



## Research article

## Global gene expression profiles in developing soybean seeds

Tomiko Asakura<sup>a,\*</sup>, Tomoko Tamura<sup>a,1</sup>, Kaede Terauchi<sup>a</sup>, Tomoyo Narikawa<sup>a</sup>, Kazuhiro Yagasaki<sup>b</sup>, Yoshiro Ishimaru<sup>a</sup>, Keiko Abe<sup>a</sup><sup>a</sup>Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan<sup>b</sup>Nagano Vegetable and Ornamental Crops Experiment Station, 1066-1 Tokoo, Souga, Shiojiri, Nagano 399-6461, Japan

## ARTICLE INFO

## Article history:

Received 18 July 2011

Accepted 20 December 2011

Available online 30 December 2011

## Keywords:

Cluster dendrogram

DNA microarray

Seed development

Soybean

Storage protein

## ABSTRACT

The gene expression profiles in soybean (*Glycine max* L.) seeds at 4 stages of development, namely, pod, 2-mm bean, 5-mm bean, and full-size bean, were examined by DNA microarray analysis. The total genes of each sample were classified into 4 clusters based on stage of development. Gene expression was strictly controlled by seed size, which coincides with the development stage. First, stage specific gene expression was examined. Many transcription factors were expressed in pod, 2-mm bean and 5-mm bean. In contrast, storage proteins were mainly expressed in full-size bean. Next, we extracted the genes that are differentially expressed genes (DEGs) that were extracted using the Rank products method of the Bioconductor software package. These DEGs were sorted into 8 groups using the hclust function according to gene expression patterns. Three of the groups across which the expression levels progressively increased included 100 genes, while 3 groups across which the levels decreased contained 47 genes. Storage proteins, seed-maturation proteins, some protease inhibitors, and the allergen Gly m Bd 28K were classified into the former groups. Lipoxygenase (LOX) family members were present in both the groups, indicating the multi-functionality with different expression patterns.

© 2011 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Soybean is one of the most important crops cultivated worldwide. It is rich in proteins and lipids, which are processed to yield vegetable oil. Every year, about 2.2 billion tons of soybeans are cultivated. Soybean is widely used in many processed foods and serves as raw material for functional foods. However, soybean contains some allergens such as Gly m Bd 30K [1–4] and trypsin inhibitor [5]. These compounds are synthesized and stored in seeds during maturation. Soybean seeds mature in about 10 weeks from flowering, although the duration to seed maturation differs according to species and surroundings. After pollination, pods elongate before seed growth. Dehydration starts 5–6 weeks after flowering, and seeds reach their maximum size in about 7–8 weeks after pollination. Subsequently, the seeds dry completely, and the maturation process is completed in about 10 weeks. During maturation, metabolic activity in the seeds drastically changes from the developing stage to the dehydration stage toward dormancy.

Changes in protein metabolism [6–8], lipid synthesis [9–12] and other functional groups such as isoflavones [13,14], proteases [15,16], and protease inhibitors [17–20] have been extensively investigated. The expression of storage proteins is first detected 14–18 d after anthesis and all proteins are expressed in 50 d after anthesis [8]. Lipid metabolism has been investigated mainly in reference to the synthesis of triglycerides and lipoxygenases (LOXs). Many authors have studied the expression patterns of individual genes during seed maturation, but a few global analyses of gene expression in seeds have been conducted.

Recently, we studied the gene expression profile of Tosan205, a soybean cultivar that almost completely lacks glycinin subunits. The gene expression profile of Tosan205 was compared with that of Tamahomare, a parent cultivar of Tosan205, at the same stages of development. Tosan205 exhibited higher expression levels of stress-related genes such as GST and APX than Tamahomare did. Several genes were identified as differentially expressed genes (DEGs), but difference in the number of DEGs between Tosan205 and Tamahomare is not high, regardless of the lack of glycinin [21]. It seems that the basic expression profile in soybean seeds is similar among different cultivars. We investigated the gene expression profiles during seed development. Though some global expression profiles were reported [22], the change of gene expression just after anthesis was not found. An understanding of the changes in gene expression from the beginning of the process of seed formation will

Abbreviations: DEG, Differentially expressed genes; LOX, Lipoxygenase; 2D, Two-dimensional.

\* Corresponding author. Tel./fax: +81 3 5841 1879.

E-mail address: [asakura@mail.ecc.u-tokyo.ac.jp](mailto:asakura@mail.ecc.u-tokyo.ac.jp) (T. Asakura).<sup>1</sup> Present address: Department of Nutritional Science, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan.

clarify seed metabolism patterns throughout seed development. Based on these observations, we conducted a global analysis of gene expression in seeds at 4 stages of development by using DNA microarray analysis.

## 2. Results

### 2.1. Numbers of genes expressing at each stage

The numbers of genes expressed in the 4 stages of development, pod, 2-mm bean, 5-mm bean, and full-size bean, were examined. The data were normalized by MAS5 and the detection calls “A,” “M” and “P” meaning “absent,” “marginal” and “present”, respectively, were used to analyze the gene expression. The results are shown in Fig. 1. Two replicated samples in each stage described with “present” were indicated in gray column. In pod stage, 29,788 genes were expressed; 29,183 in 2-mm bean, and 28,280 in 5-mm bean and 22,548 in full-size. They account for 81%, 79%, 76%, and 61%, respectively, of the total gene, 37,000 (Fig. 2A). In pod stage, genes were most expressed among others. The number of expressed genes gradually decreased as seed matured, and 21,231 genes were commonly expressed in all stages. This value accounts for the 70–95% of expressed genes in each stage. The number of genes expressed in all stages was subtracted from that of genes expressed in each stage (Fig. 2B). These genes showed stage specific expression. More than 7000 genes were expressed in pod, 2-mm bean or 5-mm bean. However, about 1000 genes were expressed in full-size bean. Thus, the gene expression is drastically changed during seed developing. The annotated genes classified in Fig. 1 are shown in Supplemental Table 1 and Table 2. Many transcription factors were found in pod, 2-mm bean. On the other hand, storage proteins were found in full-size bean. These results indicate that the gene

| Developing stage |              |              |                      | number<br>of<br>genes |
|------------------|--------------|--------------|----------------------|-----------------------|
| Pod              | 2-mm<br>bean | 5-mm<br>bean | Full<br>size<br>bean |                       |
|                  |              |              |                      | 1,628                 |
|                  |              |              |                      | 644                   |
|                  |              |              |                      | 240                   |
|                  |              |              |                      | 216                   |
|                  |              |              |                      | 927                   |
|                  |              |              |                      | 569                   |
|                  |              |              |                      | 150                   |
|                  |              |              |                      | 157                   |
|                  |              |              |                      | 364                   |
|                  |              |              |                      | 61                    |
|                  |              |              |                      | 5,164                 |
|                  |              |              |                      | 416                   |
|                  |              |              |                      | 146                   |
|                  |              |              |                      | 171                   |
|                  |              |              |                      | 21,231                |
| 29,788           | 29,183       | 28,280       | 22,548               |                       |

Fig. 1. Numbers of genes expressing at seed-developing stages gray cells represent that two replicated probes are assigned as “present”. The lowest row shows the numbers of genes expressed at 4 developing stages.

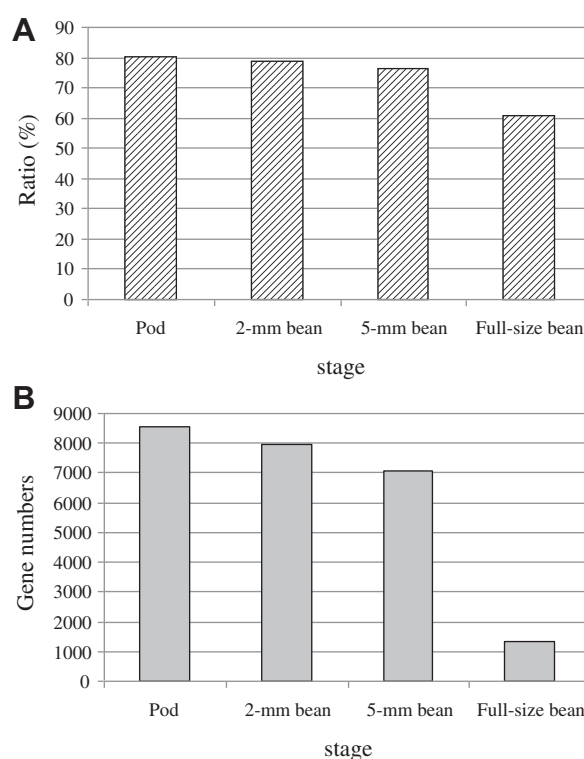


Fig. 2. Numbers of genes expressing in seeds at developing stages counted by Present/Absent call detection A, Ratio of expressed gene number “present” and total gene number at each stage. B, numbers of genes expressed at all stages subtracted from a number of genes expressed at each stage.

expression differs depending on seed size. This analytical method based on “present” or “absent” call was able to characterize the gene expression profile in each seed developing stage. However it was unable to find the change in gene expression. Therefore, we tried to analyze the gene expression by another analytical method.

### 2.2. Numbers of differentially expressed genes (DEGs) among four different stages of development

Eight samples from the 4 stages were used for hierarchical analysis, and the cluster dendrogram was constructed using the ‘pvclust’ function (Fig. 3A). The samples were clustered together based on genes that show similar expression. The clusters of the 2-mm bean and 5-mm bean samples were very closely positioned. The pod cluster was the next closest to these 2 clusters, and the full-size cluster was farthest from the other 3 clusters (Fig. 3B). These data indicate that gene expression in seeds changes as seeds develop. Next, we identified DEGs to characterize gene expression during seed development. The gene expression profiles from 2 adjacent stages were compared with each other, and genes with significant differences ( $FDR < 0.05$ ) in expression were categorized as DEGs (Table 1). A total of 1005 DEGs were identified, of which 362 were differentially expressed between the pod and 2-mm bean stages; 276, between the 2-mm and 5-mm bean stages; and 367, between the 5-mm and full-size bean stages. There were a total of 380 overlapping genes among the DEGs, which were excluded from further analyses.

### 2.3. Classification of DEGs by expression patterns

The remaining 625 DEGs were subjected to hierarchical analysis using the ‘hclust’ function. The cluster dendrogram was divided

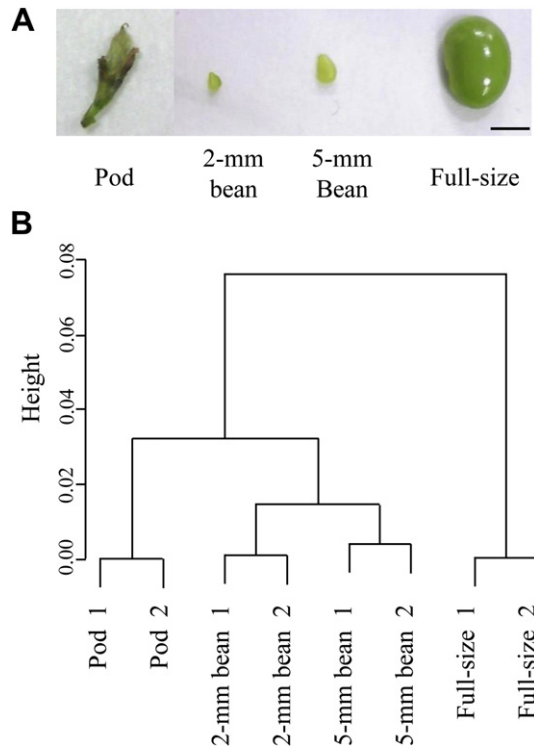


**Fig. 3.** Cluster dendrogram for genes expressed in soybean seeds A, Picture of samples used in DNA microarray. Bar = 5 mm. B, clusters for 8 samples from 4 developing stages constructed using the 'pvclust' function.

using the 'cut-tree' function to classify the genes for characterization of the change in expression. As a result, cutting the tree by the height to be classified into 8 clusters was found appropriate for the clustering to show the gene expression profiles most distinctly (Fig. 4). Annotation was made for 138 out of 625 genes (Table 2). Group 1 contained 2 genes that increased in expression during seed development, one of which was the  $\beta$ -conglycinin gene. Genes in this group showed a particularly drastic change in expression between the 2-mm and 5-mm bean stages. Group 2 contained 36 genes showing a gradual increase in expression between the 2-mm and full-size bean stages.  $\beta$ -Conglycinin  $\beta$ -subunit, glycinin subunits, and lipoxygenases belonged to this group. Group 3 contained 62 genes showing gradual increase in expression throughout development. Genes in this group included several seed-maturation proteins, as well as oleosin and  $\omega$ -6 desaturase, which are known to be involved in lipid synthesis. Seed proteins such as albumin 1, basic 7S globulin, and Gly m Bd 20K were also classified into this group. Group 4 contained 1 gene that showed a drastic decrease in expression between the pod and 2-mm bean stages, and then remained at very low levels until the full-size bean stage. Group 5 contained 28 genes showing drastic decrease in expression between the pod and 2-mm bean stages, and some expression changes during the rest of the development process. Five out of these 28 genes were annotated, including 2 haloacid dehalogenase-

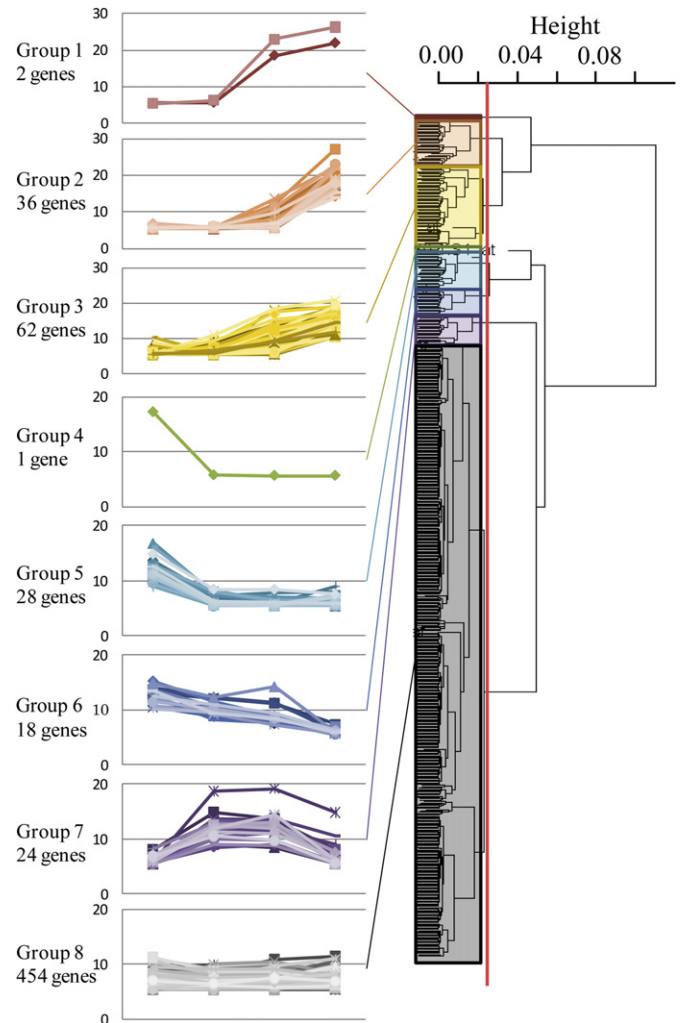


**Fig. 4.** Hierarchical clustering analysis the 625 identified DEGs (see Table 1 for list) were divided into 8 groups according to their expression patterns. The 'hclust' function was used to construct a dendrogram. The numbers of genes belonging to clusters, 1, 2, 3, 4, 5, 6, 7, and 8 were 2, 36, 62, 1, 28, 18, 24, and 454, respectively.

like hydrolases of molecular weights 28 kDa and 31 kDa, respectively. Group 6 contained 18 genes that showed gradual decrease in expression throughout the development. The genes encoding LOX5 and LOX7 were classified in this group. Group 7 contained 24 genes showing a bell-shaped expression profile, with high expression in the 2-mm and 5-mm bean stages and low expression in the pod and full-size stages. Only 1 gene, subtilisin-type protease precursor, was annotated in this group. Group 8 contained 454 genes that did not show drastic changes in expression. This group included proteases, protease inhibitors, seed-maturation proteins, LOXs, GSTs, and other enzymes. Most of the genes studied belonged to this group, including many enzymes involved in the synthesis of seed components, as well as seed-maturation proteins.

**Table 1**

The numbers of DEG extracted with FDR < 0.05 when compared between two adjacent stages. All DEG adds up to 625 genes when overlapping genes were eliminated.

| Pod vs. 2-mm bean |     | 2-mm bean vs. 5-mm bean |     | 5-mm bean vs. Full-size |           |
|-------------------|-----|-------------------------|-----|-------------------------|-----------|
| Pod > 2-mm bean   | 177 | 2-mm bean > 5-mm bean   | 113 | 5-mm bean > Full-size   | 178       |
| Pod < 2-mm bean   | 185 | 2-mm bean < 5-mm bean   | 163 | 5-mm bean < Full-size   | 189       |
| Total             | 362 | Total                   | 276 | Total                   | 367       |
|                   |     |                         |     |                         | All : 625 |

**Table 2**

Gene titles and the number of genes included in each group. The annotated genes are shown in this table and All 625 genes are listed in Supplemental Table.

|                             |                                                        |   |        |                                                      |   |
|-----------------------------|--------------------------------------------------------|---|--------|------------------------------------------------------|---|
| <b>Group 1</b>              |                                                        |   |        |                                                      |   |
| Storage proteins            | $\beta$ -conglycinin $\alpha$ -, $\alpha'$ -subunit    | 1 |        |                                                      |   |
| <b>Group 2</b>              |                                                        |   |        |                                                      |   |
| Storage proteins            | $\beta$ -conglycinin $\beta$ -subunit                  | 3 | Others | 34 kDa maturing seed vacuolar thiol protease         | 2 |
|                             | Glycinin                                               | 1 |        | Bowman-Birk protease inhibitor                       | 3 |
|                             | Glycinin G1                                            | 2 |        | Metallothionein-II protein                           | 1 |
|                             | Glycinin G2 precursor                                  | 2 |        | Napin-type 2S albumin 1 and 3 precursor              | 2 |
|                             | Glycinin G3                                            | 1 |        | Seed maturation protein PM31                         | 1 |
|                             | Glycinin G4                                            | 2 |        | Sucrose binding protein homolog S-64 (SBP)           | 1 |
| Lipoxygenase                | Lipoxygenase-1                                         | 1 |        |                                                      |   |
|                             | Lipoxygenase-2                                         | 1 |        |                                                      |   |
|                             | Lipoxygenase-3                                         | 2 |        |                                                      |   |
| <b>Group 3</b>              |                                                        |   |        |                                                      |   |
| Lipid storage               | 24 kDa oleosin isoform                                 | 3 | Others | Albumin 1                                            | 1 |
| Oxidoreductase activity     | Microsomal $\omega$ -6 desaturase                      | 2 |        | Allergen Gly m Bd 28K                                | 1 |
| Proteinase inhibitor        | Kunitz trypsin inhibitor                               | 2 |        | Basic 7S globulin isoform                            | 1 |
|                             | Protease inhibitor C-II                                | 1 |        | Class III acidic endochitinase                       | 1 |
| Seed maturation protein     | Seed maturation protein PM28                           | 1 |        | Cysteine proteinase                                  | 1 |
|                             | Seed maturation protein PM30                           | 2 |        | Eu4 (urease)                                         | 1 |
|                             | Seed maturation protein PM34                           | 1 |        | Hydrophobic seed coat protein precursor              | 1 |
|                             | Soybean seed maturation polypeptides                   | 1 |        | Metallothionein-II protein                           | 1 |
| Stress responsive           | Dehydrin                                               | 3 |        |                                                      |   |
|                             | Desiccation protective protein LEA5                    | 1 |        |                                                      |   |
| <b>Group 5</b>              |                                                        |   |        |                                                      |   |
|                             | 28 kDa protein , Haloacid dehalogenase-like hydrolases | 1 |        | Lectin                                               | 1 |
|                             | 31 kDa protein, Haloacid dehalogenase-like hydrolases  | 1 |        | Major latex protein homologue                        | 1 |
|                             |                                                        |   |        | Root nodule bacteria related protein                 | 1 |
| <b>Group 6</b>              |                                                        |   |        |                                                      |   |
| Lipoxygenase                | Lipoxygenase (lox7)                                    | 3 |        | Lipoxygenase L-5                                     | 2 |
| <b>Group 7</b>              |                                                        |   |        |                                                      |   |
|                             | Subtilisin-type protease precursor                     | 1 |        |                                                      |   |
| <b>Group 8</b>              |                                                        |   |        |                                                      |   |
| Protein biosynthetic enzyme | Asparagine synthetase                                  | 1 | Others | 24 kDa seed coat protein                             | 2 |
| Proteinase                  | Aspartic proteinase 4                                  | 1 |        | Acid phosphatase                                     | 1 |
|                             | Cysteine proteinase                                    | 1 |        | Alcohol dehydrogenase 1                              | 1 |
| Oxydation reduction         | Cytochrome P450 monooxygenase                          | 2 |        | Anthocyanidin synthase                               | 1 |
|                             | Dihydroflavonol-4-reductase (DFR2)                     | 1 |        | $\beta$ -1,3-glucanase                               | 1 |
|                             | Dihydroflavonol-4-reductase (DFR1)                     | 1 |        | $\beta$ -amylase                                     | 1 |
|                             | Fe-superoxide dismutase                                | 2 |        | BURP domain-containing protein                       | 2 |
|                             | Isoflavone reductase homolog 1                         | 1 |        | bZIP transcription factor bZIP132                    | 1 |
| Glutathione transferase     | Glutathione S-transferase GST 14                       | 1 |        | Ca <sup>2+</sup> -binding EF hand protein            | 1 |
|                             | Glutathione transferase                                | 1 |        | Chlorophyll a/b-binding protein precursor            | 2 |
| Lipoxygenase                | Lipoxygenase                                           | 2 |        | CCAAT-box binding factor HAP3 B domain               | 1 |
|                             | Lipoxygenase (lox 7)                                   | 1 |        | Cysteine proteinase inhibitor                        | 3 |
|                             | Lipoxygenase L-5                                       | 1 |        | Em protein                                           | 1 |
|                             | Lipoxygenase-9                                         | 2 |        | Endonuclease                                         | 1 |
| Proline-rich protein        | Proline-rich protein                                   | 1 |        | Endo-xyloglucan transferase                          | 1 |
|                             | Repetitive proline-rich protein                        | 2 |        | Gamma glutamyl hydrolase                             | 1 |
| Maturation protein          | LEA protein                                            | 2 |        | Mitotic cyclin                                       | 2 |
|                             | LEA protein, 51 kDa seed Maturation protein            | 1 |        | MYB transcription factor MYB172                      | 1 |
|                             | Putative 18 kDa late embryogenesis abundant protein    | 1 |        | Nodulin                                              | 1 |
|                             | 51 kDa seed maturation protein                         | 1 |        | Peroxidase                                           | 1 |
|                             | Seed maturation protein PM16                           | 1 |        | Plasma membrane Ca <sup>2+</sup> -ATPase             | 1 |
|                             | Seed maturation protein PM3                            | 2 |        | Ribulose-1,5-bisphosphate carboxylase small subunit  | 1 |
|                             | Seed maturation protein PM22                           | 1 |        | Salicylic acid induced protein                       | 1 |
|                             | Seed maturation protein PM26                           | 1 |        | Seed coat BURP domain protein 1                      | 1 |
|                             | Seed maturation protein pPM32                          | 1 |        | Selenium binding protein (sbp gene)                  | 1 |
|                             | Seed maturation protein PM35                           | 1 |        | Sorbitol-like transporter                            | 1 |
|                             | Seed maturation protein PM36                           | 1 |        | Transcription factor LEC1-A                          | 1 |
|                             | Seed maturation protein PM38                           | 1 |        | Thiamin biosynthetic enzyme                          | 3 |
|                             | Seed maturation protein PM40                           | 1 |        | WRKY43 (Salicylic acid related transcription factor) | 1 |
|                             | Seed maturation protein PM41                           | 1 |        | WRKY46 (Salicylic acid related transcription factor) | 1 |

### 3. Discussion

We examined the gene expression profile of developing soybean seeds. The comprehensive analysis revealed the characteristic gene expression profiles of soybean seeds. From pod to full-size bean stage, the numbers of expressing genes were the highest in pod and gradually decreased. To reveal the change of gene expression in detail, DEGs were extracted. 625 were identified as DEGs. Enzymes, especially LOXs, occurred most frequently among the identified DEGs. The LOX is a key enzyme which produces the specific flavor of the soybean seed. Forty five LOX genes are existed in soybean genome. The *LOX1*, *LOX2*, *LOX3*, *LOX5*, *LOX7*, and *LOX9* genes were identified as DEGs in our study. The *LOX1*, *LOX2*, and *LOX3* genes were classified into group 2, showing increased expression from the 2-mm bean stage. They are primarily expressed in soybean seeds, and their expression levels have been shown by 2D-SDS PAGE analysis to increase between bean sizes 2 mm and 10 mm [7]. On the other hand, *LOX5* and *LOX7* are classified to group 6 (expressed in pods and decrease gradually with seed maturity). These have been shown to be expressed mainly in leaves and to decrease in expression in beans of length 2 mm–10 mm [7]. Thus, the expression levels of these LOX proteins correspond to their mRNA levels. The expression levels and patterns of these LOX genes were different each other and they may share the roles accordingly.

Additionally, 3 genes related to lipid metabolism were also included in group 3. They included the 24-kDa isoform of oleosin, which accumulates on the surface of the oil body, a hydrophobic seed protein precursor involved in lipid transport, and a microsomal  $\omega$ -6 desaturase related to fatty acid biosynthesis. The expression patterns of these genes coincide with lipid accumulation, which starts about 15 d after flowering and drastically increases between 30 and 50 d after flowering. The hydrophobic seed protein level increased from early to mid stages of maturation [23]. The expression level of microsomal  $\omega$ -6 desaturase has been reported to increase during the 30–70 d period after anthesis [14]; this is consistent with our results. The expression of the 24-kDa oleosin isoform during seed maturation has been studied by RNA gel blot analysis. Accumulation of oleosin in olive showed a bell-shaped pattern of expression that began 11 weeks after flowering, reached maximum 16–19 weeks after flowering, and declined rapidly thereafter [24]. The oleosin gene expression in *Brassica* was not detectable at the torpedo stage, reached its peak at the beginning of desiccation, and declined rapidly thereafter [25]. Oleosin was thought to be expressed differently in phylogenetically distant species; however, its expression pattern had not been previously analyzed in soybean. In our results, the expression levels of oleosin in soybean increased as the seed grew, as is the case in olive.

Genes encoding certain seed-maturation proteins, which are expressed in the early stages of seed development, were present among the DEGs, as expected. The functions of these molecules remain unclear, but they are observed to regulate metabolism in various stages of seed development.

The expression of *PM28*, *PM30*, *PM31*, *PM34*, and soybean-maturation peptide increased with seed growth. Of these genes, *PM31* and *PM34* have been detected by two-dimensional (2D) electrophoresis during the maturation stage (30–70 d after pollination). This stage spans the time when the bean size is approximately 7 mm to complete desiccation [7]. Further, stress-responsive genes such as *dehydrin* and *LEA 5*, the gene encoding the drought-protective protein, were present in group 3, which showed a gradual increase in expression. It has been reported that dehydrin-like proteins are first detected in full-sized pods and that their accumulation increases during maturation and before dehydration [26]. The desiccation-protective protein gene was sorted into a group with increasing expression pattern during 30–70 d

after pollination by DNA microarray analysis [14]. The mRNA levels of these desiccation-protective genes increased with seed development. Thus, some desiccation-protective protein expression data from previous studies are consistent with our results, even though the sampling period is longer in the previous report. It is interesting that the mRNA levels of desiccation-protective protein increase from the earliest stage of seed development, much before onset of seed dehydration.

In terms of storage proteins, the mRNAs of 7S and 11S globulins were first detected 14–18 d after pollination, several days before the accumulation of the corresponding proteins [8]. The mRNAs of the 7S subunit were first detected at a seed length of 5–6 mm, and expression levels increased gradually until a seed length of 8–10 mm was reached. The mRNA of the 11S glycinin subunit was found to have a similar. Another global gene expression analysis was performed by Jones et al. [22]. They analyzed the gene expression at the stages of bean 25–50 mg, 75–100 mg, 100–200 mg, 400–500 mg, 200–300 mg yellow, and dry seed. They showed glycinin and  $\beta$ -conglycinin were already expressed in the 25–50 mg stage when both glycinin and  $\beta$ -conglycinin were adequately expressed. Compared with their experiments, our sampling stages were earlier than that of Jones. Our data started from the pod stage, so the changes in expression of glycinin and  $\beta$ -conglycinin were clearly detected. Their expressions were slightly lower in early developing stage, eventually reaching its maximum later for seed development.

Our results showed that the levels of  $\beta$ -conglycinin  $\alpha$  and  $\alpha'$ -subunits increased before glycinin subunits, which is consistent with the earlier observations [8].

Soybean seeds contain a variety of proteases and protease inhibitors. In our study, protease inhibitors were included in groups exhibiting increasing expression patterns. The Bowman-Birk protease inhibitor was included in group 2, whereas the Kunitz-type trypsin inhibitor and protease inhibitor C-II were included in group 3. Accumulation of Bowman-Birk protease inhibitor has been observed to begin around 7 weeks after flowering and to increase with seed-maturation [17,19]. Our microarray sample ended at the full-size stage (6–7 weeks after flowering). It is likely that the Bowman-Birk protease inhibitor increased in expression at the full-size stage; therefore, its genes were sorted into group 2. It has been reported that the Bowman-Birk protease inhibitor does not show any major change in expression until 60 d after pollination, but shows a drastic decrease in expression at 70 d [14]. This is because at 70 d after flowering, seed development is nearing its end and the seed starts to dehydrate. Moreover, most of the mRNA activity converges during the dehydration period. Thus, it can be concluded that the expression of the Bowman-Birk protease inhibitor increases around the full-size stage. The Kunitz-type trypsin inhibitor genes were also observed to increase in expression. This observation is in agreement with the results of a previous report that the expression of the Kunitz-type trypsin inhibitor gene appeared 25 d after flowering, reached a maximum at 77 d after flowering, and then decreased and disappeared at 93 d after flowering [27].

The presence of metallothionein-II in groups 2 and 3 is interesting, because it has been reported to be mainly expressed in the leaves and roots of soybeans [28]. In *Arabidopsis thaliana*, the metallothionein-II mRNA showed the highest expression in leaves and seeds, and was induced by metal sulfates [29]. The expression of the metallothionein-II gene in soybean seeds during seed development was reported by Jones et al. [22]. Our present work is coincident with the result in their report. Although there are many genes with unknown functions, we were able to define an overall expression pattern of numerous genes during seed development.

In this study, we were able to characterize gene expression patterns in developing soybean seeds by DNA microarray analysis.

Seed size varied across growth stages, and the clusters were created according to bean size. We identified DEGs showing increasing expression during development, such as those encoding the storage protein, seed-maturation protein, and protease inhibitor. The number of DEGs showing decreasing expression with development was smaller than that of those with increasing expression. We identified LOXs in both groups with increasing and decreasing expression, indicating their multi-functionality and suggesting stage-specific expression patterns for these proteins. Some new genes that showed expression changes during seed development, such as the 24-kDa oleosin isoform. Thus, DNA microarray analysis enabled us to observe the overall dynamics of gene expression in tissues. Compared with gene expression among neighboring stages by Rank product function, we could extract new genes with different expression patterns in seed developing stage. Recently next generation sequencing is used to reveal the global expression of genes [30]. This new technique may contribute to show RNA-Seq atlas. DNA microarray analysis and new generation sequencing will adapt to other plant seeds for characterizing stages of development. These methods are robust tool for revealing stage specific expression genes.

## 4. Material and methods

### 4.1. Plants

Soybean (*Glycine max* L.) cv. Tamahomare was cultivated at the Nagano Vegetable and Ornamental Crops Experiment Station (formerly, Nagano Chusin Agricultural Experiment Station) in Shiojiri Nagano, Japan. Samples were collected when the pods or beans grew to a specific stage or size. Pod samples were collected as the pod started to elongate, 1 week after flowering. Bean samples were collected when the beans were 2 mm and 5 mm in length; the latter stage was reached about 2 weeks after flowering. Full-size samples were collected when the soybean was at its biggest size, usually around 12–14 mm in length, before the onset of dehydration. This stage was reached approximately 6–7 weeks after flowering. All of the samples were stored at  $-80^{\circ}\text{C}$  immediately after collection and then transported to our laboratory at the University of Tokyo.

### 4.2. RNA extraction

The full-size beans were collected from different plants. Two beans of full-size were subjected to DNA microarray ( $n = 2$ ). Four or five seeds of 2-mm or 5-mm bean were collected from different plants and used for a sample of DNA microarray analysis ( $n = 2$ ). Two pods from different plants were used for a microarray sample. An RNeasy Plant Mini Kit (Qiagen, Germantown, MD) was used to extract and purify mRNA from the samples. Representative images of each sample are shown in Fig. 2A.

### 4.3. DNA microarray analysis

The samples for DNA microarray analysis were prepared and analyzed by the method described before (Narikawa et al., 2010). All microarray data were submitted to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>, GEO Series ID GSE26443). Affymetrix GCOS software was used to convert the array images with the intensity of each probe to CEL files. The CEL files were read into the personal computer and analyzed using the statistical language R and Bioconductor software. First, the files were normalized using MAS5 and then executed mas5calls by Bioconductor to make Present/Absent calls. Next the files were normalized using the Distribution Free Weighted method (DFW)

algorithm [31]. The 'pvclust' function was then used to perform the clustering analysis of 8 samples.

To investigate the expression patterns of genes, adjacent stages were compared using the 'Rank products' function. Genes with false discovery rate (FDR)  $< 0.05$  were extracted and then treated with the 'hclust' function and were divided according to their expression patterns. In order to obtain information on the identified genes, the annotation file for the Soybean Genome Array was downloaded from Affymetrix website (NetAffx) and SoyBase website (<http://soybase.org/AffyChip/>).

## Appendix. Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.plaphy.2011.12.007.

## References

- [1] R. Helm, G. Cockrell, E. Herman, A. Burks, H. Sampson, G. Bannon, Cellular and molecular characterization of a major soybean allergen, *Int. Arch. Allergy Immunol.* 117 (1998) 29–37.
- [2] R.M. Helm, G. Cockrell, C. Connaughton, C.M. West, E. Herman, H.A. Sampson, G.A. Bannon, A.W. Burks, Mutational analysis of the IgE-binding epitopes of P34/Gly m Bd 30K, *J. Allergy Clin. Immunol.* 105 (2000) 378–384.
- [3] T. Ogawa, N. Bando, H. Tsuji, H. Okajima, K. Nishikawa, K. Sasaoka, Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis, *J. Nutr. Sci. Vitaminol.* 37 (1991) 555–565 Tokyo.
- [4] X. Gu, T. Beardslee, M. Zeece, G. Sarath, J. Markwell, Identification of IgE-binding proteins in soy lecithin, *Int. Arch. Allergy Immunol.* 126 (2001) 218–225.
- [5] L.A. Moroz, W.H. Yang, Kunitz soybean trypsin inhibitor: a specific allergen in food anaphylaxis, *N. Engl. J. Med.* 302 (1980) 1126–1128.
- [6] R.B. Goldberg, S.J. Barker, L.G. Perez, Regulation of gene expression during plant embryogenesis, *Cell* 56 (1989) 149–160.
- [7] M. Hajdich, A. Ganapathy, J.W. Stein, J.J. Thelen, A systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles, and an interactive proteome database, *Plant Physiol.* 137 (2005) 1397–1419.
- [8] D.W. Meinke, J. Chen, R.N. Beachy, Expression of storage-protein genes during soybean seed development, *Planta* 153 (1981) 130–139.
- [9] B.A. Martin, R.W. Rinne, A comparison of oleic acid metabolism in the soybean (*Glycine max* [L.] Merr.) genotypes Williams and A5, a mutant with decreased linoleic acid in the seed, *Plant Physiol.* 81 (1986) 41–44.
- [10] Y. Mochizuki, M. Maebuchi, M. Kohno, M. Hirotsuka, H. Wadahama, T. Moriyama, T. Kawada, R. Urade, Changes in lipid metabolism by soy beta-conglycinin-derived peptides in HepG2 cells, *J. Agric. Food Chem.* 57 (2009) 1473–1480.
- [11] N.K. Sangwan, K. Gupta, K.S. Dhindsa, Fatty acid composition of developing soybeans, *J. Agric. Food Chem.* 34 (1986) 415–417.
- [12] L. Wang, T. Wang, W.R. Fehr, Effect of seed development stage on sphingolipid and phospholipid contents in soybean seeds, *J. Agric. Food Chem.* 54 (2006) 7812–7816.
- [13] S. Dhaubhadel, B.D. McGarvey, R. Williams, M. Gijzen, Isoflavonoid biosynthesis and accumulation in developing soybean seeds, *Plant Mol. Biol.* 53 (2003) 733–743.
- [14] S. Dhaubhadel, M. Gijzen, P. Moy, M. Farhangkhomeh, Transcriptome analysis reveals a critical role of CHS7 and CHS8 genes for isoflavonoid synthesis in soybean seeds, *Plant Physiol.* 143 (2007) 326–338.
- [15] V. Beilinson, O.V. Moskalenko, D.S. Livingstone, S.V. Reverdatto, R. Jung, N.C. Nielsen, Two subtilisin-like proteases from soybean, *Physiol. Plantarum.* 115 (2002) 585–597.
- [16] D. Gruis, J. Schulze, R. Jung, Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases, *Plant Cell* 16 (2004) 270–290.
- [17] B.J. Biermann, J.S. Banzie, J. Handelsman, J.F. Thompson, J.T. Madison, Methionine and sulfate increase a Bowman-Birk-type protease inhibitor and its messenger RNA in soybeans, *J. Agric. Food Chem.* 46 (1998) 2858–2862.
- [18] K. Liu, P. Markakis, Effect of maturity and processing on the trypsin inhibitor and oligosaccharides of soybeans, *J. Food Sci.* 52 (1987) 222–223.
- [19] J.H. Park, H.J. Jeong, B.O. Lumen, Contents and bioactivities of lunasin, Bowman-Birk inhibitor, and isoflavones in soybean seed, *J. Agric. Food Chem.* 53 (2005) 7686–7690.
- [20] L.G. Perez-Grau, R.B. Goldberg, Soybean seed protein genes are regulated spatially during embryogenesis, *Plant Cell* 1 (1989) 1095–1109.
- [21] T. Narikawa, T. Tamura, K. Yagasaki, K. Terauchi, K. Sanmiya, Y. Ishimaru, K. Abe, T. Asakura, Expression of the stress-related genes for glutathione S-transferase and ascorbate peroxidase in the most-glycinin-deficient soybean cultivar Tousan205 during seed maturation, *Biosci. Biotechnol. Biochem.* 74 (2010) 1976–1979.

- [22] S.I. Jones, D.O. Gonzalez, L.O. Vodkin, Flux of transcript patterns during soybean seed development, *BMC Genomics* 11 (2010) 136.
- [23] M. Gijzen, S.S. Miller, K. Kuflu, R.I. Buzzell, B.L. Miki, Hydrophobic protein synthesized in the pod endocarp adheres to the seed surface, *Plant Physiol.* 120 (1999) 951–959.
- [24] K. Giannoulia, G. Banilas, P. Hatzopoulos, Oleosin gene expression in olive, *J. Plant Physiol.* 164 (2007) 104–107.
- [25] S.P. Slocombe, I. Cummins, R.P. Jarvis, D.J. Murphy, Nucleotide sequence and temporal regulation of a seed-specific *Brassica napus* cDNA encoding a stearoyl-acyl carrier protein (ACP) desaturase, *Plant Mol. Biol.* 20 (1992) 151–155.
- [26] N.H. Samarah, R.E. Mullen, S.R. Cianzio, P. Scott, Dehydrin-like proteins in soybean seeds in response to drought stress during seed filling, *Crop Sci.* 46 (2006) 2141–2150.
- [27] K.D. Jofuku, R.B. Goldberg, Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants, *Plant Cell* 1 (1989) 1079–1093.
- [28] I. Kawashima, Y. Inokuchi, M. Chino, M. Kimura, N. Shimizu, Isolation of a gene for a metallothionein-like protein from soybean, *Plant Cell Physiol.* 32 (1991) 913–916.
- [29] J. Zhou, P.B. Goldsbrough, Functional homologs of fungal metallothionein genes from *Arabidopsis*, *Plant Cell* 6 (1994) 875–884.
- [30] A.J. Severin, J.L. Woody, Y.T. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Specht, M.A. Graham, S.B. Cannon, G.D. May, C.P. Vance, R.C. Shoemaker, RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome, *Plant Biol.* 10 (2010) 1471–2229.
- [31] Q. Liu, R.H. Scheuermann, A distribution free summarization method for Affymetrix genechip arrays, *Bioinformatics* 23 (2007) 321–327.