Fine Mapping of A QTL Locus (QNFSP07-1) And Analysis of Candidate Genes For Four-Seeded Pods In Soybean

Yingying Li (✉ 13009833788@163.com)

Northeast Agricultural University <https://orcid.org/0000-0003-3914-3490>

Chunyan Liu

Northeast Agricultural University

Nannan Wang

Heilongjiang Agricultural Academy of Sciences: Heilongjiang Academy of Agricultural Sciences

Zhanguo Zhang

Northeast Agricultural University

Lilong Hou

Northeast Agricultural University

Dawei Xin

Northeast Agricultural University

Zhaoming Qi

Northeast Agricultural University

Candong Li

Heilongjiang Agricultural Academy of Sciences: Heilongjiang Academy of Agricultural Sciences

Yan Yu

Changchun University of Science and Technology

Hongwei Jiang

Jilin Academy of Agricultural Sciences Changchun Branch: Jilin Academy of Agricultural Sciences

Qingshan Chen

Northeast Agricultural University

Research Article

Keywords: Candidate Genes, Four-Seeded Pod, Fine Mapping, Haplotype Analysis, QNFSP07-1, Soybean

Posted Date: August 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-695131/v1>

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

Version of Record: A version of this preprint was published at Molecular Breeding on November 1st, 2021.
See the published version at <https://doi.org/10.1007/s11032-021-01265-6>.

Abstract

The number of four-seeded pods is a quantitative trait in soybean [*Glycine max* (L.) Merr.] and is closely related to yield in terms of breeding. 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. 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 (36116118-37399738 bp) were screened, and the heterozygous individuals were subjected to inbreeding to obtain 13 F₃ populations, with a target interval of 321 kb. Gene annotation was performed on the finely mapped interval, and 27 genes were obtained. *Glyma.07G200900*, *Glyma.07G201200* were identified as candidate genes. qRT-PCR was used to measure the expression of the candidate genes at different developmental stages of soybean. 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 shows that *Glyma.07G201200* has two excellent haplotypes. Those results provide the information for gene map-based cloning and molecular marker assisted breeding of the number of four-seeded pod in 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 trying to improve the yield and quality of soybean, which is among the economically important crops species. 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 breeding new soybean varieties with multiple 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); 86 of these QTLs were located using a RIL population or an F₂ 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. used 165 recombinant inbred lines (RILs) to analyze the number of four-seeded pods and showed that three QTLs on linkage group (LG) I were associated with the number four-seeded pods. Gao et al. reported eight QTLs on LGs D1a, A1, and H were associated with the number of four-seeded pods in a RIL population. Yang et al. 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 used for QTL mapping, and 21 and 26 pod-related QTLs were detected via complete interval mapping 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 trait, 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 seed number 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 et al., 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 (Takahashi et al., 1934; Domingo et al., 1945; Johnson et al., 1962; Weiss et al., 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 et al., 2000). Three *P450/CYP78A* genes have been reported to be associated with seed size and seed number (Fang et al., 2012). *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 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 to perform gene annotation on this interval to mine candidate genes and analyze them using haplotype analysis and qRT-PCR analysis.

Materials And Methods

Plant materials, planting and experimental design

In this study, CSSLs were constructed from Suinong 14 (the main cultivar used in Heilongjiang, China) and ZYD00006 (a wild soybean resource) and subsequently backcrossed with Suinong 14. In 2006, Suinong 14 was crossed with ZYD00006 to obtain F_1 plants, and in 2007, the F_1 plants were backcrossed with Suinong 14 to obtain BC_1F_1 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 (BC_4F_4 - BC_4F_6). 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 et al., 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 et al., (2006).

The phenotypic data of NFSP were analyzed using the Segregation Analysis (SEA) software (Cao et al., 2013), which 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. Using the least squares method to estimate the gene effect value and genetic variances, and precisely calculating the probabilities in the Smirnov test (nW^2) and Kolmogorov test (Dn).

Smirnov test (nW^2) :

$$P\{nW^2 \leq x\} = \frac{1}{\sqrt{2x}} \sum_{k=0}^{\infty} \frac{F(k+\frac{1}{2})\sqrt{4k+1}}{F(k+\frac{1}{2})F(k+1)} \exp\left\{-\frac{(4k+1)^2}{16x}\right\} \times \left\{I_{-\frac{1}{4}}\left(\frac{(4k+1)^2}{16x}\right) - I_{-\frac{1}{4}}\left(\frac{(4k+1)^2}{16x}\right)\right\} \quad (1)$$

Kolmogorov test (Dn):

$$\lim_{n \rightarrow \infty} P[\sqrt{n} D_n \leq z | H_0] = K_0(z) + \frac{K_1(z)}{n^{1/2}} + \frac{K_2(z)}{n} + \frac{K_3(z)}{n^{3/2}} + O\left(\frac{1}{n^2}\right) \quad (2)$$

Identification of QTLs for the number of four-seeded pods

Parent resequencing performed according to standard protocols provided by Illumina, the amount of analysis data is 34.63 Gbp Clean Data, and the Q30 is 86.73%. The average ratio of sample to reference

genome (Williams82) was 98.64%, the average coverage depth was 10X, and the genome coverage was 97.80%.

The results of the parent resequencing combined with the SSR sequence provided by SoyBase (<https://www.soybase.org/>) were used to synthesize 1050 pairs of SSR marker primers. Primers located on Gm03, Gm06, Gm07, Gm08, Gm13, and Gm14 were selected for QTL mapping, and their sequence information is listed in Supplementary Table A1.

With respect to individual plants that are heterozygous at the segment of the initial QTL interval are screened, self-crossing results in a continuous overlapping arrangement within the separation intervals, generating a fine mapping population. All SSR markers between SSR-07-1224 and SSR-07-1257 on Gm07 were selected for primer encryption analysis, and polyacrylamide gel electrophoresis was used to analyze the genotypes. The relevant information is listed in Supplementary Table A2.

Single-marker analysis via Windows QTL Cartographer (version 2.5) (Wang et al., 2006) was used to analyze the relationships between phenotypes and genotypes by a simple linear regression model, as follows:

$$y = b_0 + b_1x + e \quad (3)$$

By judging whether b_1 and 0 are significantly different, the relationship between the molecular 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 $\Pr(F)$ 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 et al. (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://phytozome.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 Phytozome (Phytozome v12.1: Home ([doe.gov](https://www.phytozome.net/home))), and the transcriptome data of four experimental materials from the CSSLs during three periods (GI: four days after flower open, Hr: seven days after flower open, Co: ten days after flower open) were also analysed: 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%) were used to draw the heatmap with the software TBtools (Chen et al., 2020).

qRT-PCR analysis of candidate genes

With Suinong 14 (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, 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 Supplementary Table A3.

Haplotype analysis

253 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 (3000bp), the genomic sequences and the downstream (2000bp) of the four candidate genes were extracted from the Phytozome 12 database (<https://phytozome.jgi.doe.gov/pz/portal.html#>). 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 et al., 2009), the haploview (<https://haploview.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.affrc.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 9.13%, and the standard deviation was 5.61%. The average ratio of four-seeded pods to total pods of SN14 (the recurrent parent) was 18.39% (the average value of 10 standard lines), with a standard deviation of 0.07%. Statistical analysis showed that the 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, Figure. 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% (Figure. 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 43.59%, with a standard deviation of 7.77% (Figure. 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 (Figure. 1D, Table 1).

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	F ₂	1168	8.15	6	36.07	1.00	0.00	40.54
R1	F ₃	113	14.04	8.38	70.25	0.32	0.00	37.84
R2	F ₃	137	10.91	7.77	60.43	0.85	0.00	43.59
R3	F ₃	84	14.47	8.34	69.56	1.64	0.00	57.89
^a SD: Standard deviation of the phenotypic trait								
^b Skewness: Skewness of the phenotypic trait								
^c Min (cm): The minimum of the phenotypic data from the population								
^d Max (cm): The maximum of the phenotypic data from the population								

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).

Table 2
Separation and analysis of the four-seeded-pod trait in the F₂ population and F₃ population

Population name	Population	Model	AIC	Major Gene Var	Heritability (Major Gene)	P(<i>nW</i> ²)	P(<i>Dn</i>)
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

QTL analysis of the F₂ 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 (Supplementary Table A4, Figure. 2).

Sixty individuals (F₂) 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 20.00–40.54%), 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 (3957733–4224480 bp), Gm06 (48413650–48625035 bp), Gm08 (3398336–7680873 bp), Gm08 (11853745–17718597 bp) and Gm14 (8151693–12658057 bp) were used to determine the genotype of 60 individuals (F₂) (Figure. 3). 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 3).

Table 3
Single-marker analysis of the F₂ population

Linkage group	Marker	P value	Significance	Linkage group	Marker	P value	Significance ^a
Gm03	SSR_03_0222	0.504	---	Gm13	SSR_13_0853	1.000	---
Gm03	SSR_03_0235	0.778	---	Gm13	SSR_13_0854	0.318	---
Gm03	SSR_03_0239	1.000	---	Gm13	SSR_13_0864	1.000	---
Gm06	SSR_06_1716	0.369	---	Gm13	SSR_13_0867	1.000	---
Gm06	SSR_06_1718	0.320	---	Gm13	SSR_13_0868	0.307	---
Gm08	SSR_08_0363	0.816	---	Gm13	SSR_13_0877	1.000	---
Gm08	SSR_08_0398	0.491	---	Gm07	SSR_07_1183	0.788	---
Gm08	SSR_08_0617	0.215	---	Gm07	SSR_07_1212	1.000	---
Gm08	SSR_08_0652	0.413	---	Gm07	SSR_07_1224	0.000	***
Gm08	SSR_08_0758	1.000	---	Gm07	SSR_07_1248	0.000	***
Gm08	SSR_08_0881	1.000	---	Gm07	SSR_07_1254	0.000	***
Gm08	SSR_08_0895	0.337	---	Gm07	SSR_07_1257	0.000	***
Gm08	SSR_08_0923	0.567	---	Gm07	SSR_07_1260	0.135	---
Gm08	SSR_08_0941	1.000	---	Gm07	SSR_07_1262	1.000	---
Gm14	SSR_14_0603	0.803	---	Gm07	SSR_07_1358	1.000	---
Gm14	SSR_14_0616	0.465	---				
^a Significance is indicated by *, **, ***, and **** at the levels of 5%, 1%, 0.1%, and 0.01%, respectively.							

Six SSR markers with polymorphisms in the homozygous fragments of Gm13 (23349075-24544199 bp) were used to determine the genotype of 60 individuals (F₂), and a genotypic map was created (Figure. 4). 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 (23349075-24544199 bp) was associated with the NFSP was excluded (Table 3).

Nine SSR markers with polymorphisms in the homozygous fragment of Gm07 (36116118-37399738 bp) were used to determine the genotype of 60 individuals (F₂), and a genotypic map was created (Figure. 4). 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 coincides with Gm07 (36731771-37409044 bp), and the distance is 0.67 Mb. (Table 3).

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 4), and the QTL interval was thus shortened to 656 kb.

Table 4
Single-marker analysis of the R1 population

Linkage group	Marker	P value	Significance
Gm07	SSR_07_1224	0.000	***
Gm07	SSR_07_1235	0.001	***
Gm07	SSR_07_1248	0.001	***
Gm07	SSR_07_1253	0.000	****
Gm07	SSR_07_1254	0.000	***
Gm07	SSR_07_1256	0.723	—
Gm07	SSR_07_1257	0.493	—

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, SSR-07-1248 were highly significantly (Table 5). The QTL interval was thus shortened to 546 kb.

Table 5
Single-marker analysis of the R2 population

Linkage group	Marker	P value	Significance
Gm07	SSR_07_1224	0.000	***
Gm07	SSR_07_1225	0.000	****
Gm07	SSR_07_1226	0.000	****
Gm07	SSR_07_1228	0.000	****
Gm07	SSR_07_1233	0.000	****
Gm07	SSR_07_1235	0.000	****
Gm07	SSR_07_1241	0.000	****
Gm07	SSR_07_1242	0.000	****
Gm07	SSR_07_1244	0.000	***
Gm07	SSR_07_1246	0.000	***
Gm07	SSR_07_1247	0.000	***
Gm07	SSR_07_1248	0.000	****
Gm07	SSR_07_1253	0.085	—
Gm07	SSR_07_1254	0.409	—
Gm07	SSR_07_1256	0.061	—
Gm07	SSR_07_1257	0.355	—

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 significant (Table 6), and the QTL interval was therefore shortened to 321 kb. Based on the above results, as shown in Figure. 5, the QTL associated with the number of four-seeded pods located on Gm07 ranged from 36731771 bp to 37052836 bp, and the QTL (*QNFSP07-1*) interval was delimited to 321 kb.

Table 6
Single marker analysis of the R3 population

Linkage group	Marker	P value	Significance
Gm07	SSR_07_1224	0.000	****
Gm07	SSR_07_1225	0.034	*
Gm07	SSR_07_1226	0.000	****
Gm07	SSR_07_1228	0.000	****
Gm07	SSR_07_1233	0.000	***
Gm07	SSR_07_1235	0.025	*
Gm07	SSR_07_1241	1.000	---
Gm07	SSR_07_1242	1.000	---
Gm07	SSR_07_1244	1.000	---
Gm07	SSR_07_1246	1.000	---
Gm07	SSR_07_1247	1.000	---
Gm07	SSR_07_1248	1.000	---
Gm07	SSR_07_1253	1.000	---
Gm07	SSR_07_1254	0.150	---
Gm07	SSR_07_1256	1.000	---
Gm07	SSR_07_1257	1.000	---

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 (Supplementary Table A5). Fourteen genes had amino acid mutations according to the SNP sequences of the parental SN14 and ZYD00006 sequences via resequencing. *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 (Sakuma et al., 2002). *Glyma.07G199800* contains a MAC/perforin domain, which directly interacts with the cell membrane of migrating neurons (Gorbushin et al., 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 that 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), HAESAs 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 (Figure. 6A), and *Glyma.07G200900* and *Glyma.07G201200* have a positive regulatory effect in the Hr stage (Figure. 6B). *Glyma.07G200900*, *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*, *Glyma.07G201200* at 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 (Figure. 7). 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*, *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

253 germplasm resources (Supplementary Table A6) 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 Supplementary Table A7.

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, significant differences in NFSP was showed only one excellent haplotype of *Glyma.07G200900* (Supplementary Table A8).

Glyma.07G201200 has two excellent haplotypes: HA-1 and HA-2. There are nine SNP differences between HA-2, 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 gene (Figure. 8A). There are multiple SNP differences between HA-1 and HA-2, of which SNPV2713 lead to EECCRAH1 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 (Figure. 8B.)

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 agronomic 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 36731771 bp to 3740044 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 36731771 bp to 37052836 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 37079608 bp to 40919337 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 is 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 et al., 1999); AP2 (Modrusan et al., 1994); BELL (Ray et al., 1994); HLL and ANT (Schneitz et al., 1998). Many signalling pathways involved in seed number determination have also been identified, including BR signaling (Huang et al., 2013), demonstrating that BR signalling 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), protein phosphatase 2A participate 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-signalling pathways, suggesting that it can induce cell division (Toshio et al., 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 affect 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, 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 and. was used to measure the expression of the candidate genes at different developmental stages of soybean, and the result showed that Glyma.07G200900, Glyma.07G201200 were identified as candidate genes that affect NFSP. 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

D_n 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

Declarations

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’ soybean breeding technology innovation and new cultivar breeding–2019ZX16B01–The

project of science and technology for the education department of Jilin province(Project number:JJKH20211432KJ)the Natural Science Foundation of Heilongjiang (C2018017)

Author Contributions: QSC, 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.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests

References

1. 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. doi:10.1111/j.1365-313x.2005.02580.x
2. Cao XW, Liu B, Zhang YM (2013) SEA: a software package of segregation analysis of quantitative traits in plants. *Journal of Nanjing Agricultural University* 06:5–10. doi: CNKI: SUN: NJNY.0.2013-06-001
3. Chen Q, Zhang Z, Liu C, Xin D, Shan D, Qiu H, Shan C (2007) QTL analysis of major agronomic traits in soybean. *Agricultural Sciences in China* 6(4):399–405
4. Chen Y-N, Slabaugh E, Brandizzi F (2008) Membrane-tethered transcription factors in *Arabidopsis thaliana*: novel regulators in stress response and development. *Current opinion in plant biology* 11. 695–701. doi: 10.1016/j.pbi.2008.10.005
5. Chen C, Chen H, Zhang Y, Thomas HR, Xia R (2020) Tbtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 13(8):1–26. doi:10.1016/j.molp.2020.06.009
6. Corrêa L-G, Riaño-Pachón D-M, Schrago C-G, Santos R-V, Mueller R-B, Vincentz M (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PloS one* 3:e2944. doi: 10.1371/journal.pone.0002944Domingo, W. -E. (1945). Inheritance of number of seeds per pod and leaflet shape in the soybean. *J Agric Res* 70:251–268
7. Davis A-J, Yan Z, Martinez B, Munby M-C (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, 23. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, pp 16104–16114
8. Dinkins R-D, Keim K-R, Farno L, Edwards L-H (2002) Expression of the narrow leaflet gene for yield and agronomic traits in soybean. *J Hered* 93:346–351. doi:10.1093/jhered/93.5.346
9. Du J, Li WX, Dong QZ, Wang P, Su DQ, Sun MM et al (2019) *Soybean Science* 38 (3),360–370. doi: 1011861 /jissn. 1000–9841. 2019. 03. 0360
10. Eriksson S, Stransfeld L, Adamski N-M, Breuninger H, Lenhard M (2010) KLUH/CYP78A5-dependent growth signaling coordinates floral organ growth in *Arabidopsis*. *Current biology* 20. 527–532. doi: 10.4161/psb.5.8.12221Fang, C., Li, W., Li, G., Wang, Z., Zhou, Z., Ma, Y., and Yang, W. (2013). Cloning of

- Ln* gene through combined approach of map-based cloning and association study in soybean. *Journal of genetics and genomics* 40, 93–96
11. Frye C, and Microsoft Press (2016) Microsoft excel 2016 step by step
 12. Gao J-Y, Liu C-Y, Jiang H-W, Hu G-H, Chen Q-S (2012) QTL analysis of pod number per plant in soybean under multiple locations. *Chinese Journal of oil crop sciences* 34:1–7 doi: CNKI: SUN: ZGYW.0.2012-01-002
 13. Gao Z, Daneva A, Salanenka Y, Durme M-V, Huysmans M, Lin Z-C et al (2018) KIRA1 and ORESARA1 terminate flower receptivity by promoting cell death in the stigma of *Arabidopsis*. *Nat Plants* 4:365–375. doi:10.1038/s41477-018-0160-7
 14. Gorbushin AM (2016) Membrane attack complex/perforin domain-containing proteins in a dual-species transcriptome of caenogastropoda *Littorina littorea* and its trematode parasite *Himasthla elongata*. *Fish Shellfish Immunol* 54:254–256. doi:10.1016/j.fsi.2016.04.015
 15. Grafius JE (1978) Multiple characters and correlated response. *Crop Sci* 18:931–934. <https://doi.org/10.2135/cropsci1978.0011183X001800060004x>
 16. Gugger P-F, Peñaloza-Ramírez J-M, Wright J-W, Sork V-L (2017) Whole-transcriptome response to water stress in a california endemic oak, *Quercus lobata*. *Tree physiology* 37:632–644. doi:10.1093/treephys/tpw122
 17. Guzman PS, Diers BW, Neece DJ, Martin St, LeRoy SK, Grau AR, C. R., and Nelson RL (2007) QTL associated with yield in three backcross-derived populations of soybean. *Crop Sci* 47:111–122. doi:10.2135/cropsci2006.01.0003 doi
 18. He R, Guan R, Liu Z, Zhu X, Chang R, Qiu L (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
 19. Huang H, Jiang W, Hu Y, Wu P, Zhua J, Liang W et al (2013) BR Signal Influences *Arabidopsis* Ovule and Seed Number through Regulating Related Genes Expression by BZR1. *Molecular Plant* 6(2):456–469. doi:10.1093/mp/sss070
 20. Hwang EY, Song Q, Jia G, Specht JE, Cregan PB (2014) A genome-wide association study of seed protein and oil content in soybean. *BMC Genom* 15:1471–2164. doi:10.1186/1471-2164-15-1
 21. Ito T, Meyerowitz EM (2000) Overexpression of a gene encoding a cytochrome p450, *cyp78a9*, induces large and seedless fruit in *Arabidopsis*. *Plant Cell* 12(9):1541–1550
 22. Jiang Y, Lahlali R, Karunakaran C, Kumar S, Davis AR, Bueckert RA (2015) Seed set, pollen morphology and pollen surface composition response to heat stress in field pea. *Plant Cell Environ* 38:2387–2397. doi:<https://doi.org/10.1111/pce.12589>
 23. Johnson H-W, Bernard R-L (1962) Soybean genetics and breeding. *Advances in agronomy*. Academic Press 14:149–221
 24. Jeong N, Suh SJ, Kim MH, Lee S, Moon JK, Kim HS et al (2012) *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* 24(12):4807–4818. doi:10.1105/tpc.112.104968
 25. Jin X, Cao F, Xu L, Yin D, Wang H, Piao L (2017) On using wild soybean to create breeding resources and new varieties. *Agricultural Science in Northeast China* 1:12–15

26. Jun TH, Van K, Kim MY, Lee SH, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162:179–191. doi:<https://doi.org/10.1007/s10681-007-9491-6>
27. Kobayashi M, Horiuchi H, Fujita K, Takuhara Y, Suzuki S (2012) Characterization of grape C-repeat-binding factor 2 and B-box-type zinc finger protein in transgenic *Arabidopsis* plants under stress conditions. *Molecular biology reports* 39:7933–7939. doi:10.1007/s11033-012-1638-4
28. Kuroda Y, Kaga A, Tomooka N, Yano H, Takada Y, Kato S, Vaughan D (2013) QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. *Ecology evolution* 3:2150–2168. doi:<https://doi.org/10.1002/ece3.606>
29. Kurdyukov S, Song Y, Sheahan MB, Rose RJ (2014) Transcriptional regulation of early embryo development in the model legume *Medicago truncatula*. *Plant cell reports* 33:349–362. doi:10.1007/s00299-013-1535-x
30. Li CD, Jiang HW, Guo T, Wang ZX, Zheng W, Zhang Z, Y., and Zhao HH (2020) QTL mapping and interactive group validation for the number of four-seed-pod trait on soybean. 42(1): 44–50. doi:10.19802/j.issn.1007-9084.2020013
31. Li CD, Jiang HW, Zhang WB, Qiu PC, Liu CY, Li WF et al (2008) QTL analysis of seed and pod traits in soybean. *Molecular Plant Breeding* 6:1091–1100
32. Li CD, Z JN, J, H.W., Y, J.Y., H, S.Y., and W XY et al (2018) Identification and validation of number of pod - and seed - related traits qtls in soybean. *Plant Breeding*.1–16. doi:<https://doi.org/10.1111/pbr.12635>
33. Lu PL, Chen N-Z, An R, Su Z, Qi BS 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 molecular biology* 63:289–305. doi:10.1007/s11103-006-9089-8
34. Lu X-Y, Kim H-J, Zhong S-Y, Chen H-B, Hu Z-Q, Zhou B-Y (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 Genomics* 15:805–818. doi:10.1186/1471-2164-15-805
35. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
36. McCouch S (1997) Report on QTL nomenclature. *Rice Genet Newsl* 14:11–13
37. Modrusan Z, reiser L, Feldmann KA, Fischer rL, Haughn GW (1994) Homeotic transformation of ovules into carpel like structures in *Arabidopsis*. *Plant Cell* 6:333–349. doi:10.1105/tpc.6.3.333
38. Ning HL, Yuan J-Q, Dong Q-Z, Li W-B, Xue H, Wang Y-S et al (2018) Identification of QTLs related to the vertical distribution and seed-set of pod number in soybean [*Glycine max* (L.) Merri]. *Plos one* 13:1–15. doi:10.1371/journal.pone.0195830
39. Palomeque L, Li-Jun L, Li W, Hedges B, Cober ER, Rajcan I (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. *Theoretical and applied genetics* 119. 417–427. doi:<https://doi.org/10.1007/s00122-009-1049-7>
40. Pei RJ, Chen XQ, Sun N, Zhang L (2010) Application of molecular markers in soybean breeding. *Anhui Agricultural Sciences* 38(3):100–101.doi: CNKI: SUN: AHNY.0.2010-03-035

41. Peng Y-H, Zhu J-C, Yang G-B, Yuan J-Z (1994) Relation of soybean leaf shape distribution to 4-seeded pods. *Journal of Crop Science* 20:501–503
42. Qiu L-J, Chang R-Z (2006) Specification and data standard for Soybean Germplasm Resources Description[M]. China Agricultural Press, BeiJing
43. Ray A, robinson-Beers K, ray S, Baker Sc, Lang Jd (1994) The Arabidopsis floral homeotic gene BELL (BEL1) controls ovule development through negative regulation of AGAMOUS gene (AG). *Proc. Natl Acad. Sci. U S A.* 91, 5761–5765. doi: 10.2307/2365052
44. Ray, A., Lang, J.d., Golden, t., and ray, S. (1996). SHORT INTEGUMENT (SIN1), a gene required for ovule development in Arabidopsis, also controls flowering time. *Development.* 122, 2631–2638. doi: 10.1002/(SICI)1097-0320(19960901)25:1<104::AID-CYTO12>3.0.CO;2-S
45. Reinprecht Y, Poysa VW, Yu K, Rajcan I, Ablett GR, Pauls KP (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* 49:1510–1527. doi:https://doi.org/10.1139/g06-112
46. Richards R (2000) A. Selectable traits to increase crop photosynthesis and yield of grain crops. 51:447–458. doi: 10.1093/jexbot/51.suppl_1.447
47. Sakuma Y, Liu Q, Dubouzet J-G, Abe H-S, Shinozaki K, Shinozaki K-Y (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. doi:10.1006/bbrc.2001.6299
48. Santiago J, Brandt B, Wildhagen M, Hohmann U, Hothorn LA, Butenko MA, Hothorn M (2016) Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. *Elife* 5:e15075. doi:10.7554/eLife.15075.001
49. Schneitz K, Baker Sc, Gasser cS, redweik A (1998) Pattern formation and growth during floral organogenesis: HUELLENLOS and AINTEGUMENTA are required for the formation of the proximal region of the ovule primordium in Arabidopsis thaliana. *Development* 125:2555–2563. doi:10.1101/gad.195101
50. Suzuki M, Fujino K, Nakamoto Y, Ishimoto M, Funatsuki H (2010) Fine mapping and development of DNA markers for the qPDH1 locus associated with pod dehiscence in soybean. *Molecular breeding* 25:407–418. doi:https://doi.org/10.1007/s11032-009-9340-5
51. Sun D, Li W, Zhang Z, Chen Q, Ning H, Qiu L, Sun G (2006) Quantitative trait loci analysis for the developmental behavior of soybean (*Glycine max* L. Merr.). *Theoretical Applied Genetics* 112(4):665–673. doi:10.1007/s00122-005-0169-y
52. Smitha N-A, Shah A, Song J, Jameson P-E (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. doi:10.3390/ijms18020428
53. Song, J.Y., Leung, t., Ehler, L.K., Wang, c., and Liu, Z. (2000). Regulation of meristem organization and cell division by TSO1, an Arabidopsis gene with cysteine-rich repeats. *Development.* 127, 2207–2217. doi: 10.1002/(SICI)1097-0320(20000501)40:1<81::AID-CYTO11>3.0.CO
54. 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

55. Tischber T, Allphin L, Chase K, Orf JH, Lark KG (2003) Genetic of seed abortion and reproductive traits in soybean. *Crop Sci* 43:464–473. doi:<https://doi.org/10.2135/cropsci2003.0464>
56. Toshio S, Toshiyuki N. The possible involvement of a phosphate-induced transcription factor encoded by phi-2 gene from tobacco in ABA-signaling pathways. (2002). *Plant Cell Physiology*, 1,12. doi:10.1093/pcp/pcf002
57. Varshney RK, Graner A, Sorrells ME (2005) Genomics-assisted breeding for crop improvement. *Trends in plant science* 10:621-630. doi:<https://doi.org/10.1016/j.tplants.2005.10.004>
58. Wang SC, Basten CJ, Zeng ZB (2006) Windows QTL cartographer 2.5 department of statistics, North Carolina State University, Raleigh, NC. Retrieved from <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>
59. Wang X-Z, Zhang X-J, Zhou R, Sha A-H, Wu X-J, Cai S-P et al (2007) QTL analysis of pod and seed characters in soybean recombinant inbred lines. *Journal of Crop Science* 33:441–448. doi:10.3321/j.issn:0496-3490.2007.03.014
60. Wang Q, Guo L, Liu Y, Zhang G, Liu F (2002) Analysis of grey related degree between individual plant yield and main agriculture character of summer soybean. *Soybean Bulletin*, 2002 (6): 5, 12. doi: 10.3969/j.issn.1674-3547.2002.06.005
61. Wang X-B, Li Y-H, Zhang H-W, Sun G-L, Zhang W-M, Qiu L-J (2015) Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean. *Molecular biology reports*. 489–496. doi:10.1007/s11033-014-3792-3
62. Weiss M-G (1970) Genetic linkage in soybeans: linkage group I 1. *Crop Sci* 10:69–72
63. Western tL, Haughn GW (1999) BEL1 and AGAMOUS genes promote ovule identity in *Arabidopsis thaliana*. *Plant J* 18:329–336. doi:10.1046/j.1365-313X.1999.00448.x
64. Yan L, Zhang Y-Y, Song Q-J, Cregan P-B, Yang C-Y, Charles YQ et al (2017) Validation of the quantitative trait locus underlying soybean plant height using residual heterozygous lines and near-isogenic lines across multi-environments[J]. *Euphytica* 213:1–11. doi:10.1007/s10681-016-1833-9
65. Yamanaka N, Watanabe S, Toda K, Hayashi M, Fuchigami H, Takahashi R, Harada K (2005) Fine mapping of the FT1 locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line. *Theoretical and applied genetics* 110. 634–639. doi: <https://doi.org/10.1007/s00122-004-1886-3>
66. Yang Z, Sun Y-N, Qi Z-M, Xin D-W, Jiang H-W, He L et al (2013) Analysis of additive, epistasis and QE interaction of QTLs related to pod number traits in Soybean. *Journal of China Agricultural University* 18:1–13. doi:1007-4333(2013)03-0001-13
67. Zhang D, Cheng H, Wang H, Zhang H, Liu C, Yu D (2010) Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* L.). *Journal of Genetics Genomics* 37(8):545–556. doi:10.1016/s1673-8527(09)60074-6
68. Zhang WY, Wang JX, Zong CM, Fu CX, Wang MQ, Sun YJ et al (2012) QTL Analysis of Seed and Pod Traits in Soybean. *Soybean Science* 31(2):193–197
69. Zhou R, Chen H-F, Wang X-Z, Zhang X-J, Shan Z-H, Wu X-J et al (2009) QTL analysis of yield, yield components, and lodging in soybean. *Acta Agronomica Sinica* 35:821–830

Figures

Figure 1

Frequency distribution of soybean four-seeded pods in a preliminary mapping population and secondary mapping population. A: Frequency distribution of the soybean four-seeded pods in the preliminary mapping population B: Frequency distribution of the soybean four-seeded pods in the R1 population C: Frequency distribution of the soybean four-seeded pods in the R2 population D: Frequency distribution of the soybean four-seeded pods in the R3 population.

Figure 2

Distribution of the introduced ZYD00006 parental fragment and heterozygous fragment.

Figure 3

F2 generation heterozygous segmented genotype. (A) Generation genotype of Gm03 (3957733-4224480 bp). (B) Generation genotype of Gm06 (48413650-48625035 bp). (C) Generation genotype of Gm08 (3398336-7680873 bp). (D) Generation genotype of Gm08 (11853745-17718597 bp). (E) Generation genotype of Gm14 (8151693-12658057 bp)

Figure 5

Mapping interval of QNFSP07-1 for four-seeded pods in soybean

Figure 7

Relative expression of Glyma.07G200900, 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 $\pm$ SE (standard errors) of three replications. The different letters represent significant differences ($p \leq 0.05$)

A

B

Figure 8

SNP changes of the excellent haplotype and phenotypic differences of the excellent haplotype of Glyma.07G201200. (A) Phenotypic differences of the number of four-seeded pods in the excellent haplotype. (B) The significance level at $p < 0.05$

Supplementary Files

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

- [TableA1.docx](#)
- [TableA2.docx](#)
- [TableA3.docx](#)

- [TableA4.docx](#)
- [TableA5.docx](#)
- [TableA6.docx](#)
- [TableA7.docx](#)
- [TableA8.docx](#)