GM CROPS & FOOD
2021, VOL. 12, NO. 1, 497–508
https://doi.org/10.1080/21645698.2021.2000320

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Proteomics analysis reveals that foreign *cp4-epsps* gene regulates the levels of shikimate and branched pathways in genetically modified soybean line H06-698

Longguo Jin, Daoping Wang, Yongying Mu, Yong Guo, Yangjie Lin, Lijuan Qiu, and Yinghong Pan

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

ABSTRACT

Although genetically modified (GM) glyphosate-resistant soybeans with *cp4-epsps* gene have been widely planted all over the world, their proteomic characteristics are not very clear. In this study, the soybean seeds of a GM soybean line H06-698 (H) with *cp4-epsps* gene and its non-transgenic counterpart Mengdou12 (M), which were collected from two experiment fields in two years and used as 4 sample groups, were analyzed with label-free proteomics technique. A total of 1706 proteins were identified quantitatively by label-free quantification, and a total of 293 proteins were detected as common differential abundance proteins (DAPs, FC is not less than 1.5) both in two groups or more. Functional enrichment analysis of common DAPs identified from four groups, shows that most up-regulated proteins were clustered into stress response, carbon and energy metabolism, and genetic information processing. Further documentary analysis shows that 15 proteins play important roles in shikimate pathways, reactive oxygen species (ROS) and stress response. These results indicated that the change of protein abundance in different samples were affected by various factors, but except shikimate and branched pathways related proteins, only ROS and stress-related proteins were found to be stably regulated by *cp4-epsps* gene, and no unexpected and safety-related proteins such as antinutritional factors, allergenic proteins, and toxic proteins were found as DAPs. The influence of foreign genes in genetically modified plants is worthy of attention and this work provides new clues for exploring the regulated proteins and pathways in GM plants.

ARTICLE HISTORY
Received 15 June 2021
Revised 23 October 2021
Accepted 26 October 2021

KEYWORDS
Cp4-epsps gene; genetically modified; shikimate pathway; label-free proteomics; soybean

1. Introduction

Soybean (*Glycine max* L. Merril) is one of the important foods or feeds for human and animal nutrition, because of its unique characteristics as a vegetal source of proteins and oil. With the rapid growth of population, the demand for crop yield and quality is increasing.[1]Genetic modification (GM) is one of the most effective methods to solve the above problems. In 1996, glyphosate-resistant soybean variety Roundup Ready Soybean (RRS; event GTS40-3-2,) was allowed to be cultivated commercially. RRS expresses a protein of CP4 EPSPS from the bacteria *Agrobacterium tumefaciens* CP4, which makes plants tolerate exposure to herbicide glyphosate.[2] Since 1996, GM crops have been widely planted in the world, which reduce production costs and increase yield, bring huge economic benefits, but their impact on the environment and their safety as food and feed has always been concerned. And also, foreign genes integrated into the genome may lead to the insertion, deletion, or rearrangement of genes thereby affecting some biological pathways, causing some unintended side effects. We call it unintended effect which is one of the core contents of the safety assessment of transgenic plants.[3]

With the increasing use of omics profiling techniques such as transcriptomics, proteomics, metabolomics, and multi-profiling techniques,[4] we can make a more complete comparative analysis, and extend our understanding of the biological safety of GM crops. These techniques have been proposed to detect unintended effects in GM plants.[5] Among these profiling techniques, proteomics technologies develop rapidly and become a powerful tool to evaluate the unintended effects of GM plants. Proteomics has been a direct method to evaluating unintended effects at the protein level. The two-dimensional gel electrophoresis (2-DE), a first-generation proteomics technique, has been

CONTACT Yinghong Pan panyinghong@caas.cn; Lijuan Qiu qiulijuan@caas.cn Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
Supplemental data for this article can be accessed on the publisher's website.

commonly used in proteomics research for many years.[6] There have been several reports published using 2-DE analyzed unintended effects of *cp4-epsps* gene transgenic soybean line MSOY 7575. Using this method, 10 proteins showed significant differences in expression.[7] Campos[8] used more sensitive two-dimensional differential in gel electrophoresis (2D-DIGE) technique to assess different generations of transgenic (*cp4-EPSPS* gene) and non-transgenic soybean plants by proteomics, and found 24 differentially abundant protein spots by 2-D DIGE.,

Despite 2DE technology has better resolution power, it still has some shortcomings, such as laborious, poor repeatability and so on.[6] To overcome these limitations, proteomics technology coupled with mass spectrometry (MS) has been widely used in the identification of proteins in genetically modified crops. Subsequently, a label-free technique that can be used in combination with SDS-PAGE and reverse-phase chromatography to evaluate the changes in the proteome of new crop varieties.[9]

Although GM glyphosate resistant soybeans with *cp4-epsps* gene have been widely planted all over the world, the proteomic characteristics of them remains unclear. In this study, we used the transgenic soybean line H06-698 (H) and non-transgenic counterpart Mengdou12 (M) as research materials. Label-free proteomics technology was used to compare the two soybean lines to evaluate the DAPs profiles, and explore the regulated proteins in main and branched biological pathways affected by cp4-epsps gene, as well as unexpected and safety-related proteins such as anti-nutritional factors, allergenic proteins, and toxic proteins.

2. Materials and Methods

2.1. Plant Materials

The transgenic soybean line H06-698 (H), which contains the *cp4-epsps* gene, and its recurrent parental line Mengdou12 (M) were planted in two experimental fields in 2015 and 2017, respectively. The experiment fields are located in Molidawa Daur Autonomous Banner (49°23'N,124°37'E) and Zhalantun City (47°59'N,122°42'E) of Hulunbuir, Nei-Monggol Autonomous Region. H (H06-698) is GM soybean, of which H1 and H2 are seeds harvested in 2015, H3 and H4 are seeds harvested in 2017. M (Mengdou12) is a non-transgenic recurrent parental line, of which M1 and M2 are seeds harvested in 2015, M3 and M4 are seeds harvested in 2017. All samples were divided into four groups (H1/M1, H2/M2, H3/M3, and H4/M4).

2.2. Protein Extraction and Trypsin Digestion

The workflow of the soybean proteomic analysis was illustrated in Figure 1. The protein extraction method was according to the filter-aided sample preparation (FASP).[10,11] Fifty mg soybean seeds powder of each sample (the third groups of samples was treated with ice acetone at 4°C for 12 h before extraction) was extracted with 500 μL protein extraction buffer (8 M urea, 2 mM EDTA, 20 mM $CaCl_2$, 200 mM NaCl, 100 mM Tris-HCl, pH 8.2) in 2 mL centrifuge tubes and freeze-thawed three times. Then, centrifuged at 4°C and 30,000 g for 15 min, 400 mL clear supernatant was transferred into the ultrafiltration device (Amicon Ultra 0.5 mL Centrifugal Filters, 10 K,

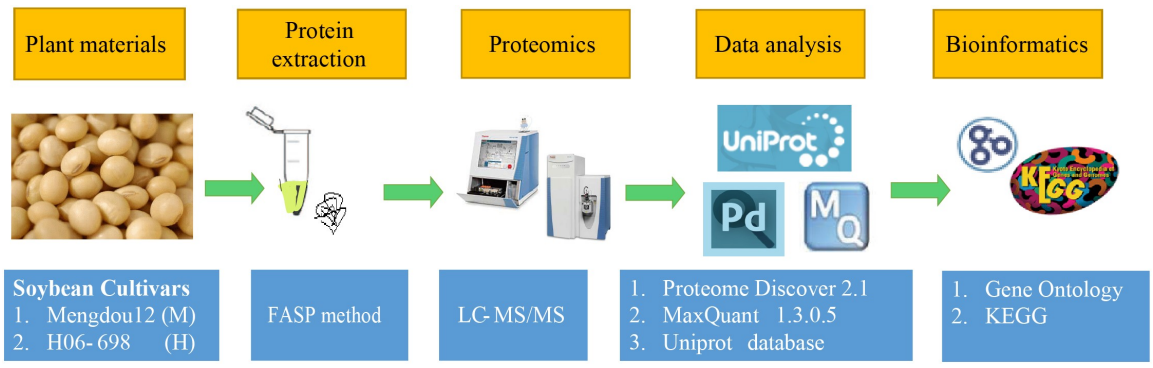

Figure 1. Experimental design and workflow for quantitative proteomic analysis of soybean seeds using label-free LC/MS technology.

Millipore). The device was centrifuged at 30,000 g for 30 min, then 200 µL DTT (500 mM DTT, 8 M urea, 100 mM Tris-HCl, PH8.2) was added and incubated at 30°C for 1 h after centrifugation at 30,000 g for 30 min. The concentrate was diluted with 200 µL UA buffer (8 M urea, 100 mM Tris-HCl, pH 8.2) and centrifuged twice at 30,000 g for 30 min. Next, 100 µL of 50 mM iodoacetamide in UA buffer (8 M urea, 100 mM Tris-HCl, pH 8.2) was added to the concentrate, incubating at room temperature for 30 min in darkness before centrifuged. The resulting concentrate was diluted three times with 100 µL of 8 M urea and twice with 100 µL 25 mM ammonium bicarbonate. Finally, 40 µL trypsin buffer (2 µg trypsin in 40 µL 25 mM NH_4HCO_3) was added and incubated for 12 h at 37°C and then centrifuged.

2.3. Mass Spectrometry (MS) Analysis

The resulting peptide mixtures (5 µL injection) were separated with a flow rate of 300 nL/min on a C18 Acclaim PepMap 100 column 3 µm 15 mm using EASY-nLC 1000 nanoHPLC system (Thermo Scientific, Bremen, Germany). The mobile phases applied were buffer A (H_2O: FA = 99.9:0.1, v/v) and buffer B (ACN: FA = 99.9:0.1, v/v). Peptide elution was performed by a gradient from 3% to 60% for buffer B over 4 min, and maintained at 60% buffer B for 2 min, followed by return to 5% in 3 min, then ramping up to 100% buffer B over 7 min, finally equilibrated in buffer A for 4 min. The fractions were transferred to Q-EXACTIVE plus tandem mass spectrometry (Thermo Fisher Scientific, San Jose, CA) by nanoelectrospray ionization. Full scan mass spectra were used to detect intact peptides with 300–1800 m/z at a resolution of 70,000.

A data dependent procedure was employed as application of spectra acquisition. The selected precursor was fragmented in the high-energy collision dissociation cell with a normalized collision energy setting of 27.0. Fragments were detected in the Orbitrap at a resolution of 17,500. TopN strategy was used to optimize the ability to select precursor ions. The setting of TopN for MS/MS was 20. Automatic gain control (AGC) was used to optimize the spectra by controlling the number of ions that are injected into the orbitrap analyzer. The value of AGC target for full MS was 3e6 and 1e5 for MS/MS.

2.4. Protein Identification and Label-Free Quantification (LFQ) Data Analysis

Mass spectrometric raw data were analyzed using Proteome Discover 2.1 (Thermo Fisher Scientific) and MaxQuant version 1.3.0.5 and Uniprot database. In the database search, full tryptic specificity was required with tolerance set at one missed cleavage. Carbamidomethyl cysteine was set as a fixed modification. N-terminal acetylation and methionine oxidation of proteins were set as variable modifications. Mass tolerance for the initial precursor was set at 20 ppm and 0.02 Da for fragmented ions. For protein identification, data was filtered with a false discovery rate (FDR) of <1%.

2.5. Bioinformatics and Statistical Analysis

To compare the intensity and sequence coverage distribution of quantitative identified protein, the average value of H and M harvested at the same year were calculated. In these groups, "H1-2" showed the average value of H1 and H2, "M1-2" showed the average value of M1 and M2, "H3-4" showed the average value of H3 and H4, "M3-4" showed the average value of M3 and M4. Then, compared *p*-value of group "H1-2" and "M1-2," *p*-value of group "H3-4" and "M3-4," *p*-value of group "H1-2" and "H3-4," *p*-value of group "M1-2" and "M3-4."

To identify functional terms of DAPs, Gene Ontology (GO) enrichment was performed by agriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php). GO terms with computed *p*-values less than 0.05 were considered significantly enriched. Principal component analysis was visualized using R package ggplot2_3.3.0 and ggfortify_0.4.9. Pearson's correlations analysis was visualized using R package corrplot_0.84. The heatmap was visualized using R package pheatmap_1.0.12. The selection criterion for the DAPs was fold change > 1.5 or fold change < 0.667.[12] Pathway enrichment analysis was performed by MapMan (https://mapman.gabipd.org/).

3. Results

3.1. Identification of Proteins from Quantitative Proteomics Analysis

To examine the repeatability of LFQ intensity in four groups of H and M samples, principal component analysis showed that the first principal component (PC1) represented 95.33% and second principal component (PC2) represented 3.56% of the total variance. The difference of samples harvested in the same year was smaller, whether H or M, the difference of samples harvested in different years was larger (Figure 2a). Pearson correlation coefficients showed a similar result with principal component analysis (Figure 2b). A number of 1706 protein groups were obtained from the label-free proteomic analysis (Table S1). And the distribution of sequence coverage and intensity are shown in Figure 2c and Figure 2d. The distribution of sequence coverage showed that the protein number of H was significantly higher than that of M in the coverage of 10–20% and 50–60%, it showed that there were differences existing in H and M samples which were harvested at the same years (Figure 2d). The distribution of signal intensity showed that although there were differences between H and M samples in different years, in general, many proteins were distributed in the range of medium and high signal intensity (Figure 2c).

3.2. Identification and Comparison of Differential Abundance Proteins (Daps)

Proteins that conformed to the following screening criteria were deemed as DAPs: fold change (FC) >1.5. Among of identified proteins, 348, 444, 391, and 304 proteins were deemed as DAPs in 4 groups of samples (H1 vs M1, H2 vs M2, H3 vs M3, H4 vs M4) (Table S2-5), respectively. Finally, we detected a total of 293 proteins as common DAPs both in two groups or more. Hierarchical cluster analysis of common DAPs was performed to better visualize

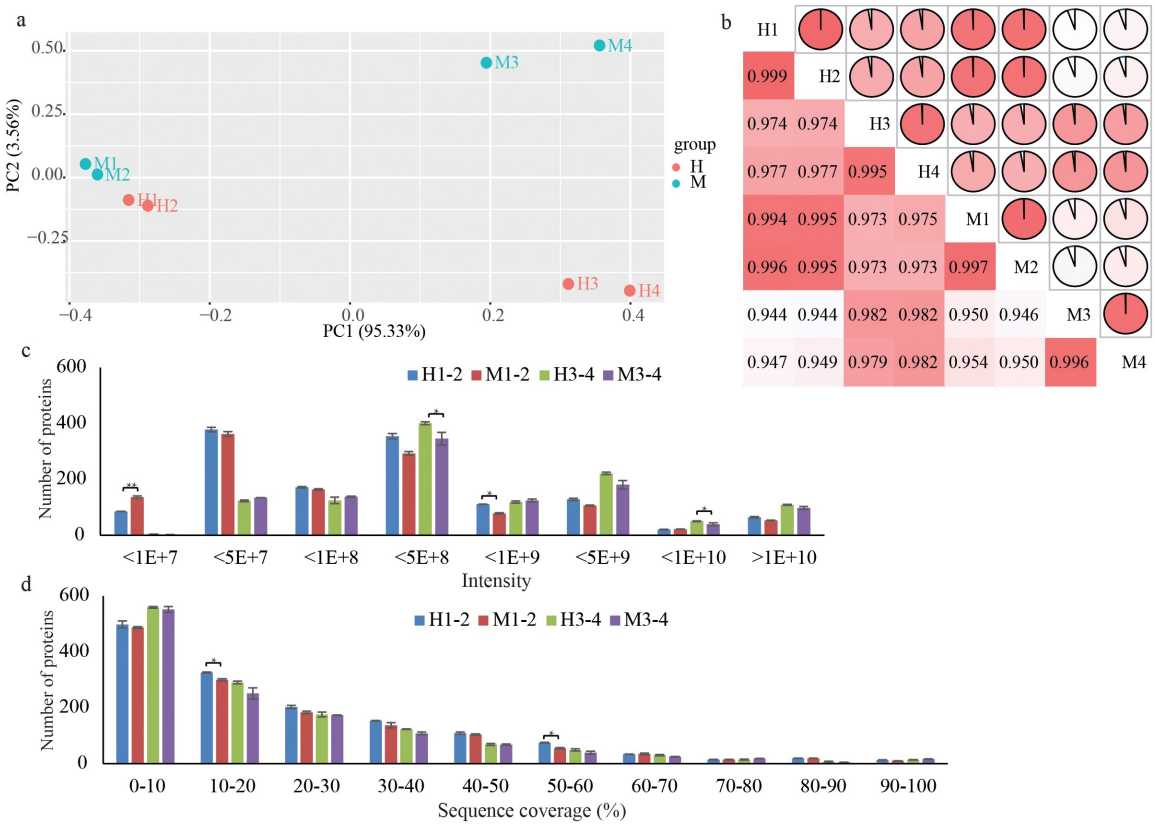

Figure 2. Identification of proteins from quantitative proteomics analysis. (a) Principal component analysis of 4 groups of H and M samples. (b) Pearson correlation analysis of 4 groups of samples. (c) The intensity distribution of quantitative identified protein. (d) The sequence coverage distribution of quantitative identified protein. "H1-2" showed the average value of H1 and H2. It was the same as "M1-2," "H3-4" and "M3-4." * Indicated *p*-value < 0.05, ** indicated *p*-value < 0.01.

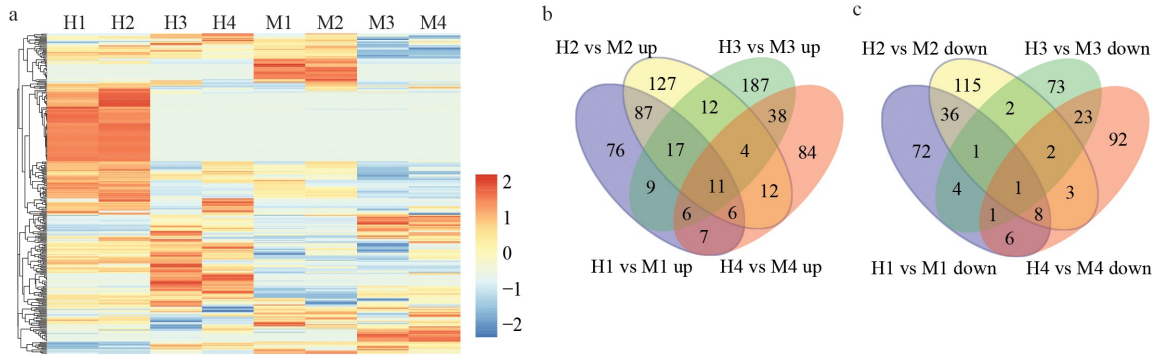

Figure 3. Comparison of proteomics data. (a) Hierarchical cluster analysis of 293 common DAPs in H and M. (b) Venn diagrams represent the overlap of up-regulated DAPs. (c) Venn diagrams represent the overlap of down-regulated DAPs.

the differences in protein abundance among four groups (Figure 3a). Among all of DAPs, 11 proteins were up-regulated in H samples of all groups (Figure 3b), and 1 was down-regulated in H samples of all groups (Figure 3c). Moreover, three proteins which had common change in four groups were also listed in Table 1. They are 26S proteasome regulatory subunit 4-like A, peroxiredoxin and cullin-associated NEDD8-dissociated protein 1 isoform A. The fold change of them in four groups were all larger than 1.2 (Figure 3b, c; Table 1). It is worth noting that a protein encoded by the foreign gene *cp4-epsps* was identified in all the GM soybean H and the intensity of CP4-EPSPS was all higher than 1E+9 (Table S1).

3.3. Bioinformatic Analysis of Differential Abundance Proteins (Daps)

To identify functional terms of DAPs, GO analysis was performed and the common enriched terms (P-value < 0.05) for up-regulated and down-regulated DAPs are shown in Figure 4a and 4b. For the up-regulated DAPs, most of the

Table 1. Fifteen up-regulated and down-regulated DAPs in four groups of samples.

Protein ID	Uniprot ID	Protein name	Function or related pathway	Intensity Ratio			
				H1/M1	H2/M2	H3/M3	H4/M4
Glyma.17G075300.1.p	I1MT35(100%, Glycine soja)	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	Reactive oxygen species	1.55	1.54	2.47	1.53
Glyma.19G171700.1.p	A0A0R0EXW6(100%, Glycine max)	26S proteasome regulatory subunit 4-like A	Ubiquitin/26s proteasome	131.17	1.97	139.31	127.53
Glyma.10G267200.1.p	A0A0R4J4C8(100%, Glycine max)	Mitochondrial ATP synthase subunit beta	Reactive oxygen species	1.62	1.66	1.53	1.79
Glyma.15G140300.1.p	I1MGE7(100%, Glycine max)	ATP citrate synthase	Reactive oxygen species	1.75	1.91	199.68	204.88
Glyma.09G178100.1.p	A0A445J2H7(100%, Glycine soja)	β–glucosidase 44 isoform A	phenylpropanoids pathway	2.29	2.24	380.73	1.87
Glyma.05G204800.1.p	A0A445KRA3 (Glycine soja)	thaumatin-like proteins	Disease/defense	1.84	1.63	632.45	2.32
Glyma.14G121200.1.p	A0A0B2PNL1 (100%, Glycine soja)	Alcohol dehydrogenase 1	Disease/defense	3.68	2.27	4.91	2.77
Glyma.05G133500.1.p	A0A0R0K508 (100%, Glycine max)	1-Cys Peroxiredoxin	phenylpropanoids pathway	185.62	198.19	245.44	147.65
EPSPS_CP4	Q587N9(100%, Glycine max)	3-phosphoshikimate 1-carboxyvinyltransferase	shikimate pathway	6666.12	8015.79	13,227.10	11,126.70
Glyma.10G067500.1.p	A0A0R0I177(100%, Glycine soja)	Monodehydroascorbate reductase	Reactive oxygen species	1.85	1.73	1.52	1.59
Glyma.18G187600.1.p	K7MT91(100%, Glycine soja)	NADPH-dependent aldo-keto reductase	Reactive oxygen species	83.56	186.74	1.85	1.83
Glyma.10G047700.1.p	I1L8Q0(100%, Glycine max)	Lambda class glutathione S-transferase	phenylpropanoids pathway	0.29	0.23	0.50	0.44
Glyma.02G042500.3.p	Q9SDY6(100%, Glycine max)	Chitinase class I	Disease/defense	1.46	1.56	116.43	162.55
Glyma.07G272400.1.p	K7L432(100%, Glycine soja)	Cullin-associated NEDD8-dissociated protein 1 isoform A	Ubiquitin/26s proteasome	71.31	54.97	69.22	1.20
Glyma.16G204600.1.p	I1MQ89 (100%, Glycine max)	Phosphopyruvate hydratase (Enolase)	shikimate pathway	1.42	1.49	371.45	2.20

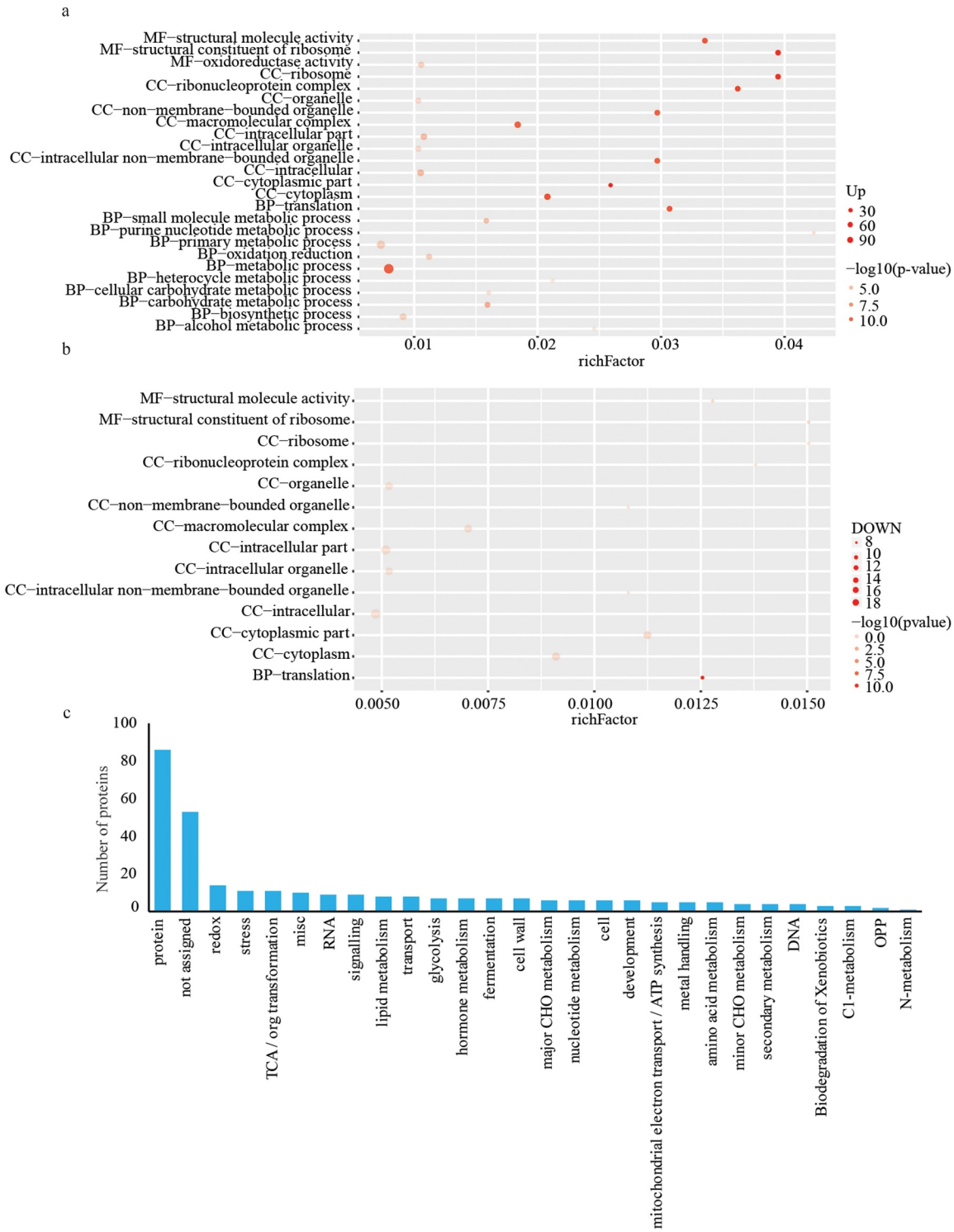

Figure 4. Function classification of differential abundance proteins (DAPs) from H and M. (a) Gene ontology analysis of up-regulated proteins. (b) Gene ontology analysis of down-regulated protein. Y-axis represented each GO term; X-axis represented rich factor, size of bubble represented the number of DAPs in each category, the depth of color represented -log10 (*p*-value) CC represented cellular component, BP represented biological process, and MF represented molecular function. (c) Pathway analysis of DAPs. TCA, tricarboxylic acid; CHO, carbohydrate.

items with rich factor greater than 0.025 and *p*-value less than 0.05 were cellular components, except translation and structural molecule activity. For the down-regulation DAPs, most of the items with *p*-value less than 0.05 were mainly classified as cellular components.

For the up-regulated DAPs, proteins were mainly distributed in translation, carbohydrate metabolic process, oxidation reduction, biosynthetic process, and primary metabolic process. Many of them had structural constituent of ribosome, structural molecule activity, and oxidoreductase activity. For the cellular component group, many of them were enriched in ribonucleoprotein complex, cytoplasm, non-membrane-bounded organelle, macromolecular complex, cytosol and so on (Figure 4a). For the down-regulation DAPs, proteins with the molecular function classification were involved mainly with structural constituent of ribosome, and structural molecule activity. Proteins with the cellular component classification were involved in ribosome, ribonucleoprotein complex, non-membrane-bounded organelle, macromolecular complex, cytoplasm and so on. In the biological process classification, proteins were involved in translation (Figure 4b).

To further understand the function of DAPs, pathway analysis was carried out on four groups of DAPs (Figure 4c). The results showed that many DAPs in the H of four groups were clustered in protein pathway. Pathways like redox, stress, and TCA/org transformation were enriched over 10 DAPs. Some pathways had greater than 5 DAPs enriched, such as RNA, signaling, lipid metabolism, transport, glycolysis, hormone metabolism, fermentation, cell wall, major CHO metabolism, nucleotide metabolism, cell, development, mitochondrial electron transport/ATP synthesis, metal handling, and amino acid metabolism. These pathways were clustered into stress response, carbon and energy metabolism, and genetic information processing.

Further analysis shows that 15 proteins play important roles in shikimate pathways, reactive oxygen species (ROS), and stress response, such as enolase (encoded by Glyma.16G204600.1.p), that involved in the pathway of glycolysis, and more, it was also an important protein in shikimate pathway.

ROSs are detoxified by antioxidants such as glutathione, ascorbate, thioredoxin, and carotenoids, and by antioxidative enzymes such as superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione peroxidase (GPx) and glutathione-S-transferase (GST).[13] In this category, we have found five ROS-related enzymes that were highly accumulated in GM Soybean H. Glyma. 10G067500.1.p encodes monodehydroascorbate reductase (MDAR), and it is one of the key enzymes in the conversion of oxidized ascorbate (AsA) back to reduce AsA in plants.[14] MDAR plays an important role in the regulation of oxidative stress in plants.[15] Glyma.10G267200.1.p encodes mitochondrial ATP synthase subunit beta. ATP synthase has also been shown to be a significant factor in the regulation of mitochondrial redox activity and ROS production.[16] GAPDH is a key enzyme in glycolysis and has multiple functions independent of its role in energy metabolism. Glyma.17G075300.1.p encodes NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (NADP-GAPDH). Previous studies have shown that GAPDH interacts with the plasma membrane-associated phospholipase D to produce the ROS in *Arabidopsis*, which can act as a signal molecule in response to ROS and water stress.[17]Glyma.18G187600.1.p encodes NADPH-dependent aldo-keto reductase (AKR), that can increase the levels of antioxidant glutathione and enhance activity of SOD, APX and glutathione reductase (GR). It suggests that AKR1 could protect these enzymes from the reactive carbonyl compounds (RCC) induced protein carbonylation by detoxification process.[18] Glyma.15G140300.1 encodes ATP-citrate synthase beta-chain protein 1, ATP-citrate lyase (ACL) catalyzes citrate to oxaloacetate (OA) and acetyl-CoA in cell cytosol. It plays a key role in the tricarboxylic (TCA) cycle. ACL also plays important roles in normal plant growth and in the biosynthesis of some secondary metabolites.[19] Thioredoxins (Trxs) are key factors in modulating ROS scavenging.[20] Trxs-targets have been identified including several enzymes of the TCA cycle such as citrate synthase in *Arabidopsis*.[21]

In addition to these proteins, some proteins have multiple functions. For instance, Glyma.10G047700.1.p (GSTLs) and Glyma.05G133500.1.p (Prxs) not only participates in the flavonoid pathway but also acts as an antioxidant that can eliminate ROS activity.[22]

We have detected three proteins highly accumulated in GM Soybean H. These proteins are involved in the disease/defense category. They belong to the chitinase class I (Glyma.02G042500.3.p), thaumatin-like proteins (Glyma.05G204800.1.p), alcohol dehydrogenase 1

(Glyma.14G121200.1.p). Plants usually respond to fungal or bacterial pathogens by expressing a group of proteins called pathogenesis-related proteins (PR). Plant chitinases are the members of the PR proteins family and protect plants from biotic and abiotic stress. The class I chitinase has a variety of functions, such as cold stress signaling pathway,[23] antifungal activity.[24] Thaumatin-like protein (TLP) is a large family in plants, and individual members play different roles in responses to various biotic and abiotic stresses.[25] TLP proteins are functionally diverging ant their genes have been shown to participate in cold, salt and drought stress responses.[26] Alcohol dehydrogenase (ADH) is a key enzyme responsible for catalyzing the reduction of pyruvate to ethanol. It also plays a critical role in various abiotic stresses, including salt, dehydration, cold, hypoxia, ABA, and H_2O_2.[27]

4 Discussion

The global seed proteins of *cp4-epsps* gene modified soybean H and its non-transgenic counterpart M were analyzed by proteomics. Principal component analysis and Pearson correlation analysis showed that the correlations in four groups (Figure 2a, Figure 2b), and it was indicated that samples which harvested at same year had higher similarity, cultivation environmental factors and proteomic experimental factors may be responsible for the large differences between four groups H and M. The distribution of sequence coverage and intensity showed that data quality was stable in different samples and experimental results were reproducible (Figure 2c, Figure 2d). Among 1706 identified proteins, over 300 DAPs with over 1.5-fold change were detected from each group, meaning that the change of protein abundance in different samples of transgenic cp4-epsps gene soybean seeds was affected by various factors. However, only 12 DAPs with over 1.5-fold change were accurately detected from all groups at same time (Figure 3b,Figure 2c), meaning that they were stably regulated by foreign cp4-epsps gene. GO and MapMan analysis of 293 common DAPs showed that they were mainly related to energy metabolism, redox, and stress. Among all of DAPs, 15 proteins were found playing important roles in pathways of shikimate, ROS, and stress response (Table 1).

EPSPS mediates the synthesis of aromatic amino acids and some secondary metabolites in shikimate pathway. The EPSPS is not only an important regulatory site for shikimic acid accumulation in plants but also a central target enzyme of herbicide glyphosate. Because a new foreign cp4-epsps gene was integrated into the soybean genome, cell metabolism may be further altered through a series reaction. In this study, we have found some changes in the expression level of up-stream and down-stream proteins involved in some metabolites originated in the shikimate pathway. Therefore, we hypothesize that GM soybean is looking for a new metabolic balance in order to maintain the normal growth under stress conditions. The proteomics results also indicate only ROS and stress-related proteins are found to be regulated, and no other unexpected proteins were found.

4.1 The Shikimate Pathway Proteins

In this study, we found that two up-regulated DAPs involved in shikimate pathway. The shikimate pathway mainly connects primary metabolism and secondary metabolism, starting from the glycolytic pathway protein phosphoenolpyruvate (PEP) and oxidative pentose phosphate pathway protein erythose-4-phosphate (E4P), are the precursor for the seven-step of shikimate pathway.[28] Enolase (ENO) is a key glycolytic enzyme, and it is a multifunctional protein.[29] Under the action of PEP and E4P, the starting enzyme 3-deoxy-D-arabinohe-ptulosonate 7-phosphate synthase (DAHPS) of shikimate pathway was produced.[30] Then through a multi-step catalytic reaction, we produce important substances chorismate, which is a precursor to a large number of secondary metabolites including flavonoids.[28] In this study, the DAP Glyma.16G204600.1.p (encoding enzymes enolase, ENO) and target protein CP4-EPSPS was highly accumulated in GM Soybean H. We suppose that increasing levels of ENO in GM soybean H may be regulated by other transcription factors. For instance, *Arabidopsis AtMYB12* can induce both primary and secondary metabolism, binding directly to the promoters of genes *DAHPS* and *ENO*, through its ability to regulate metabolism at multiple levels,[31] then ENO may lead to regulate *cp4-epsps* gene expression through shikimate pathway and branch pathway-related genes.

4.2 The Phenylpropanoid Pathway Proteins

The phenylpropanoids pathway is synthesized from the aromatic amino acid phenylalanine, which is one of the end products of the shikimate pathway. Many plant secondary metabolites are derived from phenylpropanoids pathway, such as anthocyanidin, lignin, and flavonoids. The flavonoids include flavones, flavonols, anthocyanins, and proanthocyanidins (PAs).[32] PAs as one of the final products of the flavonoid pathway, they are synthesized from the first metabolites via the shikimate and flavonoid pathways.[33] In this study, we found that three DAPs involved in phenylpropanoids pathway were up-regulated or down-regulated. The protein Glyma.09G178100.1.p encoding β-glucosidases was highly accumulated in GM Soybean H. The various defense-related β-glucosidases have the function of hydrolyzing isoflavonoid glucosides from legumes.[34] The protein Glyma.10G047700.1.p encoding lambada class glutathione S-transferase (GSTLs) was down accumulated in GM Soybean H. GSTs have been widely studied in relation to herbicide tolerance and detoxification.[35] A novel role of GSTLs has also been found in maintaining the flavonoid pool under stress conditions.[35] The protein Glyma.05G133500.1.p encoding 1-Cys peroxiredoxin (1-Cys Prx) was highly accumulated in GM Soybean H. Peroxiredoxins (Prxs) are peroxidases with multiple functions in the antioxidant defense and redox signaling. In many plants, 1-Cys Prx is a seed-specific antioxidant which eliminates reactive oxygen species (ROS).[22] For instance, the abundance of transcripts encoding 1-Cys Prx and flavonoid 3-O-galactosyltransferase correlates with the accumulation of soluble PAs in young persimmon fruits.[36]

43 Ubiquitin/26s Proteasome Pathway Proteins.

The ubiquitin/26S proteasome system (UPS) mediated protein degradation controls various developmental pathways in eukaryotes. In this study, we have found two proteins highly accumulated in GM soybean H. The first one is Glyma.07G272400.1.p, encoding Cullin-Associated Nedd8-Dissociated 1 (CAND1), Cullin-RING E3 ubiquitin ligases (CRLs) control many biological pathways through targeted ubiquitylation of signaling proteins.[37] Apple RING E3 ligase MdMIEL1 inhibits anthocyanin accumulation through ubiquitinating and degrading of MdMYB1 protein.[38] The second protein is Glyma.19G171700.1.p, which encoding 26S protease regulatory subunit. In *Arabidopsis*, loss of the function of 26S proteasome results in higher accumulation of anthocyanin in vegetative tissues.[39] Many studies have demonstrated that the turnover of several anthocyanin pathway regulators is controlled by the UPS.[40] In plants, numerous developmental and environmental factors, influence the expression of anthocyanin.[41] We suppose that the anthocyanin and the protein complex is ubiquitinated by ubiquitin E3 ligase CAND1 and then degraded through the 26S proteasome pathway.

Finally, according to the results of proteome analysis and literature analysis, we propose a model pathway of proteins related to shikimic acid (Figure 5). ENO is a multi-functional protein. ENO may lead to regulating *cp4-epsps* gene expression through shkimate pathway and branch pathway-related genes. Phenylpropanoids are synthesized from the aromatic amino acid phenylalanine as one of the end products of the shikimate pathway. A variety of secondary plant products are derived from phenylpropanoid metabolism, such as flavonoids and anthocyanidin. Then, the defense-related β-glucosidases have hydrolyzing isoflavonoid glucosides. GSTLs maintain the flavonoid pool under stress conditions. Prx and flavonoid correlates with the accumulation of PAs. PAs as one of the final products of the flavonoid pathway, they are synthesized from the first metabolites via the shikimate and flavonoid pathways. PAs entered into the ubiquitin/26S proteasome-mediated protein degradation pathways.

5. Conclusions

In seeds of this GM soybean, over one thousand proteins were quantitatively identified, and hundreds of DAPs were detected, but only 15 proteins were stably regulated by transgene *cp4-epsps*. Half of the 15 DAPs were related to some metabolites originated in the shikimate pathway, other played roles in ROS or stress response, and no unexpected and safety-related

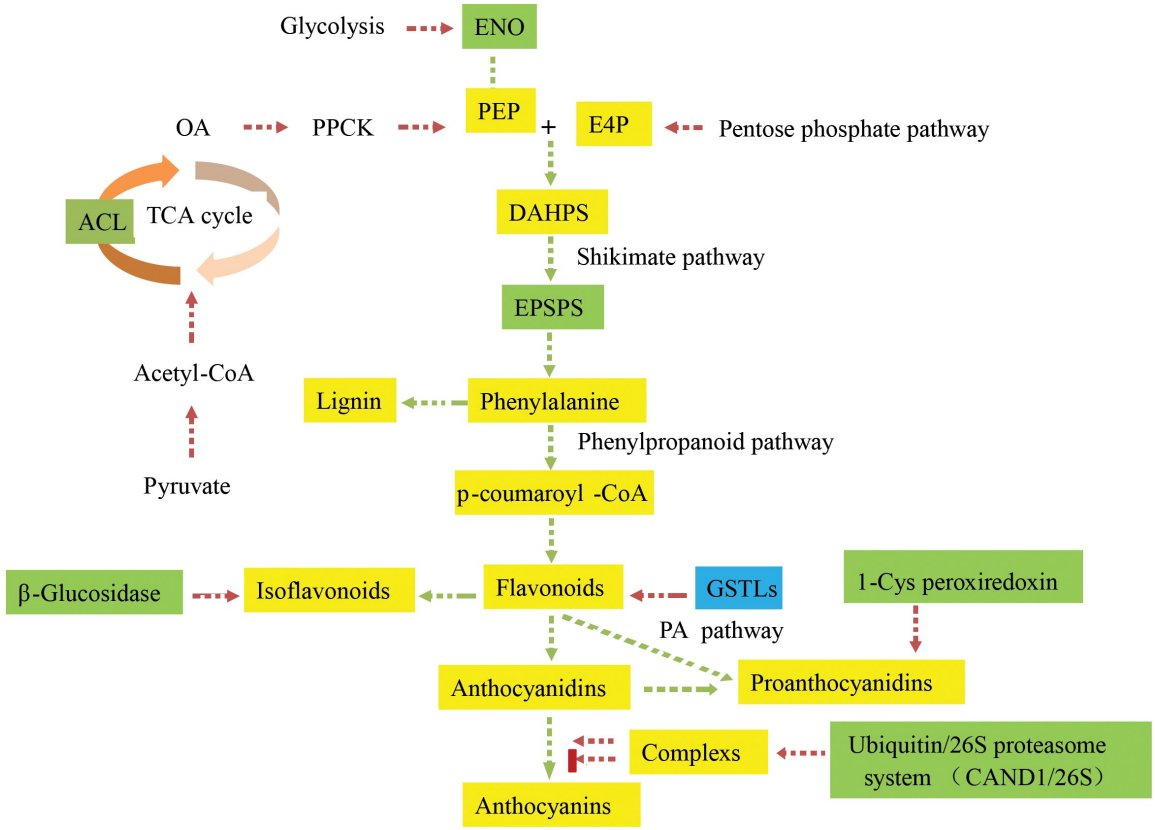

Figure 5. The schematic representation general shikimate pathway, down-stream and/or branched metabolic pathway related proteins. Among colored background boxes, green means up-accumulated DAPs, blue means down-accumulated DAPs, yellow means undetected or unchanged proteins, or metabolites. The green dashed arrow indicates shikimate and the downstream pathway, red dashed arrow indicates branched pathway. Abbreviations: ACL, ATP-citrate lyase; CAND1, CULLIN-ASSOCIATED NEDD8-DISSOCIATED 1 (CAND1); DAHPS, 3-Deoxy-D-Arabino-Heptulosonate 7-Phosphate Synthase; E4P, D-Erythrose 4-phosphate; ENO, Enolase; EPSPS, 5-Enolpyruvylshikimate-3-phosphate synthase; GSTLs, Lambada class glutathione S-transferase; OA, oxaloacetate; PEP, phosphoenolpyruvate; PA, proanthocyanidins.

proteins such as antinutritional factors, allergenic proteins, and toxic proteins were found.

Availability of data

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD025908.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by "National Key Research and Development Program of China"(2016YFD0101005) and "National Transgenic Major Program of China" (2016ZX08004-001).

ORCID

Yinghong Pan http://orcid.org/0000-0002-4025-5193

References

1. Herrera-Agudelo M 1, Miró A, Arruda, M. A. M. *In vitro* oral bioaccessibility and total content of Cu, Fe, Mn and Zn from transgenic (through cp4 EPSPS gene) and nontransgenic precursor/successor soybean seeds. Food Chemistry. 2017;15(225):125–31. doi:10.1016/j.foodchem.2017.01.017.
2. Bohn T, Cuhra M, Traavik T, Sanden M, Fagan J, Primicerio R. Compositional differences in soybeans on the market: glyphosate accumulates in roundup ready GM soybeans. Food Chemistry. 2014;153:207–15. doi:10.1016/j.foodchem.2013.12.054.
3. Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Smith M, Kärenlampi S, Kok EJ, Leguay -J-J. Unintended effects and their detection in genetically modified crops. Food and Chemical Toxicology. 2004;42(7):1089–125. doi:10.1016/j.fct.2004.02.003.

4. Barros E, Lezar S, Anttonen MJ, Van Dijk JP, Rohlig RM, Kok EJ, Engel KH. Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. Plant Biotechnology Journal. 2010;8(4):436–51. doi:10.1111/j.1467-7652.2009.00487.x.
5. Ladics GS, Bartholomaeus A, Bregitzer P, Doerrer NG, Gray A, Holzhauser T, Glenn K, Keese P, Kok E, Macdonald P. Genetic basis and detection of unintended effects in genetically modified crop plants. Transgenic Research. 2015;24(4):587–603. doi:10.1007/s11248-015-9867-7.
6. Arruda SCC, Barbosa HS, Azevedo RA, Arruda MAZ. Two-dimensional difference gel electrophoresis applied for analytical proteomics: fundamentals and applications to the study of plant proteomics. Analyst 2011;136(20):4119–26. doi:10.1039/c1an15513j.
7. Brandao AR, Barbosa HS, Arruda MA. Image analysis of two-dimensional gel electrophoresis for comparative proteomics of transgenic and non-transgenic soybean seeds. Journal of Proteomics. 2010;73(8):1433–40. doi:10.1016/j.jprot.2010.01.009.
8. Campos BK, Galazzi RM, Santos BM, Balbuena TS, Santos FN, Mokochinski JB, Arruda MA, Arruda MAZ. Comparison of generational effect on proteins and metabolites in non-transgenic and transgenic soybean seeds through the insertion of the *cp4-EPSPS* gene assessed by omics-based platforms. Ecotoxicol Environ Saf 2020;202:110918. doi:10.1016/j.ecoenv.2020.110918.
9. Mora L, Bramley PM, Fraser PD. Development and optimisation of a label-free quantitative proteomic procedure and its application in the assessment of genetically modified tomato fruit. Proteomics. 2013;13(12–13):2016–30. doi:10.1002/pmic.201200480.
10. Dixon DP, Skipsey M, Edwards, R. M. Role of glutathione transferases in plant secondary metabolism. Phytochemistry. 2010;71(4):338–50. doi:10.1016/j.phytochem.2009.12.012.
11. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nature Methods. 2009;6(5):359–62. doi:10.1038/nmeth.1322.
12. Tolin S, Arrigoni G, Trentin AR, Veljovic-Jovanovic S, Pivato M, Zechman B, Masi A. Biochemical and quantitative proteomics investigations in Arabidopsis ggt1 mutant leaves reveal a role for the gamma-glutamyl cycle in plant's adaptation to environment. Proteomics. 2013;13(12–13):2031–45. doi:10.1002/pmic.201200479.
13. Xiong L, Zhu JK. Molecular and genetic aspects of plant responses to osmotic stress. Plant Cell and Environment. 2002;25(2):131–39. doi:10.1046/j.1365-3040.2002.00782.x.
14. Yeh H, Lin L, Chen TH, Cheng CC, Chang, H. Y. TX, Lee TM. Monodehydroascorbate reductase plays a role in the tolerance of *chlamydomonas reinhardtii* to photooxidative stress. Plant and Cell Physiology. 2019;60(10):2167–79. doi:10.1093/pcp/pcz110.
15. Li F, Wu QY, Sun YL, Wang LY, Yang XH, Meng QW. Overexpression of chloroplastic monodehydroascorbate reductase enhanced tolerance to temperature and methyl viologen-mediated oxidative stresses. Physiologia Plantarum. 2010;139(4):421–34. doi:10.1111/j.1399-3054.2010.01369.x.
16. Camacho-Pereira J, Meyer LE, Machado LB, Oliveira MF, Galina A. Reactive oxygen species production by potato tuber mitochondria is modulated by mitochondrially bound hexokinase activity. Plant Physiology. 2009;149(2):1099–110. doi:10.1104/pp.108.129247.
17. Guo L, Devaiah SP, Narasimhan R, Pan X, Zhang Y, Zhang W, Wang X. Cytosolic glyceraldehyde-3-phosphate dehydrogenases interact with phospholipase Dδ to transduce hydrogen peroxide signals in the *Arabidopsis* response to stress. Plant Cell. 2012;24(5):2200–12. doi:10.1105/tpc.111.094946.
18. Vemanna RS, Babitha KC, Solanki JK, Amarnatha Reddy V, Sarangi SK, Udayakumar M. Aldo-keto reductase-1 (AKR1) protect cellular enzymes from salt stress by detoxifying reactive cytotoxic compounds. Plant Physiology and Biochemistry 2017;113:177–86. doi:10.1016/j.plaphy.2017.02.012.
19. Fatland BL, Nikolau BJ, Wurtele ES. Reverse genetic characterization of cytosolic acetyl-CoA generation by atp-citrate lyase in *Arabidopsis*. Plant Cell. 2005;17(1):182–203. doi:10.1105/tpc.104.026211.
20. Santos, C. VD, Rey P. Plant thioredoxins are key actors in the oxidative stress response. Trends in Plant Science. 2006;11(7):329–34. doi:10.1016/j.tplants.2006.05.005.
21. Schmidtmann E, König AC, Orwat A, Leister D, Hartl M, Finkemeier I. Redox regulation of *Arabidopsis* mitochondrial citrate synthase. Molecular Plant. 2014;7(1):156–69. doi:10.1093/mp/sst144.
22. Pulido P, Cazalis R, Cejudo FJ. An antioxidant redox system in the nucleus of wheat seed cells suffering oxidative stress. The Plant Journal. 2009;57(1):132–45. doi:10.1111/j.1365-313X.2008.03675.x.
23. Kashyap P, Deswal R. A novel class I Chitinase from *Hippophae rhamnoides*: indications for participating in ICE-CBF cold stress signaling pathway. Plant Science. 2017;259:62–70. doi:10.1016/j.plantsci.2017.03.004.
24. Zhao M, Ma Y, Pan YH, Zhang CH, Yuan WX. A hevein-like protein and a class I chitinase with antifungal activity from leaves of the paper mulberry. Biomedical Chromatography. 2011;25(8):908–12. doi:10.1002/bmc.1543.
25. Liu J, Sturrock R, Ekramoddoullah AK. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Reports. 2010;29(5):419–36. doi:10.1007/s00299-010-0826-8.
26. Parkhi V, Kumar V, Sunilkumar G, Campbell LM, Singh NK, Rathore KS. Expression of apoplastically secreted tobacco *osmotin* in cotton confers drought tolerance. Molecular Breeding. 2009;23(4):625–39. doi:10.1007/s11032-009-9261-3.

27. Shi H, Lium W, Yao Y, Wei Y, Chan Z. Alcohol dehydrogenase 1 (ADH1) confers both abiotic and biotic stress resistance in *Arabidopsis*. Plant Science. 2017;262:24–31. doi:10.1016/j.plantsci.2017.05.013.
28. Herrmann KM. The shikimate pathway as an entry to aromatic secondary metabolism. Plant Physiology. 1995;107(1):7–12. doi:10.1104/pp.107.1.7.
29. Diaz-Ramos A, Roig-Borrellas A, Garcia-Melero A, Lopez-Alemany R. α-Enolase, a multifunctional protein: its role on pathophysiological situations. Journal of Biomedicine and Biotechnology. 2012;2012:1–12. doi:10.1155/2012/156795.
30. Tzin V, Malitsky S, Ben Zvi MM, Bedair M, Sumner L, Aharoni A, Galili G. Expression of a bacterial feedback-insensitive 3-deoxy-Darabino-heptulosonate 7-phosphate synthase of the shikimate pathway in *Arabidopsis* elucidates potential metabolic bottlenecks between primary and secondary metabolism. New Phytologist. 2012;194(2):430–39. doi:10.1111/j.1469-8137.2012.04052.x.
31. Zhang Y, Butelli E, Alseekh S, Tohge T, Rallapalli G, Luo J, Martin C, Hill L, Santino A, Fernie AR. Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. Nature Communications. 2015;6(1):1–11. doi:10.1038/ncomms9635.
32. Winkel-Shirley B. Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiology. 2001;126(2):485–93. doi:10.1104/pp.126.2.485.
33. Dixon RA, Xie DY, Sharma SB. Proanthocyanidins – a final frontier in flavonoid research? New Phytologist. 2005;165(1):9–28. doi:10.1111/j.1469-8137.2004.01217.x.
34. Suzuki H, Takahashi S, Watanabe R, Fukushima Y, Fujita N, Noguchi A,Y, Nakayama T, Nishitani K, Nishino T, Nakayama T. An isoflavone conjugate-hydrolyzing β-Glucosidase from the roots of soybean (*Glycine max*) seedlings: purification, gene cloning, phylogenetics and cellular localization. Journal of Biological Chemistry. 2006;281(40):30251–59. doi:10.1074/jbc.M605726200.
35. Dixon DP, Edwards R. Roles for stress-inducible lambda glutathione transferases in flavonoid metabolism in plants as identified by ligand fishing. Journal of Biological Chemistry. 2010;285(47):36322–29. doi:10.1074/jbc.M110.164806.
36. Ikegami A, Akagi T, Potter D, Yamada M, Sato A, Yonemori K, Inoue K. Molecular identification of 1-Cysperoxiredoxin and anthocyanidin/flavonol 3-O-galactosyltransferase from proanthocyanidin-rich young fruits of persimmon (Diospyros kaki Thunb.). Planta. 2009;230(4):841–55. doi:10.1007/s00425-009-0989-0.
37. Lydeard JR, Schulman BA, Harper JW. Building and remodelling Cullin-RING E3 ubiquitin ligases. EMBO Reports. 2013;14(12):1050–61. doi:10.1038/embor.2013.173.
38. An JP, Liu X, Li HH, You CX, Wang X,F, Hao YJ. Apple RING E3 ligase MdMIEL1 inhibits anthocyanin accumulation by ubiquitinating and degrading MdMYB1 protein. Plant and Cell Physiology. 2017;58(11):1953–62. doi:10.1093/pcp/pcx129.
39. Fatimababy AS, Lin YL, Usharani R, Radjacommare R, Wang HT, Tsai HL, Fu H. Cross-species divergence of the major recognition pathways of ubiquitylated substrates for ubiquitin/26S proteasome-mediated proteolysis. FEBS Journal. 2010;277(3):796–816. doi:10.1111/j.1742-4658.2009.07531.x.
40. Lipford JR, Smith GT, Chi Y, Deshaies RJ. A putative stimulatory role for activator turnover in gene expression. Nature. 2005;438(7064):113–16. doi:10.1038/nature04098.
41. Jaakola L. New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Science. 2013;18(9):477–83. doi:10.1016/j.tplants.2013.06.003.