Genetics of seed protein and oil inherited from “BARC-7” soybean in two F$_2$-derived mapping populations

F. Ravelombola$^a$, P. Chen$^b$, T. Vuong$^c$, H. Nguyen$^c$, R. Mian$^d$, A. Acuña$^a$, L. Florez-Palacios$^a$, C. Wu$^a$, D. Harrison$^a$, M. De Oliveira$^a$, J. Winter$^a$, M. Ford$^a$, M.P. Da Silva$^a$, and L. Mozzoni$^a$

$^a$Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, Arkansas, USA; $^b$Fisher Delta Research Center, University of Missouri, Portageville, Missouri, USA; $^c$Division of Plant Sciences, University of Missouri, Columbia, Missouri, USA; $^d$Usda – ARS, Soybean and Nitrogen Fixation Research, Raleigh, North Carolina, USA

**ABSTRACT**

Soybean [*Glycine max* (L.) Merr.] seed protein inheritance has been extensively studied; however, genetics of high-protein “BARC-7” soybean are still unknown. In this study, we used 250 F$_2$-derived lines from each of two soybean populations for quantitative trait loci (QTL) mapping. UA 5814HP, with high-protein content, tracing to BARC-7 as maternal grandfather, was a common parent. Field experiments were conducted using a randomized complete block design with one replication across four environments. Seed protein and oil were quantified using near-infrared (NIR) instrument. Genetic linkage maps were constructed using the Infinium Soy6KSNP Beadchips. QTL analysis was performed using a composite interval mapping method. QTL for protein and oil were identified on chromosomes 6, 13, and 20. The known major QTL on chromosome 20 was not detected, but a novel QTL further downstream on chromosome 20 (only detected in population two) had high-protein alleles inherited from BARC-7-derived parent. Fine mapping efforts are currently ongoing for confirmation of these results.

**Introduction**

One of the biggest challenges for plant breeders is feeding an increasing world population, forecasted to reach ~9 billion by 2050. Soybean [*Glycine max* (L.) Merr.], a legume species native to East Asia in the *Fabaceae* family, ranks among the most important food crops in the world. It is widely grown and known for its high protein and oil contents (Birt et al. 2004). Soybean is valuable because its seeds have multiple applications in food, feed, pharmaceutical, and industrial enterprises (Birt et al. 2004; Garcia et al. 1997; Lusas 2004). Even though seed yield is the most important trait, breeders have
recently given a significant attention to quality traits, such as high protein or modified oil content (Lee et al. 2019; Carneiro et al. 2020; Singh et al. 2020). Soybean seeds contain approximately 40% protein and 20% oil (Clemente and Cahoon 2009); however, there is a negative correlation between yield and protein content in soybean (Novikova et al. 2018; Lee et al. 2019; Sobko et al. 2020; Finoto et al. 2021), and between seed protein and oil content (Lee et al. 2019). Broad-sense heritability of protein and oil content in soybean is relatively high, ranging between 0.57 and 0.97 (Chung et al. 2003; Panthee et al. 2005; Jain et al. 2018; Tian et al. 2020; Jiang et al. 2020; Zhang et al. 2021; Arnold et al. 2021).

As many as 248 QTL associated with seed protein and 320 with seed oil have been reported in Soybase (Grant et al. 2010). Some QTL for protein and oil content were detected at the same position, suggesting either closely linked QTL or QTL with pleiotropic effect on both traits. Nonetheless, according to the rules established by the Soybean Genetics Committee (error rate lower than 0.01 and confirmation study showing alleles at the same locus are segregating in all the test populations), only two of those marker associations are accepted (Fasoula, Harris, and Boerma 2004; Nichols et al. 2006). Two protein QTL, located on chromosomes (Chr.) 15 and 20, were commonly identified in several studies. Of these, QTL on Chr. 20 was considered a major QTL with the highest proportion of phenotypic variance explained. The region of the interval on Chr. 15 is between 10 and 30 centimorgan (cM), whereas that on Chr. 20 is between 20 and 40 cM (Grant et al. 2010). Warrington et al. (2015) used a recombinant inbred-line population derived from the cross Benning/Danbaekkong, and mapped a major protein QTL on Chr. 20 carrying the Danbaekkong allele that explained 55% of phenotypic variation in protein content.

Leffel (1992) released a series of high protein soybean germplasm lines, including BARC-6 (Reg.no GP-127, PI 555396), BARC-7 (Reg.no. GP-128, PI 555397), BARC-8 (Reg. no. GP-129, PI 555398) and BARC-9 (Reg. GP-130, PI 555399). Among these, BARC-7 was derived from the cross of CX797–21/D80–6931. D80–6931 is a high protein maturity group (MG) VI BC3 line, in which PI 86490 was the high protein donor parent and “Tracy” was the recurrent parent. BARC-7 is a MG IV germplasm line with purple flowers, tan and brown pods, and determinate stem growth habit. The mean seed protein of BARC-7 is 491 g kg$^{-1}$ (Leffel, 1992). BARC-7 was a parent used in the breeding program at the University of Arkansas System Division of Agriculture, and had progeny with high seed protein levels, including “UA 5814HP” (Chen et al. 2017) and “R11–7999” (Florez-Palacios et al. 2020). However, the genetic architecture of protein and oil content in many BARC-7-derived soybean elite lines is unknown. Therefore, the goal of this study was to perform QTL mapping for seed protein and oil content in two
breeding populations that traced high seed protein to BARC-7 soybean germplasm line.

Materials and methods

Plant materials and phenotyping

Four initial crosses (UA 5615C/UA 5814HP, UA 5115C/UA 5814HP, R13–359/UA 5814HP, R13–532/UA 5814HP) between high-yielding lines and the high-protein cultivar UA 5814HP were made at the Milo J. Shult Agricultural Research and Extension Center in Fayetteville, AR, in 2017. “UA 5814HP” (Chen et al. 2017) is the progeny of “R95–1705” (high protein) and “S00–9980–22” (regular protein content); “R95–1705”, in turn, is the progeny of “Hutcheson” (regular protein) and “BARC-7” (high protein) (Leffel 1992). On the other hand, “UA 5615C”, “UA 5515C” (Florez-Palacios et al. 2019), R13–359, and R13–532 are commodity MG5 soybean varieties and lines with regular seed protein levels. A total of 13, 19, 20, and 26 F1 seeds of the four populations, respectively, were sent to Costa Rica during the winter of 2018, and bulk harvested by population. Approximately 800 seeds of F2 generation for each population were then planted in 8 rows of 4.6 m length during the summer of 2018 in Fayetteville, AR. All parental lines were screened for marker polymorphisms using the Infinium Soy6KSNP Beadchips (Song et al. 2020) (data not shown). Based on parental polymorphism and agronomic field adaptation, two populations, UA 5115C/UA 5814HP (Pop1) and R13–532/UA 5814HP (Pop2), were selected for QTL analysis. A total of 250 F2 plants for each selected population were randomly selected and individually harvested for generation advancement. A sample of 50 to 100 seeds per F2:3 line was sent to a winter nursery (Costa Rica) in 2019 to advance the population and increase the number of seeds via bulk harvesting. Two-hundred fifty F2-derived lines from each population were planted in four environments (location-year combination) for phenotyping, using a randomized complete block design (RCBD) with one replication. Environments included Upala, Costa Rica (inceptisol soil order -series unknown) in 2018 (18CR); Portageville, MO (Tiptonville silt loam soil (19MO)), and Rohwer, AR (Sharkey and Desha silt loam soils (19RO)) in 2019; and Fayetteville, AR (Captina silt loam soil in 2020 (20FA)). Plots (single rows in 19CR, and two-row plots for all other environments) were 0.76 m apart with 4.6 m long with a 1.5 m alley. Entries within each population were randomly divided into four experiments, and each experiment had six checks (UA 5115C, R13–532, UA 5814HP, P53A67X, AG55X7, and AG56X8). A sub sample of 50-seed from each line was used for protein and oil estimation via Near-Infrared Spectroscopy using a DA 7250 NIR analyzer (Perten, Sweden).
Genotyping and QTL mapping

DNA was extracted from fresh young leaves using the hexadecymethylammonium bromide (CTAB) protocol. Genotyping was done using the Infinium Soy6KSNP Beadchips (Song et al. 2020) in the Soybean Improvement Laboratory USDA-ARS, Beltsville. Data were analyzed using analysis of variance (ANOVA) and treating environments as replications. Genotypes were treated as a fixed effect in JMP 16.0. The statistical model for the analysis was:

\[ y_{ij} = \mu + g_i + b_j + \varepsilon_{ij} \]

where \( y_{ij} \) is the mean response (protein content or oil content) associated with the \( i \)th genotype in the \( j \)th environment, \( \mu \) is the overall mean of protein or for oil content, \( g_i \) is the genotype effect (fixed effect), \( b_j \) is the environmental effect (fixed effect), and \( \varepsilon_{ij} \) is the experimental error associated with the \( ij \)th observation. Pearson correlation between protein and oil content across environments was computed using JMP 16.0. Broad-sense heritability was estimated using the following equation:

\[ h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \]

where \( h^2 \) the is broad-sense heritability, \( \sigma_g^2 \) is the variance of the genotype, and \( \sigma_e^2 \) is the variance of error, and \( l \) is the number of environments.

Linkage genetic maps were constructed using JoinMap v.4.1 (Kyazma B.V., Wageningen, Netherlands). The segregation distortion for single nucleotide polymorphisms (SNPs) was analyzed through a chi-square test. A total of 1,423 and 1,115 polymorphic markers for Pop1 and Pop2, respectively, were used to construct the genetic map. The genetic distance was estimated using the Kosambi mapping function to address the inference. Based on the recombination frequencies, 24 linkages were created for Pop1 and Pop2, representing the 20 haploid chromosomes in the soybean genome.

QTL analysis was conducted through WinQTL cartograph v2.5 (Wang, Basten, and Zeng 2012). Composite interval mapping (CIM) was the statistical model for the QTL search and to estimate the magnitude of their effects and their phenotypic variances. Cofactors were added on a backward regression analysis to increase the likelihood of finding a QTL. Genomic regions with a LOD (log-likelihood) >3 were considered significant QTL (Brody 2019).
Results

Phenotypic variation for seed protein and oil content

The phenotypic variation for seed protein content for each population across the four environments (18CR, 19RO, 19MO, and 20FA) showed a bell-shaped distribution, typical of quantitative traits (Supplementary Fig. S1). In 18CR, seed protein content (on a % dry matter basis) for Pop1 and Pop2 ranged from 38.12 to 46.57 and 38.96 to 47.56, with a mean of 42.16 ± 1.63% and 43.20 ± 1.47%, respectively. Similarly, for 19MO, seed protein showed a range (mean ± standard deviation) of 40.8% to 49.16% (44.8 ± 1.54%) and 39.93% to 49.29% (45.36 ± 1.56%) for Pop1 and Pop2, respectively. In 19RO, range, mean and standard deviation for Pop1 and Pop2 were 42.5%–49.42% (45.35 ± 1.23%) and 41.06%–49.78% (45.83 ± 1.35%), respectively. The observed range, mean, and standard deviation were 38.43%–45.22% (41.98 ± 1.09%) and 38.06%–45.39% (41.69 ± 1.31%) for Pop1 and Pop2, respectively, for 20FA (Table 1). Analysis of variance revealed a high significant effect of environment on protein content in both populations (Supplementary Tables S1 & S2).

Oil content also showed a bell-shaped distribution (Supplementary Fig. S2). The mean oil content was 20.47 ± 0.72%, 21.29 ± 0.85%, 21.38 ± 0.85% and 23.27 ± 0.95% for Pop1 in 20FA, 19MO, 19RO, and 18CR, respectively. For Pop2, the oil content was 22.12 ± 0.97% in 18CR, 20.28 ± 0.91% in 19MO, 20.90 ± 0.93% in 19RO, and 20.10 ± 0.74% in 20FA. Similar to the protein content, environments showed highly significant effects on oil content (Table 2).

Table 1. Descriptive statistics of seed protein content (% dry matter) in 250 F2-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy.

| Environment† | Population‡ | Mean ± Sd§ | CV(%)¶ | Minimum | Maximum | Range |
|--------------|-------------|------------|-------|---------|---------|-------|
| 18CR         | Pop1        | 42.16 ± 1.63 | 3.88 | 38.12 | 46.57 | 8.45 |
|              | Pop2        | 43.20 ± 1.47 | 3.41 | 38.96 | 47.56 | 8.6  |
| 19MO         | Pop1        | 44.8 ± 1.50  | 3.36 | 40.80 | 49.16 | 8.36 |
|              | Pop2        | 45.33 ± 1.57 | 3.47 | 39.93 | 49.29 | 9.36 |
| 19RO         | Pop1        | 45.49 ± 1.19 | 2.62 | 42.83 | 49.42 | 6.59 |
|              | Pop2        | 45.80 ± 1.31 | 2.87 | 41.06 | 49.78 | 8.72 |
| 20FA         | Pop1        | 41.98 ± 1.09 | 2.59 | 38.43 | 45.22 | 6.69 |
|              | Pop2        | 41.69 ± 1.31 | 3.16 | 38.06 | 45.39 | 7.33 |

† 18CR: 2018 Upala, Costa Rica; 19MO: 2019 Portageville, MO; 19RO: 2019 Rohwer, AR; 20FA: 2020 Fayetteville, AR
‡ Pop1: UA 5115C/UA 5814HP; Pop2: R13–532/UA 5814HP
§ Sd: standard deviation
¶ CV: Coefficient of variation
**Table 2.** Descriptive statistics of seed oil content (% dry matter) in 250 F<sub>2</sub>-derived breeding lines each of two different breeding populations when evaluated in four locations and assessed with near-infrared spectroscopy.

| Environment† | Population‡ | Mean ± Sd § | CV(%)¶ | Minimum | Maximum | Range |
|--------------|-------------|--------------|--------|---------|---------|-------|
| 18CR         | Pop1        | 23.27 ± 0.95 | 4.08   | 20.44   | 26.18   | 5.74  |
|              | Pop2        | 22.12 ± 0.97 | 4.41   | 19.2    | 24.63   | 5.43  |
| 19MO         | Pop1        | 21.29 ± 0.85 | 4.02   | 18.4    | 23.41   | 5.01  |
|              | Pop2        | 20.28 ± 0.91 | 4.48   | 17.4    | 22.7    | 5.3   |
| 19RO         | Pop1        | 21.38 ± 0.85 | 4.01   | 17.82   | 23.63   | 5.81  |
|              | Pop2        | 20.90 ± 0.93 | 4.47   | 18.14   | 23.77   | 5.63  |
| 20FA         | Pop1        | 20.47 ± 0.72 | 3.55   | 17.99   | 22.16   | 4.17  |
|              | Pop2        | 20.10 ± 0.74 | 3.72   | 18.03   | 22.18   | 4.15  |

† 18CR: 2018 Upala, Costa Rica; 19MO: 2019 Portageville, MO; 19RO: 2019 Rohwer, AR; 20FA: 2020 Fayetteville, AR
‡ Pop1: UA 5115C/UA 5814HP; Pop2: R13–532/UA 5814HP
§ Sd: standard deviation
¶ CV: Coefficient of variation

**Trait and environment correlation, and heritability**

Pearson correlation analysis showed a highly significant negative correlation (p < 0.0001) between seed oil and protein content varying from $r = -0.545$ to $-0.775$ for Pop 1 and $r = -0.587$ to $-0.655$ for Pop2 (Figure 1). Additionally,
correlation of protein levels among environments was found be significant (p < 0.05) and moderately positive, ranging from \( r = 0.28 \) to 0.47 and \( r = 0.27 \) to 0.47 for Pop1 and Pop2, respectively (Figure 2). Similarly, correlation of oil levels among environments was positive, ranging from 0.23 to 0.43 for Pop1 and from 0.30 to 0.54 for Pop2 (Figure 3).

The broad-sense heritability for protein and oil content (\( h^2 \)) was 91.24% and 90.33% for Pop1, and 93.44% and 93.91% for Pop2, respectively. These results indicate that soybean seed protein and oil contents were highly heritable and mainly influenced by genetic factors under our experimental conditions (Supplementary Table S1 and S2).

**QTL analysis of seed protein**

Analysis of QTL associated with protein content in individual environments for Pop1 showed eight QTL detected on six chromosomes (Gm03,
Gm06, Gm13, Gm15, Gm16, and Gm18). Of these eight QTL, one was identified in 19RO, one in 19MO, two in 20FA, and four in 18CR (Supplementary Table S3). These QTL had absolute additive effect that ranged from 0.15 to 0.95, and explained 4% to 15% of the phenotypic variation. Of the QTL, a region on Chr. Gm03 covering a confidence interval of 28–49 cM was observed both in 18CR and 20FA environments, albeit the actual SNP closest to peak was different. In addition, two nearby regions in Chr. Gm13 were associated with protein content in 19RO and 18CR (Supplementary Table S3). All other QTL were not consistent across environments. In an across-environment analysis, three QTL were identified on Chr. Gm06, Gm13, and Gm18, with an absolute additive effect of 0.42, 0.36, and 0.31, and explaining 12%, 9%, and 7% of phenotypic variation, respectively (Table 3). The negative value of additive effects –0.36 (Chr.13 for Pop1) and –0.24, –0.34, –0.45, –0.11, –0.50 (Chr. Gm04, Gm05, Gm13, Gm16, and Gm20, correspondingly) for Pop 2 described in Table 3 indicated that favorable alleles for increasing protein were from UA 5814HP, except for the QTL on chromosomes Gm06 and Gm18.

In Pop2, for protein content, individual-environment results showed 14 QTL located on 10 chromosomes. Two QTL were found in the environments 18CR and 20FA, whereas four QTL were identified in the environment 19RO and six in 19MO (Supplementary Table S4). Absolute additive effects ranged from 0.03 to 0.81, with 2% to 13% of phenotypic variation explained. Similar to Pop1, the negative value of the additive effects was related to the favorable allele for increasing protein coming from male parent UA 5814HP on chromosomes Gm04, Gm05, Gm6, Gm17, and Gm20. The QTL analysis across environment for Pop2 showed six QTL across six chromosomes (Gm04, Gm05, Gm06, Gm13, Gm16, and Gm20), with absolute additive effect values ranging from 0.11 to 0.51, and explaining 1% to 18% of the phenotypic variation (Table 3).

Comparing results across populations, we observed that QTL Gm06_46078974_G_A on chromosome Gm06 was consistently identified in both Pop1 and Pop2 in the across-environment analysis, with mean effects of 0.42 and 0.51 for Pop1 and Pop2, respectively (Table 3). However, the higher protein allele was not inherited from BARC-7. Additionally, a QTL on chromosome Gm13 was identified in both populations and traced to BARC-7, with 0.36 and 0.45 absolute allelic effects; however, the location in the linkage map was not the same for the Pop1 and Pop2 resulting in two different QTL (Table 3).
Table 3. QTL analysis of seed protein content for two breeding populations each consisting of 250 $F_2$-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cM) within each population. Previously-reported QTL in the reported region are presented, as available on Soybase.org as of 10 October 2021.

| Population | QTL   | Chr. | CI (cM)  | LOD at QTL_Peak | Closest Marker to QTL_Peak | Additive effect (%) | $R^2$ (%) | Previously reported QTL within CI $\dagger$ |
|------------|-------|------|----------|-----------------|---------------------------|--------------------|--------|---------------------------------|
| UA 5115C/UA 5814HP | qPROT_6 | Gm06 | 137–166.7 | 4.9             | Gm06_46078974_G_A         | 0.42               | 12%    | -                               |
|            | qPROT_13 | Gm13 | 135.9–140.3 | 3.06           | Gm13_32011502_T_C         | −0.36              | 9%     | SeedProtein 36–20, 36–21        |
| R13-532/UAS814HP | qPROT_18 | Gm18 | 0–14.9   | 3.68            | Gm18_2228646_C_T          | 0.31               | 7%     | SeedProtein 20–1, 26–12, 28–5, 36–25, 47–7 |
|            | qPROT_4  | Gm04 | 155.9–169.8 | 3.67           | Gm04_47092275_T_C         | −0.24              | 5%     | -                               |
|            | qPROT_5  | Gm05 | 46–81.2   | 4.39            | Gm05_3764264_C_T          | −0.34              | 5%     | SeedProtein 9–1, 12–1, 41–1     |
|            | qPROT_6  | Gm06 | 97.6–126.1 | 8.65            | Gm06_46078974_G_A         | 0.51               | 15%    | qSeedProtein-012, SeedProtein 13–2, 21–3, 24–1, 26–7, 28–1, 29–1, 35–1, 35–2, 36–7, 36–8 |
|            | qPROT_13 | Gm13 | 49.8–74.5  | 6.1             | Gm13_29524219_A_C         | −0.45              | 11%    | SeedProtein 21–6, 33–2, 36–18, 36–20, 36–21, 26–23, 26–24 |
|            | qPROT_16 | Gm16 | 78.1–103.5 | 4.07            | Gm16_33360539_T_C         | −0.11              | 1%     | SeedProtein 41–6                 |
|            | qPROT_20 | Gm20 | 118.7–166.6 | 6.1             | Gm20_45327121_C_A         | −0.5               | 18%    | -                               |

$\dagger$ Chromosome number

$\dagger$ Genome-wide 1000 permutation tests – LOD = 3 (a threshold value)

§ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from UA 5115C

$\dagger$ Previously-reported QTL in Soybase.org near particular QTL. A dash symbolized a new QTL.
**QTL analysis of seed oil**

A total of nine QTL for seed oil content were mapped in Pop1 (UA 5115C/UA 5814HP) in single-environment analyses (Supplementary Table S5). Of these, three QTL were mapped in 18CR, and two QTL each in 19MO, 19RO, and 20FA. These QTL were identified on chromosome Gm06 (3 QTL), Gm08 (1), Gm10 (1), Gm13 (2), Gm18 (1), and Gm20 (1). QTL on Gm06, Gm08, Gm10, Gm18 showed negative additive effects (−0.38, −0.31, −0.22, −0.02) and contributed an average of phenotypic variation of 8%, 4%, 10%, and 1%, respectively, indicating the negative effect on oil content from the UA 5814HP. The QTL on Gm06 (2 QTL), Gm13 (2), and Gm20 (1) had a positive additive effect, ranging from 0.10 to 0.64, explaining phenotypic variation from 5% to 33%, which indicated that they traced to the normal protein parent. An across-environment analysis revealed three QTL associated with oil content, found on chromosomes Gm06, Gm13, and Gm15, with an additive effect of −0.22, +0.42, and +0.34, which explained an average of 9%, 15% and 23% of phenotypic variation (Table 4).

In Pop2 (R13–532/UA 5814HP), in single-environment analyses, a total of 18 QTL on 10 chromosomes were associated with seed oil content. Seven QTL were observed in the environment 18CR, 4 each in 19MO and 19RO, and 3 in 20FA. These QTL were found on Gm01 (1 QTL), Gm04 (3), Gm06

**Table 4.** QTL analysis of seed oil content for two breeding populations each consisting of 250 F2-derived breeding lines. Analysis conducted using composite interval mapping of phenotypic data consisting of least-square means from 4-environment 1-rep trials. Confidence intervals in centimorgan (cm) within each population. Previously-reported QTL in the reported region are presented, as available on Soybase.org as of 10 October 2021.

| Population | QTL  | Chr. † | CI (cm) | LOD at QTL_Peak ‡ | Closest Marker to QTL_Peak | Additive effect (%) § | R² (%) | Previously reported QTL within CI ‖ |
|------------|------|--------|---------|-------------------|---------------------------|----------------------|--------|----------------------------------|
| UA 5115C/  | qOil_6 | Gm06  | 151.3–   | 2.65              | Gm06_46321637_C_T         | −0.22                | 9%     | SeedOil 43–7                      |
| UA 5814HP  |      |       | 153.4    |                   |                           |                       |        |                                  |
|            | qOil_13 | Gm13  | 175.5–   | 2.96              | Gm13_37339900_C_T         | 0.42                 | 15%    |                                  |
|            |      |       | 181.9    |                   |                           |                       |        |                                  |
|            | qOil_15 | Gm15  | 71.5–    | 4.74              | Gm15_12531884_A_G         | 0.34                 | 23%    | SeedOil 43–15                     |
|            |      |       | 119.9    |                   |                           |                       |        |                                  |
| R13–532/   | qOil_4 | Gm04  | 77.4–    | 3.15              | Gm04_9174100_G_A          | −0.27                | 8%     | SeedOil 23–1                      |
| UA 5814HP  |      |       | 87.4     |                   |                           |                       |        | 27–1, 30–5                         |
|            | qOil_6 | Gm06  | 83.7–    | 4.92              | Gm06_18447419_G_A         | −0.29                | 12%    | SeedOil 36–5                      |
|            |      |       | 108      |                   |                           |                       |        | 37–7                             |
|            | qOil_13 | Gm13  | 54.5–    | 4.68              | Gm13_29524129_A_C         | 0.28                 | 11%    | SeedOil 13–5                      |
|            |      |       | 74.5     |                   |                           |                       |        |                                  |
|            | qOil_20 | Gm20  | 102.6–   | 6.87              | Gm20_41356542_T_C         | 0.41                 | 22%    |                                  |
|            |      |       | 163.6    |                   |                           |                       |        |                                  |

† Chromosome number  
‡ Genome-wide 1000 permutation tests – LOD = 3 (a threshold value)  
§ Negative value: allele inherited from UA 5814HP; positive value: allele inherited from UA 5115C  
‖ Previously-reported QTL in Soybase.org near particular QTL. A dash symbolized a new QTL.
(3), Gm08 (2), Gm11 (1), Gm13 (3), Gm14 (1) Gm17 (1), Gm18 (1), and Gm20 (2); with a range between 3% and 22% of the phenotypic variation explained, and absolute additive effects varied from 0.07 to 0.52 (Supplementary Table S6). There were 10 QTL with negative additive effects, indicating that the alleles were from the parent UA 5814HP. In the across-environments analysis, we observed four QTL, mapped on Chr. Gm04, Gm06, Gm13, and Gm20, with LOD > 3.0 (Table 4). Similar to the protein results, the QTL on Chr. Gm06 and Gm13 were detected for both populations and detected in most environments. We also observed in Table 4 that the additive effect for oil for Gm06 was negative (−0.22 for Pop1 and −0.29 for Pop2), and Gm13 was positive (0.42 for Pop1 and 0.28), which was in the contrast with the additive effect for protein on the same chromosome, as the high correlation coefficient between protein and oil contents.

**Discussion**

The present study investigated the genetic control of the high protein content inherited from BARC-7 source, through QTL mapping in two F2-derived populations from a cross between UA 5115C/UA 5814HP and R13–532/UA 5814HP, which were evaluated in four different environments. Both populations exhibited a typical normal distribution, with protein and oil ranges within expected values based on parents. UA 5814HP is a line with high protein, averaging 45.5% protein, and moderate oil content (20.5%) on a dry weight basis (Chen et al. 2017); and although neither UA 5814HP nor the other parents were evaluated in the trials, the progeny showed a wide range for protein and oil values, as expected from transgressive segregation.

The average protein content of both populations was low in environments 18CR and 20FA compared to 19MO and 19RO. This is likely due to the higher temperature in 18CR and 20FA than in 19RO and 19MO during the growing season (data not shown). Conversely, higher temperature tends to increase oil content (Mourtzinis et al. 2017). That is the case in our current study, where environments 19CR and 20 FA had a high oil content compared to 19RO and 19MO. Similar observations on the effect of temperature on protein and oil were also reported in previous studies (Dornbos and Mullen 1992; Piper and Boote 1999; Specht et al. 2001; Mourtzinis et al. 2017; Novikova et al. 2018; Mertz-Henning et al. 2018; Lee et al. 2019; Sobko et al. 2020). Results also showed a negative correlation between protein and oil content. Highly negative phenotypic correlations between protein and oil are well documented in the literature (Cober and D Voldeng 2000; Assefa et al. 2018; Mertz-Henning et al. 2018; Novikova et al. 2018; Lee et al. 2019; Kambhampati et al. 2020; Yao et al. 2020; Li et al. 2021). This indicated that increasing seed protein concentration using phenotypic selection may occur at the expense of oil concentration and vice versa (Chung et al. 2003).
addition, we found a significant correlation between environments for protein and oil content. The correlation indicated that the environment is a factor that affects protein and oil; therefore, it is crucial to evaluate such traits across different environments in a mapping study. However, our study also showed high heritability for protein and oil content, which indicated that the traits were under a high level of genetic control. The high heritability of protein and oil has also been previously reported (Jain et al. 2018; Tian et al. 2020; Jiang et al. 2020; Zhang et al. 2021; Arnold et al. 2021).

Previous studies reported many QTL for protein content identified on Chr. Gm06, Gm15, Gm18, and Gm20 (Diers et al. 1992; Brummer et al. 1997; Warrington et al. 2015). In our study, we identified eight QTL associated with seed protein on Chr. Gm04, Gm05, Gm06, Gm13, Gm16, Gm18, and Gm20. The proportion of the phenotypic variance explained by a given QTL ($R^2$ value) is a parameter in deciding whether marker-assisted selection can be more efficient than conventional phenotypic selection alone (Bernardo 2001; Bernardo and Charcosset 2006). The QTL with large effect ($R^2 \geq 10\%$) were present on Chr. Gm06, Gm13 and Gm20. It is important to note that the major QTL identified in this study on Chr. Gm06 from Pop1, and on Chr. Gm13 from Pop1 and Pop2 have been reported in many previous studies (Table 3). It is crucial to highlight that the QTL on Chr. Gm06 from Pop2 has been reported as “cqSeedProtein-012” (Pathan et al. 2013). Indeed, the “cq” designation in SoyBase, indicates a “confirmed QTL.” There are only 16 QTL for protein listed as cq in SoyBase and the remaining QTL have not been confirmed to date (Grant et al. 2010). The major QTL on Chr. Gm20 (118.7–166.6 cM) from Pop2, and Chr. Gm06 (137–166.7 cM) from Pop1 have not been reported yet. Those could be potential novel QTL. In fact, the known QTL on Chr. Gm20 (20–40 cM) has been reported in multiple studies (Diers et al. 1992; Chung et al. 2003; Nichols et al. 2006; Bandillo et al. 2015; Warrington et al. 2015), but was not found to be significant in our study. Previous studies revealed that high-protein alleles at that locus have historically been associated with decreases in both seed yield and oil content. Some sources of high protein alleles include PI 437088A, PI 407780A, PI 468916, and Danbaekkong (PI 619083) (Chung et al. 2003; Warrington et al. 2015; Kim et al. 2016; Diers et al. 1992; Nichols et al. 2006). The results of our study suggest that BARC-7 may carry alleles different from Danbaekkong; this could be useful for breeders to diversify sources of higher protein.

Apart from the major QTL, some QTL had relatively small effects and were environment-specific. The inconsistency of the protein QTL could be explained by either genotype specificity or sensitivity to environmental conditions (Patil et al. 2017). This could also be elucidated by the fact that the protein is a complex, quantitative, heritable trait controlled by multiple genes and affected by environmental conditions and each of
them might express differently under given environments (Akond et al. 2014; Li et al. 2019). These inconsistent QTL across environments might bring the challenges for breeders to select by using a few markers in a breeding program.

A total of seven QTL associated with the oil content in this study were mainly detected on chromosomes Gm04, Gm06, Gm13, Gm15, and Gm20 for both populations across all environments. Of these, the QTL on Chr. Gm13 (175–181 cM) for Pop1 and on Chr. Gm04 for Pop2 (77.4–87.4 cM) are novel QTL. Akin to the protein, the inconsistency of the oil QTL can be explained by the specificity of the genotype or by its sensitivity to environmental conditions (Patil et al. 2017). The oil QTL that has the largest effect is the one on Chr. Gm15 for Pop1 ($R^2 = 23\%$). Mao et al. (2013) reported the same region on Chr. Gm15 (SeedOil 43–15) in Soybase.

In our study, not all QTL associated with oil and protein content co-located in the same exact regions, in agreement with different studies of diverse genetic backgrounds (Feng, Shi, and Wang 2009; Pathan et al. 2013; Mao et al. 2013; Rossi et al. 2013). However, the oil QTL detected on Chr. Gm20 ($R^2 = 22\%$) and on Chr. Gm13 ($R^2 = 11\%$) did overlap with protein QTL in Pop2. Moreover, the sign of the additive effect is also flipped for the oil content as compared to those of the protein QTL. This emphasizes how protein and oil are correlated, as has also been reported in many previous studies (Cober and D Voldeng 2000; Assefa et al. 2018; Mertz-Henning et al. 2018; Novikova et al. 2018; Lee et al. 2019; Kambhampati et al. 2020; Sobko et al. 2020; Yao, 2020; Finoto et al. 2021; Li et al. 2021).

**Conclusions**

A preliminary mapping using 250 $F_2$-derived lines each from two populations showed a QTL on Chr. Gm13, explaining approximately 10% of variation for seed protein content, and one QTL further downstream in Chr. 20 (only detected on population two), explaining 18% of protein variation. An ongoing fine-mapping using an advanced inbred-line mapping approach will help confirm and fine-map the regions associated with high protein and oil in BARC-7 genetic background.

**Acknowledgments**

The authors thank the generous support of the United Soybean Board. The authors thank Dr. Q. Song with USDA for the genotyping support.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).
Funding

This work was supported by the United Soybean Board.

References

Akon, M., S. Liu, M. Boney, S. K. Kantartzi, K. Meksem, N. Bellaloui, , ... and M. A. Kassem. 2014. “Identification of Quantitative Trait Loci (QTL) Underlying Protein, Oil, and Five Major Fatty Acids’ Contents in Soybean.” American Journal of Plant Sciences 5 (1): 158–167.

Arnold, B., E. Menke, M. A. Mian, Q. Song, B. Buckley, and Z. Li. 2021. “Mining QTL for Elevated Protein and Other Major Seed Composition Traits from Diverse Soybean Germplasm.” Molecular Breeding 41 (8): 1–18. doi:10.1007/s11032-021-01242-z.

Assefa, Y., N. Bajjalieh, S. Archontoulis, S. Casteel, D. Davidson, P. Kovács, ... I. A. Ciampitti. 2018. “Spatial Characterization of Soybean Yield and Quality (Amino Acids, Oil, and Protein) for United States.” Scientific Reports 8 (1): 1–11. doi:10.1038/s41598-018-32895-0.

Bandillo, N., D. Jarquin, Q. Song, R. L. Nelson, P. Cregan, J. Specht, and A. Lorenz (2015). “A Population Structure and Genome-wide Association Analysis on the USDA Soybean Germplasm Collection”.

Bernardo, R. 2001. “What if We Knew All the Genes for a Quantitative Trait in Hybrid Crops?” Crop Science 41 (1): 1–4. doi:10.2135/cropsci2001.4111.

Bernardo, R., and A. Charcosset. 2006. “Usefulness of Gene Information in Marker-Assisted Recurrent Selection: A Simulation Appraisal.” Crop Science 46 (2): 614–621. doi:10.2135/cropsci2005.05-0088.

Birt, D. F., S. Hendrich, D. L. Alekel, and M. Anthony. 2004. “Soybean and the Prevention of Chronic Human Disease.” Soybeans: Improvement, Production, and Uses 16: 1047–1117.

Brody, L. 2019. “LOD Score. In Talking Glossary of Genetic Terms. National Human Genome Research Institute.” Available online at Accessed 10/14/2021 https://www.genome.gov/genetics-glossary/LOD-Score

Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox, and R. C. Shoemaker. 1997. “Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations.” Crop Science 37 (2): 370–378. doi:10.2135/cropsci1997.0011183X003700020011x.

Carneiro, R. C., S. E. Duncan, S. F. O’Keefe, Y. Yin, C. L. Neill, and B. Zhang. 2020. “Sensory and Consumer Studies in Plant Breeding: A Guidance for Edamame Development in the US.” Frontiers in Sustainable Food Systems 4: 124. doi:10.3389/fsufs.2020.00124.

Chen, P., L. Florez-Palacios, M. Orazaly, P. Manjarrez-Sandoval, C. Wu, J. C. Rupe, R. T. Robbins, T. Kirkpatrick, and R. T. Robbins. 2017. “Registration of UA 5814HP’Soybean with High Yield and High Seed-Protein Content.” Journal of Plant Registrations 11 (2): 116–120. doi:10.3198/jpr2016.09.0046crc.

Chung, J., H. L. Babka, G. L. Graef, P. E. Staswick, D. J. Lee, P. B. Cregan, and J. E. Specht. 2003. “The Seed Protein, Oil, and Yield QTL on Soybean Linkage Group I.” Crop Science 43 (3): 1053–1067. doi:10.2135/cropsci2003.1053.

Clemente, T. E., and E. B. Cahoon. 2009. “Soybean Oil: Genetic Approaches for Modification of Functionality and Total Content.” Plant Physiology 151 (3): 1030–1040. doi:10.1104/pp.109.146282.

Cober, E. R., and H. D. Voldeng. 2000. “Developing High-protein, High-yield Soybean Populations and Lines.” Crop Science 40 (1): 39–42. doi:10.2135/cropsci2000.40139x.
Diers, B. W., P. Keim, W. R. Fehr, and R. C. Shoemaker. 1992. “RFLP Analysis of Soybean Seed Protein and Oil Content.” Theoretical and Applied Genetics 83 (5): 608–612. doi:10.1007/BF00226905.

Dornbos, D. L., and R. E. Mullen. 1992. “Soybean Seed Protein and Oil Contents and Fatty Acid Composition Adjustments by Drought and Temperature.” Journal of the American Oil Chemists Society 69 (3): 228–231. doi:10.1007/BF02635891.

Fasoula, V. A., D. K. Harris, and H. R. Boerma. 2004. “Validation and Designation of Quantitative Trait Loci for Seed Protein, Seed Oil, and Seed Weight from Two Soybean Populations.” Crop Science 44 (4): 1218–1225. doi:10.2135/cropsci2004.1218.

Feng, J. P., Q. H. Shi, and X. F. Wang. 2009. “Effects of Exogenous Silicon on Photosynthetic Capacity and Antioxidant Enzyme Activities in Chloroplast of Cucumber Seedlings under Excess Manganese.” Agricultural Sciences in China 8 (1): 40–50. doi:10.1016/S1671-2927(09)60007-9.

Finoto, E. L., M. B. B. Soares, J. D. A. A. D. Albuquerque, J. L. L. Monteiro Neto, S. D. S. Maia, S. Dona, C. Abanto-Rodriguez, E. S.-D. Silva, and C. A. Rodriguez. 2021. “Selection of Soybean Genotypes for Yield, Size, and Oil and Protein Contents.” Australian Journal of Crop Science 15 (1): 48–50. doi:10.21475/ajcs.21.15.01.2390.

Florez-Palacios, L., L. Mozzoni, M. Orazaly, P. Manjarrez-Sandoval, C. Wu, D. Dombek, and P. Chen. 2020. “Registration of Soybean Germplasm R11–7999 with High Seed Protein Content and High Yield.” J. Plant Registrations 14 (1): 82–86. doi:10.1002/plr2.20019.

Florez-Palacios, L., M. Orazaly, P. Manjarrez-Sandoval, C. Wu, L. Mozzoni, D. Dombek, and P. Chen. 2019. “Registration of ‘UA 5115C’Soybean Cultivar.” Journal of Plant Registrations 13 (1): 38–41. doi:10.3198/jpr2018.03.0014rc.

Garcia, M. C., M. Torre, F. Laborda, and M. L. Marina. 1997. “Rapid Separation of Soybean Globulins by Reversed-phase High-performance Liquid Chromatography.” Journal of Chromatography. A 758 (1): 75–83. doi:10.1016/S0021-9673(96)00717-0.

Grant, D., R. T. Nelson, S. B. Cannon, and R. C. Shoemaker. 2010. “SoyBase, the USDA-ARS Soybean Genetics and Genomics Database.” Nucleic Acids Research 38 (suppl_1): D843–D846. doi:10.1093/nar/gkp798.

Jain, R. K., A. Joshi, H. R. Chaudhary, A. Dashora, and C. L. Khatik. 2018. “Study on Genetic Variability, Heritability and Genetic Advance in Soybean [Glycine Max (L.) Merrill].” Legume Research-An International Journal 41 (4): 532–536.

Jiang, G. L., D. N. Katuuramu, Y. Xu, S. Ren, and L. K. Rutto. 2020. “Analysis and Comparison of Seed Protein, Oil, and Sugars in Edamame Dried Using Two Oven-drying Methods and Mature Soybeans.” Journal of the Science of Food and Agriculture 100 (10): 3987–3994. doi:10.1002/jsfa.10443.

Kambhampati, S., J. A. Aznar-Moreno, C. Hostetler, T. Caso, S. R. Bailey, A. H. Hubbard, D. K. Allen, and D. K. Allen. 2020. “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 (1): 18. doi:10.3390/metabo10010018.

Kim, M., S. Schultz, R. L. Nelson, and B. W. Diers. 2016. “Identification and Fine Mapping of a Soybean Seed Protein QTL from PI 407788A on Chromosome 15.” Crop Science 56 (1): 219–225. doi:10.2135/cropsci2015.06.0340.

Lee, S., K. Van, M. Sung, R. Nelson, J. LaMantia, L. K. McHale, and M. R. Mian. 2019. “Genome-wide Association Study of Seed Protein, Oil and Amino Acid Contents in Soybean from Maturity Groups I to IV.” Theoretical and Applied Genetics 132 (6): 1639–1659. doi:10.1007/s00122-019-03304-5.
Leffel, R. C. 1992. “Registration of High-protein Soybean Germplasm Lines BARC-6, BARC-7, BARC-8, and BARC-9.” *Crop Science* 32 (2): 502. doi:10.2135/cropsci1992.0011183X003200020054x.

Li, X., P. Wang, K. Zhang, S. Liu, Z. Qi, Y. Fang, . . . H. Ning. 2021. “Fine Mapping QTL and Mining Genes for Protein Content in Soybean by the Combination of Linkage and Association Analysis.” *Theoretical and Applied Genetics* 134 (4): 1095–1122. doi:10.1007/s00122-020-03756-0.

Li, D., X. Zhao, Y. Han, W. Li, and F. Xie. 2019. “Genome-wide Association Mapping for Seed Protein and Oil Contents Using a Large Panel of Soybean Accessions.” *Genomics* 111 (1): 90–95. doi:10.1016/j.ygeno.2018.01.004.

Lusas, E. W. 2004. “Soybean Processing and Utilization.” *Soybeans: Improvement, Production, and Uses* 16: 949–1045. (soybeansimprove).

Mao, T., Z. Jiang, Y. Han, W. Teng, X. Zhao, and W. Li. 2013. “Identification of Quantitative Trait Loci Underlying Soybean Protein and Oil Contents of Soybean across Multi-genetic Backgrounds and Environments.” *Plant Breeding* 132 (6): 630–641. doi:10.1111/pbr.12091.

Mertz-Henning, L. M., L. C. Ferreira, F. A. Henning, J. M. Mandarino, E. D. Santos, M. C. Oliveira, N. Neumaier, J. Farias, and N. Neumaier. 2018. “Effect of Water Deficit-induced at Vegetative and Reproductive Stages on Protein and Oil Content in Soybean Grains.” *Agronomy* 8 (1): 3. doi:10.3390/agronomy8010003.

Mourtzinis, S., A. P. Gaspar, S. L. Naeve, and S. P. Conley. 2017. “Planting Date, Maturity, and Temperature Effects on Soybean Seed Yield and Composition.” *Agronomy Journal* 109 (5): 2040–2049. doi:10.2134/agronj2017.05.0247.

Nichols, D. M., K. D. Glover, S. R. Carlson, J. E. Specht, and B. W. Diers. 2006. “Fine Mapping of a Seed Protein QTL on Soybean Linkage Group I and Its Correlated Effects on Agronomic Traits.” *Crop Science* 46 (2): 834–839. doi:10.2135/cropsci2005.05-0168.

Novikova, L. Y., I. V. Seferova, A. Y. Nekrasov, I. N. Perchuk, T. V. Shelenga, M. G. Samsonova, and M. A. Vishnyakova. 2018. “Impact of Weather and Climate on Seed Protein and Oil Content of Soybean in the North Caucasus.” *Vavilovskij Zbreve” Ural Genetiki I Selekcii/Vavilov Journal of Genetics and Breeding* 22 (6): 708–715. doi:10.18699/VJ18.414.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams. 2005. “Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean.” *Crop Science* 45 (5): 2015–2022. doi:10.2135/cropsci2004.0720.

Pathan, S. M., T. Vuong, K. Clark, J. D. Lee, J. G. Shannon, C. A. Roberts, . . . D. A. Sleper. 2013. “Genetic Mapping and Confirmation of Quantitative Trait Loci for Seed Protein and Oil Contents and Seed Weight in Soybean.” *Crop Science* 53 (3): 765–774. doi:10.2135/cropsci2012.03.0153.

Patil, G., R. Mian, T. Vuong, V. Pantalone, Q. Song, P. Chen, H. T. Nguyen, T. C. Carter, and H. T. Nguyen. 2017. “Molecular Mapping and Genomics of Soybean Seed Protein: A Review and Perspective for the Future.” *Theoretical and Applied Genetics* 130 (10): 1975–1991. doi:10.1007/s00122-017-2955-8.

Piper, E. L., and K. I. Boote. 1999. “Temperature and Cultivar Effects on Soybean Seed Oil and Protein Concentrations.” *Journal of the American Oil Chemists’ Society* 76 (10): 1233–1241. doi:10.1007/s11746-999-0099-y.

Rossi, M. E., J. H. Orf, L. J. Liu, Z. Dong, and I. Rajcan. 2013. “Genetic Basis of Soybean Adaptation to North American Vs. Asian Mega-environments in Two Independent Populations from Canadian× Chinese Crosses.” *Theoretical and Applied Genetics* 126 (7): 1809–1823. doi:10.1007/s00122-013-2094-9.
Singh, R. K., A. Prasad, M. Muthamilarasan, S. K. Parida, and M. Prasad. 2020. “Breeding and Biotechnological Interventions for Trait Improvement: Status and Prospects.” *Planta* 252 (4): 1–18. doi:10.1007/s00425-020-03465-4.

Sobko, O., A. Stahl, V. Hahn, S. Zikeli, W. Claupein, and S. Gruber. 2020. “Environmental Effects on Soybean (Glycine Max (L.) Merr) Production in Central and South Germany.” *Agronomy* 10 (12): 1847. doi:10.3390/agronomy10121847.

Song, Q., L. Yan, C. Quigley, E. Fickus, H. Wei, L. Chen, F. Dong, et al. 2020. “Soybean BARCSoySNP6K.” *An Assay for Soybean Genetics and Breeding Research the Plant J* 104 (3): 800–811. doi:10.1111/tpj.14960.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, J. P. Markwell, K. G. Lark, J. H. Orf, and K. G. Lark. 2001. “Soybean Response to Water: A QTL Analysis of Drought Tolerance.” *Crop Science* 41 (2): 493–509. doi:10.2135/cropsci2001.412493x.

Tian, X., K. Zhang, S. Liu, X. Sun, X. Li, J. Song, … H. Ning. 2020. “Quantitative Trait Locus Analysis of Protein and Oil Content in Response to Planting Density in Soybean (Glycine Max [L.] Merri.) Seeds Based on SNP Linkage Mapping.” *Frontiers in Genetics* 11: 563. doi:10.3389/fgene.2020.00563.

Wang, S., C. Basten, and Z.-B. Zeng. 2012. *Windows QTL Cartographer 2.5*. Raleigh, NC: Department of Statistics, North Carolina State University.

Warrington, C. V., H. Abdel-Haleem, D. L. Hyten, P. B. Cregan, J. H. Orf, A. S. Killam, H. R. Boerma, Z. Li, and H. R. Boerma. 2015. “QTL for Seed Protein and Amino Acids in the Benning x Danbaekkong Soybean Population.” *Theoretical and Applied Genetics* 128 (5): 839–850. doi:10.1007/s00122-015-2474-4.

Yao, Y., Q. You, G. Duan, J. Ren, S. Chu, J. Zhao, Y. Jiao, X. Zhou, and Y. Jiao. 2020. “Quantitative Trait Loci Analysis of Seed Oil Content and Composition of Wild and Cultivated Soybean.” *BMC Plant Biology* 20 (1): 1–13. doi:10.1186/s12870-019-2199-7.

Zhang, S., D. Hao, S. Zhang, D. Zhang, H. Wang, H. Du, D. Yu, and D. Yu. 2021. “Genome-wide Association Mapping for Protein, Oil and Water-soluble Protein Contents in Soybean.” *Molecular Genetics and Genomics* 296 (1): 91–102. doi:10.1007/s00438-020-01704-7.