

**PLANTPHYSIOL/2008/125633 revision**

**Running head:** Dynamic Proteomic Analysis of Filling Rice Grains

Corresponding author: Tai Wang

Research Center of Molecular and Developmental Biology, Key Laboratory of Photosynthesis and Environmental Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

Phone & Fax: +86-10-6259417

[twang@ibcas.ac.cn](mailto:twang@ibcas.ac.cn)

**Keywords:** *Oryza sativa*, proteomics, seed filling, starch, central carbon metabolism, alcoholic fermentation

# **Dynamic Proteomic Analysis Reveals a Switch Between Central Carbon Metabolism and Alcoholic Fermentation in *Oryza sativa* Filling Grains<sup>1</sup>**

Sheng Bao Xu, Tang Li, Zhu Yun Deng, Kang Chong, Yongbiao Xue and Tai Wang

Research Center of Molecular and Developmental Biology, Key Laboratory of Photosynthesis and Environmental Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China (S.B.X., T.L., Z.Y.D., K.C., T.W.); Key Laboratory of Molecular and Developmental Biology, Institute of Genetics and Developmental Biology, Beijing 10010, China (Y.B); National Center for Plant Gene Research, Beijing 100093, China (K.C., T.W.); Graduate School of Chinese Academy of Sciences, Beijing 100049, China (S.B.X., T.L.)

**Footnotes:**

<sup>1</sup>This work was supported by Chinese Ministry of Sciences and Technology (2006CB910105) and Chinese Academy of Sciences.

Corresponding author; email: [twang@ibcas.ac.cn](mailto:twang@ibcas.ac.cn); Fax:0086-10-62594170

## **Abstract**

Accumulation of reserve materials in filling grains involves coordination of different metabolic and cellular processes, and understanding the molecular mechanisms underlying the interconnections remains a major challenge for proteomics. Rice is an excellent model for studying grain filling because of its importance as a staple food and the available genome sequence database. Our observations showed that embryo differentiation and endosperm cellularization in developing rice seeds were completed approximately 6 days after flowering (DAF); and thereafter, the immature seeds mainly underwent cell enlargement and reached the size of mature seeds at 12 DAF. Grain filling began at 6 DAF and lasted until 20 DAF. Dynamic proteomic analyses revealed 396 protein spots differentially expressed throughout 8 sequential developmental stages from 6 to 20 DAF, and determined 345 identities. These proteins were involved in different cellular and metabolic processes with a prominently functional skew toward metabolism (45%) and protein synthesis/destination (20%). Expression analyses of protein groups associated with different functional categories/subcategories showed that substantially up-regulated proteins were involved in starch synthesis and alcoholic fermentation; whereas the proteins down-regulated in the process involved in central carbon metabolism and most of the other functional categories/subcategories such as cell growth/division, protein synthesis, proteolysis and signal transduction. The coordinated changes were consistent with the transition from cell growth and differentiation to starch synthesis and clearly indicated that a switch from central carbon metabolism to alcoholic fermentation may be important for starch synthesis and accumulation in the developmental process.

## INTRODUCTION

Seed development is triggered by a double fertilization process specific to plants; after double fertilization, the fertilized egg cell develops into the embryo, and the fertilized polar nuclei develop into the endosperm (Goldberg et al., 1994). In dicotyledons, the endosperm is absorbed by the embryo during development, and reserve materials are stored in embryonic cotyledons (Goldberg et al., 1994; Le et al., 2007). However, in monocots such as cereal crops, the endosperm represents the main part of the mature seed and is an important organ for reserve storage (James et al., 2003). The cereal seed (also called caryopsis) consists of the embryo, endosperm and pericarp; the outer-most endosperm cell layer differentiates into aleurone. Although seeds from different species are diverse in form, they have one common characteristic: accumulation of reserves during development, except for differences in reserve composition, such as approximately 85% of seed dry weight being starch in cereal seeds, 50% to 70% fatty acids in oilseeds and 40% proteins in soybean seeds (Ruuska et al., 2002). The reserve materials are not only essential for postembryonic growth and development by nourishing germinated embryos before the seedlings start photosynthesis, but also make seeds an important food source for humans and livestock. Therefore, the mechanism underlying the accumulation of reserves during seed development has become a hot topic and is a key scaffold for the molecular control of yield and quality of crops; however, this mechanism still remains largely unknown (Tetlow, 2006; Gutierrez et al., 2007).

Rice (*Oryza sativa* L.), one of the most important cereal crops, is the staple food for half of the world's population and has been used as an excellent model plant after Arabidopsis because of its relatively smaller genome and the completion of genome sequence, which is important for acquiring knowledge about the mechanisms of seed development and starch accumulation. Several studies have documented the cellular and morphological features of developing rice seeds (Berger, 1999; Ishimaru et al., 2003). After flowering, the fertilized egg cell undergoes fast cell division and differentiation, which leads to the formation of embryo. The embryo reaches maturity at approximately 10 days after flowering (DAF). For endosperm development, the fertilized polar nuclei undergo numerous cycles of mitosis without cellularization until 3 DAF. The cellularization of endosperm begins at 3 DAF and ceases at 6-7 DAF. As cells enlarge, the seed reaches full size at 11 to 15 DAF. During the developmental process, starch is found in pericarps from 1 to 3 DAF and begins to accumulate in endosperms from 5 DAF. Afterward,

active starch accumulation in the endosperm maintains until 20 DAF. At approximately 20 DAF, the seed enters the desiccation phase (Ishimaru et al., 2003). These findings taken together with the observations about cellular and physiological changes in developing seeds of other species (Olsen, 2001), suggest that coordinated cellular and metabolic changes that occur timely between different developmental events are related to accumulation of reserves, but the underlying mechanisms are still unknown.

Numerous studies have provided several insights into the mechanism of reserve accumulation. Mutant and transgenic analyses have identified important enzymes essential for starch synthesis and quality (James et al., 2003; Tetlow, 2006). Expression analyses by microarray and cDNA libraries in *Arabidopsis* (Girke et al., 2000; Ruuska et al., 2002), wheat (Leader, 2005) and maize (Lai et al., 2004; Verza et al., 2005) lead to identification of a large number of genes preferentially expressed in seeds or their compartments. However, increasing data have shown that in many organisms, a large proportion of proteins have faint correlation with their corresponding mRNAs in expression profiles (Greenbaum et al., 2003; Watson et al., 2003; Schmidt et al., 2007). In the *Arabidopsis* mutant *wril* which shows 80% reduction in storage reserves, a major alteration in the metabolism is accompanied by surprisingly few changes in gene expression (Ruuska et al., 2002). This lack of association between mRNA and protein (metabolism) levels indicates the importance of posttranscriptional control in regulating amounts of enzymes and metabolic fluxes through key pathways. Amylopectin, a major constituent of starch, is synthesized by the coordinated actions of AGP-glucose pyrophosphorylase (AGPase), starch synthase, starch branching enzymes and starch debranching enzymes; several types of these enzymes have multiple isoforms, and each isoform plays a distinct role in amylopectin synthesis (Tetlow, 2006). Diverse modifications are involved in regulating the activity of these enzymes (James et al., 2003; Tetlow, 2006).

Proteomic approaches based on 2-D electrophoresis (2-DE) and mass spectrometry (MS) have supplied powerful solutions for the identification of dynamic expression profiles of proteins and isoforms during development. In *Medicago truncatula*, 84 proteins differing in kinetics of expression during seed development have been identified (Gallardo et al., 2003). The proteomic analyses of filling seeds of soybean and *Brassica napus* lead to the identification of 216 and 289 unique proteins, respectively, which are differentially expressed, and reveal the properties of coordinated expression changes in proteins associated with fatty acid synthesis (Hajduch et al.,

2005; Hajduch et al., 2006). Mechin et al. identify 302 proteins differentially expressed in developing maize endosperm (Mechin et al., 2007). The changes in expression of these proteins are consistent with the important developmental shift from cellularization, cell division and cell wall deposition to storage accumulation during development (Mechin et al., 2007).

In an effort to understand the molecular regulation and metabolic network of starch synthesis and accumulation during seed development, we analyzed dynamic changes of protein expression profiles in rice during 8 sequential developmental stages associated with grain filling, from 6 to 20 DAF, and revealed five expression patterns (clusters) of differentially expressed proteins. To evaluate the dynamic functional features of categories/subcategories in the developmental process, we tried a new method, digital expression tendency, to analyze the tendency for change in expression among protein groups associated with different categories/subcategories. The cluster patterns combined with expression profiles of protein groups involved in different functional categories/subcategories clearly indicated that the switch from central carbon metabolism to alcoholic fermentation is associated with starch synthesis and accumulation. These results provide novel clues for further understanding the metabolic network involved in starch accumulation in developing seeds.

## **RESULTS**

### **Characterization of developing rice seeds**

In rice, developing seeds (also called caryopses) are classified as superior or inferior according to their location on spikes (Ishimaru et al., 2003). The superior seeds, located at the top of the spike, have a higher growth rate and uniformly reach maturity. Therefore, in this study, we selected superior seeds as experimental samples because of their distinguishing advantage of synchronous development.

To obtain basic information about rice seed development, we observed morphological features and dynamic changes of reserve accumulation in developing seeds at 2, 4, 6, 8, 10, 12, 14, 16, 18 and 20 DAF (Fig. 1, A~E). The developing seeds greatly increased in size from 2 to 8 DAF, and then had a slight increase and appeared to reach the size of mature seeds at 12 DAF (Fig. 1, A~B). Seeds after 18 DAF became translucent (Fig. 1A). In contrast, both fresh and dry weight appear to change insignificantly from 2 to 6 DAF (data not shown) but quickly increased thereafter until 18 DAF (Fig. 1E). After 18 DAF, the increase in fresh weight slowed, but the

dry weight kept increasing until 20 DAF, which indicates that developing seeds enter into the desiccation phase from 18 DAF. In general, the developmental changes in seed size, fresh and dry weight were consistent with previous observations (Ishimaru et al., 2003).

Further observation of embryos and endosperms revealed embryos entered into the globular stage before or at 4 DAF and became heart shaped from 6 to 8 DAF, when embryonic buds and roots have differentiated (Fig. 1D). The embryo showed no apparent changes in dimensions from 8 DAF (Fig. 1D). The length of endosperm at 6 DAF appeared to be close to that of the mature endosperm (Fig.1, A~B), which is in line with a report describing endosperm cells in longitudinal direction fixed around 5 DAF (Ishimaru et al., 2003). These results, in combination with the observation that the number of endosperm cell layers in the transverse direction peaked at approximately 7 DAF in rice (Ishimaru et al., 2003), suggest that cell division and differentiation in developing rice seeds occur mainly before 8 DAF. In addition, we found no starch accumulation in endosperm at 2 DAF (Fig. 1C), a small amount at 4 DAF and slightly more at 6 DAF, when starch was also observed in the pericarp (Fig. 1B). The endosperm accumulated a great amount of starch after 8 DAF, when no starch was observed in the pericarp (Fig. 1B). Taken together, these results indicate that developing seeds until 6 DAF mainly undergo active cell division and differentiation and begin grain filling at 6 DAF until 20 DAF. Thus, we tentatively divided the development process from 6 to 20 DAF into early (6 to 8 DAF), mid (8 to 12 DAF) and late (12 to 20 DAF) stages. In order to identify change in energy requirement in the process, we determined dynamic change of ATP levels in seeds. The results showed that the ATP level increased from 6 DAF until 12 DAF, and thereafter decreased dramatically (Fig. 1F), which suggested the active starch synthesis consumed less ATP than early cell enlargement. The objective of this research was to focus on characterizing protein expression profiles related to grain filling; therefore, we used the developing seeds at 6, 8, 10, 12, 14, 16, 18 and 20 DAF for further study.

### **Identification of differentially expressed proteins**

To better solve the protein expression profiles of developing rice seeds by 2-DE, we first compared protein expression patterns obtained by 2-DE with pH 3-10 and 4-7 gel strips. For example, 2-DE separation of 18 DAF seeds with pH 3-10 gel strips resolved 936 ( $\pm 42$ ) ( $n=3$ ) CBB-stained protein spots (Supplemental Fig. S1A). Among these spots, 759 ( $\pm 37$ ) were present in the sub-range of pH 4-7 and accounted for more than 90% of the total volume of all spots



detected in the pH 3-10 gel. This finding suggested that most of the proteins in rice seeds distributed around pH 4-7. 2-DE separation with pH 4-7 gel strips resolved  $1056 \pm 75$  ( $n=3$ ) CBB-stained protein spots (Fig. S1B). This situation is similar to that found in rice pollen proteomics analysis (Dai et al., 2006; Dai et al., 2007). Therefore, we used 2-DE with pH 4-7 gel strips to obtain protein expression profiles of developing rice seeds.

2-DE separation in the pH 4-7 range resolved more than 1000 protein spots from developing seeds at 6, 8, 10, 12, 14, 16, 18 and 20 DAF (Fig. 2A). Some weak spots with low relative volume (RV) in 2-DE gels are usually highly variable in different samples, even in different biological replicates of the same sample; the variation affects identification of differentially expressed proteins throughout multiple developmental stages or treatments (Vienna et al., 2000; Lewis and Currie, 2003). Therefore, to identify differentially expressed proteins, we first established a group of proteins comparable throughout the 8 distinct development stages by statistical analysis. The analysis involved evaluation of reproducible spots in triplicate biological repeats of each sample and qualification of these proteins present in at least two distinct samples by comparing the reproducible spots in each sample one by one (also see Materials and Methods). With these analyses, we established a group of 539 spots. By further quantitative and comparative analysis of the 539 spots among the 8 distinct stages, we found 396 spots groups with at least 2-fold change in expression (Supplemental Table S1).

### **Characteristics of differentially expressed proteins**

Our MS analyses under a stringent standard of a MOWSE score of more than 65 ( $P < 0.01$ ) led to identification of 309 spots (Fig.S2). Among them, 275 contained a single protein each, and the remaining 34 had 2 to 3 proteins each (32 spots with 2 proteins and 2 with 3 proteins each). In total, we obtained 345 identities representing 227 unique proteins (Table S2). According to a part and/or an instance of the parent of gene ontology term obtained for each protein, we classified these proteins into 9 functional categories: metabolism, protein synthesis/destination, defense response, cell growth/division, signal transduction, photosynthesis, transcription, intracellular traffic, and transporter (Fig. 3, Table S2). Proteins without gene ontology terms in the present database and those could not be classified into the above 9 categories were assigned as “unknown”. Interestingly, 13 identities were found to be pyruvate orthophosphate dikinases (PPDK). The protein is classically involved in C4 photosynthesis, and recently found abundant in developing seeds of cereal crops such as rice and maize (Kang et al., 2005), but the function of

PPDK in seed development remains to be elucidated. Therefore, PPDK proteins were organized as an independent category (Fig. 3).

The analysis revealed that 65% of the 345 identities were implicated in two functional groups: metabolism (45%) and protein synthesis/destination (20%), and the remaining 35% related to other 9 groups (Fig. 3, Table S2). This finding suggested the functional importance of metabolism and protein synthesis/destination in seed filling and development. To analyze dynamic changes among different processes of metabolism and protein synthesis/destination, the proteins involved in metabolism were further grouped into 11 sub-categories, including sugar conversion, glycolysis, alcoholic fermentation, pentose phosphate pathway, tricarboxylic acid (TCA) cycle, starch synthesis, amino acid metabolism, nitrogen/sulfur metabolism, nucleotide metabolism, lipid/sterol metabolism, and secondary metabolism; those related to protein synthesis/destination were grouped into 3 sub-categories: protein synthesis, protein folding/modification, and proteolysis (Table S2).

Real-time quantitative RT-PCR (qRT-PCR) analysis was employed to evaluate expression of genes. In total, 10 of the identified proteins-encoding genes were selected including 5 starch synthesis-related proteins (each with 2 to 4 isoforms), 4 glycolysis proteins and 1 tubulin protein (Table S3). In the genes, 6 showed dynamic change in mRNA levels closed (Os08g25720; OsO4g31700; Os04g08270) to or similar (Os06g46000; Os03g524600; Os03g55090) that of corresponding proteins; 4 had variable expression patterns between mRNA and protein (Os05g33380; Os08g40930; Os10g11140; Os01g05490). The results appeared be comparable with observations that about 60% protein-mRNA pairs show concordant expression (Cox et al., 2005) and thus indicate importance of our proteomic results.

### **Hierarchical clustering analysis reveals five expression patterns**

To study the expression characteristics of proteins involved in each functional category (sub-group) during seed development, we performed hierarchical clustering analysis of 275 identities that excluded spots (34) with more than one identity (Table S4). The analysis revealed five hierarchical clusters (c0, c1, c2, c3 and c4) (Table 1). The largest cluster was c0, of 76 proteins whose expression was at highest level at 6 DAF, decreased greatly at 8 DAF and remained low thereafter. The second largest clusters were c1 (67) and c4 (63). The expression of proteins in c1 changed similarly to that of proteins in c0 but showed a gradually decrease in level from a maximum at 6 DAF to a minimum at 20 DAF, whereas proteins in c4 were up-regulated from 6

to 20 DAF. Cluster c2 was composed of 35 proteins that began to up-regulate at 6 DAF, peaked in level at 10 DAF and decreased thereafter. Cluster c3 consisted of 34 proteins whose expression level began to increase at 6 DAF, peaked at 16 DAF and decreased thereafter.

Different category-/sub-category-associated proteins showed heterogeneous distribution in these clusters (Table 1). For example, most proteins involved in cell growth/division (10/13) and photosynthesis (9/11) were in c0 and c1. Protein synthesis-related proteins were distributed in c0, c1 and c2 (13/13), whereas proteolysis-related proteins were mainly in c0 and c1 (16/19). Most of the starch-synthesis-related proteins were in c4 (19/23). Glycolysis proteins mainly distributed in c0 (11/21), whereas alcoholic fermentation proteins appeared in c3 and c4 (7/9). These changes suggested switches in metabolic and/or biological processes during development from 6 to 20 DAF.

### **Composite expression profiles of the functional categories and sub-categories**

Dynamic proteomic study of developmental processes reveals temporal changes in expression levels of protein related to metabolism/cellular processes and provides important clues for further understanding the potential relation between the temporal changes in expression and the developmental events. Currently, temporal changes in expression are usually organized by composite expression profiles established with normalized total RVs of a protein group involved in a given metabolism/cellular process (Hajduch et al., 2005; Hajduch et al., 2006; Mechin et al., 2007). To evaluate the changes in expression patterns of each protein group, we analyzed composite expression profiles using the 275 spots identified as a single identity each. In the 8 analyzed categories (Fig. 4), the expression of proteins involved in metabolism and protein synthesis/destination appeared to change little during development; that of defense response-related proteins was increased up to 20 DAF; and that of PPDK proteins began to increase at 6 DAF, peaked from 8 to 14 DAF and decreased thereafter. The expression of proteins involved in cell growth/division, signal transduction, transcription and photosynthesis was decreased.

Because most of the differentially expressed proteins were involved in metabolism and protein synthesis/destination (Fig.3), we further analyzed composite expression profiles of their subgroups (Fig. 4). Among the subcategories of metabolism, the expression of starch synthesis-related proteins was increased beginning at 6 DAF, peaked at 18 DAF and slightly decreased thereafter; proteins involved in alcoholic fermentation showed little change in expression from 6 to 12 DAF and were up-regulated thereafter; proteins of the other sub-groups showed a tendency

to decrease in level (sugar conversion, glycolysis, nitrogen/sulfur metabolism), or peaked at early stage(s) (TCA cycle, lipid/sterol metabolism), or showed little change in level (amino acid metabolism, secondary metabolism) during development. In the protein synthesis/destination group, proteins involved in protein synthesis and proteolysis were decreased in level, whereas those associated with protein folding/modification were increased in level.

### **Digital expression profiles of the functional categories and sub-categories**

Although the composition profile analysis revealed the dynamic changes in abundance of protein groups, because of heterogeneity in expression pattern and/or normalized RVs among components of the same protein group, such as one or several proteins having preponderantly high RVs than most of the others, with a reverse slope in expression change, the composite expression profile of the protein group is usually represented by a few proteins with high expression level. Therefore, it is difficult to evaluate the expression changes of functional categories only by analyzing their composition profiles. To eliminate this disadvantage, here we tried to establish the expression tendency of a protein group, termed digital expression profiles in this paper (Fig. 5), by considering the up- or down-regulation feature of each protein in a given protein group (for details, see Materials and Methods). This analysis revealed 4 distinct expression patterns: (1) proteins preferentially expressed at 6 DAF, then down-regulated to a relatively constant level (proteins for metabolism and its subcategories sugar conversion, glycolysis and secondary metabolism); (2) proteins up-regulated (TCA cycle) or down-regulated (nucleotide metabolism, and protein folding/modification) at the early mid-developmental phase; (3) proteins whose expression showed a positive slope to advancing development (starch synthesis and alcoholic fermentation); and (4) proteins whose expression showed a negative slope (the remaining categories/subcategories such as protein synthesis/destination, proteolysis, cell growth/division and transcription).

To evaluate the applicability of the above method in analyzing expression change tendency of a given protein group during development, we analyzed the correlation between the digital profile and composite profile. In general, digital profiles were significantly correlated with corresponding composite profiles. The two types of profiles showed significant positive correlation for cell growth/division ( $r=0.922$ ,  $p<0.01$ ), signal transduction ( $r=0.941$ ,  $p<0.01$ ), transcription ( $r=0.862$ ,  $p<0.01$ ) and photosynthesis ( $r=0.936$ ,  $p<0.01$ ), and subcategories sugar conversion ( $r=0.984$ ,  $p<0.01$ ), glycolysis ( $r=0.979$ ,  $p<0.01$ ), TCA cycle ( $r=0.971$ ,  $p<0.01$ ), starch

synthesis ( $r=0.935$ ,  $p<0.01$ ), alcoholic fermentation ( $r=0.809$ ,  $p<0.05$ ), secondary metabolism ( $r=0.833$ ,  $p<0.05$ ), protein synthesis ( $r=0.964$ ,  $p<0.01$ ) and proteolysis ( $r=0.997$ ,  $p<0.01$ ). However, the positive correlation was not significant for the protein synthesis/destination, PPK, amino acid metabolism, nitrogen/sulfur metabolism, nucleotide metabolism, lipid/sterol metabolism, and protein folding/modification. The correlation was significantly negative for the metabolism ( $r=-0.831$ ) and defense responses ( $r=-0.882$ ), ( $p<0.05$ ). Further analysis revealed that, as discussed above, the inconsistency resulted from the presence of several highly abundant proteins whose expression profiles were contrary to most other proteins in the same category, such as starch synthesis-related proteins in metabolism, alpha-amylase inhibitor protein (spot 2296) in defense response, saposin-like type B protein (spot 2269) in lipid/sterol metabolism, Dnak-type molecular chaperone Bip (spots 1399, 1370) in protein folding/modification (Table S4), and two isoforms (spots 1196 and 1257) of DDPK. Their predominant accumulation throughout development or at some points in development caused deviation of composite expression profiles of the corresponding categories (Table S4). However, our analysis showed that the digital expression profiles explained relatively well the heterogeneous distribution of proteins related to distinct categories/subcategories in different clusters (Table 1, Fig. 5). Thus, these data suggest that digital expression profiles better reflect a real tendency for changes in expression of a protein group involved in a given category/subcategory than the currently used composite expression profiles.

In addition, starch synthesis is a main functional feature of cereal seed development; speedy accumulation of highly abundant starch synthesis-related proteins and alpha-amylase inhibitor proteins is consistent with the requirement of starch synthesis. Therefore, a comparison between the two types of expression profiles is helpful to reveal some proteins causing deviation between the two profiles of a given category/subcategory. These proteins may be important because of their high accumulation.

### **Expression characteristics of isoforms during rice seed development**

In general, a gene produces isoforms by alternative splicing of transcripts and/or posttranslational modification (PTM). In plants, isoforms have been identified in proteomes of various tissues (Dai et al., 2006; Hajduch et al., 2006; Mechin et al., 2007). In this study, we revealed 66 unique proteins ( $66/227=29\%$ ) with 184 identities (Table S2), ranging from 2 (39/66) to 13 identities (PPDK). The theoretical and experimental molecular masses (MMs) of these

identities were correlated ( $Y=0.7195x + 12118$ ,  $R^2=0.6589$ , Fig. S3A). The theoretical and experimental pI values showed a low but significant coefficient of determination ( $y=0.1964x + 4.3699$ ,  $R^2=0.13$ ,  $p<0.01$ ; Fig. S3B). The low coefficient of determination may result from 26 proteins with pI values of more than 7, even though proteins were separated by 2-DE gel with a pH 4-7 range. When these 26 proteins were excluded from the analysis, the correlation between theoretical and experimental pI values of the remaining proteins increased significantly ( $y=0.6488x + 1.8267$ ,  $R^2=0.4537$ ; Fig. S3C). Further analysis showed that the MM range of most of the 26 proteins was less than 20 kDa, indicating that a protein less than 20 kDa was more likely to change its pI value (Fig. S3D). Therefore, a modification could possibly cause a more obvious pI change for low MM proteins than for high MM proteins. Taken together, these data indicate that multiple identities of a unique protein might be PTM-generated isoforms.

Currently, the biological importance of gene-generated multiple isoforms is not fully understood. Here, we analyzed the correlation of expression profiles among isoforms of unique proteins. Spots containing more than one protein were excluded because of the difficulty in judging which protein in one spot was changed in expression. Finally, 124 identities representing 50 unique proteins were suitable for analysis (Table S5). Results showed that 52 of the 117 isoform pairs were significantly correlated in expression profile ( $P<0.05$ ) (Table S5). Most (39) of the 52 isoform pairs showed significant positive correlation, and only 13 showed negative correlation, analogous to results reported in maize (Mechin et al., 2007).

Further analysis revealed that approximately 50% of the positively correlated isoform pairs involved 5 of the 8 analyzed categories/subcategories, including starch synthesis, protein folding/modification, defense response, glycolysis and PPK (Fig. 6, Table S5). Interestingly, the expression profiles of the isoforms of amyloplast-localized proteins (isoamylase, plastidic phosphoglucomutase, pullulanase and alpha 1, 4-glucan phosphorylase, ADPGase) showed positive correlations, and most were significant. However, the isoforms of cytoplasmic ADPGase, and those involved in sugar conversion, such as sucrose synthase 3 and phosphoglucomutase, displayed remarkably negative correlation.

## DISCUSSION

In combination with previous observations (Ishimaru et al., 2003) of cellular and morphological features of developing rice seeds, we showed that the developmental process of seeds from 6

DAF represents the main events associated with the ability of seeds to fill. Further study of differentially expressed proteins identified 309 protein spots with changed expression during the process, of which 275 showed a single identity. Expression profile analysis of these identities revealed key molecular characteristics of the developmental process and supplied important clues about crucial proteins and their co-regulation in specific developmental phases.

### **Developing seeds display different proteomic characteristics at distinct stages**

The early phase (6-8 DAF) of seed development mainly involved active cell enlargement, leading to a rapid increase in seed size available for further accumulation of starch. This phase was characterized by prominent accumulation of proteins involved in cell growth (clusters c0 and c1), including all identified tubulin, actin and profilin proteins (Table 1). The former two types of proteins are assembled into microtubulins and microfilaments, respectively, with crucial roles in cellular development (Mayer and Jürgens, 2002). Ubiquitin/26S proteasome-based selective degradation of proteins contributes to the regulation of protein functions, and is considered important for different cellular and developmental events (Moon et al., 2004). Most of the identified ubiquitin/26S proteasome components (16/18) showed maximal accumulation at this phase (c0 and c1). A similar expression profile was observed in the early stage of maize seed development (Mechin et al., 2007). Accordingly, proteins involved in protein synthesis (8/12 translation factors) and amino acid metabolism (8/13) showed accumulation patterns similar to those of proteolysis-related proteins, whereas folding/modification proteins displayed distinct early and late accumulation profiles (Fig. 5), with about one-third (13/31) showing maximal accumulation at the early phase (Table 1). These lines of evidence suggest that active turn-over of proteins is crucial for normal cell growth. Interestingly, a large number of identified signal transduction-related proteins (8/12), such as GA receptor GID1L2 (spot 2093) and IAA amido synthetase GH3.8 (spot 1459), showed maximal accumulation at the early phase (Table S4). Although the importance of the two proteins in seed development remains to be identified, IAA amido synthetase is known to be involved in regulation of IAA functions (Rampey et al., 2004), and a mutation in rice GID1 locus leads to a dwarf phenotype (Ueguchi-Tanaka et al., 2007). Thus, IAA and GA signaling and/or interaction between the signal pathways is essential for normal seed development.

Seeds at the mid-phase of seed development (8–12 DAF) showed a little increase in size with faster increase in fresh and dry weight as compared with seeds at the early phase. In light of the

proteomics features that cell growth-related proteins were greatly down-regulated from the early phase and that starch synthesis-related proteins were up-regulated to the maximal level after the middle phase (see discussion below), this finding suggests that metabolism/cellular processes occurring in the middle phase involve a transition from cell growth to grain filling. In addition to 8 of the 13 identified protein synthesis-related proteins prominently accumulating in level at the early phase, the remaining 5 accumulated to the maximal level at the middle phase (Table 1), which suggests that active protein synthesis may be important for the transition. An additional feature is that half of the lipid/sterol metabolism-related proteins (3/6) were expressed at the maximal level at this stage, and the other 3 prominently accumulated at the early phase (Table 1). This result is consistent with a previous observation that developing rice seeds rapidly accumulated storage lipids between 5 and 12 DAF (Choudhury and Juliano, 1980; Ichihara et al., 2003). The lipids are mainly stored in aleurone cells (Krishnan and Dayanandan, 2003) and might be implicated in the ability of aleurone cell proliferation at the following storage phase (Kvaale and Olsen, 1986) and in the specification of periphery cells of starchy endosperms (Becraft and Asuncion-Crabb, 2000). Among these proteins, saposin-like type B protein was expressed the highest (Table S4). Saposin-like proteins can interact with lipids (Bruhn, 2005) and function as a surfactant responsible for resistance to surface tension (Cochrane and Revak, 1991). In addition, saposin B usually binds to phospholipid membranes with high affinity and appears to irreversibly cluster at the membrane surface, which leads to cell permeabilization and eventual fusion of the membranes (Poulain et al., 1992; Chang et al., 1998). Therefore, the prominently accumulated saposin B protein may function in reducing the cell surface tension resulting from cell expansion at the phase and possibly in starchy endosperm cell maturity.

Seeds in storage and desiccation phases (12-20 DAF) were defined by prominent accumulation of storage materials and finally became translucent on desiccation. For the functional skew, a large number of starch synthesis-related proteins accumulated to the maximal level at this phase (19/23, c3 and c4) (Table 1). More than one-half of proteins involved in protein folding/modification (17/31) were concurrently expressed with these starch synthesis proteins (Table 1). Thus, protein folding/modification-based regulation of protein function might be important for starch synthesis. Interestingly, 4 of the identified signal transduction proteins, including IAA amidohydrolase (spot 1718) and GA receptor DID1L3 (spot 2120), showed high accumulation at this phase (Table S2 and S4). In Arabidopsis, IAA amidohydrolase plays an



important role in IAA signaling (Rampey et al., 2004). These data suggest that IAA and GA signaling may be involved in starch synthesis.

### **Switch between central carbon metabolism and alcoholic fermentation**

Central carbon metabolism (glycolysis and TCA cycle) provides energy, cofactor regeneration, building blocks for interconversions and synthesis of metabolites, with metabolite concentration gradients usually acting as signals for regulation of diverse processes (Gutierrez et al., 2007). Analyses of intermetabolites suggest the central importance of glycolysis and TCA cycle in fruit/seed development and accumulation of reserves in different species (Rolletschek et al., 2004; Carrari and Fernie, 2006; Fait et al., 2006), but relatively little is known about their regulation (Carrari and Fernie, 2006). Our study identified most of the known glycolysis key enzymes (Table S2), more than half of which were grouped into the early stage clusters c0 and c1 (Table 1, Table S4). Expression of these proteins peaked at approximately 6 DAF and thereafter decreased to a lower constant level, which remained until 20 DAF (Fig. 5). Expression of the TCA cycle proteins increased from 6 DAF, peaked at about 10 DAF, and thereafter showed a steeper decrease (Fig. 5). Although this expression change tendency was summarized from only 3 of the 9 known enzymes involved in the TCA cycle, the pattern appeared to be supported by the dynamic change of ATP levels in developing rice seeds (Fig. 1F) and by the results of previous studies demonstrating a great decrease in TCA cycle-related protein expression (Mechin et al., 2007) and TCA cycle activity (Fait et al., 2006) as developing seeds enter the stage of reserve accumulation in maize (Mechin et al., 2007) and Arabidopsis (Fait et al., 2006). Thus, these results indicate the importance of regulation of glycolysis and TCA cycle in seed development.

A striking result is that contrary to glycolysis and TCA cycle proteins, proteins involved in alcoholic fermentation, including pyruvate decarboxylase, alcohol dehydrogenase and aldehyde dehydrogenase (Table S2), were preferentially grouped into the last two clusters (c3 and c4) (Table 1, S4) and showed great increases in expression (Fig. 5) in parallel with seed development, which indicates up-regulation of the alcoholic fermentation pathway. Alcoholic fermentation is a two-step reaction branching the glycolytic pathway at pyruvate with concomitant oxidization of NADH to NAD<sup>+</sup>, finally generating ATP without the consumption of oxygen (Tadege et al., 1999; Geigenberger, 2003). The feature is in line with increased activity of the enzyme proteins and up-regulated expression at transcriptional and translational levels under low oxygen, which suggests that the pathway is essential for pre-adaptation to anoxia (Tadege et al., 1999;

Geigenberger, 2003), but whether it is involved in seed development and accumulation of reserves is still unknown (Geigenberger, 2003). The developmentally up-regulated expression of the enzyme proteins in rice seeds provides novel clues for evaluating the importance of the pathway in seed development and accumulation of reserves.

A typical feature of developing seeds and bulky organs such as potato tubers is greatly decreased internal oxygen concentration at ambient oxygen levels (21%) (Geigenberger et al., 2000; Geigenberger, 2003; Rolletschek et al., 2004; van Dongen et al., 2004). Low oxygen concentration-mediated decrease in carbon flux through glycolysis and TCA cycle and/or in energy production have been observed in bulky organs/seeds from several species (Gibon et al., 2002; Geigenberger, 2003; van Dongen et al., 2004; Fait et al., 2006). Similarly, down-regulated expression of the enzyme proteins related to glycolysis and TCA cycle, and decreased ATP levels can be observed when storage materials were accumulating in rice seeds (Fig.1F and Fig.5). This evidence, together with low-oxygen-inducible expression of alcoholic fermentation enzymes described above, indicates that low oxygen tension seems to involve a temporal decrease in glycolysis and TCA activity and increase in alcoholic fermentation activity. However, the adaptation mechanism now known suggests that the metabolic acclimation allows a decrease in consumption of ATP and oxygen, which prevents the tissue from becoming anoxic and stops the fermentation (Geigenberger et al., 2000; Weber et al., 2005); this mechanism seems not able to explain the metabolic switches occurred in rice seeds, because the expression of alcoholic fermentation proteins were up-regulated in our observation.

Given the fact that the endosperm is the prominent part of cereal seeds and its functional skew to starch accumulation, a possible explanation for the switch is a positive mechanism underlying starch accumulation formed during evolution. Decreased oxygen tension may act as a signal, in coordination with other now-unidentified signal molecules, to regulate the switches. The switches finally decrease the flux of imported sucrose into non-reserve materials in the sink and maintain an appropriate level of energy molecules and cofactors such as PPi, thus leading to increased flux for starch synthesis. The following lines of evidence support this hypothesis. First, developing rice seeds of 6 to 20 DAF underwent sink establishment at approximately 6 to 8 DAF by active cell enlargement and thereafter active starch synthesis; the prominent activities of glycolysis and TCA cycle at the early stage were in line with the requirement for energy and synthesis of cellular components essential for cell enlargement. After the establishment of the

sink, the activity of the two pathways was greatly decreased, with glycolysis remaining at a constantly low level until 20 DAF. Second, glycolysis in rice seeds may be low hexose-consuming because the differentially expressed enzyme protein involved in conversion of fructose-6-phosphate to fructose-1,6-phosphate is pyrophosphate-dependent phosphofructokinase (PPi-PFK, spots 1518,1523 and 1526, Table S2 and S3), which uses PPi rather than ATP as a donor. Third, a restricted flux through the TCA cycle resulted in increased yield in tomato (Carrari et al., 2003; Nunes-Nesi et al., 2005), whereas increased respiration led to remarkable inhibition of starch accumulation in potato (Bologa et al., 2003). Fourth, the alcoholic fermentation pathway has been considered as the key catalytic process for recycling NAD<sup>+</sup> essential for glycolysis and TCA cycle (Tadege et al., 1999; Fernie et al., 2004). Up-regulated expression of alcoholic fermentation proteins in rice seed development may be important to maintain glycolysis at a constant level, thus maintaining an appropriate ATP level for starch synthesis under low oxygen tension. Finally, starch synthesis is very well adapted to the internal low oxygen condition (Rolletschek et al., 2005).

Together, these data clearly indicate that the coordinated switch between central carbon metabolism and alcoholic fermentation is essential for synthesis of reserves, but the mechanism underlying regulation of the switch remains poorly understood. Glycolysis in plants is a demand-driven process similar to that in *E coli*, in which the glycolic flux is controlled by the demand for ATP (Fernie et al., 2004). ABA is required for reserve accumulation, and its involvement in the control of starch accumulation appears via SNF1 kinase, which mediates the phosphorylation state of some metabolic enzymes, and a set of transcriptional factors (Gutierrez et al., 2007). ABA and GA were found to coordinately regulate the expression of alcohol dehydrogenase in barley seeds (Password et al., 2001), and cross-talk between ABA and GA signaling has been found in rice seeds (Xie et al., 2006). Two putative GA receptor proteins were identified in rice seeds; one up-regulated and the other down-regulated (Table S4). These results indicate that low oxygen tension, ABA and GA, and possibly metabolite gradient, are involved in regulation of the metabolic switches.

### **Importance of PPK in carbon skeleton and energy distribution**

PPDK catalyzes the reversible conversion of pyruvate, ATP and Pi into phosphoenolpyruvate (PEP), AMP and PPi. The rice genome has two loci encoding three types of PPK proteins: *OsPPDKA* encodes a cytosolic PPK (*OsPPDKA*), and *OsPPDKB* contributes another cytosolic

PPDK (cyOsPPDKB) and a C4-type chloroplastic PPDK (chOsPPDKB) by alternative splicing (Imaizumi et al., 1997; Moons et al., 1998). Our proteomic analysis identified 13 distinct PPDK isoforms changed in expression (Fig. 2B), all of which were from *OsPPDKB* (Table S2), representing 3.9%-8.9% of the total RV of all spots detected on 2-DE. A recent proteomic study revealed 7 PPDK isoforms differentially expressed in developing maize seeds (Mechin et al., 2007). The presence of multiple PPDK isoforms and high abundance in developing seeds revealed by these studies suggests the importance of the protein; however, the functional importance is not fully understood.

The interconversion feature of PEP and pyruvate by PPDK leads to difficulties in evaluating PPDK function in seed development (Chastain and Chollet, 2003). Expression analyses of transcripts and proteins in immature wheat and maize seeds suggested the possible functions of PPDK in refixation of respired CO<sub>2</sub> (Aoyagi and Bassham, 1984) and/or in synthesis of amino acids by pyruvate and/or by PEP, which might be required for storage protein synthesis (Mechin et al., 2007). However, the loss-of-function mutant of *OsPPDKB* showed increased lipid content and aberrant starch synthesis but no change in amino acid and storage protein contents, which suggests an important role of PPDK in lipid and starch synthesis (Kang et al., 2005). Therefore, the PPDK-mediated cycle between PEP and pyruvate and generation of metabolic signals such as PPi/Pi might have multifaceted functions in seed development, which in some ways might depend on the direction of the cycle (Chastain and Chollet, 2003).

Based on our results, functions of PPDK in seed development can be documented as follows. First, expression of PPDK in seeds was developmentally regulated: the protein peaked in level at the early stage and decreased thereafter (this study; Chastain and Chollet, 2003; Mechin et al., 2007). Second, our results suggested that glycolysis was active at early stages of the process and then remained at a relatively low and constant level, but TCA cycle showed a steeper decrease in activity; and ATP levels showed a change similar to TCA cycle, which is consistent with the observations that respiration was inhibited in part in potato tubers and developing seeds (Geigenberger, 2003). Blocking respiration is known to lead to considerable increase in pyruvate concentration (Tadege et al., 1999); the accumulated pyruvate is branched by the alcoholic fermentation pathway (see discussion above) and/or becomes available for PPDK reaction. Therefore, the PPDK reaction may function mainly to fuel pyruvate at the early stage, whereas alcoholic fermentation becomes important for consumption of pyruvate at the late

stage, because alcoholic fermentation activity appeared to be low at the early stage and increased with seed development. The PPDK reaction may have important roles in starch synthesis by recycling PEP into hexose pools (Kang et al., 2005) and by maintaining concentration of cytoplasmic PPi. Indeed, the cytosol of starch accumulation sinks contains relative high levels of PPi, which is utilized in the sinks as an alternative energy donor for conversion of imported sucrose into hexose by sucrose synthase (SuSy), which is a typical anaerobic polypeptide up-regulated in expression levels and activated under the anaerobic conditions (Guglielminetti et al., 1997). SuSy-mediated sucrose mobilization is closely associated with storage starch synthesis (Quick and Schaffer, 1996; Wobus et al., 1997; Winter and Huber, 2000). This is consistent with our observation that revealed 9 differentially expressed identities (Table S2) covering all the three known isozymes of rice SuSy (Huang et al., 1996; Wang et al., 1999). The hypothesis that DDPK in developing seeds, at least in rice, favors the reaction from pyruvate to PEP explains at least in part the cytosolic PPi pool and also supports the phenotype of the loss-of-function mutant of OsDDPKB (Kang et al., 2005). In addition, the PPi pool may be involved in maintaining constant glycolysis by the PPi-PFK reaction (see discussion above). Together, the lines of evidence indicates that the involvement of PPDK in carbon skeleton and energy partition is associated with a dynamic balance between the PPDK reaction and other metabolic actions, including glycolysis and alcoholic fermentation pathways.

### **Proteins involved in starch synthesis**

In addition to sink establishment and carbon skeleton/energy supply, starch synthesis requires coordination of multiple starch synthesis-related enzymes (Tetlow, 2006). Our study revealed a coordinated expression feature of these enzyme proteins (Table S4) and their isoforms (Fig. 7), including ADPGase, isoamylase, pullulanase, starch phosphorylase and plastidic phosphoglucomutase. Most of these proteins were grouped into late clusters c3 and c4 (Table 1).

**APGase:** the heterotetrameric enzyme consisting of small and large subunit proteins catalyzes the first-step reaction of starch synthesis by converting glucose-1-phosphate (Glc1P) to ADP-glucose (ADP-Glc), the substrate of starch synthesis (Tetlow, 2006), which in rice are encoded by six genes: 2 for the small subunits (OsAGPS1 and OsAGPS2) and 4 for the large subunit (OsAGPL1, OsAGPL2, OsAGPL3 and OsAGPL4). In addition, AGPS2 produces AGPS2a and AGPS2b by alternative splicing (Ohdan et al., 2005). OsAGPL2 and AGPS2b (48 kDa) have been found to locate in the cytoplasm and OsAGPL3 and AGPS2a (54 kDa) in

amyloplasts (Lee et al., 2007). Our proteomic study revealed 16 differentially expressed identities matching APGase: 7 for OsAGPL2, 4 for OsAGPL3 and 5 for OsAGPS2 (Table S2) (10 suitable for quantification analysis, Table S4). Most isoforms (8/10) of APGase were up-regulated at the mid and late stages of seed development, when starch synthesis was preponderant. Interestingly, the ratio of expression abundance of AGPS2b to AGS2a proteins was exactly 10:1 at 5 of 8 stages (6, 10, 12, 14 and 16 DAF) (Table S4). This indicated that OsAGPS2 produces the 2 alternative splice forms by a stringent regulation mechanism at the transcriptional and/or translational level. Accordingly, the cytoplasmic forms (OsAGPL2 and OsAGPS2b) had 3.0-4.3 times higher abundance than the amyloplast forms (OsAGPL3 and OsAGPS2a) at each development stage (Table S4). These findings supply molecular evidence for the early observation that cytoplasmic APGase contributes about 90% of total AGPase activity in rice endosperm (Sikka et al., 2001).

***Isoamylase and Pullulanase:*** Isoamylase and pullulanase belong to starch debranching enzymes and are essential for biosynthesis of normal starch (Tetlow, 2006). The rice genome encodes three types of isoamylases (IsoI, II and III) (Kubo et al., 2005) and one pullulanase (Nakamura et al., 1996). Our study revealed differential expression patterns of IsoI (2 isoforms), IsoIII (1 isoform) and pullulanase (6 isoforms) in developing seeds (Fig. 7, Table S4). Isos and pullulanase were significantly different in abundance and in expression features. Pullulanase was about 10 to 100 times higher in level than Isos at different stages (Table S4). IsoI and IsoIII peaked in level around 12 DAF; whereas the abundance of pullulanase increased until 18 DAF and decreased slightly thereafter, which is in agreement with the dynamic changes in starch accumulation. These results suggest a possible functional difference between the two types of starch debranching enzymes. A study of wheat seeds reveals that the transcriptional levels of IsoI peak at early to mid (approximately 10 DAF) and mid (15-20 DAF) stages of development (Genschel et al., 2002). In barley, Isos determined the amount of starch granules by affecting initiation of starch granules (Burton et al., 2002; Kawagoe et al., 2005). Therefore, in rice seeds, Isos may function mainly in the initiation of starch granule cores and thus be necessary for determination of the amount of starch granules, whereas pullulanase may be important for yield and structure modification of starch.

***Starch phosphorylase and plastidic phosphoglucomutase:*** Starch phosphorylase (SP) catalyzes the reversible transfer of a glucosyl unit from the  $\alpha$ -1,4-linked glucan chains to produce

Glc1P (Newgard et al., 1989). Plastidic phosphoglucomutase (PPGM) catalyzes the reversible interconversions of Glc6P and Glc1P and supplies Glc1P for AGPase in plastid. In dicotyledonous species, PPGM plays a key role in starch synthesis (Caspar et al., 1985; Fritzius et al., 2001). However, in contrast to the prominent plastidial form in non-cereal plants, AGPase in cereal seeds is available largely as cytosolic form (Tetlow, 2006); and ADP-Glc can be synthesized in the cytosol of cereal endosperms and imported into the amyloplast (Beckles et al., 2001). Therefore in cereal seeds, PPGM proteins may not be the limiting factor for starch synthesis (Beckles et al., 2001; Ohdan et al., 2005). In developing rice seeds, 3 isoforms of SP and 2 of PPGM displayed differential expression (Fig. 7). One SP isoform (spot 2461) peaked in expression level throughout 10 to 14 DAF, similar to the expression pattern of IsoIII; whereas the other 2 SP isoforms (spots 1202 and 1214) and the 2 PPGM isoforms displayed increased abundance during the entire development process, which was comparable with the expression pattern of pullulanase and APGase (Fig. 7). This suggests synchronized expression and/or activity among SP/PPGM and starch synthesis-related Isos, pullulanase and APGase is important for starch synthesis. Normal starch synthesis requires “trimming” and/or “clearing” of the disordered water-soluble polysaccharide (termed phytyglycogen), which interferes with the formation of normal starch (Tetlow, 2006). Some observations revealed that SP has a role in trimming and/or clearing (Tetlow, 2006). The co-expressed SP, Iso, pullulanase and APGase suggests roles of SP in starch granule trimming is required for starch synthesis. The reverse reaction catalyzed by PPGM produces a 20-fold excess of Glc6P over Glc1P in wheat endosperm (Davies et al., 2003). Similarly, in hypoxic maize roots, the reaction direction shows a preference for the generation of excessive Glc6P (Manjunath et al., 1998). Thus, up-regulated PPMG in developing rice seeds may function in maintaining a Glc6P/Glc1P pool which is drained for starch synthesis in amyloplasts when the source supply is deficient, which might occur at night or in the desiccation phase.

In summary, we analyzed the dynamic changes in protein expression profiles during 8 sequential developmental stages associated with grain filling from 6 to 20 DAF in rice. Our results indicate that during the developmental process, proteins involved in starch synthesis and alcoholic fermentation are up-regulated; and proteins implicated in other categories/subcategories show a tendency to decrease in expression. Importantly, our study reveals a switch from central carbon metabolism to alcoholic fermentation in the developmental

phase. Our results also suggest that coordination of different metabolism and cellular processes is associated with starch synthesis and accumulation in seed development. These results provide novel clues for further understanding the metabolic network involved in starch accumulation in developing seeds.

## **MATERIALS AND METHODS**

### ***Plant materials and sampling***

Rice plants (*Oryza sativa* L. cv. Nipponbare) were cultured during rice growing season (May–September) under natural conditions in Beijing (39° 54' N, 116° 24' E) and were fertilized (urea, 60 kg/ ha.), and watered as usual. The superior seeds (Ishimaru et al., 2003) of the top 3 spikelets were labeled at noon of anthesis when at least half of the superior seeds of the corresponding spikelets were flowering. The labeled spikelets were harvested at 6, 8, 10, 12, 14, 16, 18 and 20 days after flowering. Each sample of these stages consisted of at least 200 seeds from 30 spikes and were stored immediately at -80°C until protein extraction.

### **Observation of seed, embryo and endosperm development**

To monitor the cellular changes of embryos and endosperms, the developing seeds were fixed in FAA (50% ethanol, 5% acetic acid, 10% formalin), and then embedded in paraffin. The specimens were thin-sliced using a microtome (Leica RM2235), then mounted on a grid, and finally stained with 1% safranin and 0.5% fast green for embryo observation or 1% I<sub>2</sub>-KI for endosperm observation. The stained specimens were observed under a light microscopy.

### **Preparation of proteins**

After being dehusked, seeds (1 g) were ground with ice-cold extraction buffer (20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 10 mM PMSF, 10 mM DTT) on ice. Supernatant was collected by centrifugation at 35,000 g at 4°C for 20 min. The pellet was resuspended the extraction buffer for repeated extraction, then centrifuged at 35,000 g at 4°C for 20 min for collection of supernatant. Proteins in the combined supernatant were precipitated with 4 volumes of ice-cold trichloroacetic acid-acetone (10% trichloroacetic acid in 100% acetone) at -20°C for 4 h and then collected by centrifugation at 35,000 g for 20 min. The pelleted proteins were washed first with 80% cold acetone containing 0.07% β-mercaptoethanol and then with cold acetone containing 0.07% β-mercaptoethanol and finally vacuum dried as described (Dai et al., 2007). The resulting proteins



were dissolved in a lysis buffer (7 M urea, 2 M thiourea, 4% CHAPS, 7 mM DTT, 2% pharmalyte 3–10) at room temperature. After the removal of debris by centrifugation at 40,000g for 20 min, the proteins were quantified according to the Bradford method (Bradford, 1976) by DU640 UV-visible spectrophotometry (Beckman). Bovine serum albumin was used as a standard. The final proteins underwent 2-DE immediately or were stored in aliquots at  $-80^{\circ}\text{C}$ . For each sample, triplicate biological protein preparations were performed.

## **2-DE and image analysis**

An aliquot (1 mg proteins) of protein samples was diluted with rehydration buffer (6 M urea, 2 M thiourea, 0.5% CHAPS, 20 mM DTT, 0.5% IPG buffer 3–10, 0.002% Bromophenol Blue) to a final volume of 450  $\mu\text{L}$  and loaded onto an IPG strip holder containing a 24-cm, pH 3-10 or pH 4-7 linear gradient IPG strip (GE Healthcare). Isoelectric focusing was performed in the Ettan IPGphor Isoelectric focusing system following the protocol of the manufacturer. For SDS-PAGE, the equilibrated IPG strips were transferred onto 12.5% acrylamide gels by use of an Ettan DALT Six Electrophoresis Unit (GE Healthcare). Low-molecular-mass (relative MM) protein markers (GE Healthcare) were co-electrophoresed as MM standards. The proteins in gels were visualized by Coomassie Brilliant Blue (CBB) staining. 2-DE experiments were repeated three times by use of protein samples independently prepared from separate seed samples. Images were acquired by scanning each stained gel using an ImageScanner (GE Healthcare) at a resolution of 300 dpi and 16-bit grayscale pixel depth and then analyzed by use of ImageMaster 2D v5.0 (GE Healthcare). The experimental MM of each protein was estimated by comparison with the co-electrophoresed MM markers. The experimental pI of each protein was determined by its migration on IPG linear strips.

After spot detection, quantification, and background subtraction, the spot groups were determined. For spot profile analysis, the first replication 2-D gel of 10-DAF-seed samples was selected as the reference gel. All analyzed gels were matched individually to the reference gel, and matched spots from the different gels were assigned to a spot group. Then the spot groups were selected for profile analysis only if they were confirmed to exist in at least two independent sample sets in two stages (including the stage of 10 DAF), and all matched spots were checked manually to delete the one with lower confidence. To decrease all differences derived from 2-DE, the RV of each selected spot was chosen for analysis. For each spot, the mean RV was computed

at every stage. The spots showing a mean RV that changed more than 2 times in different stages were considered differentially expressed.

### **Protein identification with MALDI-TOF/TOF MS**

All differentially expressed spots were manually excised from the 2-DE gels. In-gel digestion and MS acquisition was performed as described previously (Dai et al., 2006). The MS spectra were created on an Ultroflex II MALDI TOF/TOF mass spectrometer (Bruker Daltonics) by use of Flexanalysis 2.4 software. The external calibration of the Ultroflex spectrometer and a two-point internal calibration for each spectrum were as described (Dai et al., 2006; Dai et al., 2007). The mass accuracy was limited to 25 ppm, and the resolution was more than 25000. After tryptic peptide masses were transferred to a BioTools 3.0 interface (Bruker Daltonics), PMFs were searched against the NCBIInr (<http://www.ncbi.nlm.nih.gov/>) and MSDB (<ftp://ftp.ncbi.nih.gov/repository/MSDB>) protein databases by use of Mascot v2.1 search engine (<http://www.matrixscience.com>; Matrix Science, London, UK). The taxonomic category was *Oryza sativa* (66,973 sequence entries in NCBI in February 2006; 66,430 entries in MSDB in December 2006). All peptide masses were considered mono-isotopic and  $[M+H]^+$  (protonated molecular ions). Searches involved a mass accuracy of  $\pm 200$  ppm, and one missing cleavage site was allowed for each search. Carbamidomethyl (C) was set as fixed modification with variable modifications: oxidation (M) and pyro-glu (N-term Q). To enhance the accuracy of identification, proteins had to meet the following: 1) the top hits of the two databases were homologous; 2) the probability-based MOWSE score of identified proteins was greater than 65 ( $P < 0.01$ ); and 3) the identified proteins matched more than 6 peptides which covered more than 15% sequence of the protein.

### **Bioinformatic analysis**

The chromosome loci of the protein-encoding genes were searched in the TIGR database (<http://tigblast.tigr.org>), and the loci with highest scores were considered positive results. Different proteins mapped in the same locus were considered unique.

Hierarchical clustering analysis was performed with the mean RV of each spot using GeneCluster 2.0 (<http://www.broad.mit.edu/cancer/software>), which allows for visualizing the profile of each cluster. After being normalized to mean 0 and variance 1, clusters were created

with default parameters, except for cluster range 3-7. Different cluster ranges were compared, and the range of five was selected because the distribution of functional categories between clusters possesses the most significant difference ( $\chi^2$  48 DF, 73.81,  $p \leq 0.01$ ).

Digital expression profiles were created on the basis of the hierarchical clustering analysis with the following equation for each category or sub-category: y (6 DAF, 8 DAF, 10 DAF, 12 DAF, 14 DAF, 16 DAF, 18 DAF, 20 DAF) = centroids (c0) \* amount (c0) + centroids (c1) \* amount (c1) + centroids (c2) \* amount (c2) + centroids (c3) \* amount (c3) + centroids (c4) \* amount (c4). Here, the centroid represents the average of normalized data of all the proteins belong to the same cluster at one stage; and centroids were all the centroids of 8 stages for a given cluster.

### **ATP measurement**

The ATP content was measured based on the luciferase reaction of the ENLITEN ATP assay kit (Promega). The ATP was extracted by following the method of Liu et al (2007) with some modifications. Briefly, seeds were ground to fine powders in liquid nitrogen. 50 mg of the powder was transferred into a tube, and then 0.2 ml of trichloroacetic acid solution (5% w/v) was added for ATP extraction. After centrifugation at 20 000 g for 5 min, 10  $\mu$ l of supernatant were transferred into a tube containing 490  $\mu$ l of 25mM Tris-acetate buffer (pH 8.0). The Bioluminescence of each sample was detected using a fluorometer (PolarSTAR, BMG Lab Technologies). A standard curve was established using the ATP standard solution supplied in the kit to quantify the ATP levels. For each sample, triplicate biological preparations were performed.

### **Real-time quantitative RT-PCR (qRT-PCR)**

Total RNA was prepared from seeds using RNAprep pure Plant Kit (TIANGEN) and treated with RNase-free DNase I. 0.5 $\mu$ g of total RNA was used for the first-strand cDNA synthesis with ReverTra Ace (TOYOBO). Triplicate quantitative assays were performed with an Mx3000P system (STRATAGENE) by use of Power SYBR® Green PCR Master Mix kit according to the manufacture's protocol (Applied Biosystems). Gene-specific primers designed by using PRIMEREXPRESS software (Applied Biosystems) were listed in the supplementary Table S3. The relative quantification method ( $2^{-\Delta\Delta C_T}$ ) was used to evaluate quantitative variation between

replicates examined.

### **Supplemental material**

**Supplemental Table S1.** Expression profile data for 396 differentially expressed protein spots.

**Supplemental Table S2.** Identities of differentially expressed proteins determined by MALDI-TOF-TOF MS.

**Supplemental Table S3.** A comparison between expression profiles of protein-mRNA pairs of the identified protein-encoding genes.

**Supplemental Table S4.** Proteins distribution in hierarchical clusters.

**Supplemental Table S5.** Expression correlation between isoform pairs.

**Supplemental Table S6.** The centroids data of each hierarchical clusters created by GeneCluster 2.0.

**Supplemental Figure S1.** Comparison of 2-DE images between pH3-10 and pH4-7 IPG stripes.

**Supplemental Figure S2 PMF-based identifications.**

**Supplemental Figure S3.** The correlation of molecular mass (MM) and isoelectric point (pI) between theoretical and experimental data.

**ACKNOWLEDGEMENT:** We thank Dr. Siqi Liu (Beijing Genomics Institute, Institute Academy of Sciences) for technical assistance in MS analysis.

### **LITERATURES CITED**

**Aoyagi K, Bassham JA** (1984) Pyruvate orthophosphate dikinase of C3 seeds and leaves as compared to the enzyme from Maize. *Plant Physiol.* **75**: 387-392

**Beckles DM, Smith AM, ap Rees T** (2001) A cytosolic ADP-glucose pyrophosphorylase is a feature of graminaceous endosperms, but not of other starch-storing organs. *Plant Physiol.* **125**: 818-827

**Becraft PW, Asuncion-Crabb Y** (2000) Positional cues specify and maintain aleurone cell fate in maize endosperm development. *Development* **127**: 4039-4048

**Berger F** (1999) Endosperm development. *Current Opinion in Plant Biology* **2**: 28-32

**Bologa KL, Fernie AR, Leisse A, Loureiro ME, Geigenberger P** (2003) A bypass of sucrose synthase leads to low internal oxygen and impaired metabolic performance in growing potato tubers. *Plant Physiol.* **132**: 2058-2072

**Bradford MM** (1976) 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

- Bruhn H** (2005) A short guided tour through functional and structural features of saposin-like proteins. *Biochem. J* **389**: 249-257
- Burton RA, Jenner H, Carrangis L, Fahy B, Fincher GB, Hylton C, Laurie DA, Parker M, Waite D, van Wegen S, Verhoeven T, Denyer K** (2002) Starch granule initiation and growth are altered in barley mutants that lack isoamylase activity. *Plant J.* **31**: 97-112
- Carrari F, Fernie AR** (2006) Metabolic regulation underlying tomato fruit development. *J Exp Bot.* **57**: 1883-1897
- Carrari F, Nunes-Nesi A, Gibon Y, Lytovchenko A, Loureiro ME, Fernie AR** (2003) Reduced expression of aconitase results in an enhanced rate of photosynthesis and marked shifts in carbon partitioning in illuminated leaves of wild species tomato. *Plant Physiol.* **133**: 1322-1335
- Caspar T, C. S, Huber, Somerville C** (1985) Alterations in growth, photosynthesis, and respiration in a starchless mutant of *Arabidopsis thaliana* (L.) deficient in chloroplast phosphoglucomutase activity. *Plant Physiol.* **79**: 11-17
- Chang R, Nir S, Poulain FR** (1998) Analysis of binding and membrane destabilization of phospholipid membranes by surfactant apoprotein B. *Biochim Biophys Acta* **1371**: 254-264
- Chastain CJ, Chollet R** (2003) Regulation of pyruvate, orthophosphate dikinase by ADP-/Pi-dependent reversible phosphorylation in C3 and C4 plants. *Plant Physiology and Biochemistry* **41**: 523-532
- Choudhury NH, Juliano BO** (1980) Lipids in developing and mature rice grain. *Phytochemistry* **19**: 1063-1069
- Cochrane CG, Revak SD** (1991) Pulmonary surfactant protein B (SP-B): structure-function relationships. *Science* **254**: 566-568
- Cox B, Kislinger T, Emili A** (2005) Integrating gene and protein expression data: pattern analysis and profile mining. *Methods* **35**: 303-314.
- Dai S, Chen T, Chong K, Xue Y, Liu S, Wang T** (2007) Proteomics identification of differentially expressed proteins associated with pollen germination and tube growth reveals characteristics of germinated *Oryza sativa* pollen. *Molecular & Cellular Proteomics* **6**: 207-230
- Dai S, Li L, Chen T, Chong K, Xue Y, Wang T** (2006) Proteomic analyses of *Oryza sativa* mature pollen reveal novel proteins associated with pollen germination and tube growth. *Proteomics* **6**: 2504-2529
- Davies EJ, Tetlow IJ, Bowsher CG, Emes MJ** (2003) Molecular and biochemical characterization of cytosolic phosphoglucomutase in wheat endosperm (*Triticum aestivum* L. cv. Axona). *J Exp Bot.* **54**: 1351-1360
- Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G** (2006) *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* **142**: 839-854
- Fernie AR, Carrari F, Sweetlove LJ** (2004) Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol.* **7**: 254-261

- Fritzius T, Aeschbacher R, Wiemken A, Wingler A** (2001) Induction of ApL3 expression by trehalose complements the starch-deficient Arabidopsis mutant *adg2-1* lacking ApL1, the large subunit of ADP-glucose pyrophosphorylase. *Plant Physiol.* **126**: 883-889
- Gallardo K, Le Signor C, Vandekereckhove J, Thompson RD, Burstin J** (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol.* **133**: 664-682
- Geigenberger P** (2003) Response of plant metabolism to too little oxygen. *Curr Opin Plant Biol.* **6**: 247-256
- Geigenberger P, Fernie AR, Gibon Y, Christ M, Stitt M** (2000) Metabolic activity decreases as an adaptive response to low internal oxygen in growing potato tubers. *Biol Chem.* **381**: 723-740
- Genschel U, Abel G, Lorz H, Lutticke S** (2002) The sugary-type isoamylase in wheat: tissue distribution and subcellular localisation. *Planta* **214**: 813-820
- Gibon Y, Vigeolas H, Tiessen A, Geigenberger P, Stitt M** (2002) Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. *Plant J.* **30**: 221-235
- Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J** (2000) Microarray analysis of developing Arabidopsis seeds. *Plant Physiol.* **124**: 1570-1581
- Goldberg RB, de Paiva G, Yadegari R** (1994) Plant embryogenesis: zygote to seed. *Science* **266**: 605-614
- Greenbaum D, Colangelo C, Williams K, Gerstein M** (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4**: 117
- Guglielminetti L, Wu Y, Boschi E, Yamaguchi J, Favati A, Vergara M, Perata P, Alpi A** (1997) Effects of anoxia on sucrose degrading enzymes in cereal seeds. *Journal of Plant Physiology* **150**: 251-258
- Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C** (2007) Combined networks regulating seed maturation. *Trends Plant Sci* **12**: 294-300
- Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ** (2006) Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. *Plant Physiol.* **141**: 32-46
- Hajduch M, Ganapathy A, Stein JW, Thelen JJ** (2005) 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**: 1397-1419
- Huang JW, Chen JT, Yu WP, Shyur LF, Wang AY, Sung HY, Lee PD, Su JC** (1996) Complete structures of three rice sucrose synthase isogenes and differential regulation of their expressions. *Biosci Biotechnol Biochem* **60**: 233-239
- Ichihara K, Kobayashi N, Saito K** (2003) Lipid synthesis and acyl-CoA synthetase in developing rice seeds. *Lipids* **38**: 881-884

- Imaizumi N, Ku MS, Ishihara K, Samejima M, Kaneko S, Matsuoka M** (1997) Characterization of the gene for pyruvate,orthophosphate dikinase from rice, a C3 plant, and a comparison of structure and expression between C3 and C4 genes for this protein. *Plant Mol Biol.* **34**: 701-716
- Ishimaru T, Matsuda T, Ohsugi R, Yamagishi T** (2003) Morphological development of rice caryopses located at the different positions in a panicle from early to middle stage of grain filling. *Functional Plant Biology* **30**: 1139-1149
- James MG, Denyer K, Myers AM** (2003) Starch synthesis in the cereal endosperm. *Curr Opin Plant Biol.* **6**: 215-222
- Kang H-G, Park S, Matsuoka M, An G** (2005) White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (OsPPDKB). *Plant J.* **42**: 901-911
- Kawagoe Y, Kubo A, Satoh H, Takaiwa F, Nakamura Y** (2005) Roles of isoamylase and ADP-glucose pyrophosphorylase in starch granule synthesis in rice endosperm. *Plant J.* **42**: 164-174
- Krishnan S, Dayanandan P** (2003) Structural and histochemical studies on grain-filling in the caryopsis of rice (*Oryza sativa* L.). *J Biosci.* **28**: 455-269
- Kubo A, Rahman S, Utsumi Y, Li Z, Mukai Y, Yamamoto M, Ugaki M, Harada K, Satoh H, Konik-Rose C** (2005) Complementation of sugary-1 phenotype in rice endosperm with the wheat isoamylase1 gene supports a direct role for isoamylase1 in amylopectin biosynthesis. *Plant Physiol.* **137**: 43-56
- Kvaale A, Olsen A** (1986) Rates of cell division in developing barley endosperms. *Ann Bot.* **57**: 829-833
- Lai J, Dey N, Kim CS, Bharti AK, Rudd S, Mayer KFX, Larkins BA, Becraft P, Messing J** (2004) Characterization of the maize endosperm transcriptome and Its comparison to the rice genome. *Genome Research* **14**: 1932-1937
- Le BH, Wagmaister JA, Kawashima T, Bui AQ, Harada JJ, Goldberg RB** (2007) Using genomics to study legume seed development. *Plant Physiol.* **144**: 562-574
- Leader DJ** (2005) Transcriptional analysis and functional genomics in wheat. *Journal of Cereal Science* **41**: 149-163
- Lee SK, Hwang SK, Han M, Eom JS, Kang HG, Han Y, Choi SB, Cho MH, Bhoo SH, An G** (2007) Identification of the ADP-glucose pyrophosphorylase isoforms essential for starch synthesis in the leaf and seed endosperm of rice (*Oryza sativa* L.). *Plant Molecular Biology* **65**: 531-546
- Lewis S, Currie I** (2003) A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **3**: 36-44
- Liu Y, Ren D, Pike S, Pallardy S, Gassmann W, Zhang S** (2007) Chloroplast-generated reactive oxygen species are involved in hypersensitive response-like cell death mediated by a mitogen-activated protein kinase cascade. *In*, Vol 51. Blackwell Synergy, pp 941-954






- Manjunath S, Lee CH, VanWinkle P, Bailey-Serres J** (1998) Molecular and biochemical characterization of cytosolic phosphoglucomutase in maize. Expression during development and in response to oxygen deprivation. *Plant Physiol.* **117**: 997-1006
- Mayer U, Jürgens G** (2002) Microtubule cytoskeleton: a track record. *Curr Opin Plant Biol.* **5**: 494-501
- Mechin V, Thevenot C, Le Guilloux M, Prioul J-L, Damerval C** (2007) Developmental analysis of maize endosperm proteome suggests a pivotal role for pyruvate orthophosphate dikinase. *Plant Physiol.* **143**: 1203-1219
- Moon J, Parry G, Estelle M** (2004) The ubiquitin-proteasome pathway and plant development. *Plant Cell* **16**: 3181-3195
- Moons A, Valcke R, Van Montagu M** (1998) Low-oxygen stress and water deficit induce cytosolic pyruvate orthophosphate dikinase (PPDK) expression in roots of rice, a C3 plant. *Plant J.* **15**: 89-98
- Nakamura Y, Umemoto T, Ogata N, Kuboki Y, Yano M, Sasaki T** (1996) Starch debranching enzyme (R-enzyme or pullulanase) from developing rice endosperm: purification, cDNA and chromosomal localization of the gene. *Planta* **199**: 209-218
- Newgard CB, Hwang PK, Fletterick RJ** (1989) The family of glycogen phosphorylases: structure and function. *Critical Reviews in Biochemistry and Molecular Biology* **24**: 69-99
- Nunes-Nesi A, Carrari F, Lytovchenko A, Smith AMO, Loureiro ME, Ratcliffe RG, Sweetlove LJ, Fernie AR** (2005) Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. *Plant Physiol.* **137**: 611-622
- Ohdan T, Francisco PB, Jr., Sawada T, Hirose T, Terao T, Satoh H, Nakamura Y** (2005) Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *J Exp Bot.* **56**: 3229-3244
- Olsen OA** (2001) Endosperm development: cellularization and cell fate specification. *Annual Review of Plant Physiology and Plant Molecular Biology* **52**: 233-267
- Password F, View ISI, Plantarum P** (2001) Regulation of alcohol dehydrogenase gene expression in barley aleurone by gibberellin and abscisic acid. *Physiologia Plantarum* **111**: 533-539
- Poulain FR, Allen L, Williams MC, Hamilton RL, Hawgood S** (1992) Effects of surfactant apolipoproteins on liposome structure: implications for tubular myelin formation. *American Journal of Physiology- Lung Cellular and Molecular Physiology* **262**: 730-739
- Quick WP, Schaffer AA** (1996) Sucrose metabolism in sources and sinks. Photoassimilate distribution in plants and crops: source-sink relationships ( Zamski,E.and Schaffer,AA,eds). New York: Marcel Dekker: 115-156
- Rampey RA, LeClere S, Kowalczyk M, Ljung K, Sandberg G, Bartel B** (2004) A family of auxin-conjugate hydrolases that contributes to free indole-3-acetic acid levels during Arabidopsis germination. *Plant Physiol.* **135**: 978-988
- Rolletschek H, Koch K, Wobus U, Borisjuk L** (2005) Positional cues for the starch/lipid balance in maize kernels and resource partitioning to the embryo. *Plant J.* **42**: 69-83

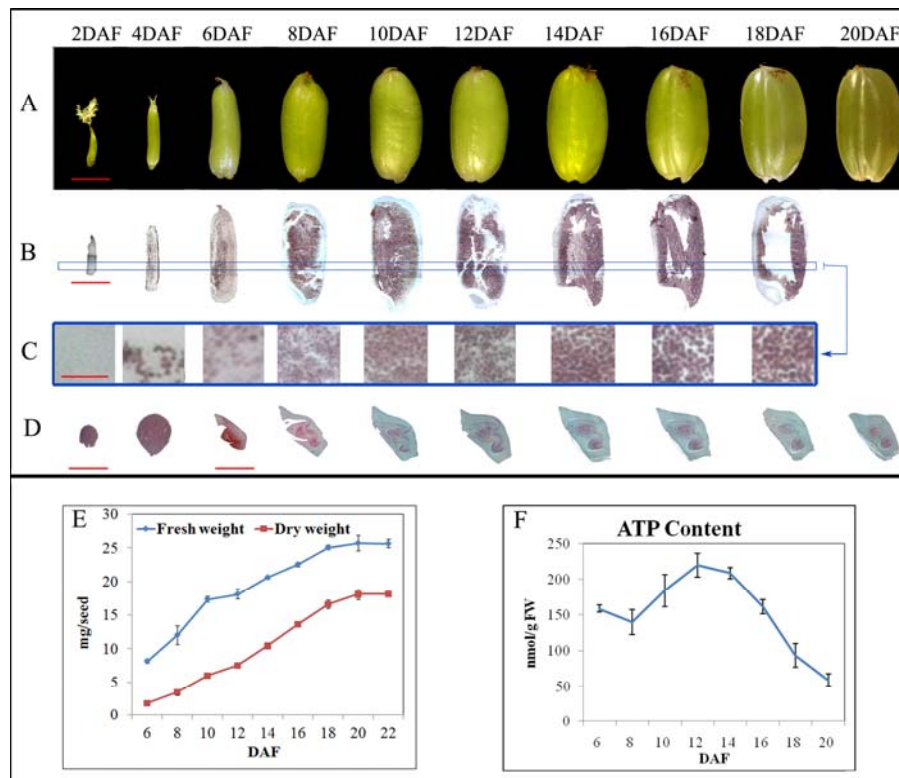


- Rolletschek H, Weschke W, Weber H, Wobus U, Borisjuk L** (2004) Energy state and its control on seed development: starch accumulation is associated with high ATP and steep oxygen gradients within barley grains. *J Exp Bot.* **55**: 1351-1359
- Ruuska SA, Girke T, Benning C, Ohlrogge JB** (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. *Plant Cell* **14**: 1191-1206
- Schmidt MW, Houseman A, Ivanov AR, Wolf DA** (2007) Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. *Mol Syst Biol.* **3**: 79
- Sikka VK, Choi SB, Kavakli IH, Sakulsingharoj C, Gupta S, Ito H, Okita TW** (2001) Subcellular compartmentation and allosteric regulation of the rice endosperm ADPglucose pyrophosphorylase. *Plant Sci.* **161**: 461-468
- Tadege M, Dupuis I, Kuhlemeier C** (1999) Ethanolic fermentation: new functions for an old pathway. *Trends Plant Sci* **4**: 320-325
- Tetlow IJ** (2006) Understanding storage starch biosynthesis in plants: a means to quality improvement. *Canadian Journal of Botany* **84**: 1167-1185
- Ueguchi-Tanaka M, Nakajima M, Katoh E, Ohmiya H, Asano K, Saji S, Hongyu X, Ashikari M, Kitano H, Yamaguchi I, Matsuoka M** (2007) Molecular interactions of a soluble gibberellin receptor, *GID1*, with a rice *DELTA* protein, *SLR1*, and gibberellin. *Plant Cell* **19**: 2140-2155
- van Dongen JT, Roeb GW, Dautzenberg M, Froehlich A, Vigeolas H, Minchin PEH, Geigenberger P** (2004) Phloem import and storage metabolism are highly coordinated by the low oxygen concentrations within developing wheat seeds. *Plant Physiol.* **135**: 1809-1821
- Verza NC, E Silva TR, Neto GC, Nogueira FTS, Fisch PH, de Rosa VE, Jr., Rebello MM, Vettore AL, da Silva FR, Arruda P** (2005) Endosperm-preferred expression of maize genes as revealed by transcriptome-wide analysis of expressed sequence tags. *Plant Mol Biol.* **59**: 363-374
- Vienna A, Vienna A, GmbH GD, Martinsried G** (2000) Observations on the reproducibility and matching efficiency of two-dimensional electrophoresis gels: Consequences for comprehensive data analysis. *Electrophoresis* **21**: 3345-3350
- Wang AY, Kao MH, Yang WH, Sayion Y, Liu LF, Lee PD, Su JC** (1999) Differentially and developmentally regulated expression of three rice sucrose synthase genes. *Plant Cell Physiol.* **40**: 800-807
- Watson BS, Asirvatham VS, Wang L, Sumner LW** (2003) Mapping the proteome of barrel medic (*Medicago truncatula*). *Plant Physiol.* **131**: 1104-1123
- Weber H, Borisjuk L, Wobus U** (2005) Molecular physiology of legume seed development. *Annual Review of Plant Biology* **56**: 253-279
- Winter H, Huber SC** (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Critical Reviews in Plant Sciences* **19**: 31-68

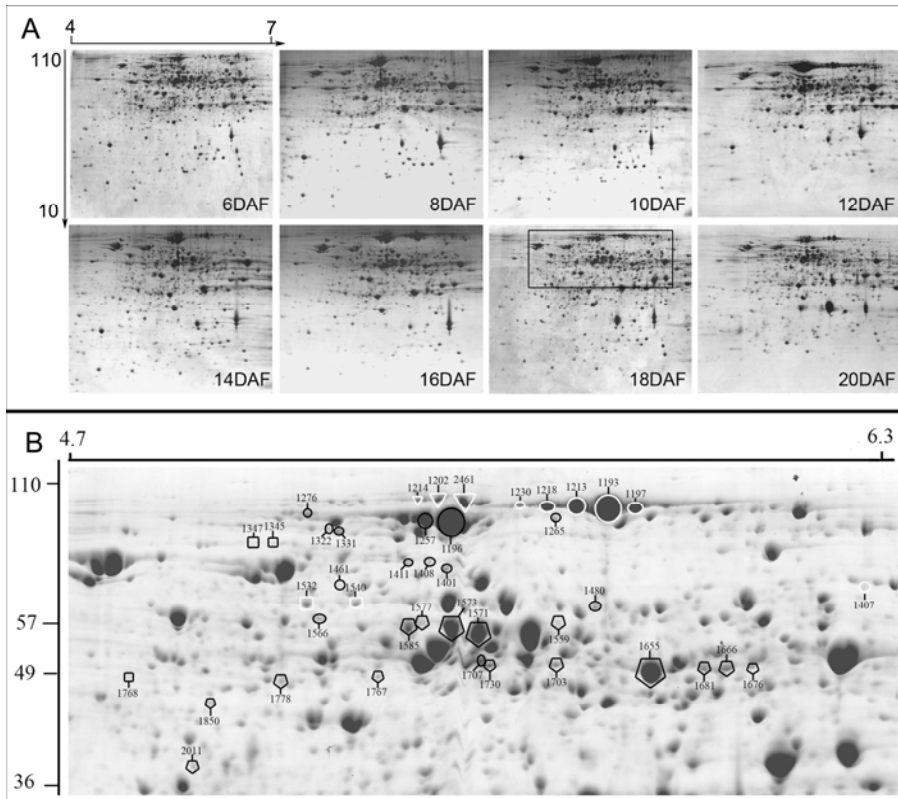
- Wobus H, Borisjuk L, Weber U** (1997) Sugar import and metabolism during seed development. *Trends Plant Sci* **2**: 169-174
- Xie Z, Zhang ZL, Zou X, Yang G, Komatsu S, Shen QJ** (2006) Interactions of two abscisic-acid induced WRKY genes in repressing gibberellin signaling in aleurone cells. *Plant J.* **46**: 231-242

Table 1. Hierarchical clusters of differentially expressed proteins and distribution of the proteins involved in each category or sub-category in different clusters. The clusters (c0~c4) was created by use of GeneCluster 2.0; raw data for the clusters were listed in Supplementary Table S4.

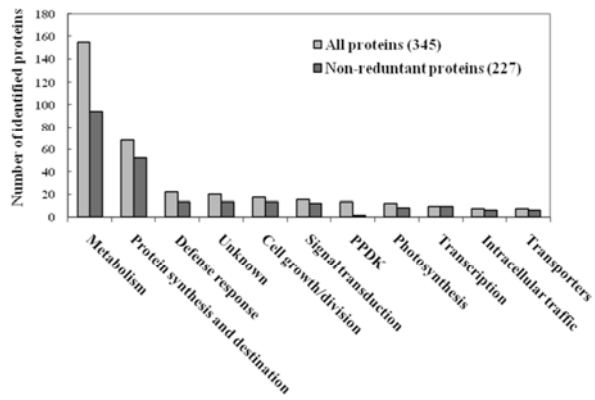
<i>Categories or Sub-categories</i>	 c0: 76	 c1: 87	 c2: 35	 c3: 34	 c4: 63	<i>total</i>
01 Metabolism	30	23	17	14	36	120
01.01 Sugars conversion	4	5	2	1	4	16
01.02 Glycolysis	11	3	1	1	5	21
01.03 Alcoholic Fermentation	1	1	0	2	5	9
01.04 Pentose phosphate	0	1	0	0	1	2
01.05 TCA pathway	0	0	3	0	0	3
01.06 starch synthesis	2	0	2	7	12	23
01.07 Amino acid metabolism	4	4	3	2	1	14
01.08 Nitrogen and sulfur metabolism	1	2	2	0	1	6
01.09 Nucleotides metabolism	2	1	1	1	3	8
01.10 Lipid and sterol metabolism	1	2	3	0	0	6
01.11 Secondary metabolism	4	3	0	0	4	11
02 Protein synthesis and destination	24	13	6	9	11	63
02.01 Protein synthesis	4	4	5	0	0	13
02.02 Protein folding and modification	11	2	1	7	10	31
02.03 Proteolysis	9	7	0	2	1	19
03 Cell growth/division	2	8	0	2	1	13
04 Signal transduction	2	6	0	2	2	12
07 Transcription	2	2	2	1	1	8
08 PPK	1	4	2	0	0	7
09 Defense response	3	6	3	3	4	19
10 Photosynthesis	5	4	1	0	1	11



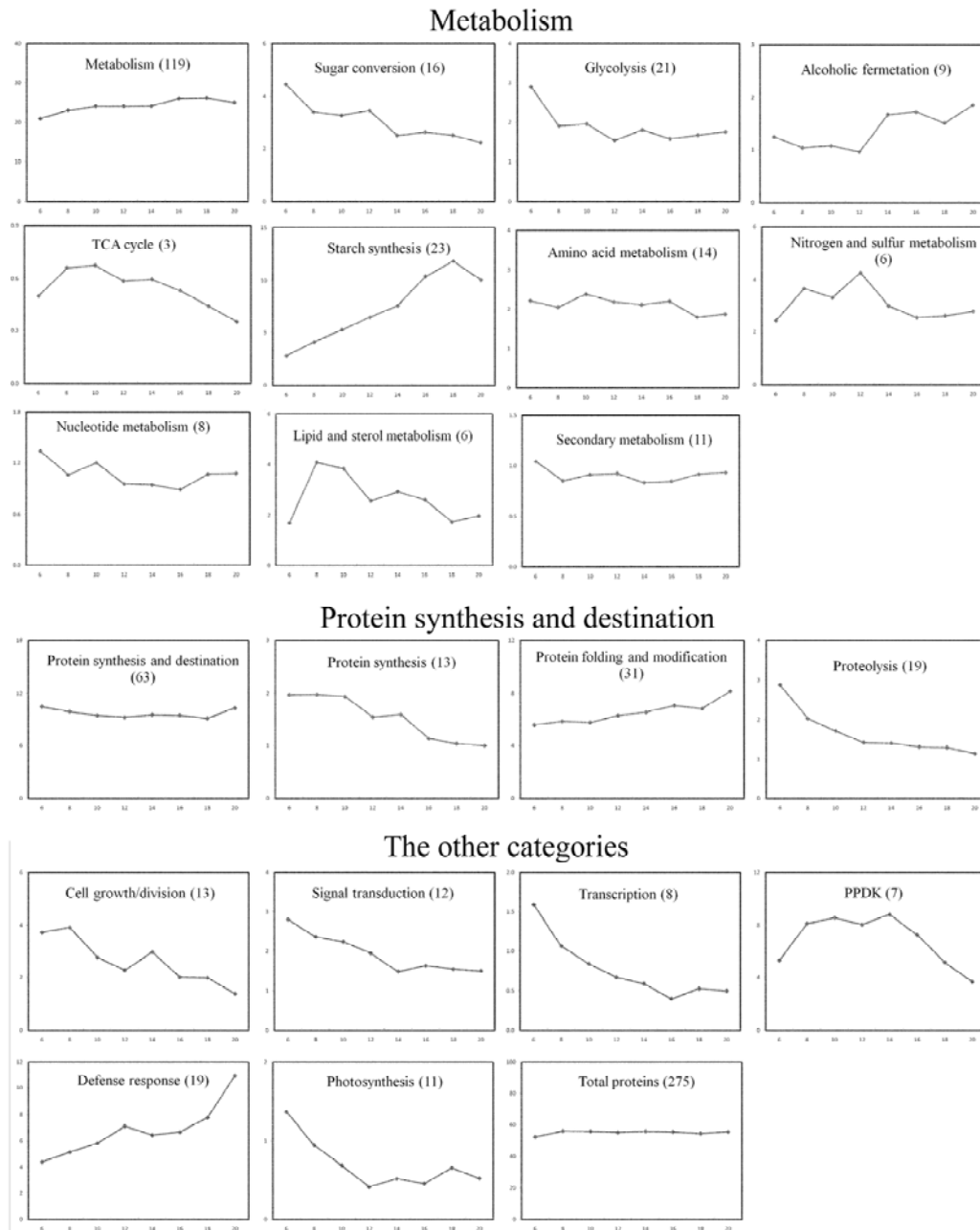
**Figure 1.** Development of rice seeds. A, whole seeds at the 10 stages (DAF-days after flowering) of seed development. B, I<sub>2</sub>-KI stained developing seed showing starch storage at 9 stages. C, Sectional profiles of blue line rectangle in B. D, Fast green stained embryos showing the development of rice embryos. Scale bar, 2 mm in A and B, 30 μm in C, 100μm (for 2 DAF and 4 DAF) and 1 mm (for 6 – 20 DAF) in D. E, Change in fresh and dry weight of developing seeds (mg/seed), 100 seeds was analyzed at each stage at least. F, Change in ATP levels in developing seeds. FW, fresh weight. Bars represent the standard deviations of three replicates.



**Figure 2.** Representative 2-DE image for each of 8 protein samples (A) and close-up of possible isoforms of some proteins (B). A, proteins prepared from developing seeds at 6, 8, 10, 12, 14, 16, 18 and 20 DAF were separated by 2-D PAGE and stained by CBB. The differentially expressed protein spots throughout the 8 stages were determined according to the method described in Materials and Methods, their relative volume values, identities and expression patterns are in Supplemental Tables S1, S2 and S4, respectively. B, Close-up of a part of a 18 DAF sample gel in (A) to show isoforms PPK and some of starch synthesis-related proteins, whose identities are in Supplemental Table S2. PPK labeled in the black-line circles, pullulanase in white-line circles, starch phosphorylase in white-line triangles, isoamylase in black-line rectangles, plastidic phosphoglucomutase in white-line rectangles and APGase in black-line pentagons. MM in kilodaltons and pI of proteins shown on the left and top of the image, respectively.

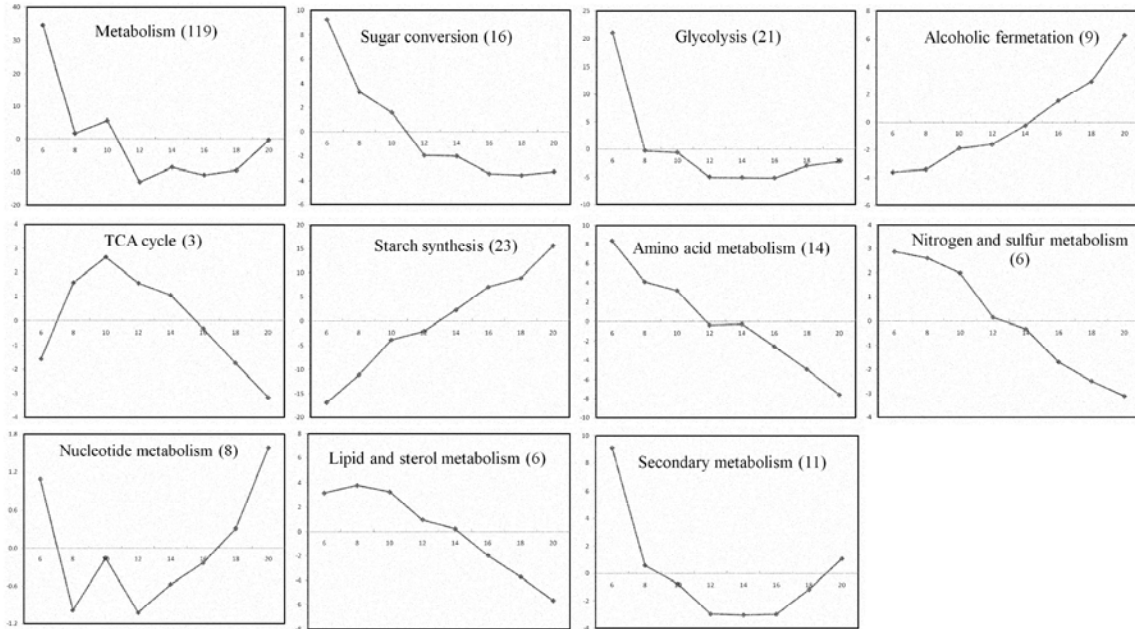


**Figure 3.** Functional classification of the identified differentially expressed proteins based on the scheme by Bevan et al. (1998). Gray bars show functional distribution of all 345 proteins (identities) with the distribution of 227 unique proteins in black bars.

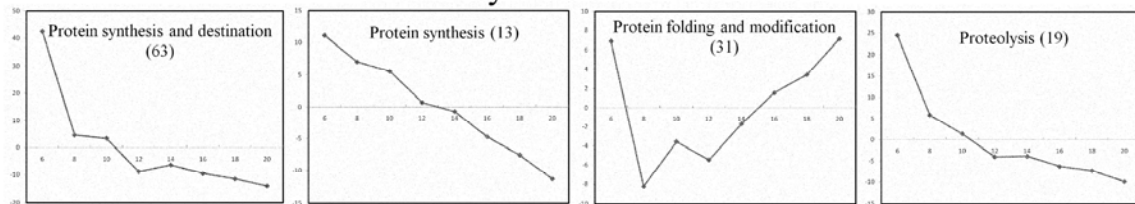


**Figure 4.** Composite expression profiles of protein groups associated with different functional categories and sub-categories. The profiles were established by the sum of all relative volume (RV) values for protein components in a given function category/sub-category (vertical axis) at each developmental stage (DAF) (horizontal axis). The total spot number used to draw the composite profiles are indicated in brackets.

## Metabolism



## Protein synthesis and destination



## The other categories

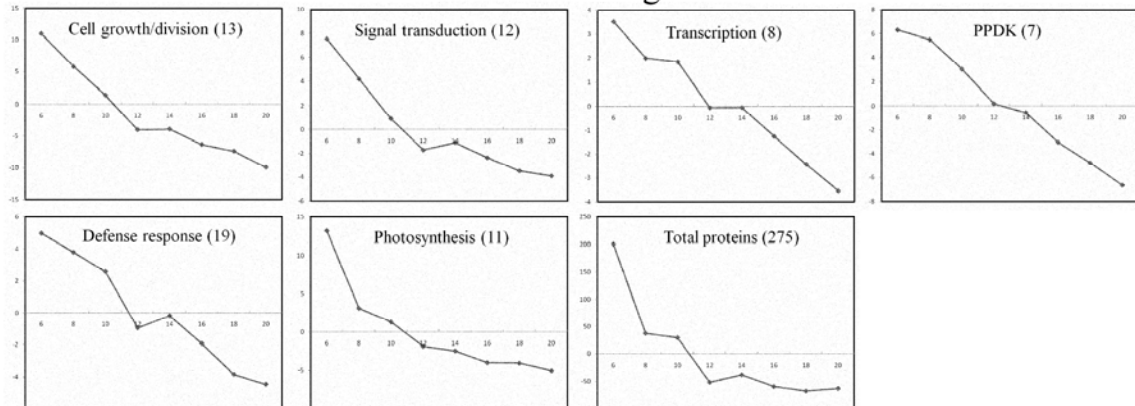
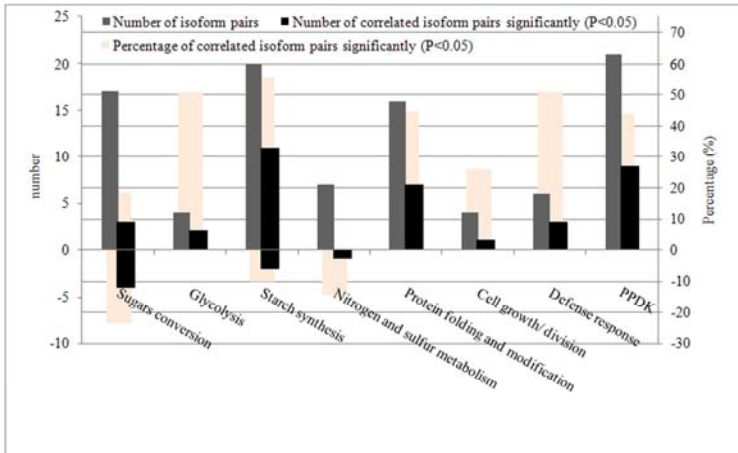
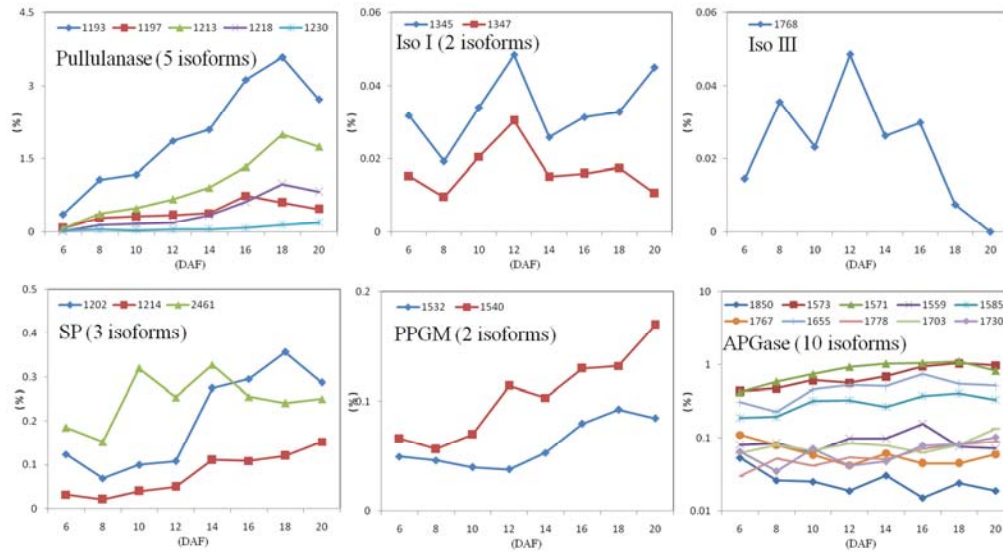


Figure 5. Digital expression tendencies of protein groups associated with different functional categories and sub-categories. The expression tendencies were constructed according to the equation:  $y(6 \text{ DAF}, 8 \text{ DAF}, 10 \text{ DAF}, 12 \text{ DAF}, 14 \text{ DAF}, 16 \text{ DAF}, 18 \text{ DAF}, 20 \text{ DAF}) = \text{centroids}(c_0) * \text{amount}(c_0) + \text{centroids}(c_1) * \text{amount}(c_1) + \text{centroids}(c_2) * \text{amount}(c_2) + \text{centroids}(c_3) * \text{amount}(c_3) + \text{centroids}(c_4) * \text{amount}(c_4)$  for each category or sub-category. Centroids data of each hierarchical cluster was showed in Supplemental Table S6. The total spot number used to construct the expression tendencies are indicated in brackets.





**Figure 6.** Correlation analysis of isoform pairs. The number of isoform pairs in each category or sub-category are in grey bars. The number of positively and negatively correlated isoform pairs are shown in black bars as positive and negative numbers in the left vertical axis. The percentage of positively and negatively correlated isoform pairs are shown in light bars as positive and negative numbers in right vertical axis. Only a category or sub-category with more than 4 isoform pairs was analyzed for statistical reasons. The statistical data are in Supplemental Table S5.



**Figure 7.** Expression profiles of proteins involved in starch metabolism. The vertical axis shows the relative volume (RV) value of each isoform (except APGase) at each developmental stage (days after flowering [DAF]) (horizontal axis). For APGase, the RV value was converted into a logarithmic scale to describe each isoform expression profile in a wider extent. Iso I, isoamylase I; Iso III, isoamylase III. SP, starch phosphorylase; PPGM, plastidic phosphoglucomutase.