Received: 3 February 2022 | Revised: 8 March 2022 | Accepted: 9 March 2022

DOI: 10.1002/fes3.381

Food and Energy Security Open Access WILEY

ORIGINAL ARTICLE

Integrated transcriptomic and proteomic characterization of a chromosome segment substitution line reveals a new regulatory network controlling the seed storage profile of soybean

Xue Han[1,2] | Jiapeng Li[1] | Yabin Zhao[1] | Zhanguo Zhang[1] | Hongwei Jiang[3] | Jinxing Wang[4] | Xuezhen Feng[1] | Yu Zhang[1] | Ziyue Du[1] | Xiaoxia Wu[1] | Qingshan Chen[1] | Zhaoming Qi[1]

[1]College of Agriculture, Northeast Agricultural University, Harbin, China

[2]Heilongjiang Academy of Land Reclamation Sciences, Harbin, China

[3]Soybean Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China

[4]Suihua Branch Institute, HeiLongJiang Academy of Agricultural Sciences, Suihua, China

Correspondence
Zhaoming Qi, Qingshan Chen, and Xiaoxia Wu, College of Agriculture, Northeast Agricultural University, Harbin 150030, Heilongjiang, China.
Emails: qizhaoming1860@126.com (ZQ); qshchen@126.com (QC); wuxiaoxia112233@163.com (XW)

Funding information
Natural Science Foundation of Heilongjiang-Outstanding Youth Foundation, Grant/Award Number: YQ2021C011; National Natural Science Foundation of China, Grant/Award Number: 31971968 and U20A2027; National Key R&D Program of China, Grant/Award Number: 2021YFF1001202; Young Scholar Bone Project

Abstract

Soybean is a major crop that provides oil and protein worldwide. Soybean products have been consumed for centuries in many different kinds of food. The regulatory network that controls the accumulation of fatty acids (FAs) and storage proteins (SSPs) in seeds is poorly understood. To gain new insights into the molecular mechanisms that contribute to high-quality seeds, seed FA and SSP contents were analyzed for the effects of chromosome segment substitution in a wild soybean line (CSSL). High-throughput transcriptomics and tandem mass tag (TMT)-based quantitative proteomics were performed on the dry seeds of the CSSL and parent lines. In total, 665 differentially expressed genes (DEGs) and 83 differentially accumulated proteins (DAPs) were identified. Of these 27 DEGs and 23 DAPs were found to regulate the seed storage profile. These genes encode proteins involved in photosynthesis, protein processing, protein sorting, and storage protein accumulation. Data are presented showing that FA synthesis was decreased by the regulation of SSP-accumulation-related genes and proteins. Taken together, the results provide new insights into the regulation network involved in the accumulation of seed storage compounds involving both source and sink processes that may be used as markers in future soybean breeding programs.

KEYWORDS

seed fatty acids, seed storage proteins, soybean seed, TMT-based quantitative proteomics, transcriptome

Xue Han, Jiapeng Li, and Yabin Zhao contributed equally to this work.

1 | INTRODUCTION

Soybean (*Glycine max* [L.] Merr.) originated in China, where it has long been grown as a major food crop. It accounts for 59% of the world's oilseed production and provides 44% of the plant proteins in the global food market (www.soystats.com). Soybean products such as tofu, soybean meal, okara, soy milk, and other soy-based beverages have been consumed for centuries, especially in Asia (Cai et al., 2021). In conjunction with the growing worldwide consumer demand for better-quality soybean products, the improvement of soybean seed quality has become increasingly important (Gong et al., 2020). As fatty acids (FAs) and seed storage proteins (SSPs) are two major components of soybean seeds, a better understanding of the regulatory networks that control their accumulation will contribute to high-quality soybean breeding and enhance the food market.

The synthesis of FAs and SSPs, which are two of the most essential components of soybean seeds, starts at pod filling and ends at seed drying (Krishnan, 2001). Five major types of FAs are present in soybean seeds: palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2), and linolenic acid (C18:3) (Clemente & Cahoon, 2009). Soybean oil is the primary plant oil in the global food supply and serves as a good source of essential FAs such as linoleic and α-linolenic acids (Blasbalg et al., 2011). Given the extensive use of different kinds of FAs in the food supply, special modification of soybean varieties could provide more choices with specific FA concentrations (Raatz et al., 2018). Glycinin (7S) and β-conglycinin (11S) are two major SSPs that account for approximately 70%–80% of the total protein in soybean seeds (Derbyshire et al., 1976). The soybean seed proteome also contains some moderately abundant proteins, such as Kunitz and Bowman–Birk trypsin inhibitors, lectin, sucrose-binding protein, and oleosin, as well as several thousand low-abundance protein enzymes (Herman, 2014; Herman & Burks, 2011). Soy protein can be widely used in the food industry as soy protein–based meat alternatives, soy flour, and soybean milk (He et al., 2019; Maforimbo et al., 2008; Zhang et al., 2021). A recent study also showed that soy protein could be selectively hydrolyzed to efficiently improve the quality of steamed bread and dough components (Li et al., 2021). In addition, soy protein is not only a nutritious food, it can also reduce the levels of plasma triacylglycerol and cholesterol in humans (Aoyama et al., 2001; Kito et al., 1993).

With the development of high-throughput sequencing, gas chromatography–mass spectrometry (GC–MS), liquid chromatography–MS (LC–MS), and tandem mass tag (TMT)-based MS, the integration of multiple omics approaches can reveal regulatory networks and provide new insights and more genetic resources for the food industry. Recently, TMT-based quantitative proteomic analyses have been performed on seed storage–related proteins in soybean seeds and soy milk (Battisti et al., 2021; Min et al., 2020). Furthermore, an integrated transcriptomic and proteomic analysis revealed the molecular mechanisms associated with *Coix* seed quality, including pathways related to flavonoid biosynthesis, starch and sucrose metabolism, lipid metabolism, and amino acid metabolism (Huang et al., 2019). Transcriptomics and proteomics have also recently been used to analyze regulatory networks in soybean, including those related to soybean cyst nematode resistance, flooding tolerance, and sprout yield and nutritional qualities. The collective findings of these studies indicate that the integration of multiple omics datasets can accelerate the speed and efficiency of molecular mechanism identification (Gu et al., 2017; Huang et al., 2019; Lin et al., 2019).

Chromosome segment substitution lines (CSSLs) are important genetic resources that were first reported in tomato nearly 30 years ago (Eshed et al., 1994). Ideally, each CSSL contains a single chromosomal segment from the donor genome against the background of the recurrent parent genome (Balakrishnan et al., 2019). CSSLs have been widely developed for the dissection of genetic loci in many crops, but only a few have been reported in soybean (Wang et al., 2013). In a previous study (Xin et al., 2016), we developed a CSSL population in which the Suinong 14 (SN14) genome was segmentally substituted into wild soybean ZYD00006 (*Glycine soja* Sieb. & Zucc.). In the present study, we screened out one of the CSSL lines by analyzing FA profiles and SSP contents. We then integrated genome-wide transcriptome and TMT-based proteome data to reveal a previously unknown network that regulates seed FAs and SSPs. We also analyzed the functions of novel candidate genes and proteins involved in the newly uncovered network. Our findings provide genetic resources for high-quality soybean breeding and offer new insights for the development of different types of soybean.

2 | MATERIALS AND METHODS

2.1 | Plant materials

A CSSL (R119) was screened out from our previously developed soybean CSSL population (Xin et al., 2016) based on the analysis of FA profiles and SSP contents during the 2019 and 2020 crop years in Harbin, China (45.75°N, 126.53°E). Field management practices were similar to those detailed in the work of Qi et al. (2018). After the soybean plant harvest, three replicate samples of dry seeds were selected from R119 and its recurrent parent, SN14.

2.2 | FA profiling and SSP content analysis

Fatty acids were analyzed according to the protocol of Song et al. (2013), with some modifications. In brief, soybean seeds were pulverized and filtered through a 60-mesh screen. Next, 5 mg of the soybean seed powder was mixed with 100 μg heptadecanoic acid and added to the extraction solution (2.5% [v/v] H_2SO_4 in CH_3OH). The sample solution was incubated at 85°C for 1 h and centrifuged at 5975 × *g* for 10 min, followed by the addition of 150 μl 0.9% (w/v) NaCl and 700 μl hexane. After air drying, the residual FA methyl esters were dissolved in 400 μl ethyl acetate for GC analysis (GC–MS; Agilent 7890B).

To measure SSP content, 50 mg of soybean seed powder was filtered through a 60-mesh screen, dried, and analyzed on an NDA702 Dumas analyzer (Velp Scientifica).

2.3 | Histological confirmation of soybean seed composition

To further confirm the differences in soybean seed composition, we compared the oil accumulation in SN14 and R119 seeds by Nile Red staining and observation under a C2 confocal microscope (Nikon) (Huang et al., 2019).

2.4 | High-throughput transcriptome sequencing and data analysis

Total RNA was isolated and purified using the TRIzol reagent (Invitrogen) following the manufacturer's protocol. After assessing the integrity and purity of the extracted RNA, mRNA was purified from 50 μg of total RNA and fragmented into small pieces. High-throughput sequencing was performed on an Illumina NovaSeq 6000 system to generate 150-bp paired-end reads. Analysis of the resulting RNA-seq data, including mapping of raw reads to the reference genome, counting of reads, and calculation of fragments per kilobase of exon per million fragments mapped, was performed as described previously (Qi et al., 2018). R119 and SN14 expression levels were normalized and used for differential expression analysis with DESeq2 (Love et al., 2014) using the criteria of log2 fold-change (FC) >1 or <−1 and adjusted $p < 0.05$; FC was calculated as R119/SN14. The soybean reference genome (Wm82.a2.v1) and annotation dataset were downloaded from Phytozome (https://phytozome-next.jgi.doe.gov/info/Gmax_Wm82_a2_v1) (Schmutz et al., 2010). The raw sequencing data have been deposited at the China National GeneBank database CNGBdb-EBB under project accession number CNP0002324 (https://db.cngb.org/ebb/bio_resources/; submission ID sub025402).

2.5 | TMT-based quantitative proteomics and data analysis

2.5.1 | Protein extraction and quantitation

R119 and SN14 soybean seed powders were passed through a 60-mesh screen and suspended on ice in 200 μl lysis buffer (4% SDS, 100 mM DTT, and 150 mM Tris-HCl [pH 8.0]). After ultrasonication and boiling, the supernatant was collected and quantified using a bicinchoninic acid protein assay kit (Bio-Rad). Finally, 15 μg of protein from each sample was resolved by 12% SDS-PAGE in 5:1 (v/v) 5× stacking buffer with Coomassie brilliant blue G (Sigma).

2.5.2 | Protein digestion

Protein digestion was performed according to the filter-aided sample preparation method (Wiśniewski et al., 2009). In brief, 300 μg of each sample was mixed with dithiothreitol (DTT). After boiling for 5 min, the solution was supplemented with 200 μl UA buffer (8 M urea and 150 mM Tris-HCl [pH 8.0]) and used for isolation by repeated ultrafiltration and centrifugation. Next, 100 μl UA buffer (with 50 mM iodoacetamide) was added, and the solution was incubated for 30 min in darkness at room temperature. The samples were then washed twice with 100 μl UA buffer and 100 μl of 25 mM NH_4HCO_3. Finally, 60 μl trypsin buffer (6 μg trypsin in 40 μl NH_4HCO_3 buffer) was added, and the solution was incubated at 37°C for 16–18 h. A desalting spin column (Thermo Fisher Scientific) was used to desalt the samples for peptide quantitation (Wiśniewski et al., 2009).

2.5.3 | TMT labeling of peptides

Peptides in each sample were labeled with TMT reagents according to the manufacturer's instructions (Thermo Fisher Scientific). After drying of labeled peptides in equivalent combinations, multiplex labeled samples were fractionated using a Pierce High-pH Reversed-Phase Peptide Fractionation kit (Thermo Fisher Scientific). The peptide portion of each fraction was evaporated to dryness and stored at −80°C for LC–MS analysis (Min et al., 2020).

2.5.4 | LC–MS/MS analysis

LC–MS/MS profiling was performed on a Q-Exactive HF-X mass spectrometer coupled to an Easy nLC 1200 liquid chromatograph (Thermo Fisher Scientific). After loading of peptides onto a trap column (100 μm × 20 mm, 5 μm, C18, Dr. Maisch GmbH), the peptides were separated on a chromatographic analysis column (75 μm × 150 mm, 3 μm, C18, Dr. Maisch GmbH) with a linear gradient of buffer A (2% acetonitrile and 0.1% formic acid) and buffer B (95% acetonitrile and 0.1% formic acid) at a flow rate of 300 nl/min over 90 min. The linear gradient was as follows: 0–2 min, 2%–8% buffer B; 2–71 min, 8%–28% buffer B; 71–79 min, 28%–40% buffer B; 79–83 min, 40%–100% buffer B; and 83–90 min, buffer B maintained at 90%. Data-dependent acquisition profile analysis was performed on the Q-Exactive HF-X mass spectrometer for 90 min with the following detection parameters: MS1 precursor ion scanning from 300 to 1800 m/z; MS1 resolution of 60,000 @ m/z 200; AGC target value of 3×10^6, and maximum injection time of 50 ms. The top 20 most abundant precursor ions from each full scan were dynamically chosen for MS2, which was performed with the following parameters: MS2 resolution of 45,000 @ m/z 200; AGC target value of 1×10^5, and maximum injection time of 50 ms. The MS2 activation type was the HCD model with a 1.2--m/z isolation window and normalized collision energy of 32 (Stryiński et al., 2019).

2.5.5 | Database searches and analysis

The LC–MS/MS raw data were imported into Proteome Discoverer (v2.4; Thermo Fisher Scientific) for protein identification against the file max_Soybean_3847_85057_202109.fasta in the Uniprot_Glycine database (downloaded from https://www.uniprot.org/taxonomy/3847, with 85,057 protein sequences, on 09/2021). Search details and parameters are shown in Table S1.

2.5.6 | Bioinformatics analysis

Differentially accumulated proteins (DAPs) were identified according to the criterion of FC >1.20 or <0.83 ($p < 0.05$) in Perseus (Tyanova et al., 2016) and R (https://www.r-project.org/), and then grouped by hierarchical clustering based on protein level as described previously (Min et al., 2020). Protein sequences were subsequently annotated against the UniProtKB/Swiss-Prot (https://www.uniprot.org/), Kyoto Encyclopedia of Genes and Genomes (KEGG; https://www.kegg.jp/), and Gene Ontology (GO; http://geneontology.org/) databases. GO and KEGG enrichment analyses were performed using Fisher's exact test with FDR correction. Protein–protein interaction networks were predicted using the STRING database (https://string-db.org/).

2.6 | Statistical analysis

Data in this study were statistically analyzed by one-way analysis of variance with Student's *t*-test ($p < 0.05$) in SPSS 17.0 (SPSS Inc.) and Microsoft Excel (2019 edition). The statistical data were visualized using R (https://www.r-project.org/).

3 | RESULTS

3.1 | Selection of a CSSL based on FA profiling and SSP analysis

CSSL_R119 was selected for detailed analysis because its average FA content was significantly lower than that of SN14, the recurrent parent of the CSSL population (Figure 1a,b). The actual palmitic acid, oleic acid, and linoleic acid contents of R119 were 0.56%–0.81%, 0.78%–1.66%, and 2.46%–2.53% lower than those of SN14, respectively, and the total FA content of R119 was 4.75% lower than that of SN14 in 2019 and 4.41% lower in 2020 (Figure 1a,b). Oil accumulation in seed cells was visualized by Nile Red staining, revealing that R119 seeds had fewer oil bodies per field of view than SN14 seeds (Figure 1c). By contrast, the actual SSP content of R119 was 3.2% higher than that of SN14 in 2019 and 2.9% higher in 2020, as confirmed by Dumas analysis (Figure 1d). Total seed protein was extracted and quantified by gradient SDS-PAGE, clearly revealing that R119 seeds contained more protein, especially 7S- and 11S-related proteins, than SN14 seeds (Figure 1e). On the basis of these results, dry seeds of R119 and SN14 were selected for high-throughput transcriptome sequencing and TMT-based quantitative proteomics analysis. A diagram of the experimental design is shown in Figure 1f.

3.2 | High-throughput RNA-seq analysis of R119 and SN14

High-throughput RNA-seq yielded approximately 310 M total clean reads, with a mapping ratio of 95.06%–95.40% after quality control. The average number of clean reads obtained from SN14 and R119 samples was approximately 50 M and 53 M, respectively (Table S2), sufficient to perform transcriptome profiling, as shown in our previous

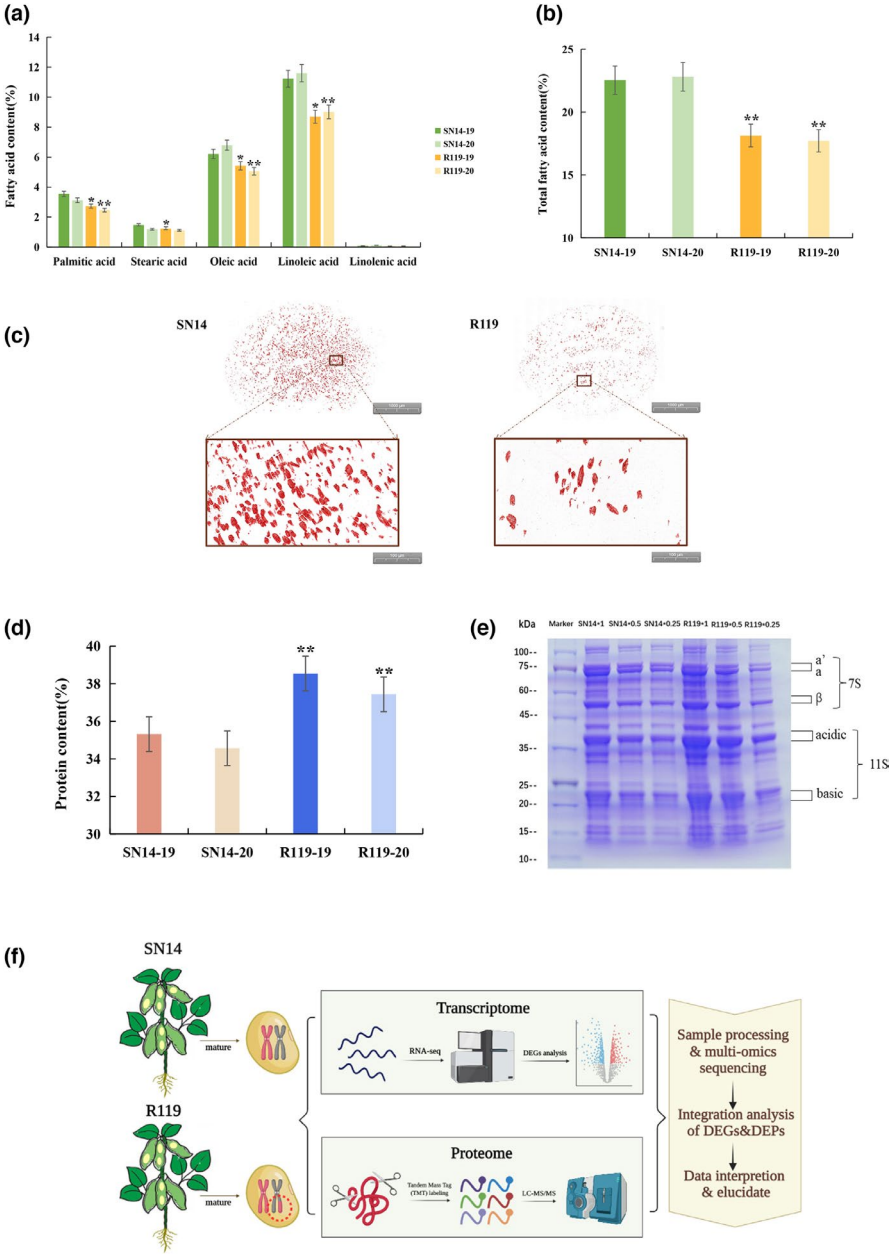

FIGURE 1 Comparison analysis of seed fatty acid and SSP between R119 and SN14 in dry seeds. (a) Comparison of seed fatty acid profile between R119 and SN14 by GC–MS during 2019 and 2020 crop year. (b) Total fatty acid content between R119 and SN14 by GC–MS during 2019 and 2020 crop year. (c) Comparison of the oil bodies (OBs) accumulation abundance stained with Nile Red between R119 and SN14, Bar = 100 and 1000 μm as marker. (d) Comparison of total SSP content between R119 and SN14 by Dumas methods during 2019 and 2020 crop year. (e) Comparison of SDS-PAGE protein profile between R119 and SN14 with concentration in gradient, with standard 1×, 0.5× and 0.25×, respectively. (f) Graphic process model of the experiment design (*$p < 0.05$; **$p < 0.01$)

study (Qi et al., 2018). A global analysis of mapped reads uncovered 33,837 expressed genes in the dry seeds (Table S3), and 665 differentially expressed genes (DEGs) were screened out using the criteria of log2 FC >1 and <−1 under an adjusted $p < 0.05$ (Table S4). Among these DEGs, 404 were upregulated and 261 were downregulated in R119 (Figure 2a). The global expression pattern of DEGs is shown as a heat map in Figure 2b. The identified DEGs were annotated with 27 GO terms at the FDR <0.05 level: 14 biological process terms, 9 cellular component terms, and 4 molecular function terms (Figure 2c). We found that DEGs related to seed storage compounds were involved in alcohol metabolism (GO:0006066), hexose metabolism (GO:0019318), glucose metabolism (GO:0006006), monosaccharide metabolism (GO:0005996), cellular carbohydrate metabolism (GO:0044262), small molecule biosynthesis (GO:0044283), and small molecule catabolism (GO:0044282). Similarly, KEGG annotation of DEGs associated with seed storage compounds revealed the enrichment of pathways related to starch and sucrose metabolism, phenylalanine metabolism, nitrogen metabolism, linoleic acid metabolism, glycolysis/gluconeogenesis, fructose and mannose metabolism, fatty acid degradation, and α-linolenic acid metabolism (Figure 2d).

Comparison of the chromosome substitution segment of R119 with SN14 revealed 86 DEGs in the segment derived from wild soybean (Table S5, Figure 2e). These DEGs could be clearly separated into various groups on the basis of their differential expression patterns (Figure 2f). Because of the limited number of genes, no significant

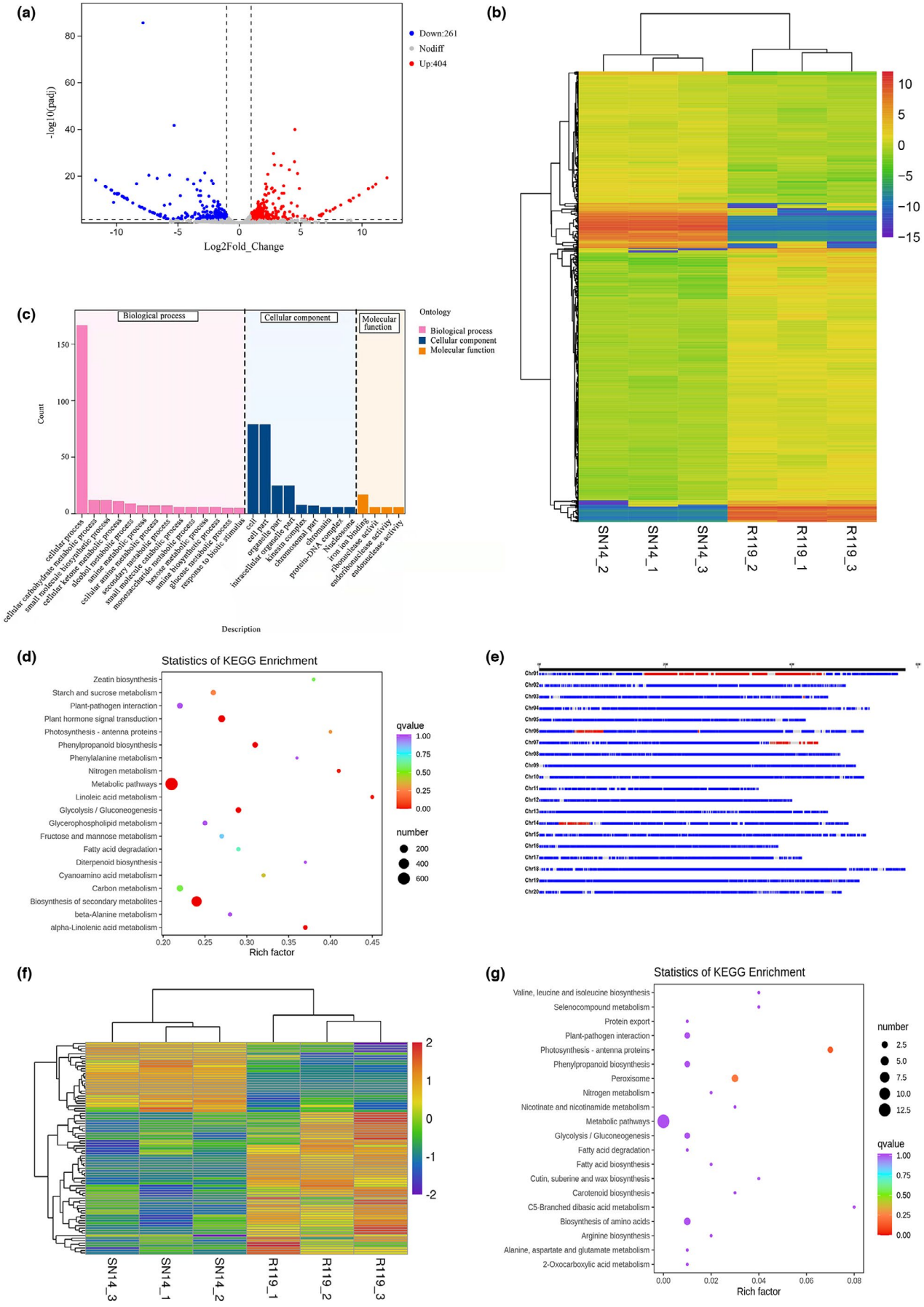
(a)
Down:261
Nodiff
Up:404
-log10(padj)
Log2Fold_Change
(b)
SN14_2
SN14_1
SN14_3
R119_2
R119_1
R119_3
(c)
Biological process
Cellular component
Molecular function
Ontology
Count
Description
(d)
Statistics of KEGG Enrichment
Zeatin biosynthesis
Starch and sucrose metabolism
Plant-pathogen interaction
Plant hormone signal transduction
Photosynthesis - antenna proteins
Phenylpropanoid biosynthesis
Phenylalanine metabolism
Nitrogen metabolism
Metabolic pathways
Linoleic acid metabolism
Glycolysis / Gluconeogenesis
Glycerophospholipid metabolism
Fructose and mannose metabolism
Fatty acid degradation
Diterpenoid biosynthesis
Cyanoamino acid metabolism
Carbon metabolism
Biosynthesis of secondary metabolites
beta-Alanine metabolism
alpha-Linolenic acid metabolism
qvalue
number
Rich factor
(e)
(f)
SN14_3
SN14_1
SN14_2
R119_1
R119_2
R119_3
(g)
Statistics of KEGG Enrichment
Valine, leucine and isoleucine biosynthesis
Selenocompound metabolism
Protein export
Plant-pathogen interaction
Photosynthesis - antenna proteins
Phenylpropanoid biosynthesis
Peroxisome
Nitrogen metabolism
Nicotinate and nicotinamide metabolism
Metabolic pathways
Glycolysis / Gluconeogenesis
Fatty acid degradation
Fatty acid biosynthesis
Cutin, suberine and wax biosynthesis
Carotenoid biosynthesis
C5-Branched dibasic acid metabolism
Biosynthesis of amino acids
Arginine biosynthesis
Alanine, aspartate and glutamate metabolism
2-Oxocarboxylic acid metabolism
number
qvalue
Rich factor

FIGURE 2 Comparative transcriptomic analysis of R119 and SN14 in dry seeds. (a) Volcano map of DEGs with upregulated and downregulated between R119 and SN14. (b) Heat map of the DEGs based on clustering between R119 and SN14. (c) GO annotation of total 665 DEGs. (d) KEGG enrichment analysis of total 665 DEGs. (e) Genome overview of substituted region by wild soybean in R119, blue represent the recurrent genome SN14, red represent the donor genome ZYD00006, gray represent the gap in genome. (f) Heat map of the 86 DEGs that located in the substituted region of R119; (g) KEGG enrichment analysis of total 86 DEGs located in the substituted region of R119

GO enrichment of the 86 DEGs was observed. According to the KEGG enrichment analysis, however, some of these genes may be involved in the following processes related to seed storage compounds: valine, leucine, and isoleucine biosynthesis; protein export; nitrogen metabolism; glycolysis/gluconeogenesis; fatty acid degradation; biosynthesis of amino acids; arginine biosynthesis; and alanine, aspartate, and glutamate metabolism (Figure 2g).

3.3 | TMT-based high-throughput proteomics analysis of DAPS between R119 and SN14

3.3.1 | Global analysis of TMT-based high-throughput proteomics data

Total protein was extracted from dry seeds of SN14 and R119 and used for TMT-based quantitative proteomics analysis. According to our quality control analysis, the lengths of identified peptides mainly ranged from 8 to 15 (Figure S1A), molecular weights primarily varied from 0 to 100 kDa (Figure S1B), and 40.29% of unique peptides had only one count (Figure S1C). These results confirm the quality and stability of the TMT-based high-throughput proteomics profile. Pearson's correlation coefficients between different replicates of the same seed type were >0.998, indicating that replicates of the same genotype were highly correlated (Figure S2A,B).

A total of 80,732 peptide spectrum matches, 18,055 unique peptides (Table S6), and 3415 modeled proteins were identified (Table S7). The number of amino acids ranged from 15 to 5031, with an average of 417.6, and molecular weights varied between 1.7 and 559.5 kDa, with an average of 46.23 kDa. All 3415 quantified proteins were annotated, mainly with the GO terms cellular process, metabolic process, localization, cellular anatomical entity, protein-containing complex, and binding and catalytic activity (Figure S3A). The top 10 enriched KEGG pathways were biosynthesis of amino acids, biosynthesis of secondary metabolites, carbon fixation in photosynthetic organisms, carbon metabolism, glycolysis/gluconeogenesis, metabolic pathways, protein processing in the endoplasmic reticulum (ER), pyruvate metabolism, and ribosome and RNA transport (Figure S3B). A subcellular localization analysis predicted that these proteins were mainly located in the cytoplasm, membranes, and ribosomes (Figure S3C).

Using the criteria of FC >1.2 or <0.83 and Student's *t*-test ($p < 0.05$), we identified 83 DAPs between SN14 and R119, including 47 and 36 DAPs with increased and decreased abundance in R119, respectively (Table S8, Figure 3a). Overall, the DAPs could be divided into distinct groups on the basis of their accumulation patterns (Figure 3b), including those annotated with the putative seed storage compound–related GO terms related to ribosome, peptide metabolic process, transcription by RNA polymerase II, β-galactosidase complex, and microtubule-associated complex (Figure 3c). In addition, KEGG enrichment analysis indicated that α-linolenic acid metabolism, galactose metabolism, and linoleic acid metabolism pathways were involved in FA metabolism in soybean (Figure 3d). The integration of DAPs and their pathways is shown in Figure 3e. A protein–protein interaction network with 47 nodes and 148 edges was constructed based on Pearson correlation analysis ($p < 0.05$). The nodes consisted of 35 DAPs and 12 pathways, namely, ribosome, linoleic acid metabolism, metabolic pathways, biosynthesis of secondary metabolites, photosynthesis, purine metabolism, α-linolenic acid metabolism, galactose metabolism, pentose phosphate pathway, C5-branched dibasic acid metabolism, glutathione metabolism, and other glycan degradation (Figure 3f).

3.3.2 | Identification of soybean dry seed–specific DAPs

DAPs whose abundance was lower in R119 were mainly associated with the GO terms ribosome related, pectinesterase activity, oxygen binding, structural molecule activity, and structural constituent of ribosome (Figure S4A). The top five decreased DAPs were A1KR24, K7MF41, A0A0R4J3X6, K7MSB5, and I1N789, and their comparative abundances between R119 and SN14 are shown in Figure S4B. Interestingly, 12 ribosomal proteins of different molecular weights were present at significantly lower abundance in R119 than in SN14. We also identified dehydrin, seipin, and ubiquitin proteins among the DAPs whose abundance was lower in R119 (Table S9).

DAPs with increased abundance in R119 were mainly enriched in the GO terms protein phosphorylation,

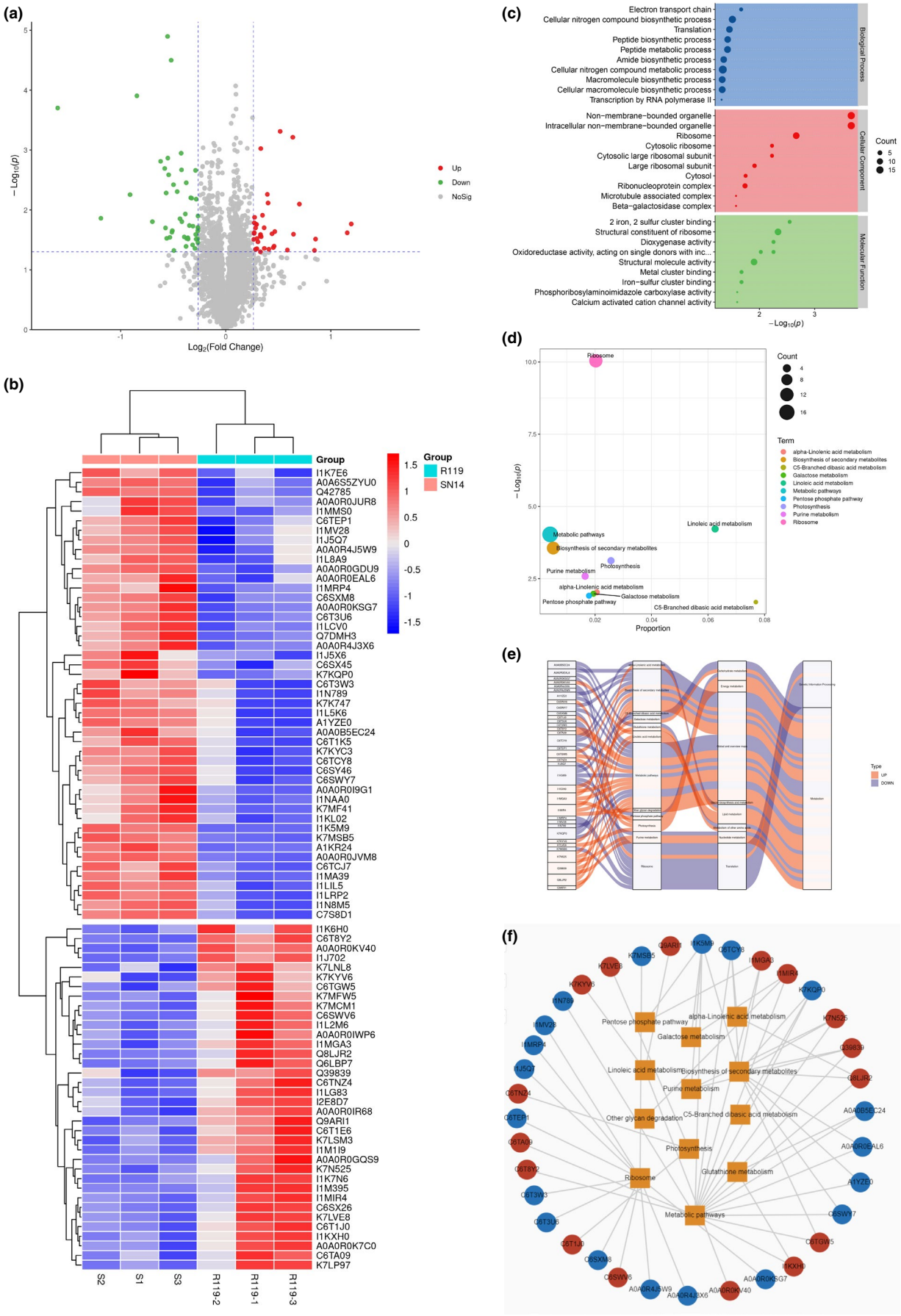
(a)
Up
Down
NoSig
$-\mathrm{Log}_{10}(p)$
Log_2(Fold Change)
(b)
Group
R119
SN14
S2
S1
S3
R119-2
R119-1
R119-3
(c)
Electron transport chain
Cellular nitrogen compound biosynthetic process
Translation
Peptide biosynthetic process
Peptide metabolic process
Amide biosynthetic process
Cellular nitrogen compound metabolic process
Macromolecule biosynthetic process
Cellular macromolecule biosynthetic process
Transcription by RNA polymerase II
Non−membrane−bounded organelle
Intracellular non−membrane−bounded organelle
Ribosome
Cytosolic ribosome
Cytosolic large ribosomal subunit
Large ribosomal subunit
Cytosol
Ribonucleoprotein complex
Microtubule associated complex
Beta−galactosidase complex
2 iron, 2 sulfur cluster binding
Structural constituent of ribosome
Dioxygenase activity
Oxidoreductase activity, acting on single donors with inc...
Structural molecule activity
Metal cluster binding
Iron−sulfur cluster binding
Phosphoribosylaminoimidazole carboxylase activity
Calcium activated cation channel activity
Biological Process
Cellular Component
Molecular Function
Count
$-\mathrm{Log}_{10}(p)$
(d)
Ribosome
Linoleic acid metabolism
Metabolic pathways
Biosynthesis of secondary metabolites
Photosynthesis
Purine metabolism
alpha-Linolenic acid metabolism
Galactose metabolism
Pentose phosphate pathway
C5-Branched dibasic acid metabolism
Proportion
Term
alpha-Linolenic acid metabolism
Biosynthesis of secondary metabolites
C5-Branched dibasic acid metabolism
Galactose metabolism
Linoleic acid metabolism
Metabolic pathways
Pentose phosphate pathway
Photosynthesis
Purine metabolism
Ribosome
(e)
Type
UP
DOWN
(f)
Pentose phosphate pathway
alpha-Linolenic acid metabolism
Galactose metabolism
Linoleic acid metabolism
Biosynthesis of secondary metabolites
Purine metabolism
Other glycan degradation
C5-Branched dibasic acid metabolism
Photosynthesis
Glutathione metabolism
Ribosome
Metabolic pathways

FIGURE 3 Comparative TMT-based quantitative proteomics analysis of R119 and SN14 in dry seeds. (a) Volcano map of DAPs with increased and decreased abundances between R119 and SN14. (b) Heat map of the DAPs based on clustering between R119 and SN14. (c) GO annotation of total 83 DAPs. (d) KEGG enrichment analysis of total 83 DAPs. (e) Overview of DAP and their pathway integration. (f) Regulation network of DAPs and metabolism pathway

generation of precursor metabolites and energy, transcription by RNA polymerase II, electron transport chain, β-galactosidase complex, nutrient reservoir activity, electron transfer activity, iron–sulfur cluster binding, oxidoreductase activity, and dioxygenase activity (Figure S4C).

The two proteins in R119 that showed the greatest increase in abundance relative to SN14 were both ferredoxin proteins and exhibited 2.28-fold and 2.23-fold higher accumulation. The genes encoding these two proteins were both located in the substituted region. A BOWMAN_BIRK domain-containing protein of the Bowman–Birk inhibitor family, which includes well-known water-soluble proteins in soybean, was 1.81-fold more abundant in R119. The DAP with the next highest increase in abundance was a perakine reductase (PR), which catalyzes NADPH-dependent reduction and is involved in the biosynthesis of monoterpenoid indole alkaloids. Lipoxygenase (LOX) showed the fifth highest increase in abundance. We also found a 1.56-fold-increase in A5A4B3 glycinin (fragment), a type of SSP, and a 1.56-fold increase in a TOG domain–containing protein that is involved in protein processing pathways and whose gene is located in the substituted region. The abundance of an albumin I domain-containing protein involved in nutrient reservoir activity, which functions in the storage of nutritional substrates, was 1.50-fold higher in R119. The top five increased DAPs and their differential abundances are shown in Figure S4D. Other DAPs with increased abundance in R119 are detailed in Table S10.

Interestingly, we detected several proteins that have not previously been identified in soybean, such as structural maintenance of chromosome protein 1 (sister chromatid cohesion complex Cohesin, subunit SMC1), PR, nutrient reservoir activity protein, embryo development ending in seed dormancy, and metal iron binding related. Significantly, 10 DAPs, including the two with the greatest increases in abundance, were encoded by genes located in the substituted region of R119.

3.4 | Integrated analysis of transcriptomics and proteomics data reveals a new regulatory network for FA and storage protein accumulation in soybean seeds

Through integrated analysis of DEGs in the substituted segment of R119, annotation of DAPs and DEGs, and pathway analysis of DEGs and DAPs involved in biochemical and cellular processes, we inferred the existence of a previously unknown pathway that extends from light energy absorption to seed storage. Although little translational activity occurs in dry soybean seeds, we nonetheless identified 665 DEGs and 83 DAPs between R119 and SN14 by high-throughput RNA-seq and TMT-based proteomics. In brief, 5 DAPs and 27 DEGs from the wild substituted genomic fragment of R119 cooperated with 18 DAPs from the cultivated soybean genomic background, and together they are presumably responsible for the differences in seed storage profile between the two soybean lines (Table 1).

The two proteins with the greatest increases in abundance were encoded by *Glyma.12G169500* and *Glyma.12G169400*, both of which are located in the substituted region. Both proteins were annotated as ferredoxin proteins that participate in the regulation of the electron transport chain and the assimilation of nitrogen via photosystem I in the chloroplasts (Hirasawa et al., 1986). Seven DEGs in the substituted region of R119 also participate in this process. Six of them were upregulated in R119 compared with the recurrent parent, suggesting that raw materials were increased for further synthesis of storage compounds in R119 (Figure 4a, Table 1).

On the basis of their 'protein export from ribosome' and 'into ER pathway' annotations, we also identified a protein with lower abundance in R119 and a protein encoded by the upregulated candidate gene *Glyma.17G073300* (located in the substituted segment of R119), both of which were annotated as signal recognition particles (SRPs). Previous research has demonstrated that overexpression of SRPs selectively limits membrane protein expression, thus indicating that the SRPs in R119 may promote this process (Yosef et al., 2010). We also identified a small nuclear ribonucleoprotein G encoded by *Glyma.11G086200* located in the substituted region of R119. This small G protein cooperates with SRP to enable proteins to exit the ribosomes and enter the ER (Bange et al., 2011) (Figure 4b, Table 1).

Among the proteins associated with protein processing in the ER, we also identified a TOG domain–containing protein encoded by *Glyma.01G103200* in the substituted segment of R119 that functions as a heat shock protein 70 (HSP70). The abundance of this protein was 1.56-fold higher in R119 than in the recurrent parent. HSP70s stabilize unfolded proteins, assist with their folding, and prevent incorrect folding by accelerating isomerization in the ER (Mehrotra et al., 2009). Five DEGs located in the

TABLE 1 The DAPs and DEGs involved the seed storage related pathway

Pathway	Protein/gene name	Log2FC	Functional annotation
Photosynthesis, energy metabolism pathway			
DAP	C6T1J0[a]	1.19430203	Ferredoxins are iron-sulfur proteins that transfer electrons in a wide variety of metabolic reactions
	K7LVE8[a]	1.155841816	Ferredoxins are iron-sulfur proteins that transfer electrons in a wide variety of metabolic reactions
DEG	*Glyma.01G115900*	1.242751144	Photosynthesis, light harvesting, CHLOROPHYLL A/B BINDING PROTEIN
	Glyma.05G212500	2.22194034	Oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor
	Glyma.06G195100	1.747047316	Thioredoxin
	Glyma.07G104500	−1.01284543	Nitrogen compound metabolic process, Glutamine synthetase, beta-Grasp domain
	Glyma.09G071400	2.760186524	Photosynthesis, light harvesting
	Glyma.13G092500	1.146710488	Trehalose-6-phosphate synthase
	Glyma.20G196900	1.841000476	Fe superoxide dismutase 2
Protein processing in ER pathway			
DAP	I1J702[a]	0.638272599	Adenyl-nucleotide exchange factor activity
	C6TCJ7	−0.566660537	SRP-dependent cotranslational protein targeting to membrane.signal recognition particle
	C6T1E6[a]	0.271919859	Small nuclear ribonucleoprotein G
DEG	*Glyma.01g118400*	−4.768738746	Di-glucose binding within endoplasmic reticulu, protein serine/threonine kinase activity
	Glyma.02G294600	−1.190108418	GRPE protein homolog, mitochondrial
	Glyma.17G073300	1.033677062	Signal recognition particle receptor alpha subunit family protein
	Glyma.02G294900	−1.25072004	Trigger factor type chaperone family protein
	Glyma.11G086200	0.03363162	Probable small nuclear ribonucleoprotein G
Portein sorting process in TGN pathway			
DAP	I1K6H0[a]	0.284483253	Protein polyubiquitination
DEG	*Glyma.01G115100*	10.90729032	Intracellular protein transport, vesicle-mediated transport, clathrin coat of trans-Golgi network vesicle
	Glyma.08G181400	1.946937577	ENTH/ANTH/VHS superfamily protein
	Glyma.11G032400	−1.082630437	DNAJ heat shock N-terminal domain-containing protein
Storage protein accumulation			
DAP	C6SX26	0.854976492	Serine-type endopeptidase inhibitor activity
	I2E8D7	0.644668437	This protein found in the seeds of many leguminous and non-leguminous plants is the source of sulfur-containing amino acids in seed meals. Seed storage protein
	Q6LBP7	0.270277	Glycinin B(1b) subunit (15 AA) (Fragment)

TABLE 1 (Continued)

Pathway	Protein/gene name	Log2FC	Functional annotation
DEG	*Glyma.13G001200*	2.041908324	Cystathionine beta-lyases/cystathionine gamma-synthases
	Glyma.05G204800	1.832630991	Osmotin 34
	Glyma.06G220800	3.517208776	blue-copper-binding protein
	Glyma.08G235300	2.313655318	Kunitz family trypsin and protease inhibitor protein
	Glyma.08G235400	1.88350536	Kunitz trypsin inhibitor 1
	Glyma.09G065000	1.480083243	Calmodulin-binding family protein
	Glyma.13G068700	6.716162155	Zinc finger C-x8-C-x5-C-x3-H type family protein
	Glyma.17G048100	1.053886686	Inter-alpha-trypsin inhibitor heavy chain-related
Ribosomal protein related pathway			
DAP	A0A0R0KSG7	−0.542892231	Structural constituent of ribosome
	A0A0R4J3X6	−0.910026231	Structural constituent of ribosome
	A0A0R4J5W9	−0.508052656	Structural constituent of ribosome
	C6SXM8	−0.496779093	Structural constituent of ribosome
	C6T3U6	−0.580412998	Structural constituent of ribosome
	C6T3W3	−0.292708028	Structural constituent of ribosome
	C6TEP1	−0.512336866	Structural constituent of ribosome
	I1J5Q7	−0.315214927	Structural constituent of ribosome
	I1MRP4	−0.572751783	Structural constituent of ribosome
	I1MV28	−0.493186244	Structural constituent of ribosome
	K7MSB5	−0.845222283	Structural constituent of ribosome
DEG	*Glyma.12G115100*	−3.787439751	Ribonuclease T2 activity
	Glyma.12G119800	−1.558780467	Ribosome biogenesis, Ribosomal L28e protein family
Lipoxygenase			
DAP	Q9ARI1	0.700918768	This protein is involved in the pathway oxylipin biosynthesis, which is part of Lipid metabolism
	A0A0R0KV40	0.518170886	This protein is involved in the pathway oxylipin biosynthesis, which is part of Lipid metabolism
	K7KYV6	0.460303501	This protein is involved in the pathway oxylipin biosynthesis, which is part of Lipid metabolism
DEG	*Glyma.12G183300*	−5.790573988	Jojoba acyl CoA reductase-related male sterility protein
	Glyma.17G075300	1.249141179	Aldehyde dehydrogenase 11A3

[a]DAP located in the substituted region of R119. All listed DEGs were located in the substituted region of R119.

substituted region of R119 participate in protein folding and are co-expressed to stabilize this process (Figure 4b, Table 1).

In general, storage proteins exiting the ER are sorted in the trans-Golgi network (TGN) and then enter multivesicular bodies (MVBs) before final storage in protein storage vacuoles. Ubiquitin proteins have an important role in MVB sorting, as the entry of several different proteins into MVBs depends on their prior ubiquitylation (Morvan et al., 2004). We also found a UBC core domain–containing protein encoded by *Glyma.05G209400* in the substituted region that functions in protein poly-ubiquitination. The

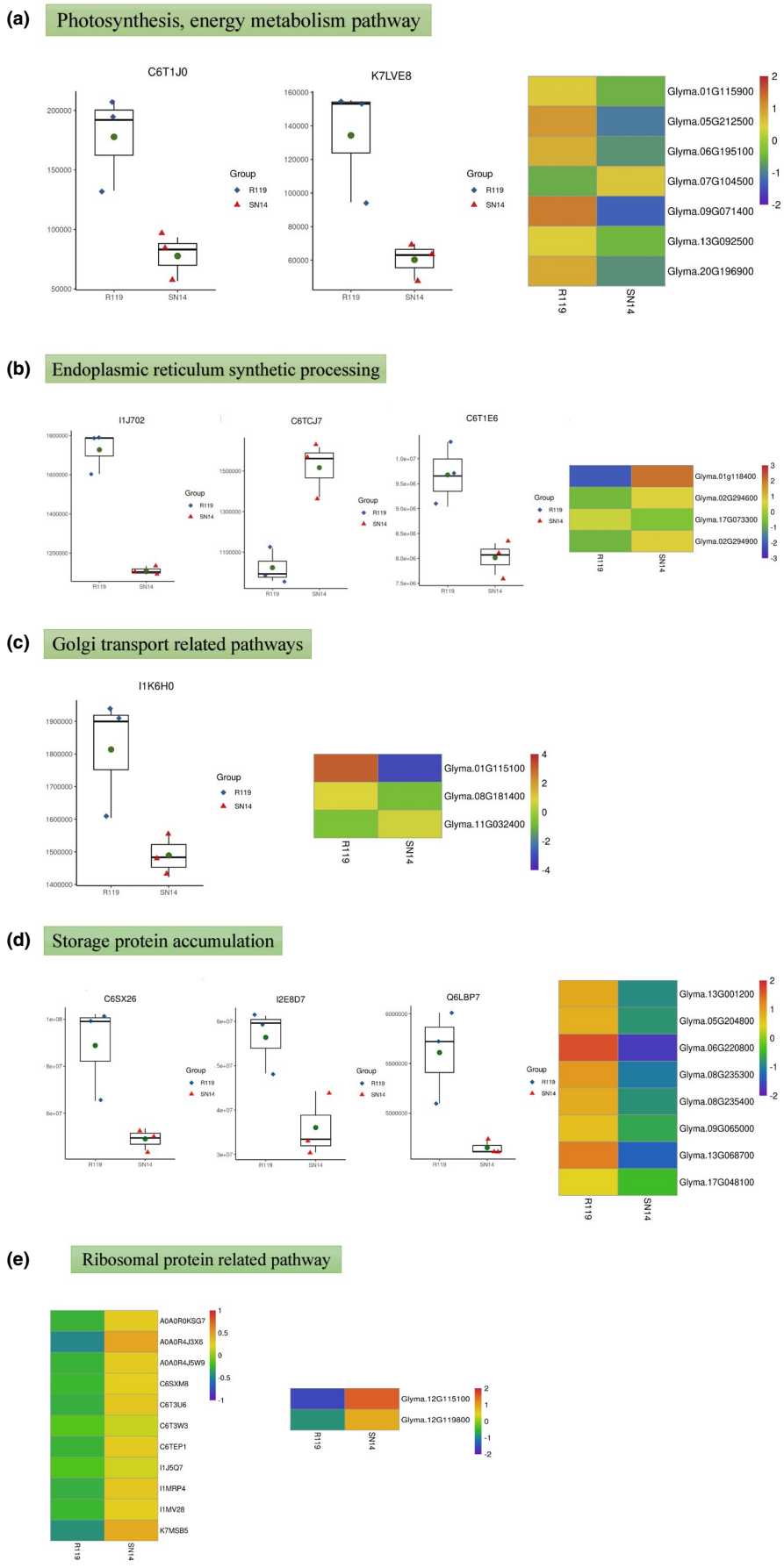

FIGURE 4 Integration analysis of DAPs and DEGs located in substituted region involved the seed storage related pathway. (a) Two DAPs and seven DEGs involved the photosynthesis, energy metabolism pathway. (b) Three DAPs and four DEGs involved the protein processing in ER. (c) A DAP and three DEGs involved the protein sorting process in TGN. (d) Three DAPs and eight DEGs involved the SSP accumulation. (e) Eleven DAPs and two DEGs involved in the ribosomal protein related pathway

increased abundance of this protein in R119 may help with the MVB sorting process. In addition, we detected three genes in the substituted region that are involved in protein sorting in the TGN (Figure 4c, Table 1).

With respect to the above-mentioned regulation of the protein storage process, some storage proteins, such as BOWMAN_BIRK domain-containing protein, A5A4B3 glycinin, and glycinin B(1b) subunit, showed increased abundance in R119 seeds. Likewise, eight SSP-related DEGs had significantly higher expression in R119 than in the cultivated soybean SN14 and encoded proteins such as cystathionine β-lyase/cystathionine γ-synthase, blue-copper-binding protein, and Kunitz family trypsin and protease inhibitor protein (Figure 4d, Table 1).

We detected three LOX proteins, which play essential roles in FA oxidation, with 1.37- to 1.62-fold increases in abundance in R119. LOX proteins have previously been found to play an important role in the degradation of storage lipids and FAs (Feussner et al., 2001). The three LOX proteins with higher abundance in R119 may therefore oxidize more FAs and reduce the FA content of R119 soybean seeds (Table 1).

In a previous investigation based on ribosomal profiling data, genes involved in lipid metabolism pathways were translationally downregulated in a ribosomal protein mutant, and lipid accumulation was reduced in the mutant plants (Li, Sun, et al., 2015; Li, Wang, et al., 2015). Similarly, in the present study, we identified 11 ribosomal proteins with significantly decreased abundance in R119, as well as a marked reduction in FA content of R119 seeds relative to SN14 seeds (Figure 4e, Table 1).

Two additional proteins with increased abundance in R119 were annotated as an RRM domain-containing protein (K7LP97) and a DNA-directed RNA polymerase subunit (K7LNL8), which are associated with mRNA binding and RNA polymerase catalyzation. These proteins typically function in the regulation of translation (Table S10).

4 | DISCUSSION

Soybean provides oil to the world market that is low in saturated fat and free from cholesterol. It is also a rich source of excellent quality protein with essential amino acids that is widely used in animal and human foods, such as infant formulas, flours, protein concentrates and isolates, and textured fibers (Cabanos et al., 2021; Chatterjee et al., 2018; Friedman & Brandon, 2001). In addition, multiple studies have reported that soy-based foods can reduce the risk of chronic diseases such as obesity, heart disease, high blood pressure, and impaired bone health (Cai et al., 2021; De Mejia & Ben, 2006; Erdmann et al., 2008; Kwon et al., 2010). A better understanding of the mechanisms that regulate soybean seed fat and protein profiles will help to improve the nutritional, health-promoting, and functional value of soy-based foods.

CSSL libraries are useful materials for mining and mapping of trait–genotype associations and also serve as valuable genetic resources (Gu et al., 2015; Li, Sun, et al., 2015; Li, Wang, et al., 2015; Takai et al., 2014). Each line in a CSSL population can be thought of as an "artificial mutant" that contains a different segment donated by the wild soybean parent and expresses specific traits differently than the recurrent parent. Although CSSL populations are essential genetic material for QTL detection, validation, and positional cloning, very few have been developed in soybean to date (Xin et al., 2016). In the present study, we screened out the introgressed line R119 with stable, high SSP and low FA contents, and microscopy analyses confirmed that R119 seeds contained significantly fewer oil bodies than SN14 seeds. By contrast, quantitative SDS-PAGE indicated that R119 seeds clearly contained more SSPs than SN14 seeds. The R119 CSSL line could therefore be used in an integrated omics analysis to uncover the essential regulatory mechanism(s) that control the accumulation of seed storage compounds. Target chromosome-segment substitution paves the way for breeding by design and for accelerating the breeding process (Zhang, 2021). This CSSL line, with its specific genomic segment from wild soybean, is an essential genetic resource that can be used for the breeding of soybean with specifically designed quality traits in the future.

TMT-based high-throughput proteomics enables highly efficient and accurate detection of proteins. Published studies have reported that proteomic analysis can increase the speed and efficiency of new protein detection and identification (Battisti et al., 2021; Min et al., 2020). Quantitative proteomics of mutants and/or transgenic lines can be used to build regulatory networks related to specific traits with great precision (Komatsu et al., 2021; Xu et al., 2021). Here, we performed TMT-based quantitative proteomics on a CSSL that differed markedly in SSP abundance from its recurrent parent, thereby obtaining insights into specific changes in its seed proteins. The total number of DAPs was relatively small because of the single substituted segment and the clear genomic background (Figure 2e), allowing the CSSL to serve as a type of precise breeding line and a genetic resource for research. We next screened out seed-associated proteins, including those related to amino acid and protein metabolism, as well as protein synthesis, folding, targeting, and degradation. The main contributors to differences in SSP accumulation in R119 seeds were DAPs enriched in categories such as photosynthesis and energy processes, protein export, protein processing in the ER, protein sorting in the TGN, and protein storage. At the same time, the

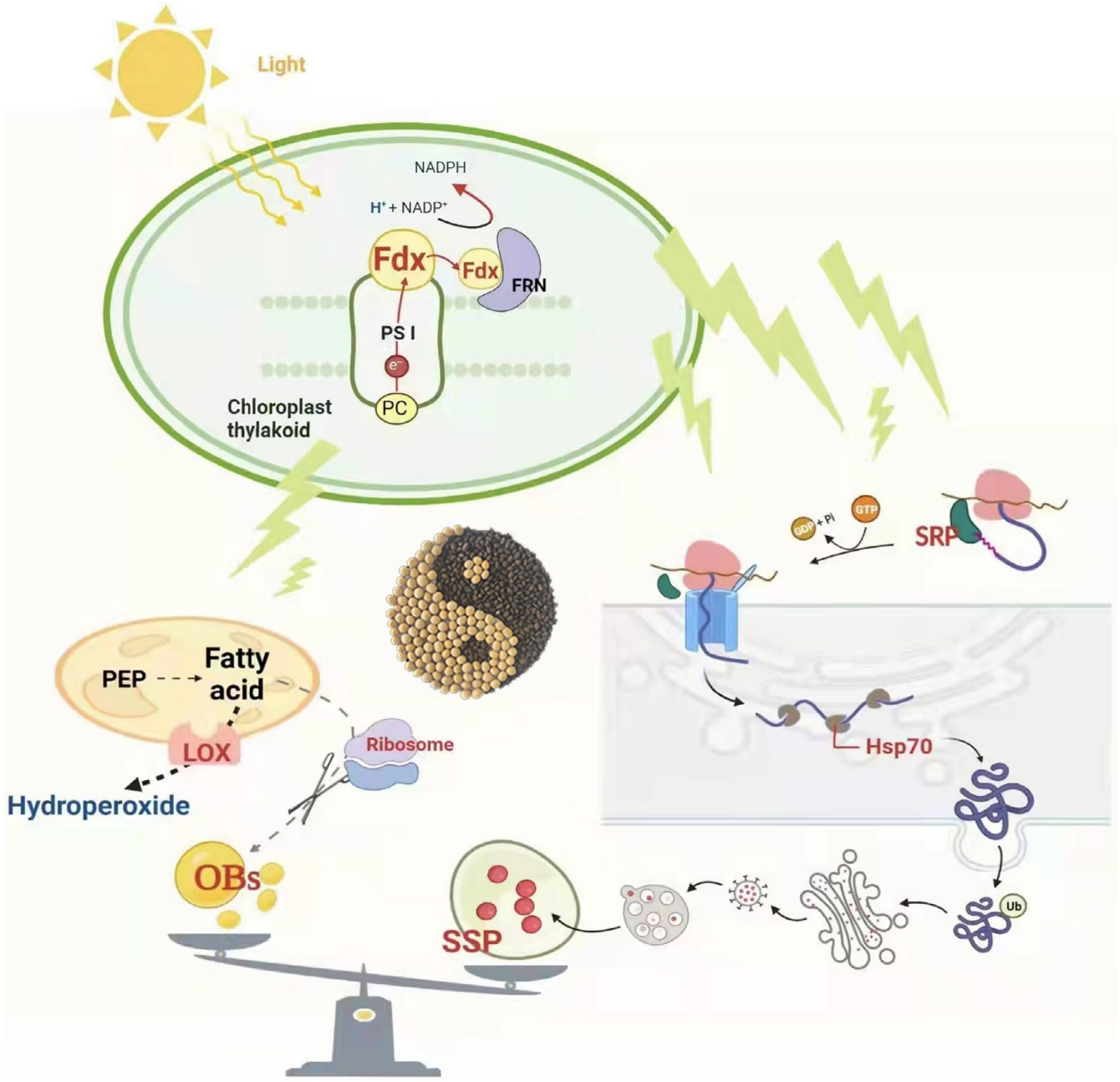

FIGURE 5 Regulation network model of seed fatty acid and SSP in soybean

accumulation of three LOX proteins increased and that of 11 ribosomal proteins decreased significantly, which may have limited FA biosynthesis. These results were consistent with the phenotype of R119, which accumulated more storage protein and less oil.

Previous integrated transcriptomic and proteomic studies provide guidance for revealing the regulatory networks associated with specific traits (Gu et al., 2017; Huang et al., 2019; Lin et al., 2019; Shi et al., 2021). Together, their findings demonstrate that the integration of multiple omics approaches is a rapid and efficient strategy for identifying regulatory networks in the era of high-throughput genomics. Although very little translational activity takes place in dry soybean seeds, the accumulation of seed storage compounds is controlled by complex mechanisms, including post-translation or expression networks. In this study, we performed transcriptomic and TMT-based proteomic analysis of dry soybean seeds from a CSSL and its recurrent parent that differed in their seed storage profiles. We identified DEGs and DAPs associated with photosynthetic production, SSP accumulation, and reductions in FA content that were predicted to affect photosystem efficiency, protein processing, protein sorting, and ribosomal-related pathways. DAPs with increased abundance in R119 were encoded by genes located in the substituted region of R119, but these genes did not exhibit

significant differential expression. By contrast, DEGs located in the substituted region of R119 were involved in the same key seed storage compound pathways as the DAPs above. In addition, some DAPs may be indirectly affected by genes (or their encoded proteins) in the substituted region and thus also influence the seed storage profile of R119. This molecular evidence was consistent with data from soybean breeding and production over the past years, in which seed oil content was significantly negatively correlated with protein content (Burton, 1987; Johnson & Bernard, 1962; Wilcox, 1998). This negative correlation has limited the breeding of soybean varieties with both high protein and oil content at the same time (Cober & Voldeng, 2000). A better understanding of their regulatory networks may contribute to balancing seed oil and protein levels in future breeding practice.

Finally, we proposed a potential model to explain why R119 stores more proteins than fatty acids. First, ferredoxin improves the photosynthetic efficiency of chloroplasts to facilitate the production of more raw materials (Hirasawa et al., 1986). Next, DAPs and DEGs involved in protein processing in the ER and sorting in the TGN are upregulated, enhancing the accumulation of SSPs. These include a small G protein that functions together with SRP to enable proteins to enter the ER (Bange et al., 2011), an HSP that assists with the folding of unfolded proteins by accelerating isomerization (Mehrotra et al., 2009), and a ubiquitin protein that participates in MVB sorting (Morvan et al., 2004). At the same time, the abundance of LOX proteins is increased and that of ribosomal-related proteins is decreased, which limits FA biosynthesis (Feussner et al., 2001; Li, Sun, et al., 2015; Li, Wang, et al., 2015). Finally, more raw materials are diverted to the formation of SSPs rather than FAs (Figure 5). In particular, we discovered that FA biosynthesis may be impaired indirectly by the regulation of genes and proteins related to SSP accumulation.

Although the further validation of these candidate DEGs and DAPs is needed, our integrated genomic, transcriptomic, and proteomic analyses have identified a previously unknown regulatory network that controls the accumulation of soybean seed storage components from source to sink. Our study findings can contribute to research on soybean seed quality and soybean breeding.

ACKNOWLEDGMENTS

This study was financially supported by the National Key R&D Program of China (2021YFF1001202), the National Natural Science Foundation of China (31971968, U20A2027), the Natural Science Foundation of Heilongjiang-Outstanding Youth Foundation (YQ2021C011), and the Young Scholar Bone Project (to Qi Zhaoming) of Northeast Agriculture University. We thank Shanghai Bioprofile Technology Company, Ltd. for assistance with the proteome experiment and A&L Scientific Editing Group for English language editing.

CONFLICT OF INTEREST

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

AUTHOR CONTRIBUTIONS

Xue Han: conceptualization, methodology, and original draft preparation; Jiapeng Li: data curation, visualization, and investigation; Yabin Zhao: data curation, visualization, and investigation; Zhanguo Zhang: data curation; Hongwei Jiang: data curation; Jinxing Wang: data curation; visualization; Xuezhen Feng: data curation; visualization; Yu Zhang: visualization; Ziyue Du: visualization; Xiaoxia Wu: supervision; Qingshan Chen: conceptualization, methodology, supervision, reviewing, and editing; Zhaoming Qi: conceptualization, methodology, supervision, writing, reviewing, and editing.

ORCID

Zhaoming Qi https://orcid.org/0000-0002-0657-9127

REFERENCES

AoYAMA, Y., Wada, M., & Morifuji, M. (2001). Orotic acid added to casein, but not to egg protein, soy protein, or wheat gluten diets increases 1, 2-diacylglycerol levels and lowers superoxide dismutase activities in rat liver. *Bioscience, Biotechnology, and Biochemistry*, *65*(10), 2166–2173. https://doi.org/10.1271/bbb.65.2166

Balakrishnan, D., Surapaneni, M., Mesapogu, S., & Neelamraju, S. (2019). Development and use of chromosome segment substitution lines as a genetic resource for crop improvement. *Theoretical and Applied Genetics*, *132*(1), 1–25. https://doi.org/10.1007/s00122-018-3219-y

Bange, G., Kümmerer, N., Grudnik, P., Lindner, R., Petzold, G., Kressler, D., Hurt, E. D., Wild, K., & Sinning, I. (2011). Structural basis for the molecular evolution of SRP-GTPase activation by protein. *Nature Structural & Molecular Biology*, *18*(12), 1376–1380. https://doi.org/10.1038/nsmb.2141

Battisti, I., Ebinezer, L. B., Lomolino, G., Masi, A., & Arrigoni, G. (2021). Protein profile of commercial soybean milks analyzed by label-free quantitative proteomics. *Food Chemistry*, *352*, 129299. https://doi.org/10.1016/j.foodchem.2021.129299

Blasbalg, T. L., Hibbeln, J. R., Ramsden, C. E., Majchrzak, S. F., & Rawlings, R. R. (2011). Changes in consumption of omega-3 and omega-6 fatty acids in the United States during the 20th century. *The American Journal of Clinical Nutrition*, *93*(5), 950–962. https://doi.org/10.3945/ajcn.110.006643

Burton, J. W. (1987). *Quantitative genetics: Results relevant to soybean breeding*. Agronomy (USA).

Cabanos, C., Matsuoka, Y., & Maruyama, N. (2021). Soybean proteins/peptides: A review on their importance, biosynthesis, vacuolar sorting, and accumulation in seeds. *Peptides*, *143*, 170598. https://doi.org/10.1016/j.peptides.2021.170598

Cai, J. S., Feng, J. Y., Ni, Z. J., Ma, R. H., Thakur, K., Wang, S., Hu, F., Zhang, J. G., & Wei, Z. J. (2021). An update on the nutritional, functional, sensory characteristics of soy products, and applications of new processing strategies. *Trends in Food Science & Technology*, *112*, 676–689. https://doi.org/10.1016/j.tifs.2021.04.039

Chatterjee, C., Gleddie, S., & Xiao, C. W. (2018). Soybean bioactive peptides and their functional properties. *Nutrients*, *10*(9), 1211. https://doi.org/10.3390/nu10091211

Clemente, T., & Cahoon, E. (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiology*, *151*, 1030–1040. https://doi.org/10.1104/pp.109.146282

Cober, E. R., & Voldeng, H. D. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Science*, *40*(1), 39–42. https://doi.org/10.2135/cropsci2000.40139x

De Mejia, E., & Ben, O. (2006). Soybean bioactive peptides: A new horizon in preventing chronic diseases. *Sexuality, Reproduction and Menopause*, *4*(2), 91–95. https://doi.org/10.1016/j.sram.2006.08.012

Derbyshire, E., Wright, D. J., & Boulter, D. (1976). Legumin and vicilin, storage proteins of legume seeds. *Phytochemistry*, *15*(1), 3–24. https://doi.org/10.1016/S0031-9422(00)89046-9

Erdmann, K., Cheung, B. W., & Schröder, H. (2008). The possible roles of food-derived bioactive peptides in reducing the risk of cardiovascular disease. *The Journal of Nutritional Biochemistry*, *19*(10), 643–654. https://doi.org/10.1016/j.jnutbio.2007.11.010

Eshed, Y., & Zamir, D. (1994). A genomic library of *Lycopersicon pennellii* in *L. esculentum*: A tool for fine mapping of genes. *Euphytica*, *79*(3), 175–179. https://doi.org/10.1007/BF00022516

Feussner, I., Kühn, H., & Wasternack, C. (2001). Lipoxygenase-dependent degradation of storage lipids. *Trends in Plant Science*, *6*(6), 268–273. https://doi.org/10.1016/S1360-1385(01)01950-1

Friedman, M., & Brandon, D. L. (2001). Nutritional and health benefits of soy proteins. *Journal of Agricultural and Food Chemistry*, *49*(3), 1069–1086. https://doi.org/10.1021/jf0009246

Gong, X., Li, X., Xia, Y., Xu, J., Li, Q., Zhang, C., & Li, M. (2020). Effects of phytochemicals from plant-based functional foods on hyperlipidemia and their underpinning mechanisms. *Trends in Food Science & Technology*, *103*, 304–320. https://doi.org/10.1016/j.tifs.2020.07.026

Gu, E.-J., Kim, D. W., Jang, G.-J., Song, S. H., Lee, J.-I., Lee, S. B., Kim, B.-M., Cho, Y., Lee, H.-J., & Kim, H.-J. (2017). Mass-based metabolomic analysis of soybean sprouts during germination. *Food Chemistry*, *217*, 311–319. https://doi.org/10.1016/j.foodchem.2016.08.113

Gu, L., Wei, B., Fan, R., Jia, X., Wang, X., & Zhang, X. (2015). Development, identification and utilization of introgression lines using Chinese endemic and synthetic wheat as donors. *Journal of Integrative Plant Biology*, *57*(8), 688–697. https://doi.org/10.1111/jipb.12324

He, L. D., Guo, X. N., & Zhu, K. X. (2019). Effect of soybean milk addition on the quality of frozen-cooked noodles. *Food Hydrocolloids*, *87*, 187–193. https://doi.org/10.1016/j.foodhyd.2018.07.048

Herman, E. M. (2014). Soybean seed proteome rebalancing. *Frontiers in Plant Science*, *5*, 437. https://doi.org/10.3389/fpls.2014.00437

Herman, E. M., & Burks, A. W. (2011). The impact of plant biotechnology on food allergy. *Current Opinion in Biotechnology*, *22*(2), 224–230. https://doi.org/10.1016/j.copbio.2010.11.003

Hirasawa, M., Boyer, J. M., Gray, K. A., Davis, D. J., & Knaff, D. B. (1986). The interaction of ferredoxin with chloroplast ferredoxin-linked enzymes. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, *851*(1), 23–28. https://doi.org/10.1016/0005-2728(86)90244-6

Huang, Q., Xu, M., Zhang, H., He, D., Kong, Y., Chen, L., & Song, H. (2019). Transcriptome and proteome analyses of the molecular mechanisms associated with coix seed nutritional quality in the process of breeding. *Food Chemistry*, *272*, 549–558. https://doi.org/10.1016/j.foodchem.2018.07.116

Johnson, H. W., & Bernard, R. L. (1962). Soybean genetics and breeding. In A. G. Norman (Ed.), *Advances in agronomy* (Vol. *14*, pp. 149–221). Academic Press. https://doi.org/10.1016/S0065-2113(08)60438-1

Kito, M., Moriyama, T., Kimura, Y., & Kambara, H. (1993). Changes in plasma lipid levels in young healthy volunteers by adding an extruder-cooked soy protein to conventional meals. *Bioscience, Biotechnology, and Biochemistry*, *57*(2), 354–355. https://doi.org/10.1271/bbb.57.354

Komatsu, S., Yamaguchi, H., Hitachi, K., Tsuchida, K., Kono, Y., & Nishimura, M. (2021). Proteomic and biochemical analyses of the mechanism of tolerance in mutant soybean responding to flooding stress. *International Journal of Molecular Sciences*, *22*(16), 9046. https://doi.org/10.3390/ijms22169046

Krishnan, H. B. (2001). Biochemistry and molecular biology of soybean seed storage proteins. *Journal of New Seeds*, *2*(3), 1–25. https://doi.org/10.1300/J153v02n03_01

Kwon, D. Y., Daily, J. W. III, Kim, H. J., & Park, S. (2010). Antidiabetic effects of fermented soybean products on type 2 diabetes. *Nutrition Research*, *30*(1), 1–13. https://doi.org/10.1016/j.nutres.2009.11.004

Li, R., Sun, R., Hicks, G. R., & Raikhel, N. V. (2015). *Arabidopsis* ribosomal proteins control vacuole trafficking and developmental programs through the regulation of lipid metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(1), E89–E98. https://doi.org/10.1073/pnas.1422656112

Li, W., Cao, W., Wang, P., Li, J., Zhang, Q., & Yan, Y. (2021). Selectively hydrolyzed soy protein as an efficient quality improver for steamed bread and its influence on dough components. *Food Chemistry*, *359*, 129926. https://doi.org/10.1016/j.foodchem.2021.129926

Li, X., Wang, W., Wang, Z., Li, K., Lim, Y. P., & Piao, Z. (2015). Construction of chromosome segment substitution lines enables QTL mapping for flowering and morphological traits in *Brassica rapa*. *Frontiers in Plant Science*, *6*, 432. https://doi.org/10.3389/fpls.2015.00432

Lin, Y., Li, W., Zhang, Y., Xia, C., Liu, Y., Wang, C., Xu, R., & Zhang, L. (2019). Identification of genes/proteins related to submergence tolerance by transcriptome and proteome analyses in soybean. *Scientific Reports*, *9*(1), 1–16. https://doi.org/10.1038/s41598-019-50757-1

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Maforimbo, E., Skurray, G., Uthayakumaran, S., & Wrigley, C. (2008). Incorporation of soy proteins into the wheat–gluten matrix during dough mixing. *Journal of Cereal Science*, *47*(2), 380–385. https://doi.org/10.1016/j.jcs.2007.01.003

Mehrotra, R., Kumar, S., Mehrotra, S., & Singh, B. D. (2009). Seed storage protein gene regulation—A jig-saw puzzle. *Indian Journal of Biotechnology*, *8*, 147–158.

Min, C. W., Gupta, R., Truong, N. V., Bae, J. W., Ko, J. M., Lee, B. W., & Kim, S. T. (2020). A TMT-based quantitative proteomic analysis provides insights into the protein changes in the seeds of high-and low-protein content soybean cultivars. *Journal of Plant Biotechnology*, *47*(3), 209–217. https://doi.org/10.5010/JPB.2020.47.3.209

Morvan, J., Froissard, M., Haguenauer-Tsapis, R., & Urban-Grimal, D. (2004). The ubiquitin ligase Rsp5p is required for modification and sorting of membrane proteins into multivesicular bodies. *Traffic*, *5*(5), 383–392. https://doi.org/10.1111/j.1398-9219.2004.00183.x

Qi, Z., Zhang, Z., Wang, Z., Yu, J., Qin, H., Mao, X., Jiang, H., Xin, D., Yin, Z., Zhu, R., Liu, C., Yu, W., Hu, Z., Wu, X., Liu, J., & Chen, Q. (2018). Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant, Cell & Environment*, *41*(9), 2109–2127. https://doi.org/10.1111/pce.13175

Raatz, S. K., Conrad, Z., Jahns, L., Belury, M. A., & Picklo, M. J. (2018). Modeled replacement of traditional soybean and canola oil with high-oleic varieties increases monounsaturated fatty acid and reduces both saturated fatty acid and polyunsaturated fatty acid intake in the US adult population. *The American Journal of Clinical Nutrition*, *108*(3), 594–602. https://doi.org/10.1093/ajcn/nqy127

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., … Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, *463*(7278), 178–183. https://doi.org/10.1038/nature08670

Shi, X., Chen, Q., Liu, S., Wang, J., Peng, D., & Kong, L. (2021). Combining targeted metabolite analyses and transcriptomics to reveal the specific chemical composition and associated genes in the incompatible soybean variety PI437654 infected with soybean cyst nematode HG1.2.3.5.7. *BMC Plant Biology*, *21*(1), 1–17. https://doi.org/10.1186/s12870-021-02998-4

Song, Q.-X., Li, Q.-T., Liu, Y.-F., Zhang, F.-X., Ma, B., Zhang, W.-K., Man, W.-Q., Du, W.-G., Wang, G.-D., Chen, S.-Y., & Zhang, J.-S. (2013). Soybean GmbZIP123 gene enhances lipid content in the seeds of transgenic Arabidopsis plants. *Journal of Experimental Botany*, *64*(14), 4329–4341. https://doi.org/10.1093/jxb/ert238

Stryiński, R., Mateos, J., Pascual, S., González, Á. F., Gallardo, J. M., Łopieńska-Biernat, E., Medina, I., & Carrera, M. (2019). Proteome profiling of L3 and L4 Anisakis simplex development stages by TMT-based quantitative proteomics. *Journal of Proteomics*, *201*, 1–11. https://doi.org/10.1016/j.jprot.2019.04.006

Takai, T., Ikka, T., Kondo, K., Nonoue, Y., Ono, N., Arai-Sanoh, Y., Yoshinaga, S., Nakano, H., Yano, M., Kondo, M., & Yamamoto, T. (2014). Genetic mechanisms underlying yield potential in the rice high-yielding cultivar Takanari, based on reciprocal chromosome segment substitution lines. *BMC Plant Biology*, *14*(1), 1–11. https://doi.org/10.1186/s12870-014-0295-2

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., & Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote) omics data. *Nature Methods*, *13*(9), 731–740. https://doi.org/10.1038/nmeth.3901

Wang, W., He, Q., Yang, H., Xiang, S., Zhao, T., & Gai, J. (2013). Development of a chromosome segment substitution line population with wild soybean (*Glycine soja* Sieb. et Zucc.) as donor parent. *Euphytica*, *189*(2), 293–307. https://doi.org/10.1007/s10681-012-0817-7

Wilcox, J. R. (1998). Increasing seed protein in soybean with eight cycles of recurrent selection. *Crop Science*, *38*(6), 1536–1540. https://doi.org/10.2135/cropsci1998.0011183X003800060021x

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–362. https://doi.org/10.1038/nmeth.1322

Xin, D., Qi, Z., Jiang, H., Hu, Z., Zhu, R., Hu, J., & Chen, Q. (2016). QTL location and epistatic effect analysis of 100-seed weight using wild soybean (*Glycine soja* Sieb. & Zucc.) chromosome segment substitution lines. *PLoS One*, *11*(3), e0149380. https://doi.org/10.1371/journal.pone.0149380

Xu, Y., Yan, F., Liu, Y., Wang, Y., Gao, H., Zhao, S., Zhu, Y., Wang, Q., & Li, J. (2021). Quantitative proteomic and lipidomics analyses of high oil content GmDGAT1-2 transgenic soybean illustrate the regulatory mechanism of lipoxygenase and oleosin. *Plant Cell Reports*, *40*(12), 2303–2323. https://doi.org/10.1007/s00299-021-02768-4

Yosef, I., Bochkareva, E. S., & Bibi, E. (2010). Escherichia coli SRP, its protein subunit Ffh, and the Ffh M domain are able to selectively limit membrane protein expression when overexpressed. *MBio*, *1*(2), e00020-10. https://doi.org/10.1128/mBio.00020-10

Zhang, G. (2021). Target chromosome-segment substitution: A way to breeding by design in rice. *The Crop Journal*, *9*(3), 658–668. https://doi.org/10.1016/j.cj.2021.03.001

Zhang, T., Dou, W., Zhang, X., Zhao, Y., Zhang, Y., Jiang, L., & Sui, X. (2021). The development history and recent updates on soy protein-based meat alternatives. *Trends in Food Science & Technology*, *109*, 702–710. https://doi.org/10.1016/j.tifs.2021.01.060

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Han, X., Li, J., Zhao, Y., Zhang, Z., Jiang, H., Wang, J., Feng, X., Zhang, Y., Du, Z., Wu, X., Chen, Q., & Qi, Z. (2022). Integrated transcriptomic and proteomic characterization of a chromosome segment substitution line reveals a new regulatory network controlling the seed storage profile of soybean. *Food and Energy Security*, *00*, e381. https://doi.org/10.1002/fes3.381