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

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

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

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

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

Genetic Diversity and Allele Mining in Soybean Germplasm

Reda Helmy Sammour
Tanta University, Faculty of Science, Botany Department
Egypt

1. Introduction

Soybean, *Glycine max* (L.) Merrill is recognized as the most important grain legume in the world in terms of total production and international trade (Golbitz, 1995), being an important source of protein and oil. There are developing thousands of breeding lines and hundreds of elite cultivars yearly in the soybean hybridization programmes over the world. The developing of these breeding lines increased genetic uniformity in the frame of species. Therefore, the genetic basis of these released cultivars is rather narrow. Generations of new and improved cultivars can be enhanced by new sources of genetic variation; therefore criteria for parental stock selection need to be considered not only by agronomic value, but also from the point of view of their genetic dissimilarity. That is why the evaluation of genetic variation is a very important task not only for population genetics but also for plant breeders. The study of genetic variation has fallen within population genetics which has focused on analyzing, measuring and partitioning genetic. The genetic diversity can be analyzed by agronomic and biochemical traits, and molecular marker polymorphisms, Analysis of gene marker data enables estimation of the mating system and monitoring of genetic changes caused by factors affecting the reproductive biology of a species. A key factor driving utilization of exotic germplasm is potential benefit. Benefit can be quite apparent for characteristics such as disease resistance or agronomic traits, but vague for yield or abiotic stress resistance.

2. Origin and diversification center of the soybean

Scholars generally agree that cultivated soybean (*Glycine max*) has originated in the eastern half of North China in the eleventh century B.C. or perhaps a bit earlier (Fukuda, 1933 and Singh, 2010). It is believed on world wide scale that soybean has been domesticated from the annual wild soybean *Glycine soja* Sieb.*et* Zucc. Many studies based on old Chinese literature, the geographic distribution of the wild ancestral species, the levels and types of genetic diversity of soybean varieties and the archeological evidence consistently indicated that China is the origin and diversification center of the cultivated soybean (Fukuda, 1933; Hymowitz, 1970; Zhuang, 1999). The evidences that China is the origin and main center of diversity of soybean are (1) the distribution of *G. soja* in China is the most extensive in terms of the numbers and diversity of types; (2) China has the earliest written records of soybean cultivation, about 4500 years ago; (3) soybean has been found in unearthed artifacts; (4)

soybeans cultivated in different countries in the world were introduced directly or indirectly from China; and (5) the pronunciation of the word of soybean in many countries is about the same as the Chinese 'Shu'; for instance, it is pronounced 'soya' in England, 'soy' in the USA, and in other languages.

Although, the origin of soybean cultivation may be China, scholars have different viewpoints on the original areas of soybean domestication. One of these views is the theory that soybean originated from northeast China (Fukuda, 1933). This theory based on the observations that semi-natural wild soybeans are extensively distributed in northeast China, that is, there are large numbers of soybean varieties that possess 'primitive' characteristics, such as small black soybean germplasm that extensively distributed in the lower and middle reaches of the yellow river North provinces. The second theory is that soybean cultivation originated in South China. In this theory, it has been thought that south China could be the origin of soybean (Wang, 1947). The evidences for that are the wide distribution of wild soybean in this area, extensive presences of primitive soybean varieties such as Nidou, Maliao Dou, Xiao Huangdou and others that have (1) the short-day character, which is considered to be the initial physiological state of soybean, and (2) the primitive agronomic characteristics related to yield and quality of soybean varieties. The other evidence supporting this theory is the close relatedness between cultivated soybeans in southern China, to wild soybeans in genetic terms based on isoenzymes, and RFLP (Restriction Fragment Length Polymorphism) markers of chloroplast and mitochondrial DNA, SSR data and botanical traits (Ding *et al.*, 2008; Guo *et al.*, 2010). In the third theory, it has been thought that the origin of soybean was the eastern part of northern China (i.e. the lower reaches the Yellow River) (Hymowitz, 1970). The evidences for his thought are the same blooming dates for both wild soybean and cultivated soybean at 35°N, confirming that cultivated soybean varieties may have been derived from local wild soybean at around 35°N. In addition, the protein content of cultivated soybean is close to that of wild soybean at 34–35°N. The fourth theory stated that the cultivated soybeans have multiple origins (Lü, 1978). The evidences for that postulation are (1) both South and North China have regions with early developed cultures, that is: the ancients in these regions used local wild soybean as food and did not domesticated wild soy-beans into cultivated ones; (2) the occurrence of wild soybean and cultivated soybean in the same regions and the similarities of both of them in morphological characters; (3) the successful cultivation of both wild and cultivated soybeans in different regions across China. In addition, the geographical distribution of the short-day character of wild soybean indicates the possibility of multiple origins of cultivated soybean.

3. Genetic diversity of soybean germplasm based on morphological traits

As we know, phenotypic traits are controlled by genes and affected by environment, but large numbers of accessions can adapt to environments. The phenotypic data has more polymorphism in genetic diversity and reveal genetic variation indirectly. On the contrary, the molecular data reveal genetic variation directly, but fewer markers have less polymorphism. It is very difficult to obtain molecular data for a large number of accessions that has enough polymorphism to show the genetic diversity of germplasm. So, the morphological traits are the suitable and practical tools for studying the genetic diversity on large numbers of accessions.

Variation in shape of plants has always been an important means of (1) distinguishing individuals; (2) controlling seed production; and (3) identifying the negative traits those

effects on yield, the genetic diversity centers of annual wild soybean and the soybean lines resistance to pod shatter, drought, pests or disease (Truong *et al.,* 2005; Malik *et al.,* 2006, 2007; Ngon *et a.,l* 2006). The studied soybean germplasm exhibited a wide range of phenotypic variation for pod number, seed number, and plant yield. It also showed that soybean developing stages had close association with agronomic traits as well as yield and yield components (Malik *et al.,* 2006, 2007; Ngon *et al.,* 2006).

Pod shape is one of the important descriptors for evaluating soybean genetic resources (IPGRI, 1998; USDA, 2001). Truong *et al.* (2005) tested the applicability of elliptic Fourier method for evaluating genetic diversity of pod shape in 20 soybean (*Glycine max* L. Merrill) genotypes. They concluded that principal component scores based on elliptic Fourier descriptors yield seemed to be useful in quantitative parameters not only for evaluating soybean pod shape in a soybean breeding program but also for describing pod shape for evaluating soybean germplasm.

The genetic diversity was evaluated for genotypes of soybean based on the yield-related traits (Rajanna *et al.,* 2000; Malik *et al.,* 2006, 2007; Ngon *et al.*, 2006). It has been reported that differences among genotypes for all the characters were highly significant and the grain yield was positively and significantly correlated with number of pods per plant. The selection for the character had positive direct effect on yield. However, some traits had negative direct effects on yield, such as the leaf area, first pod height, days to 50% flowering, days to flowering completion, days to maturity, plant height, oil content and protein content.

The study of the genetic diversity of wild soybean is invaluable for efficient utilization, conservation and management of germplasm collections. Dong *et al.* (2001) statistically analyzed the agronomic traits of the data base from the National Germplasm Evaluation Program of China to study the geographical distribution of accessions, genetic diversity of characters and genetic diversity centers of annual wild soybean. The results showed that most annual wild soybeans are distributed in northeast China, and the number of accessions decreases from the northeast to other directions in China. They proposed three genetic diversity centers for annual soybean grown in China, the northeast, the Yellow River Valley and the Southeast Coasts of China. Based on these results and Vavilov's theory of crop origination, two opposing possible models for the formation of the three centers are proposed, either these centers are independent of each other and the annual wild soybeans in these centers originated separately, or the northeast center was the primary center for annual wild soybeans in China, while the Yellow River Valley center was derived from this primary center and served as the origin for the southeast Coast center.

The genetic variability in 131 accessions of edamame soybeans (the Japanese name for a type of vegetable soybean eaten at the immature R6 stage) was analyzed using phenotypic traits e. g. maturity information, testa color, and 100-seed weight for breeding new edamame lines resistance to pod shatter (Mimura, 2001). The 131 accessions include 108 Japanese edamame, 11 Chinese maodou, 8 WSU breeding lines, 2 US edamame and 2 US grain soybeans. The obtained results indicated that Edamame genetic diversity was generally clustered around maturity groups and testa color. It was also reported that the genetic diversity among the Japanese edamame cultivars was narrow, compared to Chinese maodou; Japanese edamame and Chinese maodou soybeans may have different genetic pools.

Soybean genotypes, which exhibit genetic diversity in root system developmental plasticity in response to water deficits in order to enable physiological and genetic analyses of the regulatory mechanisms involved, were identified (Young, 2008). These genotypes can

tolerate drought stress which is the major factor that limiting soybean yield. The results showed substantial genetic diversity in the capacity for increased lateral root development (number and total length of roots produced) and in the responses of overall root and shoot growth under water deficit conditions.

The extent of between- and within-species differences in the resistance of the four commonest species of Glycine (*G. canescens*, *G. clandestina* , *G. tabacina* and *G. tomentella*) to leaf rust caused by *Phakopsora pachyrhizi* was investigated by Burdon & Marshall (1981). The results of their study showed qualitative and quantitative resistance to leaf rust, and considerable variation in a number of disease characteristics both between and within populations of each species.

4. Genetic diversity in soybean germplasm based on karyological traits

Genetic diversity based on genome size among and within plant species has been well documented in the literature (Rayburn, 1990; Bennett and Leitch, 1995; Rayburn *et al.*, 1997). The variation was pronounced in Chinese germplasm collected from diverse geographic locations. It was attributed to the environmental factors (Knight and Ackerly, 2002), cell size, minimum generation time, cell division rate and growth rate (Edwards and Endrizzi, 1975; Bennett *et al.*, 1983) and polypoid species, in species with large seeds, and habits type (Bennett *et al.*, 1998; Chung *et al.*, 1998).

Reports of genome size variation in soybean [*Glycine max* (L.)] have ranged from 40 to 0% (Rayburn *et al.*, 2004). This wide range is highly reproducible and has resulted in doubts of the existence of intra-specific DNA variation in soybean. Rayburn *et al*. (2004) determined genome size of 18 soybean lines, selected on the basis of diversity of origin, by flow cytometry. They found that genome size variation between these lines was at approximately 4%. This amount of DNA variation is lower than was originally reported (Doerschug *et al.*, 1978; Yamamota and Nagato, 1984; Hammatt *et al.*, 1991; Graham *et al.*, 1994). Doerschug *et al*. (1978) is the first to determine genome size of soybean, upon examining 11 soybean lines, reporting over a 40% variation in nuclear DNA content. Graham *et al*. (1994) observed a 15% variation among soybean cultivars while Rayburn *et al*. (1997) reported a 12% variation among 90 Chinese soybean introductions. Chung *et al*. (1998) observed among 12 soybean strains a 4.6% DNA content variation. Yamamota and Nagato (1984) stated about 60% variation, while Hammatt *et al*. (1991) reported that the variation of genome size in 14 different *Glycine* species from different parts of the world was approximately 58%. These results indicated that the variability between DNA content was varied between the different scholars. The wide variation in genome size between soybean germplasm makes these accessions good candidates for crop improvement.

5. Evaluation of genetic diversity in soybean germplasm at the biochemical level

The genetic markers have made possible a more accurate evaluation of the genetic and environmental components of variation. The biochemical markers are ones of the interesting measures of genetic diversity. They include protein techniques and isozymes. The protein techniques are practical and reliable methods for cultivars and species identification because seed storage proteins are largely independent of environmental fluctuation (Sammour, 1992, 1999; Camps *et al.*, 1994; Jha and Ohri, 1996). They are less expensive as compared to DNA

markers. SDS-PAGE is one of these techniques, widely used to describe seed protein diversity of crop germplasm (Sammour, 2007; Sammour *et al.,* 2007). Genetic diversity and the pattern of variation in soybean germplasm have been evaluated with seed proteins (Hirata *et al.,* 1999; Bushehri *et al.,* 2000; Sihag *et al.,* 2004; Malik *et al.,* 2009). SDS-PAGE (Bushehri *et al.,* 2000) and discontinuous polyacrylamide slab gel electrophoresis (Sharma and Maloo, 2009) were used very successfully in evaluating the genetic diversity and identifying soybean (*Glycine max*) cultivars. Malik *et al.,* (2009) evaluated the genetic variation in 92 accessions of soybean collected from five different geographical regions using the electrophoretic patterns of seed proteins. The accessions from various sources differed considerably, indicating that there is no definite relationship between genetic diversity and geographic diversity. Similar results were reported by (Ghafoor *et al.,* 2003). Based on the results of Ghafoor *et al.,* (2003) and Malik *et al.,* (2009), SDS-PAGE cannot be used for identification of various genotypes of wild soybean at the intra-specific level, because some of the accessions that differed on the basis of characterization and evaluation exhibited similar banding patterns. However, it might be used successfully to study inter rather than intra-specific variation (Sammour, 1989; Sammour *et al.,* 1993; Karam *et al.,* 1999; Ghafoor *et al.,* 2002). 2-D electrophoresis can be used to characterize the genotypes exhibited similar banding patterns (Sammour, 1985).

Allozyme markers have been used in soybean to evaluate genetic diversity in accessions from diverse geographic regions (Yeeh *et al.,* 1996; Chung *et al.,* 2006), wild soybean in natural populations from China, Japan and South Korea (Pei *et al.,* 1996; Fujita *et al.,* 1997), and Asian soybean populations (Hymowitz & Kaizuma, 1981; Hirata *et al.,* 1999). From an analysis of the Kunitz trypsin inhibitor (*Ti*) and beta-amylase isozyme (*Sp1* = *Amy3*), Hymowitz & Kaizuma (1981) defined seven soybean germplasm pools in Asia: (1) northeast China and the USSR, (2) central and south China, (3) Korea, (4) Japan, (5) Taiwan and south Asia, (6) north India and Nepal and (7) central India. Hirata *et al.* (1999) compared the genetic variation at 16 isozyme of 781 Japanese accessions with the genetic variations of 158 Korean and 94 Chinese accessions, detecting a number of region-specific alleles that discriminated Japanese from Chinese accessions. The presence of alleles specific to the Japanese population suggested that the present Japanese soybean population was not solely a subset of the Chinese population.

6. Evaluation of genetic diversity in soybean germplasm using molecular markers

6.1 Introduction

The soybean genome is consisting of around 1115 Mbp, much smaller than the genomes of maize and barley, but larger than the genomes of rice and *Arabidopsis* (Arumuganathan & Earle, 1991). Soybean is a tetraploid plant, evolved from a diploid ancestor (n=11), went aneuploid loss (n=10), followed by polyploidization (n=20) and diploidization (chromosome pairing behavior) (Hymowitz, 2004). As a result of polyploidization soybean has a significant percentage of internal duplicated regions distributed among its chromosomes (Pagel *et al.,* 2004). Sequence diversity in cultivated soybean is relatively low compared to other species leading to a major challenge in the improvement of this important crop. To efficiently broaden the genetic base of modern soybean cultivars, we have a detailed insight into genetic diversity of soybean germplasm. Such insight could be achieved through molecular characterization using DNA markers, which are more informative, stable and

reliable, compared to pedigree analysis and traditionally used morphological markers. The genetic markers include RFLP, RAPD, SSR and AFLP markers were used to probe the genetic differences between wild and cultivated soybeans or for the origin and dissemination of soybeans (Brown-Guedira *et al*., 2000; Tian *et al*., 2000; Li & Nelson, 2001; Xu & Zhao, 2002; Abe *et al*., 2003). These studies have revealed higher levels of genetic diversity in wild soybean.

6.2 RFLP (Restriction Fragment Length Polymorphism)
This analysis exploits variation in the occurrence of restriction sites in genomic sequences hybridizing to a cloned probe. Originally, RFLP analysis required Southern blotting and hybridization, making the method fairly slow and laborious. This technique is still used to generate ''anchor'' markers, used by many scholars to make consensus recombinational maps, though it is often implemented with the polymerase chain reaction (PCR) to generate the polymorphic fragments (Schulman, 2007).
Chung *et al*. (2006) evaluated levels of genetic diversity in USDA soybean germplasm (107 accessions), originated from six provinces in central China, using RFLP analysis. They detected significant genetic differentiation among the six provinces (mean GST = 0.133). These results suggest that Chinese germplasm accessions from various regions or provinces in the USDA germplasm collection could be used to enhance the genetic diversity of US Cultivars.

6.3 AFLP (Amplified Fragment Length Polymorphism)
AFLP is an anonymous marker method, detects restriction sites by amplifying a subset of all the sites for a given enzyme pair in the genome by PCR between ligated adapters. To some extent, it like RFLP detects single nucleotide polymorphisms (SNPs) at restriction sites.
Ude *et al*. (2003) analyzed the genetic diversity within and between Asian and North American soybean cultivars by AFLP. They found that the average genetic distance between the North American soybean cultivars and the Chinese cultivars was 8.5% and between the North American soybean cultivars and the Japanese cultivars was 8.9%, but the Chinese soybean was not completely separated from the Japanese soybean. They also revealed that Japanese cultivars may constitute a genetically distinct source of useful genes for yield improvement.

6.4 RAPD (Random Amplified Polymorphic DNA)
RAPD analysis uses conserved or general primers that amplify from many anonymous sites throughout the genome. It is indeed rapid, and need only short primers of random sequence, but suffers from low polymorphism information content (PIC), poor correlation with other marker data, and problems in reproducibility due to the low annealing temperatures in the reactions.
The genetic diversity in the wild soybean populations from the Far East region of Russia was analyzed using RAPD markers (Seitova *et al*., 2004). The results obtained suggest that (1) genetically different groups of wild soybean have active development, (2) level of polymorphism was significantly higher than in the cultivated soybean and (3) geographically isolated subpopulations showed maximum distance from the main population of wild soybean. The high level of polymorphism between the wild and cultivated soybean accessions was also reported by Kanazawa *et al*. (1998) in their study on

soybean accessions from the Far East using RAPD profiles of mitochondrial and chloroplast DNA. Xu & Gai (2003), Pham Thi Be Tu *et al.* (2003), An *et al.* (2009) confirmed the results of Kanazawa *et al.* (1998) and Seitova *et al.* (2004) in terms of the high genetic variation between the wild and cultivated soybean accessions. They also found that the diversity of *G. soja* was higher than that of *G. max*; and environmental factors may play important roles in soybean evolution. Furthermore, they revealed that accessions within each species tend to form sub-clusters that are in agreement with their geographical origins, demonstrating that an extensive geographical genetic differentiation exists in both species. Consequently, it was indicated that geographical differentiation plays a key role in the genetic differentiation of both wild and cultivated soybeans. The relationship between geographical differentiation and genetic diversity appeared in the work of Chen & Nelson (2005) who identified significant genetic differences between soybean accessions collected from different provinces in China. Their data provided pronounced evidence that primitive cultivars of China were generally genetically isolated in relatively small geographical areas. Similar results were obtained by Li & Nelson (2001, 2002) in their study on soybean accessions from 8 provinces in China using a core set of RAPD primers with high polymorphism in soybean (Thompson *et al.*, 1998). On the contrary, Brown-Guedira *et al.* (2000) did not find an association between origin and RAPD markers among soybean lines of more modern origin. It is likely that these genotypes have been dispersed by human intervention from the areas of actual origin.

The relationship between genetic differentiation and origin of 120 soybean accessions from Japan, South Korea and China was evaluated with RAPDs (Li & Nelson, 2001). They found that the Japanese and South Korean populations were more similar to each other, whereas both were genetically distinct from the Chinese population, suggesting that the S. Korean and Japanese gene pools might be probably derived from a relatively few introductions from China. Li *et al.* (2001) compared the genetic diversity of ancestral cultivars of the N. American (18) as well as the Chinese soybean germplasm pools (32) using RAPD markers, the N. American ancestors have a slightly lower level of genetic diversity. Cluster analyses generally separated the two gene pools. In particular, a great genetic variability was detected between the ancestors of northern U.S. and Canadian soybeans and the Chinese ancestors.

Chowdhury *et al.* (2002) examined the level of genetic similarity among forty-eight soybean cultivars imported out of their country Thailand using DNA (RAPD) markers. They found high level of genetic similarities between these cultivars. Cluster analysis of the obtained data classified the 48 cultivars into four groups at 0.57 similarity scale, even though the cultivars are morphologically or geographically very close. Comparing agronomic performance and RAPD analysis via dendrogram, a total of 11 cultivars can be useful to soybean breeders in Thailand who want to utilize genetically diverse introductions in soybean improvement. Baránek *et al.* (2002) evaluated the genetic diversity within 19 soybean genotypes included in the Czech National Collection of Soybean Genotypes by RAPD method. The polymorphism among the studied genotypes was 46%. Presented results enable the selection of genetically distinct individuals. Such information may be useful to breeders willing to use genetically diverse introductions in soybean improvement process.

6.5 SSRs (Simple sequence repeats)
SSRs molecular markers have been widely applied in the genetic diversity studies of the soybean germplasm (Abe *et al.*, 2003; Wang *et al.*, 2006; Fu *et al.*, 2007; Li *et al.*, 2008; Wang &

Takahata, 2007; Wang *et al.*, 2008; Yoon *et al.*, 2009). The advantages of SSR over other types of molecular markers are that they are abundant, have a high level of polymorphism, are codominant, can be easily detected with PCR and typically have a known position in the genome. High levels of polymorphism at SSR loci have been reported for both the number of alleles per locus and the gene diversity (Diwan & Cregan, 1997; Abe *et al.*, 2003; Wang *et al.*, 2006; Fu *et al.*, 2007 ; Wang *et al.*, 2010).

Wang *et al.* (2010) used 40 SSR primer pairs to study genetic variability in 40 soybean accessions of cultivars, landraces and wild soybeans collected from China. These results indicated that wild soybeans and landraces possessed greater allelic diversity than cultivars and might contain alleles not present in the cultivars which can strengthen further conservation and utilization. The UPGMA (Unweighted Pair Group Method with Arithmetic) results also exhibited that wild soybean was of more abundant genetic diversity than cultivars.

A total of 2,758 accessions of Korean soybean landraces were profiled and evaluated for genetic structure using six SSR loci (Yoon *et al.*, 2009). The accessions within collections were classified based on their traditional uses such as sauce soybean (SA), sprouted soybean (SP), soybean for cooking with rice (SCR), and others-three different Korean *Glycine max* collections and for groups distinguished by their usage, such as SA, SP, and SCR. Nei's average genetic diversity ranged from 0.68 to 0.70 across three collections, and 0.64 to 0.69 across the usage groups. The average between-group differentiation (Gst) was 0.9 among collections, and 4.1 among the usage groups. The similar average diversity among three collections implies that the genetic background of the three collections was quite similar or that there were a large number of duplicate accessions in three collections (Yoon *et al.*, 2009). The selection from the four groups classified based upon usage may be a useful way to select accessions for developing a Korean soybean landrace core collection at the RDA gene bank.

Hudcovicová *et al.* (2003) analyzed allelic profiles at 18 SSR loci of 67 soybean genotypes of various origins. Six only of SSR markers differentiated all 67 genotypes each from others successfully. Guan *et al.* (2010) investigated the genetic relationship between 205 Chinese soybean accessions that represent the seven different soybean ecotypes and 39 Japanese soybean accessions from various regions using 46 SSR loci. Cluster analysis with UPGMA separated the Chinese accessions from Japanese accessions, suggesting that soybean in these two countries form different gene pools. It also showed that (1) accessions from China have more genetic diversity than those from Japan, (2) studied germplasm was divided into three distinct groups, "corresponding to Japanese soybean, Northern China soybean, Southern China soybean and a mixed group in which most accessions were from central China", and (3) Japanese accessions had more close relationship with Chinese northeast spring and southern spring ecotypes. This study provides interesting insights into further utilization of Japanese soybean in Chinese soybean breeding.

Abe *et al.* (2003) analyzed allelic profiles at 20 SSR loci of 131 accessions introduced from 14 Asian countries. UPGMA-cluster analysis clearly separated the Japanese from the Chinese accessions, suggesting that the Japanese and Chinese populations formed different germplasm pools; showed that Korean accessions were distributed in both germplasm pools, whereas most of the accessions from south/central and southeast Asia were derived from the Chinese pool; indicated that genetic diversity in the southeast and south/central Asian populations was relatively high; and exhibited the absence of region-specific clusters in the southeast and south/central Asian populations. The relatively high genetic diversity

and the absence of region-specific clusters in the southeast and south/central Asian populations suggested that soybean in these areas has been introduced repeatedly and independently from the diverse Chinese germplasm pool. Therefore the two germplasm pools can be used as exotic genetic resources to enlarge the genetic bases of the respective Asian soybean populations.

Chotiyarnwong *et al.* (2007) evaluated the genetic diversity of 160 Thai indigenous and recommended soybean varieties by examining the length polymorphism of alleles found in 18 SSR loci from different linkage groups. UPGMA-Cluster analysis and principal component analysis (PCA) separated Thai indigenous varieties from recommended soybean varieties. However, the genetic differentiation between the indigenous and recommended soybean varieties was small.

Shi *et al.* (2010) performed genetic diversity and association analysis among 105 food-grade soybean genotypes using 65 simple sequence repeat (SSR) markers distributed on 20 soybean chromosomes. Based on the SSR marker data, the 105 soybean genotypes were divided into four clusters with six sub-groups. Thirteen SSR markers distributed on 11 chromosomes were identified to be significantly associated with oil content and 19 SSR markers distributed on 14 chromosomes with protein content. Twelve of the SSR markers were associated with both protein and oil QTL. A negative correlation was obtained between protein and oil content.

Mimura *et al.* (2007) investigated SSR diversity in 130 vegetable soybean accessions including 107 from Japan, 10 from China and 12 from the United States. Eighteen of the 130 accessions were outliers, and the rest of the accessions were grouped into nine clusters. The majority of food-grade soybean cultivars were released from Japan and South Korea because of the market availability and demands. However, the genetic diversity of South Korea food-grade soybean remains unreported (Mimura *et al.*, 2007).

Nguyen *et al.* (2007) used 20 genomic SSR and 10 EST-SSR SSR to explore the genetic diversity in accessions of soybean from different regions of the world. The selection of the thirty SSR primer-pairs was based on their distribution on the 20 genetic linkage groups of soybean, on their trinucleotide repetition unit and on their polymorphism information content. All analyzed loci were polymorphic. A low correlation between SSR and EST-SSR data was observed, thus genomic SSR and EST-SSR markers are required for an appropriate analysis of genetic diversity in soybean. They observed high genetic diversity which allowed the formation of five groups and several subgroups. They also observed a moderate relationship between genetic divergence and geographic origin of accessions.

Xie *et al.* (2005) analyzed genetic diversity of 158 Chinese summer soybean germplasm, from the primary core collection of *G. max* using 67 SSR loci. The Huanghuai and Southern summer germplasm were different in the specific alleles, allelic-frequencies and pairwise genetic similarities. UPGMA cluster analysis based on the similarity data clearly separated the Huanghuai from Southern summer soybean accessions, suggesting that they were different gene pools. The data indicated that Chinese Huanghuai and Southern summer soybean germplasm can be used to enlarge genetic basis for developing elite summer soybean cultivars by exchanging their germplasm.

Most diversity studies on cultivated soybean published by now have focused on North American (Brown-Guedira *et al.*, 2000; Narvel *et al.*, 2000; Fu *et al.*, 2007) Asian (Abe *et al.*, 2003; Xie *et al.*, 2005; Wang *et al.*, 2006; Li *et al.*, 2008; Wang *et al.*, 2008; Yoon *et al.*, 2009) as well as South American (Bonato *et al.*, 2006) soybean germplasm. In several studies only a few genotypes of European origin have been represented among germplasm studied

(Brown-Guedira *et al*., 2000; Narvel *et al*., 2000; Fu *et al.,* 2007; Hwang *et al*., 2008). Baranek *et al*. (2002) evaluated genetic diversity of 19 G*lycine max* accessions from the Czech National Collection using RAPD markers. Recently, Tavaud-Pirra *et al.* (2009) evaluated SSR diversity of 350 cultivated soybean genotypes including 185 accessions from INRA soybean collection originating from various European countries and 32 cultivars and recent breeding lines representing the genetic improvement of soybean in Western Europe from 1950 to 2000. They found the genetic diversity of European accessions to be comparable with those of the Asian accessions from the INRA collection, whereas the genetic diversity observed in European breeding lines was significantly lower. Breeding material and registered soybean cultivars in southeast European countries are strongly linked to Western breeding programs, primarily in the USA and Canada. There is little reliable information regarding the source of germplasm introduction, its pedigree and breeding schemes applied. Consequently, use of these genotypes in making crosses to develop further breeding cycles can result in an insufficient level of genetic variability. Assessing the genetic diversity of this germplasm at genomic DNA level would complement the knowledge on the European soybean gene pool (germplasm) and facilitate the utilization of the resources from southeastern Europe by soybean breeders. Ristova *et al*. (2010) therefore assess genetic diversity and relationships of 23 soybean genotypes representing several independent breeding sources from southeastern Europe and five plant introductions from Western Europe and Canada using 20 SSR markers. Cluster analysis clearly separated all genotypes from each other assigning them into three major clusters, which largely corresponded to their origin. Results of clustering were mainly in accordance with the known pedigrees.

6.6 EST (Expressed Sequence Tags)

The use of functional molecular markers, such as those developed from EST allows direct access to the population diversity in genes of agronomic interest that they represent coding sequences, facilitating the association between genotype and phenotype. Nelson and Shoemaker (2006) identified approximately 45,000 potential gene sequences (pHaps) from EST sequences of Williams/Williams 82, an inbred genotype of soybean (*Glycine max* L. Merr.) using a redundancy criterion to identify reproducible sequence differences between related genes within gene families. Analysis of these sequences revealed single base substitutions and single base indels are the most frequently observed form of sequence variation between genes within families in the dataset. Genomic sequencing of selected loci indicates that intron-like intervening sequences are numerous and are approximately 220 bp in length. Functional annotation of gene sequences indicates functional classifications are not randomly distributed among gene families containing few or many genes. The identification of potential gene sequences (pHaps) from soybean allows the scientist to get a picture of the genomic history of the organism as well as to observe the evolutionary fates of gene copies in this highly duplicated genome.

7. Allele mining in soybean germplasm

7.1 Concept

Exploitation of gene banks for efficient utilization depends on the knowledge of genetic diversity, in general, and allelic diversity at candidate gene(s) of interest, in particular. Hence, allele mining seems to be a promising in characterization of genetic diversity or allelic/genic diversity among the accessions of the collection in terms of its utility for

improving a target trait (Kaur *et al.*, 2008). The availability of sequence and sequence variation that affects the plant phenotype is of utmost importance for the utilization of genetic resources in crop improvement (Graner, 2006).

The existing allelic diversity in any crop species is caused by mutations, the evolutionary driving force (Kumar *et al.*, 2010). Mutations create new alleles or cause variations in the existing allele and allelic combinations. They take place in coding and non-coding regions of the genome either as single nucleotide polymorphism (SNP) or as insertion and deletion (InDel). As far it is known, there is no cited literatures on the effect of mutations on transcript synthesis and accumulation which in turn alter the trait expression in 5′ UTR including promoter, introns and 3′ UTR in the genome of soybean. In coding region, it may have tremendous effect on the phenotype by altering the encoded protein structure and/or function. For example, the AtAHASL protein encoded by *csr1-2* differs from the native AtAHASL protein by one amino acid substitution of a serine with an asparagine at residue 653 (S653N) which results in tolerance to imidazolinone containing herbicides. Besides the altered herbicide binding, the protein retains its biological function in the plant. Soybean line CV127 is tolerant to herbicides that contain imidazolinone. The another example is the mutations in soybean microsomal omega-3 fatty acid desaturase genes which resulted in reduce of linolenic acid concentration in soybean seeds (Bilyeu *et al.*, 2005). Alternatively, several studies suggested that many diseases resistant alleles like soybean aphid [*Aphis glycines* Matsumura (Hemiptera: Aphididae)] resistance like *Rag1* from Germplasm collection (Kim *et al.*, 2010), brown Stem Rot resistance like *Rbs1* and *Rbs3* from soybean lines L78-4049 and PI 437.833, and PI 84946-2 (Eathington *et al.*, 1995; Klos *et al.*, 2000), soybean cyst nematode (SCN) resistance genes like *rhg*1 and *Rhg*4 *from soybean* lines PI 88788, PI 437.654, Peking, PI90763 and PI209332, sudden death syndrome (SDS) resistance like Rfs1, rfs2, and rft from soybean lines PI 437654 (Meksem *et al.*, 2001).

7.2 Approaches

Two major approaches are available for the identification of sequence polymorphisms for a given gene in the naturally occurring populations: (1) modified Targeting Induced Local Lesions in Genomes (TILLING) procedure and (2) sequencing based allele mining.

7.2.1 TILLING approach

In the TILLING approach, the polymorphisms (more specifically point mutations) resulting from induced mutations in a target gene can be identified by heteroduplex analysis (Till *et al*. 2003). This technique represents a means to determine the extent of variation in mutations artificially induced. EcoTilling represents a means to determine the extent of natural variation in selected genes in the primary and secondary crop gene pools (Comai and Henikoff, 2006 and Kumar *et al.* 2010). Like TILLING, it also relies on the enzymatic cleavage of heteroduplexed DNA, formed due to single nucleotide mismatch in sequence between reference and test genotype, with a single strand specific nuclease under specific conditions followed by detection through Li-Cor genotypers. At point mutations, there will be a cleavage by the nuclease to produce two cleaved products whose sizes will be equal to the size of full length product. The presence, type and location of point mutation or SNP will be confirmed by sequencing the amplicon from the test genotype that carry the mutation.

7.2.2 Sequencing-based allele mining

This technique involves amplification of alleles in diverse genotypes through PCR followed by identification of nucleotide variation by DNA sequencing. Sequencing-based allele

mining would help to analyze individuals for haplotype structure and diversity to infer genetic association studies in plants. Unlike EcoTilling, sequencing-based allele mining does not require much sophisticated equipment or involve tedious steps, but involves huge costs of sequencing. (Kumar *et al.*, 2010)

7.3 Applications
Allele mining can be effectively and efficiently used for (1) discovery of superior alleles, through 'mining' the gene of interest from diverse genetic resources, (2) providing insight into molecular basis of novel trait variations and identifying the nucleotide sequence changes associated with superior alleles, (3) studying the rate of evolution of alleles; allelic similarity/dissimilarity at a candidate gene and allelic synteny with other members of the family, (4) paving way for molecular discrimination among related species through development of allele-specific molecular markers, and (5) facilitating introgression of novel alleles through Marker Assisted Selection (MAS) or deployment through Genetic Engineering (GE). Allele mining can also be potentially employed in the identification of nucleotide variation at a candidate gene associated with phenotypic variation for a trait. Through this, the frequency, type and the extent of occurrence of new haplotypes and the resulting phenotypic changes can be evaluated.

7.4 Challenges
The genetic resources collections, which are held collectively in various gene banks, harbour a wealth of undisclosed allelic variants. Now the challenge is how to efficiently identify and exploit the useful variation of these collections to exploit in crop improvement. The challenges stand as stampling block to make use of these collections are (1) selection of genotypes, (2) handling genomic resources, (3) demarcation of promoter region, (4) characterization of regulatory region, and (5) higher sequencing costs. The selection of germplasm to be 'mined' is one of the utmost challenges face the scholars because of the huge genetic resources collections. To overcome the aforementioned challenges, we must (1) narrow down the core collection to a manageable size while maintaining the variability, (2) refine phenotyping protocols to increase the efficiency of allele mining, (3) exploit the developments in allele mining, association genetics and comparative genomics by combining expertise from several disciplines, including molecular genetics, statistics and bioinformatics, (4) develop cheaper and faster sequencing platforms for high through put detection of allelic variations (5) develop flexible computational tools to manage genetic resources, select desirable alleles, analyze the functional nucleotide diversity to predict specific nucleotide changes responsible for altered function, accurately predict the core promoter region based on the representation/over-representation of consensus regulatory motifs, and get the snapshot of the regulatory elements which can be further examined through suitable experiments.

8. Conclusion

Soybean oil is used in many foods, industrial and fuel products. Whereas soybean meal is incorporated into animal feed. The variation in the quality and quantity of these products is basically dependent on the genetic diversity of soybean germplasm. The genetic diversity in

soybean germplasm was evolved from the dispersion of the cultivated soybean domesticated by the Chinese farmers. Many factors are affecting the dispersion of soybean including regional adaptation and selection. Morphological, cellular, biochemical (proteins and isozymes) and molecular markers have been used on the wide scale for the study of the genetic diversity of the cultivated and wild relative of soybean. These analyses were carried out to meet wide rang of objectives from simply testing the usefulness of a particular marker system to identifying exotic germplasm accessions to expand the genetic diversity of the elite germplasm pool in order to permit genetic improvement for increased soybean yield. Exploitation of soybean germplasm for efficient utilization depends on the knowledge of genetic diversity, in general, and allelic diversity at candidate gene(s) of interest, in particular. The beneficial alleles from vast soybean genetic resources existing worldwide were derived from cultivated germplasm. However, a significant portion of these beneficial alleles were still resided in the wild soybean germplasm. Nowadays, considerable attention has focused on allele mining (gene polymorphisms) and their potential use to alter protein function in ways that might prove biologically important. But increasing numbers of polymorphisms are also being identified in the regulatory and non coden regions of genes. Therefore, allele mining is a promising approach to dissect naturally occurring allelic variation at candidate genes controlling key agronomic traits which has potential applications in crop improvement programs. Allele mining can be effectively used for discovery of superior alleles, through 'mining' the gene of interest from soybean germplasm. It can also provide insight into molecular basis of novel trait variations and identify the nucleotide sequence changes associated with superior alleles. In addition, the rate of evolution of alleles; allelic similarity/dissimilarity at a candidate gene and allelic synteny with other members of the family can also be studied. Allele mining may also pave way for molecular discrimination among related species within the genus *Glycine*, development of allele-specific molecular markers, facilitating introgression of novel alleles through Marker Assisted Selection or deployment through genetic engineering. The alleles mining approaches and the challenges associated with it are also discussed.

9. References

Abe, J., Xu D. H., Suzuki, Y., Kanazawa, A. & Shimamoto, Y. (2003). Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theor. Appl. Genet.,* 106, 445-453.

An, W., Zhao, H., Dong, Y., Wang, Y., Li, Q., Zhuang, B., Gong, L. & Liu, B. (2009). Genetic diversity in annual wild soybean (*Glycine Soja* SIEB. ET ZUCC.) and cultivated soybean (*G. Max*. MERR.). from different latitudes in China. *Pak. J. Bot.*, 41. 2229-2242.

Arumuganathan, I., & Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.*, 9, 208-219.

Baranek, M., Kadlec, M., Raddova, J., Vachun, M. & Pidra, M. (2002). Evaluation of genetic diversity in 19 *Glycine max* (L.) Merr. accessions included in the Czech national collection of soybean genotypes. Czech J. Genet. Plant Breed, 38 , 69–74.

Bennett, M. D., Heslop-Harrison, J. S., Smith, J. B. & Ward, J. P. (1983). DNA density in mitotic and meiotic metaphase chromosomes of plants and animals. *J. Cell Sci.*, 63, 173-179.

Bennett, M. D. & Leitch, I. J. (1995). Nuclear DNA amounts in angiosperms. *Ann. Bot.*, 76, 113-176.

Bennett, M. D., Leitch, I. J. & Hanson, L. (1998). DNA amounts in two samples of angiosperm weeds. *Ann. Bot.*, 82, 121-134.

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

Bonato, A. L. V., Calvo E. S., Geraldi, I. O., Arias, C. A. A. (2006). Genetic similarity among soybean (Glycine max (L). Merrill) cultivars released in Brazil using AFLP markers. *Gen Mol Biol.*, 29, 692-704.

Brown-Guedira, G. L, Thompson, J. A, Nelson, R. L, Warburton, M. L (2000). Evaluation of genetic diversity of soybean introductions using RAPD and SSR markers. *Crop Sci.*, 40, 815–823.

Burdon, J. J. & Marshall, D. R. (1981). Inter- and intra-specific diversity in the disease-response of Glycine species to the leaf-rust fungus. *Phakopsora pachyrhizi. Journal of Ecology*, 69, 381-390.

Bushehri, S. A. A., Abd-Mishani, C., Yazdi-Samadi, B., Sayed-Tabatabaei, B. E. (2000). Variety specific electrophoretic profiles of soybean cultivars. *Iranian J. Agric. Sci.*, 31, 55-61.

Camps, G., Vernetti Fde, J., Augustin, E. & Irigon, D. (1994). Morphological and electrophoretical characterization of twenty soybean cultivars. *Pesqui. Agropecu. Bras.*, 29: 1779-1787.

Chen, Y. & Nelson, R. L. (2005). Relationship between Origin and Genetic Diversity in Chinese Soybean Germplasm. *Crop Sci.*, 45, 1645–1652.

Chowdhurya, A.K, Srinivesb, P., Tongpamnakc, P., Saksoongd, P. & Chatwachirawong, P. (2002). Genetic Relationship among exotic soybean introductions in Thailand: consequence for varietal registration. *ScienceAsia*, 28, 227-239.

Chotiyarnwong, O., Chatwachirawong, P., Chanprame, S. & Srinives, P. (2007). Evaluation of genetic diversity in thai indigenous and recommended soybean varieties by SSR markers. *Thai Journal of Agricultural Science,* 40, 119-126.

Chung, J., Lee J. H., Arumuganathan K., Graef G. L., & Specht J. E. (1998). Relationship between nuclear DNA content and seed and leaf size inn soybean. *Theor. Appl. Genet.*, 96, 1064–1068.

Chung, M. G., Chung, M. Y., Johnson, C. & Palmer, R. G. (2006). Restriction fragment length polymorphisms in the USDA soybean germplasm from central China. *Botanical Studies*, 47, 13-21.

Comai, L. & Henikoff, S. (2006). TILLING: practical single-nucleotide mutation discovery. Plant J., 45,684–94."

Ding, Y. L., Zhao, T. J. & Gai, J. Y. (2008). Genetic diversity and ecological diversity of soybean cultivars from China, Japan, North America, and North American Ancestral Lines determined by amplified fragment length polymorphism. *Crop Sci,* 43, 1858–1867.

Diwan, N., Cregan, P. B. (1997). Automated sizing of fluorescent labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor Appl Genet.*, 95,723– 733.

Doerschug, E. B., Miksche, J. P., Palmer, R. G. (1978). DNA content, ribosomal-RNA gene number, and protein in soybeans. *Canadian Journal of Genetics and Cytology,* 20, 5331-5338.

Dong, Y. S., Zhuang, B. C., Zhao, L.M., Sun H. & He, M.Y. (2001). The genetic diversity of annual wild soybeans grown in China. *Theor Appl Genet.*, 103, 98–103.

Earthington, S. R., Nickell, C. D. & Gray, L. E. (1995). Inheritance of brown stem rot resistance in soybean cultivar BSR 101. *J. Hered*, 86:55-60.

Edwards, G. A. & Endrizzi, J. L. (1975). Cell size nuclear size and DNA content relationships in Gossypium. *Can. J. Genet. Cytol.*, 17, 181-186.

Fu, Y. B., Peterson, G. W., Morrison, M. J. (2007). Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. *Crop Science*, 47, 1947-1954.

Fujita, R., Ohara, M., Okazaki, K., Shimamoto, Y. (1997). The extent of natural cross-pollination wild soybean (*Glycine soja*). *J Hered.*, 88,124–128.

Fukuda, Y. (1933). Cytogenetical studies on the wild and cultivated Manchurian soybeans (*Glycine*L.). *Japanese Journal of Botany* 6, 489–506.

Ghafoor, A., Gulbaaz, F. N., Afzal, M., Ashraf, M., Arshad, M. (2003). Inter-relationship between SDS-PAGE markers and agronomic traits in Chickpea (*Cicer Arietinum*). *Pak. J. Bot.*, 35, 613-624.

Ghafoor, A., Zahoor, A., Qureshi, A. S., Bashir, M. (2002). Genetic relationship in *Vigna mungo* (L.). Hepper and *V. radiata* (L.) R. Wilczek based on morphological traits and SDS-PAGE. *Euphytica*, 123, 367-2002.

Golbitz, P. (2007). *Soya and Oilseed Bluebook*, Soyatech, Inc., Bar Harbor, ME., Graham, MJ.

Graham, M.J., Nickell, C.D.& Rayburn, A.L. (1994). Relationship between genome size and maturity group in soybean. *Theoretical andApplied Genetics*, 88, 429-432.

Graner, A. (2006). Barley research at IPK, In: *Molecular markers for allele mining*, de Vicente, M.C. & Glaszmann, J.C., pp. (25-25). IPGRI, Italy.

Guan, R., Chang, R., Li, Y., Wang, L., Liu, Z., Qiu, L. (2010). Genetic diversity comparison between Chinese and Japanese soybeans (Glycine max (L.) Merr.) revealed by nuclear SSRs. *Genet Resour Crop Evol.* 57, 229–242.

Guo J., Wang, Y., Song, C., Zhou, J., Qiu, L., Huang, H., & Wang, Y. (2010). A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Annals of Botany*, 106, 505-514.

Hammatt, N., Blackwell, N. W. & Davey, M. R. (1991). Variation in the DNA content of *Glycine* species. *Journal of Experimental Botany,* 42, 659-665.

Hirata, T., Abe, J. & Shimamoto, Y. (1999). Genetic structure of the Japanese soybean population. Genet. Resour. *Crop Evol.*, 46, 441–453.

http://www.ipgri.cgiar.org/system, 1998.

Hudcovicova, M. & Kraic, J. (2003). Utilisation of SSRs for characterisation of the soybean (*Glycine max* (L.). Merr.). genetic resources. *Czech J. Genet. Plant Breed.*, 39, 120–126.

Hwang, T.Y., Nakamoto, Y., Kono, I., Enoki, H., Funatsuki, H., Kitamura, K., Ishimoto, M. (2008). Genetic diversity of cultivated and wild soybeans including Japanese elite cultivars as revealed by length polymorphism of SSR markers. *Breeding Sci.*, 58,315-323.

Hymowitz, T. (1995). Evaluation of wild perennial *Glycine* species and crosses for resistance to *Phakopsora*, In: *Proceedings of the Soybean Rust Workshop*, J.B. Sinclair & G.L.Hartman, 33–37, Urbana, IL, USA: National Soybean Rsearch Laboratory.

Hymowitz, T. (1970). On the domestication of the soybeans. *Economic Botany* 23, 408–421.

Hymowitz, T. (2004). Speciation and Cytogenetics, In Soybeans: improvement, production, and uses, H. R. Boerma & J. E. Specht, 97-136, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, Wis.

Hymowitz, T., & N. Kaizuma. (1981). Soybean protein electrophoresis profiles from 15 Asian countries or regions: Hypotheses on paths plasm for soybean yield improvement. *Crop Sci.,* 38, 1362–1368.

Jha, S. S., Ohri, D. (1996). Phylogenetic relationships of *Cajanus cajan* (L.) Millsp. (pigeonpea). and its wild relatives based on seed protein profiles. *Genet. Resour. Crop Evol.*, 43, 275–281.

Kanazawa, A., Tozuka, A., & Akimoto, S. (1998). Phylogenetic relationships of the mitochondrial genomes in the genus *Glycine* subgenus soja. Genes Genet. Syst., 73, 255–261.

Karam, M. A., Sammour, R. H., Ahmed, M. F., Ashour, F. M. & El-Sadek, L.M. (1999). Relationships among some *Medicago* species as revealed by isozyme electrophoresis. J. *Union Arab Biol.*, 9 ,269-279.

Kaur, N., Street, K., Mackay, M., Yahiaoui, N., Keller, B. (2008). Molecular approaches for characterization and use of natural disease in wheat. . *Eur J Plant Patho.,* 121,387–97.

Kim, K-S., Bellendir S., Hudson, K. A., Hill, C. B., Hartman, G. L., Hyten, D. L., Hudson, M. E & Diers, B. W. (2010). Fine mapping the soybean aphid resistance gene *Rag*1 in soybean. *Theor. Appl. Genet.,* 120,1063 – 1071.

Klos, K. L. E., Paz, M. M., Fredrick, L., Marek, Cregan, P. B. & Shoemaker, R. C. (2000). Molecular markers useful for detecting resistance to brown stem rot in soybean. Crop Sci., 40,1445–1452.

Knight, C. A. & Ackerly, D. D. (2002).Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters*, 5, 66-76.

Kumar, G. R., Sakthivel, K., Sundaram, R.M., Neeraja, C.N., Balachandran, S.M., Shobha Rani, N., Viraktamath, B.C., Madhav, M.S. (2010). Allele mining in crops: Prospects and potentials. Biotechnology Advances, 28, 451–461.

Li, C., Fatokun, C. A., Ubi, B., Singh, B. B. & Scoles, G. J. (2001). Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. *Crop Sci.,* 4, 189-97.

Li, W., Han, Y., Zhang, D., Yang, M., Teng, W., Jiang, Z., Qiu, L., Sun, G. (2008). Genetic diversity in soybean genotypes from north-eastern China and identifi cation of

candidate markers associated with maturity rating. *Plant Breeding*, 127,494-500.

Li, Z. L & Nelson, R. L (2001). Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Sci.*, 41,1337-1347.

Li, Z. & Nelson, R. L. (2002). RAPD marker diversity among cultivated and wild soybean accessions from four Chinese provinces. *Crop Sci.*, 42, 1737-1744.

Lü, S.L. (1978). Discussion on the original region of cultivated soybean in China. *Scientia Agricultura Sinica*, 4, 90–94.

Malik, M.F.A., Qureshi, A.S., Ashraf, M. & Ghafoor, A. (2006). Genetic variability of the main yield related characters in soybean. *Inter. J. Agri. & Biol.*, 8, 815-619.

Malik, M. F. A., Ashraf, M., Qureshi, A. S., & Gjafoor, A. (2007). Assessment of genetic variability, correlation and path analyses for yield and its components in soybean. *Pak. J. Bot.*, 39, 405-413.

Malik, M.F.A., Qureshi, A.S., Ashraf, M., Khan, M.R., & Javed, A. (2009). Evaluation of genetic diversity in soybean (*Glycine max*). lines using seed protein. Electrophoresis. *Australian Journal of Crop Science*, 3,107-112.

Meksem, K., Pantazopoulos, P., Njiti, V.N., Hyten, L.D., Arelli, P.R. & Lightfoot, D.A. (2001). 'Forrest' resistance to the soybean cyst nematode is bigenic: saturation mapping of the *Rhg1* and *Rhg4* loci. *Theor Appl Genet.*, 103,710–717.

Mimura, M. (2001). *Genetic Diversity of Edamame Soybean [Glycine max (L.) Merr.]*, M.S. thesis, Washington State University.

Mimura, M., Coyne, C. J., Bambuck, M. W. & Lumpkin, T. A. (2007). SSR Diversity of vegetable soybean [*Glycine max* (L.) Merr.]. *Genet. Resour. Crop Evol.*, 54, 497-508.

Narvel, J. M., Fehr, W. R., Chu, W. C., Grant, D., Shoemaker, R. C. (2000). Simple sequence repeat diversity among soybean plant introductions and elite genotypes. *Crop Sci.*, 40,1452–1458.

Nelson, R. L. & Shoemaker, R. C. (2006). Identification and analysis of gene families from the duplicated genome of soybean using EST sequences. *BMC Genomics*, 7,204 -215.

Ngon, T. T., Van, K., Kim, M. Y., & Lee, S. H. (2006). Genetic variation in flowering time and maturity and its relationship among agronomic characters in soybean. *Korean Crop Science*, 51, 163-168.

Nguyen, H. T., Wu, X. L., Vuong, T. D, Sleper, D. A, Shannon, G. J. & Stacey, G. (2007). Genome maps, genetic diversity and marker-assisted selection for soybean improvement. *Molecular Plant Breeding*, 2007, 5, 196-198.

Pagel, J., Walling, J.G., Young, N.D., Shoemaker, R.C. & Jackson, S.A. (2004). Segmental duplications within the Glycine max genome revealed by fluorescence in situ hybridization of bacterial artificial chromosomes. *Genome*, 47, 764-768.

Pei, Y. L., Wang, L., Ge S., Wang, L. Z. (1996). Study on genetic diversity of *Glycine soja*-isozyme variation in four populations (China). *Soybean Sci.*, 15, 302-309.

Pham Thi Be Tu., Nguyen Thi Lang & Bui Chi Buu. (2003). Soybean genetic diversity analysis *Omonrice*, 11, 138-142.

Rayburn, A. L. (1990). Genome size variation in southwestern Indian maize adapted to various altitudes. *Evolutionary Trends in Plants,* 4, 53-57.

Rayburn, A. L., Birdar, D. P., Bullock, D. G., Nelson, R. L., Gourmet, C. & Wetzel J. B. (1997). Nuclear DNA content diversity in chinese soybean introductions *Annals of Botany,* 80, 321-325.

Rayburn, A. L., Biradar, D. P., Nelson, R. L., McCloskey, R., & Yeater, K. M. (2004). Documenting intraspecfic genome size variation in soybean. *Crop Sci.,* 44, 261–264.

Rajanna, M. P., Viswanatha, S. R., Kulkarni, R. S. & Ramesh, S. (2000). Correlation and path analysis in soybean (*Glycine max* L. Merrill).. *Crop Res. Hisar.,* 20, 244-2447.

Ristova, D., Šarcevic, H., Šimon, S., Mihajlov, L. & Pejic, I. (2010). Genetic diversity in southeast European soybean germplasm revealed by SSR markers. *Agriculturae Conspectus Scientificus,* 75 , 21-26.

Sacks, F. M., Lichtenstein, A., Van Horn, L., Harris, W., Kris-Etherton, P., Winston, M. (2006). "Soy protein, isoflavones, and cardiovascular health: an American Heart Association Science Advisory for professionals from the Nutrition Committee". *Circulation* 113, 1034–44.

Sammour, R. H. (1985). *Use of proteins characters in the taxonomy of peas and beans.* Thesis (Ph.D.), Ph D thesis, Tanta University, Tanta, Egypt.

Sammour, R. H. (1989). Electrophoresis of the seed proteins of *Vicia faba L.* and their immediate progenitors, -a reappraisal. *Plant Breeding,* 104,196-201.

Sammour, R. H. (1991). Using electrophoretic techniques in varietal identification, biosystematics analysis, phylogenetic relations. *Journal of Islamic Academy of Sciences*, 4, 221-226, 1991.

Sammour, R. H. (1992). An electrophoretic analysis of the variation in the seed storage proteins of *Vicia faba L.* Cultivars. *Fedds Repertorium,* 103, 555-557.

Sammour, R. H., Hamoud, M. A., Haidar, A. S. & Badr, A. (1993). Electrophoretic analysis of the seed proteins of some species in genus *Lotus. Feddes Repertorium,* 104, 251-257.

Sammour, R. H. (1999). SDS-PAGE analysis of the seed protein of some *Trifolium* taxa. *Plant Varieties and Seeds*, 12, 11-210.

Sammour, R. H. (2007). Electrophoretic and immunochemical techniques for the molecular reassessment of relationships within *Vicieae. Acta Agronomica Hungarica* 55, 131-147, 2007.

Sammour, R. H., Radwans, S. A & El-Koly, A. (2007). Genetic Variability in *Phaseolus* spp. as Revealed by SDS-PAGE Markers. *Seed Technology*, 29, 50-59.

Schulman, A. H. (2007). Molecular markers to assess genetic diversity. *Euphytica*, 158, 313–321.

Sharma, S. C. & Maloo, S. R. (2009). Genetic diversity and phylogenetic relationship among soybean [*Glycine max* (L.) Merrill] varieties based on protein, evolutionary and morphological markers. *Indian Journal of Plant Genetic Resources*, 22.

Shi, A., Chen, P., Zhang, B. & Hou, A. (2010). Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. *Plant Breeding*, 129, 250–256.

Sihag, R., Hooda, J. S., Vashishtha, R. D., Malik, B. P. S. (2004). Genetic divergence in soybean [*Glycine max* (L.). Merrill]. *Annals Biol,*. 20, 17-21.

Singh, G. (2010). *The soybean: botany, production and uses: The origin and history of soybean*, L. J. Qui & R. Z. Chang, 1-23, CAB International.

Seitova, A. M., Ignatov, A. N., Suprunova, T. P., Tsvetkov, I. L., Deineko, E. V., Dorokhov, D. B., Shumnyi, V. K., & Skryabin, K. G. (2004). Genetic Variation of Wild Soybean Glycine soja Sieb. et Zucc. in the Far East Region of the Russian Federation. *Russian Journal of Genetics*, 40, 165–171.

Tavaud-Pirra, M., Sartre, P., Nelson, R. L., Santoni, S., Texier, N., Roumet, P. (2009). Genetic diversity in a soybean collection. *Crop Sci.*, 49, 895-902.

Thompson, J. A., Nelson, R. L., & Vodkin, L. O. (1998). Identification of diverse soybeangermplasm using RAPD markers. *Crop Sci.*, 38, 1348-1355.

Tian, Q. Z, Gai, J. Y, Yu, D. Y, Jia, J. Z (2000). A study of amplified fragment length polymorphism (AFLP). in soybean (Chin.).. *Soybean Sci.*, 19, 210-217.

Till, B.J., Reynolds, S.H., Green, E.A., Codomo, C.A., Enns, L.C.& Johnson, J.E., Burtner, C., Odden, A. R., Young, K., Taylor, N. E., Henikoff, J. G., Comai, I. & Henikoff, S. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res., 13, 524–30.

Truong, N. T. , Gwag, J. G., Park, Y. J, & Lee, S. H. *(2005).* Genetic Diversity of Soybean Pod Shape Based on Elliptic Fourier Descriptors. *Korean J. Crop Sci., 50,* 60-66.

Ude, G. N., Kenworthy, W. J., Costa, J. M., Cregan, P. B., & Alvernaz, J. (2003). American Ancestral Lines Determined by Amplified Fragment Length Polymorphism. *Crop Sci.*, 43,1858–1867.

USDA, (2001). Soybean Germplasm Collection Descriptors.

Wang, J.L. (1947). Evaluation of soybean traits. Agricultural Journal, 5, 6–11.

Wang, K. J.& Takahata, Y. (2007). A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*). germplasm pools using SSR markers. Genet Resour Crop Evol., 54, 157-165.

Wang, L., Guan, R., Zhangxiong, L., Chang, R. & Qiu, L. (2006). Genetic Diversity of Chinese Cultivated Soybean Revealed by SSR Markers. Crop Sci, 46,1032-1038.

Wang, M., Li, R., Yang, W. & Du, W. (2010). Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. African Journal of Biotechnology, 9, 4857-4866.

Xie, H., Guan, R., Chang, R. & Qiu, L. (2005). Genetic diversity of Chinese summer soybean germplasm revealed by SSR markers. *Chinese Science Bulletin,* 50, 526-535.

Xu, D.H. & Gai, J.Y. (2003). Genetic diversity of wild and cultivated soybeans growing in China revealed by RAPD analysis. *Plant Breeding*, 122, 503-506.

Xu, L. & Zhao, G. F. (2002). Microsatellite DNA marker and its application in genetic diversity research. *Acta Botanica Boreali Occidentalia Sinica*, 22, 714-722.

Yeeh Y., Kang, S.S. & Chung, M.G. (1996). Evaluations of the natural monument 1 populations of *Camellia japonica* (Theaceae). in Korea based on allozyme studies. *Bot. Bull. Acad. Sin.*, 37, 141-146.

Yamamoto, K. & Nagato, Y. (1984). Variation of DNA content in the genus *Glycine. Japanese Journal of Breeding*, 34, 163-170.

Yoon, M., Lee, J., Kim, C., Kang, J., Cho, E., Baek, H. (2009). DNA profi ling and genetic diversity of Korean soybean (*Glycine max* (L.). Merrill). landraces by SSR markers. *Euphytica*, 165,69-77.

Young, H. (2008). *Plasticity of soybean* (*Glycine Max* (L.). Merrill) *root development under mild water deficits*. Thesis (M.S.), University of Missouri-Columbia.

Zhuang, B. C. (1999). *Biological Studies of Wild Soybeans in China*, Science Press, Beijing (in Chinese).
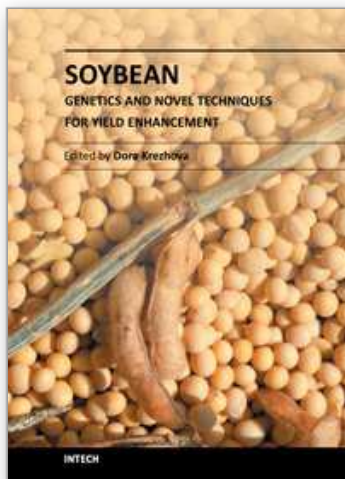
Soybean - Genetics and Novel Techniques for Yield Enhancement

Edited by Prof. Dora Krezhova

This book presents the importance of applying of novel genetics and breading technologies. The efficient genotype selections and gene transformations provide for generation of new and improved soybean cultivars, resistant to disease and environmental stresses. The book introduces also a few recent modern techniques and technologies for detection of plant stress and characterization of biomaterials as well as for processing of soybean food and oil products.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Reda Helmy Sammour (2011). Genetic Diversity and Allele Mining in Soybean Germplasm, Soybean - Genetics and Novel Techniques for Yield Enhancement, Prof. Dora Krezhova (Ed.), ISBN: 978-953-307-721-5, InTech, Available from: http://www.intechopen.com/books/soybean-genetics-and-novel-techniques-for-yield-enhancement/genetic-diversity-and-allele-mining-in-soybean-germplasm

INTECH
open science | open minds