

Research Article

Cite this article: Park C, Nguyen TT, Liu D, Wang Q, Xu D (2023). A QTL allele from wild soybean enhances protein content without reducing the oil content. *Plant Genetic Resources: Characterization and Utilization* **21**, 409–417. <https://doi.org/10.1017/S1479262123000850>

Received: 13 July 2023

Revised: 24 October 2023

Accepted: 24 October 2023

First published online: 29 November 2023

Keywords:


oil; protein; quantitative trait locus (QTL); soybean; wild soybean

Corresponding author:

Donghe Xu;

Email: xudh@jircas.affrc.go.jp

A QTL allele from wild soybean enhances protein content without reducing the oil content

Cheolwoo Park¹, Trang Thi Nguyen^{1,2}, Dequan Liu^{1,3}, Qingyu Wang³
and Donghe Xu¹ 

¹Biological Resources and Post-harvest Division, Japan International Research Center for Agricultural Science, Ohwashi, Tsukuba, Ibaraki, Japan; ²GMO Detection Laboratory, Agricultural Genetics Institute, BacTuliem, Hanoi, Vietnam and ³College of Plant Science, Jilin University, Changchun, Jilin, P.R. China

Abstract

Soybean is one of the chief crops producing protein and oil for human consumption. Wild soybean, the ancestor of cultivated soybean, possesses high seed protein content; therefore, it is a valuable genetic resource that could enhance protein content in the cultivated varieties. To identify the genes responsible for increasing protein content in wild soybean, a population comprising 113 BC₄F₆ chromosome segment substitution lines (CSSL) was developed from a cross between soybean cultivar ‘Jackson’ and wild soybean accession JWS156-1. The CSSL population was cultivated in the field conditions for 3 years (2018, 2019 and 2020), and the seeds harvested from each line were analysed for protein and oil contents by InfraTec Nova instrument. Quantitative trait locus (QTL) analysis with 243 SSR markers identified 12 QTLs associated with seed protein, oil and protein + oil contents. Among these QTLs, *qPro8* and *qPro19*, two major and stable QTLs for protein content, were detected on chromosomes 8 and 19, respectively. No QTL for oil content was detected in the vicinity of *qPro19*, indicating that *qPro19* did not influence the seed oil content. The effect of *qPro19* was validated using near-isogenic lines (NILs) of *qPro19*. By introducing the *qPro19* allele from wild soybean into another soybean variety, ‘Tachiyutaka’, a BC₄ line, T-678, that showed enhanced seed protein content, without reducing the seed oil content. This study implied that the *qPro19* allele from wild soybean could be a potential genetic resource for breeding programmes aimed to improve soybean seed quality.

Introduction

Soybean (*Glycine max* (L.) Merr.) seeds contain about 40% protein and 20% oil, making this crop one of the most important sources of protein and oil for human consumption. Currently, soybean provides >71% of the total vegetative protein and >29% of oil worldwide (<http://www.soystats.com>). Soybean protein is considered a complete protein, as it contains well-balanced essential amino acids necessary for human nutrition (Qin *et al.*, 2022). Soybean oil contains unsaturated fatty acids, particularly linoleic and oleic acids. It is also low in saturated fat and contains no cholesterol, making it a healthier alternative to other oils sourced from vegetable and animal fat. However, soybean protein content is negatively related to its oil content (Wilcox, 1998; Clemente and Cahoon, 2009; Kambhampati *et al.*, 2019). When breeding a soybean variety with one having higher protein content, it generally results in lower oil content in the variety created. Understanding the genetic mechanisms controlling protein and oil contents might enable the development of a novel soybean variety with higher protein and oil contents.

Using DNA markers, quantitative trait loci (QTLs) related to protein and oil contents in soybean have been reported. In SoyBase (<https://soybase.org>), 255 seed protein-content QTLs and 322 seed oil-content QTLs were reported in various populations and environments. These QTLs were identified using bi-parental mapping populations or natural populations. Although numerous QTLs related to protein and oil contents were reported, most of these QTLs were either population and environment dependent, duplicated or not validated.

The wild soybean (*Glycine soja* Sieb. & Zucc.), which is the ancestor of cultivated soybean, demonstrated to have higher genetic variation as compared to cultivated soybean (Zhou *et al.*, 2015), thus making wild soybean a valuable genetic resource for the improvement of cultivated soybean. Wild soybean has proved to be a potential genetic resource for improving seed yield (Concibido *et al.*, 2003), salinity tolerance (Hamwiah and Xu, 2008), disease resistance (Zhang *et al.*, 2016) and seed nutritional components (Wang *et al.*, 2007). Wild soybean is known to possess higher seed protein content than their domesticated counterparts (Sebolt *et al.*, 2000).



It might be used as an important genetic resource for improving protein content in soybean breeding.

QTLs associated with the protein content have been identified by mapping the population involved wild soybean. Diers *et al.* (1992) identified two major QTLs associated with protein and oil contents on chromosomes (Chr) 15 and 20 using RFLP markers in an F₂ population by crossing a *G. max* experimental line A81-356022 and a *G. soja* accession PI 468916. Also, Sebolt *et al.* (2000) identified two protein and one oil content-related QTLs in a BC₃ population of A81-356022 and a *G. soja* accession PI468916. The analysis indicated that the wild soybean allele of the QTL on Chr 20 was associated with a higher protein and less oil content. Patil *et al.* (2018) studied a recombinant inbred lines (RIL) population derived from a cross between cultivar ‘Williams 82’, and *G. soja* accession PI483460B and via composite interval mapping method identified 5 QTLs for seed protein content on Chr 6, 8, 13, 19 and 20 and 9 QTLs for seed oil content on Chr 2, 7, 8, 9, 14, 15, 17, 19 and 20. The major QTLs for protein and oil contents were mapped on Chr 20. Using an RIL population derived from *G. max* line Osage and a *G. soja* accession PI593983, Yang *et al.* (2022) identified two significant QTLs for oil contents on Chr 8 and 20 with the log of odds (LOD) values 9.8–25.9, and four significant QTLs for protein content on Chr 14 and 20 with LOD values of 5.3–31.7. The results showed that wild soybean allele on these QTLs could enhance protein content suggesting that wild soybean is a potential genetic resource for improving soybean protein content.

On the other hand, wild soybean possesses some undesirable agricultural traits such as small seed size, twining stem, and pod shattering that make wild soybean direct usage in soybean breeding programme difficult. To address this problem, backcrossing was used as an efficient way to eliminate the undesirable agricultural traits of wild soybean and retain the target traits in the progenies of cultivated soybean and wild soybean. With the backcrossing strategy, advanced backcrossing populations, such as chromosome segment substitution lines (CSSL), have been developed to identify favorable genes (allele) in wild soybean (Wang *et al.*, 2013; He *et al.*, 2015). In our previous study, a BC₃F₅ wild soybean CSSL population was created and was used to identify seed weight and flowering time QTLs (Liu *et al.*, 2018a; 2018b). In the present study, we created a BC₄F₆ wild soybean CSSL population. The population was cultivated in field conditions for 3 years to identify favourable protein QTL alleles from wild soybean for soybean breeding to improve seed quality.

Materials and methods

Plant materials

A CSSL population with 113 lines was derived from a cross between a cultivated soybean ‘Jackson’ and a wild soybean accession JWS156-1. The cultivated soybean variety ‘Jackson’ (PI548657) was obtained from the US National Plant Germplasm System (NPGS), and the wild soybean accession

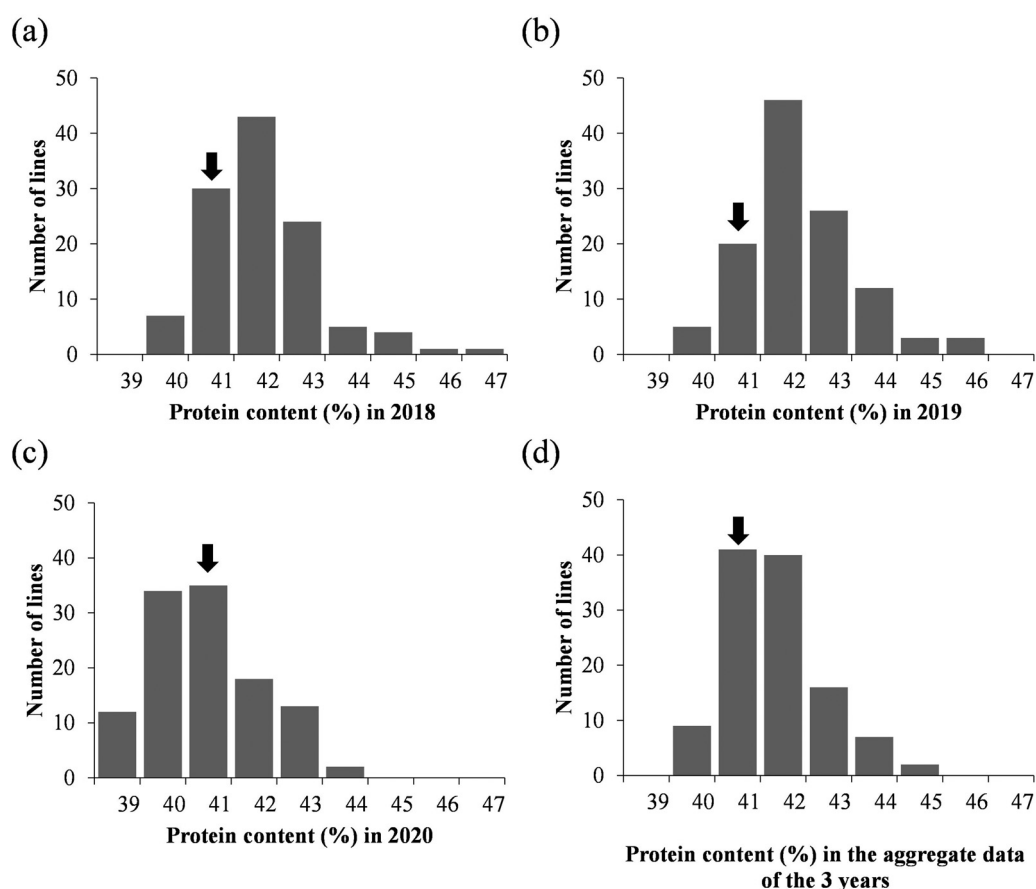


Figure 1. Frequency distribution of seed protein content in the wild soybean CSSL population ($n = 113$) in 3 consecutive years (2018 (a), 2019 (b), and 2020 (c)) and the aggregate data of the 3 years (d). Arrow: ‘Jackson’.

JWS156-1 was originally from the Kinki area of Japan provided by the National BioResource Project (*Lotus japonicus* and *G. max*) (<https://legumebase.nbrp.jp/legumebase/top.jsp>). The CSSLs were developed by crossing 'Jackson' with JWS156-1 and back-crossed with 'Jackson' for four generations, followed by successive self-pollination until the BC₄F₆ generation. The protein content, determined by Kjeldahl method (Nozawa *et al.*, 2005) and the oil content, determined by Soxhlet method (Rodrigues *et al.*, 2017) of 'Jackson' was 35.3% and 21.8%, while those of JWS156-1 was 44.0% and 10.6%.

Field experiment and measurement of protein and oil contents

The 113 BC₄F₆ CSSLs and their recurrent parent 'Jackson' were cultivated in 2018, 2019 and 2020 in the experimental farm of the Japan International Research Center for Agricultural Sciences, Japan (36.05°N, 140.08°E). All CSSLs were randomly arranged with one replication (2018 and 2019) or two replications (2020) in the field conditions. Each line was planted in a single-row plot 6 m long, with 60 cm spacing between rows and 20 cm spacing between plants. Seeds were harvested as plot bulk. Seeds were dried naturally, and seed protein and oil contents were measured using an InfraTec Nova instrument (FOSS Analytics, Hillerød, Denmark). The broad-sense heritability of protein and oil contents was calculated using the R 'inti' package (Lozano-Isla, 2021) developed to analyse multi-environment trials using the linear mixed model.

Simple sequence repeats (SSR) marker analysis

A total DNA sample was extracted from the leaf tissue using a modified CTAB method. A total of 243 SSR markers that showed polymorphism between the two original parents, 'Jackson' and JWS156-1, were analyzed for the CSSL population. All SSR markers were selected from the genetic maps of Song *et al.* (2004, 2010) and Cregan *et al.* (1999). The PCR mixture comprised 3 µl (10–50 ng) template DNA, 2 µl (10 pmol) of each primer, and 10 µl Quick Taq™ HS DyeMix (Toyobo, Osaka, Japan) in a total volume of 20 µl. PCR amplification was performed in a TAdvanced 384 thermal cycler (Biometra, Göttingen, Germany) with the following PCR reaction parameters: 94 °C for 30 s, followed by 30 cycles of 30 s at 94 °C, 30 s at 57 °C, 40 s at 72 °C, and a final extension at 72 °C for 10 min. After amplification, 10 µl of each PCR product was separated on 8% denaturing polyacrylamide gel in 1 × TBE running buffer for ~240 min at 200 V and stained using ethidium bromide. The gel was scanned using the PharosFX Molecular Imager (Bio-Rad Laboratories, Hercules, CA, USA) to detect PCR fragment polymorphism.

QTL analysis

A software QTL IciMapping (Wang *et al.*, 2016) was employed to identify QTL associated with protein content, oil content, and a sum of protein and oil (protein + oil) content in the CSSL population using the RSTEP-LRT-ADD (Stepwise regression for

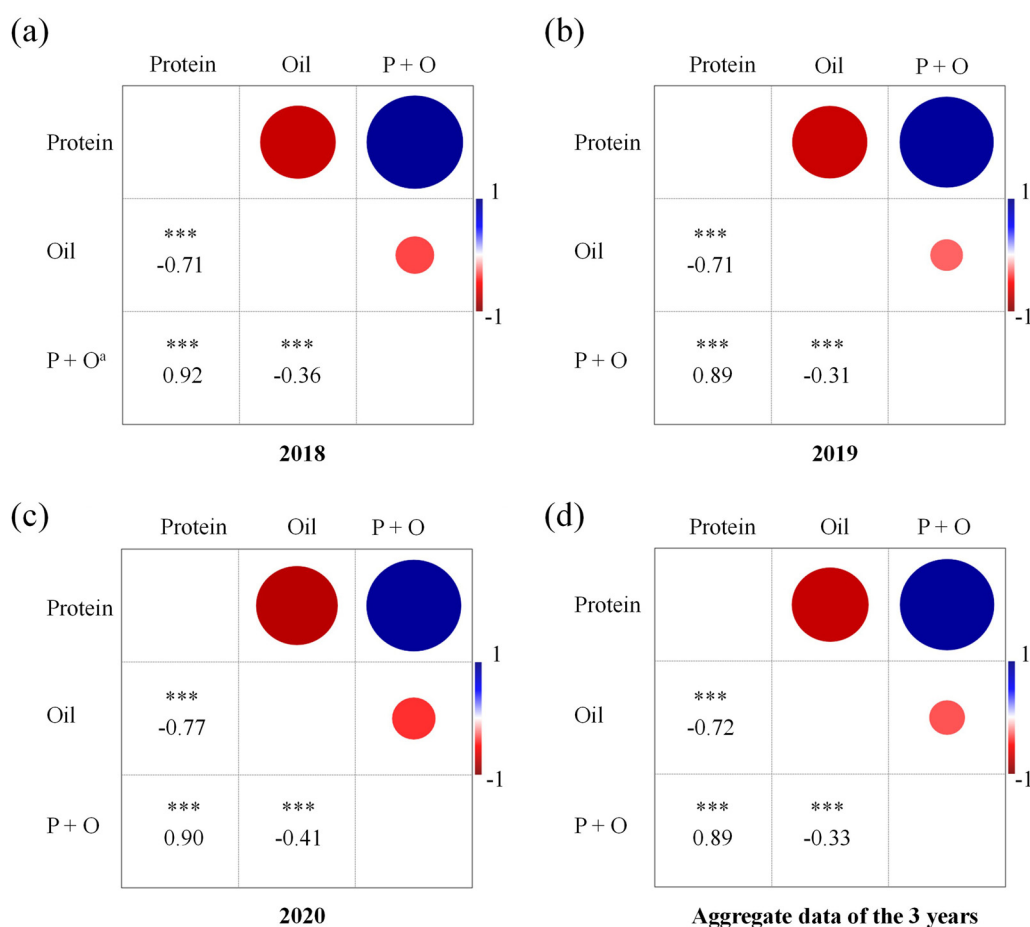


Figure 2. Pearson's correlation analysis of protein content, oil content, and a sum of protein and oil (P + O) content in the wild soybean CSSL population ($n = 113$) in 3 years (2018 (a), 2019 (b), and 2020 (c)) and the aggregate data of the 3 years (d). *** indicates significant correlations at the 0.1% level.

additive QTL) method. A threshold of LOD score >3 was set to declare the existence of a QTL.

Validation of the major protein QTL *qPro19* on Chr 19

To validate a major protein QTL on Chr 19 (*qPro19*) detected in the present study, near-isogenic lines (NIL) were developed by crossing a BC₄F₆ CSSL JJ4-188, which harboured the JWS156-1 allele on *qPro19*, with 'Jackson'. BC₅F₂ plants were selected by using SSR markers BARCSOYSSR_19_0773, BARCSOYSSR_19_0800, BARCSOYSSR_19_0826, and Satt463. Homozygous

JWS156-1 allele plants (BC₅F₂-W) and homozygous 'Jackson' allele plants (BC₅F₂-C) at *qPro19* were selected. These two lines (BC₅F₂-W and BC₅F₂-C) had similar genetic backgrounds but differed for the protein QTL *qPro19* and thus could be regarded as NILs. The NILs (BC₅F₂-W and BC₅F₂-C) were cultivated with three replicates in the field condition in 2022. After reaching maturity, 4 or 5 plants from each BC₅F₂-W and BC₅F₂-C in a replicate were harvested in bulk, and subsequently were subjected to protein and oil content measurements. The field cultivation methods and protein and oil content measurements were the same as those of the CSSL population described above.

Table 1. QTLs for seed protein, oil, and a sum of protein and oil contents detected in the wild soybean CSSL population ($n = 113$) in 3 years (2018, 2019 and 2020) and the aggregate data of the 3 years

Trait	QTL name ^a	Marker name	Marker position ^b	Year tested	LOD score ^c	PVE (%) ^d	Additive effect of JWS156-1 allele
Protein	<i>qPro1</i>	Sat_414	Chr 1: 52,336,556	2020	3.33	7.54	-0.6
	<i>qPro4</i>	SOYGPATR	Chr 4: 564,119	2020	3.24	7.27	0.65
	<i>qPro8</i>	Sat_212	Chr 8: 9,159,035	2018	6.12	19.31	1.29
				2019	5.06	15.42	1.19
				2020	4.94	11.48	1.02
				3 years ^e	6.19	18.22	1.15
	<i>qPro13</i>	Satt554	Chr 13: 39,252,658	2020	4.12	9.41	0.64
	<i>qPro19</i>	BARCSOYSSR_19_0773	Chr 19: 34,261,777	2018	3.06	9.07	0.81
2019				4.55	13.69	1.03	
3 years				5.18	14.92	0.96	
	Satt462	Chr 19: 36,736,007	2020	4.89	11.38	0.93	
Oil	<i>qOil8</i>	Sat_212	Chr 8: 9,159,035	2018	10.04	28.13	-0.67
				2019	9.39	24.29	-0.74
				2020	7.70	23.17	-0.68
				3 years	9.59	28.47	-0.69
	<i>qOil12</i>	Sat_158	Chr 12: 33,865,805	2018	3.47	8.47	-0.41
				2020	3.00	8.19	-0.45
				3 years	3.72	9.76	-0.45
	<i>qOil16.1</i>	Sat_228	Chr 16: 3,074,941	2019	3.17	7.24	0.51
	<i>qOil16.2</i>	Sat_361	Chr 16: 30,478,472	2019	4.53	10.57	-0.41
	Protein + Oil	<i>qP + O8</i>	Sat_212	Chr 8: 9,159,035	2018	4.60	11.61
2019					3.24	10.80	-0.65
<i>qP + O19</i>		BARCSOYSSR_19_0773	Chr 19: 34,261,777	2018	3.73	9.26	0.61
				2019	4.77	16.39	0.87
				2020	6.99	20.10	0.92
	Satt462	Chr 19: 36,736,007	3 years	6.60	19.47	0.84	

^aThe QTL name is defined by the trait name, chromosome number and its order on the chromosome.

^bStarting point of the marker.

^cThe log of odds (LOD) value calculated from RSTEP-LRT-ADD method.

^dPhenotypic variance explained by the QTL.

^eThe aggregate data of the 3 years.

Introducing the wild soybean allele of *qPro19* into a soybean variety 'Tachiyutaka'

To validate the effect of *qPro19* in different genetic backgrounds, the wild soybean allele of *qPro19* was introduced into a soybean variety 'Tachiyutaka' (PI594289) by crossing 'Tachiyutaka' with JWS156-1 and backcrossed with 'Tachiyutaka' for four generations, followed by successive self-pollination until the BC₄F₆ generation. A BC₄F₆ line (T-678) with homozygous JWS156-1 allele at *qPro19* was selected with the assistance of DNA markers (BARCSOYSSR_19_0773, BARCSOYSSR_19_0800, BARCSOYSSR_19_0826, and Satt463).

The T-678 line was cultivated in the field conditions for 4 years (2018, 2019, 2020 and 2022). Seed protein and oil contents were measured using the same method described above.

Results

Genotypic characterization of the BC₄F₆ wild soybean CSSL population

The graphical genotypes of the 113 CSSLs are shown in Fig. S1. All 243 markers had at least one JWS156-1 allele in the 113

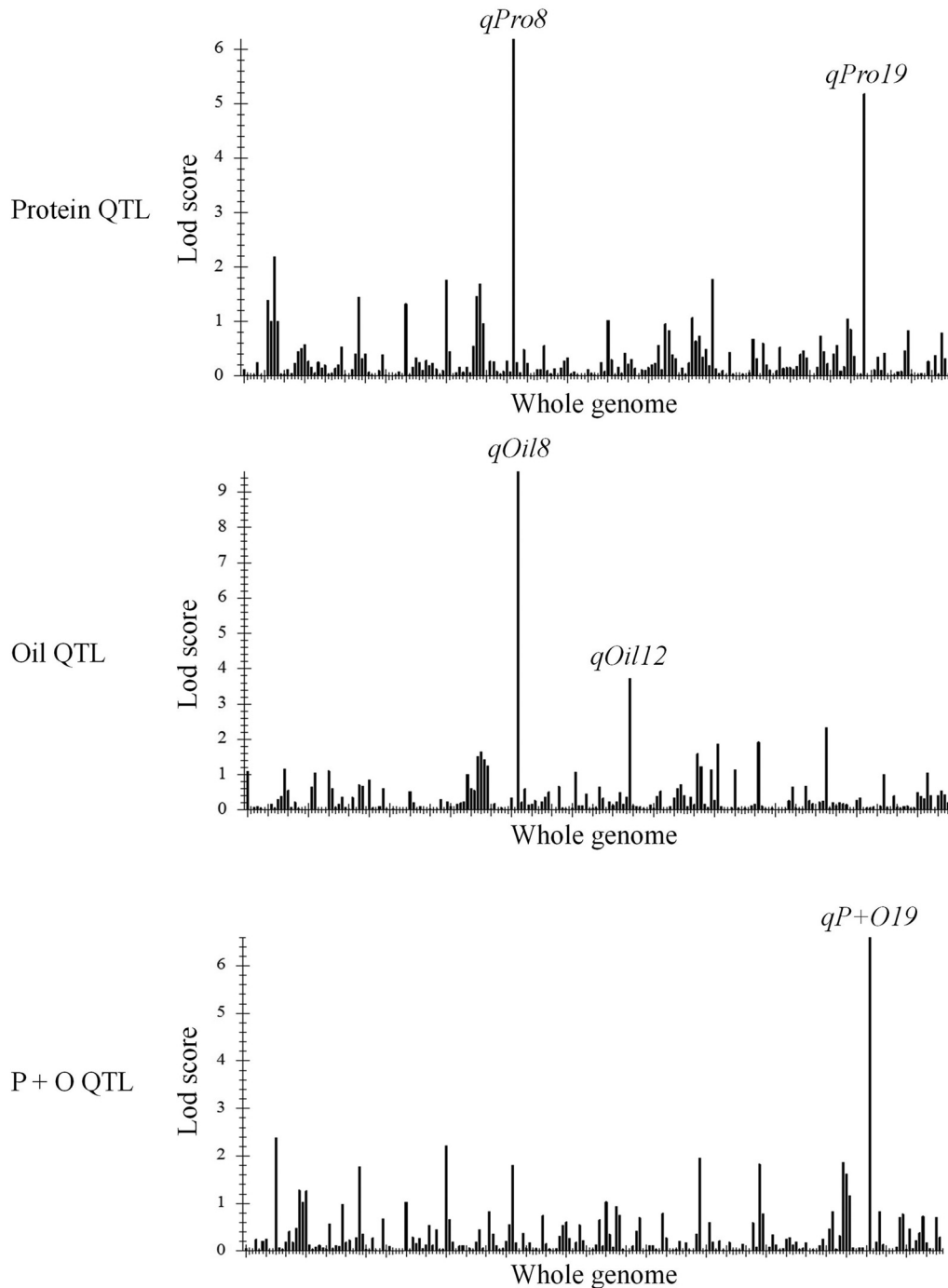


Figure 3. QTL analysis results of protein content, oil content, and a sum of protein and oil content (P + O) in the wild soybean CSSL population ($n = 113$) (the aggregate data of the 3 years).

CSSLs. The CSSLs were almost recovered by the recurrent parent 'Jackson' after backcrossing four times, and no lines with abnormal growth were observed in the 113 CSSLs. The proportion of the recurrent parent 'Jackson' alleles in each CSSL ranged from 79.4% to 99.9%, with an average of 94.2%, slightly lower than the expected value of 96.9%.

Phenotypic variations of protein, oil and sum of protein and oil contents in the CSSL population

As shown in Fig. 1, the distributions of seed protein in the CSSLs varied in different years (2018, 2019 and 2020). The range of protein content in the CSSL population was 39.3%–46.2% (average: 41.6%), 39.3%–45.6% (average: 41.9%), and 38.1%–43.8% (average: 40.4%) in 2018, 2019 and 2020, respectively. The oil and protein + oil contents also showed phenotypic variations among different years (Supplementary Figs. S2, S3). These results indicated that the seed protein, oil and protein + oil contents were affected by environmental conditions. On the other hand, protein, oil and protein + oil contents of the 113 CSSLs showed significantly positive correlations among different years (Supplementary Fig. S4). The broad heritability (H^2) for protein, oil, and protein + oil contents were 64.3%, 82.8% and 41.2%, respectively.

As depicted in Fig. 1, the CSSLs showed higher and lower values than those of the recurrent parent 'Jackson' for the seed protein, oil and protein + oil contents, suggesting transgressive segregation. Of the 113 CSSLs, 85 CSSLs (75.2%) showed higher protein content than their current variety, 'Jackson', in terms of the aggregate data of the 3 years, suggesting that wild soybean is a potential genetic resource for enhancing protein content in cultivated soybean. In contrast, only 49 CSSLs (43.4%) showed higher oil content than their current variety 'Jackson'. In the case of the protein + oil content, 89 CSSLs (78.8%) showed higher values than 'Jackson'.

The correlations among the protein, oil and protein + oil contents were consistent across the 3 years and also in the aggregate data of the 3 years (Fig. 2). The correlations between protein and oil contents were significantly negative, while those between the

protein and protein + oil contents were significantly positive. The smallest correlation (negative) was observed between oil and protein + oil contents.

QTL analysis for protein and oil contents

Five QTLs associated with protein content were identified on Chr 1, 4, 8, 13 and 19 (Table 1). The markers BARCSOYSSR_19_0773 and Satt462, which are located in a nearby region, were regarded as the same QTL. The phenotypic variance explained (PVE) by the QTL values ranged from 7.27% to 19.31%. Wild soybean alleles at all the QTLs associated with protein content contributed to increasing effects on protein content except the *qPro1* detected on Chr 1. Protein QTLs, *qPro8* and *qPro19*, were identified in the data from all 3 years and in the aggregate data of the 3 years and thus were regarded as major and stable QTLs. The other protein QTLs were only detected in 1 year, i.e. 2020.

Four QTLs associated with oil content were identified on Chr 8, 12 and 16 (Table 1). The PVE values ranged from 7.24% to 28.47%. Wild soybean allele corresponding to all the oil-content QTLs contributed to a reduction in oil content except the *qOil16.1* detected in the data from 2019. The oil-content QTL, *qOil8*, was identified in all 3 years and the aggregate data of the 3 years, and thus was regarded as a major and stable QTL. The other QTLs for oil content were only detected in 1 or 2 years. Interestingly, *qOil8* was located in the same position as the protein content QTL, *qPro8* (Table 1, Fig. 3).

Three QTLs associated with protein + oil content were detected on Chr 8, 18 and 19, with PVE values ranging from 9.26% to 20.10% (Table 1). Of these, a QTL on Chr 19 (*qP + O19*), which was located at the same position as the protein content QTL *qPro19*, was also identified across the data from the 3 years of experiments and also in the aggregate data of the 3 years.

Validation of the major protein content QTL *qPro19*

To verify the effect of the protein content QTL *qPro19* on Chr 19, two *qPro19* NILs (BC₅F₂-W and BC₅F₂-C) were evaluated in the field conditions. NIR analysis revealed that BC₅F₂-W had a

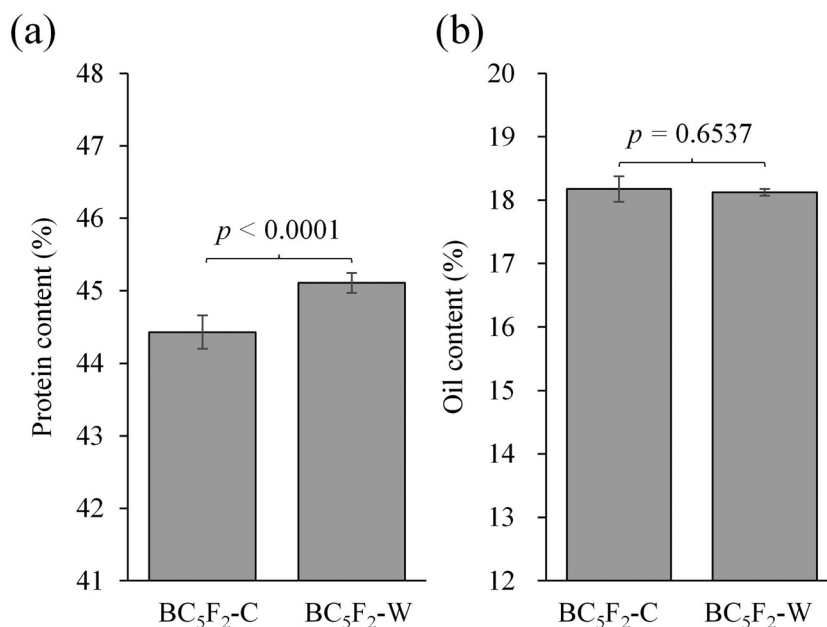


Figure 4. The allele effect of *qPro19* in the *qPro19* near-isogenic lines. BC₅F₂-W: 'Jackson' genotype; BC₅F₂-C: JWS156-1 genotype. (a): protein content; (b): oil content. Error bars: SD ($n = 3$).

protein content of $45.11\% \pm 0.14\%$, which was significantly ($P < 0.001$) higher than that of the BC₅F₂-C ($44.43\% \pm 0.23\%$) (Fig. 4). In contrast, no significant difference for oil content was observed between BC₅F₂-W ($18.13\% \pm 0.05\%$) and BC₅F₂-C ($18.18\% \pm 0.20\%$) (Fig. 4). This result confirmed the effect of *qPro19*, as an enhancer of protein content without reducing oil content in the soybean seeds.

Introducing the wild soybean allele of *qPro19* into a soybean variety 'Tachiyutaka'

To validate the effect of *qPro19* in a different genetic background, the wild soybean allele of *qPro19* was introduced into a soybean variety 'Tachiyutaka', and a BC₄F₆ line (T-678) with homozygous

JWS156-1 allele at *qPro19* was obtained based on genotypes of SSR markers BARCSOYSSR_19_0773, BARCSOYSSR_19_0800, BARCSOYSSR_19_0826 and Satt463. Field evaluations revealed that T-678 showed significantly higher protein content in comparison with its original parent 'Tachiyutaka'. In contrast, no significant difference in oil content was observed between T-678 and 'Tachiyutaka' (Fig. 5). The effect of *qPro19* was thus validated in a different genetic background.

Discussions

In soybean, protein and oil contents generally showed a negative relationship. This might be because these traits are associated with the same gene controlling both the protein and oil synthesis

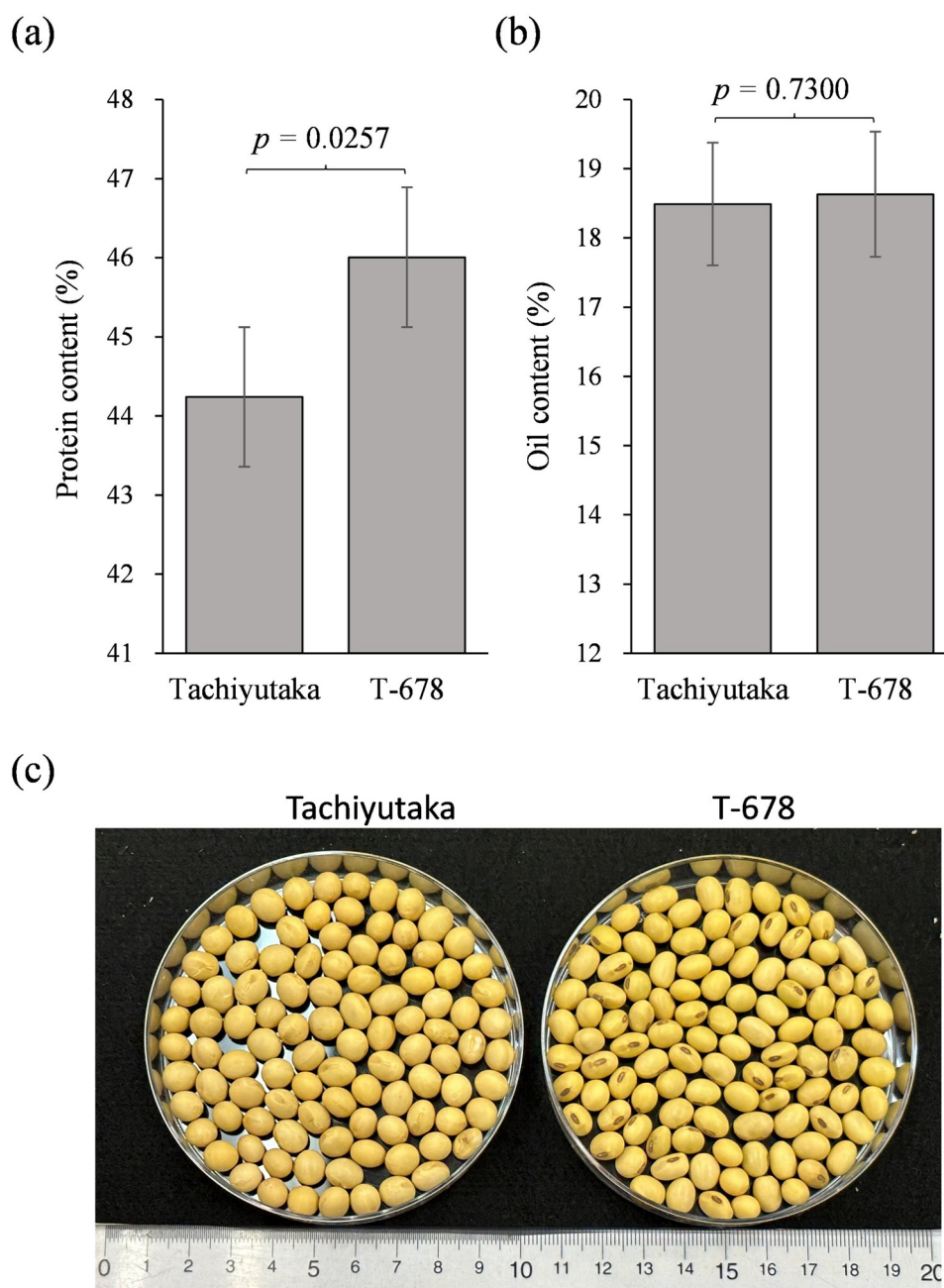


Figure 5. The allele effect of *qPro19* in a BC₄F₆ backcrossing line (T-678) in the 4 years of field trials (2018, 2019, 2020, and 2021). (a) protein content; (b) oil content; (c) appearance of seeds of 'Tachiyutaka' and T-678. Error bars: SD ($n = 4$).

pathways. The QTLs for protein and oil contents might thus be mapped in the same position. For instance, Chung *et al.*, (2003) identified a QTL associated with protein, oil, and yield on Chr 20 using a RIL population derived from a cross between a high-protein *G. max* accession PI 437088A and a high-yield cultivar 'Asgrow A3733.' This QTLs was flanked by the SSR markers Satt496 and Satt239, with the allele from PI 437088A increasing protein content but decreasing oil content and yield. Additionally, Wang *et al.* (2020) and Peng *et al.* (2021) identified and revealed that *GmSWEET10a* (*Glyma.15G049200*) caused simultaneous changes in seed size, oil content and protein content. In the present study, a chromosome segment (QTL) around SSR marker Sat_212 on Chr 8 was identified to be associated with protein, oil and protein+oil contents across 3 consecutive years (2018, 2019 and 2020) along with the aggregate data of the 3 years. The PVE values of these QTLs ranged from 11.48% to 19.31% for protein content, 23.17–28.47% for oil content and 11.61% (2018) for protein+oil content, indicating that these QTLs were the major stable QTLs controlling seed protein and oil contents. Several QTLs for protein or oil content have been already reported around this region. Warrington *et al.* (2015) reported a QTL associated with the lysine/crude protein ratio on Chr 8, which was near the QTL on Chr 8 identified in the present study. Zhang *et al.* (2021) reported a QTL for water-soluble protein content associated with an SNP marker AX-93930669 located on the physical position of 8,276,381 bp on Chr 8. In addition, Lu *et al.* (2013), Pathan *et al.* (2013), Reinprecht *et al.* (2006) and Zhang *et al.* (2017) also reported protein-related QTLs close to the QTLs identified in the present study on Chr 8.

The QTLs of *qPro8*, *qOil8*, and *qP+O8* were located in the same region and thus it may be speculated that the same gene might control them. Shook *et al.* (2021) and Zhang *et al.* (2021) identified *Glyma.08G107800*, which belongs to the aspartokinase group, as a causal gene for the protein and oil contents QTLs on Chr 8. *Glyma.08G10780* was located at approximately 0.74 Mb to the QTLs detected in the present study (*qPro8*, *qOil8* and *qPO8*). Therefore, *Glyma.08G10780* is most likely the candidate gene underlying *qPro8*, *qOil8* and *qP+O8*.

A major stable QTL (*qPro19*) for protein content was identified on Chr 19. Wild soybean allele increased protein content with additive effects ranging from 0.81% to 1.03%. A QTL for protein+oil content (*qP+O19*) was also identified in the region of *qPro19* with additive effects from 0.84% to 0.92%. This might be due to the close relationship between protein and protein+oil contents. However, no QTL for oil content was detected in the region of *qPro19*, indicating that the wild soybean allele of this QTL enhanced protein content but did not reduce oil content. Previous studies have reported several QTLs on Chr 19. Tajuddin *et al.* (2003) reported a QTL for protein content on Chr 19 flanked by SSR marker Satt156 at a distance of approximately 6.4 Mb from *qPro19* (BARCSOYSSR_19_0773), the QTL identified in our study. Orf *et al.* (1999) also reported a QTL flanked by SSR marker Satt166 for protein content on Chr 19, and this QTL was around 8.4 Mb apart from *qPro19* (BARCSOYSSR_19_0773). In addition, Chapman *et al.* (2003) reported a QTL for protein content on Chr 19 flanked by an SSR marker Satt373, and this QTL was at a distance of 15.1 Mb to *qPro19* (BARCSOYSSR_19_0773). These QTLs may not be identical to the QTL, *qPro19*, identified in the present study.

In the *qPro19* region, there was a *Glyma.19G102100*, which was a homolog gene of *Glyma.08G107800* (the candidate gene

of *qPro8*, *qOil8*, and *qPO8*). *Glyma.19G102100* might be considered as the candidate gene of *qPro19*. However, *Glyma.08G107800* was speculated to function for both protein and oil contents, but this function was not found to be consistent with the results that *qPro19* has no effect on oil content.

In soybean, genes controlling protein content generally had pleiotropic effects, particularly showing a negative effect on oil content (Chung *et al.*, 2003; Wang *et al.*, 2020; Peng *et al.*, 2021; Xu *et al.*, 2022). Up to now, nine candidate genes associated with seed protein content have been identified in soybean, as reviewed by Liu *et al.* (2023). Most of these genes impact both protein and oil contents. The candidate gene underlying the QTLs of *qPro19* appears to be functionally different from these genes involved in controlling protein biosynthesis. Since oil content remained unaffected, the gene underlying the QTL of *qPro19* might not be involved in the biosynthesis of soybean seed oil.

Some QTLs associated with protein and oil contents were identified in the data from only 1 or 2 years. For example, protein content QTLs of *qPro1*, *qPro4* and *qPro13* were detected only in the data from 2020. Oil content QTLs of *qOil16.1* and *qOil16.2* were identified in 2019, and *qOil12* was identified in 2018 and 2020. Further validations are required for such unstable QTLs.

Soybean breeders have made efforts to balance the increased seed protein content with oil content in a soybean variety. As demonstrated in the present study, by introducing the wild soybean allele of *qPro19* into a Japanese variety, 'Tachiyutaka,' a backcrossed line with enhanced protein content without decreased oil content was obtained. Identifying the QTL allele from wild soybean that increased seed protein content without reducing the oil content might open a new approach to improving soybean seed quality through breeding practices.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262123000850>.

Acknowledgements. This research was funded in-part by the Japan International Research Center for Agricultural Sciences (JIRCAS) under a research project 'Resilient crops'.

Author contributions. Xu D. H. and Wang Q. Y. conceived the study. Park C., Nguyen T. T., and Liu D. Q. performed the experiments and data analyses. Park C. and Xu D. H. wrote the manuscript. All authors read and approved the manuscript.

Competing interests. The authors declare that they have no competing interests.

References

- Chapman A, Pantalone VR, Ustun A, Allen FL, Landau-Ellis D, Triggiano RN and Gresshoff PM (2003) Quantitative trait loci for agronomic and seed quality traits in an F2 and F4: 6 soybean population. *Euphytica* **129**, 387–393.
- Chung J, Babka HL, Graef GL, Staswick PE, Lee DJ, Cregan PB, Shoemaker RC and Specht JE (2003) The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Science* **43**, 1053–1067.
- Clemente TE and Cahoon EB (2009) Soybean oil: genetic approaches for modification of functionality and total content. *Plant Physiology* **151**, 1030–1040.
- Concibido V, La Vallee B, Mclaird P, Pineda N, Meyer J, Hummel I, Yang J, Wu k and Delannay X (2003) Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. *Theoretical and Applied Genetics* **106**, 575–582.

- Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J and Specht JE (1999) An integrated genetic linkage map of the soybean genome. *Crop Science* **39**, 1464–1490.
- Diers BW, Keim P, Fehr WR and Shoemaker RC (1992) RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics* **83**, 608–612.
- Hamwiah A and Xu D (2008) Conserved salt tolerance quantitative locus (QTL) in wild and cultivated soybeans. *Breeding Science* **58**, 355–359.
- He Q, Yang H, Xiang S, Tian D, Wang W, Zhao T and Gai J (2015) Fine mapping of the genetic locus L1 conferring black pods using a chromosome segment substitution line population of soybean. *Plant Breeding* **134**, 437–445.
- Kambhampati S, Aznar-Moreno JA, Hostetler C, Caso T, Bailey SR, Hubbard AH, Durrett TP and Allen DK (2019) On the inverse correlation of protein and oil: examining the effects of altered central carbon metabolism on seed composition using soybean fast neutron mutants. *Metabolites* **10**, 18.
- Liu D, Yan Y, Fujita Y and Xu D (2018a) A major QTL (qFT12.1) allele from wild soybean delays flowering time. *Molecular Breeding* **38**, 45.
- Liu D, Yan Y, Fujita Y and Xu D (2018b) Identification and validation of QTLs for 100-seed weight using chromosome segment substitution lines in soybean. *Breeding Science* **68**, 442–448.
- Liu S, Liu Z, Hou X and Li X (2023) Genetic mapping and functional genomics of soybean seed protein. *Molecular Breeding* **43**, 29.
- Lozano-Isla F (2021) *intti: Tools and Statistical Procedures in Plant Science*. R Package Version 0.1, 3, Vienna, Austria: R Foundation.
- Lu W, Wen Z, Li H, Yuan D, Li J, Zhang H, Huang Z, Cui S and Du W (2013) Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. *Theoretical and Applied Genetics* **126**, 425–433.
- Nozawa S, Hakoda A, Sakaida K, Suzuki T and Yasui A (2005) Method performance study of the determination of total nitrogen in soy sauce by the Kjeldahl method. *Analytical Sciences* **21**, 1129–1132.
- Orf JH, Chase K, Jarvik T, Mansur LM, Cregan PB, Adler FR and Lark KG (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. *Crop Science* **39**, 1642–1651.
- Pathan SM, Vuong TD, Clark K, Lee JD, Shannon JG, Roberts CA, Eilersieck MR, Burton JW, Cregan PB, Hyten DL, Nguyen HT and Slepner DA (2013) Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Science* **53**, 765–774.
- Patil G, Vuong TD, Kale S, Valliyodan B, Deshmukh R, Zhu C, Wu X, Bai Y, Yungbluth D, Lu F, Kumpatla S, Varshney R and Nguyen HT (2018) Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal* **16**, 1939–1953.
- Peng L, Qian L, Wang M, Liu W, Song X, Cheng H, Yuan F and Zhao M (2021) Comparative transcriptome analysis during seeds development between two soybean cultivars. *PeerJ* **9**, e10772.
- Qin P, Wang T and Luo Y (2022) A review on plant-based proteins from soybean: health benefits and soy product development. *Journal of Agriculture and Food Research* **7**, 100265.
- Reinprecht Y, Poysa VW, Yu K, Rajcan I, Ablett GR and Pauls KP (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* **49**, 1510–1527.
- Rodrigues GM, Cardozo-Filho L and Silva C (2017) Pressurized liquid extraction of oil from soybean seeds. *The Canadian Journal of Chemical Engineering* **95**, 2383–2389.
- Sebolt AM, Shoemaker RC and Diers BW (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science* **40**, 1438–1444.
- Shook JM, Zhang J, Jones SE, Singh A, Diers BW and Singh AK (2021) Meta-GWAS for quantitative trait loci identification in soybean. *Genes Genomes Genetics* **11**, jkab117.
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE and Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theoretical and Applied Genetics* **109**, 122–128.
- Song Q, Jia G, Zhu Y, Grant D, Nelson RT, Hwang EY, Hyten DL and Cregan PB (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. *Crop Science* **50**, 1950–1960.
- Tajuddin T, Watanabe S, Yamanaka N and Harada K (2003) Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. *Breeding Science* **53**, 133–140.
- Wang S, Kanamaru K, Li W, Abe J, Yamada T and Kitamura K (2007) Simultaneous accumulation of high contents of α -tocopherol and lutein is possible in seeds of soybean (*Glycine max* (L.) Merr.). *Breeding Science* **57**, 297–304.
- Wang W, He Q, Yang H, Xiang S, Zhao T and Gai J (2013) Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. *Euphytica* **189**, 293–307.
- Wang J, Li H, Zhang L and Meng L (2016) Users' manual of QTL IciMapping. The Quantitative Genetics Group, Institute of Crop Science, Chinese Academy of Agricultural Sciences (CAAS) Beijing, China, and Genetic Resources Program. CIMMYT, Mexico City, Mexico.
- Wang S, Liu S, Wang J, Yokosho K, Zhou B, Yu YC, Liu Z, Frommer WB, Ma JF, Chen LQ, Guan Y, Shou H and Tian Z (2020) Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *National Science Review* **7**, 1776–1786.
- Warrington CV, Abdel-Haleem H, Hyten DL, Cregan PB, Orf JH, Killam AS, Bajjalieh N and Boerma HR (2015) QTL for seed protein and amino acids in the Benning \times Danbaekkong soybean population. *Theoretical and Applied Genetics* **128**, 839–850.
- Wilcox JR (1998) Increasing seed protein in soybean with eight cycles of recurrent selection. *Crop Science* **38**, 1536–1540.
- Xu W, Wang Q, Zhang W, Zhang H, Liu X, Song Q, Zhu Y, Cui X, Chen X and Chen H (2022) Using transcriptomic and metabolomic data to investigate the molecular mechanisms that determine protein and oil contents during seed development in soybean. *Frontiers in Plant Science* **13**, 1012394.
- Yang Y, La TC, Gillman JD, Lyu Z, Joshi T, Usovsky M, Song Q and Scaboo A (2022) Linkage analysis and residual heterozygotes derived near isogenic lines reveals a novel protein quantitative trait loci from a *Glycine soja* accession. *Frontiers in Plant Science* **13**, 938100.
- Zhang H, Li C, Davis EL, Wang J, Griffin JD, Kofsky J and Song BH (2016) Genome-wide association study of resistance to soybean cyst nematode (*Heterodera glycines*) HG Type 2.5.7 in wild soybean (*Glycine soja*). *Frontiers in Plant Science* **7**, 1214.
- Zhang D, Lü H, Chu S, Zhang H, Zhang H, Yang Y, Li H and Yu D (2017) The genetic architecture of water-soluble protein content and its genetic relationship to total protein content in soybean. *Scientific Reports* **7**, 5053.
- Zhang S, Hao D, Zhang S, Zhang D, Wang H, Du H, Kan G and Yu D (2021) Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Molecular Genetics and Genomics* **296**, 91–102.
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H and Tian Z (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* **33**, 408–414.