



# Identification of Loci and Candidate Genes Responsible for Pod Dehiscence in Soybean via Genome-Wide Association Analysis Across Multiple Environments

Dezhou Hu<sup>1†</sup>, Guizhen Kan<sup>1†</sup>, Wei Hu<sup>1</sup>, Yali Li<sup>1</sup>, Derong Hao<sup>2</sup>, Xiao Li<sup>1</sup>, Hui Yang<sup>1,3</sup>, Zhongyi Yang<sup>1</sup>, Xiaohong He<sup>1</sup>, Fang Huang<sup>1\*</sup> and Deyue Yu<sup>1,3\*</sup>

## OPEN ACCESS

### Edited by:

Roberto Papa,  
Marche Polytechnic University, Italy

### Reviewed by:

Saleh Alseekh,  
Max Planck Institute of Molecular  
Plant Physiology, Germany  
Francesca Sparvoli,  
Istituto di Biologia e Biotecnologia  
Agraria (IBBA), Italy  
Monica Rodriguez,  
University of Sassari, Italy

### \*Correspondence:

Fang Huang  
fhuang@njau.edu.cn  
Deyue Yu  
dyyu@njau.edu.cn

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

Received: 04 March 2019

Accepted: 05 June 2019

Published: 21 June 2019

### Citation:

Hu D, Kan G, Hu W, Li Y, Hao D,  
Li X, Yang H, Yang Z, He X, Huang F  
and Yu D (2019) Identification of Loci  
and Candidate Genes Responsible  
for Pod Dehiscence in Soybean via  
Genome-Wide Association Analysis  
Across Multiple Environments.  
*Front. Plant Sci.* 10:811.  
doi: 10.3389/fpls.2019.00811

<sup>1</sup> National Center for Soybean Improvement, National Key Laboratory of Crop Genetics and Germplasm Enhancement, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, China, <sup>2</sup> Jiangsu Yanjiang Institute of Agricultural Sciences, Nantong, China, <sup>3</sup> School of Life Sciences, Guangzhou University, Guangzhou, China

Pod dehiscence (shattering) is the main cause of serious yield loss during the soybean mechanical harvesting process. A better understanding of the genetic architecture and molecular mechanisms of pod dehiscence is of great significance for soybean breeding. In this study, genome-wide association analysis (GWAS) with NJAU 355K SoySNP array was performed to detect single nucleotide polymorphisms (SNPs) associated with pod dehiscence in an association panel containing 211 accessions across five environments. A total of 163 SNPs were identified as significantly associated with pod dehiscence. Among these markers, 136 SNPs identified on chromosome 16 were located in the known QTL *qPDH1*. One, one, three, eleven, three, one, three, three and one SNPs were distributed on chromosome 1, 4, 6, 8, 9, 11, 17, 18, and 20, respectively. Favorable SNPs and six haplotypes were identified based on ten functional SNPs; among those Hap2 and Hap3 were considered as optimal haplotypes. In addition, based on GWAS results, the candidate gene *Glyma09g06290* was identified. Quantitative real-time PCR (qRT-PCR) results and polymorphism analysis suggested that *Glyma09g06290* might be involved in pod dehiscence. Furthermore, a derived cleaved amplified polymorphic sequences (dCAPS) marker for *Glyma09g06290* was developed. Overall, the loci and genes identified in this study will be helpful in breeding soybean accessions resistant to pod dehiscence.

**Keywords:** soybean, pod dehiscence, GWAS, haplotype analysis, *Glyma09g06290*, polymorphism analysis

**Abbreviations:** ANOVA, analysis of variance; BLUP, best linear unbiased prediction; dCAPS, derived cleaved amplified polymorphic sequences; FDR, false discovery rate; GWAS, genome-wide association analysis;  $h^2$ , broad-sense heritability; HR, Huang-Huai ecoregion; LD, linkage disequilibrium; MAFs, minor allele frequencies; MAS, marker-assisted selection; NJ, neighbor-joining; NR, northern ecoregion; PPD, the percentage of pod dehiscence; Q-Q plots, quantile-quantile plots; qRT-PCR, quantitative real-time PCR; QTL, quantitative trait locus; RIL, recombinant inbred lines; RNA-Seq, RNA sequencing; SNP, single nucleotide polymorphisms; SR, southern ecoregion.

## INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is one of the major oil crops and provides edible oil and abundant protein for human beings. Soybean pod dehiscence (shattering) is when mature pods open along their dorsal or ventral sutures to release seeds (Kang et al., 2005). As pod dehiscence leads to significant yield loss in soybean and may cause more damage to soybean production as climatic conditions become harsher, breeding for soybean varieties with pod-shattering resistance is always one of the important goals for breeders.

In the past two decades, a few QTLs or genes controlling pod dehiscence have been identified. Previously, QTL analysis in cultivated soybean has revealed a possible major QTL and several minor QTLs that regulate pod dehiscence. Bailey et al. (1997) first identified the major QTL by using 140 restriction fragment length polymorphism (RFLP) markers in a population of 120 F4-derived lines (Bailey et al., 1997). Funatsuki et al. (2006) detected this major QTL and named it *qPDH1*; it was located between the markers Sat\_093 and Sat\_366 on chromosome 16 and accounted for 50% of the total variance (Funatsuki et al., 2006). Furthermore, this QTL was also reported by several other studies that used different populations and genetic backgrounds (Liu et al., 2007; Kang et al., 2009; Yamada et al., 2009; Suzuki et al., 2010; Ju et al., 2017), suggesting that *qPDH1* may be the major QTL associated with pod dehiscence. In addition, other QTLs that were deemed to be minor were also mapped (Bailey et al., 1997; Liu et al., 2007; Kang et al., 2009; Yamada et al., 2009).

Pod dehiscence/seed shattering has been deeply investigated in many crops (Vittori et al., 2019), several genes has been identified in rice (Konishi et al., 2006; Zhou et al., 2012), common bean (Koinange et al., 1996; Rau et al., 2019), *Medicago* (Fourquin et al., 2013), and cowpea (Suanum et al., 2016; Lo et al., 2018). In soybean, studies predicted that two to four genes were responsible for pod dehiscence (Caviness, 1969; Tsuchiya, 1986, 1987; Kang et al., 2009). To date, two genes located on chromosome 16 have been cloned and characterized. A gene, *Pdh1* (*Glyma16g25580*), detected by fine mapping (located at the *qPDH1* QTL) and encoding a dirigent-like protein, promotes pod dehiscence by increasing the torsion of dried pod walls (Funatsuki et al., 2014). In addition, Dong et al. (2014) identified a *NAC* gene, *SHAT1-5* (*Glyma16g02200*) by analyzing the changes due to domestication between cultivated soybean and wild soybean. This gene activates secondary-cell-wall biosynthesis and promotes the thickening of fiber-cap cells in pod sutures (Dong et al., 2014).

Although linkage mapping has been proved a powerful method to explore regions of the genome that co-segregate with a given trait, it is ordinarily difficult to isolate candidate genes based on a single QTL mapping experiment due to the limited markers (Zhang et al., 2014). Compared to linkage mapping, GWAS can be applied in germplasm collections or naturally occurring populations and provides higher resolution in terms of defining the genomic positions of genes or QTLs (Shirasawa et al., 2013). In recent years, with the advance of next-generation sequencing (NGS), this approach has been successfully applied to important traits in many plant species, including *Arabidopsis thaliana* (Kobayashi et al., 2016; Kalladan et al., 2017), rice

(Huang et al., 2010; Yano et al., 2016) and maize (Kump et al., 2011; Wang X. et al., 2016). In soybean, GWAS has been used to explore markers associated with many traits. For example, Hao et al. (2012) used GWAS to identify 19 SNPs and five haplotypes for soybean yield and yield components (Hao et al., 2012); Chu et al. (2017) mapped the *GmMYB29* (*Glyma20g35180*) gene related to isoflavone biosynthesis in soybean via GWAS (Chu et al., 2017); and Zhang et al. (2018) identified 87 chromosomal regions for seed composition based on 31,850 SNPs by GWAS (Zhang et al., 2018).

Although previous studies have identified many QTLs controlling pod dehiscence in soybean, these QTLs were detected using classical linkage analysis with a limited number of markers, and the allelic variation was also limited by using RILs or other linkage mapping populations. Furthermore, to date, only two pod-dehiscence-related genes have been cloned. In addition, we know little about the other QTLs related to pod dehiscence.

In our previous work, we performed primary mapping of pod-dehiscence-related QTLs due to the low density of the genetic map (Hu, 2017). In this paper, to obtain more accurate information regarding pod-dehiscence-related SNPs, GWAS with the large genome-wide NJAU 355K SoySNP array was performed across multiple environments. The aims of our study were as follows: (i) identify loci and candidate genes involved in pod dehiscence; (ii) identify favorable haplotypes relevant to pod dehiscence. These results will be helpful for MAS in soybean breeding programs.

## MATERIALS AND METHODS

### Materials and Field Trials

In this study, the association panel containing 211 soybean accessions was used (including 183 landraces, 25 improved accessions and three accessions with unknown evolution type). This population was collected from different geographic origins that exhibit variant phenotypes for pod dehiscence, and provided by the National Center for Soybean Improvement of China. Experiments to assess association panel were sowed in two places from 2013 to 2017: Jiangpu Experimental Station (32°12'N 118°37'48'E) in 2013 and 2014 (designated as environment Env1 and Env2, respectively); and the Experimental Farm of Jiangsu Yanjiang Institute of Agricultural Sciences (31°58'48'N 120°53'24'E), Nantong, China, in 2015–2017 (designated as Env3, Env4, and Env5, respectively). A complete randomized block design with three replications was used for all trials. Each line of the populations was planted in three rows, and each row was 200 cm long with 50 cm row spacing. At the seedling stage (about 2 weeks after germination), the number of plants in per row was limited to 15–20. Nutrition, water supply, weeding and insect control were maintained throughout the experiment.

### Evaluation of Pod Dehiscence

Three plants per line were harvested randomly, and twenty pods per plant were examined. The maturity pods were stored in a kraft paper bag after harvested and then put in the heating ovens as soon as possible. The degree of pod dehiscence was

identified by monitoring the percentage of dehiscent pods after heat treatment at 60°C for 3 h (Funatsuki et al., 2006; Yamada et al., 2009). As shown in **Supplementary Figure S1**, after heat treatment, the completely opened pods were used to calculate PPD.  $PPD = (\text{number of pod dehiscence}/\text{total pods}) \times 100\%$  (Yamada et al., 2009). The germplasm number, name, origin and phenotypic data of the association panel are listed in **Supplementary Table S1**.

## Phenotypic Data Analysis

The phenotypic data of the association panel was determined using the mean values for each line (Env1 to Env5) with the R software (R Development Core Team, 2015), including descriptive statistics, ANOVA, and  $h^2$ . Variable components (genotype, year, line, and location) were evaluated in R with the “lme4” package using the BLUP model. The BLUP model was based on the following formula (Merk et al., 2012):

$$Y_{ik} = \mu + G_i + Y_k + GY_{ik} + \varepsilon_{ik}$$

In this formula,  $Y_{ik}$  is the trait studied,  $\mu$  is the overall mean,  $G_i$  is the  $i$ th genotypic effect,  $Y_k$  is the effect of the  $k$ th year,  $GY_{ik}$  is the interaction of genotype  $\times$  year, and  $\varepsilon_{ik}$  is the residual error. The BLUP was also included as an environment in the GWAS analysis. The  $h^2$  of pod dehiscence was calculated using the following formula:

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$$

In which  $h^2$ , broad-sense heritability;  $\sigma_g^2$ , genotype variance;  $\sigma_{ge}^2$ , interaction variance of genotype  $\times$  environment;  $\sigma_e^2$ , variance of error components;  $n$ , the number of environments; and  $r$ , the number of replications.

## Genome-Wide Association Analysis

The 211 soybean accessions were genotyped with the NJAU 355K SoySNP array containing 282,469 SNPs (Wang J. et al., 2016). After filtering out SNPs with MAFs  $\leq 0.05$ , a total of 201,915 SNPs remained. GWAS was conducted based on the MLM model (Q + K) using the TASSEL software V5.0 (Bradbury et al., 2007). The kinship matrix (K) and population structure (Q) were calculated by TASSEL software V5.0 and STRUCTURE software version 2.3.4, respectively (Pritchard et al., 2000; Falush et al., 2003). A FDR of 0.05 was used as a threshold for significant association (Benjamini and Hochberg, 1995). Manhattan plots and Q-Q plots were generated using the R software package “qqman” (Turner, 2014).

## Digital Expression Data Analysis

Two databases were used to identify the tissue expression patterns of candidate genes. RNA-Seq data from 14 tissues, including three vegetative tissues (leaves, root, and nodules) and the whole seeds from 11 stages of reproductive tissue development (flower, pod, and seed), were downloaded from SoyBase<sup>1</sup>, and soybean microarray expression data was downloaded from Plant

Expression Database<sup>2</sup> (GEO Accession: GSE26443). The heat maps were generated by using R software packages “heatmap.”

## Candidate Gene Selection and Gene Expression Analysis

Based on the LD decay distance of significant SNPs in the GWAS results, we predicted the putative genes related to pod dehiscence according to the soybean genome annotation (Wm82.a1.v1.1<sup>3</sup>). Meanwhile, BLASTP analysis was also performed against *Arabidopsis* proteins using the amino acid sequences of these putative genes.

The soybean accession Williams 82 was selected to analyze the tissue expression patterns of the putative genes. First, seeds were grown in growth chambers under the conditions of 16/8 h (day/night), 28/25°C (day/night), and 60% relative humidity. Then, roots, stems, leaves, and flowers were sampled during the full-bloom period; pods were sampled on the 7th, 15th, 25th, and 45th days after flowering (DAF), and mature seeds were also collected. The expression levels of the putative genes were detected in 15-day pods in a subset of 18 soybean accessions, representing varieties with high and low PPD. Total RNA was isolated using a Plant RNA Extract Kit (TianGen, Beijing, China), and cDNA was synthesized using a TaKaRa Prime Script<sup>TM</sup> RT reagent kit according to the manufacturer's instructions. Gene expression in different tissues was performed by qRT-PCR assays using an ABI 7500 system (Applied Biosystems, Foster City, CA, United States) with SYBR Green Real-time Master Mix (Toyobo). Three biological and three technical replicates were used. The constitutively expressed soybean tubulin gene (GenBank accession number: AY907703) was used as a control. The normalized expression, reported as fold change, was calculated for each sample as  $\Delta\Delta CT = (CT_{Target} - CT_{Tubulin})_{genotype} - (CT_{Target} - CT_{Tubulin})_{calibrator}$  (Livak and Schmittgen, 2001). The primers used are listed in **Supplementary Table S2**.

## Gene Structure Analysis

GSDS<sup>4</sup> was used to analyze the gene structure. The protein sequences of homologous genes from different plants were obtained from NCBI database<sup>5</sup> (accession numbers are listed in **Supplementary Table S3**). Sequence alignment was analyzed using ClustalX software version 1.83 (Jeanmougin et al., 1998). A NJ phylogenetic tree was constructed based on protein sequences with MEGA 6.0 software using bootstrap method with 1,000 replications.

## Haplotype Analysis

A pair of gene-specific primers was designed (Primer 5.0) to amplify *Glyma09g06290* from 42 soybean accessions (forward: 5'-CCAAACAGAGTGAGTGACTTG-3', reverse: 5'-TGCTACTTTCTTCTTAGC-3'). These sequences were aligned using the ClustalX software version 1.8, and the SNPs with MAFs  $\geq 0.05$  were identified among these accessions. The

<sup>2</sup><http://www.plexdb.org>

<sup>3</sup><http://phytozome.jgi.doe.gov/pz/portal.html>

<sup>4</sup><http://gsds.cbi.pku.edu.cn/>

<sup>5</sup><http://www.ncbi.nlm.nih.gov/>

<sup>1</sup><http://www.soybase.org/soyseq/>

LD level was calculated by the HapView 4.0 software program (Barrett et al., 2005). The correlation between the expression of *Glyma09g06290* and pod dehiscence was calculated by SPSS 20.0 software (Pearson correlation). In addition, the haplotype block was defined by the “Solid Spine of LD” algorithm. LD analysis of 112 significant SNPs located on chromosome 16 was performed using the R software package “LDheatmap” (Shin et al., 2006).

## dCAPS Marker Development

The dCAPS marker of *Glyma09g06290* was developed based on the S<sub>-</sub>500 (C/G). The gene-specific primers were designed by dCAPS Finder 2.0<sup>6</sup> (forward: 5'-CTCATATCTTTGTGTTGTCATG-3', reverse: 5'-ATTGGAGTGATAAGGAACTGG-3'). The PCR products were digested with restriction endonuclease *NcoI* (37°C, 1 h). Digestion products were assayed by 8% polyacrylamide gel electrophoresis to distinguish the polymorphic fragments.

## RESULTS

### Phenotypic Variation in Association Panel

Descriptive statistics, ANOVA and  $h^2$  of the PPD were identified for the association panel in five environments (Table 1). The PPD for the individual accessions in the association panel ranged from 0 to 0.99, and the average value was between 0.15 and 0.27.  $h^2$  was estimated as 78.5% in association panel, which was slightly lower than that found in RILs (Kang et al., 2005). The frequency distribution showed a skewed distribution (Figure 1), not a normal distribution, which is similar to the previously reported results (Kang et al., 2009). In addition, the ANOVA results indicated that pod dehiscence was significantly influenced by the genotype effect and the interaction effects of genotype and environment ( $P < 0.001$ ) in association panel.

### Genome-Wide Association Analysis for Pod Dehiscence

QTL mapping suffers from fundamental limitations; one of these is the limited allelic diversity between recombinant inbred population and parents (Korte and Farlow, 2013). In contrast, by taking advantages of the abundant natural variations of a population, GWAS is helpful for favored allele detecting and functional SNP discovery. In this study, to mine the SNPs significantly associated with pod dehiscence, GWAS was conducted using 201,915 SNPs (NJAU 355K SoySNP array) with the MLM (Q + K) model in an association panel including 211 accessions.

A total of 163 SNPs were identified as significantly associated with pod dehiscence in five environments and BLUP (Figure 2, Table 2, and Supplementary Table S4). Among these SNPs, 136 SNPs were located on chromosome 16 in multiple environments. Among these 136 SNPs, one SNPs was detected in five environments and BLUP. In addition, 51 SNPs were identified

in four environments and BLUP; 18 SNPs were mapped in three environments and BLUP; and 19 SNPs were mapped in two environments and BLUP. These 136 SNPs were clustered on chromosome 16 from 29135922 to 29865027 (~729 kb), which completely overlapped with two QTLs (*Pod dehiscence 2-1* and *Pod dehiscence 3-4*) and partially overlapped with *Pod dehiscence 1-7* and *Pod dehiscence 4-2* (Bailey et al., 1997; Funatsuki et al., 2006; Kang et al., 2009; Yamada et al., 2009). The gene *pdh1*, controlling pod dehiscence in soybean, was also located in this cluster. In addition, a few significant SNPs were distributed on other chromosomes, such as chromosome 1, 4, 6, 8, 9, 11, 17, 18, and 20 (Figure 2, Table 2, and Supplementary Table S4). On chromosome 9, two SNPs were identified in both Env3 and Env5 as significant SNPs.

### Identification of Favorable SNPs and Haplotypes Relevant to Pod Dehiscence

One SNP was mapped in both five environments and BLUP on chromosome 16; we also found that nine SNPs were mapped in four environments and BLUP on chromosome 16 (the  $P$ -value of nine SNPs in Env3 was as low as  $9.91 \times 10^{-06}$ ); the  $R^2$  ranged from 11.9 to 38.5% (Table 2 and Supplementary Table S4). These significant SNPs have more allelic variations and can represent the quality of pod dehiscence in our association panel. Therefore, these SNPs were used to analyze the haplotypes related to pod dehiscence in soybean. We selected 189 of 211 accessions from different geographic origins for further haplotype analysis. Ten significant loci showed very strong LD level and formed one block. The number 1–10 represents these ten SNPs AX-93853844, AX-94151086, AX-93853870, AX-93853873, AX-93853874, AX-94151101, AX-93853876, AX-93853895, AX-93853896, and AX-94151124, respectively (Figure 3A and Supplementary Table S5). Of which, the SNP 8 and 3, 4, 7, 9, 10 showed complete LD. When 8 is T base, 3, 4, 7, 9, 10 is C, T, G, A, T base, respectively; when 8 is A base, 3, 4, 7, 9, 10 is A, C, A, T, C base, respectively (Figure 3B). Therefore, these ten SNPs were subdivided into five SNPs (1, 2, 5, 6, and 8). Based on the five SNPs within the LD block, six haplotype classes were observed in the 189 accessions (Hap1-Hap6) (Figure 3B). Hap2 (TGCAA) was the largest group ( $n = 83$ ); and Hap1 (CCTTT) was the second largest group ( $n = 72$ ). The other haplotypes, Hap3 (CCCAA), Hap4 (CGTTT), Hap5 (CCT0T), and Hap6 (CC0TT), comprised seven, fourteen, eight and five accessions, respectively.

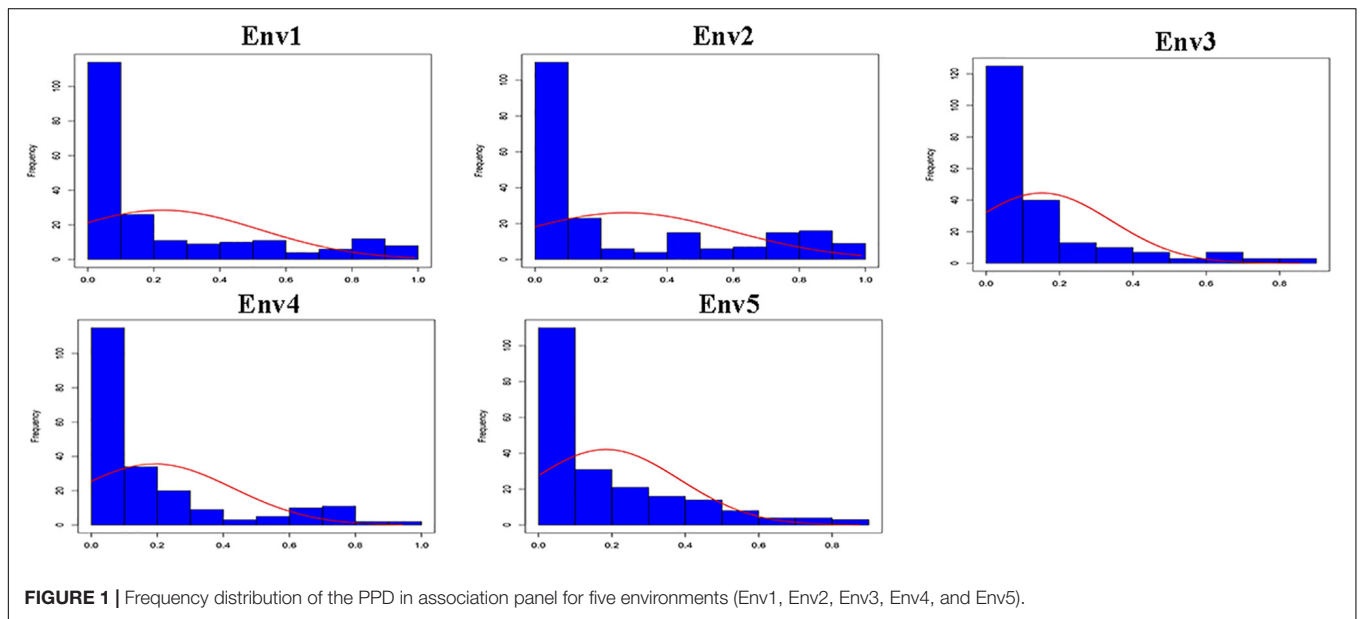
The phenotypic data of five significant SNPs and six haplotypes were further analyzed. In total, 117 accessions with 1-C exhibited significantly higher values than 1-T; 104 accessions with 2-C exhibited significantly higher values than 2-G; 101 accessions with 5-T and 12 accessions with 5-0 exhibited significantly higher values than 5-C; 98 accessions with 6-T and 14 accessions with 6-0 exhibited significantly higher values than 6-A, 109 accessions with 8-T exhibited significantly higher values than 8-A (Figures 3C–G). In addition, Hap1 and Hap5 had significantly higher PPD than Hap2 and Hap3 (Figure 3H). Considering that soybean varieties with pod-shattering resistance are one of the major goals for breeders, 1-T, 2-G, 5-C, 6-A, and

<sup>6</sup><http://helix.wustl.edu/dcaps/dcaps.html>

**TABLE 1** | Descriptive statistics, ANOVA and broad-sense heritability of the PPD in five environments of association panel (Env1, Env2, Env3, Env4, and Env5).

Env	Mean	SD	Median	Minimum	Maximum	CV	Skewness	Kurtosis	G	G × E	$h^2$
Env1	0.23	0.30	0.08	0	0.99	130.4%	1.24	0.17	***	***	78.5%
Env2	0.27	0.32	0.09	0	0.99	118.5%	0.91	-0.77	***	***	
Env3	0.15	0.19	0.08	0	0.86	126.6%	1.91	3.03	***	***	
Env4	0.19	0.24	0.09	0	0.94	126.3%	1.59	1.35	***	***	
Env5	0.20	0.23	0.09	0	0.96	115.0%	1.44	1.07	***	***	

Env, environment; SD, standard deviation; CV, coefficient of variation; G, genotype; G × E, genotype × environment;  $h^2$ , broad-sense heritability. \*\*\*Significant at  $P < 0.001$ .

**FIGURE 1** | Frequency distribution of the PPD in association panel for five environments (Env1, Env2, Env3, Env4, and Env5).

8-A were the favorable markers; Hap2 and Hap3 should be the favorable haplotypes.

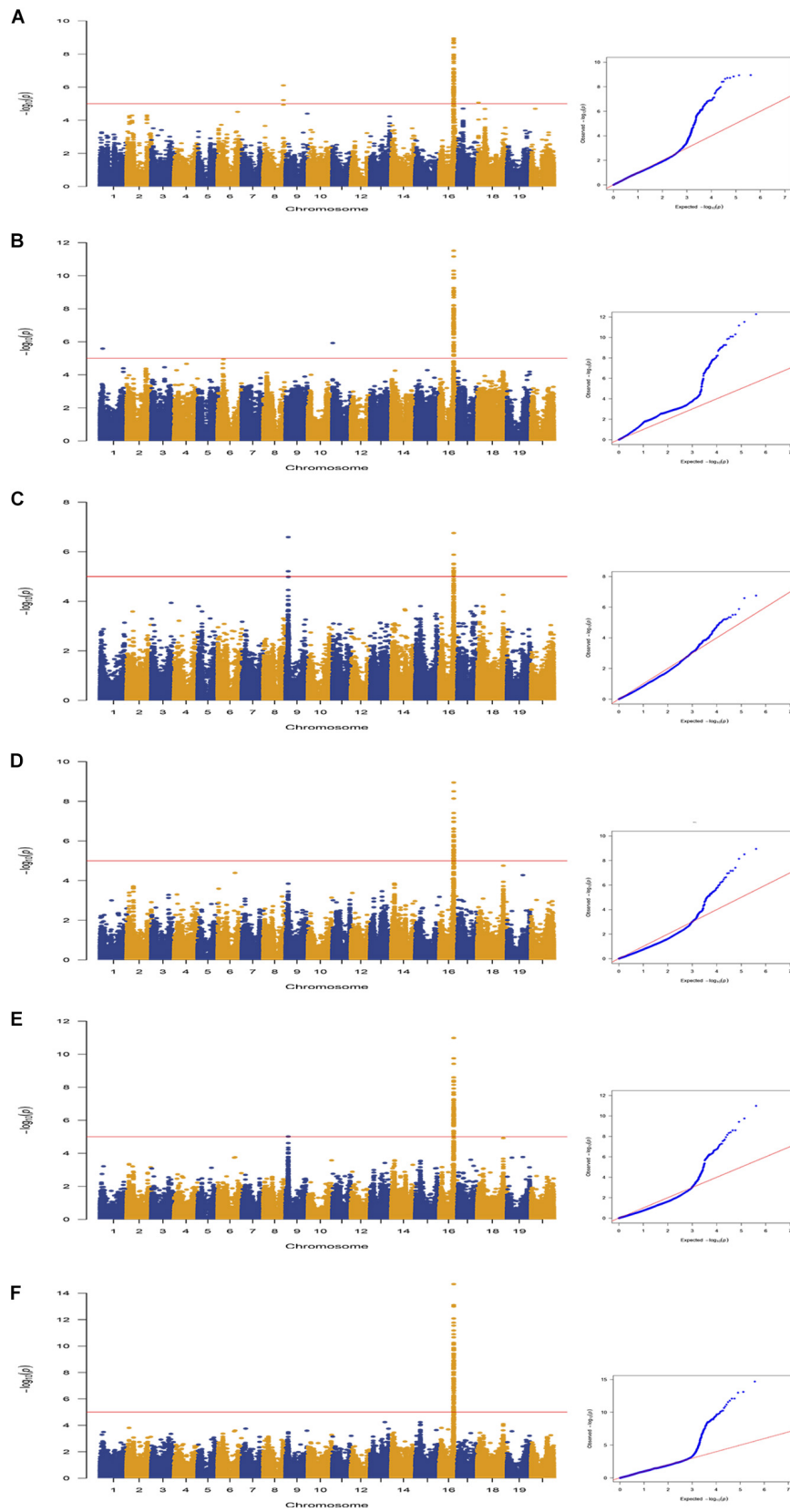
Hap1 and Hap2 were not only the unfavorable and favorable haplotypes but also the largest and second largest groups, respectively. The number of Hap3 and Hap5 was limited (seven and eight accessions, respectively). Therefore, Hap1 and Hap2 were used to analyze the geographical distribution of soybean landraces with different pod dehiscence traits. After filtering out accessions with indeterminate ecoregions, 66 and 76 accessions remained in Hap1 and Hap2, respectively. Hap1 was observed in the three main growing ecoregions in China (Tian et al., 2010) and was mainly (75.8%) distributed in the southern ecoregion (SR); Hap2-containing accessions were mostly observed in the Huang-Huai ecoregion (HR) (34.2%) and SR (34.2%) (Table 3). These results revealed that the soybean landraces of Chinese origin possessing the elite haplotype were mostly distributed in HR and SR.

## Identification of Putative Genes

In GWAS results, under the threshold of  $FDR < 0.05$ , we detected three significant SNPs on chromosome 9. Moreover, two of them were close to each other and repeatedly identified in different environments. Interestingly, in other regions, there was no significant SNP that was repeatedly detected

except on chromosome 16. Notably, one of these three SNPs, AX-93762848, was also detected by using MLM model (Supplementary Figure S2). It demonstrates that these three SNPs on chromosome 9 were essential to dehiscence trait in soybean. Therefore, the candidate genes were selected based on the SNPs on chromosome 9.

There were 26 genes detected in the 130 kb flanking regions (LD decay distance of cultivated soybeans) of the two SNPs (Table 4; Wang J. et al., 2016). We then performed a BLASTP analysis using the *Arabidopsis* genome against these 26 genes. The results showed that two genes were without any explicit biological function annotations. Among the other 24 genes with functional annotations, *Glyma09g06320* and *Glyma09g06390* were predicted to be involved in cellular differentiation and cell wall formation. Additionally, the gene *Glyma09g06290* is homologous to *At2g18969* which belongs to *bHLH* (basic helix-loop-helix) gene family. In *Arabidopsis*, two *bHLH* transcriptional factors (IND and ALC) were reported to be related to silique dehiscence (Rajani and Sundaresan, 2001; Liljegren et al., 2004). The other genes were forecasted to participate in abiotic and biotic stress, transport, metabolic process, and other functions. We also listed the detail information about the orthologs of these genes in *Medicago truncatula*, *Oryza sativa*, and *Zea mays*, respectively (Supplementary Table S6). Moreover, we identified



**FIGURE 2 |** Continued

**FIGURE 2 |** Genome-wide association study of pod dehiscence in association panel using the MLM (Q + K) model across multiple environments. Manhattan plots and Q-Q plots for pod dehiscence in the following environments: **(A)** Env1; **(B)** Env2; **(C)** Env3; **(D)** Env4; **(E)** Env5; and **(F)** BLUP. The horizontal line depicts the significance threshold when FDR = 0.05.

**TABLE 2 |** SNP loci that significantly associated with pod dehiscence based on GWAS in the association panel.

Chr	MSS	MSS Position	Mss P-value	Significant region		No. of SNPs	Env	Related QTL or gene identified in previous studies
				Start	End			
1	AX-93963018	5388850	$2.58 \times 10^{-06}$	5258850	5518850	1	Env2	–
4	AX-93708700	25726440	$2.17 \times 10^{-05}$	25596440	25856440	1	Env2	–
6	AX-93727683	11784406	$1.13 \times 10^{-5}$	10563310	41546032	3	Env2 (2), Env1 (1)	–
8	AX-93930129	42050788	$7.83 \times 10^{-07}$	41864756	42186397	11	Env1	<i>Pod dehiscence 4-1</i> (Yamada et al., 2009)
9	AX-93762848	5145974	$2.58 \times 10^{-07}$	4940686	5275974	3	Env3 Env5 (2), Env3 (1)	–
11	AX-94082716	577915	$1.19 \times 10^{-06}$	447915	707915	1	Env2	–
16	AX-93853895	29601807	$2.06 \times 10^{-15}$	29365082	29736722	1	Env1 Env2 Env3 Env4 Env5 BLUP	–
16	AX-93853825	29421300	$5.78 \times 10^{-10}$	29194421	29850274	51	Env1 Env2 Env4 Env5 BLUP	<i>Pod dehiscence 1-7</i> (Bailey et al., 1997) <i>Pod dehiscence 2-1</i> (Funatsuki et al., 2006) <i>Pod dehiscence 3-4</i> (Kang et al., 2009) <i>Pod dehiscence 4-2</i> (Yamada et al., 2009) <i>Pod wall width 1-9</i> (Guang-yu et al., 2011) <i>pdh1</i> (Funatsuki et al., 2014)
16	AX-93853894	29600010	$1.29 \times 10^{-10}$	29136226	29870913	18	Env1 Env2 Env5 BLUP (14) Env1 Env4 Env5 BLUP (4)	–
16	AX-93853796	29340570	$2.84 \times 10^{-08}$	29210570	29879081	19	Env1 Env2 BLUP (16) Env2 Env5 BLUP (3)	–
16	AX-93946746	29215338	$1.15 \times 10^{-07}$	29005922	29968491	11	Env1 BLUP	–
16	AX-93853865	29545423	$5.30 \times 10^{-07}$	29018509	29990917	36	Env1 (12), Env2 (4), BLUP (20)	–
17	AX-94289430	11868679	$1.98 \times 10^{-5}$	11738679	12008912	3	Env1	–
18	AX-94165965	2292799	$8.60 \times 10^{-06}$	2162799	54328119	3	Env1	–
20	AX-93901310	8202869	$2.03 \times 10^{-05}$	8072869	8332869	1	Env1	–

Chr, chromosome; MSS, most significant SNP; Significant region, significant region was defined as the LD decay distance flanking the significant SNPs; No. of SNPs, number of significant SNPs; Env, environment. The number in brackets indicates the number of significant SNPs detected in different environments.

the expression patterns of these 24 genes in the RNA-Seq and microarray data, respectively (**Supplementary Figure S3**). The RNA-Seq data showed that *Glyma09g06290* was predominately expressed in pod shell and *Glyma09g06390* was expressed in several tissues. However, *Glyma09g06320* was not expressed according to the RNA-seq data, and this result was confirmed by our qRT-PCR result. In the microarray data, we found that *Glyma09g06290* was highly expressed in pod elongation stage and sharply decreased in late seed growth period. By combining the above results, these two genes (*Glyma09g06390* and *Glyma09g06290*) may be related to pod dehiscence.

### Expression Patterns of Putative Genes

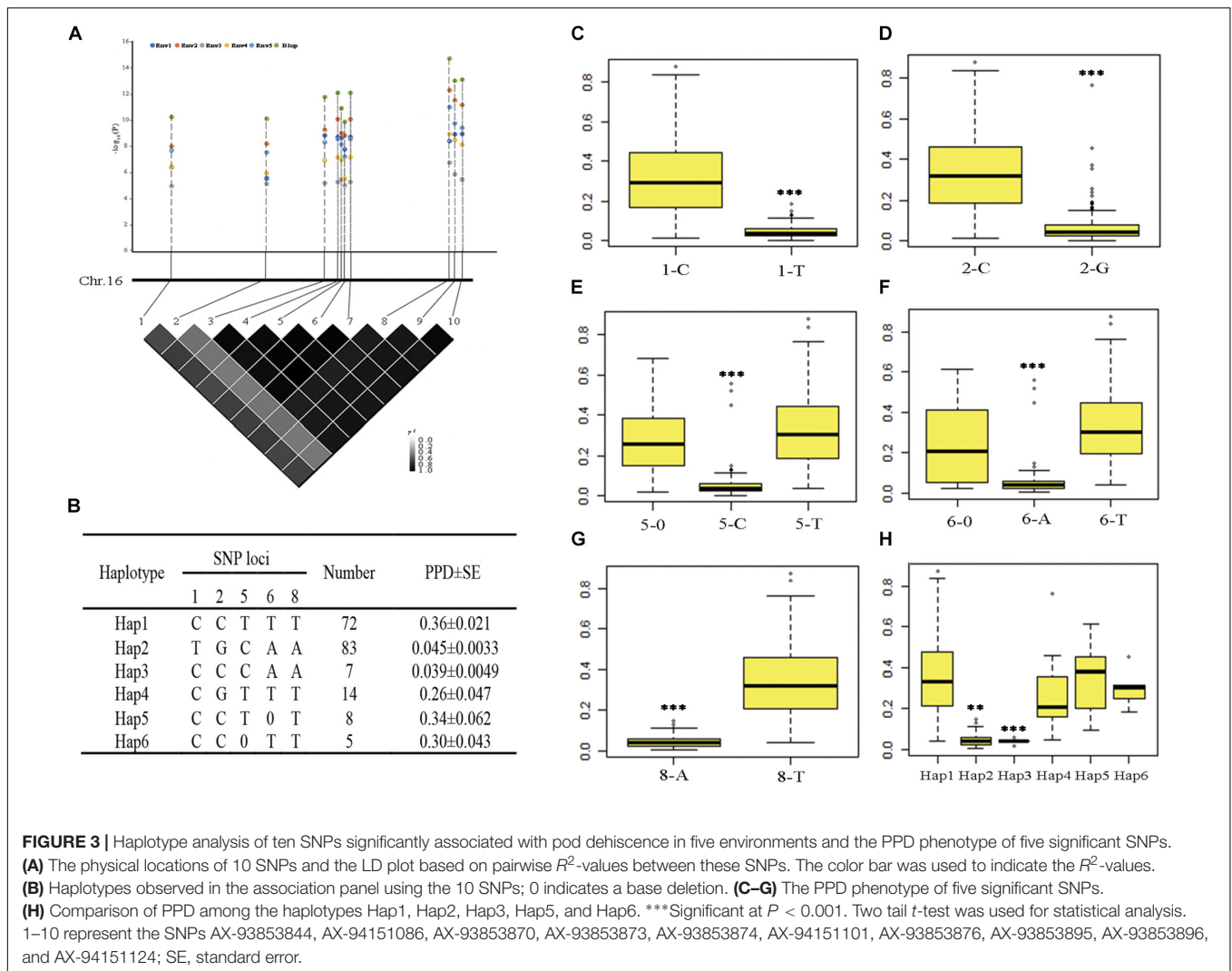
According to the Soybase<sup>7</sup>, we listed the candidate genes' coding product and metabolic processes (**Supplementary Table S7**). Moreover, to elucidate the potential functions of the two genes, qRT-PCR was performed to study the expression patterns of

these two genes. First, the soybean accession Williams 82 was selected to analyze the tissue expression pattern of the putative genes. *Glyma09g06290* was highly expressed in the pod, and the expression of *Glyma09g06290* was significantly increased in the late pod growth period, however, the expression of *Glyma09g06390* was very low in pod growth period (**Figure 4A**). These results were similar to the public RNA-Seq data (**Supplementary Figure S3**). Second, 18 soybean accessions representing varieties with high and low PPD were used to analyze the expression patterns in extreme varieties. The results showed that the expression of *Glyma09g06290* rather than that of *Glyma09g06390* in the high-PPD varieties was significantly higher than that in the low-PPD varieties in 15-day pods (**Figure 4B**). Therefore, *Glyma09g06290* may be involved in pod dehiscence in soybean.

### Sequence Analysis of *Glyma09g06290*

The full genomic length of *Glyma09g06290* is 1405 bp and only contains one exon with 624 bp. *Glyma09g06290*

<sup>7</sup><http://www.soybase.org/>



protein is composed of 207 amino acids with molecular mass of 23.23 kDa and isoelectric point of 5.39. Further, phylogenetic trees showed that *Glyma09g06290* is homologous to *Phaseolus vulgaris* Phvul.009g238800 (identity of 78.7%), *Medicago truncatula* Medtr2g036670 (identity of 61.8%), *Populus trichocarpa* Potri.018g090200 (identity of 61.8%), *M. truncatula* Medtr4g127700 (identity of 50.7%) and *A. thaliana columbia* At2g18969 (identity of 44.9%; **Figure 5**). By comparing the gene structure of *Glyma09g06290* and other plant homologous genes, we identified that they have only one exon and without intron except *Medtr4g127700* (**Figure 5**).

## Polymorphisms in the *Glyma09g06290* Gene Are Associated With Pod Dehiscence

In order to analyze the association between the allelic variation of *Glyma09g06290* and pod dehiscence, we sequenced the *Glyma09g06290* gene in a subset of 20 accessions with high PPD and 22 accessions with low PPD. An approximately 1.5-kb

genomic region, spanning the 5'- to 3'-UTR of *Glyma09g06290*, was analyzed. A total of three SNPs and two indels (insertions and deletions) were identified, including Site\_-500 (located 500 bp upstream from the translation start codon, S\_-500), Indel\_-230, S\_-128 in the 5'-UTR, Indel\_766 and S\_767 in the 3'-UTR (**Figure 6**). The association study showed that 5 probable causative sites were significantly associated with variations in pod dehiscence (**Figure 6A**). Furthermore, based on the five significant variants with strong LD, the 42 soybean genotypes

**TABLE 3 |** Geographic distribution of Hap1 and Hap2 in three soybean-growing ecoregions (NR, HR, and SR) in China.

Haplotype	Number				
	NR	HR	SR	HR/SR	NR/HR
Hap1	3 (4.6%)	2 (3.0%)	50 (75.8%)	9 (13.6%)	2 (3.0%)
Hap2	9 (11.8%)	26 (34.2%)	26 (34.2%)	8 (10.5%)	7 (9.3%)

NR, northern ecoregion; HR, Huang-Huai ecoregion; SR, southern ecoregion.



**TABLE 4 |** Genes detected in the 130 kb flanking region of the significant SNPs on chromosome 9.

Gene ID	Homolog	Functional annotation in the Phytozome database
Glyma09g06180	AT5G24080	Protein kinase superfamily protein
Glyma09g06190	AT5G24080	Protein kinase superfamily protein
Glyma09g06200	AT5G24080	Protein kinase superfamily protein
Glyma09g06220	AT5G57360	Clock-associated PAS protein ZTL
Glyma09g06230	AT2G18940	Tetratricopeptide-repeat-like superfamily protein
Glyma09g06250	AT4G30190	P-type ATPase superfamily
Glyma09g06260	AT5G41750	Disease resistance protein
Glyma09g06275	AT5G51630	Disease resistance protein
Glyma09g06290	AT2G18969	bHLH family transcription factor
Glyma09g06300	AT4G25910	NFU domain protein
Glyma09g06310	AT5G57340	Ras guanine nucleotide exchange factor Q-like protein
Glyma09g06320	AT3G62360	Carbohydrate-binding-like fold
Glyma09g06330	AT5G41540	Disease resistance protein
Glyma09g06350	AT2G18980	Peroxidase superfamily protein
Glyma09g06365	AT1G62520	Sulfated surface-like glycoprotein
Glyma09g06380	AT5G57330	Galactose mutarotase-like superfamily protein
Glyma09g06390	AT4G30160	Major actin filament bundling protein
Glyma09g06410	AT1G71840	WD-40 repeat family protein
Glyma09g06431	/	/
Glyma09g06450	AT1G17690	U3 small nucleolar RNA-associated protein
Glyma09g06460	AT3G20250	Arabidopsis Pumilio (APUM) protein
Glyma09g06470	AT2G19080	Metaxin-like protein
Glyma09g06480	AT2G19090	DUF630 family protein
Glyma09g06491	ATCG00905	/
Glyma09g06500	/	Chloroplast gene encoding ribosomal protein s12
Glyma09g06521	AT5G54780	Gyp1p superfamily protein

were classified into four haplotype classes (Hap1-Hap4). Hap 1 ( $n = 24$ ) is the largest group, and Hap2 ( $n = 13$ ) is the second largest group. Statistically, Hap2 had significantly higher PPD than Hap1 (Figures 6B,C). We then measured the expression of *Glyma09g06290* in pods from 13 of 42 soybean accessions. The expression of this gene was positively correlated with pod dehiscence ( $r^2 = 0.58$ ,  $P < 0.05$ ) (Supplementary Table S8). Therefore, these results suggested that the expression of *Glyma09g06290* could partially explain the phenotypic variation in pod dehiscence.

### Development of a Functional Marker for *Glyma09g06290*

In this study, five significant SNPs of *Glyma09g06290* were identified and associated with pod dehiscence (Figures 6A–C and Supplementary Figure S4). The S<sub>-500</sub>-C exhibited significantly lower values than S<sub>-500</sub>-G; Indel<sub>-230</sub>-0 exhibited significantly lower values than Indel<sub>-230</sub>-C; S<sub>-128</sub>-C exhibited significantly lower values than S<sub>-128</sub>-A; Indel<sub>766</sub>-T exhibited significantly lower values than Indel<sub>766</sub>-0; S<sub>767</sub>-T exhibited significantly lower values than S<sub>767</sub>-G (Figure 6B). Therefore, we developed

the dCAPS marker based on one of these SNPs (S<sub>-500</sub>). As shown in Figure 6D, we selected ten accessions with high PPD and ten accessions with low PPD. Ten accessions with high PPD produced a 166 and 143 bp amplicons by restriction enzyme digestion. The remaining ten accessions with low PPD produced 166 bp amplicons by restriction enzyme digestion. These results confirmed that dCAPS based on the S<sub>-500</sub> is a functional marker and could be useful for plant breeders.

## DISCUSSION

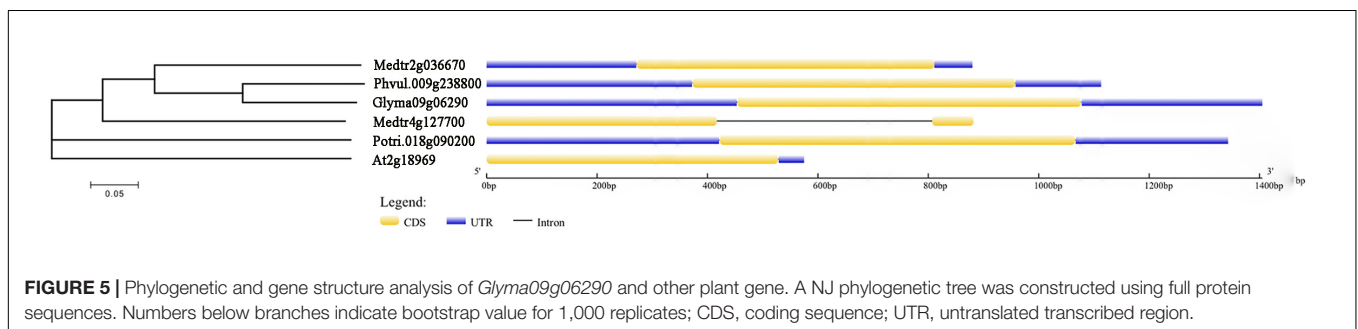
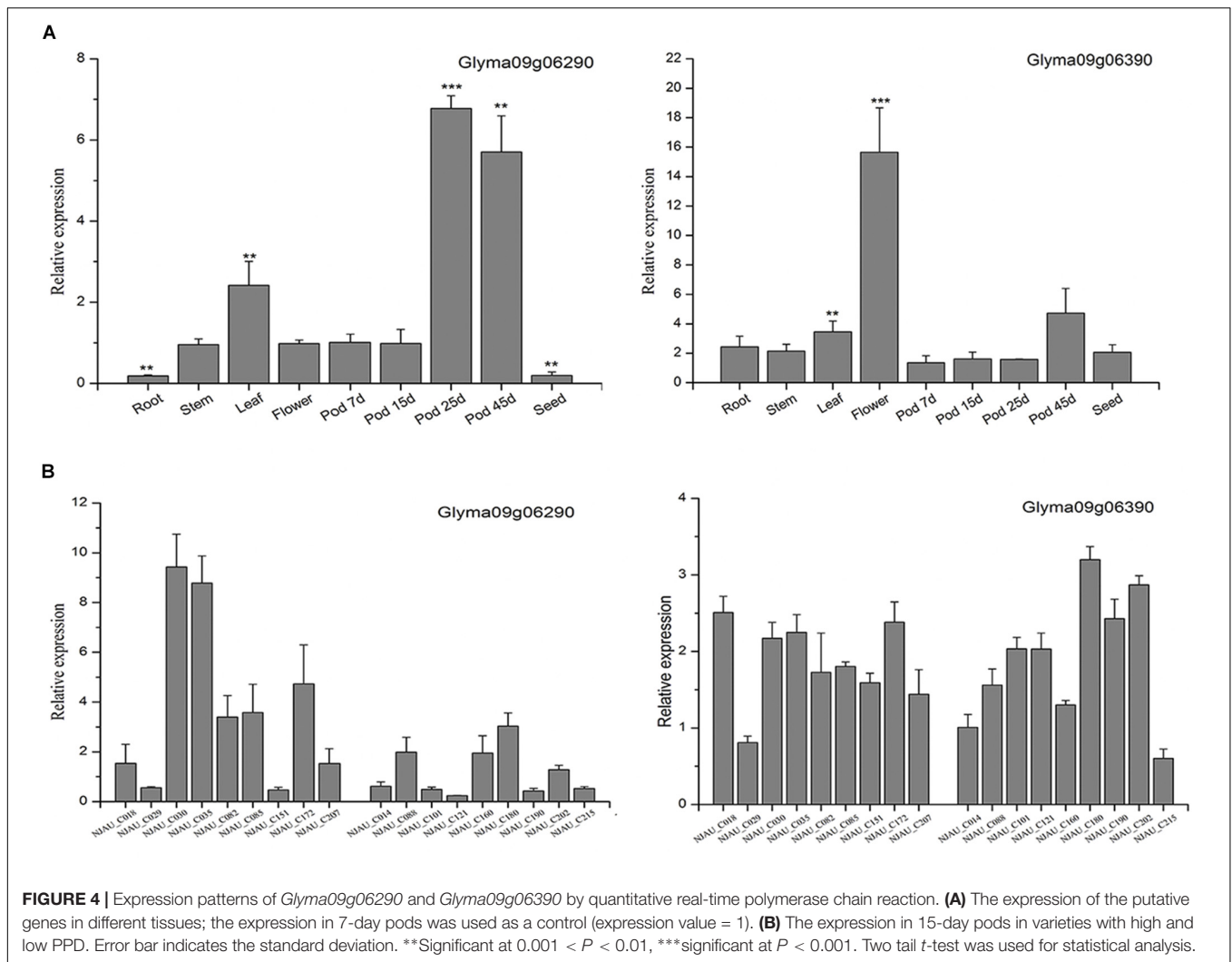
### Novel Loci Related to Pod Dehiscence Were Identified

The dehiscence of pods (shattering) prior to harvest is a major cause of yield loss in soybean production. Thus, it is extremely important to identify QTLs or genes related to pod dehiscence and apply them to MAS. In this study, by using the association panel which possessed abundant natural variations, GWAS was performed to genotype 211 cultivated soybean accessions with high-quality SNP markers. We identified 163 SNPs. Of which, 136 significant SNPs on chromosome 16 ranged from 29135922 to 29865027 (~729 kb). These results were consistent with those of a previous study, which showed that the major QTL was located on chromosome 16 (Yamada et al., 2009). Similarly, a very large GWAS also confirmed that shattering in soybean is mainly due to genes located on chromosome 16 (Zhou et al., 2015). Additionally, LD analysis exhibited tight linkage among the 136 associated SNPs (Supplementary Figure S5). One of 136 SNP (AX-93853895) with the lowest  $P$ -value and maximum  $R^2$  was in the intron of *pdh1*. Consequently, the formation of this SNP cluster on chromosome 16 may be due to the strong effect of *pdh1*. Moreover, the domestication gene related to pod dehiscence between cultivated and wild soybean, *SHAT1-5*, was not mapped in our GWAS results. Thus, the gene effect of *SHAT1-5* is likely to be weak in cultivated soybean accessions.

In addition, we also detected 25 significant SNPs except on chromosome 16 (Table 2). For example, on chromosome 8, one region contains 11 significant SNPs that were significantly associated with pod dehiscence, and this region is close to *Pod dehiscence 4-1*; On chromosome 9, 2 SNPs were co-identified in *Env3* and *Env5*; there are one, one, three, one, three, three and one significant SNPs identified on chromosome 1, 4, 6, 11, 17, 18, and 20, respectively. These SNPs or regions may be the novel loci related to pod dehiscence. Overall, the SNPs identified in this study are helpful to further understanding the genetic basis of soybean pod dehiscence.

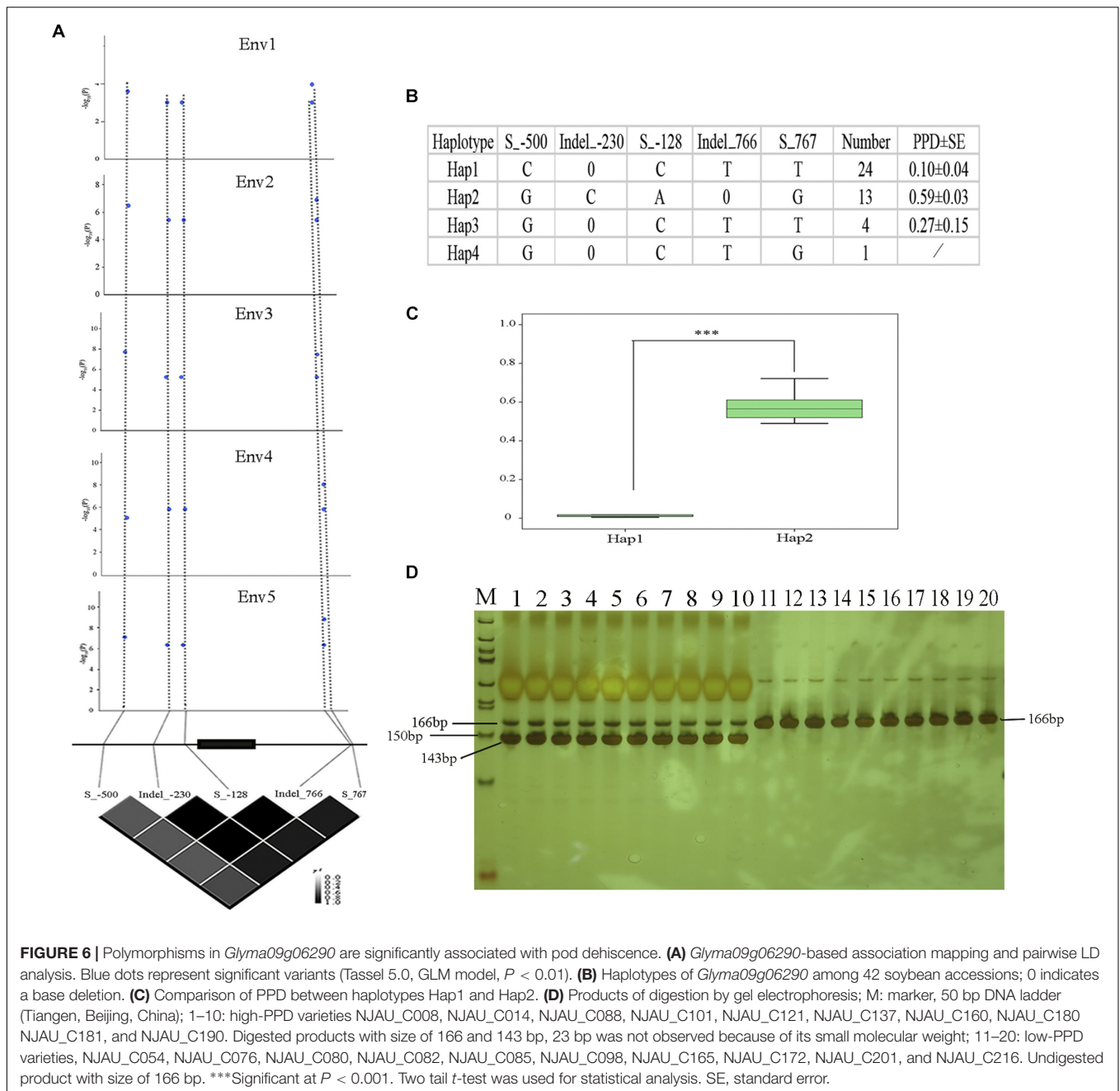
### Identification of Favorable Haplotypes Related to Pod Dehiscence in Chinese Soybean Germplasms

Haplotype analysis of a candidate gene or peak SNPs have been reported for soybean (Zhang et al., 2014), rice (Famoso et al., 2011), and cotton (Su et al., 2018). For example, in cotton, four peak SNPs located on chromosome D03 were simultaneously



associated with four plant architecture component traits. The four peak SNPs revealed five haplotypes, and Hap2 was the most favorable haplotype. In our study, ten significant SNPs were associated with pod dehiscence in Env1-Env5 and BLUP. Haplotype analysis showed that the 10 SNPs could be found as six haplotypes (Hap1-Hap6). Hap2 and Hap3 had lower PPD than the other haplotypes, demonstrating that Hap2 and Hap3 might be significant to breeding soybeans with lower PPD (Figure 3). According to the study, more than half of the

analyzed Chinese landraces possesses the *pdh1* allele (Funatsuki et al., 2014). In this study, the 189 soybean accessions were selected to represent all three ecological regions of soybean cultivation in China and soybeans with different pod dehiscence qualities. However, as more than 20,000 soybean accessions have been preserved (Wang et al., 2006), Hap2 or Hap3 could be used to discover other excellent soybean germplasms with resistance to pod dehiscence. In addition, two typical haplotypes (Hap1 and Hap2) were used to analyze the geographical



distribution of Chinese soybean landraces. We found that the favorable haplotype (Hap2) was mostly observed in the HR and SR. Taken together, the above results could facilitate the development of molecular markers for the breeding of soybean accessions with lower PPD.

## Identification of Candidate Genes for Pod Dehiscence in Soybean

Transcription factors play an important role in controlling pod dehiscence. The genetic network directing the morphogenesis of the dehiscence zone in *Arabidopsis* fruit has been identified

(Ballester and Ferrandiz, 2017). For example, two members of the MADS-box transcription factor family, SHATTERPROOF1 (SHP1) and SHATTERPROOF2 (SHP2), act redundantly to control silique dehiscence (Liljegren et al., 2000). FRUITFULL (FUL), which also belongs to the MADS-box family, negatively regulates *SHP1/2* expression (Ferrandiz et al., 2000). Moreover, two bHLH transcription factors, INDEHISCENT (IND) and ALCATRAZ (ALC), act downstream of *SHP1/2* (Dong and Wang, 2015). Briefly, the four genes *SHP1*, *SHP2*, *IND*, and *ALC* are expressed at the valve margin to direct dehiscence zone formation. In soybean, the pod-dehiscence-related gene *SHAT1-5* also encodes a NAC transcription factor. Interestingly,

previously identified genes that related to pod dehiscence or fruit shedding are homologous to those involved in seed shattering in *Arabidopsis thaliana*. However, Rau et al. (2019) identified a QTL, *qPD5.1-Pv*, represents a novel locus for the shattering trait in common bean (Rau et al., 2019).

In this study, on chromosome 9, two SNPs (AX-94059226 and AX-93762822) were observed under the threshold of  $FDR < 0.05$  in two environments (Env3 and Env5) (Table 2 and Supplementary Table S4). Moreover, we noticed that the SNP AX-93762848 was close to AX-94059226 and AX-93762822, which were observed in one environment with the  $FDR$  value  $< 0.05$  (Supplementary Table S4). Based on these SNPs, we identified a candidate gene, *Glyma09g06290*, which is located in the 130kb flanking region of two repeatedly observed SNPs, and the expression of *Glyma09g06290* was significantly increased in later pod growth stages. Furthermore, different expression patterns of *Glyma09g06290* were found in soybean varieties with high and low PPD (Figure 4). Then we sequenced the *Glyma09g06290* gene in a subset of 20 accessions with high PPD and 22 accessions with low PPD. Five SNPs were identified and used to analyze haplotypes, and the haplotype analysis showed that various haplotypes of the putative gene display variation in pod dehiscence (Figure 6). However, the biological functions of this gene should be further investigated. In summary, *Glyma09g06290* might be a putative gene involved in pod dehiscence in soybean.

In addition, the GWAS results revealed that 25 significant SNPs were identified in only one environment except for the SNPs on chromosome 16 (Table 2). However, eleven of them were consecutively distributed on chromosome 8, showed significant marker-trait association, with  $FDR$  as low as  $3.23 \times 10^{-03}$  ( $P$ -values =  $7.83 \times 10^{-07}$ ), and were close to *Pod Dehiscence 4-1* (Yamada et al., 2009) (Supplementary Table S4). Interestingly, among these eleven SNPs, seven SNPs were detected within the region of *Glyma08g42110* (one of them was detected within the exon). In addition, the *Glyma08g42110* is homologous to *ATGSL07*, which encodes callose synthase 7 (CalS7), a phloem-specific callose synthase responsible for callose deposition in developing sieve elements during phloem formation. *ATGSL07* mutant plants exhibited moderate reduction in seedling height and produced aberrant pollen grains and short siliques with aborted embryos in *A. thaliana* (Xie et al., 2011). Our further work will also focus on the biological functions of this gene.

## REFERENCES

- Bailey, M. A., Mian, M. A. R., Carter, T. E., Ashley, D. A., and Boerma, H. R. (1997). Pod dehiscence of soybean: identification of quantitative trait loci. *J. Heredity* 88, 152–154. doi: 10.3389/fpls.2016.01852
- Ballester, P., and Ferrandiz, C. (2017). Shattering fruits: variations on a dehiscent theme. *Curr. Opin. Plant Biol.* 35, 68–75. doi: 10.1016/j.pbi.2016.11.008
- Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265. doi: 10.1093/bioinformatics/bth457
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.

## CONCLUSION

In this study, 163 SNPs were identified using GWAS across multiple environments. In addition, favorable SNPs and six haplotypes relevant to pod dehiscence were identified. Furthermore, based on GWAS results, we identified a candidate gene *Glyma09g06290*. Expression patterns and allelic variation showed that *Glyma09g06290* was associated with the pod dehiscence. Additionally, we also developed a functional marker in *Glyma09g06290* for pod dehiscence in soybean, which may be useful for the breeders. Overall, these results would provide insights for molecular-assisted breeding strategies for resistance to pod dehiscence in soybean.

## DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.soybase.org/>.

## AUTHOR CONTRIBUTIONS

DY and FH designed the study. DHu, XL, and GK conducted the GWAS. DHu, DHao, WH, ZY, and XH carried out the field experiments. XL, HY, and YL carried out the qRT-PCR analysis. DHu wrote the manuscript. All authors approved the final version of the manuscript.

## FUNDING

This work was supported partly by the Ministry of Science and Technology (2017YFE0111000), the Key Transgenic Breeding Program of China (2016ZX08009003-004), and the National Natural Science Foundation of China (31571688, 31671715, and 31771821).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00811/full#supplementary-material>

- 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.
- Caviness, C. E. (1969). Heritability of pod dehiscence and its association with some agronomic characters in soybeans 1. *Crop Sci.* 9, 207–209.
- Chu, S., Wang, J., Zhu, Y., Liu, S., Zhou, X., Zhang, H., et al. (2017). An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. *PLoS Genet.* 13:e1006770. doi: 10.1371/journal.pgen.1006770
- Dong, Y., and Wang, Y. Z. (2015). Seed shattering: from models to crops. *Front. Plant Sci.* 6:476. doi: 10.3389/fpls.2015.00476
- Dong, Y., Yang, X., Liu, J., Wang, B. H., Liu, B. L., and Wang, Y. Z. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat. Commun.* 5:3352. doi: 10.1038/ncomms4352

- Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587.
- Famoso, A. N., Zhao, K., Clark, R. T., Tung, C. W., Wright, M. H., Bustamante, C., et al. (2011). Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet.* 7:e1002221. doi: 10.1371/journal.pgen.1002221
- Ferrandiz, C., Liljegren, S. J., and Yanofsky, M. F. (2000). Negative regulation of the SHATTERPROOF genes by FRUITFULL during *Arabidopsis* fruit development. *Science* 289, 436–438.
- Fourquin, C., Cerro, C. D., Victoria, F. D. C., Vialteguiraud, A., De Oliveira, A. C., and Ferrandiz, C. (2013). A change in SHATTERPROOF protein lies at the origin of a fruit morphological novelty and a new strategy for seed dispersal in *Medicago* genus. *Plant Physiol.* 162, 907–917. doi: 10.1104/pp.113.217570
- Funatsuki, H., Ishimoto, M., Tsuji, H., Kawaguchi, K., Hajika, M., and Fujino, K. (2006). Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. *Plant Breed.* 125, 195–197.
- Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., et al. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. U.S.A.* 111:17797. doi: 10.1073/pnas.1417282111
- Guang-yu, G., Rui, S., Meng, H., Yong-xin, G., Da-wei, X., Hong-wei, J., et al. (2011). Quantitative trait locus (QTL) analysis of pod related traits in different environments in soybean. *Afr. J. Biotechnol.* 10, 11848–11854.
- Hao, D., Cheng, H., Yin, Z., Cui, S., Zhang, D., Wang, H., et al. (2012). Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. *Theor. Appl. Genet.* 124, 447–458. doi: 10.1007/s00122-011-1719-0
- Hu, W. (2017). *Linkage Analysis and Association Analysis of Pod Shattering and Yield Related Factors in Soybean*. Dissertation, Nanjing Agricultural university, Nanjing.
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961. doi: 10.1038/ng.695
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem. Sci.* 23, 403–405.
- Ju, S. L., Kim, K. R., Ha, B. K., and Kang, S. (2017). Identification of SNPs tightly linked to the QTL for pod shattering in soybean. *Mol. Breed.* 37:54.
- Kalladan, R., Lasky, J. R., Chang, T. Z., Sharma, S., Juenger, T. E., and Verslues, P. E. (2017). Natural variation identifies genes affecting drought-induced abscisic acid accumulation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11536–11541. doi: 10.1073/pnas.1705884114
- Kang, S. T., Hyeunkyeung, K., Inyoul, B., Mounngun, C., Han, W. Y., Doochull, S., et al. (2005). Genetic analysis of pod dehiscence in soybean. *Kor. J. Crop.* 50, 281–285.
- Kang, S. T., Kwak, M., Kim, H. K., Choung, M. G., Han, W. Y., Baek, I. Y., et al. (2009). Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [*Glycine max* (L.) Merr.]. *Euphytica* 166, 15–24.
- Kobayashi, Y., Sadhukhan, A., Tazib, T., Nakano, Y., Kusunoki, K., Kamara, M., et al. (2016). Joint genetic and network analyses identify loci associated with root growth under NaCl stress in *Arabidopsis thaliana*. *Plant Cell Environ.* 39, 918–934. doi: 10.1111/pce.12691
- Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045.
- Konishi, S., Izawa, T., Lin, S., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396.
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. doi: 10.1186/1746-4811-9-29
- Kump, K. L., Bradbury, P. J., Wissler, R. J., Buckler, E. S., Belcher, A. R., Oropezaros, M. A., et al. (2011). Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nat. Genet.* 43:163. doi: 10.1038/ng.747
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). Shatterproof MAD5-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770.
- Liljegren, S. J., Roeder, A. H. K., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., et al. (2004). Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* 116, 843–853.
- Liu, B., Fujita, T., Yan, Z. H., Sakamoto, S., Xu, D., and Abe, J. (2007). QTL mapping of domestication-related traits in soybean (*Glycine max*). *Ann. Bot.* 100, 1027–1038.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* 25, 402–408.
- Lo, S., Muñoz-Amatriain, M., Boukar, O., Herniter, I., Cisse, N., Guo, Y.-N., et al. (2018). Identification of QTL controlling domestication-related traits in cowpea (*Vigna unguiculata* L. Walp). *Sci. Rep.* 8:6261. doi: 10.1038/s41598-018-24349-4
- Merk, H. L., Yarnes, S. C., Deynze, V., Tong, N., Menda, N., Mueller, L. A., et al. (2012). Trait diversity and potential for selection indices based on variation among regionally adapted processing tomato germplasm. *Autism Res.* 137, 427–437.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Rajani, S., and Sundaresan, V. (2001). The *Arabidopsis* myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Curr. Biol.* 11, 1914–1922.
- Rau, D., Murgia, M. L., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2019). Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant J.* 97, 693–714. doi: 10.1111/tpj.14155
- Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibrium between single nucleotide polymorphisms. *J. Stat. Softw.* 16, 1–10. doi: 10.18637/jss.v016.c03
- Shirasawa, 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
- Su, J., Li, L., Zhang, C., Wang, C., Gu, L., Wang, H., et al. (2018). Genome-wide association study identified genetic variations and candidate genes for plant architecture component traits in Chinese upland cotton. *Theor. Appl. Genet.* 131, 1299–1314. doi: 10.1007/s00122-018-3079-5
- Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kaga, A., Tomooka, N., et al. (2016). Co-localization of QTLs for pod fiber content and pod shattering in F2 and backcross populations between yardlong bean and wild cowpea. *Mol. Breed.* 36:80.
- Suzuki, M., Fujino, K., Nakamoto, Y., Ishimoto, M., and Funatsuki, H. (2010). Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. *Mol. Breed.* 25, 407–418.
- Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8563–8568.
- Tsuchiya, T. (1986). Studies on shattering resistance in soybean breeding. *Rep. Hokkaido. Prefect. Agric. Exp. Stn* 58, 1–53. doi: 10.1270/jsbbs.61.554
- Tsuchiya, T. (1987). Physiological and genetic analysis of pod shattering in soybeans. *JARQ* 21, 166–175.
- Turner, S. D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Biorxiv*
- Vittori, V. D., Gioia, T., Rodriguez, M., Bellucci, E., Bitocchi, E., Nanni, L., et al. (2019). Convergent evolution of the seed shattering trait. *Genes* 10:68.
- Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., and Yu, D. (2016). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Sci. Rep.* 6:20728. doi: 10.1038/srep20728
- Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., et al. (2016). Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48:1233. doi: 10.1038/ng.3636
- Wang, L., Guan, Y., Guan, R., Li, Y., Ma, Y., Dong, Z., et al. (2006). Establishment of Chinese soybean *Glycine max* core collections with agronomic traits and SSR markers. *Euphytica* 151, 215–223.
- Xie, B., Wang, X., Zhu, M., Zhang, Z., and Hong, Z. (2011). CalS7 encodes a callose synthase responsible for callose deposition in the phloem. *Plant J. Cell Mol. Biol.* 65, 1–14. doi: 10.1111/j.1365-313X.2010.04399.x

- Yamada, T., Funatsuki, H., Hagihara, S., Fujita, S., Tanaka, Y., Tsuji, H., et al. (2009). A major QTL, qPDH1, is commonly involved in shattering resistance of soybean cultivars. *Breed. Sci.* 59, 435–440.
- Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48:927. doi: 10.1038/ng.3596
- Zhang, D., Song, H., Cheng, H., Hao, D., Wang, H., Kan, G., et al. (2014). The acid phosphatase-encoding gene GmACP1 contributes to soybean tolerance to low-phosphorus stress. *PLoS Genet.* 10:e1004061. doi: 10.1371/journal.pgen.1004061
- Zhang, J., Wang, X., Lu, Y., Bhusal, S. J., Song, Q., Cregan, P. B., et al. (2018). Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Mol. Plant* 11:460. doi: 10.1016/j.molp.2017.12.016
- Zhou, Y., Lu, D., Li, C., Luo, J., Zhu, B., Zhu, J., et al. (2012). Genetic Control of seed shattering in rice by the APETALA2 transcription factor shattering abortion1. *Plant Cell* 24, 1034–1048. doi: 10.1105/tpc.111.094383
- 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.* 34, 441–441.

**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 © 2019 Hu, Kan, Hu, Li, Hao, Li, Yang, Yang, He, Huang and Yu. 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.