

## Article

# Determining Genetic Markers and Seed Compositions Related to High Test Weight in *Glycine max*

Zachary Shea <sup>1</sup>, William M. Singer <sup>1</sup>, Luciana Rosso <sup>1</sup>, Qijian Song <sup>2</sup> and Bo Zhang <sup>1,\*</sup>

<sup>1</sup> School of Plant & Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, USA; zachary.shea@usda.gov (Z.S.); wilmsing@vt.edu (W.M.S.); luciana@vt.edu (L.R.)

<sup>2</sup> USDA-ARS, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA; qijian.song@usda.gov

\* Correspondence: bozhang@vt.edu

**Abstract:** Test weight, one of the primary indicators of soybean seed quality, is measured as the amount of soybean seeds in kilograms that can fit into one hectoliter. The price that growers receive for their soybean is dependent on test weight. Over the past 50 years, growers have observed a decreasing trend in test weight. Therefore, it is imperative to understand better the relationship between soybean test weight and other traits to enable breeders to select parental lines with high test weights in breeding programs to ensure the grower's profitability. The objectives of the study were to identify genetic markers associated with high test weight in soybean and to determine the correlation between high test weight and five important seed composition traits (protein, oil, sucrose, raffinose, and stachyose content). Maturity group IV and V germplasms from the USDA soybean germplasm collection were grown in Blacksburg and Warsaw in Virginia from 2019 to 2021 and were measured for all of the above traits. Results show that test weight values ranged from 62–77 kg/hL over the three years. Multiple single-nucleotide polymorphisms (SNPs) significantly associated with high test weight were found on chromosome (Chr.) 15 along with a couple on chromosome 14, and 11 candidate genes were found near these SNPs. Test weight was found to be significantly negatively correlated with oil content, inconsistently correlated with protein content in all environments, and negatively correlated but not significantly with all three sugars except for raffinose in Blacksburg 2019. We concluded that the genes that underlie test weight might be on chromosome 15, and the validated associated SNPs might be used to assist breeding selection of test weight. Breeders should pay special attention to test weight while selecting for high oil content in soybean due to their negative correlation.

**Keywords:** soybean; GWAS; genetics; test weight; protein; oil



**Citation:** Shea, Z.; Singer, W.M.; Rosso, L.; Song, Q.; Zhang, B. Determining Genetic Markers and Seed Compositions Related to High Test Weight in *Glycine max*. *Plants* **2023**, *12*, 2997. <https://doi.org/10.3390/plants12162997>

Academic Editor: Zhaoshi Xu

Received: 6 July 2023

Revised: 8 August 2023

Accepted: 14 August 2023

Published: 19 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Overview of Test Weight in Soybean

Soybean (*Glycine max*) is the most important oilseed crop due to its low-cost production and its diverse uses in feed and food for its excellent nutrient profile and in paints and biofuels for its high oil content [1]. The United States has been one of the world's top two soybean producers for decades [2]. Seed quality is one of the primary factors that affect the price that growers can receive for their soybean. Multiple aspects can affect seed quality, including test weight, the amount of diseased seeds, and damaged or disfigured seeds [3]. Of these, test weight is the most important indicator of soybean seed quality because soybeans with higher test weight last longer in storage, having higher seed integrity and being less prone to mold [3]. Additionally, every crop has a standard value for test weight, and if the crop's test weight is below the standard, the grower can receive a pricing penalty [4]. The standard for soybean is at 60 lb/bu, and values below 54 lb/bu can cause a pricing penalty [4]. This is important because farmers have noticed a slight decreasing trend in test weight for soybean since the 1950s. While in the 1950s almost all soybean

had a test weight around 58–60 lb/bu (72.5–75 kg/hL), farmers are now seeing more test weight values that are closer to 56 lb/bu (70 kg/hL) [3]. Even though this is still above the 54 lb/bu (67.5 kg/hL) limit, if this trend continues, farmers will start to see soybean test weight values at or below 54 lb/bu and will be more likely to lose profit.

### 1.2. Breeding Efforts to Increase Test Weight

Although test weight is important in soybean, not much work has been conducted looking into the effect of genotypes on test weight in this crop. Some work has found that genotypes showed a significant effect on test weight in maize [5], wheat [6], and oat [7]. Recently, in soybean, genotypes were found to have a significant effect on test weight in most locations but not all depending on the environment [8,9]. These findings indicate that it is possible to improve test weight through breeding. Despite this, little to no work has been conducted regarding how to assist breeding high-test-weight soybean varieties. In order to address the concern of a decreasing trend of test weight in soybean over time, there is an urgent need to identify, validate, and utilize genetic markers associated with high test weight in soybean breeding as a quick and effective approach. Thus, breeders will shorten the period of development and selection of soybean varieties for increased test weight to ensure farmers' profitability.

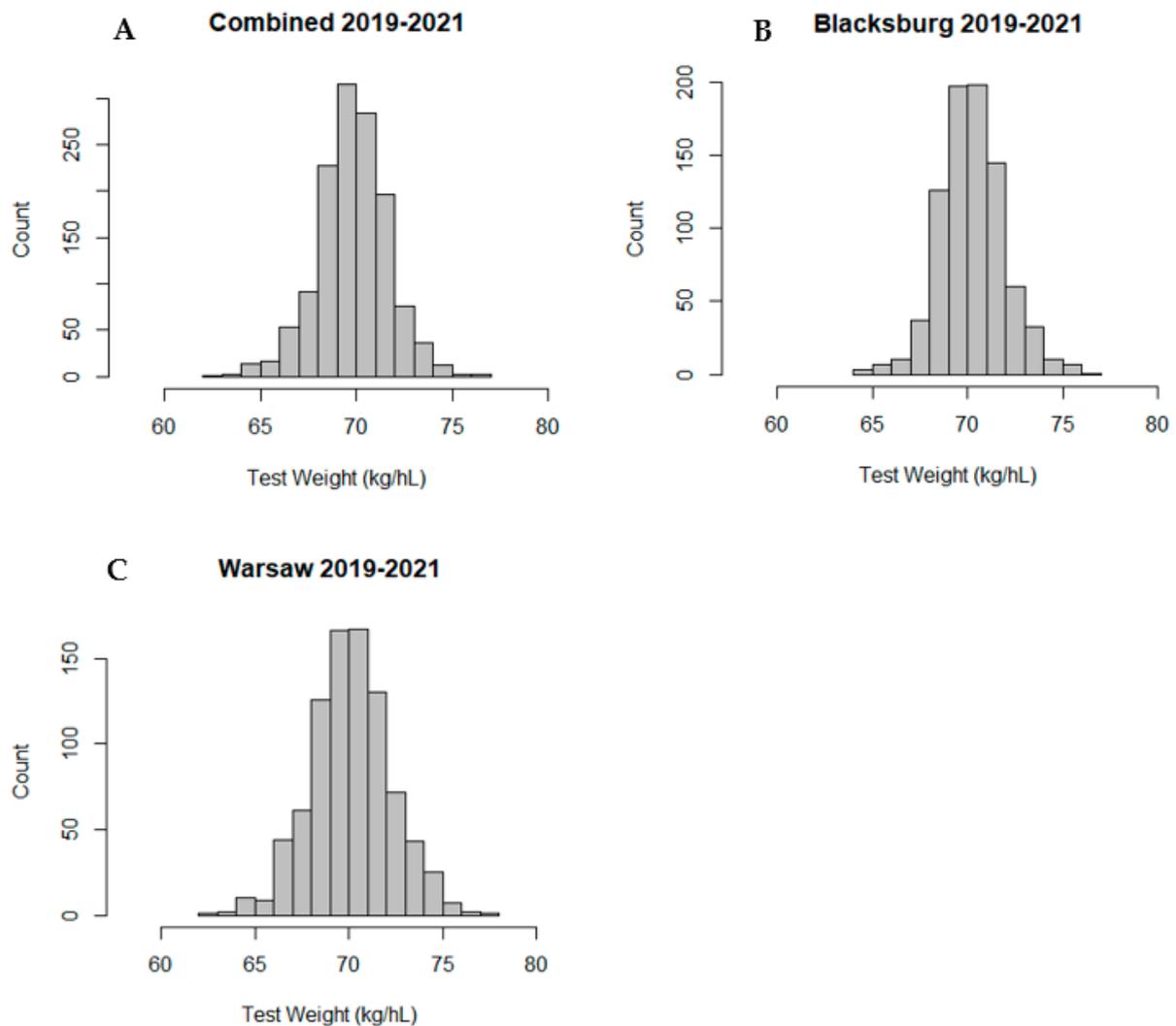
### 1.3. Correlation with Other Traits

While test weight is important in most crops, other traits are also important in plant breeding and must be taken into consideration, such as protein, oil, and sugar content in wheat and soybean. Therefore, it is necessary to gain a better understanding of the relationship between test weight and seed compositions. Although some work has been conducted previously determining this relationship, reports have indicated varying degrees of association. In wheat, some studies found that protein content had a significant, positive relationship with test weight [10,11], while others found a significant, negative correlation [12] or no significant relationship at all [13]. Oil content was found to have a significant, positive correlation with test weight in sunflowers [14] and no significant effect on test weight in oats [15]. For soybean, protein and sucrose content were found to have varying degrees of significant relationships with test weight, varying from positive to negative to not significant, while oil content and test weight had a significant, negative correlation in all experiments [8]. Test weight is known to be heavily impacted by the environment, but no research has been conducted in the Mid-Atlantic region, so it is important to determine the relationship between test weight and these traits in Virginia. Additionally, little to no work has been conducted to understand the relationship between test weight and raffinose family oligosaccharides (RFOs). While these two sugars are not as major as sucrose, reduced RFOs are a target when breeding new varieties for animal feed and human food.

## 2. Results

### 2.1. Test Weight

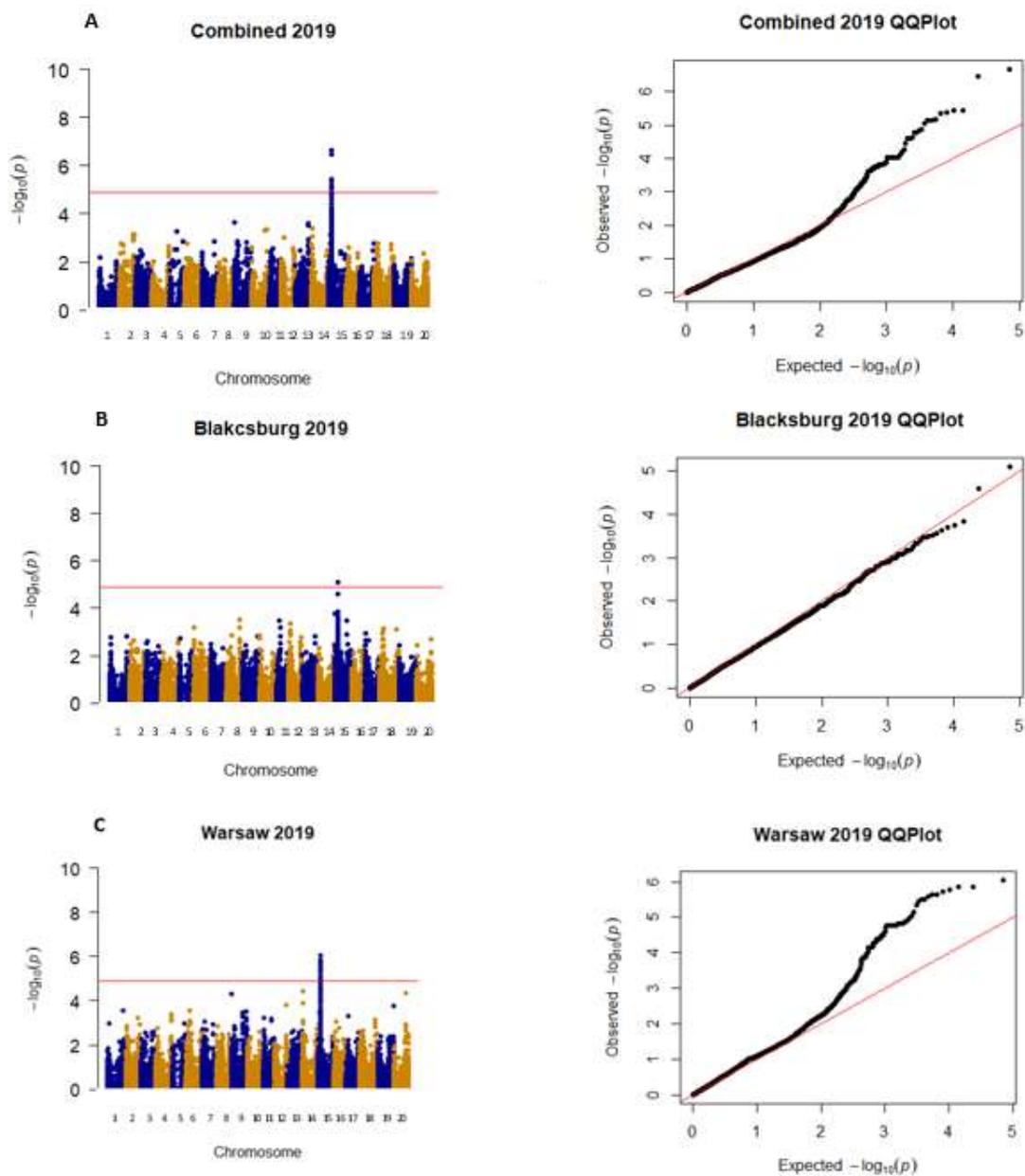
Test weight values showed a normal distribution across environments and years, ranging from 62 to 77 kg/hL with a grand mean of 70.2 kg/hL and an average standard deviation of 1.94 kg/hL. Figure 1 shows the distribution of test weight values for both locations (Figure 1A), Blacksburg (Figure 1B), and Warsaw (Figure 1C). Blacksburg ranged from 64 to 77 kg/hL with an average of 70.3 kg/hL and with a standard deviation of 1.72 kg/hL for all years, while Warsaw ranged from 62 to 77 kg/hL and had an average of 70.1 kg/hL with a standard deviation of 2.14 kg/hL. Both locations had similar averages, but apparently, Warsaw had a wider spread. Throughout all years and in both locations, there was one accession, PI87059, that consistently had high test weight.



**Figure 1.** Frequencies of test weight values across both locations (A), Blacksburg (B), and Warsaw (C) for 2019–2021.

## 2.2. Genome-Wide Association Study

All significant SNPs that were focused on for this project were on either Chr. 14 or 15, with five SNPs being significant in more than one environment, three SNPs being significant in one environment, and the other three being just under the significance threshold ( $\alpha = 4.91$ ) (Table 1). These two chromosomes were focused on because they were the only ones to have SNPs that were found to be significant or just below the significance threshold across environments. There are three exceptions to this: SNP *ss71562017* was included because it had the highest LOD score of  $-\log_{10}(p)$  in W 2020, and no other SNP from W 2020 was found to be significant. SNPs *ss715623162* and *ss715619843* were included because they were the only significant SNPs in BB 2020 on chromosomes 14 and 15. Lastly, SNP *ss715618025* was included because it was the only SNP on chromosome 14 or 15 that was significant in BB 2021. In 2019, all significant SNPs were found on chromosome 15, and significant SNPs were found on this chromosome in all locations (Figure 2). In 2020 and 2021, there were many additional SNPs on other chromosomes that were found to be significant, but they were not found in any other environments and therefore were not included for further analysis (Figures 3 and 4). Warsaw did not have any SNPs above  $-\log_{10}(p\text{-value})$  of 4.9 in 2020 and 2021 but had multiple SNPs that were close to the threshold (Figures 3 and 4). QQ plots show that the data were normally distributed (Figures 2–4).



**Figure 2.** Manhattan and QQ plots for both locations (A), Blacksburg (B), and Warsaw (C) for 2019. Chromosomes in the Manhattan plots are shown in alternating colors, each dot represents one SNP, and the significance threshold is represented by the red line at  $-\log_{10}(p)$  of 4.9. QQ plots show observed  $-\log_{10}(p)$  plotted against expected  $-\log_{10}(p)$ .

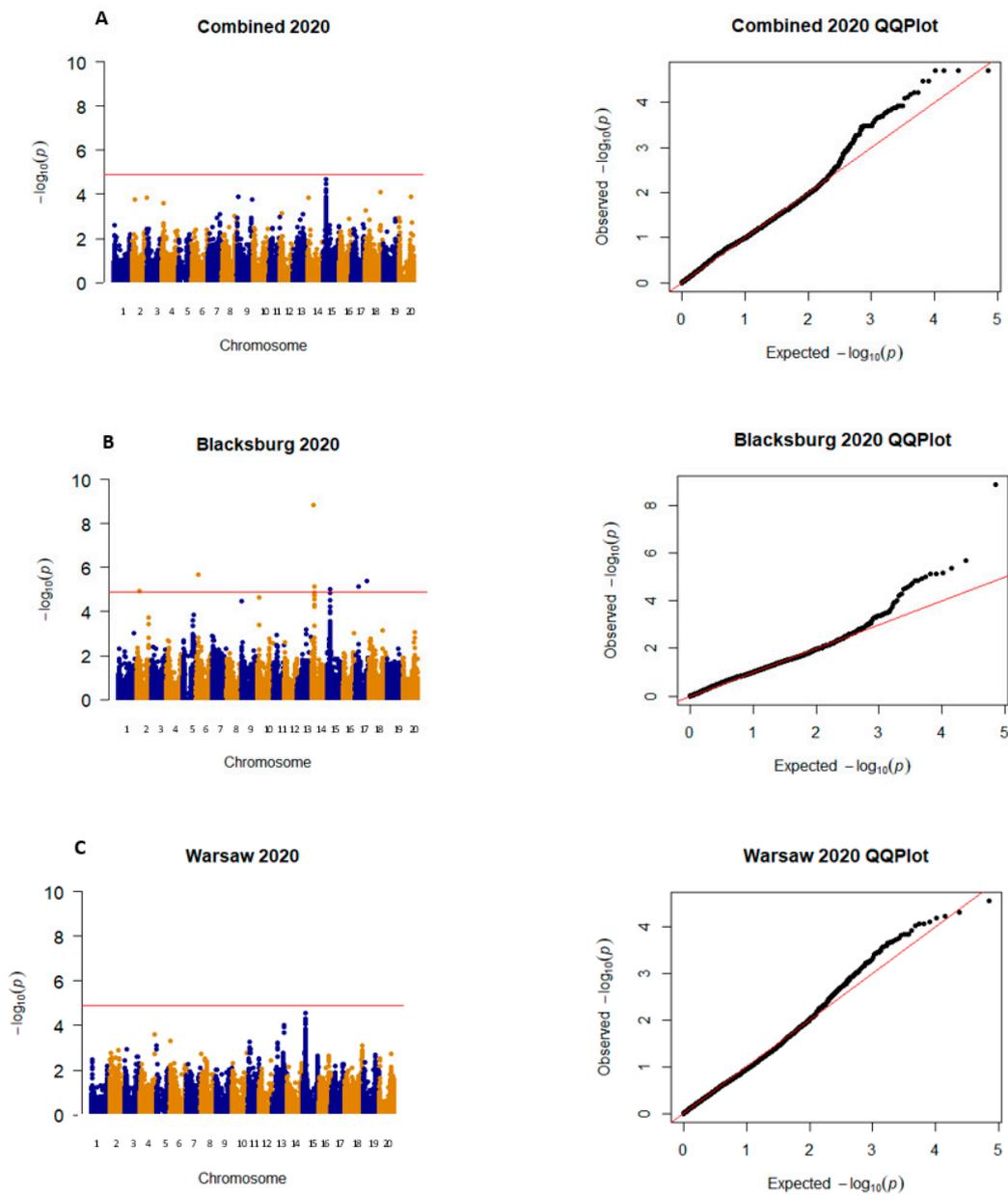
**Table 1.** Significant SNPs on chromosomes 14 and 15 associated with test weight in soybean.

Marker	Chr.	Position	Environment ( $-\log_{10}(p)$ )								
			BB 2019	BB 2020	BB 2021	W 2019	W 2020	W 2021	Combined 2019	Combined 2020	Combined 2021
ss715618482	14	3559612	NS <sup>b</sup>	NS	NS	4.41 <sup>a</sup>	NS	NS	NS	NS	NS
ss715619843	14	7207504	NS	8.86	NS	NS	NS	NS	NS	NS	NS
ss715618025	14	2201645	NS	NS	9.13	NS	NS	NS	NS	NS	NS
ss715623162	15	8758404	NS	5.37	NS	NS	NS	NS	NS	NS	NS
ss715623211	15	9205168	NS	NS	NS	5.51	NS	NS	5.13	NS	NS
ss715623224	15	9279044	NS	NS	NS	5.85	NS	NS	5.44	NS	NS

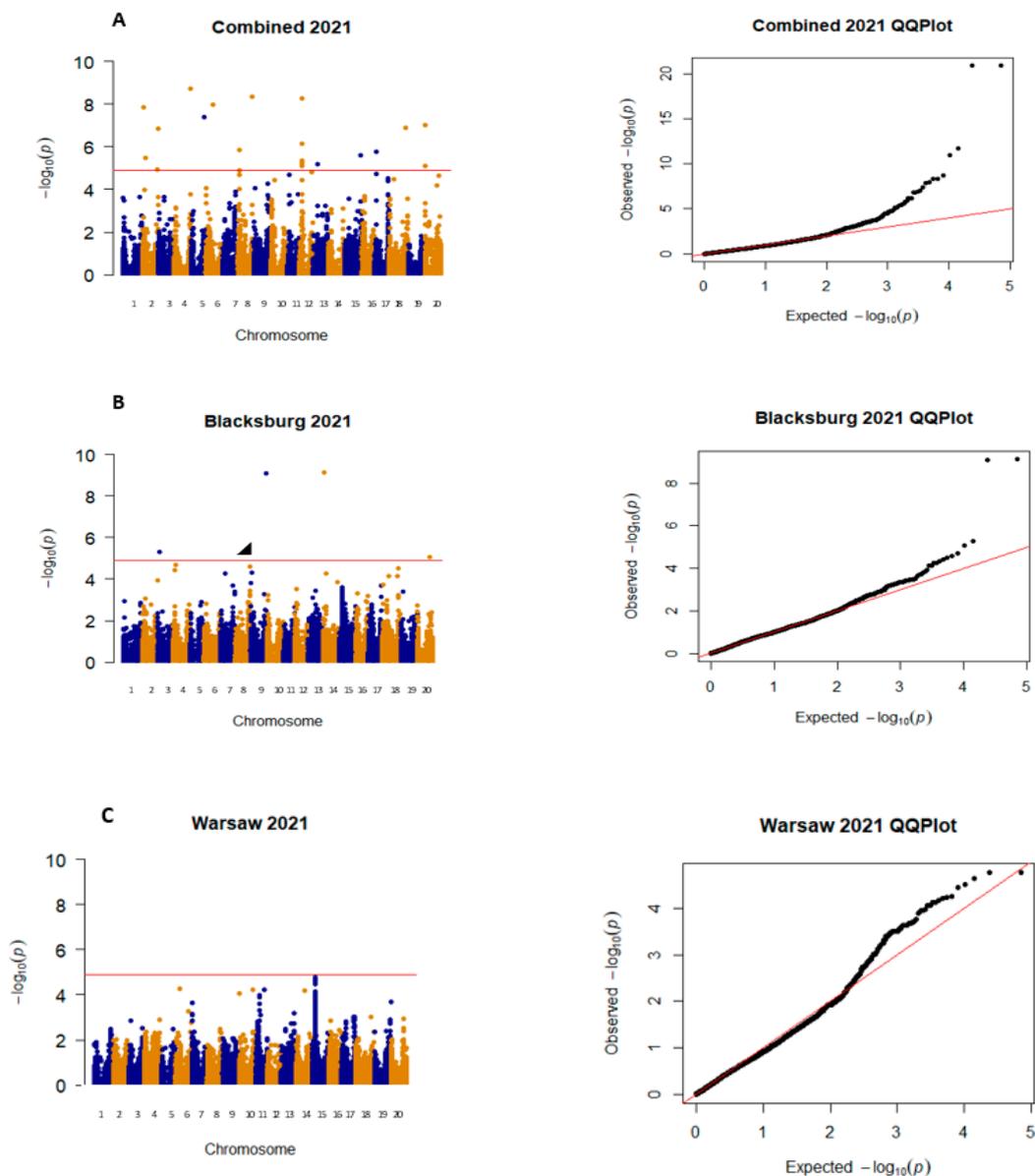
Table 1. Cont.

Marker	Chr.	Position	Environment ( $-\log_{10}(p)$ )								
			BB 2019	BB 2020	BB 2021	W 2019	W 2020	W 2021	Combined 2019	Combined 2020	Combined 2021
ss715620221	15	9383632	NS	NS	NS	5.72	NS	NS	5.33	NS	NS
ss715623250	15	9557248	NS	NS	NS	4.99	NS	NS	5.39	NS	NS
ss715623269	15	9748128	4.60 <sup>a</sup>	NS	NS	5.76	NS	NS	NS	NS	NS
ss715623270	15	9749617	5.09	NS	NS	6.04	NS	NS	6.66	NS	NS
ss715623292	15	9927090	NS	NS	NS	NS	NS	4.77 <sup>a</sup>	NS	4.70 <sup>a</sup>	NS
ss715620172	15	10176737	NS	NS	NS	NS	4.56 <sup>a</sup>	NS	NS	NS	NS

<sup>a</sup> These values are just below the significance threshold. <sup>b</sup> NS = Not significant.



**Figure 3.** Manhattan and QQ plots for both locations (A), Blacksburg (B), and Warsaw (C) for 2020. Chromosomes in the Manhattan plots are shown in alternating colors, each dot represents one SNP, and the significance threshold is represented by the red line at  $-\log_{10}(p)$  of 4.9. QQ plots show observed  $-\log_{10}(p)$  plotted against expected  $-\log_{10}(p)$ .



**Figure 4.** Manhattan and QQ plots for both locations (A), Blacksburg (B), and Warsaw (C) for 2021. Chromosomes in the Manhattan plots are shown in alternating colors, each dot represents one SNP, and the significance threshold is represented by the red line at  $-\log_{10}(p)$  of 4.9. QQ plots show observed  $-\log_{10}(p)$  plotted against expected  $-\log_{10}(p)$ .

Table 1 shows SNPs that were above the threshold or just below it in more than one location with their corresponding name, chromosome location, position, and  $-\log_{10}(p)$ -values. The SNPs ss715619843, ss715623162, and ss715620172 were included because they were the markers with the highest values for  $-\log_{10}(p)$ -value in BB 2020 (for the first two) and W 2020 (for the third), and these locations did not have any SNPs that were found to be significant in any other environment. Additionally, SNP ss715618025 was included because it was the only SNP that was significant in BB 2021 on chromosome 14 or 15.

### 2.3. Candidate Genes

A total of eight candidate genes were found on Chr. 15 and three on Chr. 14 that were located within 10 kbp of the SNPs that were found to be significant (Table 2). Only a few genes are located within 10 kbp of the significant SNPs, but no genes are located within 10 kbp of ss715623250. Most of the genes were expressed in multiple tissues

including leaf, flower, pod, and seed. Some were only expressed in specific tissues, such as Glyma.14g030400 being expressed in only flower tissue, Glyma.15g119200 being expressed in pod and seed, and Glyma.15g122800, Glyma.15g125000, and Glyma.15g127900 being expressed in root tissue. The functions of the candidate genes mostly involve RNA and/or protein binding or regulation. The gene Glyma.15g119200 has a different function that involves seed storage, and the gene Glyma.15g127900 has an unknown function.

**Table 2.** Summary of candidate genes.

Chromosome	SNP (Position)	Gene	Expression Pattern	Function
14	ss715618482 (3559612)	Glyma.14g046800	leaf, flower, pod, seed	Serine phosphatase
14	ss715619843 (7207504)	Glyma.14g082900	leaf, flower, pod	Cytochrome subfamily
14	ss715618025 (2201645)	Glyma.15g111700	leaf, flower, pod, seed,	Ribosomal protein
15	ss715623162 (8758404)	Glyma.14g030400	flower	Dioxygenase
15	ss715623211 (9205168)	Glyma.15g117100	leaf and pod	Transcriptional regulation
15	ss715623224 (9279044)	Glyma.15g118100	leaf, flower, pod, seed	Pentatricopeptide protein
15	ss715620221 (9383632)	Glyma.15g119200	pod, seed	Seed storage protein
15	ss715623269 (9748128)	Glyma.15g122800	root hair, root tip	RNA and protein binding
15	ss715623270 (9749617)	Glyma.15g122800	root hair, root tip	RNA and protein binding
15	ss715623292 (9927090)	Glyma.15g125000	root	Serves as a methyltransferase
15	ss715620172 (10176737)	Glyma.15g127900	root	Unknown

Table 2 shows the candidate genes near SNPs that were previously identified as significant or just below the threshold and their corresponding functions. Expression data were obtained from Soybase. Detailed expression data can be found in Severin et al. 2010 [16].

#### 2.4. Correlation of Test Weight with Seed Composition Traits

The Pearson's correlation between test weight and protein and oil content for 2019–2021 and between test weight and raffinose, sucrose, and stachyose content for 2019 varied depending on location and year (Table 3). The correlation between test weight and protein was not consistent due to a positive correlation in three environments and a negative correlation in another three environments. In addition, three environments showed significant correlation, namely, BB and W 2019 with negative correlations and W 2020 with a positive correlation. The correlation between test weight and oil was found to be significantly negative in all environments except for BB 2021. The strongest correlation was found in W 2021 at  $-0.387$ . The correlation between test weight and all sugars in 2019 was negative, but only raffinose had a significant correlation coefficient of  $-0.175$  in BB 2019.

**Table 3.** Correlation between test weight and five important seed composition traits.

Trait	BB 2019	W 2019	BB 2020	W 2020	BB 2021	W 2021
Protein	$-0.115^*$	$-0.136^*$	0.021	$0.390^*$	$-0.052$	0.047
Oil	$-0.174^*$	$-0.297^*$	$-0.203^*$	$-0.265^*$	$-0.024$	$-0.387^*$
Raffinose	$-0.175^*$	$-0.118$	NA	NA	NA	NA
Sucrose	$-0.086$	$-0.08$	NA	NA	NA	NA
Stachyose	$-0.003$	$-0.063$	NA	NA	NA	NA

\* These correlations were found to be significant at  $\alpha = 0.05$ .

### 3. Discussion

While test weight has not received as much attention as other traits such as protein and oil content in soybean, it is a crucial trait for farmers' profitability. This is coupled with its power to affect the pricing of soybean and the decreasing trend that farmers have observed in test weight in the past several decades [3]. In this study, we were able to identify multiple genetic markers significantly associated with test weight through GWAS,

identify potential candidate genes for test weight, and provide information regarding the relationship between test weight and important seed compositions.

Many studies have found a significant relationship between genotype and high test weight in wheat, soybean, and other crops [5–8]. However, no studies have found SNPs associated with test weight. In this study, we were able to find many significant SNPs, especially on chromosomes 14 and 15. Environments including BB 2020, BB 2021, and Combined 2021 had multiple significant SNPs that were found to be associated with test weight but were not present in any other environment. SNPs on Chr. 14 and 15 were found to be either significantly associated or close to being significantly associated with test weight across years and locations. Of these two, Chr. 15 seemed to be more consistent because it had SNPs that were significant or close to being significant in almost all environments, while Chr. 14 primarily had significant SNPs in BB 2020 and 2021, which indicates that Chr. 15 may have genes that control test weight. Furthermore, 11 of these SNPs were found to be within 10 kbp of a gene. Six of these genes had little to no expression in seed tissue, and due to this they are less likely to be related to test weight. In addition, most of these SNPs are located close to genes that encode proteins to bind and regulate RNA and other protein; ss715620221 is near Glyma.15g119200, which codes for a seed storage protein and has high expression in seed tissue. While this does not guarantee anything, ss715620221 could be a promising genetic marker because test weight is related to seed durability. Three other genes had high expression in seed tissue, so these could be promising genetic markers as well. Additionally, SNP ss715620172 is located near a gene with an unknown function, which encourages future studies to determine if Glyma.15g127900 is related to test weight.

While Chr. 15 was found to consistently have significant SNPs or SNPs close to the significance threshold, there were no SNPs that were found to be significant across all locations and years. This could be explained by the impact that the environment has on test weight. It is known that while genotypes affect test weight, the environment can also significantly influence test weight [8]. BB and W have different climates, with BB being cooler and less humid than W. Three environments, Warsaw 2020, Combined 2020, and Warsaw 2021, had no SNPs that were above the threshold of 4.91. While they did have SNPs that were just below the threshold, it is interesting that no SNPs showed significant association in those environments despite all other environments having at least one significant SNP. Additionally, BB 2020, Combined 2021, and BB 2021 had multiple SNPs on other chromosomes that were significant but not consistently found in the other environments. This contrasts with 2019, as all environments in this year only had significant SNPs on chromosome 15. These two findings could largely be explained by the time of harvesting and rainfall. It is important to note that delayed harvesting has been found to negatively affect test weight in wheat and corn [17,18]. Harvesting occurred normally in 2019 but was delayed by a couple of weeks due to the weather in 2020 and 2021. Because delaying harvesting and rainfall can impact test weight [17–19], these differences between years most likely resulted from delayed harvesting in 2020 and 2021. Other factors have been found to affect test weight as well. For instance, planting date and sulfur/phosphorus application have been found to impact test weight [20–22]. All tests were planted in mid-May, so it is unlikely that planting date had an effect on test weight for this project, and no fertilizers or nutrient supplements were applied to these tests.

Our correlation study found that the correlation between test weight and protein content was inconsistent. In BB 2019 and W 2019, protein was found to be significantly negatively correlated with test weight, but W 2020 showed significant, positive correlation. No significant correlation was found between high test weight and protein content in the other environments. Additionally, sucrose content was found to have no significant correlation with test weight. Other studies have found inconsistent correlation results between either protein or sucrose and test weight [8,11–13]. However, one study did find that sucrose could be significantly, positively related to test weight [8], but this study was conducted in Georgia, a different environment from Virginia, which might be the main cause of the different correlation results. No studies have looked at the relationships

of raffinose and stachyose with test weight. While we did find a significant, negative correlation between test weight and raffinose in BB 2019, it was found to not be significant in W 2019, and stachyose was found to not be significant in either location. Based on the inconsistent correlation between sucrose and test weight, it is not surprising that we did not find consistent significant correlations for these traits because these compounds are both sugars and similar to sucrose. On the other hand, oil was found to be significantly negatively correlated with test weight in all environments except for BB 2021. This is similar to the study conducted by Liu et al. 2019 that also found a significant, negative relationship between oil and test weight. This consistent negative relationship between oil content and high test weight could partly explain the decreasing trend in test weight that farmers have observed. One of the main traits that soybean breeders select for is higher oil content. Therefore, breeders may need to modify their selection objectives to balance oil content with increased test weight.

In summary, multiple significant SNPs associated with test weight on chromosomes 14 and 15 were identified, which could be used by breeders to quickly select high-test-weight progenies derived from high-test-weight parents in order to increase the overall test weight of a breeding program's germplasm. Additionally, a consistent negative relationship between high test weight and oil content was found. This information could be useful for breeders because while they make their selections to ensure high oil content, they should also pay attention to test weight to help offset its decreasing trend. Ultimately, incorporating parents that have high test weight into breeding schemes and taking into consideration the relationship between test weight and oil content will enable breeders to develop varieties with increased test weight to avoid harming farmers' profitability in the long run.

## 4. Materials and Methods

### 4.1. Plant Materials

All accessions were planted similarly to Singer et al. 2022 and Wang et al. 2022 [23,24]. Briefly, a total of 390 soybean accessions from a total of 17 countries from the Southern Core Collection that has been maintained at Virginia Tech were grown in Blacksburg (BB) and Warsaw (W), Virginia, for three years from 2019 to 2021. This panel contains soybean accessions that were grouped by maturity groups, and all accessions belonged to either maturity group IV or V at both locations for all years. Each sample was planted in two replications in 3 m two-row plots with 76 cm of row spacing for each plot in BB and 3 m four-row plots with 76 cm of row spacing in W, with Ellis and AG 4404 being used as commercial checks. All plots were checked for flower color and pubescent color, and any plants that did not match the correct color for that plot were removed. All plots were planted in the beginning of May in both locations and all years. In 2019, all plots were harvested around the beginning of October, but in 2020 and 2021, plots were not harvested until later in October due to weather conditions.

### 4.2. Test Weight

All samples were cleaned to remove split seeds, empty seed coats, pods, and sticks so that only intact seeds remained prior to measuring test weight. Sample seeds were also checked for seed coat and seed hilum color, which served as quality controls to remove any contaminant seeds. Blacksburg had 345, 314, and 241 accessions for 2019, 2020, and 2021, respectively. Warsaw had 314, 267, and 275 accessions for 2019, 2020, 2021, respectively. Test weight was determined by using a 2500 AGRI model. For each sample, 414 mL was used to calculate test weight. All samples were measured three times, and the resulting test weight values were averaged to obtain the final test weight value for the sample. For 2019, due to limited seed amount, the two replications had to be combined into a composite sample prior to measuring, but we were able to average two replications for a mean of the test weight of each accession in 2020 and 2021. Test weight values were adjusted to 13%

moisture content according to Liu et al. 2019. Lastly, test weight values were converted from lb/bu to kg/hL by multiplying the lb/bu values by a factor of 1.25.

#### 4.3. Genome-Wide Association Study (GWAS)

The genotypic data of all accessions screened by SoySNP50K iSelect Beadchip [25] are publicly available at soybase.org [26], with a total of 35,570 SNPs in this germplasm population. A GWAS was performed by first using the TASSEL 5.0 software to construct a kinship matrix, perform principal component analysis, and analyze association with a mixed linear model (MLM) [27]. The MLM was used to include a kinship matrix (K) with population structure (Q) to improve the statistical power using the Q + K approach [28]. A modified Šidák correction ( $\alpha_{\text{sid}} = 1 - (1 - \alpha)(1/m)$ ) for multiple testing was used to determine any significant markers, with the number of effective markers ( $M_{\text{eff}}$ ) being used instead of total number of markers ( $m$ ). The  $M_{\text{eff}}$  was determined to be 4191 through the poolr package in R with the Li and Ji method [29]. A modified significant threshold at  $\alpha = 0.05$  was constructed at  $-\log_{10}(P) > 4.91$  to determine the significance of single-nucleotide polymorphisms (SNPs). The qqman package was used to construct QQ and Manhattan plots [30]. In order to find candidate genes, genes located within 10 kbp of significant SNPs were found through the Soybase database [31]. Many SNPs were found to be significant across locations and years. To limit the search, only SNPs that were found to be significant or close to significant in multiple environments were used.

#### 4.4. Protein, Oil, and Sugar Content

To measure protein and oil content, near-infrared reflectance spectroscopy (NIRS) was performed. Each sample and replication was run twice on a DA7250 NIR Analyzer from Perten, and the protein and oil content were averaged for the sample. For each sample, about 60 mL was used. Only samples that had a yellow or green seed coat could be used. For both protein and oil content, 228 samples were used in both locations for 2019, 233 in both locations for 2020, 328 in BB 2021, and 258 in W 2021.

For sucrose, raffinose, and stachyose content, the protocol in Lord et al. 2011 was followed [32]. Briefly, 180 seed samples were first ground using a water-cooled grinder until the seeds became fine powder. All samples were weighed out to 0.1 g, and then 1.0 mL of HPLC-grade water was added to each sample. All samples were shaken for 15 min at 400 strokes per minute. Samples were centrifuged at 13.2 rpm for 15 min, and 0.5 mL of the supernatant was transferred to a new 2.0 mL centrifuge tube. Then, 0.7 mL of acetonitrile (ACN) was added, and all tubes were mixed by inverting multiple times. Samples sat for 1 h at room temperature and were then centrifuged at  $17,000 \times g$  for 15 min. After centrifuging, 100  $\mu\text{L}$  of the supernatant was mixed with 900  $\mu\text{L}$  of 65% ACN and filtered through a 0.2  $\mu\text{m}$  membrane into an HPLC sample vial. HPLC was used to determine sucrose content according to Lord et al. 2021. All samples were adjusted for moisture, and technical replicates were averaged together. Sugar analysis was only performed on 2019 samples due to there being no significance found between test weight and sugar. To determine correlation between high test weight and the three seed composition traits, protein, oil, and sugar content, R was used to calculate the Pearson's correlation coefficient. An  $\alpha$  of  $<0.05$  was used to determine if the correlations were significant.

**Author Contributions:** Z.S. was involved with methodology, formal analysis, writing—original draft preparation, and data curation. W.M.S. was involved with formal analysis and writing—revision and editing. L.R. was involved with data curation and writing—revision and editing. Q.S. was involved with writing—revision and editing. B.Z. was involved with writing—revision and editing and funding acquisition. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Virginia Soybean Board fund number 467252. Additionally, we would like to thank the Virginia Tech Open Access Subvention Fund (OASF) for their support.

**Institutional Review Board Statement:** Not applicable. This study did not involve humans or animals.

**Data Availability Statement:** All data were generated at Virginia Tech facilities and can be provided upon request. Additionally, they are available on figshare.org accessible 26 July 2023 ([https://figshare.com/articles/dataset/Test\\_Weight\\_Data\\_For\\_2019\\_to\\_2021\\_BB\\_and\\_W\\_Virginia\\_csv/23786286](https://figshare.com/articles/dataset/Test_Weight_Data_For_2019_to_2021_BB_and_W_Virginia_csv/23786286)).

**Acknowledgments:** We would like to thank all members of the Virginia Soybean breeding lab and Virginia Tech's Eastern Agricultural Research and Extension Center for help with field maintenance of plots. We would like to thank the Virginia Soybean Board for providing funding.

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

## References

- Pagano, M.C.; Miransari, M. 1—The Importance of Soybean Production Worldwide. In *Abiotic and Biotic Stresses in Soybean Production*; Miransari, M., Ed.; Academic Press: San Diego, CA, USA, 2016; pp. 1–26. Available online: <https://www.sciencedirect.com/science/article/pii/B9780128015360000013> (accessed on 23 August 2021).
- Shahbandeh. Global Major Soybean Exporting Countries Share 2020 [Internet]. Statista. 2021. Available online: <https://www.statista.com/statistics/961087/global-leading-exporters-of-soybeans-export-share/> (accessed on 23 August 2021).
- Davidson, D. Test Weight Matters in 2019 [Internet]. Ilsoyadvisor. 2018. Available online: <https://www.ilsoyadvisor.com/on-farm/ilsoyadvisor/test-weight-matters-2019> (accessed on 23 August 2021).
- Staton, M. Understanding Soybean Discount Schedules [Internet]. MSU Extension-Soybeans. 2020. Available online: <https://www.canr.msu.edu/news/understanding-soybean-discount-schedules> (accessed on 23 August 2021).
- Vázquez-Carrillo, M.G.; Rojas-Martínez, I.; Santiago-Ramos, D.; Arellano-Vázquez, J.L.; Espinosa-Calderón, A.; García-Pérez, M.; Crossa, J. Stability Analysis of Yield and Grain Quality Traits for the Nixtamalization Process of Maize Genotypes Cultivated in the Central High Valleys of Mexico. *Crop Sci.* **2016**, *56*, 3090–3099. [[CrossRef](#)]
- Jing, Q.; Jiang, D.; Dai, T.; Cao, W. [Effects of genotype and environment on wheat grain quality and protein components]. *Ying Yong Sheng Tai Xue Bao* **2003**, *14*, 1649–1653. [[PubMed](#)]
- Doehlert, D.C.; McMullen, M.S.; Hammond, J.J. Genotypic and Environmental Effects on Grain Yield and Quality of Oat Grown in North Dakota. *Crop Sci.* **2001**, *41*, 1066–1072. [[CrossRef](#)]
- Liu, L.; Prenger, E.; Zhang, J.; Little, B.; Mian, M.A.R.; Li, Z. Impact of genotype, seed composition, agronomic trait, and environment on soybean test weight. *J. Crop Improv.* **2019**, *33*, 711–729. [[CrossRef](#)]
- McNeece, B.T.; Gillenwater, J.H.; Li, Z.; Mian, M.A.R. Assessment of soybean test weight among genotypes, environments, agronomic and seed compositional traits. *Agron. J.* **2021**, *113*, 2303–2313. [[CrossRef](#)]
- Schuler, S.F.; Bacon, R.K.; Finney, P.L.; Gbur, E.E. Relationship of Test Weight and Kernel Properties to Milling and Baking Quality in Soft Red Winter Wheat. *Crop Sci.* **1995**, *35*, 949–953. [[CrossRef](#)]
- Gürsoy, S.; Sessiz, A.; Malhi, S.S. Short-term effects of tillage and residue management following cotton on grain yield and quality of wheat. *Field Crops Res.* **2010**, *119*, 260–268. [[CrossRef](#)]
- Bonfil, D.J.; Abbo, S.; Svoray, T. Sowing Date and Wheat Quality as Determined by Gluten Index. *Crop Sci.* **2015**, *55*, 2294–2306. [[CrossRef](#)]
- Rharrabi, Y.; Villegas, D.; Del Moral, L.F.G.; Aparicio, N.; Elhani, S.; Royo, C. Environmental and genetic determination of protein content and grain yield in durum wheat under Mediterranean conditions. *Plant Breed.* **2001**, *120*, 381–388. [[CrossRef](#)]
- Mijić, A.; Liović, I.; Zdunić, Z.; Marić, S.; Marjanović-Jeromela, A.; Jankulovska, M. Quantitative Analysis of Oil Yield and Its Components in Sunflower (*Helianthus annuus* L.). *Rom. Agric. Res.* **2009**, *26*, 41–46.
- Holland, J.B.; Frey, K.J.; Hammond, E.G. Correlated responses of fatty acid composition, grain quality, and agronomic traits to nine cycles of recurrent selection for increased oil content in oat. *Euphytica* **2001**, *122*, 69–79. [[CrossRef](#)]
- Severin, A.J.; Woody, J.L.; Bolon, Y.-T.; Joseph, B.; Diers, B.W.; Farmer, A.D.; Muehlbauer, G.J.; Nelson, R.T.; Grant, D.; Specht, J.E.; et al. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. *BMC Plant Biol.* **2010**, *10*, 160. [[CrossRef](#)] [[PubMed](#)]
- Cloninger, F.D.; Horrocks, R.D.; Zuber, M.S. Effects of Harvest Date, Plant Density, and Hybrid on Corn Grain Quality 1. *Agron. J.* **1975**, *67*, 693–695. [[CrossRef](#)]
- Lloyd, B.J.; Siebenmorgen, T.J.; Bacon, R.K.; Vories, E. Harvest date and conditioned moisture content effects on test weight of soft red winter wheat. *Appl. Eng. Agric.* **1999**, *15*, 525–534. [[CrossRef](#)]
- Guarienti, E.M.; Ciacco, C.F.; Cunha, G.R.D.; Del Duca, L.D.J.A.; Camargo, C.M.D.O. Effects of rainfall, relative humidity, and water excess and deficit on test weight, thousand kernel weight, and grain yield of wheat. *Food Sci. Technol.* **2005**, *25*, 412–418. [[CrossRef](#)]
- Hossain, I.; Epplin, F.M.; Krenzer, E.G. Planting Date Influence on Dual-Purpose Winter Wheat Forage Yield, Grain Yield, and Test Weight. *Agron. J.* **2003**, *95*, 1179–1188. [[CrossRef](#)]
- Agrawal, H.P.; Mishra, A.K. Sulphur nutrition of soybean. *Commun. Soil Sci. Plant Anal.* **1994**, *25*, 1303–1312. [[CrossRef](#)]
- Akter, F.; Md Islam, N.; Shamsuddoha, A.T.M.; Bhuiyan, M.S.I.; Shilpi, S. Effect of Phosphorus and Sulphur on Growth and Yield of Soybean. *Int. J. Bio-Resour. Stress Manag.* **2014**, *4*, 555–560.
- Singer, W.M.; Shea, Z.; Yu, D.; Huang, H.; Mian, M.A.R.; Shang, C.; Rosso, M.L.; Song, Q.J.; Zhang, B. Genome-Wide Association Study and Genomic Selection for Proteinogenic Methionine in Soybean Seeds. *Front. Plant Sci.* **2022**, *13*, 859109. [[CrossRef](#)]

24. Wang, Z.; Yu, D.; Morota, G.; Dhakal, K.; Singer, W.M.; Lord, N.; Huang, H.; Chen, P.; Mozzoni, L.; Li, S.; et al. Genome-wide association analysis of sucrose and alanine contents in edamame bean. *Front. Plant Sci.* **2023**, *13*, 1086007. [[CrossRef](#)]
25. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS ONE* **2013**, *8*, e54985. [[CrossRef](#)] [[PubMed](#)]
26. Song, Q.; Hyten, D.L.; Jia, G.; Quigley, C.V.; Fickus, E.W.; Nelson, R.L.; Cregan, P.B. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. *G3 Genes | Genomes | Genet.* **2015**, *5*, 1999–2006. [[CrossRef](#)] [[PubMed](#)]
27. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)] [[PubMed](#)]
28. Yu, J.; Pressoir, G.; Briggs, W.H.; Vroh Bi, I.; Yamasaki, M.; Doebley, J.F.; McMullen, M.D.; Gaut, B.S.; Nielsen, D.M.; Holland, J.B.; et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **2006**, *38*, 203–208. [[CrossRef](#)]
29. Li, J.; Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **2005**, *95*, 221–227. [[CrossRef](#)] [[PubMed](#)]
30. Turner, S.D. qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv* **2014**. [[CrossRef](#)]
31. Grant, D.; Nelson, R.T.; Cannon, S.B.; Shoemaker, R.C. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **2010**, *38* (Suppl. 1), D843–D846. [[CrossRef](#)]
32. Luciana Rosso, M.; Burleson, S.A.; Maupin, L.M.; Rainey, K.M. Development of breeder-friendly markers for selection of MIPS1 mutations in soybean. *Mol. Breed.* **2011**, *28*, 127–132. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.