crops

MDPI

Article

Identification of High-Yielding Soybean Lines with Exceptional Seed Composition Qualities

Jay Gillenwater [1,*], Rouf Mian [1,2], Mia Cunicelli [2], Brant McNeece [3] and Earl Taliercio [1,2]

[1] Department of Crop and Soil Sciences, North Carolina State University Raleigh, Raleigh, NC 27695, USA; rouf.mian@usda.gov (R.M.); earl.taliercio@usda.gov (E.T.)
[2] Soybean & Nitrogen Fixation Unit USDA-ARS, Raleigh, NC 27607, USA; mia.cunicelli@usda.gov
[3] Nutrien Ag Solutions, Greenville, MS 38703, USA; brantmcneece@gmail.com
* Correspondence: jhgille2@ncsu.edu

Abstract: In current markets, the primary uses for soybean seeds are in products derived from their oil or protein content. However, growers are compensated based on seed yield, so a more valuable crop is one that does not compromise on yield when compared with existing options, with an optimum combination of protein and oil. A negative correlation of seed protein with seed yield and oil makes the simultaneous improvement of these traits difficult but not impossible through conventional breeding. Selections of lines with exceptional yield and seed composition were made from two recombinant inbred line (RIL) soybean mapping populations to identify high protein and/or high oil lines with yields comparable to elite cultivars. The performance of these RILs was evaluated in multiple environments, and several genotypes were identified with yields comparable to those of high-yielding check cultivars with seed protein and/or oil content superior to the checks. These genotypes will provide breeders with additional sources of germplasm for continuing efforts to improve seed composition traits without compromising seed yield and provide growers with more profitable cultivars.

Keywords: soybean; protein; oil; yield

Citation: Gillenwater, J.; Mian, R.; Cunicelli, M.; McNeece, B.; Taliercio, E. Identification of High-Yielding Soybean Lines with Exceptional Seed Composition Qualities. *Crops* **2023**, *3*, 333–342. https://doi.org/10.3390/crops3040029

Academic Editor: Qing-Yao Shu

Received: 5 September 2023
Revised: 19 October 2023
Accepted: 5 December 2023
Published: 13 December 2023

1. Introduction

Seed yield, oil, and protein are all valuable traits in a soybean cultivar; however, breeding lines which have both high yield and high protein have been difficult to develop due to the negative genetic correlation between the two traits [1–3]. While considerable efforts have been made to identify loci which control these seed quality traits so that marker-assisted selection (MAS) strategies can be utilized for their improvement, to date, the applications of such markers have been few. To be effective in MAS strategies, a QTL must be stable across different environments and also different genetic backgrounds [4–6]. While many QTLs have been published for these three important traits, relatively few have been found which are stable across both environments and genetic backgrounds [7–13]. Perhaps as a consequence of this, there have been relatively few attempts to incorporate these QTLs into MAS schemes [14,15]. While there is still reason to continue this genetic research, it is important that breeders take every opportunity to identify lines with high yield, and desirable seed composition traits like oil and protein content so that new value-added varieties can be more profitable.

The protein and oil content of a soybean cultivar can vary considerably. Accessions in the USDA Soybean Germplasm collection have protein concentrations which range from 34 to 56% and oil contents which range between 8 and 27%, with means of 42.1% and 19.5%, respectively [16]. The market for soybean meals requires a minimum of 47.5% meal protein, which corresponds to approximately 41.5% protein content with 21% oil, on a dry weight basis [17]. Oil and protein content are the most important seed composition traits in soybean, so if one is decreased, the other should be correspondingly increased to

compensate. The inverse correlation between protein and oil contents is well known and is suspected to be due at least partially to the action of pleiotropic genes and competing metabolic pathways, which control the expression of each trait [18,19].

Despite the difficulty in the simultaneous improvement of all three of these traits, varieties with elevated protein contents with little or no yield reduction have been released by public sector breeders. The high-protein germplasm lines R05-1415 and R05-1772 were released recently and contain 46.9% and 46.1% protein content with 94% and 91% yield, respectively, of the high-yielding cultivar 5002T [20]. Cultivars TN03-350 and TN04-5321 contain 43.9% and 43.1% protein while having superior or comparable yields with checks cultivars [21]. The Highpro1 cultivar was released in 2016 and has a yield which is greater than or equal to that of the check cultivars, while its protein content was 5–6% higher than the checks [22].

To develop high-yielding lines with balanced protein and oil contents, two recombinant inbred line (RIL) mapping populations were screened for lines which showed desirable combinations of yield and seed composition traits. Successive rounds of selection were conducted between 2018 and 2021 to identify and characterize lines with high values for yield as well as protein and/or oil composition.

Cultivars produced through conventional breeding techniques have shown that it is possible to identify lines with both high seed protein and seed yield. Efforts to find these lines should continue to provide growers and breeders with additional high value cultivars and germplasm, which can be used to simultaneously improve protein and yield traits. The availability of these sources of germplasm is important because it will support growers by providing more valuable cultivars and will support breeders with sources of germplasm that can be used for the long-term improvement of important seed traits in soybean.

2. Materials and Methods

2.1. Population Development and Line Selection

In 2018, oil mapping populations, POP201 and POP202, were grown as plant rows at the Central Crops Research Station in Clayton, NC (CLA). These populations consisted of 273 and 237 RILs, respectively.

Pop201 was developed from a cross of LMN09-119 × N09-09 in 2015. LMN09-119 was from a cross of N6202 [23] × G98SF-114, G98SF-114 from a 'Benning' [24] × 'Danbaekkong' [25] cross; N09-09 was from a cross of N02-70 × G98-1420; N02-70 was from 'Santee' [26] × 'Holladay' [27]; G98-1420 was from the cross of 'Boggs' [28] × 'Doles' [29]. Population 202 was also developed from a cross of LMN09-19 × Nl3-47 in 2015. LMN09-19 is a sister line with the same pedigree as LMN09-119; N13-47 was derived from a recurrent selection breeding that started with intermating F1 progeny from eight crosses in 1965 at the ARS Soybean Research Unit in Raleigh, NC. These eight crosses ('Arksoy'× 'Lee', Arksoy × 'Ogden', Arksoy × D60-8107, 'Jackson' × Lee, Jackson × Ogden, Jackson × D60-8107, 'Roanoke' × Ogden, and Roanoke × D60-8107) were made in 1964. The male sterile–female fertile line with the *ms1* gene was used in this recurrent selection breeding scheme over many cycles of selections. In each cycle, the top 10% high-oil lines were selected from the progeny to start the next cycle. The N13-47 with desirable agronomic attributes and high (>24%) oil content was selected from a plant row in 2013. Schematic representations of the development of each population are given in Supplementary Figures S1 and S2 for Pop201 and Pop202, respectively. The RILs of the two mapping populations were advanced to F5 generation by the single-seed descent (SSD) method of selection. A single plant was harvested to represent each F5-derived RIL.

The agronomic traits recorded in the field were height, lodging, maturity date, and a composite agronomic score. Lodging was scored on a scale of 1–5, where 5 indicates that all plants in a plot are prostrate on the ground, and a score of 1 indicates that all plants are erect [30]. The agronomic score aimed to capture other traits of value such as the visual estimation of pod load and plot uniformity to provide a general score of a line's agronomic desirability. Agronomic score was recorded on a scale of 1–5 as well, with 1 identifying the

best lines of a population, and 5 the worst. Maturity was recorded at the R8 maturity date and was recorded as the number of days after September 1. Height was measured from the soil to the tip of the main stem.

Following harvest, seed yield, protein, and oil content were measured after seed was air-dried to approximately 7% moisture content under greenhouse conditions. Protein and oil contents were measured on a dry basis using a Perten DA 7250 NIR®instrument (Stockholm, Sweden) [31]. Yield was measured after seed had been sifted and cleaned of debris.

To select lines for the yield tests in the 2019 growing season, lines with low seed yield or extreme maturity dates in 2018 were removed from consideration to select the best performing lines for assessment. Two yield trials were then designed for each mapping population based on maturity dates, agronomic traits, and protein and oil contents. Eighty unique lines were selected from each population, which satisfied these criteria, and each yield test had 40 RILs. Three high-yielding check cultivars and the two parents of the respective population were also included in each test as checks, based on the maturity spread of the RILs. These yield tests were named Test 1 and Test 2 for RILs derived from POP201 and test 3 and 4 for RILs derived from POP202. Yield check cultivars NC-Dunphy (North Carolina State University), Osage (PI 648270) [32], and Roy [33] were used in tests 1 and 2, while NC-Dunphy, NC-Dilday (North Carolina State University), and NC-Raleigh [34] were used for tests 3 and 4 to match the maturity groups of varieties in each test. The parents for tests 1 and 2 were cultivars LMN09-119 and N09-09, and the parents for tests 3 and 4 were LMN09-19 and N13-47.

2.2. Experimental Design

The four tests were grown in two locations in 2019—the Tidewater Research Station in Plymouth, NC (PLY) and the Caswell Research Farm in Kinston, NC (CAS). The same data as previously mentioned were collected for each test in this season.

Based on the data collected from the 2019 season, further selections were conducted to identify fewer lines from the four tests for further yield testing. This was performed by identifying the RILs with a yield at or above the average yield of the checks in each test. Further selection was conducted using the seed composition traits by identifying the thirty RILs with the highest protein + oil content among the RILs which had passed the yield selection threshold.

These thirty lines were then grouped into two new tests of 15 RILs, each based on maturity date. These two new tests were named Test 1 and Test 2. Yield check cultivars were again assigned to each test to match the maturity dates of the RILs in each test. Cultivars Dunphy, Dilday, and NC-Raleigh were used as checks in Test 1 and Dunphy, Ellis, N10-697, and Osage were used as checks in Test 2.

These two tests were grown in both the 2020 and 2021 seasons. These tests were grown in CLA and CAS in 2020 and CAS and PLY in 2021. RILs were grown in a randomized complete block design with four replications in each location. The same phenotypes were evaluated for each line each season using the same methodology that was employed in the 2019 season.

The CAS location experienced severe flooding in 2020 and was subsequently removed from the analysis as the majority of the plots were severely impacted or lost by the flooding conditions.

2.3. Statistical Analysis

Phenotypic traits were analyzed with a mixed-effects model with the form:

$$y_{ijk} = \mu + E_i + B(E_i) + G_k + GE_{ik} + \epsilon_{ijk}$$

where y_{ijk} is the phenotypic measurement for rep j of genotype k in environment i; E_i is the effect of environment i; $B(E_i)$ is the effect of replication nested within environment; G_k is the effect of genotype k; GE_{ik} is the interaction effect of environment i and genotype k; and

ϵ_{ijk} is the measurement error. The genotype effect was treated as fixed and all other factors were treated as random.

Models were fit using the gamem_met function from the metan package [35] in R. Least square means (LS Means) for each genotype and trait were calculated using the above model using the emmeans package [36] in R. The emmeans package was also used to calculate contrasts as a post hoc test to compare RIL phenotype means to check means for all collected phenotypes. A sidak adjustment was used to account for multiple comparisons in the calculation of contrasts. The LS Means, their associated standard errors, and the contrasts of these means with the check means were used to determine if the mean of an RIL was significantly different from the mean of the checks, both qualitatively by comparing the overlaps of the standard errors of the means, and quantitatively by inspecting the contrasts between the RIL means and the check means.

Pearson correlation coefficients between each phenotype were calculated with the metan package as well.

The Pearson correlation is calculated for each pair of traits as:

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}$$

where x and y are measurements of the two phenotypes, m_x and m_y are the means of each phenotype, and r is the correlation coefficient.

3. Results and Discussion

3.1. Phenotypic Correlations

A strong negative correlation ($r = -0.92$ and $r = -0.91$) was observed in both Tests 1 and 2, respectively, between seed protein and seed oil content. A weak negative correlation ($r = -0.25$ and $r = -0.32$) was observed between seed protein and yield in Tests 1 and 2, respectively, which was statistically insignificant in both populations. Weak positive correlations were also observed between seed oil and seed yield in both populations. A correlation coefficient of r = 0.20 was observed in Test 1, while a coefficient of r = 0.27 was observed in Test 2. These correlation coefficients were not statistically significant in either population. The pairwise correlation coefficients for Tests 1 and 2 are presented in Supplementary Figure S3 and Figure S4, respectively. These correlations were expected and match well with the previously reported correlation behavior between these traits in soybean breeding populations.

Genotype least square means were calculated for all RILs in both tests. A visualization of these least square means and the standard error on the estimation of these means for seed protein, seed oil, and seed yield can be seen in Figure 1. Both tests tended to have more RILs with high protein rather than high oil when compared to the average of the checks, and a moderate number had yields greater than the average of the checks. Qualitatively, many RILs had a comparable yield to the checks following the relatively large standard error associated with estimating the yield, and a qualitatively larger seed protein content given the relatively small standard error associated with the estimation of seed protein content. To separate means quantitatively, we performed contrasts between each RIL mean and the average of the checks for each test. These contrasts are then used to identify subsets of RILs with desirable combinations of traits as measured in relation to the average values of the checks for each test. Specifically, we sought to identify subsets of RILs which had yields similar to, or above that of, the checks of their respective tests, which were also superior to those checks in terms of seed protein and/or seed oil content. This methodology allowed us to identify RILs with trait profiles that would allow growers or breeders to improve seed yield in combination with one or both of the seed composition traits.

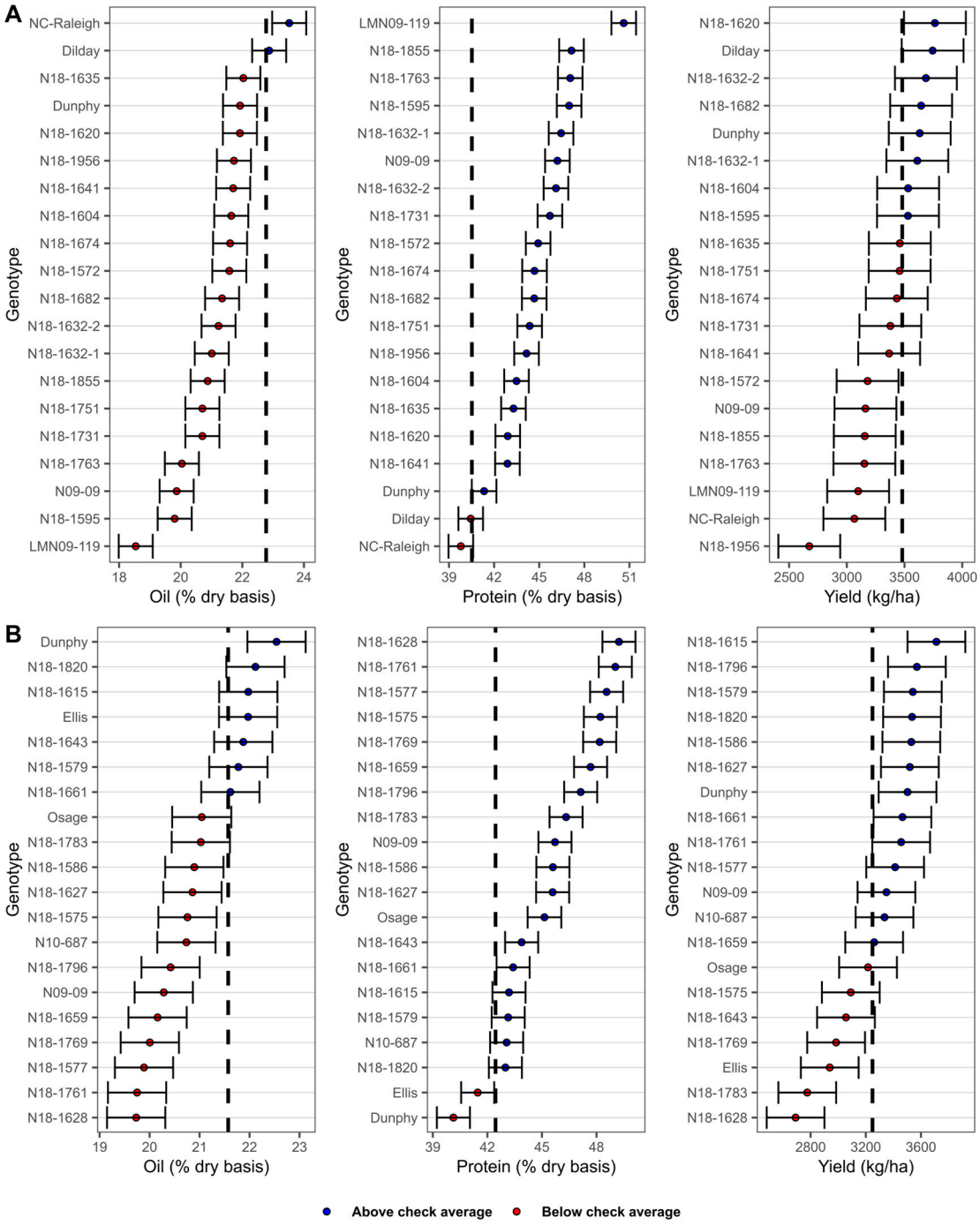

Figure 1. Genotype least square means for seed oil, seed protein, and seed yield for soybean RIL in Test 1 (**A**) and Test 2 (**B**). Points indicate that least square means and error bars indicate the standard errors in the estimation of least square means. Blue dots indicate that an RIL has a least square mean above the average of the checks in the test and a color of red indicates that the RIL has a mean value below the check average. The check average is shown with the vertical dashed line.

3.2. Yield Contrasts

Several RILs in each of the tests had yields that were comparable to that of their check cultivars as per contrasts between RIL means and the average of the check cultivars for each test. No genotypes had yields that were higher than that of the yield checks; however, many had comparable yields. Many of these genotypes with comparable yield also had protein content that was superior to the check average in each test. No genotypes had oil

that was superior to the check cultivars, but some had comparable yield and oil content as well as superior protein content. A detailed summary of all contrasts can be found in Supplementary Table S1. Soybean genotypes with superior protein content in combination with comparable or superior yield to existing high-yielding cultivars are extremely valuable for the production of soybean meal with higher protein. As mentioned previously, the meal protein of current commercial U.S. cultivars is mostly below the minimum market standard of 47.5%. The meal protein is calculated by the following formula that takes both seed protein and oil contents into account:

$$Meal\ protein = \frac{Pro13}{1 - \frac{Oil13}{100}} / 0.92$$

where *Pro*13 and *Oil*13 are seed protein and oil contents on a 13% moisture basis, respectively [37,38].

The meal protein contents of these lines with high protein and normal oil contents ranged from 104.9 to 108.7% compared to the meal protein contents of the checks. This corresponds to protein meal concentrations between 50.1 and 53.6%. Breeding line N18-1783 from Test 2 had the maximum protein meal from among these lines (Supplementary Table S2).

3.3. Genotypes with Comparable Yield and Superior Protein

Twenty-two RILs had a similar yield and superior protein content to the check cultivars. Summary data for the seed protein and seed yield of these lines are given in Table 1. From among these RILs, N18-1620 had the highest seed yield relative to the checks included in the test with an average yield that was 108.1% that of the check average of Test 1. Genotype LMN09-119 had the highest seed protein content on average with a protein content that was 124.9% that of the check average of Test 1. We observed a negative correlation between seed protein and seed yield in both tests, so RILs with seed protein content that was above the check average tended to have a seed yield content that was comparatively lower. This general tendency for RILs with high seed protein content to have a comparatively lower yield matches with previously observed trends for these traits in soybean, where this correlation structure is often reported.

However, among these genotypes are several with both average seed protein and average seed yield that are greater than the averages of the checks for each test. The genotypes that meet these criteria are N18-1620, N18-1632-1, N18-1682, and N18-1595 from Test 1 and N18-1586, N18-1627, N18-1761, N09-09, and N18-1659 from Test 2. Of particular note are genotypes N18-1632-1 from Test 1 and N18-1761 from Test 2. These two RILs have a seed protein content that is substantially above that of the average protein content of the checks while also maintaining an average yield that is above the average yield of the checks. These are notable given the previously mentioned correlation structure between seed yield and seed protein in soybean where historically it has been difficult to increase both traits simultaneously. These RILs are ideal candidates for use in breeding programs which seek to improve seed protein content without compromising seed yield or agronomics. Specifically, these RILs can provide breeders with a source of germplasm that they can use to continue to improve both seed protein content and seed yield simultaneously. This is important as growers are typically compensated for yield, but a genotype which can provide elevated protein content relative to existing cultivars with the same yield will be more valuable.

Table 1. Soybean RILs with superior protein content and comparable yield performance to high-yielding check cultivars.

		Yield [e]		Protein [f]		Meal Protein [g]		Flower Color [h]	Pubescence [i]	Maturity Group [j]
Genotype [a]	**Test [b] Name [b]**	**Value [c]**	**Rank [d]**	**Value**	**Rank**	**Value**	**Rank**			
N18-1620	Test 1	3763.47 (108.1%)	1	42.91 (105.9%)	16	50.13 (104.9%)	16	P	T	Early VI
N18-1632-2		3685.3 (105.9%)	3	46.11 (113.8%)	7	53.48 (111.9%)	6	W	T	
N18-1682		3643.81 (104.7%)	4	44.67 (110.2%)	11	51.87 (108.6%)	11	W	T	
N18-1632-1		3610.87 (103.8%)	6	46.45 (114.6%)	5	53.75 (112.5%)	4	W	T	
N18-1595		3529.96 (101.4%)	8	46.99 (116%)	4	53.68 (112.3%)	5	P	T	
N18-1635		3459.3 (99.4%)	9	43.29 (106.8%)	15	50.63 (106%)	15	W	T	
N18-1751		3457.84 (99.4%)	10	44.36 (109.5%)	12	51.16 (107.1%)	13	P	T	
N18-1674		3432.93 (98.6%)	11	44.69 (110.3%)	10	52.04 (108.9%)	10	W	T	
N18-1731		3377.4 (97%)	12	45.71 (112.8%)	8	52.72 (110.3%)	8	W	T	
N18-1641		3366.77 (96.7%)	13	42.89 (105.8%)	17	49.99 (104.6%)	17	P	T	
N18-1572		3179.97 (91.4%)	14	44.93 (110.9%)	9	52.3 (109.5%)	9	P	T	
N09-09		3161.61 (90.8%)	15	46.21 (114%)	6	52.82 (110.5%)	7	P	G	
N18-1855		3155.4 (90.7%)	16	47.14 (116.3%)	2	54.47 (114%)	2	P	T	
N18-1763		3152.16 (90.6%)	17	47.06 (116.1%)	3	53.89 (112.8%)	3	P	T	
LMN09-119		3098.26 (89%)	18	50.62 (124.9%)	1	57.07 (119.4%)	1	P	T	
N18-1586	Test 2	3530.02 (108.7%)	5	45.6 (107.6%)	10	52.7 (106.8%)	9	P	T	
N18-1627		3519.42 (108.3%)	6	45.59 (107.6%)	11	52.65 (106.7%)	10	P	T	
N18-1761		3456.31 (106.4%)	9	49.03 (115.7%)	2	55.98 (113.5%)	2	P	T	
N09-09		3349.65 (103.1%)	11	45.72 (107.9%)	9	52.48 (106.4%)	11	P	G	
N18-1659		3260.67 (100.4%)	13	47.67 (112.5%)	6	54.66 (110.8%)	6	P	T	
N18-1575		3091.25 (95.2%)	15	48.21 (113.8%)	4	55.63 (112.7%)	3	W	T	
N18-1769		2984.8 (91.9%)	17	48.17 (113.7%)	5	55.15 (111.8%)	5	W	T	
N18-1783		2775.59 (85.4%)	19	46.33 (109.3%)	8	53.61 (108.7%)	8	P	T	

[a] The genotype name. [b] The test name. [c] The genotype marginal mean for the phenotype. The percentage in the parentheses is calculated by dividing the value of each trait for each genotype by the averages of the checks in each test. [d] The ranking of this genotype for the phenotype within its test. [e] Yield in kg/ha. [f] Protein content on a percent dry basis. [g] Meal protein content on a 13-percent moisture basis. [h] Flower color: P = Purple, W = White. [i] Pubescence color: G = Gray, T = Tawny. [j] Maturity group.

3.4. Genotypes with Comparable Seed Yield and Seed Oil and Superior Seed Protein Content

Four genotypes had yield and oil that were similar to that of the yield checks, and seed protein that was greater than the average of the checks. These genotypes are N18-1620 and N18-1635 from Test 1 and genotypes N18-1586 and N18-1627 from Test 2. Summary data for these lines are given in Table 2.

Table 2. Soybean genotypes with yield and seed oil comparable to check cultivars, and seed protein superior to check cultivars.

		Yield [e]		Protein [f]		Oil [g]		Meal Protein [h]		Flower Color [i]	Pubescence [j]	Maturity Group [k]
Genotype [a]	Test [b]	Value [c]	Rank [d]	Value	Rank	Value	Rank	Value	Rank			
N18-1620	Test 1	3763.47 (108.1%)	1	42.91 (105.9%)	16	21.92 (96.3%)	5	50.13 (104.9%)	16	P	T	Early VI
N18-1635		3459.3 (99.4%)	9	43.29 (106.8%)	15	22.03 (96.8%)	3	50.63 (106%)	15	W	T	
N18-1586	Test 2	3530.02 (108.7%)	5	45.6 (107.6%)	10	20.9 (96.7%)	10	52.7 (106.8%)	9	W	T	
N18-1627		3519.42 (108.3%)	6	45.59 (107.6%)	11	20.86 (96.5%)	11	52.65 (106.7%)	10	P	T	

[a] The genotype name. [b] The test name. [c] The genotype marginal mean for the phenotype. The percentage in the parentheses is calculated by dividing the value of each trait for each genotype by the averages of the checks in each test. [d] The ranking of this genotype for the phenotype within its test. [e] Yield in kg/ha. [f] Protein content on a percent dry basis. [g] Oil content on a percent dry basis. [h] Meal protein content on a 13-percent moisture basis. [i] Flower color: P = Purple, W = White. [j] Pubescence color: G = Gray, T = Tawny. [k] Maturity group.

Traditionally, there tends to be a strong negative correlation between seed oil and seed protein content in soybean populations, where genotypes with a high seed oil content will tend to have a correspondingly lower seed protein content and vice versa. A similar inverse correlation is traditionally seen between seed protein and seed yield, where high-yielding genotypes will tend to have lower seed protein contents. These general trends were observed in both populations of this study as well.

The lines that we have identified in Table 2 are exceptional in that they had seed yield and seed oil content that were not significantly different from that of high-yielding check cultivars, but had seed protein content that was significantly greater than that of the high-yielding check cultivars. As such, these RILs can provide valuable germplasm that can be used to improve seed protein content and maintain seed oil content without compromising seed yield. This is important for soybean breeding as the value of soybean is driven through its yield, and then products derived from its seed protein and oil content. This can be of use both to breeders who are looking to improve seed composition traits in soybean without compromising seed yield and for growers who are seeking options for novel soybean varieties with seed composition that is superior to that of elite cultivars.

4. Conclusions

We have identified several genotypes with seed yield comparable to existing elite cultivars, that have seed protein or oil content that are superior to the check cultivars. These genotypes have good agronomic qualities and will provide both breeders and growers with new options for genotypes that can be used in production or used in breeding programs that seek to improve valuable soybean seed composition traits without significantly reducing seed yield.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/crops3040029/s1, Figure S1: Schematic representation of the development of Mapping Pop 201; Figure S2: Schematic representation of the development of Mapping Pop 202; Figure S3: The pairwise correlation coefficients for phenotypic observations for Test 1; Figure S4: The pairwise correlation coefficients for phenotypic observations for Test 1; Table S1: RIL-Yield check contrasts for all genotypes and phenotype combinations with relevant summary statistics.; Table S2: Protein meal values of soybean genotypes with yield and seed oil comparable to check cultivars, and seed protein superior to check cultivars.

Author Contributions: Conceptualization, R.M. and J.G.; methodology, R.M. and J.G.; software, J.G.; validation, R.M., B.M. and M.C.; formal analysis, J.G.; investigation, J.G.; resources, R.M.; data curation, J.G.; writing—original draft preparation, J.G.; writing—review and editing, R.M., M.C., B.M. and E.T.; visualization, J.G.; supervision, R.M.; project administration, R.M.; funding acquisition, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by United Soybean Board grant number 2332-203-0102.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Burton, J. Quantitative Genetics: Results Relevant to Soybean Breeding. In *Soybeans: Improvement, Production, and Uses*, 2nd ed.; Wilcox, J.R., Ed.; ASA: Madison, WI, USA; CSSA: Madison, WI, USA; SSSA: Madison, WI, USA, 1987; pp. 211–247.
2. Chaudhary, J.; Patil, G.B.; Sonah, H.; Deshmukh, R.K.; Vuong, T.D.; Valliyodan, B.; Nguyen, H.T. Expanding Omics Resources for Improvement of Soybean Seed Composition Traits. *Front. Plant Sci.* **2015**, *6*, 1021. [CrossRef] [PubMed]
3. Hartwig, E.E.; Hinson, K. Association between Chemical Composition of Seed and Seed Yield of Soybeans 1. *Crop Sci.* **1972**, *12*, 829–830. [CrossRef]
4. Singh, B.; Singh, A.K. *Marker-Assisted Plant Breeding: Principles and Practices*; Springer: New Delhi, India, 2015.
5. Brummer, E.C.; Graef, G.L.; Orf, J.; Wilcox, J.R.; Shoemaker, R.C. Mapping QTL for Seed Protein and Oil Content in Eight Soybean Populations. *Crop Sci.* **1997**, *37*, 370–378. [CrossRef]
6. Lande, R.; Thompson, R. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genetics* **1990**, *124*, 743–756. [CrossRef]
7. Rossi, M.E.; Orf, J.H.; Liu, L.J.; Dong, Z.M.; Rajcan, I. Genetic Basis of Soybean Adaptation to North American vs. Asian Mega-Environments in Two Independent Populations from Canadian x Chinese Crosses. *Theor. Appl. Genet.* **2013**, *126*, 1809–1823. [CrossRef]
8. Wang, X.Z.; Jiang, G.L.; Green, M.; Scott, R.A.; Song, Q.J.; Hyten, D.L.; Cregan, P.B. Identification and Validation of Quantitative Trait Loci for Seed Yield, Oil and Protein Contents in Two Recombinant Inbred Line Populations of Soybean. *Mol. Genet. Genom.* **2014**, *289*, 935–949. [CrossRef]
9. Pathan, S.M.; Vuong, T.; Clark, K.; Lee, J.-D.; Shannon, J.G.; Roberts, C.A.; Ellersieck, M.R.; Burton, J.W.; Cregan, P.B.; Hyten, D.L.; et al. Genetic Mapping and Confirmation of Quantitative Trait Loci for Seed Protein and Oil Contents and Seed Weight in Soybean. *Crop Sci.* **2013**, *53*, 765. [CrossRef]
10. Li, D.M.; Sun, M.M.; Han, Y.P.; Teng, W.L.; Li, W.B. Identification of QTL Underlying Soluble Pigment Content in Soybean Stems Related to Resistance to Soybean White Mold (*Sclerotinia sclerotiorum*). *Euphytica* **2010**, *172*, 49–57. [CrossRef]
11. Palomeque, L.; Liu, L.J.; Li, W.B.; Hedges, B.; Cober, E.R.; Rajcan, I. QTL in Mega-Environments: II. Agronomic Trait QTL Co-Localized with Seed Yield QTL Detected in a Population Derived from a Cross of High-Yielding Adapted x High-Yielding Exotic Soybean Lines. *Theor. Appl. Genet.* **2009**, *119*, 429–436. [CrossRef]
12. Gai, J.; Wang, Y.; Wu, X.; Chen, S. A Comparative Study on Segregation Analysis and QTL Mapping of Quantitative Traits in Plants—With a Case in Soybean. *Front. Agric. China* **2007**, *1*, 1–7. [CrossRef]
13. Chen, Q.; Zhang, Z.; Liu, C.; Xin, D.; Qiu, H.; Shan, D.; Shan, C.; Hu, G. QTL Analysis of Major Agronomic Traits in Soybean. *Agric. Sci. China* **2007**, *6*, 399–405. [CrossRef]
14. Zhang, Y.H.; Liu, M.F.; He, J.B.; Wang, Y.F.; Xing, G.N.; Li, Y.; Yang, S.P.; Zhao, T.J.; Gai, J.Y. Marker-Assisted Breeding for Transgressive Seed Protein Content in Soybean *Glycine Max* (L.) Merr. *Theor. Appl. Genet.* **2015**, *128*, 1061–1072. [CrossRef] [PubMed]
15. Wang, J.; Chen, P.; Wang, D.; Shannon, G.; Zeng, A.; Orazaly, M.; Wu, C. Identification and Mapping of Stable QTL for Protein Content in Soybean Seeds. *Mol. Breed.* **2015**, *35*, 10. [CrossRef]
16. Cunicelli, M.J.; Bhandari, H.S.; Chen, P.; Sams, C.E.; Mian, M.R.; Mozzoni, L.A.; Smallwood, C.J.; Pantalone, V.R. Effect of a Mutant Danbaekkong Allele on Soybean Seed Yield, Protein, and Oil Concentration. *J. Am. Oil Chem. Soc.* **2019**, *96*, 927–935. [CrossRef]
17. Patil, G.; Mian, R.; Vuong, T.; Pantalone, V.; Song, Q.; Chen, P.; Shannon, G.J.; Carter, T.C.; Nguyen, H.T. Molecular Mapping and Genomics of Soybean Seed Protein: A Review and Perspective for the Future. *Theor. Appl. Genet.* **2017**, *130*, 1975–1991. [CrossRef] [PubMed]
18. Gupta, M.; Bhaskar, P.B.; Sriram, S.; Wang, P.H. Integration of Omics Approaches to Understand Oil/Protein Content during Seed Development in Oilseed Crops. *Plant Cell Rep.* **2017**, *36*, 637–652. [CrossRef] [PubMed]
19. Guo, B.; Sun, L.; Jiang, S.; Ren, H.; Sun, R.; Wei, Z.; Hong, H.; Luan, X.; Wang, J.; Wang, X.; et al. Soybean Genetic Resources Contributing to Sustainable Protein Production. *Theor. Appl. Genet.* **2022**, *135*, 4095–4121. [CrossRef] [PubMed]
20. Chen, P.; Ishibashi, T.; Dombek, D.; Rupe, J. Registration of R05-1415 and R05-1772 High-Protein Soybean Germplasm Lines. *J. Plant Regist.* **2011**, *5*, 410–413. [CrossRef]
21. Panthee, D.; Pantalone, V. Registration of Soybean Germplasm Lines TN03-350 and TN04-5321 with Improved Protein Concentration and Quality. *Crop Sci.* **2006**, *46*, 2328. [CrossRef]
22. Mian, M.R.; McHale, L.; Li, Z.; Dorrance, A.E. Registration of 'Highpro1' Soybean with High Protein and High Yield Developed from a North× South Cross. *J. Plant Regist.* **2017**, *11*, 51–54. [CrossRef]

23. Carter Jr, T.; Rzewnicki, P.; Burton, J.; Villagarcia, M.; Bowman, D.; Taliercio, E.; Kwanyuen, P. Registration of N6202 Soybean Germplasm with High Protein, Favorable Yield Potential, Large Seed, and Diverse Pedigree. *J. Plant Regist.* **2010**, *4*, 73–79. [CrossRef]
24. Boerma, H.; Hussey, R.; Phillips, D.; Wood, E.; Rowan, G.; Finnerty, S. Registration of 'Benning' soybean. *Crop Sci.* **1997**, *37*, 1982. [CrossRef]
25. Kim, S.; Hong, E.; Kim, Y.; Lee, S.; Park, K.; Kim, H.; Ryu, Y.; Park, R.; Kim, Y.; Seong, Y.; et al. A New High Protein and Good Seed Quality Soybean Variety "Danbaegkong". *RDA J. Agric. Sci. Upl. Ind. Crops* **1996**, *38*, 228–232.
26. Shipe, E.; Mueller, J.; Lewis, S.; Williams, P.; Stephens, R. Registration of 'Santee' Soybean. *Crop Sci.* **2003**, *43*, 2305–2306. [CrossRef]
27. Burton, J.W.; Carter, T.; Huie, E. Registration of 'Holladay' Soybean. *Crop Sci.* **1996**, *36*, 467. [CrossRef]
28. Boerma, H.; Hussey, R.; Phillips, D.; Wood, E.; Rowan, G.; Finnerty, S. Registration of 'Boggs' Soybean. *Crop Sci.* **2000**, *40*, 294–295. [CrossRef]
29. Boerma, H.; Hussey, R.; Phillips, D.; Wood, E.; Finnerty, S. Registration of 'Doles' Soybean. *Crop Sci.* **1994**, *34*, 1411–1412. [CrossRef]
30. Fehr, I. Ch. 7. In *Soybeans: Improvement, Production and Uses*, 2nd ed.; Wilcox, J.R., Ed.; American Society of Agronomy: Madison, WI, USA, 1987; pp. 249–293.
31. Perten Instruments AB. DA 7250 Installation and Operation Manual. *Perten Instrum. AB Hagersten Swed.* **2016**, *4*, Y2.
32. Chen, P.; Sneller, C.; Mozzoni, L.; Rupe, J. Registration of 'Osage' Soybean. *J. Plant Regist.* **2007**, *1*, 89–92. [CrossRef]
33. Burton, J.; Carter, T. Registration of 'NC-Roy' Soybean. *Crop Sci.* **2005**, *45*, 2654. [CrossRef]
34. Burton, J.; Carter Jr, T.; Fountain, M.; Bowman, D. Registration of'NC-Raleigh' soybean. *Crop Sci.* **2006**, *46*, 2710. [CrossRef]
35. Olivoto, T.; Lucio, A. Metan: An R Package for Multi-Environment Trial Analysis. *Methods Ecol. Evol.* **2020**, *11*, 783–789. [CrossRef]
36. Lenth, R. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. R Package Version 1.4. 3.01. 2019. Available online: https://cran.r-project.org/web/packages/emmeans/index.html (accessed on 4 December 2023).
37. Pantalone, V.; Smallwood, C. Registration of "TN11-5102" Soybean Cultivar with High Yield and High Protein Meal. *J. Plant Regist.* **2018**, *12*, 304–308. [CrossRef]
38. Smallwood, C.; Fallen, B.; Pantalone, V. Registration of 'TN11-5140' Soybean Cultivar. *J. Plant Regist.* **2018**, *12*, 203–207. [CrossRef]