African Journal of Agricultural Research Vol. 6(4), pp. 808-816, 18 February, 2011
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR10.1059
ISSN 1991-637X ©2011 Academic Journals

Full Length Research Paper

Proteomics-based identification of storage, metabolic, and allergenic proteins in wheat seed from 2-DE gels

Liming Yang[1], Dagang Tian[2], Yuming Luo[1]*, Ruiyue Zhang[1], Chongmiao Ren[1] and Xin Zhou[1]

[1]School of Life Sciences, Huaiyin Normal University, Jiangsu Key Laboratory for Eco-Agricultural Biotechnology around Hongze Lake, Huai'an 223300, Jiangsu, China.
[2]Biotechnology Institute, Fujian Academy of Agricultural Sciences, Fuzhou 350003, China.

Accepted 21 January, 2011

Proteomic approach combining two-dimensional gel electrophoresis (2-DE), mass spectrometry and bioinformatics analysis were applied to identify major proteins in extracts of mature wheat seeds. About 920 or 700 protein spots were detected on 2-DE gels by silver staining or colloidal Coomassie Brilliant Blue staining. Eighty spots with higher abundance were selected to cut for in-gel digestion followed by matrix-assisted laser desorption/ionization-mass spectrometry-peptide map fingerprint analysis and electrospray ionization mass spectrometry-peptide sequence tags analysis. Database searches using measured peptide masses or peptide sequence tags querying wheat expressed sequence tags determined protein identities of 73 spots. These identified proteins were categorized into six classes according to the functional annotation including those with unknown functions and difficult to classify. Proteins involved in storage proteins, metabolism, defense, chaperones and allergy were the major categories. The present identification of proteins in major spots from 2-D gels includes 11 different proteins from 29 spots from wheat seed extract. It is suggested that post-translational processing or isoforms causes the same proteins to occur in different spots. In addition, we also discussed the efficiency of protein identification using species-specific EST databases.

Key words: Wheat seed, proteome profile, protein identification, expressed sequence tags, allergens, storage proteins.

INTRODUCTION

Cereal seed proteins are of importance for human and animal nutrition, plant breeding, and cultivar identification (Skylas and Wrigley, 2004; Yahata et al., 2005). The compositions of all these proteins determine the milling and processing quality of the seed. The term 'proteome' was originally coined by Wasinger et al. (1995), to study the full complement of polypeptides expressed by the genes of an organism in a specific tissue. Proteomics has the potential to contribute to cereal science and to our understanding of grain quality by elucidating the ways in which the genes are expressed during seed maturation under given environmental conditions (Skylas and Wrigley, 2004). Accordingly, there is the need to study protein composition more completely, as in the proteomic approach.

*Corresponding author. E-mail: yumingluo@163.com. Tel: 86-0517-83525086. Fax: 86- 0517-83525992.

Two-dimensional gel electrophoresis with isoelectric focusing (IEF) in the first dimension and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension allows improved resolution of proteins in complex seed extracts with high reproducibility. Current proteomics, including high resolution 2-DE gel electrophoresis, in-gel proteolytic digestion of protein spots and protein identification by mass spectrometry and database searches, is increasingly used to address questions of development, physiology and quality in major crop plants (Bahrman et al., 2004; Vensel et al., 2005). This methodology was also used for analysis of seed protein extracts from soybean (Danchenko et al., 2009), peanut (Kottapalli et al., 2008), and barley (Østergaard et al., 2004).

Wheat is one of the most important cereal crops for the global food supply, analysis of the major protein complement from wheat seed is essential and important. In the present study, the most abundant proteins found in aqueous extracts from mature wheat seeds are separated

by 2-DE and identified by matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-TOF MS) or electrospray ionization mass spectrometry (ESI-MS/MS) data querying wheat expressed sequence tags (EST) database and EST contig, the identified proteins were categorized according to the functional annotation. The present results make a central contribution to a comprehensive analysis of the protein profile of wheat seeds. In addition, the strategy of protein identification using specific-species EST or EST contig sequences is commented. It will be an efficient method for protein identification of species without full genome sequence.

MATERIALS AND METHODS

Materials and extraction of wheat seed proteins

Wheat cultivar, Ourou, a primal parent of wheat in China, was used. Protein extraction was conducted according to Damerval et al. (1986) with modification. Wheat seed was ground and homogenized, the proteins were precipitated at -20℃ with 10% (w/v) trichloroacetic acid (TCA) in acetone containing 0.07% (w/v) dithiothreitol (DTT) for 1 h. After it was centrifuged for 30 min at 15000 g at 4℃, the precipitates were washed with ice-cold acetone containing 0.07% (w/v) DTT to remove polysaccharide, pigments and lipids until the pellets were colorless. The pellets were dried by vacuum centrifugation, and resuspended for 30 min at room temperature in extraction buffer containing 7 M urea, 2% M thiourea, 4% (w/v) 3 -[(3-cholamidopropyl)-dimethylammonio]-1-propane sulfonate (CHAPS), 20 mM DTT, 0.2% (v/v) carrier ampholyte (pH 3.0 to 10.0), protease inhibitor cocktails in an amount of 1 µl per 30 mg plant tissue (Sigma). During the suspension, the mixture was sonicated five times on ice, 30 s per sonication. Supernatant was obtained by centrifuging at 40000 g for 30 min and stored at -80℃. Protein extracts were quantified using 2-D Quant Kit (Amersham Biosciences).

Two-dimensional gel electrophoresis

Protein samples prepared were separated by iso-electrophoresis using the immobilized pH gradient (IPG) strips (pH 3.0 to 10.0, NL, 17 cm) in the Protean system (BioRad). IPG strips were redydrated passively for 13 h. The voltage settings used were as follows: 250 V 1 h, 500 V 1 h, 2000 V 1 h, then 5000 V that was maintained until a total of 60000 Vh was reached. After IEF and equilibration, the second dimensional electrophoresis was conducted in 12.5% SDS-PAGE gels at 2 W/gel for 45 min and 12 W/gel for 4 h using the Multiphor system (Amersham Biosciences). Duplicate gels were prepared under the same conditions, and were subjected to silver staining and colloidal Coomassie Brilliant Blue (cCBB) staining to visualize the protein spots (Neuhoff et al., 1985; Shevchenko et al., 1996). The visualized protein spots were then isolated for mass spectrometry.

MALDI-TOF MS analysis of protein spots and protein identification

The protein profile was analyzed and quantified by PDQuest software (Bio-Rad), and protein spots with the highest abundance were cut, and subjected to digestion by trypsin (Promega, USA). In-gel digestion of protein spots, MALDI-TOF-MS analysis and database queries against the NCBInr protein database followed the procedures described in our previous researches (Yang et al., 2010). MS fingerprinting searches were also performed by Mascot program (http://www.matrixscience.com) against wheat EST sequences from NCBI dbEST databse, with a mass tolerance of 50 ppm. The EST sequence matched was again used to search EST homologs with at least 40- base overlap and 94% sequence similarity in the wheat EST database to construct cDNA contigs.

Final sequence coverage was then estimated using peptide sequence fragments from MALDI-TOF MS that matched in m/z values to those obtained by in-silico digestion of the conceptual proteins translated from the contigs using Masspeptide program (http://www.expasy.ch) with the settings of Cys by carbamidomethylation, oxidation of Met, trypsin digestion with one incomplete cleavage, and a mass tolerance of 50 ppm. Protein identification was assigned when the following criteria were met: at least four matching peptides, >20% sequence coverage. The PMF data were also used to search plant protein database (Swiss-Prot, TrEMBL and NCBInr) using ProFound (http://129.85.19.192/profound_bin/WebProFound.exe) and Mascot program (http://www.matrixscience.com) with the following parameters: monoisotopic peptide mass ([M+H]+), trypsin digestion with one incomplete cleavage, a mass tolerance of 50 ppm, cysteine treated by iodoacetamide, oxidation of methionine.

ESI-MS/MS analysis of protein spots and protein identification

High Performance Liquid Chromatography – electrospray tandem mass spectrometry (HPLC ESI-MS/MS) analysis was conducted as following: LC was performed on a surveyor LC system (Thermo Finnigan). The C18 column was obtained from Column Technology (New York, USA). Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1% formic acid in acetonitrile. The tryptic peptide mixtures were eluted using a gradient of 2 to 98% B over 180 min. The MS/MS was performed on a LTQ linear ion trap mass spectrometer (Thermo Finnigan) equipped with an electrospray interface and operated in positive ion mode. The capillary temperature was set to 170℃, and the spray voltage at 3.4 kV. Full scan mass spectra were obtained by the Orbitrap at 300 to 2000 m/z with a resolution of 30,000. The three most intense ions were determined at a threshold above the 1000 ion trap at a normalized collision energy of 35%. Automated peak recognition, dynamic exclusion, and daughter ion scanning of the top two most intense ions were performed using the Xcalibur software. The peptide tags were submitted to sequence query in a local SEQUEST or MASCOT program (http://www.matrixscience.com) for database searching against wheat EST sequences.

The search parameters used were as follows: monoisotopic peptide masses, 0.2 Da peptide mass tolerance; one missed cleavage, modifications allowed for oxidation of methionine and carboxyamidomethylation of cysteine. The criteria for a preliminary positive peptide identification for a doubly-charged peptide were a correlation factor (Xcorr) greater than 2.5, a delta cross-correlation factor (□Xcorr) greater than 0.1 (indicating a significant difference between the best match reported and the next best match), a minimum of one tryptic peptide terminus, and a high preliminary scoring. For triply-charged peptides the correlation factor threshold was set at 3.5. All matched peptides were confirmed by visual examination of the spectra.

Functional annotation of identified proteins

The functions of proteins identified were annotated by querying against the protein function database Pfam (http://www.sanger.ac.uk/Software/Pfam/; Bateman et al., 2002) or Inter-Pro (http://www.ebi.ac.uk/interpro/; Apweiler et al., 2001). The

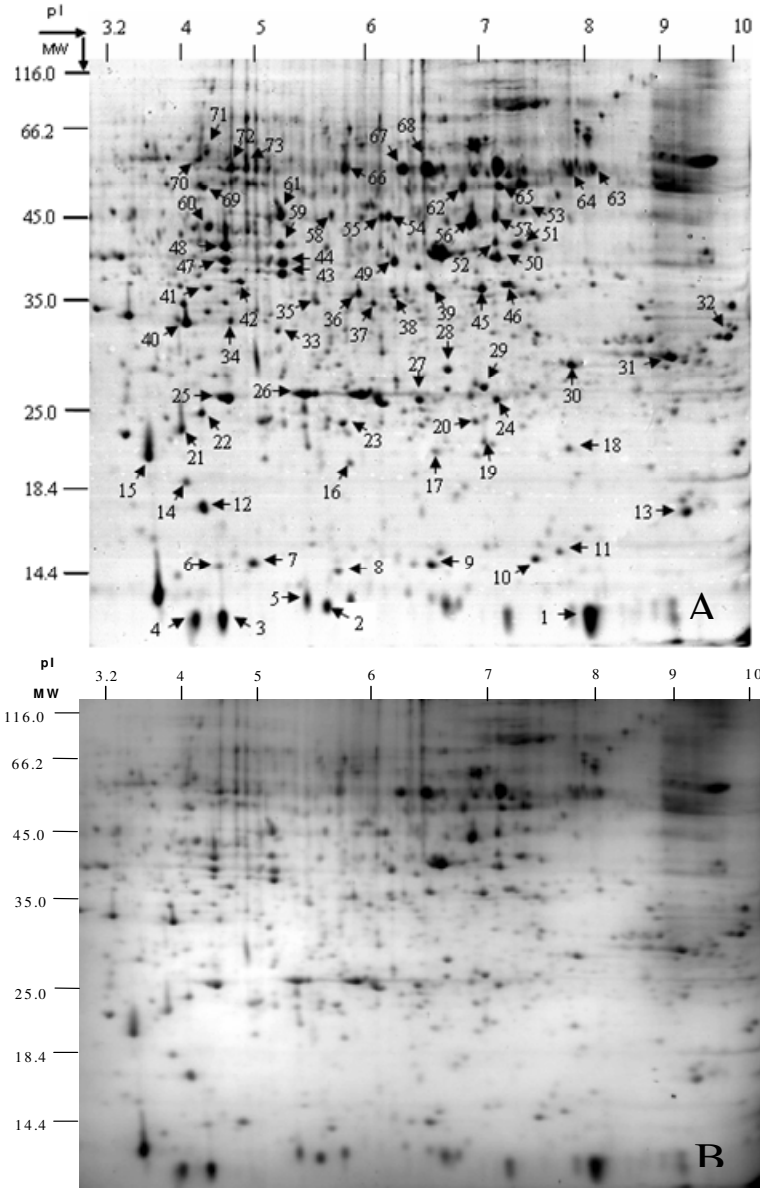

Figure 1. Proteome profile of wheat seed visualized by silver staining. (A) Profile of silver staining (B) Profile of cCBB staining. Molecular masses are indicated in kDa on the right axis and approximate pI along the top axis. Numbers correspond to the identificated spots listed in Table 1.

software KEGG (http://www.genome.jp/KEGG/pathway; Kanehisa and Goto, 2000) was used to analyze the biochemical pathways in which identified proteins participated.

RESULTS

Wheat seed proteome patterns of silver staining and colloidal coomassie brilliant blue staining

Protein extract of wheat seed was separated by 2-DE. Highly resolved 2-D gel patterns were shown by silver staining of 300 µg protein and colloidal Coomassie Brilliant Blue (cCBB) staining of 750 µg protein applied to the gels, respectively (Figure 1). The proteins were well separated in the pI range of 4.5 to 8.5 and the size range of 15 to 65 KDa. The spot pattern was highly reproducible for independent extractions for two different batches of seeds.

The well-known superior sensitivity of silver staining enabled analysis of faint spots not revealed by cCBB staining, but the resulting mass spectra of these spots were of low quality and/or did not result in successful identifications. Although the position and intensity were

comparable for the vast majority of spots on silver and cCBB-stained gels, occasional conspicuous differences in relative spot intensities were observed, most probably as a consequence of the narrow dynamic range of silver staining (Rabilloud and Charmont, 2000; Østergaard et al., 2004). Thus, we selected 80 spots with high abundance detected in cCBB-stained and silver-stained 2-DE gels for further analyses.

Protein identification using wheat EST sequences and protein sequences from viridiplantae

Protein identification was carried out from cCBB-stained and silver-stained 2-D gels. Generally, superior mass spectra were obtained using cCBB staining due to the larger amount of sample applied (Figure 2). In addition, two or more proteins were identified in the same spot in seven cases. 73 identifications were made from silver-stained and cCBB-stained gels, respectively, including 10 identifications made from both gels that verified superimposition of 2-D gel patterns. For clarity, identifications from the silver-stained and the cCBB-stained 2-DE gels are all marked on Figure 1. In the 2-DE gels of wheat seeds, 80 protein spots with high abundance were cut out, and subject to in-gel digestion and analyzed by MALDI-TOF MS or ESI-MS/MS. Protein identification using the MS data was firstly conducted by querying the Swiss-Prot, TrEMBL and NCBInr protein databases including sequences from viridiplantae (green plants). The data of most protein spots showed match to proteins of rice, maize or *Arabidopsis thaliana*. This is understandable because of the under-representation of wheat proteins in the databases. Protein identification, however, was not always successful even if good mass spectra were obtained, since a limited number of wheat protein sequences are available in protein databases. Further attempts were made to search using available EST sequences from wheat.

Protein identities were also determined by querying MS fingerprinting data or peptide sequence tags against wheat EST databases. Through wheat EST database query, the identities of 73 proteins among these spots were determined as listed in Table 1. The annotations of identified proteins were assigned by comparing their conceptual sequences to the public protein database through NCBI-BLAST server. Homologue proteins with the highest sequence similarities are listed under the corresponding annotation for each identified proteins in Table 1, and wherever possible, the well-studied homologue proteins are preferred to be included. Additionally, 13 proteins of the 73 identified proteins showed the characteristic with multiple members of a protein family (Table 1), in which one entry with the highest score was selected and described in this paper. Taken together, the 73 identified proteins represent 46 unique proteins (unipros) (Table 1). Identities of five protein Spots were further analyzed by ESI-MS/MS, and confirmed the results of PMF identification.

In addition, different searching approaches using Mascot program and ProFound program (http://prowl.rockefeller.edu/prowl-cgi/profound.exe) querying wheat EST database or plant protein databases were conducted. The identities of 29 protein spots were the same between ProFound querying, using plant protein databases and Mascot querying wheat EST database. The identities of 25 protein spots were the same using Mascot program querying plant protein databases or wheat EST database. Among them, 20 protein spots got the same identification using the aforementioned, three querying approaches (Figure 2).

Proteins identified in the extracts of wheat seeds and functional classification

All of the identities were mainly classified into six major categories based on functional database Pfam or Inter-Pro and the classification schema, including: storage proteins (8 proteins in 13 spots), metabolism (17 proteins in 22 spots), defense (5 proteins in 6 spots), chaperones (3 proteins in 4 spots), allergen-like proteins (20 proteins in 22 spots) and other proteins with unknown functions and difficult to classify (Table 1, Figure 3).

Storage proteins

The proteins involved in this category included, late embryogenesis abundant protein, which was identified in five spots (id# 2, 11, 18, 40, 48), embryo-specific protein in three spots (id# 25, 26, 38), Gamma gliadin in three spots (id# 45, 46, 56), embryo-specific protein (id# 38) and embryonic abundant protein (id# 47). These proteins have also been identified in barley seeds 2-DE gels (Østergaard et al., 2004).

Proteins involved in metabolism

Some of the spots of this class constitute the key enzymes of the metabolism processes. Many enzymes of the glycolytic pathway were identified in multiple spots. Glyceraldehyde-3-phosphate dehydrogenase was identified in three spots (id# 16, 39, 49), triosephosphate isomerase in three spots (id# 33, 54, 55), malate dehydrogenase in two spots (id# 43 and 44), two isoforms of enolase in two spots (id# 66 and 67). Many glycolytic enzymes are also involved in the conversion of sucrose to starch.

Defence proteins

Proteins involved in defence processes include Subtilisin-chymotrypsin inhibitor CI-1A (id# 3 and 4), heat

Table 1. Proteins identified on wheat seed 2-DE gels by MS querying and protein categories.

Spot No.[a]	Protein name	Observed MW (KDa) /pI[b]	Coverage/ Number of matched peptides	Accession Gi No.[c]
	Storage protein			
2	Late embryogenesis abundant protein B19.1A	12.0/5.6	41%/5	32680432
11	LEA2 protein	15.9/7.6	41%/7	23401553
18	LEA3 protein	22.3/7.8	38%/6	23150239
25	embryo-specific protein	26.5/4.7	53%/11	32684826
26	embryo-specific protein	26.5/5.4	49%/10	32684826
36	embryo globulin	36.0/5.9	34%/14	32684794
38	Embryo-specific protein	35.5/6.3	36%/8	32686512
40	Late embryogenesis abundant protein	33.5/4.1	46%/9	23090778
45	Gamma gliadin	36.0/7.0	33%/8	51528454
46	Gamma gliadin	36.8/7.3	41%/11	51528142
47	Embryonic abundant protein	40.2/4.6	28%/6	9444730
48	Late embryogenesis abundant protein	43.6/4.6	34%/9	93229646
56	Gamma gliadin	45.0/6.9	29%/7	25438817
	Metabolic protein			
16	cytosolic glyceraldehyde-3-phosphate dehydrogenase	20.3/5.8	65%/13	38983600
21	putative 1,3-beta-glucan synthase 10	23.6/4.0	26%/6	22547020
22	Soluble inorganic pyrophosphatase	25.0/4.3	34%/8	93670693
32	Protein synthesis inhibitor 1	33.7/9.7	47%/10	19095022
33	Triosephosphate isomerase, (cytosolic)	33.1/5.3	51%/14	93259044
37	Aldose reductase	35.0/6.1	35%/8	32771886
39	Glyceraldehyde-3-phosphate dehydrogenase (cytosolic)	36.1/6.6	42%/9	9839186
43	Malate dehydrogenase (cytoplasmic)	38.2/5.3	62%/17	25412915
44	Malate dehydrogenase (mitochondrial)	39.8/5.3	58%/15	11565658
49	Glyceraldehyde-3-phosphate dehydrogenase (cytosolic)	39.8/6.3	46%/14	38991119
50	esterase/lipase/thioesterase family protein-like	41.2/7.2	27%/8	93245391
51	glucan endo-1,3-beta-D-glucosidase	43.8/7.4	47%/31	45798581
54	triosephosphat-isomerase	45.4/6.3	38%/11	28824377
55	triosephosphat-isomerase	45.3/6.2	41%/13	20797113
58	Fructose bisphosphate aldolase (cytoplasmic)	46.1/5.7	57%/21	14011311
59	Glyoxalase I	43.4/5.3	31%/7	92491888
60	Phosphoglycerate kinase	44.1/4.3	29%/8	25427918
61	Alcohol dehydrogenase	46.1/5.3	37%/12	23077653
66	Enolase 1	62.1/5.8	49%/22	38879861
67	Enolase 2	62.1/6.4	43%/19	13965665
68	ATP synthase beta subunits	54.1/6.6	56%/23	12029018
69	Phosphoglycerate mutase	50.5/4.3	31%/17	93224271
	Defense protein			
3	Subtilisin-chymotrypsin inhibitor CI-1A	11.2/4.6	33%/6	19100015
4	Subtilisin-chymotrypsin inhibitor CI-1A	11.2/4.3	41%/7	19086415
20	putative thioredoxin peroxidase	24.2/7.0	36%/8	25214277
29	putative heat stress protein	27.2/7.0	41%/11	18354135
57	L-ascorbate peroxidase 6, chloroplast precursor	46.1/7.2	67%/17	38820035
65	xylanase inhibitor TAXI-IV	54.0/7.2	38%/24	56201269
	Chaperones			
15	Class I small heat shock protein	20.1/5.8	34%/9	5732974
62	Low molecular-weight glutenin subunit group 3 type II protein	52.7 /6.9	46%/17	32675644

Table 1. Contd.

70	protein disulfide isomerase	63.6/4.2	36%/21	62531778
71	Protein disulfide isomerase	64.1/4.3	42%/24	20432753
	Allergen-like proteins			
1	Non specific lipid-transfer protein	10.6/8.1	32%/5	100037539
5	α-Amylase inhibitor BDAI-I	13.0/5.5	45%/7	12004647
6	Alpha-amylase/trypsin inhibitor CM16 precursor	15.1/4.4	39%/9	32657311
7	Trypsin/α-amylase inhibitor pUP13	15.2/5.0	28%/6	15548978
8	α-Amylase inhibitor BMAI-1	14.8/5.8	43%/7	12004647
10	monomeric alpha-amylase inhibitor	15.4/7.5	37%/6	32766713
12	α-Amylase/trypsin inhibitor CMb	17.5/4.4	42%/8	13026591
13	Alpha-amylase/trypsin inhibitor CM3 precursor	17.2/9.3	36%/7	21557909
14	Superoxide dismutase (Cu-Zn) 2	18.6/4.1	47%/9	239862757
17	Thaumatin-like protein TLP4	22.0/6.7	55%/14	110237388
23	glutathione transferase	24.5/5.7	48%/13	38991833
24	manganese superoxide dismutase	26.5/7.2	32%/8	9445748
27	Glutathione S-transferase	26.0/6.5	42%/10	55676966
28	Superoxide dismutase (Mn)	29.1/6.7	27%/7	51921490
30	1,3-b-Glucanase	29.3/7.8	31%/8	177815277
31	Thioredoxin	30.4/9.1	46%/12	21997939
34	Chitinase 1	33.5/4.7	29%/9	55684961
41	Isoflavone reductase	36.8/4.4	36%/8	20051261
42	Isoflavone reductase related protein	37.2/4.7	33%/8	20051261
53	peroxidase	46.9/7.4	51%/16	38793435
72	Beta-amylase	62.9/4.7	32%/14	85846549
73	Beta-amylase	62.9/4.9	35%/15	51527739
	Proteins related to other function			
9	leucine-rich repeat protein	15.3/6.7	32%/5	21556990
19	putative GTP-binding protein yptm3	23.1/7.1	25%/5	20436439
35	20S proteasome beta 5 subunit	35.0/5.6	24%/6	21835050
52	putative NAC domain protein NAC1	43.0/7.2	37%/14	30253927
63	hypothetical protein	54.4/8.3	29%/12	38697091
64	putative microtubule-associated protein	54.1/7.8	32%/13	12013754

[a] Number of the spot in the gel.[b] observed molecular mass (MW) and pI of protein spot in the gels. [c] Accession number in National Center for Biotechnology Information (NCBI).

Figure 2. Comparison of protein identification using species-specific EST sequences and plant protein databases.

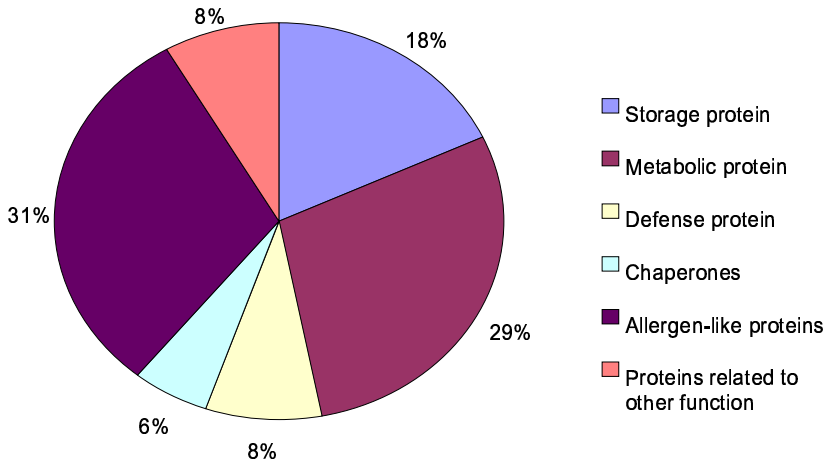

Figure 3. Functional classification of proteins identified in wheat seeds.

stress protein (id# 29), xylanase inhibitor TAXI-IV (id# 65). Peroxidase-like proteins were also identified in two spots (id# 20 and 57).

Chaperones

Two different types of heat shock proteins (HSPs) were identified in the 2-D gel pattern. The corresponding proteins have been identified from barley seeds in several spots from 2-D gels (Østergaard et al., 2004). Protein disulfide isomerase (PDI) that catalyzes correct protein disulfide bond pairing was identified in two spots (Figure 1; id# 70 and 71) varying from lower pI and Mr to higher pI and Mr.

Allergen-like proteins

In our research about wheat seed proteomic profiles, 22 allergenic-like proteins were detected, including alpha-amylase inhibitor (id# 5, 8, 10), alpha-amylase/trypsin inhibitor (id# 6, 7, 12, 13), superoxide dismutase (id# 14, 24, 28), thaumatin-like protein (id# 17), isoflavone reductase (id# 41 and 42), beta-amylase (id# 72 and 73), non-specific lipid-transfer protein (id# 1), glutathione transferase (id# 23), glutathione S-transferase (id# 27), 1,3-b-Glucanase (id# 30), thioredoxin (id# 31), chitinase 1 (id# 34), peroxidase (id# 53). In addition, some proteins, such as gamma gliadin, embryo globulin, glycerinaldehyde-3-phosphate dehydrogenase, triosephosphate isomerase, which have been considered to be major allergens in patients with bakers' asthma, a typical occupational allergic disease, is categorized into storage proteins or metabolic proteins.

Proteins with other functions and unknown functions

Several proteins were identified that did not fall into the aforementioned categories of identified proteins, or for which the function is not known. Some of the proteins are involved in essential processes carried out in living cells. GTP-binding protein yptm3 (id# 19) and 20S proteasome beta 5 subunit (id# 35) was involved in signal transduction; microtubule-associated protein (id# 64) may play a role in cell division.

DISCUSSION

Proteomics has become a major field of molecular genetics research for expression studies. Detection, identification and quantification by 2-DE of proteins in tissues, organs and organelles in a biological entity constitute an essential step in gathering genomic knowledge. Proteome analyses by 2-DE for building databases of wheat leaf and amyloplast (Bahrman et al., 2004; Balmer et al., 2006), have already permitted the identification of proteins. The aim of this work was to determine seed proteins to facilitate protein expression analysis under different conditions. Through the proteome analysis of wheat seeds showed us a frame of abundant protein components.

Many proteins of crop seeds can cause allergic disease, such as alpha-amylase, beta-amylase, alpha-amylase/ trypsin inhibitor, thaumatin-like protein, isoflavone reductase, and so on. Wheat (*Triticum aestivum* L.) is the most consumed crop in the world, but it can be responsible for IgE-mediated food allergy. Recently, proteomic strategies have been applied for the identification of allergenic proteins, and the new strategy is now referred to as "allergenomics" (Yagami et al., 2004). In allergenomics, total proteins in allergen sources are solubilized with a strong nonionic detergent and urea and effectively resolved with two dimensional gel electrophoresis. Subsequently, IgE-reactive proteins are detected by IgE immunoblotting using allergic patients' sera. To date, however, comprehensive and integrated

analysis of total proteins in wheat flour by allergenomics has not been fully demonstrated.

Peptide mass fingerprinting (PMF) acquired by MALDI-TOF MS remains the most simple and powerful technique for high-throughout protein identification. This approach can be successfully applied and is more effective for organisms when their genomes have already been sequenced and fully annotated. For species without full genome sequence, when ESTs are available, it is still possible to carry out identifications using this strategy. So, the species-specific EST databases have been used for protein identification as an alternative in species without full genome sequence information (Lisacek et al., 2001; Mathesius et al., 2001; Kim et al., 2003; Kwon et al., 2003; Coldita, 2004). Although the EST raw sequences can be used to query by MS fingerprinting data, the accuracy and efficiency of protein identification would be suffered from the out-of-frame translation and short peptide translation.

Currently, PMF data or peptide sequence tags of protein spots from wheat proteome profiles are usually used to query the Swiss-Prot, TrEMBL and NCBI protein databases to conduct cross-species protein identification (Peng et al., 2009; Kong et al., 2010). Homology or conservation in amino acid sequence is the basis for identifying proteins using the cross-species databases. However, even for the most conserved proteins, polymorphism exists at the amino acid level between different species, which could result in wrong results, if the mass spectra are affected. So, we conducted protein identification using species-specific EST sequences. In the case of wheat, there are only a limited protein sequence referenced at NCBI protein database. However, 1,050,000 EST sequences of wheat are available, and this resource can be utilized for protein identification using this strategy. Our research results further imply that species-specific EST databases are able to improve the efficiency of protein identification, and overcome the risk of cross-species databases query.

ACKNOWLEDGEMENTS

This project was sponsored by Qing Lan Project, and National Science Foundation of China (No.30900871), and Natural Science Foundation of Jiangshu Higher Education Institutions in China (No. 09KJB210001), and Excellent Youth Foundation of Huaiyin Normal University (No.08QNZCK09).

REFERENCES

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM (2001). The InterPro database, an integrated documentation resource for protein families, domains, and functional sites. Nucleic Acids Res., 29: 37-40.

Bahrman N, Le Gouis J, Negroni L, Amilhat L, Leroy P, Laine AL, Jaminon O (2004). Differential protein expression assessed by two-dimensional gel electrophoresis for two wheat varieties grown at four nitrogen levels. Proteomics, 4: 709-719.

Bahrman N, Negroni L, Jaminon O, Le Gouis J (2004). Wheat leaf proteome analysis using sequence data of proteins separated by two-dimensional electrophoresis. Proteomics, 4: 2672-2684.

Balmer Y, Vensel WH, DuPont FM, Buchanan BB, Hurkman WJ (2006). Proteome of amyloplasts isolated from developing wheat endosperm presents evidence of broad metabolic capability. J. Exp. Bot., 57: 1591-1602.

Bateman A, Birney E, Cerruti L, Drubin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002). The Pfam protein families database. Nucleic Acid Res., 29: 1-4.

Coldita F, Nyamsuren O, Niehaus K, Eubel H, Braun HP, Krajinski F (2004). Proteomic approach: identification of Medicago truncatula proteins induced in roots after infection with the pathogenic oomycete Aphanomyces euteiches. Plant Mol. Biol., 55: 109-120.

Damerval C, Devienne D, Zivy M, Thiellement A (1986). Technical improvements in two-dimensional electrophoresis increase the level of genetic variation detected in wheat-seedling proteins. Electrophoresis, 7: 52-54

Danchenko M, Skultety L, Rashydov NM, Berezhna VV, Mátel L, Salaj T, Pret'ová A, Hajduch M (2009). Proteomic analysis of mature soybean seeds from the Chernobyl area suggests plant adaptation to the contaminated environment. J. Proteome Res., 8: 2915-2922.

Kanehisa M, Goto S (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 28: 27-30.

Kim SI, Kim JY, Kim EA, Kwon KH, Kim KW, Cho K, Lee JH, Nam MH, Yang DC, Yoo JS, Park YM (2003). Proteome analysis of hairy root from Panax ginseng C.A. Meyer using peptide fingerprinting, internal sequencing and expressed sequence tag data. Proteomics, 3: 2379-2392.

Kong FJ, Oyanagi A, Komatsu S (2010). Cell wall proteome of wheat roots under flooding stress using gel-based and LC MS/MS-based proteomics approaches. Biochim. Biophys. Acta, 1804: 124-136.

Kottapalli KR, Payton P, Rakwal R, Agrawal GK, Shibato J, Burow M, Puppala N (2008). Proteomic analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. Plant Sci., 175: 321-329.

Kwon KH, Kim M, Kim JY, Kim KW, Kim SI, Park YM, Yoo JS (2003). Efficiency improvement of peptide identification for an organism without complete sequence, using expressed sequence tag database and tandem mass spectral data. Proteomics, 3: 2305-2309.

Lisacek FC, Traini MD, Sexton D, Harry JL, Wilkins MR (2001). Strategy for protein isoform identification from expressed sequence tags and its application to peptide mass fingerprinting. Proteomics, 1: 186-193.

Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG (2001). Establishment of a root proteome reference map for the model legume medicago truncatula using the expressed sequence tag database for peptide mass fingerprinting. Proteomics, 1: 1424-1440.

Neuhoff V, Arnold N, Taube D, Ehrhardt W (1985). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis, 9: 255-262.

Østergaard O, Finnie C, Laugesen S, Roepstorff P, Svennson B (2004). Proteome analysis of barley seeds: Identification of major proteins from two-dimensional gels (pI 4-7). Proteomics, 4: 2437-2447.

Peng Z, Wang M, Li F, Lv H, Li C, Xia G (2009). A proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. Mol. Cell. Proteomics, 8: 2676-2686.

Rabilloud T, Charmont S (2000). Proteome Research: Two-dimensional Gel Electrophoresis and Identification Methods, Springer Verlag, Berlin, pp: 107-126.

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem., 68: 850-858.

Skylas DJ, Wrigley CW (2004). Proteomics of grains. Encyclopedia of Grain Science (Volume 2). Oxford press, UK, pp. 480-488.

Vensel WH, Tanaka CK, Cai N, Wong JH, Buchanan BB, Hurkman WJ

(2005). Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics, 5: 1594-1611.
Wasinger VC, Cordwell SJ, Cerpa-Poljak A, Yan JX, Gooley AA, Wilkins MR, Duncan MW, Harris R, Williams KL, Humphery-Smith I (1995). Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. Electrophoresis, 16: 1090-1094.
Yagami T, Haishima Y, Tsuchiya T, Tomitaka-Yagami A, Kano H, Matsunaga K (2004). Proteomic analysis of putative latex allergens. Int. Arch. Allergy Immunol., 135: 3-11.
Yahata E, Maruyama-Funatsuki W, Nishio Z, Tabiki T, Takata K, Yamamoto Y, Tanida M, Saruyama H (2005). Wheat cultivar-specific proteins in grain revealed by 2-DE and their application to cultivar identification of flour. Proteomics, 5: 3942-3953.
Yang LM, Luo YM, Wei JF, Ren CM, Zhou X, He SH (2010). Methods for protein identification using expressed sequence tags and peptide mass fingerprinting for seed crops without complete genome sequences. Seed Sci. Res., 20: 257-262.