Chapter

Soybean Molecular Design Breeding

Xianzhong Feng and Muhammad Hafeez Ullah Khan

Abstract

Soybean is a globally important crop being rich source of edible oil and protein. Traditional phenotypic-based breeding procedures have contributed significantly to the development of several improved soybean varieties. In this context, molecular breeding technology, is seen as a viable way to address the issues and providing great opportunities to accelerate the process of soybean breeding. Hence, marker-assisted breeding (MAB) has been greatly applied in the soybean breeding to accelerate the improved soybean cultivars, transgenic breeding technology achieves great success in the soybean production. New genomics approaches and the development of genome editing technologies have increased soybean genetic diversity in its germplasm and have created new possibility to make precise genes modifications to controlling essential agronomic traits in an elite background Besides, the establishment of genotype driven phenotypic design breeding model has become a great challenge for soybean molecular breeding in the future. These approaches have the potential to expand the practical utility of molecular design breeding and speed up the germplasm and breeding materials in soybeans. This chapter goes into great detail about how current advances in genomics and phenomics can increase the efficiency and potential of MAB, transgenic technology, molecular design breeding and gene editing technology in soybean improvement.

Keywords: soybean, molecular design breeding, phenomics, genomics, genome editing

1. Introduction

Around 5000 years ago, cultivated soybean (*Glycine max* [L.] Merr.) had been domesticated from wild soybean (*Glycine soja Sieb*. & *Zucc*.). This crop has originated in China, and it spreads gradually around the different parts of the world [1]. Soybeans is now one of the most economically important oilseed and biodiesel crops, as well as a major source of protein and oil for human and animal consumption [2]. Early soybean breeding relied primarily on farmers selecting preferred seeds from the planted population. Artificial hybridization has been used since the early 1900s. In the 1940s, North American breeding programs published the first modern soybean cultivar developed through hybridization [3, 4]. Artificial hybridization became more commonly used in soybean breeding, after that it is investigated that artificial hybridization dramatically enlarged the genetic basis of established lines and increased soybean adaption as well as productivity [5]. Soybean is largely crushed into soy oil

and meal, and it can be found in a variety of edible and nonedible goods, including cooking oil, animal grains, vegan food, and milk, as well as biodiesel and other industrial applications. Soybean oil is the most widely used cooking oil in the world, second only to palm oil [6].

The major objective of the most plant breeding projects/programs in soybean is to increase the yield and quality [2]. However, in the field of plant breeding, measuring primary traits such as yield or quality, which are mostly complex quantitative traits in a large breeding populations with thousands of genotypes, is time-consuming and labor-intensive [7, 8]. Due to genetic and environmental influences, breeding for yield is recognized to be a highly complicated and nonlinear process [9]. To this end, plant breeders can efficiently identify the promising lines at early growth stages using secondary traits for selection (e.g., yield component traits and reflectance bands), which are strongly correlated with the primary trait [10, 11].

The recent advances in sequencing technology have triggered a data boom in the biology field, propelling molecular biology into a stunning postgenomic *era*. From structural characterization to functional analysis, genomic research has progressed [12]. Despite the fact that genomic mapping, bioinformatics prediction, and other technologies aid in inferring gene function; however, any theory in life science requires ultimate confirmation. This inference is required for genetic transformation and vice versa; and it appears to be a powerful tool in functional genomics. Transgenic breeding is other important approach used to introduce genetic changes for specific plant traits. This method has been successfully used to increase crop productivity, production of biofuels, improve food quality and plant resistance against severe environmental conditions by breaking species limits. Furthermore, the implementation of genome-editing tools such as CRISPR/Cas9 relies on transformation procedures, demonstrating the necessity and importance of this technology.

Marker-assisted selection (MAS) has speeded up the breeding process especially in the production of disease and insect pest-resistant cultivars [13]. Linkage and physical maps are created using various types of genetic markers [14, 15]. Consensus Map 4.0 was created to combine known genetic and physical maps [16]. Large numbers of quantitative trait loci (QTLs) associated with different crop traits have been identified in soybean using genetic markers. However, the efficiency and precision of QTL location were restricted by limited number of molecular markers and their uneven distribution. To this end, the advances in the high-throughput genotyping and phenomics have greatly enhanced the precision and resolution in the gene mapping [17, 18]. Although, the advances in high-throughput genotyping were significant to alleviate the challenges in the plant breeding [7, 19, 20], but the advances in the high-throughput field phenotyping, is far lagging behind the genomics. Hence, the phenomics is a major bottleneck in current breeding programs [19].

The mechanism of genome editing technology is to introduce double-strand breaks (DSBs) within the genome at targeted sites using sequence specific artificial nucleases (SSNs), which are then repaired using nonhomologous end joining (NHEJ) or homologous recombination repair (HR) mechanisms, resulting in targeted mutagenesis by adding, removing, or replacing DNA bases [21]. Zin finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas) are the most common SSNs at the moment [22, 23]. Despite their early development, ZFNs and TALENs are complex and expensive, which has limited their use. Since its inception, the CRISPR/Cas system has gained popularity in biological science due to its simplicity. The CRISPR/Cas9 system is the most well-known and has been

increasingly used in the crop plants in the last few years [24]. The genome editing toolset has been broadened after the CRISPR/Cas9 system by selecting Cas9 orthologs and created variations [25–27]. Dead *Cas9* and *Cas9* nickases are two of them, and they have been employed extensively in base editing, gene expression regulation, epigenome editing, cell imaging, and other domains [28].

2. Marker-assisted breeding in soybean

Marker-assisted breeding (MAB), also known as molecular-assisted breeding, is the use of molecular tools, primarily DNA markers, in conjunction with linkage maps and genomics to change and enhance plant as well as animal characteristics using genotypic tests [29]. The term MAB is used to explain the various novel strategies including MAS, marker-assisted backcrossing (MABC), marker-assisted recurrent selection (MARS), and genome-wide selection (GWS) or genomic selection (GS) [30]. MAB is recognized as a unique technique and a potent methodology for agricultural plant genetic modification, and it has been widely applied in a variety of crop species to date [29, 31].

Classical markers and DNA markers are two types of genetic markers used in plant breeding [32, 33]**.** Morphological, biochemical, and cytological indicators are examples of traditional markers. Random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellites or simple sequence repeats (SSRs), restriction fragment length polymorphism (RFLP), and single-nucleotide polymorphism (SNP) are all examples of DNA markers (SNP). Marker-assisted breeding (MAB) is the most promising of the many applications of DNA markers in plant science for cultivar creation. MAB has huge potential to increase conventional plant breeding efficiency and precision by using DNA markers that are firmly related to critical genes or loci [33]. Several allele-specific functional markers for essential soybean features such as blooming and maturity, pod dehiscence, aroma, salt tolerance, soybean cyst nematode oleic acid content, raffinose content, and Kunitz trypsin inhibitor have recently been discovered [33, 34]. Phytic acid content, glycinin, conglycinin concentration, fragrance, and lipoxygenase were also discovered as strongly connected markers for seed nutritional value, which could help with the selection of novel varieties that are free of antinutritional chemicals [33].

MAB allows selection of plant features (that are expressed late in the plants genotype) at the seedling stage based on the genotypic data; hence, reducing the time it takes to identify the phenotypic of a single plant. MAS can swiftly remove unwanted genotypes for features that are displayed at later developmental stages. This trait is very significant and valuable for backcrossing as well as recurrent selection breeding programs, which require crossing with or between chosen individuals [17, 30]. MAB is unaffected by the environment, allowing selection to take place in any setting (e.g., greenhouses and off-season nurseries). This is particularly useful for improving qualities that are only expressed in the presence of favorable environmental circumstances, such as disease/pest resistance as well as stress tolerance [30]. MAS is based on reliable markers that are strongly connected to the QTLs associated with particular trait of interest and is more effective and efficient than phenotypic selection for low-heritability traits that are highly influenced by the environment (PS). In the heterozygous state, MAB utilizes the codominant markers (e.g., SSR and SNP) to allow effective selection of recessive alleles of desirable features. To detect quality

controlled by recessive alleles, no selfing or test-crossing is required; hence, MAB may save time and speed up breeding process [29].

The MAB method significantly accelerates the accurate and efficient introgression of targeted genes into recipient varieties, as well as the recovery of the recurrent parent genetic background. With just two backcrosses ($BC_2F_{2:3}$), marker-assisted background selection in wheat was able to transfer *Yr15*, a stripe rust resistance gene in a recurrent variety and recover 97% of the genetic background of the recurrent parent, whereas phenotypic selection could only recover 82% of the background in BC_4F_7 plants [35, 36]. In this case, the MAB successfully saves the time it takes to obtain advanced breeding lines in half when compared with traditional approaches.

MAS and MABC have been frequently used to increase disease resistance and other relatively basic qualities [33, 37, 38]. MAB has been used successfully in a few soybean breeding programs to introduce single genic as well as polygenic traits into the desired genetic background (**Table 1**). Moreover, MAB has been proven to be effective in improving quantitative features that contribute to soybean nutritional value, such as seed protein content and oil quality. MAB for seed protein content (SPC) in soybeans using SSR markers yielded up to 9% transgressive segregation in the trait after two cycles [48, 49].

3. Transgenic breeding in soybean improvement

Transgenic soybeans are one of the few vegetable-based foods that contain all nine necessary amino acids. As a result, the transgenic soybean has grown in importance as a human and animal protein source, with 85% of its production going to animal feed, and the rest is going to direct human consumption [55]. Transgenic crops have been embraced by key soybean-growing countries such as the United States, Brazil,

Target trait	Gene/locus	Type of marker	References
Resistance to soybean mosaic virus	*Rsv1*, *Rsv3*, and *Rsv4*	SSR	[39, 40]
Resistance to soybean mosaic virus	R_{SC4}, R_{SC8}, and $R_{SC14Q,}$ SC_7	SSR	[41, 42]
High oleic acid content	*FAD2–1A*, *FAD2–1B*	Gene-specific Simple Probe	[43–45]
Grain yield	Yield QTL	SSR	[46]
Resistance to soybean cyst Nematode	*rhg1*, *Rhg4*	SSR	[47]
Seed protein content (SPC)	(QTL *Prot-08-1*)	SSR	[48, 49]
Salt tolerance	*GmSALT3*	SSR	[50–52].
Elimination of Kunitz trypsin inhibitor (kti)	*Ti3*	SSR	[33, 34]
Elimination of off-flavor and improvement of seed Longevity	*lox2*	lox2 specific	[53]
Low raffinose family oligosaccharides content	*RS3*	Gene-specific Simple Probe	[54]

Table 1.
Details of marker-assisted breeding conducted for improvement of soybean for various traits.

and Argentina, and they now account for about 85–95% of total soybean in terms of crop harvested area. The major markets for genetically modified agricultural seeds are North America and South America, which together account for more than 90% of the global GM seed industry. Nearly 85–95% of the soybean crop grown in North and South America is genetically modified. Demand for America's produced corn and soybean produce from other countries (particularly China) is a major factor in determining planted acreage and seed demand [56].

The landmark products of transgenic soybean's genetic composition allow it to be used for a wide range of purposes, which keeps it in high demand. Initially, manufacturers only wanted to use transgenics to grow more soy at a low cost in order to meet this demand, as well as to fix any problems in the growing process. But eventually it was discovered that soybean can be genetically modified to contain healthier components or even focus on one aspect of the soybean to produce in larger quantities. The first and second generations of genetically modified (GM) foods were named after these periods. The benefits of first-generation GM foods were oriented toward the manufacturing process and companies, whereas the second generation of GM foods offers a variety of advantages and added value for the consumer, including improved nutritional composition or even therapeutic effects [57]. The main and important landmark products of soybean are Roundup ready soybean, Generic GMO soybean, and genetic modification in soybean to improve soybean oil. Roundup Ready soybeans (the original variety was also known as GTS 40–3-2 (OECD UI: MON-04032-6)) are a series of glyphosate-resistant soybean cultivars developed by Monsanto. Glyphosate is a herbicide that kills plants by interfering with the production of phenylalanine, tyrosine, and tryptophan, which are all necessary amino acids. These amino acids are referred as "essential" since only plants and microorganisms can produce them, and mammals are dependent on the plants for these amino acids [58].

Soybean transgenic technology is an essential tool for validating the soybean gene function. Soybean genetic transformation has been explored for over two decades, but progress has been slow and inefficient, which is why some studies used Arabidopsis instead of soybean for functional validation. Several transformation systems have been developed, including shoot meristems [59], hypocotyls ([60], embryo [61], immature cotyledons, half-seed explants [62, 63], and cotyledonary nodes [64]. Agrobacterium-mediated cotyledonary node (CN) soybean transformation is currently widely employed due to its ease of usage, reproducibility, quantity of copies of foreign DNA, and low cost of experimentation [63, 65]. The overall average efficiency of transformation was 3.8–8.7% [63, 64]. Recently, the average transformation efficiency of soybean had been improved to 18.7% [66]. However, it is still lower than the 23% as reported in rice [67] and more than 30% found in maize [68].

Seed sterilization and germination followed by Agrobacterium infection, cocultivation of soybean explants and Agrobacterium, shoot induction, shoot elongation, rooting, and finally, moving the plants to pots containing soil are the different steps involved in the general transformation process. Many factors affect the transformation efficiency at the above different steps. For example, the soybean genotypes used in the transformation are the initial effector. The transformation efficiency and regeneration rate of 20 soybean varieties have been studied, and it was reported that transformation efficiency varied greatly (0.31–4.59%) among the different genotypes [69]. Second, all Agrobacterium concentrations, soybean explant status, Agrobacterium suspension medium, and cocultivation time will affect the infection efficiency during the Agrobacterium infection process, which is one of the most

significant processes. Explant regeneration is another important factor in affecting transformation efficiency. Plant hormone has been found to have a vital function in promoting explant regeneration, and the right dose could boost efficiency [64]. Previously research studies have showed that the combination of L-glutamine and L-asparagine in culture media increases the transformation efficiency by inhibiting *GmPRs* expression [70].

4. Soybean molecular design breeding

For molecular design breeding, the breeders could design the superior genotypes with particular breeding objectives based on the molecular networks regulating the agronomic traits. The breeding process can be simulated and optimized "in silico" before going to the field, which enhanced breeding accuracy and efficiency. Recent advances in genomics and phenomics have opened up new possibilities for more efficient molecular design breeding [84, 85]. Several soybean databases have been created for genomes, transcriptomics, proteomics, and germplasm analysis. SoyBase is a USDA-ARS database for genetics and genomics [78, 79]; SoyTEdb is a database of transposable elements [80]; SoyNet is a database for cofunctional networks [73]; SoyProDB is a database for seed proteins [81]; SoyPro [82]. These databases are quite useful for soybean molecular design breeding, which could provide the multiple levels of soybeans (**Table 2**).

4.1 Applications of genome selection in soybean design breeding

The genomics-assisted breeding (GAB) is one of important tools for molecular design breeding, which has allowed for higher genetic gain for complex traits at a lower cost, but it requires a molecular understanding of the trait [86]. MAS and GS are the two basic techniques used in GAB [87]. MAS is dependent on the presence of markers linked to the trait of interest, which can be discovered by linkage mapping or genome-wide association studies (GWASs). Many previous studies have shown that MAS may be successfully used in soybean by adding significant genes and large-effect QTLs for many attributes [88, 89]. Minor genes, on the other hand, control the majority of inheritance in complex characteristics, but they have never been studied because of the limitation of MAS [90]. Furthermore, the influence of the environment, epistatic interactions, and the effect of genetic background have made breeding complex traits extremely difficult. As a result, plant breeders have concluded that MAS is not an appropriate method for breeding complex plant characteristics [91].

GS uses the entire genome-wide marker profile of breeding lines to predict the genomics- estimated breeding value (GEBV) using several models, preventing the loss of a significant percentage of variation dictated by modest impact QTLs/genes [90]. However, precise genotyping and phenotyping analyses are required for accurate detection of marker-trait relationships and determination of GEBV, which determines the effectiveness of GAB. Manual low-throughput phenotyping and genotyping frequently result in the identification of false positives or negatives [37]. In this sense, high-throughput genotyping and phenotyping based on next-generation sequencing (NGS) enables for successful MAS and GS, as well as greater molecular design breeding programs success [92, 93]. The availability of high-throughput NGS-based genotyping methods has significantly speeded up the gene identification

Database	URL	Description	References
Soybean gene expression atlas	http://www.soybase.org/soyseq	A database of soybean 14 tissues specific gene expression	[71]
Soybean cDNA sequenced	http://digbio.missouriemi/soybean_atlas	A cDNA database of soybean developmental tissues specifically in root hair and meristem	[72]
SoyNet	http://www.inetbio.org/soynet	A database for network-based functional predictions	[73]
Soybean transcriptome data	http://venanciogroup.uuen.br/cgi-bin/gmax_atlas/index.cgi	A database of 1298 publicly available soybean transcriptome	[74]
Proteomics of oilseeds	http://oilseedproteomics.missurm.edu	Expression profile data for proteomics research on soybean and other oilseeds plants	[75]
Soybean Proteome Database	http://proteome.dc.affrc.go.jpSjpSoyb/	SPD a database of soybean proteomics	[76, 77]
SoyBase	http://www.soybase.org	A database of soybean genetics and genomics	[78, 79]
SoyTEdb	http://www.soytedb.org	A database of soybean transposable elements	[80]
SoyProDB	http://bioinformatics.towson.ede/Soybean_Seed_Proteins_2D_Gel_DB/Home.aspx	A database for soybean seed proteins	[81]
SoyProLow	http://bioinformatics.towson.ede/Soybean_low_abundance_prprotei_2D_Gel_DB/Gel1.aspx	A database for soybean low abundant proteins	[82]
SoyKB	http://soykb.org	A database of soybean translational genomics and for soybean molecular breeding	[83]

Table 2.
Resources and databases of soybean.

and GS, particularly in agricultural plants with bigger and more complex genomes, such as soybeans [87]. In this regard, phenomics and genomics are equally important for accurate gene identification and the development of a GS model to quantify the breeding population's GEBV (BP). Consequently, integrating these methodologies with suitable genetic diversity, soil and meteorological data, analytical tools, and databases, new varieties with improved yield, quality, and stress tolerance might be developed quickly [91].

MAS has not yielded satisfactory results in soybean for minor genes that contribute only a modest amount of obvious phenotypic variation for the complex trait [48, 87]. Most economically important soybean traits, including as yield, oil and protein content, and stress tolerance, are complex in nature, with modest effect genes controlling the majority of phenotyping variance for these traits [94]. The GS develops a prediction model by combining marker profile and phenotypic data from the training population, which is then used to estimate the GEBV of all BP individuals [95, 96]. Cross-validation on subsets of the training population is used to assess the accuracy of the prediction model before using it to select individuals from BP [87]. Following successful validation, this model can be used to select desirable plants from the BP based on GEBVs estimated solely from marker/genotypic data; hence, only genotypic data are utilized to predict the phenotypic performance of BP individuals [97].
The main benefit of GAB is that genotypic data collected at an early stage of plant development (such as seedling) can be utilized to predict phenotypic performance in mature individuals. As a result, it can significantly reduce the amount of time, money, and labor required for broad phenotypic examination across many habitats and years [98]. GAB also allows higher number of breeding selection cycles and genetic gain per unit time [87].

4.2 NGS-based genotyping for soybean design breeding

In recent decades, the total dependence on phenotypic selection has gradually shifted to a greater use of genotypic-based approaches for plant selection, facilitated by NGS-based genotyping platforms [99–101]. NGS technology has boosted throughput, speeed of genome-wide genotyping, and cost-effectiveness [102] (**Table 3**).
NGS-based genotyping technologies have tremendously aided in enhancing the resolution of gene mapping and tagging the gene/QTL extremely closer to the neighboring maker. In GWAS analysis, for example, the use of NGS has made it feasible to genotype huge populations of plants with a greater density of markers than was previously possible, which directly leads to better mapping resolution [110, 111]. In GWAS analysis, using a varied and big population allows for the discovery of more recombination break sites, which aids in the identification of candidate genes with greater precision [112]. Many studies have used NGS-based genotyping for GWAS analysis in soybean for various traits, and these studies have shown significant success in identifying candidate genes for specific traits of interest. For instance, previous study used the RAD-seq method to find a candidate gene underlying the main QTL controlling flooding tolerance in soybean [113]. Many other studies have shown that NGS-based genotyping facilitated candidate gene identification in areas such as nitrogen fixation [114], soybean plant height and primary branches [115], agronomic traits [116], disease resistance [117], and protein content [118]. The NGS-based WGRS has greatly improved the power of bulk segregant analysis (BSA) and its modified techniques, and it is now extensively employed in a variety of plant species, including soybean. For example, in another study, WGRS was employed to resequence different DNA pools

Trait	SNP genotyping platform	Number of SNPs	Population size	Training population	Prediction accuracy	Model	References
Yield and related traits	SNP Chip	2647	483	483	0.26–0.81	RR-BLUP	[97]
Amino acid contents	GBS	23,279	249	199	0.25–0.61	RR-BLUP	[100]
Chlorophyll content	SNP Chip	4089	172	100	0.31–0.76	RR-BLUP	[103, 104]
Grain yield	GBS	2395	139	55	0.64	G-BLUP	[105]
Yield-related traits	SNP Chip	4947	324	324	0.56	eBLUP	[99]
Yield-related trait	SNP Chip	5361	235	—	0.47–0.86	RR-BLUP	[106]
Yield and protein content	SNP Chip	4141	1248	252–492	0.55–0.62	G-BLUP	[107]
Yield	GBS	3000	227	227	0.6	RR-BLUP and Bayesian Models	[108]
Seed weight	SNP Chip	31,045	309	97–197	0.75–0.87	RR-BLUP	[109]
Nematode resistance	SNP Chip	3782	234	117–201	0.43–0.48	gBLUP	[103]

Table 3.
Tools of design breeding in soybean using high-throughput SNP genotyping platforms.

in the BSA study, and they discovered two significant genes that regulate cotyledon color in soybeans at the same time. Furthermore, many studies have used NGS-based techniques in the BSA approach to identify candidate genes for various soybean traits, including soybean mosaic virus [119], charcoal rot resistance [120], flowering time [121], phytophthora resistance [122], and powdery mildew resistance [123].

5. Genome editing provides the powerful tool for soybean design breeding

Genome editing has emerged as more powerful approach for functional study and molecular design breeding compared with traditional genetics approaches, namely mutagenesis, transgenic RNAi, or overexpression in obtaining plant cultivars with predictable and desired traits [124, 125]. CRISPR (clustered regularly interspaced short palindromic repeat)/Cas (CRISPR-associated) is one of the most efficient genome editing systems and has been widely used in various plant species [126]. In 2015, the first knockout and DNA homology-directed recombination (HDR) in soybean plant was successfully obtained using CRISPR/Cas9 technology [127]. Du et al. discovered that altering the AtU6–26 promoter of the CRISPR/Cas9 system to the GmU6-16g-1 promoter might considerably improve the efficacy of targeted mutagenesis in soybean [128]. Nearly 75% of the genes in soybean are duplicated, thus knocking out a single gene usually does not result in a mutant phenotype. It is critical to create a CRISPR/Cas9 system that can edit several homologous genes at the same time to obtain the desired phenotype. The CRISPR/Cas9 system that can achieve multiplex mutagenesis with better efficiency was established by refining the phases of vector synthesis, sgRNA assessment, pooled transformation, and sgRNA identification [129]. Naturally, single-nucleotide polymorphism (SNP) variants account for a major portion of phenotypic variability in agronomic traits, in addition to alleles induced by loss-of-function mutations. When a gene function is disrupted by utilizing a gene-editing technique, mostly it results in undesirable phenotype that is difficult to optimize for agronomic trait improvement [27]. As a result, in molecular breeding, generating point mutations at specific locations impacting crucial agronomic properties is extremely valuable [130]. The CRISPR/Cas9 system has recently been used to develop "base editing," which changes single bases into another without the use of DNA DSBs or a donor template [131]. Cai et al. successfully used the technique to generate the point mutants of *GmFT2a* and *GmFT4* in soybean [132].

Currently, CRISPR/Cas9 is being frequently used in soybean functional studies [128, 133, 134]. For example, to identify the genes that are responsible for flowering time, frame shift mutations created by CRISPR/Cas9 revealed that *GmFT2a* functions primarily during short day (SD), while *GmFT5a* has more substantial impacts under long day (LD) [132, 135, 136]. Similarly, knockout of *GmPRR37* by the CRISPR system suggested that it can repress flowering under LD [137]. Male sterility was seen in two CRISPR/Cas9 gene-editing mutants of *Glyma.13G114200*, showing that it was the casual gene *GmMS1* responsible for male sterility [138–140].

5.1 The advantages of genome editing for soybean molecular design breeding

A key issue and important research goal for soybean researchers since the completion of the soybean genome sequencing project is to elucidate the function of 46–56 thousand identified genes [129]. Transgenic technology is a great tool for functional

genomic research and crop genetic enhancement, but it has certain drawbacks when used in soybeans. *Agrobacterium-mediated transformation* and particle bombardment have been widely utilized to make transgenic soybean plants in recent decades. When considering its easy technique, low cost, single or low copy number of insertions, and relatively infrequent rearrangement, *Agrobacterium-mediated transformation* is a superior alternative [141]. However, because soybean is a refractory crop as far as the transformation and regeneration are considered, no sustained soybean genetic transformation has yet been developed, regardless of whatever technique is used. Furthermore, the efficacy of Agrobacterium-mediated transformation is affected by soybean tissue, cultivar, or species [142], resulting in a restricted number of soybeans types that may be enhanced directly by genetic modification. Transgenic technology is used to investigate gene function by integrating foreign DNA sequences into the plant genome, which results in either overexpression or silence of the target gene [142]. As a result of potential risks such as unintended gene insertions, endogenous gene disruption, and unpredictable gene expression that arise during transformation [127], transgenic plants frequently cause potential bio-safety issues and are subjected to regulatory restrictions on genetically modified organisms (GMOs).

Furthermore, being a diploid that developed from palaeotetraploid, soybean has a highly duplicated genome, with around 75% of projected genes possess multiple copies, resulting in substantial genetic redundancy and complicating the elucidation of soybean gene function. On the one hand, conventional random mutagenesis methods (physical, chemical mutagenesis, or T-DNA insertion) make it difficult to link genotype and phenotype because the loss of one homolog can be fully compensated by redundant homologous copies [143]. On the other hand, RNA-silencing-based technology frequently silences the entire gene family and is difficult to silence a single gene copy [144], and due to partial gene product depletion, the phenotype may be unstable [145]. To accelerate soybean gene function and breeding research, more accurate and efficient genetic engineering technology is required. Current genome editing technology has provided opportunity to overcome the aforesaid difficulties, at least in part. For starters, genome editing technology can not only alter a single gene without impacting other members of the gene family, but it can also edit many genes of interest at the same time using a single transformation, making it ideal for soybean and other polyploid crop studies [126].

Secondly, unlike transgenic technology, genome editing technology incorporates sequences that are genome editing components rather than foreign genes of interest into plant genomes. These genome editing components may be deleted once the target gene has been edited to yield transgene-free mutants, which are safer to employ in breeding and easier to commercialize under tight GMO rules. For example, in the United States, a gene-edited soybean oil with a different fatty acid profile was recently released (https://calyxt.com/first-commercial-sale-of-calyxt-high-oleic-soybean-oil-on-the-usmarket/).

Finally, because the soybean genetic transformation is inefficient, one possible approach is to do direct genome editing without transformation or tissue culture. The standard procedure is to edit a variety with a high transformation efficiency to enhance one or more characteristics and then utilize the modified plant as a donor, which contains editing components. By hybridization with donor plants and subsequent backcrossing, the modified target gene or genome editing components can be introgressed into elite lines resistant to transformation. In conclusion, genome editing technology is a strong tool that has a lot of potential for speeding up soybean breeding [137, 146].

5.2 Applications of genome editing in soybean improvement

Hairy-root transformation mediated by *Agrobacterium rhizogenes* is a simple and rapid technique for studying soybean gene function, and it only takes a few weeks to get transgenic hairy root [142, 147]. When ZFNs, TALENs, and CRISPR/Cas9 systems were initially employed in soybean, researchers preferred to use the hairy-root transformation technique to quickly assess the efficiency of these genome editing tools before undertaking the time-consuming soybean whole-plant transformation [142, 147]. As a consequence, ZFNs technology delivered the first example of genome editing in soybean in 2011, producing heritable mutations in two homologous *DICER-LIKE genes,* namely *DCL4a* and *DCL4b* [144]. In 2015, five research groups successfully evaluated the mutation efficiency of the CRISPR/Cas9 system in soybean endogenous or exogenous genes in hairy roots, establishing a precedent for using CRISPR/Cas technology to study soybean genes [127, 148, 149]. One of these teams used NHEJ to accomplish about 76% targeted mutagenesis, and they also used HR to create mutant plants with targeted gene integration at the target location, as well as a chlorsulfuron-resistant soybean with a mutated acetolactate synthase1 (ALS1) gene [127]. All of these early successes have highlighted the genome editing systems' remarkable potential to develop useful features in the near term through focused gene alterations.

Scientists evaluated the effectiveness of TALENs and CRISPR/ Cas9 in editing two phytoene desaturase genes in hairy roots, namely *GmPDS11* and *GmPDS18* [128]. The results demonstrated that CRISPR/Cas9, particularly CRISPR/Cas9 employing the GmU6-16g-1 promoter, was far more effective than TALENs at concurrently targeting two alleles. Another study was carried out to examine a variety of GmU6 promoters in soybean hairy roots and *Arabidopsis thaliana* to determine which ones were best for driving sgRNA production and discovered that GmU6-8 and GmU6-10 promoters had high activity, which improved editing effectiveness [150]. Both results are beneficial in the development of an effective CRISPR/Cas9 system for use in soybean research. In addition, the CRISPR/Cas9 method in combination with the hairy-root transformation technique has been used to edit and explore soybean storage protein genes [134], and the candidate gene governing nodulation specificity [151].

NHEJ has generally been used to delete target genes by tiny insertions or deletions, whereas HR is primarily utilized to replace or integrate targeted genes [152]. Using dual-sgRNA to cleave two neighboring loci on the same chromosome, two research studies recently showed that massive genomic deletions might be generated in soybeans [135, 153]. Due to the two editing chances, the dual-sgRNA design can not only boost gene mutation rates, but also construct substantial fragment deletion to ensure that the target gene is completely eliminated. Other fields of research will benefit from the substantial loss, such as understanding the role of regulatory elements or noncoding genes. Unlike prior HR-mediated donor integration, one study claimed that two large multigene donors (7.1 kb and 16.2 kb) were inserted into a target genomic location of soybean utilizing NHEJ and ZFNs, with donor heritability verified in T1 progeny plants [152]. This study found that NHEJ might be used instead of HR to induce accurate insertions of numerous transgenes in soybeans while avoiding the drawbacks of inefficient HR [154].

In addition to the abovementioned developments in genome editing systems to permit better application in soybean research, several researchers have used genome editing technology and entire plant transformation to investigate gene function or enhance agronomic features. The single and double mutants of the soybean genes, namely *DCL1a* and *DCL1b* using ZFNs-based mutagenesis techniques were created

to investigate their roles in the soybean miRNA system [143]. Only the *dcl1a/dcl1b* double mutants showed a morphological phenotype, which was characterized by shrunken and shriveled seed as well as seedling developmental abnormalities, whereas both *dcl1a* and *dcl1b* single mutants showed a normal phenotype, suggesting that *GmDCL1* homologs have functional redundancy. Besides, the two homologous genes, namely *GmPPD1* and *GmPPD2*, coding for Arabidopsis PEAPOD orthologs were edited with a single sgRNA [155]. As a result, double mutants with frame shift mutations displayed a severe phenotype of developmental abnormalities in leaf and pod development.

Previous results revealed that homozygous mutants produced by CRISPR/Cas9-mediated mutagenesis in *FLOWERING LOCUS T2a (GmFT2a)* delayed soybean flowering [133] and male sterile soybean lines [156]. They also used the CRISPR/Cas9 system to alter *GmFT5a* and then crossed it with *ft2a* mutants to get *ft2aft5a* double mutants. This double mutant bloomed 31.3 days later than wild-type plants and produced more pods as well as seeds under short day circumstances [136]. The researchers employed a multiplex genome editing method based on CRISPR/Cas9 to change four *SQUAMOSA PROMOTER BINDING PROTEIN-LIKE9* (*GmSPL9*) genes and generated several soybean mutants with different mutated locus combinations. These mutants revealed altered node number on the main stem and branch number [157]. CRISPR/Cas9 technology has also enhanced soybean seed oil profile [158], offensive beany flavor of soybean seed product [159], and isoflavone content and resistance to soybean mosaic virus [160]. The experiments mentioned above showed that genome editing technology has a lot of potential for improving soybeans.

Despite the fact that CRISPR/Cas9 technology has become the standard for genome editing, NGG PAM's fundamental restriction limits its use in highly precise genomic areas. ZFNs and TALENs have crucial tools as a complement to the CRISPR/Cas9 system since their target range is infinite. The combined usage of these systems will aid the soybean breeding and functional genomics projects. For example, a group of genes were mutated that encode the main machinery proteins involved in small RNA processing in soybean and *Medicago truncatula* using CRISPR/Cas9 and TALENs [161]. Together with the previously reported mutants induced by ZFNs, the resultant mutant plants established a collection of mutant resources for future studies of short RNA biology in legume plants [143].

Many scientists used CRISPR/Cas9-based high-throughput mutagenesis to create a genome-wide mutant library or mutant collection related to certain functions for pooled functional screen, which provided a wealth of resource in mammalian cell, rice, and tomato [162–164]. The success of creating these CRISPR libraries, however, is strongly reliant on a reliable genetic transformation mechanism. The creation of a CRISPR library in soybean is not only a massive task, but it is also extremely important for research and breeding. A first soybean CRISPR library targeting over 100 candidate genes was recently developed. A collection of mutant soybean lines was also created utilizing an enhanced process at several important steps, demonstrating the feasibility and usefulness of using CRISPR/Cas9 technology to execute large-scale mutagenesis in soybeans [129]. Following the CRISPR/Cas9 system, new genomic editing tools such as the CRISPR/Cas12a system, BE systems, and other CRISPR/Cas variants have been developed. However, to our knowledge, most of these have yet to be employed in soybean, with the exception of one work [165] that used Cas12a-RNP in soybean protoplasts to induce gene editing. RNP-based genome editing may not be the best option for soybean research due to a lack of updated information on successful protoplast regeneration (**Figure 1**).

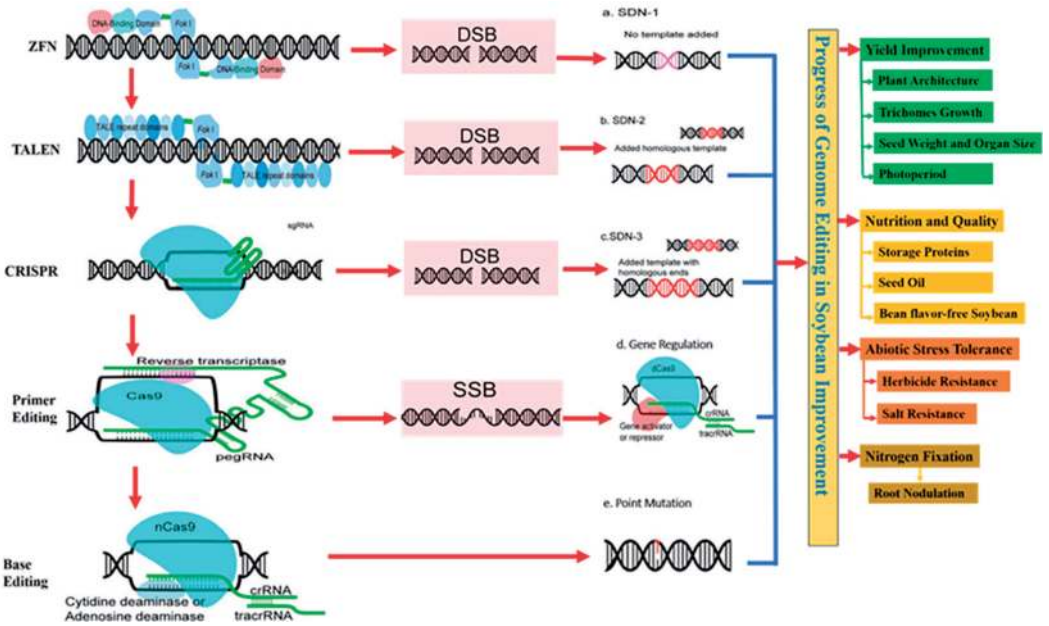

Figure 1.
Genome editing platforms and editing outcomes. Each editing platform (arrow) and its outcomes (rectangular) are coded with the same color. ZFN, zinc-finger nuclease; TALEN, transcription activator-like effector nuclease; CRISPR, clustered regulatory interspaced short palindromic repeat; DSB, double-strand breaks; SSB, single-strand breaks; outcomes of GE created by site-directed nucleases (SDN) include: SDN1 – The approach involves DNA breaks repair through DNA repair mechanisms in the host cellular without using an added repair template; SDN2 – The approach involves the break repair via HR using an added homologous repair template; and SDN3 – The approach involves DNA break repair via either HDR or NHEJ pathway using an added DNA template containing no homologous sequences but with homologous ends: Progress and application of genome editing in soybean improvement.

6. Conclusion and future prospective

Traditional breeding has made important contributions to soybean improvement and the generation of soybean varieties with enhanced yield, quality, and tolerance to numerous stressors throughout the last century. Traditional crop development technologies, on the other hand, are unable to keep up with the world's rapidly rising population and climate change [166, 167]. Reduced generation time allows for faster soybean breeding, which may be accomplished by the quick creation of homozygous lines employing the doubled-haploid (DH) production process. The creation of a high-throughput DH production program in soybean would be tremendously beneficial in achieving the crop's targeted genetic gain. Soybean androgenesis, root development, and unusual shoot induction have all shown slight advancements. However, there is currently no efficient, repeatable way for producing doubled haploids in soybean. One of the primary impediments to the development of a commercial DH production procedure in soybeans may be the tissue's resistance to *in vitro* regeneration [168, 169]. To achieve a sustainable yield, it is necessary to identify genetic resources in the form of water-stress-resistant soybean genotypes and genomics-assisted water-stress mitigation approaches. Several techniques, such as QTL mapping, genome-wide association mapping, and comparative transcriptome analyses, are being used to determine the genetic basis for water-stress tolerance in soybean [170, 171].

Recent advances in NGS-based genotyping technologies and powerful computational pipelines have significantly reduced the cost of WGS/WGRS, allowing the

discovery, sequencing, and genotyping of hundreds of thousands of markers in one step. For large-scale marker identification, NGS-marker technologies based on reduced representation sequencing are the ideal solution, especially for the huge and complex soybean genome. These NGS-based marker approaches represent the soybean's partial genome, and they can even be used without a reference sequence. RAD-seq (or its variations) and GBS are two NGS technologies that have previously proven to be efficient and effective procedures for GAB and have been widely employed for GS investigations in various agricultural plants. Furthermore, the NGS has enabled the fabrication of high-density SNP chips for HTG in soybean. The low cost, genome-wide marker coverage, better speed and throughput, and higher marker density of NGS-marker technologies have allowed geneticists to explore the inheritance of numerous traits at the nucleotide level accuracy.

On the other hand, GS employs a number of markers spread over the entire genome to forecast the breeding value of a breeding line for selection. GS can quantify Mendelian sampling without phenotyping the entire population thanks to genome-wide dense markers. It shortens cycles to save time while also increasing genetic gain per unit of time. GS was compared with traditional phenotypic selection in soybean to see if it has any advantages in terms of accuracy and time savings.

GE technologies, particularly CRISPR-based systems, have advanced quickly, with the majority of them being implemented to give effective tools for soybean improvement. If this technique is properly implemented in plant breeding programs, a recent field trial of high oleic soybean employing TALENs has indicated the bright future of soybean improvement. Currently, the discovery of more GE target genes associated with agronomically important traits, the adoption of newly developed GE technologies, the simplification and renovation of editing reagent delivery, and the improvement of target mutant recovery method in soybean will improve editing outcomes, save time, and lower product development costs. The development of GE products will be aided by the cost-effective preparation of intellectual property for GE technologies, as well as breeders' and farmers' comprehension of GE-related government regulation. In several countries, transgene-free or DNA-free edited plants are considered nongenetically modified events, making GE soybean production easier. In future, more applications of "base editing" for single genes or several genes at once would substantially aid functional research and molecular design breeding in soybean.

Next-generation GAB in agricultural plants has been enabled by recent advances in crop phenomics and genomics, which have provided several high-throughput platforms, as well as statistical approaches and computational tools. When these modern technologies are integrated, they can precisely and accurately identify genes/QTLs, as well as their beneficial usage in soybean breeding [172, 173]. Despite the fact that high-throughput SNP genotyping technologies have completely revolutionized marker application in soybean breeding, they have enabled research groups to apply GWAS and GS for soybean improvement on a regular basis. These marker technologies, however, must be paired with HTP to produce meaningful genetic gain from complex features in order to reap the full benefits of genomic investigations. So far, only a few studies involving the use of both HTP and HTG in soybean have been reported. This is because large-scale field-based HTP has a greater cost. New advances in crop phenotyping technology are expected to make HTP more inexpensive for commercial application in soybean breeding projects in the near future. This would undoubtedly increase the scope of germplasm assessment and facilitate the development of better soybean cultivars. WGRS-based genotyping will become increasingly

viable and cost-effective as the cost of DNA sequencing falls. Sequencing-based genotyping employing genome-reduction methods such as GBS and RAD-sequencing appears to be more cost-efficient for breeding-based applications such as GS at the moment. Since the cultivated soybean has a limited genetic basis, genome editing and TILLING can be used to produce a variety of changes in these orthologs, from knock-down to knockout alleles. For quick deployment of these alleles in breeding programs, it should be combined with the speed breeding facility.

The ZFNs, TALENs, CRISPR/Cas9, CRISPR/Cas12a, BEs, and other CRISPR/Cas variations provide a robust genome editing toolkit that will aid future functional genomic and genetic improvement studies in soybeans and other plants. CRISPR/Cas9 technology may become the preferred method for soybean breeding due to its efficiency, multiplex editing, and high-throughput mutagenesis capabilities, as well as its maturity. With the progress of additional genome editing methods, however, soybean genome editing will become more versatile. Despite the fact that substantial effort may be required to employ these techniques, given the enormous potential of genome editing and the economic importance of soybean, we anticipate that these issues will be resolved in the near future.

Acknowledgements

This research was funded by the Zhejiang Lab (Grant No. 2021PE0AC04) and Yazhou Bay Seed Lab (Grant No. B21HJ0101).

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author details

Xianzhong Feng* and Muhammad Hafeez Ullah Khan
Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

*Address all correspondence to: fengxianzhong@iga.ac.cn

References

[1] Wilson RF. Soybean: Market driven research needs. In: Genetics and Genomics of Soybean. Springer; 2008. pp. 3-15

[2] Hartman GL, West ED, Herman TK. Crops that feed the world 2. Soybean worldwide production, use, and constraints caused by pathogens and pests. Food Security. 2011;**3**:5-17

[3] Rincker K, Nelson R, Specht J, Sleper D, Cary T, Cianzio SR, et al. Genetic improvement of US soybean in maturity groups II, III, and IV. Crop Science. 2014;**54**:1419-1432

[4] Wolfgang G, An YC. Genetic separation of southern and northern soybean breeding programs in North America and their associated allelic variation at four maturity loci. Molecular Breeding. 2017;**37**:1-9

[5] Anderson EJ, Ali ML, Beavis WD, Chen P, Clemente TE, Diers BW, et al. Soybean [*Glycine max* (L.) Merr.] breeding: History, improvement, production and future opportunities. In: Advances in Plant Breeding Strategies: Legumes. Springer; 2019. pp. 431-516

[6] Didinger C, Foster MT, Bunning M, Thompson HJJDB. Pulses: Production P, nutrition. Nutrition and human health benefits of dry beans and other pulses. Dry beans pulses: Production, processing. Nutrition. 2022:481-504

[7] Araus JL, Cairns JE. Field high-throughput phenotyping: The new crop breeding frontier. Trends in Plant Science. 2014;**19**:52-61

[8] Xiong Q, Tang G, Zhong L, He H, Chen X. Response to nitrogen deficiency and compensation on physiological characteristics, yield formation, and nitrogen utilization of rice. Frontiers in Plant Science. 2018;**9**:1075

[9] Collins NC, Tardieu F, Tuberosa R. Quantitative trait loci and crop performance under abiotic stress: Where do we stand? Plant Physiology. 2008;**147**:469-486

[10] Jin J, Liu X, Wang G, Mi L, Shen Z, Chen X, et al. Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China. Field Crops Research. 2010;**115**:116-123

[11] Montesinos-López OA, Montesinos-López A, Crossa J, de Los Campos G, Alvarado G, Suchismita M, et al. Predicting grain yield using canopy hyperspectral reflectance in wheat breeding data. Plant Methods. 2017;**13**:1-23

[12] Medini D, Serruto D, Parkhill J, Relman DA, Donati C, Moxon R, et al. Microbiology in the post-genomic era. Nature Review Microbiology. 2008;**6**:419-430

[13] Li MW, Wang Z, Jiang B, Kaga A, Wong FL, Zhang G, et al. Impacts of genomic research on soybean improvement in East Asia. Theoretical and Applied Genetics. 2020;**133**:1655-1678

[14] Chan C, Qi X, Li MW, Wong FL, Lam HM. Recent developments of genomic research in soybean. Journal of Genetics and Genomics. 2012;**39**:317-324

[15] Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, et al. A new integrated genetic linkage

map of the soybean. Theoretical and Applied Genetics. 2004;**109**:122-128

[16] Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, et al. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics. 2010;**11**:1-8

[17] Song X, Wei H, Cheng W, Yang S, Zhao Y, Li X, et al. Development of INDEL markers for genetic mapping based on whole genome resequencing in soybean. G3: Genes, Genomes, Genetics. 2015;**5**:2793-2799

[18] Bhat JA, Yu D, Bohra A, Ganie SA, Varshney RK. Features and applications of haplotypes in crop breeding. Communications Biology. 2021;**4**:1-12

[19] Araus JL, Kefauver SC, Zaman-Allah M, Olsen MS, Cairns JE. Translating high-throughput phenotyping into genetic gain. Trends in Plant Science. 2018;**23**:451-466

[20] Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M. Plant Phenomics, from sensors to knowledge. Current Biololgy. 2017;**27**:R770-RR83

[21] Pacher M, Puchta H. From classical mutagenesis to nuclease-based breeding–directing natural DNA repair for a natural end-product. The Plant Journal, Wiley Online Library. 2017:819-833

[22] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;**339**:819-823

[23] Khan MHU, Khan SU, Muhammad A, Hu L, Yang Y, Fan C. Induced mutation and epigenetics modification in plants for crop improvement by targeting CRISPR/Cas9 technology. Joournal of Cellular Physiology. 2018;**233**:4578-4594

[24] Chen K, Gao C. Targeted genome modification technologies and their applications in crop improvements. Plant Cell Reports. 2014;**33**:575-583

[25] Cebrian-Serrano A, Davies B. CRISPR-Cas orthologues and variants: Optimizing the repertoire, specificity and delivery of genome engineering tools. Mammalian Genome. 2017;**28**:247-261

[26] Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;**163**:759-771

[27] Wang T, Xun H, Wang W, Ding X, Tian H, Hussain S, et al. Mutation of *GmAITR* genes by CRISPR/Cas9 genome editing results in enhanced salinity stress tolerance in soybean. Frontiers in Plant Science. 2021:2752

[28] Moradpour M, Abdulah SNA. CRISPR/dC as9 platforms in plants: Strategies and applications beyond genome editing. Plant Biotechnology Journal. 2020;**18**:32-44

[29] Jiang GL. Molecular markers and marker-assisted breeding in plants. In: Plant Breeding from Laboratories to Fields. Vol. 3. London, UK: IntechOpen; 2013. pp. 45-83

[30] Ribaut JM, de Vicente MC, Delannay X. Molecular breeding in developing countries: Challenges and perspectives. Current Opinion in Plant Biology. 2010;**13**:213-218

[31] Moose SP, Mumm RH. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology. 2008;**147**:969-977

[32] Xu Y. Molecular Plant Breeding. CABI; 2010

[33] Mundhe S, Patil R, Oak M, Jaybhay S. Accelerating soybean improvement through genomics-assisted breeding. In: Accelerated Plant Breeding. Vol. 4. Springer; 2022. pp. 41-62

[34] Kumawat G, Gupta S, Ratnaparkhe MB, Maranna S, Satpute GK. QTL omics in soybean: A way forward for translational genomics and breeding. Frontiers in Plant Science. 2016;**7**:1852

[35] Randhawa HS, Mutti JS, Kidwell K, Morris CF, Chen X, Gill KS. Rapid and targeted introgression of genes into popular wheat cultivars using marker-assisted background selection. PLoS One. 2009;**4**:e5752

[36] Badri J, Lakshmidevi G, JaiVidhya L, Prasad MS, Laha GS, Lakshmi VJ, et al. Multiparent-derived, Markerassisted introgression lines of the elite Indian rice cultivar,'Krishna Hamsa' show resistance against bacterial blight and blast and tolerance to drought. Plants. 2022;**11**:622

[37] Tuberosa R. Marker-assisted breeding in crops. In: Encyclopedia of Sustainability Science Technology. 2012. pp. 6425-6448

[38] Swarup S, Cargill EJ, Crosby K, Flagel L, Kniskern J, Glenn KC. Genetic diversity is indispensable for plant breeding to improve crops. Crop Science. 2021;**61**:839-852

[39] Shi A, Chen P, Li D, Zheng C, Zhang B, Hou A. Pyramiding multiple genes for resistance to soybean mosaic virus in soybean using molecular markers. Molecular Breeding. 2009;**23**:113-124

[40] Widyasari K, Tran P-T, Shin J, Son H, Kim K-H. Overexpression of purple acid phosphatase *GmPAP2*. 1 confers resistance to soybean mosaic virus in a susceptible soybean cultivar. Journal of Experimental Botany. 2022;**73**:1623-1642

[41] Wang D, Lin Z, Kai L, Ying M, Wang L, Yang Y, et al. Marker-assisted pyramiding of soybean resistance genes RSC4, RSC8, and RSC14Q to soybean mosaic virus. Journal of Integrative Agriculture. 2017;**16**:2413-2420

[42] Zhang Y, Song J, Wang L, Yang M, Hu K, Li W, et al. Identifying quantitative trait loci and candidate genes conferring resistance to soybean mosaic virus SC7 by quantitative trait loci-sequencing in soybean. Frontiers in Plant Science. 2022;**13**

[43] Pham AT, Lee JD, Shannon JG, Bilyeu KD. Mutant alleles of *FAD2-1A* and *FAD2-1B* combine to produce soybeans with the high oleic acid seed oil trait. BMC Plant Biology. 2010;**10**:1-13

[44] Zhang H, Jiang H, Hu Z, Song Q, An YC. Development of a versatile resource for post-genomic research through consolidating and characterizing 1500 diverse wild and cultivated soybean genomes. BMC Genomics. 2022;**23**:1-13

[45] Pham AT, Lee JD, Shannon JG, Bilyeu KD. A novel *FAD2-1A* allele in a soybean plant introduction offers an alternate means to produce soybean seed oil with 85% oleic acid content. Theoretical Applied Genetics. 2011;**123**:793-802

[46] Sebastian S, Streit L, Stephens P, Thompson J, Hedges B, Fabrizius M, et al. Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Science. 2010;**50**:1196-1206

[47] Santana FA, Silva MFD, Guimarães JKF, Ferreira MFDS,

Pereira WD, Piovesan ND, et al. Marker-assisted selection strategies for developing resistant soybean plants to cyst nematode. Crop Breeding Applied Biotechnology. 2014;**14**:180-186

[48] Zhang YH, Liu MF, He JB, Wang YF, Xing GN, Li Y, et al. Marker-assisted breeding for transgressive seed protein content in soybean [*Glycine max* (L.) Merr.]. Theoretical Applied Genetics. 2015;**128**:1061-1072

[49] Zhang S, Du H, Ma Y, Li H, Kan G, Yu D. Linkage and association study discovered loci and candidate genes for glycinin and β-conglycinin in soybean (*Glycine max L. Merr*.). Theoretical Applied Genetics. 2021;**134**: 1201-1215

[50] Liu Y, Yu L, Qu Y, Chen J, Liu X, Hong H, et al. *GmSALT3*, which confers improved soybean salt tolerance in the field, increases leaf Cl^- exclusion prior to Na^+ exclusion but does not improve early vigor under salinity. Frontiers in Plant Science. 2016;**7**:1485

[51] Akanbi M, Egbichi I, Ludidi N. Comparative analysis of responses to field salinity stress in contrasting soybean accessions highlights NaCl exclusion in leaves as a key mechanism for salinity stress tolerance. Journal of Oasis Agriculture and Sustainable Development. 2021:19-24

[52] Karahara I, Horie T. Functions and structure of roots and their contributions to salinity tolerance in plants. Breeding Science. 2021;**71**:89-108

[53] Rawal R, Kumar V, Rani A, Gokhale SM. Genetic elimination of off-flavour generating lipoxygenase2 gene of soybean through marker assisted backcrossing and its effect on seed longevity. Plant Breeding Biotechenolgy. 2020;**8**:163-173

[54] Hagely KB, Jo H, Kim JH, Hudson KA, Bilyeu K. Molecular-assisted breeding for improved carbohydrate profiles in soybean seed. Theoretical Applied Genetics. 2020;**133**:1189-1200

[55] Qin P, Wang T, Luo Y. A review on plant-based proteins from soybean: Health benefits and soy product development. Journal of agriculture. Food Research. 2022:100265

[56] Baghbani-Arani A, Poureisa M, Alekajbaf H, Borz-Abad RK, Khodadadi-Dashtaki K. Current status of genetically modified crops in Iran and the world: Overview of production and consumption challenges. 2021. DOI: https://doi.org/10.21203/rs.3.rs-141395/v1

[57] Celec P, Kukučková M, Renczésová V, Natarajan S, Pálffy R, Gardlík R, et al. Biological and biomedical aspects of genetically modified food. Biomedicine Pharmacotherapy. 2005;**59**:531-540

[58] Zhang M, Liu S, Wang Z, Yuan Y, Zhang Z, Liang Q, et al. Progress in soybean functional genomics over the past decade. Plant Biotechnology Journal. 2022;**20**:256

[59] Rech EL, Vianna GR, Aragao FJ. High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. Nature Protocols. 2008;**3**:410-418

[60] Wang G, Xu Y. Hypocotyl-based Agrobacterium-mediated transformation of soybean (*Glycine max*) and application for RNA interference. Plant Cell Reports. 2008;**27**:1177-1184

[61] Trick H, Finer J. Sonication-assisted Agrobacterium-mediated transformation of soybean [*Glycine max (L.) Merr*] embryogenic suspension culture tissue. Plant Cell Reports. 1998;**17**:482-488

[62] Liu HK, Yang C, Wei ZM. Efficient Agrobacterium tumefaciens-mediated transformation of soybeans using an embryonic tip regeneration system. Planta. 2004;**219**:1042-1049

[63] Paz MM, Martinez JC, Kalvig AB, Fonger TM, Wang K. Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. Plant Cell Reports. 2006;**25**:206-213

[64] Li S, Cong Y, Liu Y, Wang T, Shuai Q, Chen N, et al. Optimization of Agrobacterium-mediated transformation in soybean. Frontiers in Plant Science. 2017;**8**:246

[65] Somers DA, Samac DA, Olhoft PM. Recent advances in legume transformation. Plant Physiology. 2003;**131**:892-899

[66] Pareddy D, Chennareddy S, Anthony G, Sardesai N, Mall T, Minnicks T, et al. Improved soybean transformation for efficient and high throughput transgenic production. Transgenic Research. 2020;**29**:267-281

[67] Ge X, Chu Z, Lin Y, Wang S. A tissue culture system for different germplasms of indica rice. Plant Cell Reports. 2006;**25**:392-402

[68] Yang A, He C, Zhang K. Improvement of Agrobacterium-mediated transformation of embryogenic calluses from maize elite inbred lines. In Vitro Cellular Developmental Biology-Plant. 2006;**42**:215-219

[69] Song ZY, Tian JL, Fu WZ, Li L, Lu LH, Zhou L, et al. Screening Chinese soybean genotypes for Agrobacterium-mediated genetic transformation suitability. Journal of Zhejiang University SCIENCE B. 2013;**14**:289-298

[70] Chen L, Cai Y, Liu X, Yao W, Guo C, Sun S, et al. Improvement of soybean *Agrobacterium-mediated transformation* efficiency by adding glutamine and asparagine into the culture media. International Journal of Molecular Sciences. 2018;**19**:3039

[71] Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, et al. RNA-Seq atlas of *Glycine max*: A guide to the soybean transcriptome. BMC Plant Biology. 2010;**10**:1-16

[72] Libault M, Zhang XC, Govindarajulu M, Qiu J, Ong YT, Brechenmacher L, et al. A member of the highly conserved FWL (tomato FW2. 2-like) gene family is essential for soybean nodule organogenesis. The Plant Journal. 2010;**62**:852-864

[73] Kim E, Hwang S, Lee I. SoyNet: A database of co-functional networks for soybean *Glycine max*. Nucleic Acids Research. 2017;**45**: D1082-D10D9

[74] Machado FB, Moharana KC, Almeida-Silva F, Gazara RK, Pedrosa-Silva F, Coelho FS, et al. Systematic analysis of 1298 RNA-Seq samples and construction of a comprehensive soybean (Glycine max) expression atlas. The Plant Journal. 2020;**103**:1894-1909

[75] Agrawal GK, Hajduch M, Graham K, Thelen JJ. In-depth investigation of the soybean seed-filling proteome and comparison with a parallel study of rapeseed. Plant Physiology. 2008;**148**:504-518

[76] Ohyanagi H, Sakata K, Komatsu S. Soybean proteome database 2012: Update on the comprehensive data repository for soybean proteomics. Frontiers in Plant Science. 2012;**3**:110

[77] Komatsu S, Wang X, Yin X, Nanjo Y, Ohyanagi H, Sakata K. Integration of gel-based and gel-free proteomic data for functional analysis of proteins through soybean proteome database. Journal of Proteomics. 2017;**163**:52-66

[78] Grant D, Nelson RT, Cannon SB, Shoemaker RC. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research. 2010;**38**:D843-D8D6

[79] Brown AV, Conners SI, Huang W, Wilkey AP, Grant D, Weeks NT, et al. A new decade and new data at SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research. 2021;**49**:D1496-DD501

[80] Du J, Grant D, Tian Z, Nelson RT, Zhu L, Shoemaker RC, et al. SoyTEdb: A comprehensive database of transposable elements in the soybean genome. BMC Genomics. 2010;**11**:1-7

[81] Tavakolan M, Alkharouf NW, Khan FH, Natarajan S. SoyProDB: A database for the identification of soybean seed proteins. Bioinformation. 2013;**9**:165

[82] Tavakolan M, Alkharouf NW, Matthews BF, Natarajan SS. SoyProLow: A protein database enriched in low abundant soybean proteins. Bioinformation. 2014;**10**:599

[83] Joshi T, Fitzpatrick MR, Chen S, Liu Y, Zhang H, Endacott RZ, et al. Soybean knowledge base (SoyKB): A web resource for integration of soybean translational genomics and molecular breeding. Nucleic Acids Research. 2014;**42**:D1245-D1D52

[84] Mir RR, Reynolds M, Pinto F, Khan MA, Bhat MA. High-throughput phenotyping for crop improvement in the genomics era. Plant Science. 2019;**282**:60-72

[85] van Bezouw RF, Keurentjes JJ, Harbinson J, Aarts MG. Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. The Plant Journal. 2019;**97**:112-133

[86] Steiner B, Michel S, Maccaferri M, Lemmens M, Tuberosa R, Buerstmayr H. Exploring and exploiting the genetic variation of fusarium head blight resistance for genomic-assisted breeding in the elite durum wheat gene pool. Theoretical Applied Genetics. 2019;**132**:969-988

[87] Bhat JA, Ali S, Salgotra RK, Mir ZA, Dutta S, Jadon V, et al. Genomic selection in the era of next generation sequencing for complex traits in plant breeding. Frontiers in Genetics. 2016;**7**:221

[88] Gupta PK, Balyan HS, Gahlaut V. QTL analysis for drought tolerance in wheat: Present status and future possibilities. Agronomy. 2017;**7**:5

[89] Saghai Maroof M, Jeong SC, Gunduz I, Tucker D, Buss G, Tolin S. Pyramiding of soybean mosaic virus resistance genes by marker-assisted selection. Crop Science. 2008;**48**:517-526

[90] Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redona E, et al. Genomic selection and association mapping in rice (*Oryza sativa*): Effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics. 2015;**11**:e1004982

[91] Xu K, Zhang X-M, Chen H, Zhang C, Zhu J, Cheng Z, et al. Fine-tuning florigen increases field yield through improving photosynthesis in soybean. Frontiers. Plant Science. 2021;**12**

[92] Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, et al. Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder's equation. Theoretical Applied Genetics. 2019;**132**:627-645

[93] Khan A, Sovero V, Gemenet D. Genome-assisted breeding for drought resistance. Current Genomics. 2016;**17**:330-342

[94] Hina A, Cao Y, Song S, Li S, Sharmin RA, Elattar MA, et al. High-resolution mapping in two RIL populations refines major "QTL hotspot" regions for seed size and shape in soybean (*Glycine max* L.). International Journal of Molecular Sciences. 2020;**21**:1040

[95] Gao H, Wang Y, Li W, Gu Y, Lai Y, Bi Y, et al. Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. Journal of Experimental Botany. 2018;**69**:5089-5104

[96] Yin L, Zhang H, Zhou X, Yuan X, Zhao S, Li X, et al. KAML: Improving genomic prediction accuracy of complex traits using machine learning determined parameters. Genome Biology. 2020;**21**:1-22

[97] Stewart-Brown BB, Song Q, Vaughn JN, Li Z. Genomic selection for yield and seed composition traits within an applied soybean breeding program. G3: Genes, Genomes, Genetics. 2019;**9**:2253-2265

[98] Bohra A, Saxena K, Varshney RK, Saxena RK. Genomics-assisted breeding for pigeonpea improvement. Theoretical Applied Genetics. 2020;**133**:1721-1737

[99] Matei G, Woyann LG, Milioli AS, de Bem OI, Zdziarski AD, Zanella R, et al. Genomic selection in soybean: Accuracy and time gain in relation to phenotypic selection. Journal Molecular Breeding. 2018;**38**:1-13

[100] Qin J, Shi A, Song Q, Li S, Wang F, Cao Y, et al. Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. Frontiers in Plant Science. 2019:1445

[101] Ravelombola WS, Qin J, Shi A, Nice L, Bao Y, Lorenz A, et al. Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean *cyst* nematode infestation. PLoS One. 2020;**15**:e0235089

[102] Getachew T. Pulse crops production opportunities, challenges and its value chain in Ethiopia: A review article. Journal of Environment Earth Science. 2019;**9**:20-29

[103] Ravelombola WS, Qin J, Shi A, Nice L, Bao Y, Lorenz A, et al. Genome-wide association study and genomic selection for soybean chlorophyll content associated with soybean *cyst* nematode tolerance. BMC Genomics. 2019;**20**:1-18

[104] Shi A, Gepts P, Song Q, Xiong H, Michaels TE, Chen S. Genome-wide association study and genomic prediction for soybean cyst nematode resistance in USDA common bean (*Phaseolus vulgaris*) core collection. Frontiers in Plant Science. 2021;**12**

[105] Jarquín D, Howard R, Graef G, Lorenz A. Response surface analysis of genomic prediction accuracy values using quality control covariates in soybean. Evolutionary Bioinformatics. 2019;**15**:1176934319831307

[106] Ma Y, Reif JC, Jiang Y, Wen Z, Wang D, Liu Z, et al. Potential of marker selection to increase prediction accuracy of genomic selection in soybean

(*Glycine max* L.). Molecular Breeding. 2016;**36**:1-10

[107] Duhnen A, Gras A, Teyssèdre S, Romestant M, Claustres B, Daydé J, et al. Genomic selection for yield and seed protein content in soybean: A study of breeding program data and assessment of prediction accuracy. Crop Science. 2017;**57**:1325-1337

[108] Đorđević V, Ćeran M, Miladinović J, Balešević-Tubić S, Petrović K, Miladinov Z, et al. Exploring the performance of genomic prediction models for soybean yield using different validation approaches. Molecular Breeding. 2019;**39**:1-15

[109] Zhang J, Song Q, Cregan PB, Jiang GL. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). Theoretical Applied Genetics. 2016;**129**:117-130

[110] Jing Y, Zhao X, Wang J, Teng W, Qiu L, Han Y, et al. Identification of the genomic region underlying seed weight per plant in soybean (*Glycine max L. Merr*.) via high-throughput single-nucleotide polymorphisms and a genome-wide association study. Frontiers in Plant Science. 2018:1392

[111] Li D, Zhao X, Han Y, Li W, Xie F. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. Genomics. 2019;**111**:90-95

[112] Alqudah AM, Sallam A, Baenziger PS, Börner A. GWAS: Fast-forwarding gene identification and characterization in temperate cereals: Lessons from barley. A review. Journal of Advanced Research. 2020;**22**:119-135

[113] Hu D, Kan G, Hu W, Li Y, Hao D, Li X, et al. Identification of loci and candidate genes responsible for pod dehiscence in soybean via genome-wide association analysis across multiple environments. Frontiers in Plant Science. 2019:811

[114] Torkamaneh D, Chalifour F-P, Beauchamp CJ, Agrama H, Boahen S, Maaroufi H, et al. Genome-wide association analyses reveal the genetic basis of biomass accumulation under symbiotic nitrogen fixation in African soybean. Theoretical Applied Genetics. 2020;**133**:665-676

[115] Borah J, Singode A, Talukdar A, Yadav RR, Sarma RN. Genome-wide association studies (GWAS) reveal candidate genes for plant height and number of primary branches in soybean [*Glycine max (L.) Merr.*]. Indian Journal of Genetics Plant Breeding. 2018;**78**:460-469

[116] Sonah H, O'Donoughue L, Cober E, Rajcan I, Belzile F. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soybean. Plant Biotechnology Journal. 2015;**13**:211-221

[117] Zhao X, Han Y, Li Y, Liu D, Sun M, Zhao Y, et al. Loci and candidate gene identification for resistance to Sclerotinia sclerotiorum in soybean (*Glycine max* L. Merr.) via association and linkage maps. The Plant Journal. 2015;**82**:245-255

[118] Sui M, Jing Y, Li H, Zhan Y, Luo J, Teng W, et al. Identification of loci and candidate genes analyses for tocopherol concentration of soybean seed. Frontiers in Plant Science. 2020:1366

[119] Yang Q, Jin H, Yu X, Fu X, Zhi H, Yuan F. Rapid identification of soybean resistance genes to soybean mosaic virus by SLAF-seq bulked segregant analysis. Plant Molecular Biology Reporter. 2020;**38**:666-675

[120] da Silva MP, Zaccaron AZ, Bluhm BH, Rupe JC, Wood L, Mozzoni LA, et al. Bulked segregant analysis using next-generation sequencing for identification of genetic loci for charcoal rot resistance in soybean. Physiological Molecular Plant Pathology. 2020;**109**:101440

[121] Watanabe S, Tsukamoto C, Oshita T, Yamada T, Anai T, Kaga A. Identification of quantitative trait loci for flowering time by a combination of restriction site–associated DNA sequencing and bulked segregant analysis in soybean. Breeding Science. 2017:17013

[122] Cheng Y, Ma Q, Ren H, Xia Q, Song E, Tan Z, et al. Fine mapping of a Phytophthora-resistance gene RpsWY in soybean (*Glycine max* L.) by high-throughput genome-wide sequencing. Theoretical Applied Genetics. 2017;**130**:1041-1051

[123] Jiang B, Li M, Cheng Y, Cai Z, Ma Q, Jiang Z, et al. Genetic mapping of powdery mildew resistance genes in soybean by high-throughput genome-wide sequencing. Theoretical Applied Genetics. 2019;**132**:1833-1845

[124] Gao C. The future of CRISPR technologies in agriculture. Nature Reviews Molecular Cell Biology. 2018;**19**:275-276

[125] Gao C. Genome engineering for crop improvement and future agriculture. Cell. 2021;**184**:1621-1635

[126] Bao A, Zhang C, Huang Y, Chen H, Zhou X, Cao D. Genome editing technology and application in soybean improvement. Oil Crop Science. 2020;**5**:31-40

[127] Li Z, Liu Z-B, Xing A, Moon BP, Koellhoffer JP, Huang L, et al. Cas9-guide RNA directed genome editing in soybean. Plant Physiology. 2015;**169**:960-970

[128] Du H, Zeng X, Zhao M, Cui X, Wang Q, Yang H, et al. Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. Journal of Biotechnology. 2016;**217**:90-97

[129] Bai M, Yuan J, Kuang H, Gong P, Li S, Zhang Z, et al. Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soyabean. Plant Biotechnology Journal. 2020;**18**:721-731

[130] Mishra R, Joshi RK, Zhao K. Base editing in crops: Current advances, limitations and future implications. Plant Biotechnology Journal. 2020;**18**:20-31

[131] Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;**533**:420-424

[132] Cai Y, Chen L, Zhang Y, Yuan S, Su Q, Sun S, et al. Target base editing in soybean using a modified CRISPR/Cas9 system. Plant Biotechnology Journal. 2020;**18**:1996

[133] Cai Y, Chen L, Liu X, Guo C, Sun S, Wu C, et al. CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soybean. Plant Biotechnology Journal. 2018;**16**:176-185

[134] Li C, Nguyen V, Liu J, Fu W, Chen C, Yu K, et al. Mutagenesis of seed storage protein genes in soybean using CRISPR/Cas9. BMC Research Notes. 2019;**12**:1-7

[135] Cai Y, Chen L, Sun S, Wu C, Yao W, Jiang B, et al. CRISPR/Cas9-mediated deletion of large genomic fragments in soybean. International Journal of Molecular Sciences. 2018;**19**:3835

[136] Cai Y, Wang L, Chen L, Wu T, Liu L, Sun S, et al. Mutagenesis of *GmFT2a* and *GmFT5a* mediated by CRISPR/Cas9 contributes for expanding the

regional adaptability of soybean. Plant Biotechnology Journal. 2020;**18**:298-309

[137] Wang L, Sun S, Wu T, Liu L, Sun X, Cai Y, et al. Natural variation and CRISPR/Cas9-mediated mutation in *GmPRR37* affect photoperiodic flowering and contribute to regional adaptation of soybean. Plant Biotechnology Journal. 2020;**18**:1869-1881

[138] Fang X, Sun X, Yang X, Li Q, Lin C, Xu J, et al. MS1 is essential for male fertility by regulating the microsporocyte cell plate expansion in soybean. Science China Life Sciences. 2021;**64**:1533-1545

[139] Jiang B, Chen L, Yang C, Wu T, Yuan S, Wu C, et al. The cloning and CRISPR/Cas9-mediated mutagenesis of a male sterility gene MS1 of soybean. Plant Biotechnology Journal. 2021;**19**:1098

[140] Nadeem M, Chen A, Hong H, Li D, Li J, Zhao D, et al. GmMs1 encodes a kinesin-like protein essential for male fertility in soybean (*Glycine max L.*). Journal of Integrative Plant Biology. 2021;**63**:1054-1064

[141] Li S, Cong Y, Liu Y, Wang T, Shuai Q, Chen N, et al. Optimization of Agrobacterium-mediated transformation in soybean. Frontiers in Plant Science. 2017;**8**:246

[142] Homrich MS, Wiebke-Strohm B, Weber RLM, Bodanese-Zanettini MH. Soybean genetic transformation: A valuable tool for the functional study of genes and the production of agronomically improved plants. Genetics Molecular Biology. 2012;**35**:998-1010

[143] Curtin SJ, Michno JM, Campbell BW, Gil-Humanes J, Mathioni SM, Hammond R, et al. MicroRNA maturation and microRNA target gene expression regulation are severely disrupted in soybean dicer-like1 double mutants. G3: Genes, Genomes, Genetics. 2016;**6**:423-433

[144] Curtin SJ, Zhang F, Sander JD, Haun WJ, Starker C, Baltes NJ, et al. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiology. 2011;**156**:466-473

[145] Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;**343**:84-87

[146] Kelliher T, Starr D, Su X, Tang G, Chen Z, Carter J, et al. One-step genome editing of elite crop germplasm during haploid induction. Nature Biotechnology. 2019;**37**:287-292

[147] Rahman SU, McCoy E, Raza G, Ali Z, Mansoor S, Amin I. Improvement of soybean; a way forward transition from genetic engineering to new plant breeding technologies. Molecular Biotechnology. 2022:1-19

[148] Cai Y, Chen L, Liu X, Sun S, Wu C, Jiang B, et al. CRISPR/Cas9-mediated genome editing in soybean hairy roots. PLoS One. 2015;**10**:e0136064

[149] Jacobs TB, LaFayette PR, Schmitz RJ, Parrott WA. Targeted genome modifications in soybean with CRISPR/Cas9. BMC Biotechnology. 2015;**15**:1-10

[150] Di YH, Sun XJ, Hu Z, Jiang QY, Song GH, Zhang B, et al. Enhancing the CRISPR/Cas9 system based on multiple GmU6 promoters in soybean. Biochemical Biophysical Research Communications. 2019;**519**:819-823

[151] Tang F, Yang S, Liu J, Zhu H. Rj4, a gene controlling nodulation specificity in soybeans, encodes a thaumatin-like protein but not the one previously

reported. Plant Physiology. 2016;**170**:26-32

[152] Bonawitz ND, Ainley WM, Itaya A, Chennareddy SR, Cicak T, Effinger K, et al. Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. Plant Biotechnology Journal. 2019;**17**:750-761

[153] Do PT, Nguyen CX, Bui HT, Tran LT, Stacey G, Gillman JD, et al. Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous *GmFAD2-1A* and *GmFAD2-1B* genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean. BMC Plant Biology. 2019;**19**:1-14

[154] Khan MHU, Khan SU, Muhammad A, Hu L, Yang Y, Fan C. Induced mutation and epigenetics modification in plants for crop improvement by targeting CRISPR/Cas9 technology. Journal of Cellular Physiology. 2018;**233**:4578-4594

[155] Kanazashi Y, Hirose A, Takahashi I, Mikami M, Endo M, Hirose S, et al. Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. Plant Cell Reports. 2018;**37**:553-563

[156] Chen X, Yang S, Zhang Y, Zhu X, Yang X, Zhang C, et al. Generation of male sterile soybean lines with the CRISPR/Cas9 system. The Crop Journal. 2021;**9**:1270-1277

[157] Bao A, Chen H, Chen L, Chen S, Hao Q, Guo W, et al. CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. BMC Plant Biology. 2019;**19**:1-12

[158] Al Amin N, Ahmad N, Wu N, Pu X, Ma T, Du Y, et al. CRISPR-Cas9 mediated targeted disruption of FAD2-2 microsomal omega-6 desaturase in soybean (*Glycine max.* L). BMC Biotechnology. 2019;**19**:1-10

[159] Wang J, Kuang H, Zhang Z, Yang Y, Yan L, Zhang M, et al. Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. The Crop Journal. 2020;**8**:432-439

[160] Zhang P, Du H, Wang J, Pu Y, Yang C, Yan R, et al. Multiplex CRISPR/Cas9-mediated metabolic engineering increases soybean isoflavone content and resistance to soybean mosaic virus. Plant Biotechnology Journal. 2020;**18**:1384-1395

[161] Curtin SJ, Xiong Y, Michno JM, Campbell BW, Stec AO, Čermák T, et al. CRISPR/Cas9 and talen s generate heritable mutations for genes involved in small rna processing of glycine max and medicago truncatula. Plant Biotechnology Journal. 2018;**16**:1125-1137

[162] Jacobs TB, Zhang N, Patel D, Martin GB. Generation of a collection of mutant tomato lines using pooled CRISPR libraries. Plant Physiology. 2017;**174**:2023-2037

[163] Lu Y, Ye X, Guo R, Huang J, Wang W, Tang J, et al. Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Molecular Plant. 2017;**10**:1242-1245

[164] Meng X, Yu H, Zhang Y, Zhuang F, Song X, Gao S, et al. Construction of a genome-wide mutant library in rice using CRISPR/Cas9. Molecular Plant. 2017;**10**:1238-1241

[165] Kim H, Kim S-T, Ryu J, Kang BC, Kim J-S, Kim S-G. CRISPR/Cpf1-mediated DNA-free plant genome editing. Nature Communications. 2017;**8**:1-7

[166] Gosal SS, Wani SH, Kang MS. Biotechnology and drought tolerance. Journal of Crop Improvement. 2009;**23**:19-54

[167] Détain A, Bhowmik P, Leborgne-Castel N, Ochatt S. Latest biotechnology tools and targets for improving abiotic stress tolerance in protein legumes. Environmental. Experimental Botany. 2022:104824

[168] Croser J, Lülsdorf M, Davies P, Clarke H, Bayliss K, Mallikarjuna N, et al. Toward doubled haploid production in the Fabaceae: Progress, constraints, and opportunities. Critical Reviews in Plant Sciences. 2006;**25**:139-157

[169] Hale B, Ferrie AM, Chellamma S, Samuel JP, Phillips GC. Androgenesis-based doubled Haploidy: Past, present, and future perspectives. Frontiers in Plant Science. 2022:2944

[170] Valliyodan B, Ye H, Song L, Murphy M, Shannon JG, Nguyen HT. Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. Journal of Experimental Botany. 2017;**68**:1835-1849

[171] Ratnaparkhe MB, Satpute GK, Kumawat G, Chandra S, Kamble VG, Kavishwar R, et al. Genomic designing for abiotic stress tolerant soybean. In: Genomic Designing for Abiotic Stress Resistant Oilseed Crops. Springer; 2022. pp. 1-73

[172] Patil G, Do T, Vuong TD, Valliyodan B, Lee J-D, 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**:1-13

[173] Moreira FF, Oliveira HR, Volenec JJ, Rainey KM, Brito LF. Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. Frontiers in Plant Science. 2020;**11**:681