CrossMark

REVIEW

# Molecular approaches for genetic improvement of seed quality and characterization of genetic diversity in soybean: a critical review

**Niraj Tripathi** · **Dhirendra Khare**

**Abstract** Soybean is an economically important leguminous crop. Genetic improvements of soybeans have focused on enhancement of seed and oil yield, development of varieties suited to different cropping systems, and breeding resistant/tolerant varieties for various biotic and abiotic stresses. Plant breeders have used conventional breeding techniques for the improvement of these traits in soybean. The conventional breeding process can be greatly accelerated through the application of molecular and genomic approaches. Molecular markers have proved to be a new tool in soybean breeding by enhancing selection efficiency in a rapid and time-bound manner. An overview of molecular approaches for the genetic improvement of soybean seed quality parameters, considering recent applications of marker-assisted selection and 'omics' research, is provided in this article.

**Keywords** Breeding · Crop improvement · Genetic improvement · Molecular markers · Omics · Soybeans

N. Tripathi (✉)
Biotechnology Centre, Jawaharlal Nehru Agricultural University, Jabalpur 482004, India
e-mail: tripathi.niraj@gmail.com

D. Khare
Department of Plant Breeding and Genetics, Jawaharlal Nehru Agricultural University, Jabalpur 482004, India

## Introduction

Soybean (*Glycine max*) is an important leguminous crop and a rich source of protein (40 %); it also provides more than 50 % of the edible oil consumed worldwide (Soy Stats; www.soyohio.org/wp-content/uploads/2013/07/ASA_SoyStats_fnl.pdf). It enhances soil fertility with efficient symbiotic $N_2$ fixing ability. Further, the crop holds promise due to the possible production of soybean protein-based biodegradable materials (Song et al. 2011) and anti-cancerous formulations (Ko et al. 2013). Understanding the process of soybean genetic improvement is vital for future gains in breeding improved soybean varieties (Zhou et al. 2015).

Advancements in technologies have enhanced soybean production at a global level. Soybean belongs to the self-pollinated category with a natural out-crossing of 0.5–1 %. Due to their autogamous nature, plant-breeding practices, such as pedigree, backcross, single pod descent and population breeding, have been applied for the genetic improvement of quality attributes and development of high-yielding varieties of this important oil-seed crop. In addition to conventional breeding, application of molecular tools can enhance the precision of soybean breeding. In particular, application of molecular tools will be helpful in the improvement of soybean seed traits, which are considered complex due to multi-gene control, and are highly

influenced by the environment and hence difficult to manipulate.

Among the molecular tools, the application of molecular markers has gained predominance in the improvement of many crop species. Molecular markers are stretches of DNA sequence with a precisely defined nucleotide arrangement and distribution for different organisms. They are of immense use in the estimation of the diversity and genetic organization of the available genotype and in the prediction of the heterosis, based on polymorphism/genetic distance present between the parents, in hybrid development programmes. The use of PCR-based markers has simplified the process of marker-based gene tagging, cloning of agronomically superior genes, polymorphism studies, and phylogenetic analysis, culminating in marker-assisted selection (MAS) of desired genes or feasible genotypes. DNA markers provide a number of benefits over conventional morphological markers, as the information provided by these markers can be used to analyze the target. The different types of markers used for tagging and mapping agronomically important genes and their marker-assisted selection are detailed below.

## Markers

These are of three types: morphological, biochemical and dna.

### Morphological markers

Morphological markers, such as leaf size and shape, pubescence color, flower color, pod color, hilum color, seed shape, awn type, seed coat color, fruit shape and stem length, etc. have been used traditionally to verify the genetics, association, varietal verification, seed production, maintenance and certification of genetic purity of a variety. Morphological markers are limited in number; their expression is often influenced by environment fluctuation, and many of them are not closely linked with economic traits and even have adverse effects on the development and growth of plants. However, morphological markers have been used for diversity analysis in various plant species (Lira-Medeiros et al. 2015).

### Biochemical markers

Biochemical markers are protein markers and may be separated into two groups: storage proteins and functional proteins or isozymes. Most commonly used protein markers are the isozymes. These markers are co-dominant in nature. However, due to their limited numbers, their use is restricted. In soybean, biochemical markers are mostly used for cultivar identification, hybrid seed testing, divergence analysis, as well as determining uniformity in seed production and genetic purity of cultivars. Evaluation of the banding pattern of storage proteins of soybean cultivars using SDS-PAGE can characterize soybean cultivars compared to isozyme patterns (Liu et al. 2007). Seed storage protein polymorphism has been used successfully for detecting variability among soybean genotypes, which show great geographical diversity (Iqbal et al. 2015).

### DNA markers

DNA markers are patterns of arrangement of nucleotides and the polymorphism among them, which could be used to identify the differences between two individuals or crop varieties, and they follow a simple Mendelian pattern of inheritance. A high level of polymorphism, co-dominant inheritance, frequent occurrence in the genome, selective neutral behavior, easy and fast assay, high reproducibility and easy exchange of data between laboratories are characteristics of an ideal DNA marker. DNA markers are of the following types:

Restriction fragment length polymorphism (RFLP markers): Among all molecular markers, RFLP are the first generation used for plant genome analysis. RFLP depends upon restriction enzyme (endonuclease) digestion of template DNA. These enzymes recognize short DNA fragments (3–6 base pairs) and cut the DNA at sequence-specific sites. Genomic DNA digestion with restriction enzymes gives polymorphism through DNA fragment length. After digestion, fragments of different lengths between genotypes can be identified by Southern blotting using a suitable probe. In soybean, RFLP markers were applied in the 1980s and contributed to the construction of the first genetic map of the soybean genome. RFLP markers for the evaluation of genetic diversity in other legume species have been reported (Islam et al. 2015).

Randomly amplified polymorphic DNA (RAPD markers): The RAPD markers are the first DNA markers based on PCR technology. The RAPD markers have been applied extensively for genetic diversity studies in soybean germplasm (Mladenovic et al. 2008). Gwata and Wofford (2013) identified the potential of RAPD analysis of the promiscuous nodulation trait in soybean (*Glycine max* L) and Khare et al. (2013) used RAPD markers for genetic diversity analysis among Indian soybean genotypes.

Simple sequence repeats (SSR markers): SSRs are usually 2–6 bp long single-locus markers with a high level of allelic variation and acceptability (Sahu et al. 2012). The most common repeats in soybean are: AT, ATT, TA, TAT, CT, CTT. In plant genome analyses, the first applications of SSRs were in soybean. SSR markers have been subjected to continuous development and are used for high throughput molecular mapping in soybean. Hu et al. (2014) performed association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb.) and Bisen et al. (2015) used SSR markers for diversity analysis of Indian soybean germplasm.

Amplified fragment length polymorphism (AFLP markers): AFLP markers are based on the use of restriction enzymes (endonucleases), and pre-selective amplifications with PCR. In soybean, fewer reports are available on the development of AFLP markers than in other plant species, mostly because of the successful application of SSRs. AFLP analysis, however, was used to study the genetic relationship among 25 soybean varieties from Japan and Thailand (Nimnual et al. 2014).

Single nucleotide polymorphism (SNP markers): differences in individual DNA bases with point mutations are referred to as SNP. SNP are potentially useful as genetic markers because they enable the distinction of one haplotype from another (Hyten et al. 2010). In a study, 20 soybean chromosomes and corresponding linkage groups of the 1536 SNP markers comprising the Universal Soy Linkage Panel 1.0 and distribution of the 1006 SSR markers were mapped and Consensus Map 4.0 scaled in Kosambi centimorgans by Hyten et al. (2010). The Illumina Infinium array (SoySNP50 KiSelectBeadChip) for ∼50,000 SNPs has been successfully developed and used for the genotyping of several soybean plant introduction lines (Song et al. 2013). Technological advances like this makes it possible to re-sequence hundreds of lines in a cost effective manner and these developments have started a new era of genotyping by re-sequencing. A total of 55,159 SNPs were genotyped using various methods including Illumina Infinium and Golden Gate assays and 31,954 markers with minor allele frequency >0.10 were used to estimate linkage disequilibrium in heterochromatic and euchromatic regions related to seed protein and oil content in soybean. Twenty five SNPs have been detected in 13 different genomic regions associated with seed oil. Of these markers, seven SNPs were found to be associated with both protein and oil (Hwang et al. 2014). The molecular analysis of the cultivated and wild accessions in the USDA soybean germplasm collection includes 18,480 domesticated and 1168 wild soybean accessions introduced from 84 countries or developed in the United States; they were genotyped with the SoySNP50 K BeadChip containing more than 50K SNPs. Redundant accessions were identified in the collection, and distinct genetic backgrounds of soybean from different geographic origins were observed, which could be a unique resource for soybean genetic improvement (Song et al. 2015). High-throughput and functional SNP detection assays for oleic and linolenic acids in soybean have been performed recently (Shi et al. 2015).

## Marker-assisted selection (MAS)

Integrating molecular marker technologies, such as MAS, into crop improvement programmes could become increasingly important to achieve increased genetic gains with greater speed and precision. The promise of MAS for improving polygenic traits in a quick time frame and in a cost-effective manner remains to be fulfilled. There is a wider appreciation that simply demonstrating that a complex trait can be dissected into quantitative trait loci (QTL) and mapped to approximate genomic locations using DNA markers would not serve the ultimate goal of trait improvement. In facing the challenge of improving several lines for quantitative traits, MAS strategies use DNA markers in one key selection step to maximize their impact.

Marker-assisted breeding has become sophisticated with the availability of complete soybean genome sequences and subsequent development of locus-specific molecular markers. Recent development in

genomics provides the power to predict genetic factors, their evolution, distribution, and interactions. The platform of soybean core, mini core and integrated applied core collections developed by systematic researches has provided powerful materials for the evaluation of germplasm, identification of trait-specific accessions, gene discovery, allele mining, genomic study, maker development and molecular breeding. It is helpful for enhancing the use of soybean genetic resources in sustainable crop improvement (Qiu et al. 2013). Availability of genome-wide high density markers facilitates the haplotype analysis and identification of alleles of agronomically important traits (Tardivel et al. 2014). The application areas of MAS for improvement of seed quality traits in soybean are described below.

## Seed quality parameters

### Seed longevity

Seed longevity is a major inherited problem of soybean in subtropical hot and humid regions that adversely affects sowing seed quality (Khare et al. 1996). At physiological maturity, the seed reaches its maximum potential for germination and vigor. Longevity of soybean seeds has been affected by a number of traits shown to be under genetic control, including seed size, hardness, coat thickness, permeability, hull percentage and oil content. These traits have been used in breeding programmes to improve soybean seed quality and longevity. The wide occurrence of unfavorable weather conditions during soybean harvest result in poor quality seed and rapid deterioration during storage (Singh et al. 2008).

The maternal plant genome can influence the longevity of soybean seed. Four segregating SSR markers (Satt538, Satt600, Satt434 and Satt285) are significantly associated with seed longevity in soybean and located on linkage groups A2 (158.63 cM), D1b (75.4 cM), H (105 cM) and J (25.57 cM), respectively. Individually, these markers explained between 6.3 % (Satt285) and 7.5 % (Sat_434) of the total phenotypic variation for the trait (Singh et al. 2008, Table 1). Specific bands produced by SSR markers Satt371, Satt453 and Satt618 make them candidate markers for linkage with seed storability and testa colour (Hosamani et al. 2013). The last decade has also

seen research directed towards the inference of the genetic basis of seed deterioration through investigation of quantitative trait loci associated with seed longevity (Han et al. 2014). Seed ageing is a complex biological trait and difficult to monitor. These efforts have helped to identify favorable longevity alleles for better prediction of seed longevity in plant germplasm collections (Fu et al. 2015).

### Seed coat permeability

Seed coat permeability and electrolyte leaching are important traits that have been negatively associated with seed longevity in soybean. Association of four independent SSR markers, Satt434, Satt538, Satt281, and Satt598, with seed coat permeability and electrolyte leaching with an explanation for 3.9 % (Satt434)–4.5 % (Satt538) of the total phenotypic variation for seed coat permeability has been reported (Singh et al. 2008).

### Hard seededness

This is a quantitative trait in soybean that affects the germination rate, viability and quality of stored seeds. Several major QTLs influence hard seededness in soybean. SSR markers from the molecular linkage group A2, D1b, H and J of soybean are associated with traits, particularly hard seededness, seed oil concentration, seed protein and seed size, which have a positive association with seed longevity (Singh et al. 2008).

### Seed coat cracking

Elliptical cracks in the soybean seed coat that separates epidermal and hypodermal tissues and exposes the underlying parenchyma tissues is termed seed coat cracking. It adversely affects the external appearance and reduces the commercial value, providing a path for the entrance of pathogenic and adverse environmental factors to affect seed quality. Prevalence of low temperatures during flowering adversely affects appearance and quality of soybean seeds at maturity.

Two types of seed coat cracking have been observed in soybean: Type 1 (irregular cracks) and Type 2 (net-like cracks). Nakamura et al. (2003) suggested that net-like cracking is controlled primarily by a major gene SoyPRPI, and genes contributing to

**Table 1** SSR markers reported to be linked with seed characteristics in soybean

| Sr. no. | Characteristic | Locus | LG | Linked markers | Reference |
|---|---|---|---|---|---|
| 1. | Seed longevity | | J | Satt285, | Singh et al. (2008) |
| | | | H | Satt434 | |
| 2. | Seed storability | | C2 | Satt371, | Hosamani et al. (2013) |
| | | | B1 | Satt453, | |
| | | | M | Satt618 | |
| 3. | Seed coat permeability | | H | Satt434, | Singh et al. (2008) |
| | | | A2 | Satt538, | |
| | | | C2 | Satt281, | |
| | | | E | Satt598 | |
| 4. | Seed coat cracking | SoyPRPI | K | Satt264 | Nakamura et al. (2003) |
| 5. | Tocopherol | | MLG | Sat_167, | Dwiyanti et al. (2007) |
| | | | K | Sat_243 | |
| | | | C2 | Satt376, | Li et al. (2010) |
| | | | C2 | Satt286 | |
| | | | D1b | Satt266 | |
| | | γ-TMT3 | K | Sat_243 | Dwiyanti et al. (2011) |

*LG* Linkage group

seed weight may have minor effects on the intensity of cracking. SSR locus SoyPRPI associated with seed coat cracking was found approx. 1 kb upstream to the ORF of the PRPI gene. The SoyPRPI locus was subsequently assigned to MLG K. SSR marker Satt264 is closely related to the SoyPRPI locus for proline-rich cell wall protein and has a minor effect on the average cracking index (Nakamura et al. 2003).

### Tocopherol content

Four tocopherols (α-, β-, γ-, δ-) occur in nature and can be distinguished by the number and position of methyl groups on the aromatic hormonal head group. The primary function of tocopherols in plants is to limit non-enzymatic lipid oxidation during seed storage, germination and early seedling development (Sattler et al. 2003). A defined role of tocopherol in seed longevity has been identified, showing a correlation of tocopherol content or degradation with longevity. Genetic analysis in soybean revealed that α-tocopherol content is a highly heritable trait (Dwiyanti et al. 2007) and can be simultaneously increased in soybean seed (Wang et al. 2007). Genotypes COSOYA2 and Ankur exhibited comparatively higher values for tocopherol content and seem to be good donors (Rani et al. 2007). Hybridization between varieties with a high α-tocopherol content (20–30 %)

and low tocopherol content (<10 %) showed that SSR markers Sat_167 and Sat_243 were located on MLG K, flanked by an area with genes controlling a high α-tocopherol concentration (Table 1). This suggested that α-tocopherol and total tocopherol may be regulated independently (Dwiyanti et al. 2007).

Association analysis is an alternative to conventional methods to detect the location of genes or QTL and provides relatively high resolution in terms of defining the genome position of a gene or QTL. A genome-wide association study was performed to identify QTL controlling seed protein and oil concentration in 298 soybean germplasm accessions exhibiting a wide range of seed protein and oil content (Hwang et al. 2014).

### Pod shattering

The extent of yield loss due to pod shattering in soybean may range from 34 to 100 % depending upon the extent to which harvesting may be delayed after maturity. Morphological architecture of plant, anatomical structures of the pod, chemical composition of the pod wall, and genetic constitution of the variety and environmental conditions at maturity determine the degree of pod shattering (Gulluoglu et al. 2006).

Pod shattering in soybean is a highly heritable trait with narrow sense heritability. Segregation of pod shattering is highly complex, with a quantitative response. The major QTL for pod shattering, designated as qPDH1, is located between SSR markers, Sat_093 and Sat_366. Moreover, the shattering resistance allele at qPDH1 proved useful in various genetic backgrounds at multiple locations (Funatsuki et al. 2006). Analysis of the relationship between the degree of pod dehiscence and graphical genotype of soybean lines confined the location of qPDH1 to a 134 kb region on chromosome 16 (formerly linkage group J), where ten putative genes were predicted to be present (Suzuki et al. 2010).

## Use of markers in varietal characterization and verification

In the era of intellectual property rights , it is important to characterize the various traits of crop germplasm available in the country. The Government of India has enacted its *sui generis* system designated as "Protection of Plant Varieties and Farmers' Right Act 2001" under a general agreement on trade and tariff, which undertakes registration of extant and new plant varieties through the plant variety registry on the basis of varietal characteristics. To protect the plant varieties, the genetic material has to be registered based on the Distinctness, Uniformity and Stability test, as well as novelty, which requires descriptors derived from morphological traits, biochemical reactions, electrophoresis and DNA fingerprinting of various crop varieties. Cultivar verification and purity assessment are important for maintenance, multiplication and seed certification of varieties, which is of critical importance for a sustained increase in agricultural productivity. Under the New Seed Policy Act, 2001, all new varieties have to be registered based on the criteria of novelty, distinctness, uniformity and stability (N-DUS) as per the PPV and FR authority and national test guidelines.

Use of molecular markers may provide a solution to the problem of protection of new soybean varieties, offering many considerable potential benefits for DUS testing within the variety registration system. A combination of SSR and morphological descriptors has been recognized as the best option to study genetic relationships and for a clear classification of soybean

varieties for their protection and to establish minimum genetic distances for distinctness. The International Union for the protection of new varieties of plants (UPOV) has recognized the utility of molecular markers associated with descriptive phenotypic characteristics. UPOV (2011) has constituted a working group on biochemical and molecular techniques, and DNA profiling in particular (BMT), to study the utility of molecular markers in the variety registration system. The BMT in its thirteenth session (BMT/13/13) provided "International guidelines on molecular methodologies" for developing harmonized methodologies with the aim of generating high quality molecular data for a range of applications. The 44th session of the technical committee of the ISTA confirmed that ISTA would consider molecular markers for variety verification (ISTA 2012).

## Association of molecular markers with morphological traits

Expression of hilum color in soybean seed is genetically controlled but is also influenced by the environment. Therefore variation/deviation from normal expression is not necessarily related to genetic variation. Rabel et al. (2010) reported Satt070 present at B2 as a potent marker to distinguish soybean varieties based on hilum color. The working group of UPOV evaluated soybean seeds for variation in hilum color using variety CD 222 (black hilum) and CD 02RV-8444 and CD01RV-7618 (brown hilum) involving 16 microsatellite molecular markers. Marker Satt020 at position B2 was found to be suitable to differentiate between purple and white flowered plants in population segregation (Jian et al. 2012). Markers Satt331.1.2, Satt 371.5, Satt367.3, and Satt 212.1 were recognized competent enough to distinguish soybean plants based on pubescence density (Daymann et al. 2009), markers Sat_268 and Sat_105 for narrow leaflets (Jeong et al. 2011), Satt_105 for the presence and absence of four-seeded pods, and marker *GmF35H* for flower color (white and purple) in soybean (Jeong et al. 2011). Discrimination tests showed a high percentage of accurate classification of growth habit (95.8 %) and pubescence color (80.6 %) traits with the SSR Sat286 and *SoyF3H,* and for pod color (74.2 %) and leaflet size (73.5 %) with the SSR

GMES1173 and Satt571 (GmPin1), respectively (Mariela et al. 2011) (Table 2).

In the present scenario, advanced omics approaches are important in the field of crop improvement, and many of them have been applied for the improvement of soybean.

## Advanced omics approaches in soybean

The biochemical and molecular mechanisms that control the compound metabolic system of soybean seed development decide the definitive equilibrium of protein, lipid and carbohydrate stored in the developed seed. Many of the genes and metabolites that participate in seed metabolism are unknown or poorly defined, and much more remains to be understood about the regulation of metabolic networks. A global omics analysis can provide insights into the regulation of seed metabolism, even without a priori assumptions about the structure of these networks (Li et al. 2015). Some of the advanced approaches (Table 3) that have potential applications for soybean improvement are discussed below.

### Transcriptomics

A cost effective and high-throughput RNA-seq makes analysis of transcriptomes possible in crops such as soybean where genome sequence information is available. It has several advantages over the microarray technology, where available genomic information is used to design probe sets, thus limiting the number of genes whose expression can be analyzed. RNA-seq does not require gene information and is capable of identifying novel transcripts that were previously unknown and also provides opportunities to analyze non-coding and differentially spliced RNA. Rhodanese and RPS4 probably play important roles in regulating the photoperiod to control flowering time in soybean. Transcript-derived fragments isolated from leaves of soybean growing under short-day length, long-day length and a neutral photoperiod were identified as regulated by the photoperiod (Ai-Hua et al. 2014).

### Proteomics

Proteomics deals with structural and functional features of all the proteins in an organism. A non-gel-based protein identification and quantification isotope labeling-based technique using isobaric tags for relative and absolute quantitation (iTRAQ) is one of the major quantification tools in differential proteomic research. Qin et al. (2013) performed iTRAQ-based proteomic analysis of an elite soybean variety (Jidou17) and its parents to evaluate the parental contributions to its elite traits.

**Table 2** Association of SSR markers with expression of morphological traits in soybean

| Sr. | Characteristics | Linked markers | LG | QTL | Gene associated | Reference |
|---|---|---|---|---|---|---|
| 1. | Flower color | Satt020 | B2 | *W1* | *GmF3'5'H Glyma13g04210* | Jian et al. (2012) |
|  |  | GmF35H | F |  |  | Mariela et al. (2011) |
| 2. | Pubescence color | SoyF3H | C2 | *T* | *SoyF3'H Glyma06g21920* | Mariela et al. (2011) |
| 3. | Pubescence density | Satt331.1.2, Satt 371.5, Satt367.3, Satt 212.1 |  |  |  | Daymann et al. (2009) |
| 4. | Leaflet shape | Sat_268, Satt_105 | I |  |  | Song et al. (2004) |
|  |  | Satt369 |  |  |  | Jeong et al. (2011) |
| 5. | Leaflet size | Satt571 (GmPin1) | I | *Ln* | Glyma20g01760 | Mariela et al. (2011) |
| 6. | Presence of four seeded pods | Sat_268, Satt_105 | I |  |  | Jeong et al. (2011) |
| 7. | Hilum color | Satt070 | B2 |  |  | Rabel et al. (2010) |
| 8. | Growth habit | Sat286 | L | *Dt1* | *TFL1 Glyma19g37890* | Mariela et al. (2011) |
| 9. | Pod color | GMES1173 | N | *L2* | Glyma03g00410 | Mariela et al. (2011) |
| 10. | Flowering and maturity | Satt229 | L | *E3/FT3* | *GmPhyA3 Glyma19g41210* | Mariela et al. (2011) |

LG-Linkage Group

**Table 3** Advanced approaches using different technological platforms in soybean

| Sr. no. | Approach | Platform | Achievement | Reference |
|---|---|---|---|---|
| 1. | Transcriptomics | Expression profiling | Combined approach with proteomics | Komatsu et al. (2009) |
| | | Custom array containing cDNAs | Identification of genes involved in DNA repair and RNA stability | O'Rourke et al. (2007) |
| | | cDNA microarray | Expression of genes related to cell recovery | Zhu et al. (2009) |
| | | Affymetrix soybean Array Gene Chip | Alteration in expression of many *GmNAC* and hormone-related genes | Le et al. (2012) |
| | | Affymetrix Soybean Gene Chip | Genes with altered expression | Ge et al. (2010) |
| 2. | Proteomics | iTRAQ | Understanding of the mechanism of leaf development | Qin et al. (2013) |
| | | 2-D-DIGE, mass spectrometry | Role of Organic acid biosynthesis, anti-oxidation and de-toxification enzymes | Duressa et al. (2011) |
| | | MALDI-TOF | Identification of novel proteins viz., a translation initiation factor, apyrase, auxin-amidohydrolase and coproporphyrinogen oxidase | Alam et al. (2010) |
| 3. | Ionomics | Ion coupled plasma mass spectrometry | Identification of elemental composition of potential mutants | Ziegler et al. (2013) |

## Metabolomics

Metabolomics provide a better way to understand the biochemical pathways by identifying and quantifying the complete range of primary and secondary metabolites, providing knowledge of genes, transcripts and proteins involved in biological processes, and also helps in understanding their molecular mechanisms in plants. In a study, out of 169 metabolites in soybean detected using GC/MS and UHPLC-MS/MS, 104 were significantly variable in their levels across tested cultivars (Lin et al. 2014). Metabolite markers were identified to distinguish genetically-related soybean cultivars, and significant associations were reported within the same or among different metabolite groups. This approach potentially provides the basis for further studies on seed metabolism and metabolic engineering to improve the seed quality and yield of soybean.

## Ionomics

Ionomics is important for understanding element composition and their role in biochemical, physiological functionality and nutritional requirements of plants. P and K are the two key elements used as macronutrients in fertilizer to ensure a better crop yield. However, plants require many other elements and those are not uniformly distributed among different soil types. Plants have evolved with diverse element uptake ability at different locations because of diverse soil types (Fujita et al. 2013). This justifies the need for integrating ionomics with genomics to explore existing genetic differences. The plant ionome has become very sensitive and specific so that the element profile reflects different physiological states.

## Phenomics

Phenomics is concerned with high-throughput exploration of phenotypes. Precise phenotyping is important to recognize any genetic system. In plants, a fastidious phenotype is used to realize the biological status, such as infection, pest invasion or physiological confusions. The success of genomics is based on how reliable the connection is between a genetic marker and the phenotype. Therefore, phenomics integrated with other omics approaches has the most potential in plant breeding. Marker tracking approaches (Martrack leaf) have also been used to facilitate accurate analysis of two-dimensional leaf expansion with high temporal resolution (Mielewczik et al. 2013).

## Conclusion

*Glycine max* (L.) Merr, a crop with a balanced content of protein, fat and carbohydrate, serves as an important food, feed and bio-feedstock. Fast advancement of molecular techniques promises to fulfill the clear opportunities that lie ahead for soybean cultivation on a global level. The applications of different omics approaches have made the handling of segregating populations more scientific and the selection process precise. Scientific discoveries in the area of structural and functional plant genomics would lead to the evolution of new soybean varieties with the required uniqueness to enhance agronomic properties alongside conventional breeding methods. Linking biotechnology with plant breeding can certainly lead to a yellow revolution in oil-seed crops like soybean with concomitant improvement in seed quality.

## References

Ai-Hua S, Yin-Hua C, Zhi-Hui S et al (2014) Identification of photoperiod-regulated gene in soybean and function analysis in *Nicotiana benthamiana*. J Genet 93:43–51

Alam I, Lee DG, Kim KH et al (2010) Proteome analysis of soybean roots under waterlogging stress at an early vegetative stage. J Biosci 35:49–62

Bisen A, Khare D, Nair P et al (2015) SSR analysis of 38 genotypes of soybean [Glycine Max (L.) Merr.) genetic diversity in India. Physiol Mol Biol Plants 21:109–115

BMT/13/13 (2011) SSR markers in Brazilian soybeans; working group on biochemical and molecular techniques, and DNA-profiling in particular. Thirteenth Session Brasilia, November 22–24, 2p

Daymann V, Senthil N, Ravnidran M et al (2009) Density analysis in selected Indian soybean using morphological and SSR markers. IJIB 5:125–129

Duressa D, Soliman K, Taylor R et al (2011) Proteomic analysis of soybean roots under aluminum stress. Int J Plant Genom. doi:10.1155/2011/282531

Dwiyanti MS, Ujiie A, Thuy LTB et al (2007) Genetic analysis of high a-tocopherol content in soybean seeds. Breed Sci 57:23–28

Dwiyanti MS, Yamada T, Sato M et al (2011) Genetic variation of g-tocopherol methyltransferase gene contributes to elevated a-tocopherol content in soybean seeds. BMC Plant Biol 11:152

Fu Y-Bi, Ahmad Z, Diederichsen A (2015) Towards a better monitoring of seed ageing under ex situ seed conservation. Conserv Physiol. doi:10.1093/conphys/cov026

Fujita Y, Venterink HO, vanBodegom PM et al (2013) Low investment in sexual reproduction threatens plants adapted to phosphorus limitation. Nature 505:82–86

Funatsuki H, Ishimoto M, Suji HT et al (2006) Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. Plant Breed 125:195–197

Ge Y, Li Y, Zhu YM et al (2010) Global transcriptome profiling of wild soybean (*Glycinesoja*) roots under NaHCO$_3$ treatment. BMC Plant Biol 10:153

Gulluoglu L, Halis A, Mehmet A (2006) Effects of some plant growth regulators and nutrient complexes on pod shattering and yield losses of soybean under hot and dry conditions. J Plant Sci 5:368–372

Gwata ET, Wofford DS (2013) Potential of RAPD analysis of the promiscuous nodulation trait in soybean (Glycine max L). Biol Fertil Soil 49:241–244

Han Z, Ku L, Zhang Z et al (2014) QTLs for seed vigor-related traits identified in maize seeds germinated under artificial aging conditions. PloS one 9:e92535

Hosamani J, Arun Kumar MB, Talukdar A et al (2013) Molecular characterization and identification of candidate markers for seed longevity in soybean [*Glycine max* (L.) Merill]. Indian J Genet 73(1):64–71

Hu Z, Zhang D, Zhang G et al (2014) Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). Breed Sci 63:441–449

Hwang EY, Song Q, Jia G et al (2014) Genome-wide association study of seed protein and oil content in soybean. BMC Genom 15:1–12

Hyten DL, Choi IY, Song Q et al (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci 50:960–968

Iqbal Z, Naeem R, Ashraf M et al (2015) Genetic diversity of soybean accessions using seed storage proteins. Pak J Bot 47:203–209

Islam S, Mir JI, Kudesia R (2015) Evaluation of genetic diversity in *Vigna radiata* (L.) using protein profiling and molecular marker (RFLP). Int J Plant Breed Genet 9:238–246

ISTA (2012) Updates on progress towards development of DNA based approach for testing variety identity. Report by DNA working group held on 11–14 June 2012, Venlo, The Netherlands

Jeong N, Moon JK, Kim HS et al (2011) Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean. Theor Appl Genet 122:865–874

Jian S, Lu WG, Wen ZX et al (2012) Analysis of protein subunit variation in the mini core collection of cultivated soybean. J Henan Agric Sci 41:42–45

Khare D, Bhale MS, Raut ND et al (1996) Factors influencing germination and field emergence of soybean. JNKVV Res J 30:1–15

Khare D, Bisen A, Nair P et al (2013) Genetic diversity in soybean germplasm identified by RAPD markers. AsPac J Mol Biol Biotechnol 21:121–123

Ko KP, Park SK, Yang JJ et al (2013) Intake of soy products and other foods and gastric cancer risk: a prospective study. J Epidemiol 23:337

Komatsu S, Yamamoto R, Nanjo Y et al (2009) A comprehensive analysis of the soybean genes and proteins expressed under flooding stress using transcriptome and proteome techniques. J Proteome Res 8:4766–4778

Le DT, Nishiyama R, Watanabe Y et al (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. PloS one 7:e49522

Li YH, Li W, Zhang C et al (2010) Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. New Phytol 188:242–253

Li L, HurM Lee JY et al (2015) A systems biology approach toward understanding seed composition in soybean. BMC Genom 16(Suppl 3):S9

Lin H, Rao J, Shi J et al (2014) A seed metabolomic study reveals significant metabolite variations and correlations among different soybean cultivars. J Integr Plant Biol. doi:10.1111/jipb.12228

Lira-Medeiros CF, Cardoso MA, Fernandes RA et al (2015) Analysis of genetic diversity of two mangrove species with morphological alterations in a natural environment. Diversity 7:105–117

Liu S, Zhou R, Tian S et al (2007) A study on subunit groups of soybean protein extracts under SDS-PAGE. J Am Oil Chem Soc 84:793–801

Mariela I, Castro AL, Capdevielle F (2011) Development of functional markers associated with phenotypic traits for identification in soybean. Working group on biochemical and molecular techniques, and DNA-profiling in particular. Thirteenth session Brasilia, November 22–24, 2011 BMT/13/9 2p

Mielewczik M, Friedli M, Kirchgessner N et al (2013) Diel leaf growth of soybean: an ovel method to analyze two-dimensional leaf expansion in high temporal resolution based on a marker tracking approach (Martrack Leaf). Plant Methods 9:30

Mladenovic DS, Nikolic A, Peric V (2008) Cluster analysis of soybean genotypes based on RAPD markers. Proc 43rd Croatian and 3rd int symp agricul, pp 367–370, ISBN 978-953-6135-67-7, opatija, February 2008, Faculty of agronomy, Zagreb, Croatia

Nakamura T, Yang D, Kalaiselvi S et al (2003) Genetic analysis of net like cracking in soybean seed coats. Euphytica 133:179–184

Nimnual A, Romkaew J, Chukeatirote E et al (2014) Evaluation of genetic relationship among some important Japanese and Thai soybean varieties using AFLP analysis. Aust J Crop Sci 8:481–485

O'Rourke J, Charlson D, Gonzalez D et al (2007) Microarray analysis of iron deficiency chlorosis in near-isogenic soybean lines. BMC Genom 8:476

Qin J, Gu F, Liu D et al (2013) Proteomic analysis of elite soybean Jidou17 and its parents using iTRAQ-based quantitative approaches. Proteome Sci 11:12

Qiu LJ, Xing LL, Guo Y et al (2013) A platform for soybean molecular breeding: the utilization of core collections for food security. Plant Mol Biol 83:41–50

Rabel M, Vieira ESN, Lana UGP et al (2010) Marcadores-moleculares micros satellites naavaliação de sementes de soma comvariaçãonacoloração do hilo. Rev Brasil Sement 32:19–25

Rani A, Kumar V, Verma SK et al (2007) Tocopherol content and profile of soybean: genotypic variability and correlation studies. J Am Oil Chem Soc 84:377–383

Sahu P, Khare D, Tripathi N et al (2012) Molecular screening for disease resistance as strategic and tactical gene pool in soybean. J Food Leg 25:200–205

Sattler SE, Cahoon EB, Coughlan SJ et al (2003) Characterization of tocopherol cyclases from higher plants and cyano-bacteria. Evolutionary implications for tocopherol synthesis and function. Plant Physiol 132:2184–2195

Shi Z, Bachleda N, Pham AT et al (2015) High-throughput and functional SNP detection assays for oleic and linolenic acids in soybean. Mol Breed 35:176

Singh RK, Raipuria RK, Bhatia VS et al (2008) SSR markers associated with seed longevity in soybean. Seed Sci Technol 36:162–167

Song QJ, Marda LF, Shoemaker RC, Lark KG, Concibido VC et al (2004) A new integrated genetic linkage map of the Soybean. Theor Appl Genet 109:122–128

Song F, Tang DL, Wang XL et al (2011) Biodegradable soy protein isolate-based materials: a review. Biomacromol 12:3369–3380

Song Q, Hyten DL, Jia G et al (2013) Development and evaluation of SoySNP50 K, a high-density genotyping array for soybean. PloS one 8:e54985

Song Q, Hyten DL, Jia G et al (2015) Fingerprinting soybean germplasm and its utility in genomic research. Genetics 5:1999–2006

Suzuki M, Fujino K, Nakamoto Y et al (2010) Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. Mol Breed 25:407–418

Tardivel A, Sonah H, Belzile F et al (2014) Rapid identification of alleles at the soybean maturity gene E3 using genotyping by sequencing and a haplotype-based approach. Plant Gen 7:1–9

UPOV (2011) Working group on biochemical and molecular techniques, and DNA profiling in session thirteenth held at Brazil. November 22–24, BMT/13/25

Wang LX, Guan RX, Li YH et al (2007) Genetic diversity of chinese spring soybean germplasm revealed by SSR markers. Plant Breed 127:56–61

Zhou Z, Jiang Y, Wang Z et al (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol 33:408–414

Ziegler G, Terauchi A, Becker A et al (2013) Ionomic screening of field-grown soybean identifies mutants with altered seed elemental composition. Plant Gen 6:1–9