Identification and Validation of Major QTLs, Epistatic Interactions, and Candidate Genes for Soybean Seed Shape and Weight Using Two Related RIL Populations

Mahmoud A. Elattar¹,², Benjamin Karikari¹†, Shuguang Li¹, Shiyu Song¹, Yongce Cao¹, Muhammed Aslam¹, Aiman Hina¹, Salah Fatouh Abou-Elwafa²* and Tuanjie Zhao¹*  

¹ National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean (Ministry of Agriculture), State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, ² Agronomy Department, Faculty of Agriculture, Minia University, Minia, Egypt, ³ Agronomy Department, Faculty of Agriculture, Assiut University, Assiut, Egypt

Understanding the genetic mechanism underlying seed size, shape, and weight is essential for enhancing soybean cultivars. High-density genetic maps of two recombinant inbred line (RIL) populations, LM6 and ZM6, were evaluated across multiple environments to identify and validate M-QTLs as well as identify candidate genes behind major and stable quantitative trait loci (QTLs). A total of 239 and 43 M-QTLs were mapped by composite interval mapping (CIM) and mixed-model-based composite interval mapping (MCIM) approaches, from which 180 and 18, respectively, are novel QTLs. Twenty-two QTLs including four novel major QTLs were validated in the two RIL populations across multiple environments. Moreover, 18 QTLs showed significant AE effects, and 40 pairwise of the identified QTLs exhibited digenic epistatic effects. Thirty-four QTLs associated with seed flatness index (FI) were identified and reported here for the first time. Seven QTL clusters comprising several QTLs for seed size, shape, and weight on genomic regions of chromosomes 3, 4, 5, 7, 9, 17, and 19 were identified. Gene annotations, gene ontology (GO) enrichment, and RNA-seq analyses of the genomic regions of those seven QTL clusters identified 47 candidate genes for seed-related traits. These genes are highly expressed in seed-related tissues and nodules, which might be deemed as potential candidate genes regulating the seed size, weight, and shape traits in soybean. This study provides detailed information on the genetic basis of the studied traits and candidate genes that could be efficiently implemented by soybean breeders for fine mapping and gene cloning, and for marker-assisted selection (MAS) targeted at improving these traits individually or concurrently.

Keywords: *Glycine max*, QTL mapping, QTL clusters, marker assisted breeding, epistatic interactions, candidate genes
INTRODUCTION

Soybean (*Glycine max* L. Merr.) is one of the most important food crops, being a rich source of dietary protein (69%) and providing over 50% edible oil globally, and has a significant role in health and biofuel (Hoeck et al., 2003). Besides improving soil fertility by integrating atmospheric nitrogen in the soil through a synergistic interaction with microorganisms, because of its high nutritional value, soybean is used in human food and animal feed (Wang et al., 2019). Throughout the last five decades, soybean production in China slightly increased. To meet domestic demands, China imports almost 80% of its requirements of soybean; therefore, improving soybean production is a major aim of soybean breeders to make the country self-sufficient (Liu et al., 2018). Most plant breeders are targeting yield-related traits to improve soybean production.

Seed size is an essential trait in flowering plants and plays a critical role in adaptation to the environment (Tao et al., 2017). However, these traits are complex quantitative traits regulated by polygenes and strongly influenced by environment and genotype × environment (G × E) interaction, and hence it is more difficult to select for based on phenotype (Yao et al., 2014). All soybean varieties developed in tropical and subtropical countries have small seed size compared to the temperate-region varieties. Besides, seed size, shape, and weight are important seed quality traits with significant influence on seed use (Basra, 1995; Teng et al., 2017; Wu et al., 2018). A positive correlation between seed size/weight and seed yield has been reported in several studies. Seed size/weight revealed a positive association with seed germination capability and vigor, thereby significantly affecting the competitive capability of the seedling for nutrient and water resources and light, hence enhancing stress tolerance (Edwards and Hartwig, 1971; Haig, 2013). Dissecting the genetic factors underlying seed size, shape, and weight and their relationship to the ambient environment is essential for improving soybean yield and quality-related traits. In addition, understanding the additive and additive × environment (AE) effects of quantitative trait loci (QTLs) and their contribution to the phenotypic variations would facilitate the application marker-assisted selection (MAS) because it will prominently lead the breeders in the QTL selection and expectation of the outcomes of MAS (Jannink et al., 2009).

A major aim of utilizing linkage mapping in plant breeding is to deepen our understanding of the inheritance and genetic architecture of quantitative traits and detect markers that can be employed as indirect selection tools in plant breeding (Bernardo, 2008; Abou-Elwafa, 2016a). In this regard, QTL mapping has been regularly used for detecting the QTL/gene underlying the quantitative traits such as seed size, shape, and weight in crop plants. As known, parental diversity and marker density greatly influence the accuracy and precision of QTL mapping. Besides, the population size used in most of the previously published reports for genetic mapping studies usually varied from 50 to 250 individuals; however, larger populations are needed for high-resolution mapping. A high-density genetic map facilitates the detection of closely linked markers associated with QTLs and provides an effective base for investigating quantitative traits (Mohan et al., 1997; Galal et al., 2014; Tewodros and Zelalem, 2016). The statistical difference between phenotypic data obtained from various environments could enhance the accuracy to detect QTL position (Zhao and Xu, 2012). Previous studies identified important seed size and shape QTLs, which were also associated with hundred seed weight (HSW); however, most of the studies used low-density genetic maps based on restriction fragment length polymorphism, simple-sequence repeat markers, and biochemical and morphological markers which have a large confidence interval with a low resolution of QTLs that are not suitable for candidate gene identification (Bernardo, 2008; Han et al., 2012; Abou-Elwafa, 2016a,b). Therefore, it is crucial to employ high-density genetic maps to detect more new recombination in a population, which will increase the accuracy of QTL mapping, candidate gene identification, and MAS (Mahmoud et al., 2018; Cao et al., 2019; Hina et al., 2020). Recent advances in genetic and genomic tools and approaches have facilitated the identification of QTLs associated with various agronomic traits in different crop species, including soybean, and the identification of candidate genes underlying these genomic regions (Peterson et al., 2012; Sun et al., 2012; Xie et al., 2018; Zhang F. et al., 2019; Hina et al., 2020; Kajiya-Kanega et al., 2020). Although epistatic interaction has a stronger effect on inbreeding depression, heterosis, adaptation, speciation, and reproductive isolation (Ma et al., 2015), previous studies focused mostly on identifying main-effect QTLs associated with seed sizes, shapes, and 100-seed weight in soybean. To date, at least 441, 52, and 297 QTLs for seed size, shape, and HSW have been reported1 based on various genetic contexts, advances in marker technology, statistical methods, and multiple environments. However, most of these QTLs are minor (R² < 10%), not stable, and with larger genomic regions/confidence intervals (Han et al., 2012; Hu et al., 2013; Kato et al., 2014). Recently, there have been limited studies on detecting QTLs with epistatic effects and their interactions with the environment (QEs) (Panthee et al., 2005; Zhang et al., 2011, 2018; Liang et al., 2016). Knowledge about the molecular mechanisms underlying soybean seed size, shape, and weight is still limited. So far, only two seed sizes/weight-related genes have been cloned and characterized from the soybean, i.e., the *Glyma20g25000* (In) gene that has a significant impact on seed size and number of seeds per pod (Jeong et al., 2012) and the *PP2C-I* gene that enhances seed size/weight (Lu et al., 2017). Therefore, it is essential to identify major and stable QTLs and candidate genes related to seed size, shape, and weight to improve our understanding of genetic mechanisms controlling these important traits in soybean (Kato et al., 2014; Zhang et al., 2018). The present study aimed to (i) map main-effect QTLs (M-QTLs), additive × additive (AA) QTLs, and QE for seed size, shape, and weight traits; (ii) employ two QTL mapping approaches to validate the identified QTLs in two mapping populations across multiple environments; (iii) analyze the epistatic QTL pairs and their interactions with the environment for further utilization of

---

1www.soybase.org
these QTLs in soybean genetic improvement; and (iv) mine potential candidate genes for the major ($R^2 > 10\%$) and stable (identified across multiple environments or populations) QTLs. We hypothesize that the results of this study would provide comprehensive knowledge on the genetic bases for these traits and mined candidate genes would serve as a foundation for functional validation and verification of some genes for seed size, shape, and weight in soybean. Besides, the results would be useful for the application of marker-assisted breeding (MAB) in soybean.

**MATERIALS AND METHODS**

**Plant Materials and Experiments**

Two recombinant inbred line (RIL) populations, i.e., ZM6 and LM6, comprising 126 and 104 lines, respectively, were used in the present study. The two populations were developed by single seed descent (SSD) from crosses between the genotypes Zhengyang (Z) and Linhefengdingou (L) as female parents and the M8206 (M6) genotype as the male parent. The two female parents, Z and L, have an average 100-seed weight of 17.1 and 35 g, respectively, whereas the male parent has an average 100-seed weight of 13.7 g.

The two RIL populations along with their parents were evaluated for seed size, shape and HSW across multiple environments. Experiments were conducted in the Jiangpu Experimental Station (33° 030’ N and 63° 118’ E), Nanjing, Jiangsu Province, in the 2012, 2013, 2014, and 2017 growing seasons (designated as 12JP, 13JP, 14JP, and 17JP, respectively), the Fengyang Experimental Station, Chuzhou, Anhui Province (32° 870’ N and 117° 560’ E), in the 2012 growing season (designated as 12FY), and the Yancheng Experimental Station, Yancheng, Jiangsu Province (33° 410’ N and 120° 200’ E), in 2014 (designated as 14YC). Plants were sown in June and harvest was done in October of the same year. Experiments were designed in a randomized complete blocks design (RCBD) with three replications. The experimental plot was one row of 2-m-long at 5-cm plant-to-plant distance and 50-cm row-to-row distance. Planting and post-planting operations were carried out following the recommended agronomic practices.

**Phenotypic Evaluation and Statistical Analysis**

Eight seed-related traits including seed length (SL), seed width (SW), seed thickness (ST), seed length/width (SLW), seed length/thickness (SLT), seed width/thickness (SWT), flatness index (FI), and 100-seed weight (HSW) were evaluated in LM6 and ZM6 populations under all environments. Phenotypic data were measured and recorded according to standard procedures (Tomooka et al., 2002; Cheng et al., 2006). In brief, seeds harvested from 10 guarded plants in the middle of each row were used for estimating SL, SW, ST, and HSW. The SL was measured as the longest dimension over the seed equivalent to the hilum. SW was measured as the longest dimension across the seed vertical to the hilum. ST was measured as the longest dimension from top to bottom of the seed. The SL, SW, and ST were estimated in millimeters (mm) using the Vernier caliper instrument, according to Kaushik et al. (2007) (Figure 1). Seed shape was identified by calculating three different ratios, i.e., SL/SW (SLW), SL/ST (SLT), and SW/ST (SWT), and FI. The ratios between the SL, SW, and ST were estimated from the individual values of the length, width, and thickness of the seeds according to Omokhafe and Alika (2004), while FI was calculated following the formula elaborated by Cailleux (1945) and Cerdà and García-Fayos (2002) to describe seed shape:

$$FI = \frac{(L+W)}{2T}$$

where $L$ is the SL, $W$ is the SW, and $T$ is the ST.

It extended from a value of 1 for the round seeds to more than 2 for skinny seeds. The HSW was expressed as an average of five measurements of 100 randomly selected seeds.

Descriptive statistics of the seed size, seed shape, and HSW traits were calculated using the SPSS software, version 24. Analysis of variance (ANOVA) for each environment and the combined overall environments (CE) was performed using the PROC GLM procedure in SAS software based on the random model (SAS Institute Inc. v. 9.02, 2010, Cary, NC, United States). Broad-sense heritability ($h^2$) in individual environments was estimated as:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

2http://www.spss.com

**FIGURE 1** | Measuring seed width (SW), length (SL), and thickness (ST).
where the $h^2$ was estimated as follows:

$$h^2 = \frac{\sigma^2_g}{\sigma^2_g + \sigma^2_e + \sigma^2_ge/nr}$$

where $\sigma^2_g$, $\sigma^2_e$, and $\sigma^2_ge$ are the variance components estimated from the ANOVA for the genotypic, error, and genotype × experiment variances, respectively, with $r$ as the number of replicates and $n$ as the number of environments. All the parameters were assessed from the expected mean squares in ANOVA. The Pearson correlation coefficient ($r$) between seed size, seed shape, and HSW traits was calculated from the mean data using the SAS PROC CORR with data obtained for CE (average across environments) for each population.

**Construction of Genetic Maps and QTL Analysis**

Genetic map information was obtained from the National Center for Soybean Improvement, Nanjing Agricultural University. High-density genetic maps of the ZM6 and LM6 populations comprise 2601 and 2267 bin markers by using the RAD-seq technique, respectively (Supplementary Table 1), which were constructed as previously reported (Karikari et al., 2019; Zhang X. et al., 2019). The total lengths of the ZM6 and LM6 genetic maps were 2630.22 and 2453.79 cM, with an average distance between the markers 1.01 and 1.08 cM, respectively (Supplementary Table 1). The average marker per chromosome was 130 and 113 for the ZM6 and LM6 maps, respectively, with average genetic distances per chromosome of 131.51 and 122.69 cM, respectively (Figure 2 and Supplementary Table 1).

**Mapping of Main- and Epistatic-Effect QTLs**

The WinQTLCart 2.5 software (Wang et al., 2006) was employed to identify the M-QTLs using the average values of seed size, seed shape, and 100-seed weight from the individual environments and overall environments with the composite interval mapping model (CIM) (Zeng, 1994). The software running features were 10 cM window size, 1 cM running speed, the logarithm of odds (LOD) (Morton, 1955) threshold which was computed using 1000 permutations because of an experiment-wide error proportion of $P < 0.05$ (Churchill and Doerge, 1994), and the confidence interval which was determined using a 1-LOD support interval, which was controlled by finding the local on the two sides of a QTL top that is compatible with a reduction of the 1 LOD score. QTLs detected within overlapping intervals in different environments were considered the same (Qi et al., 2017). To identify the genetic effects of the QTLs, i.e., additive QTLs, additive × additive (AA), AE, and AA × environment (AAE), the mixed-model-based composite interval mapping (MCIM) procedure was employed in the QTLNetwork V2.1 software (Yang et al., 2008). The critical F-value was calculated by a permutation test with 1000 permutations for MCIM. The effects of QTLs were assessed using the Markov Chain Monte Carlo (MCMC) approach. Epistatic effects, candidate interval selection, and putative QTL detection were estimated with an experiment-wide error proportion of $P < 0.05$ (Yang et al., 2007; Xing et al., 2012).

**In silico Identification of Candidate Genes**

QTLs identified in two or more environments with an $R^2 > 10\%$ were considered as stable and major QTLs (Qi et al., 2017). Genomic regions with several M-QTLs related to different studied traits were identified as a QTL cluster.
The Phytozome<sup>1</sup> and SoyBase (<sup>see</sup> text footnote 1) online platform repositories were employed to retrieve all model genes within the physical interval position of the QTL clusters. Potential candidate genes were identified based on gene annotations (<sup>see</sup> text footnotes 1 and 3) and the reported putative function of genes implicated in these traits. Gene ontology (GO) enrichment analysis was performed for the identified candidate genes within each QTL cluster region using AgriGO V2.0<sup>0</sup> (Tian et al., 2017). Gene classification was then carried out using the Web Gene Ontology (WeGO) Annotation Plotting tool, Version 2.0 (Ye et al., 2006). The publicly available RNA-Seq database on the SoyBase website was used to analyze the expression of the candidate genes in various soybean tissues and developmental stages. A heat map was used to visualize the fold-change patterns of these candidate genes was constructed using the TBtools_JRE 1.068 software (Chen C. et al., 2020).

**RESULTS**

Phenotypic Variations in RIL Populations in Multiple Environments

All measured (SL, ST, SW, and HSW) and calculated (SLW, SLT, SWT, and FI) phenotypic traits exhibited significant differences among the three parental lines across all environments as indicated by the ANOVA (<sup>Supplementary Tables 2</sup> and 3). ANOVA revealed that all studied traits were significantly (<p> < 0.001 or < 0.05) influenced by the environment, genotypes, and the genotype × environment interaction (<sup>Supplementary Tables 4, 5</sup>), indicating the differential response of the genotypes to the changes in environmental cues. The two populations showed continuous phenotypic variations in all studied traits, implying a polygenic inheritance of these traits (<sup>Figure 3</sup>). The differences in mean phenotypic values among the three parental lines for seed size, seed shape, and HSW traits were constantly high across all studied environments, and their multi-environment means for both populations (<sup>Figure 4</sup>). Compared to the male parent, the female parent of the LM6 population, Linhefenqingdou, exhibited an average increase of 27.80, 28.19, 31.10, and 41.37% in SL, ST, SW, and HSW, respectively. Meanwhile, in the ZM6 population the female parent Zhengyang surpassed the male parent M8206 by an average of 11.00, 9.66, 7.65, and 17.53% in SL, ST, SW, and HSW across all environments, respectively (<sup>Figure 4</sup>, <sup>Supplementary Tables 2, 3</sup>). In both populations, several lines overstep their parents in both directions in all studied traits across all environments, suggesting the occurrence of transgressive segregations within the two populations (<sup>Figures 3, 4</sup>). The broad-sense heritability (h<sup>2</sup>) under individual environments ranged from 66.75 to 98.08%, 64.39 to 95.72%, and 81.45 to 99.36% for seed size, HSW, and seed shape, respectively (<sup>Supplementary Tables 2, 3</sup>). Meanwhile, h<sup>2</sup> under combined environments (CE) ranged from 78.25 to 87.31%, 65.70 to 90.80%, and 92.95 to 95.72% for seed size, shape, and HSW, respectively. The correlation coefficient (r<sup>2</sup>) among SL, ST, and SW exhibited significant positive correlations with each other and with two of the seed shape traits (SLT and SLW) in both populations with r<sup>2</sup> values ranging from 0.79 to 0.91. Meanwhile, SL, ST, and SW exhibited significant negative correlations with the other two seed shape traits (SWT and FI) (<sup>Supplementary Table 6</sup>). Except for the correlation between SLW and SWT, all the seed shape traits showed significant positive correlations with each other in both populations with r<sup>2</sup> values ranging from 0.33 to 0.95. All seed size traits, i.e., SL, SW, and ST, showed significant positive correlations with HSW with r<sup>2</sup> values ranging from 0.29 to 0.70 in both populations.

**Validation of Identified QTLs Employing Two Mapping Approaches**

A total of 92, 99, and 48 M-QTLs associated with seed size, seed shape, and HSW, respectively, were mapped by the CIM approach (<sup>Supplementary Tables 7–9</sup>). Meanwhile, forty-three QTLs were identified for seed size, shape, and HSW by using MCIM approach (<sup>Tables 1, 2</sup>). Among these, 22 QTLs were identified and validated by both approaches within the same physical chromosomal position, indicating the dependability and stability of these QTLs. A comparison of the physical chromosomal regions of the QTLs detected by both approaches revealed the identification and validation of four QTLs, i.e., qSL-7-1<sub>LM6</sub>, qSW-19-2<sub>LM6</sub>, qFI-3-1<sub>LM6</sub>, and qHSW-3-2<sub>LM6</sub>, for the first time in the two populations (LM6 and ZM6) with an <r><sup>2</sup></r> > 10%. Therefore, we considered these QTLs as novel stable and major QTLs that could be used for map-based cloning, candidate gene identification, and QTL stacking into elite cultivars targeted at improving seed size, shape, and HSW in soybean.

**Identification of the Main Effects of the Stable Additive Seed Size QTLs**

A total of 92 M-QTLs were mapped for seed size-related traits, i.e., SL, SW, and ST, on all soybean chromosomes, except chromosomes 1 and 12, with LOD scores and <r><sup>2</sup></r> values ranging from 2.5 to 10.3 and 5.0 to 19.7%, respectively, in the two populations (<sup>Supplementary Table 7</sup> and <sup>Supplementary Figures 1a–c</sup>). Out of these, 30 M-QTLs for SL, 35 for SW, and 27 for ST with alleles underlying QTLs inherited from either of the parents. Seventy-two M-QTLs were mapped in one environment while the remaining 20 were mapped within overlapping regions in at least one environment with or without CE. Forty-seven QTLs exhibiting <r><sup>2</sup></r> > 10% were considered as major QTLs. The most prominent QTL was the qSW-17-2<sub>LM6</sub> (LOD = 6.70–10.29, and <r><sup>2</sup></r> = 16.60–18.30%), which was detected within the physical position 6,844,412–9,645,325 bp in 14JP and CE (<sup>Figure 5b</sup>). Likewise, the qSL-10-2<sub>ZM6.LM6</sub> (LOD = 6.08–6.89, and <r><sup>2</sup></r> = 15.4–17.1%) in ZM6 (17JP) and LM6...
Main Effects of the Stable Additive Seed Shape QTLs

In total, 99 M-QTLs related to seed shape traits (SLT, SLW, SWT, and FI) were mapped to 19 soybean chromosomes excluding chromosome 2 in both populations across four environments and the CE with LOD scores of 2.50–10.44 and $R^2$ of 5.12–31.56% by the CIM approach (Supplementary Table 8 and Supplementary Figures 1d–g). From the 99 M-QTLs, 22, 33, 11, and 22 were detected for SLT, SLW, SWT, and FI, respectively (Supplementary Table 8). Among them, 71 M-QTLs were detected in multiple environments, while 28 were mapped in at least one environment either with or without the CE. Eight M-QTLs for SLW ($qSLW-3-2_{LM6}$, $qSLW-5-3_{ZM6}$, $qSLW-9-3_{LM6}$, $qSLW-13-4_{ZM6}$, $qSLW-15-1_{LM6}$, $qSLW-15-3_{LM6}$, $qSLW-16-2_{ZM6}$, and $qSLW-16-3_{ZM6}$) were mapped in at least one environment with or without the CE. Similarly, seven M-QTLs for SLT ($qSLT-1-3_{LM6}$, $qSLT-5-3_{ZM6}$, $qSLT-11-1_{LM6}$, $qSLT-13-1_{LM6}$, $qSLT-14-1_{LM6}$, $qSLT-16-1_{LM6}$, and $qSLT-20-1_{ZM6}$) were mapped in at least one environment with or without the CE (Supplementary Table 8 and Supplementary Figure 1e). Likewise, four M-QTLs ($qSWT-8-1_{LM6}$, $qSWT-11-2_{ZM6}$, $qSWT-13-1_{LM6}$, and $qSWT-17-1_{ZM6}$) were mapped for SWT in at least one environment with or without the CE (Supplementary Table 8 and Supplementary Figure 1f). Also, a total of 10 M-QTLs ($qFI-1-1_{ZM6}$, $qFI-1-2_{LM6}$, $qFI-1-3_{LM6}$, $qFI-1-4_{ZM6}$, $qFI-3-1_{LM6}$, $qFI-3-3_{ZM6}$, $qFI-5-2_{ZM6}$, $qFI-11-1_{LM6}$, $qFI-11-1_{LM6}$, $qFI-11-1_{LM6}$) was located to the physical position between 41,454,163 and 43,944,243 bp.
**FIGURE 4** | Performance of the parents of the two RIL populations, i.e., LINHE and M8206 (A) and Zhengyang and M8206 (B) along with the two derived RIL populations, LM6 and ZM6, respectively for seed size and shape traits as well as 100-seed weight among multiple environments. 12JP, 13JP, 14JP, and 17JP indicate phenotyping at the Jiangpu Experimental Station in the 2012, 2013, 2014, and 2017 growing seasons, respectively. 12FY indicates the Fengyang Experimental Station, Chuzhou in the 2012 growing season. 14YC indicates the Yancheng Experimental Station in 2014. SL, seed length (mm); SW, seed width (mm); ST, seed thickness (mm); SLW, seed length to width ratio; SLT, seed length to thickness ratio; SWT, seed width to thickness ratio; FI, flatness index; HSW, 100-seed weight (g).

$q\text{SLT-1-2}_{ZM6}$, $q\text{SWT-1-1}_{ZM6}$, and $q\text{FI-1-2}_{ZM6}$, $q\text{SLW-16-2}_{ZM6}$, $q\text{SLT-16-1}_{LM6}$, and $q\text{SLW-16-3}_{LM6}$, respectively (Supplementary Figure 2 and Supplementary Table 8). The $q\text{SLT-5-3}_{LM6}$, $q\text{SLW-5-3}_{LM6}$, and $q\text{FI-5-1}_{LM6}$, $q\text{SLT-11-1}_{LM6}$, and $q\text{FI-11-1}_{LM6}$, respectively (Supplementary Figure 2 and Supplementary Table 8)

### Main Effects of the Stable Additive Seed Weight QTLs

A total of 48 M-QTLs for HSW were detected, from which 27 were detected in a specific environment and 21 were mapped in at least one environment with or without the CE (Supplementary Table 9 and Supplementary Figure 1h). The LOD scores and $R^2$ values of these M-QTLs ranged from 2.51 to 10.61 and 4.8 to 24.5%, respectively. The highest number of M-QTLs (six QTLs) was mapped on Chr 04 followed by Chr10 with five M-QTLs, and the lowest number of one M-QTL was mapped...


| QTL1 | Marker interval | Physical position (bp) | Position (cM) | Additive – effect (A)3 | Additive x environment effect (AE)4 | References |
|------|-----------------|------------------------|---------------|------------------------|---------------------------------------|------------|
| qSL-7-1_{LM6} | bin744-bin745 | 3,324,836–3,459,470 | 15.25 | 0.16** | 17.45 | NS | NS | NS | NS | 0.07 Hu et al., 2013 |
| qSL-13-9_{LM6} | bin1536-bin1536 | 43,244,220–44,026,819 | 140.15 | 0.13** | 6.08 | NS | NS | NS | NS | 0.28 Salas et al., 2006 |
| qSW-19-2_{LM6} | bin2100-bin2101 | 34,493,194–34,882,495 | 42.57 | 0.18** | 19.47 | −0.13** | −0.11* | 0.06* | 0.09* | 0.18 New |
| qST-18-4_{LM6} | bin1979-bin1980 | 9,222,099–10,402,370 | 43.69 | 0.06** | 11.34 | NS | NS | NS | NS | 0.07 Fang et al., 2017 |
| qST-20-1_{LM6} | bin2461-bin2462 | 41,009,636–41,399,912 | 130.75 | 0.05** | 3.78 | NS | NS | NS | NS | 0.24 New |
| qFI-17-6_{LM6} | bin2177-bin2178 | 41,009,636–41,399,912 | 130.75 | 0.05** | 3.78 | NS | NS | NS | NS | 0.24 New |
| qFI-20-1_{LM6} | bin2461-bin2462 | 1–6,62,752 | 0.08** | 5.45 | NS | NS | NS | NS | 0.28 Li et al., 2013 |

**p < 0.05; ***p < 0.01; NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parent Zhengyang and Linhefengqindou, while those with negative values show beneficial alleles from parent Meng 8206.
to Chr 02, 09, 17, and 18. The most prominent M-QTLs were \textit{qHSW-14-2}_{LM6}, \textit{qHSW-10-3}_{LM6}, and \textit{qHSW-10-4}_{LM6} with LOD scores and $R^2$ values of 10.61 and 24.50% (Figure 5), 7.57 and 17.60%, and 7.20 and 16.90%, respectively. Among those 21 M-QTLs, \textit{qHSW-4-3}_{LM6}, \textit{ZM6}, \textit{qHSW-6-2}_{LM6}, \textit{qHSW-10-1}_{LM6}, \textit{qHSW-13-1}_{ZM6}, \textit{qHSW-15-2}_{LM6}, and \textit{qHSW-15-4}_{LM6} were mapped in at least three environments with an average $R^2$ of 13.01%.

**Analysis of Additive-Effect QTLs and Additive x Environment QTL Interactions**

The mixed-MCIM approach implemented in the QTL Network V2.1 software for both RIL populations across multiple environments identified 35 AA QTLs on 17 chromosomes related to seven seed size and seed shape traits. These comprise 9, 3, 7, 3, 4, 1, and 8 A QTLs associated with SL, SW, ST, SLW, SLT, SWT, and FI, respectively, in the LM6 and ZM6 populations across all environments (Table 1). The contributed allele of 11 QTLs of them which reduces seed size and seed shape values through significant additive effects is inherited from the M8206 parent. Meanwhile, the contributed allele of the remaining 24 QTLs, which enhances seed size and shape values through significant additive effects, is inherited from either Zhengyang or Linhefenqingdong parent of the ZM6 or LM6 population, respectively (Table 1). Thirteen out of 35 QTLs revealed significant AE effects in at least one environment. However, five QTLs, i.e., \textit{qSW-13-5}_{LM6}, \textit{qSW-19-2}_{LM6}, \textit{qFI-8-6}_{LM6}, \textit{qSL-10-2}_{LM6}, and \textit{qST-13-5}_{LM6}, showed significant or highly significant AE among all studied environments (Table 1). The influence of AE effects on seed size and seed shape values was environmentally dependent (Table 1). Eight AA QTLs associated with HSW were identified on six chromosomes, i.e., Chr 03, 08, 09, 13, 14, and 16, in LM6 and ZM6 populations across six environments (Table 2). Six of those eight QTLs displayed a positive additive effect with the beneficial allele that could increase HSW which is inherited from the female parents (Linhefenqingdong or Zhengyang). Meanwhile, the remaining two QTLs, i.e., \textit{qHSW-13-3}_{LM6} and \textit{qHSW-14-2}_{LM6}, revealed negative additive effects with the allele that reduces HSW which is inherited from the common male parent (Meng8206) (Table 2). Two QTLs associated with HSW, i.e., \textit{qHSW-14-3}_{LM6} and \textit{qHSW-8-3}_{LM6}, displayed significant AE effects in two environments, whereas \textit{qHSW-13-3}_{LM6} showed a significant AE only in one environment (the 13JP environment). The \textit{qHSW-14-4}_{LM6} QTL revealed significant AE effects across three different environments, i.e., 12FY, 12JP, and 17JP (Table 2).

**Validation of QTLs and Identification of QTL Clusters**

A chromosomal region comprising several identified M-QTLs for different studied seed phenotypic traits was designated as a QTL cluster. Accordingly, 24 QTL clusters on 17 chromosomes with the exception to Chr 02, 12, and 18 were identified (Supplementary Table 10 and Supplementary Figure 2). Among the identified 24 clusters, seven clusters harbored QTLs related
to seed size, seed shape, and HSW, five clusters harbored QTLs related only to seed size and seed shape traits, nine clusters comprised QTLs related to seed size and HSW traits, and three clusters harbored QTLs for only seed shape traits (Supplementary Table 10). QTLs within 15 clusters revealed positive additive effects with the beneficial alleles which are inherited from the big seed size and heavy seed weight parents (Zhengyang or Linhafenqindou). Seven out of 24
clusters contain QTLs that have been detected and validated in the low RIL populations (Supplementary Table 10). The most prominent M-QTL (qFI-1-3LM6) with a LOD score of 3.71–10.44 and $R^2$ (10.45–31.50%) was located to Cluster-01. Each cluster comprised a different number of QTLs, with the highest number of QTLs, i.e., seven, associated with seed size, shape, and HSW traits which were in cluster-03 at the physical position of 1,509,548–6,780,840 bp allocated as two QTLs related to seed size (qSL-3-1LM6 and qSL-3-2LM6), four QTLs for seed shape (qSLW-3-2LM6, qFI-3-1LM6, qSLT-3-1LM6, and qFI-3-2LM6), and one QTL HSW (qHSW-3-1LM6). Except for qHSW-3-1LM6, all QTLs in this cluster were major QTLs (with an $R^2$ > 10%). Each of the clusters: 13, 16.2, and 17.1 comprises five to six M-QTLs related to seed size and HSW traits identified in one of the two RIL populations and displayed $R^2$ values of 8.85–13.43, 5.96–11.26, and 6.8–18.30%, respectively (Supplementary Figure 2 and Supplementary Table 10). Another rich region of QTLs was cluster-20 on chr20 that harbors five seed size and shape M-QTLs, i.e., qFI-20-1LM6, qSLT-20-1LM6, qSLW-20-1LM6, qST-20-1LM6, and qSW-20-1LM6, from which three are major QTLs with $R^2$ of 11.2–19.2% within a physical region of 1.2 Mb (Supplementary Table 10). Cluster-09 comprises five stable (identified in the two populations) seed size, shape, and HSW QTLs with $R^2$ values ranging from 12.5 to 16.3%. Cluster-14.1 comprises four major M-QTLs in both populations with $R^2$ values ranging from 10.4 to 18.4% within the chromosomal region between 5,834,015 and 9,844,637 bp, one from which qSW-14-2LM6 is associated with seed size traits, whereas the other three (qSLW-14-1LM6, qFI-14-1LM6, and qSLT-14-1LM6) were associated with seed shape traits (Supplementary Table 10). Six clusters harboring four M-QTLs each were identified, from which four clusters, i.e., cluster-07, cluster-19.1, cluster-08, and cluster-14.2, comprise QTLs associated with HSW, seed size, and shape trait cluster-10.2 that comprises M-QTLs for seed size and shape traits, and cluster-16.1 that comprises only M-QTLs related to seed shape traits (Supplementary Table 10). The remaining nine clusters have three QTLs each; out of them, cluster-01 and cluster-17.2 comprise major QTLs related only to seed shape traits. Meanwhile, cluster-04.1 and cluster-19.2 contain minor M-QTLs associated with SW, SL, and HSW. Another two clusters comprise M-QTLs for both seed size and shape traits (Supplementary Table 10). The other three M-QTL clusters, i.e., cluster-10.1, cluster-11, and cluster-15, comprise both major and minor QTLs for seed size traits and HSW. Cluster-04.2 comprises the two QTLs qHSW-4-3LM6.ZM6 and qSL-4-1LM6 with $R^2$ values of 13.1–17.7%.

Analysis of Epistatic-Effect QTLs and Epistatic × Environment QTL Interactions

Analysis of the seed size and shape trait data under all environments identified 38 pairwise epistatic effects (AA) QTLs, from which 2, 13, 6, 2, 3, 5, and 7 pairs were related to SL, SW, ST, SLW, SLT, SWT, and FI traits, respectively, with $R^2$ values ranging 0.51–11.35% (Table 3). All QTL pairs displayed a high significant AA effect. Further analyses revealed that 20 AA QTLs showed significant or highly significant pairwise additive–additive–environment (AAE) interaction effects in at least one environment with $R^2$ values ranging from 0.13 to 5.31% (Table 3). Ten pairs showed significant AAE in two environments, i.e., 12FY (AAE1) and 12JP (AAE2), while three pairs displayed significant AAE in 12JP (AAE2) and 14JP (AAE3) environments (Table 3). This shows the effect of the environment on gene expression on phenotype development through epistatic effects. Out of the 38 QTLs, 16 pairwise interactions exhibited negative epistatic effects (AA) that decrease the values of seed size and shape traits, whereas 22 pairwise interactions exhibited positive epistatic effects (AA) that increase the values of seed size and shape traits (Table 3). The pairwise interaction between qFI-1-1LM6 and qFI-7-3LM6 revealed the strongest positive epistatic effect (0.65), whereas the pairwise qSL-6-1LM6 and qSLW-9-1LM6 revealed the weakest positive epistatic effect (0.02). Conversely, qSWT-3-1LM6 and qSWT-13-1LM6 resulted in the strongest negative epistatic effect (−0.71), whereas the pairwise qSL-2-6ZM6 and qSLW-18-3ZM6 pairwise resulted in the weakest negative epistatic effect (−0.02) (Table 3). Two digenic positive pairwise epistatic QTLs for HSW with highly significant additive × additive (AA) effects were identified on four chromosomes (Table 4). The first pairwise is composed of two QTLs, qHSW-11-1LM6 and qHSW-20-1LM6, with an $R^2$ of 3.46%, whereas the second pairwise comprises the two QTLs qHSW-9-1LM6 and qHSW-16-3ZM6 with an $R^2$ of 1.38%. However, the two pairs did not show any significant AAE interaction effects across all six environments (Table 4).

Candidate Gene Mining of the Main-Effect QTLS

The 24 M-QTL clusters were filtered based on the richness in QTLs associated with all or some of the seed size, shape, and HSW traits. As a result, seven QTL clusters, i.e., cluster-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, were used to identify candidate genes based on publicly available databases such as SoyBase and Phytozome and published papers. According to the physical intervals of the seven QTL clusters, 242, 190, 444, 367, 437, 523, and 116 genes were identified within cluster-03, 04.1, 05.1, 07, 09, 17.1, and 19, respectively, which were retrieved from the SoyBase database (see text footnote 1; Supplementary Table 11). GO enrichment analyses via AgriGO V2.0 (see text footnote 4) (Tian et al., 2017) were used to classify the model genes in each cluster. The classification was based on molecular function, biological process, and cellular components visualized on the Web-based GO (WeGO) V2.0 https://wego.genomics.cn (Ye et al., 2006). In all seven QTL clusters, high percentages of genes were related to catalytic activity, cell part, cell, cellular process, binding, and metabolic process terms, besides the response to stimulus in cluster-03 (Figure 6). These indicate the essential roles of these terms in the seed size, shape, and seed weight development in soybean. Probable candidate genes underlying these QTL clusters responsible for seed size, shape, and HSW in soybean were further predicted based on gene annotations, GO enrichment
### TABLE 3 | Estimated epistatic effects (AA) and environmental (AAE) interaction of QTLs for soybean seed size traits (SL, SW, and ST) and seed shape (SLW, SLT, SWT, and FI) traits across all environments.

| Trait | RIL | QTL_i | Chr_i | Interval_i | Pos_i | QTL_j | Chr_j | Interval_j | Pos_j | Epistasis – effect (AA) | Epistasis × environment effect (AAE) |
|-------|-----|--------|-------|------------|-------|--------|-------|------------|-------|------------------------|---------------------------------|
|       |     |        |       |            |       |        |       |            |       | H² | AA | AAE1 | AAE2 | AAE3 | AAE4 | H² % |
| ST    | qST-3-1 | 2 | bin237-bin238 | qST-3-1 | 3 | bin344-bin345 | 114.06 | 0.05** | 0.51 |
|       | qST-6-3 | 6 | bin674-bin678 | qST-6-3 | 6 | bin625-bin626 | 82.38 | -0.04** | 5.40 |
|       | qST-7-2 | 7 | bin749-bin750 | qST-7-2 | 7 | bin783-bin784 | 53.32 | 0.03** | 5.74 |
|       | qST-7-3 | 7 | bin783-bin784 | qST-7-3 | 7 | bin783-bin784 | 53.32 | 0.03** | 5.74 |
|       | qST-16-1 | 16 | bin1744-bin1745 | qST-16-1 | 16 | bin1744-bin1745 | 72.28 | 0.04** | 11.35 |
|       | qSLW-6-2 | 6 | bin625-bin626 | qSLW-6-2 | 6 | bin625-bin626 | 82.38 | -0.04** | 5.40 |
|       | qSLW-9-1 | 9 | bin1084-bin1085 | qSLW-9-1 | 9 | bin1084-bin1085 | 186.90 | 0.3** | 8.61 |
|       | qST-10-5 | 10 | bin1153-bin1154 | qST-10-5 | 10 | bin1153-bin1154 | 97.99 | 0.03** | 2.56 |
|       | qSL-12-4 | 12 | bin1553-bin1554 | qSL-12-4 | 12 | bin1553-bin1554 | 97.99 | 0.03** | 2.56 |
|       | qSL-15-5 | 15 | bin1919-bin1920 | qSL-15-5 | 15 | bin1919-bin1920 | 85.99 | -0.06** | 1.88 |
|       | qSL-2-3 | 2 | bin214-bin215 | qSL-2-3 | 2 | bin214-bin215 | 96.57 | 0.02** | 8.61 |
|       | qSL-2-4 | 2 | bin214-bin215 | qSL-2-4 | 2 | bin214-bin215 | 96.57 | 0.02** | 8.61 |
|       | qSL-3-2 | 3 | bin258-bin259 | qSL-3-2 | 3 | bin258-bin259 | 97.79 | -0.12** | 6.68 |
|       | qST-10-6 | 10 | bin1334-bin1335 | qST-10-6 | 10 | bin1334-bin1335 | 105.85 | 0.05** | 0.05 |
|       | qSLW-6-3 | 6 | bin684-bin685 | qSLW-6-3 | 6 | bin684-bin685 | 86.36 | 0.05** | 0.05 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSL-14-3 | 14 | bin1627-bin1628 | qSL-14-3 | 14 | bin1627-bin1628 | 85.96 | 0.02** | 1.88 |
|       | qST-10-6 | 10 | bin1334-bin1335 | qST-10-6 | 10 | bin1334-bin1335 | 105.85 | 0.05** | 0.05 |
|       | qSLW-6-3 | 6 | bin684-bin685 | qSLW-6-3 | 6 | bin684-bin685 | 86.36 | 0.05** | 0.05 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSL-14-3 | 14 | bin1627-bin1628 | qSL-14-3 | 14 | bin1627-bin1628 | 85.96 | 0.02** | 1.88 |
|       | qST-10-6 | 10 | bin1334-bin1335 | qST-10-6 | 10 | bin1334-bin1335 | 105.85 | 0.05** | 0.05 |
|       | qSLW-6-3 | 6 | bin684-bin685 | qSLW-6-3 | 6 | bin684-bin685 | 86.36 | 0.05** | 0.05 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |
|       | qSLW-8-2 | 8 | bin206-bin207 | qSLW-8-2 | 8 | bin206-bin207 | 158.24 | 0.05** | 2.26 |

Chr_j and Chr_j indicate the two sites involved in epistatic interactions; Pos indicates genetic position for each of the sites, *p < 0.05; **p < 0.01; NS, non-significant. AA indicates epistatic effects between two QTLs; those with positive values showing two-locus genotypes being the same as those in parent Linhengfengqingdou and Zhengyang (or Meng 8206) have the beneficial effects, while the two-locus recombinants take the negative effects. The case of negative values is the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; and AE4, JP2017.
TABLE 4 | Estimated epistatic effects (AA) and environmental (AAE) interaction of QTLs for soybean 100-seed weight across all environments.

| QTL_i | Chr_i | Interval_i | Pos_i | Physical position (bp)_i | QTL_j | Chr_j | Interval_j | Pos_j | Physical position (bp)_j |
|-------|-------|------------|-------|--------------------------|-------|-------|------------|-------|--------------------------|
|       |       |            |       | AA                       |       |       |            |       | AAE                       |
|       |       |            |       | AAH²                   |       |       |            |       | AAE1 0.94**         |
|       |       |            |       | AAH²                   |       |       |            |       | AAE2 0.93**         |
|       |       |            |       | AAH²                   |       |       |            |       | AAE3 0.93**         |
|       |       |            |       | AAH²                   |       |       |            |       | AAE4 NS             |
|       |       |            |       | AAH²                   |       |       |            |       | AAE5 NS             |
|       |       |            |       | AAH²                   |       |       |            |       | AAE6 NS             |

Chr_i and Chr_j indicate the two sites involved in epistatic interactions; Pos indicates genetic position for each of the sites. *p < 0.05; **p < 0.01; NS, non-significant. AA indicates epistatic effects between two QTLs; those with positive values showing two-locus genotypes being the same as those in parent Linhefenqingdou and Zhengyang (or Meng 8206) have the beneficial effects, while the two-locus recombinants take the negative effects. The case of negative values is the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, YC2014; AE4, JZ2015; AE5, JP2016.

DISCUSSION

The present study has implemented high-density genetic maps constructed from two-related RIL populations LM6 and ZM6 comprising 2,267 and 2,601 bin markers, respectively (Li et al., 2017), to validate QTLs associated with seed size, shape, and weight. To minimize the environmental errors, the two RIL populations were evaluated in four environments. The transgressive segregation and continuous variations observed in the two populations in all studied phenotypic traits facilitate the identification of a high number of both major and minor effect QTLs including some novel QTLs associated with all studied traits (Teng et al., 2009; Xu et al., 2011; Zhang et al., 2018). All measured and calculated traits in both populations were significantly (P < 0.01) influenced by genotype (G), environment (E), and their interactions (G × E), suggesting that the seed size, shape, and weight traits are not only governed by both genetic and environment; however, there was an effect of the G × E interaction as well (Sun et al., 2012; Hu et al., 2013; Liang et al., 2016). This explains the observed high \( h^2 \) (99.04%) and accordingly deduces that these traits are amenable to manipulation by selection without the help of molecular markers. Except for SL, SW, and ST that exhibited a highly significant correlation between each other and with HSW, our data showed that seed size, shape, and weight traits are not correlated, which is favorable when breeding for a round type with smaller or bigger seed size (Cober et al., 1997; Salas et al., 2006).

For validation of identified QTLs, a comparative QTL analysis using the CIM QTL mapping approach with the SoyBase database identified 69, 82, and 29 novel QTLs for seed size, shape, and HSW, respectively, indicating the distinct genetic architecture analysis, and the previously known putative biological function of the gene. Based on these, 19, 12, 26, 18, 22, 30, and 16 candidate genes were identified within the QTL clusters-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, respectively (Supplementary Table 12). These genes may function directly or indirectly in regulating seed development in soybean, which regulates seed size, shape, and HSW. These genes are involved in response to brassinosteroid stimulus, regulation of cell proliferation and differentiation, regulation of transcription, secondary metabolism and signaling, storage of proteins and lipids, hormone-mediated signaling pathway, regulation of the cell cycle process, transport, ubiquitin-dependent protein catabolic process, embryonic pattern specification, and response to auxin stimulus (Table 5). However, the RNS-seq data of genes in the soybean genome (Severin et al., 2010) that is publicly available on SoyBase was used to heatmap the expression of those candidate genes in the young leaf, flower, pod, seed, root, and nodule (Figure 7 and Supplementary Table 13). From the heatmaps, 47 genes out of the identified 143 candidate genes are highly expressed during seed developmental stages and in seed-related tissues (Figure 7 and Supplementary Table 13); hence, they could be potential seed size, shape, and HSW regulatory genes.
of the LM6 and ZM6 populations. These novel QTLs together explain over 88.00% of phenotypic variance for seed size, shape, and weight, signifying their potential value for improving soybean cultivars. Besides, the identification of novel QTLs in the present study suggests that more germplasms are required for unraveling the complex genetic basis for seed size and shape traits in soybean. Among these novel QTLs, eight novel major QTLs associated with HSW where their physical intervals did not overlap with any of the previously reported HSW QTLs, suggesting them as potential loci for HSW and major QTLs for...

FIGURE 6 | WeGO analysis of the genes located within the seven major QTL clusters: (A) Cluster-03; (B) Cluster-4.1; (C) Cluster-5.1; (D) Cluster-07; (E) Cluster-09; (F) Cluster-17.1; and (G) Cluster-19.1.
future fine mapping to delimit the physical interval. Numerous QTLs associated with SW, HSW, SLW, and SLT identified in this study are co-localized with previously reported corresponding QTLs (Salas et al., 2006; Li et al., 2010; Moongkanna et al., 2011; Hu et al., 2013; Jun et al., 2014; Fang et al., 2017; Hacisalihoglu et al., 2018; Hina et al., 2020). Our study identified for the first time 13 major QTLs ($R^2 > 10\%$) related to FI; thus, we considered them as novel QTLs. Besides, Chr01 and Chr03 harbored four and three FI QTLs, suggesting crucial roles of Chr 01 and 03 in controlling the inheritance of seed FI in soybean. The

### TABLE 5 | Candidate genes identified within the seven QTL clusters that are highly expressed in soybean seed.

| QTL clusters | Gene               | Start   | Stop    | Gene functional annotation                                      |
|--------------|--------------------|---------|---------|-----------------------------------------------------------------|
| Cluster-03   | Glyma03g01880      | 1668601 | 1674475 | Seed dormancy process; protein ubiquitination; lipid storage    |
|              | Glyma03g03210      | 3001933 | 3005806 | Pollen development; embryo sac egg cell differentiation; DNA-dependent |
|              | Glyma03g03760      | 3581308 | 3584468 | Maintenance of shoot apical meristem identity; cell differentiation |
|              | Glyma03g04330      | 3581308 | 3584468 | Embryo development; regulation of seed maturation               |
|              | Glyma03g04620      | 4798029 | 4801122 | Regulation of meristem growth; protein deubiquitination          |
| Cluster-04.1 | Glyma04g02970      | 2146489 | 2152500 | Embryo sac egg cell differentiation                               |
|              | Glyma04g03210      | 2347024 | 2349849 | Fatty acid beta-oxidation; response to auxin stimulus; ovule development |
|              | Glyma04g03810      | 2630227 | 2632308 | Brassinosteroid mediated signaling pathway; seed development; ovule development |
|              | Glyma04g04460      | 3305860 | 3308715 | Response to cytokinin stimulus; response to brassinosteroid stimulus; seed development |
|              | Glyma04g04540      | 3395831 | 3397238 | Response to ethylene stimulus; seed dormancy process; floral organ morphogenesis |
|              | Glyma04g04870      | 3628743 | 3634478 | Embryo development ending in seed dormancy                       |
| Cluster-05.1 | Glyma05g28950      | 34669156| 3467593 | Nucleotide biosynthetic process; embryo development ending in seed dormancy |
|              | Glyma05g29700      | 3522628 | 3524029 | Brassinosteroid biosynthetic process; starch biosynthetic process |
|              | Glyma05g30380      | 35754306| 3575603 | Embryo development; protein ubiquitination; lipid storage; anther development |
|              | Glyma05g31450      | 36578952| 36583516| Post-embryonic development                                       |
|              | Glyma05g31490      | 36611301| 36615160| Embryo development ending in seed dormancy                       |
|              | Glyma05g31830      | 36870586| 36873840| Ubiquitin-dependent protein catabolic process                    |
|              | Glyma05g32030      | 37026301| 37031440| Ubiquitin-dependent protein catabolic process; multicellular organismal development |
|              | Glyma05g33790      | 38337126| 38341140| Phosphatidylcholine biosynthetic process; metabolic process      |
|              | Glyma05g34070      | 38511154| 38513219| Cellular response to abscisic acid stimulus                     |
| Cluster-07   | Glyma07g13230      | 11764552| 11784123| Embryo sac egg cell differentiation; protein ubiquitination; lipid storage |
|              | Glyma07g13730      | 12749034| 12753558| Embryo development; positive regulation of gene expression       |
|              | Glyma07g14460      | 13900307| 13906228| Embryo development ending in seed dormancy                       |
|              | Glyma07g15050      | 14900705| 14909235| Seed dormancy process; regulation of cell cycle process          |
|              | Glyma07g15640      | 1537898 | 15384642| Response to hormone stimulus and auxin stimulus; response to brassinosteroid stimulus |
|              | Glyma07g15840      | 15528948| 15544150| Ubiquitin-dependent protein catabolic process; regulation of lipid catabolic process |
| Cluster-09   | Glyma09g28640      | 3557335 | 35579018| Embryo development ending in seed dormancy; cellular response to abscisic acid stimulus |
|              | Glyma09g29030      | 35969729| 35993075| Ubiquitin-dependent protein catabolic process; fatty acid beta-oxidation |
|              | Glyma09g29720      | 36540972| 36548174| Response to auxin stimulus; auxin metabolic process              |
|              | Glyma09g30130      | 37014420| 37022621| Protein import into nucleus; embryo sac egg cell differentiation |
|              | Glyma09g30650      | 37426876| 37433118| Phosphatidylcholine biosynthetic process; metabolic process; pollen development |
|              | Glyma09g31620      | 38298193| 38307446| Response to abscisic acid stimulus; embryo development           |
|              | Glyma09g32600      | 39104082| 39107332| Translational elongation; embryo development ending in seed dormancy |
|              | Glyma09g32680      | 39173955| 39183935| Regulation of protein phosphorylation                            |
|              | Glyma09g33630      | 40063507| 40067999| Response to auxin stimulus; seed dormancy process               |
| Cluster-17.1 | Glyma17g09320      | 6889969 | 6894069 | Seed maturation; histone deacetylation; response to abscisic acid stimulus |
|              | Glyma17g09690      | 7171761 | 7180015 | Seed maturation; protein ubiquitination; lipid storage          |
|              | Glyma17g10290      | 7077775 | 711360 | Pollen tube growth; seed dormancy process; ovule development    |
|              | Glyma17g10380      | 7768561 | 7778131 | Ubiquitin-dependent protein catabolic process                    |
|              | Glyma17g10990      | 8262700 | 8267178 | Carbohydrate metabolic process                                  |
|              | Glyma17g11410      | 8557013 | 8563158 | Regulation of embryo sac egg cell differentiation               |
|              | Glyma17g12950      | 9873806 | 9891206 | Protein folding; embryo development response to starvation      |
|              | Glyma17g15490      | 12218497| 12226562| Ubiquitin-dependent protein catabolic process                    |
|              | Glyma17g15550      | 12302621| 12306143| N-terminal protein myristoylation; pollen development; pollen tube growth |
| Cluster-19.1 | Glyma19g32990      | 4066918 | 40669847| Glucose catabolic process; response to auxin stimulus           |
|              | Glyma19g33620      | 41194146| 41196743| Maltese metabolic process; starch biosynthetic process; glucosinolate biosynthetic process |
|              | Glyma19g33650      | 41237306| 41242657| Abscisic acid biosynthetic process; plant-type cell wall modification; pollen tube growth |
positive alleles for seed size, shape, and HSW traits were inherited from both parents of the two RIL populations. Therefore, it is likely that not only the higher seed size and heavyweight parent (Linhefenqingdou or Zhengyang) contributed favorable alleles but also the lighter seed weight parent (M8206) might play a role (Cao et al., 2019; Hina et al., 2020).

Mapping of QTLs associated with seed size, shape, and weight-related traits using the MCIM approach was performed to (i) dissect the additive effect QTLs and Q × E interactions, which is essential for selecting the most compatible varieties adapted to particular environments, and (ii) further validate the QTLs identified by the CIM approach. The MCIM approach identified 18 QTLs for seed sizes, shapes, and weight traits that are co-localized in the same physical interval of the CIM-mapped QTLs. Therefore, these QTLs could also be stable QTLs for further fine mapping and map-based cloning to uncover the genetic control and mechanisms of seed size, shape, and weight traits in soybean, and molecular markers tightly linked to these QTLs could be used for MAS.

Dissecting the epistatic and QTL × environment effects are crucial for understanding the genetic mechanisms that contributed to the phenotypic variations of complex traits (Kaushik et al., 2007). Disregarding intergenic interactions will lead to the overestimation of individual QTL effects, and the underestimation of genetic variance resulting in a large drop in the genetic response to MAS especially in late generations (Nyquist and Baker, 1991; Zhang et al., 2004). The identified 40 pairwise digenic epistatic QTLs for seed size, shape, and weight-related traits in the present study could be considered as modifying genes that do not exhibit only additive effects but could affect the expression of seed size, shape, and weight-related genes through epistatic interactions. Similar results for the epistatic interaction of seed size, shape, and weight QTLs have been also previously reported by Xin et al. (2016) and Zhang et al. (2018). The appearance of epistatic interactions for a specific trait makes selection difficult. Noteworthily, all main-effect QTLs detected in our study had no epistatic effect, which raises the heritability of the trait guiding to easier selection.

Genomic regions were identified as QTL clusters based on the presence of several QTLs related to all or some of the seed size, shape, and HSW traits. Accordingly, 24 QTL clusters were identified on 17 chromosomes each containing three or more QTLs related to seed size, shape, and HSW traits. These QTL clusters have not been previously reported, which enhances the developing knowledge of the genetic control of these traits. The co-localization of QTLs for seed size, shape, and HSW and how they have exceptionally corresponded support the highly significant correlation with each other (Cai and Morishima, 2002) (Supplementary Table 10). Besides, the occurrence of the QTL clustering could signify a linkage of...
QTLs/genes or outcome from the multiple effects of one QTL in the same genomic region (Wang et al., 2006; Cao et al., 2017; Liu et al., 2017). The QTL clusters reveal that the QTL linkage/gathering could make the enhancement of seed size and shape easier than single QTLs (Hina et al., 2020). Significant positive correlations of soybean seed protein and oil contents and seed yield with seed size and seed shape have been shown; therefore, these traits are directly associated with seed size and shape in soybean (Qi et al., 2011; Hacisalihoglu and Settles, 2017; Wu et al., 2018). This notion would explain the co-localization of QTLs associated with seed protein and oil contents in the genomic regions of several QTL (Panthee et al., 2005; Salas et al., 2006; Vieira et al., 2006; Moongkanna et al., 2011; Yang et al., 2011). The position of the first flower and the number of days to flowering have large effects on seed number per plant in soybean (Tasma et al., 2001; Yamanaka et al., 2001; Khan et al., 2008), which affects seed size and HSW indicating the existence of common genetic factors for these traits. QTLs associated with the position of the first flower identified previously (Tasma et al., 2001; Han et al., 2012) are located to the genomic region of clusters 16.1, 19.2, and 20 (Hyten et al., 2004). The extensive analysis of QTL clusters in our study suggests that breeding programs aiming to improve seed size, shape, and weight with enhanced quality should focus on QTL clustering and select QTLs within these regions. Besides, the existence of QTL clusters provides evidence that some traits-related genes are more densely concentrated in specific genomic regions of crop genomes than others (Fang et al., 2017).

Identification of candidate genes underlying QTL regions is of great interest for breeding programs (Abou-Elwafa, 2018; Abou-Elwafa and Shehzad, 2018). A bioinformatics pipeline implementing genomic sequences of identified QTL clusters was employed to identify candidate genes. The pipeline comprises three complementary steps, i.e., (1) retrieving candidate genes from the SoyBase database, (2) visualizing the molecular function of candidate genes by GO enrichment analyses and gene classification, and (3) implying candidate genes in seed size, shape, and weight based on their expression profiles. Accordingly, 47 genes were considered as potential candidates. Most of the identified candidate genes are related to the terms of catalytic activity, cell part, cell, cellular process, and binding and metabolic process as indicated by GO enrichment and gene classification analyses. These terms have functions related/involved in seed development, which influence the size, shape, and weight of seeds (Mao et al., 2010; Li and Li, 2014). For example, the Glyma07g14460 gene underlying QTL cluster-7 belongs to the oxygenase (CYP51G1) protein class, which has been confirmed to regulate seed size in soybean (Zhao et al., 2016). Furthermore, 10 candidate genes were identified as a regulator of ubiquitin-dependent protein catabolic process, RING-type E3 ubiquitin ligases, and lipid catabolic process (Table 5). Several components of the ubiquitin pathway such as the ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin protein ligase (E3) have been reported to play important roles in regulation seed and organ size (Li and Li, 2014). Similarly, 16 candidate genes have functions in pollen tube development, embryo sac egg cell differentiation, post-embryonic development, regulation of seed maturation, positive regulation of gene expression, regulation of cell cycle process, ovule development, anther development, seed dormancy process, and seed maturation (Table 5), and hence they are likely to participate in regulating seed size, shape, and weight in plants, including soybean (Meng et al., 2016). Ten candidate genes are involved in response to auxin stimulus, response to ethylene stimulus, and abscisic acid biosynthetic process which are known to be implicated in promoting seed size and weight in Arabidopsis (Table 5) (Xie et al., 2014). Six genes play functions in the glucose catabolic process, phosphatidylcholine biosynthetic process, carbohydrate metabolic process, maltose metabolic process, and starch biosynthetic process and are implicated in the partitioning and translocation of photoassimilates and grain filling in rice (Table 5) (Chen J. et al., 2020; Zhang et al., 2020).

CONCLUSION

QTLs associated with seed size, shape, and weight in soybean were identified and validated using two mapping approaches in two populations across multiple environments. This is the first comprehensive investigation of the identification and validation of QTLs for the FI as a seed shape trait in soybean. Employing a bioinformatics pipeline identified candidate genes behind genomic regions harboring major and stable QTL clusters underlying the inheritance of seed size, shape, and weight. The implemented bioinformatics pipeline delimits the number of the identified candidate genes to 47-gene genomic regions involved directly or indirectly in seed size, shape, and weight. These genes are highly expressed in seed-related tissues and nodules, indicating that they may be involved in regulating these traits in soybean. Furthermore, some of the potential 47 candidate genes have been included in our ongoing projects for functional validation to confirm their effect on seed size, shape, and weight. Our study provides detailed information for genetic bases of the studied traits and candidate genes that could be efficiently implemented by soybean breeders for fine mapping and gene cloning and for MAS targeted at improving seed size, shape, and weight.

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 designed the project. ME performed the experiments and drafted the manuscript. ME, BK, SS, SL, YC, MA, and AH analyzed the data. TZ and SA-E revised the manuscript.
All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Key R&D Program of China (2018YFD0100800), the National Natural Science Foundation of China (31571691 and 31871646), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT_17R55), the Fundamental Research Funds for the Central Universities (KY201801), and the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCICMCP) Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.666440/full#supplementary-material

REFERENCES

Abou-Elwafa, S. F. (2016a). Association mapping for drought tolerance in barley at the reproductive stage. Comptes Rendus Biol. 339, 51–59. doi: 10.1016/j.crvi.2015.12.002

Abou-Elwafa, S. F. (2016b). Association mapping for yield and yield-contributing traits in barley under drought conditions with genome-based SSR markers. C. R. Biol. 339, 153–162. doi: 10.1016/j.crvi.2016.03.001

Abou-Elwafa, S. F. (2018). Identification of genes associated with drought tolerance in barley. Biol. Plant. 62, 299–306. doi: 10.1007/s10211-017-0765-0

Abou-Elwafa, S. F., and Shehzad, T. (2018). Genetic identification and expression profiling of drought responsive genes in sorghum. Environ. Exp. Bot. 155, 12–20. doi: 10.1016/j.envexpbot.2018.06.019

Basra, A. S. (1995). Seed quality: Basic Mechanisms and Agricultural Implications. New York, NY: Food Products Press.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. 48, 1649–1664. doi: 10.2135/cropsci2008.03.0131

Cai, H., and Morishima, H. (2002). QTL clusters reflect character associations in wild and cultivated rice. Theor. Appl. Genet. 104, 1217–1228. doi: 10.1007/s00122-001-0819-7

Cailleaux, A. (1945). Distinction des galets marins et fluviales. Bull. de la Soc. Géologique de France 5, 375–404. doi: 10.2113/gssgfbull.s5-xv.7-8.375

Cao, Y., Li, S., Wang, Y., Bhat, J. A., Karikari, B., et al. (2019). Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants 8:373. doi: 10.3390/plants8100373

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

Cerdà, A., and Garcia-Fayos, P. (2002). The influence of seed size and shape on their removal by water erosion. Catena 48, 293–301.

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TTools-an integrative toolkit developed for interactive analyses of big biological data. bioRxiv [preprint], 13, 1194–1202.

Chen, J., Cao, F., Li, H., Shan, S., Tao, Z., Lei, T., et al. (2020). Genotypic variation in the grain photosynthetic contribution to grain filling in rice. J. Plant Physiol. 253:153269. doi: 10.1016/j.jplph.2020.153269
Cheng, X., Wang, S., Wang, L., Wang, S., Zong, X., and Wang, P. (2006). *Descriptors and Data Standard for Azuki Bean [Vigna angularis (Willd.) Ohwi & Ohashi]*. Beijing: China Agriculture Press.

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971. doi: 10.1093/genetics/138.3.963

Cober, E., Voldeng, H., and Frégeau-Reid, J. (1997). Heritability of seed shape and seed size in soybean. *Crop Sci.* 37, 1767–1769. doi: 10.2135/cropsci1997.0011183x000730060017x

Edwards, C. J. Jr., and Hartwig, E. E. (1971). Effect of seed size upon rate of germination in soybeans. *Agron. J.* 63, 429–450. doi: 10.2134/agronj1971.00029162006300060024x

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. *Genet. Biol.* 18, 1–14.

Funatsuki, H., Kawaguchi, K., Matsuba, S., Sato, Y., and Ishimoto, M. (2005). *Functional genomics of soybean root resistance to charcoal rot*. *Theor. Appl. Genet.* 112, 1366–1374. doi: 10.1007/s00122-005-0179-7

Hacisalihoglu, G., and Settles, A. M. (2017). Quantification of seed ionome associated with chilling tolerance during reproductive growth in soybean. *Theo. Appl. Gen.* 111, 851–861.

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., et al. (2012). QTL analysis of cleistogamy in soybean. *Theor. Appl. Genet.* 117, 479–487. doi: 10.1007/s00122-008-0792-5

Hu, Z., Zhang, H., Kan, G., Ma, D., Zhang, D., Shi, G., et al. (2013). Determination of soybean seed weight across multi-genetic backgrounds and environments. *J. Int. Plant Biol.* 35, 341–346. doi: 10.1007/s13353-013-0036-z

Jannink, J.-L., Moreau, L., Charmet, G., and Charcosset, A. (2009). Overview of association studies dissect the genetic networks underlying agronomical traits in soybean (*Glycine max*). *Theor. Appl. Genet.* 117, 213–228. doi: 10.1007/s00122-008-0098-x

Khan, N. A., Githiri, S. M., Benitez, E. R., Abe, J., Kawasaki, S., Hayashi, T., et al. (2008). QTL analysis of cleistogamy in soybean. *Theor. Appl. Genet.* 117, 479–487. doi: 10.1007/s00122-008-0792-5

Kin, C. (2013). Kin conflict in seed development: an interdependent but fratricidal collective. *Ann. Rev. Cell Dev. Biol.* 29, 189–211. doi: 10.1146/annurev-cellbio-101512-122324

Khan, N. A., Githiri, S. M., Benitez, E. R., Abe, J., Kawasaki, S., Hayashi, T., et al. (2008). QTL analysis of cleistogamy in soybean. *Theor. Appl. Genet.* 117, 479–487. doi: 10.1007/s00122-008-0792-5

Kli, N., and Li, Y. (2014). Ubiquitin-mediated control of seed size in plants. *Front. Plant Sci.* 5:332. doi: 10.3389/fpls.2014.00332

Li, S., Cao, Y., He, J., Zhao, T., and Gai, J. (2017). Detecting the QTL-allele system conferring flowering date in a nested association mapping population of soybean using a novel procedure. *Theor. Appl. Genet.* 130, 2297–2314. doi: 10.1007/s00122-017-4790-0

Li, W., Zheng, D.-H., Van, K., and Lee, S.-H. (2008). QTL mapping for major agronomic traits across two years in soybean (*Glycine max L. Merr.*). *J. Crop. Sci. Biotechnol.* 11, 171–190.

Liang, H., Xu, L., Yu, Y., Yang, H., Dong, W., and Zhang, H. (2016). Identification of QTLs with main, epistatic and QTL by environment interaction effects for seed shape and hundred-seed weight in soybean across multiple years. *J. Genet.* 95, 473–477. doi: 10.1007/s10929-016-0648-8

Liu, D., Yan, Y., Fujita, Y., and Xu, D. (2018). Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breed. Science* 68, 442–448. doi: 10.1270/jsbs.17127

Liu, N., Li, M., Hu, X., Ma, Q., Mu, Y., Tan, Z., et al. (2017). Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. *BMC Genom.* 18:466.

Lu, X., Xiong, Q., Cheng, T., Li, Q.-T., Liu, X.-L., Bi, Y.-D., et al. (2017). A P2PC-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10, 670–684. doi: 10.1093/mp/mtx132

Ma, J., Wingen, L. U., Orford, S., Fenwick, P., Wang, J., and Griffiths, S. (2015). Using the UK reference population Avalon® Cadenza as a platform to compare breeding strategies in elite Western European bread wheat. *Mol. Breeding* 35:70.

Mahnoud, A. F., Abou-Elwafa, S. F., and Sheshadri, T. (2018). Identification of charcoal rot resistance QTLs in sorghum using association and in silico analyses. *J. Appl. Genet.* 59, 243–251. doi: 10.1007/s13353-018-0446-5

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking differential domain functions of the G3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19579–19584. doi: 10.1073/pnas.1014419107

Meng, Y., Chen, F., Shuai, H., Luo, X., Ding, J., Tang, S., et al. (2016). Karrikins delay soybean seed germination by mediating abscisic acid and gibberellin biogenesis under shaded conditions. *Sci. Rep.* 6, 1–12.

Mohan, M., Nair, S., Bhagwat, A., Krishna, T., Yano, M., Bhatia, C., et al. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol. Breeding* 3, 87–103.

Moongkanna, J., Nakasathien, S., Novitzky, W., Kwanyuen, P., Sinchaisri, P., and Srinives, P. (2011). SSR markers linking to seed traits and total oil content in soybean. *Thai. J. Agric. Sci.* 44, 233–241.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7:277.

Nyquist, W. E., and Baker, R. (1991). Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10, 235–322. doi: 10.1080/07352689109382313

Omokhafe, K., and Alika, J. (2004). Clonal variation and correlation of seed traits and total oil content in *Vigna angularis* L. *Biomass and Bioenergy* 31, 497–502. doi: 10.1016/j.biombioe.2004.07.020

Panthee, D., Pantalone, V., West, D., Saxton, A., and Sams, C. (2005). Comparative QTL analysis of root lesion nematode resistance to *Sclerotium rolfsii* in barley. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17597–17598. doi: 10.1073/pnas.1014419107

Panthee, D., Pantalone, V., West, D., Saxton, A., and Sams, C. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Sci.* 45, 2015–2022. doi: 10.2135/cropsci2004.07.020

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., and Hoekstra, H. E. (2012). Double digest RAD-seq: an inexpensive method for de novo SNP discovery.
Xing, G., Zhou, B., Wang, Y., Zhao, T., Yu, D., Chen, S., et al. (2012). Genetic components and major QTL confer resistance to bean pyralid (Lamprosena indicata Fabricius) under multiple environments in four RIL populations of soybean. *Theor. Appl. Genet.* 125, 859–875. doi: 10.1007/s00122-012-1878-7

Xu, Y., Li, H.-N., Li, G.-J., Wang, X., Cheng, L.-G., and Zhang, Y.-M. (2011). Mapping quantitative trait loci for seed size traits in soybean (Glycine max L. Merr.). *Theor. Appl. Genet.* 122, 581–594. doi: 10.1007/s00122-010-1471-x

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

Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X., et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24, 721–723. doi: 10.1093/bioinformatics/btm494

Yang, J., Zhu, J., and Williams, R. W. (2007). Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23, 1527–1536. doi: 10.1093/bioinformatics/btm143

Yang, K., Moon, J.-K., Jeong, N., Chun, H.-K., Kang, S.-T., Back, K., et al. (2011). Novel major quantitative trait loci regulating the content of isoflavone in soybeans. *Genes Genom.* 33, 685–692. doi: 10.1016/j.sggg.2011-04-0343

Yao, D., Wang, P., Zhang, J., Liu, Z., Guan, S., Liu, S., et al. (2014). A QTL mapping analysis of main yield traits in soybean. *J. South China Agric. Univ.* 35, 41–46.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WFGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–W297.

Zhang, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics* 136, 1457–1468. doi: 10.1093/genetics/136.4.1457

Zhang, F., Kang, J., Long, R., Yu, L.-X., Wang, Z., Zhao, Z., et al. (2019). High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC Plant Biol.* 19:165.

Zhang, H., Chen, J., Shan, S., Cao, F., Chen, G., Zou, Y., et al. (2020). Proteomic profiling reveals differentially expressed proteins associated with amylose accumulation during rice grain filling. *BMC Genom.* 21:714.

Zhang, L., Xu, X., Zhao, C., Shan, F., Yuan, S., and Sun, H. (2011). QTL analysis of plant height based on doubled haploid (DH) population derived from PTMS wheat. *Mol. Plant Breeding* 2, 92–97.

Zhang, W.-K., Wang, Y.-J., Luo, G.-Z., Zhang, J.-S., He, C.-Y., Wu, X.-L., et al. (2004). QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers. *Theor. Appl. Genet.* 108, 1131–1139. doi: 10.1007/s00122-003-1527-2

Zhang, X., Hina, A., Song, S., Kong, J., Bhat, J. A., and Zhao, T. (2019). Whole-genome mapping identified novel “QTL hotspots regions” for seed storability in soybean (Glycine max L). *BMC Genom.* 20:499.

Zhang, Y., Li, W., Lin, Y., Zhang, L., Wang, C., and Xu, R. (2018). Construction of a high-density genetic map and mapping of QTLs for soybean (Glycine max) agronomic and seed quality traits by specific length amplified fragment sequencing. *BMC Genom.* 19:641.

Zhao, B., Dai, A., Wei, H., Yang, S., Wang, B., Jiang, N., et al. (2016). ArabidopsisKL homologue GmCYPR7A2 regulates seed size in soybean. *Plant Mol. Biol.* 90, 33–47. doi: 10.1007/s11103-015-0392-0

Zhao, F., and Xu, S. (2012). Genotype by environment interaction of quantitative traits: a case study in barley. *G3 Genes Genom. Genet.* 2, 779–788. doi: 10.1534/g3.112.002980

**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.

Copyright © 2021 Elattar, Karikari, Li, Song, Cao, Aslam, Hina, Abou-Elwa and Zhao. 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.