Genotyping of Soybean Cultivars With Medium-Density Array Reveals the Population Structure and QTNs Underlying Maturity and Seed Traits

Ya-ying Wang¹,²,3†, Yu-qiu Li¹,²,³†, Hong-yan Wu¹, Bo Hu¹,², Jia-jia Zheng¹,², Hong Zhai¹, Shi-xiang Lv¹,4, Xin-lei Liu¹, Xin Chen⁵, Hong-mei Qiu³, Jiayin Yang⁶, Chun-mei Zong⁷, De-zhi Han⁸, Zi-xiang Wen⁹, De-chun Wang⁹ and Zheng-jun Xia¹

¹Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China; ²University of Chinese Academy of Sciences, Beijing, China; ³Soybean Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China; ⁴Heilongjiang Academy of Agricultural Sciences, Harbin, China; ⁵Jiangsu Academy of Agricultural Sciences, Nanjing, China; ⁶Huaixin Institute of Agricultural Sciences in Xuhuai Region of Jiangsu Province, Huairan, China; ⁷Mudanjiang Branch of Heilongjiang Academy of Agricultural Sciences, Mudanjiang, China; ⁸Heihe Branch of Heilongjiang Academy of Agricultural Sciences, Heihe, China; ⁹Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI, United States

Soybean was domesticated about 5,000 to 6,000 years ago in China. Although genotyping technologies such as genotyping by sequencing (GBS) and high-density array are available, it is convenient and economical to genotype cultivars or populations using medium-density SNP array in genetic study as well as in molecular breeding. In this study, 235 cultivars, collected from China, Japan, USA, Canada and some other countries, were genotyped using SoySNP8k iSelect BeadChip with 7,189 single nucleotide polymorphisms (SNPs). In total, 4,471 polymorphic SNP markers were used to analyze population structure and perform genome-wide association study (GWAS). The most likely K value was 7, indicating this population can be divided into 7 subpopulations, which is well in accordance with the geographic origins of cultivars or accession studied. The LD decay rate was estimated at 184 kb, where $r^2$ dropped to half of its maximum value (0.205). GWAS using FarmCPU detected a stable quantitative trait nucleotide (QTN) for hilum color and seed color, which is consistent with the known loci or genes. Although no universal QTNs for flowering time and maturity were identified across all environments, a total of 30 consistent QTNs were detected for flowering time (R1) or maturity (R7 and R8) on 16 chromosomes, most of them were corresponding to known E1 to E4 genes or QTL region reported in SoyBase (soybase.org). Of 16 consistent QTNs for protein and oil contents, 11 QTNs were detected having antagonistic effects on protein and oil content, while 4 QTNs soly for oil content, and one QTN soly for protein content. The information gained in this study demonstrated that the usefulness of the medium-density SNP array in genotyping for genetic study and molecular breeding.

Keywords: soybean, GWAS, flowering time, protein content, oil content, population structure, FarmCPU
INTRODUCTION

Soybean [Glycine max (L.) Merr.] is one of important crops worldwide, providing a sustainable source of high-quality protein feed and vegetable oil. Soybean was domesticated in China more than 5,000–6,000 years ago. Soybean can grow across a wide range of latitudes from 50° N to 35° S (Norman, 1978). Soybean yield related traits such as flowering, maturity, and protein/oil contents are quantitatively inherited traits controlled by internal and external factors (Xia et al., 2013).

Each soybean cultivar adapts to a limited latitudinal region for its maximal yield since soybean is a short day plants with photoperiod sensitivity (Xia et al., 2012b). Flowering time and maturity are important agronomic traits related to soybean adaptability and productivity. More than 200 loci or genes have been mapped to control flowering time in soybean (SoyBase, www.soybase.org). Previous studies identified eleven major-effect loci affecting flowering and maturity in soybean, which have been designated as E1 to E10, and the J locus for “long juvenile period” (Bernard, 1971; Buzzell, 1971; Buzzell and Voldeng, 1980; McBlain and Bernard, 1987; Ray et al., 1995; Bonato and Vello, 1999; Cober and Voldeng, 2001; Cober et al., 2010; Kong et al., 2014; Samanfar et al., 2017). Of these genes, E1, E2, E3, E4, E6, E9, E10, and J have been cloned and functionally characterized (Liu et al., 2008; Watanabe et al., 2009, 2011; Xia et al., 2012a; Zhai et al., 2014a; Zhao et al., 2016; Lu et al., 2017; Samanfar et al., 2017). E1 encodes a nuclear-localized B3 domain-containing protein, suppresses both GmFT2a and GmFT5a expression, two FT orthologs promoting early flowering in soybean (Xia et al., 2012a). E1 expression is suppressed in short day, which is regarded as the main factor for soybean being a short day plant (Xia et al., 2012a; Zhai et al., 2015; Zhang et al., 2016). E2 encodes a homolog of GIGANTEA, controls soybean flowering through regulation of GmFT2a expression but not GmFT5a (Watanabe et al., 2011). E3 and E4 are Phytochrome A (PHYA) genes of GmPHYA3 and GmPHYA2 (Liu et al., 2008; Watanabe et al., 2009). Various allelic combinations of E1, E3 or E4 lead to various photoperiod insensitivity, enabling soybean to adapt to high-latitude environments (Zhai et al., 2014b). J loci is identified as the ortholog of Arabidopsis thaliana EARLY FLOWERING 3 (ELF3), which control flowering time through regulation of E1 expression (Lu et al., 2017). Higher E1 expression in short day enables soybean to grow in the area of lower latitude near equator. E9 and E10 are GmFT2a and GmFT4, FT homolog of Arabidopsis (Zhai et al., 2014a; Zhao et al., 2016). Apart from negative report on existence of E5 loci (Dissanayaka et al., 2016), molecular identities of E7 and E8 are still unknown. Many quantitative trait loci (QTL) or quantitative trait nucleotide (QTN) related to soybean flowering time (first flowering, R1) and maturity have also been documented at SoyBase (http://soybase.org). Many genes or QTL might regulate flowering time through regulation of the expression of the E1 gene (Zhai et al., 2015).

Soybean seed compositions traits such as protein and oil contents are important quality traits in breeding programs. Patil et al. (2017) reviewed molecular mapping and genomic of soybean seed protein, and concluded genetic improvement of soybean protein meal is a complex process because of negative correlation with oil, yield, and the temperature (Patil et al., 2017). Major QTL were repeatedly detected on chromosome (20 (LG I) and 15 (LG E) (Patil et al., 2017). Leamy et al. (2017) studied seed composition traits in wild soybean (Glycine soja) and found 29 SNPs located on ten different chromosomes that are significantly associated with the seven seed composition traits, of which eight SNPs co-localized with QTLs previously uncovered in linkage or association mapping studies conducted with cultivated soybean samples (Leamy et al., 2017). Zhou et al. (2015) mapped major QTN for protein on chromosome 13, 3, 17, 12, 11, and 15 using a 302 accessions (Zhou et al., 2015). More than 100 quantitative trait loci (QTLs) for soybean oil content have been documented at SoyBase (https://www.soybase.org). Cao et al. (2017) found 8 QTLs explained a range of phenotypic variance from 6.3 to 26.3% using RIL population, and qOil-5-1, qOil-10-1, and qOil-14-1 were detected in different environments (Cao et al., 2017). And qOil-5-1 was also detected using natural population and further localized to a linkage disequilibrium block region of approximately 440 kb (Zhang et al., 2017).

Recently, researchers have been applied GWAS in soybean (Bandillo et al., 2015; Wen et al., 2015; Zhang et al., 2015, 2016; Zhou et al., 2015; Contreras-Soto et al., 2017; Fang et al., 2017). Zhang et al. (2015) revealed that genetic loci underlying some agronomically important traits, such as days to flowering, days to maturity, duration of flowering-to-maturity, and plant height in early maturity soybean (Zhang et al., 2015). The ability of GWAS to capture one trait often depends on the frequency of the accessions with contrast phenotypic value in the population being investigated. Recently, as the great advance in sequencing technology, genotyping by sequencing (GBS) has been a choice over other genotyping method, SNP array and traditional SSR markers.

In comparison of traditional lineage analysis, genome-wide association study (GWAS) takes advantage of more historic recombination events that have occurred within natural populations. GWAS has been widely applied to crop plants such as maize (Tian et al., 2011), rice (Huang et al., 2010; Ma et al., 2016). However, in rice, recently studies demonstrates the power of GWAS in combination of biparental association mapping and fine-mapping in dissect agronomic important trait (Huang et al., 2010; Ma et al., 2016).

In this study, we genotyped 235 cultivars using Illumina SoySNP8k iSelect BeadChip; and 4471 core SNP markers were selected. A relatively complex population structure (K = 7) was revealed. GWAS were performed to identify the QTN associated with flowering time and the protein/oil contents using FarmCPU. More than 30 QTN were identified under multiple environments for flowering time and maturity; while 16 consistent QTNs were detected for protein and oil contents.
MATERIALS AND METHODS

Cultivars and Growth Condition
A set of 235 cultivars collected from China, Japan, USA, and Canada were mainly obtained from the Gene Resource Center of Jilin Academy of Agricultural Sciences, China. The origin and other traits for these cultivars are listed in Table S1.

Phenotypic Observation
Soybean accessions were evaluated for photoperiodic responses at six geographic locations: (1) Harbin (hereafter termed as HRB): Research field at the Campus of Northeast Institute of Geography and Agroecology, Harbin, Heilongjiang (45°70’N, 126°64’E); (2) Mudanjiang (hereafter termed as MDJ): Mudanjiang Research Station, Heilongjiang Academy of Agricultural Science (44°42’N, 129°52’E); (3) Gongzhuling (hereafter termed as GZL): Gongzhuling Research Station, Jilin Academy of Agricultural Science, Gongzhuling, Jilin (43°53’N, 124°84’E); (4) Jinan (JN): Campus of Shandong Normal University, Jinan, Shandong (36°66’N, 117°17’E); (5) Huaian (hereafter termed as HA): Huaiyin Research Station, Jiangsu Academy of Agricultural Science, Huaian, Jiangsu (33°57’N, 119°04’E); (6) Nanjing (hereafter termed as NJ): Luhe Research Station, Jiangsu Academy of Agricultural Science, Nanjing, Jiangsu (32°31’N, 118°82’E). At least 15 plants for each cultivar or accession per geographic location were grown in a single row with 20 cm apart for phenotypic evaluation. Days from planting to flowering (R1) and maturity (R7 and R8) were recorded according to Fehr's description (Fehr et al., 1971). R1 refers the beginning of bloom (the opening of the first flower at any node on the main stem). R7 represents the beginning of maturity (one normal pod on the main stem has reached its mature pod color, normally brown or tan); R8 stands for full maturity (95 percent of the pods having reached their mature pod color). For a given cultivar, each specific R stage is defined only when at least 50% of individual plants reached that stage.

Seed were harvested upon maturity. In HRB, GZL, MDJ locations, cultivars that did not reach mature stage (R8) were precluded for maturity and protein/oil content.

Seed coat or hilum color were classified into four groups and coded as follows: (1) yellow or yellowish; (2) green or light brown; (3) brown; (4) black. Seed-weight (100-seedweight) was determined by weighing 3 different set of randomly selected 100 seeds for each cultivar or accession. Seed protein and oil contents were measured using MATRIX-I FT-NIR spectrometer (Bruker). The protein or oil contents were measured three times using different bulk seeds of a given cultivar.

The heritability estimates were calculated using variance components obtained by lm4 of R package (Fehr, 1987).

Genotyping With SNP Markers
DNA was extracted from fresh leaves using the hexadecyltrimethylammonium bromide (CTAB) method with slight modification (Murray and Thompson, 1980; Xia et al., 2007). Due to availability of financial budget, cultivars were divided into two batches (95 cultivars and 140 cultivars) to proceed genotyping. Genotyping using Illumina SoySNP8k iSelect BeadChip (Akond et al., 2013; Yang et al., 2017), which contained a total of 7,189 SNPs and was specifically manufactured by Infinium HD Ultra. SNP genotyping was performed with the Illumina Iscan platform (Illumina, Inc., San Diego, CA). A series of procedures, such as incubation, DNA amplification, preparation of bead assay, hybridization of samples for the bead assay, extension, staining of samples, and imaging of the bead assay, were conducted following previously reported methods (Song et al., 2013). The SNP alleles were called with the Genome Studio Genotyping module (Illumina, Inc.) (Song et al., 2013), and SNP data is available at ftp://159.226.208.134/public/SNP_data.zip (Data Sheet 1).

Population Structure Analysis and GWAS
Population structure analysis was performed using STRUCTURE (Pritchard et al., 2000) and to choose the appropriate number of inferred clusters to model the data, 5 independent runs were performed for each K cluster (2 < K < 3, the length of the burn-in is 10,000, the length of MCMC(Markov chain Monte Carlo) is 10,000). After several attempts, we found that our parameter set was sufficient, longer length of burn-in and MCMC did not change the result significantly. Furthermore, population structure was assessed for K values ranging from 2 to 13 on the entire panel using high quality SNPs. The calculation method of STRUCTURE is based on the Bayesian model. For the simulation result of each K value, STRUCTURE will correspondingly produce the log maximum likelihood value, “LnP(D).” As LnP(D) increases, the K value is closer to the real case. The simulation result with largest LnP(D) and smallest K value is the optimal result (Evanno et al., 2005). The neighboring-joining tree was analyzed using the TASSEL (Version 5.2.38) (Bradbury et al., 2007).

By analyzing r2 value of all pairs of SNPs located within 1 Mb of physical distance, the LD decay trend was found following the regression of negative natural logarithm. Heterozygosis, linkage disequilibrium decade, and kinship plot were generated using GAPIT (Ipka et al., 2012) with default parameters. For kinship plot, a heat map of the values in the values in the kinship matrix is created. Kinship matrix was using the VanRaden kinship algorithm (Tang et al., 2016).

GWAS was conducted the Fixed and random model Circulating Probability Unification (FarmCPU; Liu X. L. et al., 2016) with Bonferroni-corrected threshold with 0.01. This recently developed model selection algorithm takes into account the confounding problem between covariates and test marker by using both Fixed Effect Model (FEM) and a Random Effect Model (REM) (Arora et al., 2017). The first three principal components calculated using GAPIT were used as covariates. The quantile-quantile (Q–Q) plot was used for assessing how fit the model was to account for population structure.

RESULT AND DISCUSSION
Polymorphic SNPs Among the Tested Accessions
Of total 5,039 polymorphic SNP makers, 4,961 were mapped into 20 chromosome (Chr) and 31 scaffolds. Apart from unmapped 78 markers, 4,930 SNP markers were successfully mapped onto 20
chromosomes of the soybean genome (Gmax_275_Wm82.a2.v1; http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Gmax) using the stand-alone BLAST applications (BLAST+) (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST) (Data Sheet 1, ftp://159.226.208.134/public/SNP_data.zip). In order to delimit the influence of batch specific or biased SNP markers on GWAS and population structure analysis, we deleted 459 batch specific or biased SNPs. The unbiased SNP was defined as the frequency of two homogenous nucleotide identities (e.g., AA, GG, or AG) at a given locus in a batch was 0.85 or higher. An unbiased marker having the same two nucleotide identities in two batches were kept for further analysis. According to this threshold of 0.85, 4,471 polymorphic SNP markers were enclosed for population structure and GWAS analysis (Data Sheet 1, ftp://159.226.208.134/public/SNP_data.zip).

Rare SNPs other than two majority nucleotide identities were treated as unknown. Heterozygosis was calculated for both individuals and makers (Figure S1A). By analyzing $r^2$ value of all pairs of SNPs located within 1 Mb of physical distance, the LD decay trend was found following the regression of negative natural logarithm (Figure 1D). The LD decay rate was estimated at 184 kb, where $r^2$ drop to half of its maximum value (0.205). Also this trend was confirmed using GAPIT (Figure S1B). This LD rate calculated is well consistent with previous studies (Zhang et al., 2015; Song et al., 2016).

**Population Structures**

Two hundred thirty five cultivars were originally obtained from different geographic origins, e.g., different latitudinal regions of China, Japan, USA. Apart from 5 landraces, the majority of set of germplasms are modern cultivars (Table S1). According to the population structure, the most likely value of K was 7 and such a portioning of the population was consistent with the significant delta K value (Figures 1A,B). Moreover, this result is also well in accordance with the neighbor-joining tree (Figure 1A). All cultivars are classified into 7 subgroups, which are generally in accordance with their geographic origins, Japan, Northern America, central China, Huang-huai region.
China, Northern area China, landraces (wild soybean) (Figure 1). This classification was also supported by the VanRaden kinship algorithm (Figure 2).

In this study, a relatively complex population structure ($K = 7$) was revealed in comparison of previous reports in which population structures ($K = 2, 4, 9$) were disclosed (Sonah et al., 2015; Liu Z. X. et al., 2016; Fang et al., 2017). After eliminating batch specific or biased markers, the set of 4471 markers might represents the core markers for this set of germplasm (Data Sheet 1, ftp://159.226.208.134/public/SNP_data.zip).

**GWAS on Hilum Color and Seed Coat Color**

Genetic control of seed hilum color has been well documented (Githiri et al., 2007; Oyoo et al., 2011; Cho et al., 2017). We used this trait as a control to monitor the accuracy of our GWAS analysis (Sonah et al., 2015). In this study, only one significant QTN peaked at Gm08_8571052_A_G-0_T_F_2177931718 (Chr08:8601055) was detected (Figure 3A, Table S2). *Chalcone synthase (CHS)* gene has been proved to regulate the hilum color. The significant QTN overlapped a *CHS* gene clustered region in chromosome 8 (Githiri et al., 2007;...
FIGURE 3 | GWAS of seed hilum color, seed coat color, flowering time (R1), protein and oil using FarmCPU. Manhattan plots (bottom) and Quantile-quantile (upper right) plot. Negative log10 P-values from a genome-wide scan are plotted against SNP positions of 20 chromosomes. The horizontal dash line indicates the significant threshold ($2 \times 10^{-5}$). (A) Hilium color at Harbin in 2011; (B) Seed coat color at Harbin in 2011; (C) Flowering time (R1) at Harbin in 2011; (D) Flowering time (R1) at Huaian in 2011; (E) Oil content at Harbin in 2011; (F) Oil content at Huaian in 2011; (G) Protein content at Harbin in 2011; (H) Protein content at Nanjing in 2011.
These CHS genes are mainly controlled by Bloom1 (B1) on chromosome 13, which encodes a transmembrane transporter-like protein for biosynthesis of the bloom in pod endocarp (Zhang et al., 2018). Interestingly, this gene also elevated seed oil contents in domesticated soybeans.

**GWAS on Flowering Time and Maturity**

In this study, flowering time R1 and maturity R7 and R8 were evaluated in six geographic locations. For flowering time, the basic statistics of flowering time (R1) of cultivars were presented in **Table 1**. It took longer days to reach R1 in the northern locations, HRB, MDJ, and GZL (**Figure 4**). Other parameters such as Skewness, Kurtosis, K-S distance, K-S probability, SWilk W, SWilk probability indicated these traits were quantitatively inherited (**Table 1**). The correlation coefficients with a range of 0.592 to 0.978 between R1 of soybean cultivars grown at different locations in 2011 or 2012 (**Table 2**) were all statistically significant.

---

### **Table 1** | The basic statistics of flowering time (R1) of cultivars grown at different locations in 2011 or 2012.

|     | N   | Mean  | Std dev | Std. error | Max  | Min  | Skewness | Kurtosis | K-S dist. | K-S Prob. | SWilk W | SWilk prob |
|-----|-----|-------|---------|------------|------|------|----------|----------|-----------|-----------|---------|-----------|
| HRB_11 | 154 | 66.182 | 16.937 | 1.365 | 111.00 | 47.00 | 0.62 | 0.87 | 0.17 | <0.001 | 0.89 | <0.001 |
| HRB_12 | 156 | 66.622 | 19.244 | 1.541 | 115.00 | 45.00 | 0.89 | 0.38 | 0.17 | <0.001 | 0.87 | <0.001 |
| MDJ_11 | 158 | 51.076 | 17.892 | 1.423 | 96.00 | 27.00 | 0.78 | 0.46 | 0.14 | <0.001 | 0.91 | <0.001 |
| MDJ_12 | 164 | 54.848 | 18.49 | 1.444 | 131.00 | 28.00 | 1.39 | 2.23 | 0.20 | <0.001 | 0.88 | <0.001 |
| GZL_11 | 150 | 46.84 | 18.684 | 1.526 | 91.00 | 26.00 | 0.77 | 0.79 | 0.18 | <0.001 | 0.86 | <0.001 |
| GZL_12 | 147 | 54.455 | 13.179 | 1.087 | 78.67 | 26.33 | 0.10 | 1.27 | 0.14 | <0.001 | 0.93 | <0.001 |
| HA_11 | 188 | 47.417 | 16.306 | 1.258 | 101.00 | 23.00 | 1.47 | 1.64 | 0.17 | <0.001 | 0.83 | <0.001 |
| HA_12 | 150 | 40.053 | 10.301 | 0.819 | 62.00 | 22.00 | 1.28 | 0.51 | 0.26 | <0.001 | 0.80 | <0.001 |
| JA_11 | 173 | 32.52 | 7.599 | 0.578 | 63.00 | 23.00 | 1.35 | 1.70 | 0.22 | <0.001 | 0.85 | <0.001 |
| JA_12 | 150 | 34.529 | 7.338 | 0.556 | 63.00 | 25.00 | 1.22 | 1.45 | 0.18 | <0.001 | 0.88 | <0.001 |
| NJ_11 | 174 | 45.546 | 8.302 | 0.629 | 71.00 | 31.00 | 0.93 | 1.51 | 0.22 | <0.001 | 0.89 | <0.001 |
| NJ_12 | 174 | 31.489 | 8.796 | 0.667 | 61.00 | 16.00 | 0.87 | 1.17 | 0.16 | <0.001 | 0.93 | <0.001 |
Although no consistent QTNs for flowering time and maturity were identified across all environments, a total of 30 consistent QTNs were detected for flowering time (R1) or maturity (R7 and R8) on 16 chromosomes (Figures 3C–H; Table 4; Figures S3–S6; Table S3). In Table 4 and Table S3, we only listed the QTN that has been detected more than three environments. In Table 4, we listed the corresponding QTLs listed in SoyBase or known genes with a physical distance less than 5 Mb.

In chromosome 10 (LG O), we detected a QTN at 45054578 with effect of 7.40 (Table 4; Table S3), which is about 240 kb away from the reported E2 gene (Watanabe et al., 2009). This gene is a major genetic factor controlling flowering time, maturity, geographic adaption in Chinese cultivars (Zhai et al., 2014a; Wang et al., 2016; Fang et al., 2017; Langewisch et al., 2017). In chromosome 19 (LG L), 4 QTN were detected to be significantly associated with flowering time or maturity (Table 4; Table S3). Three QTN at 44839670, 46634511, 46730237 were detected in 5, 12, and 5 environments respectively. QTN at position of 44839670 on chromosome 19 exhibited consistent effect on flowering time or maturity with average of 2.17 day. QTN at 46634511, displayed homogeneous effect on flowering or maturity with average of 3.21 days. In this region, E3 gene, encoding phytochrome A (PHY A), is located from 47633059 to 47641958. The QTN (Gm19_46611973_C_T-1_B_F_2179344248) at 46730237 were detected having four location with positive (suppressing flowering) effect (average of, in QTN for R1 in GZL in 2011 displayed an opposing effect of −6.27 days. In generally, the E3 region is strongly associated with flowering time and domestication (Watanabe et al., 2009; Zhai et al., 2014a; Zhou et al., 2015; Langewisch et al., 2017). The QTN disclosed in this study might this region is very important in term of regulation of flowering time or maturity. However, the authenticity of these QTNs or the relationship with the E3 gene merits further investigation.

On chromosome 6, a QTN (Gm06_10891060_T_C-1_B_F_2179335984) was detected at 10919417 with effect

**Table 3.** The heritability estimates were calculated using variance components obtained by lme4 of R package.

| Groups name            | Variance | Std. dev. | F   | Heritability |
|------------------------|----------|-----------|-----|--------------|
| **STASTICAL ANALYSIS FOR FLOWERING TIME (R1)** |           |           |     |              |
| Cultivar*YEAR          | 0.9737   | 0.9868    | 0.4869 |
| Cultivar*LOC           | 244.9000 | 15.6600   | 48.9800 |
| Cultivar               | 72.8300  | 8.5340    |     |              |
| YEAR                   | 0.0000   | 0.0000    | 0.0000 |
| REP in LOC*YEAR        | 2.6090   | 1.6150    | 0.2609 |
| LOC                    | 47.4000  | 6.8840    |     |              |
| Residual               | 23.0200  | 4.7980    | 2.3020 |
| **STASTICAL ANALYSIS FOR OIL CONTENT (OL)** |           |           |     |              |
| Cultivar*YEAR          | 0.3205   | 0.5661    | 0.16025 |
| Cultivar*LOC           | 2.8742   | 1.6963    | 0.57484 |
| Cultivar               | 1.4405   | 1.2002    |     |              |
| YEAR                   | 0.1168   | 0.3417    | 0.0584 |
| REP in LOC*YEAR        | 0.1153   | 0.3396    | 0.01153 |
| LOC                    | 0.1414   | 0.376     |     |              |
| Residual               | 0.7645   | 0.8744    | 0.07645 |
| **STASTICAL ANALYSIS FOR PROTEIN CONTENT (PR)** |           |           |     |              |
| Cultivar*YEAR          | 3.11     | 1.7635    | 1.555 |
| Cultivar               | 1.6388   | 1.2801    | 0.32776 |
| YEAR                   | 1.6875   | 1.209     |     |              |
| REP in LOC*YEAR        | 0.4832   | 0.6951    | 0.2416 |
| LOC                    | 4.6175   | 2.1488    | 0.46175 |
| Residual               | 1.0955   | 1.0466    |     |              |
| Residual               | 2.4393   | 1.5618    | 0.24393 |

As can be seen in Table 3, the heritability of flowering time and oil content is 0.5833 and 0.6364, respectively. However, the heritability of protein content is only 0.3947, which indicates that these traits are less influenced by genetic factors. The results obtained in this study demonstrate that GWAS can be a powerful tool for identifying genes associated with flowering time and maturity.
of 2.47 day. The E1 gene is located in the pericentromeric region from 2007253 to 2007829 (Xia et al., 2012b) of chromosome 6. Glyma.06G2078001 in phytosome is physically corresponding to the E1 gene, however, this coding region of this gene was annotated from 2007077 to 200794. The lack of polymorphic SNP in the E1 region might account for not being able to detect this major gene. Another Phytochrome A gene, E4, located at Chr20:33236018..33241692 (forward), was reported to be less diversified among Chinese and American cultivars (Zhai et al., 2014b; Langewisch et al., 2017). A QTN (Gm20_34881595_C-T_1_B_F_2179344630) was detected about 3 Mb away from E4 gene. GmFT5a, an FT homolog, located at Chr16:435885..437742 (reverse) about 89 kb from the QTN (Gm16_3598173_C-T_1_B_F_2179342018 with average effect of 3.09) detected (Table 4; Table S3). Other QTNs detected over 3 environments were mapped on chromosome 3, 4, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19 (Figures 3C-H; Table 4; Figures S3–S6; Table S3). Among them, QTN (Gm11_10721006_A_G-1_T_F_2179339194) at 10752436 bp on Chr 11 (LG B1), QTN (Gm12_37315664_A_G-1_T_F_2179339946) at 37271658 on Chr 12 (LG H), QTN(Gm15_1349135_T_C-1_B_F_2179341354) at 1348441 of on Chr 15 (LG E); QTL (Gm18_34401760_G_A-1_T_F_2179343324) at 24606904 on Chr 18 (LG G) were identified in 7 or more environments. Fang et al. (2017) also reported a QTN on chromosome 18 (Fang et al., 2017), whether QTN (Gm18_34401760_G_A-1_T_F_2179343324) is the same as the QTLs reported by other researchers (SoyBase, www.soybase.org) merits further investigation.

In our previous study, the genotypes at E1, E2, E3, and E4 of 180 cultivars revealed great allelic variations at E1 and E3 genes (Zhai et al., 2014b). The power of GWAS to capture a certain trait often depends on the frequency of the accessions with contrast phenotypic value in the population being investigated (Yan et al., 2017). In the previous GWAS studies, fewer QTNs were detected for this trait. When the modern cultivars only a QTN corresponding to E3 was detected at a natural population of 304 short-season soybean lines (K = 9) (Sonah et al., 2015).

Table 4: Physical position, P-value, effect, and distance to known QTL or known genes of QTN for flowering time (R1) and maturity (R7 and R8) detected using FarmCPU.

| Chr | Position | LG | Average of P.value | Average of effect | Distance to known QTL or gene (Kb) | QTL in SoyBase or known gene |
|-----|----------|----|--------------------|-----------------|----------------------------------|----------------------------|
| 3   | 1094352  | N  | 4.05 × 10^-3       | -2.38           | 4,570                            | Pod maturity 19-3 (Guzman et al., 2007) |
| 4   | 6130517  | C1 | 4.38 × 10^-3       | 2.94            | 266                              | Pod maturity 1-1 (Keim et al., 1990) |
| 4   | 36583411 | C1 | 1.78 × 10^-3       | -2.66           |                                  |                            |
| 4   | 39484122 | C1 | 4.23 × 10^-3       | 0.46            |                                  |                            |
| 6   | 10919417 | C2 | 1.81 × 10^-3       | 2.47            | 2,130                            | Pod maturity 13-3 (Specht et al., 2001) |
| 7   | 4918268  | M  | 2.20 × 10^-6       | -5.89           | 92                               | First flower 2-2 (Mansur et al., 1993) |
| 7   | 4928246  | M  | 4.40 × 10^-6       | 8.22            | 82.45                            | First flower 2-2 (Mansur et al., 1993) |
| 8   | 8251563  | M  | 3.16 × 10^-3       | 4.01            | 2,260                            | First flower 6-2 (Orf et al., 1999) |
| 18  | 18036672 | A2 | 3.92 × 10^-3       | 3.74            |                                  |                            |
| 9   | 49446558 | K  | 1.02 × 10^-3       | -2.07           | 4,730                            | First flower 24-4 (Kuroda et al., 2013) |
| 10  | 45054578 | O  | 7.56 × 10^-6       | 7.40            | 240                              | E2 (Watanabe et al., 2011) |
| 11  | 10752436 | B1 | 2.97 × 10^-4       | 3.66            | 83.7                             | First flower 11-2 (Gai et al., 2007) |
| 11  | 28002684 | B1 | 2.70 × 10^-3       | 2.42            | 966                              | First flower 8-4 (Yamanaka et al., 2001) |
| 12  | 37271658 | H  | 9.74 × 10^-4       | 3.74^3          | 535                              | Pod maturity 37-3 (Panthee et al., 2007) |
| 14  | 5766604  | B2 | 8.81 × 10^-4       | 5.52            |                                  |                            |
| 14  | 44255110 | B2 | 2.98 × 10^-4       | -4.11           | 540                              | First flower 21-1 (Reinprecht et al., 2006) |
| 15  | 1348441  | E  | 1.22 × 10^-3       | -4.14           | 1,170                            | Pod maturity 34-4 (Yao et al., 2015) |
| 16  | 2643365  | J  | 3.94 × 10^-3       | -1.58           | 995                              | Pod maturity 19-6 (Guzman et al., 2007) |
| 17  | 3623089  | J  | 4.29 × 10^-3       | 3.09            | 89                               | GmFT5a (Takehama et al., 2018) |
| 17  | 5422636  | D2 | 4.14 × 10^-3       | -2.19           |                                  |                            |
| 18  | 1883973  | G  | 2.18 × 10^-5       | -4.97           | 87.5                             | First flower 21-4 (Reinprecht et al., 2006) |
| 18  | 3737376  | G  | 3.52 × 10^-3       | 2.42            | 3,290                            | Pod maturity 16-2 (Kabelka et al., 2004) |
| 24  | 24606904 | G  | 3.03 × 10^-3       | 2.44            | 2,230                            | Pod maturity 34-5 (Yao et al., 2015) |
| 18  | 45935966 | G  | 3.68 × 10^-3       | -3.09           | 3,240                            | First flower 10-2 (Tasma et al., 2001) |
| 19  | 35744249 | L  | 9.82 × 10^-6       | -4.16           | 1,440                            | First flower 15-2 (Komatsu et al., 2007) |
| 19  | 44839670 | L  | 2.48 × 10^-3       | 2.17            | 343                              | First flower 2-3 (Mansur et al., 1993) |
| 19  | 46634511 | L  | 2.84 × 10^-3       | -3.21           | 125                              | Pod maturity 4-3 (Mansur et al., 1996); First flower 16-4 (Khan et al., 2008) |
| 19  | 46730237 | L  | 2.77 × 10^-3       | 0.27^4          | 437                              | E3 (Watanabe et al., 2009) |
| 20  | 36021032 | I  | 4.61 × 10^-4       | 2.48            | 821                              | E4 (Lu et al., 2008) |

Only QTN that was detected more than three environments were listed.

^3Effect of -3.524 for R1_MDJF_2012 was not counted due to the opposing effect; ^4effect of -6.267737 for R1_QZI_2011 was not counted due to the opposing effect.
While using 892 cultivars (K = 4), only a QTN corresponds to E2 locus was identified (Fang et al., 2017).

No universal QTN was detected over all environments in this study. Common QTNs detected in three or more environments are also informative for us to understand this trait, although authenticity of these QTNs detected in this study need to be verified. GWAS and biparental linkage mapping are commentary each other in mapping and thereafter gene cloning. At present, around 50 biparental populations were generated using the cultivars in this study. We will use these populations to verify the QTN obtained in this study. Fine-mapping or positional cloning will be performed when a novel gene or QTN is verified.

**GWAS of Protein and Oil Contents of Cultivar Seeds**

In this study, protein and oil contents were simultaneously measured in 5 geographic location in 2011 and 2012. The basic statistics of two traits were listed in Table 5 and presented in Figure 5. The parameters such as Skewness, Kurtosis, K-S distance, K-S probability, SWilk W, SWilk probability indicated this trait were quantitatively inherited (Table 5). The correlation coefficients between protein and oil were presented in Table 6. From the correlation coefficients, the protein contents were negatively and significantly correlated to oil content in the same environments or different environments; while the protein contents in an environments was positively correlated to protein contents in other environments (Table 6, Figure 5). The trend was the same for oil contents. According to statistical analysis, the broad sense heritability for oil and protein were 0.6364 and 0.3947. When we used data for protein and oil contents obtained in 9 environments for GWAS using FarmCPU, 16 consistent QTNs for protein and oil contents were detected for oil or protein over 3 environments (Table 7, Table S4: Figures 3G,H, Figures S7, S8). Eleven QTNs were detected having antagonistic effects on protein and oil content, while 4 QTNs soly for oil content, and one QTN soly for protein content. Of eleven QTN for both traits detected over 3 environments, each QTN showed antagonistic effects on protein.

**TABLE 5** | The basic statistics of protein and oil contents of cultivars grown at different locations in 2011 or 2012.

|       | N  | Mean  | Std dev | Std. error | Max  | Min  | Skewness | Kurtosis | K-S dist. | K-S Prob. | SWilk W | SWilk prob |
|-------|----|-------|---------|------------|-------|------|----------|----------|-----------|-----------|---------|----------|
| PR_HRB_11 | 143 | 40.979 | 2.59    | 0.217      | 51.08 | 32.47 | 0.61     | 2.43     | 0.06      | 0.145     | 0.96    | <0.001   |
| OL_HR_B_11 | 143 | 18.952 | 2.325   | 0.194      | 23.61 | 11.83 | -0.56    | 0.05     | 0.07      | 0.103     | 0.98    | 0.016    |
| PR_HRB_12 | 145 | 39.955 | 3.297   | 0.274      | 50.18 | 29.28 | 0.10     | 0.41     | 0.05      | 0.391     | 0.99    | 0.673    |
| OL_HR_B_12 | 145 | 18.347 | 2.28    | 0.188      | 22.53 | 12.13 | -0.74    | 0.12     | 0.12      | <0.001    | 0.95    | <0.001   |
| PR_MDJ_11 | 126 | 39.908 | 3.184   | 0.284      | 50.57 | 32.93 | 0.55     | 0.37     | 0.08      | 0.033     | 0.98    | 0.025    |
| OL_MDJ_11 | 126 | 20.334 | 2.262   | 0.201      | 25.11 | 13.28 | -0.67    | 0.54     | 0.09      | 0.01      | 0.97    | 0.006    |
| PR_MDJ_12 | 129 | 40.299 | 2.789   | 0.246      | 50.06 | 32.79 | 0.51     | 1.38     | 0.07      | 0.164     | 0.98    | 0.018    |
| OL_MDJ_12 | 129 | 20.184 | 2.28    | 0.201      | 24.24 | 12.82 | -0.66    | 0.06     | 0.09      | 0.015     | 0.96    | 0.001    |
| PR_NJ_11  | 140 | 39.679 | 2.672   | 0.226      | 47.21 | 32.75 | 0.20     | -0.21    | 0.04      | 0.653     | 0.99    | 0.712    |
| OL_NJ_11  | 140 | 21.187 | 2.31    | 0.195      | 25.16 | 14.44 | -0.74    | 0.14     | 0.10      | <0.001    | 0.96    | <0.001   |
| PR_NJ_12  | 150 | 42.474 | 2.717   | 0.222      | 50.63 | 36.61 | 0.60     | 0.19     | 0.07      | 0.109     | 0.98    | 0.008    |
| OL_NJ_12  | 150 | 19.612 | 2.033   | 0.166      | 23.54 | 12.86 | -0.71    | 0.45     | 0.10      | 0.001     | 0.96    | <0.001   |
| PR_HA_11  | 164 | 42.222 | 2.949   | 0.23       | 51.19 | 34.47 | 0.14     | -0.08    | 0.04      | 0.649     | 1.00    | 0.953    |
| OL_HA_11  | 164 | 20.383 | 1.918   | 0.15       | 25.09 | 14.09 | -0.55    | 0.84     | 0.05      | 0.273     | 0.98    | 0.011    |
| PR_HA_12  | 168 | 40.091 | 3.002   | 0.232      | 50.72 | 32.13 | 0.33     | 0.52     | 0.04      | 0.651     | 0.99    | 0.175    |
| OL_HA_12  | 168 | 19.928 | 2.986   | 0.177      | 24.18 | 9.78  | -1.02    | 2.44     | 0.07      | 0.066     | 0.95    | <0.001   |
| PR_NJ_12  | 159 | 41.598 | 2.523   | 0.2        | 48.59 | 35.23 | 0.06     | -0.29    | 0.06      | 0.264     | 0.99    | 0.478    |
| OL_NJ_11  | 159 | 20.867 | 1.676   | 0.133      | 24.51 | 16.11 | -0.36    | -0.29    | 0.07      | 0.039     | 0.99    | 0.091    |

Name in the first column or the first row is composed of tran, location, and year. For trait, PR, protein content; OL, oil content; For location, HRB, Harbin; MDJ, Mudanjiang; NJ, Jinan; HA, Huaian; NJ, Nanjing. For years, 11, 2011; 12, 2012. For protein or oil contents, PR, protein content; OL, oil content.
and oil contents, which indicated these QTNs are involved in biological pathway affecting both oil and protein. Major QTL were repeatedly detected on Chromosome 20 (LG I) and 15 (LG E) using America cultivars (Patil et al., 2017). In this study, we detected three QTNs on Chromosome 20 (LG O). Two QTNs were identified for both traits. QTN (Gm20_2372509_T_C-L_T_R_2179344425) at position of 2366428 with antagonistic effects on protein (0.431691) and oil (−0.45203) and QTN (Gm20_7927513_A_G-1_T_F_2179344472) with antagonistic effects on protein(0.76146) and oil (−0.47998). Another QTN (Gm20_38151772_C_T-1_T_R_2179344711) for oil with effect of −0.53353 was identified on chromosome 20. We did not detect any consistent QTN on Chr 15 (LG E). All 16 QTNs mapped in this study (Table 7) were physically near (less than 5 Mb) QTL reported in SoyBase.

### Conclusion and Further Consideration

Instead of traditional molecular markers, e.g., SSR, AFLP, advances in sequencing technologies have enabled high-density array and GBS to be widely applied to genomic and genetic study to dissect genetic population structure and GWAS (Sonah et al., 2013; Bandillo et al., 2015; Wen et al., 2015; Zhang et al., 2015; Contreras-Soto et al., 2017; Fang et al., 2017; Yan et al., 2017). However, this study employed a medium density array to reveal population genetic structure, the result showed the quality of the population genetic study has been improved by elimination of some batch specific or biased SNPs. Also the GWAS quality has been monitored using hilum color and seed coat color. Fast genotyping method e.g., using a set of core SNP array is in high demand for genetic study or molecular breeding (Chaudhary et al., 2015).

The information gained in this study demonstrated that the usefulness of the medium-density SNP array in genotyping for genetic study and molecular breeding.

Up to date, there are a large number of loci or QTL have been identified by GWAS using different set of natural population or by linkage or association mapping using biparental populations under different environments in different years. In generally, the effect of each locus is rather small, its detection might be influenced by population size, population structure, accuracy of phenotyping, physical location of the causal gene (e.g., pericentromeric region), epistatic association between QTLs as well as environmental factors. High negative correlation coefficients between oil and protein content in soybean was revealed in this study, which is consistent with previous reports (Böydak et al., 2002; Karamaslan et al., 2008); common regions or loci might have favorable effect on one and unfavorable effect on the other. The higher negative correlation coefficients of two traits might reflect that we might be able to detect QTL or QTN with higher effect on both traits. Hwang et al. (2014) found seven of 13 regions associated with oil content also have effect on protein content (Hwang et al., 2014). Similarly, in this study, we have detected 11 common QTNs associated with oil and antagonistically associated with protein, although no universal QTN detected over all environments. However, the overall oil and protein content can be varied to a great extent, also the environmental effect e.g., latitudinal
### TABLE 7

| Trait     | Chr | LG  | Position | P-value | Effect on PR | Effect on OL | Distance to known QTL or gene | QTL information from SoyBase |
|-----------|-----|-----|----------|---------|--------------|--------------|--------------------------------|------------------------------|
| PR/OL     | 1   | D1a | 8869097  | 0.002549| 0.00955      | −0.62331     | 1,140                          | Seed protein 3-5 (Brummer et al., 1997) |
|           | 5   | A1  | 37361373 | 0.002501| −0.5009      | 0.387996     | 346                            | Seed protein 41-1 (Jun et al., 2003) |
|           | 8   | A2  | 8613057  | 0.001182| 2.506472     | −1.08523     | 17                             | Seed protein 26-1 (Reinprecht et al., 2006) |
|           | 13  | F   | 13865497 | 0.000118| 1.267005     | −0.75341     | 382                            | Seed protein 36-22 (Mao et al., 2013) |
|           | 16  | J   | 4582681  | 0.002254| 1.1395       | −0.50493     | 302                            | Seed protein 24-4 (Qi et al., 2011) |
|           | 17  | D2  | 11939572 | 0.000118| 1.267005     | −0.75341     | 382                            | Seed protein 36-22 (Mao et al., 2013) |
|           | 18  | G   | 3737376  | 0.004217| 0.477983     | −0.31451     | 111                            | Seed protein 20-1 (Panthee et al., 2005) |
|           | 18  | G   | 43143230 | 0.000556| 1.004969     | −0.44868     | 1,612                          | Seed oil 42-33 (Qi et al., 2015) |
| PR        | 5   | A1  | 37987063 | 0.002457| 0.621571     | −0.87557     | 31                             | Seed oil 23-6 (Hyten et al., 2004) |
| OL        | 7   | M   | 8251563  | 0.000152| −0.93037     | 0.502982     | 423                            | Seed oil 43-27 (Mao et al., 2013) |
|           | 8   | A2  | 3823489  | 0.00048 | −0.45203     | 0.420013     | 319                            | Seed oil 14-3 (Csanádi et al., 2001) |
|           | 11  | B1  | 10752436 | 0.001861| −0.77553     | 0.420013     | 319                            | Seed oil 14-3 (Csanádi et al., 2001) |
|           | 20  | I   | 20469935 | 0.002656| 0.76146      | −0.47998     | 3,710                          | Seed oil 2-2 (Csanádi et al., 2001) |
|           | 20  | I   | 39264676 | 0.002104| −0.53353     | 0.420013     | 1,002                          | Seed oil 24-1 (Qi et al., 2011) |

Only QTN that was detected more than three environments were listed.

GWAS in combination with biparental populations such as RIL, NIL, CSSL, is very powerful for QTL identification and their gene cloning. As high throughput sequencing data aggregate, the important QTL or QTN detected by traditional linkage mapping or GWAS will be verified and subsequently cloned. As most components of a molecular or signaling pathway have been identified (Gentzbittel et al., 2015), information of gene regulation or crosstalk with different pathways will enable us to build a genetic network that can be used in molecular design breeding.

**AUTHOR CONTRIBUTIONS**

ZX conceived this project; YW, YL performed the most experiments in the laboratory; HW, BH, HZ, SL, XL, XC, HQ, JY, CZ, DH conducted field experiment and phenotypic observation; JZ, ZW, ZX: performed data analysis including GWAS; ZX, YW, and YL wrote the article; DW contributed to scientific discussions and critical revision of manuscript. All authors reviewed the final manuscript.
This work was supported by National Key R&D Program of China (2016YFD0101902 and 2016YFD0100201); by Strategic Priority Research Program of the Chinese Academy of Sciences (XDA0801010503); and by Programs (31471518, 31771869, 31771818) from National Natural Science Foundation of China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Professor Kyuya Harada (Dept. of Biotechnology, Osaka University, Japan) and Professor Liuling Yan (Dept. of Plant and Soil Sciences, Oklahoma State University, USA) for critical comments and English editing. Also thanks to Scientific Data Center of Northeast Black Soil, National Earth System Science Data Sharing Infrastructure, National Science and Technology Infrastructure of China, (http://northeast.geodata.cn).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00610/full#supplementary-material

REFERENCES

Bonato, E. R., and Vello, N. A. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22, 229–232. doi: 10.1590/S1415-47571999000200016
Boyda, E., Alpaslan, M., Hayta, M., Gereck, S., and Simsek, M. (2002). Seed composition of Soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. J. Agric. Food Chem. 50, 4718–4720. doi: 10.1021/jf0253331
Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramirez, Y., Breseme, S. R., et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
Brummer, E. C., Graef, G. L., Orf, J., Wilcox, J. R., and Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37, 370–378. doi: 10.2135/cropsci1997.0011183X003700020011x
Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. Newsletter 7, 26–29.

Bonato, E. R., and Vello, N. A. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22, 229–232. doi: 10.1590/S1415-47571999000200016
Boyda, E., Alpaslan, M., Hayta, M., Gereck, S., and Simsek, M. (2002). Seed composition of Soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. J. Agric. Food Chem. 50, 4718–4720. doi: 10.1021/jf0253331
Bradbury, P. J., Zhang, Z., Kroum, 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
Brumner, E. C., Graef, G. L., Orf, J., Wilcox, J. R., and Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37, 370–378. doi: 10.2135/cropsci1997.0011183X003700020011x
Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. Newsletter 7, 26–29.

Bonato, E. R., and Vello, N. A. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22, 229–232. doi: 10.1590/S1415-47571999000200016
Boyda, E., Alpaslan, M., Hayta, M., Gereck, S., and Simsek, M. (2002). Seed composition of Soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. J. Agric. Food Chem. 50, 4718–4720. doi: 10.1021/jf0253331
Bradbury, P. J., Zhang, Z., Kroum, 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
Brumner, E. C., Graef, G. L., Orf, J., Wilcox, J. R., and Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37, 370–378. doi: 10.2135/cropsci1997.0011183X003700020011x
Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. Newsletter 7, 26–29.

Bonato, E. R., and Vello, N. A. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22, 229–232. doi: 10.1590/S1415-47571999000200016
Boyda, E., Alpaslan, M., Hayta, M., Gereck, S., and Simsek, M. (2002). Seed composition of Soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. J. Agric. Food Chem. 50, 4718–4720. doi: 10.1021/jf0253331
Bradbury, P. J., Zhang, Z., Kroum, 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
Brumner, E. C., Graef, G. L., Orf, J., Wilcox, J. R., and Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37, 370–378. doi: 10.2135/cropsci1997.0011183X003700020011x
Buzzell, R. I., and Voldeng, H. D. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. Newsletter 7, 26–29.
Buzzell, R. I. (1971). Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can. J. Genet. Cytol.*, 13, 703–707. doi: 10.1139/g71-100

Cao, Y., Li, S., Wang, Z., Chang, F., Kong, J., Gai, J., et al. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Front. Plant Sci.* 8:1222. doi: 10.3389/fpls.2017.01222

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyudan, B., et al. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Front. Plant Sci.* 6:1021. doi: 10.3389/fpls.2015.01021

Cho, Y. B., Jones, S. L., and Vodkin, L. O. (2017). Mutations in Argonaute5 illuminate epistatic interactions of the K1 and k loci leading to saddle seed color patterns in *Glycine max*. *Plant Cell* 29, 708–725. doi: 10.1105/tpc.17.00162

Cober, E. R., and Voldeng, H. D. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Sci.* 41, 698–701. doi: 10.2135/cropsci2001.411698x

Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. (2010). A new locus for early maturity in soybean. *Crop Sci.* 50, 524–527. doi: 10.2138/cropsci2009.04.0174

Contreras-Soto, R. L., Mora, F., de Oliveira, M. A. R., Higashi, W., Scapim, C. A., and Chen, S. (2007). A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in *soybean* (*Glycine max*). *Theor. Appl. Genet.* 114, 300–309. doi: 10.1007/s00122-008-0792-5

Dissanayaka, A., Rodriguez, T. O., Di, S., Yan, F., Githiri, S. M., Rodas, F. R., et al. (2016). Quantitative trait locus mapping of soybean maturity gene ES. *Breed. Sci.* 66, 407–415. doi: 10.1270/jsbbs.15160

Eskandari, M., Cober, E. R., and Rajcan, I. (2013). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RI populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126, 483–495. doi: 10.1007/s00122-012-1995-3

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

Fang, C., Ma, Y. M., Wu, S. W., 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

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.011813X001000060051x

Fehr, W. R. (1987). *Principles of Cultivar Development*. A. W. Fehr, C. E. Caviness, D. T. Burmood, and J. S. Pennington. (1971). Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can. J. Genet. Cytol.*, 13, 703–707. doi: 10.1139/g71-100

Gai, J., Wang, Y., Wu, X., and Chen, S. (2007). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. *Crop Sci.* 41, 698–701. doi: 10.2135/cropsci2001.411698x

Hyten, D. L., Pantalone, V. R., Sams, C. E., Saxton, A. M., Landau-Ellis, D., Stefinaki, T. R. et al. (2004). Seed quality QTL in a prominent soybean population. *Theor. Appl. Genet.* 109, 552–561. doi: 10.1007/s00122-004-1661-5

Jun, T. H., Van, K., Kim, M. Y., Lee, S. H., and Walker, D. R. (2008). Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162, 179–191. doi: 10.1007/s10681-007-9491-6

Keim, P., Diers, B. W., Fehr, W. R., LeRoy, A. R., Baianu, I. C., You, T., et al. (2004). Putative alleles for increased yield from soybean plant introductions. *Crop Sci.* 44, 784–791. doi: 10.2135/cropsci2004.7840

Kurasalas, D., Boydak, E., Turkgolu, H., and Hakam, M. (2008). Effect of different seed rates on oil and protein content and fatty acid composition of soybean seeds. *Asian J. Chem.* 20, 2115–2124.

Liang, H. Z., Yu, Y. L., Wang, S. F., Lian Y., Wang, T. F., Wei, Y. L., et al. (2010). Naturally occurring diversity helps to reveal genes of adaptive trait variation in soybean. *Theor. Appl. Genet.* 121, 949–959. doi: 10.1007/s00122-008-0792-5

Lee, S. H., Bailey, M. A., Mian, M. A. R., Carter, T. E., Shipe, E. R., Ashley, D. A., et al. (1996). RFLP loci associated with soybean seed protein and lipid content across populations and locations. *Theor. Appl. Genet.* 93, 649–657. doi: 10.1007/BF00224058

Li, X. L., Hu, M., Fan, B., Buckler, E. S., and Zhang, Z. W. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:e1005767. doi: 10.1371/journal.pgen.1005767

Liu, B. H., Kanazawa, A., Matsumura, H., Takahashi, R., Murakami, K., and Abe, J. (2008). QTL analysis of cleistogamy in soybean. *Theor. Appl. Genet.* 117, 479–487. doi: 10.1007/s00122-007-0792-5

Liu, X. L., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. W. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12:e1005767. doi: 10.1371/journal.pgen.1005767

Mansur, L. M., Lark, K. G., Kross, H., and Oliveira, A. (1993). Interval mapping for QTL in *soybean* (*Glycine max*). *Theor. Appl. Genet.* 86, 907–913. doi: 10.1007/BF00211040
Mansur, L. M., Orf, J. H., Chase, K., Jarvik, T., Cregan, P. B., and Lark, K. G. (1996). Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci.* 36, 1327–1336. doi: 10.2135/cropsci1996.0011183X003600050042x

Mao, T., Jiang, Z., Han, Y., Teng, W., Zhao, X., and Li, W. (2013). Identification of quantitative trait loci controlling seed protein and oil contents of soybean across multi-genetic backgrounds and environments. *Plant Breed.* 132, 630–641. doi: 10.1111/pbr.12091

McBlain, B. A., and Bernard, R. L. (1987). A new gene affecting the timing of flowering and maturity in soybean. *J. Hered.* 78, 160–162. doi: 10.1093/oxfordjournals.jhered.a110349

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321

Norman, A. G. (1978). *Soybean Physiology, Agronomy, and Utilization.* New York, NY: Academic Press. doi: 10.1097/00006094-19790400-00003

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

Oyoo, M. E., Benitez, E. R., Kurosaki, H., Ohnishi, S., Miyoshi, T., Kiribuchi-Otobe, C., et al. (2011). QTL Analysis of soybean seed coat discolouration associated with II TT genotype. *Crop Sci.* 51, 464–469. doi: 10.2135/cropsci2010.02.0121

Panthee, D. R., Pantalone, V. R., Saxton, A. M., West, D. R., and Sams, C. E. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45, 2015–2022. doi: 10.2135/cropsci2004.0720

Pathan, S. M., Vuong, T., Clark, K., Lee, J. D., Shannon, J. G., Roberts, C. A., 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

Patil, G., Rain, R., Vuong, T., Panthee, V., Song, Q., Chen, P., et al. (2017). Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* 130, 1975–1991. doi: 10.1007/s00122-017-2955-8

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *Am. J. Hum. Genet.* 67, 170–181. doi: 10.1086/302939

Qi, Z. M., Xue, H. A. N., Sun, Y. N., Qiong, W. U., Shan, D. P., Du, X. Y., et al. (2011). An integrated quantitative trait locus map of oil content in soybean. *Glycine max* (L.) Merr., generated using a meta-analysis method for mining genetic loci for oil protein in two recombinant inbred line populations of soybean. *Mol. Genet. Genomics* 289, 935–949. doi: 10.1007/s00438-014-0665-x

Watanabe, S., Hideshima, R., Xia, Z. J., Tsukobuka, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics* 182, 1251–1262. doi: 10.1053/j.genetics.2010.03.008

Watanabe, S., Xia, Z. J., Hideshima, R., Tsukobuka, Y., Sato, S., Yamanaoka, N., et al. (2011). A Map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *GIGANTEA* 138, 395–407. doi: 10.1053/j.genetics.2011.02.002

Xia, Z., Tsukobuka, Y., Hoshii, M., Hanawa, M., Yano, C., Okamura, K., et al. (2007). An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F2 population. *DNA Res.* 14, 257–269. doi: 10.1093/dnares/dsm027

Xia, Z., Watanabe, S., Yamada, T., Tsukobuka, Y., Nakashima, H., Zhai, H., et al. (2012a). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21255–21264. doi: 10.1073/pnas.1117982109

Xia, Z. J., Zhai, H., Liu, B. H., Kang, F. J., Yuan, X. H., Wu, H. Y., et al. (2012b). Molecular identification of genes controlling flowering time, maturity, and photoperiod response in soybean. *Plant Syst. Evol.* 298, 1217–1227. doi: 10.1007/s10535-012-0656-9

Xia, Z., Zhai, H., Hu, L., Wu, H., and Zhang, Y. (2013). Recent achievement in gene cloning and functional genomics in soybean. *Sci. World J.* 2013:281367. doi: 10.1155/2013/281367

Yamanaka, N., Ninomiya, S., Hoshi, M., Tsubokura, Y., Nagamura, Y., et al. (2012). An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F2 population. *DNA Res.* 14, 257–269. doi: 10.1093/dnares/dsm027

Yamanaoka, N., Ninomiya, S., Hoshi, M., Tsukobuka, Y., Yano, M., Nagamura, Y., et al. (2001). An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA Res.* 8, 61–72. doi: 10.1093/dnares/8.2.61

Yao, D., Hofmann, N., Li, S. X., Ferreira, M. E., Song, B. H., Jiang, G. L., et al. (2017). Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. *BMC Genomics* 18:529. doi: 10.1186/s12864-017-3922-0

Yang, G., Zhai, H., Hu, H. Y., Zhang, X. Z., Lu, S. X., Wang, Y. Y., et al. (2017). QTL effects and epistatic interaction for flowering time, leaflet morphology and branching number in a soybean mapping population of JapaneseChinese cultivars. *J. Integr. Agric.* 16, 990–1912. doi: 10.1016/S2075-319X(16)61539-3

Yao, D., Liu, Z., Zhang, J., Liu, S. Y., Qu, J., Guan, S., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14, 6101–6109. doi: 10.4238/2015.June.8.8

Zhai, H., Lu, S., Liang, S., Wu, H., Zhang, X., Liu, B., et al. (2014a). GmFT4, a homolog of FLOWERING LOCUS T, is positively regulated by
E1 and functions as a flowering repressor in soybean. *PLoS ONE* 9:e89030. doi: 10.1371/journal.pone.0089030

Zhai, H., Lu, S., Wang, Y., Chen, X., Ren, H., Yang, J., et al. (2014b). Allelic variations at four major maturity E genes and transcriptional abundance of the E1 gene are associated with flowering time and maturity of soybean cultivars. *PLoS ONE* 9:e97636. doi: 10.1371/journal.pone.0097636

Zhai, H., Lu, S., Wu, H., Zhang, Y., Zhang, X., Yang, J., et al. (2015). Diurnal expression pattern, allelic variation, and association analysis reveal functional features of the E1 gene in control of photoperiodic flowering in soybean. *PLoS ONE* 10:e0135909. doi: 10.1371/journal.pone.0135909

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4

Zhang, D., Sun, L., Li, S., Wang, W., Ding, Y., Swarm, S., et al. (2018). Elevation of soybean seed oil content through selection for seed coat shininess. *Nat. Plants* 4, 30–35. doi: 10.1038/s41477-017-0084-7

Zhang, X., Zhai, H., Wang, Y., Tian, X., Zhang, Y., Wu, H., et al. (2016). Functional conservation and diversification of the soybean maturity gene E1 and its homologs in legumes. *Sci. Rep.* 6:29548. doi: 10.1038/srep29548

Zhang, D., Zhao, M., Li, S., Sun, L., Wang, W., Cai, C., et al. (2017). Plasticity and innovation of regulatory mechanisms underlying seed oil content mediated by duplicated genes in the palaeopolyploid soybean. *Plant J.* 90, 1120–1133. doi: 10.1111/tpj.13533

Zhao, C., Takeshima, R., Zhu, J., Xu, M., Sato, M., Watanabe, S., et al. (2016). A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a FLOWERING LOCUS T ortholog. *BMC Plant Biol.* 16:20. doi: 10.1186/s12870-016-0704-9

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 © 2018 Wang, Li, Wu, Hu, Zheng, Zhai, Lv, Liu, Chen, Qiu, Yang, Zong, Han, Wen, Wang and Xia. 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) and the copyright owner 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.