

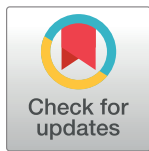
RESEARCH ARTICLE

iTRAQ protein profile analysis of developmental dynamics in soybean [*Glycine max* (L.) Merr.] leaves

Jun Qin^{1,2}, Jianan Zhang³, Fengmin Wang¹, Jinghua Wang¹, Zhi Zheng³, Changcheng Yin⁴, Hao Chen⁴, Ainong Shi², Bo Zhang⁵, Pengyin Chen², Mengchen Zhang^{1*}

1 National Soybean Improvement Center Shijiazhuang Sub-Center, North China Key Laboratory of Biology and Genetic Improvement of Soybean Ministry of Agriculture, Cereal & Oil Crop Institute, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, P.R. China, **2** Department of Crop, Soil and Environmental Sciences, University of Arkansas, Fayetteville, AR, United States of America, **3** National Foxtail Millet Improvement Center, Minor Cereal Crops Laboratory of Hebei Province Institute of Millet Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, P. R. China, **4** Beijing Protein Innovation, B-8, Beijing Airport Industrial Zone, Beijing, People's Republic of China, **5** Department of Crop, Soil and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States of America

* mengchenzhang@hotmail.com



OPEN ACCESS

Citation: Qin J, Zhang J, Wang F, Wang J, Zheng Z, Yin C, et al. (2017) iTRAQ protein profile analysis of developmental dynamics in soybean [*Glycine max* (L.) Merr.] leaves. PLoS ONE 12(9): e0181910. <https://doi.org/10.1371/journal.pone.0181910>

Editor: Hon-Ming Lam, Chinese University of Hong Kong, HONG KONG

Received: August 24, 2016

Accepted: July 10, 2017

Published: September 27, 2017

Copyright: © 2017 Qin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by the Natural Science Fund for Distinguished Young Scholars of Hebei Province (C2014301035), National Natural Science Foundation of China (31100880), the Key Project of the Natural Science Foundation of Hebei Province (C2012301020).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Zao5241 is an elite soybean [*Glycine max* (L.) Merr.] line and backbone parent. In this study, we employed iTRAQ to analyze the proteomes and protein expression profiles of Zao5241 during leaf development. We identified 1,245 proteins in all experiments, of which only 45 had been previously annotated. Among overlapping proteins between three biological replicates, 598 proteins with 2 unique peptides identified were reliably quantified. The protein datasets were classified into 36 GO functional terms, and the photosynthesis term was most significantly enriched. A total of 113 proteins were defined as being differentially expressed during leaf development; 41 proteins were found to be differently expressed between two and four week old leaves, and 84 proteins were found to be differently expressed between two and six week old leaves, respectively. Cluster analysis of the data revealed dynamic proteomes. Proteins annotated as electron carrier activity were greatly enriched in the peak expression profiles, and photosynthesis proteins were negatively modulated along the whole time course. This dataset will serve as the foundation for a systems biology approach to understanding photosynthetic development.

Introduction

Soybean [*Glycine max* (L.) Merr.] is an important oil and grain crop and is also the world's most important source of edible vegetable oil and vegetable protein[1]. Soybean seeds are rich in protein, fat, and other nutritional compounds including isoflavones and oligosaccharides [2]. With the increasing gulf between supply and demand, fully tapping yield potential and producing elite, high yielding, high quality soybean cultivars is essential for production[3].

Photosynthetic efficiency directly affects soybean production, so understanding the dynamic development of the leaf is of singular importance.

As the leaf is so important to the basic functions of plant growth including photosynthesis, transpiration, and respiration, it has been the focus of widespread and continuing research. The cell-specific transcriptomes of successive developmental stages were compared in bundle sheath (BS) and mesophyll (M) cells of maize leaves [4], the results showed that the number of genes preferentially transcribed in one or another cell type varies greatly between the different stages of leaf development. The transcriptome of maize leaves were analyzed using Illumina sequencing, the results quantified transcript abundance along a leaf developmental gradient in mature BS and M cells [5]. Differential gene expression analysis was performed in soybean leaf tissues at late developmental stages under drought stress showing that the down regulation of many photosynthesis-related genes can contribute to retardation of growth under drought stress which may serve as an adaptive mechanism for plant survival [6]. However, the research relevant to the dynamic development of soybean leaves is still lacking.

Studying proteins to reveal gene function and the nature of biological phenomena is important. However, biological proteins are complex, and each cell generally has thousands of proteins. Therefore, a technique capable of separating large number of proteins simultaneously is required. Proteomics is a recent developing technology which could be useful for the large-scale, comprehensive study of protein structure, modification, action and expression in cells and tissue [7]. Comparative proteomics is the identification of differences and changes in the proteome between states, tissue types, and environments. On one level, this provides the tools and methods for studying the function of life and physiological and pathological phenomena, but these same tools can also be applied to examine the basic laws of life [8–10]. There are many methods for proteomics research. Two-dimensional electrophoresis (2-DE) is the main method for separating proteins due to its simplicity and is widely used in various plants proteomics. However, 2-DE results are not sufficiently accurate with some proteins being easily lost and others hard to detect in low abundance [11]. Compared with 2-DE, isotope affinity tag (ICAT) technology significantly improves the ease of separating membrane proteins, can be combined with high performance liquid chromatography, and provides protein isolation results which are more accurate and reliable. However, ICAT technology applies only to proteins containing cysteine residues and identifying small molecule peptides is difficult [12]. iTRAQ (Isobaric Tags for Relative and Absolute Quantitation) is a high-throughput method which can be used to study the relative and absolute quantification of 2–8 samples at the same time with good accuracy and repeatability. iTRAQ is one of the most widely used markers in comparison proteomics [13]. The research technology of iTRAQ quantitative protein has high sensitivity and can identify any type of protein. iTRAQ does not affect protein physical and chemical properties such as the isoelectric point, abundance of the protein itself, molecular weight, and hydrophobicity. At the same time, iTRAQ can detect alkaline, hydrophobic, high or low molecular weight, and low abundance proteins which are difficult to detect using 2-DE. In protein quantitative experiments using iTRAQ, more than 97% of the peptides can be marked efficiently and completely with high repeatability. In addition, through direct tandem application with the mass spectrum, iTRAQ can be automated saving time and effort with a fast reaction rate. All in all, iTRAQ is a good technique to study the plant protein quantity.

In our previously study, we compared leaf proteome specifically the photosynthetic protein expression patterns, between Hobbit and JD17 at same developmental stage [14]. In this research, we compared the protein different expression profiles in three important growth periods of Zao 5241, like V2 (second Trifoliate), R1 (beginning flowering) and R3 (beginning pod). Moreover, we analyzed enriched expressed proteins and revealed dynamic proteomes by cluster analysis. The results reported here represent a necessary step toward providing a useful

tool and information for isolating proteins and better understanding the molecular basis underlying developmental dynamics in soybean.

Materials and methods

Plant growth conditions and tissue collection

Zao 5241 is one of the most important elite germplasm lines for soybean breeding in the Huang-Huai-Hai region incorporating foreign descent such as Yanli from Japan and Williams from the United States. It is a versatile line with excellent plant type traits, early maturity, high oil content (23.77%), good resistance to four different strains of Soybean Mosaic Virus (SMV), and combining ability. It has been used to develop new soybean varieties and germplasm for twenty years in summer soybean breeding.

The plants were grown in a 50:50 mix of vermiculite and soil in a greenhouse with 12hr/12hr light/dark, 30°C light/22°C dark, and 60% relative humidity in Shijiazhuang (114°26'E, 38°03'N). Three growth stages, V2 (second trifoliate), R1 (beginning flowering) and R3 (beginning pod) were chosen as time points to analyze the protein expression profiles in order to analyze and detect different protein expression profiles associated with leave, flower and pod development. Leaves were collected at 2, 4, and 6-weeks 2 hours into the light period. Three biological replicates were carried out, and each sample was pooled from 5 plants.

Protein extraction

Total protein samples were extracted using the cold-acetone method described by Meyer et al. [15]. Leaves were ground in liquid nitrogen and suspended in precooled acetone (-20°C) containing 10% (v/v) TCA. Then, samples were incubated at -20°C for 2 h after thorough mixing. Proteins were collected by centrifuging at 30,000 g and 4°C for 30 min, and the supernatant was carefully removed. To reduce acidity, the protein pellets were re-suspended with precooled acetone and centrifuged again at 20,000 g and 4°C for 30 min. They were then washed 3 times with cold acetone and re-suspended in 1 ml protein extraction reagent [8 M urea, 4% (w/v) CHAPS, 30 mM HEPES, 1 mM PMSE, 2 mM EDTA and 10 mM DTT], the protein was resolubilized through sonication in an ultrasonic cleaner (SCIENTZ, SB4200D). The supernatant was collected after centrifuging at 20,000 g at 4°C for 30 min. Protein concentration was determined using a 2-D Quant Kit (General Electric Company, USA). SDS-PAGE was used to verify the protein quality and concentration.

Protein digestion and iTRAQ labeling

Total protein (100 µL) was diluted by equal volume tetraethylammonium bicarbonate (TEAB, pH 8.5; Sigma, St. Louis, MO). Subsequently, modified trypsin (Promega, Madison, WI) was added to the mixture (3.3 µg trypsin/100 µg protein) to digest the protein at 37°C for 24 hour. The solvent was then removed in a refrigerated centrifuge (Eppendorf, 5417R), and MALDI TOF/TOF (Bruker limited, Coventry, UK) was used to verify the efficiency of protein digestion.

Peptides were labeled with an iTRAQ labeling kit (Applied Biosystems) according to the manufacturer's manual. 2, 4, and 6-week-old samples were labeled with reagent 114–119, respectively. Nine Peptide samples labeled were divided into two iTRAQ groups, each iTRAQ groups were carried out independently (S1 Table). Three independent biological experiments were performed. The peptides were then combined and further fractionated offline using strong cation exchange (SCX) with high performance liquid chromatography (HPLC) system (Shimadzu, Japan) connected to a SCX column (Luna 5 µ column, 4.6 mm I.D. × 250 mm,

5 μm , 100 \AA ; Phenomenex, Torrance, CA). The retained peptides were eluted using Buffer A (10 mM KH_2PO_4 in 25% ACN, pH 3.0) and Buffer B (2 M KCl, 10 mM KH_2PO_4 in 25% ACN, pH 3.0) with a flow rate of 1 mL/min. In total, 38 fractions were collected and combined into 17 fractions to reduce peptide complexity. Subsequently, eluted fractions were lyophilized in a centrifugal speed vacuum concentrator and dissolved with 0.1% formic acid prior to reverse-phase nanoflow liquid chromatography (nLC) and tandem mass spectrometry (nLC-MS/MS).

MS/MS analysis

MS/MS analysis was performed on an Easy Nano-LC system (Proxeon Biosystems, Odense, Denmark) connected to a hybrid quadrupole/time-of-flight mass spectrometer (MicroTOF-Q II, Bruker, Germany) equipped with a nano electrospray ion source. Peptides of each fraction were equalized before being injected into the Nano-LC-system. The peptides were separated on a C18 analytical reverse phase column of 75 μm i.d.x100mm length with Solution A (5% acetonitrile, v/v, 0.1% formic acid, v/v), and Solution B (95% acetonitrile, v/v, 0.1% formic acid, v/v) at a flow rate of 300 nL/min. After equilibrating with 5% Solution B for 10 min, the following gradient schedule would start: 45% Solution B at 80 min, 80% Solution B at 85 min, maintained for 15 min, 5% at 105 min and held for 15 min before ramping back down to the initial solvent condition. The MicroTOF-Q II hybrid MS was used to analyze the fractions. The MS/MS survey scan for all experiments between 50 and 2000 mass-to-charge ratio (m/z). The ionization tip voltage and interface temperature was run as 1250 V and 150°C.

Protein identification and quantification

All MS data were collected using MicroTOFcontrol and analyzed using DataAnalysis (Bruker Daltonics). Mascot v2.3.01 (Matrix Science) [16] was used to identify proteins. The Universal Protein Resource (UniProt) [17] database was used as a reference. For biological repeats, spectra from 17 fractions were combined into one file and searched. The parameters were set as follows: specifying trypsin as the digestion enzyme; cysteine carbamidomethylation as fixed modification; iTRAQ8-Plex on N-terminal residue, iTRAQ-8Plex on tyrosine (Y), iTRAQ-8Plex on lysine (K), glutamine-pyroglutamic acid and oxidation on methionine (M) as the variable modification; peptide tolerance was set at 0.05 Da, and MS/MS tolerance was set at 0.05Da. Finally, an additional filter before exporting the data was set: significance threshold $P < 0.05$ (with 95% confidence) and ion score or expected cut-off less than 0.05 (with 95% confidence).

The protein search results were exported by Mascot then normalized and quantified using Scaffold v3.0 [18]. 4- and 6-week-old protein profiles were quantified based on the 2-week-old results.

Protein annotation

Based on the results of the protein search against the UniProt database, DAVID Bioinformatics Resources v6.7 [19] was employed to obtain the functional classification of the proteins based on Gene Ontology (GO) [20] and Clusters of Orthologous Groups (COG) terms [21] and determine the distribution of the gene functions at the macro level. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>) [22, 23] was used to annotate the pathway of these proteins. The enrichment test of functional category proteins was performed by a chi-square test with the defining cutoff as 0.01 under an *Arabidopsis thaliana* background. A false discovery rate (FDR) significance threshold of 0.05 was used as false-positive control.

Differentially expressed protein definition and cluster analysis

Proteins with significant changes during leaf development were selected using the method described before [24, 25]. 90% confidence (Z -score = 1.645) of \log^2 ratios was used to select the proteins whose distribution was removed from the main distribution. The mean and SD from those proteins overlapping in three biological repeats was calculated. The profiles from different time points were calculated independently, and a broad threshold was selected. The result showed the cut-off value for up-regulated proteins was 1.75-fold (mean + Z -score \times SD) and down-regulated proteins was 0.69-fold (mean - Z -score \times SD). Protein ratios outside this range were defined as being significantly different at $P = 0.01$.

The expression profiles of the differentially expressed proteins were determined by cluster analysis based on the k-means method with Pearson's correlation distance in Genesis v1.7.6 [26, 27]. The number of clusters was determined using the Figures of Merit (FOM) algorithm [28].

Results

Identification of leaves proteins

A total of 1,245 proteins were identified in all experiments with the number of unique peptides ranging from 0 to 40. Of all proteins, only 405 (32.53%) had been well annotated previously under a *Glycine max* background. These were mostly composed of metabolic pathway, biosynthesis of secondary metabolites pathway, and cytoplasm component proteins. Remarkably, a total of 49 *Glycine* annotated proteins related to photosynthesis were found among the abundant proteins. All of the proteins were also compared against the UniprotKB/Swiss-Prot *Arabidopsis thaliana* protein database using blastx with an E-value cut-off of $1e-5$ revealing 1,212 proteins. For the three biological replicates, 1,002, 982, and 970 proteins (80.48% 78.88%, and 77.91% of total proteins) were identified with an overlap of 727 (72.55%, 74.03%, and 74.95%) proteins. Although the lack of protein information limited identification, this investigation demonstrated the capability and applicability of the iTRAQ technology to quantify the leaf proteome of soybean.

Functional classification and enrichment analysis of proteins

For the iTRAQ dataset, 598 proteins, with at least 2 unique peptides, were used for further analysis. The iTRAQ identified proteins were classified into 36 GO functional terms which are distributed into three main categories according to their functional role: biological process (434), cellular components (460), and molecular function (439) (Fig 1).

There were 17 GO terms in the biological process category. Four out of those GO terms were significantly enriched ($P < 0.01$): metabolic process (366, $P = 4.61E-20$), response to stimulus (191, $P = 1.99E-17$), cellular process (352, $P = 1.78E-13$), and cellular component biogenesis (35, $P = 2.11E-04$). Within the cellular component category, there were 9 GO terms, six of which were significantly enriched ($P < 0.01$): organelle part (270, $P = 4.87E-100$), organelle (400, $P = 1.68E-54$), envelope (101, $P = 2.05E-41$), macromolecular complex (157, $P = 1.16E-33$), membrane-enclosed lumen (77, $P = 9.09E-27$), and extracellular region (76, $P = 3.71E-11$). Four out of the ten GO Terms were enriched in the molecular functional category: structural molecule activity (67, $P = 2.17E-26$), antioxidant activity (19, $P = 3.22E-08$), electron carrier activity (40, $P = 1.56E-06$), and catalytic activity (267, $P = 1.59E-05$). A total of 58 proteins out of 598 were annotated with photosynthesis (GO: 0015979), a child term of the biological process category, which was significantly enriched in this dataset ($P = 1.70E-44$). 31 proteins were classified as electron transport chain (GO: 0022900) which is the father category of photosynthetic electron transport chain (GO: 0009767) and respiratory electron transport chain (GO: 0022904).

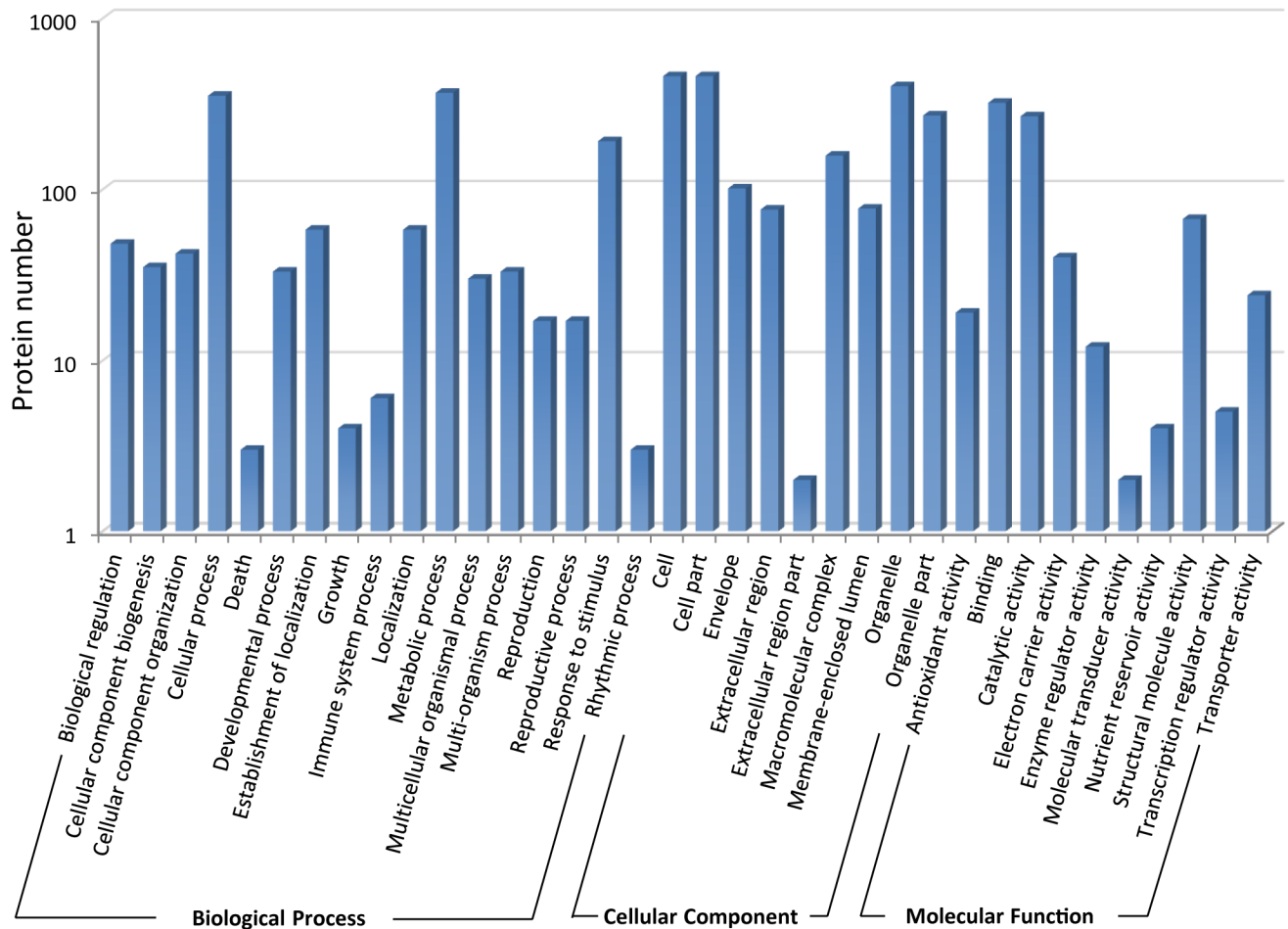


Fig 1. Functional classification of the iTRAQ identified proteins of soybean leaves.

<https://doi.org/10.1371/journal.pone.0181910.g001>

Based on the KEGG database, 394 proteins were annotated with 85 pathways. Among these pathways, metabolism (gmx01100) and biosynthesis of secondary metabolites (gmx01110) involved the most proteins (163 and 85 proteins, respectively); however, photosynthesis (gmx00195) was the most enriched pathway ($P = 2.68E-15$) followed by carbon fixation in photosynthetic organisms (gmx00710) ($P = 8.70E-10$) and photosynthesis—antenna proteins (gmx00195) ($P = 1.04E-5$). In addition, another pathway related to photosynthesis, citrate cycle (TCA cycle; gmx00020), was also enriched ($P = 1.73E-3$) in the iTRAQ protein dataset (S2 Table).

Differentially expressed proteins and enrichment analysis

For 598 proteins, at least 2 unique peptides were identified which permitted quantification [24]. Within the dataset, 544, 578, and 596 proteins were identified in the two, four, and six-week stages, respectively (Fig 2).

Based on a 90% confidence level, cutoff values of 1.75-fold for up-regulated proteins and 0.69-fold for down-regulated proteins were used to define the differential expression during soybean development. A total of 113 proteins were defined as differentially expressed proteins (S3 Table). 41 proteins were found to be differentially expressed between four-week and two-week old leaves; of these, 21 were decreased and 20 increased during leaf development. 84

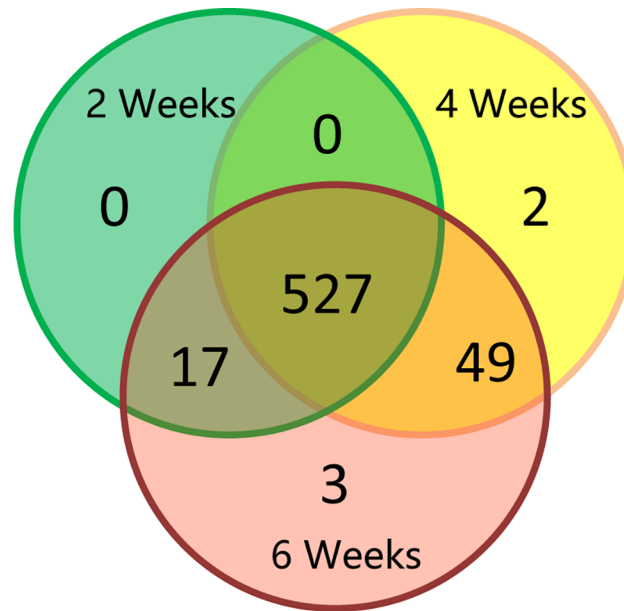


Fig 2. Venn diagram showing the proteins expressed in three stages of soybean development.

<https://doi.org/10.1371/journal.pone.0181910.g002>

proteins were defined as being differentially expressed between six- and two-week old leaves; 60 of which were decreased and 24 increased during leaf development. The proteins which were differentially expressed at either four-/two- or six-/two- week old leaves were separated into categories (Fig 3). Two-week-old results serve as a reference, developmental process (GO:0032502), Organelle (GO:0043226), and organelle part (GO:0044422) protein were both enriched in four- ($P = 8.39E-01$, $P = 2.35E-03$, and $P = 2.57E-06$) and six-week old leaves ($P = 9.70E-01$, $P = 4.23E-11$, and $P = 7.44E-20$). Response to stimulus (GO: 0050896), catalytic activity (GO: 0003824) and electron carrier activity (GO: 0009055) proteins were individually enriched ($P = 9.35E-3$, $P = 7.99E-3$, and $P = 7.83E-5$) in four -week old leaves. Cellular component biogenesis (GO:0044085), metabolic process (GO:0008152), envelope (GO:0031975), extracellular region (GO:0005576), macromolecular complex (GO:0032991), membrane-enclosed lumen (GO:0031974), and structural molecule activity (GO:0005198) were individually enriched ($P = 6.61E-02$, $P = 2.50E-05$, $P = 1.52E-10$, $P = 1.57E-03$, $P = 2.75E-10$, $P = 1.82E-03$, and $P = 1.00E-15$) in six -week old leaves.

For KEGG analysis, those proteins were significantly enriched in Photosynthesis pathway ($P = 4.18E-4$ for diff-expressed proteins of four/two weeks old leaves, $P = 5.45E-6$ for diff-expressed proteins of six/two weeks old leaves).

Cluster analysis of differentially expressed proteins and validation with immunoassays

Clusters of differentially expressed proteins were analyzed based on the k-means methods using Pearson's correlation distance in Genesis v1.76. The proteins were divided into seven clusters based on their expression modulation representing the number of profiles indicated by FOM analysis (Fig 4A). Cluster 1 and 2 contained proteins positively or negatively modulated along the whole time course; cluster 3 and cluster 4 contained proteins modulated only during the four to six weeks developmental stage; cluster 5 and cluster 6 contained proteins positively and negatively modulated only at four weeks old leaves, and proteins up-regulated

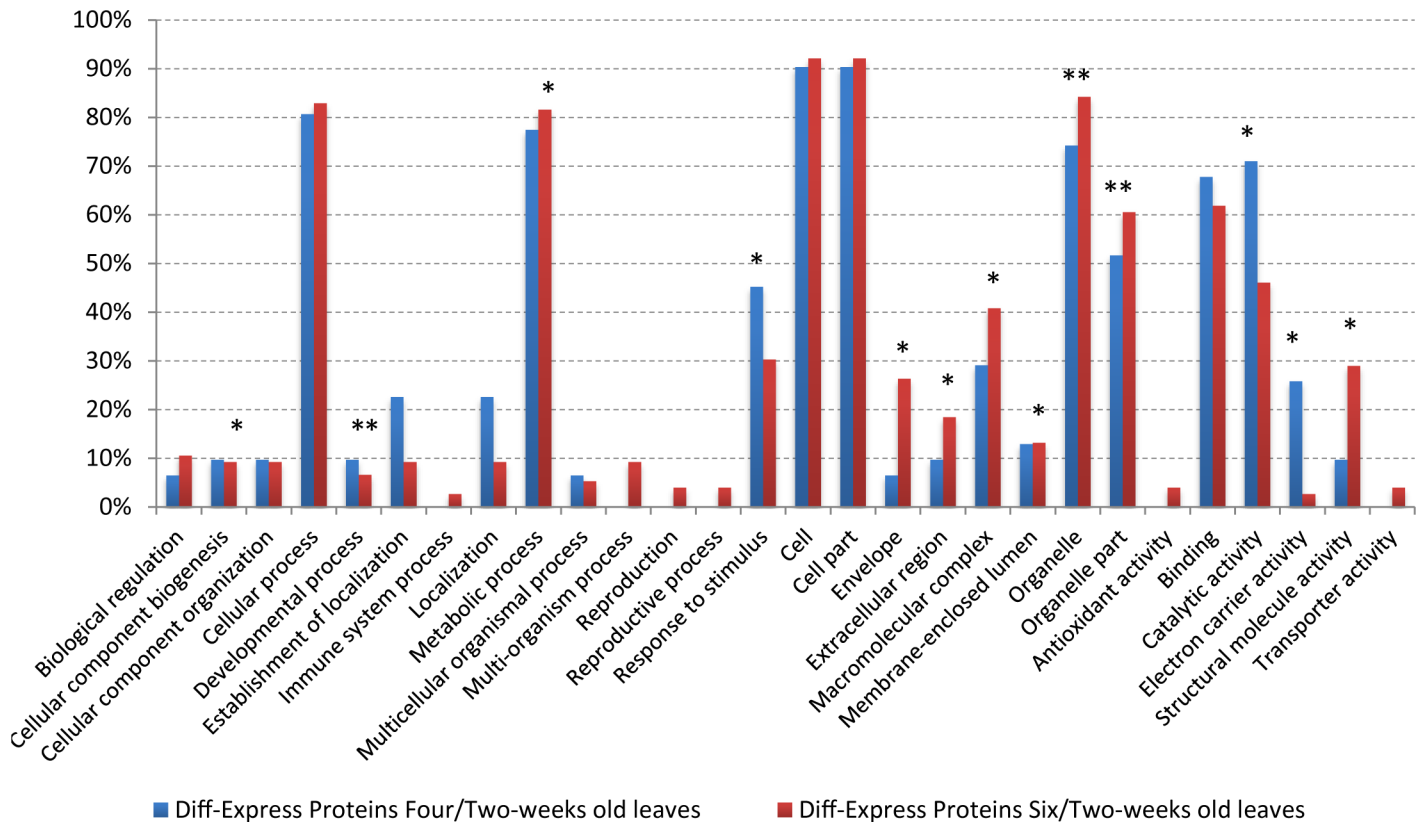


Fig 3. The functional classification of differentially expressed proteins in four/two and six/two weeks old leaves. * The diff-expressed proteins were enriched in this functional category.

<https://doi.org/10.1371/journal.pone.0181910.g003>

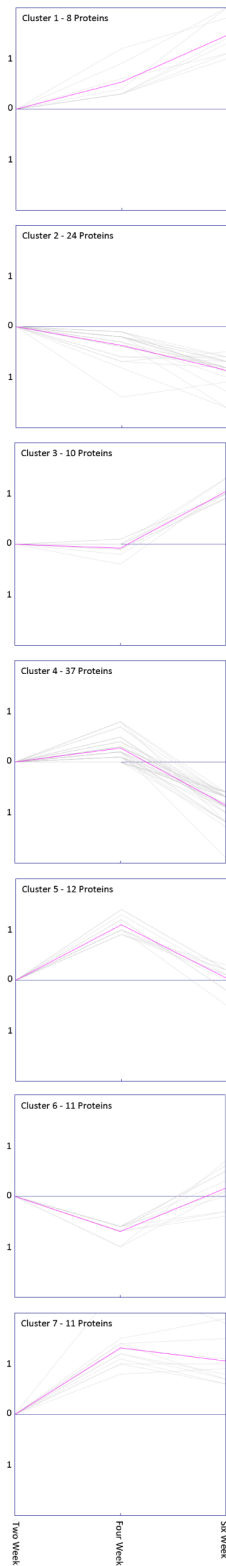
during two to four weeks stage fell into cluster 7 (Fig 4B). Functional categories enrichment were showed in Fig 4C. For example, proteins classified into catalytic activity were greatly enriched in cluster 3, with the proteins being upregulated during the second stage. Differentially-expressed proteins annotated as electron carrier activity were greatly enriched in cluster 6 with peak expression of proteins in the four week old leaves. For each cluster, the enrichment of KEGG analysis based on the DAVID database (Fig 4C) were showed Photosynthesis pathway was significantly enriched in cluster 2 and cluster 5.

For further validation, the expression patterns of four proteins, CAB1, OEE1, LOX, and SBP were identified by western blotting and the results were consistent with the iTRAQ results, this verified the reliability of the iTRAQ system and the data analysis methods (Fig 5).

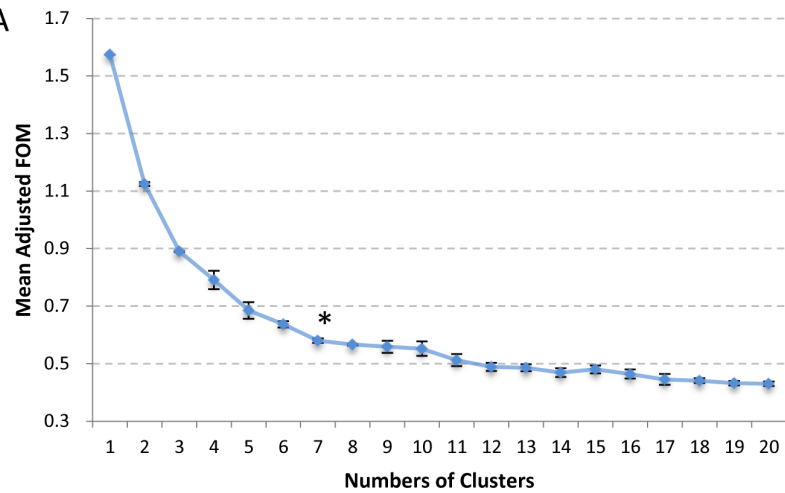
Discussion

iTRAQ is a technique with high sensitivity and accuracy, demonstrating a remarkable advantage in simultaneous analysis of multiple samples and subsequently providing the relative quantification on hundreds of proteins at one time[29]. In the present study, an iTRAQ-based proteomic technique was used to assess proteomic changes and identify proteins which were differentially expressed during soybean leaf development. In total, 1,245 proteins were identified. Data from three biological replicates were analyzed and proteins detected by querying data with a soybean protein database. Although soybean protein annotation is limited, a total of 58 proteins out of 598 were annotated with photosynthesis, and 31 proteins were classified as electron transport chain. Pathway analysis showed that photosynthesis was the most

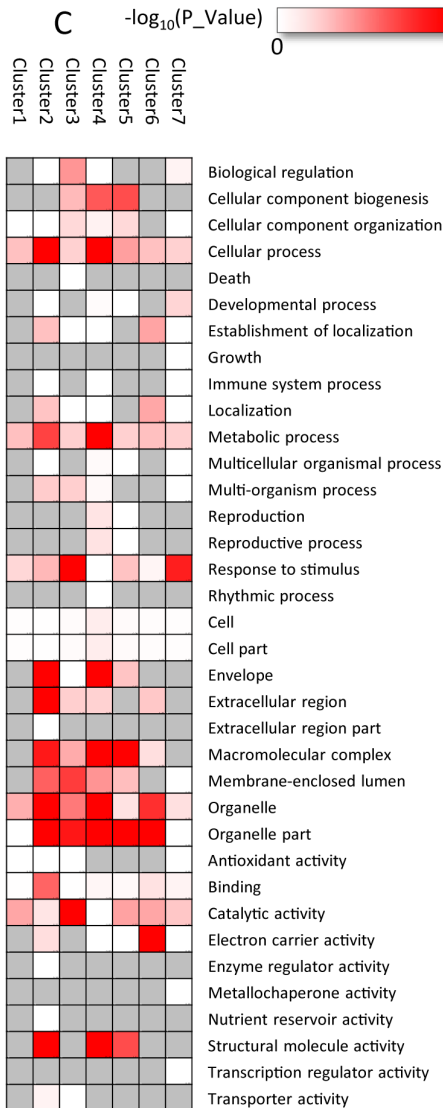
B



A



C



D

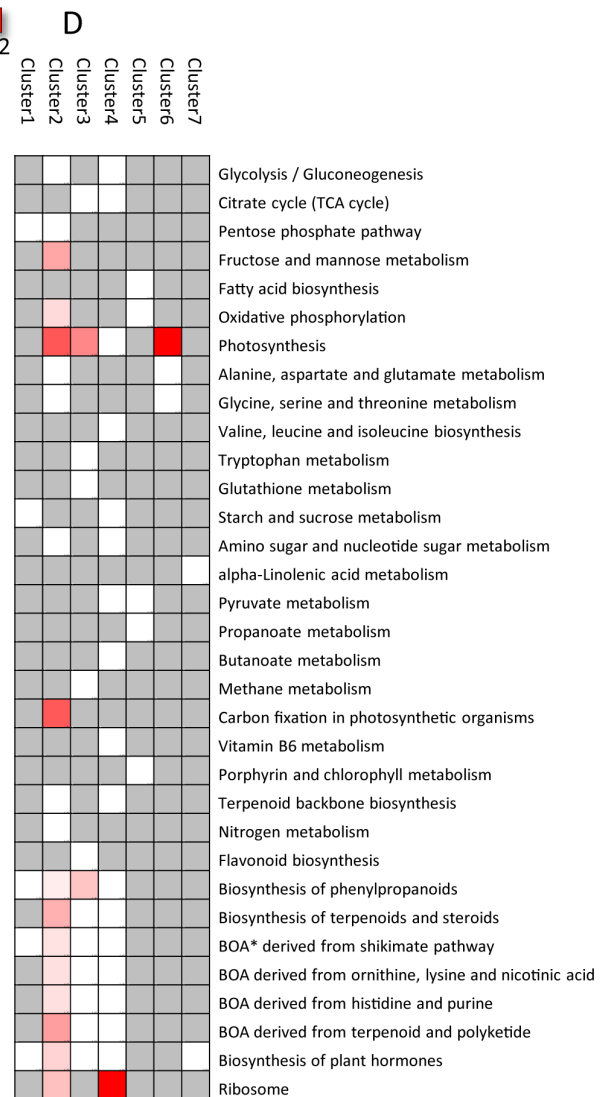


Fig 4. Categories distribution in the seven expression clusters. A) Figure of merit analysis indicated the number of cluster; B) Clusters were obtained by *K*-means method on the gene expression profiles from 113 diff-expressed proteins. C) Enrichment of KEGG analysis among the seven clusters. * BOA, Biosynthesis of Alkaloids. Red, significant enrichment; white, non-significant; gray, not-detected.

<https://doi.org/10.1371/journal.pone.0181910.g004>

enriched pathway followed by carbon fixation in photosynthetic organisms and photosynthesis. Differential proteomic analysis of grapevine leaves revealed responses to heat stress and subsequent recovery by iTRAQ. One hundred and seventy-four proteins were differentially expressed under heat stress and/or during the recovery phase, in comparison to unstressed

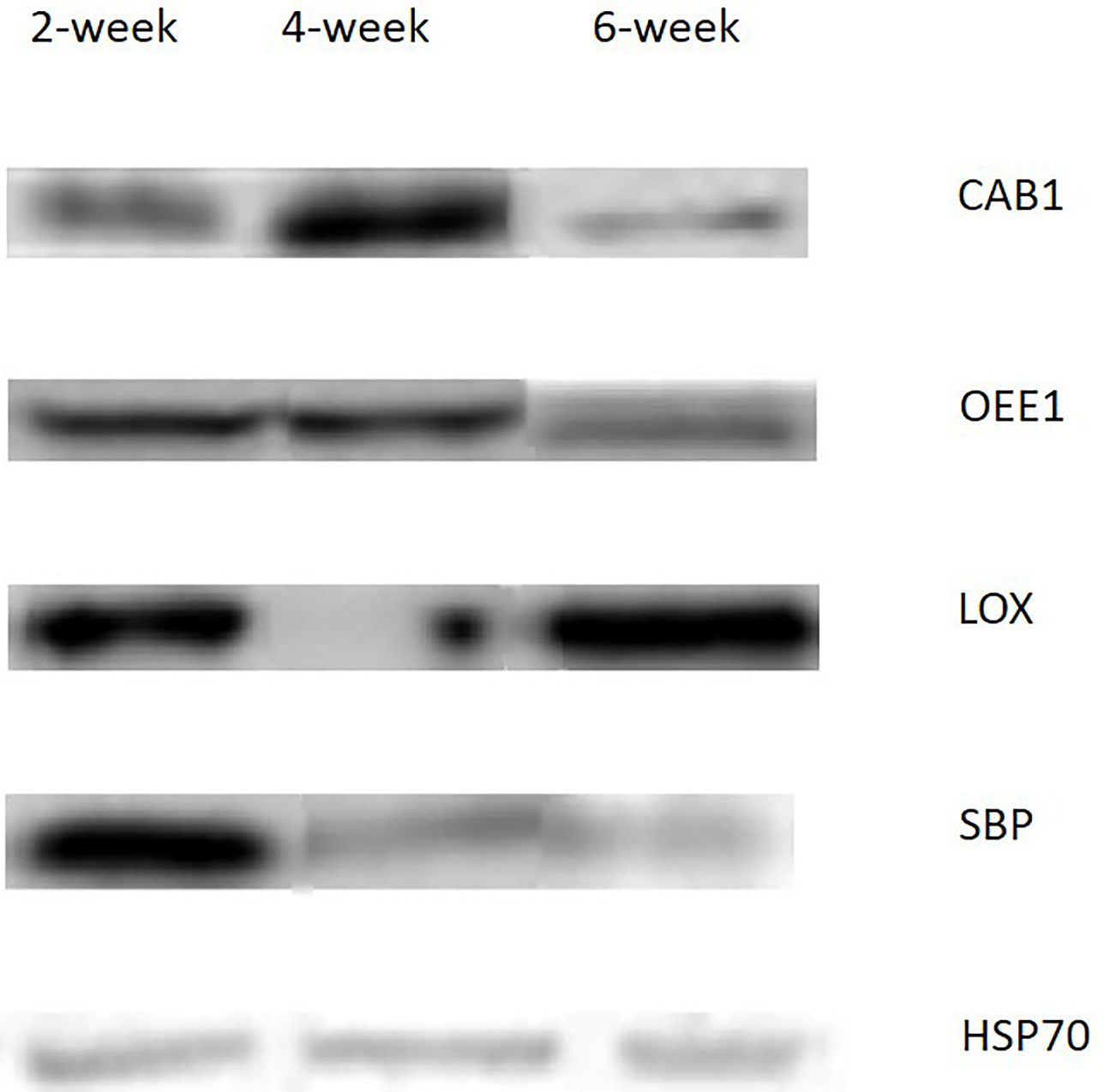


Fig 5. Western blotting detection of soybean proteins CAB1 OEE1 LOX SBP HSP70.

<https://doi.org/10.1371/journal.pone.0181910.g005>

controls. Based on MapMan ontology, functional categories for these dysregulated proteins included mainly photosynthesis (about 20%), proteins (13%), and stress (8%). Some proteins related to electron transport chain of photosynthesis, antioxidant enzymes, HSPs and other stress response proteins, and glycolysis may play key roles in enhancing grapevine adaptation to and recovery capacity from heat stress. In our preliminary research, proteomic analysis of two soybean cultivars were performed using iTRAQ-based quantitative. Enrichment factor analysis indicated that proteins involved in photosynthesis comprised an important category [14].

To estimate how many and what proteins expressed during the leave developing, the 2-week Protein Profile of Zao5241 was used to produce the reference, 41 and 84 of 1,245 differentially expressed proteins were found in 4 weeks and 6 weeks, respectively. On the basis of GO and pathway enrichment analysis, we concluded that the “photosynthesis was most significantly enriched. KEGG and enrichment analysis of differentially expressed proteins showed that electron carrier activity proteins were significantly enriched in the photosynthesis pathway ($P = 4.18E-4$ for differentially expressed proteins of four-/two- weeks old leaves, $P = 5.45E-6$ for diff-expressed proteins of six/two weeks old leaves). The proteomes of soybean leaves and roots under salt treatment were investigated using iTRAQ-based proteomic approach. Protein-protein interaction analysis revealed metabolism, carbohydrate and energy metabolism, protein synthesis and redox homeostasis could be assigned to four high salt stress response networks[30].

Moreover, dynamic proteomes were revealed by cluster analysis. Electron carrier activity proteins were greatly enriched in the peak expression profiles, and photosynthesis proteins were negatively modulated along the whole time course. Further studies for protein function analysis were needed to further clarify the molecular mechanism during leaf development in soybean.

Supporting information

S1 Table. The information of nine peptide samples labeled.

(XLSX)

S2 Table. Proteins annotated with three pathways related to photosynthesis.

(XLSX)

S3 Table. Differentially expressed proteins between four-week and two-week old leaves.

(XLSX)

Acknowledgments

This study was supported by the Natural [science Fund for Distinguished Young Scholars](#) of Hebei Province (C2014301035), National Natural Science Foundation of China (31100880), the Key Project of the Natural Science Foundation of Hebei Province (C2012301020).

Author Contributions

Data curation: Jun Qin, Jianan Zhang.

Formal analysis: Zhi Zheng.

Funding acquisition: Jun Qin, Mengchen Zhang.

Investigation: Bo Zhang.

Methodology: Jinghua Wang.

Project administration: Mengchen Zhang.

Software: Jianan Zhang, Ainong Shi.

Validation: Fengmin Wang, Changcheng Yin, Hao Chen.

Writing – review & editing: Jun Qin, Pengyin Chen.

References

1. Sinclair JB. (1982) Compendium of soybean diseases: American Phytopathological Society and University of Illinois.
2. Messina MJ. (1999) Legumes and soybeans: overview of their nutritional profiles and health effects. *The American journal of clinical nutrition*. 70(3):439s–450s.
3. Tester M, Langridge P. (2010) Breeding technologies to increase crop production in a changing world. *Science*. 327(5967):818–822. <https://doi.org/10.1126/science.1183700> PMID: 20150489
4. Tausta SL, Li P, Si Y, Gandotra N, Liu P, Sun Q, et al. (2014) Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset of C4-related processes. *Journal of Experimental Botany*. 65(13):3543–5555. <https://doi.org/10.1093/jxb/eru152> PMID: 24790109
5. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, et al. (2010) The developmental dynamics of the maize leaf transcriptome. *Nature genetics*. 42(12):1060–1067. <https://doi.org/10.1038/ng.703> PMID: 21037569
6. Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, Yamaguchi-Shinozaki K, et al. (2012) Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS One*. 7(11):e49522. <https://doi.org/10.1371/journal.pone.0049522> PMID: 23189148
7. Fountoulakis M. (2004) Application of proteomics technologies in the investigation of the brain. *Mass spectrometry reviews*. 23(4):231–258. <https://doi.org/10.1002/mas.10075> PMID: 15133836
8. Wang P, Li Y, Han S, Li B. (2012) Identification of differential expressed proteins responding to low temperature stress in pistils of kernel apricot. *Journal of Food, Agriculture & Environment*. 10(3&4):550–554.
9. Chen S, Harmon AC. (2006) Advances in plant proteomics. *Proteomics*. 6(20):5504–5516. <https://doi.org/10.1002/pmic.200600143> PMID: 16972296
10. Liang Y, Jing Y, Shen S. (2004) Advances in plant proteomics. *Acta Phytoecological Sinica*. 28(1):114–125.
11. Tacken MG, Rottier PJ, Gielkens AL, Peeters BP. (2000) Interactions in vivo between the proteins of infectious bursal disease virus: capsid protein VP3 interacts with the RNA-dependent RNA polymerase, VP1. *Journal of General Virology*. 81(1):209–218.
12. Peng Z, He S, Gong W, Sun J, Pan Z, Xu F, et al. (2014) Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics*. 15(1):760.
13. Wu WW, Wang G, Baek SJ, Shen RF. (2006) Comparative study of three proteomic quantitative methods, DIGE, cI-CAT, and iTRAQ, using 2D gel-or LC-MALDI TOF/TOF. *Journal of Proteome Research*. 5(3):651–658. <https://doi.org/10.1021/pr050405o> PMID: 16512681
14. Qin J, Zhang J, Liu D, Yin C, Wang F, Chen P, et al. (2016) iTRAQ-based analysis of developmental dynamics in the soybean leaf proteome reveals pathways associated with leaf photosynthetic rate. *Molecular Genetics and Genomics*. 291(4):1595–1605. <https://doi.org/10.1007/s00438-016-1202-3> PMID: 27048574
15. Meyer Y, Grosset J, Chartier Y, Cleyet-Marel JC. (1988) Preparation by two-dimensional electrophoresis of proteins for antibody production: antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. *Electrophoresis*. 9(11):704–712. <https://doi.org/10.1002/elps.1150091105> PMID: 3250873
16. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Research*. 33(Database issue):154–159.
17. Searle BC. (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics*. 10(6):1265–1269. <https://doi.org/10.1002/pmic.200900437> PMID: 20077414
18. Huang D, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.

- Nucleic Acids Research. 35(Web Server issue):W169–W175. doi: [10.1093/nar/gkm415](https://doi.org/10.1093/nar/gkm415). PMID: [17576678](https://pubmed.ncbi.nlm.nih.gov/17576678/); PubMed Central PMCID: PMC1933169.
19. Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research. 32(Database issue):D258–D261. <https://doi.org/10.1093/nar/gkh036> PMID: [14681407](https://pubmed.ncbi.nlm.nih.gov/14681407/)
 20. Natale DA, Shankavaram UT, Galperin MY, Wolf YI, Aravind L, Koonin EV. (2000) Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). Genome Biology. 1(5):RESEARCH0009.1–0009.19.
 21. Kanehisa M, Goto S. (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research. 28(1):27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
 22. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 27(1):29–34. PMID: [9847135](https://pubmed.ncbi.nlm.nih.gov/9847135/)
 23. Cox J, Mann M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 6(12):1367–1372. Epub 2008/11/26. doi: <https://doi.org/10.1038/nbt.1511> PMID: [19029910](https://pubmed.ncbi.nlm.nih.gov/19029910/).
 24. Lan P, Li W, Wen TN, Shiao JY, Wu YC, Lin W, et al. (2011) iTRAQ protein profile analysis of Arabidopsis roots reveals new aspects critical for iron homeostasis. Plant Physiology. 155(2):821–834. <https://doi.org/10.1104/pp.110.169508> PMID: [21173025](https://pubmed.ncbi.nlm.nih.gov/21173025/)
 25. Sturn A, Quackenbush J, Trajanoski Z. (2002) Genesis: cluster analysis of microarray data. Bioinformatics. 18(1):207–208. PMID: [11836235](https://pubmed.ncbi.nlm.nih.gov/11836235/)
 26. Zenoni S, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G, et al. (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. Plant Physiol. 152(4):1787–1795. <https://doi.org/10.1104/pp.109.149716> PMID: [20118272](https://pubmed.ncbi.nlm.nih.gov/20118272/)
 27. Yeung KY, Haynor DR, Ruzzo WL. (2001) Validating clustering for gene expression data. Bioinformatics. 17(4):309–318. PMID: [11301299](https://pubmed.ncbi.nlm.nih.gov/11301299/)
 28. Xie X, Wang X, Liu L, Dong S, Pi X, Liu W. (2011) iTRAQ technology and its application in proteomics. Chinese Journal of Biochemistry & Molecular Biol. 27(7):616–621.
 29. Liu GT, Ma L, Duan W, Wang BC, Li JH, Xu HG, et al. (2014) Differential proteomic analysis of grapevine leaves by iTRAQ reveals responses to heat stress and subsequent recovery. BMC plant biology. 14(1):110.
 30. Ji W, Cong R, Li S, Li R, Qin Z, Li Y, et al. (2016) Comparative Proteomic Analysis of Soybean Leaves and Roots by iTRAQ Provides Insights into Response Mechanisms to Short-Term Salt Stress. Frontiers in Plant Science. 7.