



Comparison of Genetic Diversity between Chinese and American Soybean (*Glycine max* (L.)) Accessions Revealed by High-Density SNPs

Zhangxiong Liu^{1†}, Huihui Li^{1†}, Zixiang Wen^{2†}, Xuhong Fan³, Yinghui Li¹, Rongxia Guan¹, Yong Guo¹, Shuming Wang³, Dechun Wang^{2*} and Lijuan Qiu^{1*}

¹ National Key Facility for Gene Resources and Genetic Improvement, Key Laboratory of Crop Germplasm Utilization, Ministry of Agriculture, Institute of Crop Sciences, Chinese Academy of Agricultural Science, Beijing, China, ² Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States, ³ Institute of Soybean Research, Jilin Academy of Agricultural Sciences, Changchun, China

OPEN ACCESS

Edited by:

Guo-Bo Chen,
Zhejiang Provincial People's Hospital,
China

Reviewed by:

Suhong Bu,
Fujian Agriculture and Forestry
University, China
Yuan-Ming Zhang,
Huazhong Agricultural University,
China

*Correspondence:

Dechun Wang
wangdech@msu.edu
Lijuan Qiu
qiu lijuan@caas.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 06 August 2017

Accepted: 13 November 2017

Published: 30 November 2017

Citation:

Liu Z, Li H, Wen Z, Fan X, Li Y,
Guan R, Guo Y, Wang S, Wang D and
Qiu L (2017) Comparison of Genetic
Diversity between Chinese and
American Soybean (*Glycine max* (L.))
Accessions Revealed by High-Density
SNPs. *Front. Plant Sci.* 8:2014.
doi: 10.3389/fpls.2017.02014

Soybean is one of the most important economic crops for both China and the United States (US). The exchange of germplasm between these two countries has long been active. In order to investigate genetic relationships between Chinese and US soybean germplasm, 277 Chinese soybean accessions and 300 US soybean accessions from geographically diverse regions were analyzed using 5,361 SNP markers. The genetic diversity and the polymorphism information content (PIC) of the Chinese accessions was higher than that of the US accessions. Population structure analysis, principal component analysis, and cluster analysis all showed that the genetic basis of Chinese soybeans is distinct from that of the USA. The groupings observed in clustering analysis reflected the geographical origins of the accessions; this conclusion was validated with both genetic distance analysis and relative kinship analysis. F_{ST} -based and EigenGWAS statistical analysis revealed high genetic variation between the two subpopulations. Analysis of the 10 loci with the strongest selection signals showed that many loci were located in chromosome regions that have previously been identified as quantitative trait loci (QTL) associated with environmental-adaptation-related and yield-related traits. The pattern of diversity among the American and Chinese accessions should help breeders to select appropriate parental accessions to enhance the performance of future soybean cultivars.

Keywords: soybean, single nucleotide polymorphism (SNP), genetic diversity, population structure, US soybean accessions, Chinese soybean accessions

INTRODUCTION

Soybean originated in China, and has a history of planting for more than 4,000 years (Hymowitz and Newell, 1981). It was introduced to the USA in 1765 as animal feed firstly (Hymowitz and Harlan, 1983) and has been widely planted in the USA since 1922. To date, soybean is one of the most important economic crops for both China and the USA.

There are extensive soybean breeding programs in both China and the USA, most of which rely on a genetic base of Chinese origin (Cui et al., 2001). There have been over 400 publicly released cultivars in the USA, derived from ~80 soybean ancestral lines (Gizlice et al., 1994), most of which were originally introduced from China in the early twentieth century (Li et al., 2001). In addition, during past decades, the exchange and utilization of soybean germplasm between the two countries has been and continues to be active. The Chinese Gene Bank contains 28,580 soybean accessions, of which 1,718 (6.01%) were introduced from the USA. There are 14,330 soybean germplasm accessions that are conserved in USDA Soybean Germplasm Collection in the USA, of which 5,216 (36.40%) are from China. Some elite cultivars from North America were used in Chinese soybean breeding programs for broadening the genetic base of modern soybean cultivars. During the period from 1923 to 2005, 1,300 soybean cultivars were released in China; according to a study of the pedigree of these cultivars, the genetic contribution from US cultivars to Chinese cultivars was 139.83, accounting for 10.76% of the total (Gai et al., 2015). Moreover, Carter et al. (2000) suggested that Chinese cultivars should also be viewed as an important reservoir of genetic diversity that can be used to yet further expand the genetic base for North American soybean breeding efforts. Therefore, considering the ongoing increases in exchange and utilization of soybean germplasm between the two countries, assessing genetic relationships is a worthy area for further study.

The pattern of genetic variation in soybean germplasm resources between China and USA has been evaluated through pedigree information (Cui et al., 2000a,b), phenotypic observation (Cui et al., 2001), and low-density polymorphism markers (Qiu et al., 1997; Carter et al., 2000; Li et al., 2001). Gizlice et al. (1994) analyzed the pedigrees of 258 publicly developed cultivars in US released between 1947 and 1988, and determined that >84% of their parentage could be traced back to only 17 ancestors, and most of which were selected from the introduced landraces from China. Based on the pedigree analysis, Cui et al. (2000a,b) found that accessions from the two countries had few identifiable ancestors in common probably because of different environments and breeders' selection preference between two countries. Using 25 biochemical, morphological, and agronomic traits, Cui et al. (2001) found much more extensive phenotypic diversity among the Chinese cultivars than among North America cultivars, with obvious distinctness between the two groups. Using evaluation based on random amplified polymorphic DNA (RAPD) markers, Li et al. (2001) demonstrated that the ancestors of soybean cultivars from North America and China were clearly different. To date, previous reported results in this research area have been based on a relatively small number of progenitor soybean cultivars using relative low-density molecular markers.

Significant progress has been made using high throughput genotyping technologies to detect variability in DNA sequences, and these technologies are now used regularly in crop germplasm research and breeding (Akond et al., 2013; Li et al., 2014; Lee et al., 2015; Zhou et al., 2015, 2016; Han et al., 2016; Wang et al., 2016; Chang and Hartman, 2017; Fang et al., 2017; Liu et al.,

2017a). In the present study, a total of 577 soybean cultivars or advanced breeding lines from the main production areas of China and the USA were analyzed using the 5,361 SNP markers on the Illumina SoySNP6k iSelect BeadChip. The objectives were (i) to compare the genetic diversity between the American and Chinese soybeans; (ii) to investigate the population structures for these subpopulations; and (iii) to identify loci that have undergone selection based on differences in allele frequencies between the subpopulations. The results in present study will provide useful information to future soybean breeding for both China and the USA.

MATERIALS AND METHODS

Plant Materials

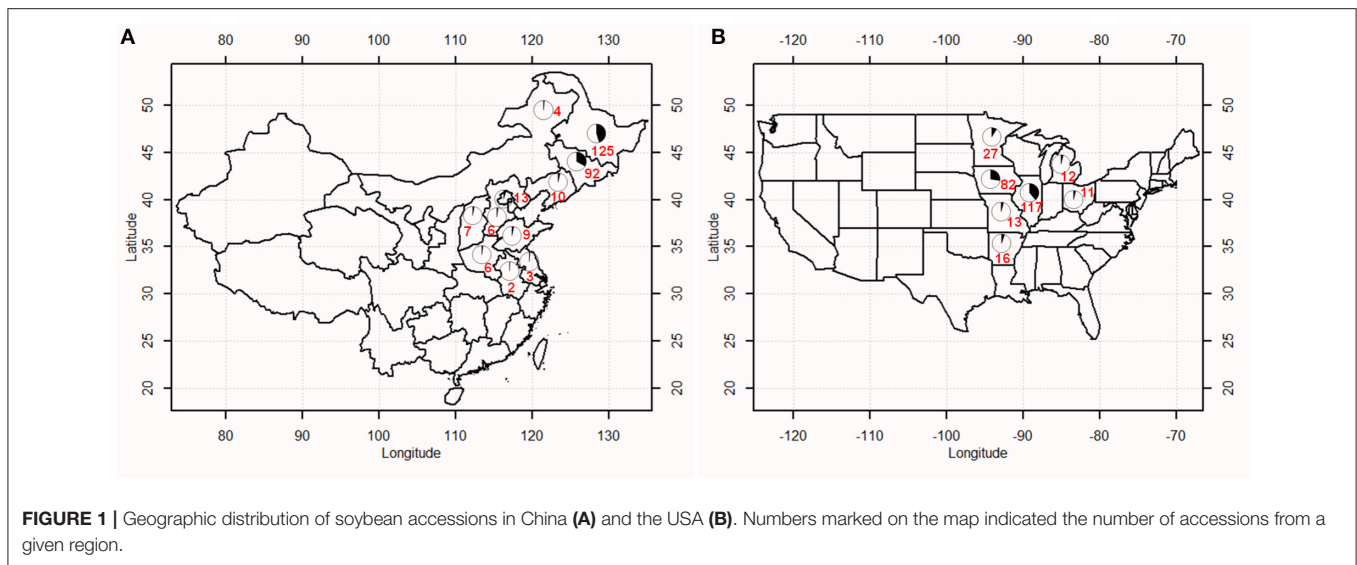
This study examined soybean cultivars and advanced breeding lines: 277 from China (hereafter termed as CN-set) and 300 from the USA (hereafter termed as US-set). Of the 277 Chinese accessions collected from 11 provinces, 231 accessions were derived from the Northern spring (Nsp) ecotype and 46 were derived from the Huang-huai-hai summer (Hsu) ecotype. Detailed information on the 277 accessions can be found in Supplementary Table S1. The 300 diverse accessions of US-set represent a range of materials developed by public breeders in the North Central soybean production area of the USA (Wen et al., 2014); detailed information on these 300 accessions was listed in Supplementary Table S2. The geographic distribution of the 577 soybean accessions was presented in **Figure 1**.

DNA Extraction and SNP Genotyping

Genomic DNA was extracted from soybean seedlings (leaf) following the protocol presented by Kisha et al. (1997). All of the accessions in the present study were genotyped with the Illumina SoySNP6k iSelect BeadChip (Illumina, USA), which can be used to distinguish different genotypes among 5,361 SNPs (Akond et al., 2013); the chromosomal distributions and quality control for these SNPs was demonstrated in Wen et al. (2014).

Basic Population Genetic Parameters

Using software of PowerMarker 3.25 (Liu and Muse, 2005), minor allele frequency (MAF), genetic diversity, polymorphism information content (PIC), and heterozygosity, were evaluated for each population pool (i.e., the CN-set or the US-set) across the soybean genome. The average and the range of the Roger distance within and between subpopulations, and across all genotypes, were calculated. The degree of linkage disequilibrium (LD) was evaluated by LD parameter r^2 , calculated by TASSEL 5.0 (Bradbury et al., 2007). The decay distance of LD at $r^2 = 0.1$ was determined as the length of a LD block. Kinship matrix analysis (Loiselle et al., 1995) in TASSEL 5.0 was conducted to uncover the genetic identity between two given accessions by $K_{ij} = \sum_l [\sum_a (p_{ila} - p_{la})(p_{jla} - p_{la}) + \sum_a \frac{p_{la}(1-p_{la})}{(n_l-1)}] / \sum_l \sum_a (p_{la}(1-p_{la}))$, where p_{ila} is the frequency of allele a at locus l in individual i ; p_{la} is the frequency of allele a at locus l in the reference sample; n_l is the number of alleles defined in the sample at locus l (the number of



individuals times the ploidy level minus the number missing alleles); and $(n_1 - 1)$ is a sampling bias correction. Negative values between two accessions, indicating the existence of a weaker relationship than would be expected between two random individuals, were replaced by zero. Analysis of molecular variance (AMOVA) was performed to estimate the variance between populations and among accessions within populations based on analyses of variance of allele frequencies (Excoffier et al., 1992) using Arlequin 3.5 software (Excoffier and Lischer, 2010). The population fixation statistic F_{ST} between the CN-set and US-set was calculated genome-widely using Arlequin 3.5.

Population Structure Analysis

Three multivariate analyses, including model-based population structure analysis, principal component analysis (PCA), and cluster analysis with a neighbor-joining algorithm, were employed to divide the soybean accessions into subgroups. The Bayesian model-based program STRUCTURE 2.3 (Pritchard et al., 2000) was used to infer the population structure and to assign the 577 genotypes into subpopulations based on 5,195 polymorphic SNP markers (there were 166 SNPs of the 5,361 SNPs on the Illumina SoySNP6k iSelect BeadChip for which no data points were obtained in more than 20% of the accessions; see the first section of the results). Ten independent analysis instances, based on 100,000 MCMC replications and 100,000 burn-ins, were performed, with the hypothetical number of subpopulations (k) ranging from 1 to 10. The number of subpopulations was determined when Δk reached its highest value (an *ad hoc* statistic; Evanno et al., 2005). PCA and cluster analysis were implemented in TASSEL 5.0.

Identification of Loci under Selection

Three statistical methods were used to detect the loci under selection. (1) Differences in allele frequency between the CN-set and the US-set were tested by Student's t -test: $t = \frac{f_1 - f_2}{\sqrt{\left(\frac{1}{2n_1} + \frac{1}{2n_2}\right) f_{exp}(1 - f_{exp})}}$, where $f_{exp} = \frac{f_1 n_1 + f_2 n_2}{n_1 + n_2}$, f_1 and f_2 were

the allele frequencies in the CN-set and US-set, respectively, and n_1 and n_2 were the sample sizes in the CN-set and US-set, respectively. The population-specific alleles were determined for each subpopulation based on zero allele frequency in one subpopulation and non-zero in another subpopulation; and different allele frequency between two subpopulations reaching at significance level $P < 0.001$. (2) F_{ST} between the CN-set and US-set was calculated for individual SNP using VCFtools (<https://vcftools.github.io/index.html>) (Weir and Cockerham, 1984) and EigenGWAS (Chen et al., 2016). The VCFtools was conducted with a sliding window of 100 kb and a step size of 10 kb (Schmutz et al., 2014) over the whole genome, and the regions with the top 5% of F_{ST} -values were regarded as highly diverged across the two groups. (3) Finding loci under selection through genome-wide association studies of eigenvectors were implemented by EigenGWAS. There were three steps included. Firstly, genetic relationship matrix was generated for the 577 accessions; secondly, the top 10 eigenvalues and eigenvectors were calculated; and then linear model for selected eigenvectors from the second step was conducted.

RESULTS

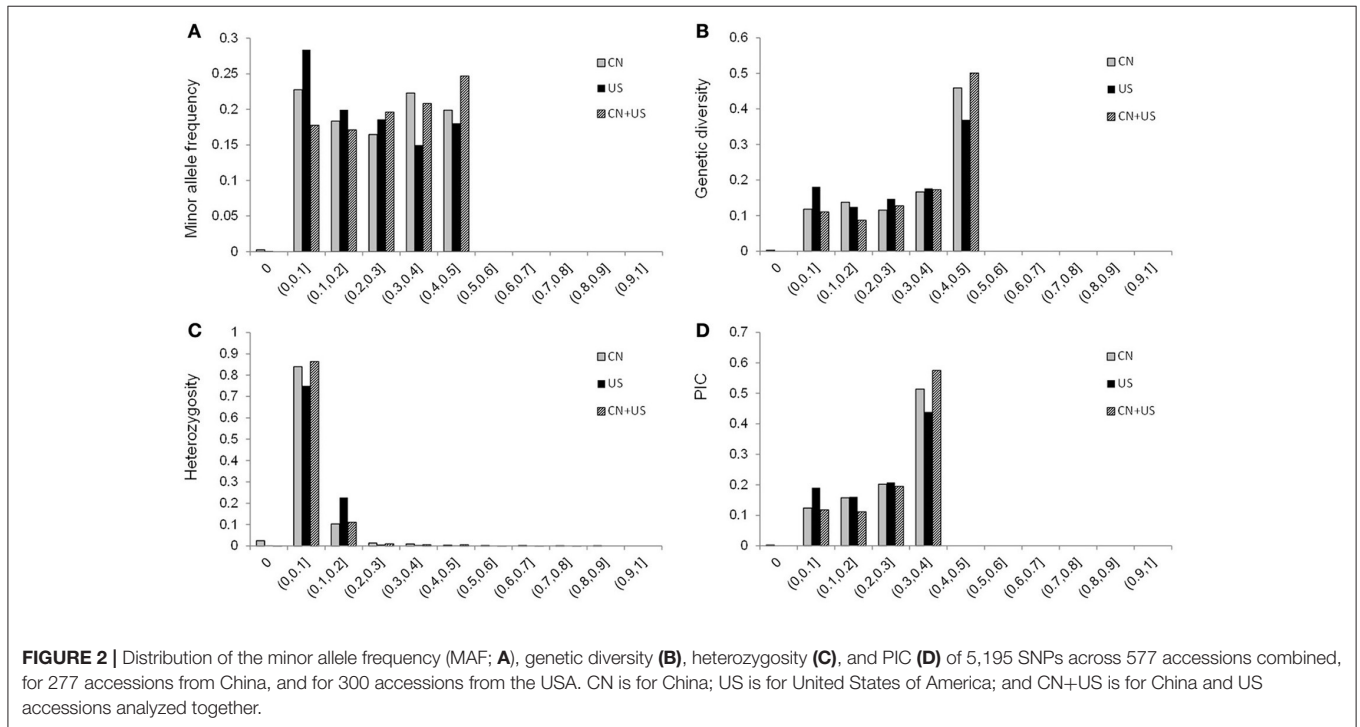
Genetic Diversity

A total of 577 soybean accessions were analyzed using the 5,361 SNP markers of the Illumina SoySNP6k iSelect BeadChip. SNP markers with missing data points for more than 20% (166 SNPs) of the accessions were not used for further analysis, so a total of 5,195 SNPs (96.90%) were used in our study. In the CN-set, the average PIC was 0.2643, ranging from 0 to 0.3750, and the average genetic diversity was 0.3307, ranging from 0 to 0.5000. In the US-set, the PIC ranged from 0 to 0.3750, with an average of 0.2408, and genetic diversity ranged from 0 to 0.5000, with an average of 0.2988 (Table 1 and Figure 2). These results indicated that the Chinese soybean accessions had a higher level of genetic diversity than the US soybean accessions. As expected, there was an increase in the estimates of PIC and genetic diversity for the

TABLE 1 | Genetic parameters revealed by the analysis of 5,195 polymorphic SNP markers among the soybean accessions of the diversity panel.

Origin	Minor allele frequency	Genetic diversity	Heterozygosity	PIC
CN	0.2498 (0~0.5000)	0.3307 (0~0.5000)	0.0658 (0~0.8654)	0.2643 (0~0.3750)
US	0.2209 (0~0.5000)	0.2988 (0~0.5000)	0.0762 (0~0.9588)	0.2408 (0~0.3750)
CN+US	0.2679 (0.0026~0.5000)	0.3489 (0.0052~0.5000)	0.0711 (0~0.7750)	0.2769 (0.0052~0.3750)

CN is for China; and US is for United States of America.

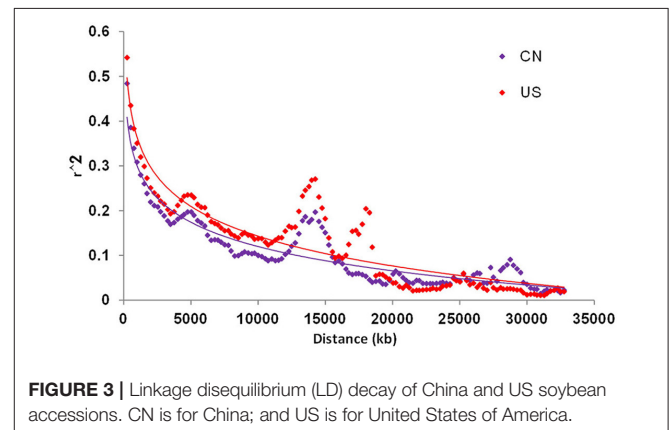


entire diversity collection (**Table 1** and **Figure 2**), as compared to either the CN-set or the US-set.

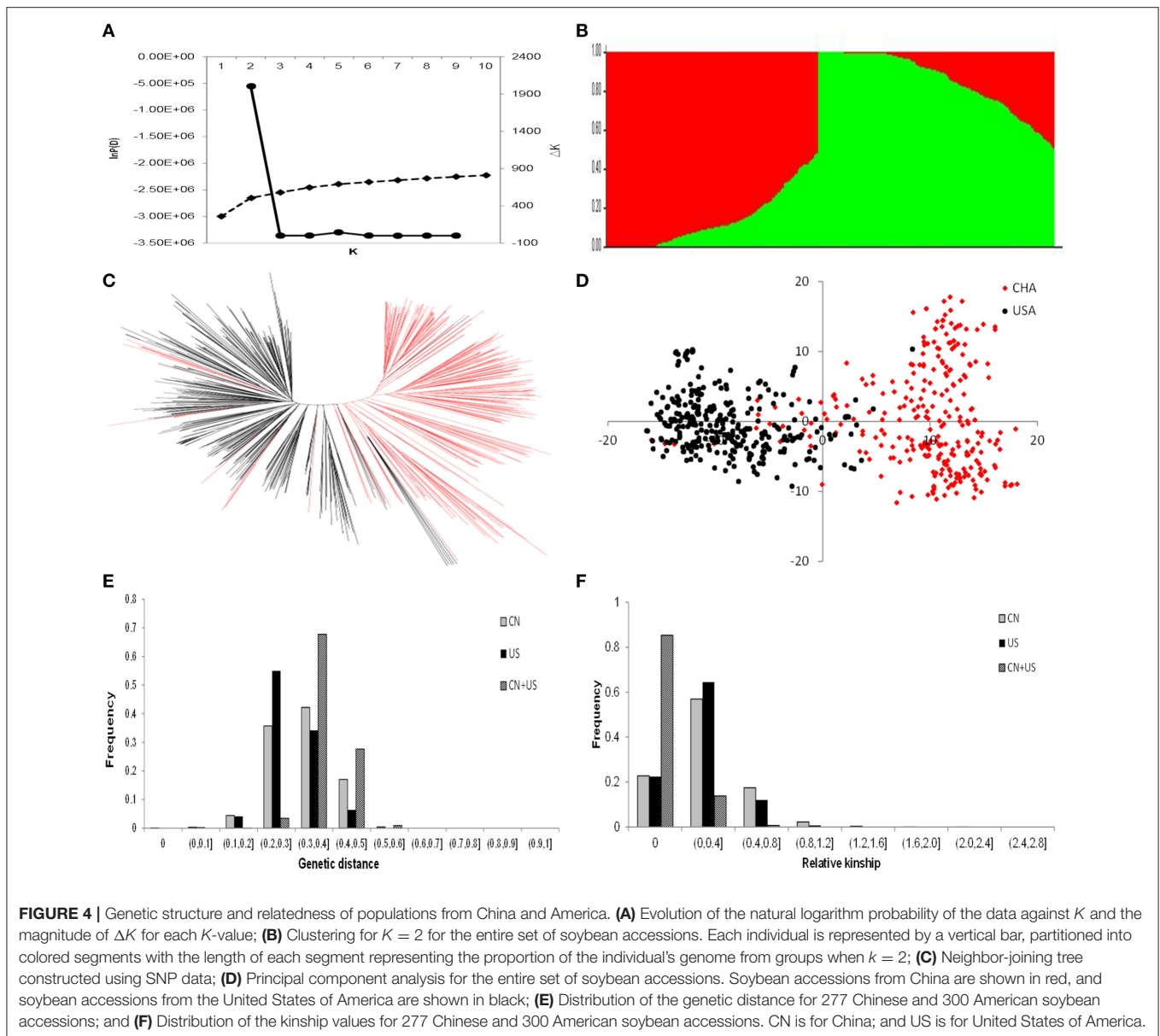
Linkage disequilibrium (LD) analysis was performed based on 5,195 SNPs in both the CN-set and the US-set. The r^2 statistic was calculated and tested for each pairwise SNP to measure the degree of LD (**Figure 3**). In the CN-set, we found that r^2 for 58.41% of the pairwise SNPs were significant level at the $\alpha = 0.01$. Comparatively, in the US-set, r^2 for 59.96% of the pairwise SNPs were significant level at the $\alpha = 0.01$. In both sets, LD gradually declined with increased physical distance. The distance over which LD decays to $r^2 = 0.1$ was 8,500 kb in the CN-set and 15,500 kb in the US-set. Thus, our LD analysis also indicated that the Chinese soybeans have a higher level of genetic diversity than the American soybeans.

Unbiased Population Structure Analysis

We investigated the possible population structure without introducing any prior information or assumptions. Population clustering was performed using the STRUCTURE program.



Likelihood (\ln) increased continuously, with no obvious inflection point (**Figure 4A**), which implies that the accessions included in the analysis were very diverse. In addition, the



Evanno criterion supported the choice of $K = 2$ as the highest level of structure (Figure 4B). There were 273 accessions in subpopulation 1, among which 257 accessions were from China and 16 accessions were from US. By comparison, there were 304 accessions in subpopulation 2, among which 20 accessions were from China and 284 accessions were from the USA (Supplementary Table S3). Of the 20 Chinese accessions in subpopulation 2, the ancestors of 15 of these accessions came from US. These results are consistent with the geographical origin and pedigree information of the 577 accessions.

In order to validate and gain further insight into the genetic diversity of the soybean germplasm panel, we constructed a neighbor-joining tree based on the frequency of shared alleles among the accessions. The 577 soybean accessions were classified into two major groups (Figure 4C). One group was composed

largely of accessions from the CN-set, and the other group was mainly composed of accessions from the US-set. These groupings were almost the same as the two subpopulations identified in our analysis conducted with the STRUCTURE program. PCA has been proposed as an alternative to population structure analysis for studying population stratification from genotypic data (Patterson et al., 2006). A PCA of the entire set of 577 accessions with the 5,195 SNPs (Figure 4D and Supplementary Table S4) also showed a clear separation of the same two groups that were identified in our population structure and neighbor-joining tree analyses.

Our unbiased population structure, neighbor-joining tree, and PCA analyses all clearly indicated the existence of two distinct subpopulations among the 577 accessions of our study. Clear divergence existed between the Chinese and US soybean

accessions. With the exception of a very small number of accessions that were grouped into the subpopulation for the other country (<7%), most of the accessions from the same origin (country) were clustered into the same genetic group. Therefore, the following studies were based on the two pre-defined genetic pools (i.e., the original two soybean populations from China and US).

Genetic Differences of the Two Subpopulations

Genetic differences between the pre-defined genetic pools were evaluated using four types of analysis of population differentiation: AMOVA, Roger genetic distance, genetic relatedness, and allele frequency. AMOVA indicated that 19.33% of the total genetic variation occurred among the subpopulations, whereas 80.67% was within the subpopulations. The population pairwise F_{ST} was 0.1933 ($P < 0.001$) between the two subpopulations (Supplementary Table S5), which indicates a high level of difference. The average Roger genetic distances within the CN-set, the US-set, and the combined dataset were 0.3223, 0.2932, and 0.3794, respectively (Figure 4E and Supplementary Table S6). For the CN-set, 42.18% of the distance values were between 0.300 and 0.400. For the US-set 54.99% of the distance values were between 0.200 and 0.300. Viewing the CN-set and the US-set collectively, 67.75% of the distance values were between 0.3 and 0.4 (Figure 4E). These results clearly demonstrate that the genetic distance within the CN-set is larger than that in the US-set, and the genetic distance of pairwise genotypes between subpopulations was larger than that of within subpopulations.

Relative kinship reflects the approximate degree of identity between two given accessions (Figure 4F and Supplementary Table S7). In the CN-set, the kinship coefficients between pairs of accessions varied from 0 to 1.4108, with an overall average of 0.0477; 22.76% of the estimates had a value of zero, which means there is almost no relationship. Comparatively, for the US-set, the kinship coefficients ranged from 0 to 1.5637, with an overall average of 0.0416; 22.43% of the pairwise kinship estimates had a value of zero. For combined analysis of all 577 accessions, the kinship coefficients ranged from 0 to 1.0204, with an overall average of 0.0195; 85.37% of the pairwise kinship estimates had a value of zero, indicating that most of the genotypes were not highly related (Figure 4F).

The distribution of the differences of allele frequencies (denoted by $f_1 - f_2$) ranged from 0 to 88.63% between the CN-set and the US-set and was almost symmetrical (Figure 5A), indicating that selection has occurred in both directions. Absolute values of $f_1 - f_2$ for 35.78% of the loci were lower than 0.1, while the values for 7.72% of the loci were higher than 0.5 (Figure 5B). Across the soybean genome, the F_{ST} for 56.28% of the loci was lower than 0.1, while the value for 2.73% of the loci was higher than 0.5 (Figure 6). These results imply that a large proportion of genomic regions have maintained similar frequencies in both subpopulations. We performed the statistical tests to determine if the difference between the allele frequencies was significant. Results showed that the allele frequencies of 3,223

SNPs (62.04% of the total 5,195 SNPs) reached the significance level at $P = 10^{-6}$. F_{ST} -values, absolute values of difference between the allele frequencies ($|f_1 - f_2|$), and $-\log(P)$ of $f_1 - f_2$ values (P is the P -value from statistical test) are presented in Figure 7. The loci with selection signals identified based on F_{ST} , test of $f_1 - f_2$ values, and results from EigenGWAS were highly consistent over the 20 soybean chromosomes (Figure 7).

The top 10 candidate loci under selection are presented in Table 2. The strongest selection signal was found at SNP locus ss245627275 on chromosome 5. Four loci with selection signal overlapped with previously identified QTLs for photoperiod traits: ss246094102 on chromosome 6 (Liu et al., 2011), ss246490128 on chromosome 8 (Pooprompan et al., 2006; Reinprecht et al., 2006), ss247294954 on chromosome 10 (Li et al., 2008; Kuroda et al., 2013), and ss250485410 on chromosome 20 (Reinprecht et al., 2006). Seven loci with selection signals overlapped with previously identified QTLs for seed quality. For example, ss250485410 on chromosome 20 is in a region associated with seed protein content (Lu et al., 2013), and seed oil (Qi et al., 2011). Four loci with selection signals overlapped with QTLs associated with defense traits: SNP locus ss246490128 on chromosome 8 is in a region associated with SCN (Wu et al., 2009) and flood tolerance (Sayama et al., 2009); SNP locus ss247294954 is on chromosome 10 and is in a region associated with drought (Carpentieri-Pipolo et al., 2011) and flood tolerance (Githiri et al., 2006); SNP locus ss249030246 on chromosome 16 is in a region associated with whitefly resistance (Zhang et al., 2013); and SNP locus ss250485410 on chromosome 20 is in a regions associated with Sudden Death Syndrome (SDS, Swaminathan et al., 2016). Four loci with selection signals were overlapped with yield-related traits: SNP locus ss246490128 on chromosome 8 is in a region associated with seed weight (Kim et al., 2010), and branching (Liu et al., 2017a); SNP locus ss247294954 on chromosome 10 is in a region associated with seed weight (Sun et al., 2012), and branching (Li et al., 2008); SNP locus ss247790225 on chromosome 12 is in a region associated with branching (Liu et al., 2017a); and SNP locus ss250485410 on chromosome 20 is in a region associated with seed weight (Kato et al., 2014).

Three loci with selection signals (ss246490128 on chromosome 8, ss247294954 on chromosome 10, and ss250485410 on chromosome 20) had pleiotropic effects on all of the following physiological aspects: photoperiod, seed quality, defense, and yield-related traits. Among the top 10 loci ranked based on the strength of their selection signals, ss245627275 on chromosome 5, ss246165779 on chromosome 7, and ss249429323 on chromosome 17 have not been previously reported to be associated with any traits, and one gene near SNP locus ss245627275 was annotated as RAS-related nuclear protein; one gene near SNP locus ss246165779 was annotated as having a function relating to nucleoside diphosphate kinase activity; and one gene near SNP locus ss249429323 was annotated as having a function relating to protein domain specific binding (Table 2). The calculation of the allele frequency for the two subpopulations allowed the identification of subpopulation-specific alleles. Of the 5,191 SNP markers, four specific alleles for the CN-set were

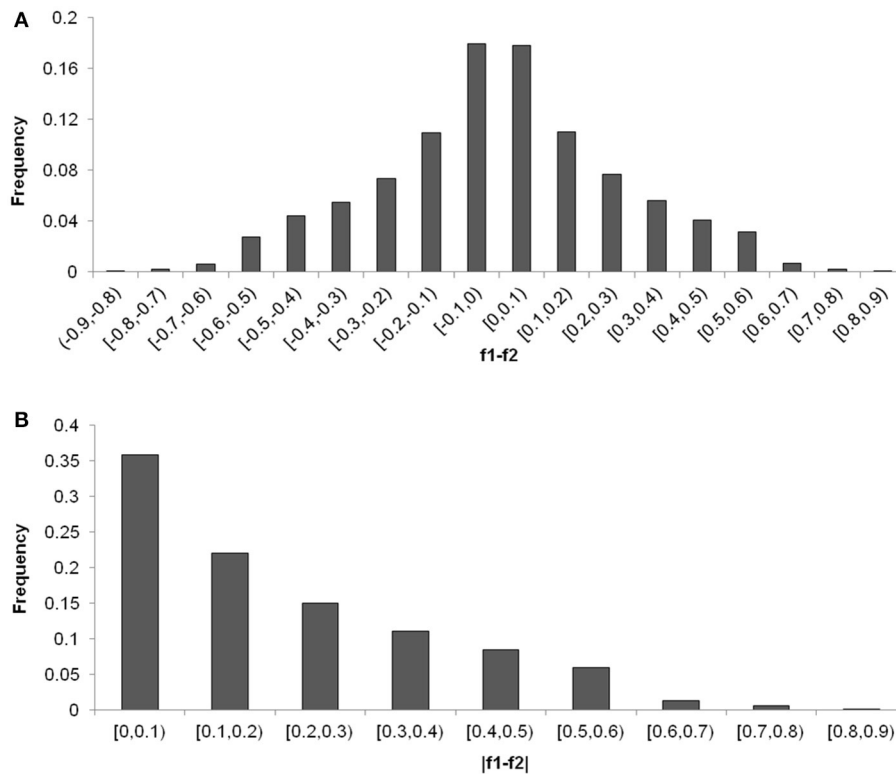


FIGURE 5 | Distribution of the differentiation of allele frequencies, $f_1 - f_2$ (A) and $|f_1 - f_2|$ (B, absolute values of $f_1 - f_2$) between Chinese and American soybean accessions.

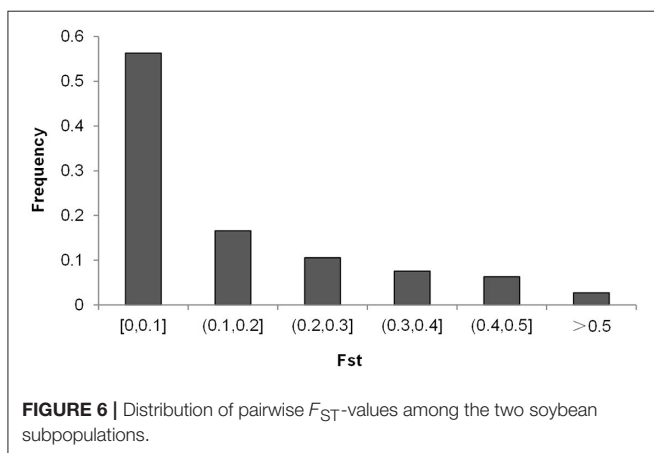


FIGURE 6 | Distribution of pairwise F_{ST} -values among the two soybean subpopulations.

distributed on chromosomes 5, 7, 10, and 12, but no specific allele existed in the US-set (Table 3).

DISCUSSION

Fang et al. (2017) dissected the genetic architecture of 84 agronomic traits and investigated the genetic networks underling their phenotypic correlations by 809 soybean accessions with diverse geographic distribution. Of 809 accessions, only 67

accessions derived from US, and majority of accessions (683 accessions) came from China, accounting for 84.4%. While in current study, to particularly compare the genetic diversity of soybean cultivars from China and US, and to identify loci that have undergone selection based on different statistical algorithms, we collected fairly balanced samples (277 for China vs. 300 for US; Figure 1) from the major production areas in both countries to reach a high statistical power. The planting area of soybean in northeast and Huang-huai-hai regions is account for more than 80% of total soybean planting area in China. Of 1,300 cultivars released in China from 2005 to 2013, 1,077 cultivars (82.85%) were from northeast and Huang-huai-hai regions, and 656 cultivars (50.46%) were the progenies of cultivars imported from abroad. While for the southern region, 223 cultivars were released accounting for 17.15% of the total cultivars released in China, of which 82 cultivars (6.31%) were released with exotic ancestors. As for the materials collected in USA, the situation is similar as those in China. The main production areas in USA are in the central and northern regions, and the released cultivars from southern region of the USA are comparatively small. Considering the reasons above, the 277 accessions in the CN-set were selected from the geographic regions typified by the Northern spring (Nsp) ecotype and the Huang-huai-hai summer (Hsu) ecotype. The 300 accessions in the US-set covered accessions in maturity groups I–IV, which include accessions used in the main soybean

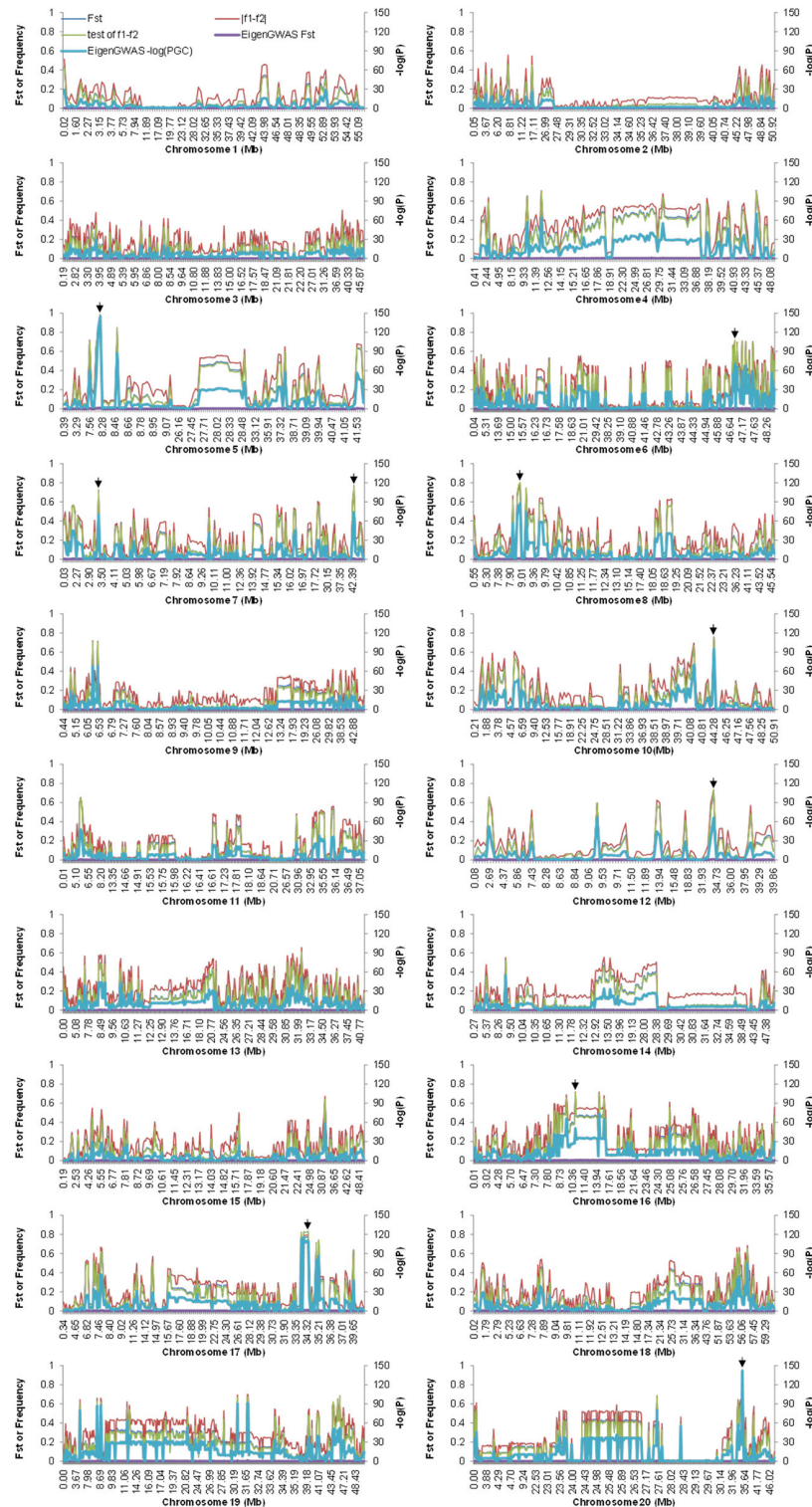


FIGURE 7 | F_{ST} -values and the allele difference for the whole genome among Chinese and American soybean accessions. F_{ST} were calculated from VCFtool and EigenGWAS, denoted by F_{ST} and EigenGWAS F_{ST} , respectively. PGC were P -values corrected by genomic control, which were calculated by EigenGWAS. To demonstrate the plot clearly, the values of $-\log(PGC)$ were multiplied by 20. Arrows indicate the physical positions of the top 10 candidate loci under selection.

TABLE 2 | The top 10 SNPs with significantly ($P < 10^{-6}$) different allele frequencies between China and the US germplasms.

SNP name	Chr.	Allele	Position	Allele frequency		Allele frequency difference	F_{ST}	Test of $f_1 - f_2$	-log(P) from EigenGWAS	Associated traits
				CN	US					
ss245627275 [#]	5	A	8234203	0.8967	0.0103	0.8863	0.8866	148.3306	342.8146	
ss246094102	6	A	46849787	0.7955	0.0765	0.7189	0.6925	107.7709	210.3497	Photoperiod insensitivity (Liu et al., 2011); cqSeed protein (Pathan et al., 2013)
ss246165779*	7	C	3376487	0.1586	0.8869	0.7283	0.6942	108.3583	209.3325	
ss246411658	7	C	42514762	0.2539	0.9883	0.7344	0.7409	116.1999	221.5186	Seed daidzein (Gutierrez-Gonzalez et al., 2010)
ss246490128	8	C	8917276	0.8971	0.1160	0.7810	0.7567	120.5933	269.0964	Pod maturity (Reinprecht et al., 2006); First flower (Pooprompan et al., 2006); Seed weight (Kim et al., 2010); Branching (Liu et al., 2017a); Seed coat and hilum color (Zhou et al., 2015); Seed protein (Lu et al., 2013); SCN (Wu et al., 2009); Flood tolerance (Sayama et al., 2009)
ss247294954	10	C	44278194	0.0879	0.8426	0.7546	0.7256	114.3218	296.2328	First flower (Kuroda et al., 2013); Reproductive stage length (Li et al., 2008); Pod maturity (Li et al., 2008); Seed weight (Sun et al., 2012); Branching (Li et al., 2008); Seed protein (Chen et al., 2007); Drought tolerance (Carpentieri-Pipolo et al., 2011); Flood tolerance (Githiri et al., 2006)
ss247790225	12	A	34375177	0.1350	0.8672	0.7322	0.6971	108.8819	196.7905	Branching (Liu et al., 2017a); Seed oil (Liu et al., 2017a); Seed protein (Liu et al., 2017a); Lodging (Liu et al., 2017b)
ss249030246	16	A	10794262	0.9233	0.2040	0.7193	0.6777	106.5537	142.4584	Seed oil (Mao et al., 2013); Whitefly resistance (Zhang et al., 2013)
ss249429323 ^{**}	17	A	34433642	0.1097	0.9078	0.7982	0.7783	124.9834	301.3143	
ss250485410	20	C	34577475	0.0967	0.9680	0.8713	0.8671	143.8360	294.0404	First flower (Reinprecht et al., 2006); Seed weight (Kato et al., 2014); Seed protein (Lu et al., 2013); Seed oil (Qi et al., 2011); Lodging (Liu et al., 2017b); Sudden Death Syndrome (SDS, Swaminathan et al., 2016)

[#]Gene annotation information for this locus is "RAS-related nuclear protein," and "Cytochrome P450 superfamily protein"; *Gene annotation information for this locus is "nucleoside diphosphate kinase," "Small nuclear ribonucleoprotein splicing factor," and "Lung seven transmembrane receptor"; **Gene annotation information for this locus is "Ubiquinol cytochrome c reductase, subunit QCR7," and "protein domain specific binding." CN is for China; and US is for United States of America.

production regions of the USA. It is important to consider the strong influence of maturity on phenotypic traits in soybean: large pleiotropic effects are well-documented for plant height, lodging, and seed yield, and these large effects tend to overwhelm other phenotypic differences among genotypes (Cui et al., 2001). However, importantly, identification based on genomic data is not influenced by maturity or other phenotypic factors. In the present study, 5,361 SNP markers from the Illumina SoySNP6k iSelect BeadChip were employed to genotype all of the accessions in our diversity panel.

Genetic Diversity among Soybean Groups

The average genetic diversity and PIC-values for the combined set of all accessions were, respectively, 0.3489 and 0.2769. Compared to the previously reported results based on SNP (Hao et al., 2012; Zhou et al., 2015) or SSR (Li et al., 2011) data, the

level of genetic diversity that we observed here is lower, almost certainly because the materials used in the present study were cultivars and advanced breeding lines, while the materials used in Hao et al. (2012) and Li et al. (2011) were mainly landraces. We observed that the accessions from the US had lower genetic diversity and PIC-values than the accessions from China. Our conclusion that the Chinese cultivars are more diverse than the American cultivars is consistent with the conclusions of previous studies, including molecular analyses using RAPD markers to predict the likely ancestors of Chinese and US soybean cultivars (Qiu et al., 1997; Li et al., 2001) and a phenotypic diversity study of modern Chinese and North American soybean cultivars conducted by Cui et al. (2001). The lower LD-values and larger genetic distance within the Chinese accessions as compared to the American accessions again highlights that the Chinese germplasm has relatively greater genetic diversity.

TABLE 3 | Population-specific alleles for Chinese soybean genotypes.

Marker	Allele	Chr.	Position	Allele frequency in CN	Associated trait
ss245630590	C	5	8603833	0.0939	
ss246176034	G	7	4488650	0.0421	First flower (Orf et al., 1999); Pod maturity (Wang et al., 2004); Seed yield (Wang et al., 2014)
ss247314990	G	10	47212732	0.0722	Pod maturity (Specht et al., 2001); Seed weight (Han et al., 2012)
ss247667029	G	12	8693064	0.0618	Seed isoflavone (Han et al., 2015)

CN is for China.

Population Structure Analysis of the Chinese and US Soybean Accessions

AMOVA, model-based population structure analysis, NJ-cluster analysis, and PCA were used to examine whether or not the 577 soybean accessions were from highly diverse origins and/or whether or not the accessions from China and the USA are homogeneous or represent two genetically distinct subgroups. We used a variety of methods and consistently obtained similar results for both the number of accessions and the particular membership of accessions within groups. The American accessions were genetically distinct from the Chinese accessions. There were clearly two different subpopulations: an American one and a Chinese one. Cluster analyses grouped accessions with similar geographical origins together, and these findings were in accordance with previous studies (Li et al., 2011; Hao et al., 2012; Sun et al., 2014; Zhou et al., 2015; Fang et al., 2017; Liu et al., 2017a). We noted that 93.76% of the soybean accessions were assigned into the subgroup corresponding to their expected geographical origin (Supplementary Tables S1–S3), implying that the geographic distribution of soybean cultivars and advanced breeding lines is a reliable parameter to use for understanding the overall population structure.

Regarding kinship, the 577 accessions are distantly related: For combined analysis of the two subpopulations, 85.37% of the pairwise kinship estimates were equal to zero (Figure 4F). The distance between the two subpopulations (0.3794; Supplementary Table S6) was larger than that within either of the two individual groupings (0.3223 for the CN-set and 0.2932 for the US-set; Supplementary Table S6). Our conclusion that the accessions are distantly related is also supported by the significant allelic frequency differences observed among the accessions based on AMOVA and F_{ST} analyses. Significant structuring of genetic variation was found between the Chinese and American soybean accessions (Supplementary Table S5). Viewed collectively, our findings indicate that even though the soybean genetic base in the USA is ultimately derived from only a few soybean cultivars that were introduced from China (Li et al., 2001), and Chinese plant breeders have successfully employed elite Northern American cultivars as breeding materials (Gai et al., 2015), breeding efforts of breeders of two countries in the time have generated a quite distinct genetic base of soybeans. Another influence that may underlie the observed genetic diversity between the two subpopulations is the efforts of breeders over many decades that have been focused on improving local adaptation to different environmental conditions in the soybean growth areas of the two countries.

Given the recognized importance of including materials from diverse genetic groups in breeding programs (Thompson and Nelson, 1998; Li et al., 2001), it is important to carefully support the exchange, identification, and utilization of exotic soybean germplasm in soybean improvement efforts.

Implications for Soybean Breeding

The identification of loci with selective signals is a centrally important step for understanding how various populations have adapted to particular agronomic practices and/or unique environments. The F_{ST} approach for detecting loci that have been under selection has been applied to many crops, including wild emmer wheat (Ren et al., 2013), common bean (Papa et al., 2007), tomato (Corrado et al., 2013), and soybean (Han et al., 2016), among many others, and markers that are identified using an F_{ST} -outlier method often reveal genome regions that have previously been associated with quantitative trait loci (QTL) related to domestication (Ren et al., 2013).

In the present study, we used an F_{ST} -based statistical analysis and EigenGWAS to identify selection signals at SNP loci. Although, the majority of the loci had low F_{ST} -values in pairwise comparisons, our data revealed the presence of genomic regions with high genetic variation between the two subpopulations (Figure 6), and the results were consistent with that of EigenGWAS. Of the top 10 candidate loci under selection, four loci overlapped with previously identified QTLs for photoperiod traits, seven loci overlapped with previously identified QTLs for seed quality-related traits, four loci were in regions associated with abiotic and biotic stress related traits, and four loci were in regions associated with yield traits. Three of the loci with selection signals were in regions associated with multiple physiological aspects (photoperiod, seed quality, abiotic and biotic-related, and yield-related traits). One locus ss246490128 on chromosome 8 overlapped with previously identified QTL Sat_215, around which a candidate domestication gene *Glymo08g09310* controlled seed size was identified (Zhou et al., 2016). Zhou et al. (2015) also reported four loci related to soybean domestication, which were in the regions of loci ss246490128 on Chromosome 8, ss247294954 on Chromosome 10, ss247790225 on chromosome 12, and ss249429323 on chromosome 17 in present study, respectively. The trait-related loci under selection can result in phenotypic variation between populations from two countries. For example, Wang et al. (2005) evaluated the yield-related traits for accessions from China and the USA that were from similar latitudes, and found that cultivars from the USA had higher plant heights, more branching, higher

numbers of pods, and higher yields than those from China. The loci with selection signals that we observed may be (or have been) under direct selection, but it is more likely that they are located in chromosome regions that were selected during the domestication process. The putative functions associated with these loci are that related to the role in adaptation: these loci are putatively involved in processes that are vital for plant growth and survival. The loci with selection signals that we identified in the present study are of potential interest for plant breeders, as they likely contribute to the existing crop performance differences between Chinese and US soybeans.

AUTHOR CONTRIBUTIONS

LQ, DW, ZL, and SW: conceived and designed the experiments; ZL, ZW, and XF: performed the experiments; ZL, HL, YL, RG, and YG: analyzed the data; ZL, HL, and ZW: wrote the paper.

REFERENCES

- Akond, M., Liu, S., Schoener, L., Anderson, J. A., Kantartz, S. K., Meksem, K., et al. (2013). SNP-based genetic linkage map of soybean using the SoySNP6K Illumina Infinium BeadChip genotyping array. *J. Plant Genome Sci.* 1, 80–89. doi: 10.5147/jpgs.2013.0090
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Carpentieri-Pipolo, V., Pipolo, A., Abdel-Haleem, H., Boerma, H., and Sinclair, T. (2011). Identification of QTLs associated with limited leaf hydraulic conductance in soybean. *Euphytica* 186, 679–686. doi: 10.1007/s10681-011-0535-6
- Carter, T. E., Nelson, R. L., Cregan, P. B., Boerma, H. R., Manjarrez-Sandoval, P., Zhou, X., et al. (2000). “Project SAVE (Soybean Asian Variety Evaluation)—potential new sources of yield genes with no strings from USB, public, and private cooperative research,” in *Proceedings of The Twenty-Eighth Soybean Seed Research Conference 1998*, ed B. Park (Washington, DC: American Seed Trade Association, Inc.), 68–83.
- Chang, H. X., and Hartman, G. L. (2017). Characterization of insect resistance loci in the USDA soybean germplasm collection using genome-wide association studies. *Front. Plant Sci.* 8:670. doi: 10.3389/fpls.2017.00670
- Chen, G. B., Lee, S. H., Zhu, Z. X., Benyamin, B., and Robinson, M. R. (2016). EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations. *Heredity* 117, 51–61. doi: 10.1038/hdy.2016.25
- Chen, Q., Zhang, Z., Liu, C., Xin, D., Qiu, H., Shan, D., et al. (2007). QTL Analysis of major agronomic traits in soybean. *Agric. Sci. China* 6, 399–405. doi: 10.1016/S1671-2927(07)60062-5
- Corrado, G., Piffanelli, P., Caramante, M., Coppola, M., and Rao, R. (2013). SNP genotyping reveals genetic diversity between cultivated landraces and contemporary varieties of tomato. *BMC Genomics* 14:835. doi: 10.1186/1471-2164-14-835
- Cui, Z. L., Carter, T. E., Burton, J. W., and Wells, R. (2001). Phenotypic diversity of modern Chinese and North American soybean cultivars. *Crop Sci.* 41, 1954–1967. doi: 10.2135/cropsci2001.1954
- Cui, Z., Carter, T. E., and Burton, J. W. (2000a). Genetic base of 651 Chinese soybean cultivars released during 1923 to 1995. *Crop Sci.* 40, 1470–1481. doi: 10.2135/cropsci2000.4051470x
- Cui, Z., Carter, T. E., and Burton, J. W. (2000b). Genetic diversity patterns in Chinese soybean cultivars based on coefficient of parentage. *Crop Sci.* 40, 1780–1793. doi: 10.2135/cropsci2000.4061780x
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Excoffier, L., Smouse, P. E., and Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18:161. doi: 10.1186/s13059-017-1289-9
- Gai, J. Y., Xiong, D. J., and Zhao, T. J. (2015). *The Pedigrees and Germplasm Bases of Soybean Cultivars Released in China (1923-2005)*. Beijing: China Agricultural Press.
- Githiri, S., Watanabe, S., Harada, K., and Takahashi, R. (2006). QTL analysis of flooding tolerance in soybean at an early vegetative growth stage. *Plant Breed.* 125, 613–618. doi: 10.1111/j.1439-0523.2006.01291.x
- Gizlice, Z., Carter, T. E., and Burton, J. W. (1994). Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34, 1143–1151. doi: 10.2135/cropsci1994.0011183X003400050001x
- Gutierrez-Gonzalez, J. J., Wu, X., Gillman, J. D., Lee, J. D., Zhong, R., Yu, O., et al. (2010). Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. *BMC Plant Biol.* 10:105. doi: 10.1186/1471-2229-10-105
- Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theor. Appl. Genet.* 125, 671–683. doi: 10.1007/s00122-012-1859-x
- Han, Y., Teng, W., Wang, Y., Zhao, X., Wu, L., Li, D., et al. (2015). Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. *Plant Breed.* 134, 300–309. doi: 10.1111/pbr.12259
- Han, Y., Zhao, X., Liu, D., Li, Y., Lightfoot, D. A., Yang, Z., et al. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytol.* 209, 871–884. doi: 10.1111/nph.13626
- Hao, D., Cheng, H., Yin, Z. T., Cui, S. Y., Zhang, D., Wang, H., et al. (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. doi: 10.1007/s00122-011-1719-0
- Hymowitz, T., and Harlan, J. R. (1983). Introduction of soybean to North America by Samuel Bowen in 1765. *Econ. Bot.* 37, 371–379. doi: 10.1007/BF02904196

FUNDING

This work was supported by the National Key R & D Program for Crop Breeding (grant 2016YFD0100304), the Development of Novel Elite Soybean Cultivars and Lines with High Oil Content program (grant Z161100000916005-06), the Crop Germplasm Resources Protection program (2017NWB036-05, 2016NWB030-05), the Platform of National Crop Germplasm Resources of China (2017-004, 2016-004), the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS), and the Michigan Soybean Promotion Committee.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02014/full#supplementary-material>

- Hymowitz, T., and Newell, C. (1981). Taxonomy of the genus *Glycine*, domestication and uses of soybeans. *Econ. Bot.* 35, 272–288. doi: 10.1007/BF02859119
- Kato, S., Sayama, T., Fujii, K., Yumoto, S., Kono, Y., Hwang, T., et al. (2014). A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor. Appl. Genet.* 127, 1365–1374. doi: 10.1007/s00122-014-2304-0
- Kim, H., Kim, Y., Kim, S., Son, B., Choi, Y., Kang, J., et al. (2010). Analysis of quantitative trait loci (QTLs) for seed size and fatty acid composition using recombinant inbred lines in soybean. *J. Life Sci.* 20, 1186–1192. doi: 10.5352/JLS.2010.20.8.1186
- Kisha, T. J., Sneller, C. H., and Diers, B. W. (1997). Relationship between genetic distance among parents and genetic variance in populations of soybean. *Crop Sci.* 37, 1317–1325. doi: 10.2135/cropsci1997.001183X003700040048x
- Kuroda, Y., Kaga, A., Norihiko, T., Yano, H., Takada, Y., Kato, S., et al. (2013). QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. *Ecol. Evol.* 3, 2150–2168. doi: 10.1002/ece3.606
- Lee, Y. G., Jeong, N., Kim, J. H., Lee, K., Kim, K. H., Pirani, A., et al. (2015). Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* 81, 625–636. doi: 10.1111/tpj.12755
- Li, W., Zheng, D., Van, K., and Lee, S. (2008). QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). *J. Crop Sci. Biotech.* 11, 171–190. Available online at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.538.6031&rep=rep1&type=pdf>
- Li, Y. H., Zhou, G., Ma, J., Jiang, W., Jin, L. G., Zhang, Z., et al. (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* 32, 1045–1052. doi: 10.1038/nbt.2979
- Li, Y., Smulders, M. J. M., Chang, R. Z., and Qiu, L. J. (2011). Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. *Conserv. Genet.* 12, 1145–1157. doi: 10.1007/s10592-011-0216-y
- Li, Z., Qiu, L., Thompson, J. A., Welsh, M. M., and Nelson, R. L. (2001). Molecular genetic analysis of U.S. and Chinese soybean ancestral lines. *Crop Sci.* 41, 1330–1336. doi: 10.2135/cropsci2001.4141330x
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Liu, W., Kim, M., Kang, Y., Van, K., Lee, Y., Srinives, P., et al. (2011). QTL identification of flowering time at three different latitudes reveals homeologous genomic regions that control flowering in soybean. *Theor. Appl. Genet.* 4, 545–553. doi: 10.1007/s00122-011-1606-8
- Liu, Z., Li, H., Fan, X., Huang, W., Yang, J., Wen, Z., et al. (2017a). Phenotypic characterization and genetic dissection of nine agronomic traits in Tokachi nagaha and its derived cultivars in soybean (*Glycine max* (L.) Merr.). *Plant Sci.* 256, 72–86. doi: 10.1016/j.plantsci.2016.11.009
- Liu, Z., Li, H., Fan, X., Huang, W., Yang, J., Zheng, Y., et al. (2017b). Selection of soybean elite cultivars based on phenotypic and genomic characters related to lodging tolerance. *Plant Breed.* 136, 526–538. doi: 10.1111/pbr.12495
- Loiselle, B. A., Sork, V. L., Nason, J., and Graham, C. (1995). Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am. J. Bot.* 82, 1420–1425. doi: 10.2307/2445869
- Lu, W., Wen, Z., Li, H., Yuan, D., Li, J., Zhang, H., et al. (2013). Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *Theor. Appl. Genet.* 126, 425–433. doi: 10.1007/s00122-012-1990-8
- Mao, T., Jiang, Z., Han, Y., Teng, W., Zhao, X., and Li, W. (2013). Identification of quantitative trait loci underlying seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breed.* 132, 630–641. doi: 10.1111/pbr.12091
- Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., et al. (1999). Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Sci.* 39, 1642–1651. doi: 10.2135/cropsci1999.3961642x
- Papa, R., Bellucci, E., Rossi, M., Leonardi, S., Rau, D., Gepts, P., et al. (2007). Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Ann. Bot.* 100, 1039–1051. doi: 10.1093/aob/mcm151
- Pathan, S., Vuong, T., Clark, K., Lee, J., Shannon, J., Roberts, C., et al. (2013). Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Sci.* 53, 765–774. doi: 10.2135/cropsci2012.03.0153
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190
- Pooprompan, P., Wasee, S., Toojinda, T., Abe, J., Chanprame, S., and Srinives, P. (2006). Molecular marker analysis of days to flowering in vegetable soybean (*Glycine max* (L.) Merrill). *Kase. J.* 40, 573–581.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959. Available online at: <http://www.genetics.org/content/155/2/945>
- Qi, Z., Wu, Q., Han, X., Sun, Y., Du, X., Liu, C., et al. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica* 179, 499–514. doi: 10.1007/s10681-011-0386-1
- Qiu, L., Nelson, R. L., and Vodkin, L. O. (1997). Evaluation of soybean germplasm with random amplification polymorphic DNA (RAPD) markers. *Sci. Agric. Sin.* 23, 408–417 (in Chinese).
- Reinprecht, Y., Poysa, V. W., Yu, K., Rajcan, I., Ablett, G. R., and Pauls, K. p., (2006). Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* 49, 1510–1527. doi: 10.1139/g06-112
- Ren, J., Chen, L., Sun, D., You, F. M., Wang, J., Peng, Y., et al. (2013). SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. *BMC Evol. Biol.* 13:169. doi: 10.1186/1471-2148-13-169
- Sayama, T., Nakazaki, T., Ishikawa, G., Yagasaki, K., Yamada, N., Hirota, N., et al. (2009). QTL analysis of seed-flooding tolerance in soybean (*Glycine max* [L.] Merr.). *Plant Sci.* 176, 514–521. doi: 10.1016/j.plantsci.2009.01.007
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Specht, J. E., Chase, K., Macrander, M., Graef, G. L., Chung, J., Markwell, J. P., et al. (2001). Soybean response to water: a QTL analysis of drought tolerance. *Crop Sci.* 41, 493–509. doi: 10.2135/cropsci2001.412493x
- Sun, J., Guo, N., Lei, J., Li, L., Hu, G., and Xing, H. (2014). Association mapping for partial resistance to *Phytophthora sojae* in soybean (*Glycine max* (L.) Merr.). *J. Genet.* 93, 355–363. doi: 10.1007/s12041-014-0383-y
- Sun, Y. N., Pan, J. B., Shi, X. L., Du, X. Y., Wu, Q., Qi, Z. M., et al. (2012). Multi-environment mapping and meta-analysis of 100-seed weight in soybean. *Mol. Biol. Rep.* 39, 9435–9443. doi: 10.1007/s11033-012-1808-4
- Swaminathan, S., Abeysekara, N. S., Liu, M., Cianzio, S. R., and Bhattacharyya, M. K., (2016). Quantitative trait loci underlying host responses of soybean to *Fusarium virguliforme* toxins that cause foliar sudden death syndrome. *Theor. Appl. Genet.* 129, 495–506. doi: 10.1007/s00122-015-2643-5
- Thompson, J. A., and Nelson, R. L. (1998). Core set of primers to evaluate genetic diversity in soybean. *Crop Sci.* 38, 1356–1362. doi: 10.2135/cropsci1998.001183X003800050034x
- Wang, D., Graef, G. L., Procopiuk, A. M., and Diers, B. W., (2004). Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor. Appl. Genet.* 108, 458–467. doi: 10.1007/s00122-003-1449-z
- Wang, H., Zhang, H., Liu, C., Xie, F., and Steven, S. K. M. (2005). Comparison on the plant type and yield of soybean varieties from close latitude of China and America. *Soybean Sci.* 24, 286–290 (in Chinese). doi: 10.3969/j.issn.1000-9841.2005.04.009
- Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., and Yu, D. (2016). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci. Rep.* 6:20728. doi: 10.1038/srep20728
- Wang, X., Jiang, G. L., Green, M., Scott, R. A., Song, Q., Hyten, D. L., et al. (2014). Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Mol. Genet. Genomics* 289, 935–949. doi: 10.1007/s00438-014-0865-x
- Weir, B. S., and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., et al. (2014). Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics* 15:809. doi: 10.1186/1471-2164-15-809

- Wu, X., Blake, S., Sleper, D. A., Shannon, J. G., Cregan, P., and Nguyen, H. T., (2009). QTL, additive and epistatic effects for SCN resistance in PI 437654. *Theor. Appl. Genet.* 118, 1093–1105. doi: 10.1007/s00122-009-0965-x
- Zhang, J., Li, W., Zhang, L., Dai, H., Ci, D., and Xu, R. (2013). QTL mapping of soybean resistance to whitefly (*Bemisia tabaci* Gennadius) under multi-environment conditions. *Austral. J. Crop Sci.* 7, 1212–1218. Available online at: <https://search.informit.com.au/documentSummary;dn=409760226953952;res=IELHSS>
- Zhou, L., Luo, L., Zuo, J. F., Yang, L., Zhang, L., Guang, X., et al. (2016). Identification and validation of candidate genes associated with domesticated and improved traits in soybean. *Plant Genome* 9, 1–17. doi: 10.3835/plantgenome2015.09.0090
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi: 10.1038/nbt.3096
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Liu, Li, Wen, Fan, Li, Guan, Guo, Wang, Wang and Qiu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.