Fine mapping of a QTL locus (\textit{QNFSP07-1}) and analysis of candidate genes for four-seeded pods in soybean

Yingying Li · Chunyan Liu · Nannan Wang · Zhanguo Zhang · Lilong Hou · Dawei Xin · Zhaoming Qi · Candong Li · Yan Yu · Hongwei Jiang · Qingshan Chen

Received: 7 July 2021 / Accepted: 6 November 2021 / Published online: 20 November 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract Soybean [\textit{Glycine max} (L.) Merr.] is an important grain and oil crop in the world, and it is the main source of high-quality protein. The number of four-seeded pods is a quantitative trait in soybean and is closely related to yield in terms of breeding. Therefore, it is of great significance to study the inheritance of four-seed pods and to excavate related genes for improving soybean yield. In this study, individuals with high ratio of four-seed pods which from chromosome segment substitution lines (CSSLs) that can be stably inherited were selected as the parent, and Suinong 14 (SN14) was used as recurrent parent to construct secondary mapping population via marker-assisted selection. From 2006 to 2017, QTL analysis was performed using secondary mapping populations, and the initial QTL mapping interval was 0.67 Mb and was located on Gm07. Based on the initial QTL mapping results, individuals that were heterozygous at the interval (36,116,118–37,399,738 bp) were screened in 2018, and the heterozygous individuals were subjected to inbreeding to obtain 13 \(F_3\) populations, with...
a target interval of 321 kb. Gene annotation was performed on the fine mapping interval, and 27 genes were obtained. Among 27 genes, Glyma.07G200900 and Glyma.07G201200 were identified as candidate genes. qRT-PCR was used to measure the expression of the 2 candidate genes at different developmental stages of soybean, and the expression levels of the 2 candidate genes in terms of cell division (axillary buds, COTs, EMs) were higher than those in terms of cell expansion (MM, LM), and these genes play a positive regulatory role in the formation of four-seeded pods. Haplotype analysis of 2 candidate genes which shows that Glyma.07G201200 has two excellent haplotypes, and the significance level between the two excellent haplotypes at \( p < 0.05 \). Those results provide the information for gene map–based cloning and molecular marker–assisted breeding of the number of four-seeded pod in soybean.

**Keywords**  Candidate genes · Four-seeded pod · Fine mapping · Haplotype analysis · QNFSP07-1 · Soybean

**Introduction**

Soybean [*Glycine max* (L.) Merr.] originated in China and has enriched oil and protein contents (Hwang et al. 2014). Soybean breeders have been working hard to improve the yield and quality of soybean, which is among the economically important crop varieties. As the number of four-seeded pods is closely related to yield, mapping QTLs related to four-seeded pods of soybean via marker-assisted selection will provide new ideas for cultivating new soybean varieties with multi pods and high yields (Yang et al. 2013; Kumar et al. 2014).

More than 401 QTLs for the number of pods per plant have been identified in soybean (Sun et al. 2006; Chen et al. 2007; Zhang et al. 2010); Among them, 86 of these QTLs were located by RIL population or \( F_2 \) population, and IM and CIM methods were used to detect QTLs for four-seeded pods in LGs B2, D2, H, I, and M (Wang et al. 2007; Li et al. 2020). Three QTLs associated with the number four-seeded pods have been detected, located on Gm02, Gm11, and Gm17 (Li et al. 2008). Zhou et al. (2009) used 165 recombinant inbred lines (RILs) to analyze the number of four-seeded pods and showed that 3 QTLs on linkage group (LG) I were associated with the number of four-seeded pods. Gao et al. (2012) reported 8 QTLs on LGs D1a, A1, and H were associated with the number of four-seeded pods in a RIL population. Yang et al. (2013) used a RIL population comprising 147 \( F_{2.14}^{-}-F_{2.19} \) individuals and detected QTLs for the number of four-seeded pods under 11 environmental conditions; these QTLs were located on LGs A1, B1, and F. Two others set of RILs (RIL 3613 and RIL 6013) have been for QTL mapping, and 21 and 26 pod-related QTLs were detected via complete interval mapping (CIM) and could explain the phenotypic variation percentages of 1.25% and 7.91%, respectively (Ning et al. 2018).

The number of four-seeded pods is a complex quantitative character, is highly heritable, and is susceptible to environmental conditions (Pei et al. 2010). Studies have shown that the number of four-seeded pods has a significant positive correlation with the number of pods per plant, and the number of pods per plant is highly correlated with yield (Peng et al. 1994). Moreover, the number of pods per plant and the number of seeds per plant are highly positively correlated, showing that an increase in the number of pods increases the number of seeds. In addition, the number of four-seeded pods and the number of seeds per pods are positively correlated (Li et al. 2018). The number of four-seeded pods is one of the important breeding goals for breeding new soybean varieties with high and stable yields (Richards, 2000). A relationship between lanceolate leaves and the number of pods in soybean has also been reported. Compared with that of plants with oval leaves, the ratio of four-seeded pods to the number of pods per plant of those with the lanceolate leaves was higher, suggesting that the alleles controlling lanceolate leaves play a major pleiotropic role in increasing the ratio of four-seeded pods in soybean (Takashashi, 1934; Domingo, 1945; Johnson and Bernard, 1962; Weiss, 1970). \( Ln \) has been widely explored, its expression has been found in lanceolate leaves, it is considered to be closely related to the number of four-seeded pods, and it has a significant pleiotropic effect (Jeong et al. 2012). \( Ln \) controls lanceolate leaves of soybean and is stably inherited in offspring, and the ratio of plants with four-seeded pods with lanceolate leaves (\( ln/ln \)) is higher than the ratio of plants with four-seeded pods with the
oval leaves (Ln/Ln) (Dinkins et al. 2002). In Arabidopsis, overexpression of CYP78A9 can induce large and seedless fruits and has been reported to be associated with the number of seeds (Ito and Meyerowitz, 2000). Three P450/CYP78A genes (AtCYP78A5, AtCYP78A6, and AtCYP78A9) have been reported to be associated with seed size and pod number (Li et al. 2016). KLU/CYP78A5 has been shown to promote both leaf and floral organ growth and seed development by increasing cell proliferation in ovules and increasing KLU activity in ovules; this activity is sufficient to increase seed size and number (Eriksson et al. 2010).

GmCYP78A10, which belongs to the P450/CYP78A family, was cloned, and it was shown that the proportion of four-seeded pods in soybean plants carrying the GmCYP78A10a allele was significantly higher than that carrying the GmCYP78A10b allele (Wang et al. 2015).

In this study, we selected a line with a high proportion of four-seeded pods from chromosome segment substitution lines (CSSLs), which were crossed with Suinong 14 (SN14) to construct a fine mapping population. The phenotype of this population was investigated. This information was combined with that determined via SSR markers to determine the genotype. The main objectives of this study were to define a more precise QTL interval and perform gene annotation on this interval to mine candidate genes and analyze them using qRT-PCR analysis and haplotype analysis.

Materials and methods

Plant materials, planting, and experimental design

In this study, CSSLs were constructed from SN14 (the main cultivar used in Heilongjiang, China) and ZYD00006 (a wild soybean resource) and subsequently backcrossed with SN14. In 2006, SN14 was crossed with ZYD00006 to obtain F1 plants, and in 2007, the F1 plants were backcrossed with SN14 to obtain BC1F1 plants. The experimental material CSSL860-13 was selected as the parent of the fine mapping population and has a stable and high ratio of the number of four-seeded pods to the number of pods per plant from 2012 to 2014 (BC3F6-BC4F6). T-tests showed significant differences between the phenotypic data of CSSL860-13 and SN14. In 2015, CSSL860-13 (BC4F6) was backcrossed with SN14 to obtain a CSSL860-13/SN14 backcross population. In 2016, all F1 seeds harvested the previous year were planted in an experimental field. In 2017, 1168 lines of the CSSL860-13/SN14 (F2) population were obtained in the experimental field, constituting a preliminary mapping population. In 2018, individual F2 plants were screened for heterozygous intervals, after which all were planted in the experimental field; 492 lines (F3) were ultimately obtained for the fine mapping population.

From 2006 to 2018, the mapping population was planted in Xiangyang Farm, Xiangfang District, Harbin, Heilongjiang Province (Harbin, latitude 45° 72′ N, longitude 126° 68′ E). A random block design with three replicates was used from 2006 to 2014, the row length was 5 m, a row spacing of 60 cm, and a plant spacing of 5 cm, 80 seeds were sown per row, and fields were under general management.

Data collection and segregation analysis

The phenotypic data of the number of four-seeded pods (NFSP) were recorded with Microsoft® Excel 2016 (Frye, 2016) and analyzed with SPSS 17.0 (Grafius.1978), including the frequency distribution, standard deviation, skewness, and extremum. The phenotypic data for the number of pods traits of each line were recorded from 2006 to 2018 as described by Qiu and Chang (2006).

The phenotypic data of NFSP were analyzed using the Segregation Analysis (SEA) software (Cao et al. 2013), which was based on the Microsoft Visual Studio 2010 platform and C++ programming language to identify the main genes and multi-gene mixed genetic models of quantitative trait. All genetic models are analyzed, and the Akaike Information Criterion (AIC) value is calculated by using the likelihood function. According to the AIC minimum value to determine the optimal genetic model, least-squares method was used to estimate the gene effect value and genetic variances, and precisely calculating the probabilities in the Smirnov test (nW2) and Kolmogorov test (Dn).

Smirnov test (nW2):
markers and QTLs can be determined. \( F \) statistical analysis is used to compare the following hypotheses: \( H_0, b_1 = 0; H_1, b_1 \) is not zero. The smaller the \( \text{Pr}(F) \) value is, the smaller the support for \( H_0 \), and the more relevant the tag is to the QTL.

The QTL nomenclature of soybean pod-seed number traits refers to the naming standard of QTLs of McCouch (1997), specifically: Q + trait name + linkage group number + “- “ + QTL sequence number.

Candidate gene mining

The “Williams 82.a2.v1” genome was used as a reference genome for predicting the candidate genes, and SoyBase (https://www.soybase.org/) provided the physical positions of the major QTL intervals. To determine the gene sequence related to the number of four-seeded pods, BLAST (https://phytozone.jgi.doe.gov/pz/portal.html#) was used to perform a homology analysis between the extracted gene sequence and the sequence of the candidate gene interval, and candidate genes related to the number of four-seeded pods were obtained.

The expression data of 27 candidate genes in 26 tissues of soybean were downloaded from Phytozone (https://phytozone-next.jgi.doe.gov/), and the transcriptome data of four experimental materials from the CSSLs during three periods (Gl: 4 days after flower open, Hr: 7 days after flower open, Co: 10 days after flower open) were also analyzed. A high ratio of four-seeded pods to the number of pods per plant (HNFSP) consisting of HNFSP-1 (24.62%) and HNFSP-2 (12.84%) and a low ratio of four-seeded pods to the number of pods per plant (LNFSP) consisting of LNFSP-1 (4.85%) and LNFSP-2 (0.40%) was used to draw the heatmap with the software TBtools (Chen et al. 2020).

qRT-PCR analysis of candidate genes

With SN14 (the ratio of four-seeded pods to the number of pods per plant: 18.39%) used as a control, experimental materials with a high ratio of
four-seeded pods to the number of pods per plant (line 953: 31.00%) and materials with a low ratio of four-seeded pods to the number of pods per plant (line 32: 0.00%) were screened for qRT-PCR. Florescence, the axillary bud stage, COT (cotyledon) stage, EM (early maturity) stage, MM (mid-seed maturity) stage, and LM (late seed maturity) stage were selected for sample collection according to soybean development stage criteria published on SoyBase (https://www.soybase.org/). The axillary bud and COT and EM stages are cell division stages, while the MM and LM stages are cell expansion stages (SoyBase ontology). Total RNA was extracted from different tissues of soybean by TRIzol reagent (Ambion). Oligo 7.57 software was used to design gene primers, and GmACTIN4 (GenBank login number: AF049106) was used as an internal control. Information concerning the primers is listed in Table S3.

Haplotype analysis

Two hundred fifty-three soybean germplasm resources from across China were used for haplotype analysis; the planting methods and field management practices were the same as those for the mapping population. The results of the germplasm resource resequencing (10×sequencing) and the genomic sequence information were used for this study, and the upstream (3000 bp), the genomic sequences, and the downstream (2000 bp) of the four candidate genes were extracted from the Phytozome 13 database (https://phytozome-next.jgi.doe.gov/). The DNAsp 5.0 software was used for analyzing the haplotype distribution of the genomic sequence of the candidate genes in the germplasm resource population (Librado and Rozas, 2009); the haplovie (https://haplview.software.informer.com/4.2/) was used for screening the excellent haplotypes (the number of cultivars belonging to the haplotype exceeding 5% of the total). SPSS 17.0 software (Armonk, New York, USA) was used to determine the effects of each haplotype on the phenotype. PLACE (plant cis-acting regulatory DNA elements (https://www.dna.afrc.go.jp/PLACE/)) was used to query the function of plant promoter elements.

Results

Phenotypic analysis and segregation analysis

In 2017, the phenotypic data of 1168 individuals in the backcross population were obtained. The ratio of four-seeded pods to total pods ranged from 0 to 40.54%, with an average of 8.15%, and the standard deviation was 6.00%. The average ratio of four-seeded pods to the number of pods per plant ranged from 35.00 ~ 41.00%; there were 512 individuals whose ratio of four-seeded pods was within 8 ~ 25% and 124 individuals without four-seeded pods (Table 1, Fig. 1A).

There were 113 individuals in the R1 fine mapping population (line 337; line 1160), and the ratio of the number of four-seeded pods to the number of pods per plant ranged from 0.00 to 37.84%, with a standard deviation of 8.38% (Fig. 1B). There were 137 individuals in the R2 population (line 18, line 108, line 554, line 614), and the ratio of the number of four-seeded pods to the number of pods per plant ranged from 0.00 to 37.84%, with a standard deviation of 8.38% (Fig. 1B). There were 137 individuals in the R3 population (line 18, line 108, line 554, line 614), and the ratio of the number of four-seeded pods to the number of pods per plant ranged from 35.00 ~ 41.00%; there were 512 individuals whose ratio of four-seeded pods was within 8 ~ 25% and 124 individuals without four-seeded pods (Table 1, Fig. 1A).

Table 1 Statistics table of the phenotypes of the preliminary mapping population and secondary mapping population

| Population name      | Population | Quantity | Average (%) | SD   | Variance  | Skewness | Min (%) | Max (%) |
|----------------------|------------|----------|-------------|------|-----------|----------|---------|---------|
| CSSL860-13/SN14 F2   | 1168       | 8.15     | 6.00        | 36.07| 1.00      | 0.00     | 40.54   |
| R1 F3                | 113        | 14.04    | 8.38        | 70.25| 0.32      | 0.00     | 37.84   |
| R2 F3                | 137        | 10.91    | 7.77        | 60.43| 0.85      | 0.00     | 43.59   |
| R3 F3                | 84         | 14.47    | 8.34        | 69.56| 1.64      | 0.00     | 57.89   |

aSD: standard deviation of the phenotypic trait.
bSkewness: skewness of the phenotypic trait.
cMin (%): the minimum of the phenotypic data from the population.
dMax (%): the maximum of the phenotypic data from the population.
pods to the number of pods per plant ranged from 0.00 to 43.59%, with a standard deviation of 7.77% (Fig. 1C). There were eighty-four individuals in the R3 population (line 271; line 345; line 712) and the ratio of the number of four-seeded pods to the number of pods per plant ranged from 0.00 to 57.89%, with a standard deviation of 8.34%. These results are consistent with a normal distribution and are suitable for phenotypic and genotype association analyses (Fig. 1D, Table 1).

The results show that the optimal model of the F2 population was 2MG-AD, indicating that there are two major genes in the population. The variance of the main gene is 8.58, and the heritability of the main gene is 0.75; the Smirnov test ($nW^2$) result was 0.00, and the probability of the Kolmogorov test ($Dn$) was 1. The optimal model of R1 was 2MG-A, indicating that there are two major genes in this population. The variance of the main gene was 57.44, and the heritability of the main gene is 0.81; the Smirnov test ($nW^2$) result was 0.73, and the probability of the
Kolmogorov test \((Dn)\) was 1. The optimal model of \(R2\) was 2MG-AD, indicating that there are two major genes in the population. The variance of the main gene is 37.31, and the heritability of the main gene is 0.61; the Smirnov test \((nW^{2})\) result was 0.73, and the Kolmogorov test \((Dn)\) was 1. The optimal model of \(R3\) was 2MG-EEAD, indicating that there are two major genes in the population. The variance of the main gene is 47.51, and the heritability of the main gene is 0.01; the Smirnov test \((nW^{2})\) result was 0.71, and the Kolmogorov test \((Dn)\) was 1 (Table 2).

QTL analysis of the \(F_2\) population

The resequencing results of the population parents (CSSL860-13 and SN14) showed that there were heterozygous segments on Gm03, Gm06, Gm08, and Gm14. The heterozygous fragments are 0.26 Mb, 0.21 Mb, 4.28 Mb, 5.86 Mb, and 4.50 Mb, respectively. ZYD00006 homozygous fragments were present in Gm07 and Gm13, and the homozygous fragments were 1.28 Mb and 1.95 Mb, respectively (Figure. S1).

Sixty individuals \((F_2)\) were screened for genotypic analysis; among them, 30 individuals had a high ratio of four-seeded pods to the number of pods per plant (ranging from 20.00 to 40.54%), and another 30 individuals did not have four-seeded pods. Sixty-five pairs of SSR markers were screened from within the ZYD00006 homozygous fragment and the heterozygous fragment, and the intervals were approximately 0.13–0.41 Mb. SN14 and ZYD00006 were used for identifying polymorphic SSR markers, and 29 pairs of SSR markers with polymorphisms between the parents were identified. Sixteen pairs of SSR markers with polymorphisms in five heterozygous fragments, of Gm03 (3,957,733–4,224,480 bp), Gm06 (48,413,650–48,625,035 bp), Gm08 (3,398,336–7,680,873 bp), Gm08 (11,853,745–17,718,597 bp), and Gm14 (8,151,693–12,658,057 bp) were used to determine the genotype of 60 individuals \((F_2)\) (Figure. S2). The results showed that there was no significant correlation between the genotype of individuals with a high ratio and the genotype of the individuals with a low ratio. Single-marker analysis showed that the \(P\) values of all the 16 SSR markers were greater than 5%, which were not significant, and the possibility of the five heterozygous segments associated with the NFSP could be excluded (Table S4).

Six SSR markers with polymorphisms in the homozygous fragments of Gm13 (23,349,075–24,544,199 bp) were used to determine the genotype of 60 individuals \((F_2)\), and a genotypic map was created (Figure. S3). The results showed that there was no significant correlation between the genotype of the individuals with a high ratio and the genotype of the individuals with a low ratio. Single-marker analysis showed that the \(P\) values of all six SSR markers were greater than 5%, which was not significant, and the possibility that the Gm13 homozygous fragment (23,349,075–24,544,199 bp) was associated with the NFSP was excluded (Table S4).

Nine SSR markers with polymorphisms in the homozygous fragment of Gm07 (36,116,118–37,399,738 bp) were used to determine the genotype of 60 individuals \((F_2)\), and a genotypic map was created (Figure. S3). The results showed that there was a significant correlation between the genotype of the individuals with a high ratio and the genotype of the individuals with a low ratio. Single-marker analysis showed that four SSR markers, SSR-07–1224, SSR-07–1248, SSR-07–1254, and SSR-07–1257, have high significance. The interval

| Population name | Population | Model | AIC  | Major gene var | Heritability (Major gene) | \(P\) \((nW^{2})\) | \(P\) \((Dn)\) |
|-----------------|------------|-------|------|----------------|--------------------------|----------------|----------------|
| CSSL860-13/SN14 | \(F_2\)    | 2MG-AD| 5968.00 | 8.58          | 0.75                     | 0.00            | 1              |
| R1              | \(F_3\)    | 2MG-A | 800.87  | 57.44         | 0.81                     | 0.96            | 1              |
| R2              | \(F_3\)    | 2MG-AD| 954.53  | 37.31         | 0.61                     | 0.73            | 1              |
| R3              | \(F_3\)    | 2MG-EEAD| 36.59  | 47.51         | 0.01                     | 0.71            | 1              |

Table 2: Separation and analysis of the four-seeded-pod trait in the \(F_2\) population and \(F_3\) population
coincides with Gm07 (36,731,771–37,409,044 bp), and the distance is 0.67 Mb. (Table S4).

QTL analysis of the F₃ population

Seven SSR markers with polymorphisms, SSR-07–1224, SSR-07–1235, SSR-07–1248, SSR-07–1253, SSR-07–1254, SSR-07–1256, and SSR-07–1257, were screened for genotypic analysis of 13 subpopulations (F₃). Single-marker analysis showed that SSR-07–1224, SSR-07–1235, SSR-07–1248, SSR-07–1253, and SSR-07–1254 are highly significant in the R1 (line 337, line 1160) subpopulation (Table S5), and the QTL interval was thus shortened to 656 kb.

Based on the above mapping results, nine SSR markers with polymorphisms, SSR-07–1225, SSR-07–1226, SSR-07–1228, SSR-07–1233, SSR-07–1241, SSR-07–1242, SSR-07–1244, SSR-07–1246 and SSR-07–1247, were added, and a total of sixteen SSR markers were analyzed for the remaining 11 subpopulations. Among them, R2 (line 18, line 108, line 554, line 614) has significance in the interval from SSR-07–1224 to SSR-07–1248, and single-marker analysis showed that SSR-07–1224, SSR-07–1225, SSR-07–1226, SSR-07–1228, SSR-07–1233, SSR-07–1235, and SSR-07–1248 were highly significantly (Table S6). The QTL interval was thus shortened to 546 kb.

To shorten the QTL interval further, the remaining population was analyzed. The results showed that R3 (line 271, line 345, line 712) has a significant mapping interval (SSR-07–1224 ~ SSR-07–1235). Single-marker analysis showed that SSR-07–1224, SSR-07–1225, SSR-07–1226, SSR-07–1228, SSR-07–1233, and SSR-07–1235 were

---

Fig. 2 Mapping interval of QNFSP07-I for four-seeded pods in soybean
significant (Table S7), and the QTL interval was therefore shortened to 321 kb. Based on the above results, as shown in Fig. 2, the QTL associated with the number of four-seeded pods located on Gm07 ranged from 36,731,771 bp to 37,052,836 bp, and the QTL (QNFSP07-1) interval was delimited to 321 kb.

Candidate gene mining in QNFSP07-1 for soybean NFSP

The marker sequences of QNFSP07-1 were obtained using SoyBase (http://www.soybase.org/), and candidate genes were detected and annotated based on a BLAST search of publicly available resources, and 27 genes were identified (Table S8). Fourteen genes had amino acid mutations according to the SNP sequences of the parental SN14 and ZYD000006 sequences via resequencing, and 13 genes have indels between parents in the Intron region and 5' UTR region, no amino acid mutations among these indels (Table S9) Glyma.07G200400 and Glyma.07G201000 were not expressed. Glyma.07G200200 encodes an HSP20-like protein of the chaperone superfamily, whose members are involved in the drought stress response (Gugger et al. 2017). Glyma.07G201500 contains the DUF679 protein domain and decreases downstream gene expression during flower senescence (Gao et al. 2018). Glyma.07G199700 contains the ERF protein 9 domain, and the amino acids valine and glutamate, which are retained in the ERF protein domain, play an important role in drought and low-temperature stress responses (Gao et al. 2018). Glyma.07G199800 contains a MAC/perforin domain, which directly interacts with the cell membrane of migrating neurons (Gorbushin, 2016). Glyma.07G200800 contains a zf-Di19 superfamily domain, which is related to the plant drought stress response (Lu et al. 2007). Glyma.07G201300 was annotated as being associated with the glucose-6-phosphate/phosphoric acid transporter; the homologous gene in Arabidopsis is AT5G46110, which is downregulated gene in response to VVZFPL overexpression, and overexpression of VVZFPL affects plant photosynthesis and slows plant growth (Kobayashi et al. 2012). Glyma.07G200900 was annotated as an OB-folding protein according to KEGG pathway analysis (pathway_id k07466); this protein is related to litchi flowering defects (Lu et al. 2014). Glyma.07G201200 was annotated as homeobox 1, and the homologous gene in Arabidopsis is AT1G28420, which is a membrane-tethered transcription factor (MTTF) belonging to the bZIP family (Chen et al. 2008). Previous studies have shown that members of the bZIP gene family are TFs (transcription factors); bZIPS are ubiquitous within the genome of green plants, and they affect organ and tissue differentiation and regulate cell elongation (Smitha et al. 2017; Corrêa et al. 2008). Glyma.07G201400 was annotated as a cysteine-tRNA synthase, and the homologous gene in Arabidopsis is AT2G31170, which is a regulatory gene encoding aminoacyl-tRNA synthase. Studies have shown that interruption of chloroplast translation affects gametophytes in the seed during embryogenesis, and interruption of mitochondrial translation leads to ovule abortion and cessation of male–female gametogenesis (Berg et al. 2005). Glyma.07G201600 was annotated as encoding the HAESA-like protein 1, and the homologous gene in Arabidopsis is AT1G28440. The HAESA coreceptor SERK1 is a positive regulator of the flower abscission pathway, which allows high-affinity sensing of peptide hormones by binding to the Arg-His-Asn motif in IDA. This sequence pattern is conserved among various plant peptides, indicating that plant peptide hormone receptors share common ligand binding patterns and activation mechanisms. Leucine-rich repeat receptor kinases (LRR-RKs), HAE-SAs, and the peptide hormone IDA control flower abscission (Santiago et al. 2016). Combinations of the results of previous studies have shown that the development of ovules, embryos, pollen, stamens, and flowers has a great influence on the number of pods and seeds (Kurdyukov et al. 2014; Jiang et al. 2015).

Glyma.07G200900 and Glyma.07G201200 have higher expression in pod (Fig. 3A), and Glyma.07G200900 and Glyma.07G201200 have a positive regulatory effect in the Hr stage (Fig. 3B). Glyma.07G200900 and Glyma.07G201200 were selected as candidate genes in this study, and additional studies were carried out.

Candidate gene expression analysis

The expression of two candidate genes, Glyma.07G200900 and Glyma.07G201200, at
Fig. 3  Heatmap profiles of the candidate genes in tissues and different stages. 
A Heatmap profiles of the candidate genes in tissues. 
B Heatmap profiles of the candidate genes in different stages.
different growth and development stages of soybean (florescence, axillary bud, COT, EM, MM, and LM stages) was measured via qRT-PCR. The selected materials included line 953 (high ratio of four-seeded pods to the number of pods per plant), line 32 (low ratio of four-seeded pods to the number of pods per plant), and SN14 (control). The results showed that the expression of the two candidate genes in the cell division stage (axillary bud, COT, and EM stages) was higher than that in the cell expansion stage (MM and LM stages). The highest expression occurred in the EM stage, followed by the COT stage, axillary bud stage, florescence stage, and MM stage. The lowest expression occurred in the LM stage, indicating that the expression decreased with the development of organs (Fig. 4). The candidate genes may affect cell division rather than cell expansion and affect ovule number by regulating cell division, which is related to the formation of four-seeded pods in soybean. The expression of the two candidate genes in the three experimental materials was the highest in the EM stage and lowest in the LM stage. The expression was associated with a high ratio of the number four-seeded pods to the number of pods per plant (higher than that of SN14) and a low ratio of the number four-seeded pods to the number of pods per plant, indicating that these two candidate genes have a positive regulatory effect on the NFSP of soybean. Therefore, Glyma.07G200900 and Glyma.07G201200 can be used as candidate genes for the number of four-seeded pods of soybean, laying a foundation for future gene cloning and functional verification.

Haplotype analysis of candidate genes

Two hundred fifty-three germplasm resources (Table S10) were used for haplotype analysis, which differed significantly in their NFSP and met the requirements for haplotype analysis of the candidate genes. Whole-genome resequencing of the ninety-two germplasm resources was carried out in 2019, and the phenotype statistics are shown in Table S11.

We selected the full-length sequences of four candidate genes and their upstream 3000 bp promoter region sequence and 2000 bp downstream in the Phytozome database as analysis objects, and DNAsp 5.0 was used to detect haplotype polymorphisms. The results showed that Glyma.07G201200 has 2 excellent haplotypes, while Glyma.07G200900 has only one excellent haplotype in the germplasm resource. ANOVA was used to analyze the differences between the phenotypes of the excellent haplotypes, and significant differences in NFSP were showed only one excellent haplotype of Glyma.07G200900 (Table S12).

Glyma.07G201200 has two excellent haplotypes: HA-1 and HA-2. There are nine SNP differences between HA-2, nine SNP differences in the promoter region sequence (V186, V581, V2356, V2625, V2656, V2713, V2720, V2787, V2940), four SNP differences in the CDS region (V8713, V10660, V12131, V12448), and one difference in the genomic intron region (V13150). The change in SNP V8713 resulted in an amino acid change (D > E) and the change in SNP V12131 resulted in an amino acid change (K > Q), potentially leading to changes in the structure and function of the

Fig. 4 Relative expression of Glyma.07G200900 and Glyma.07G201200, in different stages. A qRT-PCR analysis of Glyma.07G200900 expression. B qRT-PCR analysis of Glyma.07G201200 expression. The bars represent the means ± SE (standard errors) of three replications. The different letters represent significant differences (p ≤ 0.05)
gene (Fig. 5A). There are multiple SNP differences between HA-1 and HA-2, of which SNPV2713 leads to EECCRCAH1 element according to PLACE. With respect to NFSP based on phenotypic values and haplotypes, ANOVA showed that there were significant differences between HA-1 and HA-2, and HA-2 was the haplotype with a high number of four-seeded pods. HA-1 was the haplotype with a low number of four-seeded pods, with \( P \) values of 0.046 (Fig. 5B.)

**Discussion**

Molecular marker–assisted selection has resulted in the detection of many QTLs of different quantitative traits, including physiological, morphological, and quality-related traits, as well as yield-related traits such as the soybean seed filling rate, salt tolerance, seed yield, and protein content (Varshney et al. 2005; Reinprecht et al. 2006; Guzman et al. 2007; Jun et al. 2008; Palomeque et al. 2009). Soybean NFSP is the key factor in determining yield, and it has also been highly conserved in the process of crop domestication. Molecular markers were used to locate the QTLs and genes related to the NFSP in soybean, determine the chromosomes and positions where they are located, clarify their genetic effects, analyze the gene functions, and determine the relationship between QTLs and the genes to determine their role in the formation of four-seeded pods to increase the NFSP for soybean genetic improvement.

Eighty-six QTLs related to the NFSP of soybean have been published, and the mapping results are distributed across all 20 chromosomes of soybean (Zhang et al. 2012). During research involving marker-assisted breeding, the main effect of QTLs is very important for the study of soybean agro-nomic characteristics, and most researchers have focused on determining how to obtain main-effect QTLs. In this study, an NFSP-related QTL was preliminarily mapped onto Gm07 between BARCSOYSSR_07_1224 and BARCSOYSSR_07_1257, ranging from 36,731,771 to 3,740,044 bp. SSR primer encryption was applied to delineate the target segment; the QTL segment was shortened to the SSR marker between BARCSOYSSR_07_1224 and BARCSOYSSR_07_1235, ranging from 36,731,771 to 37,052,836 bp, and the segment was 321 kb. A previous study mapped a QTL related to the number of four-seeded pods, which was located on Gm07, from 37,079,608 to 40,919,337 bp; our region overlapped with this initial QTL (Kuroda et al. 2013), and it was concluded that the QTL identified in this study had certain authenticity. To date, the \( Ln \) gene controlling four-seeded pods has been found and studied extensively. In 2013, using map-based, \( Ln \) was mapped to 57.3 kb of soybean Gm20 (Fang et al. 2013). The QTL obtained with \( Ln \) is important for further study of the number of four-seeded pods of soybean in this study and provides excellent population material support for gene mining and functional research.

The use of a large number of plant populations segregating for the initially mapped QTL in high-density
linkage analysis is an effective method for QTL fine mapping. There are two main advantages of using an RHL population for fine mapping. First, when developing a population for fine mapping, only one line from the F2 population needs to be selected based on its genotype, and backcrossing and selection based on DNA markers or phenotypes are not needed. The genotype composition of the RHL can then be determined by checking only the genotypic data of the linkage map, and there is no need to analyze the genotype of any marker. In 1997, an RHL population was first used for QTL fine mapping. At present, this technique has been widely used for fine mapping of related traits in soybean (Yamanaka et al. 2015; He et al. 2009; Suzuki et al. 2010; Yan et al. 2017).

The maximum number of seeds per pod is determined by the number of ovules, and each stage of ovule development also affects seed formation. Many genes are reported to be related to ovule development, including BEL1 (Western and Haughn, 1999); AP2 (Modrusan et al. 1994); BELL (Ray et al. 1994); HLL; and ANT (Schneitz et al. 1998). Many signal pathways involved in seed number determination have also been identified, including BR signaling (Huang et al. 2013), demonstrating that BR signal affects ovule and seed development by regulating BZR1. The number of ovules in ovaries affects the number of seeds per pod, and cell division affects the seed setting rate (Tischber et al. 2003). SHORT INTEGUMENTS1 and 2 (SIN1 and SIN2) are key genes for cell division (Ray et al. 1996), and TSO1 plays an important role in cell elongation (Song et al. 2000). *Glyma.07G200900* is involved in the KEGG pathway (pathway_id k07466), and protein phosphatase 2A participates in cell division, which is the molecular mechanism of DNA: replication, repair, recombination, and chromosome dynamics (Davis et al. 2008). *Glyma.07G201200* was annotated as homeobox 1, and it belongs to the bZIP family; the bZIP protein was reported to be involved in ABA-signaling pathways, suggesting that it can induce cell division (Toshio and Toshiyuki, 2002). In the present study, by analyzing the transcriptome data of four experimental materials from CSSLs and the expression levels in 26 soybean tissues, we identified two key genes that may be affected NFSP. The expression of two genes in the cell division stage was higher than that in the cell expansion stage. These results suggest that the gene may be involved in the cell division stage. The haplotypes of the two genes were analyzed by 253 germplasm resources, and the result shows that *Glyma.07G201200* has two excellent haplotypes, there are differences in the promoter region sequence and CDS region, and resulted the amino acid change. ANOVA showed that the number of four-seeded pods (the ratio of the number of four-seeded pods) among 253 germplasm resources with different excellent haplotypes has significant differences, laying the foundation for our study of the molecular mechanisms of NFSP.

Wild soybean is abundant and widely distributed, with a large number of pods (Jin et al. 2017), and it is not isolated from cultivated soybeans. The excellent traits or genes of wild soybean can be introduced into cultivated soybean during the process of cross breeding. In this study, residual heterozygous individuals were selected from the CSSL population generated by crossing SN14 and ZYD00006. The theoretical background was the SN14 genotype with a short heterozygous interval and few heterozygous loci, which provided a material basis for rapidly mapping the QTL for the soybean four-seeded-pod trait (*QNFSP07-1*). The primary mapping material was developed in 2015, and *QNFSP07-1* was detected in the primary mapping material. Plants heterozygous for the target segment were screened, and a secondary isolated population was developed to gradually complete the QTL fine mapping of the four-seeded-pod trait in soybean.

**Conclusions**

We used CSSL860-13/SN14 (F3) population for QTL mapping, and obtained one QTL, which the QTL interval is 321 kb. According to the gene annotation of the QTL interval, 27 genes were obtained. qRT-PCR was used to measure the expression of the candidate genes at different developmental stages of soybean, and the result showed that *Glyma.07G200900* and *Glyma.07G201200* were identified as candidate genes that affect NFSP. Haplotype analysis of two candidate genes, *Glyma.07G201200*, has two excellent haplotypes, and the significance level between the two excellent haplotypes at \( p < 0.05 \). Those results provide the information for gene map–based cloning and
molecular marker–assisted breeding of the number of four-seeded pod in soybean.

**Abbreviations** CSSL: Chromosome segment substitution line; Dr: The Kolmogorov test; LG: Linkage group; Max: Maximum; Min: Minimum; NFSP: The number of four-seeded pods; *nW*²: The Smirnov test; QTL: Quantitative trait loci; RIL: Recombinant inbred line; SD: Standard deviation

**Author contribution** QSC and HWJ designed the study. YYL, CYL, and NNW analyzed the data and wrote the paper. ZGZ, LLH, DWX, ZMQ, YY, and CDL participated in correcting the manuscript. All authors have read and approved the final manuscript.

**Funding** This study was financially supported by the National Nature Science Foundation of China (31771882, 31801389), “Hundred-thousand and million project of Heilongjiang province for engineering and technology science” (C2018017), the project of science and technology innovation and new cultivar breeding (2019ZX16B01), the project of science and technology for the education department of Jilin province (project number: JJKH20211432KJ), and the National Science Foundation of Heilongjiang (C2018017).

**Declarations**

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare no competing interests.

**References**

Berg M, Rogers R, Muralla R, Meinke D (2005) Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in *Arabidopsis*. Plant J 44:866–878. https://doi.org/10.1111/j.1365-313x.2005.02580.x

Cao XW, Liu B, Zhang YM (2013) SEA: a software package of segregation analysis of quantitative traits in plants. Journal of Nanjing Agricultural University. 6: 5–10. CNKI: SUN: ZGYW.0.2012–01–002

Chen QS, Zhang ZC, Liu CY et al (2007) QTL analysis of major agronomic traits in soybean. Agricultural Sciences in China 6:399–405

Chen YN, Slabaugh E, Brandizzi F (2008) Membrane-tethered transcription factors in *Arabidopsis thaliana*: novel regulators in stress response and development. Curr Opin Plant Biol 11:695–701. https://doi.org/10.1016/j.pbi.2008.10.005

Chen CJ, Chen H, Zhang Y et al (2020) Thtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1–26. https://doi.org/10.1016/j.molp.2020.06.009

Corrêa LG, Riaño-Pachón DM, Schrago CG et al (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. PLoS ONE 3:e2944. https://doi.org/10.1371/journal.pone.0002944

Domingo WE (1945) Inheritance of number of seeds per pod and leaflet shape in the soybean. J Agric Res 70:251–268

Davis AJ, Yan Z, Martinez B, Munby MC (2008) DNA: replication, repair, recombination, and chromosome dynamics: protein Phosphatase 2A is targeted to cell division control protein 6 by a calcium-binding regulatory subunit. J Biol Chem 23:16104–16114

Dinkins RD, Keim KR, Farno L, Edwards LH (2002) Expression of the narrow leaflet gene for yield and agronomic traits in soybean. J Hered 93:346–351. https://doi.org/10.1093/jhered/93.5.346

Eriksson S, Stransfeld L, Adamski NM et al (2010) KLUH/ CYP78A5-dependent growth signaling coordinates floral organ growth in *Arabidopsis*. Curr Biol 20:527–532. https://doi.org/10.1016/j.cub.2010.04.012

Fang C, Li WY, Li GQ et al (2012) QTL analysis of soybean branching using residual heterogeneity in a segregating population. J Genet Genomics 40:93–96. https://doi.org/10.1016/j.jgg.2013.01.002

Frye C, Microsoft Press (2016). Microsoft excel 2016 step by step.

Gao JY, Liu CY, Jiang HW et al. (2012) QTL analysis of pod number per plant in soybean under multiple locations. Chinese J. Catal 34: 1–7. CNKI: SUN: ZGYW.0.2012–01–002

Gao Z, Danev A, Salanenka Y et al (2018) KIRA1 and ORESARA1 terminate flower receptivity by promoting cell death in the stigma of *Arabidopsis*. Nat Plants 4:365–375. https://doi.org/10.1038/s41477-018-0160-7

Gorbushin AM (2016) Membrane attack complex/perforin domain-containing proteins in a dual-species transcriptome of caenogastropoda Littorina littorea and its trematoide parasite Himasthla elongata. Fish Shellfish Immunol 54:254–256. https://doi.org/10.1016/j.fsi.2016.04.015

Grafius JE (1978) Multiple characters and correlated response. Crop Sci 18:931–934. https://doi.org/10.2135/crops ci1978.0011183x001800060004x

Gugger PF, Peñaloza-Ramírez JM, Wright JW, Sork VL (2017) Whole-transcriptome response to water stress in a californian endemic oak. Quercus Lobata Tree Physiol 37:632–644. https://doi.org/10.1093/treephy/tpw122

Guzman PS, Diers BW, Neece DJ et al (2007) QTL associated with yield in three backcross-derived populations of soybean. Crop Sci 47:111–122. https://doi.org/10.2135/crops ci2006.01.0003

He R, Guan R, Liu ZX et al (2009) Mapping the qBN-c1-1 locus to LG C1 for soybean branching using residual heterozygous lines derived from a segregation population. Scientia Agricultura Sinica 42:1152–1157

Huang HY, Jiang WB, Hu YW et al (2013) BR Signal Influences *Arabidopsis* ovule and seed number through regulating related genes expression by BZR1. Mol Plant 6:456–469. https://doi.org/10.1093/mp/sss070

Hwang EY, Song QJ, Jia GF et al (2014) A genome-wide association study of seed protein and oil content in soybean.
BMC Genom 15:1471–2164. https://doi.org/10.1186/s12866-017-0831-1

Ito T, Meyerowitz EM (2000) Overexpression of a gene encoding a cytochrome p450, cry78a9, induces large and seedless fruit in Arabidopsis. Plant Cell 12:1541–1550. https://doi.org/10.1105/tpc.12.1541

Jiang YF, Lahlali R, Karunakaran C et al (2015) Seed set, pollen morphology and pollen surface composition response to heat stress in field pea. Plant Cell Environ 38:2387–2397. https://doi.org/10.1111/pce.12589

Johnson HW, Bernard RL (1962) Soybean genetics and breeding. Advances in Agronomy Academic Press 14:149–221

Jeong N, Suh SJ, Kim M et al (2012) QTL affecting fitness of hybrids between wild and cultivated soybeans in China for accelerating soybean breeding in northeast China 1:12–15

Jin X, Cao F, Xu L et al (2017) On using wild soybean to create breeding resources and new varieties. Agricultural Science in Northeast China 1:12–15

Kobayashi M, Horiuchi H, Fujita K et al (2012) Characterization of grape C-repeat-binding factor 2 and B-box-type zinc finger protein in transgenic Arabidopsis plants under stress conditions. Mol Biol Rep 39:7933–7939. https://doi.org/10.1007/s10053-012-1638-4

Kuroda Y, Kaga A, Tomooka N et al (2013) QTL affecting leaflet shape and number of seeds per pod in soybean. Plant Cell 24:4807–4818. https://doi.org/10.1105/tpc.112.104968

Li CD, Jiang HW, Guo T et al. (2020) QTL mapping and interactive group validation for the number of four-seed-pod trait on soybean. Ecol Evol 3:2150–2168. https://doi.org/10.1002/ece3.606

Kumar B, Talukdar A, Balà I et al (2014) Population structure and association mapping studies for important agronomic traits in soybean. J Genet 93:775–784. https://doi.org/10.1007/s12041-014-0454-0

Kurdyukov S, Song Y, Sheahan MB, Rose RJ (2014) Transcriptional regulation of early embryo development in the model legume Medicago truncatula. Plant Cell Rep 33:349–362. https://doi.org/10.1007/s00122-013-1535-x

Li CD, Jiang HW, Guo T et al. (2020) QTL mapping and interactive group validation for the number of four-seed-pod trait on soybean. 42: 44–50. https://doi.org/10.19802/j.issn.1007-9084.2020013.

Li CD, Jiang HW, Zhang WB et al (2008) QTL analysis of seed and pod traits in soybean. Molecular Plant Breeding 6:1091–1100

Li CD, Z JN, J HW, et al (2018) Identification and validation of number of pod-and seed-related traits qtls in soybean. Plant Breeding 137:730–745. https://doi.org/10.1111/pbr.12635

Lu PL, Chen NZ, An R et al (2007) A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. Plant Mol Biol 63:289–305. https://doi.org/10.1007/s11103-006-9089-8

Li Y, Chang R, Qiu L (2016) Utilizing the diversity of wild soybeans in China for accelerating soybean breeding in the genome era[M].

Lu XY, Kim HJ, Zhong SY et al (2014) De novo transcriptome assembly for rudimentary leaves in litchi chinesis sonn. and identification of differentially expressed genes in response to reactive oxygen species. BMC Genom 15:805–818. https://doi.org/10.1186/s12866-017-0831-1

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

McCouch S (1997) Report on QTL nomenclature. Rice Genet News14:11–13

Modrusan Z, Reiser L, Feldmann KA et al (1994) Homeotic transformation of ovules into carpel like structures in Arabidopsis. Plant Cell 6:333–349. https://doi.org/10.1105/tpc.6.3.333

Ning HL, Yuan JQ, Dong QZ et al (2018) Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [Glycine max (L.) Merril]. PLoS ONE 13:1–15. https://doi.org/10.1371/journal.pone.0195830

Palomeque L, Li JL, Li W et al (2009) QTL in mega-environments: I. universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted×high-yielding exotic soybean lines. Theor Appl Genet 119:417–427. https://doi.org/10.1007/s00122-009-1049-6

Pei RJ, Chen XQ, Sun N, Zhang L (2010) Application of molecular markers in soybean breeding. Anhui Agricultural Sciences. 38: 100–101. CNKI: SUN: AHNY.0.2010–03–035.

Peng YH, Zhu JC, Yang GB, Yuan JZ (1994) Relation of soybean leaf shape distribution to 4-seeded pods. J Crop Sci 20:501–503

Qiu LJ, Chang RZ (2006) Specification and data standard for soybean germplasm resources description[M]. China Agricultural Press, BeiJing

Ray A, Robinson-Beers K, Ray S et al (1994) The Arabidopsis floral homeotic gene BELL (BELI) controls ovule development through negative regulation of AGAMOUS gene (AG). Proc Natl Aca Sci USA 91:5761–5765. https://doi.org/10.1023/b/evol.000002365052

Ray A, Lang Jd, Golden T, Ray S (1996) SHORT INTEGU MENT (SIN1), a gene required for ovule development in Arabidopsis, also controls flowering time. Development 122:2631–2638. https://doi.org/10.1002/(SICI)1097-0248(19990601)25:1%3C104:aid-CYTO12%3E3.0.CO;2-S

Reinprecht Y, Poya VV, Yu K et al (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (Glycine max (L.) Merrill) germplasm. Genome 49:1510–1527. https://doi.org/10.1139/g06-112

Richards R (2000) A. Selectable traits to increase crop photosynthesis and yield of grain crops. J Exp Bot 51:447–458. https://doi.org/10.1093/jexbot/51.suppl_1.447

Sakuma Y, Liu Q, Dubouzet JG et al (2002) DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. Biochem Biophys Res Commun 290:998–1009. https://doi.org/10.1006/bbrc.2001.6299

Santiago J, Brandt B, Wildhagen M et al (2016) Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. Elife 5:e15075. https://doi.org/10.7554/eLife.15075.001
Schneitz K, Baker SC, Gasser CS, Redweik A (1998) Pattern formation and growth during floral organogenesis: heleniumos and antetegmina are required for the formation of the proximal region of the ovule primordium in Arabidopsis thaliana. Development 125:2555–2563. https://doi.org/10.1016/gad.195101

Suzuki M, Fujino K, Nakamoto Y et al (2010) Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. Mol Breeding 25:407–418. https://doi.org/10.1007/s11032-009-9340-5

Sun DS, Li WX, Zhang ZC et al (2006) Quantitative trait loci analysis for the developmental behavior of soybean (Glycine max L. Merr.). Theor Appl Genet 112:665–673. https://doi.org/10.1007/s00122-005-0169-y

Smitha NA, Shah A, Song J, Jameson PE (2017) Differential gene expression in the meristem and during early fruit growth of Pisum sativum L. identifies potential targets for breeding. Int J Mol Sci 18:428–440. https://doi.org/10.3390/ijms18020428

Song JY, Leung T, Ehler LK et al (2000) Regulation of meristem organization and cell division by TSO1, an Arabidopsis gene with cysteine-rich repeats. Development 127:2207–2217. https://doi.org/10.1002/df.1007-0320(200007)127:2207–2217

Takahashi N (1934) Linkage relation between the genes for the forms of leaves and the number of seeds per pod of soybeans. Jpn J Genet 9:208–225. https://doi.org/10.1266/jjg.9.208

Tischber T, Allphin L, Chase K et al (2003) Genetic of seed abortion and reproductive traits in soybean. Crop Sci 43:464–473. https://doi.org/10.2135/cropsci2003.0464

Toshio S, Toshiyuki N (2002) The possible involvement of a phosphate-induced transcription factor encoded by phi-2 gene from tobacco in ABA-signaling pathways. Plant Cell Physiol 1:12. https://doi.org/10.1093/pch/pcf002

Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci 10:621–630. https://doi.org/10.1016/j.tplants.2005.10.004

Wang S C, Basten CJ, Zeng Z B (2006) Windows QTL cartographer 2.5 department of statistics, North Carolina State University, Raleigh, NC. Retrieved from http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Wang XZ, Zhang XJ, Zhou R et al (2007) QTL analysis of pod and seed characters in soybean recombinant inbred lines. J Crop Sci 3:441–448. https://doi.org/10.3321/j.issn:0496-3490.2007.03.014

Wang Q, Guo L, Liu Y et al (2002) Analysis of grey related degree between individual plant yield and main agriculture character of summer soybean. Soybean Bulletin 6:5–12. https://doi.org/10.3969/j.issn.1674-3547.2002.06.005

Wang XB, Li YH, Zhang HW et al (2015) Evolution and association analysis of GmCYP78A10 gene with seed size/weight and pod number in soybean. Mol Biol Rer 42(2):489–496. https://doi.org/10.1007/s11033-014-3792-3

Weiss MG (1970) Genetic linkage in soybeans: linkage group I. Crop Sci 10:69–72

Western TL, Haughn GW (1999) BEL1 and AGAMOUS genes promote ovule identity in Arabidopsis thaliana. Plant J 18:329–336. https://doi.org/10.1046/j.1365-313X.1999.00448.x

Yan L, Zhang YY, Song QJ et al (2017) Validation of the quantitative trait locus underlying soybean plant height using residual heterozygous lines and near-isogenic lines across multi-environments. Euphytica 213:1–11. https://doi.org/10.1007/s10681-016-1833-9

Yamanaka N, Watanabe S, Toda K et al (2015) Fine mapping of the FT1 locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line. Theor Appl Genet 110:634–639. https://doi.org/10.1007/s00122-004-1886-3

Yang Z, Sun YN, Qi ZM et al (2013) Analysis of additive, epistasis and QE interaction of QTLs related to pod number traits in Soybean. J China Agricul Univ 18(1):13. https://doi.org/10.1007/s10681-016-1833-9

Zhang D, Cheng H, Wang H et al (2010) Identification of genomic regions determining flower and pod numbers development in soybean (Glycine max L.). J Genet Genomics 37:545–556. https://doi.org/10.1016/s1673-8527(09)60074-6

Zhang WY, Wang JX, Zong CM et al (2012) QTL analysis of seed and pod traits in soybean. Soybean Science 31:193–197

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.