Available online at www.sciencedirect.com

ScienceDirect

Identification of a novel seed size associated locus SW9-1 in soybean☆

Jiajia Li[a], *Jinghui Zhao*[a], *Yinghui Li*[b], *Yali Gao*[a], *Sunan Hua*[a], *Muhammad Nadeem*[a], *Genlou Sun*[a,c], *Wenming Zhang*[a], *Jinfeng Hou*[d], *Xiaobo Wang*[a,*], *Lijuan Qiu*[b,*]

[a]*School of Agronomy, Anhui Agricultural University, Hefei 230036, Anhui, China*
[b]*The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/Key Laboratory of Crop Gene Resource and Germplasm Enhancement (MOA), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[c]*Biology Department, Saint Mary's University, Halifax, NS B3H 3C3, Canada*
[d]*College of Horticulture, Anhui Agricultural University, Hefei 230036, Anhui, China*

ARTICLE INFO

Article history:
Received 16 August 2018
Received in revised form 3 December 2018
Accepted 12 January 2019
Available online 30 January 2019

Keywords:
Soybean [*Glycine max* (L.) Merr.]
Seed size/weight
SW9-1
Molecular breeding

ABSTRACT

Seed size is one of the vital traits determining seed appearance, quality, and yield. Untangling the genetic mechanisms regulating soybean 100-seed weight (100-SW), seed length and seed width across environments may provide a theoretical basis for improving seed yield. However, there are few reports related to QTL mapping of 100-SW across multiple ecological regions. In this study, 21 loci associated with seed size traits were identified using a genome-wide association of 5361 single nucleotide polymorphisms (SNPs) across three ecoregions in China, which could explain 8.12%–14.25% of the phenotypic variance respectively. A new locus, named as *SW9-1* on chromosome 9 that explained 10.05%–10.93% of the seed weight variance was found significantly related to seed size traits, and was not previously reported. The selection effect analysis showed that *SW9-1* locus has a relatively high phenotypic effect (13.67) on 100-SW, with a greater contribution by the accessions with bigger seeds (3.69) than the accessions with small seeds (1.66). Increases in seed weight were accompanied by increases in the frequency of *SW9-1T* allele, with >90% of the bred varieties with a 100-SW >30 g carrying *SW9-1T*. Analysis of *SW9-1* allelic variation in additional soybean accessions showed that *SW9-1T* allele accounting for 13.83% of the wild accessions, while in 46.55% and 51.57% of the landraces and bred accessions, respectively, this results indicating that the *SW9-1* locus has been subjected to artificial selection during the early stages of soybean breeding, especially the utilization of *SW9-1T* in edamame for big seed. These results suggest that *SW9-1* is a novel and reliable locus associated with seed size traits, and might have an important implication for increasing soybean seed weight in molecular design breeding. Cloning this locus in future may provide new insights into the genetic mechanisms underlying soybean seed size traits.

☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding authors.*
E-mail addresses: wangxiaobo@ahau.edu.cn, (X. Wang), qiulijuan@caas.cn. (L. Qiu).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

https://doi.org/10.1016/j.cj.2018.12.010

1. Introduction

The main objectives of soybean breeding are to increase seed yield and improve the quality of seed products to satisfy the requirements for human and animal consumption. Seed size is one of the major traits targeted by soybean breeders because it is vital for determining seed appearance, quality, and yield [1,2]. Soybean seed weight is often expressed as 100-seed weight (100-SW), and is affected by seed size, which is measured based on seed length, width, and thickness as well as their corresponding ratios [1]. Seed size is an important soybean yield component [3–6]. Thus, selecting soybean lines with increasing 100-SW is one of the main goals for soybean breeding programs [7,8]. However, soybean seed yield and 100-SW are influenced by multiple quantitative trait loci (QTL) or genes as well as by various environmental conditions [2,9,10]. Thus, it is difficult to improve soybean yield *via* traditional breeding methods [11].

The application of molecular markers represents a fast and accurate approach for identifying new sources of genetic variation and for investigating the genetic factors controlling quantitative traits [12,13]. The use of molecular markers to indirectly select important agronomic traits (*e.g.*, seed size and weight) can improve the efficiency of traditional plant breeding programs [14]. Along with the development of soybean consensus genetic maps [15–23], molecular markers have facilitated the identification of QTL controlling important quantitative traits in many soybean populations [24]. Additionally, the construction of genetic maps and analyses of QTL have enabled researchers to estimate the number of QTL regulating genetic variations in soybean [25]. This strategy has also been widely applied for detecting and identifying QTL for soybean seed size (weight) traits. Consequently, many QTL, single nucleotide polymorphism (SNP) markers, and genes associated with these traits have been identified over the past two decades [1,2,8,11,12,25–39]. However, most of the identified QTL and SNP markers have rarely been applied in soybean breeding programs.

Genome-wide association study (GWAS) has become a powerful approach for elucidating complex agronomic traits and identifying causal variants with modest effects on target traits in crops [6,40–44]. An earlier investigation indicated that SNP genotyping by genome-wide identification effectively promotes the applicability of association analyses for QTL mapping in crops [45]. Specifically, a GWAS with SNP markers has been applied to analyze various important traits in many plant species, resulting in the identification of causal genes for a broad range of complex traits in diverse plant species, including *Arabidopsis thaliana* [41], rapeseed (*Brassica napus*) [43], rice (*Oryza sativa* L.) [46], maize (*Zea mays*) [47,48], sorghum [49], and soybean (*Glycine* max) [2,6,11,34,39,50–55]. Therefore, identifying SNPs associated with seed size-related traits may enable soybean breeders to combine the causal genes during the breeding of new lines.

In our previous study, we observed that the *GmCYP78A10* gene associated with seed size may have undergone artificial selection during soybean breeding [2]. In *A. thaliana*, the *P450/CYP78A* gene family, which includes *AtCYP78A5*, *AtCYP78A6*, and *AtCYP78A9*, has very important effects on seed size [56–58]. Moreover, Du et al. [59] recently confirmed that overexpressed *GmCYP78A5* (*i.e.*, a close homolog of *GmCYP78A10*) in transgenic soybean plants significantly increased seed size and weight. Our previous results also revealed two highly conserved *GmCYP78A10* alleles, *GmCYP78A10a* allele mainly distributed in wild soybean (*G. soja*), while *GmCYP78A10b* predominantly presented in cultivated soybean (*G. max*) [2]. We speculated that soybean breeders initially selected big seeds (cultivated soybean) carrying *GmCYP78A10b*, however, this artificial selection did not significantly affect soybean seed yield. To further elucidate the evolution of loci or genes related to soybean seed size, we analyzed genome regions associated with soybean seed size traits including 100-SW, seed length (SL), and seed width (SW) using GWAS and SNP genotyping of 146 soybean accessions under multiple environments from three major ecoregions in China, and found a novel locus *SW9-1* associated with soybean seed size. *SW9-1* allelic variation in wild and cultivated soybean accessions was characterized. Artificial selection effect on this locus and its application in molecular breeding to develop new soybean lines with increased seed weight were discussed.

2. Materials and methods

2.1. Plant materials and phenotypic data collection

A total of 146 soybean accessions mainly from three major ecoregions in China were selected for association mapping including 49 accessions from Huanghuai (HH), 49 from Northeast South (NS), and 48 from Northeast North (NN) (Table S1). As soybean is much more sensitive to photoperiod than other field crops like wheat, rice and corn, some of soybean varieties from northern part of China could not reach maturity when planting in Huanghuai geographic region [31]. To ensure the normal maturation, and obtain soybean seeds, thus, 49, 49, and 48 different soybean accessions were grown at 6, 6, and 2 experiment locations (environments) in HH, NS, and NN in China, respectively, with single-row of 3.00 m length and 0.50 m spacing between rows in a randomized complete-block design with three replicates (Table S1). One hundred and eighty-four bred accessions, 94 wild accessions, and 116 landrace from the Chinese mini-core soybean collection were used to validate the association of a derived cleaved amplified polymorphic sequence (dCAPS) marker for *SW9-1* locus. All these accessions were grown in single-row with 3.00 m length and 0.50 m spacing between rows in a randomized complete-block design with three replicates at Dayangdian experiment station of Anhui Agricultural University (31°56′6.65″N, 117°11′49.39″E) and the experimental farm of Suzhou Academy of Agricultural Sciences (33°38′30.61″N, 117°04′36.32″E) in 2016 and 2017, respectively.

After the soybean seeds were harvested and dried naturally until reaching a stable seed weight, 100-SW (g), SL (mm), and SW (mm) of all soybean accessions were centrally evaluated using High Precision Electronic Balance (APTP456 series) and Multifunction Seed Analyzer (SC-G type, Hangzhou Wanshen Detection Technology Co., Ltd., Hangzhou, China, http://www.wseen.com/).

2.2. DNA extraction and SNP genotyping

Genomic DNA was extracted from individual dry seeds as described by Kang et al. [60]. All soybean accessions were genotyped by the Illumina SoySNP6k iSelect BeadChip (Illumina, San Diego, CA, USA), which consists of 5361 SNPs [21]. Analyses of the chromosomal distributions, coding, and quality of the SNPs were described by Wen et al. [51]. The exclusion of SNPs with a missing data rate >0.25 and a minor allele frequency <0.05 resulted in 4987 (93.67%) SNP markers being retained for the subsequent analysis.

2.3. Population structure analysis

The population structure of the 146 soybean varieties was investigated using 4987 SNPs according to the Bayesian model-based program, STRUCTURE 2.3.4 [61,62]. The length of the burn-in period and the number of Markov Chain Monte Carlo replications after the burn-in were set at 100,000, with an admixture and allele frequencies-correlated model. Five independent runs were completed with the hypothetical number of subpopulations (k) ranging from 1 to 10. The correct k estimation was determined by combining the log probability of data [ln $P(D)$] from the STRUCTURE output and an *ad hoc* statistic, Δk, which was based on the rate of change in the log probability of data between successive k values [62]. On the basis of the correct k value, each soybean variety was assigned to a subpopulation for which the membership value (Q-value) was >0.5 [61,63], and the population structure matrix (Q) was generated for further analyses.

2.4. Phenotypic data analysis

Phenotypic data across all environments were analyzed with SAS 9.3 (SAS Institute, Inc., Cary, NC, USA). Additionally, location within in a year was treated as a single environment. The analysis of variance (ANOVA) for each trait was performed to evaluate the effect of genotype and environment on phenotypic variance with R 'lm' function (R Core Team, 2012, https://www.r-project.org/). The model is $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$, where i is for each line, j is for each environment. τ_i is for genotype effect and β_j is for environmental effect, μ is for average value and ε_{ij} is for error. The broad-sense heritability (h^2) for each trait was calculated as:

$$h^2 = \delta^2_g/(\delta^2_g + \delta^2_e/n)$$

where, δ^2_g is the genetic variance, δ^2_e is the residual variance and n is the number of environments. The estimates of δ^2_g and δ^2_e were obtained by the mixed linear model using lme4 package in R (R Core Team, 2012, https://www.r-project.org/), treating genotype and environment as random effects, as follow: model = lmer (trait ~ (1|Line) + (1|Env), data = data) [64].

The effects of the population structure on the phenotypic variation were calculated based on the mean values for each trait using PROC general linear model (GLM). The model statement included one of the two components of the Q matrix (k = 3) [65]. Correlation coefficients between soybean yield and yield components were estimated with PROC CORR.

2.5. Genome-wide association study

The SPAGeDi [66] and TASSEL 4.0 [67] programs were used to calculate the pair-wise relatedness coefficients (K, kinship matrix) for estimating the genetic relatedness among individuals, with the negative value of kinship set as 0. A GWAS based on the GLM and mixed linear model (MLM) was conducted for three soybean seed size traits with the TASSEL 4.0 program and 4987 SNPs [68]. For the MLM analysis, a marker-based kinship matrix (K) obtained with TASSEL 4.0 was used along with the Q matrix to correct for family and population structures. By Bonferroni correction, a P-value threshold of 1×10^{-3} was used to declare the significance of marker trait associations. The SNP sites with relatively high and stable contribution rates were selected for further analyses.

2.6. Development of molecular markers

The Primer Premier 5.0 software (Premier Bio-soft International, Palo Alto, CA, USA) was used to develop dCAPS markers for three SNP loci related to non-synonymous mutations associated with soybean seed size traits identified by the GWAS. Enzyme digestion information for the SNP loci was confirmed using the dCAPS Finder 2.0 (http://helix.wustl.edu/dapaps/dcaps.html) online program, and Eam1105I, *Hinf* I, and *Xba* I were selected as candidate enzymes. Details regarding the developed dCAPS markers were provided in Table S2. A PCR was completed in a 20 μL reaction volume consisting of 2.0 μL template DNA (100 ng μL^{-1}), 2.0 μL 10× buffer, 2.0 μL dNTPs, 0.5 μL each primer (10 μmol L^{-1}), 0.2 μL 1.0 U *Taq* DNA polymerase, and 12.8 μL ddH_2O. The PCR was completed using the program: 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 55 °C for 30 s with a decrease of 0.2 °C per cycle, and 72 °C for 30 s; 72 °C for 5 min. A quantitative real-time PCR was completed using an ABI PCR System (Applied Biosystems, Foster City, CA, USA). The PCR products were digested with restriction endonucleases for 3 h following the manufacturer's recommended protocols (Takara, Dalian, China).

2.7. Statistical analysis

Data were analyzed using SAS, SPSS, and Microsoft Excel. Additionally, Student's t-test was completed using the independent-samples t-test (version 19.0; http://www.spss.com/). The phenotypic effect values for the allelic variations were calculated with the following equation:

$$a_c = \sum x_{cd}/n_c - \sum N_e/n_e$$

where, a_c is the phenotypic effect value; x_{cd} is the phenotypic data for the d-trait of the variety carrying the c-allelic variation type; n_c is the number of varieties containing the c-allelic variation type; N_e is the phenotypic data of the e-individual carrying the non-allelic variation type; n_e is the number of varieties containing the non-allelic variation type.

An a_c value >0 indicated that allelic variation type had additive effects on soybean seed size, while a value <0 indicated the allelic variation had negative effects.

3. Results

3.1. Phenotypic variation and population structure analysis

Each trait varied widely among the different ecoregions (Table 1). The 100-SWs in HH, NS, and NN ecoregions were 12.71–23.15, 11.60–24.07, and 13.19–20.71 g, with average values of 18.27, 20.16, and 17.29 g, respectively (Table 1). Compared to NS and NN ecoregions, large range of variation for 100-SW (3.06), SL (0.41), and SW (0.33) in the 6 environments of HH ecoregion was observed (Table 1). The h^2 value of the different soybean accessions across three major ecoregions was 70.13%–93.87% (Table 1). The highest h^2 value (93.87%) was observed in HH, indicating that the majority of the 100-SW variation was due to genetic effects. In contrast, the 100-SW in the other two ecoregions might be affected primarily by environmental factors.

To avoid false-positive associations because of population stratification, STRUCTURE and a neighbor-joining tree-based method were used to estimate the relatedness among 146 soybean accessions using 4987 SNPs (Fig. 1). The distribution of the ln $P(D)$ value for each given k value did not show a clear peak trend (Fig. 1-A). The *ad hoc* quantity (Δk) exhibited a much higher likelihood at $k = 3$ than at $k = 2$ and 4–10 (Fig. 1-B), suggesting that the 146 soybean accessions could be divided into three major subpopulations (Fig. 1-C). The neighbor-joining tree results (Fig. 1-D) were consistent with the STRUCTURE results (Fig. 1-C). Of the 86 accessions in subpopulation P_3, 49 were from HH, and exhibited relatively late maturity. The subpopulation P_1 (26 accessions) and P_2 (34 accessions) were mainly from NS and NN, and exhibited relatively early maturity. An earlier study suggested that diversity in the photoperiod response among groups that differed regarding maturity might be the primary factor mediating the differentiation of cultivated soybean [69]. The co-existence or overlapping of genetic and species differentiation were likely due to natural and artificial selection based on fitness [70]. The Q matrices obtained from the STRUCTURE program for the three subpopulations were used to control the variation of the population structure in the following genetic analysis.

Table 1 – Analysis of soybean 100-seed weight as well as seed length and width across three ecoregions.

Trait	Ecoregion	Min	Max	Mean	SD	CV (%)	h^2 (%)
100-SW	Huanghuai (6E)	12.71	23.15	18.27	3.06	16.75	93.87
	Northeast South (6E)	11.60	24.07	20.16	1.85	9.18	89.79
	Northeast North (2E)	13.19	20.71	17.29	1.14	6.59	70.13
SL	Huanghuai (6E)	6.82	8.88	7.73	0.41	5.30	94.87
	Northeast South (6E)	5.46	8.60	7.43	0.28	3.77	90.02
	Northeast North (2E)	6.43	7.81	7.05	0.21	2.98	69.93
SW	Huanghuai (6E)	5.62	7.28	6.49	0.33	5.08	95.53
	Northeast South (6E)	5.17	7.45	6.88	0.22	3.20	95.99
	Northeast North (2E)	6.12	6.90	6.55	0.14	2.14	60.40

E, environment; 100-SW, 100-seed weight (g); SL, seed length (mm); SW, seed width (mm); Min, minimum value (g); Max, maximum value (g); CV, coefficient of variation; h^2, broad-sense heritability.

3.2. Identification of SNP loci related to soybean seed size

A total of 21 SNPs including 3 SNPs associated with 100-SW, 16 SNPs associated with SL, and 2 SNPs associated with SW were identified by GWAS using 4987 SNP markers (Table S3). Eighteen of the 21 SNPs, each explained >10% of the phenotypic variance, while the other three SNPs each explained >8% of the phenotypic variance. On average, 11.34% of the phenotypic variance could be explained by each SNP (Table S3). All of these SNPs were significantly associated with seed size traits in all 14 environments in three soybean ecoregions of China. To better identify significant SNP loci associated with soybean seed size traits, two environmental locations with higher relevance were selected for each of three ecoregions. An association analysis of eight combinations representing six locations was then conducted. Accordingly, 3, 17, and 1 SNPs were identified significantly associated ($P < 0.01$) with 100-SW, SL, and SW, respectively, in at least four combinations for three ecoregions (Table S4). One SNP (ss248666643) on chromosome 15 was related to 100-SW and SL in at least five combinations (Table S4). Moreover, two SNPs (ss246792949 and ss244709037) were significantly associated with 100-SW, SL, and SW in at least four combinations, and explained an average of 10.01% of the phenotypic variance of three traits (Table S4).

3.3. Identification of loci with significant effects on seed size and development of molecular marker

To compare the allelic variation of three SNP loci (ss246792949T/C, ss244709037A/G, and ss250164298G/A) significantly associated with seed size (100-SW), the dCAPS markers corresponding to three loci were developed. Sequence analysis results revealed that the PCR products of the three loci were all target fragments, with mismatched bases at the corresponding SNP positions, which were consistent with the expected results. A total of 50 bred soybean varieties with extreme 100-SW (25 100-SW ≤ 13.60 g; 25 100-SW ≥ 28.90 g) were further used to verify the three loci with the corresponding dCAPS markers. The sizes of the digested DNA fragments were consistent with the predicted sizes (Table S2, Fig. 2). These results confirmed the suitability of the dCAPS markers developed for three loci.

3.4. Identification and validation of the SW9-1 (ss246792949T/C) locus

Based on the average value of 100-SW, we divided soybean accessions into big seeds (BSs) and small seeds (SSs) group. The genotyping of the dCAPS marker revealed that the frequencies of the ss244709037A and ss244709037G alleles were 40% and 60%, respectively. No obvious differences in the selection efficiencies of ss244709037A and ss244709037G were observed (Table 2). However, the frequency of the ss250164298A allele was 28%, and the selection efficiency in SSs was about twice higher than that in BSs. The frequency of the ss250164298G allele was 72%, which was mainly expressed in BSs (55.56%).

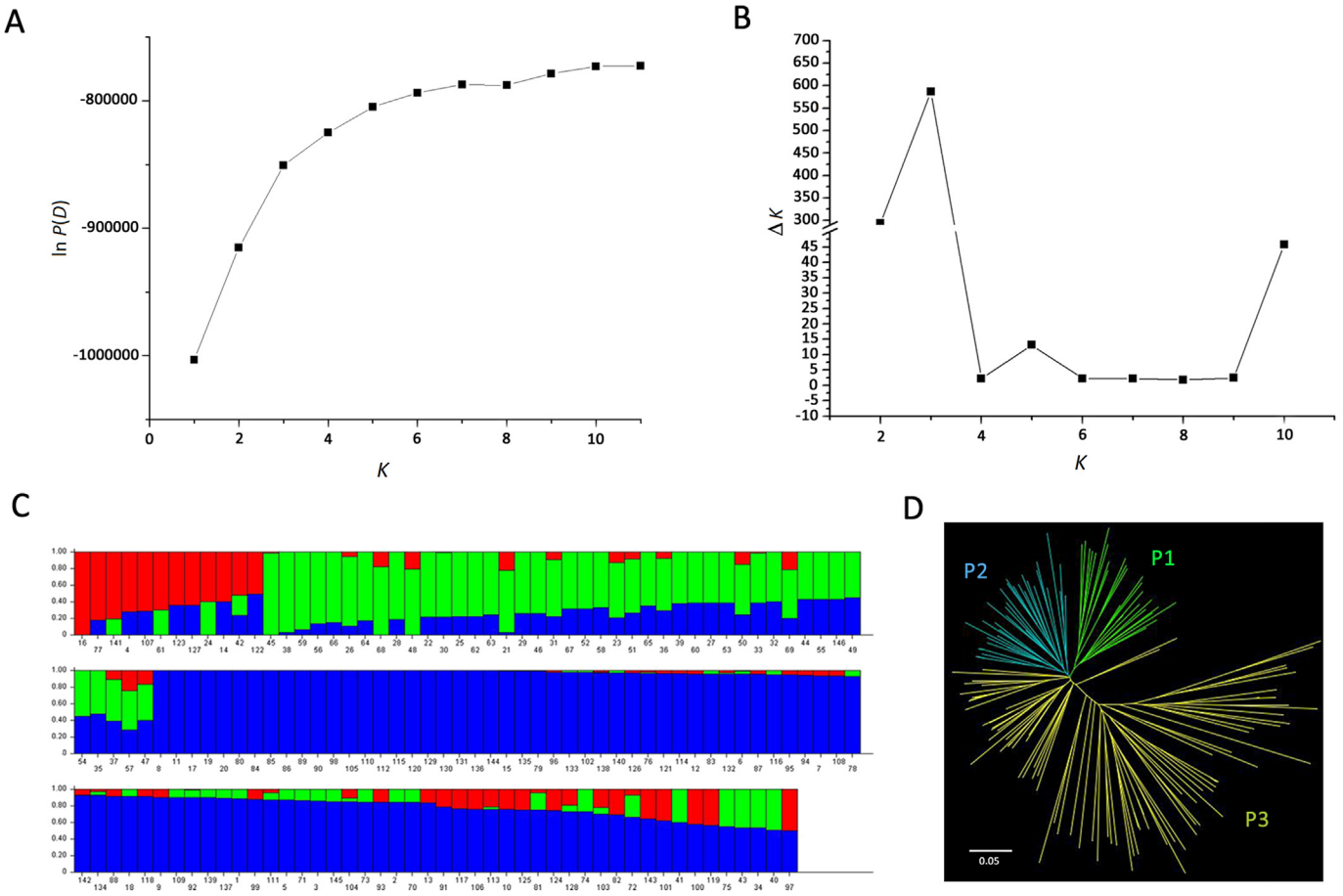

Fig. 1 – Analysis of the population structure of 146 soybean accessions. (A) Distribution of the ln *P*(*D*) values corresponding to each hypothetical *k*-value. (B) Calculated *k*-value over five runs with the putative *k* ranging from 1 to 10. The highest Δ*k* value occurred when *k* = 3, indicating that all the tested soybean accessions could be divided into three major subpopulations. (C) Model-based population structure for all 146 soybean accessions. (D) Neighbor-joining tree based on Nei's genetic distance matrix.

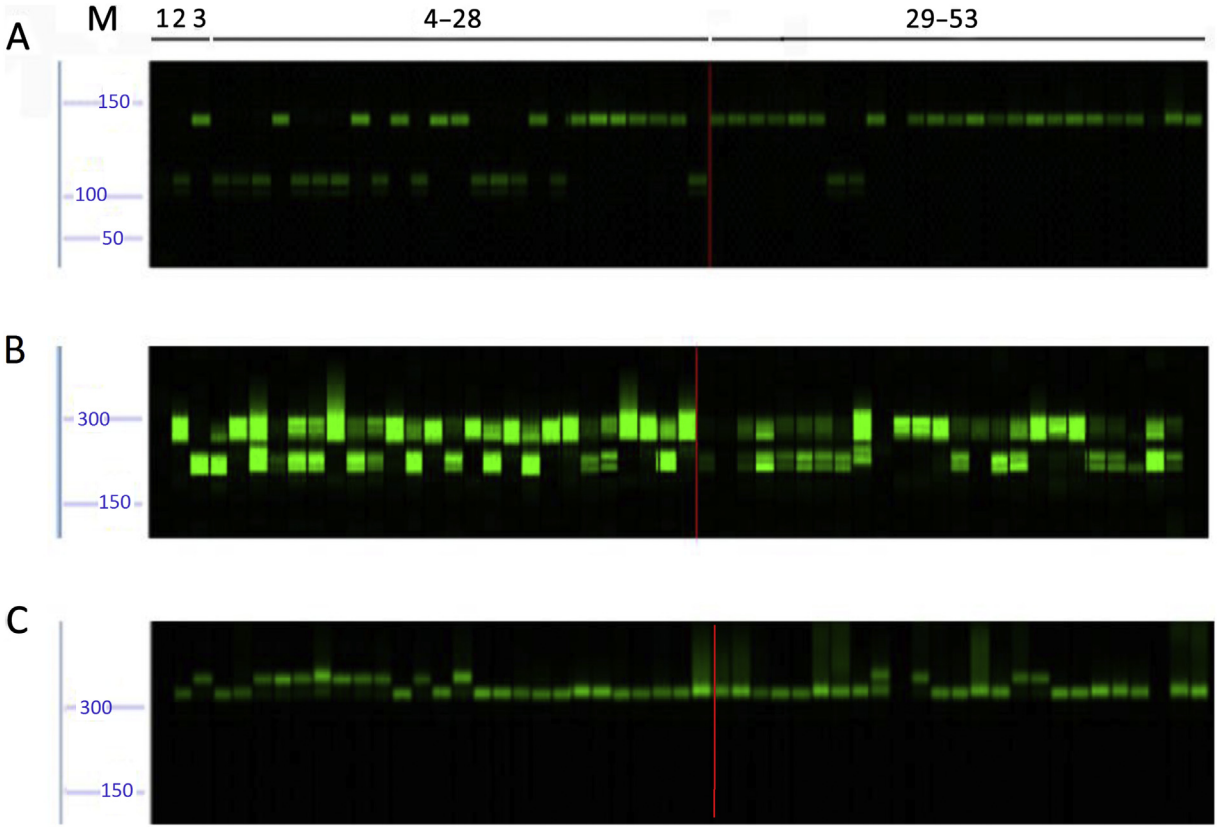

Fig. 2 – Polymorphism analysis involving dCAPS markers and soybean cultivars. (A) Polymorphisms of dCAPS marker ss246792949 in 52 soybean accessions. (B) Polymorphisms of dCAPS marker ss244709037 in 52 soybean accessions. (C) Polymorphisms of dCAPS marker ss250164298 in 52 soybean accessions. M: marker; 1: ddH_2O; 2–3: soybean accessions with a known genotype (A: 2 represents the CC genotype and 3 represents the TT genotype; B: 2 represents the AA genotype and 3 represents the GG genotype; C: 2 represents the GG genotype and 3 represents the AA genotype); 4–53: 50 soybean accessions with an unknown genotype, of which 4–28: 100-seed weight of 7.40–13.60 g and 29–53: 100-seed weight of 28.90–44.25 g.

Additionally, the allele variation at the ss250164298 (A/G) locus did not significantly affect soybean seed weight (Table 2). The frequency of the ss246792949T allele (named *SW9-1T*) was 70% in BSs, and the selection efficiency in BSs was approximately twice higher than that in SSs. In contrast, the frequency of the ss246792949C allele (named *SW9-1C*) was 30%, which was mainly expressed in SSs (86.67%) (Table 2). The 100-SW of soybean varieties with *SW9-1T* allele was significantly heavier than those with *SW9-1C* allele ($P < 0.01$) (Table 2). These results indicated that the allelic variation at the *SW9-1* locus significantly affected soybean seed weight. Furthermore, the *SW9-1* locus had relatively high phenotypic effect (13.67) on 100-SW, with greater contribution by the accessions with BSs (3.69) than those with SSs (1.66). The *SW9-1T/C* locus had more significant effect on 100-SW than the ss244709037 (A/G) and ss250164298 (A/G) loci (Table 2). Meanwhile, the quantile-plots indicated that

Table 2 – Allelic variation at three loci related to soybean seed weight corresponding to the type of seed distribution, frequency, selection efficiency, and phenotypic analysis of the allelic variation effects.

Allelic variation	Sample number	Frequency (%)	Select efficiency (%)		Mean ± SD (g)	a_c (T–C; A–G; A–G)			P-value
			Small seed	Big seed		Total a_c	Small seed	Big seed	
ss246792949-T	35	70.00	12 (34.29)	23 (65.71)	26.61 ± 11.44	13.67	1.66	3.69	9.92E
ss246792949-C	15	30.00	13 (86.67)	2 (13.33)	12.94 ± 7.42		−1.66	−3.69	−05 **
ss244709037-A	20	40.00	11 (55.00)	9 (45.00)	21.46 ± 11.31	−1.74	1.16	−0.88	6.23E−01
ss244709037-G	30	60.00	14 (46.67)	16 (53.33)	23.20 ± 12.75		−1.16	0.88	
ss250164298-A	14	28.00	9 (64.29)	5 (35.71)	18.47 ± 12.72	−5.61	−2.42	0.85	1.43E−01
ss250164298-G	36	72.00	16 (44.44)	20 (55.56)	24.08 ± 11.67		2.42	−0.85	

** Significant at $P < 0.01$.

Table 3 – Verification of the effectiveness of SW9-1T/C based on dCAPS markers as well as bred varieties and wild accessions.

Type	Allelic variation	Sample number	Frequency (%)	Select efficiency (%)		Mean ± SD (g)	$a_{c\ (T-C)}$			P-value
				Small seed	Big seed		Total a_c	Small seed	Big seed	
Bred variety	SW9-1T	110	0.60	53 (48.18)	57 (51.82)	22.28 ± 8.30	3.73	0.86	3.11	1.33E–03**
	SW9-1C	74	0.40	46 (62.16)	28 (37.84)	18.55 ± 6.44		–0.86	–3.11	
Wild type	SW9-1T	13	0.14	/	/	2.89 ± 1.72	0.73	/	/	7.75E–02
	SW9-1C	81	0.86	/	/	2.16 ± 1.31		/	/	

** Significant at $P < 0.01$; /, not calculated.

the allelic variation at the SW9-1 locus had significant effect (Δm = 2.36 g; $P = 5.748E{-}05$) on 100-SW (Fig. S1).

3.5. Distribution of SW9-1 alleles in soybean accessions

To confirm the effect of SW9-1 on soybean seed size (100-SW), we analyzed total of 184 accessions including additional 134 newly bred soybean accessions. The frequency of the SW9-1T allele was 60%, with higher selection efficiency for the accessions with BSs (51.82%) than for the accessions with SSs. The frequency of the SW9-1C allele was 40%, which was mainly detected in SSs (62.16%) (Table 3). Additionally, the SW9-1T/C alleles had a significant effect on 100-SW, with a phenotypic effect value of 3.73 (Table 3). The results suggested that SW9-1C allele was primarily in the accessions with SSs, while SW9-1T was mainly in the accessions with BSs (Tables 2 and 3). To confirm the association of these alleles with seed size, 94 wild-type accessions with averaged 100-SW of 2.26 g were used for an additional genotyping with these dCAPS markers. We observed that 81 of 94 wild-type accessions (86.17%) mainly contained the SW9-1C allele, while the remaining 13 accessions mainly comprised the SW9-1T allele (Table 3).

3.6. Genetic and selective effects of SW9-1T/C alleles

We observed that SW9-1T occurred more frequently than SW9-1C in elite cultivars or accessions with BSs, in contrast to the allelic distribution in wild soybean accessions with SSs (Tables 2 and 3). This suggested that SW9-1T might have been artificially selected during the early stages of soybean breeding. To verify this hypothesis, the frequencies of the two alleles in 94 wild accessions, 116 landraces, and 318 bred accessions were further determined. The SW9-1T allele was detected in 13.83% of the wild accessions, while in 46.55% landraces and 51.57% bred accessions (Table 4, Fig. 3). The frequency difference between wild and cultivated (bred varieties and especially landraces) soybean populations implied that the SW9-1 locus was under artificial selection during the early stages of soybean domestication. An analysis of the distribution of SW9-1 alleles in bred soybean seeds with different 100-SWs revealed that SW9-1T was in the accessions with heavier seeds, while SW9-1C was in the accessions with lighter seeds (Fig. 4). Bred soybean varieties were further divided into four groups according to the 100-SW range. The SW9-1C allele was detected in 77.78% of varieties with a 100-SW <10 g, while only 22.22% carried SW9-1T. Increases in 100-SW were accompanied by increases in the frequency of SW9-1T. >90.00% bred varieties with a 100-SW >30 g carrying SW9-1T. Moreover, there was a significantly difference ($P < 0.01$) between the 100-SWs associated with SW9-1T and SW9-1C (Table 4, Fig. 4).

3.7. Novelty of SW9-1 locus for soybean weight

On the basis of our results, the SW9-1 locus was localized to chromosome 9 in soybean genome (Table S5). Comparison of SW9-1 position with QTL positions reported in bi-parental mapping studies revealed that 13 QTL/loci on chromosome 9 were associated with soybean seed weight (100-SW) (Table S6). However, the physical location of the SW9-1 locus was not

Table 4 – Distribution and selective effect of SW9-1T/C alleles in soybean accessions.

Sample type	Range of 100-SW (g)	Allelic variation		Mean ± SD		P-value
		SW9-1C	SW9-1T	SW9-1C	SW9-1T	
Wild soybean (94)	0.50–7.30	81 (86.17%)	13 (13.83%)	2.16 ± 1.31	2.89 ± 1.72	7.80E–02
Landrace (116)	7.65–39.05	62 (53.45%)	54 (46.55%)	15.18 ± 3.94	16.32 ± 5.14	1.20E–04**
Bred variety (318)	7.45–45.00	154 (48.43%)	164 (51.57%)	18.27 ± 4.72	21.46 ± 7.02	2.30E–05**
Bred variety (318)	≤10	7 (77.78%)	2 (22.22%)	8.95 ± 0.99	8.78 ± 0.88	8.70E–01
	10–20	101 (56.11%)	79 (43.89%)	16.23 ± 1.98	16.04 ± 1.99	3.20E–02*
	20–30	44 (40.74%)	64 (59.26%)	23.87 ± 2.65	24.38 ± 2.66	6.50E–03**
	>30	2 (9.52%)	19 (90.48%)	30.65 ± 0.57	35.52 ± 4.37	1.20E–04**

* Significant at $P < 0.05$.
** Significant at $P < 0.01$.

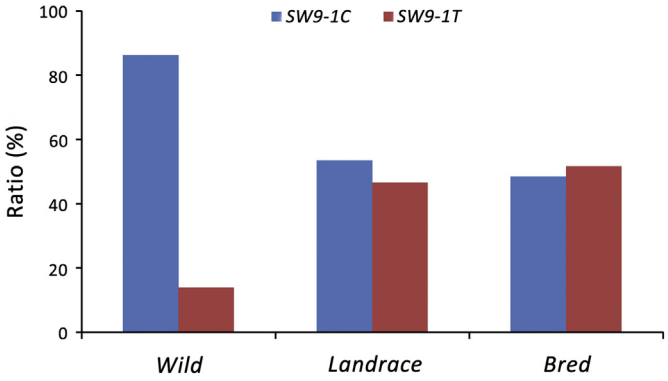

Fig. 3 – Distribution of the SW9-1T/C alleles in different soybean accessions.

within and did not overlap with the physical location interval for known QTL. Thus, *SW9-1T/C* may be a new allele determining soybean seed size (weight).

4. Discussion

4.1. *Analyses of population structure and genetic diversity*

Previous studies have indicated that SNPs can be used to obtain accurate sequence information *via* high-throughput sequencing technology to compensate for bi-allelic shortcomings [70–72]. SNP markers have also widely been used in soybean genotyping [6,11,52–55,73]. These studies suggested that achieving a similar accuracy to that of simple sequence repeats (SSRs) in analyses of population structures and genetic diversity required about 10 times more SNPs than SSRs [74–75]. In this study, 146 soybean accessions from three different ecoregions in China were classified into three subpopulations, which the highest genetic variability (CV) and h^2 associated with 100-SW (16.82% and 93.87%, respectively), SL (5.30% and 94.87%, respectively), and SW (5.11% and 95.53%, respectively) were observed in the HH ecoregion (P_3). Meanwhile, the lowest values were observed for the NN (P2) ecoregion (Table 1). These results indicated that there were significant genetic differences among the soybean varieties grown in the three ecoregions (subpopulations).

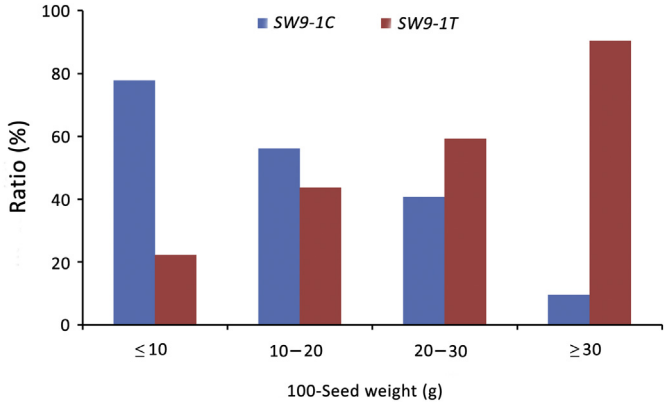

Fig. 4 – Distribution of the SW9-1T/C alleles in different 100-SW ranges for bred soybean varieties.

4.2. *SW9-1 is a reliable locus for soybean 100-SW*

Because of functional markers (CAPS/dCAPS markers) can quickly and effectively identify SNP alleles, they have been widely applied to detect mutations causing phenotypic changes in many crops, including wheat (*Triticum aestivum* L.) [76,77], *Brassica oleracea* [78], maize [79], rice [80,81], and barley (*Hordeum vulgare* L.) [82]. Here, the reliability of *SW9-1* locus was evaluated by our newly developed dCAPS marker. The results showed that *SW9-1* locus has the higher phenotypic effect value (13.67) for 100-SW than the ss244709037(A/G) and ss250164298(A/G) loci, more effect on accessions with BSs (3.69) than on accessions with SSs (1.66) (Table 2). Therefore, we speculated that the *SW9-1T/C* alleles have more significant effect on soybean seed weight. Further analysis results confirmed this speculation (Table 3, Fig. S1). Important soybean quantitative traits, including seed size, are largely affected by multiple genes and environmental conditions [2,9,10]. Consequently, related QTL or loci that can be detected in different environments or genetic backgrounds may be useful because of their apparent stability [83,84]. To verify the stability of *SW9-1* locus in multiple environments, 184 bred accessions involved in genotyping were divided into different ecoregion according to their sources. The effect of allelic variation at *SW9-1* locus on 100-SW showed different distribution in different ecoregion, 100-SW of accessions with *SW9-1T* type was greater than that of varieties with *SW9-1C* type in different ecoregion, especially in Huanghuai ecoregion, there were significant differences in 100-SW between accessions with *SW9-1T* and *SW9-1C* loci. Thus, we concluded that the *SW9-1* locus was reliable and persistent across different environments, and significantly influenced soybean 100-SW. The high reliability of this locus across multiple environments might have considerable implication for marker-based breeding of soybean cultivars with enhanced seed size (100-SW).

4.3. *Domestication and artificial selection for SW9-1 locus*

Seed size is an important determinant of the evolutionary fitness of plant species [85]. It is also a key agronomic trait

during crop domestication because of its important role in determining the seed yield of crops [85–87]. Previous investigations uncovered many QTL or genes correlated with domestication-related traits, including seed size, seed weight, plant height, flower color, stem determinacy, seed coat color, hilum color, and pubescence form [1,2,32–36,39,88–92]. In this study, we observed that the *SW9-1C* allele was mainly expressed in soybean populations with small seeds, in contrast, *SW9-1T* was the predominant allele in soybean varieties with big seeds. Further analysis of additional accessions indicated that *SW9-1T* was gradually selected for use (13.83% wild accessions, 46.55% landraces and 51.57% bred accessions), while *SW9-1C* was gradually eliminated. Increases in seed weight were associated with increases in the frequency of *SW9-1T*, >90.00% bred soybean accessions (mostly were vegetable soybeans (edamame)) with a 100-SW >30 g carried this allele. The results indicated that the allelic variation of *SW9-1* loci had been subjected to domestication or artificial selection during the early stage of soybean breeding programme, especially *SW9-1T* in edamame with heavier seeds. Considering that the *SW9-1T/C* alleles significantly affect soybean seed size (100-SW), the positive effects of *SW9-1T* on seed size should be applied in soybean modern breeding programs.

4.4. *Potential implication of* SW9-1 *locus for soybean breeding*

Currently, >400 QTL or markers for soybean seed size traits have been detected on 20 chromosomes, including >280 QTL or markers associated with seed weight (100-SW) (SoyBase, http://www.soybase.org/; Table S7). A QTL region controlling plant height, lodging, and seed yield spanning approximately 1.2 Mbp on chromosome 9 was detected in a GWAS [88]. Li et al. [34] subsequently used a pan-genome analysis to identify four genes [*Glyma09g06950* (Gm9: 5,783,764–5,788,819), *Glyma09g07090* (Gm9: 5,932,308–5,935,017), *Glyma09g07206* (Gm9: 6,072,782–6,078,905), and *Glyma09g07760* (Gm9: 6,694,656–6,699,276)] in this region with large-effect SNPs and/or indels related to soybean growth and seed yield. Moreover, Zhou et al. [39] also detected a genetic loci associated with seed weight, which was consistent with the previously reported *qSW* locus on chromosome 17 [88]. There are four genes and 13 QTL related to soybean seed weight (100-SW) reported on chromosome 9, however, the physical distance gap between *SW9-1* and its nearest locus (*Sdwt 15-6*) is >600 kb [24]. The physical location of the *SW9-1* locus was not overlapped with a known QTL or a gene interval listed in SoyBase (http://www.soybase.org/). Therefore, the *SW9-1T/C* locus represents a novel locus.

The fact that soybean 100-SW is regulated by multiple genes and is influenced by various environmental conditions makes it difficult to improve soybean yield *via* traditional breeding methods, however, the development of modern molecular biology technologies (genomics, gene editing technology, molecular design breeding (MDB), *etc.*) has provided alternative tools to enhance breeding efficiency in soybean breeding programs [2,9–11]. Compared with other breeding techniques, MDB can precisely modify agronomic traits at the genetic level, and avoid the problem associated with traditional breeding methods regarding the transfer of undesirable genes. Thus, MDB can greatly increase breeding efficiency and shorten the breeding cycle [93]. Previous studies confirmed that highly consistent marker loci across environments can be used for MDB to enhance soybean seed weight [12]. Therefore, applying specific genes or allele variations for soybean breeding *via* MDB may be an effective way to genetically improve soybean seed size and yield. This method has been successfully applied to enhance chilling tolerance, nitrogen use efficiency, grain shape, yield and grain quality, and grain size and yield in rice [94–97]. In this study, *SW9-1* (*SW9-1T*) locus was linked with the reported large-effect QTL to form a "molecular module" according to the MDB theme, which may be relevant for the breeding of new soybean varieties that produce large seeds, as well as for elucidating the theoretical basis for the genetic improvement and prediction of soybean seed weight in different ecoregions.

Supplementary data for this article can be found online at https://doi.org/10.1016/j.cj.2018.12.010.

Conflict of interest

Authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2016YFD0100201), the National Natural Science Foundation of China (31771819), the China Postdoctoral Science Foundation (2017M621990), the Introduced Leading Talent Research Team for Universities in Anhui Province, and the Natural Science Foundation of Anhui Province, China (1608085QC66).

REFERENCES

[1] Y. Xu, H.N. Li, G.J. Li, X. Wang, L.G. Cheng, Y.M. Zhang, Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* (L.) Merr.), Theor. Appl. Genet. 122 (2011) 581–594.
[2] X.B. Wang, Y.H. Li, H.W. Zhang, G.L. Sun, W.M. Zhang, L.J. Qiu, Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean, Mol. Biol. Rep. 42 (2015) 489–496.
[3] J.S. Burris, O.T. Edje, A.H. Wahab, Effects of seed size on seedling performance in soybeans: II. Seedling growth and photosynthesis and field performance, Crop Sci. 13 (1973) 207–210.
[4] T.J. Smith, H.M. Camper, Effect of seed size on soybean performance, Agron. J. 67 (1975) 681–684.
[5] J.W. Burton, Quantitative genetics: results relevant to soybean breeding, in: J.R. Wilcox (Ed.), Soybeans: Improvement, Production and Uses, 2nd editionAmerican Society of Agronomy, Madison, WI, USA 1987, pp. 211–247.
[6] J.P. Zhang, Q.J. Song, P.B. Cregan, G.L. Jiang, Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*), Theor. Appl. Genet. 129 (2016) 117–130.
[7] H.Z. Liang, W.D. Li, H. Wang, X.J. Fang, Genetic effects on seed traits in soybean, Acta Genet. Sin. 32 (2005) 1199–1204 (in Chinese with English abstract).
[8] W.L. Teng, Y.P. Han, Y.P. Du, D.S. Sun, Z.C. Zhang, L.J. Qiu, G.L. Sun, W.B. Li, QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.), Heredity 102 (2009) 372–380.

[9] L.M. Mansur, J.H. Orf, K.G. Lark, Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* (L). Merr.), Theor. Appl. Genet. 86 (1993) 914–918.
[10] J.Z. Li, X.Q. Huang, F. Heinrichs, M.W. Ganal, M.S. Röder, Analysis of QTLs for yield, yield components, and malting quality in a BC_3-DH population of spring barley, Theor. Appl. Genet. 110 (2005) 356–363.
[11] D.R. Hao, H. Cheng, Z.T. Yin, S.Y. Cui, D. Zhang, H. Wang, D.Y. Yu, 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 (2012) 447–458.
[12] M.A.R. Mian, M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, H.R. Boerma, Molecular markers associated with seed weight in two soybean populations, Theor. Appl. Genet. 93 (1996) 1011–1016.
[13] Y.N. Sun, J.B. Pan, X.L. Shi, X.Y. Du, Q. Wu, Z.M. Qi, H.W. Jiang, D.W. Xin, C.Y. Liu, G.H. Hu, Q.S. Chen, Multi-environment mapping and meta-analysis of 100-seed weight in soybean, Mol. Biol. Rep. 39 (2012) 9435–9443.
[14] F.L. Allen, Usefulness of plant genome mapping to plant breeding, in: P. Gresshoff (Ed.), Plant Genome Analysis, CRC Press, Boca Raton, Florida, USA 1994, pp. 11–18.
[15] Q.J. Song, L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, P.B. Cregan, A new integrated genetic linkage map of the soybean, Theor. Appl. Genet. 109 (2004) 122–128.
[16] Q.J. Song, G.F. Jia, Y.L. Zhu, D. Grant, R.T. Nelson, E.Y. Hwang, D.L. Hyten, P.B. Cregan, Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean, Crop Sci. 50 (2010) 1950–1960.
[17] Q.J. Song, J. Jenkins, G.F. Jia, D.L. Hyten, V. Pantalone, S.A. Jackson, J. Schmutz, P.B. Cregan, Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly *Glyma1.01*, BMC Genomics 17 (2016) 3.
[18] I.Y. Choi, D.L. Hyten, L.K. Matukumalli, Q.J. Song, J.M. Chaky, C.V. Quigley, K. Chase, K.G. Lark, R.S. Reiter, M.S. Yoon, E.Y. Hwang, S.I. Yi, N.D. Young, C.P. Randy, C. van Tassell, E.J. Specht, P.B. Cregan, A soybean transcript map: gene distribution, haplotype and single nucleotide polymorphism analysis, Genetics 176 (2007) 685–696.
[19] T.Y. Hwang, T. Sayama, M. Takahashi, Y. Takada, Y. Nakamoto, H. Funatsuki, H. Hisano, S. Sasamoto, S. Sato, S. Tabata, I. Kono, M. Hoshi, M. Hanawa, C. Yano, Z. Xia, K. Harada, K. Kitamura, M. Ishimoto, High-density integrated linkage map based on SSR markers in soybean, DNA Res. 16 (2009) 213–225.
[20] D.L. Hyten, I.Y. Choi, Q.J. Song, J.E. Specht, T.E. Carter Jr., R.C. Shoemaker, E.Y. Hwang, L.K. Matukumalli, P.B. Cregan, A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping, Crop Sci. 50 (2010) 960–968.
[21] M. Akond, S.M. Liu, L. Schoener, J.A. Anderson, S.K. Kantartzi, K. Meksem, Q.J. Song, D.C. Wang, Z.X. Wen, A SNP-based genetic linkage map of soybean using the SoySNP6K Illumina infinium bead chip genotyping array, J. Plant Genome Sci. 1 (2013) 80–89.
[22] B. Li, L. Tian, J.Y. Zhang, L. Huang, F.X. Han, S.R. Yan, L.Z. Wang, Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in *Glycine max*, BMC Genomics 15 (2014) 1086.
[23] S. Lee, K.R. Freewalt, L.K. Mchal, Q.J. Song, T.H. Jun, A.P. Michel, A.E. Dorrance, M.A.R. Mian, A high-resolution genetic linkage map of soybean based on 357 recombinant inbred lines genotyped with BARCSoySNP6K, Mol. Breed. 35 (2015) 28.
[24] D.L. Hyten, V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, M.E. Schmidt, Seed quality QTL in a prominent soybean population, Theor. Appl. Genet. 109 (2004) 552–561.
[25] Y.P. Han, D.M. Li, D. Zhu, H.Y. Li, X.P. Li, W.L. Teng, W.B. Li, QTL analysis of soybean seed weight across multi-genetic backgrounds and environments, Theor. Appl. Genet. 125 (2012) 671–683.
[26] J.H. Orf, K. Chase, F.R. Adler, L.M. Mansur, K.G. Lark, Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean, Crop Sci. 39 (1999) 1652–1657.
[27] M.D. Smalley, W.R. Fehr, S.R. Cianzio, F. Han, S.A. Sebastian, L.G. Streit, Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm, Crop Sci. 44 (2004) 436–442.
[28] P. Salas, J.C. Oyarzo-Llaipen, D. Wang, K. Chase, L. Mansur, Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max* L. Merr.), Theor. Appl. Genet. 113 (2006) 1459–1466.
[29] J.Y. Gai, Y.J. Wang, X.L. Wu, S.Y. Chen, A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean, Front. Agric. China 1 (2007) 1–7.
[30] P. Guzman, B. Neece, D.J.S. Martin, S. LeRoy, A. Grau, C. Hughes, T. Nelson, QTL associated with yield in three backcross-derived populations of soybean, Crop Sci. 47 (2007) 111–122.
[31] Y.H. Li, R.X. Guan, Z.X. Liu, Y.S. Ma, L.X. Wang, L.H. Li, F.Y. Lin, W.J. Luan, P.Y. Chen, Z. Yan, Y. Guan, L. Zhu, X.C. Ning, M.J.M. Smulders, W. Li, R.H. Piao, Y.H. Cui, Z.M. Yu, M. Guan, R.Z. Chang, A.F. Hou, A.N. Shi, B. Zhang, S.L. Zhu, L.J. Qiu, Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China, Theor. Appl. Genet. 117 (2008) 857–871.
[32] Y.H. Li, W. Li, C. Zhang, L. Yang, R.Z. Chang, B.S. Gaut, L.J. Qiu, 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 (2010) 242–253.
[33] Y.H. Li, S.C. Zhao, J.X. Ma, D. Li, L. Yan, J. Li, X.T. Qi, X.S. Guo, L. Zhang, W.M. He, R.Z. Chang, Q.S. Liang, Y. Guo, C. Ye, X.B. Wang, Y. Tao, R.X. Guan, J.Y. Wang, Y.L. Liu, L.G. Jin, X.Q. Zhang, Z.X. Liu, L.J. Zhang, J. Chen, K.J. Wang, R. Nielsen, R.Q. Li, P.Y. Chen, W.B. Li, J.C. Reif, M. Purugganan, J. Wang, M.C. Zhang, J. Wang, L.J. Qiu, Molecular footprints of domestication and improvement in soybean revealed by whole genome re-sequencing, BMC Genomics 14 (2013) 579.
[34] Y.H. Li, G.Y. Zhou, J.X. Ma, W.K. Jiang, L.G. Jin, Z.H. Zhang, Y. Guo, J.B. Zhang, Y. Sui, L.T. Zheng, S.S. Zhang, Q.Y. Zuo, X.H. Shi, Y.F. Li, W.K. Zhang, Y.Y. Hu, G.Y. Kong, H.L. Hong, B. Tan, J. Song, Z.X. Liu, Y.S. Wang, H. Ruan, C.K.L. Yeung, J. Liu, H.L. Wang, L.J. Zhang, R.X. Guan, K.J. Wang, W.B. Li, S.Y. Chen, R.Z. Chang, Z. Jiang, S.A. Jackson, R.Q. Li, L.J. Qiu, *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits, Nat. Biotechnol. 32 (2014) 1045–1052.
[35] J.K. Yang, M.K. Wak, M.Y. Kim, K. Van, S.H. Lee, Association analysis of candidate gene for seed weight regulation in soybean, Br. J. Pharmacol. 169 (2009) 462–476.
[36] Y. Niu, Y. Xu, X.F. Liu, S.X. Yang, S.P. Wei, F.T. Xie, Y.M. Zhang, Association mapping for seed size and shape traits in soybean cultivars, Mol. Breed. 31 (2013) 785–794.
[37] T.H. Jun, K. Freewalt, A.P. Michel, R. Mian, Identification of novel QTL for leaf traits in soybean, Plant Breed. 133 (2014) 61–66.

[38] S. Kato, T. Sayama, K. Fujii, S. Yumoto, Y. Kono, T.Y. Hwang, A. Kikuchi, Y. Takada, Y. Tanaka, T. Shiraiwa, M. Ishimoto, A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds, Theor. Appl. Genet. 127 (2014) 1365–1374.
[39] Z.K. Zhou, Y. Jiang, Z. Wang, Z.H. Guo, J. Lyu, W.Y. Li, Y.J. Yu, L.P. Shu, Y.J. Zhao, Y.M. Ma, C. Fang, Y.T. Shen, T.F. Liu, C.C. Li, Q. Li, M. Wu, M. Wang, Y.S. Wu, Y. Dong, W.T. Wan, X. Wang, Z.L. Ding, Y.D. Gao, H. Xiang, B.G. Zhu, S.H. Lee, W. Wang, Z.X. Tian, Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean, Nat. Biotechnol. 33 (2015) 408–414.
[40] T.H.E. Meuwissen, B.J. Hayes, M.E. Goddard, Prediction of total genetic value using genome-wide dense marker maps, Genetics 157 (2001) 1819–1829.
[41] S. Atwell, Y.S. Huang, B.J. Vilhjálmsson, G. Willems, M. Horton, Y. Li, D.Z. Meng, A. Platt, A.M. Tarone, T.T. Hu, R. Jiang, N.W. Muliyati, X. Zhang, M.A. Amer, I. Baxter, B. Brachi, J. Chory, C. Dean, M. Debieu, J. de Meaux, J.R. Ecker, N. Faure, J. M. Kniskern, J.D.G. Jones, T. Michael, A. Nemri, F. Roux, D.E. Salt, C.L. Tang, M. Todesco, B. Traw, D. Weigel, P. Marjoram, J. O. Borevitz, J. Bergelson, M. Nordborg, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines, Nature 465 (2010) 627–631.
[42] J.D. Burridge, H.M. Schneider, B.L. Huynh, P.A. Roberts, A. Bucksch, J.P. Lynch, Genome-wide association mapping and agronomic impact of cowpea root architecture, Theor. Appl. Genet. 130 (2017) 419–431.
[43] Q.H. Zhou, D.P. Han, A.S. Mason, C. Zhou, W. Zheng, Y.Z. Li, C. J. Wu, D.H. Fu, Y.J. Huang, Earliness traits in rapeseed (*Brassica napus*): SNP loci and candidate genes identified by genome-wide association analysis, DNA Res. 25 (2017) 229–244.
[44] Y.Y. Shi, L.L. Gao, Z.C. Wu, X.J. Zhang, M.M. Wang, C.S. Zhang, F. Zhang, Y.L. Zhou, Z.K. Li, Genome-wide association study of salt tolerance at the seed germination stage in rice, BMC Plant Biol. 17 (2017) 92.
[45] J.A. Rafalski, Association genetics in crop improvement, Curr. Opin. Plant Biol. 13 (2010) 174–180.
[46] X.H. Huang, X.G. Wei, T. Sang, Q. Zhao, Q. Feng, Y. Zhao, C.R. Zhu, T.T. Lu, Z.W. Zhang, M. Li, D.L. Fan, Y.L. Guo, A.H. Wang, L. Wang, L.W. Deng, W.J. Li, Y.Q. Lu, Q.J. Weng, K.Y. Liu, T. Huang, T.Y. Zhou, Y.F. Jing, W. Li, Z. Lin, E.S. Buckler, Q. Qian, Q.F. Zhang, J.Y. Li, B. Han, Genome-wide association studies of 14 agronomic traits in rice landraces, Nat. Genet. 42 (2010) 961–967.
[47] J.S. Lai, R.Q. Li, X. Xu, W.W. Jin, M.L. Xu, H.N. Zhao, Z.K. Xiang, W.B. Song, K. Ying, M. Zhang, Y.P. Jiao, P.X. Ni, J.G. Zhang, D. Li, X.S. Guo, K.X. Ye, M. Jian, B. Wang, H.S. Zheng, H.Q. Liang, X.Q. Zhang, S.C. Wang, S.J. Chen, J.S. Li, Y. Fu, N.M. Springer, H.M. Yang, J. Wang, J.R. Dai, P.S. Schnable, J. Wang, Genome-wide patterns of genetic variation among elite maize inbred lines, Nat. Genet. 42 (2010) 1027–1030.
[48] H. Li, Z.Y. Peng, X.H. Yang, W.D. Wang, J.J. Fu, J.H. Wang, Y.J. Han, Y.C. Chai, T.T. Guo, N. Yang, J. Liu, M.L. Warburton, Y.B. Cheng, X.M. Hao, P. Zhang, J.Y. Zhao, Y.J. Liu, G.Y. Wang, J.S. Li, J.B. Yan, Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels, Nat. Genet. 45 (2013) 43–50.
[49] G.P. Morris, P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, O. Riera-Lizarazu, P.J. Brown, C.B. Acharya, S.E. Mitchell, J. Harriman, J.C. Glaubitz, E.S. Buckler, S. Kresovich, Population genomic and genome-wide association studies of agroclimatic traits in sorghum, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 453–458.
[50] E.Y. Hwang, Q.J. Song, G.F. Jia, J.E. Specht, D.L. Hyten, J. Costa, P.B. Cregan, A genome-wide association study of seed protein and oil content in soybean, BMC Genomics 15 (2014) 1.
[51] Z.X. Wen, R.J. Tan, J.Z. Yuan, C. Bales, W.Y. Du, S.C. Zhang, M. I. Chilvers, C. Schmidt, Q.J. Song, P.B. Cregan, D.C. Wang, Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean, BMC Genomics 15 (2014) 809.
[52] J.P. Zhang, Q.J. Song, P.B. Cregan, R.L. Nelson, X.Z. Wang, J.X. Wu, G.L. Jiang, Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm, BMC Genomics 16 (2015) 217.
[53] Y.H. Zhang, J.B. He, Y.F. Wang, G.N. Xing, J.M. Zhao, Y. Li, S.P. Yang, R.G. Palmer, T.J. Zhao, J.Y. Gai, Establishment of a 100-seed weight quantitative trait locus-allele matrix of the germplasm population for optimal recombination design in soybean breeding programmes, J. Exp. Bot. 66 (2015) 6311–6325.
[54] K.P. Kulkarni, S. Asekova, D.H. Lee, K. Bilyeu, J.T. Song, J.D. Lee, Mapping QTLs for 100-seed weight in an interspecific soybean cross of Williams 82 (*Glycine max*) and PI 366121 (*Glycine soja*), Crop Pasture Sci. 68 (2017) 148–155.
[55] X. Lu, Q. Xiong, T. Cheng, Q.T. Li, X.L. Liu, Y.D. Bi, W. Li, W.K. Zhang, B. Ma, Y.C. Lai, W.G. Du, W.Q. Man, S.Y. Chen, J.S. Zhang, A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight, Mol. Plant 10 (2017) 670–684.
[56] N.M. Adamski, E. Anastasiou, S. Eriksson, C.M. O'Neill, M. Lenhard, Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 20115–20120.
[57] W.J. Fang, Z.B. Wang, R.F. Cui, J. Li, Y.H. Li, Maternal control of seed size by *EOD3/CYP78A6* in *Arabidopsis thaliana*, Plant J. 70 (2012) 929–939.
[58] J.W. Wang, R. Schwab, B. Czech, E. Mica, D. Weigel, Dual effects of miR156-targeted *SPL* genes and *CYP78A5/KLUH* on plastochron length and organ size in *Arabidopsis thaliana*, Plant Cell 20 (2008) 1231–1243.
[59] L. Du, S.D. Wang, C.M. He, B. Zhou, Y.L. Ruan, H.X. Shou, Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis, J. Exp. Bot. 68 (2017) 1955–1972.
[60] H.W. Kang, Y.G. Cho, U.H. Yoon, M.Y. Eun, A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed, Plant Mol. Biol. Report. 16 (1998) 90.
[61] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[62] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 14 (2005) 2611–2620.
[63] F. Breseghello, M.E. Sorrells, Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars, Genetics 172 (2006) 1165–1177.
[64] H. Funatsuki, K. Kawaguchi, S. Matsuba, Y. Sato, M. Ishimoto, Mapping of QTL associated with chilling tolerance during reproductive growth in soybean, Theor. Appl. Genet. 111 (2005) 851–861.
[65] X.H. Yang, J.B. Yan, T. Shah, M.L. Warburton, Q. Li, L. Li, Y.F. Gao, Y.C. Chai, Z.Y. Fu, Y. Zhou, S.T. Xu, G.H. Bai, Y.J. Meng, Y. P. Zheng, J.S. Li, Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection, Theor. Appl. Genet. 121 (2010) 417–431.
[66] O.J. Hardy, X. Vekemans, SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels, Mol. Ecol. Notes 2 (2002) 618–620.
[67] P.J. Bradbury, Z.W. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler, TASSEL: software for association mapping of complex traits in diverse samples, Bioinformatics 23 (2007) 2633–2635.

[68] J. Yu, G. Pressoir, W.H. Briggs, B.I. Vroh, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, E.S. Buckler, A unified mixed-model method for association mapping that accounts for multiple levels of relatedness, Nat. Genet. 38 (2006) 203–208.
[69] E.H. Roberts, A. Qi, R.H. Ellis, R.J. Summerfield, R.J. Lawn, S. Shanmugasundaram, Use of field observations to characterize genotypic flowering responses to photoperiod and temperature: a soybean exemplar, Theor. Appl. Genet. 93 (1996) 519–533.
[70] Y.L. Lu, J.B. Yan, C.T. Guimares, S. Taba, Z.F. Hao, B.S. Gao, J.H. Chen, J. Li, J.S. Zhang, B.S. Vivek, C. Magorokosho, S. Mugo, D. Makumbi, S.N. Parentoni, T. Shah, T.Z. Rong, J.H. Crouch, Y.B. Xu, Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms, Theor. Appl. Genet. 120 (2009) 93–115.
[71] J.B. Yan, X.H. Yang, T. Shah, H. Sánchez-Villeda, J.S. Li, M. Warburton, Y. Zhou, J.H. Crouch, Y.B. Xu, High-throughput SNP genotyping with the GoldenGate assay in maize, Mol. Breed. 25 (2010) 441–451.
[72] W.W. Wen, S. Taba, T. Shah, T.V.H. Chavez, J.B. Yan, Detection of genetic integrity of conserved maize (*Zea mays* L.) germplasm in genebanks using SNP markers, Genet. Resour. Crop. Evol. 58 (2011) 189–207.
[73] D.L. Hyten, I.Y. Choi, Q.J. Song, R.C. Shoemaker, R.L. Nelson, J. M. Costa, J.E. Specht, P.B. Cregan, Highly variable patterns of linkage disequilibrium in multiple soybean populations, Genetics 175 (2007) 1937–1944.
[74] J.M. Yu, Z.W. Zhang, C.S. Zhu, D.A. Tabanao, G. Pressoir, M.R. Tuinstra, S. Kresovich, R.J. Todhunter, E.S. Buckler, Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping, Plant Genome 2 (2009) 63–77.
[75] D. Van Inghelandt, A. Melchinger, C. Lebreton, B. Stich, Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers, Theor. Appl. Genet. 120 (2010) 1289–1299.
[76] Y.L. Zhu, S.X. Wang, H.P. Zhang, L.X. Zhao, Z.Y. Wu, H. Jiang, J. J. Cao, K. Liu, M. Qin, J. Lu, G.L. Sun, X.C. Xia, C. Chang, C.X. Ma, Identification of major loci for seed dormancy at different post-ripening stages after harvest and validation of a novel locus on chromosome 2AL in common wheat, Mol. Breed. 36 (2016) 174.
[77] T. Yanagisawa, C. Kiribuchi-Otobe, H. Hirano, Y. Suzuki, M. Fujita, Detection of single nucleotide polymorphism (SNP) controlling the waxy character in wheat by using a derived cleaved amplified polymorphic sequence (dCAPS) marker, Theor. Appl. Genet. 107 (2003) 84–88.
[78] Y.Y. Zhang, Z.Y. Fang, Q.B. Qang, Y.M. Liu, L.M. Yang, M. Zhuang, P.T. Sun, Chloroplast subspecies-specific SNP detection and its maternal inheritance in *Brassica oleracea* L. by using a dCAPS marker, J. Hered. 103 (2012) 606–611.
[79] H. Di, X.J. Liu, Q.K. Wang, J.F. Weng, L. Zhang, X.H. Li, Z.H. Wang, Development of SNP-based dCAPS markers linked to major head smut resistance quantitative trait locus *qHS2.09* in maize, Euphytica 202 (2015) 69–79.
[80] F.S. Song, J.L. Ni, Y.L. Qian, L. Li, D.H. Ni, J.B. Yang, Development of SNP-based dCAPS markers for identifying male sterile gene *tms5* in two-line hybrid rice, Genet. Mol. Res. 15 (2016) (gmr.15038512).
[81] G.A. Lee, H.J. Koh, H.K. Chung, A. Dixit, J.W. Chung, K.H. Ma, S. Y. Lee, J.R. Lee, G.S. Lee, J.G. Gwag, T.S. Kim, Y.J. Park, Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice, Mol. Breed. 24 (2009) 93–101.
[82] A. Bungartz, M. Klaus, B. Mathew, J. Léon, A.A. Naz, Development of new SNP derived cleaved amplified polymorphic sequence marker set and its successful utilization in the genetic analysis of seed color variation in barley, Genomics 107 (2016) 100–107.
[83] E.C. Brummer, G.L. Graef, J. Orf, J.R. Wilcox, R.C. Shoemaker, Mapping QTL for seed protein and oil content in eight soybean populations, Crop Sci. 37 (1997) 370–378.
[84] T.M. Fulton, T. Beck-Bunn, D. Emmatty, Y. Eshed, J. Lopez, V. Petiard, J. Uhlig, D. Zamir, S.D. Tanksley, QTL analysis of an advanced backcross of *Lycopersicon peruvianum* to the cultivated tomato and comparisons with QTL found in other wild species, Theor. Appl. Genet. 95 (1997) 881–894.
[85] N. Li, Y.H. Li, Maternal control of seed size in plants, J. Exp. Bot. 66 (2015) 1087–1097.
[86] V.M. Davis, K.D. Gibson, T.T. Bauman, S.C. Weller, W.G. Johnson, Influence of weed management practices and crop rotation on glyphosate-resistant horse weed (*Conyza canadensis*) population dynamics and crop yield years III and IV, Weed Sci. 57 (2009) 417–426.
[87] F. Huang, Y.J. Chi, J.Y. Gai, D.Y. Yu, Identification of transcription factors predominantly expressed in soybean flowers and characterization of *GmSEP1* encoding a SEPALLATA1-like protein, Gene 438 (2009) 40–48.
[88] B.H. Liu, T. Fujita, Z.H. Yan, S. Sakamoto, D.H. Xu, J. Abe, QTL mapping of domestication-related traits in soybean (*Glycine max*), Ann. Bot. 100 (2007) 1027–1038.
[89] Z.X. Tian, X.B. Wang, R. Lee, Y.H. Li, J.E. Spechtd, R.L. Nelsone, P.E. McCleanc, L.J. Qiu, J.X. Ma, Artificial selection for determinate growth habit in soybean, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 8563–8568.
[90] Z.B. Hu, H.R. Zhang, G.Z. Kan, D.Y. Ma, D. Zhang, G.X. Shi, D.L. Hong, G.Z. Zhang, D.Y. Yu, Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.), Genetica 141 (2013) 247–254.
[91] F.T. Xie, Y. Niu, J. Zhang, S.H. Bu, H.Z. Zhang, Q.C. Geng, J.Y. Feng, Y.M. Zhang, Fine mapping of quantitative trait loci for seed size traits in soybean, Mol. Breed. 34 (2014) 2165–2178.
[92] S. Manan, M.Z. Ahmad, G.Y. Zhang, B.B. Chen, B.U. Haq, J.H. Yang, J. Zhao, Soybean *LEC2* regulates subsets of genes involved in controlling the biosynthesis and catabolism of seed storage substances and seed development, Front. Plant Sci. 8 (2017) 1604.
[93] X.Z. Feng, B.H. Liu, S.X. Yang, Progress and perspective of soybean molecular design breeding research, Soil Crop 3 (2014) 123–131 (in Chinese with English abstract).
[94] Y. Ma, X.Y. Dai, Y.Y. Xu, W. Luo, X.M. Zheng, D.L. Zeng, Y.J. Pan, X.L. Lin, H.H. Liu, D.J. Zhang, J. Xiao, X.Y. Guo, S.J. Xu, Y.D. Niu, J.B. Jin, H. Zhang, X. Xu, L.G. Li, W. Wang, Q. Qian, S. Ge, K. Chong, *COLD1* confers chilling tolerance in rice, Cell 160 (2015) 1209–1221.
[95] B. Hu, W. Wang, S.J. Ou, J.Y. Tang, H. Li, R.H. Che, Z.H. Zhang, X.Y. Chai, H.R. Wang, Y.Q. Wang, C.Z. Liang, L.C. Liu, Z.Z. Piao, Q.Y. Deng, K. Deng, C. Xu, Y. Liang, L.H. Zhang, L.G. Li, C.C. Chu, Variation in *NRT1.1B* contributes to nitrate-use divergence between rice subspecies, Nat. Genet. 47 (2015) 834–838.
[96] S.K. Wang, S. Li, Q. Liu, K. Wu, J.Q. Zhang, S.S. Wang, Y. Wang, X.B. Chen, Y. Zhang, C.X. Gao, F. Wang, H.X. Huang, X.D. Fu, The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality, Nat. Genet. 47 (2015) 949–954.
[97] L.Z. Si, J.Y. Chen, X.H. Huang, H. Gong, J.H. Luo, Q.Q. Hou, T.Y. Zhou, T.T. Lu, J.J. Zhu, Y.Y. Shanguan, E.W. Chen, C.X. Gong, Q. Zhao, Y.F. Jing, Y. Zhao, Y. Li, L.L. Cui, D.L. Fan, Y.Q. Lu, Q.J. Weng, Y.C. Wang, Q.L. Zhan, K.Y. Liu, X.H. Wei, K. An, G. An, B. Han, *OsSPL13* controls grain size in cultivated rice, Nat. Genet. 48 (2016) 447–456.