Genome-Wide Association Study to Map Genomic Regions Related to the Initiation Time of Four Growth Stage Traits in Soybean

Wenliang Yan1,2, Benjamin Karikari3, Fangguo Chang1, Fangzhou Zhao1, Yinghu Zhang4, Dongmei Li1,2, Tuanjie Zhao1* and Haiyan Jiang2*

1 Key Laboratory of Biology and Genetics and Breeding for Soybean, Ministry of Agriculture, State Key Laboratory for Crop Genetics and Germplasm Enhancement, National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China, 2 College of Artificial Intelligence, Nanjing Agricultural University, Nanjing, China, 3 Department of Crop Science, Faculty of Agriculture, Food and Consumer Sciences, University for Development Studies, Tamale, Ghana, 4 Institute of Agricultural Sciences in Jiangsu Coastal Region, Yancheng, China

The time to flowering (DF), pod beginning (DPB), seed formation (DSF), and maturity initiation (DMI) in soybean (Glycine max [L.] Merr) are important characteristics of growth stage traits (GSTs) in Chinese summer-sowing soybean, and are influenced by genetic as well as environmental factors. To better understand the molecular mechanism underlying the initiation times of GSTs, we investigated four GSTs of 309 diverse soybean accessions in six different environments and Best Linear Unbiased Prediction values. Furthermore, the genome-wide association study was conducted by a Fixed and random model Circulating Probability Unification method using over 60,000 single nucleotide polymorphism (SNP) markers to identify the significant quantitative trait nucleotide (QTN) regions with phenotypic data. As a result, 212 SNPs within 102 QTN regions were associated with four GSTs. Of which, eight stable regions were repeatedly detected in least three datasets for one GST. Interestingly, half of the QTN regions overlapped with previously reported quantitative trait nucleotide (QTN) regions with phenotypic data. As a result, 212 SNPs within 102 QTN regions were associated with four GSTs. Of which, eight stable regions were repeatedly detected in least three datasets for one GST. Interestingly, half of the QTN regions overlapped with previously reported quantitative trait loci or well-known soybean growth period genes. The hotspots associated with all GSTs were concentrated on chromosome 10. E2 (Glyma10g36600), a gene with a known function in regulating flowering and maturity in soybean, is also found on this chromosome. Thus, this genomic region may account for the strong correlation among the four GSTs. All the significant SNPs in the remaining 7 QTN regions could cause the significant phenotypic variation with both the major and minor alleles. Two hundred and seventy-five genes in soybean and their homologs in Arabidopsis were screened within ± 500 kb of 7 peak SNPs in the corresponding QTN regions. Most of the genes are involved in flowering, response to auxin stimulus, or regulation of seed germination, among others.

The findings reported here provide an insight for genetic improvement which will aid in breeding of soybean cultivars that can be adapted to the various summer sowing areas in China and beyond.

Keywords: soybean, growth stage traits, environmental and genetic factors, quantitative trait nucleotide, genetic improvement
INTRODUCTION

Soybean (Glycine max [L.] Merr.) is a versatile crop with numerous nutritional qualities and uses. It is the highest source of plant protein and second highest oil yielder among oilseed crops (Kole et al., 2015). As a photoperiod-sensitive and self-pollinated species, the growth stage traits (GSTs) play important roles in the adaptability and yield improvement of soybean (Li M. et al., 2019). Among the GSTs, flowering time and maturity are important agronomic traits related to soybean productivity (Wang et al., 2018). Furthermore, characteristics related to pods and seeds are also often regarded as yield-related traits, or important components of yield in soybean (Du et al., 2017; Li M. et al., 2019; Song et al., 2020), which emphasizes the importance of understanding the genetics underlying the initiation time of GSTs.

The GSTs are controlled by both genetic (G) and environmental (E) factors, which make breeding efforts utilizing traditional, conventional methods both ineffective and inefficient, leading to limited progress having been made on soybean adaptability, biomass, and economic yield (Cober and Morrison, 2010). In addition, the genetic mechanisms governing soybean flowering time to maturity are complex, regulated by polygenes with both major and minor effects (Zhang et al., 2015). So far, several locigenes have been reported to regulate flowering in soybean, prominent among them include the E-series (E1–11) (Bernard, 1971; Buzzell and Voldeng, 1980; Mcllain and Bernard, 1987; Kang and Kim, 1996; Bonato and Vello, 1999; Molnar et al., 2003; Watanabe et al., 2009; Cober et al., 2010; Kong et al., 2014; Samanfar et al., 2017; Wang et al., 2019), J (Ray et al., 1995), D1 (Liu et al., 2008), and D2 (Zhang D. et al., 2019). It has been reported that these loci interact with photoperiod in modulating flowering and maturity. Under natural day length, the dominant alleles tend to delay flowering time and maturity, but the magnitude of effect of each gene can be different.

To date, SoyBase database1 currently lists 356 quantitative trait loci (QTLs) for first flower, 11 QTLs for pod beginning (Tasma et al., 2001), and 5 QTLs for seed formation (Mansur et al., 1993; Palomeque et al., 2010). Genetic studies on the initiation traits of the soybean pod and seed are relatively few. The pod-bearing period is the most vigorous growth stage of soybean and also the maximum phase when dry matter forms, partitions, and accumulates (Hu et al., 2019). Reasonable pod-beginning time and as many pods as possible will directly lead to a high yield, especially in vegetable soybean varieties (Li X. et al., 2012; Zhang et al., 2021). Asakura et al. (2012) and Zhang et al. (2016) found more than one hundred genes that were highly expressed exclusively at young seed stages, whereas Karikari et al. (2020) identified three hub genes (Glyma06g44510, Glyma08g06420, and Glyma19g28070) that were expressed relatively more highly in large-seeded cultivars than small-seeded cultivars at seed formation stage.

With the advance of next-generation sequencing in recent years, the genome-wide association study (GWAS) can be more efficiently applied in germplasm collections or naturally occurring populations and provides higher resolution in terms of defining the genomic positions of QTLs/genes (Shirasawa et al., 2013). In addition, the first whole-genome sequence of the variety Williams 82, completed in 2010, provides a powerful resource for soybean functional genomic research (Schmutz et al., 2010). Moreover, multi-locus GWAS methods, e.g., Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al., 2016), are more powerful and robust in detecting loci associated with polygenic and/or multi-factorial complex traits than single-locus methods (Zhang Y. et al., 2019; Sertse et al., 2021), where single-locus models tend to only be able to detect large-effect QTNs (He et al., 2019b).

Favorable alleles could be successfully introgressed by using marker-assisted selection (MAS), if they are independent of the environmental and genetic background (Palomeque et al., 2010). This means that genetic studies of the same trait in multiple environments seems more scientifically viable to identify stable quantitative trait nucleotides (QTNs) for practical plant breeding. In this study, we evaluated 309 diverse soybean accessions across six different agricultural experimental environments to record four GSTs [i.e., days to flowering (DF), pod beginning (DPB), seed formation (DSF), and maturity initiation (DMI)], then conducted a FarmCPU GWAS with individual environment and best linear unbiased prediction (BLUP) values by utilizing 61,174 single nucleotide polymorphism (SNP) markers to identify QTN regions. Also, we mined candidate genes around the stable QTN regions for functional genomics studies in order to unravel the molecular mechanisms underlying the studied GSTs. Our overall objective was to provide a valuable insight into the genetic architecture of soybean GSTs and genetic resources for accelerating soybean genomic breeding, aimed at yield and yield stability as well as wider adaptability to multiple ecological zones within China and beyond.

MATERIALS AND METHODS

Plant Materials and Experiment Design

A set of 309 breeding lines adapted to the Chinese Yangtze-Huai region, primarily utilized for grain and vegetable were used in this experiment. This population is designated as Yangtze-Huai Soybean Breeding Line Population (YHSBLP). Seeds for the panel were obtained from the National Center for Soybean

---

1https://www.soybase.org/
Improvement, Nanjing Agricultural University, Nanjing, in Jiangsu Province, China.

All soybean materials were planted via a randomized complete block design with three replications in six different environments across 2 years (2018 and 2019) viz., Jiangpu (JP) Experimental Station (32°12′N and 118°37′E), and Yancheng (YC) City (33°21′N and 120°09′E) in Jiangsu province, and Dangtu (DT) Experimental Station (31°34′N and 118°29′E) in Anhui province. 2018 planting was done on June 22 in YC and July 3 in JP and these were designated 18YC and 18JP, respectively. 2019 planting was done on June 21 in DT, June 24 in JP, June 19 in YC (19YC6), and July 4 (19YC7), according to the local soybean cropping calendar. Standard cultural and agronomic practices were followed similarly in each environment.

Phenotypic Evaluation
Four traits related to GSTs viz., DF, DPB, DSF, and DMI were recorded following the Fehr’s classification method (Fehr et al., 1971). Briefly, DF was recorded at the R1 stage (the day when 50% of the plants on a plot have an open flower at one of the top four nodes with a fully expanded leaf); DPB represents days from emergence to a 1/4 inch (0.5 cm) long pod at one of the four uppermost nodes on the main stem with a fully developed leaf at the R3 stage; DSF refers to days from emergence to a 1/8 inch (0.3 cm) long seed in one pod present at any of the four uppermost nodes on the main stem with a fully developed leaf at the R5 stage, and DMI were recorded at the R7 stage (days from emergence to 50% of the leaves turning yellow and one pod having reached full mature coloration).

Statistical Analysis
Descriptive statistics such as mean, standard deviation (SD), maximum and minimum trait value, and coefficient of variation (CV) (%) were calculated using SAS version 9.3 (SAS Institute, 2010, Inc., Cary, NC, United States). Analysis of variance (ANOVA) was conducted to evaluate the effects of genotype (G), environment (E), and genotype-by-environment interaction (G × E) on the RGS traits in the joint environments using the PROC general linear model (GLM) in SAS software. The statistical model for the ANOVA was $y_{ijk} = \mu + G_i + E_j + G_i E_j + R_{k(i)} + \epsilon_{ijk}$, where $y_{ijk}$ stands for the individual observation of $ijk$th experimental unit; $\mu$ is the total mean, $G_i$ is the effect of the $i$th genotype, $E_j$ is the effect of the $j$th environment, $G_i E_j$ is the interaction effect between the $i$th genotype and the $j$th environment, $R_{k(i)}$ is the effect of the $k$th block within the $j$th environment, and $\epsilon_{ijk}$ is a random error following $N(0, \sigma^2_e)$. Broad-sense heritability ($h^2$) was calculated as: $h^2 = \frac{\sigma^2_g}{\sigma^2_g + \frac{\sigma^2_e}{n} + \frac{\sigma^2_r}{nr}}$ for the combined environments, where $\sigma^2_g$ is the genotypic variance, $\sigma^2_E$ is the G × E interaction variance, $\sigma^2_e$ is the error variance, $n$ is the number of environments, and $r$ is the number of replications. Pearson correlation coefficients among the traits were estimated and visualized with Corrplot package in R (Taiyun and Viliam, 2017).

To minimize the effects of environmental variation, the best linear unbiased predictions (BLUPs) of individual lines for each trait were calculated using the R package lme4 (Bates et al., 2014). Frequency distribution of phenotypic BLUPs was plotted by using OriginPro 2020 Statistical Software (Origin Corporation, Northampton, MA, United States).

SNP Genotyping
The present study used restriction-site-associated DNA sequencing (RAD-seq) approach to sequence the genomic DNA of 309 lines of YHSBLP, and this sequencing was carried out by Beijing Genomics Institution, Shenzhen, China. First, the genomic DNA of all 309 soybean lines was extracted from young and healthy leaf samples with a modified CTAB method (Allen et al., 2006). Taq I enzyme was used to digest this genomic DNA for the construction of a genomic DNA library. The DNA fragments of 400–700 bp were selected and sequenced using an Illumina HiSeq 2,000 standard protocol for multiplexed shotgun genotyping, and 90-mer paired-end reads were generated (Andolfatto et al., 2011). All the sequence reads were aligned to the reference genome of Williams 82 (genome assembly I annotation version 1.1) (Schmutz et al., 2010), using the SOAP2 software (Li et al., 2009). Based on the Bayesian estimation of the site frequency, RealSFS was utilized for the SNP calling (Yi et al., 2010). The SNP data was screened at a rate of missing and heterozygous allele calls ≤ 30%, and then the missing genotypes were imputed using fastPHASE software (Scheet and Stephens, 2006). A total of 61,174 SNPs with minor allele frequencies (MAF) ≥ 5% were selected from 87,308 SNPs and were used for the GWAS analysis. The SNP dataset used in the current study is available in the Sequence Read Archive (SRA) at NCBI (SRA accession: PRJNA648781) repository, and on the website of the National Center for Soybean Improvement.2

Population Structure and Linkage Disequilibrium Analyses
SNPs were first pruned by the indep-pairwise command option of pLINK v1.07 (Purcell et al., 2007), and the pruned SNPs were used to infer population structure using the Bayesian Markov Chain Monte Carlo (MCMC) model in STRUCTURE v2.3.4 software (Pritchard et al., 2000). The number of presumed populations (K) was set from 1 to 11 using a burn-in of 10,000, and a run length of 100,000 and, each K-value was obtained with seven independent runs. The statistic “ΔK,” which indicates the most likely number of subpopulations, was calculated by following the Evanno method (Evanno et al., 2005), and then the Q matrix under the highest ΔK-value was obtained. A pairwise distance matrix derived from the Nei’s genetic distance for all polymorphic SNPs was calculated to construct a Neighbor-joining (NJ) tree using TASSEL 5.0, and the Kinship was also analyzed in this software (Bradbury et al., 2007). Principal component analysis (PCA) was performed by using GCTA v1.92 software (Yang et al., 2011).

Linkage disequilibrium (LD), estimated as the $r^2$ between SNPs, was calculated for each chromosome based on the LD tools option of RTM-GWAS V1.1 (He et al., 2017). LD was

\[ http://ncsi.njau.edu.cn/zygx.htm \]
visualized using the mean $r^2$ within bin sizes of 100 kb within the 5 Mb for each chromosome (Chr.) and then averaged across the whole genome. LD decay was calculated as the point at which the chromosomes reached 50% of their original LD value (Huang et al., 2010). Only $r^2$ for SNPs with pairwise distances less than 5 Mb in every chromosome was used to draw the average LD decay figure by GraphPad Prism version 5.01 (GraphPad Software, San Diego California United States).³

### Genome-Wide Association and Haplotype Block Analyses

A total of 61,174 SNPs with MAF >5% were utilized for the association mapping with the 309 soybean lines. The FarmCPU model was used to detect significant marker-trait associations in GAPIT version 3.0 (Lipka et al., 2012). The Kinship matrix (K) and pseudo QTNs were used as covariates to minimize false positives as well as to increase statistical power. The significance threshold used in the present study was set at $-\log_{10}(1/m)$ where $m =$ the number of markers, thus $-\log_{10}(1/61,174) =$ 4.787 as the Bonferroni correction line. The phenotypic variation explained (PVE) was calculated by mixed liner model in TASSEL (Bradbury et al., 2007). Manhattan plots were generated using the R software package agman (Turner, 2018). Haplotype block analysis was carried out for the significant and stable QTN regions in Haplovie software version 4.2 with four gamete rule using default settings (Barrett, 2009). The GSTs of accessions in each significant SNP were compared with Duncan Range Multiple test via SPSS 25.0 (IBM Corp. Released 2017, Armonk, NY: IBM Corp) at significant level of 0.05 to determine the extent of variation among the accessions grouped.

### Candidate Gene Prediction Within Major QTNs

For each of the four growth stage traits, there will be seven phenotypic datasets, including the measured data from six different environments plus the BLUP values. We treated SNPs that were co-localized in at least three datasets and exceeded the significance threshold as QTNs. Based on the LD distance, we extended and selected the region of 500 kb upstream and downstream of the QTNs as 1.0 Mb QTN regions (Brodie et al., 2016; Karikari et al., 2020). According to the functional annotations available in the ‘TAIR’³ and SoyBase (see text footnote 1) databases as well as available literature, the possible candidate genes were predicted within major significant QTN regions.

### RESULTS

#### Natural Variation Available in YHSBLP Population for RGS Traits

All the GSTs exhibited large phenotypic variation across the six environments. Descriptive statistics and F-values from ANOVA

---

³www.graphpad.com

４https://www.arabidopsis.org/
TABLE 1 | Descriptive statistics for four growth stage traits across six different environments in the 309 soybean accessions.

| Trait | Mean | SD  | Min | Max  | CV(%) | $F_G^b$  | $F_E^b$ | $F_{RE}^b$ | $F_{GXE}^b$ | $h^2(\%)^c$ |
|-------|------|-----|-----|------|-------|---------|---------|-----------|------------|-------------|
| DF    | 44   | 5.54| 27  | 65   | 12.54 | 25.39*** | 5.63*** | 1.02      | 4.79***    | 96.74       |
| DPB   | 58   | 6.72| 40  | 79   | 11.62 | 24.2***  | 2.76**  | 0.4       | 4.6***     | 96.33       |
| DSF   | 67   | 6.87| 47  | 87   | 10.82 | 21.85*** | 1.73    | 0.97      | 4.53***    | 96.25       |
| DMI   | 103  | 8.71| 82  | 124  | 8.56  | 11.33*** | 1.12    | 0.28      | 3.93***    | 94.17       |

$^a$DF, days to flowering; DPB, days to pod beginning; DSF, days to seed formation; DMI, days to maturity initiation.
$^b$From ANOVA Table; $F_G$, $F_E$, $F_{RE}$, and $F_{GXE}$ represent the F-values for genotypic, environmental, block effects and genotype × environment interaction, respectively.
$^c$**p < 0.01, ***p < 0.001.

DF, DPB, DSF, and DMI, respectively, was named $qGm-10-1$. Another region of $qGm15-1$, was associated with DF and DSF, which had the average PVE for the two GSTs of 1.54 and 2.49%, respectively.

Furthermore, the regions of $qGm06-3$ and $qGm17-1$ were mostly detected to associate with the DF trait within the remaining 6 QTN regions, which could contribute an average PVE of DF as 3.50 and 2.85%, respectively. Two QTN regions, $qGm02-1$ and $qGm06-1$, were mostly associated with the DMI trait, and had average positive genetic effect as 1.15 and 1.86. Unprecedentedly, the $qGm06-2$ region was identified to associate mostly with the DPB trait, and caused an average PVE of 4.05%, with the average negative genetic effect of $-2.01$. The final QTN region was $qGm11-1$, which associated with DMI in three datasets. It contributed an average negative genetic effect of $-0.83$ for this trait (Table 2 and Supplementary Table 3).

**Candidate Genes Within GSTs-Related QTN Regions**

The most prominent SNP—i.e., Gm10_44378814 in the $qGm10-1$ region with the highest $-\log_{10}(P)$ values (21.76) and associated with all four GSTs on Chr.10—overlapped with E2 (Glyma10g36600), a well-known gene regulating flowering and maturation duration in soybean, and encodes a homolog

FIGURE 1 | Frequency distribution and correlation analysis of four growth stage traits (GSTs) best linear unbiased prediction values (BLUPs). (A–D) represent the frequency distribution of time to flowering, pod beginning, seed formation and maturity initiation, respectively.
FIGURE 2 | Population structure of 309 soybean accessions. (A) Calculation of the true K of the YHSBLP following procedure outlined by Evanno et al. (2005). (B) A neighbor-joining tree of the tested accessions that could be divided into four subpopulations. (C) Population structure derived from STRUCTURE software. Four colors (blue, yellow, green, and red) represent the four subpopulations (subpopulation 1, 2, 3, and 4, respectively). Each vertical column represents one individual and each colored segment in each column represents the percentage of the individual in the population. (D) Principal component analysis plot of the 309 accessions; two-dimensional scales were used to reveal population stratification.

FIGURE 3 | Linkage disequilibrium decay plots across soybean chromosomes and the average decay across the genome.

of GIGANTEA (GI) gene (Watanabe et al., 2011). This gene is located 338 kb downstream from Gm10_44378814 (Supplementary Table 4). A total of 275 Arabidopsis homologs genes within the remaining 7 QTN regions (except for the qGm10-1 region) were screened (Supplementary Figure 2) and listed in Supplementary Table 5.
For the DF trait, the $qGm06-3$, $qGm15-1$, and $qGm17-1$ regions harbor 47, 33, and 87 candidate genes, with expression data available for 22, 16, and 55 genes in the SoyBase transcriptome database (see text footnote 1). Among these genes, 6, 3, and 17 genes expressed more highly (by more than 50-fold) in flower tissues with the highest expression of fatty acyl-ACP thioesterases B protein ($Glyma06g23560$; 1,587-fold), followed by CTC-interacting domain 7 ($Glyma17g33510$; 287-fold), and Transcription factor TFIIE proteins ($Glyma06g23400$; 192-fold). For the DPB trait, $qGm6-2$ region obtained 65 model genes, of which 28 had expression data on SoyBase (see text footnote 1). Six genes ($Glyma06g22410$, $Glyma06g22220$, $Glyma06g22260$, $Glyma06g22840$, $Glyma06g22260$, and $Glyma06g22800$) expressed higher in one-cm pod tissues with highest expression of COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family ($Glyma06g22410$; 216-fold), followed by a protein of unknown function ($Glyma06g22220$; 87-fold) and WD-40 repeat family proteins ($Glyma06g22840$; 66-fold). One DSF related QTN region of $qGm15-1$ harbors 14 genes with available expression data from 33 model genes. Only 2 genes expressed more highly in seed tissues 21 days after flowering (according to the Table 1), with the highest expression being of NAD(P)-binding Rossmann-fold superfamily protein ($Glyma15g27630$; 126-fold), followed by a protein of unknown function ($Glyma15g27790$; 51-fold). Two DMI related QTN regions was $qGm02-1$ and $qGm06-1$. 63 and 93 candidate genes were screened from them, respectively. 9 and 8 genes expressed higher in seed tissues 42 days after flowering. Two kinds of ribosomal protein ($Glyma06g20540$; 1,181-fold and $Glyma02g00540$; 335-fold) were the two most highly expressed proteins, followed by an organic cation/carnitine transporter 3 protein ($Glyma06g20500$; 220-fold).

**DISCUSSION**

Four Soybean Growth Stage Traits Were Significantly Affected by Genetic-Environment Interaction

As a model crop for photoperiodic response research (Garner and Allard, 1920), the phenotypic variation in the GSTs of soybean...
TABLE 2 | Eight QTN regions associated with four growth stage traits via FarmCPU method across different environments in summer soybean.

| Regionb | Chr | SNPb | Pos.(bp)c | No| −log10(P)d | PVE(%)e | Position(Mb)f | Effc | Trait_environmentg |
|---------|-----|------|-----------|---|------------|--------|---------------|------|-------------------|
| qGm02-1 | 2   | Gm02_150932 | 150,932   | 2 | 6.27       | 1.97−4.91 | 0.15−0.16 | 1.56 | DPB_6; DMI_1,2,3,5 |
| qGm06-1 | 6   | Gm06_16710123 | 16,710,123 | 1 | 5.65       | 1.28−3.12 | 16.71 | 2.11 | DMI_1,4,7         |
| qGm06-2 | 6   | Gm06_19332290 | 19,332,290 | 1 | 12.93      | 1.44−5.77 | 19.33 | −2.31 | DF_7; DPB_1,4,5; DSF_4 |
| qGm06-3 | 6   | Gm06_21072696 | 21,072,696 | 2 | 10.91      | 2.74−4.13 | 20.99−21.07 | −2.02 | DF_2,3,6         |
| qGm10-1 | 10  | Gm10_44378814 | 44,378,814 | 3 | 21.76      | 2.39−7.66 | 44.35−44.56 | −2.05 | DF_1,2,4,5,6,7; DPB_1,2,3,4,5,6; DSF_1,4,5,6; DMI_1,3,4,5,6,7 |
| qGm11-1 | 11  | Gm11_15963231 | 15,963,231 | 3 | 7.92       | 2.04−3.70 | 15.95−15.96 | −1.29 | DPB_6; DSF_1,4; DMI_1,3,4 |
| qGm15-1 | 15  | Gm15_29990587 | 29,990,587 | 4 | 11.71      | 0.51−4.61 | 29.75−29.99 | 1.95  | DF_1,4,5,6; DPB_1,5; DSF_1,4,5; DMI_4 |
| qGm17-1 | 17  | Gm17_37676700 | 37,676,700 | 2 | 7.41       | 2.40−3.26 | 37.68−37.79 | 1.06  | DF_1,6,7; DPB_1   |

a QTN regions which was detected at least in three environments.
b SNPs that were most significantly associated with the trait.
c QTN position (bp) on soybean genome assembly Wm82.a2.v1.
d The number of significant QTNs identified in the region.
e Phenotypic variation explained (PVE) calculated by mixed linear model.

was considerably influenced by environmental conditions. The four GSTs (DT, DPB, DSF, and DMI) of 309 summer soybean genetic resources in this study varied widely (Table 1) across the six different environments and BLUPs. The G × E played a significant role, suggesting that in addition to genetic factors, environmental factors such as photoperiod and temperature have direct influence on the four GSTs. Similar observations have been reported by Li M. et al. (2019). While the genetic analysis of G × E for target traits in crop breeding has become the emerging research direction, the genetic background of important GSTs in soybean need to be paid more attention for increasing the adaptivity of the crop to diverse environments.

Selection Basis of Linkage Disequilibrium Decay Distance and GWAS Method

LD is one of the important factors affecting mapping resolution of GWAS analysis (Zhang et al., 2015). In the present study, a considerable number of SNP markers (>61,000) were used for the detection of marker-trait associated QTNs linked with various GSTs in the YHSBLP mapping panel. We observed a diverse range of LD among various soybean chromosomes with an average LD decay distance of 1.5 Mb (Figure 3), which is consistent with previous studies (Vuong et al., 2015; Zhang et al., 2015; Copley et al., 2018). Our study also confirmed that soybean possesses long LD decay distance, and suggested that marker-trait association identified in soybean can be directly used in MAS rather than map-based cloning. The high LD may be due to cleistogamous characteristics of soybean which have a strong influence on genomic homogeneity and reduced genomic variation, and this characteristic might become more sensitive during domestication practices, resulting in low genetic diversity and high LD (Lam et al., 2010). In addition, LD in a population is affected by historical recombination, mutations, allele frequency, genetic drift, founder effects, artificial selection during domestication, or breeding (Alqudah et al., 2020). The high LD may lead to numerous unrelated genes being screened out, therefore, some studies have proposed to use a cutoff of 500 kb, since enhancers and repressors may be as distant as 500 kb from their genes (Brodie et al., 2016; Karikari et al., 2020). By keeping this in view, we defined the range of LD for target traits in crop breeding has become the emerging research direction, the genetic background of important GSTs in soybean need to be paid more attention for increasing the adaptivity of the crop to diverse environments.

Genetic Bases of Four GSTs in YHSBLP Population

Fifty percent of the QTN regions in this study overlapped with the previously reported QTLs or genes with known functions related to GSTs of soybean (Supplementary Table 4). E2 is a homologous gene GmGIA of Arabidopsis thaliana GIGANTEA (GI). The flowering time of E2 and its near-isogenic lines (e2)
was similar at high latitude 43°N and middle latitude 36°N, indicating that the regulation of E2 gene on flowering time of soybean might not only depend on photoperiod, so that it had a stronger geographical adaptability (Watanabe et al., 2011). Therefore, the E2 gene has broad application prospects in breeding practice. In this study, the same location on Chr.10 near the E2 gene was identified among the four GSTs in both BLUPs and measured data, which may be the basis for the strong genetic correlation among GSTs. Considering the role of the E2 gene in pod and seed is rarely reported, the results of this study have a guiding significance for the further functional study of E2 gene.

The allele variation of the most significant SNPs in each QTN region could cause significant changes in the corresponding phenotype, indicating that there may be candidate genes affecting soybean GSTs in the remaining 7 QTN regions (Supplementary Figure 3). On Chr. 06, two QTN regions qGm06-1 and qGm06-2 were associated with DMI and DPB, respectively. Zhang et al. (2015) identified two SNPs, Gm06_16723946 and Gm06_16167243, that are associated with flowering time and days from flowering to maturity, respectively, whose location was close to the most significant SNP—Gm06_16710123 of qGm06-1—in this study. A similar result was also found in the genetic study on the reproductive to vegetative growth period ratio in soybean by Wang et al. (2015). Fang et al. (2017) found a region from 19,178,035 to 20,299,454 bp on Chr.06 was associated with full bloom date traits, which was overlapped with the region of qGm06-2 detected in this study. Another QTN region, qGm17-1 was located on Chr.17 associated with DF, the most significant SNP Gm17_37676700 in this region was only 100 kb (Gm17_37574384) away from the genetic location results of soybean flowering time under various photo-thermal conditions reported by Mao et al. (2017).

Some candidate genes around the QTN regions are also noteworthy: ARF4 (auxin response factor 4) and GCR1 (G-protein-coupled receptor 1) are well-known genes modulating flowering, whose mutants affect the response to auxin and shorten time to flowering (Colucci et al., 2002; Pashkovskiy et al., 2016). We identified two orthologs of ARF4 and GCR1, Glyma06g23830 (At5g60450) on Chr.06, and Glyma17g33480 (At1g48270) on Chr.17, which are located at the DF-related QTN regions of qGm06-3 and qGm17-1, respectively. In addition, the AGL8 (At5g60910.1) ortholog Glyma06g22650 was detected on Chr.06 around the peak SNP Gm06_19332290 in the qGm06-2 region, which was associated with DPB. AGL8 is a MADS box gene and involved in the early step of specifying floral meristem identity as well as the later step of determining the fate of floral organ primordia (Mandel and Yanofsky, 1995; Balanža et al., 2019), hence Glyma06g22650 may be involved in regulating GSTs. A plant-specific HD2 histone deacetylase gene, HD2B (At5g22650), was identified around the peak SNP Gm11_15963231 on the qGm11-1 region associated with DSF. HD2B was regarded as a genetic factor associated with seed dormancy and plant growth, development, and response to abiotic stresses (Yano et al., 2013; Han et al., 2016). Another Arabidopsis trehalose biosynthesis gene, TPS1 (At1g78580), has an expression analysis and a spatial and temporal activity of promoter which suggests that this gene is active within the seed-filling stage of development, TPS1 was related to Glyma15g27480 in the qGm15-1 region, which may have an influence on the seed formation time of soybean (Dijiken et al., 2004; Vuong et al., 2015).

In the two remaining QTN regions associated with DMI, the LFR (At3g22990) gene coded an ARM repeat superfamily protein and was identified in the qGm02-1 region. Its ortholog in rice, OsLFR, resulted in homozygous lethality in the early seed stage through endosperm and embryo defects via its depletion (Qi et al., 2020). Another gene, ARR6 (At3g47620) was reported as a response regulator of cytokinins, which could regulate the response to cytokinin stimulus and thus delay leaf senescence. This gene was found as an ortholog of Glyma06g20660 in the qGm06-1 region (Hallmark and Rashotte, 2020).

The markers detected in this present study could be validated and used for MAS geared at improving GSTs in soybean, and provide foundation for future projects, including the designing of Kompetitive Allele Specific PCR (KASP) markers for practical breeding programs. The candidate genes predicted provide valuable information for functional validation to elucidate the molecular mechanism underlying the studied traits.

**SUMMARY**

In this study, we evaluated 309 diverse soybean breeding lines across five different environments for four GSTs (DF, DPB, DSF, and DMI), and applied a FarmCPU GWAS with high-density SNPs (>61 k). Fifty-eight SNPs within 8 QTN regions were detected to associate with the four GSTs. In addition, one QTN region on Chr.10 was detected repeatedly among all four GSTs and overlapped with the well-known E2 gene. Seven remaining QTN regions contained the peak SNPs of Gm02_150932, Gm06_16710123, Gm06_19332290, Gm06_21072696, Gm11_15963231, Gm15_29990587, and Gm17_37676700 on Chr.02, Chr.06, Chr.11, Chr.15, and Chr.17 were identified uniquely or commonly associated with the four GSTs and caused significant phenotypic variation with both major and minor alleles. Two hundred and seventy-five Arabidopsis genes involved in defense response, flowering, embryo development and so on with their homologs in soybean were screened, and some predicted as candidate genes (Glyma06g23830, Glyma17g33480, Glyma06g22650, Glyma15g27480, and Glyma06g20660) for DF, DPB, DSF, and DMI of summer soybean. However, the predicted candidate genes need further screening and functional validation to ascertain their actual roles in modulating the GSTs in soybean.

**DATA AVAILABILITY STATEMENT**

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.
AUTHOR CONTRIBUTIONS

TZ and HJ conceived and designed the experiments. FC, FZ, YZ, and WY performed the experiments. DL and WY analyzed the data. WY drafted the manuscript. TZ and BK revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Natural Science Foundation of China (Nos. 31872847 and 31871646) and the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP) Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.715529/full#supplementary-material

Supplementary Figure 1 | Pearson correlation among the four growth stage traits of the BLUPs (A) and measured data from six different environments (B). **p < 0.001. DF- days to flowering; DPB-days to pod beginning; DSF- days to seed formation; DMI- days to maturity initiation. 1–7 represent the environment code of BLUP: 18JP, 18YC, 19DT, 19JP, 19YC6, and 19YC7, respectively.

Supplementary Figure 2 | Manhattan plot and linkage disequilibrium block of 7 peak SNPs in different environments. Linkage disequilibrium blocks associated with flowering, pod beginning, seed formation and maturity initiation time near Gm02_150932, Gm06_16710123, Gm06_19322990, Gm06_21023766, Gm11_15963231, Gm15_25990587, and Gm17_37676700. Significance threshold is denoted by the red line which was set as –log10(p) = 4.787. The up panel was the Manhattan plots of negative log10 transformed p-values vs. QTMs. The down panel was haplotype block based on pairwise linkage disequilibrium r2-values. DF- days to flowering; DPB-days to pod beginning; DSF- days to seed formation; DMI- days to maturity initiation. 1–7 represent the environment code of BLUP: 18JP, 18YC, 19DT, 19JP, 19YC6, and 19YC7, respectively.

Supplementary Figure 3 | Phenotypic variation between cultivars carrying different alleles of the QTMs significantly associated with four growth stage traits in various environments. The box plot shows the significant difference of days to flowering, pod beginning, seed formation and maturity initiation of the cultivars carrying two alleles of the QTMs. The significant QTMs were Gm02_150932, Gm06_16710123, Gm06_19322990, Gm06_21023766, Gm11_15963231, Gm15_25990587, and Gm17_37676700. The major allele of significant loci was marked by gray, and the minor allele was marked by pink. Significant differences tested by the t-test are also given (**p < 0.001, *p < 0.01, *p < 0.05). DF- days to flowering; DPB-days to pod beginning; DSF- days to seed formation; DMI- days to maturity initiation. 18 and 19 represent the years of 2018 and 2019, DT : Dangui; JP : Jiangpu; 2019YC6 and 2019YC7 means the planting date of June and July in Yancheng city, respectively.

Supplementary Table 1 | Phenotypic data of four growth stage traits in 309 summer soybeans across different datasets.

Supplementary Table 2.1 | SNP associated with days to flowering of summer soybeans.

Supplementary Table 2.2 | SNP associated with days to pod beginning in summer soybeans.

Supplementary Table 2.3 | SNP associated with days to seed formation in summer soybeans.

Supplementary Table 2.4 | SNP associated with days to maturity initiation in summer soybeans.

Supplementary Table 3 | Details of eight QTN regions associated with four growth stage traits via FarmCPU method in summer soybeans.

Supplementary Table 4 | The overlapped QTLs or genes between present and previous studies within ± 500 kbp upstream or downstream from the peak SNPs.

Supplementary Table 5 | The detailed functional annotation information of 275 candidate genes in or near the QTN regions.

REFERENCES

Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetlytrimethylammonium bromide. Nat. Protoc. 1, 2320–2325. doi: 10.1038/nprot.2006.384

Alqudah, A. M., Sallam, A., Stephen Baenziger, P., and Borner, A. (2020). GWAS: fast-forwarding gene identification and characterization in temperate cereals: lessons from barley—a review. J. Adv. Res. 22, 119–135. doi: 10.1016/j.jare.2019.10.013

Andolfatto, P., Davison, D., Erezylmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., et al. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21, 610–617. doi: 10.1101/gr.115402.110

Asakura, T., Tamura, T., Terauchi, K., Narikawa, T., Yagasaki, K., Ishimaru, Y., et al. (2012). Global gene expression profiles in developing soybean seeds. Plant Physiol. Bioch. 52, 147–153. doi: 10.1016/j.plaphy.2011.12.007

Balanzà, V., Martínez-Fernández, I., Sato, S., Yanofsky, M. F., and Ferrandiz, C. (1999). Inflorescence meristem fate is dependent on seed development and fruit full in Arabidopsis thaliana. Front. Plant. Sci. 10:1622. doi: 10.3389/fpls.2019.01622

Barrett, J. C. (2009). Haploviz: visualization and analysis of SNP genotype data. Cold Spring Harb. Protoc. 2009:i71. doi: 10.1101/pdb.top71

Bates, D., Mächler, M., and Bolker, B. (2014). Fitting linear mixed-effects models using lme4. J. Stat. Softw. arXiv [Preprint] 67, 1–48. doi: 10.18637/jss.v067.i01 arXiv:1406.

Bernard, R. (1971). Two major genes for time of flowering and maturity in soybeans. Crop Sci. 11, 242–244. doi: 10.2135/cropsci1971.0011833X001100020022x

Bonato, E. R., and Vello, N. A. (1999). Es, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22, 229–232. doi: 10.1590/S1415-47511999000200016

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

Brodie, A., Azaria, J. R., and Ofran, Y. (2016). How far from the SNP may the causative genes be? Nucleic Acids. Res. 44, 6046–6054. doi: 10.1093/nar/gkw500

Bussell, R., and Voldeng, H. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. News. 7, 26–29.

Chen, Z., Lancon-Verdier, V., Le Signor, C., She, Y., Kang, Y., and Verdier, J. (2021). Genome-wide association study identified candidate genes for seed size and seed composition improvement in M. truncatula. Sci. Rep. UK 11:4224. doi: 10.1038/s41598-021-83581-7

Cober, E. R., Molnar, S. J., Charette, M., and Voldeng, H. D. (2010). A new locus affecting seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. Theor. Appl. Genet. 120, 1005–1012. doi: 10.1007/s00122-009-1228-6
Colucci, G., Acone, F., Ayleshmerini, N., Chalmers, D., and Chrispeels, M. J. (2002). GCR1, the putative Arabidopsis G protein-coupled receptor gene is cell cycle-regulated, and its overexpression abolishes seed dormancy and shortens time to flowering. *PNAS* 99, 4736–4741. doi: 10.1073/pnas.070420899

Copley, R. T., Dupepe, M., and O’Donoughue, L. S. (2016). Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. *BMC Genomics* 19:167. doi: 10.1186/s12866-018-4458-4

Dijken, A. J. H. V., Schleupmann, H., and Smeekens, S. C. M. (2004). *Arabidopsis* trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. *Plant Physiol.* 135 (2), 969–977. doi: 10.1104/pp.104.039743

Du, J., Wang, S., He, C., Zhou, B., Ruan, Y., and Shou, H. (2017). Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *J. Exp. Bot.* 68, 1955–1972. doi: 10.1093/jxb/erw460

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., 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

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/cropsci1971.00113934001100060051x

Garnier, W. W., and Allard, H. A. (1920). Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J. Agric. Res.* 2, 157–158.

Hallmark, H. T., and Rashotte, A. M. (2020). Cytokinin isopentenyladenine and cytokinin-associated genes. *Plant Direct.*

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies for pasmo resistance in flax (*P. sativum* L.) Merrill. *Am. J. Hum. Genet.* 86, 907–913. doi: 10.1010/BF0211040

Mao, T., Li, J., Wen, Z., Wu, T., Wu, C., Sun, S., et al. (2017). Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photothermal conditions. *BMC Genomics* 18:415. doi: 10.1186/s12864-017-3778-3

Mcblain, B. A., and Bernard, R. L. (1987). A new gene affecting the time of flowering in soybean (*Glycine max* L.). *Front. Plant Sci.* 3, 1763–1771. doi: 10.3389/fpls.2017.011763

Mansur, L. M., Lark, K. G., Kross, H., and Oliveira, A. (1993). Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). *Theor. Appl. Genet.* 86, 907–913. doi: 10.1007/BF02211040

Ray, J., Hinson, K., Mankono, J., and Malo, M. (1995). Genetic control of a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* 92, 901–916. doi: 10.1007/BF00257439

Scheet, P., and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78, 629–644. doi: 10.1086/502802

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qi, D., Wen, Q., Meng, Z., Yuan, S., Guo, H., Zhao, H., et al. (2020). OsLFR is a novel factor genes in *Oryza sativa* SWI/SNF complex members in *Arabidopsis* photoresponses associated with duplication of the phytocrome a gene. *Genetics* 180, 995–1007.

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2003). Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. *Genome* 46, 1024–1036. doi: 10.1139/g03-079

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theor. Appl. Genet.* 120, 997–1003. doi: 10.1007/s00122-009-1227-7

Pashkovskiy, P. P., Kartashov, A. V., Zlobin, I. E., Pogosyan, S. I., and Kuznetsov, V. V. (2016). Blue light alters MIR167 expression and microRNA-targeted auxin response factor genes in *Arabidopsis thaliana* plants. *Plant Physiol. Bioch.* PPB 104, 146–154. doi: 10.1016/j.plaphy.2016.03.018

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Qi, D., Wen, Q., Meng, Z., Yuan, S., Guo, H., Zhao, H., et al. (2020). OsLFR is a novel factor genes in *Oryza sativa* SWI/SNF complex members in *Arabidopsis* photoresponses associated with duplication of the phytocrome a gene. *Genetics* 180, 995–1007.

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2003). Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. *Genome* 46, 1024–1036. doi: 10.1139/g03-079

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: 10.1086/302959

Ray, J., Hinson, K., Mankono, J., and Malo, M. (1995). Genetic control of a long-juvenile trait in soybean. *Crop Sci.* 35, 1001–1006.

Ruan, Y., Patrick, J. W., Bouzyan, M., Osorio, S., and Fernie, A. R. (2012). Molecular regulation of seed and fruit set. *Trends Plant. Sci.* 17, 656–665. doi: 10.1016/j.tplants.2012.06.005

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* 130, 377–390. doi: 10.1007/s00122-016-4289-7
Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670

Sertse, D., You, F. M., Ravichandran, S., Soto-Cerda, B. J., Duguid, S., and Cloutier, S. (2021). Loci harboring genes with important role in drought and related abiotic stress responses in flax revealed by multiple GWAS models. *Theor. Appl. Genet.* 134, 191–212. doi: 10.1007/s00122-020-03691-0

Shirasaka, K., Fukuoka, H., Matsunaga, H., Kobayashi, Y., Kobayashi, I., Hirakawa, H., et al. (2013). Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato. *DNA Res.* 20, 593–603. doi: 10.1093/dnares/dst033

Song, J., Sun, X., Zhang, K., Liu, S., Wang, J., Yang, C., et al. (2020). Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. *Mol. Breeding* 40:60. doi: 10.1007/s11032-020-01140-w

Taiyun, W., and Viliam, S. (2017). *r* package "Corrplot": *Visualization of a Correlation Matrix* (version 0.84). Available online at: https://github.com/taiyun/corrplot (accessed April 20, 2021).

Tasma, I. M., Lorenzen, L. L., Green, D. E., and Shoemaker, R. C. (2001). Mapping dominant locus, E11, controls early flowering time and maturity in soybean. *Mol. Breeding* 10.1186/s12864-015-1811-y

Wang, F., Nan, H., Chen, L., Fang, C., Zhang, H., Su, T., et al. (2019). A new locus E3. *Am. J. Hum. Genet.* 188, 395–407. doi: 10.1016/j.ajhg.2019.11.011

Wang, Y., Cheng, L., Leng, J., Wu, C., Shao, G., Hou, W., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4

Wang, Y., Li, Y., Wu, H., Hu, B., Zheng, J., Zhai, H., et al. (2018). Genotyping of new multi-locus gwas methodologies in the genetic dissection of complex traits. *Front. Plant. Sci.* 10:hhh. 10.3389/fpls.2018.00100

Watanabe, S., Hideshima, R., Xia, Z., Tsukubocura, 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: 10.1534/genetics.108.098772

Watanabe, S., Xia, Z., Hideshima, R., Tsukubocura, Y., Sato, S., Yamanaka, N., et al. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that thegigantea gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062

Xu, Y., Yang, T., Zhou, Y., Yin, S., Li, P., Liu, J., et al. (2018). Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. *Front. Plant. Sci.* 9:1311. doi: 10.3389/fpls.2018.01311

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Yano, R., Takebayashi, Y., Nambara, E., Kamiya, Y., and Sog, M. (2013). Combining association mapping and transcriptomics identify HD2B histone deactylase as a genetic factor associated with seed dormancy in Arabidopsis thaliana. *Plant J.* 74, 815–828. doi: 10.1111/jtp.12167

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X. P., Pool, J. E., et al. (2010). Sequencing of 50 human Exomes reveals adaptation to high altitude. *Science* 329:7578. doi: 10.1126/science.1190371

Zhang, D., Wang, X., Li, S., Wang, C., Gosney, M. J., Mickelbart, M. V., et al. (2019). A post-domestication mutation, Dt2, triggers systemic modification of divergent and convergent pathways modulating multiple agronomic traits in soybean. *Mol. Plant* 12, 1366–1382. doi: 10.1016/j.molp.2019.05.010

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Xu, 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, W., Xu, W., Zhang, H., Liu, X., Cui, X., Li, S., et al. (2021). Comparative selective signature analysis and high-resolution GWAS reveal a new candidate gene controlling seed weight in soybean. *Theor. Appl. Genet.* 134, 1329–1341. doi: 10.1007/s00122-021-03774-6

Zhang, Y., Jia, Z., and Dunwell, J. M. (2019). Editorial: the applications of new multi-locus gwas methodologies in the genetic dissection of complex traits. *Front. Plant. Sci.* 10:hhh. 10.3389/fpls.2019.00100

Conflict of Interest: 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.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yan, Karikari, Chang, Zhao, Zhang, Li, Zhao and Jiang. 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(s) 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.