QTL mapping and candidate gene mining for soybean seed weight per plant

Meng Yu¹, Zhangxiong Liu¹, Shanshan Jiang¹, Ning Xu³, Qingshan Chen¹, Zhaoming Qi¹ and Wenhe Lv¹

¹College of Agriculture Department, Northeast Agricultural University, Harbin, Heilongjiang, PR China; ²Institute of Crop Sciences Department, Chinese Academy of Agricultural Sciences, Beijing, PR China; ³Heilongjiang Academy of Land Reclamation Sciences, Harbin, Heilongjiang, PR China

ABSTRACT
In this study, 147 recombinant inbred soybean lines constructed from the parents Charleston and Dongnong594, were used to map quantitative trait loci (QTLs) for seed weight per plant in multiple years (2006 to 2010 and 2013). QTL mapping was done using a simple sequence repeat (SSR) map combined with specific-locus amplified fragment (SLAF) map and the composite interval mapping (CIM), multiple interval mapping (MIM) and inclusive complete interval mapping method (ICIM) algorithms. By combining QTLs with the QTL physical locations, two QTL intervals located in the D1a and C2 linkage groups were found in different years. KEGG and GO annotations of the 413 genes located within these QTL intervals were used to screen for candidate genes potentially involved in seed production. Based on WeGo online genetic classification, we ultimately selected four genes related to yield traits, Glyma.01G158700, Glyma.01G147800, Glyma.01G125400 and Glyma.01G156800, as candidate genes.

Introduction
Soybean, which is rich in both protein and oil, is one of the most economically important crops [1]. Over the past 50 years, soybean production has declined year after year in China, and yields have increased slowly. Soybean imports account for 80% of the total demand, so improving domestic soybean production has become the primary goal of soybean breeding. The most important traits targeted by soybean breeding programmes are yield-related traits. One of the main yield components is grain weight per plant, which is a complex quantitative trait controlled by multiple genes and influenced by environmental conditions, and therefore more difficult to select for compared with monogenic traits [2].

Numerous yield-related quantitative trait loci (QTLs) have been identified in different genetic backgrounds and environments [3]. For example, Keim et al. [4] first began to perform soybean QTL analysis, in such aspects as the soybean yield and quality traits. Mian et al. [5] used two soybean populations to identify seed weight per plant QTLs in different environments and different years; they found two consistent QTLs located on linkage groups F and K, respectively. The QTL qSw13–1 was detected only once in each recombinant inbred population; the alleles of this QTL could affect seed weight [5]. Additional QTLs have also been identified in the vicinity of qSw13–1 [6,7]. Chen et al. [8] mapped QTLs for grain weight per plant in a single environment and timepoint, that were distributed on the M, A1 and A2 linkage groups. Zhou et al. [9] located 38 QTLs for yield, yield components and lodging properties that were mainly concentrated in the C2, F and I linkage groups; the QTL on linkage group I was associated with the 100-grain weight. Liu et al. [10] found two significant seed yield per plant (SYPP) QTLs potentially related to QTLs for seed numbers per pod (SNPP) and pod numbers per plant (PNPP) at the same loci on Chr 8 (Satt390) and 10 (Sat...108). Kato et al. [11] used two populations of 225 and 250 recombinant inbred lines (RILs) developed from crosses between Japanese and US cultivars of soybean; and identified 15 significant QTLs for single seed weight (SSW) dispersed among 11 chromosomes in the two populations. One QTL located between Sat...284 and Sat...292 on chromosome 17 was detected (3.6 < LOD < 14.1) in both populations grown in all environments [11]. Yan et al. [12] used two populations to identify seven over-dominant QTLs associated with seed weight. Teng et al. [13] constructed a molecular genetic map including 213 simple sequence repeat (SSR) markers distributed over 18 of 20 chromosomes (linkage groups) and identified nine QTLs associated with seed weight. QTLs controlling the seed weight have also been identified in...
other species, including rapeseed, chickpea and lentil [14–17].

Many researchers have focused on mapping QTLs for seed weight per plant [4–17], but QTLs are largely located in a small number of environments. Few studies have focused on mapping grain weight per plant QTLs under multiple environments, so it is difficult to discover the main effect QTL loci. Meanwhile, without a high density map, it is difficult to guarantee the accuracy of the results. Therefore, in this study, we mapped the grain weight per plant QTLs at multiple years using a Charleston × Dongnong594 recombinant inbred line (RIL) population that has previously been sequenced by specific-locus amplified fragment (SLAF)-seq and for which a high-density genetic map has been constructed [18]. This RIL population was grown in the field for six years and multiple mapping methods were employed to (1) identify the major grain weight per plant QTLs and environment-specific QTLs, (2) mine the major QTL candidate genes, and (3) compare QTL positions in different years and in different studies to obtain high-confidence QTLs, in order to be used in soybean breeding programmes.

Materials and methods

QTL analysis of soybean seed weight per plant

Experimental material

A total of 147 RILs derived from a cross between the American cultivar ‘Charleston’ (♂) and the Chinese line ‘Dongnong594’ (♀) (developed at the Northeast Agricultural University, Harbin, Heilongjiang, China) were used for QTL mapping studies [19]. The RIL population was constructed using single-seed descent methods by Chen et al. [20]. The 147 RILs and their parents were planted in Harbin (HRB; longitude 126°37E, latitude 45°45N) each year from 2006 to 2010, and in 2013. The plants were arranged in a randomized complete block experimental design in plots 1 m in length and 0.5 m in width, with two replications in 2006–2007 and three replications in 2008–2010 and 2013.

Measurement of grain weight per plant

Seeds from five randomly harvested plants were taken per plot for each line to measure the grain weight per plant. For each RIL, grain weight per plant data from the two (2006–2007) or three replication plots (2008–2010, 2013) were averaged.

Genetic map

The genetic map for this RIL population was first constructed by Chen et al. [20]. The total length of the genetic map is 1913.5 cM, and consists of 167 SSR markers with an average distance between markers of 11.89 cM.

A high density genetic map of this RIL population was also constructed by Qi et al. [18] using 5308 SLAF markers identified by SLAF-seq. The total length of this map is 2655.68 cM with an average distance of 0.5 cM between adjacent markers.

QTL analysis

The CIM and MIM models in Windows QTL Cartographer V.2.5 and ICIM in QTL ICIMapping were used to jointly analyze QTLs to identify the major loci for grain weight per plant in each year. With the likelihood ratio (LR) greater than 11.5 (LOD value > 2.5) as the threshold for QTL, we determined the confidence interval by the peak of LOD values intersection with 2.5, in order to find the corresponding identification interval. All favourable alleles were derived from the parent ‘Charleston’ as indicated by the positive values of the additive effect and the parent ‘Dongnong594’, as indicated by the negative values of the additive effects.

Mining of major SWPP QTL for candidate genes

Soybean genomic data were downloaded from the Phytozome website (http://phytozome.jgi.doe.gov), according to the physical position of the major QTLs, and candidate genes were extracted from the predicted gene list. Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/) and gene ontology (GO; http://geneontology.org) databases were used to obtain detailed pathway, gene ontology and annotation information. We organized the functional annotation information for all genes within the QTL interval which had been identified in different environments, screened out the fake genes and used Web Gene Ontology Annotation Plot (WeGo; http://wego.genomics.org.cn) to implement the gene classification. A Perl script was used to extract the gene sequences located within the QTL region, and gene annotation was carried out on the BMKCloud (http://www.biocloud.net).

Results and discussion

Grain weight per plant data analysis

The data of grain weight per plant for six years were continuously distributed and the skewness was small. The grain weight per plant ranged from 3.92 to 44.63 g. This result indicated that QTL mapping was suitable for the whole population (Table 1).
**QTL mapping of grain weight per plant using the SSR genetic map**

A total of 25 QTLs controlling grain weight per plant located on 10 linkage groups were identified by CIM and MIM using the SSR genetic map. The seven QTLs detected by the CIM method were located on the D1a, A1, B1 and C2 soybean linkage groups. The LOD scores ranged from 2.53 to 4.20, the phenotype contribution rates from 7% to 13%, and the additive effects ranged from −2.42 to 1.6. The 18 QTLs detected by MIM were located on the D1a, G, B2, D1b, N, B1, H, C2, E and A1 linkage groups. The LOD scores ranged from 1.14 to 4.98, the phenotype contribution rates from 2.70% to 38.30%, and the additive effects ranged from −4.64 to 4.07. Two of these QTLs were mapped to the same physical location in different environments. One QTL was mapped to the D1b linkage group by MIM in 2008 and 2009, with a LOD score of 1.9 and 1.14, additive effect of 0.4215 and 0.8172, and phenotype contribution rate of 6.7% and 4.5%, respectively, and a confidence interval of 108.4 cM to 141.6 cM. The other QTL was mapped to linkage group G by MIM in 2006 and 2013, with a LOD score of 1.3 and 1.2, additive effects of −0.8565 and −2.265, and contribution rates of 4.1% and 4.6%, respectively, and a confidence interval of 1.25 cM to 14.1 cM (Table 2).

**QTL mapping of grain weight per plant using the SLAF genetic map**

A total of 29 QTLs controlling grain weight per plant located on 13 linkage groups were identified by ICIM and CIM using the SLAF map. The 22 QTLs detected by CIM were located on linkage groups D1a, E, L, A2, C2, H, M, O, J and L with LOD scores ranging from 2.57 to 6.3, phenotype contribution rates from 6% to 15%, and additive effects ranging from −3.38 to 5. Seven QTLs were detected by ICIM. These QTLs were located on linkage groups I, E, C1, C2, D2 and H and had LOD scores ranging from 2.0104 to 4.5389, phenotype contribution rates

---

**Table 1. Information about grain weight per plant.**

| Year | Environment | Parent | RIL population |
|------|-------------|--------|----------------|
| 2006 | Harbin      | Charleston | 13.90 – 17.58 |
| 2007 | Harbin      | Dongnong594 | 16.60 – 19.65 |
| 2008 | Harbin      | Average   | 15.54          |
| 2009 | Harbin      | Range     | 6.02 – 30.84   |
| 2010 | Harbin      | Kurtosis  | 0.10           |
| 2013 | Harbin      | Skewness  | 0.85           |

---

**Table 2. QTL mapping of grain weight per plant in SSR mapping.**

| Year | Method | Linkage | Confidence interval | Marker interval | QTL position | LOD | Additive effect | R² (%) |
|------|--------|---------|---------------------|----------------|--------------|-----|----------------|--------|
| 2006 | CIM    | D1a     | 199.8 – 210.1       | Sat1 (216-21) | 207.40       | 3.50| –2.31          | 100.0  |
|      | MIM    | D1a     | 211.0 – 216.8       | Sat2 (370-370)| 212.80       | 3.40| –2.42          | 13.00  |
|      |        | D1a     | 231.1 – 232.5       | Sat1 (254-254)| 232.00       | 2.70| 1.6            | 10.00  |
|      |        | D1a     | 198.0 – 217.7       | Sat1 (370-482)| 207.40       | 4.98| –2.25          | 13.60  |
|      |        | D1a     | 228.9 – 245.0       | Sat1 (273-273)| 232.00       | 4.90| 1.68           | 2.80   |
|      |        | G       | 0.5 – 14.1          | Sat1 (273-273)| 11.20        | 1.30| 0.86           | 4.10   |
| 2007 | CIM    | A1      | 205.9 – 219.3       | Sat1 (270-270)| 212.90       | 2.95| –1.25          | 8.00   |
|      | MIM    | D1a     | 185.9 – 191.3       | Sat1 (273-482)| 188.10       | 2.70| –1.12          | 8.00   |
|      |        | D1a     | 178.7 – 187.2       | Sat1 (273-273)| 186.80       | 2.50| 4.07           | 4.80   |
|      |        | D1a     | 187.2 – 192.1       | Sat1 (273-482)| 188.20       | 2.90| –4.64          | 38.30  |
| 2008 | CIM    | B2      | 82.8 – 97.2         | Sat1 (200-200)| 96.90        | 1.60| 0.73           | 2.70   |
|      | MIM    | D1a     | 58.4 – 66.6         | Sat1 (273-273)| 61.50        | 4.20| –2.01          | 13.00  |
|      |        | B1      | 59.3 – 63.6         | Sat1 (273-273)| 61.50        | 2.89| –1.45          | 7.10   |
|      |        | D1b     | 108.4 – 141.6       | Sat1 (273-273)| 117.50       | 1.14| 0.82           | 4.50   |
| 2009 | CIM    | C2      | 69.8 – 72.6         | Sat1 (273-273)| 71.10        | 2.53| –1             | 7.00   |
|      | MIM    | C2      | 44.6 – 103.2        | Sat1 (273-273)| 71.40        | 2.40| –0.99          | 6.20   |
|      |        | D1a     | 53.7 – 84.7         | Sat1 (273-273)| 64.00        | 1.80| –0.97          | 4.40   |
|      |        | E       | 10.1 – 29.7         | Sat1 (273-273)| 20.00        | 1.54| 0.96           | 6.50   |
| 2013 | CIM    | –       | –                   | –              | –            | –   | –              | –      |
|      | MIM    | A1      | 72.5 – 110          | Sat1 (200-200)| 89.00        | 2.00| 3.05           | 5.40   |
|      |        | C2      | 193.4 – 230.3       | Sat1 (273-273)| 212.90       | 1.50| –3.3           | 7.40   |
|      |        | G       | 1.25 – 14.5         | Sat1 (273-273)| 2.40         | 1.20| –2.27          | 4.60   |
Grain weight per plant of SSR map and SLAF map QTL mapping integration.

Integration of grain weight per plant QTLs identified using the SSR and SLAF maps
Integration of grain weight per plant QTLs results using the SLAF and SSR maps revealed two QTL intervals that were identified in in two years. One QTL was located on the D1a linkage group by CIM in 2006, and the other was located on the C2 linkage group by both MIM and CIM in 2010 (Table 4).

In this study, the parents were distant geography hybridization; the traits had more differences and the RILs were planted in the same location for many years. This approach allowed us to evaluate the stability of QTLs, which is very necessary in QTL mapping. Fulton et al. [21] suggested that QTLs found in a variety of environments are more useful than those found only in one environment. Several researchers have used this approach in order to identify QTLs controlling soybean seed weight [22–27]. In this study, grain weight per plant QTLs were mainly concentrated in the A1, A2, D1b, G and I linkage groups. Fan et al. [28] used CIM and MIM to map QTLs controlling grain weight per plant in multiple years and multiple time points. They identified 13 QTLs located on linkage groups D1a, B1, B2, C2, F, G and A1 that had basically identical locations in each year [28]. This was supported by the results from our study. In this study, however, we found no QTLs on the F linkage group. This may be due to differences in population materials and the corresponding linkage group markers and mapping methods. The contribution rates of the same QTLs mapped by different mapping methods were also different. A true model of QTL distribution and population parameters is not known; whether a QTL is present at a certain location is yet to be tested repeatedly. Li

Table 3. QTL mapping of grain weight per plant in SLAF mapping.

| Year | Method | Linkage | Confidence interval | Marker interval | QTL position | LOD | Additive effect | $R^2$ (%) |
|------|--------|---------|---------------------|----------------|--------------|-----|----------------|----------|
| 2006 | CIM    | D1a     | 68.2–70.9           | Mark14282-Mark665297 | 69.50         | 4.40 | 2.10           | 10.00    |
|      |        | E       | 20.8–34.1           | Mark25731-Mark404   | 23.70         | 3.60 | –1.37          | 8.00     |
|      |        | L       | 100.3–101.0         | Mark139999-Mark139963 | 100.70        | 2.80 | 2.22           | 10.00    |
|      | ICIM   | A1      | 67.60–69.2          | Mark17626-Mark672315 | 69.00         | 4.54 | –1.71          | 12.92    |
|      |        | J       | 117.73–118.414     | Mark75643-Mark62195 | 118.00        | 3.65 | 1.72           | 10.15    |
| 2007 | CIM    | A2      | 35.7–47.1           | Mark1007458-Mark1006709 | 41.30         | 3.90 | –1.07          | 9.00     |
|      |        | C2      | 3.1–6.6             | Mark51325-Mark471932 | 5.20          | 3.30 | 1.02           | 7.00     |
|      |        | E       | 59.0–59.35          | Mark43156-Mark371795 | 59.30         | 3.00 | 0.90           | 7.00     |
| 2008 | CIM    | A2      | 37.7–47.1           | Mark1007458-Mark1006709 | 36.00         | 3.70 | 0.62           | 8.00     |
|      |        | C2      | 3.1–6.6             | Mark51325-Mark471932 | 47.60         | 4.20 | 0.68           | 10.00    |
|      |        | E       | 59.0–59.5           | Mark43156-Mark371795 | 63.50         | 3.00 | 0.06           | 7.00     |
| 2009 | CIM    | H       | 56.1–57.2           | Mark106308-Mark88685 | 56.70         | 3.30 | 1.21           | 8.00     |
|      |        | L       | 58.7–59.3           | Mark110365-Mark121828 | 58.70         | 2.57 | 1.08           | 7.00     |
|      |        | H       | 60.6–65.0           | Mark132546-Mark108560 | 62.30         | 3.36 | 1.32           | 8.00     |
| 2010 | CIM    | C2      | 101.1–105.7         | Mark45104-Mark461991 | 104.10        | 3.20 | 1.21           | 6.00     |
|      |        | M       | 114.2–117.6         | Mark191274-Mark210279 | 115.70        | 2.80 | 1.01           | 6.00     |
|      |        | O       | 91.6–92.5           | Mark1127092-Mark1149226 | 91.90         | 2.60 | –1.20          | 6.00     |
|      | ICIM   | I       | 90.677–92.057       | Mark750587-Mark608491 | 91.00         | 3.05 | 1.08           | 9.23     |
| 2013 | CIM    | A2      | 27.5–29.7           | Mark1012456-Mark1008258 | 28.00         | 3.30 | –3.38          | 8.00     |
|      | ICIM   | A2      | 33.1–36.6           | Mark999886-Mark1007458 | 35.10         | 3.09 | –3.10          | 7.00     |
|      |        | J       | 62.0–66.0           | Mark705177-Mark688381 | 64.40         | 3.90 | 4.45           | 8.00     |
|      |        | L       | 68.6–69.3           | Mark738466-Mark742249 | 68.90         | 2.70 | 3.54           | 6.00     |
|      |        | L       | 60.4–60.8           | Mark1436175-Mark1387850 | 55.80         | 6.30 | 5.00           | 15.00    |
|      |        | L       | 63.0–67.0           | Mark1419993-Mark1378243 | 64.90         | 3.80 | 3.83           | 9.00     |
|      |        | C1      | 50.534–51.219       | Mark30158-Mark326398 | 51.00         | 2.22 | –2.78          | 7.09     |
|      |        | C2      | 25.665–26.203       | Mark446204-Mark497260 | 26.00         | 2.02 | –2.91          | 6.49     |
|      |        | D2      | 82.133–83.168       | Mark277972-Mark237794 | 83.00         | 2.23 | –2.92          | 7.14     |
|      | H      | 90.876–91.216      | Mark121541-Mark119247 | 91.00         | 2.01 | –3.52          | 6.40     |

Table 4. Grain weight per plant of SSR map and SLAF map QTL mapping integration.

| Year | Method | Map   | Linkage | Confidence interval | Marker interval | QTL position | LOD | Additive effect | $R^2$ (%) |
|------|--------|-------|---------|---------------------|----------------|--------------|-----|----------------|----------|
| 2010 | MIM    | SSR   | C2      | 44.6–103.2          | Satt076-Satt002 | 71.4          | 2.4 | –0.99          | 6.2      |
| 2010 | CIM    | SLAF  | C2      | 101.1–105.7         | Mark45104-Mark481991 | 104.1         | 3.2  | 1.21           | 6.0      |
| 2010 | MIM    | SSR   | D1a     | 53.7–84.7           | Satt526-Satt05 | 64            | 1.8  | –0.97          | 4.4      |
| 2006 | CIM    | SLAF  | D1a     | 66.2–70.9           | Mark614262-Mark6655297 | 69.5          | 4.4  | 2.1            | 10.0     |
et al. [29] performed a comparative study of four different QTL mapping algorithms and found that QTL mapping methods differed in their ability to map QTLs. The CIM method had the strongest ability to map, but it remains to be confirmed whether all of the QTLs found are real [29]. It is possible that the difference in the method of estimation can give different positioning of the same QTL and a different contribution rate [29]. When a real trait QTL distribution model and general parameters are previously unknown, the experiment test has to be repeated to confirm whether a QTL exists in a certain position [29]. Kao et al. [30] compared the performance of the CIM, IM and MIM methods to map QTLs in an F2 population of pine, and in contrast to Li et al. [29], found that better QTL detection and more accurate positioning was obtained with MIM. Different QTL positions obtained using CIM and MIM are due to differences in the estimation method, the type of cross, different selection of groups, selection of tag density and selection method. Finding the appropriate model between different traits of different crops and methods to locate related QTL still needs further study.

To combine the SLAF and SSR map in this study, we used WinQTLcart and QTL IciMapping to position the material for many years. We obtained the grain weight per plant traits QTL integrating interval and the distribution on the two linkage groups, D1a and C2, respectively. Our results are consistent with Jiang et al. [31], who also identified a QTL for grain weight per plant on linkage group C2, which is believed to control the yield-related traits. In this study, two QTLs of grain weight per plant were obtained with the SSR map. Compared with the SLAF map, only one QTL interval for controlling the grain weight per plant was selected, and the other QTL interval was screened out. This could improve the accuracy of integrated analysis. Simultaneous detection using three methods could also improve the accuracy of QTLs.

Overall, the QTLs that we identified in this study are consistent with those identified by other researchers. We have also identified large-effect grain weight per plant QTLs in different genetic backgrounds and in different genetic environments, indicating that they are stable. Our work lays the foundation for precise positioning of QTL and marker-assisted breeding.

**Mining integrated QTLs for candidate genes**

**QTL interval gene prediction**

According to the position of the QTL interval which had been identified in different years for the grain weight per plant traits on the whole genome, candidate genes were screened from the whole soybean genome. In total, 413 candidate genes were identified in different gene databases; 235 genes had GO annotations and 78 genes had KEGG annotations. In the GO database, 21 gene lengths were between 300 and 1000 bp, and the lengths of 212 genes were over 1000 bp. In the KEGG database, 10 genes were between 300 and 1000 bp, and 68 genes were longer than 1000 bp. Of the 413 candidate genes, more than 10% were annotated to the cell composition categories, cell structure and organelles; more than 10% were annotated to the molecular function categories, binding and catalytic reaction and more than 10% were annotated to the biological process categories of

![Figure 1. WeGo annotation information.](image-url)
biological regulation, cellular processes and metabolism. There were few genes encoding proteins with an extra-cellular region and genes involved in transcription regulation and cell apoptosis (Figure 1).

QTL interval gene annotations
Based on gene annotation information, we selected four candidate genes predicted as being related to the process of biomolecules (Table 5). All these four candidate genes were annotated as (leucine-rich repeat receptor-like kinase, LRK). LRK1, which includes the protein kinase-like domain (IPR011009), protein kinase domain (IPR000719), serine/threonine/dual specificity protein kinase, catalytic domain (IPR002290), which restricts gibberellin biosynthesis during the internode elongation process by down-regulation of the gibberellin biosynthetic gene coding for ent-kaurene oxidase, could influence the yield of rice [32].

This study selected four candidate genes related to production which had the same structure domain with LRK1. The identification of the LRK1 gene is in agreement with Yang et al. [32], who reported that constitutively expressed genes could promote plant growth, could improve the spike grain number and increase production. Therefore this mining study suggested four candidate genes that could be considered related to production traits.

Conclusions
In this study, we mapped QTLs for grain weight per plant over six years, and identified two QTLs that were detected in two or more years by CIM and MIM. We screened 413 candidate genes located within these QTL regions and proposed four candidate genes related to yield traits based on KEGG and GO annotations. QTL mapping and gene prediction of grain weight traits in soybean are important to increase yield, and also may lay an important foundation for marker assisted breeding.

Acknowledgments
We acknowledge the Key Laboratory of soybean biology of Ministry of Education. This research was financed by the National Key R & D Program for Crop Breeding [grant no. 2016YFD0100300]; the National Key R & D Program for Crop Breeding [grant no. 2016YFD0100500 and 2016YFD0100201-21]; the SIPT Project of Northeast Agricultural University [grant no. 201610224146]; the outstanding academic leaders projects of Harbin [grant no. 2015RQXXJ018]; National Natural Science Foundation of China [grant no. 31401465, 31471516 and 31501332]; Young Talent project of Northeast agriculture university [grant no. 518062]; Heilongjiang Funds for Distinguished Young Scientists [grant no. JC2016004]; the China Post Doctoral Project [grant no. 2015MS81419]. The authors thank Dr Chen Qingshan from the Northeast Agricultural University for kindly providing the RIL population. We also appreciate the anonymous reviewers for their valuable comments and constructive suggestions on this manuscript.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by the National Key R & D Program for Crop Breeding [grant no. 2016YFD0100300]; the National Key R & D Program for Crop Breeding [grant no. 2016YFD0100500 and 2016YFD0100201-21]; the SIPT Project of Northeast Agricultural University [grant no. 201610224146]; the outstanding academic leaders projects of Harbin [grant no. 2015RQXXJ018]; National Natural Science Foundation of China [grant no. 31401465, 31471516 and 31501332]; Young Talent project of Northeast agriculture university [grant no. 518062];

Table 5. Predictive gene annotation information.

| Candidate genes | Gene functional annotation | Functional structure domain | Corresponding gene structure domain | Corresponding gene function | Reference |
|-----------------|----------------------------|-----------------------------|------------------------------------|-----------------------------|-----------|
| Glyma.01G158700 | Transducin/WD40 repeat-like superfamily protein | Protein kinase-like domain (IPR011009), Protein kinase domain (IPR000719), Serine/threonine/dual specificity protein kinase, catalytic domain (IPR002290) | LRK1 | Rice internode elongation | Yang et al. 2013 [32] |
| Glyma.01G156800 | Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase family protein | Ribonuclease H-like domain (IPR012337) | LRK1 | Rice internode elongation | Yang et al. 2013 [32] |
| Glyma.01G125400 | Disease resistance protein (TIR-NBS-LRR class) family | Leucine-rich repeat (IPR001611) | LRK1 | Rice internode elongation | Yang et al. 2013 [32] |
| Glyma.01G147800 | ENHANCED DISEASE RESISTANCE 2 | Ribonuclease H-like domain (IPR012337) | LRK1 | Rice internode elongation | Yang et al. 2013 [32] |
Heilongjiang Funds for Distinguished Young Scientists [grant no. JC2016004]; the China Post Doctoral Project [grant no. 2015MS81419].

References

[1] Zhang JP, Song QJ, Cregan P. B, et al. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). Theor Appl Genet. 2016;129(1):117–130.

[2] Yao D, Wang PW, Zhang J, et al. [A QTL mapping analysis of main yield traits in soybean]. J South China Agric Univ. 2014;35(3):41–46. Chinese.

[3] Tanksley SD, Miller J, Paterson A, et al. Molecular mapping of plant chromosomes. In: Gustafson JP, Appels R. editors. Chromosome structure and function. Stadler Genetics Symposia Series. Boston: Springer; 1988. p. 157–173.

[4] Keim P, Diers BW, Olson TC, et al. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genet. 1990;126(3):735–742.

[5] Mian MAR, Bailey MA, Tamulonis JP, et al. Molecular markers associated with weight in two soybean populations. Theor Appl Genet. 1996;93(7):1011–1016.

[6] Hoeck JA, Fehr WR, Shoemaker RC, et al. Molecular marker analysis of seed size in soybean. Crop Sci. 2003;43(1):68–74.

[7] Hyten DL, Pantalone VR, Sams CE, et al. Seed quality QTL in a prominent soybean population. Theor Appl Genet. 2004;109(3):552–561.

[8] Chen QS, Zhang ZC, Liu CY, et al. QTL analysis of major agronomic traits in soybean. Agr Sci China. 2007;6(4):399–405.

[9] Zhou R, Chen HF, Wang XZ, et al. [QTL analysis of yield, yield components, and lodging in soybean]. Acta Agron Sin. 2009;35(5):821–830. Chinese.

[10] Liu WY, Kim MY, Van K, et al. QTL identification of yield-related traits and their association with flowering and maturity in soybean. Crop Sci Biotech. 2011;14(1):65–70.

[11] Kato S, Sayama T, Fujii K, et al. A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. Theor Appl Genet. 2014;127(6):1365–1374.

[12] Yan L, Li YH, Yang CY, et al. Identification and validation of an overdominant QTL controlling soybean seed weight using populations derived from Glycine max × Glycine soja. Plant Breed. 2014;133(5):632–637.

[13] Teng WL, Feng L, Li W, et al. Dissection of the genetic architecture for soybean seed weight across multiple environments. Crop Pasture Sci. 2017;68(4):358–365.

[14] Ni N, Shi QJ, Wang XF, et al. A combined linkage and regional association mapping validation and fine mapping of two major pleiotropic QTLs for seed weight and silique length in rapeseed (Brassica napus L.). BMC Plant Biol. 2014; cited 2017 Aug 29] 14:114. DOI: 10.1186/1471-2229-14-114.

[15] Verma P, Goyal R, Chahota RK, et al. Construction of a genetic linkage map and identification of QTLs for seed weight and seed size traits in lentil (Lens culinaris Medik.). PlosOne. 2015; cited 2017 Aug 29] 10(10):e0139666. DOI: 10.1371/journal.pone.0139666.

[16] Dhaka N, Rout K, Yadava SK, et al. Genetic dissection of seed weight by QTL analysis and detection of allelic variation in Indian and east European gene pool lines of Brassica juncea. Theor Appl Genet. 2015;130(2):293–307.

[17] Singh VK, Khan AW, Jaganathan D, et al. QTL-seq for rapid identification of candidate genes for 100-seed weight and root/root total plant dry weight ratio under rainfed conditions in chickpea. Plant Biotechnol J. 2016;14(11):2110–2119.

[18] Qi ZM, Huang L, Zhu RS, et al. A high-density genetic map for soybean based on specific length amplified fragment sequencing. Plos One. 2014; cited 2017 Aug 29] 9(8): e104871. DOI: 10.1371/journal.pone.0108047.

[19] Cooper RL, Martin RJ, St Martin SK, et al. Registration of ‘Charleston’ soybean. Crop Sci. 1995;35(2):593–593.

[20] Chen QS, Zhang ZC, Liu CY, et al. [Construction and analysis of soybean genetic map using recombinant inbred line of Charleston × Dongnong 594]. Sci Agric Sin. 2005;38(7):1312–11316. Chinese.

[21] Fulton TM, Beck-Bunn T, Emmatty D, et al. QTL analysis of an advanced backcross of lycopersicon peruvianum to the cultivated tomato and comparisons with QTLs found in other wild species. Theor Appl Genet. 1997;95(5):881–894.

[22] Pathan SM, Vuong T, Clark K, et al. Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. Crop Sci. 2013;53(3):765–774.

[23] Liu YL, Li YH, Reifb JC, et al. Identification of quantitative trait loci underlying plant height and seed weight in soybean. Plant Genome. 2013; cited 2017 Aug 29] 6(3): 841–856. DOI: 10.3835/plantgenome2013.03.0006.0.

[24] Kulkarni KP, Kim M, Shannon JG, et al. Identification of quantitative trait loci controlling soybean seed weight in recombinant inbred lines derived from PI 483463 (Glycine soja × ‘Hutcheson’ (G. max). Plant Breed. 2016;135(5):614–620.

[25] Xin DW, Qi ZM, Jiang HW, et al. QTL location and epistatic effect analysis of 100-seed weight using wild soybean (Glycine soja Sieb. & Zucc.) chromosome segment substitution lines. Plos One. 2016; cited 2017 Aug 29] 11: e0149380. DOI: 10.1371/journal.pone.0149380.

[26] Yang Z, Xin DW, Liu CY, et al. Identification of QTLs for seed and pod traits in soybean and analysis for additive effects and epistatic effects of QTLS among multiple environments. Mol Genet Genomics. 2013;288:651–667.

[27] Hu ZB, Zhang HR, Kan GZ, et al. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (Glycine max L. Merr.). Genetica. 2013;141:247–254.

[28] Fan DM, Sun DJ, Ma ZZ, et al. [QTL mapping and interaction analysis of seed weight per plant in soybean among different environments]. Acta Agron Sin. 2013;39(6):1021–1029. Chinese.

[29] Li JQ, Zhang QJ, Ye SP, et al. [Comparative research on four mapping methods of QTLS]. Acta Agron Sin. 2005;11:1473–1477. Chinese.

[30] Kao CH, Zeng ZB, Teasdaler RD. Multiple interval mapping restricts internode elongation by down-regulating OsKO2. Biotechnol Lett. 2013;35(1):121–128.