

## Article

# Integrated Transcriptomic and Proteomic Characterization of a Chromosome Segment Substitution Line Reveals the Regulatory Mechanism Controlling the Seed Weight in Soybean

Siming Wei <sup>1,†</sup>, Zhenhai Yu <sup>2,†</sup>, Fangfang Du <sup>1,†</sup>, Fubin Cao <sup>1</sup>, Mingliang Yang <sup>1</sup>, Chunyan Liu <sup>1</sup>, Zhaoming Qi <sup>1</sup> , Qingshan Chen <sup>1</sup>, Jianan Zou <sup>1,\*</sup> and Jinhui Wang <sup>1,\*</sup>

- <sup>1</sup> National Key Laboratory of Smart Farm Technology and System, Key Laboratory of Soybean Biology in Chinese Ministry of Education, College of Agriculture, Northeast Agricultural University, Harbin 150030, China; wsm20016@126.com (S.W.); s220301027@neau.edu.cn (F.D.); fbincao@126.com (F.C.); mlyang@neau.edu.cn (M.Y.); cyliucn@neau.edu.cn (C.L.); qizhaoming1860@126.com (Z.Q.)
- <sup>2</sup> Heilongjiang Province Green Food Science Institute, Harbin 150028, China; hyzh1188@sina.com
- \* Correspondence: zoujianan@neau.edu.cn (J.Z.); wangjinhui113@neau.edu.cn (J.W.)
- † These authors contributed equally to this work.

**Abstract:** Soybean is the major global source of edible oils and vegetable proteins. Seed size and weight are crucial traits determining the soybean yield. Understanding the molecular regulatory mechanism underlying the seed weight and size is helpful for improving soybean genetic breeding. The molecular regulatory pathways controlling the seed weight and size were investigated in this study. The 100-seed weight, seed length, seed width, and seed weight per plant of a chromosome segment substitution line (CSSL) R217 increased compared with those of its recurrent parent ‘Suinong14’ (SN14). Transcriptomic and proteomic analyses of R217 and SN14 were performed at the seed developmental stages S15 and S20. In total, 2643 differentially expressed genes (DEGs) and 208 differentially accumulated proteins (DAPs) were detected at S15, and 1943 DEGs and 1248 DAPs were detected at S20. Furthermore, integrated transcriptomic and proteomic analyses revealed that mitogen-activated protein kinase signaling and cell wall biosynthesis and modification were potential pathways associated with seed weight and size control. Finally, 59 candidate genes that might control seed weight and size were identified. Among them, 25 genes were located on the substituted segments of R217. Two critical pathways controlling seed weight were uncovered in our work. These findings provided new insights into the seed weight-related regulatory network in soybean.

**Keywords:** seed weight; seed size; regulatory pathway; transcriptomic; proteomic; soybean



**Citation:** Wei, S.; Yu, Z.; Du, F.; Cao, F.; Yang, M.; Liu, C.; Qi, Z.; Chen, Q.; Zou, J.; Wang, J. Integrated Transcriptomic and Proteomic Characterization of a Chromosome Segment Substitution Line Reveals the Regulatory Mechanism Controlling the Seed Weight in Soybean. *Plants* **2024**, *13*, 908. <https://doi.org/10.3390/plants13060908>

Academic Editors: Igor G. Loskutov and Jiangqi Wen

Received: 10 February 2024  
Revised: 13 March 2024  
Accepted: 18 March 2024  
Published: 21 March 2024



**Copyright:** © 2024 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

Globally, soybean is widely used as a crucial food, oil, and forage crop. In 2023, the planting area of soybeans in China was 157 million mu, and the total production was 20.84 million tons, with a self-sufficiency rate of only 17.3% (<http://www.stats.gov.cn/>, accessed on 12 March 2024). Improving soybean yield has always been a critical goal worth continuously pursuing. Seed yield is determined by the number of seeds produced and seed weight and is a complex quantitative trait [1,2]. The 100-seed weight, seed length, and seed width are critical factors determining soybean yield. Seed-related traits are often correlated with each other. In previous studies, the 100-seed weight and seed size were generally positively correlated [3–7]. Cell development determines the final seed size and weight, and cell proliferation and expansion co-regulate the final seed size [2]. Genes regulating seed weight and size were recently cloned in soybean [2–9]. Revealing the seed weight regulation pathway contributes to the genetic improvement in soybean.

Transcriptomic analysis is an effective method for understanding the molecular genetic relationship of seed weight based on the transcriptional variation. The genes related to

seed weight and size control *GmGA200X*, *SoyWRKY15a*, and *GmCYP78A5* were cloned based on the results of the transcriptomic analysis [3,10,11]. The potential genetic variation underlying the developmental evolution of soybean seeds was revealed based on the transcriptomic comparison between the cultivar and wild soybean [12]. Proteomic analysis is generally used for investigating the effects of seed development on the yield, thereby predicting potential proteins related to seed weight and size and revealing the interaction network between seed size-related proteins in major crops [13–17]. Multi-omics are more efficient and comprehensive for revealing the regulatory mechanism underlying seed development. Integrated transcriptomic and proteomic analyses were recently used to investigate the potential pathways related to seed weight and size [11,18,19].

The molecular networks of seed size have been preliminarily established in *Arabidopsis* and rice [20,21]. Mitogen-activated protein kinase (MAPK) signaling pathways are the critical seed size-regulating pathways [21]. The relatively complete MAPK cascade, OsMKKK10-OsMKK4-OsMPK6 module, has been reported to control grain size in rice, and this module was identified through the genetic screening of the small grain phenotype and map-based cloning [22–24]. The loss of function of *OsMKKK10*, *OsMKK4*, or *OsMPK6* results in small and light grains, and *OsMKKK10* or *OsMKK4* expression results in heavy and large grains [25]. Similarly, in *Arabidopsis*, *MKK4*, and *MKK5* control seed size upstream of *MPK6*, which is consistent with the regulatory pattern in rice [26]. *ERECTA1* upstream of the OsMKKK10-OsMKK4-OsMPK6 cascade is involved in controlling seed size by regulating cytokinin homeostasis in rice [27]. Moreover, the ERECTA-MAPK-DA1-UBP15 signaling module regulates seed size in *Arabidopsis* [28]. A lectin RLK, LecRK-VIII.2, controls the seed size, silique number, seed number, and seed yield in *Arabidopsis*, and the positive role of LecRK-VIII.2 in seed coat is a maternal effect exerted through the simultaneous control of cell expansion and proliferation [29]. The OsMKK-OsMAPK6-OsWRKY53 signaling module was identified to control seed size in rice through cell expansion [30,31]. In addition, Map Kinase Phosphatase 1 (OsMKP1) was involved in seed size control by negatively regulating the OsMKKK10-OsMKK4-OsMPK6 cascade [27]. These reports suggest that the OsMKKK10-OsMKK4-OsMPK6 module for seed size control is conserved among multiple species.

Cell walls are essential for plant organ development [32]. Plant cell walls are basically composed of three major polysaccharide types, including pectin, cellulose, and hemicellulose [33]. The cell wall structure and composition determine its various functions, and multiple functions have been identified for cell wall biosynthesis and modification pathways [34]. One of the crucial functions of cell walls is to participate in seed development, and ultimately determine seed size [35–39]. Cell wall biosynthesis and modification-related genes that control seed weight have been identified in rice, *Arabidopsis*, maize, barley, wheat, tomato, and soybean [5,35–46]. Few studies have reported that structural variations in the cell wall affect seed development, and ultimately determine seed size in soybean. The seed thickness-regulating gene *ST1* (*Seed Thickness 1*) encodes uridine diphosphate (UDP)-D-glucuronate 4-epimerase. *ST1* can catalyze UDP-galacturonic acid production and promote pectin biosynthesis [5]. In addition, the major function of polygalacturonase is to decompose pectin or pectinase enzymes. *PG031* is a pleiotropic *polygalacturonase* (*PG*) gene that regulates seed coat permeability and 100-seed weight in soybean. Subcellular localization has demonstrated that *PG031* is localized to the cell wall of the seed coat [46]. These studies have shown that cell wall-related pathways are crucial for controlling seed size. Therefore, as more seed size-regulating genes are cloned, the molecular genetic network of the seed size needs to be revealed.

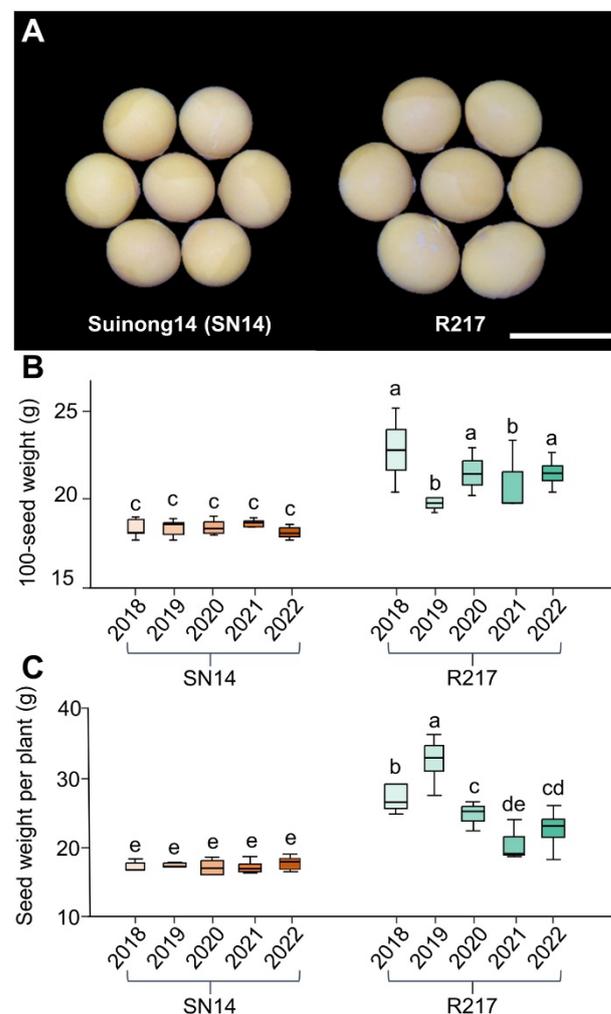
CSSLs are ideal genetic resources for studying genetic variations because each CSSL carries a single chromosomal segment. CSSLs are used to identify the quantitative trait loci (QTL), cloning genes, identifying the genetic pattern of agronomic traits, and investigating the interacting relationships between genes and the environment. CSSLs were first developed from tomato plants [47] and were later widely used to study multiple crops such as rice, barley, and cotton [48–50]. A CSSL population was developed by substituting

the chromosome segments of ‘ZYD00006’ into the SN14 genome [51,52]. A large-seed CSSL R217 was selected from among the CSSLs. In this study, the regulatory pathways of seed weight were illuminated based on the results of the integrated transcriptomic and proteomic analyses.

## 2. Results

### 2.1. Phenotypic Variation of the Large Seed CSSL R217

A large-seed CSSL R217 was selected from a wild soybean CSSLs population. The CSSL was developed from the wild soybean varieties ‘ZYD00006’ (ZYD06) and ‘Suinong14’ (SN14), which were used as the donor and recurrent parents, respectively. The substituted segments of CSSLs covered over 90% of the wild soybean genome [51,52]. Phenotypic statistics of seed weight and size for multiple years indicated that 100-seed weight, seed length, seed width, and seed weight per plant of R217 increased compared to SN14 (Figure 1A–C and Figure S1). Therefore, R217 is a suitable material for studying the genetic variation mechanism of seed weight. These results suggest that chromosomal segments on R217 are the main reason for the variation in seed weight between R217 and SN14.



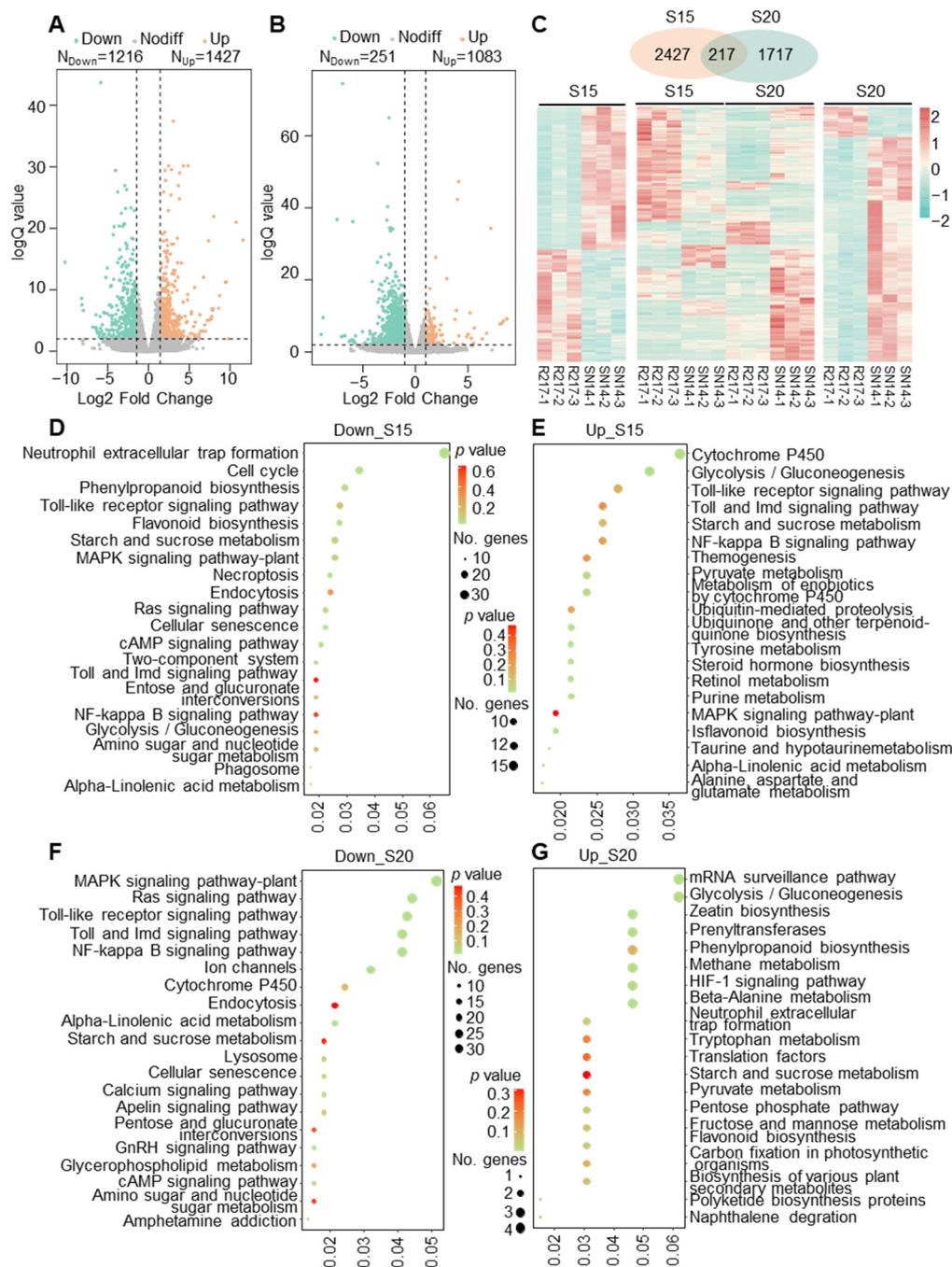
**Figure 1.** The comparison of 100-seed weight and seed weight per plant between SN14 and large-seed CSSL R217 for 5 years. (A) The seed sizes of SN14 and R217. Bar = 1 cm. (B,C) The phenotypic data of 100-seed weight and seed weight per plant for SN14 and R217 from 2018 to 2022. In (B,C), different lowercase letters (a, b, c, d, and e) indicate significant differences, the threshold is set to  $p < 0.05$ , and identical letters suggest insignificant differences in one-way ANOVA.

## 2.2. Transcriptomic Comparison between R217 and SN14

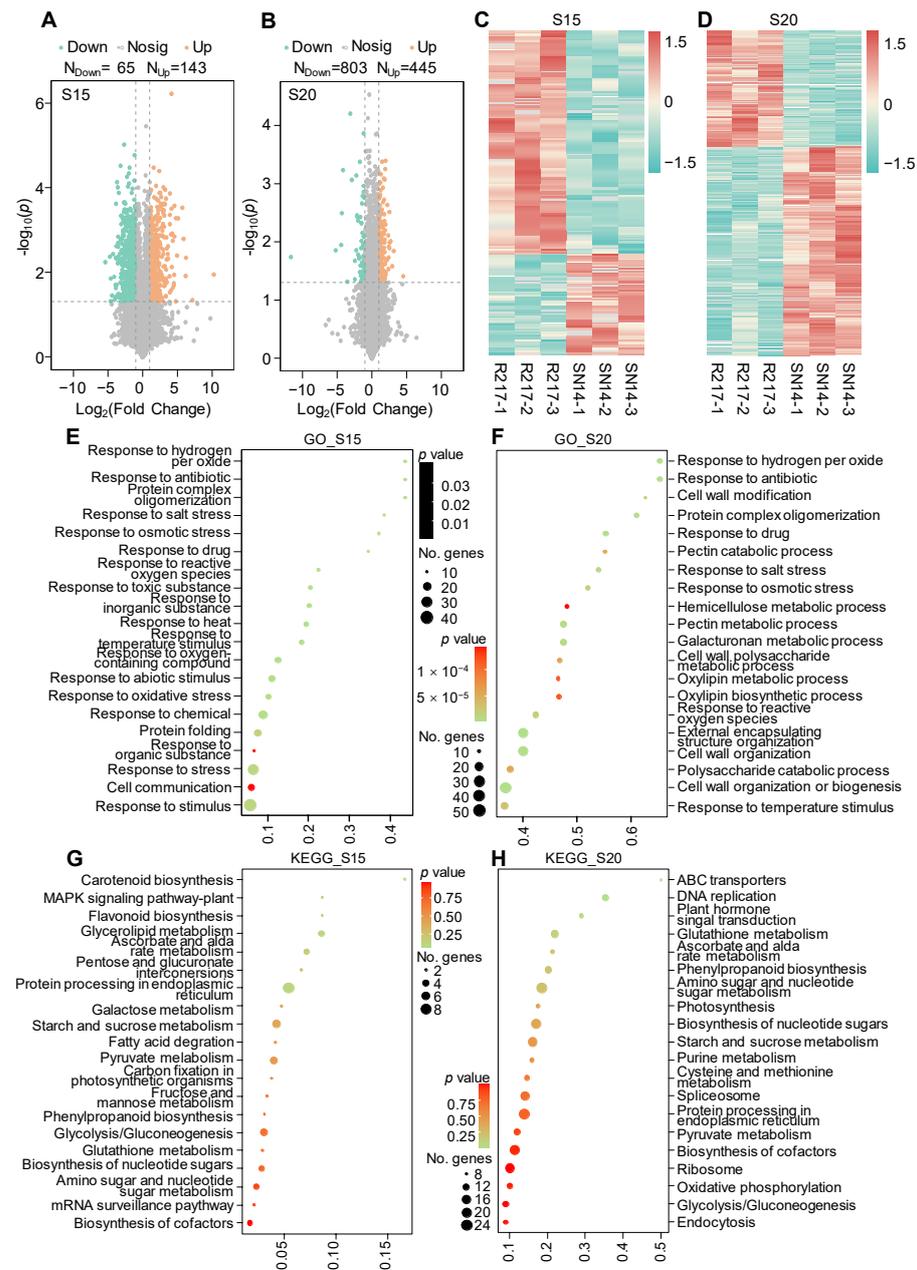
To reveal the transcriptional variations between R217 and SN14, the transcriptomic analysis was performed at S15 and S20 seed developmental stages. A total of 2643 DEGs were detected, including 1216 upregulated and 1427 downregulated DEGs between R217 and SN14 at S15 (Figure 2A). A total of 1934 DEGs were estimated at S20, including 251 upregulated and 1083 downregulated DEGs, respectively (Figure 2B). Overall, 2427 specific DEGs at S15, 1717 specific DEGs at S20, and 217 DEGs were shared between R217 and SN14 at both S15 and S20 (Figure 2C). The results of the KEGG enrichment analysis indicated that the downregulated DEGs at S15 were chiefly enriched in neutrophil extracellular trap formation, cell cycles, and phenylpropanoid biosynthesis (Figure 2D). Moreover, the upregulated DEGs at S15 were enriched in cytochrome P450, glycolysis/gluconeogenesis, Toll-like receptor signaling pathway (Figure 2E). At S20, the downregulated DEGs were enriched in the MAPK signaling pathway, RAS signaling pathway, and Toll-like receptor signaling pathway (Figure 2F), and the upregulated DEGs were enriched in the mRNA surveillance pathway, glycolysis/gluconeogenesis, zeatin biosynthesis (Figure 2G). The pathways, including Toll-like receptor signaling pathway, toll and IMD signaling pathway, starch and sucrose metabolism, MAPK signaling pathway-plant, starch and sucrose metabolism, NF-kappa B signaling pathway, and alpha-linolenic acid metabolism, were enriched in both upregulated and downregulated DEGs at S15 (Figure 2D,E). The processes of MAPK signaling pathway, phenylpropanoid biosynthesis, starch and sucrose metabolism, pyruvate metabolism, flavonoid biosynthesis, toll and IMD signaling pathway, Toll-like receptor signaling pathway, cytochrome P450, glycolysis/gluconeogenesis, NF-kappa B signaling pathway, amino sugar metabolism, nucleotide sugar metabolism, and RAS signaling pathway were together enriched at S15 and S20 (Figure 2D–G). Additionally, according to the GO enrichment analysis, histone H3-K9 modification and methylation, DNA modification, ion transport and methylation, plant-type secondary cell wall biogenesis, and cell cycle-related GO terms were enriched at S15 (Figure S2A,B). Immune and plant-type cell wall modification-related GO terms were enriched at S20 (Figure S2C,D). These results were worth mentioning that both GO and KEGG enrichment analyses were enriched in pathways and GO terms related to cell wall biosynthesis and modification, including phenylpropanoid biosynthesis, plant-type secondary cell wall biogenesis, and plant-type cell wall modification at two seed development stages (Figure 2 and Figure S2).

## 2.3. Proteomic Analysis of R217 and SN14

Proteomic analysis was performed to compare the variation in protein expression between R217 and SN14. In total, 65 and 143 downregulated and upregulated DAPs were detected at S15, respectively (Figure 3A). At S20, 803 and 445 downregulated and upregulated DAPs were detected, respectively (Figure 3B). Protein expression during soybean seed developmental stages is a dynamic process. Protein expression differed at S15 and S20 (Figure 3C,D). At S15, the response to stress-related GO terms was mainly enriched (Figure 3E). However, at S20, cell wall biosynthesis- and modification-related GO terms were mainly enriched, including pectin catabolic or metabolic process, cell wall modification, and cell wall organization or biogenesis at S20, as determined based on the GO enrichment analysis of DAPs (Figure 3F). Similarly, the cell wall-related pathways, including flavonoid biosynthesis and phenylpropanoid biosynthesis, were also enriched at both S15 and S20, as determined based on the KEGG enrichment analysis (Figure 3G,H); the MAPK signaling pathway was also enriched (Figure 3H). These results are consistent with the analysis of DEGs, which indicates that DAPs are also enriched in the MAPK signaling pathway and cell wall biosynthesis and modification. These results suggest that the variation in the seed weight in R217 is caused by differences in the expression patterns of genes related to the MAPK signaling pathway and cell wall biosynthesis and modification during seed developmental stages.



**Figure 2.** The DEG analysis and KEGG enrichment analysis in R217 and SN14 at S15 and S20 seed developmental stages. (A,B) Volcano map of DEGs at S15 and S20 seed developmental stages, respectively. Down—DEGs with downregulated expression. Up—DEGs with upregulated expression. Nodiff—genes without differential expression in RNA-seq analysis. (C) The heatmap of all DEGs in R217 and SN14 at S15 and S20 seed developmental stages. (D,E) The KEGG enrichment analysis of the top downregulated and upregulated DEGs at the S15 seed developmental stage. (F,G) The KEGG enrichment analysis of the top downregulated and upregulated DEGs at the S20 seed developmental stage. In (D–G), the vertical coordinates represent the GO terms enriched with genes from the substituted segments. The horizontal axis represents the gene ratio. No. genes—the number of genes.

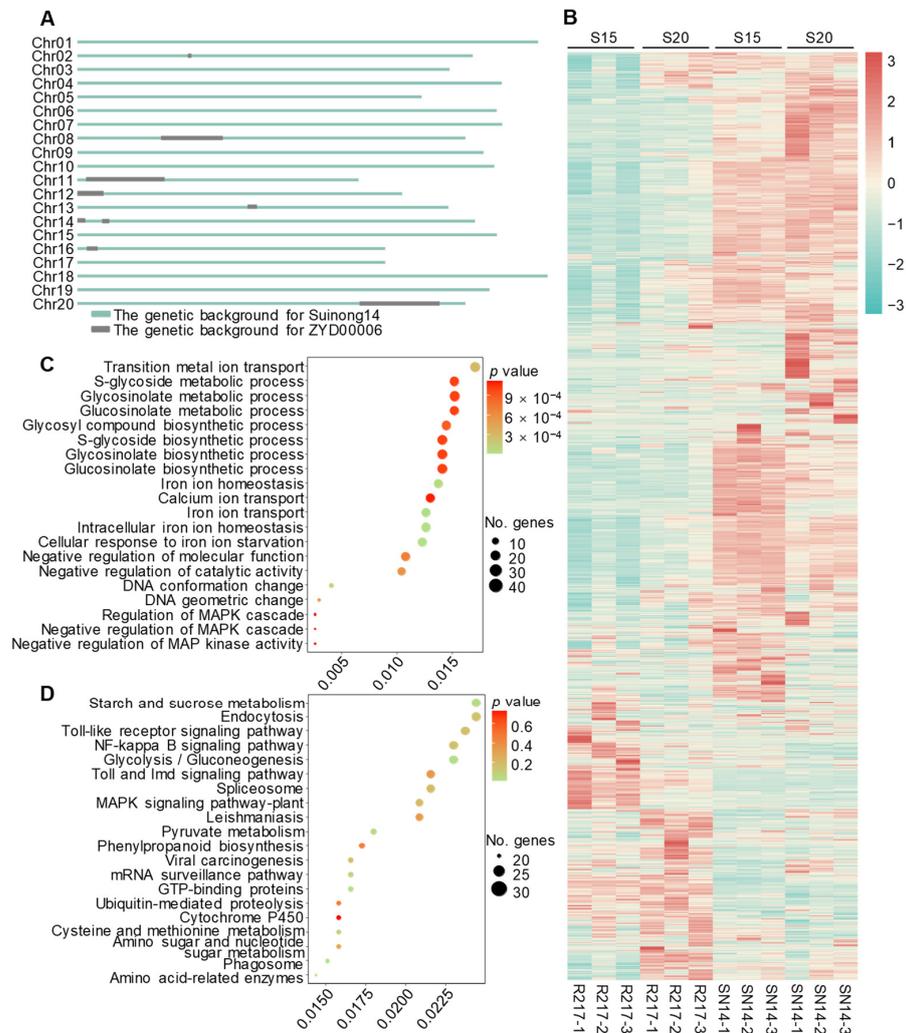


**Figure 3.** The DAP analysis, GO enrichment analysis, and KEGG enrichment analysis of DAPs in R217 and SN14 at S15 and S20 seed developmental stages. (A,B) Volcano maps of all DAPs at the S15 and S20 seed developmental stages. Down—DEGs with downregulated expression. Up—DEGs with upregulated expression. Nosig—genes without differential expression in RNA-seq analysis.  $N_{Down}$ —number of downregulated DAPs.  $N_{Up}$ —number of upregulated DAPs. (C,D) The heatmap of all DAPs in R217 and SN14 at S15 and S20. (E,F) The GO enrichment analysis of the top DAPs at S15 and S20. (G,H) The KEGG enrichment analysis of the top DAPs at S15 and S20. In (E–H), the vertical coordinates represent the GO terms enriched with genes from the substituted segments. The horizontal axis represents the gene ratio.

#### 2.4. Analysis of Genes on the Substituted Chromosomal Segments of R217

To further explain the seed weight variation of R217, the bin-map data of SN14 and ZYD00006 were used to obtain the information of substituted segments on genome of R217 [52]. Ten substituted segments were distributed on eight chromosomes of R217 (Figure 4A; Table S1). The expression patterns of all genes on the substituted segments were analyzed based on the RNA-seq and proteomic data (Figure 4B–D). A total of 3203 genes

were located on the substituted segments of R217 at S15 and S20 seed developmental stages (Figure 4B). These genes are mainly enriched in ion transport, glycometabolism and MAPK signaling pathways by performing GO enrichment analysis (Figure 4C). At the same time, the pathways of starch and sucrose metabolism, endocytosis, Toll-like receptor signaling pathway, and MAPK signaling pathway were mainly enriched based on KEGG enrichment analysis (Figure 4D). The enrichment analysis results of both methods were enriched in the MAPK signaling pathway.



**Figure 4.** Gene analysis and enrichment analyses on the substituted segments of R217. (A) Genome diagram of R217. The distribution of ZYD00006 substituted segments on the R217 genome. The green lines represent the SN14 genome and the gray boxes represent the ZYD00006 genome. (B) The heatmap of all genes on the substituted region of R217 at S15 and S20 seed developmental stages. S15 and S20 represent the 15th and 20th days after fertilization, respectively. (C) The GO enrichment analysis of genes on the substituted segments of R217. (D) The KEGG enrichment analysis of all genes on the substituted region of R217. In (C,D), the vertical coordinates represent the GO terms enriched with genes from the substituted segments. The horizontal axis represents the gene ratio. No. genes—the number of genes.

### 2.5. Potential Molecular Genetic Mechanism of Seed Weight

Integrating transcriptomic and proteomic analyses provides a more comprehensive understanding of the molecular genetic mechanism of seed weight. The results of transcriptomic and proteomic analyses indicate that multiple pathways are involved in regulating seed weight. The variations in gene expression on the substituted segments of R217 caused

changes in the expression patterns of DAPs and DEGs outside the substituted segments. Therefore, according to the results of gene analyses on the substituted segments, the MAPK signaling pathway might be a critical pathway regulating seed weight. Fourteen DEGs involved in the MAPK signaling pathway were selected from the substituted segments of R217 (Figure 5A). Furthermore, the analysis results of the DEGs and DAP expression patterns based on FPKM and LFQ intensity data indicate that 2 DAPs and 18 DEGs are involved in the MAPK signaling pathway in regions outside of the substituted segments (Figure 5A). The expression patterns of most MAPK signaling pathway-related genes were similar. The expression of these genes was significantly lower at S15 than at S20. The expression of the 18 DEGs decreased in R217 compared with SN14 (Figure 5A). Conversely, I1KVS8 and *Glyma.08G004500* were highly expressed at S15, but their expression was lower at S20 (Figure 5A). The expression level was higher in R217 than in SN14 at both S15 and S20 (Figure 5A).

The cell wall biosynthesis process might be another crucial pathway that participates in seed weight regulation. Two DAPs and nine DEGs involved in cell wall biosynthesis and modification were selected from the substituted segments of R217 (Figure 5B). The expression patterns of DEGs and DAPs were quite different between R217 and SN14 at S15 and S20. This indicates that DAPs and DEGs involved in the MAPK signaling pathway and cell wall biosynthesis and modification were direct factors affecting the variation in seed weight. Furthermore, the analysis results of gene expression patterns based on FPKM and LFQ intensity data indicate that nine DAPs and five DEGs outside the substituted segments were involved in cell wall biosynthesis and modification (Figure 5B). Among the aforementioned five DEGs, the expression patterns of four DEGs, namely *Glyma.06G225500*, *Glyma.09G051100*, *Glyma.15G157100*, and *Glyma.19G033600* were consistent, with the expression being higher at S15 than at S20 (Figure 5B). The expression level of all five DEGs significantly increased in R217 compared with SN14 at S15 and S20 (Figure 5B). The expression of I1JQY5, I1KAS1, K7LZ56, I1LVD2, and A0A0R0G1U8 was consistent with that of the aforementioned four DEGs (Figure 5B). On the contrary, the expression of four DAPs, namely I1N6E4, I1KDI6, I1L158, and I1L5A3, was lower at S15 than at S20, and the expression levels of the four proteins significantly decreased in R217 compared with SN14 (Figure 5B). The detailed expression and protein abundance data of these candidate genes are also listed (Table S2). The results revealed that DEGs and DAPs related to the MAPK signaling pathway and the cell wall biosynthesis and modification pathway may be the main genes and proteins involved in seed weight regulation. Moreover, these results suggest that DEGs and DAPs located on the substituted segments are direct factors causing the variation in seed weight in R217, including the genes related to the MAPK signaling pathway and the cell wall biosynthesis and modification pathway.

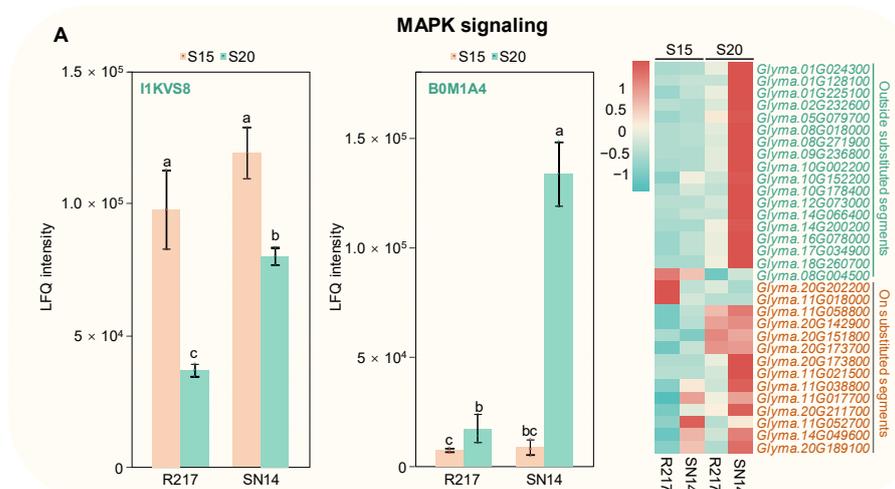
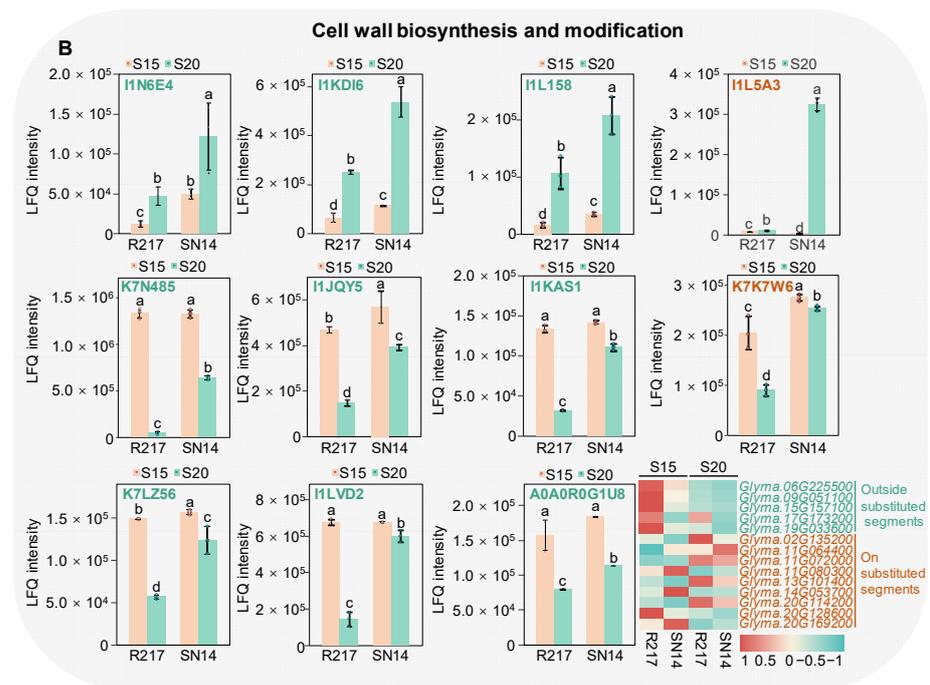


Figure 5. Cont.



**Figure 5.** The expressed pattern of DAPs and DEGs involved in the 100-seed weight-related pathway at S15 and S20 seed developmental stages. (A,B) DAPs and DEGs involved in the MAPK signaling pathway at S15 and S20 seed developmental stages. The green font represents the genes or protein on the substituted segments. The tangerine font represents the genes or protein outside the substituted segments. Different lowercase letters (a, b, c, and d) indicate significant differences and identical letters suggest insignificant differences in one-way ANOVA.

### 3. Discussion

Soybean is considered a dual functional crop of great economic value because of its high protein and oil content [53]. Seed weight is one of the traits that significantly affect soybean yield [3,9]. Establishing the genetic molecular mechanism underlying seed weight can offer strong clues for the breeding and genetic improvement of modern soybean varieties. In this study, a large seed CSSL R217 was selected from a CSSL population. The 100-seed weight, seed length, seed width, and seed weight per plant increased compared with its parent SN14. The substitution of ZYD06 chromosomal segments increased the seed size and weight of R217. The substituted segments of R217 were distributed on eight chromosomes, and 3203 genes were located on the substituted segments. These genes were mainly enriched in the MAPK signaling pathway, phenylpropanoid metabolic pathway, glycometabolism, and ubiquitination pathway. The transcriptomic analysis-based DEGs and proteomic analysis-based DAPs were also enriched in the MAPK signaling pathway and the cell wall biosynthesis and modification pathway at both S15 and S20 seed developmental stages, although the expression patterns of these DAPs and DEGs differed at the two stages. We thus uncovered two major potential pathways involved in seed weight control, including the MAPK signaling pathway and the cell wall biosynthesis and modification pathway.

#### 3.1. CSSLs Facilitate Identifying Seed Size-Regulating Genes in Soybean

Because of its highly consistent genetic backgrounds, the CSSL is an efficient population for identifying the QTL related to seed weight, analyzing the interaction between the QTL, and improving varieties [54,55]. A few substituted CSSL segments can cause variations in seed size and weight. Identifying seed weight-related genes on the substituted segments directly contributes to reducing the workload compared to analyzing the whole soybean genome. CSSLs are widely used for studying multiple species, includ-



### 3.2. Integrated Transcriptomic and Proteomic Analyses Reveal the Seed Weight-Related Regulatory Pathways

Seed weight and size are complex quantitative traits that are controlled by many minor genes [1,2]. Therefore, cloning as many genes as possible that regulate seed weight, revealing the regulatory mechanism controlling seed weight, and establishing a seed weight regulatory network are challenging. Multi-omics analysis was recently used to identify and uncover essential components of stress responses, senescence, and yields in multiple economically crucial crops, including rice, maize, peanut, wheat, cotton, and soybean [58–66]. Thus, multi-omics analysis is among the effective methods for identifying seed weight-related genes and revealing their associated regulatory mechanisms. However, multi-omics analysis is currently only applied to study pollen development, seed storage profiles, and abiotic stress in soybean [18,66,67], and its application to seed weight in soybean has not been reported. In this study, transcriptome and proteomic analyses of R217 and SN14 were performed at S15 and S20 stages, respectively. In total, 2643 DEGs and 208 DAPs were detected at S15 and 1934 DEGs and 1248 DAPs were detected at S20. To more accurately and comprehensively reveal the seed weight-associated regulatory mechanism, we integrated the transcriptomic and proteomic data. Two seed weight-regulating pathways were uncovered, namely the MAPK signaling pathway and cell wall biosynthesis pathway. Moreover, 14 genes involved in the MAPK signaling pathway and 11 genes involved in the cell wall biosynthesis pathway were considered as seed weight-controlling candidate genes. These genes were all located on the substituted segments of R217. Genomic sequencing analysis will be performed based on the whole-genome re-sequencing data of CSSL [51,52], which will help further identify crucial seed weight-controlling genes from the aforementioned candidate genes.

### 3.3. Potential Regulatory Mechanisms Involved in Seed Size Control

The MAPK signaling pathway plays a crucial role in eukaryotes. This pathway has been identified to be involved in plant immunity, response to environmental stresses, and normal growth and development [21]. One of the crucial functions of the aforementioned pathway is controlling seed size [20,68]. Three tiers of protein kinases collectively compose a complete MAPK cascade [20,69]. The MAPK cascade OsMKKK10–OsMKK4–OsMPK6 module was revealed in rice [22–24], and the homologous module has been uncovered in *Arabidopsis* [26]. MAPK cascades are conserved among multiple species. The seed size-controlling pathways were further complemented based on the OsMKKK10–OsMKK4–OsMPK6 modules, including the ERECTA1–OsMKKK10–OsMKK4–OsMPK6–DST–cytokinin [70], ERECTA1–MAPK–DA1–UBP15 [22], OsMKK4–OsMAPK6–OsWRKY53 [30,31,71], MPK1–OsMKKK10–OsMKK4–OsMPK6 [25,27], and LecRK–VIII.2–OsMKKK10–OsMKK4–OsMPK6 modules [29]. Many genes were enriched in the MAPK signaling pathway in this study. In addition, several top DEGs and DAPs were also enriched in BR signaling and ubiquitin–proteasome pathways. Studies have reported that OsWRKY53 in rice and its orthologous gene WRKY33 in *Arabidopsis* are key genes regulating seed size that are involved in the BR signaling pathway [30,71]. The DA1–UBP15 module plays a crucial role in controlling seed size by participating in ubiquitin–proteasome pathways [72,73]. On combining the results of the previous studies and the present study, we noted that the candidate genes detected might control the seed size and weight by regulating the expression of MAPK–WRKY and MAPK–BR-related genes.

The main components of cell walls are pectin, cellulose, hemicellulose, and lignin, with lignin present between two cellulose molecules [74,75]. Cellulose is mainly composed of glucose, pectin is composed of a group of galacturonic acids, and lignin is mainly composed of phenylpropane. Cell walls are essential for vegetative development [32,76]. They provide structural support for plant cells, protect cell walls from abiotic and biotic stresses, and determine the final cell morphology [32,44]. GW2 controlled the grain size of wheat by regulating the expression of cell wall synthesis-related genes and affected the synthesis of cell walls in wheat [36]. In barley, two pectin methylesterase genes *OVULE PECTIN*

*MODIFIER 1 (OPM1)* and *OPM2* controlled grain size by regulating homogalacturonan accumulation [42]. In *Arabidopsis*, the  $\alpha$ -XYLOSIDASE1 (*XYL1*) gene encodes the only  $\alpha$ -xylosidase acting on xyloglucans, which are the major hemicellulose polysaccharides in the primary cell walls of dicots and non-graminaceous monocots. *XYL1* was regulated by the MADS box transcription factor *SEEDSTICK* to affect cell wall thickness and ultimately determine seed size [39]. *GS3.1* regulates the grain size of rice by controlling metabolic-flux distribution between flavonoid and lignin metabolons [77]. In summary, the cell wall composition and modification regulate seed size. However, details of the regulatory mechanisms underlying these cell wall-related genes have not been reported. Several top DEGs and DAPs in this study were enriched in cell wall modification, the cell wall polysaccharide metabolic process, the pectin metabolic process, and phenylpropanoid biosynthesis. These results suggested that genes encoding cell wall biological process-related top DEGs and DAPs are crucial seed size-controlling candidate genes. The mechanism regulating seed weight through cell wall regulation will be further explored in future work.

Moreover, MAPK cascades are involved in cell wall integrity regulation [32]. The MAP-kinase 3 phosphorylates UDP-glucose dehydrogenase and provides UDP-sugar for cell wall biosynthesis in *Arabidopsis* [78]. MPK4 affects cellulose content by negatively regulating L-arabinose synthesis for cell wall formation in *Arabidopsis* [79]. Based on these reported works, we inferred that MAPK cascades regulate cell wall biosynthesis, but the cooperative regulation of these cascades and the cell wall process controlling seed size remain unclear. Therefore, based on the results of our integration of transcriptomic and proteomic analyses, we believed that the MAPK cascades and cell wall biosynthesis and modification may co-regulate seed weight and size (Figure 6; Table S3).

## 4. Materials and Methods

### 4.1. Plant Materials

A large seed CSSL R217 was selected from the soybean CSSL population. The donor parent of the CSSL population was wild soybean ZYD00006, and the recurrent parent was the cultivated soybean SN14 [51,52]. These are varieties from Heilongjiang Province, China. The CSSLs have been re-sequenced [52], and were grown at Xiangyang experimental Farm of Northeast Agricultural University in Harbin, China (45.75° N, 126.53° E). Each row was 5 m in length and the spacing distance was between plants was 6 cm and that between the rows was 0.6 m. After the seeds naturally matured, the harvest was completed. The materials used for sequencing were planted in a greenhouse under the same conditions, and each material was replanted three times.

### 4.2. Seed Weight and Size Evaluation

Three individuals with dry seeds were randomly selected from R217 and SN14 after the soybean plants matured. The 100-seed weight, seed length, seed width, and seed weight per plant of R217 and SN14 were estimated from 2018 to 2022. The 100-seed weight was measured by randomly selecting 100 high-quality seeds for each individual. The seed length and width were estimated using an average of 5 seeds. The seed weight per plant was evaluated by weighing all the seeds produced by a single plant.

### 4.3. RNA-Seq Analysis

Total RNA was extracted from three independent biological samples of R217 and SN14 at S15 and S20. Total RNA was isolated using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. mRNA was purified and fragmented into small pieces from 50  $\mu$ g of total RNA. mRNA was reverse transcribed to produce cDNA for RNA-seq research. Applied Protein Technology (APT BIO, Shanghai, China) performed RNA-seq analysis by using an Illumina NovaSeq 6000 system to produce 150-bp pair-end (PE150) reads. fastq was used to perform basic statistics on raw seed quality. The clean reads were mapped to the soybean genome by using Hisat2 (version 2.2.1) [80]. To measure the gene expression levels, we calculated the Fragments Per Kilobase of the

transcript sequence per Millions base pairs sequenced values [81]. DESeq2 was further used to analyze the gene expression profile and detect differentially expressed genes (DEGs) [82]. The criteria were  $|\log_2(\text{fold change})| > 1.0$  and corrected  $p$  value ( $p_{\text{adj}}$ )  $< 0.05$ .

#### 4.4. Protein Extraction and Digestion

Lysis and protein extraction in R217 and SN14 were performed using the lysis buffer SDT (4% SDS, 100 mM Tris-HCl, 1 mM DTT, pH 7.6). The BCA Protein Assay Kit (Bio-Rad, Hercules, CA, USA) was used to quantify the protein amount. The filter-aided sample preparation procedure described by Matthias Mann was used for performing trypsin-mediated protein digestion. The digested peptides of each sample were desalted on C18 cartridges (Empore™ SPE Cartridges C18 (standard density), volume 3 mL, bed I.D. 7 mm, Sigma, St. Louis, MO, USA), concentrated through vacuum centrifugation, and reconstituted in 40  $\mu\text{L}$  of 0.1% ( $v/v$ ) formic acid. The mixture of 20  $\mu\text{g}$  of each protein sample and 5 $\times$  loading buffer was boiled for 5 min. The proteins were separated using a 12.5% SDS-PAGE gel at a constant current of 14 mA for 90 min. Coomassie Blue R-250 staining was performed to visualize the protein bands.

#### 4.5. LC-MS/MS Analysis

The LC-MS/MS analysis was performed using the timsTOF Pro mass spectrometer (Bruker, Billerica, MA, USA), which was coupled with Nanoelute (Bruker Daltonics) for 60/120/240 min. A reverse phase trap column (Thermo Scientific, Waltham, MA, USA, Acclaim PepMap100, nanoViper C18, 100  $\mu\text{m} \times 2$  cm) was applied to load the peptides, and the reverse phase trap column was connected to the C18-reversed phase analytical column (Thermo Scientific Easy Column, inner diameter, 75  $\mu\text{m}$ ; length, 10 cm; resin, 3  $\mu\text{m}$ ) in 0.1% formic acid. The IntelliFlow technology was used to separate a linear gradient of the mixture of 0.1% formic acid and 84% acetonitrile at a flow rate of 300 nL/min. The mass spectrometer was used with the positive ion mode, and ion mobility MS spectra were collected over a mass range of  $m/z$  100–1700 and  $1/k_0$  of 0.6–1.6. The mass spectrometer operated 10 cycles of PASEF MS/MS with a target intensity of 1.5 k, and the threshold was set as 2500. Release occurred for 40 min under active exclusion.

#### 4.6. Proteomic Data Processing

The MS raw data for the R217 and SN14 samples were searched and combined using MaxQuant 1.5.3.17 software for the identification and quantitation analyses [83]. A maximum of two missed cleavages were allowed for tryptic peptides. The modification pattern, including carbamidomethyl (C) and oxidation (M), was added, and the MS/MS tolerance was set to 20 ppm. Proteins were annotated by searching the Uniprot\_Glycine database (downloaded from <https://www.uniprot.org/taxonomy/3847>, with 85,057 protein sequences, accessed on 4 September 2021), and the database pattern was set to Target-Reverse. The protein FDR  $\leq 0.01$  was required for positive protein hits. The intensity-based absolute quantification method and the label-free quantification method in MaxQuant 1.5.3.17 software were used for protein quantification [83]. The criterion of FC  $> 2$  or  $< 0.5$  ( $p < 0.05$ ) was established to identify differentially accumulated proteins (DAPs) in Perseus [84] and R (<https://www.r-project.org/>, accessed on 22 October 2022).

#### 4.7. Bioinformatics Analysis

Principal component analysis was performed to evaluate variations in DEG expression by using the stats package in R. The information on GO annotation was obtained from the database of SoyBase (<https://soybase.org>, accessed on 28 October 2022). The annotated information on DAPs was based on the UniProtKB/Swiss-Prot (<https://www.uniprot.org/>, accessed on 28 October 2022). The results of enrichment analyses of both DEGs and DAPs were from Gene Ontology (GO; <http://geneontology.org/>, accessed on 8 November 2022) and Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.kegg.jp/>, accessed on 8 November 2022) databases. All quantified proteins were considered the

background dataset, and the enrichment analyses were performed using Fisher's exact test. To further adjust the derived  $p$  values, the Benjamini–Hochberg correction for multiple testing was performed. The GO and KEGG enrichment analyses were performed using the ClusterProfiler (version 4.0.5) [85].

#### 4.8. Statistical Analysis

Significance analyses were performed using one-way ANOVA in SPSS Statistics17.0 software (International Business Machines Corporation, New York, NY, USA), and the threshold was set to  $p < 0.05$ . The box plots, volcano plots, and histograms were drawn using the “ggplot2” [86] and “ggpubr” (<http://mysql.orst.edu/pub/cran/web/packages/ggpubr/index.html>, accessed on 20 July 2023) packages in R version 4.0.2 (<https://www.r-project.org/>, accessed on 20 July 2023). The heatmaps were plotted using the “pheatmap” (<https://cran.r-project.org/web/packages/pheatmap/index.html>, accessed on 26 July 2023) in R version 4.0.2 (<https://www.r-project.org/>, accessed on 26 July 2023).

## 5. Conclusions

This study reveals the potential seed weight regulatory pathways in soybean, including the MAPK signaling pathway and cell wall biological processes (Figure 6; Table S3). CSSL R217 was selected from the CSSL population. The 100-seed weight, seed length, seed width, and seed weight per plant of R217 increased compared to SN14. Integrated analysis of the transcriptomic and proteomic results indicated that these DEGs and DAPs were primarily enriched in the MAPK signaling pathway and cell wall biosynthesis pathway, and 59 candidate genes associated with 100-seed weight were predicted. Moreover, the gene analysis on the instituted segments of R217 revealed 25 candidate genes associated with 100-seed weight located on these segments. In conclusion, a potential seed weight regulatory module was uncovered in this work, and 25 candidate genes were predicted to be involved in the two seed weight control pathways (Figure 5; Table S2). These results provide new insights into the molecular mechanisms of soybean seed weight.

**Supplementary Materials:** The following supporting information is included in the provided supplementary files: <https://www.mdpi.com/article/10.3390/plants13060908/s1>, Figure S1. The comparison of seed size between SN14 and R217 for 5 years. (A,B) The comparison of seed length and seed width between SN14 and R217 from 2018 to 2022. Different lowercase letters (a, b, c, d, and e) indicate significant differences, the threshold is set to  $p < 0.05$ , and identical letters suggest insignificant differences in one-way ANOVA; Figure S2. The GO enrichment analysis of top DEGs in R217 and SN14 at S15 and S20 seed developmental stages. (A–D) The GO enrichment analysis of the top down-regulated and upregulated DEGs at S15 and S20 seed developmental stage. Down—downregulated DEGs. Up—upregulated DEGs. S15 and S20 are the developmental stages of soybean seeds on the 15th and 20th days after fertilization. The vertical axis represents GO terms. The GO enrichment analysis information of DEGs were from Gene Ontology (GO; <http://geneontology.org/>, accessed on 8 November 2022) databases. The horizontal axis represents the gene ratio; Table S1. The physical positions of the R217 genome substituted segments based on the bin-map analysis results; Table S2. Expression and protein abundance data for candidate genes of 100-seed weight; Table S3. Expression and protein abundance data for candidate genes involved in pathways related to 100-seed weight.

**Author Contributions:** Conceptualization, J.W., J.Z., Q.C. and S.W.; methodology, J.W., J.Z., Q.C. and Z.Q.; investigation, S.W., F.D., Z.Y., J.Z. and F.C.; data curation, S.W., F.D., Z.Y., F.C., C.L. and M.Y.; writing—original draft preparation, S.W., F.D., Z.Y. and J.W.; writing—review and editing, S.W., F.D., J.Z., Z.Q. and J.W.; visualization, J.W., J.Z. and S.W.; supervision, J.W. and Z.Q.; funding acquisition, C.L. and J.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the National Natural Science Foundation of China (grant numbers: 32201809, 32272072, 32372106, and U20A2027) and the National Key R&D Program of China (2021YFF1001202).

**Data Availability Statement:** All data are available in the manuscript or the Supplementary Materials. The RNA-seq data have been deposited in the NCBI at the Sequence Read Archive (SRA) database

under the accession numbers of PRJNA1031595. The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>, the announcement date is 25 October 2023.) through the iProX partner repository with the data set identifier PXD046403. All plant materials reported in this study are available upon completion of the material transfer agreement.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## References

- Karikari, B.; Wang, Z.; Zhou, Y.; Yan, W.; Feng, J.; Zhao, T. Identification of quantitative trait nucleotides and candidate genes for soybean seed weight by multiple models of genome-wide association study. *BMC Plant Biol.* **2020**, *20*, 404. [[CrossRef](#)]
- Nguyen, C.X.; Paddock, K.J.; Zhang, Z.Y.; Stacey, M.G. GmKIX8-1 regulates organ size in soybean and is the causative gene for the major seed weight QTL qSw17-1. *New Phytol.* **2021**, *229*, 920–934. [[CrossRef](#)]
- Lu, X.; Xiong, Q.; Cheng, T.; Li, Q.T.; Liu, X.L.; Bi, Y.D.; Li, W.; Zhang, W.K.; Ma, B.; Lai, Y.C.; et al. A *PP2C-1* allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* **2017**, *10*, 670–684. [[CrossRef](#)]
- Duan, Z.B.; Zhang, M.; Zhang, Z.F.; Liang, S.; Fan, L.; Yang, X.; Yuan, Y.Q.; Pan, Y.; Zhou, G.A.; Liu, S.L.; et al. Natural allelic variation of *GmST05* controlling seed size and quality in soybean. *Plant Biotechnol. J.* **2022**, *20*, 1807–1818. [[CrossRef](#)]
- Li, J.; Zhang, Y.H.; Ma, R.R.; Huang, W.X.; Hou, J.J.; Fang, C.; Wang, L.S.; Yuan, Z.H.; Sun, Q.; Dong, X.H.; et al. Identification of *ST1* reveals a selection involving hitchhiking of seed morphology and oil content during soybean domestication. *Plant Biotechnol. J.* **2022**, *20*, 1110–1121. [[CrossRef](#)]
- Cai, Z.; Xian, P.; Cheng, Y.; Zhong, Y.; Yang, Y.; Zhou, Q.; Lian, T.; Ma, Q.; Nian, H.; Ge, L. *MOTHER-OF-FT-AND-TFL1* regulates the seed oil and protein content in soybean. *New Phytol.* **2023**, *239*, 905–919. [[CrossRef](#)]
- Wei, S.M.; Yong, B.; Jiang, H.W.; Wang, Y.; Li, B.B.; Yang, C.; Zhu, W.W.; An, Z.H.; Chen, Q.S.; He, C.Y. A loss-of-function mutant allele of a glycosyl hydrolase gene has been co-opted for seed weight control during soybean domestication. *J. Integr. Plant Biol.* **2023**, *65*, 2469–2689. [[CrossRef](#)]
- Wang, S.D.; Liu, S.L.; Wang, J.; Yokosho, K.; Zhou, B.; Yu, Y.C.; Liu, Z.; Frommer, W.B.; Ma, J.F.; Chen, L.Q.; et al. Simultaneous changes in seed size, oil content and protein content driven by selection of *SWEET* homologues during soybean domestication. *Natl. Sci. Rev.* **2020**, *7*, 1776–1786. [[CrossRef](#)]
- Zhu, W.W.; Yang, C.; Yong, B.; Wang, Y.; Li, B.B.; Gu, Y.Z.; Wei, S.M.; An, Z.H.; Sun, W.K.; Qiu, L.J.; et al. An enhancing effect attributed to a nonsynonymous mutation in *SOYBEAN SEED SIZE 1*, a *SPINDLY*-like gene, is exploited in soybean domestication and improvement. *New Phytol.* **2022**, *236*, 1375–1392. [[CrossRef](#)]
- Gu, Y.Z.; Li, W.; Jiang, H.W.; Wang, Y.; Gao, H.H.; Liu, M.; Chen, Q.S.; Lai, Y.C.; He, C.Y. Differential expression of a *WRKY* gene between wild and cultivated soybeans correlates to seed size. *J. Exp. Bot.* **2017**, *68*, 2717–2729. [[CrossRef](#)]
- Du, D.; Gao, X.; Geng, J.; Li, Q.; Li, L.; Lv, Q.; Li, X. Identification of key proteins and networks related to grain development in wheat (*Triticum aestivum* L.) by Comparative Transcription and Proteomic Analysis of Allelic Variants in TaGW2-6A. *Front. Plant Sci.* **2016**, *7*, 922. [[CrossRef](#)]
- Gao, H.; Wang, Y.; Li, W.; Gu, Y.; Lai, Y.; Bi, Y.; He, C. Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. *J. Exp. Bot.* **2018**, *69*, 5089–5104. [[CrossRef](#)]
- Ma, C.; Zhou, J.; Chen, G.; Bian, Y.; Lv, D.; Li, X.; Wang, Z.; Yan, Y. iTRAQ-based quantitative proteome and phosphoprotein characterization reveals the central metabolism changes involved in wheat grain development. *BMC Genom.* **2014**, *15*, 1029. [[CrossRef](#)]
- Yu, T.; Li, G.; Dong, S.; Liu, P.; Zhang, J.; Zhao, B. Proteomic analysis of maize grain development using iTRAQ reveals temporal programs of diverse metabolic processes. *BMC Plant Biol.* **2016**, *16*, 241. [[CrossRef](#)]
- Zhang, L.; Dong, Y.; Wang, Q.; Du, C.; Xiong, W.; Li, X.; Zhu, S.; Li, Y. iTRAQ-based proteomics analysis and network integration for kernel tissue development in maize. *Int. J. Mol. Sci.* **2017**, *18*, 1840. [[CrossRef](#)]
- Zeng, D.; Cui, J.; Yin, Y.; Zhang, M.; Shan, S.; Liu, M.Y.; Cheng, D.; Lu, W.; Sun, Y. Proteomic analysis in different development stages on SP0 generation of rice seeds after space flight. *Life Sci. Space Res.* **2020**, *26*, 34–45. [[CrossRef](#)]
- Qing, D.; Zhou, Y.; Pan, Y.; Yang, X.; Li, J.; Zhou, W.; Liang, H.; Chen, W.; Chen, L.; Lu, C.; et al. TMT-based quantitative proteomic analysis of indica rice cultivars reveals that novel components of the signaling pathways might play a role in grain length regulation. *J. Proteom.* **2023**, *270*, 104745. [[CrossRef](#)]
- Han, X.; Li, J.; Zhao, Y.; Zhang, Z.; Jiang, H.; Wang, J.; Feng, X.; Zhang, Y.; Du, Z.; Wu, X.; et al. Integrated transcriptomic and proteomic characterization of a chromosome segment substitution line reveals a new regulatory network controlling the seed storage profile of soybean. *Food Energy Secur.* **2022**, *11*, e381. [[CrossRef](#)]
- Jin, X.; Zhai, H.; Wang, P.; Zhang, X.; Wu, X.; Zhang, H.; Chen, S.; Huang, Z. Physiological and omics analysis of maize inbred lines during late grain development. *Genes Genom.* **2022**, *44*, 993–1006. [[CrossRef](#)]
- Li, N.; Xu, R.; Li, Y.H. Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* **2019**, *70*, 435–463. [[CrossRef](#)]
- Zhang, J.; Zhang, X.; Liu, X.; Pai, Q.; Wang, Y.; Wu, X. Molecular network for regulation of seed size in plants. *Int. J. Mol. Sci.* **2023**, *24*, 10666. [[CrossRef](#)]

22. Duan, P.; Rao, Y.; Zeng, D.; Yang, Y.; Xu, R.; Zhang, B.; Dong, G.; Qian, Q.; Li, Y. *SMALL GRAIN 1*, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. *Plant J.* **2014**, *77*, 547–557. [[CrossRef](#)]
23. Liu, S.; Hua, L.; Dong, S.; Chen, H.; Zhu, X.; Jiang, J.; Zhang, F.; Li, Y.; Fang, X.; Chen, F. OsMAPK6, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J.* **2015**, *84*, 672–681. [[CrossRef](#)]
24. Xu, R.; Duan, P.; Yu, H.; Zhou, Z.; Zhang, B.; Wang, R.; Li, J.; Zhang, G.; Zhuang, S.; Lyu, J.; et al. Control of grain size and weight by the OsMKKK10-OsMKK4-OsMAPK6 Signaling pathway in rice. *Mol. Plant* **2018**, *11*, 860–873. [[CrossRef](#)]
25. Xu, R.; Yu, H.; Wang, J.; Duan, P.; Zhang, B.; Li, J.; Li, Y.; Xu, J.; Lyu, J.; Li, N.; et al. A mitogen-activated protein kinase phosphatase influences grain size and weight in rice. *Plant J.* **2018**, *95*, 937–946. [[CrossRef](#)]
26. Zhang, M.; Wu, H.; Su, J.; Wang, H.; Zhu, Q.; Liu, Y.; Xu, J.; Lukowitz, W.; Zhang, S. Maternal control of embryogenesis by MPK6 and its upstream MKK4/MKK5 in *Arabidopsis*. *Plant J.* **2017**, *92*, 1005–1019. [[CrossRef](#)]
27. Guo, T.; Chen, K.; Dong, N.Q.; Shi, C.L.; Ye, W.W.; Gao, J.P.; Shan, J.X.; Lin, H.X. *GRAIN SIZE AND NUMBER<sub>1</sub>* negatively regulates the OsMKKK10-OsMKK4-OsMPK6 cascade to coordinate the trade-off between grain number per panicle and grain size in rice. *Plant Cell* **2018**, *30*, 871–888. [[CrossRef](#)]
28. Wu, X.; Cai, X.; Zhang, B.; Wu, S.; Wang, R.; Li, N.; Li, Y.; Sun, Y.; Tang, W. *ERECTA* regulates seed size independently of its intracellular domain via MAPK-DA1-UBP15 signaling. *Plant Cell* **2022**, *34*, 3773–3789. [[CrossRef](#)]
29. Xiao, W.; Hu, S.; Zou, X.; Cai, R.; Liao, R.; Lin, X.; Yao, R.; Guo, X. Lectin receptor-like kinase LecRK-VIII.2 is a missing link in MAPK signaling-mediated yield control. *Plant Physiol.* **2021**, *187*, 303–320. [[CrossRef](#)] [[PubMed](#)]
30. Tian, X.; Li, X.; Zhou, W.; Ren, Y.; Wang, Z.; Liu, Z.; Tang, J.; Tong, H.; Fang, J.; Bu, Q. Transcription factor OsWRKY53 positively regulates brassinosteroid signaling and plant architecture. *Plant Physiol.* **2017**, *175*, 1337–1349. [[CrossRef](#)]
31. Liu, Z.; Mei, E.; Tian, X.; He, M.; Tang, J.; Xu, M.; Liu, J.; Song, L.; Li, X.; Wang, Z.; et al. OsMKKK70 regulates grain size and leaf angle in rice through the OsMKK4-OsMAPK6-OsWRKY53 signaling pathway. *J. Integr. Plant Biol.* **2021**, *63*, 2043–2057. [[CrossRef](#)]
32. Baez, L.A.; Tichá, T.; Hamann, T. Cell wall integrity regulation across plant species. *Plant Mol. Biol.* **2022**, *109*, 483–504. [[CrossRef](#)]
33. Gigli-Bisceglia, N.; Engelsdorf, T.; Hamann, T. Plant cell wall integrity maintenance in model plants and crop species-relevant cell wall components and underlying guiding principles. *Cell Mol. Life Sci.* **2020**, *77*, 2049–2077. [[CrossRef](#)] [[PubMed](#)]
34. Iwai, H. Virtual issue: Cell wall functions in plant growth and environmental responses. *J. Plant Res.* **2021**, *134*, 1155–1158. [[CrossRef](#)] [[PubMed](#)]
35. Di Marzo, M.; Viana, V.E.; Banfi, C.; Cassina, V.; Corti, R.; Herrera-Ubaldo, H.; Babolin, N.; Guazzotti, A.; Kiegle, E.; Gregis, V.; et al. Cell wall modifications by  $\alpha$ -XYLOSIDASE1 are required for control of seed and fruit size in *Arabidopsis*. *J. Exp. Bot.* **2022**, *73*, 1499–1515. [[CrossRef](#)]
36. Li, B.; Liu, H.; Zhang, Y.; Kang, T.; Zhang, L.; Tong, J.; Xiao, L.; Zhang, H. Constitutive expression of cell wall invertase genes increases grain yield and starch content in maize. *Plant Biotechnol. J.* **2013**, *11*, 1080–1091. [[CrossRef](#)] [[PubMed](#)]
37. Yang, X.; Wilkinson, L.G.; Aubert, M.K.; Houston, K.; Shirley, N.J.; Tucker, M.R. Ovule cell wall composition is a maternal determinant of grain size in barley. *New Phytol.* **2023**, *237*, 2136–2147. [[CrossRef](#)]
38. Mira, J.P.; Arenas-M, A.; Calderini, D.F.; Canales, J. Integrated transcriptome analysis identified key expansin genes associated with wheat cell wall, grain weight and yield. *Plants* **2023**, *12*, 2868. [[CrossRef](#)]
39. Jin, Y.; Ni, D.A.; Ruan, Y.L. Posttranslational elevation of cell wall invertase activity by silencing its inhibitor in tomato delays leaf senescence and increases seed weight and fruit hexose level. *Plant Cell* **2009**, *21*, 2072–2089. [[CrossRef](#)]
40. Ishimaru, K.; Hirotsu, N.; Madoka, Y.; Murakami, N.; Hara, N.; Onodera, H.; Kashiwagi, T.; Ujiie, K.; Shimizu, B.; Onishi, A.; et al. Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* **2013**, *45*, 707–711. [[CrossRef](#)]
41. Sestili, F.; Pagliarello, R.; Zega, A.; Saletti, R.; Pucci, A.; Botticella, E.; Masci, S.; Tundo, S.; Moschetti, I.; Foti, S.; et al. Enhancing grain size in durum wheat using RNAi to knockdown *GW2* genes. *Theor. Appl. Genet.* **2019**, *132*, 419–429. [[CrossRef](#)]
42. Wang, S.; Yang, H.; Mei, J.; Liu, X.; Wen, Z.; Zhang, L.; Xu, Z.; Zhang, B.; Zhou, Y. Rice homeobox protein *KNAT7* integrates the pathways regulating cell expansion and wall stiffness. *Plant Physiol.* **2019**, *181*, 669–682. [[CrossRef](#)]
43. Dong, N.Q.; Sun, Y.; Guo, T.; Shi, C.L.; Zhang, Y.M.; Kan, Y.; Xiang, Y.H.; Zhang, H.; Yang, Y.B.; Li, Y.C.; et al. UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. *Nat. Commun.* **2020**, *11*, 2629. [[CrossRef](#)] [[PubMed](#)]
44. Hu, M.J.; Zhang, H.P.; Cao, J.J.; Zhu, X.F.; Wang, S.X.; Jiang, H.; Wu, Z.Y.; Lu, J.; Chang, C.; Sun, G.L.; et al. Characterization of an IAA-glucose hydrolase gene *TaTGW6* associated with grain weight in common wheat (*Triticum aestivum* L.). *Mol. Breed.* **2016**, *36*, 25. [[CrossRef](#)]
45. Terao, A.; Hyodo, H.; Satoh, S.; Iwai, H. Changes in the distribution of cell wall polysaccharides in early fruit pericarp and ovule, from fruit set to early fruit development, in tomato (*Solanum lycopersicum*). *J. Plant Res.* **2013**, *126*, 719–728. [[CrossRef](#)] [[PubMed](#)]
46. Wang, F.; Sun, X.; Liu, B.; Kong, F.; Pan, X.; Zhang, H. A polygalacturonase gene *PG031* regulates seed coat permeability with a pleiotropic effect on seed weight in soybean. *Theor. Appl. Genet.* **2022**, *135*, 1603–1618. [[CrossRef](#)]
47. Eshed, Y.; Zamir, D. A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica* **1994**, *79*, 175–179. [[CrossRef](#)]
48. Pillen, K.; Zacharias, A.; Léon, J. Advanced backcross QTL analysis in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *107*, 340–352. [[CrossRef](#)]

49. Shen, G.J.; Xing, Y.Z. Two novel QTLs for heading date are identified using a set of chromosome segment substitution lines in rice (*Oryza sativa* L.). *J. Genet. Genom.* **2014**, *41*, 659–662. [[CrossRef](#)]
50. Zhu, D.; Li, X.M.; Wang, Z.M.; You, C.Y.; Nie, X.H.; Sun, J.; Zhang, X.L.; Zhang, D.W.; Lin, Z.X. Genetic dissection of an allotetraploid interspecific CSSLs guides interspecific genetics and breeding in cotton. *BMC Genom.* **2020**, *21*, 431. [[CrossRef](#)]
51. Xin, D.W.; Qi, Z.M.; Jiang, H.W.; Hu, Z.B.; Zhu, R.S.; Hu, J.H.; Han, H.Y.; Hu, G.H.; Liu, C.Y.; Chen, Q.S. QTL Location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. *PLoS ONE* **2016**, *11*, e0149380.
52. Zheng, H.Y.; Hou, L.L.; Xie, J.G.; Cao, F.B.; Wei, R.; Yang, M.L.; Qi, Z.M.; Zhu, R.S.; Zhang, Z.G.; Xin, D.W.; et al. Construction of chromosome segment substitution lines and inheritance of seed-pod characteristics in wild soybean. *Front. Plant Sci.* **2022**, *13*, 869455. [[CrossRef](#)]
53. Clemente, T.E.; Cahoon, E.B. Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiol.* **2009**, *151*, 1030–1040. [[CrossRef](#)] [[PubMed](#)]
54. Liu, G.M.; Li, W.T.; Zeng, R.Z.; Zhang, Z.M.; Zhang, G.Q. Identification of QTLs on substituted segments in single segment substitution lines of rice. *J. Genet. Genom.* **2004**, *31*, 1395–1400.
55. Thomson, M. Substitution mapping of *dth1.1*, a flowering-time quantitative trait locus (QTL) associated with transgressive variation in rice, reveals multiple sub-QTL. *Genetics* **2006**, *172*, 2501–2514. [[CrossRef](#)] [[PubMed](#)]
56. Eshed, Y.; Zamir, D. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* **1995**, *141*, 1147–1162. [[CrossRef](#)]
57. Yang, H.Y.; Wang, W.B.; He, Q.Y.; Xiang, S.H.; Gai, J.Y. Identifying a wild allele conferring small seed size, high protein content and low oil content using chromosome segment substitution lines in soybean. *Theor. Appl. Genet.* **2019**, *132*, 2793–2807. [[CrossRef](#)]
58. Deshmukh, R.; Sonah, H.; Patil, G.; Chen, W.; Prince, S.; Mutava, R.; Vuong, T.; Valliyodan, B.; Nguyen, H.T. Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* **2014**, *5*, 244. [[CrossRef](#)]
59. Peng, Z.; He, S.; Gong, W.; Xu, F.; Pan, Z.; Jia, Y.; Geng, X.; Du, X. Integration of proteomic and transcriptomic profiles reveals multiple levels of genetic regulation of salt tolerance in cotton. *BMC Plant Biol.* **2018**, *18*, 128. [[CrossRef](#)] [[PubMed](#)]
60. Shah, T.; Xu, J.; Zou, X.; Cheng, Y.; Nasir, M.; Zhang, X. Omics approaches for engineering wheat production under abiotic stresses. *Int. J. Mol. Sci.* **2018**, *19*, 2390. [[CrossRef](#)]
61. Xu, Y.; Zhao, Y.; Wang, X.; Ma, Y.; Li, P.; Yang, Z.; Zhang, X.; Xu, C.; Xu, S. Incorporation of parental phenotypic data into multi-omic models improves prediction of yield-related traits in hybrid rice. *Plant Biotechnol. J.* **2021**, *19*, 261–272. [[CrossRef](#)]
62. Zhou, Q.; Fu, Z.; Liu, H.; Wang, J.; Guo, Z.; Zhang, X.; Tian, R.; Liu, Y.; Qu, J.; Li, W.; et al. Mining novel kernel size-related genes by pQTL mapping and multi-omics integrative analysis in developing maize kernels. *Plant Biotechnol. J.* **2021**, *19*, 1489–1491. [[CrossRef](#)]
63. Liu, Y.; Yi, C.; Liu, Q.; Wang, C.; Wang, W.; Han, F.; Hu, X. Multi-omics profiling identifies candidate genes controlling seed size in peanut. *Plants* **2022**, *11*, 3276. [[CrossRef](#)]
64. Wang, Z.; Tao, S.; Liu, S.; Jia, M.; Cui, D.; Sun, G.; Deng, Z.; Wang, F.; Kong, X.; Fu, M.; et al. A multi-omics approach for rapid identification of large genomic lesions at the wheat dense spike (wds) locus. *Front. Plant Sci.* **2022**, *13*, 850302. [[CrossRef](#)] [[PubMed](#)]
65. Cai, Q.; Jiao, F.; Wang, Q.; Zhang, E.; Song, X.; Pei, Y.; Li, J.; Zhao, M.; Guo, X. Multiomics comparative analysis of the maize large grain mutant *tc19* identified pathways related to kernel development. *BMC Genom.* **2023**, *24*, 537. [[CrossRef](#)] [[PubMed](#)]
66. Wang, D.; Wang, Y.; Zhang, L.; Yang, Y.; Wu, Q.; Hu, G.; Wang, W.; Li, J.; Huang, Z. Integrated transcriptomic and proteomic analysis of a cytoplasmic male sterility line and associated maintainer line in soybean. *Front. Plant Sci.* **2023**, *14*, 1098125. [[CrossRef](#)]
67. Li, J.; Chen, L.; Zhi, X.; Wang, J.; Lu, Y.; Tian, Z.; Wu, M.; Shan, Y.; Chen, H.; Liao, W.; et al. Integrated transcriptome and proteome analysis reveals molecular responses of soybean anther under high-temperature stress. *Front. Plant Sci.* **2023**, *14*, 1187922. [[CrossRef](#)]
68. Zhang, M.; Zhang, S. Mitogen-activated protein kinase cascades in plant signaling. *J. Integr. Plant Biol.* **2022**, *64*, 301–341. [[CrossRef](#)] [[PubMed](#)]
69. Xu, J.; Zhang, S. Mitogen-activated protein kinase cascades in signaling plant growth and development. *Trends Plant Sci.* **2015**, *20*, 56–64. [[CrossRef](#)]
70. Guo, T.; Lu, Z.Q.; Shan, J.X.; Ye, W.W.; Dong, N.Q.; Lin, H.X. *ERECTA1* acts upstream of the OsMKKK10-OsMKK4-OsMPK6 cascade to control spikelet number by regulating cytokinin metabolism in rice. *Plant Cell* **2020**, *32*, 2763–2779. [[CrossRef](#)] [[PubMed](#)]
71. Mao, G.; Meng, X.; Liu, Y.; Zheng, Z.; Chen, Z.; Zhang, S. Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in *Arabidopsis*. *Plant Cell* **2011**, *23*, 1639–1653. [[CrossRef](#)]
72. Du, L.; Li, N.; Chen, L.; Xu, Y.; Li, Y.; Zhang, Y.; Li, C.; Li, Y. The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in *Arabidopsis*. *Plant Cell* **2014**, *26*, 665–677. [[CrossRef](#)]
73. Dong, H.; Dumenil, J.; Lu, F.H.; Na, L.; Vanhaeren, H.; Naumann, C.; Klecker, M.; Prior, R.; Smith, C.; McKenzie, N.; et al. Ubiquitylation activates a peptidase that promotes cleavage and destabilization of its activating E3 ligases and diverse growth regulatory proteins to limit cell proliferation in *Arabidopsis*. *Genes Dev.* **2017**, *31*, 197–208. [[CrossRef](#)]
74. Kumar, M.; Campbell, L.; Turner, S. Secondary cell walls: Biosynthesis and manipulation. *J. Exp. Bot.* **2016**, *67*, 515–531. [[CrossRef](#)]
75. Zhong, R.; Ye, Z.H. Regulation of cell wall biosynthesis. *Curr. Opin. Plant Biol.* **2007**, *10*, 564–752. [[CrossRef](#)]

76. Lin, F.; Manisseri, C.; Fagerström, A.; Peck, M.L.; Vega-Sánchez, M.E.; Williams, B.; Chiniquy, D.M.; Saha, P.; Pattathil, S.; Conlin, B.; et al. Cell wall composition and candidate biosynthesis gene expression during rice development. *Plant Cell Physiol.* **2016**, *57*, 2058–2075. [[CrossRef](#)]
77. Zhang, Y.M.; Yu, H.X.; Ye, W.W.; Shan, J.X.; Dong, N.Q.; Guo, T.; Kan, Y.; Xiang, Y.H.; Zhang, H.; Yang, Y.B.; et al. A rice QTL *GS3.1* regulates grain size through metabolic-flux distribution between flavonoid and lignin metabolons without affecting stress tolerance. *Commun. Biol.* **2021**, *4*, 1171. [[CrossRef](#)] [[PubMed](#)]
78. Kohlberger, M.; Thalhamer, T.; Weiss, R.; Tenhaken, R. *Arabidopsis* MAP-Kinase 3 phosphorylates UDP-glucose dehydrogenase: A key enzyme providing UDP-sugar for cell wall biosynthesis. *Plant Mol. Biol. Rep.* **2018**, *36*, 870–877. [[CrossRef](#)] [[PubMed](#)]
79. Liu, Y.; Li, Y. MPK4 negatively regulates the l-arabinose synthesis of cell wall in *Arabidopsis*. *Biochem. Biophys. Res. Commun.* **2022**, *613*, 7–11. [[CrossRef](#)] [[PubMed](#)]
80. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
81. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
82. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
83. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [[CrossRef](#)]
84. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. Methods* **2016**, *13*, 731–740. [[CrossRef](#)]
85. Wu, T.Z.; Hu, E.Q.; Xu, S.B.; Chen, M.J.; Guo, P.F.; Dai, Z.Z.; Feng, T.Z.; Zhou, L.; Tang, W.L.; Zhan, L.; et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2021**, *2*, 100141. [[CrossRef](#)] [[PubMed](#)]
86. Galensia, K. ggplot2: Elegant graphics for data analysis (2nd ed.). *Comput. Rev.* **2017**, *58*, 457–458.

**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.