

Article

Identification and Candidate Gene Evaluation of a Large Fast Neutron-Induced Deletion Associated with a High-Oil Phenotype in Soybean Seeds

William R. Serson ^{1,*}, Mohammad Fazel Soltani Gishini ², Robert M. Stupar ³, Adrian O. Stec ³, Paul R. Armstrong ⁴ and David Hildebrand ⁵

¹ Department of Biology, Penn State University, Lehigh Valley, Center Valley, PA 18034, USA

² Department of Plant Pathology, University of Kentucky, Lexington, KY 40546, USA; mohammad.soltani@uky.edu

³ Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA; stup0004@umn.edu (R.M.S.); stecx002@umn.edu (A.O.S.); paul.armstrong@usda.gov (P.R.A.)

⁴ United States Department of Agriculture-Agricultural Research Service, Manhattan, KS 66502, USA

⁵ Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY 40546, USA; dhild@uky.edu

* Correspondence: wrs5272@psu.edu

Abstract: Since the dawn of agriculture, crops have been genetically altered for desirable characteristics. This has included the selection of natural and induced mutants. Increasing the production of plant oils such as soybean (*Glycine max*) oil as a renewable resource for food and fuel is valuable. Successful breeding for higher oil levels in soybeans, however, usually results in reduced seed protein. A soybean fast neutron population was screened for oil content, and three high oil mutants with minimal reductions in protein levels were found. Three backcross F2 populations derived from these mutants exhibited segregation for seed oil content. DNA was pooled from the high-oil and normal-oil plants within each population and assessed by comparative genomic hybridization. A deletion encompassing 20 gene models on chromosome 14 was found to co-segregate with the high-oil trait in two of the three populations. Eighteen genes in the deleted region have known functions that appear unrelated to oil biosynthesis and accumulation pathways, while one of the unknown genes (*Glyma.14G101900*) may contribute to the regulation of lipid droplet formation. This high-oil trait can facilitate the breeding of high-oil soybeans without protein reduction, resulting in higher meal protein levels.

Keywords: fast neutron mutagenesis; renewable oil; triacylglyceride; comparative genomics hybridization



Citation: Serson, W.R.; Gishini, M.F.S.; Stupar, R.M.; Stec, A.O.; Armstrong, P.R.; Hildebrand, D. Identification and Candidate Gene Evaluation of a Large Fast Neutron-Induced Deletion Associated with a High-Oil Phenotype in Soybean Seeds. *Genes* **2024**, *15*, 892. <https://doi.org/10.3390/genes15070892>

Academic Editors: Prakit Somta and Xingxing Yuan

Received: 6 June 2024

Revised: 29 June 2024

Accepted: 3 July 2024

Published: 8 July 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

There is a strong genetic component that is well understood for oil content and quality via the oil biosynthetic pathway and its regulation. Seed oil is biosynthesized during the second main stage of seed maturation [1–3], at which time the relevant biosynthetic enzymes are highly expressed. For instance, in studies of expression profiles of triacylglyceride (TAG) biosynthetic enzymes and oil accumulation in developing soybeans (*Glycine max*), *DGAT1* shows an expression profile suggesting a dominant role in soybean oil biosynthesis, but *DGAT2* and *PDAT* do not [4–6].

It is becoming increasingly possible to alter hydrocarbon flux in soybeans. Multiple studies indicate that oil content increases with a higher expression of TAG biosynthetic genes [7–9]. Increased expression of regulatory genes that upregulate multiple enzymes for fatty acid biosynthesis also can result in higher oil levels [9]. Co-expression of the transcription factor *WRI1* with *DGAT1*, a key rate-limiting enzyme, is shown to have a synergistic effect on TAG biosynthesis in plants [10–12]. Overall, increasing sink strength

results in increased oil and protein, with a strong pronounced effect on protein and less on oils [13].

Cultivar effects are well understood to affect protein content and amino acids in soybeans, most likely due to heritable differences in TAG biosynthetic genes and regulatory factors. For instance, germplasm line N6202 produced seeds with 45.7% protein content and a 10% reduction in grain yield compared to a control variety, NC-Roy [14]. In contrast, TN03-350 and TN04-5321 achieved 43.1–43.9% protein content without sacrificing seed yield. Considering the importance of grain yields in commercial varieties, such a result is more desirable than a decrease in yield with improved grain quality. In soybean genotypes of early maturity groups, average-to-high protein content ($399\text{--}476\text{ g/kg}^{-1}$) was found in years with high air temperature and moderate rates of rainfall during the seed-filling period, whereas seed protein content was drastically reduced ($265\text{--}347\text{ g/kg}^{-1}$) in seasons of insufficient nitrogen fixation or higher amounts of precipitation during seed filling [15].

In plant breeding, random mutagenesis is a common way to generate mutations and increase genetic diversity for traits with limited natural variation. Common examples include the use of a chemical mutagen like ethyl methane sulfonate (EMS) or bombardment with gamma radiation, and today even site-directed mutagenesis is possible [16]. While these methods can produce useful traits, such as an early flowering mutant or increased oil content, it is possible that other important genes could be mutated, causing undesirable phenotypes, such as altered seed composition. However, mutagenized populations are not brought under the same scrutiny as transgenic approaches, and therefore traits induced this way are much easier to incorporate into the existing germplasm.

While utilization of mutagens can introduce much-needed levels of genetic and phenotypic diversity, it is imperative to understand the nature of the mutations induced. The mutagen used in this study, fast neutron bombardment (FN), typically induces deletion and/or chromosomal rearrangement events from several base pairs-long to several megabases [17–19]. The phenotypic variation can be used to study the association of genes with specific traits [20] or as a source of new variation for breeding purposes [21–23]. Comparative genomic hybridization (CGH) is one of the fastest and most effective ways to assess duplications or deletions caused by irradiation mutagenesis, such as FN. This technology utilizes oligonucleotide probes affixed to slides, known as microarrays. Fluorescently labeled DNA samples can then be hybridized to the probes, thus emitting a fluorescence intensity proportional to the DNA sample copy number for each probe sequence. The mutant DNA sample can be compared to a control sample (the non-mutagenized parent DNA) to identify tracts of sequence that have been duplicated or deleted in the mutant compared to its parent line [17,18,24]. In the case of the current study, the soybean microarray consists of 700,000+ features, allowing for a probe spacing of approximately 1000 base pairs across the euchromatic portion of the genome. With the advances of next generation sequencing and microarray technology, combined with intimate knowledge of the soybean genome, we can now harness CGH technology to quickly and precisely assess copy number variants (CNVs) in segregating mutant backcrosses.

Here, we utilized a fast neutron population of soybeans which exhibit 3-to-4% higher oil content than the parent variety with only a minor decrease in protein, resulting in seeds with increased oil plus protein content compared to the parent varieties. We hypothesize that the loss of a specific gene or genes within the deleted 300 kb region on chromosome 14 is responsible for the high-oil phenotype. Our analysis provides new insight into which of these genes may be the most likely to cause this change and may be the best candidate for future functional analyses.

2. Materials and Methods

2.1. Genetic Material

Seeds from three mutant lines, 1R22C28Cgadbr355aMN13, 5R12C21Dar387dMN13, and 5R16C01Dar388eMN13, and their parent, M92-220, from the University of Minnesota, were requested from their fast neutron (FN) mutant library in April 2014 and planted in

the field in Lexington, KY, during the summer of 2014 to verify high-oil content in that environment using a single-line, randomized complete block design with border seed of the “Jack” cultivar. Ten plants were in each line, with three total replications, with each plant individually bulked and analyzed via non-destructive NIRS [25–27].

M92-220 is a maturity group “I” soybean produced by the University of Minnesota soybean breeding program. It is derived from the cultivar “MN1302” (PI 616498), which was originally selected from a cross between “Hendricks” × “Archer” [28]. M92-220 exhibits an indeterminate growth habit, purple flowers, grey pubescence, brown pods, yellow seeds, and a buff hilum.

2.2. Backcrossing

Two mutant lines (1R22C28Cgadbr355aMN13 and 5R12C21Dar387dMN13) in the M8 and M5 generations, respectively, exhibited high oil content in the KY environment and were thus planted and backcrossed to the parent line M92-220 in 2015, similar to other methodologies [18]. Successful crosses were harvested, and eleven F1 crosses were assessed for oil content via single-seed NMR. These were then grown in a greenhouse in the winter of 2015–2016 and in the spring of 2016. Thirty F2 seeds from each of the F1 plants were assessed for single-seed oil content and then planted at Spindletop Farm in Lexington, KY. In addition, 30 other randomly selected seeds were planted for each line, bringing the total to 60 total F2 seeds from F1 plants planted for each line.

2.3. Tissue Sampling, Seed Composition, and DNA Extraction

Leaf tissue was collected and frozen at $-80\text{ }^{\circ}\text{C}$ [29] for comparative genomic hybridization (CGH) analysis from F2 plants. Seeds were harvested from each plant in the fall, and each F2 plant’s F3 seed was analyzed in bulk with a Perten DA7200 NIRS for oil, protein, and moisture content. These data were used to generate scatter plots of the F2 sibs. Strong inverse correlations of oil and protein content suggested that segregations of the mutant trait were found in three F2 populations.

2.4. NMR Methods

Oil content was determined by single-seed NMR, using a Minspec 20 (Bruker Biospin, The Woodlands, TX, USA) [25]. The instrument accommodates a 20 mm sampling-tube diameter. Seeds were weighed and placed into the tube and allowed to warm to $40\text{ }^{\circ}\text{C}$ before insertion into the instrument. The standard oil seed measurement procedure supplied with the instrument controller was used. Calibration of the NMR instrument was performed using weighed amounts of extracted soybean oil encompassing the range of oil weights of the seed samples. Four different weights were used and were expressed on tissue paper at the bottom of the 20 mm sample tube.

2.5. CGH Analysis

DNA was extracted on a per-plant basis using a QIAGEN Plant DNeasy Kit from the leaves of three F2 populations (known here as A1, A2, and A3) expected to segregate for seed-oil content. The F2 plants that gave rise to F2:3 seeds with the highest seed oil (“high-oil” plants) and lowest seed oil (“normal-oil” plants) were respectively identified using NIRS. The DNA from the high-oil F2 plants was bulked, and the DNA for the normal-oil F2 plants was bulked for each of the three populations. The bulked DNAs were then subjected to CGH analyses, as previously described [17], using a custom NimbleGen CGH microarray with over approximately 700,000 probes, approximately 1 probe every 1 kb [17]. The CGH probe positions were designed according to the soybean cultivar Williams 82 genome version 2 assembly (Wm82.a2.v1). Each CGH was performed as a comparison within each population (e.g., A1 high-oil versus A1 normal-oil, etc.) From these comparisons, a large deletion was detected on chromosome 14, and the genes in that region (according to the Williams 82 version Wm82.a2.v1 gene annotation set) were analyzed on soybase.org and cross-referenced to homologous genes in the TAIR database.

2.6. Functional Analysis

SignalP-6.0 (<https://services.healthtech.dtu.dk/services/SignalP-6.0/>, accessed on 1 January 2024) was used to investigate possible signal peptides. The topology analysis was performed by Protter (<https://wlab.ethz.ch/protter/start/>, accessed on 1 January 2024) [30], and the protein network was established using the STRING database (<https://string-db.org>, accessed on 1 January 2024) [31]. Interactions in STRING are derived from genomic-context predictions, high-throughput lab experiments, co-expression, automated text-mining, and previous knowledge in databases. Phyre2, also known as Protein Fold Recognition Server, is a web portal for protein modeling, prediction, and analysis (<http://www.sbg.bio.ic.ac.uk>, accessed on 1 January 2024) [32]. Phyre2 was used to study the function of unknown genes in this study, and Chimera software version 1.11.2 was used to visualize the structural protein models [33].

3. Results

3.1. Oil and Protein Content of Backcrosses

Three mutant lines were selected for analysis based on previously observed high seed-oil content when grown in MN. The full names of these lines (1R22C28Cgadbr355aMN13, 5R12C21Dar387dMN13, and 5R16C01Dar388eMN13) are abbreviated herein as “1R22”, “5R12”, and “5R16” for simplicity. When grown in KY in 2015, these lines again showed a higher mean seed oil content, though not as pronounced compared to their wild-type parent line, M92-220, as was previously observed in MN (Table 1). We attempted to backcross the mutants to M92-220, resulting in eleven successful crosses. However, successful crosses were only observed for two of the mutants, 1R22 and 5R12.

Table 1. Mean oil content (on a dry weight basis) and the standard error in parentheses of three FN lines from the field in Minnesota (2013) and Lexington, KY (2015), and the parental line, determined on a dry-weight basis via bulk-seed NIRS. $n = 3$ plots in one location. Data are displayed as mean \pm SE.

ID	Gen	MN Oil (2013)	KY Oil (2015)	Notes
1R22C28Cgadbr355aMN13	M8	22.0 \pm 0.6	20.4 \pm 1.3	High-oil, short, bushy, indeterminate, late maturity
5R12C21Dar387dMN13	M5	22.4 \pm 0.8	20.1 \pm 0.9	High-oil, erect petioles and lateral branches, slightly chlorotic small lanceolate leaves, late maturity
5R16C01Dar388eMN13	M5	21.9 \pm 1.1	22.3 \pm 1.2	High-oil, short, slightly chlorotic smaller lanceolate leaves, petioles long compared to plant ht., late maturity
M92-220	Parent	19.0 \pm 0.7	19.4 \pm 0.6	Parent of mutant lines

Successful crosses were obtained between 1R22 \times M92-220 and 5R12 \times M92-220. The oil content of the resulting F1 seeds ranged from 17.3 to 22.5% (Table 2), as was determined by single-seed NMR. These F1 seeds were grown in a greenhouse over winter and designated as A1 through B3 plants. A5, A6, A7, and B3 seeds were planted but were not viable or did not survive to maturity. The F2 seeds resulting from the greenhouse planting were harvested and bulked to measure oil and protein content for all the seeds from each F2 population via NIRS (Table 3). Furthermore, thirty seeds were randomly selected from each of these populations for single-seed NMR analysis. Table 4 shows data from the A1 population. These 30 F2 seeds analyzed via NMR were then planted in the field for each respective population, along with an additional 30 F2 seeds that did not contain single-seed data per population. Leaf tissue was collected during the two-leaf stage for the F2 plants, and the F3 seeds were harvested at maturity. NIRS bulk data were collected on each F2:3 sample, using approximately 100 seeds. The protein and oil estimates for these families are shown for lines derived from the A1, A2, and A3 F1 lineages, which are all from crosses between 1R22 and the M92-220 WT (Table 5). Scatter plots of the protein and oil estimates for a deeper sampling of F2:3 families are shown (Figure 1); each data

point represents the values for a given F2:3 sample of seeds. The populations generally showed good spread in their data, with inverse correlations between protein and oil. This indicates that a genetic factor may have segregated in the F2 plants, perhaps an FN-induced mutation that is associated with the high-oil trait.

Table 2. Oil content of individual F1 seeds. Each F1 seed was assigned a population name for reference to the future generations raised from that seed. The first parent mentioned is the maternal plant, and the second mentioned is the paternal. Oil content was determined by single-seed NMR using a Minspec 20 (Bruker Biospin, The Woodlands, TX, USA). Seeds were weighed and placed into the tube and allowed to warm to 40 °C before insertion into the instrument. The standard oil seed measurement procedure supplied with the instrument controller was used. $n = 3$ technical replications on each seed, and SEs are in parentheses. Herein, the lines are referred to by the population name rather than parents for clarity.

Name	Parents	Mass (g)	OIL (% db)
A1	1R22×WT	0.2544	22.5 (0.19)
A2	1R22×WT	0.2181	22.8 (0.30)
A3	1R22×WT	0.2401	22.5 (0.26)
A4	5R12×WT	0.1742	19.3 (0.26)
A5	5R12×WT	0.2092	19.6 (0.33)
A6	5R12×WT	0.2027	21.6 (0.11)
A7	5R12×WT	0.2741	18.9 (0.27)
A8	WT×1R22	0.2734	19.3 (0.21)
B1	WT×1R22	0.1562	17.3 (0.37)
B2	WT×1R22	0.2771	18.9 (0.22)
WT	N/A	0.2051	20.5 (0.23)
1R22	N/A	0.1716	22.9 (0.36)
5R12	N/A	0.1706	20.5 (1.3)

Table 3. Protein and oil content of bulked F2 seeds. The “Cross” column shows the parents in the original cross; the first parent shown is the maternal plant, and the second mentioned is the paternal. Oil and protein content was determined by bulk-seed NIRS using a Perten DA7200 Spectrometer with $n = 3$ technical replications on each seed batch.

Plant ID	Cross	Mean Protein	SE	Mean Oil	SE
A1	1R22×WT	43.6	0.8	22.4	0.2
A2	1R22×WT	43.1	0.1	22.3	0.3
A3	1R22×WT	42.5	0.3	22.6	0.3
A4	5R12×WT	41.3	0.5	25.1	0.3
A7	5R12×WT	42.4	0.0	22.1	0.0
A8	WT×1R22	44.3	0.1	19.1	0.2
B1	WT×1R22	42.7	0.8	22.1	0.4
B2	WT×1R22	43.8	0.2	21.6	0.2
M92-220	Parent	42.6	0.2	22.5	0.2
5R12	Mutant	42.7	0.5	21.8	0.1
1R22	Mutant	38.6	0.5	24.6	0.4

Table 4. Single-seed oil and protein of 30 randomly selected F2 seeds (from F1 plants) selected for field planting, all of which are sibs. The field # (number) serves as an identifier related to seeds retained in long-term storage and used to track lineages. Oil and protein were determined for each seed using single-seed NIRS with 3 technical reps per seed. Single seeds were also surveyed in a similar way for lines A2, A3, A4, A8, B1, and B2, whose genetic lineages are described in Table 3.

Line	Field # (ID)	AVG Prot	SE Prot	AVG Oil	SE Oil
A1	1	45.8	0.5	20.9	0.2
A1	2	44.8	0.5	20.7	0.6

Table 4. Cont.

Line	Field # (ID)	AVG Prot	SE Prot	AVG Oil	SE Oil
A1	3	46.1	0.6	21.1	0.2
A1	4	45.2	0.9	18.1	0.3
A1	5	45.1	0.6	20.7	0.1
A1	6	38.5	0.7	22.4	0.3
A1	7	42.6	0.4	22.2	0.5
A1	8	46.1	0.3	20.0	0.0
A1	9	41.2	0.3	22.3	0.4
A1	10	50.6	0.4	17.5	0.2
A1	11	47.3	0.3	19.6	0.1
A1	12	42.0	0.2	22.7	0.2
A1	13	38.6	1.0	22.4	0.6
A1	14	48.2	0.5	19.1	0.4
A1	15	44.0	1.2	21.0	0.5
A1	16	46.3	0.2	20.1	0.2
A1	17	53.1	0.9	17.3	0.6
A1	18	45.7	0.0	20.9	0.0
A1	19	48.3	0.2	19.7	0.7
A1	20	45.6	0.4	21.1	0.5
A1	21	45.5	0.4	19.8	0.4
A1	22	37.1	0.8	24.3	0.3
A1	23	46.1	0.3	19.6	0.4
A1	24	51.1	1.1	18.0	0.0
A1	25	46.9	0.4	19.7	0.2
A1	26	43.5	0.7	20.4	0.2
A1	27	43.3	0.3	23.7	0.5
A1	28	42.2	1.0	22.2	0.3
A1	29	49.8	1.0	19.0	0.4
A1	30	40.8	0.1	22.1	0.2

Table 5. Seed oil and protein from F2:3 lines A1, A2, and A3 (all originally derived from 1R22×WT crosses). Each row represents data from a sample of F3 seeds derived from an individual F2 plant. Oil and protein were determined via NIRS using a Pertene DA7200 spectrometer and calculated on a 0% moisture basis. Each population (A1, A2, and A3) was split into a high-oil (bold text) and normal-oil (non-bold text) group. These groupings were used to determine which F2 DNA samples were bulked together for CGH analyses (high-oil bulk versus normal-oil bulk).

Plant ID	Mean Protein	Mean Oil
MNA1: 37.5	40.1	24.2
MNA1: 11	39.3	24.2
MNA1: 14	40.2	24.1
MNA1: 16	39.0	24.0
MNA1: 44	38.6	24.0
MNA1: 33	39.0	23.3
MNA1: 15	43.9	21.7
MNA1: 3	45.3	21.4
MNA1: 26	43.9	21.3
MNA1: 12	45.2	21.0
MNA1: 3	43.8	20.6
MNA1: 41.5	46.3	20.2
MNA2: 8	40.1	24.1
MNA2: 28	39.9	23.9
MNA2: 29	39.6	23.8
MNA2: 40	38.9	23.8
MNA2: 4	39.2	23.7
MNA2: 43	39.8	23.7
MNA2: 31	38.6	23.4
MNA2: 5	39.5	23.4

Table 5. Cont.

Plant ID	Mean Protein	Mean Oil
MNA2: 25	40.2	23.2
MNA2: 33	39.8	23.2
MNA2: 21	45.3	22.4
MNA2: 30	43.9	21.9
MNA2: 44.5	45.1	21.7
MNA2: 10	44.9	21.7
MNA2: 33.5	43.8	21.5
MNA2: 15	45.0	21.5
MNA2: 42	45.2	21.0
MNA2: 44	44.5	20.4
MNA3: 7	39.5	24.3
MNA3: 42.5	39.0	24.0
MNA3: 3	40.2	23.6
MNA3: 8	40.9	23.5
MNA3: 14	39.1	23.4
MNA3: 34	44.0	21.2
MNA3: 38.5	44.5	21.1
MNA3: 30.5	44.4	20.9
MNA3: 4	45.1	20.7
MNA3: 37.5	44.5	20.4
MNA3: 25	44.3	20.4
MNA3: 16	44.9	20.2
MNA3: 33	46.0	19.2
MNA3: 34.5	45.1	19.2

3.2. CGH Analysis Reveals a Strong Candidate Deletion for the High-Oil Mutant Phenotype

We chose to further investigate the A1, A2, and A3 populations to determine if an FN-induced deletion was co-segregating with the high-seed-oil phenotype in the 1R22 mutant populations. We collected DNA for each of the F2 plants grown in the field in KY. We performed a bulk segregant analysis using CGH to see if there were any deletions that co-segregated with the high-oil phenotype. Based on the F2:3 NIRS data (Table 5), we binned each F2 into those that gave rise to high seed oil (high-oil, bold) and those that gave rise to normal seed oil levels (normal-oil, non-bold) within each respective population. We then bulked the DNA for the high-oil and normal-oil plants, respectively. These bulked DNA samples were subjected to CGH analyses to compare the deletion/duplication profile of the high-oil and normal-oil bulks for each of the three populations. The deletion/duplication profile of the high-oil and normal-oil bulks for each of the three populations was determined. We expected that the CGH profile would identify a differential hybridization between any deletion that was enriched in the high-oil segregating bulk of individuals compared to the normal-oil segregating bulk.

One large deletion event (>300 kb) on chromosome 14 was enriched in the high-oil bulks of the A1 and A3 populations (Figure 2), as evidenced by the string of data points below the log₂ value of zero. (Zero indicates no difference between the high-oil bulk and the normal bulk, whereas values below zero indicate an enrichment for an FN deletion in the high-oil bulks compared to the normal bulks.) Specifically, the deletion was found on Chr14, from bp 9,994,086 to 10,301,954, a span of approximately 308 kb. The deletion is more pronounced in the A1 population, indicating that a greater proportion of individuals in that population were likely homozygous for the large deletion. Nonetheless, the A3 population also shows an enrichment for the large deletion in the high-oil bulk compared to the normal bulk. We know from previous experience that it is difficult to determine perfect bulks based on phenotypes for seed composition traits [18]. Thus, it is not surprising that our three populations did not all show similar enrichment. In fact, the A2 population did not show enrichment for this deletion in its high-oil bulk. It is worth noting, however, that the spread of the oil data in the A2 population was less distinct than the spread from the

A1 and A3 populations (Figure 1), indicating that the A2 plants were more likely to have their DNA samples bulked into the wrong group. Thus, we tentatively conclude that the large deletion on chromosome 14 is likely co-segregating with the high-seed-oil phenotype in these mutant populations.

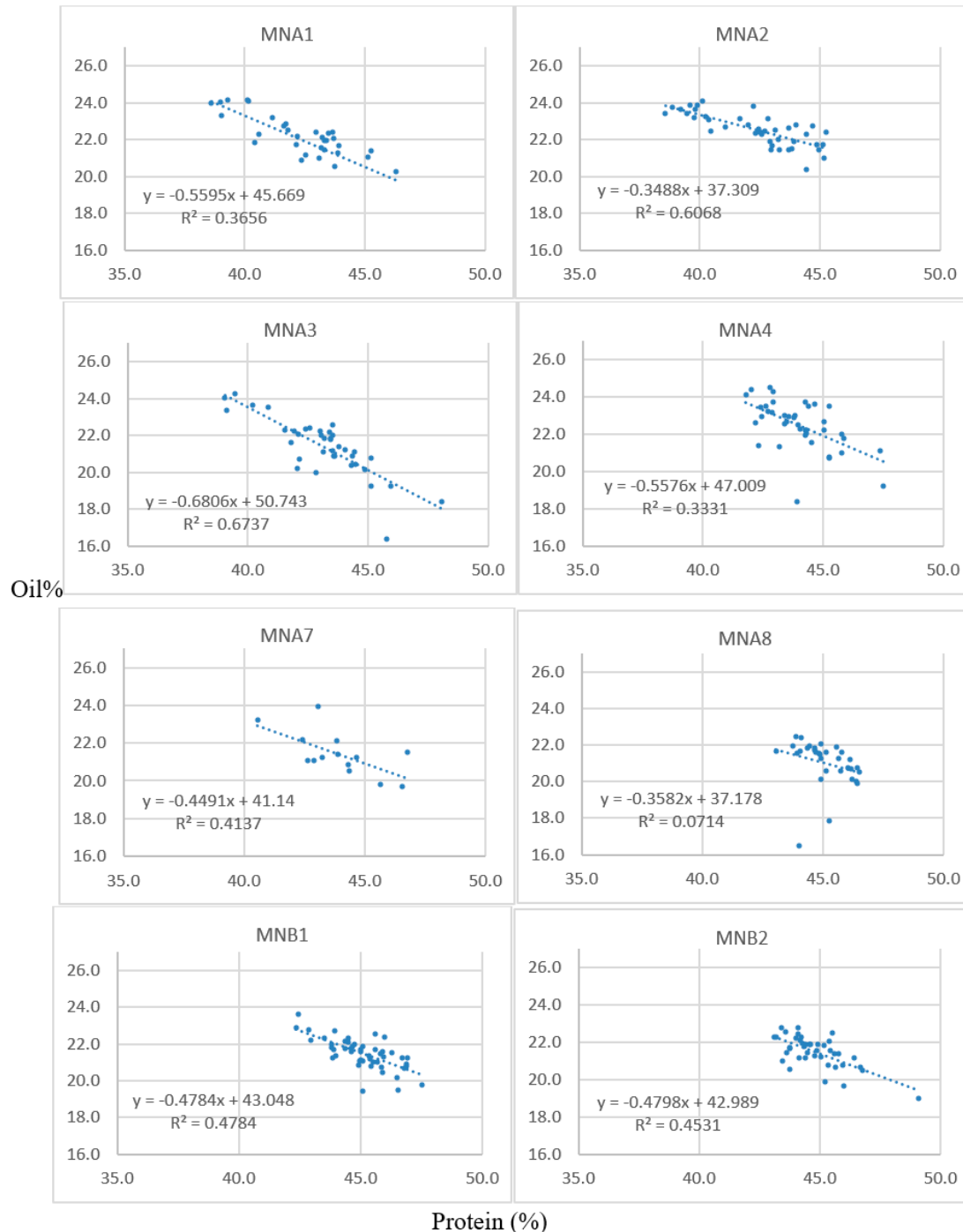


Figure 1. Scatter plots of the protein (x-axes) and oil (y-axes) content of F2:3 seeds. Each data point represents the NIRS values for each respective F2:3 family (e.g., corresponding to the rows in Table 5 for the A1, A2, and A3 populations). Seeds were analyzed for percent oil and protein content on a dry-weight basis via bulk-seed NIRS with $n = 3$ technical replications for each F2:3 seed lot.

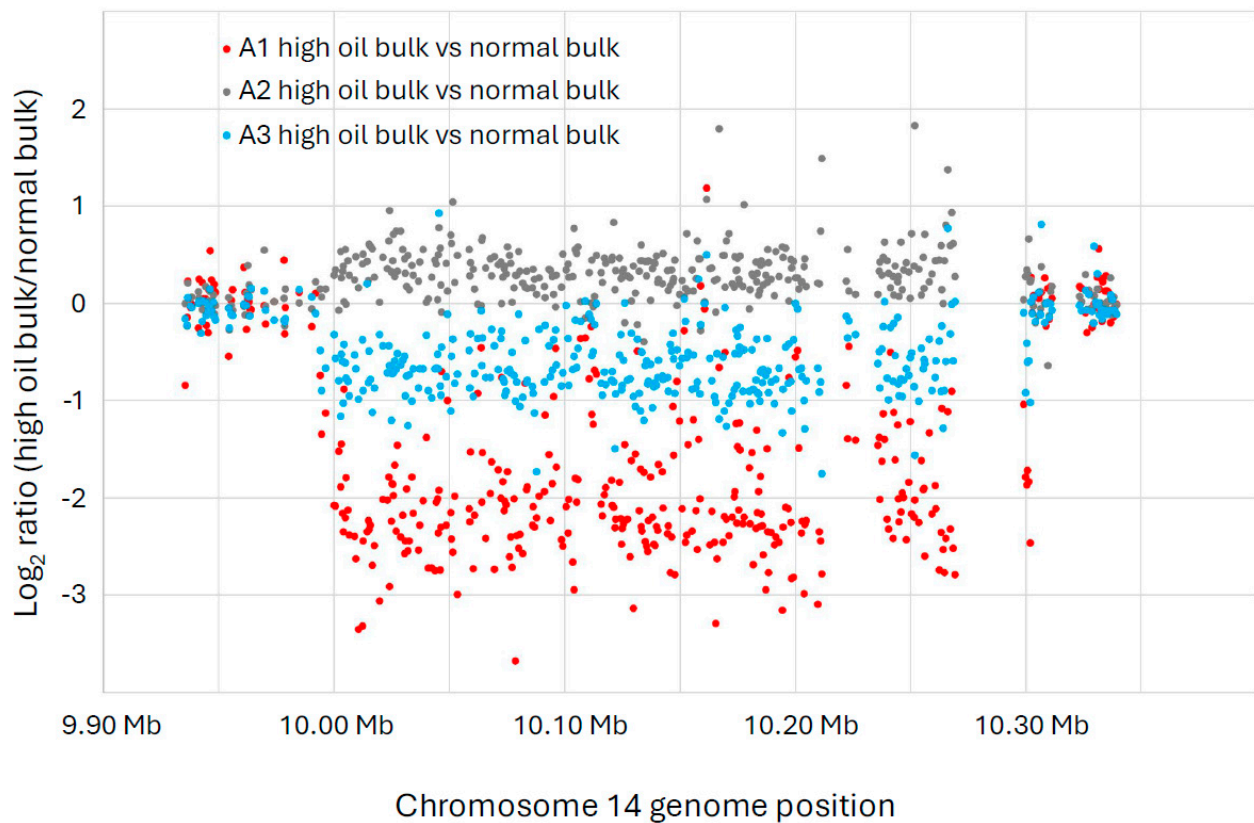


Figure 2. A large copy number variation (CNV) event detected in chromosome 14 of the A1 and A3 populations (both derived from separate crosses between 1R22 \times WT) exhibited strong inverse correlations of oil and protein content. The x -axis indicates the location of each microarray feature along chromosome 14, according to the Williams 82 genome version 2 assembly (Wm82.a2.v1). The y -axis shows the \log_2 ratio of the CGH intensity from the high-oil versus the normal-oil bulk for each microarray feature. The blue dots show the CGH comparison between A1 high-oil versus A1 normal-oil. The orange dots show the CGH of high-oil versus normal-oil for A2, and the grey dots show the CGH of high-oil versus normal-oil for A3. These data indicate that a large deletion is enriched in the high-oil plants of the A1 and A3 populations, while the A2 population does not show this enrichment. This deletion event was detected on Chr14, from bp 9,994,086 to 10,301,954, a span of approximately 308 kb.

3.3. Functional Analysis

The deletion on chromosome 14 encompasses 20 soybean gene models. All annotated functions of the 20 putative genes in the deleted region are summarized (Table 6). Of the 20 genes, 18 have a readily predicted function, but none appears to have an obvious connection to seed oil phenotypes. Tables 6–8 contain information about all genes within the deletion region. Among genes in this table, *Glyma.14G102100* and *Glyma.14G101900* had no readily predicted function. Therefore, bioinformatic analyses were performed to reveal possible functions for these genes. *Glyma.14G101900* with 82 amino acids is composed of 61% disordered protein (lacks an ordered three-dimensional structure) (Table 7) and is predicted to be a transmembrane protein by TAIR. The membrane topology of this gene was illustrated and predicted by Protter (Figure 3). The protein network for *Glyma.14G101900* (Figure 4) and the PDB model of both *Glyma.14G101900* and *Glyma.14G102100* were predicted by Phyre2 and illustrated by Chimera (Figures 5 and 6).

Table 6. A large deletion event was detected on Chr14, from position 9,994,086 to 10,301,954 (coordinates based on the Williams 82 genome version 2 assembly (Wm82.a2.v1), a span of approximately 308 kb in the 1R22-derived populations. * Putative transcription factors.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14	Predicted Functions			
			Pfam, KEGG	GO, KOG, AT	PantherFam	UniProt
<i>Glyma.14G101000</i>	<i>AT5G10810.1</i> , NP_001241623, gmx:548093	Start: 10254421 Stop: 10258523, 128 orthologues 3 paralogues 7 domains and features, 45 oligo probes	Enhancer of rudimentary (PF01133), Glutaredoxin 2, C terminal domain (PF04399).	Positive regulation of Notch signaling pathway (GO:0045747), cell cycle (GO:0007049), pyrimidine nucleotide biosynthetic process (GO:0006221), enhancer of rudimentary (KOG1766)	Enhancer of rudimentary (PTHR12373)	Enhancer of rudimentary homolog (C6TKU9)
<i>Glyma.14G101100</i>	<i>AT5G10820.1</i> , XP_003545404, gmx:100806186	Start: 10268501 Stop: 10272945, 101 orthologues 12 paralogues. 23 domains and features, 49 oligo probes	Major Facilitator Superfamily (PF07690), The bioplerin/folate transporter (PF03092)	Transmembrane transport (GO:0055085), integral component of membrane (GO:0016021)	Folate bioplerin transporter 1, chloroplastic (PTHR31585)	Folate-bioplerin transporter 6 (I1M978)
<i>Glyma.14G101200</i>	<i>AT4G31870.1</i> , KAH1212471	Start: 10274782 Stop: 10276509, 138 orthologues 13 paralogues. 12 domains and features, 11 oligo probes.	Glutathione peroxidase (PF00255)	Response to oxidative stress (GO:0006979), obsolete oxidation–reduction process (GO:0055114), Glutathione peroxidase activity (GO:0004602)	Glutathione peroxidase (PTHR11592)	Glutathione peroxidase (A0A0R0GIC5), Biological process: response to oxidative stress
<i>Glyma.14G101300</i>	<i>AT4G31870.1</i> KAH1093928	Start: 10277857 Stop: 10278815, 1 orthologue 7 domains and features, 20 oligo probes.	Glutathione peroxidase (PF00255)	Response to oxidative stress (GO:0006979), obsolete oxidation–reduction process (GO:0055114), Glutathione peroxidase activity (GO:0004602)	Glutathione peroxidase, (PF00255)	Glutathione peroxidase (A0A0R0GNK4), Biological process: response to oxidative stress
<i>Glyma.14G101400</i>	<i>AT4G31860.1</i> , KAH1212475 gmx: K17499	Start: 10285154 Stop: 10291084, 63 orthologues 153 paralogues. 13 domains and features, 34 oligo probes	Protein phosphatase 2C (PF00481)	Protein dephosphorylation (GO:0006470), catalytic activity (GO:0003824), protein serine/threonine phosphatase activity (GO:0004722), serine/threonine protein phosphatase (KOG0699)	Protein phosphatase 2C (PTHR13832)	protein serine/threonine phosphatase (I1M981), Mg ²⁺ Mn ²⁺ metal-ion binding, myosin phosphatase activity, Biological process: protein dephosphorylation

Table 6. Cont.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14	Predicted Functions			
<i>Glyma.14G101500</i>	<i>AT4G31180.1</i> , <i>XP_003544559</i> , gmx:100788164	Start: 10297083 Stop: 10303442, 255 orthologues 14 paralogues 21 domains and features, 56 oligo probes.	tRNA synthetases class II (D, K and N) (PF00152), tRNA-synt_2d (PF01409), tRNA anti-codon (PF01336), KEGG: Mitochondrial biogenesis (K01876)	tRNA aminoacylation for protein translation (GO:0006418), nucleotide binding (GO:0000166), aminoacyl-tRNA ligase activity (GO:0004812), ATP binding (GO:0005524)	Aspartyl/lysyl-tRNA synthetase (PTHR22594)	Aspartate—tRNA ligase (I1M984), [ATP + L-aspartate + tRNA (Asp) = AMP + diphosphate + L-aspartyl-tRNA (Asp)] Cellular component: aminoacyl-tRNA synthetase multi-enzyme complex, cytosol Molecular function: aspartate-tRNA ligase activity, ATP binding, DNA binding, RNA binding Biological process: aspartyl-tRNA aminoacylation ATPase AAA-type core domain-containing protein (A0A0R0GCB5), Molecular function: ATP binding, ATP hydrolysis activity Nuclear nucleic acid-binding protein C1D (C6T2F3). Plays a role in the recruitment of the exosome to pre-rRNA to mediate the 3'-5' end processing of the 5.8S rRNA. Cellular component: cytoplasm, exosome, nucleolus Molecular function: DNA binding, RNA binding Biological process: maturation of 5.8S rRNA, regulation of gene expression
<i>Glyma.14G101600</i>	<i>AT2G18193.1</i> , <i>KAH1212479.1</i>	Start: 10306650 Stop: 10310815, 492 orthologues 34 paralogues 8 domains and features, 6 oligo probes.	ATPase family associated with various cellular activities (AAA) (PF00004)	ATP binding (GO:0005524)	BCS1 AAA-type ATPase (PTHR23070)	
<i>Glyma.14G101700</i>	<i>AT5G25080.1</i> <i>NP_001237643.1</i> , gmx:100500667	Start: 10314547 Stop: 10319982, 13 orthologues. Five domains and features, 53 oligo probes	Sas10/Utp3/C1D family (PF04000) KEGG: Eukaryotic RNA degradation (100500667), Messenger RNA biogenesis (K12592)	KOG: DNA-binding protein C1D involved in regulation of double-strand break repair (KOG4835),AT: Sas10/Utp3/C1D family (<i>AT5G25080.1</i>)	Sun-cor steroid hormone receptor co-repressor, (PTHR15341), Nuclear nucleic acid-binding protein (PTHR15341:S3)	

Table 6. Cont.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14		Predicted Functions	
<i>Glyma.14G101800</i>	<i>AT4G31840.1</i> , <i>KAG4962690.1</i>	Start: 10327216 Stop: 10328729, 57 orthologues and 5 paralogues 14 domains and features, 40 oligo probes	Plastocyanin-like domain (PF02298)	Electron transfer activity (GO:0009055) AT: Early nodulin-like protein 15 (<i>AT4G31840.1</i>)	Blue copper protein JGI N/A IEA (PTHR33021), Phytocyanin domain-containing protein (I1M987) Cellular component: plasma membrane. Molecular function: electron transfer activity
<i>Glyma.14G101900</i>	<i>AT4G31830.1</i> , <i>KAG4962691.1</i>	Start: 10329690 Stop: 10330667, 101 orthologues and 1 paralogue 5 domains and features, 26 oligo probes		AT: Transmembrane protein (<i>AT4G31830.1</i>)	OS09G0127700 PROTEIN (PTHR33919) Transmembrane protein (I1M988). Cellular component: membrane
<i>Glyma.14G102000</i>	<i>AT5G10840.1</i> , XP_003544560.1, gmx:100788693	Start: 10332624 Stop: 10337739, 286 orthologues and 28 paralogues. Twenty domains and features, 34 oligo probes	Endomembrane protein 70 (PF02990), Major Facilitator Superfamily (PF07690) KEGG: Exosome (K17086)	Integral component of membrane (GO:0016021) AT: Endomembrane protein 70 protein family (<i>AT5G10840.1</i>)	Transmembrane 9 superfamily member (I1M989) Cellular component: endosome membrane, Golgi membrane, membrane Biological process: protein localization to membrane
<i>Glyma.14G102100</i>	None <i>KAH1093938.1</i>	Start: 10093844 Stop: 10094237, 34 paralogues. One domain and feature			Uncharacterized protein (A0A0R0GBJ6)
<i>Glyma.14G102200</i> *	<i>AT1G18400.1</i> , <i>KAH1093939.1</i>	Start: 10350687 Stop: 10352402, 119 orthologues and 79 paralogues. Ten domains and features, 9 oligo probes		AT: Encodes the brassinosteroid signaling component <i>BEE1</i> (<i>BR-ENHANCED EXPRESSION 1</i>). Positively modulates the shade avoidance syndrome in Arabidopsis seedlings. (<i>AT1G18400.1</i>)	Sterol regulatory element-binding protein (PTHR12565) Transcription factor bee 3 1 hit (PTHR12565:SF340) Retrotran_gag_3 domain-containing protein 1 hit (PTHR47481:SF19)
<i>Glyma.14G102300</i>	<i>AT1G21280.1</i>	Start: 10115205 Stop: 10116250, 1 orthologue and 4 paralogues. Four domains and features	Retrotran_gag_2 1 hit (PF14223)	AT: Copia-like polyprotein/retrotransposon (<i>AT1G21280.1</i>)	Retrotran_gag_3 domain-containing protein (A0A0R0GBC5)

Table 6. Cont.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14		Predicted Functions		
<i>Glyma.14G102400</i>	<i>AT1G19260.1</i> XP_014622176.2 gmx:102660685	Start: 10409336 Stop: 10412210, 53 orthologues and 83 paralogues. Seven domains and features.	Domain of unknown function DUF4371 (PF14291) hAT family C-terminal dimerisation region (PF05699) KEEG: LOW QUALITY PROTEIN: zinc finger MYM-type protein 1 (102660685)	AT: Encodes a ceramide synthase that uses very long chain fatty acyl-CoA and trihydroxy LCB substrates (<i>AT1G19260.1</i>)	General transcription factor 2-related zinc finger protein (PTHR11697) Zinc finger, mym domain-containing 1 1 hit (PTHR11697:SF227)	TTF-type domain-containing protein (K7M603)
<i>Glyma.14G102500</i>	<i>AT4G31820.1</i> , XP_003544563.1, gmx:100790291	Start: 10424224 Stop: 10429543, 159 orthologues and 75 paralogues, 18 domains and features and maps to 29 oligo probes.	BTB/POZ domain (PF00651) NPH3 family (PF03000) KEEG: BTB/POZ domain-containing protein NPY1 (100790291)	Animal organ development (GO:0048513) auxin transport (GO:0060918) obsolete signal transducer activity (GO:0004871) protein binding (GO:0005515) AT: A member of the <i>NPY</i> family genes (<i>NPY1/AT4G31820</i> , <i>NPY2/AT2G14820</i> , <i>NPY3/AT5G67440</i> , <i>NPY4/AT2G23050</i> , <i>NPY5/AT4G37590</i>). Encodes a protein with similarity to <i>NHP3</i> . Contains BTB/POZ domain. Promoter region has canonical auxin response element-binding site and Wus-binding site. Co-localizes to the late endosome with PID. Regulates cotyledon development through control of <i>PIN1</i> polarity in concert with PID. Also involved in sepal and gynoecia development. <i>AT4G31820.1</i>	OS12G0117600 Protein (PTHR32370) BTB/POZ Domain-containing protein <i>NPY1</i> (PTHR32370:SF7)	<i>NPH3</i> domain-containing protein (K7M604), Pathway: Protein modification, protein ubiquitination Biological process: protein ubiquitination

Table 6. Cont.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14		Predicted Functions		
<i>Glyma.14G102600</i>	<i>AT4G31810.1</i> , <i>XP_028198380.1</i>	Start: 10431488 Stop: 10441267 117 orthologues and 19 paralogues. Eight domains and features, 31 oligo probes	Enoyl-CoA hydratase/isomerase (PF00378), ECH_2 1 hit (PF16113) KEGG: beta-Alanine metabolism (K05605), Valine, leucine and isoleucine degradation (map00280), beta-Alanine metabolism (map00410), Propanoate metabolism (map00640), Metabolic pathways (map01100), Carbon metabolism (map01200)	Metabolic process (GO:0008152), Catalytic activity (GO:0003824) AT: ATP-dependent caseinolytic (Clp) protease/crotonase family protein (AT4G31810.1)	Enoyl-coa hydratase-related (PTHR11941)	3-hydroxyisobutyryl-CoA hydrolase (I1M992) Hydrolyzes 3-hydroxyisobutyryl-CoA (HIBYL-CoA), a saline catabolite. Has high activity toward isobutyryl-CoA. Could be an isobutyryl-CoA dehydrogenase that functions in valine catabolism. Catalytic activity 3-hydroxy-2-methylpropanoyl-CoA + H ₂ O = 3-hydroxy-2-methylpropanoate + CoA + H ⁺ ,
<i>Glyma.14G102700</i>	<i>AT5G10870.1</i> , <i>XP_006596036.1</i>	Start: 10445611 Stop: 10448775 174 orthologues and 9 paralogues, 9 domains and features 19 oligo probes.	KEGG: Phenylalanine, tyrosine and tryptophan biosynthesis (K01850), Phenylalanine, tyrosine and tryptophan biosynthesis (map00400), metabolic pathways (map01100), Biosynthesis of secondary metabolites (map01110), biosynthesis of amino acids (map01230)	Aromatic amino acid family biosynthetic process (GO:0009073), Chorismate mutase activity (GO:0004106), Chorismate mutase (KOG0795) AT: Encodes chorismate mutase AtCM2 (AT5G10870.1)	Chorismate mutase (PTHR21145)	Chorismate mutase (A0A0R0GIU8), Cellular component: cytoplasm, Molecular function: Chorismate mutase activity. Biological process: amino acid biosynthetic process, aromatic amino acid family biosynthetic process, chorismate metabolic process

Table 6. Cont.

	Arabidopsis Homologue ID, NCBI Protein ID, KEGG ID	Chr: Gm14		Predicted Functions		
<i>Glyma.14G102800</i>	<i>AT2G25050.1</i> , <i>KAG5110165.1</i>	Start: 10486626 Stop: 10508579, 240 orthologues and 36 paralogues, 57 domains and features, 28 oligo probes.	C2 domain of PTEN tumor-suppressor protein (PF10409), DUF4283 1 (PF14111), <i>FH2 2</i> (PF02181), <i>PTEN_C2 1</i> (PF10409), <i>RVT_1 1</i> (PF00078)	AT: Class II formin; modulator of pollen tube elongation (<i>AT5G58160.1</i>)	Formin-related (PTHR23213), FORMIN-J 1 (PTHR45733) FORMIN-J 1 (PTHR45733)	Formin-like protein A0A368UH40 Cellular component: membrane Molecular function: phosphoprotein phosphatase activity
<i>Glyma.14G102900</i> *	<i>AT1G80840.1</i> , <i>XP_014622429.1</i> , <i>gmx:100791870</i>	Start: 10511024 Stop: 10513622 80 orthologues and 44 paralogues. 10 domains and features, 47 oligo probes.	<i>WRKY</i> DNA-binding domain (PF03106), <i>Takusan</i> (PF04822) <i>NUDE_C</i> (PF04880)	Regulation of transcription, DNA-templated (GO:0006355), DNA-binding transcription factor activity (GO:0003700), Sequence-specific DNA binding (GO:0043565), AT: <i>WRKY DNA-BINDING</i> <i>PROTEIN 40</i> (<i>AT1G80840.1</i>) Pathogen-induced transcription factor	<i>WRKY</i> transcription factor 36-related (PTHR31429), <i>WRKY</i> transcription factor 40-related 1 (PTHR31429:SF38)	<i>WRKY</i> domain-containing protein (I1M995), Cellular component: nucleus Molecular function: DNA-binding transcription factor activity, sequence-specific DNA binding

Table 7. Phyre2 results for unknown genes (*Glyma.14G101900* and *Glyma.14G102100*).

Gene ID	Protein Seq		Confidence and Coverage	PDB Molecule	PDB Header	Details
<i>Glyma.14G101900</i>	MDPQKAQAEASKRPPGHGATEV- LHQKKS L PFSFTTMTIAGLLI- TAAVGYSVLYVKKKPEASA- KDVTKVSVGVAKPEETHPEN	82	Confidence: 5.2% Coverage: 33%	na(+)/h(+) antiporter subunit b	Structure of bacillus pseudofirmus Mrp antiporter complex, monomer	61% of this sequence is predicted to be disordered. Disordered-region structures cannot be meaningfully predicted.
<i>Glyma.14G102100</i>	MVGE E E P D W M T P Y K N F L T Q G - V L P S H D N E V R C L K W K A N Y Y I I L D G - E L L K R G L I A S L L K C L N N Q Q T D Y V I R - E L H E G I C A L Y I G G R S L A T K V T L L T L Q - R D V D D A R S L Q T F R A P L L T I S I V	118	Confidence: 99.4% Coverage: 88%	Transposon ty3-g gag-pol polyprotein	DNA-binding protein	

Table 8. Review of putative function of 19 Arabidopsis homologues of the genes in the deleted region of the high-oil soybean mutant.

Name of Gene	Function Reported	References
<i>AT5G10810</i>	As one of 25 candidate <i>AtPNP</i> -As which showed weak interaction strength in the yeast two-hybrid (Y2H) analysis. <i>AtPNP</i> -A = plant natriuretic peptides (PNPs), which comprise a novel class of hormones that systemically affect salt and water balance and responses to plant pathogens.	[34]
<i>AT5G10820</i>	A Folate–Biopterin Transporter (<i>FBT</i>) family member. <i>FBT</i> s are essential cofactors in one-carbon metabolism. The <i>FBT</i> family belongs to the major facilitator superfamily (<i>MFS</i>) and contains 12 transmembrane α -helices. <i>Glutathione peroxidases 7(GPX7)</i> is one of the major ROS-scavenging enzymes which catalyze the reduction of H ₂ O ₂ in order to prevent potential H ₂ O ₂ -induced cellular damage.	[35] [36]
<i>AT4G31870</i>	<i>GPX7</i> (Cys-108, Gln-143, and Trp-197) residues are potential catalytic residues found to be strictly conserved. <i>GPX7</i> is linked to the establishment of the photooxidative stress tolerance and the basal resistance to <i>P. syringae</i> infection. <i>GPX7</i> belongs to a family of thiol-based glutathione peroxidases that catalyzes the reduction of H ₂ O ₂ and hydroperoxides to H ₂ O or alcohols using glutathione as an electron donor. Plant GPXs are implicated in redox signal transduction. <i>VaAQ</i> (a putative GARP-type transcription factor of Amur grape (<i>Vitis amurensis</i>) overexpression increases antioxidant enzyme activities and upregulates ROS scavenging-related genes such as <i>GPX7</i> under cold stress. Strongly induced in carotenoid-accumulating Arabidopsis roots. Expressed highly in response to oxidation–reduction processes Molecular analysis indicates that <i>glutathione peroxidase 7 (GPX7)</i> is specifically induced to compensate for the absence of APx-R (ascorbate peroxidase-related). (Peroxidases are enzymes that catalyze the reduction of hydrogen peroxide, thus minimizing cell injury and modulating signaling pathways in response to this reactive oxygen species.) The transcript abundance of the <i>GPX7 (At4g31870)</i> was increased by a cryoprotectant treatment. <i>GPX7 (At4g31870)</i> is increased upon auxin application. A protein phosphatase DEGs involved in Cold Response. AP2C18, highly ABA-inducible.	[37] [38] [39] [40] [41] [42] [43] [44] [45] [46,47]
<i>AT4G31860</i>	Potentially involved in sucrose-induced atrazine tolerance. Protein phosphatase 2C, putative/PP2C, putative. One of the differentially expressed genes related to plant hormone signal transduction pathways. Putative function: abscisic acid (aba) signal transduction. ABA-induced genes in guard cells of Arabidopsis.	[48] [49] [50]

Table 8. Cont.

Name of Gene	Function Reported	References
AT4G31180	Mutations in <i>At4g31180</i> cause the <i>ibi1</i> (induced disease immunity) phenotype and can block BABA-IR in the background of SA-producing Col-0. (β -aminobutyric acid (BABA) is a priming agent that provides broad-spectrum disease protection.)	[51]
	An aspartyl tRNA synthase (AspRS) orthologue (<i>At4g31180</i>) that improves tolerance to biotic stress.	
	The <i>At4g31180</i> is a target of a synthetic isomer of GABA, called BABA (β -Amino Butyric Acid).	
	Aspartyl tRNA synthetase (AspRS) <i>IBI1</i> in <i>Arabidopsis thaliana</i> (<i>Arabidopsis</i>) acts as an enantiomer-specific receptor of BABA.	
	The primary function of AspRS enzymes is the charging of tRNA ^{Asp} with L-aspartic acid (L-Asp) for protein biosynthesis.	
	One of the genes that is sensitive to infection through the NPR1- or JAR1-dependent pathways.	
	AT4G31180 (<i>IBI1</i> : impaired in BABA-Induced Disease Immunity 1) is one of the proteins identified in the seed monosome and polysome fractions that, based on their annotation, have been associated with RNA binding.	
AT2G18193	The <i>IBI1</i> Receptor of β -Aminobutyric Acid interacts with VOZ transcription factors to regulate abscisic acid signaling and callose-associated defense.	[54]
	AT4G31180 is a putative interaction partner of <i>AtGRXS17</i> (<i>Arabidopsis</i> Glutaredoxin S17).	[55]
	One of the fifty-three predicted genes with similarities to aminoacyl-tRNA synthetases identified in <i>A. thaliana</i> .	[56]
	AT2G18193 is a P-loop nucleoside triphosphate hydrolases that only displayed significant induction upon X-irradiation in wild-type seedlings and not in imbibed seeds or <i>sog1</i> mutants.	[57]
	(The transcription factor <i>SUPPRESSOR OF GAMMA 1 (SOG1)</i> , which is unique to plants but functionally similar to the mammalian tumor suppressor p53.)	[58]
	Expression level of (AAA-ATPase <i>At2g18193</i> -like, P-loop containing nucleoside triphosphate hydrolases superfamily protein) genes that are elevated in MSC	[59]
	(Transcriptome Analyses of Mosaic) mitochondrial mutants of cucumber lines. (DNA repair mechanisms are the regulation of proteolytic processes.)	[59]
AT5G25080	A gene regulated by sulfur deficiency encoding a protein with possible ATPase activity and metal-ion binding.	[60]
	A gene whose expression is ABA-inducible in the wild type of <i>Arabidopsis</i> but not in the <i>ros1-4 (REPRESSOR OF SILENCING 1 (ROS1))</i> mutant.	[61]
	A gene that is significantly upregulated in <i>pTSPO-PDC1</i> under drought stress compared to WT plants. (TSPO: tryptophan-rich sensory protein.)	[62]
	One of the 50 genes upregulated by TBM treatment. (Acetolactate synthase (ALS)-inhibiting herbicide tribenuron methyl (TBM).)	[63]
	Encodes a protein with high homology to Rrp47p and is encoded by <i>At5g25080</i> . (Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs, Mitchell et al., 2003.)	[64]
	Candidate gene that includes significantly associated SNPs for traits involved in drought tolerance. Sas10/Utp3/C1D family.	[65]
	AT5G25080 (<i>RRP47</i>) is a known RNA exosome component with the RRP6 cofactor in plants that mediates protein–protein interactions.	[66]
AT4G31840	Oxido-reductases family: blue copper-binding proteins.	[66]
	A CWP or CW transcript differentially accumulated at a given growth temperature in floral stems.	[67]
	Highly repressed under <i>FPS</i> and <i>SQS</i> expression on the transcription of nuclear genes.	[67]
	(Squalene biosynthesis genes <i>FARNESYL DIPHOSPHATE SYNTHASE (FPS)</i> and <i>SQUALENE SYNTHASE (SQS)</i> were engineered via the <i>Nicotiana tabacum</i> .	[68]
	Early nodulin-like protein 15, increased abundance in HS (Humic Substances) treated vs. untreated roots of <i>Arabidopsis</i> .	[68]
AT4G31840	AT4G31840 (<i>ENODL15</i>). Protein–protein interaction networks linked to aliphatic and indole glucosinolate biosynthetic pathways in <i>Arabidopsis</i> .	[69]
	One of the top 30 genes that are most downregulated in <i>myb3r1 myb3r4</i> seedlings.	[70]
	(Mutations in <i>MYB3R1</i> and <i>MYB3R4</i> cause pleiotropic developmental defects and preferential downregulation of multiple G2/M-specific genes in <i>Arabidopsis</i> .)	[70]

Table 8. Cont.

Name of Gene	Function Reported	References
AT4G31830	A gene that may be involved in the drought response and upregulated in leaf tissue.	[64]
	A gene upregulated in the NAE (<i>N</i> -Acylethanolamines) (and ABA-seedling arrays) annotated as “embryo associated” (e.g., late embryogenesis abundant genes, dehydrins, globulins, oleosins, and vicilins).	[71]
	A conserved drought-adaptive gene whose function is unclassified.	[72]
	A gene that shows tissue specificity, as well as expression conservation in rice and Arabidopsis seeds.	[73]
	Homologue to (Mtr.17894.1.S1_at) a putative ABI3 regulon of <i>Medicago truncatula</i> .	[74]
AT5G10840	Endomembrane protein (70 protein family) that is downregulated by <i>Colletotrichum acutatum</i> in strawberry crown tissue.	[75]
	AT5G10840 encodes a highly altered redox-regulated protein in response to 3 mM bicarbonate treatment in <i>A. thaliana</i> var. <i>Landsberg erecta</i> .	[76]
	A conserved syntenic region that pairs between the pseudo-ancestral <i>Arabidopsis</i> genome and <i>Prunus</i> genetic maps. Endomembrane protein 70, putative TM4 family.	[77]
	The transmembrane proteins identified from the plasma membrane of poplar differentiating xylem and phloem. Endomembrane protein 70, putative. BRASSINOSTEROID ENHANCED EXPRESSION1 (<i>BEE1</i>) (<i>At1g18400</i>) is a low-temperature regulator of flavonoid accumulation.	[78]
	<i>BEE1</i> and <i>GFR</i> (<i>G2-LIKE FLAVONOID REGULATOR</i>) were both shown to negatively regulate anthocyanin accumulation by inhibiting anthocyanin synthesis genes via the suppression of the bHLH (<i>TRANSPARENT TESTA8</i> (<i>TT8</i>) and <i>GLABROUS3</i> (<i>GL3</i>)) and/or the MYB (<i>PRODUCTION OF ANTHOCYANIN PIGMENTS2</i> (<i>PAP2</i>)) components of the MBW complex.	[79]
AT1G18400	<i>Arabidopsis BEE1</i> (<i>AT1G18400</i>) is the orthologue of bHLH056 in papaya. bHLH056 may be involved in the process of ABA stress but has different function compared to <i>Arabidopsis</i> .	[80]
	A positive regulator of flavonoid accumulation at low temperatures.	[79]
	A brassinosteroid signaling component and a positive regulator of shade avoidance syndrome.	[81]
	The putative <i>BEE1</i> (c42857_g1_i1_AT1G18400) showed lower expression in ovules at 16 DAA (days after anthesis) in small-seeded litchi.	[82]
	A gene encoding the BR (brassinosteroid) signaling components <i>EE3</i> (<i>AT1G73830</i>) and <i>EE1</i> (<i>AT1G18400</i>) that are significantly upregulated by ethanol treatment, suggesting that the BR pathway is also involved in plant responses to ethanol.	[83]
	<i>BEE1</i> , BR-related transcription factor that are upregulated by BR which encode putative <i>AtMYC2</i> (bHLH) proteins in <i>A. thaliana</i> .	[84]
	One of the three redundant brassinosteroid early response genes that encode putative bHLH transcription factors required for normal growth.	[84]
	One of differential expression genes related to flowering in the Photoperiod pathway.	[85]
	<i>BEE1</i> is a positive regulator of photoperiod flowering and promotes flowering by directly binding to the floral integrator FT.	[86]
	<i>BEE1</i> , -2, and -3 are negative regulators of photomorphogenesis.	[87]
AT1G21280	Involved in the response to iron deficiency.	[88]
	A SNP predicted to be associated with brown rot resistance in peach.	[89]
	Homologue to Tp57577_TGAC_v2_mRNA41271.v2 gene transcription involved in regrowth influenced by location and environmental conditions response after mowing of red clover (<i>Trifolium pratense</i>).	[90]
	Is a duplicated region in Chr 10 of soybean associated with seed protein content.	[91]

Table 8. Cont.

Name of Gene	Function Reported	References
<i>AT1G19260</i>	<p><i>LOH3</i> (<i>At1g19260</i>)-encoded ceramide synthases use very long chain fatty acyl-CoA and trihydroxy long-chain base) LCB (substrates. Overexpression of <i>LOH1</i> and <i>LOH3</i> resulted in a significant increase in plant size. <i>LOH1</i> and <i>LOH3</i> overexpression results in a significant increase in cell number of root meristems. In contrast to results from <i>LOH2</i> overexpression lines, <i>LOH1</i> and <i>LOH3</i> overexpression results in little change in total sphingolipid content and composition of plants relative to wild-type controls, although small but significant reductions in C16 fatty acid-containing sphingolipids were detected as a result of minor changes throughout the sphingolipidome.</p> <p>Is a zinc finger protein-coding gene. Transposes transcription factor-type zinc finger protein with a HAT dimerization domain.</p> <p>Upregulation related to either auxin metabolism, transport, signaling, or response.</p> <p>Predicted target genes of <i>Gnetum gnemon</i> miRNAs against <i>Arabidopsis thaliana</i>.</p> <p>One of the shade-regulated <i>KAN1</i> target genes</p>	<p>[92]</p> <p>[93]</p> <p>[94]</p> <p>[95]</p> <p>[96]</p>
<i>AT4G31820</i>	<p>Functions redundantly in auxin-mediated organogenesis and root gravitropism with the <i>AGC3</i> (protein kinase A, cGMP-dependent protein kinase, and protein kinase C) kinase family</p> <p><i>AT4G31820</i> (<i>ENP1/NPY1</i>) is in the <i>PGP</i> Family of Auxin Transport Facilitators. Role in regulation of Auxin Pathway Genes by <i>REV</i> and <i>KAN</i>. (<i>KANAD11</i> (<i>KAN1</i>), a member of the <i>GARP</i> family of transcription factors, a key regulator of abaxial identity, leaf growth, and meristem formation in <i>Arabidopsis thaliana</i>., <i>REVOLUTA</i> (<i>REV</i>).)</p> <p>All five <i>NPY</i> genes (<i>NPY1</i> = <i>At4g31820</i>, <i>NPY2</i> = <i>At2g14820</i>, <i>NPY3</i> = <i>At5g67440</i>, <i>NPY4</i> = <i>At2g23050</i>, and <i>NPY5</i> = <i>At4g37590</i>) were expressed in tips of <i>Arabidopsis</i> primary roots, but they displayed unique and overlapping patterns. <i>NPY</i> genes play an essential role in root gravitropic responses in <i>Arabidopsis</i>.</p> <p><i>NPY</i> genes and <i>AGC</i> kinases define 2 key steps in a pathway that controls <i>YUC</i>-mediated organogenesis in <i>Arabidopsis</i>.</p> <p>Plays a critical role in auxin-regulated organogenesis in <i>Arabidopsis</i>.</p> <p>(<i>CHY4</i>, <i>At4g31810</i>) is a putative mitochondrial enzyme in valine degradation. A null mutant of 3-hydroxyisobutyryl-CoA hydrolase (<i>CHY4</i>, <i>At4g31810</i>) resulting in an embryo lethal phenotype. <i>CHY4</i> is essential for embryo development.</p> <p><i>CHY4</i> involved in leucine degradation and exhibited a strong association with leucine levels in dark-related datasets.</p>	<p>[97]</p> <p>[98]</p> <p>[99]</p> <p>[100]</p> <p>[101]</p> <p>[102]</p> <p>[103]</p>
<i>AT4G31810</i>	<p>Candidate gene tagged by the associated SNPs related to four important fatty acid (erucic acid, oleic acid, linoleic acid, and linolenic acid) biosynthesis and metabolism in <i>Brassica napus</i>.</p> <p>Homologue to <i>BnaA08g12350D</i>. Enoyl-CoA hydratase/isomerase family protein.</p> <p>Downregulated as substrates of the <i>AtICP55</i> protein. <i>AtICP55</i> is a secondary processing mitochondrial peptidase.</p> <p>Identified as one of conserved syntenic regions pairs between the pseudo-ancestral <i>Arabidopsis</i> genome and <i>Prunus</i> genetic maps. Function: enoyl-CoA hydratase/isomerase family protein.</p> <p>Candidate stigma-specific gene from <i>S. squalidus</i>. Function: Cys protease</p> <p>Involved in Leu degradation. Function: enoyl-CoA hydratases.</p>	<p>[104]</p> <p>[105]</p> <p>[77]</p> <p>[106]</p> <p>[107]</p>

Table 8. Cont.

Name of Gene	Function Reported	References
AT5G10870	Chorismate mutase gene involved in VTE (vitamin E) biosynthesis. Chloroplastic.	[108]
	Upregulated gene in the resistant genotype (<i>Myzus persicae</i>) after GPA infestation. Shikimate pathway, chorismate mutase 2	[109]
	<i>AthCM2</i> (AT5G10870) involved in the shikimate pathway that directs bulk carbon flow toward biosynthesis of aromatic amino acids.	[110]
	Contributes to phenylalanine biosynthesis in Arabidopsis.	[111]
	Involved in shikimate and phenylalanine biosynthesis in plants and algae.	[112]
	The activity of <i>AtCM2</i> appears to be insensitive to Phe and Tyr. (The first committed step of Phe biosynthesis from chorismate is catalyzed by chorismate mutase (CM).)	[113]
AT2G25050	Involved in metabolic pathways of amino acids and their associated genes.	[114]
	Belonging to the Asp family and the aromatic amino acid (AAA) networks.	[115]
	Encoding the actin-binding formin homology <i>FH2</i> protein.	[116]
	Around one-third of these <i>CHG-DMCs</i> (cytosine methylation sequence) located within a 3.3 kb region on chromosome 2 within the gene <i>At2g25050</i> .	[117]
AT1G80840	<i>AT2G25050</i> (<i>AtFH18</i>) is 81% close phylogenetic relationships to <i>GmFH3</i> in <i>G. max</i> .	[118]
	Associated with the cell cycle classification and involves the division of specific cells to form the final apple fruit shape.	[119]
	<i>AT1G80840</i> (<i>WRKY40</i>) encodes a pathogen-induced TF and harbors five associated distal peaks with its promoter.	[120]
	A conserved double- <i>W</i> box in the promoter of <i>CaWRKY40</i> mediates autoregulation during response to pathogen attack and heat stress in pepper.	[121]
	Associated with plant defense response.	[122]
	Codes for a pathogen-induced transcription factor.	[123]
AT1G80840	Common downstream gene of SA and upregulation of this gene diminished at high temperature by SA.	[124]
	Overexpression of <i>AtWRKY40</i> enhanced drought stress responses, presumably by interfering with the reactive oxygen species (ROS)-scavenging pathway and osmolyte accumulation process.	[125]
AT1G80840	Involved in the response to abiotic stresses.	[125]

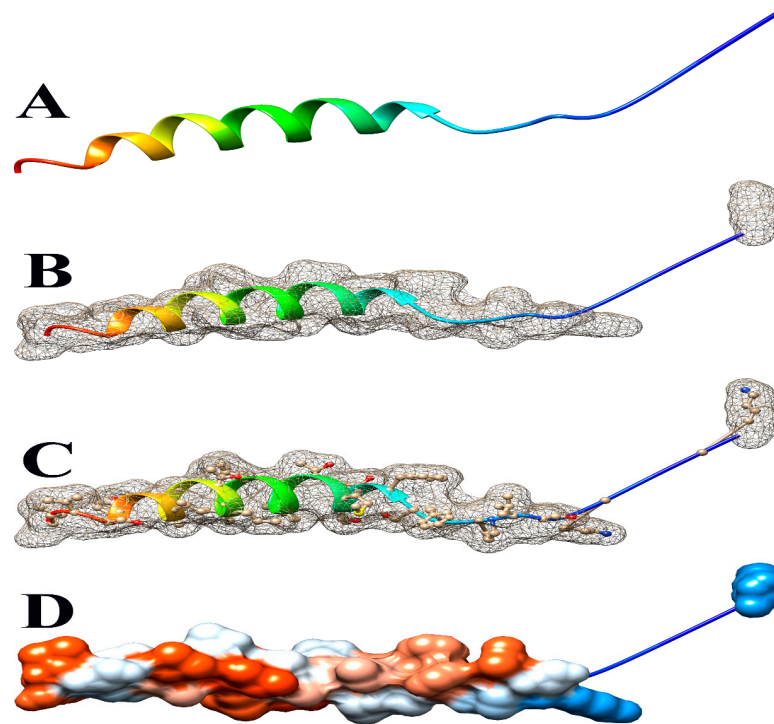


Figure 5. Some views of *Glyma.14G101900* PDB model predicted by Phyre2 and illustrated by Chimera. (A) Ribbon view, (B) mesh surface view, (C) mesh surface with ball and stick view, and (D) hydrophobicity surface (hydrophobicity surface preset from dodger blue for the most hydrophilic to white to orange-red for the most hydrophobic).

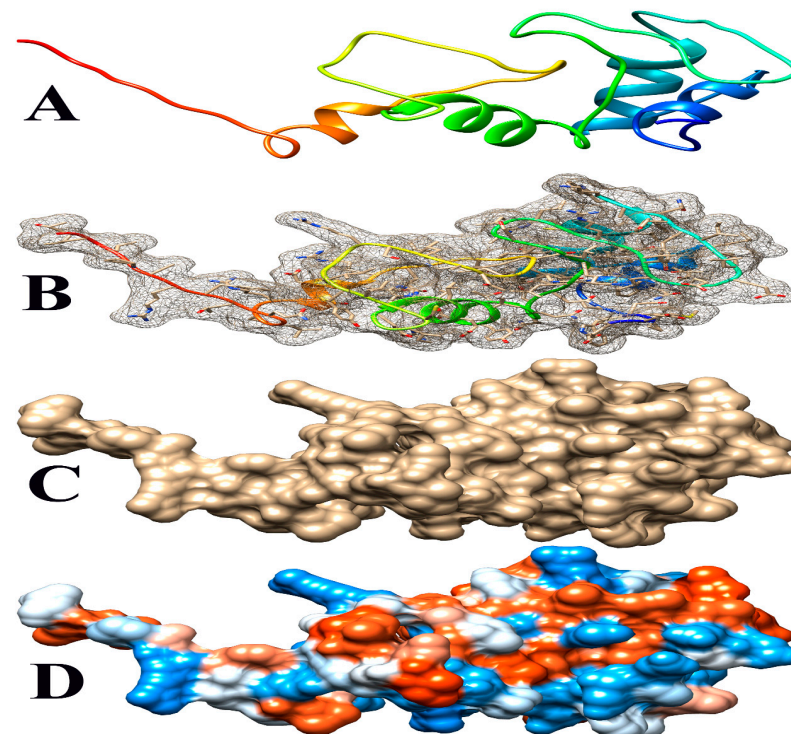


Figure 6. Some views of *Glyma.14G102100* PDB model predicted by Phyre2 and illustrated by Chimera. (A) Ribbon view, (B) mesh surface with ball and stick view, (C) surface view, and (D) hydrophobicity surface (hydrophobicity surface preset from dodger blue for the most hydrophilic to white to orange-red for the most hydrophobic).

4. Discussion

Forward screening of mutant populations for seed composition traits has been utilized with success in many crop species, including soybean seed composition changes induced by FN [18,21,22]. Soybean FN mutant families are powerful due to large variations in mutation sizes, from several bp deletions up to Mb-sized deletion events [17–19,24]. We utilized this resource and a forward screening approach to identify three mutants with high seed oil content and lower decreases in protein than are found in varieties produced using conventional breeding techniques [126–128]. These mutants were then backcrossed to the parent variety, which, in theory, should produce heterozygous offspring for the mutant traits in the F1 population. After a generation of self-pollination, we would expect to observe a genotypic segregation ratio of 1:2:1 for the homozygous mutant/heterozygous/homozygous wild type.

In the populations A1, A2, and A3, phenotype segregation was apparent in the F2:3 seeds, with oil content ranging from 19 to 24%, spread over a somewhat normal distribution. Assuming that the high-oil phenotype is caused by an FN-induced deletion, CGH may be able to show the deletion event when comparing the highest and lowest oil F2 bulked DNA samples, which was clearly observed in the A1 population. Furthermore, the size of the deletion detected, about 300 kb, is within the size range we anticipate and is frequently observed in CGH on soybean FN mutants [17,18,24].

While the size of the deletion event may imply that it is the source of observed phenotypic variation, it is also important to examine the deleted genes to begin to hypothesize about the mechanism which causes the observed variability. To date, many efforts have been made through mutagenesis, conventional breeding, and biotechnology to increase oil content of seeds. It has been well established that the ratio of sucrose/asn+gln from the mother plant significantly alters oil and protein content but results in an inverse correlation of these traits [129–131].

There is a genetic component to this, as high oil is heritable when selected in this manner, but usually results in a corresponding and roughly equal loss of protein. Metabolic engineering efforts have been effective at elevating oil content without the corresponding loss in protein by utilizing push, pull, and protect mechanisms. First, the “push” mechanism directs hydrocarbon resources toward the oil biosynthetic pathway, creating an abundant source of metabolic precursors. The transcription factor WRI1 is known to operate in this fashion [12,132,133]. Next, “pull” mechanisms occur later in the pathway and use downstream metabolites at a faster rate, thereby causing upstream resources to be re-directed into the pathway [11].

The enzyme diacylglycerol acyltransferase (DGAT) catalyzes the final and only dedicated step to TAG synthesis via the Kennedy pathway by combining a diacylglycerol molecule with an acyl-CoA [134]. Biochemical studies have also determined that this is a rate-limiting step in many species, so increasing the speed of TAG formation via DGAT increases the speed of the entire pathway [135]. DGAT-overexpression studies confirm this phenomenon, and DGAT-overexpressed plants have significantly higher oil levels, with no decrease in protein [136,137]. Lastly, “protect” mechanisms ensure that TAGs already formed do not degrade. For example, lipase knockouts [138] exhibit increased oil content, as do oleosin overexpressors, which are proteins that stabilize storage oil bodies [11]. Better yet, plants which are engineered with two or more of these steps exhibit synergistic increases in oil, such as tobacco leaves with up to 30% on a dry-weight basis of storage lipids [12].

In 1R22, the main FN mutant from this study, 20 gene models are located within a deletion that co-segregates with the high-oil phenotype. Eighteen of these genes have known functions but do not have an obvious fit to the high-seed-oil mutant phenotype. One of the unknown genes, *Glyma.14G102100*, is predicted to be a transposon ty3-g gag-pol polyprotein with 99.4% confidence and classified as a DNA-binding protein by Phyre2 (Table 7). We have no reason to suspect that this gene is involved in the mutant high-oil trait. The other unknown gene, *Glyma.14G101900*, has a COOH terminus predicted to be

inside the plasma membrane, while the H2N terminal is predicted to be extra-cellular. A signal peptide analysis also showed that *Glyma.14G101900* has no signal peptide (Figure 3). Some articles state that disordered proteins mainly trigger cellular stress responses or affect protein interaction networks. Ma et al. [139] stated that the deletion of such a disordered region enhances oil accumulation in *Arabidopsis*. The hydrophobicity surface and other views of the *Glyma.14G101900* and *Glyma.14G102100* PDB model predicted by Phyre2 and is illustrated via Chimera (Figures 5 and 6).

Protein network analysis performed by the STRING database indicates that *Glyma.14G101900* has an interaction with four main proteins (STRING identifiers: I1L6A3, I1MQK0, A0A0R0F173, and I1L3W9), as shown in Figure 4. I1L3W9 is an uncharacterized protein that belongs to the short-chain dehydrogenase/reductase (SDR) family. A0A0R0F173 is AB hydrolase-1 domain-containing protein. I1MQK0 is also an uncharacterized protein that belongs to the short-chain dehydrogenases/reductases (SDR) family. SDR enzymes have critical roles in lipid metabolism [140].

I1L6A3 is a Seipin 1A that has a role in lipid droplet formation and storage, and it is necessary for both adipogenesis and lipid droplet (LD) organization [141]. At this stage, we do not know how a deletion of *Glyma.14G101900* would change the interactions between these proteins to produce high oil. It is possible that *Glyma.14G101900* has a negative interaction with these proteins, specifically Seipin, such that its deletion increases oil production.

We also speculate that one of the proteins in the deleted section may have a role in reducing triacylglycerol biosynthesis. Thus, many pathway analyses were performed, but no clear role of any of the gene products in triacylglycerol biosynthesis has been uncovered so far. The investigations of possible functions of these putative genes (especially from available RNA-seq data) are summarized in Table 8.

There seem to be several plausible hypotheses for how this deletion event may be influencing oil content in seeds; however, further studies are needed to confirm this. First, it would be useful to establish a genetic marker for this deletion to confirm that this location is the source of the high-oil phenotype. This could be a simple PCR amplicon across the deletion boundaries, which would provide a PCR product in plants that carry the deletion and no product in plants that do not carry the deletion. This would be analogous to a *VgDGAT* marker that was used to track a transgene conferring a high-oil phenotype [136,137]. In other CGH mutant lines, this method was used to confirm a *FAD2* gene deletion, and the high oleic acid content correlated directly with the presence or absence of a PCR marker [17]. In addition, cloning of the genes in this region and inserting them into the mutant via transgenesis may be able to rescue the wild-type phenotype and precisely pinpoint the gene responsible for increased oil content. In the future, a full functional analysis could systematically reveal the candidate gene or genes within this deletion that underly the changes in phenotype. Ultimately, breeders and growers are not very concerned with the exact gene or mechanism which increases oil content; they know only that change is established as heritable, is easily identifiable with standard assays, and has limited other detrimental effects on the phenotype. Ultimately, it seems that the loss of a specific gene or genes within the deleted 300 kb region on chromosome 14 is responsible for the high-oil phenotype, and a further analysis could pinpoint the ultimate cause of these changes.

Author Contributions: Conceptualization, W.R.S. and D.H.; methodology, D.H., R.M.S. and W.R.S.; investigation, W.R.S., M.F.S.G., A.O.S. and P.R.A.; data curation W.R.S., A.O.S. and P.R.A.; writing—original draft preparation, W.R.S. and M.F.S.G.; writing—review and editing, W.R.S., D.H. and R.M.S.; visualization, R.M.S.; supervision, D.H. and R.M.S.; project administration, R.M.S. and D.H.; funding acquisition, D.H. and R.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for DNA extraction and the oil and protein analysis was provided by the Kentucky Soybean Board, and CGH analysis was provided by the United Soybean Board, Project #1520-532-5603.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All relevant and pertinent data are made available in the manuscript. We encourage anyone with questions to contact William Serson for further information: wrs5272@psu.edu.

Acknowledgments: We would like to thank Evie Beckert, Kai Su, and Jeff Roessler for their assistance in the maintenance and harvesting of these soybeans in the field, as well as in the oil and protein analysis, and the University of Kentucky farm crew at Spindletop Farm for preparing the land. The mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal-opportunity provider and employer.

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

References

- Goldberg, R.B.; Barker, S.J.; Perez-Grau, L. Regulation of gene expression during plant embryogenesis. *Cell* **1989**, *56*, 149–160. [[CrossRef](#)] [[PubMed](#)]
- Harwood, J.L.; Page, R.A. Biochemistry of oil synthesis. In *Designer Oil Crops*; Murphy, D.J., Ed.; VCH: Weinheim, Germany, 1994; pp. 165–194.
- Le, B.H.; Wagmaister, J.A.; Kawashima, T.; Bui, A.Q.; Harada, J.J.; Goldberg, R.B. Using genomics to study legume seed development. *Plant Physiol.* **2007**, *144*, 562–574. [[CrossRef](#)] [[PubMed](#)]
- Li, R.; Hatanaka, T.; Yu, K.; Wu, Y.; Fukushige, H.; Hildebrand, D. Soybean oil biosynthesis: Role of diacylglycerol acyltransferases. *Funct. Integr. Genom.* **2013**, *13*, 99–113. [[CrossRef](#)] [[PubMed](#)]
- Flyckt, K.S.; Roesler, K.; Haug Collet, K.; Jaureguy, L.; Booth, R.; Thatcher, S.R.; Everard, J.D.; Ripp, K.G.; Liu, Z.-B.; Shen, B. A Novel Soybean Diacylglycerol Acyltransferase 1b Variant with Three Amino Acid Substitutions Increases Seed Oil Content. *Plant Cell Physiol.* **2023**, *65*, 872–884. [[CrossRef](#)] [[PubMed](#)]
- Torabi, S.; Sukumaran, A.; Dhaubhadel, S.; Johnson, S.E.; LaFayette, P.; Parrott, W.A.; Rajcan, I.; Eskandari, M. Effects of type I Diacylglycerol O-acyltransferase (DGAT1) genes on soybean (*Glycine max* L.) seed composition. *Sci. Rep.* **2021**, *11*, 2556. [[CrossRef](#)] [[PubMed](#)]
- Rao, S.; Hildebrand, D. Changes in Oil Content of Transgenic Soybeans Expressing the Yeast SLC1 Gene. *Lipids* **2009**, *44*, 945–951. [[CrossRef](#)] [[PubMed](#)]
- Taylor, D.C.; Yan, Z.; Kumar, A.; Francis, T.; Giblin, E.M.; Barton, D.L.; Ferrie, J.R.; Laroche, A.; Shah, S.; Weiming, Z.; et al. Molecular modification of triacylglycerol accumulation by over-expression of DGAT1 to produce canola with increased seed oil content under field conditions. *Botany* **2009**, *87*, 533–543. [[CrossRef](#)]
- Andrianov, V.; Borisjuk, N.; Pogrebnyak, N.; Brinker, A.; Dixon, J.; Spitsin, S.; Flynn, J.; Matyszczyk, P.; Andryszak, K.; Laurelli, M.; et al. Tobacco as a production platform for biofuel: Overexpression of *Arabidopsis* DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.* **2010**, *8*, 277–287. [[CrossRef](#)] [[PubMed](#)]
- Van Erp, H.; Kelly, A.A.; Menard, G.; Eastmond, P.J. Multigene engineering of triacylglycerol metabolism boosts seed oil content in *Arabidopsis*. *Plant Physiol.* **2014**, *165*, 30–36. [[CrossRef](#)]
- Vanhercke, T.; El Tahchy, A.; Shrestha, P.; Zhou, X.-R.; Singh, S.P.; Petrie, J.R. Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants. *FEBS Lett.* **2013**, *587*, 364–369. [[CrossRef](#)]
- Vanhercke, T.; El Tahchy, A.; Liu, Q.; Zhou, X.-R.; Shrestha, P.; Divi, U.K.; Ral, J.-P.; Mansour, M.P.; Nichols, P.D.; James, C.N.; et al. Metabolic engineering of biomass for high energy density: Oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnol. J.* **2014**, *12*, 231–239. [[CrossRef](#)] [[PubMed](#)]
- Rotundo, J.L.; Borrás, L.; Westgate, M.E. Linking assimilate supply and seed developmental processes that determine soybean seed composition. *Eur. J. Agron.* **2011**, *35*, 184–191. [[CrossRef](#)]
- Carter, T.E.; Rzewnicki, P.E.; Burton, J.W.; Villagarcia, M.R.; Bowman, D.T.; Taliercio, E.; Kwanyuen, P. Registration of N6202 soybean germplasm with high protein, favorable yield potential, large seed, and diverse pedigree. *J. Plant Regist.* **2010**, *4*, 73–79. [[CrossRef](#)]
- Vollmann, J.; Fritz, C.N.; Wagenstrahl, H.; Ruckebauer, P. Environmental and genetic variation of soybean seed protein content under Central European growing conditions. *J. Sci. Food Agric.* **2000**, *80*, 1300–1306. [[CrossRef](#)]
- Bezie, Y.; Tilahun, T.; Atnaf, M.; Taye, M. The potential applications of site-directed mutagenesis for crop improvement: A review. *J. Crop Sci. Biotechnol.* **2021**, *24*, 229–244. [[CrossRef](#)]
- Bolon, Y.-T.; Haun, W.J.; Xu, W.W.; Grant, D.; Stacey, M.G.; Nelson, R.T.; Gerhardt, D.J.; Jeddeloh, J.A.; Stacey, G.; Muehlbauer, G.J.; et al. Phenotypic and Genomic Analyses of a Fast Neutron Mutant Population Resource in Soybean. *Plant Physiol.* **2011**, *156*, 240–253. [[CrossRef](#)] [[PubMed](#)]

18. Dobbels, A.A.; Michno, J.-M.; Campbell, B.W.; Viridi, K.S.; Stec, A.O.; Muehlbauer, G.J.; Naeve, S.L.; Stupar, R.M. An induced chromosomal translocation in soybean disrupts a KASI ortholog and is associated with a high-sucrose and low-oil seed phenotype. *G3 Genes Genomes Genet.* **2017**, *7*, 1215–1223. [[CrossRef](#)] [[PubMed](#)]
19. Wyant, S.R.; Rodriguez, M.F.; Carter, C.K.; Parrott, W.A.; Jackson, S.A.; Stupar, R.M.; Morrell, P.L. Fast neutron mutagenesis in soybean enriches for small indels and creates frameshift mutations. *G3 Genes Genomes Genet.* **2022**, *12*, jkab431. [[CrossRef](#)]
20. Campbell, B.W.; Hofstad, A.N.; Sreekanta, S.; Fu, F.; Kono, T.J.Y.; O'Rourke, J.A.; Vance, C.P.; Muehlbauer, G.J.; Stupar, R.M. Fast neutron-induced structural rearrangements at a soybean NAP1 locus result in gnarled trichomes. *Theor. Appl. Genet.* **2016**, *129*, 1725–1738. [[CrossRef](#)]
21. Prenger, E.M.; Ostezan, A.; Mian, M.A.R.; Stupar, R.M.; Glenn, T.; Li, Z. Identification and characterization of a fast-neutron-induced mutant with elevated seed protein content in soybean. *Theor. Appl. Genet.* **2019**, *132*, 2965–2983. [[CrossRef](#)]
22. Ostezan, A.; Prenger, E.M.; Rosso, L.; Zhang, B.; Stupar, R.M.; Glenn, T.; Mian, M.A.R.; Li, Z. A chromosome 16 deletion conferring a high sucrose phenotype in soybean. *Theor. Appl. Genet.* **2023**, *136*, 109. [[CrossRef](#)] [[PubMed](#)]
23. Islam, N.; Stupar, R.M.; Qijian, S.; Luthria, D.L.; Garrett, W.; Stec, A.O.; Roessler, J.; Natarajan, S.S. Genomic changes and biochemical alterations of seed protein and oil content in a subset of fast neutron induced soybean mutants. *BMC Plant Biol.* **2019**, *19*, 420. [[CrossRef](#)] [[PubMed](#)]
24. Bolon, Y.-T.; Stec, A.O.; Michno, J.-M.; Roessler, J.; Bhaskar, P.B.; Ries, L.; Dobbels, A.A.; Campbell, B.W.; Young, N.P.; Anderson, J.E.; et al. Genome Resilience and Prevalence of Segmental Duplications Following Fast Neutron Irradiation of Soybean. *Genetics* **2014**, *198*, 967–981. [[CrossRef](#)] [[PubMed](#)]
25. Armstrong, P.R.; Tallada, J.G.; Hurburgh, C.; Hildebrand, D.F.; Specht, J.E. Development of single-seed near-infrared spectroscopic predictions of corn and soybean constituents using bulk reference values and mean spectra. *Trans. ASABE* **2011**, *54*, 1529–1535. [[CrossRef](#)]
26. Jiang, G.-L. Comparison and application of non-destructive NIR evaluations of seed protein and oil content in soybean breeding. *Agronomy* **2020**, *10*, 77. [[CrossRef](#)]
27. Serson, W.; Armstrong, P.; Maghirang, E.; Al-Bakri, A.; Phillips, T.; Al-Amery, M.; Su, K.; Hildebrand, D. Development of Whole and Ground Seed Near-Infrared Spectroscopy Calibrations for Oil, Protein, Moisture, and Fatty Acids in *Salvia hispanica*. *J. Am. Oil Chem. Soc.* **2020**, *97*, 3–13. [[CrossRef](#)]
28. Orf, J.H.; Denny, R.L. Registration of 'MN1302' Soybean. *Crop Sci.* **2004**, *44*, 693. [[CrossRef](#)]
29. Till, B.J.; Jankowicz-Cieslak, J.; Huynh, O.A.; Beshir, M.M.; Laport, R.G.; Hofinger, B.J.; Till, B.J.; Jankowicz-Cieslak, J.; Huynh, O.A.; Beshir, M.M. Sample collection and storage. In *Low-Cost Methods for Molecular Characterization of Mutant Plants: Tissue Desiccation, DNA Extraction and Mutation Discovery: Protocols*; Till, B.J., Jankowicz-Cieslak, J., Huynh, O.A., Beshir, M.M., Laport, R.G., Hofinger, B.J., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 9–11.
30. Omasits, U.; Ahrens, C.H.; Müller, S.; Wollscheid, B. Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **2014**, *30*, 884–886. [[CrossRef](#)]
31. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S. The STRING database in 2023: Protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [[CrossRef](#)] [[PubMed](#)]
32. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
33. Yang, Z.; Lasker, K.; Schneidman-Duhovny, D.; Webb, B.; Huang, C.C.; Pettersen, E.F.; Goddard, T.D.; Meng, E.C.; Sali, A.; Ferrin, T.E. UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *J. Struct. Biol.* **2012**, *179*, 269–278. [[CrossRef](#)] [[PubMed](#)]
34. Turek, I.; Wheeler, J.; Bartels, S.; Szczurek, J.; Wang, Y.H.; Taylor, P.; Gehring, C.; Irving, H. A natriuretic peptide from *Arabidopsis thaliana* (AtPNP-A) can modulate catalase 2 activity. *Sci. Rep.* **2020**, *10*, 19632. [[CrossRef](#)] [[PubMed](#)]
35. Eudes, A.; Kunji, E.R.; Noiri, A.; Klaus, S.M.; Vickers, T.J.; Beverley, S.M.; Gregory, J.F.; Hanson, A.D. Identification of transport-critical residues in a folate transporter from the folate-biopterin transporter (FBT) family. *J. Biol. Chem.* **2010**, *285*, 2867–2875. [[CrossRef](#)] [[PubMed](#)]
36. Ozyigit, I.I.; Filiz, E.; Vatanserver, R.; Kurtoglu, K.Y.; Koc, I.; Öztürk, M.X.; Anjum, N.A. Identification and comparative analysis of H₂O₂-scavenging enzymes (ascorbate peroxidase and glutathione peroxidase) in selected plants employing bioinformatics approaches. *Front. Plant Sci.* **2016**, *7*, 301. [[CrossRef](#)] [[PubMed](#)]
37. Chang, C.C.; Slesak, I.; Jordá, L.; Sotnikov, A.; Melzer, M.; Miszalski, Z.; Mullineaux, P.M.; Parker, J.E.; Karpinska, B.; Karpinski, S. Arabidopsis chloroplastic glutathione peroxidases play a role in cross talk between photooxidative stress and immune responses. *Plant Physiol.* **2009**, *150*, 670–683. [[CrossRef](#)] [[PubMed](#)]
38. Santamaría, M.E.; Arnaiz, A.; Velasco-Arroyo, B.; Grbic, V.; Diaz, I.; Martinez, M. Arabidopsis response to the spider mite *Tetranychus urticae* depends on the regulation of reactive oxygen species homeostasis. *Sci. Rep.* **2018**, *8*, 9432. [[CrossRef](#)] [[PubMed](#)]
39. Sun, X.; Matus, J.T.; Wong, D.C.J.; Wang, Z.; Chai, F.; Zhang, L.; Fang, T.; Zhao, L.; Wang, Y.; Han, Y. The GARP/MYB-related grape transcription factor AQUILO improves cold tolerance and promotes the accumulation of raffinose family oligosaccharides. *J. Exp. Bot.* **2018**, *69*, 1749–1764. [[CrossRef](#)] [[PubMed](#)]

40. Koschmieder, J.; Wüst, F.; Schaub, P.; Álvarez, D.; Trautmann, D.; Krischke, M.; Rustenholz, C.; Mano, J.i.; Mueller, M.J.; Bartels, D. Plant apocarotenoid metabolism utilizes defense mechanisms against reactive carbonyl species and xenobiotics. *Plant Physiol.* **2021**, *185*, 331–351. [[CrossRef](#)] [[PubMed](#)]
41. Zhang, Q.; Dai, X.; Wang, H.; Wang, F.; Tang, D.; Jiang, C.; Zhang, X.; Guo, W.; Lei, Y.; Ma, C. Transcriptomic Profiling Provides Molecular Insights into Hydrogen Peroxide-Enhanced Arabidopsis Growth and Its Salt Tolerance. *Front. Plant Sci.* **2022**, *13*, 866063. [[CrossRef](#)]
42. Lazzarotto, F.; Wahni, K.; Piovesana, M.; Maraschin, F.; Messens, J.; Margis-Pinheiro, M. Arabidopsis APx-R is a plastidial ascorbate-independent peroxidase regulated by photomorphogenesis. *Antioxidants* **2021**, *10*, 65. [[CrossRef](#)]
43. Stock, J.; Bräutigam, A.; Melzer, M.; Bienert, G.P.; Bunk, B.; Nagel, M.; Overmann, J.; Keller, E.J.; Mock, H.-P. The transcription factor WRKY22 is required during cryo-stress acclimation in Arabidopsis shoot tips. *J. Exp. Bot.* **2020**, *71*, 4993–5009. [[CrossRef](#)] [[PubMed](#)]
44. Passaia, G.; Queval, G.; Bai, J.; Margis-Pinheiro, M.; Foyer, C.H. The effects of redox controls mediated by glutathione peroxidases on root architecture in *Arabidopsis thaliana*. *J. Exp. Bot.* **2014**, *65*, 1403–1413. [[CrossRef](#)] [[PubMed](#)]
45. Luo, M.; Liu, X.; Su, H.; Li, M.; Li, M.; Wei, J. Regulatory Networks of Flowering Genes in *Angelica sinensis* during Vernalization. *Plants* **2022**, *11*, 1355. [[CrossRef](#)] [[PubMed](#)]
46. Jia, F.; Gampala, S.S.; Mittal, A.; Luo, Q.; Rock, C.D. Cre-lox univector acceptor vectors for functional screening in protoplasts: Analysis of Arabidopsis donor cDNAs encoding ABSCISIC ACID INSENSITIVE1-like protein phosphatases. *Plant Mol. Biol.* **2009**, *70*, 693–708. [[CrossRef](#)]
47. Suzuki, M.; Ketterling, M.G.; Li, Q.-B.; McCarty, D.R. Viviparous1 alters global gene expression patterns through regulation of abscisic acid signaling. *Plant Physiol.* **2003**, *132*, 1664–1677. [[CrossRef](#)] [[PubMed](#)]
48. Ramel, F.; Sulmon, C.; Cabello-Hurtado, F.; Tacconat, L.; Martin-Magniette, M.-L.; Renou, J.-P.; El Amrani, A.; Couée, I.; Gouesbet, G. Genome-wide interacting effects of sucrose and herbicide-mediated stress in *Arabidopsis thaliana*: Novel insights into atrazine toxicity and sucrose-induced tolerance. *BMC Genom.* **2007**, *8*, 450. [[CrossRef](#)] [[PubMed](#)]
49. Dong, P.; Xiong, F.; Que, Y.; Wang, K.; Yu, L.; Li, Z.; Ren, M. Expression profiling and functional analysis reveals that TOR is a key player in regulating photosynthesis and phytohormone signaling pathways in Arabidopsis. *Front. Plant Sci.* **2015**, *6*, 677. [[CrossRef](#)]
50. Leonhardt, N.; Kwak, J.M.; Robert, N.; Waner, D.; Leonhardt, G.; Schroeder, J.I. Microarray expression analyses of Arabidopsis guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. *Plant Cell* **2004**, *16*, 596–615. [[CrossRef](#)]
51. Luna, E.; Van Hulten, M.; Zhang, Y.; Berkowitz, O.; López, A.; Pétriacq, P.; Sellwood, M.A.; Chen, B.; Burrell, M.; Van De Meene, A. Plant perception of β -aminobutyric acid is mediated by an aspartyl-tRNA synthetase. *Nat. Chem. Biol.* **2014**, *10*, 450–456. [[CrossRef](#)]
52. Fabro, G.; Di Rienzo, J.A.; Voigt, C.A.; Savchenko, T.; Dehesh, K.; Somerville, S.; Alvarez, M.E. Genome-wide expression profiling Arabidopsis at the stage of Glovinomyces cichoracearum haustorium formation. *Plant Physiol.* **2008**, *146*, 1421–1439. [[CrossRef](#)]
53. Bai, B.; Van Der Horst, S.; Cordewener, J.H.; America, T.A.; Hanson, J.; Bentsink, L. Seed-stored mRNAs that are specifically associated to monosomes are translationally regulated during germination. *Plant Physiol.* **2020**, *182*, 378–392. [[CrossRef](#)] [[PubMed](#)]
54. Schwarzenbacher, R.E.; Wardell, G.; Stassen, J.; Guest, E.; Zhang, P.; Luna, E.; Ton, J. The IB1 receptor of β -aminobutyric acid interacts with VOZ transcription factors to regulate abscisic acid signaling and callose-associated defense. *Mol. Plant* **2020**, *13*, 1455–1469. [[CrossRef](#)] [[PubMed](#)]
55. Knesting, J.; Riondet, C.; Maria, C.; Kruse, I.; Bécuwe, N.; König, N.; Berndt, C.; Tourrette, S.; Guillemot-Montoya, J.; Herrero, E. Arabidopsis glutaredoxin S17 and its partner, the nuclear factor Y subunit C11/negative cofactor 2 α , contribute to maintenance of the shoot apical meristem under long-day photoperiod. *Plant Physiol.* **2015**, *167*, 1643–1658. [[CrossRef](#)] [[PubMed](#)]
56. Duchêne, A.-M.; Giritch, A.; Hoffmann, B.; Cognat, V.; Lancelin, D.; Peeters, N.M.; Zaepfel, M.; Maréchal-Drouard, L.; Small, I.D. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16484–16489. [[CrossRef](#)] [[PubMed](#)]
57. Waterworth, W.M.; Latham, R.; Wang, D.; Alsharif, M.; West, C.E. Seed DNA damage responses promote germination and growth in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2202172119. [[CrossRef](#)] [[PubMed](#)]
58. Mróz, T.L.; Eves-van den Akker, S.; Bernat, A.; Skarzyńska, A.; Prysycz, L.; Olberg, M.; Havey, M.J.; Bartoszewski, G. Transcriptome analyses of mosaic (MSC) mitochondrial mutants of cucumber in a highly inbred nuclear background. *G3 Genes Genomes Genet.* **2018**, *8*, 953–965. [[CrossRef](#)] [[PubMed](#)]
59. Dietzen, C.; Koprivova, A.; Whitcomb, S.J.; Langen, G.; Jobe, T.O.; Hoefgen, R.; Kopriva, S. The transcription factor EIL1 participates in the regulation of sulfur-deficiency response. *Plant Physiol.* **2020**, *184*, 2120–2136. [[CrossRef](#)] [[PubMed](#)]
60. Kim, J.-S.; Lim, J.Y.; Shin, H.; Kim, B.-G.; Yoo, S.-D.; Kim, W.T.; Huh, J.H. ROS1-dependent DNA demethylation is required for ABA-inducible NIC3 expression. *Plant Physiol.* **2019**, *179*, 1810–1821. [[CrossRef](#)] [[PubMed](#)]
61. Rasheed, S.; Bashir, K.; Kim, J.-M.; Ando, M.; Tanaka, M.; Seki, M. The modulation of acetic acid pathway genes in Arabidopsis improves survival under drought stress. *Sci. Rep.* **2018**, *8*, 7831. [[CrossRef](#)]
62. Lian, J.-l.; Ren, L.-S.; Zhang, C.; Yu, C.-Y.; Huang, Z.; Xu, A.-X.; Dong, J.-G. How exposure to ALS-inhibiting gametocide tribenuron-methyl induces male sterility in rapeseed. *BMC Plant Biol.* **2019**, *19*, 124. [[CrossRef](#)]

63. Lange, H.; Holec, S.; Cognat, V.; Pieuchot, L.; Le Ret, M.; Canaday, J.; Gagliardi, D. Degradation of a polyadenylated rRNA maturation by-product involves one of the three RRP6-like proteins in *Arabidopsis thaliana*. *Mol. Cell. Biol.* **2008**, *28*, 3038–3044. [[CrossRef](#)] [[PubMed](#)]
64. Guo, X.; Wang, Y.; Hou, Y.; Zhou, Z.; Sun, R.; Qin, T.; Wang, K.; Liu, F.; Wang, Y.; Huang, Z. Genome-wide dissection of the genetic basis for drought tolerance in *Gossypium hirsutum* L. races. *Front. Plant Sci.* **2022**, *13*, 876095. [[CrossRef](#)] [[PubMed](#)]
65. Lange, H.; Gagliardi, D. Catalytic activities, molecular connections, and biological functions of plant RNA exosome complexes. *Plant Cell* **2022**, *34*, 967–988. [[CrossRef](#)] [[PubMed](#)]
66. Duruflé, H.; Ranocha, P.; Balliau, T.; Zivy, M.; Albenne, C.; Burlat, V.; Déjean, S.; Jamet, E.; Dunand, C. An integrative study showing the adaptation to sub-optimal growth conditions of natural populations of *Arabidopsis thaliana*: A focus on cell wall changes. *Cells* **2020**, *9*, 2249. [[CrossRef](#)] [[PubMed](#)]
67. Pasoreck, E.K.; Su, J.; Silverman, I.M.; Gosai, S.J.; Gregory, B.D.; Yuan, J.S.; Daniell, H. Terpene metabolic engineering via nuclear or chloroplast genomes profoundly and globally impacts off-target pathways through metabolite signalling. *Plant Biotechnol. J.* **2016**, *14*, 1862–1875. [[CrossRef](#)]
68. Roomi, S.; Masi, A.; Conselvan, G.B.; Trevisan, S.; Quaggiotti, S.; Pivato, M.; Arrigoni, G.; Yasmin, T.; Carletti, P. Protein profiling of *Arabidopsis* roots treated with humic substances: Insights into the metabolic and interactome networks. *Front. Plant Sci.* **2018**, *9*, 1812. [[CrossRef](#)]
69. Nintemann, S.J.; Vik, D.; Svozil, J.; Bak, M.; Baerenfaller, K.; Burow, M.; Halkier, B.A. Unravelling protein-protein interaction networks linked to aliphatic and indole glucosinolate biosynthetic pathways in *Arabidopsis*. *Front. Plant Sci.* **2017**, *8*, 2028. [[CrossRef](#)]
70. Haga, N.; Kobayashi, K.; Suzuki, T.; Maeo, K.; Kubo, M.; Ohtani, M.; Mitsuda, N.; Demura, T.; Nakamura, K.; Jürgens, G. Mutations in MYB3R1 and MYB3R4 cause pleiotropic developmental defects and preferential down-regulation of multiple G2/M-specific genes in *Arabidopsis*. *Plant Physiol.* **2011**, *157*, 706–717. [[CrossRef](#)]
71. Teaster, N.D.; Motes, C.M.; Tang, Y.; Wiant, W.C.; Cotter, M.Q.; Wang, Y.-S.; Kilaru, A.; Venables, B.J.; Hasenstein, K.H.; Gonzalez, G. N-Acylethanolamine metabolism interacts with abscisic acid signaling in *Arabidopsis thaliana* seedlings. *Plant Cell* **2007**, *19*, 2454–2469. [[CrossRef](#)]
72. Shaar-Moshe, L.; Hübner, S.; Peleg, Z. Identification of conserved drought-adaptive genes using a cross-species meta-analysis approach. *BMC Plant Biol.* **2015**, *15*, 111. [[CrossRef](#)]
73. Movahedi, S.; Van de Peer, Y.; Vandepoele, K. Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in *Arabidopsis* and rice. *Plant Physiol.* **2011**, *156*, 1316–1330. [[CrossRef](#)] [[PubMed](#)]
74. Verdier, J.; Lalanne, D.; Pelletier, S.; Torres-Jerez, I.; Righetti, K.; Bandyopadhyay, K.; Leprince, O.; Chatelain, E.; Vu, B.L.; Gouzy, J. A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* **2013**, *163*, 757–774. [[CrossRef](#)] [[PubMed](#)]
75. Amil-Ruiz, F.; Garrido-Gala, J.; Gadea, J.; Blanco-Portales, R.; Muñoz-Mérida, A.; Trelles, O.; de Los Santos, B.; Arroyo, F.T.; Aguado-Puig, A.; Romero, F. Partial activation of SA- and JA-defensive pathways in strawberry upon *Colletotrichum acutatum* interaction. *Front. Plant Sci.* **2016**, *7*, 1036. [[CrossRef](#)] [[PubMed](#)]
76. Yin, Z.; Balmant, K.; Geng, S.; Zhu, N.; Zhang, T.; Dufresne, C.; Dai, S.; Chen, S. Bicarbonate induced redox proteome changes in *Arabidopsis* suspension cells. *Front. Plant Sci.* **2017**, *8*, 58. [[CrossRef](#)] [[PubMed](#)]
77. Jung, S.; Main, D.; Staton, M.; Cho, I.; Zhebentyayeva, T.; Arús, P.; Abbott, A. Synteny conservation between the *Prunus* genome and both the present and ancestral *Arabidopsis* genomes. *BMC Genom.* **2006**, *7*, 81. [[CrossRef](#)] [[PubMed](#)]
78. Song, D.; Xi, W.; Shen, J.; Bi, T.; Li, L. Characterization of the plasma membrane proteins and receptor-like kinases associated with secondary vascular differentiation in poplar. *Plant Mol. Biol.* **2011**, *76*, 97–115. [[CrossRef](#)] [[PubMed](#)]
79. Petridis, A.; Döll, S.; Nichelmann, L.; Bilger, W.; Mock, H.P. *Arabidopsis thaliana* G2-LIKE FLAVONOID REGULATOR and BRASSINOSTEROID ENHANCED EXPRESSION1 are low-temperature regulators of flavonoid accumulation. *New Phytol.* **2016**, *211*, 912–925. [[CrossRef](#)] [[PubMed](#)]
80. Yang, M.; Yang, H.; Kuang, R.; Zhou, C.; Huang, B.; Wei, Y. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in papaya (*Carica papaya* L.). *PeerJ* **2019**, *8*, e9319. [[CrossRef](#)] [[PubMed](#)]
81. Cifuentes-Esquivel, N.; Bou-Torrent, J.; Galstyan, A.; Gallemí, M.; Sessa, G.; Salla Martret, M.; Roig-Villanova, I.; Ruberti, I.; Martínez-García, J.F. The bHLH proteins BEE and BIM positively modulate the shade avoidance syndrome in *Arabidopsis* seedlings. *Plant J.* **2013**, *75*, 989–1002. [[CrossRef](#)]
82. Pathak, A.K.; Singh, S.P.; Sharma, R.; Nath, V.; Tuli, R. Transcriptome analysis at mid-stage seed development in litchi with contrasting seed size. *3 Biotech* **2022**, *12*, 47. [[CrossRef](#)]
83. Yuan, L.-B.; Chen, L.; Zhai, N.; Zhou, Y.; Zhao, S.-S.; Shi, L.-L.; Xiao, S.; Yu, L.-J.; Xie, L.-J. The anaerobic product ethanol promotes autophagy-dependent submergence tolerance in *Arabidopsis*. *Int. J. Mol. Sci.* **2020**, *21*, 7361. [[CrossRef](#)] [[PubMed](#)]
84. Friedrichsen, D.M.; Nemhauser, J.; Muramitsu, T.; Maloof, J.N.; Alonso, J.; Ecker, J.R.; Furuya, M.; Chory, J. Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* **2002**, *162*, 1445–1456. [[CrossRef](#)] [[PubMed](#)]

85. Liu, H.; Xiao, S.; Sui, S.; Huang, R.; Wang, X.; Wu, H.; Liu, X. A tandem CCCH type zinc finger protein gene CpC3H3 from *Chimonanthus praecox* promotes flowering and enhances drought tolerance in *Arabidopsis*. *BMC Plant Biol.* **2022**, *22*, 506. [[CrossRef](#)] [[PubMed](#)]
86. Wang, F.; Gao, Y.; Liu, Y.; Zhang, X.; Gu, X.; Ma, D.; Zhao, Z.; Yuan, Z.; Xue, H.; Liu, H. BES1-regulated BEE1 controls photoperiodic flowering downstream of blue light signaling pathway in *Arabidopsis*. *New Phytol.* **2019**, *223*, 1407–1419. [[CrossRef](#)] [[PubMed](#)]
87. Cao, J.; Liang, Y.; Yan, T.; Wang, X.; Zhou, H.; Chen, C.; Zhang, Y.; Zhang, B.; Zhang, S.; Liao, J. The photomorphogenic repressors BBX28 and BBX29 integrate light and brassinosteroid signaling to inhibit seedling development in *Arabidopsis*. *Plant Cell* **2022**, *34*, 2266–2285. [[CrossRef](#)] [[PubMed](#)]
88. Zhang, X.-Y.; Qiu, J.-Y.; Hui, Q.-L.; Xu, Y.-Y.; He, Y.-Z.; Peng, L.-Z.; Fu, X.-Z. Systematic analysis of the basic/helix-loop-helix (bHLH) transcription factor family in pummelo (*Citrus grandis*) and identification of the key members involved in the response to iron deficiency. *BMC Genom.* **2020**, *21*, 233. [[CrossRef](#)] [[PubMed](#)]
89. Martínez-García, P.J.; Parfitt, D.E.; Bostock, R.M.; Fresnedo-Ramirez, J.; Vazquez-Lobo, A.; Ogundiwin, E.A.; Gradziel, T.M.; Crisosto, C.H. Application of genomic and quantitative genetic tools to identify candidate resistance genes for brown rot resistance in peach. *PLoS ONE* **2013**, *8*, e78634. [[CrossRef](#)]
90. Herbert, D.B.; Gross, T.; Rupp, O.; Becker, A. Transcriptome analysis reveals major transcriptional changes during regrowth after mowing of red clover (*Trifolium pratense*). *BMC Plant Biol.* **2021**, *21*, 95. [[CrossRef](#)] [[PubMed](#)]
91. Lestari, P.; Van, K.; Lee, J.; Kang, Y.J.; Lee, S.-H. Gene divergence of homeologous regions associated with a major seed protein content QTL in soybean. *Front. Plant Sci.* **2013**, *4*, 176. [[CrossRef](#)]
92. Luttgaharm, K.D.; Chen, M.; Mehra, A.; Cahoon, R.E.; Markham, J.E.; Cahoon, E.B. Overexpression of *Arabidopsis* ceramide synthases differentially affects growth, sphingolipid metabolism, programmed cell death, and mycotoxin resistance. *Plant Physiol.* **2015**, *169*, 1108–1117. [[CrossRef](#)]
93. Corral, J.M.; Vogel, H.; Aliyu, O.M.; Hensel, G.; Thiel, T.; Kumlehn, J.; Sharbel, T.F. A conserved apomixis-specific polymorphism is correlated with exclusive exonuclease expression in premeiotic ovules of apomictic *Boechera* species. *Plant Physiol.* **2013**, *163*, 1660–1672. [[CrossRef](#)] [[PubMed](#)]
94. Vargas, J.; Gómez, I.; Vidal, E.A.; Lee, C.P.; Millar, A.H.; Jordana, X.; Roschztardt, H. Growth Developmental Defects of Mitochondrial Iron Transporter 1 and 2 Mutants in *Arabidopsis* in Iron Sufficient Conditions. *Plants* **2023**, *12*, 1176. [[CrossRef](#)]
95. Krishnatreya, D.B.; Ray, D.; Baruah, P.M.; Dowarah, B.; Bordoloi, K.S.; Agarwal, H.; Agarwala, N. Identification of putative miRNAs from expressed sequence tags of *Gnetum gnemon* L. and their cross-kingdom targets. *BioTechnologia* **2021**, *102*, 179. [[CrossRef](#)]
96. Xie, Y.; Straub, D.; Eguen, T.; Brandt, R.; Stahl, M.; Martínez-García, J.F.; Wenkel, S. Meta-analysis of *Arabidopsis* KANADI1 direct target genes identifies a basic growth-promoting module acting upstream of hormonal signaling pathways. *Plant Physiol.* **2015**, *169*, 1240–1253. [[CrossRef](#)] [[PubMed](#)]
97. Sakai, T.; Haga, K. Molecular genetic analysis of phototropism in *Arabidopsis*. *Plant Cell Physiol.* **2012**, *53*, 1517–1534. [[CrossRef](#)] [[PubMed](#)]
98. Huang, T.; Harrar, Y.; Lin, C.; Reinhart, B.; Newell, N.R.; Talavera-Rauh, F.; Hokin, S.A.; Barton, M.K.; Kerstetter, R.A. *Arabidopsis* KANADI1 acts as a transcriptional repressor by interacting with a specific cis-element and regulates auxin biosynthesis, transport, and signaling in opposition to HD-ZIP III factors. *Plant Cell* **2014**, *26*, 246–262. [[CrossRef](#)]
99. Li, Y.; Dai, X.; Cheng, Y.; Zhao, Y. NPY genes play an essential role in root gravitropic responses in *Arabidopsis*. *Mol. Plant* **2011**, *4*, 171–179. [[CrossRef](#)]
100. Cheng, Y.; Qin, G.; Dai, X.; Zhao, Y. NPY genes and AGC kinases define two key steps in auxin-mediated organogenesis in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 21017–21022. [[CrossRef](#)]
101. Cheng, Y.; Qin, G.; Dai, X.; Zhao, Y. NPY1, a BTB-NPH3-like protein, plays a critical role in auxin-regulated organogenesis in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18825–18829. [[CrossRef](#)]
102. Gipson, A.B.; Morton, K.J.; Rhee, R.J.; Simo, S.; Clayton, J.A.; Perrett, M.E.; Binkley, C.G.; Jensen, E.L.; Oakes, D.L.; Rouhier, M.F. Disruptions in valine degradation affect seed development and germination in *Arabidopsis*. *Plant J.* **2017**, *90*, 1029–1039. [[CrossRef](#)]
103. Zhu, F.; Alseekh, S.; Koper, K.; Tong, H.; Nikoloski, Z.; Naake, T.; Liu, H.; Yan, J.; Brotman, Y.; Wen, W. Genome-wide association of the metabolic shifts underpinning dark-induced senescence in *Arabidopsis*. *Plant Cell* **2022**, *34*, 557–578. [[CrossRef](#)] [[PubMed](#)]
104. Zhu, Q.; King, G.J.; Liu, X.; Shan, N.; Borpatragohain, P.; Baten, A.; Wang, P.; Luo, S.; Zhou, Q. Identification of SNP loci and candidate genes related to four important fatty acid composition in *Brassica napus* using genome wide association study. *PLoS ONE* **2019**, *14*, e0221578. [[CrossRef](#)] [[PubMed](#)]
105. Carrie, C.; Venne, A.S.; Zahedi, R.P.; Soll, J. Identification of cleavage sites and substrate proteins for two mitochondrial intermediate peptidases in *Arabidopsis thaliana*. *J. Exp. Bot.* **2015**, *66*, 2691–2708. [[CrossRef](#)] [[PubMed](#)]
106. Allen, A.M.; Lexer, C.; Hiscock, S.J. Comparative analysis of pistil transcriptomes reveals conserved and novel genes expressed in dry, wet, and semidry stigmas. *Plant Physiol.* **2010**, *154*, 1347–1360. [[CrossRef](#)] [[PubMed](#)]
107. Binder, S. *Branched-Chain Amino Acid Metabolism in Arabidopsis thaliana*; The *Arabidopsis* Book/American Society of Plant Biologists: Rockville, MD, USA, 2010; Volume 8.
108. Almeida, J.; Quadrana, L.; Asís, R.; Setta, N.; De Godoy, F.; Bermudez, L.; Otaiza, S.N.; Correa da Silva, J.V.; Fernie, A.R.; Carrari, F. Genetic dissection of vitamin E biosynthesis in tomato. *J. Exp. Bot.* **2011**, *62*, 3781–3798. [[CrossRef](#)] [[PubMed](#)]

109. Le Boulch, P.; Poëssel, J.-L.; Roux, D.; Lugan, R. Molecular mechanisms of resistance to *Myzus persicae* conferred by the peach *Rm2* gene: A multi-omics view. *Front. Plant Sci.* **2022**, *13*, 992544. [[CrossRef](#)] [[PubMed](#)]
110. Yokoyama, R.; de Oliveira, M.V.; Kleven, B.; Maeda, H.A. The entry reaction of the plant shikimate pathway is subjected to highly complex metabolite-mediated regulation. *Plant Cell* **2021**, *33*, 671–696. [[CrossRef](#)] [[PubMed](#)]
111. Qian, Y.; Lynch, J.H.; Guo, L.; Rhodes, D.; Morgan, J.A.; Dudareva, N. Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. *Nat. Commun.* **2019**, *10*, 15. [[CrossRef](#)]
112. Tohge, T.; Watanabe, M.; Hoefgen, R.; Fernie, A.R. Shikimate and phenylalanine biosynthesis in the green lineage. *Front. Plant Sci.* **2013**, *4*, 62. [[CrossRef](#)]
113. Tzin, V.; Galili, G. *The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in Arabidopsis thaliana*; The Arabidopsis Book/American Society of Plant Biologists: Rockville, MD, USA, 2010; Volume 8.
114. Less, H.; Galili, G. Principal transcriptional programs regulating plant amino acid metabolism in response to abiotic stresses. *Plant Physiol.* **2008**, *147*, 316–330. [[CrossRef](#)]
115. Less, H.; Galili, G. Coordinations between gene modules control the operation of plant amino acid metabolic networks. *BMC Syst. Biol.* **2009**, *3*, 14. [[CrossRef](#)] [[PubMed](#)]
116. Schmid, M.W.; Heichinger, C.; Coman Schmid, D.; Guthörl, D.; Gagliardini, V.; Bruggmann, R.; Aluri, S.; Aquino, C.; Schmid, B.; Turnbull, L.A. Contribution of epigenetic variation to adaptation in Arabidopsis. *Nat. Commun.* **2018**, *9*, 4446. [[CrossRef](#)]
117. Cvrčková, F.; Novotný, M.; Pícková, D.; Žárský, V. Formin homology 2 domains occur in multiple contexts in angiosperms. *BMC Genom.* **2004**, *5*, 44. [[CrossRef](#)] [[PubMed](#)]
118. Janssen, B.J.; Thodey, K.; Schaffer, R.J.; Alba, R.; Balakrishnan, L.; Bishop, R.; Bowen, J.H.; Crowhurst, R.N.; Gleave, A.P.; Ledger, S. Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. *BMC Plant Biol.* **2008**, *8*, 16. [[CrossRef](#)] [[PubMed](#)]
119. Shang, G.-D.; Xu, Z.-G.; Wan, M.-C.; Wang, F.-X.; Wang, J.-W. FindIT2: An R/Bioconductor package to identify influential transcription factor and targets based on multi-omics data. *BMC Genom.* **2022**, *23*, 272. [[CrossRef](#)] [[PubMed](#)]
120. Liu, Z.Q.; Shi, L.P.; Yang, S.; Qiu, S.S.; Ma, X.L.; Cai, J.S.; Guan, D.Y.; Wang, Z.H.; He, S.L. A conserved double-W box in the promoter of CaWRKY40 mediates autoregulation during response to pathogen attack and heat stress in pepper. *Mol. Plant Pathol.* **2021**, *22*, 3–18. [[CrossRef](#)]
121. Bergmann, T.; Menkhaus, J.; Ye, W.; Schemmel, M.; Hasler, M.; Rietz, S.; Leckband, G.; Cai, D. QTL mapping and transcriptome analysis identify novel QTLs and candidate genes in Brassica villosa for quantitative resistance against Sclerotinia sclerotiorum. *Theor. Appl. Genet.* **2023**, *136*, 86. [[CrossRef](#)]
122. Xu, X.; Chen, C.; Fan, B.; Chen, Z. Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell* **2006**, *18*, 1310–1326. [[CrossRef](#)] [[PubMed](#)]
123. Kashima, M.; Kamitani, M.; Nomura, Y.; Mori-Moriyama, N.; Betsuyaku, S.; Hirata, H.; Nagano, A.J. DeLTa-Seq: Direct-lysate targeted RNA-Seq from crude tissue lysate. *Plant Methods* **2022**, *18*, 99. [[CrossRef](#)]
124. Che, Y.; Sun, Y.; Lu, S.; Zhao, F.; Hou, L.; Liu, X. AtWRKY40 functions in drought stress response in *Arabidopsis thaliana*. *J. Plant Physiol.* **2018**, *54*, 456–464.
125. Chen, H.; Lai, Z.; Shi, J.; Xiao, Y.; Chen, Z.; Xu, X. Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol.* **2010**, *10*, 281. [[CrossRef](#)] [[PubMed](#)]
126. Carrera, C.; Martínez, M.J.; Dardanelli, J.; Balzarini, M. Environmental Variation and Correlation of Seed Components in Nontransgenic Soybeans: Protein, Oil, Unsaturated Fatty Acids, Tocopherols, and Isoflavones. *Crop Sci.* **2011**, *51*, 800–809. [[CrossRef](#)]
127. Ergo, V.V.; Veas, R.E.; Vega, C.R.C.; Lascano, R.; Carrera, C.S. Ecophysiological mechanisms underlying the positive relationship between seed protein concentration and yield in soybean under field heat and drought stress. *J. Agron. Crop Sci.* **2024**, *210*, e12703. [[CrossRef](#)]
128. Fields, J.; Saxton, A.M.; Beyl, C.A.; Kopsell, D.A.; Cregan, P.B.; Hyten, D.L.; Cuvaca, I.; Pantalone, V.R. Seed Protein and Oil QTL in a Prominent Glycine max Genetic Pedigree: Enhancing Stability for Marker Assisted Selection. *Agronomy* **2023**, *13*, 567. [[CrossRef](#)]
129. Allen, D.K.; Young, J.D. Carbon and nitrogen provisions alter the metabolic flux in developing soybean embryos. *Plant Physiol.* **2013**, *161*, 1458–1475. [[CrossRef](#)] [[PubMed](#)]
130. Pavlovic, T.; Margarit, E.; Müller, G.L.; Saenz, E.; Ruzzo, A.I.; Drincovich, M.F.; Borrás, L.; Saigo, M.; Wheeler, M.C.G. Differential metabolic reprogramming in developing soybean embryos in response to nutritional conditions and abscisic acid. *Plant Mol. Biol.* **2023**, *113*, 89–103. [[CrossRef](#)]
131. Schwender, J. Walking the ‘design–build–test–learn’ cycle: Flux analysis and genetic engineering reveal the pliability of plant central metabolism. *New Phytol.* **2023**, *239*, 1539–1541. [[CrossRef](#)] [[PubMed](#)]
132. Focks, N.; Benning, C. wrinkled1: A Novel, Low-Seed-Oil Mutant of Arabidopsis with a Deficiency in the Seed-Specific Regulation of Carbohydrate Metabolism. *Plant Physiol.* **1998**, *118*, 91–101. [[CrossRef](#)] [[PubMed](#)]
133. Jo, L.; Pelletier, J.; Harada, J.J. Genome-Wide Profiling of Soybean WRINKLED1 Transcription Factor Binding Sites Provides Insight into the Regulation of Fatty Acid and Triacylglycerol Biosynthesis Program in Seeds. *bioRxiv* **2024**. [[CrossRef](#)]
134. Sidorov, R.A.; Tsydendambaev, V.D. Biosynthesis of fatty oils in higher plants. *Russ. J. Plant Physiol.* **2014**, *61*, 1–18. [[CrossRef](#)]

135. Weselake, R.J.; Shah, S.; Tang, M.; Quant, P.A.; Snyder, C.L.; Furukawa-Stoffer, T.L.; Zhu, W.; Taylor, D.C.; Zou, J.; Kumar, A.; et al. Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (*Brassica napus*) to increase seed oil content. *J. Exp. Bot.* **2008**, *59*, 3543–3549. [[CrossRef](#)] [[PubMed](#)]
136. Roesler, K.; Shen, B.; Bermudez, E.; Li, C.; Hunt, J.; Damude, H.G.; Ripp, K.G.; Everard, J.D.; Booth, J.R.; Castaneda, L.; et al. An Improved Variant of Soybean Type 1 Diacylglycerol Acyltransferase Increases the Oil Content and Decreases the Soluble Carbohydrate Content of Soybeans. *Plant Physiol.* **2016**, *171*, 878–893. [[CrossRef](#)] [[PubMed](#)]
137. Hatanaka, T.; Serson, W.; Li, R.; Armstrong, P.; Yu, K.; Pfeiffer, T.; Li, X.-L.; Hildebrand, D. A *Vernonia* Diacylglycerol Acyltransferase Can Increase Renewable Oil Production. *J. Agric. Food Chem.* **2016**, *64*, 7188–7194. [[CrossRef](#)] [[PubMed](#)]
138. Kelly, A.A.; Shaw, E.; Powers, S.J.; Kurup, S.; Eastmond, P.J. Suppression of the SUGAR-DEPENDENT1 triacylglycerol lipase family during seed development enhances oil yield in oilseed rape (*Brassica napus* L.). *Plant Biotechnol. J.* **2013**, *11*, 355–361. [[CrossRef](#)] [[PubMed](#)]
139. Ma, W.; Kong, Q.; Grix, M.; Mantyla, J.J.; Yang, Y.; Benning, C.; Ohlrogge, J.B. Deletion of a C-terminal intrinsically disordered region of WRINKLED 1 affects its stability and enhances oil accumulation in Arabidopsis. *Plant J.* **2015**, *83*, 864–874. [[CrossRef](#)] [[PubMed](#)]
140. Kavanagh, K.L.; Jörnvall, H.; Persson, B.; Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell. Mol. Life Sci.* **2008**, *65*, 3895–3906. [[CrossRef](#)]
141. Fei, W.; Du, X.; Yang, H. Seipin, adipogenesis and lipid droplets. *Trends Endocrinol. Metab.* **2011**, *22*, 204–210. [[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.