A platform for soybean molecular breeding: the utilization of core collections for food security

Li-Juan Qiu · Li-Li Xing · Yong Guo ·
Jun Wang · Scott A. Jackson · Ru-Zhen Chang

Abstract Soybean is an important crop not only for human consumption but also for its addition of nitrogen to the soil during crop rotation. China has the largest collection of cultivated soybeans (*Glycine max*) and wild soybeans (*Glycine soja*) all over the world. The platform of soybean core, mini core and integrated applied core collections has been developed in the past decade based on systematic researches which included the sampling strategies, statistical methods, phenotypic data and SSR markers. Meanwhile, intergrated applied core collections including accessions with single or integrated favorite traits are being developed in order to meet the demand of soybean breeding. These kinds of core collections provide powerful materials for evaluation of germplasm, identification of trait-specific accessions, gene discovery, allele mining, genomic study, maker development, and molecular breeding. Some successful cases have proved the usefulness and efficiency of this platform. The platform is helpful for enhancing utilization of soybean genetic resources in sustainable crop improvement for food security. The efficient utilization of this platform in the future is relying on accurate phenotyping methods, abundant functional markers, high-throughput genotyping platforms, and effective breeding programs.

Keywords Plant genetic resources · Soybean · Utilization of core collection · Molecular breeding · Food security

L.-J. Qiu (✉) · L.-L. Xing · Y. Guo · J. Wang · R.-Z. Chang
The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: qiulijuan@caas.cn

S. A. Jackson
Center for Applied Genetic Technology, University of Georgia, Athens, GA 30602, USA

Introduction

Food security exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food (FAO 2003). The sufficiency of human food is dependent in large on the availability and nutrient sufficiency of the plants consumed directly, or indirectly through animals (Bruulsema et al. 2012). Of the 7,000 plant species used worldwide in food and agriculture, only 30 crops 'feed the world', which provide 95 % of global plant-derived energy-intake (calories) and proteins (Schmidt and Wei 2006). The soybean is a crop of global importance and is one of the most frequently cultivated crops worldwide. It is important for both protein meal and vegetable oil and is used for both human and animal consumption as well as for industrial purposes, such as biofuels (Hartman et al. 2011). In addition, the soybean also plays an important role in crop diversification and benefits other crops due to its addition of nitrogen to the soil during crop rotation (Singh 2010). Soybean seeds contain about 20 % oil and 40 % protein. Of the oil fraction, 95 % is consumed as edible oil with the rest used for industrial products from cosmetics and hygiene products to paint removers and plastics. In addition, about 98 % of the soybean meal is used in livestock and aquaculture feeds due to its high protein level (Liu 2008). Soybeans are unique among crop plants in that they supply protein equal in quality to that of animal sources but with less saturated fat and no cholesterol (Young 1991). For this reason, soybeans have long been consumed in Asia and greatly increased in popularity outside of Asia as a primary source of protein in such traditional foods as tofu, soymilk, tempeh, natto, sprouts, green vegetable soybeans, and many others (www.soyinfocenter.com).

Soybean was domesticated in China about 4,500 years ago, during the ancient Huangdi period (Qiu et al. 2011a).

Fig. 1 The development and utilization of soybean core collections
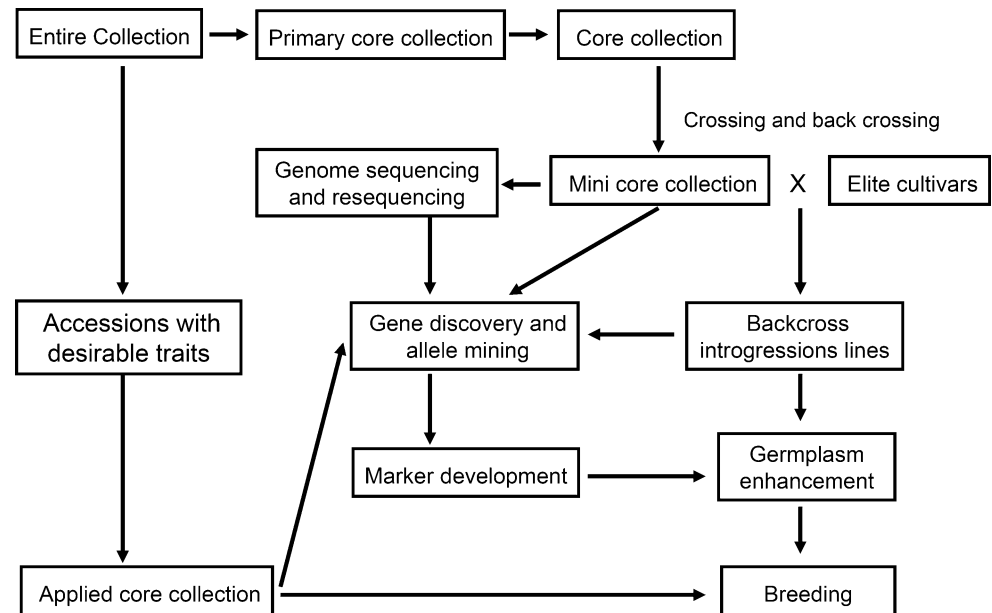
By the 16th century A.D. soybean was transported to Japan, Indonesia, Philippines and Vietnam, and later introduced to Europe and America (Hymowitz and Newell 1980). It is now cultivated in more than 60 countries across five continents. As the center of cultivated soybean (*Glycine max*), China has the most abundant genetic resources for soybean (>23,000 accessions) and its wild relatives, *Glycine soja* (>7,000 accessions). Chinese soybean germplasm is used widely throughout the world and these accessions are useful for small farmers that have to cope with heterogeneous microclimates and for advanced soybean improvement programs and to deal with needs in the future such as climate change. Although a large number of soybean accessions are now conserved either in various genebanks or in situ, <1 % of them have been used for breeding. In order to increase the efficiency of utilization of soybean genetic resources, different types of core collections have been developed from the whole collection of Chinese soybean germplasm, using a combination of passport data and morphological traits (Qiu et al. 2003; Song et al. 2010; Wang et al. 2006; Zhao et al. 2005). In this paper, the development and utilization of different core collections for Chinese soybean germplasm as a platform were reviewed. This platform is helpful for enhancing utilization of soybean genetic resources for sustainable crop improvement for food security (Fig. 1). The challenges and prospects of this platform for molecular breeding and food security were also discussed.

Development of soybean core collections

Plant germplasm is the lifeblood of plant breeding without which breeding is impossible to conduct. It is important that the diversity within major crops not only conserved but managed wisely and mined for useful traits. Although these abundant resources provide a rich genetic base for breeding, they also bring the difficulties for conservation and utilization. The approaches of core collection and mini core collection provided a rational approach to use a manageable number of accessions from large germplasm collections (Brown 1989; Frankel and Brown 1984; Upadhyaya and Oritz 2001). A core collection is a subset of accessions that represents the genetic diversity of a species and its relatives. Those samples not included in the core collections are not abandoned, but rather maintained as "reserved collection", the total number of resources referred here is over 3,000 accessions, take 10 % core collections represent 70 % of the whole collection (Frankel and Brown 1984). A mini core collection consists of 10 % accessions of the core collection, and hence only 1 % of the entire collection (Upadhyaya and Oritz 2001). The mini core collection still represents the diversity of the entire collection but the number is largely reduced compared with that of core collection. Due to the reduced size, core collection and mini core collection could be studied extensively and the information derived can be used to guide more efficient utilization of the much larger reserved collection. Up to now, more than 35 core collections and mini core collections have been developed for nearly 30 species, including rice, wheat, maize, soybean, chickpea, peanut, common bean, and so on (Balfourier et al. 2007; Blair et al. 2009; Hao et al. 2008; Holbrook and Dong 2005; Li et al. 2004; Upadhyaya et al. 2001; Upadhyaya and Oritz 2001; Wang et al. 2006; Zhang et al. 2011a).

In soybean, the first core collection was developed in 1987 based on the passport data but without any follow-up

reports (Brown et al. 1987). Since then, core collections of cultivated and wild soybean were developed based on different gene pools all over the world. A core collection of soybean accessions was used for analysis of the distribution of the major soybean seed allergens (Yaklich et al. 1999). The Chinese cultivated soybean core collection has been developed from 23,587 accessions conserved in the Chinese National Soybean GeneBank (CNSGB) (Wang et al. 2006) and 652 accessions were selected to represent the total 6,172 accessions of Chinese anneal wild soybean (Zhao et al. 2005). A core set of Korean landraces was developed from approximately 7,000 accessions conserved in the National Genebank of Rural Development Administration, Korea (Cho et al. 2008). A core collection (1,600 accessions) was also selected from 16,999 accessions in USDA Soybean Germplasm Collection (Oliveira et al. 2010). Recently, based on passport data records, a Japanese landrace subset consisting of 832 accessions was proportionally selected from 3,994 accessions of the National Institute of Agrobiological Sciences Genebank (Kaga et al. 2012). Among all these core collections of soybean, Chinese soybean core collection (2,794 accessions of cultivated soybeans and 652 accessions of wild soybeans) was the largest and systemically studied one.

The strategy for development of soybean core collections

The strategy for development of soybean core collections included the following aspects. First, the samples which were mainly domestic soybean accessions supplemented with exotic germplasm to broaden the diversity within the core. Second, data acquisition was based on passport data, taking into consideration of agronomic traits and molecular markers. Third, the selection was based on stratification and statistics as well as experienced artificial selection. The representativeness was tested on primary core collection (Cui et al. 2003, 2004), core collection (Wang et al. 2006) and mini core collection (Wang et al. 2008) levels with different criteria (Qiu et al. 2009). Therefore, the soybean core and mini core collections were developed through selection, test, supplement, retest, and systematic representativeness analyses. In addition, single and integrated applied core collections were also developed with desirable traits based on breeding demand. The characteristic of all this collections were broaden representative, high diversity but meet different demand of study and breeding.

From entire germplasm collection to core and mini core collections

More than 23,000 cultivated soybean accessions and 7,000 wild soybean accessions are conserved in the CNSGB now.

These accessions include local landraces and wild soybeans collected in China and overseas, improved varieties and breeding lines developed by Chinese soybean breeders or introduced from overseas agricultural research institutes. In order to detect the redundancy of these accessions, some accessions including six accessions with the same name "Man cang jin" were characterized by agronomic traits and SSR markers. The results showed that all of these accessions had obvious difference at both agronomic and molecular level, indicating the extremely low redundancy of accessions stored in CNSGB (Yan et al. 2003). In order to select proper core collection of cultivated soybean for further molecular evaluation, 20 sampling methods were compared using passport data and agronomic traits, and the optimal strategy and sample size were selected (Qiu et al. 2003). The samples were analyzed with core set of SSR markers that were selected and tested using small amount of samples (Wang et al. 2003; Xie et al. 2003). By integration of the SSR marker data and agronomic traits, a candidate core collection of cultivated soybean was developed. Then the selected candidate core collection was supplemented with additional 130 accessions with various agronomic traits and 494 accessions with disease resistances and stress tolerance (Wang et al. 2003, 2006). The core collection of Chinese cultivated soybean consisting of 2,794 accessions was developed, which accounted for 11.8 % representative of the entire collection and 73.6 % of genetic diversity separately (Qiu et al. 2009). At the same time, core collection of annual Chinese wild soybean was also developed by comparing five sampling strategies with genetic representation of 83.6 % for the entire collection (Zhao et al. 2005).

In order to easy evaluation of soybean core accessions for replication at different locations, mini core collections of soybean were also developed due to the large size of entire collection of soybean. The mini core collection of cultivated soybean consisting of 248 accessions was developed based on the primary core collection and it represented 94.5 % of the phenotypic diversity and 63.5 % of the genetic diversity respectively (Qiu et al. 2009; Wang et al. 2006). SSR marker analysis also showed that this mini core collection had high genetic diversity (Song et al. 2010). Furthermore, a set of mini core collection of 96 wild soybean accessions were also evaluated for the extent of genetic diversity (Liu et al. 2009).

From trait-specific core collection to integrated applied core collection

The trait-specific accessions in core collection and mini core collection were scare due to the need of including as many accessions with various traits under the premise of the appropriate sample size as possible. Therefore, the elite germplasm identified in the mini core collection could only

be used as an indicator for directional identification of entire collection. Since soybean is a short-day plant and it adapts to a narrow geographic range, accessions adapted certain geography would have difficulties in adapting to varied eco-regions in breeding, especially for yield-related traits. Therefore, selecting more accessions with the same desirable traits for evaluation will be valuable for soybean improvement. The concept of applied core collection concerned representative germplasm with targeted desirable traits and high level of genetic diversity. Applied core collection of soybean could fulfill the needs of accessions adapted different eco-regions and broadening the genetic base of breeding programs. Up to know, different desirable traits, including resistance to soybean cyst nematode (SCN) (Duan et al. 2008; Ma et al. 2006), resistance to soybean mosaic virus (SMV) (Mi et al. 2004), high phosphorus efficiency (Zhao et al. 2004) etc., have been used to develop applied core collections of soybean. These core collections all encompassed more than 70 % of the genetic variation present in the overall collection with targeted traits.

However, the sample size would increase suddenly if developing one applied core collection for each specific desirable trait, which would be inconvenient for research and utilization and go against the principle of core collection. Therefore, soybean integrated applied core collection consisting of accessions with cold tolerance, drought tolerance, salt tolerance, resistance to SCN, resistance to SMV, high protein content, and high fat content was developed (Guo et al. 2013). These accessions showed a high level of diversity in terms of target traits, non-target traits and molecular markers. Meanwhile, the integrated applied core collection of soybean not only had increased the number of soybean accessions with specific desirable traits but also had similar diversity as mini core collection. Therefore, the concept of integrated applied core collection resolves the conflict between reducing sample size and concentrating genetic diversity and solves the problem of establishing many core collections for different traits. Accessions with integrated excellent traits in breeding would be conducive in improving varieties and shortening the time for application in breeding.

Utilization of core collections: from phenotypes to genes/markers and breeding

Due to the proper sample size, high diversity, and broad representative, core collections are ideal resources for phenotypic, genetic, genomic researches and breeding program. After development of different kinds of core collection, the utilization of them have become more and more important. As a model crop, rice research is always ahead of other crops including wheat, maize, soybean, and so on. Many important genes including *GS3*, *PROG1*, *IPA1*, *DTH8*, and others, were cloned successfully using introgression lines (ILs) or near-isogenic lines (NILs) based on germplasm collections of rice (Fan et al. 2006; Jiao et al. 2010; Jin et al. 2008; Tan et al. 2008; Wei et al. 2010). Using 21 elite varieties of rice as recurrent parents, and 188 donor parents originating from 24 countries or regions, approximately 60,000 ILs and over 17,000 NILs based on the mini core collection of rice have been developed systemically (Li et al. 2011; Long et al. 2009). Similarly, soybean core collections have been played and will play an important role in basic research of soybean biology and soybean breeding for food security.

Germplasm evaluation and identification

The core and mini core collections of the germplasm were evaluated at diverse locations to identify trait-specific diverse accessions. Due to the reduced size, the core and the mini core collections have been characterized more precisely and some very useful trait specific accessions have been identified (Upadhyaya et al. 2008). These accessions are similar to or better than the control cultivar(s) for a particular trait, agronomically good and genetically diverse. Some plant germplasm with desirable agronomic and nutritional traits had been identified using core collection and mini core collection, such as chickpea with drought-avoidance root traits (Kashiwagi et al. 2005), pigeonpea with drought resistance (Upadhyaya et al. 2007), pigeonpea with resistance to multiple disease (Sharma et al. 2012), peanut with high-quality (Chu et al. 2007) and wheat with high-molecular-weight glutenin subunits (Zhang et al. 2002). Soybean core collection and mini core collections also played an important role in the identification of desirable agronomic and nutritional traits. Although only part of the core collections was used in traits identification, variation was observed and extensively existed for both yield and quality related traits by testing on one site or multi-sites (Table 1). For example, there was threefold variation in the amount of allergic protein 28 K (Zhang et al. 2006) to over 12-fold for protein subunit (Wang et al. 2008). Except seed quality and yield, the variations were also observed for identifying the disease resistance. Using four SMV strains to test 99 accessions in soybean core collections from Huanghuai and South part of China, accessions with all six resistant levels were identified (Chen et al. 2009). One accession was resistant to the four strains, three accessions were resistant to three strains except SC-7, and six accessions with ability to keep disease lesion from spreading, all of which are valuable in breeding (Chen et al. 2009). The identification of these trait-specific diverse accessions also proved that core collections were

representative of the larger collection and adequately captured useful genetic variation.

Gene discovery and allele mining

Due to abundant in diversity and practical in scale, core collections and mini core collections could be used for gene discovery and allele mining (Table 2). One way for gene discovery using core collection is association analysis of core collections and their derived lines. With the development of new sequencing technology and availability of core collections, it is now feasible to establish an open-source translational research platform for genome-wide association studies (GWAS). GWAS have contributed tremendously to discovering genes controlling natural variations of complex traits in rice (Huang et al. 2010,

2012). Using SSR and SNP markers to genotype the core collections of soybean, linkage disequilibrium and association for plant height, maturity, 100-seed weight, protein content and oil content were also carried out. Twenty-one SSR markers on 16 chromosomes were found significantly ($P < 0.01$) associated with these traits (Li et al. 2010). Genome-wide association analysis of 191 accessions based on the unified mixed model also identified 19 SNPs and five haplotypes associated with soybean yield and yield components in three or more environments (Hao et al. 2012).

The other powerful tool to explore target genes controlling important traits is applying ILs and NILs developed from core collections. In a random backcross population, the allelic and genotypic frequencies at a given locus are known. A significant deviation of donor allele

Table 1 Summary of desirable traits evaluation with core or mini core collections in soybean

Traits	Germplasm		Evaluation			References
	Collection	Number	Unit	Average	Range	
Quality						
28 K allergen deficient	Core	220	%	33	21.1–61.5	Zhang et al. (2006)
11S/7S	Mini core	205	Ratio	1. 56	0. 55–4. 95	Wang et al. (2008)
Isoflavone	Mini core	100	mg/g	3.68	0.62–7.66	Wang et al. (2011a)
Isoflavone	Core	300	mg/g	2.39	0.18–6.07	Yuan et al. (2011)
α′ subunit	Mini core	236	%	8.7	0.2–16.7	Jian et al. (2012)
α subunit	Mini core	236	%	5.4	0.5–13.1	Jian et al. (2012)
β subunit	Mini core	236	%	15.0	5.8–29.3	Jian et al. (2012)
A3 subunit	Mini core	236	%	7.8	3.0–15.3	Jian et al. (2012)
A1 A2 A4 subunit	Mini core	236	%	22.8	13.4–32.0	Jian et al. (2012)
B subunit	Mini core	236	%	40.4	20.5–52.6	Jian et al. (2012)
11S/7S	Mini core	236	Ratio	2.94	1.228–13.745	Jian et al. (2012)
Yield						
Plant height	Mini core	60	cm	–	65.9–154.83	Liu et al. (2011)
Branches	Mini core	60	No.	–	0.85–5.16	Liu et al. (2011)
100-seed weight	Mini core	60	g	–	8.96–25.65	Liu et al. (2011)
Seed weight	Mini core	60	kg/hm	–	1,647.35–3,070.07	Liu et al. (2011)
Plant height	Core	195	cm	89.9	24.0–263.4	Hu et al. (2009)
Effective branches	Core	195	No.	4.2	0–12.0	Hu et al. (2009)
100-seed weight	Core	195	g	13.4	4.3–25.3	Hu et al. (2009)
Growth duration	Core	195	day	94.8	71–125	Hu et al. (2009)
Nodes on main stem	Core	195	No.	19.1	9.5–29.4	Hu et al. (2009)
Effective pods per plant	Core	195	No.	110.4	22.8–333.0	Hu et al. (2009)
Total pods per plant	Core	195	No.	119.2	25.3–338	Hu et al. (2009)
Pods per node	Core	195	No.	6.3	1.5–15.1	Hu et al. (2009)
Podding height	Core	195	cm	9.3	0–28.2	Hu et al. (2009)
Seed length	Mini core	235	mm	7.5	4.8–10.5	Wang et al. (2011b)
Seed width	Mini core	235	mm	5.9	3.4–8.5	Wang et al. (2011b)
Seed thickness	Mini core	235	mm	4.4	1.9–7.2	Wang et al. (2011b)
100-seed weight	Mini core	235	g	15.1	4.84–34.21	Wang et al. (2011b)

frequency at a single locus in ILs from the expected Mendelian level implied a positive selection favoring the donor allele and inferred a possible putative QTL locus. In soybean, ILs and NILs also becoming important to isolate genes controlling important traits such as photoperiodic flowering, leaflet shape and number of seeds per pod (Jeong et al. 2012; Xia et al. 2012). QTLs related to cold and drought stress had also been identified by using backcross introgression lines developed from mini core collection (Jiang et al. 2009; Qiu et al. 2011b).

Plant accessions from wild or locally adapted landrace genepools conserved in genebanks contain a rich repertoire of alleles that have been left behind by the selective processes of domestication, selection and cross-breeding that paved the way to today's elite cultivars (Kilian and Graner 2012). Most of accession in core collection and mini core collection of cultivated and wild soybean are either wild or landrace. The mini core collection or the reference sets (subset of mini core) could constitute the starting materials for dissecting the naturally occurring allelic variation at candidate genes through allele mining efforts (Kumar et al. 2010). Mini core collection of soybean was used to identify allele distribution of *GmTFL1* gene, which regulated the determinate growth habit in soybean. The finding that four *Gmtfl1* alleles were observed among cultivated soybeans whereas wild soybeans only contained one *GmTfl1* allele reflected the effects of genetic bottlenecks created by soybean germplasm introduction and modern breeding (Tian et al. 2010). Three low frequency new alleles of *GmF3′H* and *GmF3′5′H* were also identified using partial mini core collection, which indicated that the materials used for genotyping had abundant genetic variation (Guo and Qiu 2013).

Genomic study and marker development

With the advent of reference genome sequences for many crops, genomic approaches can be used to find specific genes and alleles and to clarify the origin, differentiation, domestication, and genome evolution of crops. Completion of the soybean genome sequence opens a new and exciting chapter for soybean biology study (Schmutz et al. 2010). A total of 17 wild and 14 cultivated soybean genomes were re-sequenced to an average of approximately ×5 depth and >90 % coverage. A set of 205,614 tag SNPs that may be useful for QTL mapping, association studies and marker development were identified (Lam et al. 2010). A subset of 25 accessions from the mini core collection were also used for resequencing and about 25.5 % SNPs were novel SNPs when compared them with the resequencing data of 31 accessions because the former set of accessions has more genetic diversity than the later set of acceesions (Li et al. 2013). Furthermore, re-sequencing of wild soybean *G. soja* and a subsequent comparative genomic analysis with reference to the *G. max* genome also identified a large degree of nucleotide and structural variation between the wild and domesticated soybean (Kim et al. 2010). The variation identified by resequencing of soybean germplasm could be used for marker development.

Molecular breeding and germplasm enhancement

With advances in identification of QTLs/genes responsible for important agronomic traits, molecular breeding is gradually becoming an actuality. Molecular markers developed from functional genes were used to introgression soybean cultivars for improving important agronomic

Table 2 Marker-trait association by directly and indirectly using core collections in soybean

Name of sample	No. of sample	Trait	No. of QTLs	References
Core collections				
Partial mini core collection	189	Yield and yield component	19	Hao et al. (2012)
Partial mini core collection and intergated applied core colletion	159	High oil content	6	Li et al. (2010)
		High protein content	1	
		Drought tolerance	5	
		SCN resistance	6	
		SMV resistance	3	
Backcrossing lines				
Backcrossing lines of Clark with Hongfeng 11	46	Low temperature resistant	13	Jiang et al. (2009)
Backcrossing lines of Clark with Hongfeng 11	72	Salt tolerance	3	Li et al. (2009)
		Drought tolerance	5	
		Agronomic	17	
Backcrossing lines of Clark with Hongfeng 11	23	Salt tolerance germination stage	22	Qiu et al. (2011b)
		Low temperature tolerance	15	

traits, such as yield, quality, disease resistance and tolerance abiotic stresses. Molecular breeding has been used in breeding of SCN-resistance cultivar (Zhonghuang 57), cultivar with *Lox2*-deficient (Zhonghuang 18), cultivar with trypsin inhibitor-deficient (Zhongdou 28), and cultivar with double deficient of *Lox2* and *Lox3* (Wuxing No.1) .

Most of local varieties/landraces in soybean core collections are no longer used for production and in breeding. In order to better exploit the important agronomic traits of these germplasm in soybean breeding, backcross introgression lines were made by cross and backcross of advanced varieties with mini core collections, which could broaden the genetic base of soybean and conduct germplasm enhancement for improvement. Since most of landraces usually have undesirable traits such as low yield, late maturity and easy lodging, they are difficulty to be used for genetic improvement via traditional breeding. Backcrossing can shorten the time to achieve improved varieties by taking superior yielding varieties and introgressing new traits. The contribution analysis of the one accessions to its derivatives showed that more than 77 % of the varieties derived from soybean accession Tokachi-Nagaha had <25 % genetic base of this accession (Guo et al. 2007). This indicated that multiple crossing or backcrossed could eliminate undesirable traits and have good adaptability. Recent studies showed that backcrossing twice using modern varieties as receptor crossing with accessions from mini core collection, most of the undesirable traits could be improved remarkably (Yan et al. 2012; Zhang et al. 2011b). It was found that BC_2F_2 progenies had higher protein content that was positively correlated to pods per plant, kernels per plant and 100-seed weight. Especially, the yield potential of lines developed by backcrossing the mini core collections appeared to be promising. This strategy of breeding, based on the backcrossing of modern cultivar with accessions from core collections, is significant in sustainable maintaining and developing useful variation and providing new genes to improve soybean cultivars continuously.

Challenges and prospects

Considering the huge number of accessions that are held collectively in various gene banks, genetic resources collections are deemed to harbor a wealth of undisclosed allelic variants. The rich genetic diversity of plant genetic resources is a prerequisite for germplasm improvement for food security. The systematic study of soybean core collections also provides a platform for soybean molecular breeding in the future. Now the challenge is how to efficiently identify and exploit the useful genes, alleles, molecular markers and variation for cultivar improvement.

Accurate phenotyping methods

Phenotyping would be the key for success as phenotypes are considered as the best clues for genotypes. Increase in precision of phenotyping techniques maximizes the chances of 'mining' the prospective genotypes. Also, accurate evaluation of phenotypes of soybean core collations in different regions with duplication will accelerate the course of gene discovery and marker development. Hence, there is a need to refine phenotyping protocols and equipments to increase the efficiency of phenotyping.

Abundant functional genes and markers

Identification of functional genes and markers is the base of molecular breeding. More and more functional genes will be isolated and functional markers will be developed by using the soybean germplasm resources, which also rely on nucleotide and/or structural variation on whole genomic level. Moreover, wild relatives of crops are a reservoir of genetic variation that can be utilized to increase yield and response different stress. Therefore, it is necessary to integrate genomes of a set of representative individuals (pan-genome) by de novo sequencing wild/cultivated soybeans and comparison of the reference genome.

High-throughput genotyping platforms

As technology advances and the cost drops, next-generation sequencing technologies revolutionized our ways on germplasm characterization. Accessions from core collections will be able to be genotyped with large number of markers or even sequenced at whole genome level in the foreseeable future. High-throughput genotyping platforms will be used for effective genome-wide gene-trait association studies to identify genes useful for genetic improving crop productivity.

Effective breeding programs

Molecular markers of single gene/locus could be used for molecular breeding in soybean. The advances in soybean research and the enrichment of large-scale marker data sets provide us tools to determine the genetic basis of all important agronomic traits. Based on genes uncovered for these traits, breeders can design and combine all the most favorable alleles into one variety, and then use molecular markers closely linked to these loci to develop the desired varieties. Meanwhile, breeding genetically modified soybean using favorable genes identified from soybean germplasm resources will also be another choise for molecular breeding. The molecular design breeding will

play an important role in soybean breeding for food security.

References

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265–1275

Blair MW, Diaz LM, Buendia HF, Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). Theor Appl Genet 119:955–972

Brown AHD (1989) Core collections: a practical approach to genetic resources management. Genome 31:818–824

Brown AHD, Grace JP, Speer SS (1987) Designation of a "core" collection of perennial *Glycine*. Soyb Genet Newsl 14:59–70

Bruulsema T, Heffer P, Welch R, Cakmak I, Moran K (2012) Fertilizing crops to improve human health: a scientific review. IPNI/IFA, Norcross, GA

Chen S-Y, Zheng G-J, Yang Z-L, Liu R-M, Zhi H-J (2009) Evaluation of resistance to SMV of soybean core collection from Southern China. Chin J Oil Crop Sci 31:513–516

Cho GT, Yoon MS, Lee J, Baek HJ, Kang JH, Kim TS, Paek NC (2008) Development of a core set of Korean soybean landraces [*Glycine max* (L.) Merr.]. J Crop Sci Biotech 11:157–162

Chu Y, Ramos L, Holbrook CC, Ozias-Akins P (2007) Frequency of a loss-of-function mutation in oleoyl-PC desaturase (ahFAD2A) in the mini-core of the US peanut germplasm collection. Crop Sci 47:2372–2378

Cui Y-H, Qiu L-J, Chang R-Z, Lv W-H (2003) Examination of representativeness of the primary core collection in Huanghuai summer sowing soybean (*Glycine max*) using SSR. J Plant Genet Resour 4:9–15

Cui Y-H, Qiu L-J, Chang R-Z, Lv W-H (2004) Representative test for primary core collection of summer sowing soybeans in Huanghuai region of China. Acta Agron Sin 1:284–288

Duan Y-X, Zhou B, Chen L-J, Wu H-Y (2008) Representative analysis of the establishment of a core collection focused on race 3 of soybean cyst nematode. Soybean Sci 27:366–372

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet 112:1164–1171

FAO (2003) Trade reforms and food security: conceptualizing the linkages. Food and Agriculture Organization of the United Nations, Rome

Frankel OH, Brown AHD (1984) Current plant genetic resources—a critical appraisal. Genetics: New Frontiers Vol. IV. Oxford and IBH Publishing, New Delhi

Guo Y, Qiu L-J (2013) Allele-specific marker development and selection efficiencies for both flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase genes in soybean subgenus *soja*. Theor Appl Genet. doi:10.1007/s00122-013-2063-3

Guo J–J, Chang R-Z, Zhang J-X, Zhang J-S, Guan R-X, Qiu L-J (2007) Contribution analysis of Japanese soybean germplasm Tokachi-Nagaha to Chinese soybean cultivars. Soybean Sci 26:807–819

Guo Y, Li Y-H, Hong H-L, Qiu L-J (2013) Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). Crop J (accepted)

Hao C-Y, Dong Y-C, Wang L-F, You G-X, Zhang H-N, Gai H-M, Jia J-Z, Zhang X-Y (2008) Establishment and genetic analysis of a wheat core collection in China. Chin Sci Bull 53:908–915

Hao D, Cheng H, Yin Z, Cui S, Zhang D, Wang H, Yu D (2012) Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor Appl Genet 124:447–458

Hartman GL, West ED, Herman TK (2011) Crops that feed the world 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. Food Secur 3:5–17

Holbrook C, Dong W (2005) Development and evaluation of a mini core collection for the US peanut germplasm collection. Crop Sci 45:1540–1544

Hu G-Y, Zhang L, Zhang L-Y, Huang Z-P, Li J-K, Hu C (2009) Agronomic characteristics and adaptability of Chinese soybean (*Glycine max*) core collection in Huanghuai summer sowing ecological region. J Plant Genet Resour 10:448–452

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44:32–39

Hymowitz T, Newell CA (1980) Toxonomy, speciation, domestication, dissemination, germplasm resources, and variation in the genus *Glycine*. In: Summerfeld RJ, Bunting AH (eds) Advances in legume science. Royal Botanic Gardens, Kew, pp 251–264

Jeong N, Suh SJ, Kim MH, Lee S, Moon JK, Kim HS, Jeong SC (2012) *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. Plant Cell 24:4807–4818

Jian S, Lu W-G, Wen Z-X, Li H-C, Yuan D-H, Li J-Y, ZHang H, Ye Y-Z (2012) Analysis of protein subunit variation in the mini core collection of cultivated soybean. J Henan Agric Sci 41:42–45

Jiang H-W, Li C-D, Liu C-Y, Zhang W-B, Qiu P-C, Li W-F, Gao Y-L, Hu G-H, Chen Q-S (2009) Genotype analysis and QTL mapping for tolerance to low temperature in germination by introgression lines in soybean. Acta Agron Sin 35:1268–1273

Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. Nat Genet 42:541–544

Jin J, Huang W, Gao JP, Yang J, Shi M, Zhu MZ, Luo D, Lin HX (2008) Genetic control of rice plant architecture under domestication. Nat Genet 40:1365–1369

Kaga A, Shimizu T, Watanabe S, Tsubokura Y, Katayose Y, Harada K, Vaughan DA, Tomooka N (2012) Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. Breed Sci 61:566–592

Kashiwagi J, Krishnamurthy L, Upadhyaya HD, Krishna H, Chandra S, Vadez V, Serraj R (2005) Genetic variability of drought-avoidance root traits in the mini-core germplasm collection of chickpea (Cicer arietinum L.). Euphytica 146:213–222

Kilian B, Graner A (2012) NGS technologies for analyzing germplasm diversity in genebanks. Brief Funct Genomics 11:38–50

Kim MY, Lee S, Van K, Kim TH, Jeong SC, Choi IY, Kim DS, Lee YS, Park D, Ma J, Kim WY, Kim BC, Park S, Lee KA, Kim DH, Kim KH, Shin JH, Jang YE, Kim KD, Liu WX, Chaisan T, Kang YJ, Lee YH, Moon JK, Schmutz J, Jackson SA, Bhak J, Lee SH (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proc Natl Acad Sci USA 107:22032–22037

Kumar GR, Sakthivel K, Sundaram RM, Neeraja CN, Balachandran SM, Rani NS, Viraktamath BC, Madhav MS (2010) Allele mining in crops: prospects and potentials. Biotechnol Adv 28:451–461

Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B, Li J, Jian M, Wang J, Shao G, Sun SS, Zhang G (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059

Li Y, Shi Y, Cao Y, Wang T (2004) Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. Genet Resour Crop Evol 51:845–852

Li C-D, Jiang H-W, Liu C-Y, Qiu P-C, Zhang W-B, Li W-F, Gao Y-L, Chen Q-S, Hu G-H (2009) Genotype and QTL analysis of drought tolerance loci for directional population in soybean. Chin J Oil Crop Sci 31:285–292

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (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 J, Zhang H, Wang D, Tang B, Chen C, Zhang D, Zhang M, Duan J, Xiong H, Li Z (2011) Rice omics and biotechnology in China. Plant Omics 4:302–317

Li Y-H, Zhao S-C, Ma J-X, Li D, Yan L, Li J, Qi X-T, Guo X-S, Zhang L, He W-M, Chang R-Z, Liang Q-S, Guo Y, Ye C, Wang X-B, Tao Y, Guan R-X, Wang J-Y, Liu Y-L, Jin L-G, Zhang X-Q, Liu Z-X, Zhang L-J, Chen J, Wang K-J, Nielsen R, Li R-Q, Chen P-Y, Li W-B, Reif JC, Purugganan M, Wang J, Zhang M-C, Wang J, Qiu L-J (2013) Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing. BMC Genomics (accepted)

Liu K (2008) Food use of whole soybeans. In: Johnson L, White PJ, Galloway R (eds) Soybeans: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL, pp 441–482

Liu Y, Li X, Wang K (2009) Analysis of the genetic variability for the mini-core collection of Chinese wild soybean (Glycine soja) collection in the National Gene Bank based on SSR markers. J Plant Genet Resour 10:211–217

Liu Z-X, Yang C-Y, Rea Xu (2011) Analysis of adaptability of soybean mini core collection in huang-huai rivers region in China. Acta Agron Sin 37:443–451

Long P, Yang H, Yu S-B, Mei H-W, Li Z-K, Luo L-J (2009) Development and conservation of introgression lines in rice. J Plant Genet Resour 10:51–54

Ma Y-S, Wang W-H, Wang L-X, Ma F-M, Wang P-W, Chang R-Z, Qiu L-J (2006) Genetic diversity of soybean and the

establishment of a core collection focused on resistance to soybean cyst nematode. J Integr Plant Biol 48:722–731

Mi S-J, Qiu L-J, Chang R-Z, Hao Z-B, Guan R-X (2004) Genetic diversity analysis of varieties of Glycine max L. Merr resistant to soybean mosaic virus by SSR fingerprints. Acta Phytopathol Sin 34:244–253

Oliveira MF, Nelson RL, Geraldi IO, Cruz CD, de Toledo JFF (2010) Establishing a soybean germplasm core collection. Field Crops Res 119:277–289

Qiu L-J, Cao Y-S, Chang R-Z, Zhou X-A, Wang G-X, Sun J-Y, Xie H, Zhang B, Li X-H, Xu Z-Y, Liu L-F (2003) Establishment of Chinese soybean (G. max) core collection: sampling strategy. Sci Agri Sin 36:1442–1449

Qiu L-J, Li Y-H, Guan R-X, Liu Z-X, Wang L-X, Chang R-Z (2009) Establishment representative testing and research progress of soybean core collection and mini core collection. Acta Agron Sin 35:571–579

Qiu L-J, Chen P-Y, Liu Z-X, Li Y-H, Guan R-X, Wang L-H, Chang R-Z (2011a) The worldwide utilization of the Chinese soybean germplasm collection. Plant Genet Resour Charact 9:109–122

Qiu P-C, Zhang W-B, Li C-D, Jiang H-W, Liu C-Y, Fan D-M, Zeng Q-L, Hu G-H, Chen Q-S (2011b) Genetic overlap of drought-tolerance loci between germination stage and seedling stage analyzed using introgression lines in soybean. Acta Agron Sin 37:477–483

Schmidt MR, Wei W (2006) Loss of agro-biodiversity, uncertainty, and perceived control: a comparative risk perception study in Austria and China. Risk Anal 26:455–470

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Sharma M, Rathore A, Mangala UN, Ghosh R, Sharma S, Upadhyay HD, Pande S (2012) New sources of resistance to fusarium wilt and sterility mosaic disease in a mini-core collection of pigeonpea germplasm. Eur J Plant Pathol 133:707–714

Singh G (2010) The soybean: botany, production and uses. CABI Publishing, Wallingford

Song X-E, Li Y-H, Chang R-Z, Guo P-Y, Qiu L-J (2010) Population structure and genetic diversity of mini core collection of cultivated soybean (Glycine max (L.) Merr.) in China. Sci Agric Sin 43:2209–2219

Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang X, Xie D, Sun C (2008) Control of a key transition from prostrate to erect growth in rice domestication. Nat Genet 40:1360–1364

Tian Z, Wang X, Lee R, Li Y, Specht JE, Nelson RL, McClean PE, Qiu L, Ma J (2010) Artificial selection for determinate growth habit in soybean. Proc Natl Acad Sci USA 107:8563–8568

Upadhyaya HD, Oritz R (2001) A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. Theor Appl Genet 102:1292–1298

Upadhyaya HD, Bramel PJ, Singh S (2001) Development of a chickpea core subset using geographic distribution and quantitative traits. Crop Sci 41:206–210

Upadhyaya HD, Reddy KN, Gowda CLL, Singh S (2007) Phenotypic diversity in the pigeonpea (Cajanus cajan) core collection. Genet Resour Crop Evol 54:1167–1184

Upadhyaya HD, Gowda CLL, Sastry DVSSR (2008) Plant genetic resources management: collection, characterization, conservation and utilization. J SAT Agric Res 6:1–15

Wang B, Chang R-Z, Tao L, Guan R-X, Yan L, Zhang M-H, Feng Z-F, Qiu L-J (2003) Identification of SSR primer numbers for analyzing genetic diversity of Chinese soybean cultivated soybean. Mol Plant Breed 1:82–88

Wang L-X, Guan Y, Guan R-X, Li Y-H, Ma Y-S, Dong Z-M, Liu X, Zhang H-Y, Zhang Y-Q, Liu Z-X, Chang R-Z, Xu H-M, Li L-H, Lin F-Y, Luan W-J, Ya N-Z, Ning X-C, Zhu L, Cui Y-H, Piao R-H, Liu Y, Chen P-Y, Qiu L-J (2006) Establishment of Chinese soybean (*Glycine max*) core collections with agronomic traits and SSR markers. Euphytica 151:215–223

Wang L–L, Guan R-X, Qi Z, Qiu L-J, Luo S-P (2008) Analysis of 11S/7S ratio between soybean mini core collection and cultivars. J Plant Genet Resour 9:68–72

Wang L-L, Duan C-X, Shuang S-M, Ren G-X (2011a) Determine the isoflavones content in soybean and analysis by near-infrared reflectance spectroscopy (NIRS). Food Sci Technol 36:242–246

Wang X, Fu G, Yuan S, Zhang H (2011b) Genetic diversity analysis of seed appearance quality of Chinese soybean mini core collection. Procedia Eng 18:392–397

Wei X, Xu J, Guo H, Jiang L, Chen S, Yu C, Zhou Z, Hu P, Zhai H, Wan J (2010) *DTH8* suppresses flowering in rice, influencing plant height and yield potential simultaneously. Plant Physiol 153:1747–1758

Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lu S, Wu H, Tabata S, Harada K (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc Natl Acad Sci USA 109:E2155–E2164

Xie H, Chang R-Z, Cao Y-S, Zhang M-H, Feng Z-F, Qiu L-J (2003) Selection of core SSR loci by using Chinese autumn soybean. Sci Agric Sin 36:360–366

Yaklich RW, Helm RM, Cockrell G, Herman EM (1999) Analysis of the distribution of the major soybean seed allergens in a core collection of *Glycine max* accessions. Crop Sci 39:1444–1447

Yan Z, Chang R-Z, Guan R-X, Liu Z-X, Qiu L-J (2003) Analysis of similarity and difference of various collections of soybean variety Mancangjin by using agronomic traits and SSR markers. J Plant Genet Resour 4:128–133

Yan L, Feng Y, Yang C-Y, Shi X-L, Zhang M-C (2012) The character of protein and oil content distribution among introgression lines using Jidou12 as recipient parent. Acta Agric Boreali Sin 27:87–92

Young VR (1991) Soy protein in relation to human protein and amino acid nutrition. J Am Diet Assoc 91:828–835

Yuan F-J, Jiang Y, Dong D-K, Zhu D-H, Li B-Q, Zhu S-L, Fu X-J (2011) Analysis of isoflavone content in soybean core collection. J Chin Cereals Oils Association 26:5–8

Zhang X-Y, Pang B-S, You G-X, Wang L-F, Jia J-Z, Dong Y-C (2002) Allelic variation and genetic diversity at *Glu-1* loci in Chinese wheat (*Triticum aestivum* L.) germplasms. Sci Agri Sin 35:1302–1310

Zhang Y-Q, Guan R-X, Liu Z-X, Chang R-Z, Yao Y-S, Qiu L-J (2006) Identification of Gly m Bd 28 K and Gly m Bd 30 K lacking soybean by random sampling of core collection in soybean. Acta Agron Sin 32:324–329

Zhang H, Zhang D, Wang M, Sun J, Qi Y, Li J, Wei X, Han L, Qiu Z, Tang S, Li Z (2011a) A core collection and mini core collection of *Oryza sativa* L. in China. Theor Appl Genet 122:49–61

Zhang J-W, Han F-X, Sun J-M, Yu F-K, Ma L, Chen M-Y, Zhang J-Y, Yan S-R, Yang H (2011b) Genetic variation of the protein content and correlation between protein content and main agronomic characters of soybean. J Plant Genet Resour 12:501–506

Zhao J, Fu J-B, Liao H, He Y, Nian H, Hu Y-M, Qiu L-J, Dong Y-S, Yan X-L (2004) Characterization of root architecture in an applied core collection for phosphorus efficiency of soybean germplasm. Chinese Sci Bull 49:1611–1620

Zhao L, Dong Y, Liu B, Hao S, Wang K, Li X (2005) Establishment of a core collection for the Chinese annual wild soybean (*Glycine Soja*). Chin Sci Bul 50:989–996