Accepted Manuscript

Integration of gel-based and gel-free proteomic data for functional analysis of proteins through Soybean Proteome Database

Setsuko Komatsu, Xin Wang, Xiaojian Yin, Yohei Nanjo, Hajime Ohyanagi, Katsumi Sakata

PII: S1874-3919(17)30167-7
DOI: doi: [10.1016/j.jprot.2017.05.009](https://doi.org/10.1016/j.jprot.2017.05.009)
Reference: JPROT 2848

To appear in: *Journal of Proteomics*

Received date: 27 February 2017
Revised date: 20 April 2017
Accepted date: 8 May 2017

Please cite this article as: Setsuko Komatsu, Xin Wang, Xiaojian Yin, Yohei Nanjo, Hajime Ohyanagi, Katsumi Sakata , Integration of gel-based and gel-free proteomic data for functional analysis of proteins through Soybean Proteome Database, *Journal of Proteomics* (2017), doi: [10.1016/j.jprot.2017.05.009](https://doi.org/10.1016/j.jprot.2017.05.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Integration of gel-based and gel-free proteomic data for functional analysis of proteins through Soybean Proteome Database

Setsuko Komatsu^{a,b,*}, Xin Wang^{a,b}, Xiaojian Yin^{a,b}, Yohei Nanjo^a, Hajime Ohyanagi^c,
Katsumi Sakata^{d,*}

^a Institute of Crop Science, National Agriculture and Food Research Organization,
Tsukuba 305-8518, Japan

^b Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba 305-8572,
Japan

^c King Abdullah University of Science and Technology (KAUST), Computational
Bioscience Research Center (CBRC), Thuwal 23955-6900, Kingdom of Saudi Arabia

^d Department of Life Science and Informatics, Maebashi Institute of Technology,
Maebashi 371-0816, Japan

*To whom correspondence should be addressed:

Setsuko Komatsu, Faculty of Life and Environmental Sciences, University of Tsukuba,
Tennodai 1-1-1, Tsukuba 305-8572, Japan. Tel: 81-29-853-5900. E-mail address:
komatsu.setsuko.fu@u.tsukuba.ac.jp

Katsumi Sakata, Department of Life Science and Informatics, Maebashi Institute of
Technology, Maebashi 371-0816, Japan. Tel: 81-27-265-7388. E-mail address:
ksakata@maebashi-it.ac.jp

Running title: Soybean Proteome Database

Abbreviations: LC, liquid chromatography; MS, mass spectrometry.

ABSTRACT

The Soybean Proteome Database (SPD) stores data on soybean proteins obtained with gel-based and gel-free proteomic techniques. The database was constructed to provide information on proteins for functional analyses. The majority of the data is focused on soybean (*Glycine max* ‘Enrei’). The growth and yield of soybean are strongly affected by environmental stresses such as flooding. The database was originally constructed using data on soybean proteins separated by two-dimensional polyacrylamide gel electrophoresis, which is a gel-based proteomic technique. Since 2015, the database has been expanded to incorporate data obtained by label-free mass spectrometry-based quantitative proteomics, which is a gel-free proteomic technique. Here, the portions of the database consisting of gel-free proteomic data are described. The gel-free proteomic database contains 39,212 proteins identified in 63 sample sets, such as temporal and organ-specific samples of soybean plants grown under flooding stress or non-stressed conditions. In addition, data on organellar proteins identified in mitochondria, nuclei, and endoplasmic reticulum are stored. Furthermore, the database integrates multiple omics data such as genomics, transcriptomics, metabolomics, and proteomics. The SPD database is accessible at <http://proteome.dc.affrc.go.jp/Soybean/>.

Keywords: Soybean, Database, Temporal-specific protein profile, Organ-specific protein profile, Proteomics, Abiotic stress

1. Introduction

Soybean is an important food crop that is rich in vegetable oil and protein [1], as well as phytochemicals such as isoflavones and phenolic compounds [2]. Soybean has a complex genome because multiple neopolyploid speciation events are superimposed on the paleopolyploid genome [3]. Given the agro-economic importance of soybean, the genome of the cultivar ‘Williams 82’ was sequenced using a whole-genome shotgun approach and assembled with physical and high-density genetic maps [4]. Most of the genome sequence was assembled into 20 chromosomes and 46,430 protein-coding genes were predicted [4]. The genome sequence of the cultivar ‘Enrei’ was also elucidated to provide a reference for characterization of Japanese domestic soybean cultivars [5]. High-resolution genetic linkage maps challenged the false joins or misplaced scaffolds and unanchored scaffolds to improve the soybean genome sequence [6]. The genomic information for soybean has indicated the complexity of the soybean genome, facilitated genetic/genomic research, and contributed to diverse biological studies of soybean.

Genomic databases such as SoyBase [7], SoyDB [8], and SoyTEdb [9] provide resources for genetic sequences, transcription factors, and translational information. With the availability of genomic information, a number of soybean databases containing transcriptomic, proteomic, functional network, and integrated data have been developed. SoyFN provides comprehensive information on genes and miRNAs, which are connected on the notion of functional similarities [10]. SoyKB is an integrated database for bridging translational genomics and molecular breeding research [11]. The proteomic databases SoyProDB [12] and SoyProLow [13] were constructed based on gel-based proteomic analysis of the soybean seed. Collectively, these databases broaden knowledge of the soybean genome and provide useful resources for breeding.

To provide information on proteins for functional analyses, the Soybean Proteome

Database (SPD) was originally constructed using data for soybean proteins separated by two-dimensional polyacrylamide gel electrophoresis, which is classified as a gel-based proteomic technique [14]. In addition, the data were integrated with the information from transcriptomics, proteomics, and metabolomics [15]. The database has been improved by incorporation of data from label-free mass spectrometry (MS)-based quantitative proteomics, which is classified as a gel-free proteomic technique. This technique is directly applied to peptide samples after digestion of proteins for liquid chromatography (LC)-MS. In this report, the portions of the SPD derived from gel-free proteomics are described. Although soybean is a widely cultivated crop, its growth and yield are markedly affected by adverse environmental conditions, especially flooding stress [16]. The SPD also stores information on the flooding-responsive proteins, which were analyzed in temporal and organ-specific protein profiles of soybean grown in the presence or absence of flooding stress in the present study.

2. Materials and methods

2.1. Plant material and treatments

Seeds of soybean (*Glycine max* L.) 'Enrei' were sterilized with 3% sodium hypochlorite solution and rinsed in water. After sowing the seeds in 450 mL of silica sand moistened with 100 mL of water in a seedling case (145 mm x 55 mm x 95 mm), the seeds were incubated in a growth chamber illuminated with white fluorescent light ($160 \mu\text{mol m}^{-2} \text{s}^{-1}$, 16 h light period/day) at 25°C up to 12 days. Two-day-old soybeans were exposed to flooding stress for 6 days by adding excess water to the silica sand. To represent a recovery stage after flooding stress, 2-day-old soybeans were flooded for 2 or 4 days and excess water was removed for 6 days. The cotyledon and embryo were collected from the seed; and the root, hypocotyl, cotyledon, and leaf were daily sampled for proteomic analysis. In

addition, the root of 2-day-old soybean was collected for isolation of organelles such as mitochondria, nuclei, and endoplasmic reticulum. For all proteomic analyses, three independent experiments were performed as biological replicates. For biological replicates, the seeds for each biological replicate were sown on different days.

2.2. Protein extraction and enrichment

A portion (500 mg) of each sample was ground to powder in liquid nitrogen using a mortar and pestle. The powder was added to a solution of 10% trichloroacetic acid and 0.07% 2-mercaptoethanol in acetone, and the resulting suspension was mixed thoroughly by vortexing. The mixture was sonicated for 10 min and incubated for 1 h at -20°C. After centrifugation at 9,000 x *g* for 20 min at 4°C, the supernatant was removed and the remaining pellet was washed twice with 0.07% 2-mercaptoethanol in acetone. The pellet was dried using a Speed-Vac concentrator (Savant Instruments, Hicksville, NY, USA) and resuspended in lysis buffer containing 8 M urea, 2 M thiourea, 5% CHAPS, and 2 mM tributylphosphine by vigorous vortexing for 1 h. The suspension was centrifuged at 20,000 x *g* for 20 min and the resulting supernatant was collected as the protein extract [17].

Extracted proteins (100 µg) in lysis buffer were adjusted to a final volume of 100 µL. Methanol (400 µL) was added to each sample and mixed before addition of 100 µL of chloroform and 300 µL of water. After mixing and centrifugation at 20,000 x *g* for 10 min to achieve phase separation, the upper phase was discarded and 300 µL of methanol was added to the lower phase, and then centrifuged at 20,000 x *g* for 10 min. The pellet was collected as the soluble fraction [17].

2.3. Isolation of organelles

2.3.1. Mitochondria

A portion (8 g) of the samples was ground in 12 mL of grinding buffer containing 0.45 M mannitol, 50 mM sodium pyrophosphate, 2 mM EGTA, 0.5% PVP-40, 0.5% BSA, and 20 mM cysteine (pH 8.0) with a mortar and pestle. The homogenate was filtered through four layers of Miracloth (Calbiochem, Darmstadt, Germany) via a funnel into a beaker. The filtered homogenate was centrifuged at $1,500 \times g$ for 5 min. The supernatant was gently removed, without disturbing the pellet and other cell debris, and transferred to a set of ultracentrifuge tubes. The supernatant (8 mL) was layered over a Percoll gradient prepared in an ultracentrifuge tube; from top to bottom, the gradient consisted of 7 mL of 40% Percoll, 10 mL of 23% Percoll, and 7 mL of 18% Percoll in mannitol wash buffer containing 0.3 M mannitol, 10 mM MOPS (pH 7.5), and 0.1% BSA. The tube was centrifuged at $34,000 \times g$ for 1 h at 4°C with the brake turned off during deceleration. Mitochondria were visible as a series of dense bands at the 23-40% Percoll interface. The mitochondrial layer was collected using a Pasteur pipette, and transferred to a new ultracentrifuge tube. The samples were diluted with sucrose wash buffer containing 0.3 M sucrose and 10 mM MOPS (pH 7.5), and then centrifuged at $24,000 \times g$ for 30 min at 4°C twice. The supernatant was discarded. Nine times the sample volume of cold acetone was added to a new tube and incubated at -20°C overnight. The precipitated sample was centrifuged at $20,000 \times g$ for 15 min at 4°C . The supernatant was discarded and the pellet was dried for 15 min at room temperature. The pellet was collected as the mitochondrial fraction [18,19].

2.3.2. Endoplasmic reticulum

A portion (5 g) of the samples was ground in grinding buffer containing HEPES (pH 7.5), sucrose, and KCl with a mortar and pestle. Endoplasmic reticulum was isolated with

the Endoplasmic Reticulum Enrichment Kit (Imgenex, San Diego, CA, USA) following the manufacturer's instructions. The homogenate was centrifuged at $3,000 \times g$ for 10 min at 4°C . The supernatant was centrifuged at $12,000 \times g$ for 15 min at 4°C , and the resulting supernatant was centrifuged at $90,000 \times g$ for 1 h at 4°C . The pellet was collected as the total endoplasmic reticulum-enriched fraction. To purify the rough endoplasmic reticulum-enriched fraction, the supernatant was transferred to a beaker and CaCl_2 solution of 15 times the volume of the supernatant was added dropwise while stirring the supernatant for 15 min. The supernatant with CaCl_2 solution was transferred from the beaker to a centrifuge tube and centrifuged at $8,000 \times g$ for 10 min at 4°C . The pellet was collected as rough endoplasmic reticulum-enriched fraction [20,21].

2.3.3. Nuclei

A portion (1.5 g) of the samples was ground in 5 mL Nuclei Isolation Buffer (Sigma, St. Louis, MO, USA) with a mortar and pestle. Nuclei were isolated with the Plant Nuclei Isolation/Extraction Kit (Sigma) in accordance with the manufacturer's instructions with some modifications. The homogenate was filtered through a double layer of Filter Mesh 100 and transferred to a 15 mL tube, and then centrifuged at $1,300 \times g$ for 10 min at 4°C . The resulting pellet was resuspended in Nuclei Isolation Buffer containing protease inhibitor mixture (Roche, Werk Penzberg, Germany) and phosphatase inhibitor mixture (Sigma), and then layered on top of a 1.5 M sucrose cushion prepared in Nuclei Isolation Buffer. After centrifugation at $12,000 \times g$ for 10 min at 4°C , the resulting supernatant was removed and the pellet was resuspended in Nuclei Isolation Buffer containing protease inhibitor mixture. The suspension was again layered on a 1.5 M sucrose cushion and centrifuged at $12,000 \times g$ for 10 min at 4°C to pellet the nuclei-enriched fraction. The purified nuclei were washed twice using Nuclei Isolation Buffer containing protease

inhibitor mixture. The pellet was collected as the nuclei fraction [22,23].

2.4. Protein reduction, alkylation, and digestion for proteomic analysis

Proteins were resuspended in 50 mM NH_4HCO_3 , reduced with 50 mM dithiothreitol for 30 min at 56°C, and alkylated with 50 mM iodoacetamide for 30 min at 37°C in the dark. Alkylated proteins were digested with trypsin and lysyl endopeptidase (Wako, Osaka, Japan) at a 1:100 enzyme/protein ratio for 16 h at 37°C. Peptides were acidified with formic acid ($\text{pH} < 3$) and analyzed by nanoLC-MS/MS.

2.5. Measurement of protein and peptide concentrations

The method of Bradford [24] was used to determine the protein concentration with bovine serum albumin used as the standard. A Direct Detect Spectrometer (Millipore, Billerica, MA, USA) equipped with the Direct Detect software (version 3.0.25.0) was used to determine the peptide concentration.

2.6. Protein identification using nanoLC-MS/MS

The peptides were loaded onto the LC system (Ultimate 3000 nanoLC; Dionex, Germering, Germany) equipped with a trap column (C18 PepMap, 300 μm ID x 5 mm; Dionex) equilibrated with 0.1% formic acid and eluted with a linear acetonitrile gradient in 0.1% formic acid at a flow rate of 200 nL/min. The eluted peptides were loaded and separated on the column (C18 capillary tip column, 75 μm ID x 120 mm; Nikkyo Technos, Tokyo, Japan) with a spray voltage of 1.5 kV. The peptide ions were detected using MS (nanospray LTQ Orbitrap Discovery MS; Thermo Fisher Scientific, San Jose, CA, USA) in the data-dependent acquisition mode with the installed Xcalibur software (version 2.1; Thermo Fisher Scientific). Full-scan mass spectra were acquired in the MS

over 400-1,500 m/z with resolution of 30,000. A lock mass function was used to obtain high mass accuracy [25]. The ten most intense precursor ions were selected for collision-induced fragmentation in the linear ion trap at a normalized collision energy of 35%. Dynamic exclusion was employed within 90 sec to prevent repetitive selection of peptides [26].

2.7. MS data analysis

Identification of proteins was performed using the Mascot search engine (version 2.5.1; Matrix Science, London, UK) with a soybean peptide database (55,787 sequences) constructed from the soybean genome database of Phytozome (version 9.1, <http://www.phytozome.net/soybean>) [4]. Proteome Discoverer software (version 1.4.0.288; Thermo Fisher Scientific) was used to process the acquired raw data files. For the Mascot search, the carbamidomethylation of cysteine was set as a fixed modification and oxidation of methionine was set as a variable modification. Trypsin was specified as the proteolytic enzyme and one missed cleavage was allowed. Peptide mass tolerance was set at 10 ppm, fragment mass tolerance was set at 0.8 Da, and peptide charges were set at +2, +3, and +4. An automatic decoy database search was performed as part of the search. Mascot results were filtered with the Percolator function to improve the accuracy and sensitivity of peptide identification.

2.8. Calculation of protein abundance using MS data

Abundance of the identified proteins was analyzed on the basis of exponentially modified protein abundance index (emPAI) values. Briefly, the emPAI value of each identified protein was divided by the sum of the emPAI values for all identified proteins and multiplied by 100. Protein content was estimated as the molar fraction percentage

(mol%) [27].

2.9. Differential analysis of proteins using MS data

For differential analysis of the relative abundances of peptides and proteins between samples, the commercial label-free quantification package SIEVE (version 2.1.377; Thermo Fisher Scientific) was used. The chromatographic peaks detected by MS were aligned, and the peptide peaks were detected as a frame on all parent ions scanned by MS/MS using 5 min of frame time width and 10 ppm of frame m/z width. Chromatographic peak areas within a frame were compared for each sample, and the ratios between samples were determined for each frame. The frames detected in the MS/MS scan were matched to the imported Mascot results. The peptide ratios between samples were determined from the variance-weighted average of the ratios in frames that matched the peptides in the MS/MS spectrum. The ratios of peptides were further integrated to determine the ratios of the corresponding proteins. In the differential analysis of protein abundance, total ion current was used for normalization. The minimum requirement for identification of a protein was two matched peptides. Significant changes in the abundance of proteins between samples were determined ($p < 0.05$).

2.10. Functional categorization

Functional categorization of identified proteins was performed using MapMan bin codes (<http://mapman.gabipd.org/>) [28].

2.11. Cluster and in silico protein–protein interaction analyses using protein abundance

Protein abundance ratios at different time points of flooding stress were subjected

to hierarchical clustering analysis (a centroid linkage method based on a Euclidean distance metric). Interacting proteins were identified by comparisons of changes in abundance ratios over time. Clustered proteins were analyzed for *in silico* protein-protein interactions that were estimated by temporal expression profiling utilizing an *S*-system differential equation [29] as a mathematical model. Each interaction between proteins was tested based on a goodness-of-fit, which indicated how well the *S*-system differential equation simulated expression of the corresponding target protein. The interactions that showed a coefficient of determination (r^2) > 0.98 were considered as candidate interactions. In the model protein interaction diagram, a red arrow indicates an inductive interaction and corresponds to $g_{ij} > 0$ in the *S*-system differential equation, and a blue T-bar indicates a suppressive interaction and corresponds to $g_{ij} < 0$ in the *S*-system differential equation [30].

3. Results and discussion

3.1. Construction and usage of the SPD database

The SPD was originally constructed using data for soybean proteins separated by two-dimensional polyacrylamide gel electrophoresis, which is a gel-based proteomic technique [14,15]. Since 2015, data from label-free MS based quantitative proteomics, which is a gel-free proteomic technique, have been incorporated in the SPD. The status of the web server environment is as follows: i) the operating system is CentOS 6.6 (32 bit); ii) the web daemon is Apache 2.2.15; iii) the database system is PostgreSQL 8.4.20; and iv) the programming language is Perl 5.10.1. The data are downloadable for the portion of the database that comprises protein data collected by a gel-free proteomic technique. The Make2D-DB II package (version 2.50.1) was used to develop the portion of the database consisting of protein data collected by a gel-based proteomic technique, and

additional source scripts were developed by the authors.

In this study, the portions of the database obtained from a gel-free proteomic technique are described. The gel-free proteomic data comprises 39,212 proteins identified in 63 sample sets, such as temporal and organ-specific samples of soybean grown under flooding stress or the non-stressed condition. In addition, data on organellar proteins identified in mitochondria, nuclei, and endoplasmic reticulum are stored. Specifically, the following samples were used as gel-free proteomic data: i) cotyledon and embryo in seed, ii) leaf, cotyledon, hypocotyl, and root in seedling at 1, 2, 3, 4, 5, 6, 8, 10, and 12 days; iii) cotyledon and root in seedling under flooding stress for 0, 1, 2, 3, and 4 days; iv) cotyledon, hypocotyl, and root in seedling which is 2 days flooding and water removal (recovery stage); and v) cotyledon, hypocotyl, and root in seedling which is 4 days flooding and water removal (recovery stage). Furthermore, proteins from mitochondria, nucleolus, and endoplasmic reticulum were used as gel-free proteomic data. On the SPD website homepage (<http://proteome.dc.affrc.go.jp/Soybean/>), there are options to access “Gel-based Proteomics” and “Gel-free Proteomics” data. On selection of “Gel-free Proteomics”, images of soybean plants (Fig. 1A) and the list of organs (Fig. 1B) together with the time series and organelles from which the data were obtained are shown. The “Enter search keyword” box enables the user to search for protein information (Fig. 1C) in conjunction with the associated soybean or *Arabidopsis* accession numbers or description of proteins/genes. AC Soybean and AC *Arabidopsis* mean Protein ID in genome of soybean as “Glyma20g28550” and *Arabidopsis* as “AT3G22490” according to the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). Furthermore, “Download all data in spread sheet style” is used to download all data.

When the image of a specific organ of soybean plant at a specific time point is selected, a list of proteins is shown. The data tables and search results are compatible

with Microsoft Excel. The result file can be downloaded onto a personal computer, and the file dragged/dropped into a spreadsheet. In detail, the explanation on “How to use search result” is in the web site as follows: (1) Search data, (2) Save “Result” as a text file, and then (3) Open the file “Result cgi.txt” with Microsoft Excel (Fig.2). The data table includes the soybean accession number, *Arabidopsis* accession number, description of proteins, protein concentration (mol%), emPAI, score, protein mass, number of matched peptides, growth stage of the plant, the name of the organ, and the site on the genome map. When a protein is selected, information on the chromosomal position from DAIZUbase (<http://daizu.dna.affrc.go.jp/>) is shown (Fig. 3).

Proteomic data for the cotyledon of seed were used for improvement of ‘Enrei’ soybean genomic data [5]. Using gene models associated with the seed proteomic data, a total of 164 proteins corresponded to genes associated with storage proteins, lipid synthesis/degradation enzymes, sorting/folding-related proteins, late embryogenesis abundant proteins, glycolysis pathway enzymes, protease/protease inhibitors, and other proteins [5]. Based on this knowledge, the SPD is linked to the soybean genome database DAIZUbase, which was constructed with data for the cultivar ‘Enrei’ [5], which is same as this proteome database. It is indicated that this protein profile may provide the basis for understanding cultivar diversity and adaptation to environmental conditions.

The SPD constructed with gel-based proteomic data contains 23 reference maps collected from a number of organs, tissues, and organelles of the cultivar ‘Enrei’ [14]. The database also contains 3,399 and 2,019 proteins from seven organs and four subcellular compartments, respectively. Furthermore, the database integrates multiple omics data such as information from transcriptomics, metabolomics, and proteomics [15]. The present database has been enhanced with gel-free proteomic data consisting of 39,212 proteins identified in 63 sample sets. Voluntary deposition of proteomic

information in this database might be necessary for effective utilization of available knowledge for the management of abiotic stress tolerance in soybean.

3.2. Application for functional analyses

Although soybean is a widely cultivated crop, it is sensitive to various abiotic stresses, particularly waterlogging stress during the germination, early vegetative, and early reproductive stages of development [31,32]. Several proteomic studies have examined the changes that occur in the root system of soybean under flooding stress, because the root is highly sensitive to waterlogging stress [33]. The SPD also stores data on the flooding-responsive proteins, which were analyzed in temporal and organ-specific samples of soybean grown under flooding stress or non-stressed conditions. To confirm the utility of the data in the SPD, a portion of the data (Table 1) were analyzed by functional categorization (Fig. 4), cluster separation (Fig. 5), and *in silico* protein-protein interaction analysis (Fig. 6).

For these analyses, the following data were used. Two-day-old soybeans were treated with flooding stress or the non-stressed condition for 1, 2, 3, and 4 days. Proteins in the roots were extracted, reduced, alkylated, and digested. Digested peptides were directly analyzed by nanoLC-MS/MS, which is a gel-free proteomic technique. A total of 228 proteins were significantly changed in abundance in root between non-treated and flooded soybeans in a time-dependent manner (Table 1). To explore the biological processes in the soybean root exposed to flooding stress, the identified proteins were functionally classified using MapMan bin codes (Fig. 4). Strongly affected proteins were involved in protein synthesis/degradation (26.75%), cell organization (11.40%), glycolysis (6.58%), stress (5.70%), tricarboxylic acid cycle (5.70%), and amino acid metabolism (5.25%). Using a gel-based proteomic technique, Komatsu et al. [34]

reported that the abundance of 28 proteins associated with protein destination/storage, energy, and disease/defense was significantly changed in the root between non-treated and flooded soybeans in a time-dependent manner. Although gel-based quantification methods are still useful for proteomic analyses, gel-free proteomic methods have become mainstream in recent years, because a wide range of proteins, including extremely basic/acidic and low-abundance or hydrophobic proteins, are detected. In particular, it is suggested that gel-free proteomic techniques are superior when comparing many sampling points over a time series.

To understand the function of the significantly changed proteins, cluster analysis was performed using the same data in the SPD (Fig. 5). A total of 228 significantly changed proteins were grouped into four groups. Group I comprised proteins that were significantly increased in non-treated soybean and mildly increased under flooding stress. Group II consisted of proteins that were increased in non-treated soybean and decreased under flooding stress. Group III comprised proteins that were significantly decreased in non-treated soybean, but mildly decreased under flooding stress. Group IV consisted of four proteins that were decreased in non-treated soybean, but were significantly increased by flooding stress. The latter four proteins comprised two sucrose synthases (ID317, Glyma03g37441.1; ID205, Glyma19g40041.1), thiamine pyrophosphate-dependent pyruvate decarboxylase (ID329), and pyruvate decarboxylase 2 (ID334, Glyma18g43460.1). Hashiguchi et al. [35] subjected 51 differentially changed protein spots identified with a gel-based proteomic technique to cluster analysis and recognized six groups of proteins; however, all of the proteins analyzed were storage proteins, which are abundant proteins, such as late embryogenesis-abundant protein, maturation polypeptide, and albumin precursor. This knowledge and the present results indicate that a gel-free proteomic technique is also useful for identification of unique proteins for

understanding the stress-response mechanism.

Using the same data in the SPD, *in silico* protein-protein interaction analysis was performed (Fig. 6). In the non-treated soybean, more than two apparent interaction networks were recognized; however, a single large interaction network was generated for flooding-treated soybean. Although the number of significantly changed proteins in the present study was different compared with the results reported by Hashiguchi et al. [35], the tendency for integration of protein-protein interaction networks into a single large interaction network in flooding-treated soybean was identical. It is suggested that this type of interaction analysis is useful for screening important factors in plant response to environmental stresses. Importin alpha isoform 1 (ID197, Glyma10g29000.1) and importin alpha isoform 2 (ID129, Glyma09g04430.1) were positively affected from majority proteins identified in non-treated soybean and were positioned in the center of the protein-protein interaction network under flooding stress (Fig. 6). On the other hand, cytochrome P450 family 93 (ID346, Glyma07g32330.1) and MA3 domain-containing protein (ID368, Glyma05g02910.1) were negatively affected from majority proteins identified in non-treated soybean; however, these proteins were isolated from the protein network under flooding stress (Fig. 6).

Importin alpha and beta isoforms were positioned in the center of the interaction networks both in non-treated and flooded soybeans (Fig. 6). Nuclear import of proteins that contain classical nuclear localization signals is initiated by importin alpha, which recognizes and binds to the nuclear localization signals in the cytoplasm [36]. Importin abundance is decreased under flooding stress supplemented with abscisic acid treatment, which improved the survival ratio of soybean under flooding stress [11]. It is suggested that even under flooding, the function of protein nuclear transfer may be maintained. Cytochrome P450 and MA3 domain-containing protein are associated with abscisic acid

[37] and ethylene [38], respectively. It is suggested that hormonal signaling pathways are important for the functioning of these two proteins under normal, non-stress conditions; however, the proteins cannot function under flooding stress. Furthermore, these proteins are candidate proteins for improvement of flooding tolerance of soybean. Taken together, the above-mentioned analyses indicate that application of these data stored in the SPD will be useful for functional analyses.

4. Conclusion and future directions

The SPD, originally constructed with gel-based proteomic data in 2009 [14], was enhanced by incorporation of gel-free proteomic data in the present study. The gel-based proteomic technique provides a visual representation of the proteins including intact protein maps, in which protein isoforms with modifications resulting from molecular weight and isoelectric point changes can be clearly displayed [39]. The SPD stores data obtained from both gel-based and gel-free proteomic techniques. The gel-free proteomic database comprises 39,212 proteins identified in 63 sample sets, such as different organs of soybean plants grown under flooding stress or non-stressed conditions in a time-dependent manner. In addition, organellar proteins identified in mitochondria, nuclei, and endoplasmic reticulum are stored in the gel-free proteomics database. A total of 44,704 proteins, including 5,490 proteins identified using a gel-based proteomic technique, are stored in the SPD. It accounts for approximately 80% of all predicted proteins from genome sequences, though there are overlapped proteins. Based on the demonstrated application of data stored in the database for functional analyses, it is suggested that these data will be useful for analyses of biological mechanisms in soybean. Furthermore, coupled with recent advances in information and communication technology, the usefulness of this database would increase in the analyses of biological mechanisms.

Information on the proteins significantly altered under other stresses such as drought, salt, and nanoparticle stresses, will be stored in the database in future; furthermore, information on additional organellar proteins, such as those associated with the plasma membrane and chloroplasts, will be stored. Information on post-translational modifications such as phosphorylation, glycosylation, and other modifications obtained experimentally by gel-free proteomic analysis will be added to the SPD and the number of identified proteins will be increased.

Conflict of interest

All of the authors declare no conflict of interest.

Acknowledgements

We thank Mr. Shunsuke Fujieda of the Maebashi Institute of Technology, Japan for construction of the database. The authors also thank Dr. Kazuhiko Sugimoto of the Institute of Crop Science, NARO, Japan for support of the database. This work was supported by JSPS KAKENHI Grant Number JP15H04445. The Soybean Proteome Database (SPD) has been constructed using the “Virtual Lab System” in the Tsukuba Business-Academia Cooperation Support Center, Japan.

References

- [1] Sugiyama, A., Ueda, Y., Takase, H., Yazaki, K. Do soybeans select specific species of *Bradyrhizobium* during growth? *Commun. Integr. Biol.* **8**, e992734 (2015).
- [2] Lee, S. J., Kim, J. J., Moon, H. I., Ahn, J. K., Chun, S. C, Jung, W. S., Lee, O. K., Chung, I. M. Analysis of isoflavones and phenolic compounds in Korean soybean [*Glycine max* (L.) Merrill] seeds of different seed weights. *J. Agric. Food Chem.* **56**,

- 2751-2758 (2008).
- [3] Shoemaker, R. C., Schlueter, J., Doyle, J. J. Paleopolyploidy and gene duplication in soybean and other legumes. *Curr. Opin. Plant Biol.* **9**, 104-109 (2006).
- [4] 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, T. U., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abemathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X. C., Shinozaki, K., Nguyen, H. T., Wing, R. A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Shoemaker, R. C., Jackson, S. A. Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-183 (2010).
- [5] Shimomura, M., Kanamori, H., Komatsu, S., Namiki, N., Mukai, Y., Kurita, K., Kamatsuki, K., Ikawa, H., Yano, R., Ishimoto, M., Kaga, A., Katayose, Y. The *Glycine max* cv. Enrei Genome for Improvement of Japanese Soybean Cultivars. *Int. J. Genomics* **2015**, 358127 (2015).
- [6] Song, Q., Jenkins, J., Jia, G., Hyten, D. L., Pantalone, V., Jackson, S. A., Schmutz, J., Cregan, P. B. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genomics* **17**, 33 (2016).
- [7] Grant, D., Nelson, R. T., Cannon, S. B., Shoemaker, R. C. SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* **38**, D843-846 (2010).
- [8] Wang, Z., Libault, M., Joshi, T., Valliyodan, B., Nguyen, H. T., Xu, D., Stacey, G., Cheng, J. SoyDB: a knowledge database of soybean transcription factors. *BMC Plant Biol.* **10**, 14 (2010).
- [9] Du, J., Grant, D., Tian, Z., Nelson, R. T., Zhu, L., Shoemaker, R. C., Ma, J. SoyTEDb: a comprehensive database of transposable elements in the soybean genome. *BMC*

- Genomics* **11**, 113 (2010).
- [10] Xu, Y., Guo, M., Liu, X., Wang, C., Liu, Y. SoyFN: a knowledge database of soybean functional networks. *Database (Oxford)*, bau019 (2014).
- [11] Joshi, T., Fitzpatrick, M. R., Chen, S., Liu, Y., Zhang, H., Endacott, R. Z., Gaudiello, E. C., Stacey, G., Nguyen, H. T., Xu, D. Soybean knowledge base (SoyKB): a web resource for integration of soybean translational genomics and molecular breeding. *Nucleic Acids Res.* **42**, D1245-1252 (2014).
- [12] Tavakolan, M., Alkharouf, N. W., Khan, F. H., Natarajan, S. SoyProDB: A database for the identification of soybean seed proteins. *Bioinformation* **9**, 165-167 (2013).
- [13] Tavakolan, M., Alkharouf, N. W., Matthews, B. F., Natarajan, S. S. SoyProLow: A protein database enriched in low abundant soybean proteins. *Bioinformation* **10**, 599-601 (2014).
- [14] Sakata, K., Ohyanagi, H., Nobori, H., Nakamura, T., Hashiguchi, A., Nanjo, Y., Mikami, Y., Yunokawa, H., Komatsu, S. Soybean proteome database: a data resource for plant differential omics. *J. Proteome Res.* **8**, 3539-3548 (2009).
- [15] Ohyanagi, H., Sakata, K., Komatsu, S. Soybean Proteome Database 2012: update on the comprehensive data repository for soybean proteomics. *Front. Plant Sci.* **3**, 110 (2012).
- [16] Komatsu, S., Hiraga, S., Yanagawa, Y. Proteomics techniques for the development of flood tolerant crops. *J. Proteome Res.* **11**, 68-78 (2012).
- [17] Komatsu, S., Han, C., Nanjo, Y., Altaf-Un-Nahar, M., Wang, K., He, D., Yang, P. Label-free quantitative proteomic analysis of abscisic acid effect in early-stage soybean under flooding. *J. Proteome Res.* **12**, 4769-4784 (2013).
- [18] Huang, S., Jacoby, R. P., Millar, H. A., Taylor, N. L. Plant Proteomics: Methods and Protocols, 2nd Ed, pp 499-526, Springer, NY (2014).

- [19] Mustafa, G., Komatsu, S. Insight into the response of soybean mitochondrial proteins to various sizes of aluminum oxide nanoparticles under flooding stress. *J. Proteome Res.* **15**, 4464-4475 (2016).
- [20] Szarka, A., Stadler, K., Jenei, V., Margittai, E., Csala, M., Jakus, J., Mandl, J., Bańhegyi, G. Ascorbyl free radical and dehydroascorbate formation in rat liver endoplasmic reticulum. *J. Bioenerg. Biomembr.* **34**, 317-323 (2002).
- [21] Wang, X., Komatsu, S. Gel-free/label-free proteomic analysis of endoplasmic reticulum proteins in soybean root tips under flooding and drought stresses. *J. Proteome Res.* **15**, 2211-2227 (2016).
- [22] Yin, X., Komatsu, S. Quantitative proteomics of nuclear phosphoproteins in the root tip of soybean during the initial stages of flooding stress. *J. Proteomics* **119**, 183-195 (2015).
- [23] Yin, X., Komatsu, S. Nuclear proteomics reveals the role of protein synthesis and chromatin structure in root tip of soybean during the initial stage of flooding stress. *J. Proteome Res.* **15**, 2283-2298 (2016).
- [24] Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* **72**, 248-254 (1976).
- [25] Olsen, J. V., de Godoy, L. M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010-2021 (2005).
- [26] Zhang, Y., Wen, Z., Washburn, M. P., Florens, L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. *Anal. Chem.* **81**, 6317-6326 (2009).

- [27] Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265-1272 (2005).
- [28] Usadel, B., Nagel, A., Thimm, O., Redestig, H., Blaesing, O. E., Palacios-Rofas, N., Selbig, J., Hannemann, J., Piques, M. C., Steinhauser, D., Scheible, W. R., Gibon, Y., Morcuende, R., Weicht, D., Meyer, S., Stitt, M. Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known databases. *Plant Physiol.* **138**, 1195-1204 (2005).
- [29] Voit, E. O. *Computational analysis of biochemical systems*; Cambridge University Press: Cambridge, UK (2000).
- [30] Tanaka, N., Mitsui, S., Nobori, H., Yanagi, K., Komatsu, S. Expression and function of proteins during development of the basal region in rice seedlings. *Mol. Cell. Proteomics* **4**, 796-808 (2005).
- [31] Hou, F. F., Thseng, F. S. Studies on the flooding tolerance of soybean seed: varietal differences. *Euphytica* **57**, 169-173 (1991).
- [32] Githiri, S. M., Watanabe, S., Harada, K., Takahashi, R. QTL analysis of flooding tolerance in soybean at an early vegetative growth stage. *Plant Breed.* **125**, 613-618 (2006).
- [33] Hossain, Z., Komatsu, S. Potentiality of soybean proteomics in untying the mechanism of flood and drought stress tolerance. *Proteomes* **2**, 107-127 (2014).
- [34] Komatsu, S., Sugimoto, T., Hoshino, T., Nanjo, Y., Furukawa, K. Identification of flooding stress responsible cascades in root and hypocotyl of soybean using proteome analysis. *Amino Acids* **38**, 729-738 (2010).
- [35] Hashiguchi, A., Sakata, K., Komatsu, S. Proteome analysis of early-stage soybean

- seedlings under flooding stress. *J. Proteome Res.* **8**, 2058-2069 (2009).
- [36] Jiang, C. J., Shoji, K., Matsuki, R., Baba, A., Inagaki, N., Ban, H., Iwasaki, T., Imamoto, N., Yoneda, Y., Deng, X. W., Yamamoto, N. Molecular cloning of a novel importin alpha homologue from rice, by which constitutive photomorphogenic 1 (COP1) nuclear localization signal (NLS)-protein is preferentially nuclear imported. *J. Biol. Chem.* **276**, 9322-9329 (2001).
- [37] Yin, X., Hiraga, S., Hajika, M., Nishimura, M., Komatsu, S. Transcriptomic analysis reveals the flooding tolerant mechanism in flooding tolerant line and abscisic acid treated soybean. *Plant Mol. Biol.* **93**, 479-496 (2017).
- [38] Yin, X., Sakata, K., Komatsu, S. Phosphoproteomics reveals the effect of ethylene in soybean root under flooding stress. *J. Proteome Res.* **13**, 5618-5634 (2014).
- [39] Nanjo, Y., Nouri, M. Z., Komatsu, S. Quantitative proteomic analyses of crop seedlings subjected to stress conditions; a commentary. *Phytochemistry* **72**, 1263-1272 (2011).

Figure Legends

Fig. 1. Web presentation of the Soybean Proteome Database (SPD). After opening the Soybean Proteome Database homepage (<http://proteome.dc.affrc.go.jp/Soybean/>), the lefthand menu includes options for “Gel-based Proteomics” and “Gel-free Proteomics”. When “Gel-free Proteomics” is selected, images of soybean plants (A) and a list of organs or organelles over a sequence of time points (B) are displayed. The “Enter Search keyword” box allows the user to search for protein information and “Download all data in spread sheet style” is used to download all data (C).

Fig. 2. How to use search result. The explanation for “How to use search result” is in the bottom of first page of “Gel-free Proteomics Database for Soybean Proteins”. (1) Search data, (2) Save “Result” as a text file, and then (3) Open the file “Result cgi.txt” with Microsoft Excel.

Fig. 3. Example of a search for soybean protein information in the Soybean Proteome Database. When an image of a specific soybean plant or an organ or organelle is selected, a list of proteins is displayed. The data tables and search results are compatible with Microsoft Excel spreadsheet. The result file can be downloaded onto a personal computer, and dragged/dropped the file into a spreadsheet. When one of the proteins is selected, information on its chromosomal position from DAIZUbase (<http://daizu.dna.affrc.go.jp/>) is shown.

Fig. 4. Application of the Soybean Proteome Database for functional categorization of proteins in the soybean root identified in a time-dependent manner. The Soybean

Proteome Database stores data for organs of soybean plants grown under flooding stress with time course manner. Using these data, a set of 228 proteins significantly changed in abundance in the root under flooding stress, compared with those of non-treated soybean, were detected and analyzed using MapMan bin codes. The proteins used in this analysis are listed in Table 1.

Fig. 5. Application of the Soybean Proteome Database for cluster analysis of significantly changed proteins in soybean during exposure to flooding stress. The same set of 228 proteins used for functional categorization was subjected to cluster analysis using a centroid linkage method based on a Euclidean distance metric. Green, red, and yellow shading indicate an increase, decrease, and no change, respectively, compared with the starting point of treatment, which was 2-day-old soybean seedlings. The number of groups resolved (I with I' to IV) is shown to the right of the heat map. The protein numbers correspond to those for each protein listed in Table 1.

Fig. 6. Application of the Soybean Proteome Database for *in silico* protein-protein interaction analysis of proteins with significant changes in abundance. A set of 228 proteins that changed in abundance under flooding stress compared with the starting point of treatment, which was 2-day-old soybean seedlings, were analyzed for *in silico* protein-protein interactions using the S-system differential equation. The interacting proteins in the network were estimated based on the time-course differential analysis. Red arrows indicate inductive interactions and blue T-bars indicate suppressive interactions. Green arrows show the positions of protein IDs 129, 197, 346, and 368. The protein numbers correspond to those for each protein listed in Table 1.

Figure 1

Figure 2

Figure 3

Figure 5

Figure 6

Table 1. The list of commonly changed proteins in soybean roots without and with flooding stress with time dependent manner

					Control					Flooding				Function
					R	R	R	R	R	R	R	R	R	
No.	ID	Protein Description			at	at	at	at	at	io	io	io	io	
					3(	3(	4(	5(	6(	1)	2)	3)	4)	
No.	ID	Protein Description			0)	0)	0)	0)	0)	F	F	F	F	on
					0)	0)	0)	0)	0)	F	F	F	F	
1	1	Glyma03g2885	beta_1_3_glucanase 1	1	0.	4.	2.	3.	0.	0.	0.	0.	0.	misc
					8	8	7	5	6	5	6	5	5	
					5	7	5	3	5	1	7	5	5	
2	3	Glyma05g3370	receptor_like kinase 1	1	8.	1.	3.	0.	0.	0.	0.	0.	0.	signalling
					0	7	1	9	5	3	2	2	2	
					4	8	5	9	6	6	1	4	4	
3	4	Glyma19g4093	Class II aaRS and biotin synthetases	1	6.	0.	1.	2.	0.	0.	1.	1.	1.	protein
					9	7	9	8	8	9	0	5	5	
					4	8	8	1	8	2	6	5	5	
4	4	Glyma12g3328	ADP/ATP carrier 3	1	8	9.	6.	7.	0.	1.	1.	6.	6.	transport
					8	1	6	2	9	1	4	0	0	
					3	4	7	6	1	4	4	3	3	
5	4	Glyma18g5265	heat shock cognate protein 70_1	1	6.	4.	8.	3.	0.	0.	0.	2.	2.	stress
					7	2	0	1	8	8	9	0	0	
					5	9	4	9	6	4	0	0	0	
6	4	Glyma13g3714	ADP/ATP carrier 3	1	6.	8.	6.	4.	0.	1.	1.	5.	5.	transport
					6	7	6	6	8	0	3	9	9	
					0	7	7	7	5	7	9	0	0	
7	5	Glyma18g5261	heat shock cognate protein 70_1	1	6.	3.	7.	1	0.	0.	0.	2.	2.	stress
					0	8	2	9.	9	9	9	0	0	
					7	7	1	8	1	1	7	1	1	

[illegible]

		1.1			1	1	7	7	5	3	5	9	
								0					
								1					
1		Glyma1			4.	2.	2.	9.	1.	1.	3.	2.	
7	8	3g4313			6	8	4	5	2	6	3	4	
	3	0.1	NADP_malic enzyme 4	1	8	2	3	4	3	8	2	1	TCA
								1					
1		Glyma0			4.	2.	1.	6.	0.	0.	1.	1.	
8	9	8g4459	HEAT SHOCK		4	3	5	7	9	9	0	6	
	0	0.1	PROTEIN 81.4	1	3	3	9	1	8	5	1	3	stress
		Glyma0			4.	1.	1.	0.	1.	1.	0.	2.	
1	9	8g1085			4	1	0	9	0	0	8	6	amino
9	1	0.2	urease	1	3	1	1	0	5	9	4	6	acid
								2					
2		Glyma0			4.	1.	2.	5.	1.	1.	2.	3.	
0	9	6g1811	glyceraldehyde_3_phosp		3	5	1	5	5	5	5	3	glycol
	2	0.1	hate dehydrogenase C2	1	4	3	2	5	2	3	1	8	ysis
		Glyma1			4.	0.	0.	1.	0.	0.	0.	1.	
2	9	5g1125			1	6	7	5	4	9	7	3	
1	4	0.1	peroxiredoxin IIF	1	6	1	3	2	4	8	7	6	redox
		Glyma1			4.	2.	2.	3.	1.	1.	1.	2.	
2	9	9g4177			0	2	4	6	3	4	3	6	protei
2	7	0.1	importin alpha isoform 2	1	9	5	1	5	2	1	0	1	n
								1					
2	1	Glyma1			4.	1.	3.	2.	1.	1.	1.	4.	
3	0	3g1514	Cobalamin_independent		0	6	5	1	0	4	5	8	amino
	2	0.1	synthase	1	3	2	3	2	7	5	3	2	acid
								2					
2	1	Glyma0			3.	2.	2.	4.	1.	1.	2.	4.	
4	0	4g3687	glyceraldehyde_3_phosp		8	0	3	2	3	4	2	2	glycol
	3	0.3	hate dehydrogenase C1	1	8	1	6	1	0	4	7	6	ysis
		Glyma0			3.	0.	0.	1.	0.	0.	0.	2.	
2	0	5g2784			7	9	9	3	9	8	8	0	amino
5	5	0.1	urease	1	4	0	2	9	6	0	9	3	acid

				2									
2	1	Glyma0			3.	1.	2.	3.	1.	1.	2.	4.	
6	1	4g3686	glyceraldehyde_3_phosp		3	9	2	9	2	4	1	0	glycol
	1	0.1	hate dehydrogenase C2	1	0	2	3	3	1	1	4	8	ysis
2	1	Glyma0			3.	0.	3.	0.	1.	1.	1.	7.	N-
7	1	7g1181	glutamine synthase clone		2	7	8	8	1	7	6	2	metab
	2	0.1	R1	1	8	5	7	9	4	4	4	7	olism
2	1	Glyma0			3.	0.	0.	1.	1.	1.	1.	0.	
8	1	2g1164	UDP_glucosyl		2	7	9	8	1	0	0	8	
	5	0.1	transferase 73B3	1	5	2	3	5	6	6	6	7	misc
2	1	Glyma1			3.	2.	3.	3.	0.	1.	1.	2.	
9	1	9g2273	Cytosol aminopeptidase		2	2	0	2	8	0	0	8	protei
	6	0.1	family protein	1	3	9	9	8	9	8	6	4	n
3	1	Glyma0			3.	3.	2.	6.	0.	0.	0.	0.	
0	1	1g3136	20S proteasome alpha		2	6	0	4	9	9	9	3	protei
	7	0.1	subunit C1	1	2	0	7	3	5	9	5	7	n
3	1	Glyma1			3.	0.	0.	1.	0.	0.	0.	1.	
1	1	3g2771			1	6	6	2	4	9	7	2	
	9	0.1	peroxiredoxin IIF	1	7	3	8	5	1	8	3	9	redox
3	1	Glyma0			3.	2.	2.	8.	1.	0.	1.	3.	
2	2	1g2594	proteasome subunit		0	5	3	2	2	9	2	0	protei
	0	0.2	PAB1	1	9	2	9	2	6	7	0	1	n
				3									
3	1	Glyma1			3.	1.	2.	9.	1.	1.	2.	5.	
3	2	0g4133	ATP synthase alpha/beta		0	8	4	0	3	3	3	5	mitoE
	4	0.1	family protein	1	2	6	0	6	1	3	5	5	TC
3	1	Glyma1			2.	1.	2.	3.	0.	0.	0.	0.	
4	2	9g2120	ATPase_ AAA_type_		9	9	1	3	8	7	5	4	
	5	0.1	CDC48 protein	1	7	4	5	2	8	0	3	2	cell
				1 1									
3	1	Glyma0			2.	1.	2.	4.	0.	1.	1.	3.	
5	2	5g1084	Cobalamin_independent		9	5	0	5	9	4	3	1	amino
	6	0.2	synthase	1	6	3	8	0	9	0	8	9	acid
3	1	Glyma1	Pyruvate kinase family	1	2.	2.	1.	6.	1.	1.	1.	1.	glycol

6	2	9g0087	protein		9	4	2	3	0	3	6	8	ysis
	7	0.2			6	7	1	8	8	4	1	5	
								2					
3	1	Glyma0			2.	2.	2.	0.	0.	1.	1.	1.	
7	2	7g0043	vacuolar ATP synthase		9	4	6	3	9	2	4	6	transp
	8	0.1	subunit A	1	4	2	7	4	2	1	1	6	ort
	1	Glyma0			2.	2.	2.	5.	1.	1.	0.	0.	
3	2	9g0443			8	9	6	5	0	0	7	5	protei
8	9	0.1	importin alpha isoform 2	1	3	8	7	0	9	3	6	7	n
	1	Glyma1			2.	1.	1.	5.	0.	0.	0.	0.	
3	3	9g3674	ATPase_ AAA_type_		8	6	7	0	8	6	6	8	
9	0	0.1	CDC48 protein	1	3	8	1	3	8	9	4	1	cell
	1	Glyma0			2.	1.	3.	8.	0.	1.	1.	9.	
4	3	3g2797			8	8	5	1	9	2	3	8	
0	1	0.1	beta_6 tubulin	1	1	8	0	9	1	8	9	1	cell
								1					
4	1	Glyma0			2.	1.	2.	3.	1.	1.	2.	5.	
1	3	6g1812	glyceraldehyde_3_phosp		7	5	0	3	4	4	2	0	glycol
	4	0.1	hate dehydrogenase C1	1	8	8	5	4	8	7	5	2	ysis
	1	Glyma2			2.	1.	1.	5.	1.	1.	1.	2.	
4	3	0g1423			7	5	5	2	0	0	2	1	
2	6	5.1	heat shock protein 81_2	1	6	2	1	8	8	6	7	6	stress
	1	Glyma0			2.	2.	1.	5.	1.	1.	1.	1.	
4	3	5g0931	Pyruvate kinase family		7	3	1	6	0	2	4	4	glycol
3	9	0.2	protein	1	4	6	9	6	6	9	6	7	ysis
								2				1	
4	1	Glyma1			2.	1.	4.	1.	1.	1.	1.	1.	
4	4	4g4038			7	9	1	5	1	3	4	1	
	2	0.1	tubulin beta_1 chain	1	0	8	1	1	1	5	0	4	cell
	1	Glyma1			2.	2.	2.	6.	1.	0.	0.	1.	
4	4	1g2006	ATPase_ AAA_type_		6	1	1	3	1	7	5	0	
5	4	0.2	CDC48 protein	1	8	6	5	6	1	9	9	2	cell
	1	Glyma0			2.	2.	3.	6.	0.	1.	1.	7.	
6	4	5g2561	tubulin beta_7 chain	1	6	0	6	1	9	0	1	2	cell

	6	0.1			6	6	4	9	0	9	9	0	
	1	Glyma0			2.	1.	1.	2.	0.	0.	0.	0.	
4	4	6g1900	ATPase_ AAA_type_		6	3	7	1	7	6	4	6	
7	7	0.1	CDC48 protein	1	3	8	4	6	1	4	9	8	cell
	1	Glyma0			2.	1.	2.	3.	0.	0.	0.	2.	
4	4	3g3913	thioredoxin family		6	9	3	3	5	5	5	4	
8	8	0.1	protein	1	2	9	5	2	7	9	1	2	redox
								1					
4	1	Glyma0			2.	1.	2.	0.	1.	1.	1.	5.	
9	4	5g3580			6	6	5	1	3	4	9	1	
	9	0.1	NADP_malic enzyme 4	1	1	0	7	9	4	0	2	4	TCA
	1	Glyma0			2.	2.	5.	7.	0.	0.	0.	0.	
5	5	5g0146			6	8	6	0	8	7	9	5	transp
0	0	0.1	H-ATPase 11	1	0	7	1	3	0	7	2	8	ort
								1				1	
5	1	Glyma0			2.	2.	4.	5.	1.	1.	1.	0.	
1	5	4g0261			5	0	0	1	1	3	3	8	
	2	0.1	tubulin beta_1 chain	1	8	3	7	0	1	4	9	1	cell
	1	Glyma0			2.	1.	3.	6.	0.	0.	1.	6.	
5	5	9g0302			5	8	4	3	8	9	1	4	
2	3	0.1	tubulin beta 8	1	8	8	2	5	6	9	0	4	cell
	1	Glyma1			2.	1.	1.	2.	0.	0.	0.	0.	
5	5	2g0841	ATPase_ AAA_type_		5	4	9	3	9	7	5	8	
3	5	0.2	CDC48 protein	1	5	5	6	2	8	9	8	9	cell
	1	Glyma1			2.	1.	2.	3.	0.	0.	1.	1.	
5	5	8g2278			5	9	8	8	7	8	0	5	glycol
4	7	0.1	enolase 1	1	4	4	1	8	4	5	6	5	ysis
	1	Glyma0			2.	2.	1.	1.	0.	0.	0.	0.	
5	5	7g3661	Mevalonate/galactokinas		5	8	7	6	7	6	7	7	minor
5	9	0.1	e family protein	1	1	1	5	1	2	6	9	7	CHO
	1	Glyma1			2.	1.	1.	5.	0.	0.	0.	0.	
5	6	0g0648	ATPase_ AAA_type_		5	7	7	1	8	6	6	8	
6	0	0.1	CDC48 protein	1	0	4	0	1	7	5	2	0	cell
5	1	Glyma1	glyceraldehyde_3_phosp	1	2.	0.	1.	1	1.	1.	2.	6.	glycol

7	6	8g0133	hate dehydrogenase C1		4	9	7	4.	6	9	7	1	ysis
	3	0.1			9	3	4	0	4	3	2	8	
								0					
	1	Glyma0			2.	1.	1.	1.	0.	0.	0.	0.	
5	6	4g3595	ATPase_ AAA_type_		4	2	5	9	7	6	5	6	
8	4	0.1	CDC48 protein	1	9	8	6	2	4	7	5	7	cell
	1	Glyma0			2.	0.	0.	8.	0.	0.	0.	0.	
5	6	4g4072	60S acidic ribosomal		4	9	6	7	4	6	5	4	protei
9	5	0.1	protein family	1	8	4	3	6	6	2	6	9	n
	1	Glyma0			2.	1.	3.	5.	0.	0.	1.	5.	
6	6	1g2695			4	5	4	7	9	9	1	4	
0	6	0.1	tubulin beta chain 3	1	8	8	4	1	0	9	0	8	cell
	1	Glyma0			2.	3.	1.	7.	0.	1.	0.	0.	
6	6	6g0838			4	0	2	4	8	0	7	7	protei
1	8	0.1	non_ATPase subunit 9	1	8	3	1	5	8	1	6	7	n
	1	Glyma1			2.	1.	3.	6.	0.	1.	1.	5.	
6	6	0g3796			4	8	2	0	9	0	1	7	
2	9	0.1	tubulin beta 8	1	6	1	5	4	0	1	2	5	cell
	1	Glyma1			2.	1.	3.	1.	0.	0.	0.	0.	
6	7	1g1217	basic transcription factor		4	4	3	5	4	5	4	2	
3	1	0.2	3	1	4	2	0	5	9	2	4	2	RNA
	1	Glyma1			2.	1.	1.	3.	0.	0.	0.	0.	
6	7	1g0672			2	0	2	2	8	8	8	7	
4	3	0.1	Clathrin_ heavy chain	1	7	5	7	3	6	2	6	4	cell
	1	Glyma1			2.	1.	2.	1.	0.	0.	0.	0.	
6	7	2g0438	basic transcription factor		2	2	7	4	4	4	4	2	
5	4	3.1	3	1	5	0	1	7	8	6	1	1	RNA
	1	Glyma0			2.	1.	2.	1.	0.	0.	0.	1.	
6	7	9g3801	Aldolase_type TIM		2	8	2	1	8	8	6	0	cell
6	5	0.2	barrel family protein	1	1	3	8	3	0	1	8	0	wall
	1	Glyma1			2.	0.	0.	4.	0.	0.	0.	1.	not
6	7	7g0360	Haem oxygenase_like_		1	9	7	3	7	8	6	2	assign
7	6	0.1	multi_helical	1	9	8	7	5	4	0	0	2	ed
6	1	Glyma0	Clathrin_ heavy chain	1	2.	1.	4.	2.	0.	0.	0.	0.	cell

8	7	1g3858			1	3	0	8	8	8	8	7	
	7	0.1			9	4	7	7	6	3	7	3	
								1					
6	1	Glyma1			2.	0.	1.	3.	1.	1.	2.	6.	
9	7	1g3736	glyceraldehyde_3_phosp		1	9	7	2	5	8	6	0	glycol
	8	0.1	hate dehydrogenase C1	1	7	4	7	6	9	9	6	9	ysis
	1	Glyma1			2.	1.	1.	3.	0.	0.	0.	0.	
7	7	3g4363			1	3	3	5	7	9	7	5	
0	9	0.1	heat shock protein 91	1	7	1	4	5	1	7	7	8	stress
	1	Glyma1			2.	1.	1.	3.	1.	1.	0.	0.	
7	8	5g1548			1	9	9	1	0	0	8	7	protei
1	0	0.1	importin alpha isoform 2	1	6	7	1	9	3	2	2	6	n
	1	Glyma1			2.	1.	2.	9.	0.	0.	0.	1.	
7	8	2g1058			1	3	6	0	7	8	9	6	
2	1	0.1	aconitase 3	1	3	3	2	7	6	1	1	0	TCA
	1	Glyma1			2.	0.	1.	8.	0.	0.	0.	0.	
7	8	8g4371	20S proteasome alpha		0	7	0	6	6	7	6	8	protei
3	3	0.2	subunit C1	1	8	5	2	9	1	3	9	0	n
	1	Glyma0			2.	1.	2.	8.	0.	0.	0.	1.	
7	8	6g4619			0	3	5	8	7	8	8	4	
4	4	0.2	aconitase 3	1	7	3	9	0	6	4	9	6	TCA
	1	Glyma1			2.	1.	1.	2.	0.	1.	0.	1.	
7	8	2g3382	phosphoenolpyruvate		0	4	2	8	9	0	8	1	glycol
5	5	0.1	carboxylase 2	1	5	0	4	7	5	4	6	5	ysis
	1	Glyma1			2.	1.	1.	3.	0.	0.	0.	0.	
7	8	4g0634			0	0	1	7	9	8	8	6	
6	6	0.1	Clathrin_ heavy chain	1	5	3	6	5	2	0	4	5	cell
	1	Glyma0			2.	1.	4.	2.	0.	0.	0.	0.	
7	8	2g3936			0	3	1	5	8	7	8	4	
7	8	0.1	Clathrin_ heavy chain	1	3	0	0	0	0	8	1	8	cell
	1	Glyma1			2.	0.	1.	4.	1.	1.	1.	1.	
7	8	0g3449	Pyruvate kinase family		0	6	7	2	1	3	7	1	glycol
8	9	0.1	protein	1	2	6	1	8	1	8	8	6	ysis
7	1	Glyma1	Protein of unknown	1	2.	1.	1.	1.	0.	0.	0.	1.	not

9	9	5g0137	function_ DUF642		0	2	3	6	8	3	5	8	assign
	0	0.1			0	2	6	3	9	9	6	1	ed
8	1	Glyma1			1.	0.	0.	2.	0.	0.	0.	0.	
0	9	2g0698			9	8	9	0	9	8	7	8	cell
	1	0.1	UDP_XYL synthase 6	1	3	6	2	1	3	4	5	2	wall
8	1	Glyma0			1.	0.	1.	3.	0.	0.	0.	3.	
1	9	7g0172	Insulinase (Peptidase		9	8	4	2	7	8	8	2	protei
	2	0.1	family M16) protein	1	2	6	6	8	9	8	6	8	n
8	1	Glyma1			1.	0.	0.	3.	0.	0.	0.	1.	
2	9	0g2563			9	6	6	3	7	8	8	3	protei
	3	0.1	heat shock protein 60	1	0	0	0	6	0	6	6	3	n
					1								
8	1	Glyma0			1.	2.	2.	1.	0.	1.	1.	6.	
3	9	5g3786			8	2	5	2	8	1	1	9	
	5	0.1	beta_6 tubulin	1	5	2	1	5	1	0	4	4	cell
8	1	Glyma1			1.	2.	2.	2.	1.	1.	1.	2.	
4	9	0g2900			8	2	0	8	2	3	2	2	protei
	7	0.1	importin alpha isoform 1	1	5	1	9	7	1	3	2	0	n
8	2	Glyma1			1.	1.	4.	2.	0.	0.	0.	0.	
5	0	4g3751			8	4	1	5	8	7	8	5	
	0	0.1	Clathrin_ heavy chain	1	4	5	8	4	1	8	1	4	cell
8	2	Glyma1			1.	1.	2.	2.	0.	0.	0.	0.	
6	0	8g0296			8	1	3	0	8	8	8	6	
	1	0.1	Clathrin_ heavy chain	1	3	5	4	2	7	3	5	7	cell
8	2	Glyma0			1.	0.	0.	5.	0.	0.	0.	0.	
7	0	5g0315	homolog of nucleolar		7	8	6	7	4	4	5	4	
	3	0.1	protein NOP56	1	7	8	9	2	4	1	6	9	RNA
8	2	Glyma0	Nascent		1.	0.	1.	3.	1.	1.	1.	0.	
8	0	3g2757	polypeptide_associated		7	8	1	7	0	0	0	8	protei
	4	0.1	complex alpha	1	6	2	6	2	4	7	1	7	n
8	2	Glyma1			1.	0.	0.	1.	1.	3.	2.	3.	
9	0	9g4004			7	4	6	3	6	0	6	0	major
	5	1.1	sucrose synthase 3	1	6	6	9	2	9	2	5	0	CHO

9	2	Glyma0			1.	0.	1.	5.	1.	1.	1.	2.	
0	0	7g0513	alanine aminotransferase		7	8	3	2	1	3	3	4	amino
	6	0.1	2	1	2	9	6	0	7	2	3	0	acid
9	2	Glyma0	S adenosyl L methionine		1.	0.	0.	0.	0.	0.	0.	0.	
0	0	5g0605	dependent		7	4	6	4	1	1	1	0	
1	7	0.3	methyltransferases	1	1	4	4	4	7	6	2	5	Stress
9	2	Glyma1			1.	0.	1.	3.	0.	0.	0.	0.	
0	0	7g1377	homolog of nucleolar		7	9	1	0	7	4	5	5	
2	8	0.1	protein NOP56	1	0	0	3	8	3	4	7	5	RNA
9	2	Glyma0			1.	0.	0.	1.	0.	0.	0.	0.	
0	0	8g4789			6	7	8	3	7	8	6	7	protei
3	9	0.1	glycyl_tRNA synthetase	1	9	3	1	2	2	3	9	5	n
9	2	Glyma2			1.	1.	1.	4.	0.	0.	0.	1.	
1	0	0g0981	phosphoenolpyruvate		6	2	6	2	9	8	8	2	glycol
4	0	0.1	carboxylase 3	1	6	0	1	7	5	4	0	9	ysis
9	2	Glyma1			1.	1.	2.	5.	0.	0.	1.	2.	
5	1	3g3848			6	4	1	8	8	8	2	1	
1	0	0.1	aconitase 3	1	3	4	2	1	5	3	8	4	TCA
9	2	Glyma1			1.	1.	1.	5.	0.	0.	1.	2.	
1	0	2g3200			6	3	9	4	8	7	1	0	
6	2	0.1	aconitase 3	1	1	9	3	4	0	9	5	5	TCA
9	2	Glyma0			1.	0.	0.	9.	0.	0.	0.	0.	
1	0	5g2206			6	7	9	0	8	7	7	2	protei
7	3	0.1	Subtilase family protein	1	1	0	9	6	1	1	5	1	n
9	2	Glyma1			1.	1.	2.	2.	1.	0.	0.	0.	
1	0	7g1088			5	9	2	1	0	9	8	9	
8	5	0.1	malate dehydrogenase	1	7	1	9	2	8	9	9	9	TCA
9	2	Glyma1			1.	0.	2.	5.	0.	0.	0.	1.	
1	0	1g1922	TCP_1/cpn60 chaperonin		5	8	1	2	8	7	9	6	protei
9	6	0.1	family protein	1	6	8	3	3	3	9	7	4	n
1	2	Glyma1			1.	0.	1.	1.	1.	1.	0.	0.	
0	1	0g2925	regulatory particle		5	7	3	4	1	2	9	8	protei
0	8	0.1	triple_A ATPase 5A	1	2	6	4	0	4	8	4	7	n
1	2	Glyma0	HEAT SHOCK	1	1.	1.	1.	1	0.	0.	0.	0.	stress

0	2	2g4721	PROTEIN 81.4		4	1	6	4.	9	9	9	8	
1	0	0.1			9	1	6	4	3	0	0	1	
								1					
								1					
1	2	Glyma1			1.	1.	2.	6.	0.	0.	0.	0.	not
0	2	6g0553	translocon_associated		4	1	0	7	5	6	6	9	assign
2	1	0.1	protein beta	1	9	5	3	3	8	2	9	9	ed
1	2	Glyma1			1.	0.	2.	5.	0.	0.	0.	1.	
0	2	2g0925	TCP_1/cpn60 chaperonin		4	8	1	2	8	8	9	8	protei
3	2	0.1	family protein	1	9	8	2	3	3	1	9	5	n
1	2	Glyma1			1.	0.	0.	0.	0.	0.	0.	1.	
0	2	3g1170	AMP_dependent		4	4	2	4	5	5	7	7	
4	3	0.2	synthetase and ligase	1	8	4	9	0	9	6	2	8	lipid
1	2	Glyma1			1.	1.	1.	2.	0.	0.	0.	0.	
0	2	7g1417	nucleotide_rhamnose		4	2	5	1	9	9	7	6	cell
5	5	0.1	synthase	1	8	7	3	8	2	6	6	4	wall
1	2	Glyma1			1.	1.	0.	2.	1.	1.	1.	0.	
0	2	4g0269	endoribonuclease L_PSP		4	5	6	5	1	2	0	6	
6	6	0.4	family protein	1	7	0	5	7	7	7	6	6	RNA
1	2	Glyma1			1.	1.	0.	3.	1.	0.	0.	0.	
0	2	2g2419	Aldolase superfamily		4	3	8	0	0	8	9	6	
7	7	0.1	protein	1	6	8	3	1	6	7	3	6	PS
1	2	Glyma1			1.	0.	1.	5.	1.	1.	1.	2.	
0	2	6g0163	alanine aminotransferase		4	8	2	1	1	1	1	0	amino
8	8	0.1	2	1	4	7	6	3	2	6	0	2	acid
1	2	Glyma0			1.	0.	1.	2.	0.	0.	0.	0.	
0	2	4g0169			4	6	8	7	9	7	8	4	protei
9	9	0.1	Ribophorin I	1	1	6	7	6	4	6	0	0	n
1	2	Glyma0			1.	1.	1.	1.	0.	0.	0.	0.	
1	3	7g0042	regulatory particle		3	1	9	2	9	9	7	5	protei
0	0	0.1	triple_A ATPase 6A	1	7	4	7	6	5	0	8	6	n
1	2	Glyma1			1.	1.	1.	1.	1.	1.	0.	0.	
1	3	8g0573	regulatory particle		3	0	9	4	0	0	9	5	protei
1	4	0.1	triple_A ATPase 3	1	4	5	3	7	9	3	7	6	n

1	2	Glyma0		1.	0.	1.	1.	1.	1.	1.	0.	not
1	3	8g3939	DC1 domain_containing		3	9	3	8	0	0	3	assign
2	5	0.1	protein	1	3	3	8	3	6	4	8	ed
1	2	Glyma2		1.	1.	0.	1.	0.	0.	0.	0.	
1	3	0g2209	Proteasome component		3	0	8	0	8	4	9	protei
3	6	0.1	domain protein	1	2	6	1	7	8	4	2	n
								3			1	
1	2	Glyma0		1.	0.	0.	0.	0.	1.	1.	7.	
1	3	6g2129	eukaryotic elongation		3	6	6	7	9	1	0	protei
4	7	0.1	factor 5A_1	1	2	9	9	2	3	0	1	n
1	2	Glyma1		1.	1.	0.	1.	0.	1.	0.	0.	
1	3	8g3155	Insulinase (Peptidase		3	0	6	4	9	7	8	protei
5	8	0.3	family M16) protein	1	0	4	7	0	8	9	7	n
1	2	Glyma0		1.	1.	1.	1.	0.	0.	0.	0.	
1	3	8g0679	Pyridoxal phosphate		3	1	2	6	9	9	7	amino
6	9	0.1	dependent transferases	1	0	9	8	0	3	0	7	acid
1	2	Glyma0		1.	0.	0.	1.	0.	0.	0.	0.	
1	4	3g3295	eukaryotic translation		2	5	6	1	8	8	8	protei
7	2	0.4	initiation factor 3A	1	9	8	8	7	2	0	2	n
1	2	Glyma0		1.	2.	1.	3.	0.	1.	1.	4.	not
1	4	3g0276			2	2	0	5	8	4	7	assign
8	4	0.1	N/A	1	7	1	5	8	5	4	5	ed
1	2	Glyma0		1.	1.	0.	0.	0.	0.	1.	0.	
1	4	8g3942	N_terminal nucleophile		2	1	4	5	6	9	4	protei
9	5	0.1	aminohydrolases	1	7	0	2	2	3	7	2	n
								1				
1	2	Glyma1		1.	2.	3.	1.	0.	1.	1.	1.	
2	4	1g0855			2	0	8	6	8	0	2	
0	6	0.1	aconitase 1	1	6	1	8	1	8	0	7	TCA
1	2	Glyma0		1.	1.	0.	2.	0.	0.	0.	1.	
2	4	6g4305	phosphoenolpyruvate		2	0	9	6	8	8	6	glycol
1	7	0.1	carboxylase 1	1	5	0	5	3	4	5	8	ysis
1	2	Glyma1	aspartate		1.	1.	0.	0.	1.	1.	1.	amino
2	4	7g3305	aminotransferase 5	1	2	4	6	9	0	2	1	acid

2	9	0.1			3	1	2	3	7	3	5	6	
1	2	Glyma1			1.	0.	0.	7.	0.	0.	0.	0.	
2	5	7g1785			2	7	9	7	7	7	6	2	protei
3	0	0.1	Subtilase family protein	1	3	3	4	2	7	0	5	6	n
1	2	Glyma0			1.	2.	3.	7.	0.	1.	1.	1.	
2	5	1g3675			2	0	2	6	9	0	2	2	
4	1	0.1	aconitase 1	1	3	9	2	0	4	4	3	4	TCA
1	2	Glyma1	Thiamine pyrophosphate		1.	1.	4.	0.	1.	0.	1.	0.	Ferme
2	5	4g3886	dependent pyruvate		2	8	0	9	0	9	2	6	ntatio
1	2	Glyma0	S adenosyl L methionine		1.	0.	0.	0.	0.	0.	0.	0.	
2	5	8g4122	dependent		2	5	2	2	7	6	5	3	
6	3	0.4	methyltransferases	1	2	4	9	1	0	3	3	8	stress
1	2	Glyma1			1.	0.	0.	3.	0.	0.	0.	0.	
2	5	6g0509			1	3	5	3	6	5	6	9	protei
7	4	0.1	ribosomal protein S27	1	8	5	3	6	6	9	2	4	n
1	2	Glyma0			1.	2.	0.	1.	1.	0.	0.	0.	
2	5	7g1000	protein phosphatase 2A		1	0	7	6	1	6	7	3	protei
8	5	0.1	subunit A2	1	8	4	0	1	0	5	1	6	n
1	2	Glyma0	Thiamine pyrophosphate		1.	1.	0.	0.	0.	1.	1.	0.	ferme
2	5	2g4055	dependent pyruvate		1	2	9	8	8	0	2	5	ntatio
9	6	0.1	decarboxylase	1	8	0	3	1	3	0	8	8	n
1	2	Glyma1			1.	0.	0.	0.	0.	1.	0.	0.	
3	5	9g4249	Eukaryotic aspartyl		1	4	5	3	6	2	5	6	protei
0	7	0.1	protease family protein	1	8	8	4	6	3	3	8	1	n
1	2	Glyma0			1.	0.	0.	1.	0.	0.	0.	0.	
3	5	9g2525			1	7	5	6	8	9	8	8	protei
1	8	0.1	tripeptidyl peptidase ii	1	6	7	8	9	5	3	6	6	n
1	2	Glyma0	S_adenosyl_L_methioni		1.	1.	1.	1.	1.	1.	1.	3.	2nd
3	5	1g0073	ne_dependent		1	1	4	8	3	4	1	6	metab
2	9	0.5	methyltransferases	1	5	1	7	8	0	2	6	2	olism
1	2	Glyma1	aspartate		1.	2.	0.	0.	1.	1.	1.	0.	amino
3	6	4g1348	aminotransferase 5	1	1	1	8	7	0	2	1	8	acid

3	1	0.1			2	0	2	6	7	6	7	8	
1	2	Glyma0			1.	1.	1.	2.	0.	0.	0.	0.	
3	6	8g3670			0	2	8	5	8	9	7	4	protei
4	3	0.1	ARF_GAP domain 8	1	9	4	1	3	5	9	2	1	n
1	2	Glyma0			1.	0.	0.	0.	1.	0.	0.	0.	
3	6	2g0386			0	5	5	9	1	9	7	4	polya
5	4	0.1	spermidine synthase 1	1	9	4	4	4	0	8	9	3	mine
1	2	Glyma1			1.	0.	0.	6.	0.	0.	0.	0.	
3	6	3g3949	60S acidic ribosomal		0	9	9	8	7	7	8	9	protei
6	5	0.2	protein family	1	6	3	3	2	3	2	4	3	n
1	2	Glyma1			1.	0.	0.	5.	0.	1.	1.	0.	
3	6	2g3080	60S acidic ribosomal		0	8	8	9	8	0	0	9	protei
7	6	0.3	protein family	1	5	6	3	4	8	0	9	7	n
1	2	Glyma1	Late embryogenesis		1.	0.	1.	3.	0.	1.	0.	0.	
3	6	4g0418	abundant protein_ group		0	9	0	3	9	0	8	5	develo
8	7	0.1	2	1	5	6	4	5	7	4	4	6	pment
1	2	Glyma1	Cystathionine		1.	0.	1.	0.	0.	1.	1.	2.	not
3	6	3g4426	beta_synthase (CBS)		0	7	4	9	8	0	8	5	assign
9	8	1.1	protein	1	3	3	3	2	3	5	8	1	ed
1	2	Glyma1			1.	1.	0.	0.	0.	0.	0.	0.	
4	6	4g0552			0	0	4	9	8	7	6	7	
0	9	0.1	PDI_like 2_2	1	3	8	7	6	8	2	9	6	redox
1	2	Glyma1			1.	0.	0.	0.	0.	0.	0.	0.	
4	7	2g0939	NAD_		0	5	4	2	6	7	6	7	protei
1	0	0.1	ADP_ribosyltransferases	1	0	5	0	5	8	9	2	6	n
1	2	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
4	7	3g3508	60S acidic ribosomal		9	1	3	9	4	4	4	1	protei
2	1	0.1	protein family	1	9	4	9	0	4	0	8	8	n
1	2	Glyma0			0.	0.	1.	4.	1.	1.	2.	1.	
4	7	4g0562			9	9	2	9	0	5	4	7	protei
3	2	0.5	aminopeptidase M1	1	9	0	2	7	8	6	3	4	n
1	2	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
4	7	7g1357	MA3 domain_containing		9	4	5	5	8	8	8	6	
4	3	0.1	protein	1	8	0	6	2	4	0	5	1	DNA

1	2	Glyma0	nascent		0.	0.	0.	1.	0.	0.	0.	0.	
4	7	4g4055	polypeptide_associated		9	5	8	8	7	9	7	4	protei
5	4	0.1	complex subunit alpha	1	8	6	8	5	3	9	2	1	n
1	2	Glyma1			0.	0.	0.	4.	1.	1.	1.	1.	
4	7	9g0238	voltage dependent anion		9	8	8	5	1	3	6	2	transp
6	5	0.1	channel 1	1	7	5	2	1	9	3	2	0	ort
1	2	Glyma1			0.	2.	0.	2.	0.	0.	1.	0.	
4	7	2g0429			9	1	7	9	7	9	1	6	
7	6	0.1	Coatomer_ alpha subunit	1	7	9	3	9	6	3	7	6	cell
1	2	Glyma0			0.	0.	0.	0.	1.	1.	0.	0.	
4	7	4g3762	ubiquitin_conjugating		9	4	4	6	1	2	9	1	protei
8	8	0.5	enzyme 27	1	6	4	2	7	8	3	7	8	n
1	2	Glyma2			0.	2.	0.	2.	0.	0.	0.	0.	
4	8	0g2479	protein phosphatase 2A		9	5	6	0	9	5	7	3	protei
9	0	0.1	subunit A2	1	3	0	4	0	7	9	0	6	n
1	2	Glyma1			0.	0.	0.	0.	0.	0.	0.	1.	
5	8	1g3763	TCP_1/cpn60 chaperonin		9	9	7	6	9	7	7	4	protei
0	1	0.1	family protein	1	2	6	6	4	5	9	6	3	n
1	2	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
5	8	3g3214	ribophorin II (RPN2)		9	5	4	3	6	6	6	4	develo
1	3	0.3	family protein	1	1	8	2	8	3	3	6	8	pment
1	2	Glyma1			0.	0.	1.	1.	0.	0.	0.	0.	
5	8	1g1261	coatomer gamma_2		9	8	5	5	7	6	6	5	
2	4	0.1	subunit	1	0	1	4	6	4	6	6	7	cell
1	2	Glyma0			0.	0.	0.	0.	1.	0.	0.	0.	
5	8	4g3657	MA3 domain_containing		8	4	5	4	0	8	7	5	
3	6	0.1	protein	1	9	5	8	8	8	8	9	7	DNA
1	2	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	not
5	8	2g0129	transducin family protein		8	6	5	5	7	8	8	6	assign
4	7	1.1	/ WD_40 repeat	1	8	1	2	6	3	6	8	6	ed
1	2	Glyma1	NOP56_like pre RNA		0.	0.	0.	0.	0.	0.	0.	0.	
5	8	1g1220	processing		8	4	3	4	8	4	6	2	protei
5	8	0.1	ribonucleoprotein	1	8	5	1	7	3	8	3	9	n
1	2	Glyma2	Coatomer_ alpha subunit	1	0.	2.	0.	2.	0.	0.	0.	0.	cell

5	8	0g3327			8	4	8	4	7	6	6	5	
6	9	0.1			7	4	2	9	3	7	7	3	
1	2	Glyma0			0.	0.	0.	2.	0.	0.	0.	0.	
5	9	8g3911	Pyruvate phosphate		8	7	6	3	5	6	5	3	major
7	0	0.1	dikinase_ PEP	1	7	4	3	9	3	2	8	2	CHO
1	2	Glyma0			0.	1.	1.	0.	0.	1.	1.	0.	
5	9	8g4521	alpha_glucan		8	0	1	1	8	1	1	3	major
8	2	0.1	phosphorylase 2	1	6	9	9	1	3	4	9	6	CHO
1	2	Glyma1			0.	0.	1.	2.	1.	1.	0.	0.	
5	9	1g1077	NAD(P)_binding		8	6	3	0	0	0	9	8	
9	3	0.3	Rossmann_fold	1	4	9	7	9	4	4	1	3	lipid
1	2	Glyma0			0.	0.	0.	1.	1.	1.	0.	1.	
6	9	5g0319	arginosuccinate synthase		8	6	7	0	0	0	9	1	amino
0	5	0.1	family	1	3	7	5	4	3	8	8	0	acid
1	2	Glyma1			0.	0.	0.	1.	0.	0.	0.	0.	
6	9	5g4320	sterol methyltransferase		8	9	9	2	8	6	2	2	hormo
1	7	0.3	1	1	3	8	2	8	0	1	2	4	ne
1	2	Glyma0			0.	0.	0.	1.	1.	0.	1.	0.	
6	9	9g3349	Zincin_like		8	9	8	1	0	8	0	5	protei
2	9	0.1	metalloproteases	1	1	0	5	4	9	7	2	7	n
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	not
6	0	8g1476			8	7	6	7	9	7	7	5	assign
3	0	0.1	binding	1	1	7	1	7	4	2	7	6	ed
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
6	0	9g0638			8	6	7	5	7	7	5	5	minor
4	1	0.2	Galactose mutarotase	1	1	1	7	9	5	5	3	2	CHO
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	not
6	0	1g3615			8	7	6	7	6	5	6	3	assign
5	5	0.2	binding	1	0	6	1	3	1	5	2	2	ed
1	3	Glyma1			0.	1.	0.	1.	0.	0.	0.	0.	
6	0	0g3431			8	0	6	8	7	6	6	5	
6	6	0.1	Coatomer_ alpha subunit	1	0	5	5	8	4	8	9	6	cell
1	3	Glyma1	Class I glutamine		0.	0.	1.	0.	0.	1.	0.	0.	not
6	0	2g3575	amidotransferase	1	7	5	0	6	8	0	7	3	assign

7	8	0.1			8	0	3	1	1	9	5	5	ed
1	3	Glyma1			0.	0.	0.	0.	0.	0.	1.	0.	
6	0	2g1768	flavodoxin_like quinone		7	8	2	1	6	7	1	5	
8	9	0.1	reductase 1	1	6	0	7	9	8	9	0	0	lipid
1	3	Glyma0			0.	0.	0.	3.	1.	0.	0.	0.	
6	1	3g3280	regulatory particle		7	7	6	0	1	9	9	9	protei
9	0	0.1	AAA_ATPase 2A	1	5	7	2	5	2	5	8	0	n
1	3	Glyma1			0.	0.	0.	1.	0.	0.	0.	0.	
7	1	2g0207	methionine__tRNA		7	5	9	0	7	7	7	7	protei
0	4	6.1	ligase	1	2	8	4	1	6	7	2	6	n
1	3	Glyma1			0.	0.	0.	1.	0.	1.	0.	0.	
7	1	1g0725	succinate dehydrogenase		7	4	7	7	8	0	8	7	
1	5	0.1	1_1	1	2	9	1	0	6	0	1	4	TCA
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
7	1	3g1020	methionine__tRNA		7	4	5	4	4	5	4	4	protei
2	6	0.1	ligase	1	1	1	6	9	9	3	7	8	n
1												1	
7	3	Glyma0			0.	0.	0.	1.	1.	2.	2.	3.	
3	1	3g3744			6	4	6	0	7	8	8	6	major
7	1.1	sucrose synthase 3	1	8	6	8	6	2	0	8	4	4	CHO
1	3	Glyma2			0.	0.	0.	2.	0.	0.	0.	0.	
7	1	0g2216			6	5	5	9	3	2	2	1	signall
4	8	0.1	phytochrome A	1	8	6	4	0	5	4	6	9	ing
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	not
7	1	4g0871			6	8	9	5	5	5	4	3	assign
5	9	0.1	catalytics	1	5	1	9	7	0	5	9	3	ed
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
7	2	3g4387	pleiotropic drug		6	4	3	2	8	8	9	5	transp
6	1	0.1	resistance 12	1	4	2	5	2	8	3	8	9	ort
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	2nd
7	2	1g0295	hydroxymethylglutaryl_		6	4	4	6	6	5	5	1	metab
7	4	0.1	CoA synthase	1	3	5	5	3	7	1	0	6	olism
1	3	Glyma1	NAD_		0.	0.	0.	0.	0.	0.	1.	0.	protei
7	2	1g1907	ADP_ribosyltransferases	1	6	2	2	5	7	9	7	7	n

8	5	0.2			3	0	3	2	0	7	6	5	
1	3	Glyma1			0.	1.	0.	1.	0.	0.	0.	2.	
7	2	8g0662	ubiquitin_activating		6	0	9	2	8	8	8	5	protei
9	6	0.1	enzyme 1	1	3	6	0	8	8	1	5	5	n
1	3	Glyma0			0.	1.	0.	0.	1.	0.	0.	1.	
8	2	5g0526	RAB GTPase homolog		6	0	3	4	1	9	9	5	signall
0	7	0.1	1C	1	1	4	9	4	6	5	1	2	ing
1	3	Glyma0	Thiamine pyrophosphate		0.	0.	0.	0.	3.	3.	4.	5.	ferme
8	2	7g1857	dependent pyruvate		5	4	4	3	6	6	1	9	ntatio
1	9	0.1	decarboxylase	1	8	9	4	8	8	0	5	5	n
1	3	Glyma0			0.	1.	0.	0.	1.	0.	0.	0.	
8	3	9g3176	protein phosphatase 2A		5	6	6	9	0	6	5	3	protei
2	0	0.1	subunit A2	1	8	4	0	1	0	3	9	0	n
1	3	Glyma0			0.	0.	0.	0.	1.	1.	0.	1.	
8	3	8g0239	vacuolar ATP synthase		5	4	2	3	0	0	8	2	transp
3	1	0.1	subunit E1	1	6	5	0	9	1	8	4	5	ort
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	2nd
8	3	1g4245	hydroxymethylglutaryl_		5	5	5	7	7	5	5	2	metab
4	2	0.1	CoA synthase	1	6	9	0	1	4	7	5	7	olism
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
8	3	6g2703			5	5	5	8	5	7	7	4	
5	3	0.1	tubulin alpha_3	1	4	5	5	1	3	2	3	5	cell
1	3	Glyma1			0.	0.	0.	0.	3.	3.	4.	6.	ferme
8	3	8g4346	pyruvate		5	3	3	3	6	6	1	6	ntatio
6	4	0.1	decarboxylase_2	1	3	1	4	4	2	2	1	2	n
1	3	Glyma1			0.	0.	0.	0.	1.	1.	0.	1.	
8	3	7g0738	xylose isomerase family		5	5	2	5	0	2	9	0	minor
7	5	0.1	protein	1	2	7	8	9	1	3	2	6	CHO
1	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
8	3	2g0177	Zinc_binding alcohol		5	4	3	7	8	8	7	3	
8	6	0.3	dehydrogenase	1	2	5	1	5	9	4	4	1	misc
1	3	Glyma1	hydroxymethylglutaryl_		0.	0.	0.	0.	0.	0.	0.	0.	2nd
8	3	7g1415	CoA synthase	1	5	3	3	4	3	3	2	0	metab

9	8	0.1			1	2	1	3	6	6	3	8	olism
1	3	Glyma0			0.	0.	0.	0.	0.	1.	0.	0.	
9	3	1g3820	succinate dehydrogenase		5	5	3	4	7	0	6	5	
0	9	0.1	1_1	1	1	6	2	7	5	7	4	2	TCA
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
9	4	6g0179			5	4	6	0	9	7	8	3	protei
1	0	0.1	Ribophorin I	1	1	9	4	9	2	4	1	3	n
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
9	4	4g0722			5	5	3	3	2	2	1	1	cell
2	1	0.1	cellulose synthase 1	1	0	0	6	9	7	5	3	1	wall
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
9	4	5g3502	Class II aaRS and biotin		4	5	5	4	5	7	9	5	protei
3	3	1.1	synthetases	1	8	8	5	7	6	6	7	8	n
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
9	4	2g0640	succinate dehydrogenase		4	2	2	6	6	8	3	5	
4	5	0.1	1_1	1	8	2	2	3	3	4	9	1	TCA
1	3	Glyma0	cytochrome P450_		0.	0.	0.	0.	0.	0.	0.	0.	
9	4	7g3233	family 93_ subfamily D_		4	4	3	1	5	5	6	3	
5	6	0.1	polypeptide 1	1	7	6	1	8	9	1	0	4	misc
1	3	Glyma0			0.	0.	0.	0.	1.	1.	0.	1.	
9	4	5g3719	vacuolar H_ ATPase		4	5	2	3	0	0	8	1	transp
6	8	0.1	subunit E isoform 3	1	7	0	4	8	2	3	0	7	ort
1	3	Glyma0	cytochrome P450 family		0.	0.	0.	0.	0.	0.	0.	0.	
9	4	8g2595	72 subfamily A		4	3	4	4	3	1	1	1	
7	9	0.1	polypeptide 15	1	7	9	2	7	0	3	1	5	misc
1	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
9	5	7g3846	UDP_Glycosyltransferas		4	5	3	7	8	7	6	4	
8	0	0.1	e	1	6	9	5	2	2	0	6	9	misc
1	3	Glyma0			0.	3.	1.	1.	0.	2.	4.	6.	not
9	5	3g0285			4	8	1	3	6	0	1	3	assign
9	1	2.1	N/A	1	6	5	2	3	3	4	9	6	ed
2	3	Glyma1			0.	0.	0.	2.	1.	1.	0.	2.	
0	5	5g1617			4	5	8	6	0	1	9	8	major
0	2	1.1	sucrose synthase 2	1	6	3	3	4	7	2	9	1	CHO

2	3	Glyma2			0.	0.	0.	0.	0.	0.	0.	0.	
0	5	0g2516	CysteinyI_tRNA		4	4	5	3	8	7	7	4	protei
1	3	0.3	synthetase_ class Ia	1	5	6	3	8	1	7	8	7	n
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
0	5	4g3969	GTP_binding		4	7	2	1	8	7	9	6	signall
2	4	0.1	protein_related	1	3	7	5	6	0	9	4	0	ing
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
0	5	1g4244	nucleotide_rhamnose		4	5	5	6	5	4	2	1	cell
3	6	1.1	synthase	1	2	2	8	3	5	1	5	7	wall
2	3	Glyma0			0.	0.	0.	0.	0.	1.	0.	0.	
0	5	5g3663	ARM repeat superfamily		4	5	6	6	8	0	9	4	protei
4	7	0.1	protein	1	1	5	1	0	0	6	0	9	n
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
0	5	1g0690			3	5	5	7	4	6	6	2	protei
5	8	0.1	aminopeptidase P1	1	8	1	9	4	0	2	2	6	n
2	3	Glyma1			0.	0.	0.	0.	1.	1.	1.	0.	
0	5	1g1587	RmlC_like cupins		3	1	1	1	8	1	1	2	develo
6	9	0.1	superfamily protein	1	8	1	5	7	5	1	7	2	pment
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
0	6	6g0229	Pyridoxal_dependent		3	1	1	5	2	2	0	1	amino
7	0	0.1	decarboxylase	1	7	0	1	9	2	0	8	0	acid
2	3	Glyma1	Cystathionine		0.	0.	0.	0.	0.	0.	0.	0.	not
0	6	5g0100	beta_synthase (CBS)		3	1	1	0	6	5	6	3	assign
8	5	0.1	protein	1	5	4	8	8	3	5	8	6	ed
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	not
0	6	3g1714			3	2	1	1	5	6	4	1	assign
9	6	0.1	N/A	1	3	0	9	6	3	8	5	9	ed
2	3	Glyma0			0.	0.	0.	0.	1.	2.	3.	2.	
1	6	6g0565			3	4	3	2	1	1	1	4	protei
0	7	0.1	aminopeptidase M1	1	2	2	7	6	3	3	9	8	n
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
1	6	5g0291	MA3 domain_containing		3	3	2	1	6	6	7	4	
1	8	0.1	protein	1	2	1	1	0	5	8	8	2	DNA
2	3	Glyma0	auxin transport protein	1	0.	0.	0.	0.	0.	0.	0.	0.	hormo

1	6	8g2799	(BIG)		2	3	2	3	5	3	4	2	ne
2	9	3.1			9	3	7	9	1	9	0	6	
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
1	7	6g2308			2	2	3	1	4	4	5	2	glycol
3	1	1.1	enolase 1	1	8	7	3	5	7	4	3	1	ysis
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
1	7	7g0378	pleiotropic drug		2	3	2	2	4	4	4	1	transp
4	2	0.1	resistance 12	1	7	0	1	4	5	0	0	4	ort
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
1	7	4g1010	L_Aspartase_like family		2	3	3	2	5	5	4	2	amino
5	4	0.1	protein	1	6	8	3	1	0	6	9	2	acid
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
1	7	8g0748	alpha_glucan		2	2	2	1	4	4	6	3	major
6	6	0.1	phosphorylase 2	1	4	0	1	0	6	5	0	5	CHO
2	3	Glyma0			0.	0.	0.	0.	1.	0.	0.	0.	
1	7	8g3840	regulatory particle		2	5	6	4	3	9	9	2	protei
7	7	3.1	AAA_ATPase 2A	1	4	6	5	0	0	8	6	2	n
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
1	7	6g1492			2	3	4	1	6	5	4	2	minor
8	8	0.1	Melibiose family protein	1	4	7	5	7	4	1	7	7	CHO
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
1	8	4g0715	Glucose_1_phosphate		2	2	2	1	4	3	3	1	cell
9	0	0.1	adenylyltransferase	1	1	2	5	5	6	7	8	0	wall
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
2	8	2g0180	GroES_like zinc_binding		1	0	0	0	3	1	9	1	
0	2	1.1	dehydrogenase	1	8	4	2	3	1	4	5	8	misc
2	3	Glyma2			0.	0.	0.	0.	1.	1.	1.	0.	
2	8	0g3878	peptidase M1 family		1	5	1	2	0	3	2	4	protei
1	3	0.1	protein	1	6	9	7	9	0	2	4	7	n
2	3	Glyma1			0.	0.	0.	0.	0.	0.	0.	0.	
2	8	9g4224			1	8	8	1	3	3	6	3	
2	4	0.2	histone H2A 7	1	3	0	1	5	6	8	3	1	DNA
2	3	Glyma1	protein phosphatase 2A		0.	0.	0.	1.	0.	0.	0.	0.	protei
2	8	0g4225	subunit A2	1	1	4	4	1	9	7	6	4	n

3	5	0.1			3	4	3	1	5	6	9	6	
2	3	Glyma0			0.	0.	0.	0.	1.	0.	1.	0.	
2	8	1g4347	AMP_dependent		1	2	3	3	0	4	7	1	
4	6	0.1	synthetase and ligase	1	2	7	7	4	1	5	7	6	lipid
2	3	Glyma1			0.	0.	0.	0.	0.	1.	0.	0.	
2	8	5g0610			1	2	1	1	7	2	7	3	nucleo
5	7	0.1	pyrimidine 2	1	2	8	9	7	3	7	5	7	tide
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	2nd
2	8	7g1534	caffeoyl_CoA		1	1	0	0	5	6	4	3	metab
6	8	1.1	3_O_methyltransferase	1	1	1	5	2	6	1	9	0	olism
2	3	Glyma0			0.	0.	0.	0.	0.	0.	0.	0.	
2	8	169s00	60S acidic ribosomal		0	0	0	0	5	6	3	0	protei
7	9	200	protein family	1	7	7	2	8	0	4	6	2	n
2	3	Glyma2			0.	0.	0.	0.	0.	1.	0.	0.	
2	9	0g3203	phosphoglucosamine		0	0	0	1	6	0	9	2	glycol
8	0	0.1	mutase	1	5	2	6	2	4	7	1	5	ysis

N/A means that there is no description. Abbreviations are as follows: misc, miscellaneous; TCA, *tricarboxylic acid cycle*; mitoETC, mitochondrial electron transport chains; PS, photosynthesis; 2nd metabolism, second metabolism.

Biological significance

The Soybean Proteome Database stores data obtained from both gel-based and gel-free proteomic techniques. The gel-free proteomic database comprises 39,212 proteins identified in 63 sample sets, such as different organs of soybean plants grown under flooding stress or non-stressed conditions in a time-dependent manner. In addition, organellar proteins identified in mitochondria, nuclei, and endoplasmic reticulum are stored in the gel-free proteomics database. A total of 44,704 proteins, including 5,490 proteins identified using a gel-based proteomic technique, are stored in the SPD. It accounts for approximately 80% of all predicted proteins from genome sequences, though there are over lapped proteins. Based on the demonstrated application of data stored in the database for functional analyses, it is suggested that these data will be useful for analyses of biological mechanisms in soybean. Furthermore, coupled with recent advances in information and communication technology, the usefulness of this database would increase in the analyses of biological mechanisms.

Highlights

- Soybean Proteome Database stores data obtained from both gel-based and gel-free proteomic techniques.
- The gel-free proteomic database comprises 39,212 proteins identified in 63 sample sets in soybean.
- Database stores different organs of soybean grown without or with stress in a time-dependent manner.
- Organellar proteins identified in mitochondria, nuclei, and endoplasmic reticulum are also stored
- Based on functional analyses, these data are useful for analyses of biological mechanisms in soybean.