



Detection of quantitative trait loci controlling soybean protein, oil and amino acid content in a recombinant inbred line population

Elizabeth M. Clevinger · Ruslan Biyashev ·
M. A. Saghai Maroof

Received: 21 October 2025 / Accepted: 6 April 2026 / Published online: 21 April 2026
© The Author(s) 2026

Abstract Soybean is an important source of protein and oil globally with soymeal providing amino acids and metabolizable energy to both humans and animals. The objective of this study was to identify quantitative trait loci (QTL) for amino acids, crude protein and near-infrared reflectance spectroscopy (NIR)-measured seed protein and oil content. To achieve this objective, a recombinant inbred line (RIL) population was genotyped using 12,761 molecular markers and phenotyped for 18 amino acids, crude protein and NIR-measured protein and oil content. Multiple major effect QTL were identified for all three traits on chromosomes 2 and 15. A major effect QTL for one amino acid, cysteine, was identified on chromosome 20. The QTL identified in this study for soybean composition traits and the associated molecular markers may be useful in breeding to improve the nutritional quality of soybeans.

Keywords Soybean · QTL mapping · Amino acids · Protein · RIL population

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10681-026-03724-7>.

E. M. Clevinger · R. Biyashev · M. A. Saghai Maroof (✉)
School of Plant and Environmental Sciences, Virginia
Tech, Blacksburg, VA 24061, USA
e-mail: smaroof@vt.edu

Soybean (*Glycine max* (L.) Merr.) is a source for essential amino acids and metabolizable energy for both human and animal consumption (Patil et al. 2017; Ma et al. 2019). One of the most important uses of soybean is protein rich soymeal for poultry and swine feed, since it has the highest level of crude protein among plant-based protein sources (Brzostowski and Diers 2017; Cromwell and Coffey 1991). The profile of the amino acid constituents which comprise protein is one of the more important aspects of soymeal (Friedman and Brandon 2001).

Amino acids have been divided into two groups, essential and non-essential, based on their requirement and availability during metabolic processes. Non-essential amino acids do not need to be supplied from an outside source, whereas essential amino acids are not produced by animals and are needed for vital metabolic processes (Panthee et al. 2006). All nine essential amino acids including histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine are present in soymeal (Boisen et al. 2000).

Soybean seed protein, amino acids, and oil content are complex quantitatively inherited polygenic traits (Wehrmann et al. 1987; Wilcox 1998; Cober and Voldeng 2000; Singer et al. 2023). Increasing seed protein becomes difficult as there is a strong negative correlation between oil and protein accumulation in seeds (Patil et al. 2017). The relationship between these QTL and the environment adds to the complexity of developing improved soybeans,

especially under unfavorable growing conditions, like drought and heat, which normally decrease protein and increase oil content (Nakagawa et al. 2020). Panthee et al. (2006) used the population N87-984-16×TN93-99 and identified at least one QTL for each of the 18 amino acids studied in their experiment. QTLs were found on almost every chromosome in their study except chromosomes 6, 11 and 12. Li et al. (2018) identified QTL for multiple amino acids in the RIL population of Zheng 92,116×Liaodou 14. They identified pleiotropic amino acid QTL clusters on chromosomes 6, 9, 10, 13, 17 and 20 in this population. Qin et al. (2019) evaluated 15 amino acids using 249 soybean accessions. They identified 231 SNPs significantly associated with variations in amino acid contents. Lee et al. (2019) conducted a genome-wide association study (GWAS) using data from 621 soybean accessions in five environments and identified eight SNPs significantly associated with QTL for methionine, cysteine, lysine and threonine content. Kim et al. (2023) performed GWAS on protein, oil and amino acid content using 203 wild soybean accessions. A set of 44 significant SNPs were identified that were associated with these three traits. Van et al. (2025) utilized multiple approaches to combine network analysis with GWAS to identify significant SNPs for single amino acids.

The identification of unique QTLs responsible for high protein content and essential amino acids is a critical step in the improvement of the nutritional quality of soymeal. The goal of this study was to identify QTL associated with amino acids, protein and oil content. The population of PI 507429×PI 399084 was used previously in Clevinger et al. (2023) to identify large effect QTLs for protein and oil content on chromosomes 2 and 15 using four years of field replicated data. In the present study, additional data from the same population was collected and analyzed on the seed amino acid, protein and oil contents from an additional year of field study in Blacksburg, Virginia, in 2022 with two replications. For this study, the two replications and their average at the F_{12} generation were used to analyze seed amino acids, protein and oil content data.

A complete amino acid profile and crude protein assay were conducted at the University of Missouri-Columbia Agricultural Experiment Station Chemical Laboratories (AOAC Official Methods 982.30E (a,b,c) and 990.03) for both replications based on

dry seed weight. The 18 amino acids analyzed were aspartic acid (Asp), threonine (Thr), serine (Ser), glutamic acid (Glu), proline (Pro), glycine (Gly), alanine (Ala), cysteine (Cys), valine (Val), methionine (Met), isoleucine (Ile), leucine (Leu), tyrosine (Tyr), phenylalanine (Phe), lysine (Lys), histidine (His), arginine (Arg), tryptophan (Trp). These same lines were also assayed for seed protein and oil content utilizing near-infrared reflectance (NIR) spectroscopy at Virginia Tech (Blacksburg, VA) using a DA 7250 NIR Analyzer (Pertin Instruments, Springfield, IL) as described in Clevinger et al. (2023).

To identify chromosomal regions associated with protein, oil and amino acids from the 2022 harvest, the segregating population was genotyped with a total of 12,761 markers including 10,526 SNP loci from GBS, 2217 SNP markers from SoySNP6k Illumina Beadchip (Song et al. 2020), 26 SSR and one indel type marker using DNA from F_7 RILs as described previously in Clevinger et al. (2023). Linkage map construction was done using SNP data from GBS, SSR and one indel type marker and SoySNP6K of the RIL population. JoinMap 4.0 (van Ooijen 2006) with a LOD threshold of 4.0 and a maximum recombination frequency of 50% for the original grouping was employed. Marker order and their positions within each linkage group were determined using the maximum likelihood algorithm and Kosambi mapping function; those unassigned to any linkage group were excluded.

The RIL population was planted in 2022 with two replications. The data from each replication (Table S1A and S1B) and the mean of those two replications (Table S2) resulted in three phenotypic data sets. After performing correlation analysis of the composition trait values between the two replications (Table S3), it was determined they were in agreement, so the average values (Table S2) were used for subsequent analysis. The 18 amino acids assayed were positively correlated with one another, crude protein and NIR protein and negatively correlated with NIR oil (Table S4). All amino acids were negatively skewed except Cys and Trp, which were positively skewed (Table S2). The crude protein content ranged from 31.51 to 48.52 (w/w%), whereas protein content measured via NIR ranged from 32.05 to 46.74% for the RILs (Table S2) in the population. Transgressive segregation in this RIL population was observed in all the traits assayed in this study. There were no lower

end transgressive lines for Cys and Met and no higher end transgressive lines for Gly and Val (Figure S1).

The average value of two replications of the population was used for phenotypic trait data analysis. Major effect QTLs for seventeen amino acids were identified on chromosome 2 in this population (Table 1). Met had the lowest LOD at 6.26 that explained 26.1% of the phenotypic variation. LODs for the remaining 16 amino acid QTLs ranged from 6.97 to 14.39 with explained variations of 28.5% and 50.7% for Trp and Leu, respectively (Table 1). The chromosome 2 QTL region for 12 out of the 17 amino acids shared the same marker, BARC-SoySSR_02_1672, at the maximum LOD (Fig. 1a, Table 1). QTLs for all 17 amino acid contents were

flanked by KASP markers GBS02_47122206K and GBS02_47432297K (KASP converted) on chromosome 2. The physical positions of these flanking markers were 47,122,206 to 47,432,297 bp (Wm82.a4).

Crude protein shared the same flanking markers as the 17 amino acids, GBS02_47122206K and GBS02_47432297K (Fig. 1a), and maximum LOD marker of Gm02_48583877_A_C with an LOD of 12.58 and explained 45.9% of the variation. Protein and oil content measured via NIR also shared the same flanking markers but differed from each other for the marker at the maximum LOD, Gm02_48583877_A_C and BARCSoySSR_02_1672, respectively (Table 1).

Table 1 Quantitative trait loci for amino acid, crude protein, NIR-measured protein and oil content detected on chromosome 2 based on composite interval mapping in the RIL population using the average of data from two replications

Trait	LOD	% Explained variation	Marker at max LOD	Flanking markers
Asp	10.7	40.7	BARCSoySSR_02_1672	GBS02_47122206K and GBS02_47432297K
Thr	12.53	45.8	BARCSoySSR_02_1672	
Ser	11.16	42	BARCSoySSR_02_1672	
Glu	13.51	48.2	Gm02_48583877_A_C	
Pro	10.67	40.6	BARCSoySSR_02_1672	
Gly	13.02	47.1	BARCSoySSR_02_1672	
Ala	12.45	45.7	BARCSoySSR_02_1672	
Val	14.24	50.3	BARCSoySSR_02_1672	
Met	6.26	26.1	BARCSoySSR_02_1672	
Ile	13.59	48.7	Gm02_48583877_A_C	
Leu	14.39	50.7	BARCSoySSR_02_1672	
Tyr	13.38	48.2	BARCSoySSR_02_1672	
Phe	14.05	50.1	BARCSoySSR_02_1672	
Lys	12.74	46.4	BARCSoySSR_02_1672	
His	11.39	42.6	Gm02_48583877_A_C	
Arg	10.89	41.6	Gm02_48583877_A_C	
Trp	6.97	28.5	Gm02_48702118_T_G	
Crude protein	12.58	45.9	Gm02_48583877_A_C	
NIR protein	12.84	46.4	Gm02_48583877_A_C	
NIR oil	7.74	31.5	BARCSoySSR_02_1672	

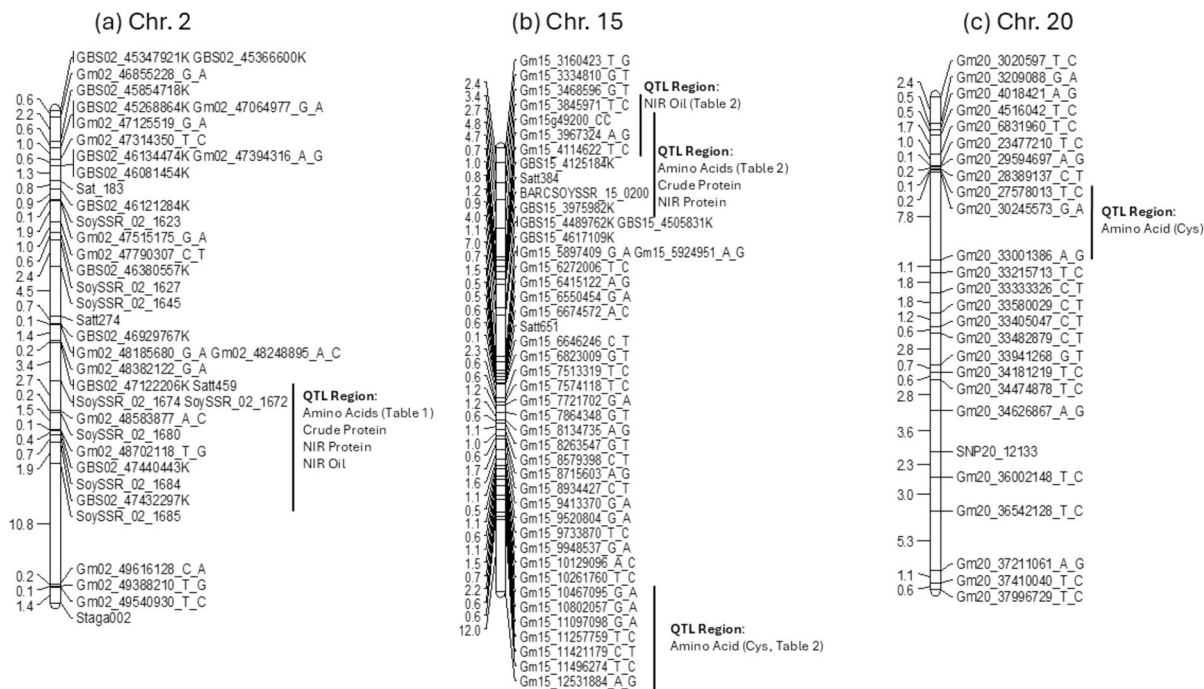


Fig. 1 Regional chromosomal maps for the PI 507429×PI 399084 RIL population with the QTL region for amino acids, crude protein content, NIR-measured protein and oil content shown by the black bar region to the right. **a** regional map for

chromosome 2 **(b)** regional map for chromosome 15, and **(c)** regional map for chromosome 20. Genotypic data is the same as used in Clevinger et al. (2023)

QTLs for 17 amino acids were also identified on chromosome 15. The LODs for these QTL ranged from 3.92 to 6.22 for Ile and Pro, respectively, with the explained phenotypic variation ranging from 18.1 to 27.5% (Table 2). Sixteen out of the seventeen detected amino acid QTL regions had the same marker at the maximum LOD, Gm15_3967324_A_G, with flanking markers of Gm15g49200_CC (based on *Glyma.15g049200*) and GBS15_3975982K (Fig. 1b). The physical positions of these markers on chromosome 15 were (3,861,273–3,863,135) to 3,975,982 bp, respectively (Wm82.a4). A QTL for Cys was identified further down chromosome 15 with marker, Gm15_11257759_T_C, at the maximum LOD and flanking markers, Gm15_10467095_G_A and Gm15_12531884_A_G (Table 2, Fig. 1b). These flanking markers are at the physical positions of 10,545,687 to 12,553,029 bp, respectively (Wm82.a2).

The crude protein QTL on chromosome 15 explained 22.8% of the phenotypic variation with the same marker at the maximum LOD,

Gm15_3967324_A_G, and the same flanking markers, Gm15g49200_CC and GBS15_3975982K, as the majority of the amino acids (Fig. 1b). Protein and oil content measured via NIR explained 21.2% and 29.7% of the phenotypic variation (Table 2). Seed protein content shared the same flanking markers and the marker at the maximum LOD as all the amino acids aside from Cys, whereas seed oil content shared the same marker at the maximum LOD but differed in its flanking markers (Gm15_3845971_T_C and Gm15_4114622_T_C) (Table 2). The physical positions of the flanking markers of the seed oil QTL on chromosome 15 were 3,863,922 to 4,133,547 bp (Wm82.a2).

A QTL for Cys content was identified on chromosome 20. This was the only amino acid QTL mapping onto this chromosome in this population. This QTL explained 19.5% of the variation of this amino acid with an LOD of 4.39. The marker at the maximum LOD is Gm20_30245573_G_A with flanking markers of Gm20_27578013_T_C and Gm20_33001386_A_G (Fig. 1c) with physical

Table 2 Quantitative trait loci for amino acid, crude protein, NIR-measured protein and oil content detected on chromosome 15 based on composite interval mapping in the RIL population using the average of data from two replications

Trait	LOD	% Explained variation	Marker at max LOD	Flanking markers
Asp	5.24	23.6	Gm15_3967324_A_G	Gm15g49200_CC and GBS15_3975982K
Thr	5.21	22.9	Gm15_3967324_A_G	
Ser	5.61	25	Gm15_3967324_A_G	
Glu	4.24	19	Gm15_3967324_A_G	
Pro	6.22	27.5	Gm15_3967324_A_G	
Gly	5.16	23.3	Gm15_3967324_A_G	
Ala	4.68	21.3	Gm15_3967324_A_G	
Val	4.56	20.7	Gm15_3967324_A_G	
Met	4.42	20.2	Gm15_3967324_A_G	
Ile	3.92	18.1	Gm15_3967324_A_G	
Leu	4.56	20.6	Gm15_3967324_A_G	
Tyr	5.57	24.1	Gm15_3967324_A_G	
Phe	4.19	19.2	Gm15_3967324_A_G	
Lys	5.12	22.6	Gm15_3967324_A_G	
His	4.71	21.2	Gm15_3967324_A_G	
Arg	4.78	21.1	Gm15_3967324_A_G	Gm15_10467095_G_A and Gm15_12531884_A_G
Cys	4.45	19.2	Gm15_11257759_T_C	
Crude protein	5.07	22.8	Gm15_3967324_A_G	Gm15g49200_CC and GBS15_3975982K
NIR protein	4.79	21.2	Gm15_3967324_A_G	
NIR oil	6.83	29.7	Gm15_3967324_A_G	Gm15_3845971_T_C and Gm15_4114622_T_C

positions of 28,741,905 to 34,140,804 bp (Wm82.a2).

Soybean composition traits such as seed amino acid, protein and oil content are quantitatively inherited traits. These traits are influenced by multiple genes and environmental conditions such as heat and drought, that complicate the breeding process. Few studies have been conducted to identify genes controlling amino acid content, while many have been performed focusing on protein and oil content (Duan et al. 2023; Patil et al. 2017). The previous mapping studies involving protein, oil and amino acids have identified multiple QTLs across many chromosomes for these traits which make breeding for improvement difficult using conventional breeding methods since so many QTL are involved across multiple chromosomes (Fallen et al. 2013; Khandaker et al. 2015; Warrington et al. 2015; Ma et al. 2019; Singer et al.

2022; Cunicelli et al. 2022). Breeding for improved amino acid content has been hindered due to the lack of genetic variability and the absence of high-throughput and cost-effective phenotyping screening for large numbers of samples (Patil et al. 2017). Our study aimed to look at the relationship between these three soybean composition traits. Nodulation is another trait that influences protein and oil content. The high protein phenotype in soybeans relies on nodules. Protein content levels are reduced in soybean lines that are non-nodulating (Indrasumunar et al. 2011). Patibandla et al. (2025) identified a mutant soybean line that showed a high oil phenotype with consequently low protein levels due to a mutation in the *Glyma.02g270800* gene associated with the *rjl* locus controlling nodulation. This gene, *Glyma.02g270800*, is in the QTL region of chromosome 2 of this study.

These *rjl* mutants are associated with reduced yield and protein. Patibandla et al. (2025) also assayed the PI 507429×PI 399084 population for nodule formation and hypothesized that the PI 507429 parent may also carry an allele of *rjl* that impacts nodule formation causing lower protein levels. Their work determined that the low protein/high oil phenotype was associated with a lack of nodules. They resequenced the gene in PI 507429 and discovered that it has the same polymorphism as the *rjl* reference allele (Patibandla et al. 2025).

In this study, the RIL population was evaluated with two replications in 2022 and identified large effect QTLs on chromosomes 2 and 15 for multiple amino acids, protein and oil contents and on chromosome 20 for Cys amino acid content. These findings align with the seed protein and oil content QTLs detected on chromosomes 2 and 15 (Clevinger et al. 2023) and show a strong association of amino acids with both seed protein and oil contents. The QTL regions identified on chromosomes 2 and 15 for amino acids, crude protein and NIR protein and oil content in this study are located within the QTL regions detected for protein and oil content by Clevinger et al. (2023), aside from the QTL region for Cys content on chromosomes 15 and 20. Multiple candidate genes in the QTL regions on chromosomes 2 and 15 are potentially related to protein activity (Table S5). There were 50 genes identified in the QTL region on chromosome 2 (Table S5a) and 30 and 114 genes, respectively, in the QTL regions of chromosome 15 (Table S5b) identified. A study by Panthee et al. (2006) identified QTL for multiple amino acids in the region of Satt274, above the QTL region for this population on chromosome 2 (Fig. 1a). Van et al. (2026) identified linkage disequilibrium blocks with significant associations with Phe on chromosome 2 (46,612,167 — 46,852,340 bp) below the flanking markers for our area of interest on chromosome 2. Li et al. (2024) predicted 16 candidate genes associated with soybean essential amino acids and 8 of these genes are located in the same QTL region of chromosome 2 in our population.

QTL were detected on chromosome 2 for all individual amino acids assayed, except Cys, in the same QTL region as protein and oil content. It is interesting to note that no Cys QTL was detected on chromosome 2 in this population. The QTL region detected on chromosome 2 for all the traits assayed was within

a narrower region than we previously reported for protein and oil content (Clevinger et al. 2023). A QTL for Cys was detected on chromosome 15 in this population but in a different region than the QTL for other individual amino acids, protein and oil contents. A QTL for Cys was the only trait looked at in this study that was detected on chromosome 20. There are 72 genes in the QTL region on chromosome 20 for Cys (Table S5c). Amino acid and protein content are negatively correlated with oil and as we and others have reported in previous studies, they are intrinsically linked which makes breeding for increased protein or oil difficult. The molecular markers identified for composition traits in this study should have applications in breeding for improved soybean cultivars. Transgressive segregation in the RIL population indicates that favorable alleles are inherited from both parental lines for all traits analyzed. These lines could be useful as breeding line germplasm.

Author contributions Conceptualization: M.A. Saghai Maroof; Methodology: M.A. Saghai Maroof and Ruslan Biyashev; Formal analysis: Elizabeth M. Clevinger and Ruslan Biyashev; Investigation: Elizabeth M. Clevinger and Ruslan Biyashev; Resources: M.A. Saghai Maroof; Data curation: Elizabeth M. Clevinger and Ruslan Biyashev; Writing—original draft preparation: Elizabeth M. Clevinger; Writing—review and editing: Elizabeth M. Clevinger, Ruslan Biyashev, M.A. Saghai Maroof; Supervision: M.A. Saghai Maroof; Project administration: M.A. Saghai Maroof; Funding acquisition: M.A. Saghai Maroof. All authors read and approved the final manuscript.

Funding This project was funded in part by the Virginia Soybean Board and Virginia Agricultural Council. Additional funding was provided by the Agricultural Experiment Station Hatch Program (VA-160136) at Virginia Tech.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not

included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Boisen S, Hvelplund T, Weisbjerg MR (2000) Ideal amino acid profiles as a basis for feed protein evaluation. *Livest Prod Sci* 64(2–3):239–251. [https://doi.org/10.1016/S0301-6226\(99\)00146-3](https://doi.org/10.1016/S0301-6226(99)00146-3)
- Brzostowski LF, Diers BW (2017) Agronomic evaluation of a high protein allele from PI407788A on chromosome 15 across two soybean backgrounds. *Crop Sci* 57(6):2972–2978. <https://doi.org/10.2135/cropsci2017.02.0083>
- Clevinger EM, Biyashev R, Haak D, Song Q, Pilot G, Saghai Maroof MA (2023) Identification of quantitative trait loci controlling soybean seed protein and oil content. *PLoS ONE* 18(6):e0286329. <https://doi.org/10.1371/journal.pone.0286329>
- Cober ER, Voldeng HD (2000) Developing high-protein, high-yield soybean populations and lines. *Crop Sci* 40(1):39–42. <https://doi.org/10.2135/cropsci2000.40139x>
- Cromwell G, Coffey R (1991) Phosphorous-a key essential nutrient, yet a possible major pollutant-its central role in animal nutrition. In: Lyons TP (ed) *Biotechnology in the feed industry*. Alltech Tech Publishers, Nicholasville, KY, pp 133–145
- Cunicelli M, Olukolu BA, Sams C, Schneider L, West D, Pantalone V (2022) Mapping and identification of QTL in 5601T x U99–310255 RIL population using SNP genotyping: soybean seed quality traits. *Mol Biol Rep* 49(7):6623–6632. <https://doi.org/10.1007/s11033-022-07505-y>
- Duan Z, Li Q, Wang H, He X, Zhang M (2023) Genetic regulatory networks of soybean seed size, oil and protein contents. *Front Plant Sci* 14:1160418. <https://doi.org/10.3389/fpls.2023.1160418>
- Fallen BD, Hatcher CN, Allen FL, Kopsell DA, Saxton AM, Chen PY, Kantartzi SK, Cregan P, Hyten D, Pantalone V (2013) Soybean seed amino acid content QTL detected using the universal soy linkage panel 1.0 with 1,536 SNPs. *J Plant Genome Sci* 1(3):68–79
- Friedman M, Brandon DL (2001) Nutritional and health benefits of soy proteins. *J Agric Food Chem* 49(3):1069–1086. <https://doi.org/10.1021/jf0009246>
- Indrasumunar A, Searle I, Lin MH, Kereszt A, Men A, Carroll BJ, Gresshoff PM (2011) Nodulation factor receptor kinase 1alpha controls nodule organ number in soybean (*Glycine max* L. Merr.). *Plant J* 65(1):39–50. <https://doi.org/10.1111/j.1365-3113.2010.04398.x>
- Khandaker L, Akond M, Liu S, Kantartzi SK, Meksem K, Belaloui N, Lightfoot DA, Kassem MA (2015) Mapping of QTL associated with seed amino acids content in “MD96-5722” by “Spencer” RIL population of soybean using SNP markers. *Food Nutr Sci* 6(11):974–984. <https://doi.org/10.4236/fns.2015.611101>
- Kim WJ, Kang BH, Kang S, Shin S, Chowdhury S, Jeong SC, Choi MS, Park SK, Moon JK, Ryu J, Ha BK (2023) A genome-wide association study of protein, oil, and amino acid content in wild soybean (*Glycine soja*). *Plants Basel*. <https://doi.org/10.3390/plants12081665>
- Lee S, Van K, Sung M, Nelson R, LaMantia J, McHale LK, Mian MAR (2019) Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theor Appl Genet* 132(6):1639–1659. <https://doi.org/10.1007/s00122-019-03304-5>
- Li X, Tian R, Kamala S, Du H, Li W, Kong Y, Zhang C (2018) Identification and verification of pleiotropic QTL controlling multiple amino acid contents in soybean seed. *Euphytica*. <https://doi.org/10.1007/s10681-018-2170-y>
- Li X, Tan P, Xiong G, Ma R, Gao W, Jiang A, Liu J, Du C, Zhang J, Zhang X, Zhang L, Yi Z, Fang X, Zhang J (2024) Densification of genetic map and stable quantitative trait locus analysis for amino acid content of seed in soybean (*Glycine max* L.). *Plants Basel*. <https://doi.org/10.3390/plants13152020>
- Ma Y, Ma W, Hu D, Zhang X, Yuan W, He X, Kan G, Yu D (2019) QTL mapping for protein and sulfur-containing amino acid contents using a high-density bin-map in soybean (*Glycine max* L. Merr.). *J Agric Food Chem* 67(44):12313–12321. <https://doi.org/10.1021/acs.jafc.9b04497>
- Nakagawa ACS, Ario N, Tomita Y, Tanaka S, Murayama N, Mizuta C, Iwaya-Inoue M, Ishibashi Y (2020) High temperature during soybean seed development differentially alters lipid and protein metabolism. *Plant Prod Sci* 23(4):504–512. <https://doi.org/10.1080/1343943x.2020.1742581>
- Panthee DR, Pantalone VR, Saxton AM, West DR, Sams CE (2006) Genomic regions associated with amino acid composition in soybean. *Mol Breed* 17(1):79–89. <https://doi.org/10.1007/s11032-005-2519-5>
- Patibandla SV, Carrero-Colon M, Song Q, Qin Q, Clevinger E, Zhu H, Maroof MAS, Hudson K (2025) Elevated Soybean Seed Oil Phenotype Associated with a Single Nucleotide Polymorphism in GmNFR1alpha. *Plants (Basel)* 14(23):3676
- Patil G, Mian R, Vuong T, Pantalone V, Song Q, Chen P, Shannon GJ, Carter TC, Nguyen HT (2017) Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor Appl Genet* 130(10):1975–1991. <https://doi.org/10.1007/s00122-017-2955-8>
- Qin J, Shi A, Song Q, Li S, Wang F, Cao Y, Ravelombola W, Song Q, Yang C, Zhang M (2019) Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. *Front Plant Sci* 10:1445. <https://doi.org/10.3389/fpls.2019.01445>
- Singer WM, Shea Z, Yu D, Huang H, Mian MAR, Shang C, Rosso ML, Song QJ, Zhang B (2022) Genome-wide association study and genomic selection for proteinogenic methionine in soybean seeds. *Front Plant Sci* 13:859109. <https://doi.org/10.3389/fpls.2022.859109>
- Singer WM, Lee YC, Shea Z, Vieira CC, Lee D, Li X, Cunicelli M, Kadam SS, Khan MAW, Shannon G, Mian MAR, Nguyen HT, Zhang B (2023) Soybean genetics, genomics, and breeding for improving nutritional value and reducing

- antinutritional traits in food and feed. *Plant Genome* 16(4):e20415. <https://doi.org/10.1002/tpg2.20415>
- Song Q, Yan L, Quigley C, Fickus E, Wei H, Chen L, Dong F, Araya S, Liu J, Hyten D, Pantalone V, Nelson RL (2020) Soybean BARCSoySNP6K: an assay for soybean genetics and breeding research. *Plant J* 104(3):800–811. <https://doi.org/10.1111/tpj.14960>
- Van K, Clevinger EM, Biyashev RM, Collakova E, Pilot G, Graef G, McHale LK, Saghai Maroof MA (2026) Network analysis identifies QTLs and relationships between seed protein and free amino acids in soybean. *Euphytica* 222(4). <https://doi.org/10.1007/s10681-026-03712-x>
- Van K, Lee S, Mian MAR, McHale LK (2025) Network analysis combined with genome-wide association study helps identification of genes related to amino acid contents in soybean. *BMC Genomics* 26(1):21. <https://doi.org/10.1186/s12864-024-11163-8>
- van Ooijen JW (2006) JoinMap 4.0: software for the calculation of genetic linkage maps in experimental populations. Kyasma B.V., Wageningen, Netherlands
- Warrington CV, Abdel-Haleem H, Hyten DL, Cregan PB, Orf JH, Killam AS, Bajjalieh N, Li Z, Boerma HR (2015) QTL for seed protein and amino acids in the Benning x Danbaekkong soybean population. *Theor Appl Genet* 128(5):839–850. <https://doi.org/10.1007/s00122-015-2474-4>
- Wehrmann VK, Fehr WR, Cianzio SR, Cavins JF (1987) Transfer of high seed protein to high-yielding soybean cultivars. *Crop Sci* 27(5):927–931. <https://doi.org/10.2135/cropsci1987.0011183X002700050020x>
- Wilcox JR (1998) Increasing seed protein in soybean with eight cycles of recurrent selection. *Crop Sci* 38(6):1536–1540. <https://doi.org/10.2135/cropsci1998.0011183X003800060021x>

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