

Comparative QTL analysis and candidate genes identification of seed size, shape and weight in soybean (*Glycine max* L.)

Mahmoud A Elattar

Nanjing Agricultural University

Benjamin Karikari

Nanjing Agricultural University

Shuguang Li

Nanjing Agricultural University

Shiyu Song

Nanjing Agricultural University

Yongce Cao

Nanjing Agricultural University

Muhammed Aslam

Nanjing Agricultural University

Aiman Hina

Nanjing Agricultural University

Salah Fatouh Abou-Elwafa (✉ elwafa75@aun.edu.eg)

Assiut University Faculty of Agriculture <https://orcid.org/0000-0002-9018-598X>

Tuanjie Zhao

Nanjing Agricultural University

Research Article

Keywords: Glycine max, Soybean seed, QTL mapping, QTL clusters, epistatic interactions, marker assisted breeding

Posted Date: February 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-206236/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Dissecting the genetic mechanism underlying seed size, shape and weight is essential to these traits for enhancing soybean cultivars. High-density genetic maps of two recombinant inbred line populations, LM6 and ZM6, evaluated in multiple environments to identify candidate genes behind seed-related traits major and stable QTLs. A total of 239 and 43 M-QTL were mapped by composite interval mapping and mixed-model based composite interval mapping approaches, respectively, from which 22 common QTLs including four major and novel QTLs. CIM and MCIM approaches identified 180 and 18 novel M-QTLs, respectively. Moreover, 18 QTLs showed significant AE effects, and 40 pairwise of the identified QTLs exhibited digenic epistatic effects. Seed flatness index QTLs (34 QTLs) were identified and reported for the first time. Seven QTL clusters underlying the inheritance of 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 identified 47 candidate genes for seed-related traits within the genomic regions of those 7 QTL clusters. These genes are highly expressed in seed-related tissues and nodules, that might be deemed as potential candidate genes regulating the above traits in soybean. This study provides detailed information for the genetic bases of the studied traits and candidate genes that could be efficiently implemented by soybean breeders for fine mapping and gene cloning as well as for MAS targeted at improving these traits individually or concurrently.

Introduction

Soybean (*Glycine max* L. Merr.) is one of the most important food crops, being a rich source of dietary protein (69 %) and provides more than 50 % edible oil globally, as well as has a significant role in health and biofuel (Hoeck et al. 2003). It is used in human food and animal feed due to its high nutritional value and improves soil fertility by integrating atmospheric nitrogen in the soil through a synergetic interaction with microorganisms (Wang et al. 2019). Throughout the last five decades, soybean production in China has suffered continuous annual decline and reduced yields. To meet domestic demands, China imports nearly 80 % to meet its domestic demand, therefore improving soybean production has been the main objective of soybean breeders to make the country self-sufficient (Liu et al. 2018). Accordingly, most plant breeders are targeting yield-related traits to improve soybean production. In this regard, seed size traits, i.e., seed length (SL), thickness (ST) and width (SW), and seed shape traits, i.e., length-to-thickness (SLT), length-to-width (SLW), width-to-thickness (SWT) ratios and flatness index (FI) determine seed vigor, quality and yield in soybean (Cha-um and Kirdmanee 2011; Salas et al. 2006). Flatness index contributes a lot towards seed vigor which impacts the seedling emergence time by determining the development of Soybean plant throughout their growth cycle as contributing factor to higher leaf area index from early vegetative stages (V_1) to reproductive stages (R_2) (Ebone et al. 2020). Vigor reduced meaning flatness index reduced, showing the increase in variability among the plants in the field. It generates differences among distinct groups as a result of higher trifoliolate leaf area, allowing the plants to use more resources (light) for their development. Thus, shooting the yield of the plant because of higher leaf area index, higher accumulation of photoassimilates and finally contributed to seed yield. These plants therefore not need to increase their average internode length (Wu et al. 2017) and have a higher number of total nodes and a higher pod number which explains the differences in grain mass (Ainsworth et al. 2012).

Moreover, 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 \times environment (G \times E) interaction, and hence it is more difficult to select for based on phenotype compare with monogenic traits (Yao et al. 2014). A positive correlation between seed size/weight and seed yield has been reported in several studies (Burriss et al. 1973; Smith and Camper Jr 1975). Besides, seed weight/size has been demonstrated to be associated with seed germination capability and vigor, thereby significantly affecting the competitive capability of the seedling for nutrient and water resources and light, hence enhance stress tolerance (Edwards Jr and Hartwig 1971; Haig 2013). All soybean varieties evolved in tropical and subtropical countries, such as Indonesia and India have small seed size compared to the temperate region varieties, such as China, the USA and Japan. Besides, seed size, shape and weight are important seed quality traits with great influence on seed use, for example, round seeds are often desirable for food-type soybean, and large-seeded cultivars are typically used for green soybeans (edamame), soymilk, miso, boiled soybean (*nimame*) and soybean curd (*tofu*), whereas, small-grained cultivars are desirable for sprout production and fermented soybean (*natto*) (Basra 1995; Teng et al. 2017; Wu et al. 2018). Therefore, depending on the end-use or biographical area of growth, many soybean varieties with different seed sizes and shapes with a 100-seed weight ranged from 3.0 - 77.5 g have been developed (Panthee et al. 2005).

Quantitative trait locus (QTL) analysis provides an efficient tool for crop breeders to study the genetic factors underlying quantitatively inherited traits and search for new sources of variation. The key influence of quantitative trait loci (QTLs), phenotypic variance (PV) of quantitative traits is often regulated by epistasis and QTL \times environment (QEs) interactions, which greatly lead to variations in complex traits (Yang et al. 2005). Furthermore, QTL analysis of complex traits may increase the accuracy of QTL mapping if QTL by QTL interactions are considered. However, epistatic interaction has a stronger effect on inbreeding depression, heterosis, adaptation, speciation and reproductive isolation (Ma et al. 2015). Most of the previous studies focused on the detection of main-effect QTLs for seed sizes, shapes and 100-seed weight in soybean. To date, at least 441, 52 and 297 QTLs for seed size, shape and 100-seed weight (HSW) have been reported (www.soybase.org) based on various genetic contexts, advances in marker technology, statistical methods and specific environments. However, most of these QTLs are minor ($R^2 < 10\%$), not stable and with larger genomic region/confidence interval. Salas et al. (Salas et al. 2006) found that 26 QTLs for seed size and shape on 13 soybean linkage groups. Han et al. (Han et al. 2012) mapped 46 QTLs associated with HSW in 3 RIL soybean populations with one common male parent (Hefeng25). Furthermore, Hu et al. (Hu et al. 2013) detected 10 QTLs for seed shape on 6 chromosomes in soybean. Moreover, Kato et al. (Kato et al. 2014) identified 15 major QTLs for single seed weight among 11 chromosomes by using two RIL soybean populations developed from crosses between the US and Japanese cultivars of soybean. Recently, there have been limited studies on detecting QTLs with epistatic effects and their interactions with the environment (QEs) (Liang et al. 2016; Panthee et al. 2005; Xu et al. 2011; Zhang et al. 2018).

Most of the previous QTL analysis researches have relied on low-density genetic maps used biochemical and morphological, simple sequence repeats (SSRs), restriction fragment length polymorphisms (RFLPs), or other low-quantity markers, which resulted in large QTLs confidence intervals with a low resolution compared with high-density SNP markers that are valuable for high-throughput QTL mapping (Hu et al. 2013; Moongkanna et al. 2011; Salas et al. 2006). With the recent advancement in next-generation sequencing (NGS) techniques, a number of techniques of high-throughput sequencing have been developed to

generate large scale markers. Among them include genotyping-by-sequencing (GBS) (Poland et al. 2012), restriction-site associated DNA sequencing (RAD-seq) (Miller et al. 2007; Peterson et al. 2012) and specific length amplified fragment sequencing (SLAF-seq) (Sun et al. 2013). These sequencing technologies have facilitated the production of hundreds to millions of single-nucleotide polymorphisms (SNPs) throughout the whole genome that promote the development of high-density linkage maps. The RAD-seq produces markers that have been proved to be a promising tool for SNP detection and genetic map construction (Chutimanitsakun et al. 2011; Xie et al. 2018). A number of genetic maps produced by RAD-seq have been generated and used for QTL mapping in several crops such as barley (Chutimanitsakun et al. 2011), soybean (Hina et al. 2020), cowpea (Pan et al. 2017), jute (Kundu et al. 2015), sorghum (Kajiy-Kanegae et al. 2020), alfalfa (Zhang et al. 2019a) among others.

In addition to the above, knowledge of molecular mechanisms underlying soybean seed size, shape and weight is still limited. So far, only two seed sizes/weight-related genes have been isolated from the soybean. The gene *Glyma20g25000 (ln)* has a significant impact on seed size and the number of seeds per pod (Jeong et al. 2012), and the PP2C-1 allele underlying *Glyma17g33690* from the wild soybean accession 'ZYD7' was reported and demonstrated to increase 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). Due to the shortage in the available molecular markers and the lack of high-density linkage maps which resulted in low resolution and large confidence intervals of identified QTLs, the present study used RAD-seq to generate over 2200 bin-markers for either of the two-related recombinant inbred line (RIL) populations for QTL mapping and candidate gene(s) identification for seed size, shape and weight. The two-related RIL populations (ZM6 and LM6) were derived from a common male parent *Meng 8206 (M8206)* crossed with either *Zhengyang (Z)* and *Linhefenqingdou (L)* and RILs and their parents were evaluated across multiple environments. The study was aimed to: (i). map main-effect QTLs (M-QTLs), additive by additive QTLs and QE for seed size, shape and weight traits, (ii). analyze epistatic QTL pairs and their interactions with the environment for further utilization of these QTLs in soybean genetic improvement, and (iii). mine potential candidate genes for the major and stable QTLs. These findings would be useful for the application of marker-assisted breeding (MAB) in soybean and provide comprehensive knowledge on the genetic bases for these traits as well as mined candidate genes would serve as a foundation for functional validation and verification of some genes for seed size, shape and weight in soybean.

Materials And Methods

Plant materials and experiments

Two recombinant inbred lines (RIL) populations, i.e., ZM6 and LM6 (Karikari et al. 2019; Zhang et al. 2019b), consisting of 126 and 104 lines, respectively, were used in the present study. The two populations were developed by single seed descent (SSD) with the genotypes *Zhengyang (Z)* and *Linhefenqingdou (L)* were used as female parents and the *M8206 (M6)* genotype was used 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 and shape across multiple environments. Experiments were conducted in the Jiangpu Experimental Station (33°030 N and 63°1180 E), Nanjing, Jiangsu Province, in 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 2012 growing season (designated as 12FY) and the Yancheng Experimental Station, Yangcheng, Jiangsu Province, (33°410 N and 120°200 E) in 2014 (designated as 14YC). Plants were sown in June and harvest were 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 agronomical practices.

Phenotypic Measurement and Statistical Analyses

Eight seed-related traits including seed length (SL), seed width (SW), seed thickness (ST), seed length/seed width (SLW), seed length/seed thickness (SLT), seed width/seed 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 (Cheng et al. 2006; Tomooka et al. 2002). 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) utilizing the Vernier caliper instrument, according to (Kaushik et al. 2007) (Fig. 1). The seed shape was identified by calculating three different ratios, i.e., SL/SW (SLW), SL/ST (SLT), and SW/ST (SWT), as well as flatness index (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 (Omokhafa and Alika 2004). Flatness index is an indicator for high or lower seed vigor which impacts the seedling uniformity index as it reduces if the seed vigor reduces. Thus, showing that the plants with FI near to one has good seed vigor concluding higher leaf area index at early vegetative stages and hence higher yield per plant as they have a good number of pods due to the higher accumulation of photoassimilates. These plants, therefore, do not need to increase their average internode length and have a higher number of total nodes and a higher pod number. While flatness index (FI) was calculated following the formula elaborated by (Cailleux 1945) and (Cerdà and Garcia-Fayos 2002) to describe seed shape:

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

where *L* is the seed length, *W* is seed width and *T* is seed thickness.

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.

The descriptive statistics of the seed size, seed shape, and HSW traits were calculated using the SPSS software, version 24 (<http://www.spss.com>). The analysis of variance (ANOVA) for each environment and the combined overall environments (CE) were performed using the PROC GLM procedure in SAS software based on the random model (SAS Institute Inc. v. 9.02, 2010, Cary, NC, USA). The broad-sense heritability (h^2) in individual environments was estimated as:

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2)$$

whereas in the CE h^2 was estimated as follow: $h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_{ge}^2/n + \sigma_e^2/nr)$

where σ_g^2 , σ_e^2 and σ_{ge}^2 are the variance components estimated from the analysis of variance for the genotypic, error and genotype \times 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. Pearson correlation coefficient (r) between seed size, seed shape, and HSW traits was calculated from the mean data utilizing the SAS PROC CORR with data obtained for CE (average across environments) for each population.

Construction of Genetic Maps and QTL Analysis

High-density genetic maps of the ZM6 and LM6 populations consist of 2601 and 2267 bin markers by using RAD-seq technique, respectively (Karikari et al. 2019; Zhang et al. 2019b) (Suppl. Table 1). The total length of the ZM6 and LM6 maps were 2630.22 and 2453.79 cM, with an average distance between the markers 1.01 and 1.08 cM, respectively (Suppl. Table 1). The Average marker per chromosome was 130 and 113 for ZM6 and LM6 linkage maps, respectively, with an average genetic distance per chromosome 131.51 and 122.69 cM (Suppl. Table 1).

Main- and Epistatic-Effect QTLs Mapping

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 was computed using 1000 permutations due to an experiment-wide error proportion of $P < 0.05$ (Churchill and Doerge 1994), and the confidence interval was determined utilizing a 1-LOD support interval, which was controlled by finding the local on the two sides of a QTL top that compatible with a reduction of 1 LOD score. The QTL detected within the overlapping intervals in different environments were considered the same (Palomeque et al. 2009; Palomeque et al. 2010; Zhaoming et al. 2017). Moreover, to identify the genetic effects of the QTLs, i.e., additive QTLs, additive \times additive (AA), additive \times environment (AE) and AA \times environment (AAE), the mixed-model based composite interval mapping (MCIM) procedure was employed in the QTLNetwork V2.1 software (Yang et al. 2008). 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 $P < 0.05$ (Wang et al. 1994; Xing et al. 2012; Yang et al. 2007).

Mining of Candidate Genes for QTL Clusters

QTLs identified in two or more environments with an $R^2 > 10\%$ were considered as major and stable QTLs (Zhaoming et al. 2017). Regions on chromosomes with several M-QTLs related to different traits studied in this research were termed as a QTL Cluster. The Phytozome (<http://phytozome.jgi.doe.gov>) and SoyBase (<http://www.soybase.org>) online platform repositories, were employed to retrieve all model genes within the physical interval position of the main "QTL Clusters". Possible candidate genes were predicted based on gene annotations (<http://www.soybase.org> and <https://phytozome.jgi.doe.gov>) as well as the reported putative function of genes implicated in these traits. Gene ontology (GO) information was obtained from SoyBase via the online resources, i.e., The Center for Biotechnology Information (NCBI: <https://www.ncbi.nlm.nih.gov>), GeneMania (<http://genemania.org/>) and the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.kegg.jp). These online tools were employed to further screen the predicted candidate genes. Gene ontology (GO) enrichment analysis was conducted for all the genes within each QTL cluster region using AgriGO V2.0 (<http://systemsbiology.cau.edu.cn>) (Tian et al. 2017). Gene classification was then carried out using 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 predicted candidate genes in various soybean tissues and the development stages. A heatmap to visualize the fold-change patterns of these candidate genes was developed using the TBtools_JRE 1.068 software (Chen et al. 2020a).

Results

Phenotypic evaluation of seed size, seed shape, and 100-seed weight traits

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 analysis of variance (Tables S2 and S3). Analysis of variance (ANOVA) revealed that all studied traits were significantly ($P < 0.001$ or $P < 0.05$) influenced by the environment, genotypes and the genotype \times environment interaction (Tables S4 and S5), indicating the differential response of the genotypes to the changes in environmental cues. Furthermore, the two populations showed continuous phenotypic variations in all studied traits, implying a polygenic inheritance of these traits (Fig. 2). Besides, the estimation of skewness, kurtosis and coefficient of variation (CV) for all studied traits across all environments showed that most of the recorded skewness and kurtosis values were < 1 with CV values of $> 3\%$, emphasizing that

these traits in both populations are controlled by polygenes and are fit for QTL mapping (Tables S2 and S3). 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 (Fig. 3). 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, compared to the male parent, M8206. 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 (Fig. 3, Tables S2 and S3). 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 (Fig. 2, 3). The broad-sense heritability (h^2) under individual environments ranged from 69.58–99.04%, 69.08–97.9% and 78.49–98.68% for seed size, HSW and seed shape (Tables S2 and S3). Meanwhile, h^2 under combined environments (CE) ranged from 90.82–94.48%, 87.51–94.46% and 97.58–98.39% for seed size, shape and HSW, respectively. The correlation coefficient (r^2) 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^2 values ranged from 0.79–0.91. Meanwhile, SL, ST and SW exhibited significant negative correlations with the other two seed shape traits (SWT and FI) (Suppl. Table 6). 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^2 values ranged from 0.33–0.95. Furthermore, all seed size traits, i.e., SL, SW, and ST showed significant positive correlations with HSW with r^2 values ranged from 0.29 to 0.70 in both populations.

Mapping of seed size main-effect QTLs in the two-related RIL populations

A total of 92 M-QTLs were mapped for seed size related traits, i.e., SL, SW and ST, on all chromosomes in soybean, except chromosomes 1 and 12, with logarithm of odd (LOD) scores and phenotypic variations (R^2) ranged from 2.5–10.3 and 5.0–19.7 %, respectively, in the two populations (Suppl. Table 7, Suppl. Figure 1a, b, c). Out of these, 30 M-QTLs for SL, 35 for SW and 27 for ST with alleles underlying QTLs emanated from either of parents. Seventy-two M-QTLs were mapped in a specific environment while the remaining 20 were mapped within overlapping regions in at least one specific environment together with or without CE. Forty-seven QTLs that exhibited $R^2 > 10$ %, hence were considered as major QTLs. The most prominent QTL was the *qSW-17-2_{LM6}* (LOD = 6.70–10.29, and $R^2 = 16.60–18.30$ %) was detected within the physical position 6844412–9645325 bp in 14JP and CE (Fig. 4b). Likewise, the *qSL-10-2_{ZM6, LM6}* (LOD = 6.08–6.89, and $R^2 = 15.4–17.1$ % in ZM6 (17JP) and LM6 (14JP) populations) was located to the physical position between 41454163–43944243 bp. Aside, *qSL-10-2_{ZM6, LM6}*, in terms of stability across at least two specific environments with or without the CE, *qSL-10-1_{LM6}*, *qSL-11-1_{ZM6}*, *qSL-18-3_{LM6}*, *qSL-20-1_{ZM6}*, *qSW-10-2_{ZM6}*, *qSW-16-1_{LM6}*, *qST-4-1_{LM6, ZM6}*, *qST-16-1_{LM6}* and *qST-20-1_{ZM6}* accounted for averages of 11.17, 14.00, 10.70, 7.65, 11.30, 6.75, 8.85, 13.30 and 11.13 %, respectively (Suppl. Table 7, Suppl. Figure 1a, b, c).

Mapping of seed shape main-effect QTLs in the two populations.

In total, ninety-nine M-QTLs related to seed shape traits (SLT, SLW, SWT, and FI) were mapped to 19 soybean chromosomes excluding chromosome (Chr02) in both populations (ZM6 and LM6) among four environments plus CE with LOD scores of (2.50–10.44) and R^2 (5.12–31.56 %) by the composite interval mapping (CIM) approach (Suppl. Table 8, Suppl. Figure 1d, e, f, g). From the 99 M-QTLs, 22, 33, 11, and 22 were detected for SLT, SLW, SWT, and FI, respectively (Suppl. Table 8). Among them, seventy-one M-QTLs were detected in specific environments, while 28 were mapped in at least one specific environment together either with or without the CE. Eight M-QTLs for SLW (*qSLW-3-2_{LM6}*, *qSLW-5-3_{ZM6, LM6}*, *qSLW-9-3_{LM6, ZM6}*, *qSLW-13-4_{ZM6, LM6}*, *qSLW-15-1_{LM6}*, *qSLW-15-3_{LM6}*, *qSLW-16-2_{ZM6}* and *qSLW-16-3_{ZM6}*) were mapped in at least one specific 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 specific environment with or without CE (Suppl. Table 8, Suppl. Figure 1e). Likewise, four M-QTLs (*qSWT-8-1_{LM6}*, *qSWT-11-2_{ZM6, LM6}*, *qSWT-13-1_{ZM6}* and *qSWT-17-1_{ZM6}*) were mapped for SWT (Suppl. Table 8, Suppl. Figure 1f). Also, a total of 10 M-QTLs (*qFI-1-1_{ZM6}*, *qFI-1-2_{ZM6, LM6}*, *qFI-1-3_{LM6}*, *qFI-1-4_{ZM6, LM6}*, *qFI-3-1_{LM6}*, *qFI-3-3_{ZM6}*, *qFI-5-2_{ZM6}*, *qFI-11-1_{LM6}*, *qFI-14-1_{LM6}*, *qFI-17-1_{ZM6}* and *qFI-20-1_{ZM6}*) were considered as stable.

Several physical regions identified harbored at least two seed shape related traits, i.e., 1730667–3014518 bp harbored *qSLT-1-2_{ZM6}*, *qSWT-1-1_{ZM6}* and *qFI-1-2_{ZM6, LM6}*, and 4946300–35955471 bp had *qSLT-1-3_{LM6}*, *qSWT-1-2_{LM6}* and *qFI-1-3_{LM6}* on Chr01 (Suppl. Figure 2, Suppl. Table 8). Two M-QTLs (*qSLW-3-2_{LM6}* and *qFI-3-1_{LM6}*) colocalized within the physical region of 1509548–3515594 bp. Moreover, *qSLT-5-3_{ZM6}*, *qFI-5-2_{ZM6}* and *qSLW-5-3_{ZM6, LM6}* overlapped within the physical region of 38035798–41186985 bp (Suppl. Figure 2, Suppl. Table 8). Furthermore, *qSLT-11-1_{LM6}* and *qFI-11-1_{LM6}* (Chr11); *qSLT-13-1_{LM6}*, *qSLW-13-4_{ZM6, LM6}* and *qSWT-13-1_{ZM6}* (Chr13); *qFI-14-1_{LM6}*, *qSLT-14-1_{LM6}* and *qFI-14-1_{LM6}* (Chr14) located within regions of 17145381–23469672, 33303067–39562563 and 3468251–8668367 bp, respectively (Suppl. Figure 2, Suppl. Table 8). *qSLW-16-2_{ZM6}*, *qSLT-16-1_{LM6}* and *qSLW-16-3_{ZM6}* QTLs *qFI-17-1_{ZM6}* and *qSWT-17-1_{ZM6}* and *qSLT-20-1_{ZM6}* and *qFI-20-1_{ZM6}* were located to the chromosomal regions of 26903205–31959397 on Chr16, 40207655–41672092 bp on Chr17 and 1–1115156 bp on Chr20, respectively (Suppl. Figure 2, Tables S8).

Mapping of hundred-seed weight main-effect QTLs in the two populations

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 specific environment and/without CE (Suppl. Table 9, Suppl. Figure 1h). The LOD scores and R^2 values of these M-QTLs ranged from 2.51–10.61 and 4.8–24.5 %, respectively. The highest number of M-QTLs of 6 were mapped on Chr04 followed by Chr10 with 5 M-QTLs and the lowest number of one M-QTL was mapped to Chr02, Chr09, Chr17 and Chr18. The most prominent M-QTLs were *qHSW-14-2_{ZM6}*, *qHSW-10-3_{LM6}* and *qHSW-10-4_{LM6}* with LOD scores and R^2 values of 10.61 and 24.50 % (Fig. 4), 7.57 and 17.60 %, and 7.20 and 16.90 %, respectively. Among those 21 M-QTLs, *qHSW-4-3_{LM6, ZM6}*, *qHSW-6-2_{LM6}*, *qHSW-10-1_{LM6}*, *qHSW-13-1_{ZM6}*, *qHSW-15-2_{LM6}* and *qHSW-15-4_{LM6}* were mapped in at least three environments with an average R^2 of 13.01 %.

Comparative analysis of main-effect QTLs for seed size, shape and weight in the two-related populations

Regions on chromosomes with several identified M-QTLs for different studied seed phenotypic traits were designated as a QTL cluster. Accordingly, twenty-four QTL clusters located on 17 chromosomes with exception of Chr02, Chr12 and Chr18, were identified (Suppl. Table 10, Suppl. 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 3 clusters harbor QTLs for only seed shape traits (Suppl. Table 10). The majority of these clusters contained major QTLs. Furthermore, QTLs within 15 clusters revealed positive additive effects with the beneficial alleles inherited from the big seed size and heavy seed weight parents (either *Zhengyang* or *Linhefenqingdou*). Eight clusters out of 24 contain QTLs that have been detected in both populations (Suppl. Table 10). The most prominent M-QTL (*qFI-1-3_{LM6}*) with a LOD score of 3.71–10.44 and R^2 (10.45–31.50 %) was located in 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 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-3-1_{LM6}* and *qSL-3-2_{LM6}*), four QTLs for seed shape (*qSLW-3-2_{LM6}*, *qFI-3-1_{LM6}*, *qSLT-3-1_{LM6}*, and *qFI-3-2_{LM6}*) and one QTL HSW (*qHSW-3-1_{LM6}*). In addition, except for *qHSW-3-1_{LM6}*, all QTLs in this cluster were major QTLs with $R^2 > 10$ %. Furthermore, each of clusters-13, 16.2, and 17.1 contained five or six QTLs only related to seed size and HSW traits and displayed R^2 of 8.85–13.43 %, 5.96–11.26 %, and 6.8–18.30 %, respectively, and these clusters comprised M-QTLs from only one of the two RIL populations (Suppl. Figure 2, Suppl. Table 10). Another rich region of QTLs was cluster-20 on Chr20 with a physical length of 1.2Mb. This region harbored 5 M-QTLs related to seed size and shape, i.e., *qFI-20-1_{ZM6}*, *qSLT-20-1_{ZM6}*, *qSLW-20-1_{ZM6}*, *qST-20-1_{ZM6}*, and *qSW-20-1_{ZM6}*, out of these five QTLs, three are major QTLs with R^2 of 11.2–19.2% (Suppl. Table 10). Cluster-09 contained 5 QTLs related to seed size, shape, and HSW and displayed R^2 of 16.3 % and 12.5 % for *qHSW-9-1_{ZM6}* and *qSLW-9-3_{LM6}*, respectively, across the two populations. Cluster-14.1 consisted of four major M-QTLs within the physical region between 5834015-9844637bp in both populations with R^2 values ranged from 10.4-18.4%, one from which (*qSW-14-2_{ZM6}*) are related to seed size traits, and three (*qSLW-14-1_{LM6}*, *qFI-14-1_{LM6}*, and *qSLT-14-1_{LM6}*) for seed shape traits all were (Suppl. Table 10). Six clusters harbored 4 M-QTLs each, were identified, from which four clusters harbored QTLs associated with HSW as well as some seed size and seed shape traits, i.e., cluster-07 and cluster-19.1, cluster-08 and cluster-14.2 cluster-10.2 mapped on Chr10 and contained M-QTLs for seed size and seed shape traits, and cluster-16.1 that contained only M-QTLs related to seed shape traits (Suppl. Table 10). The remaining 9 clusters have three QTLs each, out of them cluster-01 and cluster-17.2 that contain major QTLs related only to seed shape traits. Conversely, cluster-04.1 and cluster-19.2 contain minor M-QTLs associated with SW, SL and HSW. Another two clusters contained M-QTLs for both seed size and seed shape traits (Suppl. Table 10). Moreover, the other three clusters of M-QTLs, i.e., cluster-10.1, cluster-11 and cluster-15 contained both major and minor QTLs for seed size traits and HSW. The Cluster-04.2 had two QTLs for two traits, *qHSW-4-3_{LM6, ZM6}*, and *qSL-4-1_{ZM6}* with R^2 of 13.1–17.7 %. Among the identified 24 clusters, seven clusters harbored QTLs related to seed size, seed shape, and HSW, five clusters harbored QTLs related to only seed size and seed shape traits, nine clusters had QTLs related to seed size and HSW traits, and three clusters harbored QTLs for only seed shape traits (Suppl. Table 10). The majority of these clusters had major QTLs. Furthermore, QTLs within 15 clusters revealed positive additive effects with the beneficial alleles inherited from the big seed size and heavy seed weight parents (either *Zhengyang* or *Linhefenqingdou*). Eight clusters out of twenty-four contained QTLs that have been detected in both populations (Suppl. Table 10).

Analyses of additive effect QTL and additive QTL × environment interactions for seed size, shape and weight.

The mixed-model based composite interval mapping (MCIM) method was implemented in QTL Network V2.1 software to map for additive effect (A) QTLs and their interactions with the environment (AE) was performed for both RIL populations across multi- environments. In total, thirty-five AA on 17 chromosomes related to seven seed size and seed shape traits were identified. These comprised 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). Furthermore, the contributed allele of 11 QTLs of them was inherited from the *M8206* parent that decreased seed size and seed shape values through significant additive effects. Meanwhile, the contributed allele of the remaining 24 QTLs descended from either *Zhengyang* or *Linhefenqingdou* parent of the ZM6 or LM6 population, respectively, that increased seed size and shape values through significant additive effects (Table 1). On the other hand, thirteen out of 35 QTLs revealed significant AE effects in at least one environment. However, five QTLs, i.e., *qSW-13-5_{LM6}*, *qSW-19-2_{LM6}*, *qFI-8-6_{LM6}*, *qSL-10-2_{ZM6}* and *qST-13-5_{ZM6}*, showed significant or highly significant AE among all studied environments (Table 1). Furthermore, 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 6 chromosomes, i.e., Chr03, Chr08, Chr09, Chr13, Chr14 and Chr16 in LM6 and ZM6 populations across six environments (Table 2). Six of those 8 QTLs displayed a positive additive effect with the beneficial allele inherited from the female parent (*Linhefenqingdou* or *Zhengyang* in the LM6 or ZM6 population) which could increase HSW. Meanwhile, the remaining two QTLs, i.e., *qHSW-13-3_{ZM6}* and *qHSW-14-2_{ZM6}*, revealed negative additive effects with the alleles are inherited from the common male parent (*Meng8206*) which could reduce HSW (Table 2). Additionally, two QTLs, i.e., *qHSW-14-3_{LM6}* and *qHSW-8-3_{ZM6}*, displayed significant AE effects in two individual environments. Whereas, the *qHSW-13-3_{ZM6}* showed a significant AE only in the 13JP environment. In addition, the *qHSW-14-4_{ZM6}* revealed significant or highly significant AE effects across three different environments, i.e., 12FY, 12JP, and 17JP (Table 2).

Table 1

Additive and additive × environment interaction effect of QTLs associated with seed size traits (SL, SW and ST) and seed shape (SLW, SLT, SWT and FI) traits in soybean seeds.

QTL ¹	Marker interval	Position (cM) ²	Physical position (bp)	Additive -Effect (A) ³		Additive × Environment Effect (AE) ⁴					Reference
				A	H ² %	AE1	AE2	AE3	AE4	H ² %	
<i>qSL-7-1</i> _{LM6}	bin744-bin745	15.25	3324836–3459470	0.16**	17.45	NS	NS	NS	NS	0.07	Hu et al. 2013
<i>qSL-13-6</i> _{LM6}	bin1535-bin1536	140.15	43244220–44026619	0.13**	6.08	NS	NS	NS	NS	0.28	Salas et al. 2006
<i>qSW-13-5</i> _{LM6}	bin1536-bin1537	143.04	43953331–44408971	0.51**	10.52	0.18*	-0.18*	0.16*	0.21**	0.15	Salas et al. 2006
<i>qSW-19-2</i> _{LM6}	bin2100-bin2101	42.57	34493194–34882495	0.18**	19.47	-0.13**	-0.11*	0.06*	0.09*	0.18	New
<i>qST-9-5</i> _{LM6}	bin1022-bin1023	51.88	7308659–7459924	0.12**	8.6	NS	NS	NS	NS	0.03	Salas et al. 2006
<i>qST-18-4</i> _{LM6}	bin1979-bin1980	43.69	9222099–10402370	0.06**	11.34	NS	NS	NS	NS	0.07	Fang et al. 2017
<i>qSLT-3-1</i> _{LM6}	bin247-bin248	19.49	3119582–3515594	0.12**	2.67	NS	-.22**	NS	NS	4.93	New
<i>qSLT-14-1</i> _{LM6}	bin1586-bin1587	42.38	7850227–8143522	-0.09**	1.96	NS	0.27**	NS	NS	6.73	Li et al. 2010
<i>qSLT-17-5</i> _{LM6}	bin1883-bin1884	70.26	13441932–13696232	-0.13**	2.41	NS	NS	NS	NS	1.05	New
<i>qSWT-7-5</i> _{LM6}	bin816-bin817	79.69	29822346–35034728	-0.67**	12.58	NS	NS	NS	NS	1.27	Fang et al. 2017
<i>qFI-3-3</i> _{LM6}	bin244-bin245	17.30	2790829–2980527	0.16**	11.48	NS	NS	NS	NS	0.13	New
<i>qFI-5-1</i> _{LM6}	bin476-bin477	0.21	1-529217	0.17**	21.22	NS	NS	NS	NS	0.15	New
<i>qFI-8-6</i> _{LM6}	bin954-bin955	95.88	35158414–37964850	0.17**	15.74	0.21*	-0.19*	0.14**	0.21*	0.11	New
<i>qFI-9-5</i> _{LM6}	bin1030-bin1031	56.45	20192294–27035074	-0.13**	4.27	NS	NS	NS	NS	0.21	New
<i>qFI-11-3</i> _{LM6}	bin1290-bin1291	69.93	18546688–18767705	0.1**	2.27	NS	NS	NS	NS	1.47	New
<i>qFI-16-1</i> _{LM6}	bin1745-bin1746	1.66	697999–908917	-0.08**	14.74	NS	NS	NS	NS	0.11	New
<i>qSL-1-4</i> _{ZM6}	bin4-bin5	3.23	754691–1375000	0.05**	3.33	NS	NS	NS	NS	0.08	New
<i>qSL-9-2</i> _{ZM6}	bin1174-bin1175	90.55	38507474–38736001	0.04*	2.06	NS	NS	0.07*	NS	2.29	New
<i>qSL-10-1</i> _{ZM6}	bin1236-bin1237	24.62	3150454–3297961	0.05**	13.67	NS	NS	NS	NS	1.4	New
<i>qSL-10-2</i> _{ZM6}	bin1334-bin1335	106.35	44226599–44378813	0.05**	5.57	-0.1**	0.06*	0.09**	-0.07*	3.58	Li et al. 2010
<i>qSL-12-4</i> _{ZM6}	bin1553-bin1554	97.99	38615116–38812896	-0.05**	3.17	NS	NS	NS	NS	0.83	New
<i>qSL-13-2</i> _{ZM6}	bin1612-bin1613	71.65	25830321–26065585	-0.08**	6.17	NS	NS	NS	NS	1.33	Fang et al. 2017
<i>qSL-15-5</i> _{ZM6}	bin1918-bin1919	85.59	17503517–17963129	0.14**	3.56	NS	NS	NS	NS	0.08	Salas et al. 2006
<i>qSW-8-5</i> _{ZM6}	bin959-bin960	73.74	11970511–12228336	0.04**	3	NS	NS	NS	NS	0.65	New
<i>qST-10-5</i> _{ZM6}	bin1334-bin1335	106.35	44226599–44378813	0.42**	3.99	NS	NS	NS	NS	3.07	Hu et al. 2013

Chr., chromosome. * p < 0.05; ** p < 0.01; NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parents Zhengyang and Linhefenqingdou while those with negative values show beneficial alleles from parent Meng 8206.H2 indicates phenotypic variation explained by additive effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; AE4, JP2017.

QTL ¹	Marker interval	Position (cM) ²	Physical position (bp)	Additive -Effect (A) ³		Additive × Environment Effect (AE) ⁴						Reference
				A	H ² %	AE1	AE2	AE3	AE4	H ² %		
<i>qST-10-6</i> _{ZM6}	bin1336-bin1337	107.17	44378814–44741960	-0.41**	2.25	NS	NS	0.15**	NS	3.17	New	
<i>qST-13-5</i> _{ZM6}	bin1609-bin1610	67.36	24985496–25641179	-0.41**	4.1	0.8**	-0.6*	-0.7**	0.61*	3.41	New	
<i>qST-14-3</i> _{ZM6}	bin1809-bin1810	104.68	47489495–47717306	0.04**	2.87	NS	NS	0.05**	NS	2.42	New	
<i>qST-20-1</i> _{ZM6}	bin2463-bin2464	4.37	662753–1045131	-0.06**	3.78	NS	NS	NS	NS	0.71	New	
<i>qSLW-9-4</i> _{ZM6}	bin1172-bin1173	89.06	38139739–38507473	0.99**	2.51	NS	NS	NS	NS	1.25	Li et al. 2010	
<i>qSLW-10-2</i> _{ZM6}	bin1275-bin1279	60.40	14218565–17808941	0.82**	12.74	NS	-0.92*	0.95*	NS	2.3	New	
<i>qSLW-13-5</i> _{ZM6}	bin1653-bin1654	102.63	32704220–33303066	-0.1**	1.15	NS	0.3**	NS	NS	3.95	Salas et al. 2006	
<i>qSLT-5-3</i> _{ZM6}	bin600-bin599	93.89	40328493–40882874	0.013**	11.18	NS	NS	NS	NS	0.32	Salas et al. 2006	
<i>qFI-17-6</i> _{ZM6}	bin2177-bin2178	130.75	41009636–41399912	0.06**	14.81	NS	NS	NS	NS	0.6	New	
<i>qFI-20-1</i> _{ZM6}	bin2461-bin2462	2.80	1- 662752	0.08**	5.45	NS	NS	0.12*	0.09*	0.28	New	

Chr., chromosome. * p < 0.05; ** p < 0.01; NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parents Zhengyang and Linhefenqingdou while those with negative values show beneficial alleles from parent Meng 8206.H2 indicates phenotypic variation explained by additive effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; AE4, JP2017.

Table 2
Additive and additive × environment interaction effect of QTLs associated with 100-seed weight trait in soybean seeds.

QTL ¹	Marker interval	Position (cM) ²	Physical position (bp)	Additive -Effect (A) ³		Additive × Environment Effect (AE) ⁴							Reference
				A	H ² %	AE1	AE2	AE3	AE4	AE5	AE6	H ² %	
<i>qHSW-3-2</i> _{LM6}	bin255-bin256	30.64	5833775–6780840	0.61**	10.75	NS	NS	NS	NS	NS	NS	0.16	Li et al. 2008
<i>qHSW-14-3</i> _{LM6}	bin1640-bin1641	101.33	48267526–48523627	0.48**	0.28	0.52*	NS	NS	NS	-0.16*	NS	1.99	New
<i>qHSW-8-3</i> _{ZM6}	bin963-bin964	80.64	12871276–13803222	0.29**	3.21	NS	NS	NS	0.34*	-0.41*	NS	0.42	Han et al. 2012
<i>qHSW-9-1</i> _{ZM6}	bin1162-bin1163	77.51	35758796–36561550	0.40**	2.47	NS	NS	NS	NS	NS	NS	0.88	New
<i>qHSW-13-3</i> _{ZM6}	bin1611-bin1612	69.99	25641180–26012595	-0.33**	5.7	NS	NS	-0.38*	NS	NS	NS	1.13	Funatsuki et al. 2005
<i>qHSW-14-2</i> _{ZM6}	bin1746-bin1747	28.67	4176245–4861311	-0.15**	1.66	NS	NS	NS	NS	NS	NS	0.17	New
<i>qHSW-14-4</i> _{ZM6}	bin1809-bin1810	104.68	47489495–47717306	0.46**	2.78	0.54**	NS	NS	0.43*	NS	-0.69**	4.75	New
<i>qHSW-16-3</i> _{ZM6}	bin2043-bin2044	103.81	35441262–35607069	0.21**	1.67	NS	NS	NS	NS	NS	NS	0.29	New

Chr., chromosome. * p < 0.05; ** p < 0.01; NS, non-significant. A indicates additive effects, those with positive values show beneficial alleles from parents Zhengyang and Linhefenqingdou while those with negative values show beneficial alleles from parent Meng 8206.H2 indicates phenotypic variation explained by additive effects. AE1, FY2012; AE2, JP2012; AE3, JP2013; AE4, JP2014; AE5, YC2014; AE6, JP2017.

Comparison Of 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 (Tables S7-S9). Meanwhile, forty-three QTLs were identified for seed size, shape and HSW by using MCIM approach (Tables 1 and 2). Among these, twenty-two QTLs were identified by both approaches within the same physical chromosomal position, indicating the dependability and stability of these QTLs. Moreover, a comparison of the physical chromosomal regions of the QTLs detected by both approaches revealed that four QTLs, i.e., *qSL-7-1_{LM6}*, *qSW-19-2_{LM6}*, *qFI-3-1_{LM6}*, and *qHSW-3-2_{LM6}*, were identified for the first time in the two populations (LM6 and ZM6) with an $R^2 > 10\%$. Therefore, we considered these QTLs as novel and most stable QTLs that could be validated and utilized for map-based cloning, candidate genes identification and QTL stacking into elite cultivars targeted at improving seed size, shape and HSW in soybean.

Analyses of epistatic-effect QTLs and their interaction with the environment

Analysis of the seed size and shape traits data under all four 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 ranged 0.51–11.35 % (Table 3). All QTL pairs displayed a high significant additive \times additive (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 ranged 0.13–5.31 % (Table 3). Furthermore, 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 indicates the effect of the environment on gene expression on phenotype development through epistatic effects. Out of 38 QTLs, sixteen pairwise interactions exhibited negative epistatic effects (AA) that decreased the values of seed size and shape traits, whereas 22 pairwise interactions exhibited positive epistatic effects (AA) that increased the values of seed size and shape traits (Table 3). The pairwise interaction between *qFI-1-1_{ZM6}* and *qFI-7-3_{ZM6}* revealed the strongest positive epistatic effect (0.65), whereas the pairwise *qSLT-6-1_{LM6}* and *qSLW-9-1_{LM6}* revealed the weakest positive epistatic effect (0.02). Conversely, *qSWT-3-1_{LM6}* and *qSWT-13-1_{LM6}* resulted in the strongest negative epistatic effect (-0.71), whereas the pairwise *qSLW-2-6_{ZM6}* and *qSLW-18-3_{ZM6}* pairwise resulted in the weakest negative epistatic effect (-0.02) (Table 3).

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 STW) across all environments.

RIL	Trait	QTL_i	Chr_i	Interval_i	Pos_i	QTL_j	Chr_j	Interval_j	Pos_j	Epistasis - Effect (AA)		Epistasis × Environment Effect (AAE)			
										AA	H ² %	AAE1	AAE2	AAE3	AAE4
LM6	SW	qSW-2-2 _{LM6}	2	bin124-bin125	23.92	qSW-16-2 _{LM6}	16	bin1757-bin1758	18.19	0.08**	7.24	NS	NS	NS	NS
		qSW-16-2 _{LM6}	3	bin229-bin230	1.96	qSW-13-1 _{LM6}	13	bin1509-bin1510	107.72	-0.1**	5.40	NS	NS	NS	NS
		qSW-4-2 _{LM6}	4	bin353-bin354	13.86	qSW-15-3 _{LM6}	15	bin1738-bin1739	106.86	-0.53**	2.85	NS	NS	NS	NS
		qSW-4-2 _{LM6}	4	bin353-bin354	13.86	qSW-15-4 _{LM6}	15	bin1740-bin1741	110.00	0.33**	7.30	0.11**	-0.11**	NS	NS
		qSW-5-1 _{LM6}	5	bin525-bin526	74.76	qSW-12-1 _{LM6}	12	bin1352-bin1353	21.59	-0.11**	3.96	NS	NS	NS	NS
		qSW-7-1 _{LM6}	7	bin784-bin785	53.84	qSW-15-4 _{LM6}	15	bin1740-bin1741	110.00	-0.10**	5.83	NS	NS	NS	NS
		qSW-8-1 _{LM6}	8	bin984-bin985	130.62	qSW-10-3 _{LM6}	10	bin1219-bin1220	97.61	-0.21**	3.20	-0.13*	0.12*	NS	NS
		qSW-10-1 _{LM6}	10	bin1184-bin1185	62.20	qSW-20-1 _{LM6}	20	bin2231-bin2232	69.64	-0.09**	2.20	NS	NS	NS	NS
		qSW-10-3 _{LM6}	10	bin1219-bin1220	97.61	qSW-16-4 _{LM6}	16	bin1810-bin1811	87.22	0.14**	1.47	NS	NS	NS	NS
		qSW-11-1 _{LM6}	11	bin1291-bin1292	70.56	qSW-15-1 _{LM6}	15	bin1715-bin1716	85.15	-0.17**	4.79	NS	NS	NS	NS
qSW-11-2 _{LM6}	11	bin1292-bin1296	71.58	qSW-15-2 _{LM6}	15	bin1717-bin1718	85.58	0.23**	9.22	NS	NS	NS	0.10**		
ST		qST-3-1 _{LM6}	3	bin237-bin238	8.12	qST-3-3 _{LM6}	3	bin344-bin345	114.06	0.05**	0.51	NS	NS	NS	NS
		qST-6-3 _{LM6}	6	bin647-bin648	109.3	qST-11-1 _{LM6}	11	bin1274-bin1275	54.45	0.1**	5.55	NS	NS	NS	0.12**
		qST-7-2 _{LM6}	7	bin749-bin750	19.81	qST-16-3 _{LM6}	16	bin1755-bin1756	17.12	0.09**	6.31	NS	NS	NS	NS
		qST-7-3 _{LM6}	7	bin783-bin784	53.32	qST-15-2 _{LM6}	15	bin1741-bin1742	110.5	-0.12**	5.74	NS	NS	NS	NS
		qST-16-1 _{LM6}	16	bin1744-bin1745	1.66	qST-17-2 _{LM6}	17	bin1886-bin1887	72.87	-0.14**	9.36	0.06*	-0.06*	NS	NS
SLW		qSLW-8-2 _{LM6}	8	bin941-bin942	86.24	qST-14-3 _{LM6}	14	bin1625-bin1626	84.91	0.03**	11.35	NS	NS	NS	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 show two loci genotypes being the same as those in parent Linhefenqingdou, Zhengyang (or Meng 8206) have the beneficial effects, while the two-loci recombinants take the negative effects. The case of negative value the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; AE4, JP2017.

SLT	<i>qSLT-5-3_{LM6}</i>	5	bin560-bin543	92.74	<i>qSLW-6-2_{LM6}</i>	6	bin625-bin626	82.38	-0.04**	5.40	0.03*	NS	NS	NS
	<i>qSLT-6-1_{LM6}</i>	6	bin584-bin585	28.72	<i>qSLW-9-1_{LM6}</i>	9	bin1095-bin1096	128.66	0.02**	1.61	-0.03*	0.03*	NS	NS
SWT	<i>qSWT-3-1_{LM6}</i>	3	bin236-bin237	7.79	<i>qSWT-13-1_{LM6}</i>	13	bin1435-bin1434	15.23	-0.71**	5.19	0.81**	-0.8**	NS	NS
	<i>qSWT-6-1_{LM6}</i>	6	bin588-bin589	33.52	<i>qSWT-18-1_{LM6}</i>	18	bin2036-bin2037	95.49	-0.62**	3.88	-0.64*	0.63*	NS	NS
	<i>qSWT-11-1_{LM6}</i>	11	bin1262-bin1263	40.91	<i>qSWT-20-1_{LM6}</i>	20	bin2177-bin2178	19.84	0.14**	8.03	-0.11**	0.15**	NS	NS
	<i>qSWT-16-1_{LM6}</i>	16	bin1744-bin1745	1.66	<i>qSWT-17-3_{LM6}</i>	17	bin1886-bin1887	72.87	0.11**	11.08	-0.15**	0.17**	NS	NS
FI	<i>qFI-1-5_{LM6}</i>	1	bin59-bin60	55.61	<i>qFI-14-4_{LM6}</i>	14	bin1627-bin1628	85.96	0.17**	2.89	NS	NS	0.09*	NS
	<i>qFI-5-2_{LM6}</i>	5	bin516-bin517	65.79	<i>qFI-10-2_{LM6}</i>	10	bin1223-bin1224	108.26	0.04**	10.02	NS	NS	NS	NS
	<i>qFI-16-2_{LM6}</i>	16	bin1745-bin1746	1.66	<i>qSWT-17-2_{LM6}</i>	17	bin1886-bin1887	38.76	0.02**	6.76	-0.18*	0.02*	NS	0.03*
SL	<i>qSL-12-4_{ZM6}</i>	12	bin1553-bin1554	97.99	<i>qSL-15-5_{ZM6}</i>	15	bin1919-bin1920	85.99	-0.06**	1.88	NS	NS	NS	NS
	<i>qSL-2-3_{ZM6}</i>	2	bin214-bin211	96.57	<i>qSL-8-6_{ZM6}</i>	8	bin1084-bin1085	186.90	0.2**	8.61	NS	NS	NS	NS
ZM6 SW	<i>qSW-4-4_{ZM6}</i>	4	bin434-bin435	62.42	<i>qSW-20-5_{ZM6}</i>	20	bin2590-bin2591	97.79	-0.12**	6.68	NS	NS	NS	0.07*
	<i>qSW-6-3_{ZM6}</i>	6	bin684-bin685	86.36	<i>qSW-6-5_{ZM6}</i>	6	bin703-bin704	109.84	0.05**	0.66	-0.08**	0.08**	NS	NS
ST	<i>qST-10-5_{ZM6}</i>	10	bin1334-bin1335	106.35	<i>qST-10-6_{ZM6}</i>	10	bin1336-bin1337	107.17	-0.7**	2.36	NS	NS	NS	NS
SLW	<i>qSLW-2-6_{ZM6}</i>	2	bin260-bin261	158.24	<i>qSLW-18-3_{ZM6}</i>	18	bin2336-bin2337	123.90	-0.02**	1.14	NS	0.04**	-0.03**	NS
SLT	<i>qSLT-1-4_{ZM6}</i>	1	bin53-bin54	43.95	<i>qSLT-7-2_{ZM6}</i>	7	bin884-bin885	103.56	0.03**	4.22	NS	-0.03*	0.03**	NS
SWT	<i>qSWT-1-3_{ZM6}</i>	1	bin62-bin63	47.73	<i>qSWT-8-1_{ZM6}</i>	8	bin914-bin915	15.98	0.13**	5.30	NS	NS	NS	NS
FI	<i>qFI-17-6_{ZM6}</i>	17	bin2177-bin2178	130.75	<i>qFI-20-1_{ZM6}</i>	20	bin2461-bin2462	2.79	0.51**	1.75	NS	NS	NS	NS
	<i>qFI-1-1_{ZM6}</i>	1	bin58-bin59	46.05	<i>qFI-7-3_{ZM6}</i>	7	bin884-bin885	103.56	0.65**	5.14	NS	NS	1.01**	NS
	<i>qFI-1-3_{ZM6}</i>	1	bin72-bin73	66.36	<i>qFI-7-1_{ZM6}</i>	7	bin872-bin873	96.62	0.3**	6.01	NS	NS	NS	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 show two loci genotypes being the same as those in parent Linhefenqingdou, Zhengyang (or Meng 8206) have the beneficial effects, while the two-loci recombinants take the negative effects. The case of negative value the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; AE4, JP2017.

<i>qFl-3-1_{ZM6}</i>	3	bin289- bin290	27.39	<i>qFl-18-2_{ZM6}</i>	18	bin2313- bin2314	112.03	0.12**	7.29	NS	0.9*	-0.86*	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 show two loci genotypes being the same as those in parent Linhefenqingdou, Zhengyang (or Meng 8206) have the beneficial effects, while the two-loci recombinants take the negative effects. The case of negative value the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, FY2012; AE2, JP2012; AE3, JP2014; AE4, JP2017.

Two digenic pairwise epistatic QTLs for HSW with highly significant additive × additive (AA) effects were identified on 4 chromosomes (Table 4). The first pairwise is composed of 2 QTLs, *qHSW-11-1_{LM6}* and *qHSW-20-1_{LM6}* with an R² of 3.46 %, whereas the second pairwise comprises the two QTLs *qHSW-9-1_{ZM6}* and *qHSW-16-3_{ZM6}* with an R² of 1.38 %. In addition, the two pairwise interactions exhibited positive epistatic effects that could increase the HSW in both populations. Meanwhile, the two pairs did not show any significant AAE interaction effects across all six environments (Table 4).

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	Epistasis - Effect (AA)				
										AA	H ² %	AAE1	AAE2	AAE3
<i>qHSW-11-1_{LM6}</i>	11	bin1245- bin1246	27.25	6135584- 6494224	<i>qHSW-20-1_{LM6}</i>	20	bin2175- bin2176	17.48	1272590- 1470471	0.51**	3.46	NS	NS	NS
<i>qHSW-9-1_{ZM6}</i>	9	bin1162- bin1163	77.51	35758796- 36561550	<i>qHSW-16-3_{ZM6}</i>	16	bin2043- bin2044	103.81	35441262- 35607069	0.34**	1.38	NS	NS	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 effects between two QTLs, those with positive values show two loci genotypes being the same as those in parent Linhefenqingdou, Zhengyang (or Meng 8206) while the two-loci recombinants take the negative effects. The case of negative values is the opposite. H² indicates phenotypic variation explained by epistatic effects. AE1, JP2012; AE2, JP2013; AE3, JP2014; AE4, JP2014; AE5, YC2014; AE6, JP2017.

Mining Major M-QTLs Clusters For Candidate Genes Identification

The 24 M-QTL clusters were filtered based on the richness in QTLs related to all or some of the seed size, shape and HSW or those with at least one QTL for 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 predict candidate genes based on publicly available databases such as SoyBase and Phytozome, and published papers. According to the physical intervals of the 7 QTL clusters, a total of 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, were retrieved from the SoyBase database (www.soybase.org; Suppl. Table 11). Gene ontology (GO) enrichment analyses via AgriGO V2.0 (<http://systemsbiology.cau.edu.cn>) (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, in addition to the response to stimulus in cluster-03 (Fig. 5). These indicate 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 analysis and the previously known putative biological function of the gene. Based on these, a total of 19, 12, 26, 18, 22, 30, and 16 candidate genes were identified within QTL clusters-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, respectively (Suppl. Table 12). These genes may work directly or indirectly in regulating seed development in soybean, which in turn regulating 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 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 expression of genes in soybean genome developed by (Severin et al. 2010) and publicly available on SoyBase was used to heatmap the expression of 19, 12, 26, 18, 22, 30, and 16 candidate genes were identified within QTL clusters-03, 04.1, 05.1, 07, 09, 17.1, and 19.1, respectively, in the young leaf, flower, pod, seed, root and nodule (Fig. 6, Suppl. Table 13). From the heatmaps, forty-seven genes out of the identified 143 candidate genes are highly expressed during seed developmental stages as well as in seed-related tissues (Fig. 6, Suppl. Table 13), hence these could be considered as potential candidate genes, however, they need screening and validation for utilization for seed size, shape, and weight improvement in soybean.

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	3005606	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	4798039	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
	Glyma04g03610	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	34678593	Nucleotide biosynthetic process; embryo development ending in seed dormancy
	Glyma05g29700	35236284	35242029	Brassinosteroid biosynthetic process; starch biosynthetic process
	Glyma05g30380	35754306	35755603	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	38341410	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	13903037	13906228	Embryo development ending in seed dormancy
	Glyma07g15050	14900705	14909235	Seed dormancy process; regulation of cell cycle process
	Glyma07g15640	15378798	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	35573357	35579018	Embryo development ending in seed dormancy; cellular response to abscisic acid stimulus
	Glyma09g29030	35989729	35993075	Ubiquitin-dependent protein catabolic process; fatty acid beta-oxidation
	Glyma09g29720	36540972	36548174	Response to auxin stimulus; auxin metabolic process
	Glyma09g30130	37014420	37023261	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	39100482	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	7186015	Seed maturation; protein ubiquitination; lipid storage
	Glyma17g10290	7707775	7711360	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

QTL Clusters	Gene	Start	Stop	Gene Functional Annotation
	Glyma17g12950	9873806	9891306	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	40666918	40669847	Glucose catabolic process; response to auxin stimulus
	Glyma19g33620	41194146	41196743	Maltose metabolic process; starch biosynthetic process; glucosinolate biosynthetic process
	Glyma19g33650	41237306	41242657	Abscisic acid biosynthetic process; plant-type cell wall modification; pollen tube growth

Discussion

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 effects of 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 objective 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 (Abou-Elwafa 2016; Bernardo 2008). 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 50–250 individuals, but larger populations are needed for high-resolution mapping. Moreover, a high-density genetic map facilitates the detection of narrow linked markers associated with QTLs and provide a good base for investigating quantitative traits (Galal et al. 2014; Mohan et al. 1997; Tewodros and Zelalem 2016). Besides, 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 classified to be associated with HSW, however, most of the studies utilized low-density genetic maps based on RFLP, SSR markers, biochemical and morphological markers which have large confidence interval with low resolution of QTLs not suitable for detecting candidate gene (Abou-Elwafa 2016; Bernardo 2008; Han et al. 2012). Therefore, it is crucial to employ high-density genetic maps to detect more new recombination in a population, which in turn will increase the accuracy of QTL mapping and MAS (Cao et al. 2019; Hina et al. 2020).

In the present study, high-density genetic maps constructed from the two-related RIL populations LM6 and ZM6 consist of 2267 and 2601 bin markers, respectively, were used. The markers in the LM6 and ZM6 linkage maps were distributed to all 20 linkage groups and covered the length of 2453.79 and 2630.22 cM, with 1.08 cM and 1.01 cM average distance between adjacent markers, respectively (Li et al. 2017). To minimize the environmental errors, the two RIL populations were evaluated in four environments (including different geographical areas and years). The high-density linkage maps of LM6 and ZM6 RIL populations across multi-environment as well as the combined environment were employed to map major main-effect, additive-effect and epistatic-effect QTLs together with interactions with environments and the candidate genes underly seed size, shape and seed weight traits. The parents of the two mapping populations showed high phenotypic variations across all environments in all studied traits, i.e., SL, SW, ST, SLW, SLT, SWT, FI and HSW. Consequently, 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 but also there is an effect of the G×E interaction (Hu et al. 2013; Liang et al. 2016; Sun et al. 2012). This explains the observed high h^2 (99.04 %), and accordingly deduced that these traits are amenable to manipulation by selection without the assistance of molecular markers, indicating that these traits may produce the same phenotypic values when evaluated in the same geographical area. 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).

Comparative QTL results using the CIM QTL mapping approach with SoyBase database identified 69, 82 and 29 novel QTLs for seed size, shape, and HSW, respectively, indicating the distinct genetic architecture of the LM6 and ZM6 populations. These novel QTLs together explain more than 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 needed to be used for unraveling the complex genetic basis for seed size and shape traits in soybean. Among these novel QTLs, the *qFI-1-3_{LM6}* showed the highest R^2 and LOD values and therefore may be the major QTL underlies flatness index (FI).

Noteworthy, all FI QTLs were identified for the first time in this study, thus we considered them as novel QTLs. Chr01 and Chr03 harbored 4 and 3 FI QTLs, suggesting crucial roles of Chr01 and Chr03 in controlling the inheritance of seed FI in soybean. Round soybean seeds are required for soybean varieties used for food. Furthermore, the data revealed that the average flatness indices across all environments in both the LM6 and ZM6 populations have sphere seeds ($FI \approx 1.0$), indicating that it is essential to start with at least one round seeded parent to get a segregant with round seed shape. Remarkably, eight novel major QTLs for HSW, including *qHSW-4-1_{LM6}*, *qHSW-6-2_{LM6}*, *qHSW-7-3_{LM6}*, *qHSW-9-1_{ZM6}*, *qHSW-10-2_{LM6}*, *qHSW-10-3_{LM6}*, *qHSW-14-1_{LM6, ZM6}* and *qHSW-14-2_{ZM6}* 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 future fine mapping to delimit the physical interval. The *qHSW-4-3_{LM6, ZM6}* was detected in the physical interval of 42.7–45.8 Mb on Chr04 that overlapped with the previously identified seed-weight QTLs, i.e., *seed weight 47 - 1* and *seed wtQTL4.1* (Hacisalihoglu et al. 2018; Li et al. 2010). Nine QTLs for SL identified in this

study were colocalized as previously reported (Hu et al. 2013; Jun et al. 2014; Salas et al. 2006). Two major QTLs related to SW, i.e., *qSW-4-4_{ZM6}* and *qSW-6-1_{LM6}*, both residing a genomic sequence of approximate 128 kb colocalized with the previously identified SW QTL *A063-1* (Salas et al. 2006) and SW QTL detected previously in another independent soybean populations (Hina et al. 2020; Moongkanna et al. 2011), respectively. Thus, *qSW-4-4_{ZM6}* and *qSW-6-1_{LM6}* could be considered as stable QTLs for further fine mapping and map-based cloning to clarify the genetic mechanisms underlying SW. The *qSLW-5-1_{LM6}* overlapped with *qSLW-5-2* (Satt449) which was previously reported by (Jun et al. 2014). Furthermore, the three QTLs associated with SLT (*qSLT-5-3_{ZM6}*, *qSLT-16-1_{LM6}*, and *qSLT-20-1_{ZM6}*) are colocalized with previously reported QTLs (Fang et al. 2017; Jun et al. 2014; Salas et al. 2006). The SWT QTL between *Satt508* and *Satt421* on Chr08 that have been previously mapped (Salas et al. 2006) are colocalized to the *qSWT-8-1_{LM6}* QTL that was mapped to the physical interval 93–96.3 cM of Chr08. Additionally, our study identified for the first time 13 major QTLs ($R^2 > 10\%$) related to FI, thus we considered them as novel QTLs. Among these novel QTLs, the *qFI-1-3_{LM6}* showed the highest R^2 and LOD values and therefore might be the major QTL underlies FI. Besides, Chr01 and Chr03 harbored 4 and 3 FI QTLs, suggesting crucial roles of Chr01 and Chr03 in controlling the inheritance of seed FI in soybean. Moreover, the 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 heavy weight parent (*Linhafenqingdou* 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 method approach identified 18 QTLs for seed sizes, shapes, and weight traits that are colocalized in the same physical interval of the CIM-mapped QTLs as previous studies. The major SL QTL *qSL-7-1_{LM6}* is colocalized with *Satt150* QTL (Salas et al. 2006). Furthermore, the SW QTLs *qSW-13-5_{LM6}*, *qST-18-4_{LM6}*, *qSWT-7-5_{LM6}*, and *qSLT-5-3_{ZM6}* were mapped in the same position as reported in previous studies (Fang et al. 2017; Salas et al. 2006). Additionally, the *qHSW-3-2_{LM6}*, *qHSW-8-3_{ZM6}* and *qHSW-13-3_{ZM6}* QTLs are colocalized to the previously identified SW QTL *Seed weight 32 - 3* (*Satt675*), *Seed weight 35 - 1* and *Seed weight 19 - 2* (*Satt114*) QTLs, respectively (Funatsuki et al. 2005; Han et al. 2012; Li et al. 2008). Therefore, these QTLs could also be considered as 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 greatly contributed to the phenotypic variations of complex traits (Kaushik et al. 2007). The genetic construction of seed size, shape, and weight also contains epistatic interactions between QTLs (Kato et al. 2014; Liang et al. 2016). Therefore, disregarding intergenic interactions will lead to the over-estimation of individual QTL effects, and the under-estimation of genetic variance (Nyquist and Baker 1991). Consequently, this might result in a large drop in the genetic response to MAS especially in late generations (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; Zhang et al. 2018). The appearance of epistatic interactions for a specific trait makes selection difficult. Noteworthy, 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 a large number of QTLs related to all or some of seed size, shape and HSW. The identification of 24 QTL clusters located on 17 different chromosomes. Accordingly, twenty-four QTL clusters were identified on 17 chromosomes each contained three or more QTLs related to seed size, shape, and HSW traits. These QTL clusters have not been reported previously and added to the developing knowledge of the genetic control of these traits. Moreover, the colocalization of QTLs for seed size, shape, and HSW and the way that they have exceptionally corresponded support the highly significant correlation with each other (Cai and Morishima 2002) (Suppl. 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 (Cao et al. 2017; Liu et al. 2017; Wang et al. 2006). Furthermore, the QTL clusters displayed that the QTLs linkage/ gathering could make the enhancement of seed size and shape more easily than single QTLs (Hina et al. 2020). Significant positive correlations between soybean seed protein and oil contents and seed size and seed shape have been demonstrated, therefore, both traits are directly associated with seed size and shape in soybean (Hacisalihoglu and Settles 2017; Qi et al. 2011; Wu et al. 2018). This notion would explain the colocalization of QTLs associated with seed protein and oil contents in the genomic regions of several QTL clusters including clusters 1, 04.1, 04.2, 06, 07, 09, 10.1, 14.1, 14.2, 17.1, and 17.2. (Moongkanna et al. 2011; Panthee et al. 2005; Salas et al. 2006; Vieira et al. 2006; Yang et al. 2011). Additionally, the position of the first flower and the number of days to flowering have large effects on seed number per plant in soybean (Khan et al. 2008; Tasma et al. 2001; Yamanaka et al. 2001), which in turn 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 (Han et al. 2012; Tasma et al. 2001), are located to the genomic region of clusters 16.1, 19.2, and 20 (Hyten et al. 2004). The extensive analysis of QTLs 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). So far, only two seed sizes/weight-related genes have been cloned from the soybean, i.e., the *Glyma20g25000* (*ln*) gene that has a significant impact on seed size and the number of seeds per pod (Jeong et al. 2012), and the PP2C-1 allele underlying *Glyma17g33690* has been reported to increase seed size/weight (Lu et al. 2017). In our study, a bioinformatics pipeline implementing genomic sequences of identified QTL clusters was employed to identify candidate genes. The pipeline consists of three complementary steps, i.e., 1) retrieving candidate genes from the SoyBase database (www.soybase.org) visualize the molecular function of candidate genes by GO enrichment analyses and gene classification, and 3) the implication of candidate genes in seed size, shape and weight based on their expression profiles. Accordingly, one-hundred forty-three genes were considered as potential candidates. The GO enrichment and gene classification analyses showed that most of the identified candidate genes behind QTL clusters are related to the terms of catalytic activity, cell part, cell,

cellular process, binding and metabolic process terms in addition to the response to stimulus in cluster-03, and these terms are reported as being vital elements in seed development (Li and Li 2014; Mao et al. 2010). For example, *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). These predicted 143 genes have functions that are related/involved in seed development, which in turn influences the size, shape, and weight of seeds, such as brassinosteroid mediated signaling pathway, regulation of cell differentiation and proliferation, fatty acid beta-oxidation, peroxisome organization, double fertilization forming a zygote and endosperm, lipid transport and storage, regulation of hormone levels transport and metabolic processes, ubiquitin-dependent protein catabolic process, and sugar mediated signaling pathway (Li and Li 2014; Mao et al. 2010). Furthermore, ten 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). Additionally, ten candidate genes are involved in response to auxin stimulus, response to ethylene stimulus, abscisic acid biosynthetic process that are known to be implicated in promoting seed size and weight in Arabidopsis (Table 5) (Xie et al. 2014). Furthermore, six genes are known to play functions in glucose catabolic process, phosphatidylcholine biosynthetic process, carbohydrate metabolic process, maltose metabolic process and starch biosynthetic process which some of them are known to be involved in partitioning and translocation of photo-assimilates and grain filling in rice (Table 5) (Chen et al. 2020b; Zhang et al. 2020).

Conclusions

The present study employed high-density maps of two-related RIL populations, LM6 and ZM6 evaluated in multiple environments to identify M-QTLs as well candidate genes controlling seed size, shape, and weight in soybean. Besides, this is the first detailed and comprehensive investigation of QTLs for flatness index as a seed shape trait in soybean. A total of 180 and 18 M-QTLs were reported for the first time in this study using the CIM and MCIM QTL mapping approaches, respectively. Besides, sixty-nine QTLs were considered as stable as detected in more than one specific environment or one individual environment together with CE. All the positive alleles of 282 identified QTLs were inherited from the female parents. Our data revealed 7 major and stable QTL clusters underlying the inheritance of seed size, shape and weight located to genomic regions on chromosomes 3, 4, 5, 7, 9, 17 and 19 in soybean. The implemented bioinformatics pipeline delimits the number of the identified candidate genes to 47 genes within the physical interval of the previously mentioned 7 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 the above traits in soybean. Furthermore, some of the potential 47 candidate genes have been included in our on-going 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 as well as for MAS targeted at improving seed size, shape and weight.

Declarations

Author Contributions

T.Z. designed the project. M.A.E. performed the experiments. M.A.E., B.K., S.S., S.L., Y.C., M.A. and A.H. analyzed the data. M.A.E. drafted the manuscript. T.Z. and S.F.A. revised the paper. 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 (31871646), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT_17R55), the Fundamental Research Funds for the Central Universities (KYT201801), the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIMCPC) Program.

Conflicts of Interest

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Availability of data and material

All data are included within the manuscript and its supplementary material.

Code availability

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

References

- Abou-Elwafa SF (2016) Association mapping for drought tolerance in barley at the reproductive stage. *Cr Biol* 339:51-59
- Abou-Elwafa SF (2018) Identification of genes associated with drought tolerance in barley. *Biologia Plantarum* 62:299-306
- Abou-Elwafa SF, Shehzad T (2018) Genetic identification and expression profiling of drought responsive genes in sorghum. *Environmental and Experimental Botany* 155:12-20
- Ainsworth EA, Yendrek CR, Skoneczka JA, Long SP (2012) Accelerating yield potential in soybean: potential targets for biotechnological improvement. *Plant, cell & environment* 35:38-52
- Basra AS (1995) Seed quality: basic mechanisms and agricultural implications
- Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop science* 48:1649-1664
- Burris J, Edje O, Wahab A (1973) Effects of Seed Size on Seedling Performance in Soybeans: II. Seedling Growth and Photosynthesis and Field Performance 1. *Crop Science* 13:207-210
- Cai H, Morishima H (2002) QTL clusters reflect character associations in wild and cultivated rice. *Theoretical and Applied Genetics* 104:1217-1228
- Cailleux A (1945) Distinction des galets marins et fluviaux. *Bulletin de la Société géologique de France* 5:375-404
- Cao Y, Li S, Chen G, Wang Y, Bhat JA, Karikari B, Kong J, Gai J, Zhao T (2019) Deciphering the Genetic Architecture of Plant Height in Soybean Using Two RIL Populations Sharing a Common M8206 Parent. *Plants* 8:373
- Cao Y, Li S, Wang Z, Chang F, Kong J, Gai J, Zhao T (2017) Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Frontiers in plant science* 8:1222
- Cerdà A, Garcia-Fayos P (2002) The influence of seed size and shape on their removal by water erosion. *Catena* 48:293-301
- Cha-um S, Kirdmanee C (2011) Remediation of salt-affected soil by the addition of organic matter: an investigation into improving glutinous rice productivity. *Scientia Agricola* 68:406-410
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020a) TBtools-an integrative toolkit developed for interactive analyses of big biological data. *bioRxiv*:289660
- Chen J, Cao F, Li H, Shan S, Tao Z, Lei T, Liu Y, Xiao Z, Zou Y, Huang M, Abou-Elwafa SF (2020b) Genotypic variation in the grain photosynthetic contribution to grain filling in rice. *Journal of Plant Physiology* 253:153269
- Cheng X, Wang S, Wang L, Wang S, Zong X, Wang P (2006) Descriptors and Data Standard for Adzuki Bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. China Agriculture Press, Beijing
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971
- Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM (2011) Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC genomics* 12:4
- Cober E, Voldeng H, Frégeau-Reid J (1997) Heritability of seed shape and seed size in soybean. *Crop Science* 37:1767-1769
- Ebone LA, Caverzan A, Tagliari A, Chiomento JLT, Silveira DC, Chavarria G (2020) Soybean Seed Vigor: Uniformity and Growth as Key Factors to Improve Yield. *Agronomy* 10:545
- Edwards Jr CJ, Hartwig EE (1971) Effect of seed size upon rate of germination in soybeans 1. *Agronomy Journal* 63:429-450
- Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Hu G, Zhou Z, Yu H, Zhang M (2017) Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome biology* 18:1-14
- Funatsuki H, Kawaguchi K, Matsuba S, Sato Y, Ishimoto M (2005) Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. *Theoretical and Applied Genetics* 111:851-861
- Galal A, Sharma S, Abou-Elwafa SF, Sharma S, Kopisch-Obuch F, Laubach E, Perovic D, Ordon F, Jung C (2014) Comparative QTL analysis of root lesion nematode resistance in barley. *Theor Appl Genet* 127:1399-1407
- Hacisalihoglu G, Burton AL, Gustin JL, Eker S, Asikli S, Heybet EH, Ozturk L, Cakmak I, Yazici A, Burkey KO (2018) Quantitative trait loci associated with soybean seed weight and composition under different phosphorus levels. *Journal of integrative plant biology* 60:232-241

Hacisalihoglu G, Settles AM (2017) Quantification of seed ionome variation in 90 diverse soybean (*Glycine max*) lines. *Journal of Plant Nutrition* 40:2808-2817

Haig D (2013) Kin conflict in seed development: an interdependent but fractious collective. *Annual review of cell and developmental biology* 29:189-211

Han Y, Li D, Zhu D, Li H, Li X, Teng W, Li W (2012) QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theoretical and Applied Genetics* 125:671-683

Hina A, Cao Y, Song S, Li S, Sharmin RA, Elattar MA, Bhat JA, Zhao T (2020) High-Resolution Mapping in Two RIL Populations Refines Major "QTL Hotspot" Regions for Seed Size and Shape in Soybean (*Glycine max* L.). *International Journal of Molecular Sciences* 21:1040

Hoeck JA, Fehr WR, Shoemaker RC, Welke GA, Johnson SL, Cianzio SR (2003) Molecular marker analysis of seed size in soybean. *Crop Science* 43:68-74

Hu Z, Zhang H, Kan G, Ma D, Zhang D, Shi G, Hong D, Zhang G, Yu D (2013) Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* 141:247-254

Hyten D, Pantalone V, Sams C, Saxton A, Landau-Ellis D, Stefaniak T, Schmidt M (2004) Seed quality QTL in a prominent soybean population. *Theoretical and Applied Genetics* 109:552-561

Jannink J-L, Moreau L, Charmet G, Charcosset A (2009) Overview of QTL detection in plants and tests for synergistic epistatic interactions. *Genetica* 136:225

Jeong N, Suh SJ, Kim M-H, Lee S, Moon J-K, Kim HS, Jeong S-C (2012) Ln is a key regulator of leaflet shape and number of seeds per pod in soybean. *The Plant Cell* 24:4807-4818

Jun TH, Freewalt K, Michel AP, Mian R (2014) Identification of novel QTL for leaf traits in soybean. *Plant Breeding* 133:61-66

Kajiya-Kanegae H, Takanashi H, Fujimoto M, Ishimori M, Ohnishi N, Fiona WW, Omollo EA, Kobayashi M, Yano K, Nakano M (2020) RAD-seq-Based High-Density Linkage Map Construction and QTL Mapping of Biomass-Related Traits in Sorghum Using a Japanese Landrace Takakibi NOG. *Plant and Cell Physiology*

Karikari B, Li S, Bhat JA, Cao Y, Kong J, Yang J, Gai J, Zhao T (2019) Genome-wide detection of major and epistatic effect QTLs for seed protein and oil content in soybean under multiple environments using high-density bin map. *International journal of molecular sciences* 20:979

Kato S, Sayama T, Fujii K, Yumoto S, Kono Y, Hwang T-Y, Kikuchi A, Takada Y, Tanaka Y, Shiraiwa T (2014) A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theoretical and Applied Genetics* 127:1365-1374

Kaushik N, Kumar K, Kumar S, Kaushik N, Roy S (2007) Genetic variability and divergence studies in seed traits and oil content of *Jatropha* (*Jatropha curcas* L.) accessions. *Biomass and Bioenergy* 31:497-502

Khan NA, Githiri SM, Benitez ER, Abe J, Kawasaki S, Hayashi T, Takahashi R (2008) QTL analysis of cleistogamy in soybean. *Theoretical and Applied Genetics* 117:479-487

Kundu A, Chakraborty A, Mandal NA, Das D, Karmakar PG, Singh NK, Sarkar D (2015) A restriction-site-associated DNA (RAD) linkage map, comparative genomics and identification of QTL for histological fibre content coincident with those for retted bast fibre yield and its major components in jute (*Corchorus olitorius* L., Malvaceae sl). *Molecular Breeding* 35:19

Li D, Sun M, Han Y, Teng W, Li W (2010) Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*). *Euphytica* 172:49-57

Li N, Li Y (2014) Ubiquitin-mediated control of seed size in plants. *Frontiers in plant science* 5:332

Li S, Cao Y, He J, Zhao T, Gai J (2017) Detecting the QTL-allele system conferring flowering date in a nested association mapping population of soybean using a novel procedure. *Theoretical and Applied Genetics* 130:2297-2314

Li W, Zheng D-H, Van K, 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, 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. *Journal of genetics* 95:475-477

Liu D, Yan Y, Fujita Y, Xu D (2018) Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breeding science* 68:442-448

Liu N, Li M, Hu X, Ma Q, Mu Y, Tan Z, Xia Q, Zhang G, Nian H (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 genomics* 18:466

Lu X, Xiong Q, Cheng T, Li Q-T, Liu X-L, Bi Y-D, Li W, Zhang W-K, Ma B, Lai Y-C (2017) A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Molecular Plant* 10:670-684

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

Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proceedings of the National Academy of Sciences* 107:19579-19584

Meng Y, Chen F, Shuai H, Luo X, Ding J, Tang S, Xu S, Liu J, Liu W, Du J (2016) Karrikins delay soybean seed germination by mediating abscisic acid and gibberellin biogenesis under shaded conditions. *Scientific reports* 6:1-12

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research* 17:240-248

Mohan M, Nair S, Bhagwat A, Krishna T, Yano M, Bhatia C, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular breeding* 3:87-103

Moongkanna J, Nakasathien S, Novitzky W, Kwanyuen P, Sinchaisri P, Srinives P (2011) SSR markers linking to seed traits and total oil content in soybean. *Thai Journal of Agricultural Science* 44:233-241

Morton NE (1955) Sequential tests for the detection of linkage. *American journal of human genetics* 7:277

Nyquist WE, Baker R (1991) Estimation of heritability and prediction of selection response in plant populations. *Critical reviews in plant sciences* 10:235-322

Omokhafa K, Alika J (2004) Clonal variation and correlation of seed characters in *Hevea brasiliensis* Muell. Arg. *Industrial Crops and Products* 19:175-184

Palomeque L, Li-Jun L, Li W, Hedges B, Cober ER, Rajcan I (2009) QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted× high-yielding exotic soybean lines. *Theoretical and Applied Genetics* 119:429-436

Palomeque L, Liu L-J, Li W, Hedges BR, Cober ER, Smid MP, Lukens L, Rajcan I (2010) Validation of mega-environment universal and specific QTL associated with seed yield and agronomic traits in soybeans. *Theoretical and applied genetics* 120:997-1003

Pan L, Wang N, Wu Z, Guo R, Yu X, Zheng Y, Xia Q, Gui S, Chen C (2017) A high density genetic map derived from RAD sequencing and its application in QTL analysis of yield-related traits in *Vigna unguiculata*. *Frontiers in plant science* 8:1544

Panthee D, Pantalone V, West D, Saxton A, Sams C (2005) Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Science* 45:2015-2022

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS one* 7:e37135

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS one* 7:e32253

Qi Z-m, Wu Q, Han X, Sun Y-n, Du X-y, Liu C-y, Jiang H-w, Hu G-h, Chen Q-s (2011) Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica* 179:499-514

Salas P, Oyarzo-Llaipen J, Wang D, Chase K, Mansur L (2006) Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics* 113:1459-1466

Severin AJ, Woody JL, Bolon Y-T, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE (2010) RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC plant biology* 10:160

Smith T, Camper Jr H (1975) Effects of seed size on soybean performance. *Agronomy Journal* 67:681-684

Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS one* 8:e58700

Sun Y-n, Pan J-b, Shi X-l, Du X-y, Wu Q, Qi Z-m, Jiang H-w, Xin D-w, Liu C-y, Hu G-h (2012) Multi-environment mapping and meta-analysis of 100-seed weight in soybean. *Molecular Biology Reports* 39:9435-9443

Tao Y, Mace ES, Tai S, Cruickshank A, Campbell BC, Zhao X, Van Oosterom EJ, Godwin ID, Botella JR, Jordan DR (2017) Whole-genome analysis of candidate genes associated with seed size and weight in sorghum bicolor reveals signatures of artificial selection and insights into parallel domestication in cereal crops. *Frontiers in plant science* 8:1237

Tasma I, Lorenzen L, Green D, Shoemaker R (2001) Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Molecular Breeding* 8:25-35

Teng W, Feng L, Li W, Wu D, Zhao X, Han Y, Li W (2017) Dissection of the genetic architecture for soybean seed weight across multiple environments. *Crop and Pasture Science* 68:358-365

- Teng W, Han Y, Du Y, Sun D, Zhang Z, Qiu L, Sun G, Li W (2009) QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). *Heredity* 102:372-380
- Tewodros M, Zelalem B (2016) Advances in Quantitative Trait Loci, mapping and importance of Markers Assisted Selection in plant breeding research. *International Journal of Plant Breeding and Genetics* 10:58-68
- Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z (2017) agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic acids research* 45:W122-W129
- Tomooka N, Vaughan D, Maxted N, Moss H (2002) The Asian Vigna: genus *Vigna* subgenus *ceratotropis* genetic resources. Springer Science & Business Media
- Vieira AJD, Oliveira DAd, Soares TCB, Schuster I, Piovesan ND, Martínez CA, Barros EGd, Moreira MA (2006) Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Brazilian Journal of Plant Physiology* 18:281-290
- Wang C, Rutledge J, Gianola D (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics Selection Evolution* 26:91-115
- Wang D, Bales-Arcelo C, Zhang Z, Gu C, DiFonzo CD, Zhang G, Yang Z, Liu M, Mensah C (2019) Sources of aphid resistance in soybean plants. Google Patents
- Wang S, Basten C, Zeng Z (2006) Windows QTL Cartographer 2.5. Dep. of Statistics, North Carolina State Univ., Raleigh. Windows QTL cartographer 25 Dep of Statistics, North Carolina State Univ, Raleigh:-
- Wu D, Zhan Y, Sun Q, Xu L, Lian M, Zhao X, Han Y, Li W (2018) Identification of quantitative trait loci underlying soybean (*Glycine max* [L.] Merr.) seed weight including main, epistatic and QTL× environment effects in different regions of Northeast China. *Plant Breeding* 137:194-202
- Wu Y, Gong W, Yang W (2017) Shade inhibits leaf size by controlling cell proliferation and enlargement in soybean. *Scientific reports* 7:1-10
- Xie F-T, Niu Y, Zhang J, Bu S-H, Zhang H-Z, Geng Q-C, Feng J-Y, Zhang Y-M (2014) Fine mapping of quantitative trait loci for seed size traits in soybean. *Molecular Breeding* 34:2165-2178
- Xie M, Ming Y, Shao F, Jian J, Zhang Y, Peng Z (2018) Restriction site-associated DNA sequencing for SNP discovery and high-density genetic map construction in southern catfish (*Silurus meridionalis*). *Royal Society open science* 5:172054
- Xin D, Qi Z, Jiang H, Hu Z, Zhu R, Hu J, Han H, Hu G, Liu C, Chen Q (2016) QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. *PLoS one* 11:e0149380
- Xing G, Zhou B, Wang Y, Zhao T, Yu D, Chen S, Gai J (2012) Genetic components and major QTL confer resistance to bean pyralid (*Lamprosema indicata* Fabricius) under multiple environments in four RIL populations of soybean. *Theoretical and Applied Genetics* 125:859-875
- Xu Y, Li H-N, Li G-J, Wang X, Cheng L-G, Zhang Y-M (2011) Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics* 122:581-594
- Yamanaka N, Ninomiya S, Hoshi M, Tsubokura Y, Yano M, Nagamura Y, Sasaki T, Harada K (2001) An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. *DNA research* 8:61-72
- Yang BZ, Zhao H, Kranzler HR, Gelernter J (2005) Practical population group assignment with selected informative markers: characteristics and properties of Bayesian clustering via STRUCTURE. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* 28:302-312
- Yang J, Hu C, Hu H, Yu R, Xia Z, Ye X, Zhu J (2008) QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24:721-723
- Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. *Bioinformatics* 23:1527-1536
- Yang K, Moon J-K, Jeong N, Chun H-K, Kang S-T, Back K, Jeong S-C (2011) Novel major quantitative trait loci regulating the content of isoflavone in soybean seeds. *Genes & Genomics* 33:685-692
- Yao D, Wang P, Zhang J, Liu Z, Guan S, Liu S, Qu J (2014) A QTL mapping analysis of main yield traits in soybean. *Journal of South China Agricultural University* 35:41-46
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L (2006) WEGO: a web tool for plotting GO annotations. *Nucleic acids research* 34:W293-W297
- Zeng Z-B (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468
- Zhang F, Kang J, Long R, Yu L-X, Wang Z, Zhao Z, Zhang T, Yang Q (2019a) High-density linkage map construction and mapping QTL for yield and yield components in autotetraploid alfalfa using RAD-seq. *BMC plant biology* 19:165

Zhang H, Chen J, Shan S, Cao F, Chen G, Zou Y, Huang M, Abou-Elwafa SF (2020) Proteomic profiling reveals differentially expressed proteins associated with amylose accumulation during rice grain filling. *BMC Genomics* 21:714

Zhang W-K, Wang Y-J, Luo G-Z, Zhang J-S, He C-Y, Wu X-L, Gai J-Y, Chen S-Y (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theoretical and Applied Genetics* 108:1131-1139

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

Zhang Y, Li W, Lin Y, Zhang L, Wang C, 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 genomics* 19:641

Zhao B, Dai A, Wei H, Yang S, Wang B, Jiang N, Feng X (2016) Arabidopsis KLU homologue GmCYP78A72 regulates seed size in soybean. *Plant molecular biology* 90:33-47

Zhao F, Xu S (2012) Genotype by environment interaction of quantitative traits: a case study in barley. *G3: Genes, Genomes, Genetics* 2:779-788

Zhaoming Q, Xiaoying Z, Huidong Q, Dawei X, Xue H, Hongwei J, Zhengong Y, Zhanguo Z, Jinzhu Z, Rongsheng Z (2017) Identification and validation of major QTLs and epistatic interactions for seed oil content in soybeans under multiple environments based on a high-density map. *Euphytica* 213:162

Figures

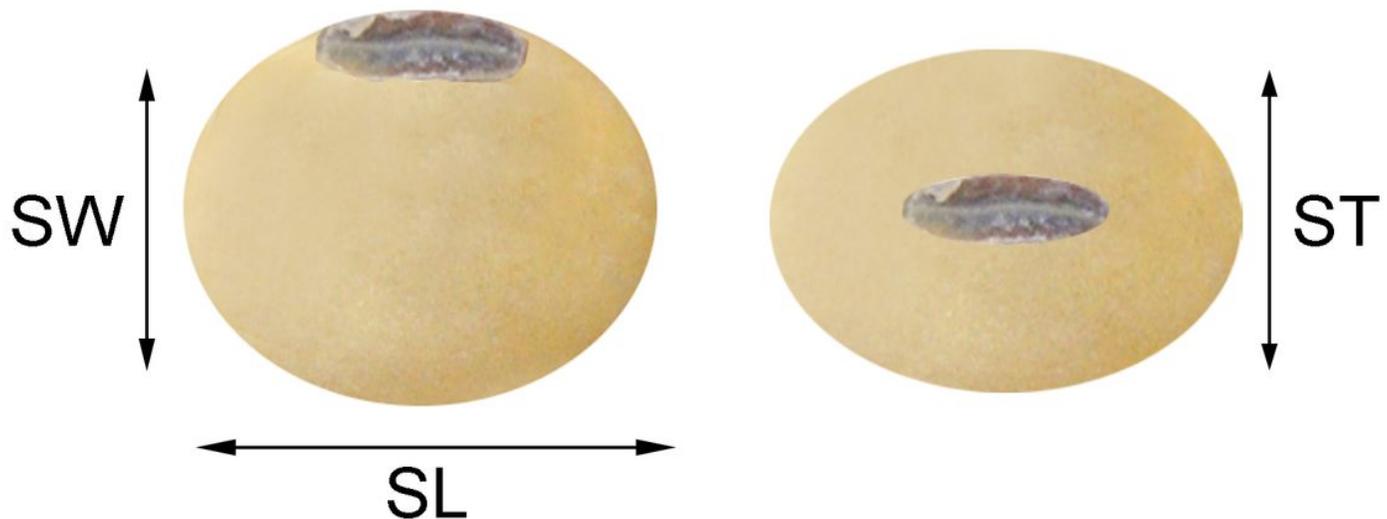


Figure 1

Measuring seed width (SW), length (SL) and thickness (ST).

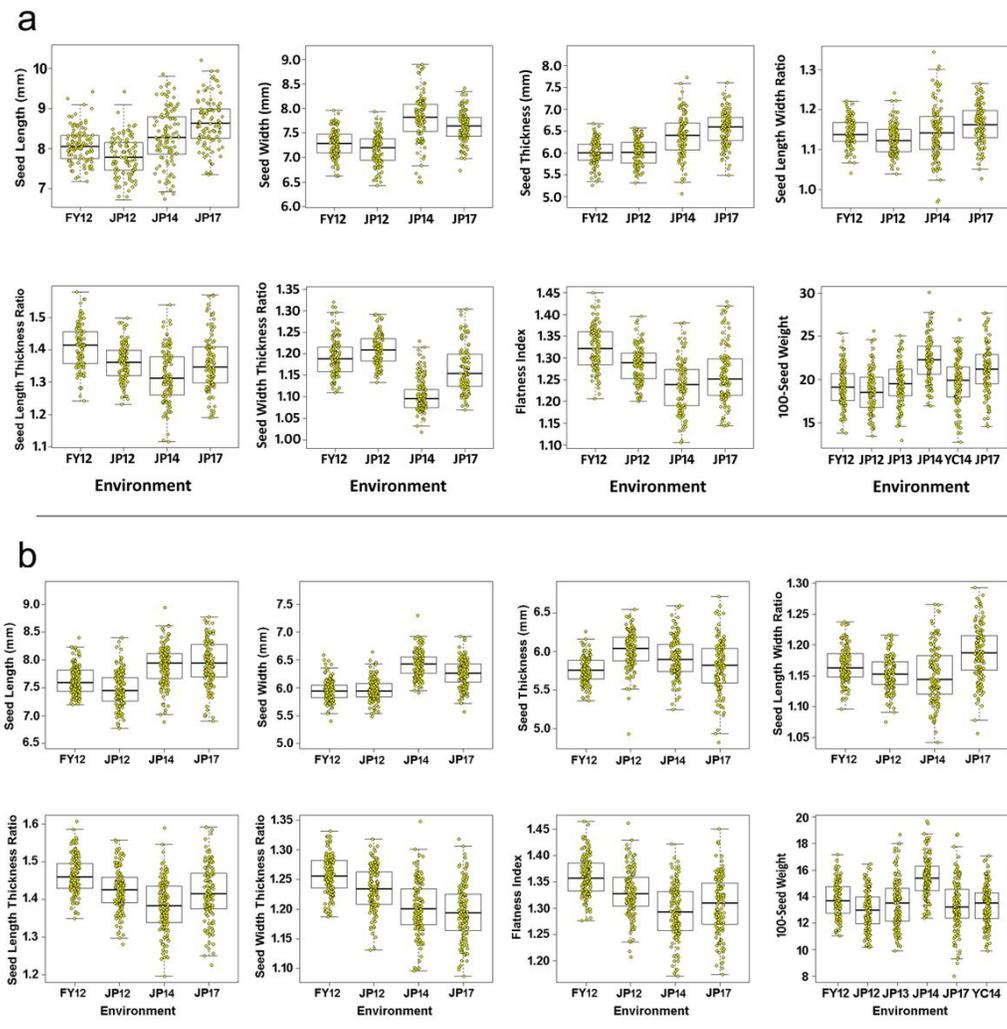


Figure 2
 Boxplot for seed size, seed shape and 100-seed weight traits. The black line in the middle of the box shows the median, the white box indicates the range from the lower quartile to the upper quartile, the dashed black line and yellow dots represent the dispersion and frequency distribution of the phenotypic data in each of the six environments, i.e., 12FY, 12JP, 13JP, 14JP, 14YC, and 17JP. While a and b represent LM6 and ZM6 populations

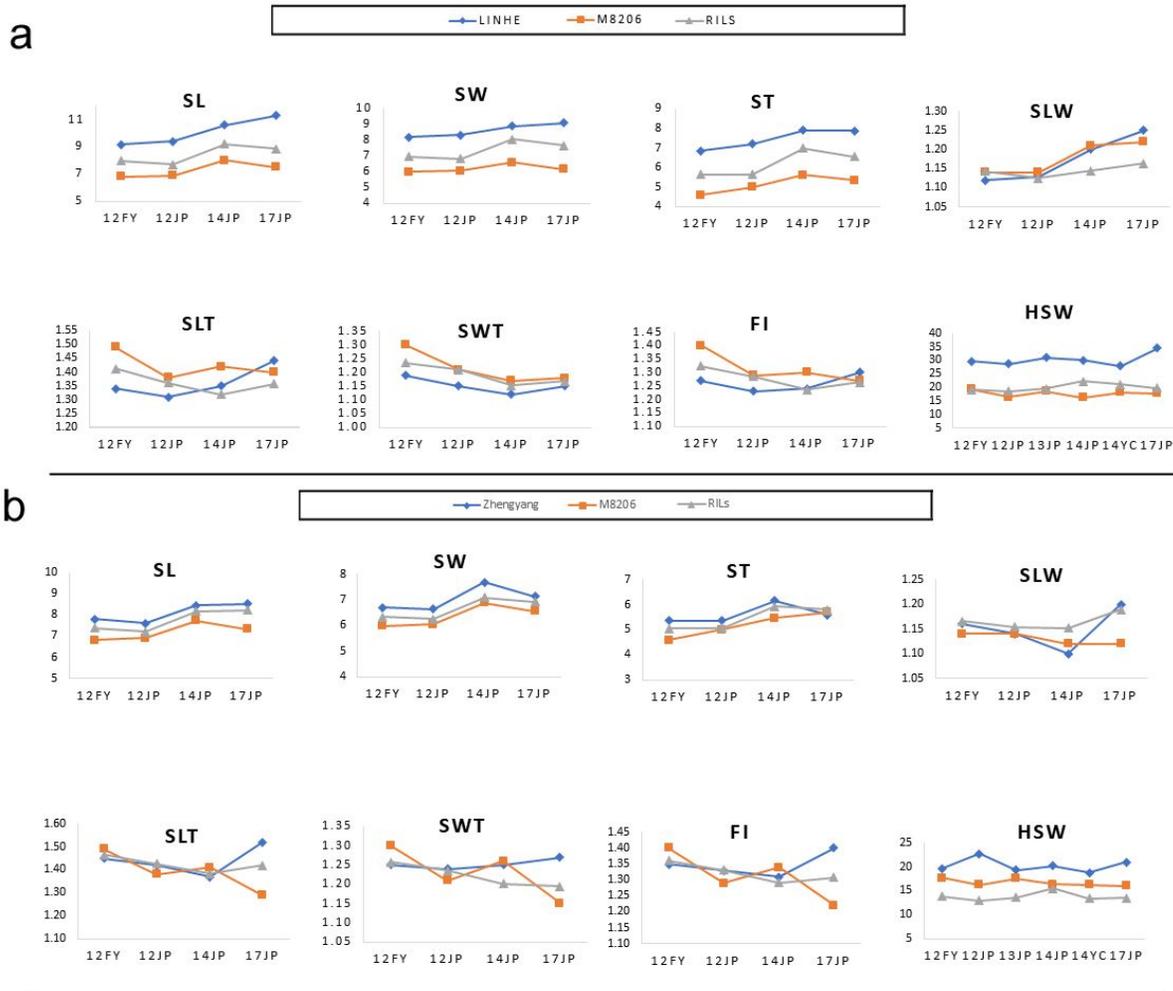


Figure 3
 Performance of the three parents of the RILs mapping populations, i.e., Meng8206, Linhefenqingdou, and Zhengyanghuangdou along with the two derived RIL populations, LM6 and ZM6, for seed sizes and shapes traits as well as 100-seed weight among multiple environments. 12JP, 13JP, 14JP and 17JP indicate phenotyping at the Jiangpu Experimental Station in 2012, 2013, 2014 and 2017 growing seasons, respectively. 12FY indicates the Fengyang Experimental Station, Chuzhou in 2012 growing season. 14YC indicates at the Yancheng Experimental Station in 2014.

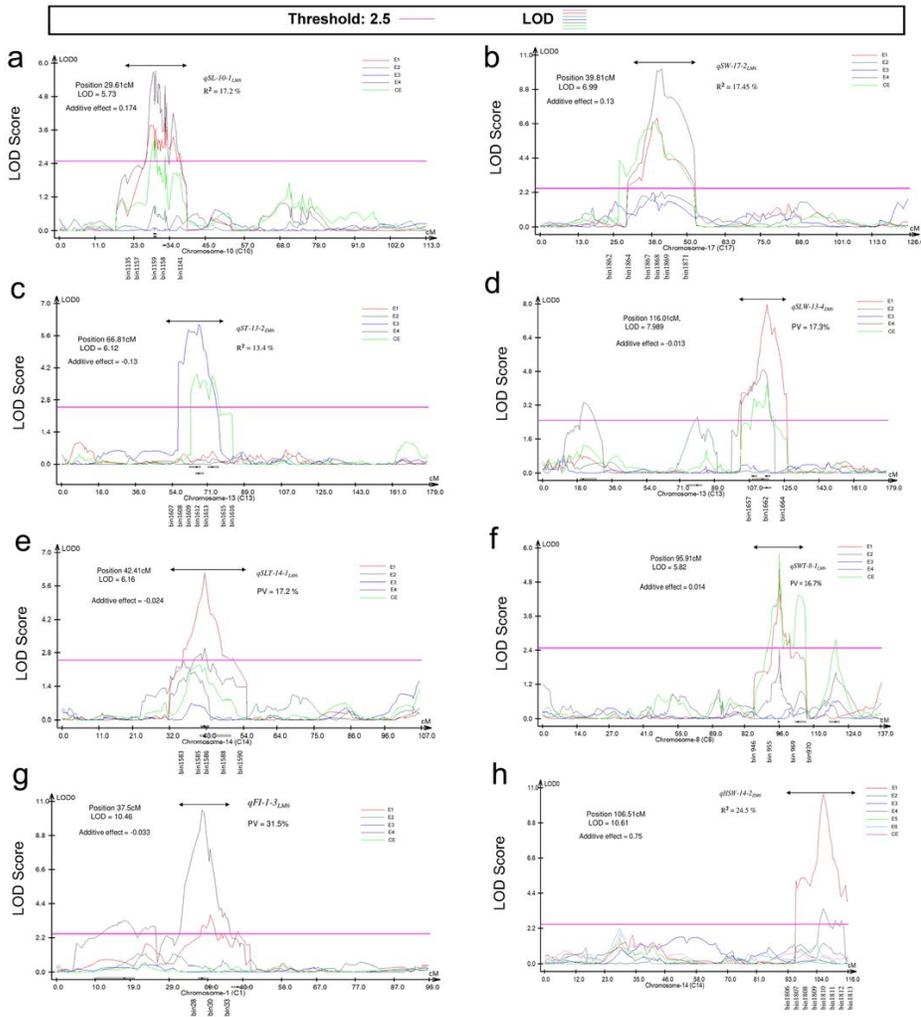


Figure 4

Position of most prominent QTL detected by CIM approach associated with seed size and seed shape traits in the LM6 and ZM6 RIL populations grown in multiple environments indicated as with E1, FY2012; E2, JP2012; E3, JP2013; E4, JP2014; E5, YC2014; E6, JP2017 respectively, in addition to the combined environment (CE). (a) LOD curve for qSL-10-1LM6, (b) LOD curve for qSW-17-2LM6, (c) LOD curve for qST-13-2ZM6, (d) LOD curve for qSLW-13-4ZM6, (e) LOD curve for qSLT-14-1LM6, (f) LOD curve for qSWT-8-1LM6, (g) LOD curve for qFI-1-3LM6, and (h) LOD curve for qHSW-14-2ZM6. The LOD threshold (2.5) is indicated by a pink line. The double-headed arrow denotes the location of prominent QTL. The X and Y-axis represent chromosome and LOD score, respectively.

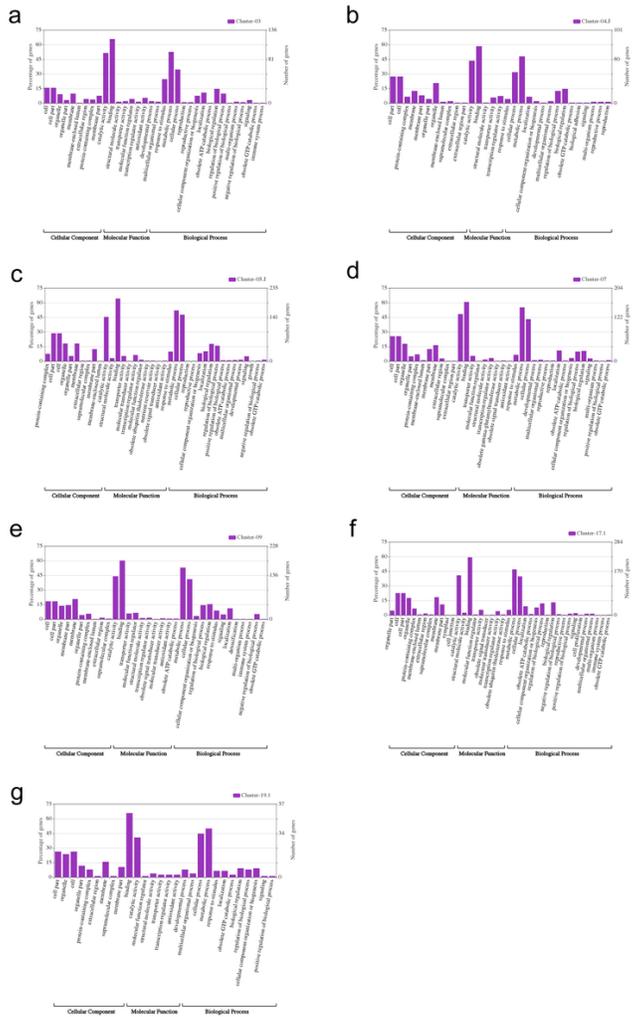


Figure 5
 WeGO analysis of the genes located within the seven major QTL clusters: (a) Cluster-03; (b) Cluster- 4.1; (c) Cluster-5.1; (d) Cluster-07; (e) Cluster-09; (f) Cluster-17.1; (g) Cluster-19.1

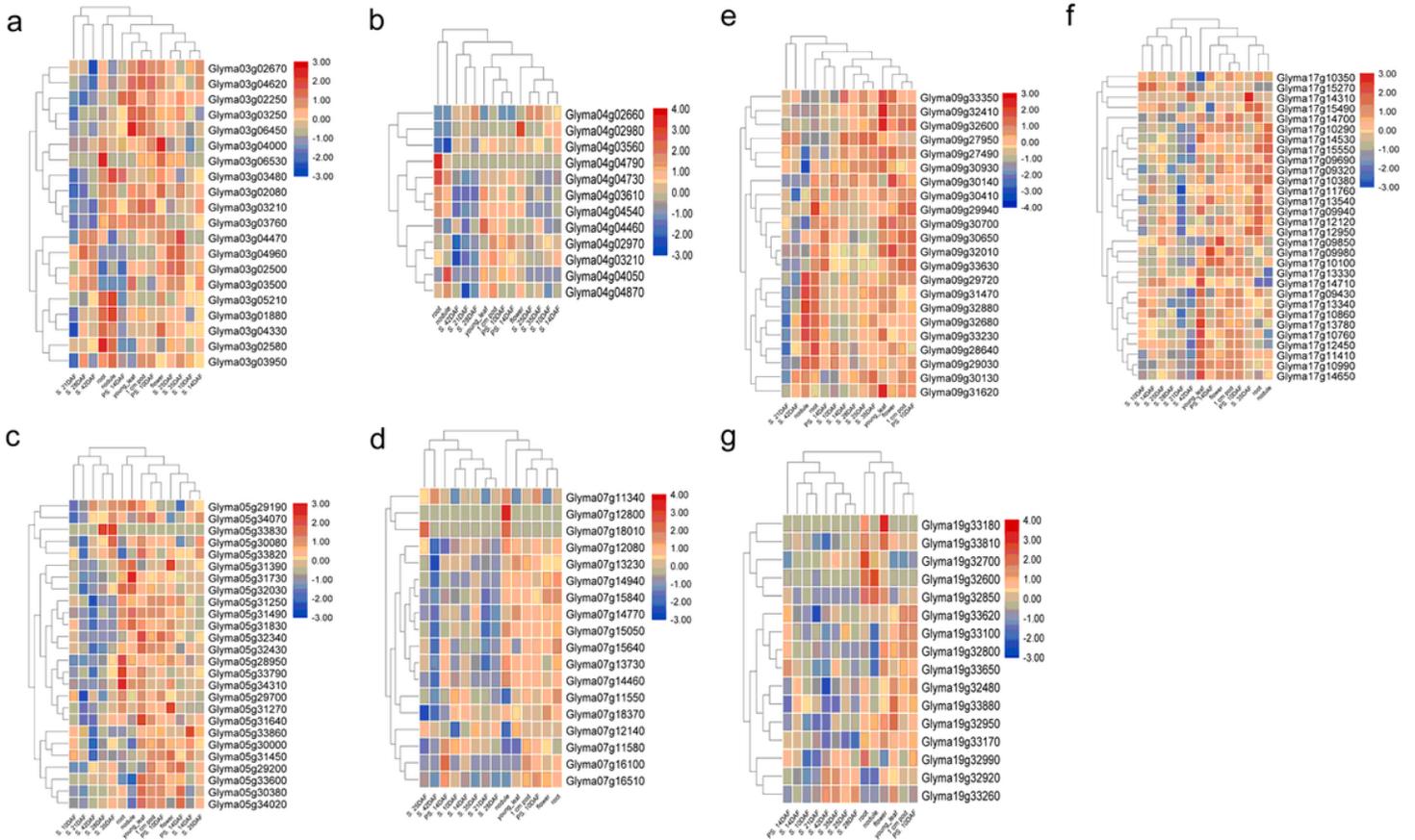


Figure 6
 WeGO analysis of the genes located within the seven major QTL clusters: (a) Cluster-03; (b) Cluster- 4.1; (c) Cluster-5.1; (d) Cluster-07; (e) Cluster-09; (f) Cluster-17.1; (g) Cluster-19.1

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Suppl.Fig.1.tif](#)
- [Suppl.Fig.2.tif](#)
- [Suppl.Table1.xlsx](#)
- [Suppl.Table2.xlsx](#)
- [Suppl.Table3.xlsx](#)
- [Suppl.Table4.xlsx](#)
- [Suppl.Table5.xlsx](#)
- [Suppl.Table6.xlsx](#)
- [Suppl.Table7.xlsx](#)
- [Suppl.Table8.xlsx](#)
- [Suppl.Table9.xlsx](#)
- [Suppl.Table10.xlsx](#)
- [Suppl.Table11.xlsx](#)
- [Suppl.Table12.xlsx](#)
- [Suppl.Table13.xlsx](#)