*Glycine max* L. Merr.). Journal of Agricultural and Food Chemistry. 2016;**64**:3473-3483

[99] Panthee D, Pantalone V, Sams C, Saxton A, West D, Orf J, et al. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theoretical and Applied Genetics. 2006;**112**:546-553

[100] Chaudhary J, Patil GB, Sonah H, Deshmukh RK, Vuong TD, Valliyodan B, et al. Expanding omics resources for improvement of soybean seed composition traits. Frontiers in Plant Science. 2015;**6**:1021

[101] Krishnan HB, Kim WS, Oehrle NW, Alaswad AA, Baxter I, Wiebold WJ, et al. Introgression of leginsulin, a cysteine-rich protein, and high-protein trait from an Asian soybean plant introduction genotype into a North American experimental soybean line. Journal of Agricultural and Food Chemistry. 2015;**63**:2862-2869

[102] Deshmukh R, Sonah H, Patil G, Chen W, Prince S, Mutava R. Integrating omic approaches for abiotic stress tolerance in soybean. Frontiers in Plant Science. 2014;**5**:244

[103] Yates JL. Use of diverse germplasm to improve peanut root-knot nematode resistance and seed protein content in soybean [PhD dissertation]. Athens: University of Georgia; 2006

[104] Mian MA, McHale L, Li Z, Dorrance AE. Registration of 'Highpro1' soybean with high protein and high yield developed from a north × south cross. Journal of Plant Registrations. 2017;**11**:51-54

[105] Bolon Y-T, Stec AO, Michno J-M, Roessler J, Bhaskar PB, Ries L, et al. Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics. 2014;**198**:967-981

[106] Sandhu D, Alt JL, Scherder CW, Fehr WR, Bhattacharyya MK. Enhanced oleic acid content in the soybean mutant M23 is associated with the deletion in the Fad2-1a gene encoding a fatty acid desaturase. Journal of the American Oil Chemists' Society. 2007;**84**:229-235

[107] Schmidt MA, Herman EM. Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. Molecular Plant. 2008;**1**:910-924

[108] Gillman JD, Stacey MG, Cui Y, Berg HR, Stacey G. Deletions of the SACPD-C locus elevate seed stearic acid levels but also result in fatty acid and morphological alterations in nitrogen fixing nodules. BMC Plant Biology. 2014;**14**:1

[109] Kitamura K. Genetic improvement of nutritional and food processing quality in soybean. JARQ. 1995;**29**:1-8

[110] Song B, Shen L, Wei X, Guo B, Tuo Y, Tian F, et al. Marker-assisted backcrossing of a null allele of the α-subunit of soybean (*Glycine max*) β-conglycinin with a Chinese soybean cultivar (a). The development of improved lines. Plant Breeding. 2014;**133**:638-648

[111] Brim C, Burton J. Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science. 1979;**19**:494-498

[112] Holbrook CC, Burton JW, Carter TE. Evaluation of recurrent restricted index selection for increasing yield while holding seed protein constant in soybean. Crop Science. 1989;**29**:324-329

[113] Lewers K, Palmer R. Recurrent selection in soybean. Plant Breeding Reviews. 1997;**15**:275-314

[114] Varshney RK, Mohan SM, Gaur PM, Gangarao N, Pandey MK, Bohra A, et al. Achievements and prospects of genomics-assisted breeding

in three legume crops of the semi-arid tropics. Biotechnology Advances. 2013;**31**:1120-1134

[115] Jarquin D, Specht J, Lorenz A. Prospects of genomic prediction in the USDA soybean germplasm collection: Historical data creates robust models for enhancing selection of accessions. G3: Genes, Genomes, Genetics. 2016;**116**:031443

[116] Patil G, Do T, Vuong TD, Valliyodan B, Lee JD, Chaudhary J, et al. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Scientific Reports. 2016;**6**:19199

[117] Tanksley S, Nelson J. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theoretical and Applied Genetics. 1996;**92**:191-203

[118] Mozzoni L, Shi AN, Chen PY. Genetic analysis of high sucrose, low Raffinose, and low Stachyose content in V99-5089 soybean seeds. Journal of Crop Improvement. 2013;**27**:606-616

[119] Axelord B, Cheesborough TM, Laasko S. Lipoxygenase from soybean. Methods in Enzymology. 1981;**71**:441

[120] Xu SJ, Singh RJ, Kalipara KP, Hymowitz T. Primary trisomics in soybean: Origin, identification, breeding behaviour and use in linkage mapping. Crop Science. 2000;**40**:1543

[121] Kumar V, Rani A, Rawal R, Hussain SM. Lipoxygenase-2 free Indian soybean genotype. Current Science. 2013;**104**:586

[122] Reinprecht Y, Luke-Labey SY, Yu K, Poysa VW, Rajcan I, Ablett GR, et al. Molecular basis of seed lipoxygenase-null traits in soybean line OX948. Theoretical and Applied Genetics. 2011;**122**:1247

[123] Gutierrez-Gonzalez JJ, Vuong TD, Zhong R, Yu O, Lee JD, Shannon G, et al. Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. Theoretical and Applied Genetics. 2011;**123**:1375-1385

[124] Li B, Tian L, Zhang J, Huang L, Han F, Yan S, et al. Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*. BMC Genomics. 2014;**15**:1086

[125] Akond M, Liu S, Kantartzi SK, Meksem K, Bellaloui N, Lightfoot DA, et al. Quantitative trait loci for seed Isoflavone contents in "MD96-5722" by "Spencer" recombinant inbred lines of soybean. Journal of Agricultural and Food Chemistry. 2014;**62**:1464-1468

[126] Dong YS, Zhao LM, Liu B, Wang ZW, Jin ZQ, Sun H. The genetic diversity of cultivated soybean grown in China. Theoretical and Applied Genetics. 2004;**108**:931-936

[127] Cui Z, Carter TE, Burton JW, Wells R. Phenotypic diversity of modern Chinese and north American soybean cultivars. Crop Science. 2001;**41**:1954-1967

[128] Liu K. Soybeans: Chemistry, Technology and Utilization. New York: Chapman & Hall; 1997. p. 523

[129] Bernard RL, Hymowitz T, Cremeens CR. Registration of "Kunitz" soybean. Crop Science. 1991;**31**:232-233

[130] Schmidt MA, Hymowitz T, Herman EM. Breeding and characterization of soybean triple null; a stack of recessive alleles of Kunitz trypsin inhibitor, soybean agglutinin, and P34 allergen nulls. Plant Breeding. 2015;**134**(3):310-315

[131] Kim MS, Park MJ, Jeong WH, Nam KC, Chung JI. SSR marker tightly linked to the Ti locus in soybean [*Glycine max* (L.) Merr.]. Euphytica. 2006;**152**:361-366

[132] Rani A, Kumar V, Mourya V, Singh RK, Husain SM. Validation of SSR markers linked to null kunitz trypsin inhibitor allele in Indian soybean [*Glycine max* (L.) Merr.] population. Journal of Plant Biochemistry and Biotechnology. 2011;**20**:258-261

[133] Kumar V, Rani A, Rawal R, Mourya V. Marker assisted accelerated introgression of null allele of kunitz trypsin inhibitor in soybean. Breeding Science. 2015;**65**:447-452

[134] Jegadeesan S, Yu KF, Poysa V, Gawalko E, Morrison MJ, Shi C, et al. Mapping andvalidation of simple sequence repeat markers linked to a major gene controlling seed cadmium accumulation in soybean [*Glycine max* (L.) Merr]. Theoretical and Applied Genetics. 2010;**121**:283-294

[135] Ishikawa S, Ae N, Sugiyama M, Murakami M, Arao T. Genotypic variation in shoot cadmium concentration in rice and soybean in soils with different levels of cadmium contamination. Soil Science & Plant Nutrition. 2005;**51**:101-108

[136] Grant CA, Clarke JM, Duguid S, Chaney RL. Selection and breeding of plant cultivars to minimize cadmium accumulation. Science of the Total Environment. 2008;**390**:301-310

[137] Benitez ER, Hajika M, Takahshi R. Single-base substitution in P1B-ATPase. Gene is associated with a major QTL for seed cadmium concentration in soybean. Journal of Heredity. 2012;**103**:278-286