Food Sci. Biotechnol. Vol. 17, No. 5, pp. 912 ~ 918 (2008)

Food Science and Biotechnology

ⓒ The Korean Society of Food Science and Technology

Proteomic Characterization of the ‘Agakong’, a Small-seeded Recombinant Inbred Line Derived from ‘Eunhakong’ (*Glycine max*) × ‘KLG10084’ (*Glycine soja*)

Ung-Kyu Choi[1,2], Hyun Su Ryu[1], Hyun Tae Kim[1], Sun Mi Yun[1], Su-Jin Lee[1], Jae Dek Choi[1], Young-Hyun Hwang[3], Soo Young Choi[4], and Oh-Shin Kwon[1]*

[1]*School of Life Sciences and Biotechnology, College of Natural Sciences, Kyungpook National University, Daegu 702-701, Korea*
[2]*Department of Oriental Medicinal Food and Nutrition, Asia University, Gyeongsan, Gyeongbuk 712-220, Korea*
[3]*Department of Agronomy, Kyungpook National University, Daegu 702-701, Korea*
[4]*Department of Biomedical Science and Research Institute for Bioscience and Biotechnology, Hallym University, Chuncheon, Gangwon 200-702, Korea*

Abstract This study was conducted to identify the differences in proteomic characteristics of ‘Agakong’, recombinant inbred line, and its parental genotypes ‘Eunhakong’ (*Glycine max*) and ‘KLG10084’ (*G. soja*). The isoflavone content of ‘Agakong’ was 3 times higher than that of its parental lines. A combined high-throughput proteomic approach was employed to determine the expression profile and identity of proteins using 2-dimensional gel electrophoresis and matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry. The overall distribution patterns of proteins are quite similar, but lots of protein spot intensities varied among the genotypes. A total of 41 proteins, representing significant difference in the quantities of protein among the lines, were successfully identified. Among them, more than 50% of the proteins identified were subunits of glycinin and β-conglycinin, 2 major storage proteins. This study showed that the proteomic analysis could help to define specific changes in protein level and composition, which can occur in the generation of new soybean varieties.

Keywords: proteomics, isoflavone, ‘Agakong’, ‘Eunhakong’, ‘KLG10084’

Introduction

Soybeans are present in almost every diet throughout the world because they are good sources of starch, dietary fiber, protein, lipids, and minerals (1-3). Peculiarly, soybeans supply a major portion of the world's demand for vegetable protein (4). Soybean seeds contain about 40% protein. Two major storage proteins, β-conglycinin and glycinin, account for approximately 70-80% of the total proteins. Proteins that accumulate during seed development are essential reserves for supporting germination and early seedling growth in plant. By contrast, approximately 20% of the soybean seed dry weight is oil in the form of triacylglycerol. Soybeans also contain a variety of bioactive phytochemicals including bioactive peptides, saponin, and isoflavones. Particularly soy isoflavones are known to exhibit various health-beneficial effects including preventive effects in the development of cardiovascular diseases and cancers (5-7).

In previous study, we selected new small-seeded recombinant inbred line (RIL: ‘Agakong’) derived from an interspecific cross between ‘Eunhakong’ (*Glycine max*) ×wild soybean (*Glycine soja*: ‘KLG10084’), and verified agronomic characteristics (8-10). However, limited information is available on the biochemical and genetic mechanism of the new soybean line, ‘Agakong’. For a better understanding of the genetypic variation, elucidation of the protein composition is necessary because of its direct relationship to phenotype.

Proteomic tools are powerful methodologies for accurately detecting changes in protein composition. This analysis can be used to visualize and compare complex mixtures of proteins and to characterize biological responses. Recent technical improvements in 2-dimensional (2-D) gel electrophoresis and mass spectrometry (MS) have made it possible to identify hundreds of proteins rapidly and investigate levels of protein expression, subcellular localization, and post-translational modifications (11). In recent years, proteomic studies using plant specimens such as soybean (12) and ginseng (11) has become popular for the purpose of identification of species and cultivars. Natarajan *et al*. (13) established a combined proteomic approach for the thorough analysis of storage protein compositions of wild (*G. soja*) and cultivated (*G. max*) genotypes, but the research of proteins which have various biological function was not reported yet.

In this study, we applied 2-D electrophoresis and combined with matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) to compare the protein profile of 3 genotypes ‘Eunhakong’, ‘KLG10084’, and ‘Agakong’ including storage proteins, metabolic proteins. The main aim of this work was to extend our comparative proteomic approach to identify differentially accumulated proteins and study their functional alteration following specific crossing.

*Corresponding author: Tel: +82-53-950-6356; Fax: +82-53-943-2762
E-mail: oskwon@knu.ac.kr
Received September 6, 2007; Revised November 28, 2007;
Accepted November 29, 2007

Materials and Methods

Soybeans 'Eunhakong' (*G. max*) were obtained from National Youngnam Agricultural Experiment Station (Gyeongnam, Korea) and the 'KLG10084' (*G. soja*) (10), which has a green seed coat color, and their new recombinants inbred line, 'Agakong' (*G. max*) (8) were obtained from the Soyventure Co., Ltd. (Daegu, Korea). Seeds were stored at −80°C until used.

Chemicals Chemicals for electrophoresis, including acrylamide bis-acrylamide, sodium dodecyl sulfate (SDS), *N,N,N,N*-tetramethylethylendiamine (TEMED), ammonium persulfate (Sigma-Aldrich, St. Louis, MO, USA). Urea, dithiothreitol (DTT), CHAPS, and pharmalyte (pH 3.5-10) immobilized pH gradiant (IPG) DryStrips (pH 3-10, pH 4-7) were purchased from Amersham Pharmacia Biotech (Piscataway, NJ, USA). Tris aminomethane, 2-mercaptoethanol, α-cyanohydroxycinnamic acid (CHCA) matrix, and trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich. Coomassie brilliant blue (CBB) R-250 and Bradford solution were purchased from Bio-Rad (Hercules, CA, USA). Modified porcine sequencing grade trypsin was purchased from Promega (Madison, WI, USA). Protein inhibitor cocktail tablet (CompleteTM) was purchased from Roche Diagnostics (Basel, Switzerland). All other chemicals were reagent grade.

Quantitative analysis of isoflavone by high performance liquid chromatography (HPLC) Quantitative analysis of isoflavone was performed following the method described by Wang *et al*. (14) with slight modification. The 20 μL filtrate was injected to an HPLC system equipped with a Bondapack C18 column after the system had been equilibrated at ambient temperature. Also, the ultraviolet (UV) detector was stabilized with mobile phase (methanol-1 mM ammonium acetate, 6:4) at a flow rate of 1 mL/min for 30 min. Effluent was detected at 254 nm and chromatogram was recorded for 20 min. Isoflavones were identified by their retention times of standard addition, and their contents were calculated by comparing their peak areas with those of standards.

Extraction of proteins from seeds: Modified trichloroacetic acid/acetone precipitation/urea solubilization extraction This protocol was performed according to Natarajan *et al*. (13). For this method, soybean seeds were powdered in liquid nitrogen using a mortar and pestle. An 100 mg of the soybean seed powder was homogenized with 2 mL of a solution containing 10%(w/v) trichloroacetic acid (TCA) in acetone with 0.07%(v/v) 2-mercaptoethanol (2-ME). The total protein was precipitated for overnight at −20°C. The extract was centrifuged at 14,000×g for 20 min at 4°C. The pellet was washed 2-3 times with acetone containing 0.07%(v/v) 2-ME. Then, the pellet was dried under vacuum for 30 min and the acetone dry powder was resuspended in 500 μL of lysis buffer [9.5 M urea, 2% CHAPS, 0.8%(w/v) one of the Pharmalyte (pH 3.5-10.0)] and 1% DTT and protein inhibitor cocktail tablet. After sonication on ice for 30 min, the samples were centrifuged for 15 min at 10,000×g and the supernatants were directly applied to an 18 cm IPG strip. The protein concentration was determined according to Bradford method (15) using a commercial dye reagent from Bio-Rad. Seeds were stored at −70°C until use.

2-D electrophoresis analysis A 2-D electrophoresis was performed in an IPGphor isoelectric focusing (IPG-IEF) system (Amersham Pharmacia Biotech) using 180 mm pH 3-10 and 4-7 Immobiline DryStrips (180×3×0.5 mm) for the first dimension (1-D), and 12% SDS-polyacylamide gels (PAGE) for the second dimension (2-D). SDS-PAGE was performed in an Ettan DALTsix Larger Vertical system (Amersham Pharmacia Biotech).

IPG-IEF can be simplified by the use of the integrated IPGphor system, in which rehydrated with 340 μL of rehydration buffer [9.5 M urea, 2% CHAPS, 0.8%(w/v) one of the pharmalyte, 1% DTT, 0.005% bromophenol blue] containing 100 μg (pH 3-10) and 120 μg (pH 4-7) of protein. And IEF are performed in a one-step procedure. Initial rehydration was for 12 hr at 20°C. IEF was then carried out with the following voltage program: 500 (gradient over 1 hr), 1,000 (gradient over 1 hr), and 8,000 V (fixed for 10 hr) at 50 μA/strip. The IPG strips were equilibrated for 30 min in 125 mM Tris (pH 6.8) containing 6 M urea, 30%(v/v) glycerol, 2%(w/v) SDS, and 65 mM DTT, and then for a further 30 min in the same buffer, except that DTT was replaced with 260 mM iodoacetamide. The IPG strips were then sealed with 0.5% agarose in SDS running buffer at the top of slab gels (260 ×200×1.5 mm^3).

The 2-D electrophoresis was performed in the Ettan DALTsix Larger Vertical system at 120 V/gel and room temperature for 10 hr. The gels were then with 50% ethanol containing, 3% phosphoric acid for 30 min, and stained with 0.02% CBB R-250 (Bio-Rad), 3% phosphoric acid, 17% ammonium sulfate, 34% ethanol. Molecular masses were determined by simultaneously running standard protein markers (MBI Fermentas, Amherst, NY, USA) in the range 11-170 kDa. The pI values used were those given by the supplier of the immobilized pH gradient strips. Excess dye was washed from the gels with distilled water and the gels were scanned with a UMAX PowerLook 1120 (UMAX Technologies, Dallas, TX, USA) scanner. Protein spots were outlined (first automatically and then manually) and quantified using PDQuest software (Bio-Rad).

MALDI-TOF mass spectrometer To identify the protein spots on gel pieces, preparative 2-D gels were excised, and cut into 1-2 mm pieces. These were added to 100 μL of 25 mM NH_4HCO_3/50% acetonitrile (ACN), incubated for 10 min to remove CBB, then rinsed thoroughly. The gel pieces were then dehydrated and dried thoroughly in a vacuum centrifuge for a few min. The dried gel pieces were rehydrated with 20 μL 50 mM NH_4HCO_3 (pH 8.0) containing 20 μg/mL trypsin (Promega) and the proteins in the gel pieces were digested overnight at 37°C. After digestion was complete, the supernatant was transferred to another tube, and the gel piece incubated with 20 μL of 50% ACN/5% formic acid for 10 min at room temperature. The extract was transferred to the primary supernatant, and the extraction repeated once more. The extracted digests were evaporated to dryness in a vacuum centrifuge. The digests were redissolved in 2 μL of 0.1% TFA. The

solution was vortexed and centrifuged. The 1 μL of supernatant was mixed with 1 μL of 10 mg/mL CHCA in 50% ACN/0.1% TFA and the mixed solution was spotted onto a 96-spot MALDI target. MALDI-TOF analyses were performed on a Voyager DESTR MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA, USA). The MALDI-TOF mass spectrometer was operated in positive-ion, delayed-extraction (200 nsec delay time) reflector mode. And spectra were acquired with 100 laser shots of a 337 nm nitrogen laser operating at 20.0 Hz. Spectra were calibrated using the Angiotensin autolysis peaks as internal standard at *m/z* 974.01 and 1,297.07.

Results were analyzed with Data Explorer software (Applied Biosystems) to obtain accurate masses (±50-100 ppm) for all the peptides in the tryptic digest. The results from the Protein Mass Fingerprinting (PMF), together with the pI and molecular mass values (estimated from 2-D electrophoresis gels), were used to search the Swiss-Prot or NCBInr protein databases with a special search tool [MS-FIT from Protein Prospector V 4.0.4 (http://prospector.ucsf.edu)], which compares the experimentally determined tryptic peptide masses with theoretical peptide masses calculated for proteins contained in these databases. Search parameters were ±50-100 ppm peptide mass tolerance and one maximum missed cleavage.

Results and Discussion

Shape and the isoflavone content of three genotypes of soybean seeds The size and the shape of 3 genotypes of soybeans used in this experiment were shown in Fig. 1. The means of 100-seed weight of 'Eunhakong', 'KLG10084', and 'Agakong' were 13.1, 2.3, and 6.3 g, respectively. This result was similar with the report of Kim *et al*. (16) in which the 100-seed weight of the Korean *G. soja* germplasm accessions was ranged from 1.2-2.9 g. It was indicated that 'KLG10084' displayed typical characters of wild soybean. The 'Agakong', of which 100-seeds weight is lighter than 10 g, was suitable variety for soybean spout and *cheonggukjang*.

There are 3 types of isoflavones in soybean, and each type exists in 4 different chemical forms, which include aglycones (daidzein, genistein, and glycitein) and their β-glucoside conjugates: glucosides (daidzin, genistin, and glycitin), malonylglucosides (6-*O*-malonyldaidzin, 6-*O*-malonylgenistin, and 6-*O*-malonyl glycitin), and acetylglucosides (6-*O*-acetyldaidzin, 6-*O*-acetylgenistin, and 6-*O*-acetylglycitin). Most of them exist as β-glucoside, the malonylglucoside, and acetylglucoside forms (17). The

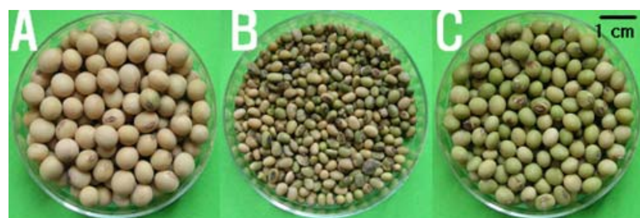

Fig. 1. The shape of 3 genotype of soybeans. (A) 'Eunhakong' (*G. max*), (B) 'KLG10084' (*G. soja*: a wild type soybean line used as one of the parents in the experiments), (C) 'Agakong' (*G. max*: a recombinant inbred line derived from 'Eunhakong'× 'KLG10084').

Table 1. Composition of isoflavones in the 3 genotypes of soybean seeds (unit: μg/g)

Isoflavone	'Eunhakong'	'KLG10084'	'Agakong'
Daidzin	694.2	409.9	2,117.8
Glycitin	50.0	100.0	150.0
Genistin	708.1	568.6	2,357.5
A-Daidzin	90.8	47.7	155.6
M-Genistin	1,578.2	641.1	4,663.5
A-Genistin	23.3	18.4	89.1
Daidzein	62.7	182.8	69.6
Genistein	78.9	117.8	81.7
Total	3,286.2	2,086.3	9,684.8

bioavailable isoflavones (aglycones) are formed by the hydrolysis of glycosides through β-glucosidase present in soybean. Most soybean isoflavones which is increased with germination progressed (18,19) are present in the glycoside form and are converted to aglycones during fermentation due to the β-glucosidase activity of microorganisms (20). In our investigation, the isoflavone content of 3 kinds of soybean seeds ('Eunhakong', 'KLG10084', 'Agakong') were determined with HPLC (Table 1). Total isoflavone content of 'Agakong' was 9,684.8 μg/g which is more than 3 times higher than that in the 'Eunhakong' (3,286.2 μg/g) and 'KLG10084' (2,086.3 μg/g). Zhu *et al.* (17) reported that the isoflavone contents of dried 'Hutcheson' and 'Caviness' soybean seeds were 2,190 and 2,286 μg/g, respectively, and is gradually increased after soaking. The maximum total isoflavone content was obtained when hypocotyls of the seeds were 0.5-2.5 mm.

2-D electrophoresis patterns and identification of differentially accumulated proteins by MS We used 2-D electrophoresis in conjunction with quantitative image analysis and sequencing mass spectrometry to investigate discrepancies in the protein expression profiles of the 'Eunhakong', 'KLG10084', and 'Agakong' (RIL of 'Eunhakong' ×'KLG10084'). Three gels per sample were simultaneously separated with broad range pH 3.0-10.0 immobilized pH gradient and narrow pH 4.0-7.0 immobilized pH gradient for the 1-D, and analyzed with PDQuest 2-D software. Figure 2 showed 2-D patterns of tested 3 kinds of soybean seeds. More than hundreds spots were detected. Of those proteins with molecular masses ranging from 11-95 kDa, 70% had acidic pIs, whereas 30% of polypeptide spots fell within the alkaline region. There were a few protein spots over 95 kDa, and some spots were detected with a pI>9.5. Overall, the seed protein composition of all 3 soybean lines was strikingly similar. Moreover, these 2-D patterns share some similarities with the patterns observed for indigenous Nepalese soybean and North American cultivars (21).

Protein spots from the 2-D gel were subjected to trypsin digestion and MALDI-TOF analysis. Although the protein profile of 3 varieties is quite similar, at least more than 100 spot differences were detected. Consequently, 41 proteins which revealed obvious differences (more than 3 fold) in the quantities among the soybean lines were successfully

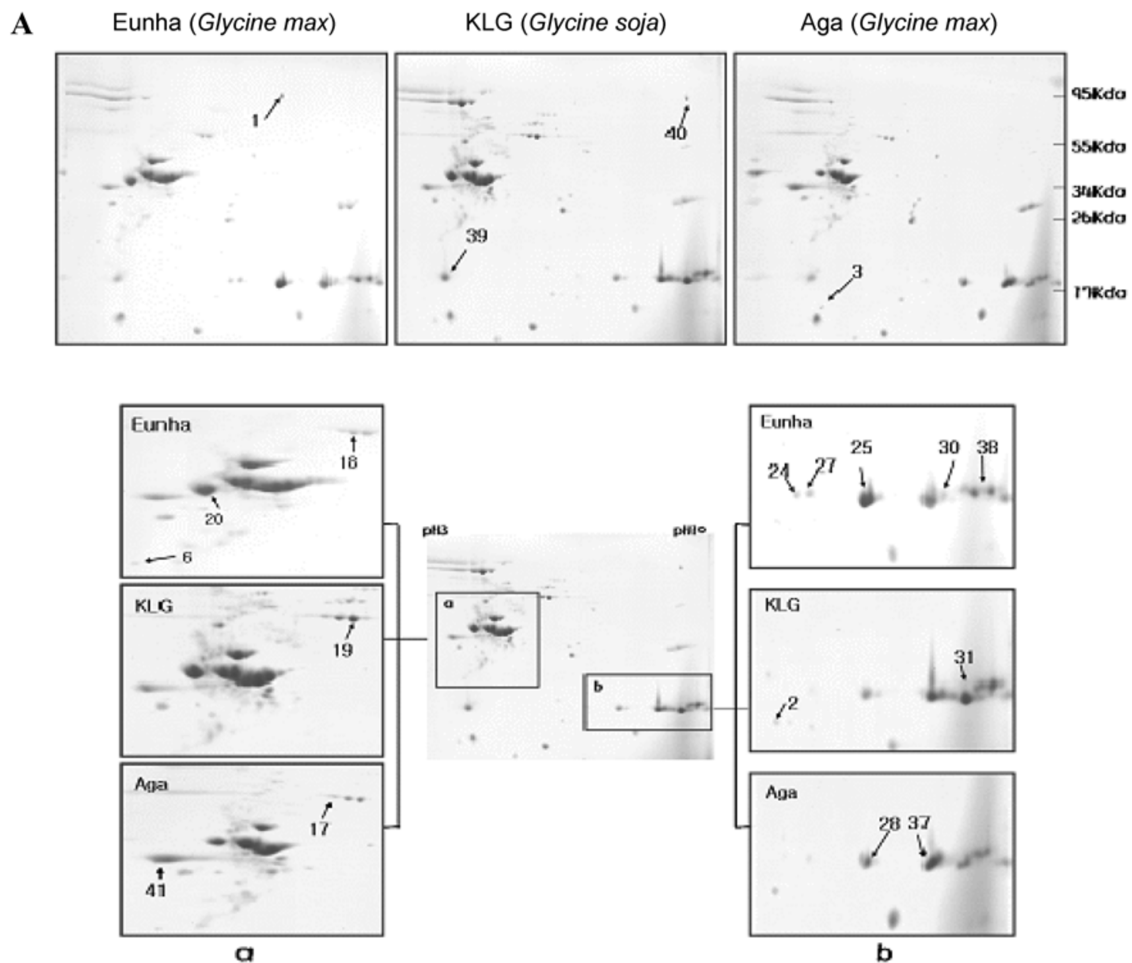

Fig. 2. Proteomic comparison of the proteins in 'Eunhakong', 'KLG10084', and 'Agakong'. (A) Proteins were separated on a pH 3-10 IPG strip in the first dimension and on an 12% SDS-polyacrylamide gel in the second dimension. (B) Protein extracts were separated on a pH 4-7 IPG strip followed by 12% SDS-PAGE. Gels were stained with Coomassie brilliant blue G-250. Arrows indicate the spots showing abundant or less abundant proteins. The numbers indicated on the gels correspond to the gel numbers given in Table 2.

identified. The relative abundance of the protein spots was is shown in Table 2. More than 97% of spots had a sequence coverage exceeding 10%. Identification was validated by agreement between the apparent Mw and pI determined from 2-D gels, and the theoretical values of the identified proteins (ΔMw <20% or ΔpI <0.5).

Classification of the differentially accumulated proteins in the soybean seed The differentially accumulated proteins were classified in terms of their physiological functions using information from PubMed (22) and the swiss Prot/ TrEMBL protein knowledgebase (23). As shown in Fig. 3, half of the proteins identified were storage proteins, and other groups were involved in metabolic, defense (reactive oxygen species: ROS), cell growth and division, transporter, and the many other proteins.

Storage proteins are largely responsible for the nutritional and physicochemical properties of soybeans. Recently, Natarajan *et al.* (13) applied a combined proteomic approach for separation, identification, and comparison of 2 major storage proteins, β-conglycinin and glycinin, in wild (*G. soja*) and cultivated (*G. max*) soybean seeds. They reported that the total number of storage protein spots detected in wild and cultivated genotypes was approximately 44 and 34, respectively. In this study, 54% of the proteins identified are involved in storage protein, such as β-conglycinin β chain precursor, glycinin G1 precursor, glycinin G2 precursor, glycinin G3 precursor, glycinin G3 subunit, and glycinin G4 precursor. Many of storage proteins were multiply spotted, which is similar to other reports (4,24). The variation in the distribution of the protein spots could be due to post-translational modifications or proteolysis. However, the conglycinin and glycinin classes of storage protein belong to multigene families, thus the genetic redundancy is another possible explanation for the multiple isoelectric species observed within these protein classes (25). The β-conglycinin is encoded by a gene family of at least 15 highly homologous genes. Glycinin has 5 nonallelic genes, which code for 5 glycinin protein precursors, G1, G2, G3, G4, and G5, respectively (25). Interestingly, significant differences in the accumulation of the subunits of β-conglycinin and glycinin were evident between 'Agakong' and each of the parental lines. Most of the subunits of the storage proteins accumulated at higher amounts in 'Eunhakong' or 'KLG10084' line when compared to the 'Agakong'. In contrast, 2 protein spots (#35 and #37)

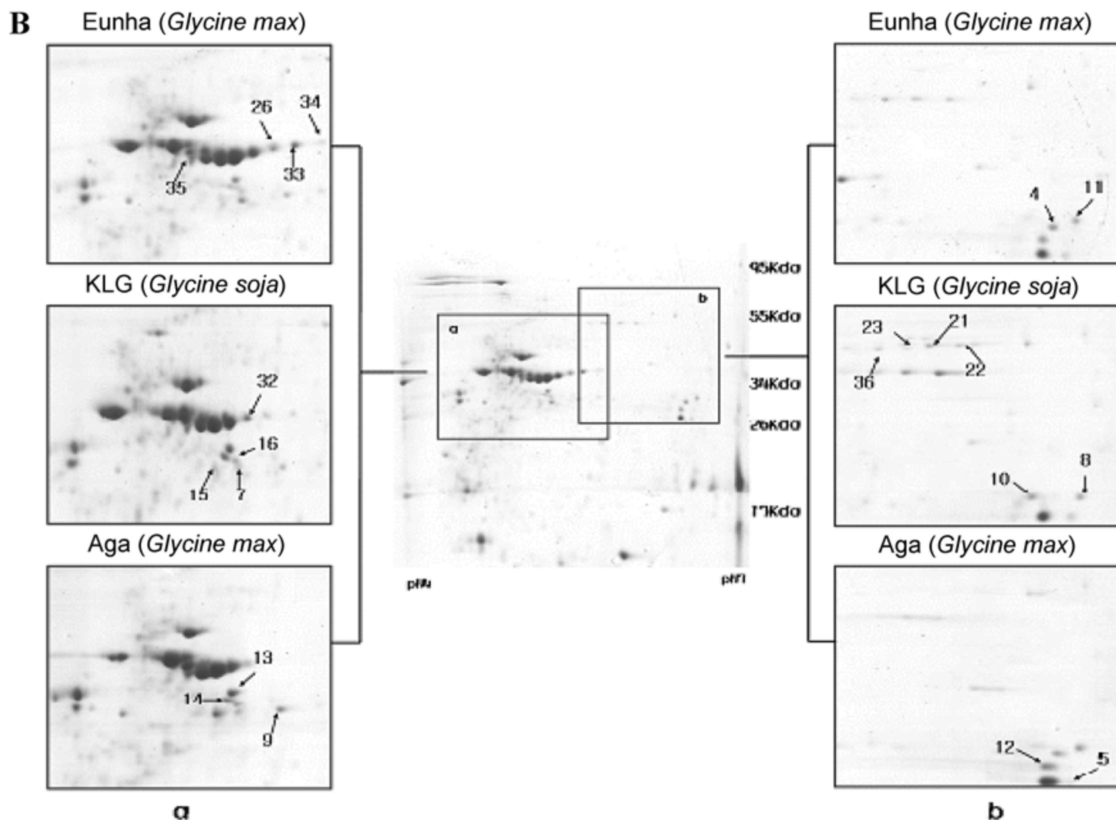

Fig. 2. Continued

were more abundant in 'Agakong' than in the parental lines. The large amount of variation in seed protein composition among 3 genotypes may be due to a different complement of genes in the wild genotype that control expression of β-conglycinin and glycinin as compared to the cultivated genotype (26).

We observed that 12% of the proteins identified are involved in cell growth and division, including seed maturation protein and dehydrin. Both seed maturation protein PM36 and PM31 content of the 'Agakong' were higher than those of the 'Eunhakong' and the 'KLG10084'. In contrast, dehydrin is relatively more abundant in 'Ennhakong' when compared to the other lines. Dehydrin, a group of highly hydrophilic proteins, are distributed in a wide rage of organism including the higher plants, yeast and cyanobacteria. The distinct differential accumulation may influence in plant growth and in stress tolerance.

A cluster of glutathione *S*-transferase (GST) was also identified during the course of this work. The abundance of the GSTs was varied, with 'Agakong' showing the least accumulation of GST17. In contrast, intensities of GST12 and GST7 were higher in 'Agakong' as compared to the parental genotypes. In plants, GST activity protects cells form a wide range of biotic and a biotic stresses including pathogen attack, xenobiotic toxicity, and oxidative stress. A large diversity of isoforms accounts for both the wide array of substrates recognized and the varied catalytic properties of different GSTs.

In conclusion, we used 2-D electrophoresis in conjunction with quantitative image analysis and sequencing mass spectrometry to investigate discrepancies in the protein

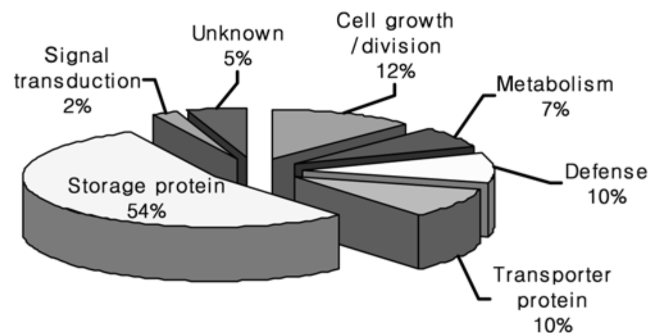

Fig. 3. Classification of the differentially accumulated proteins identified. Pie charts representing the distribution of the 41 identified proteins according to their biological functions. Assignments were made on the basis of information from the NCBI (www.ncbi.nlm.nih.gov.PubMed) websites.

profiles of 'Agakong' as compared to parental genotypes 'Eunhakong' and 'KLG10084'. The overall distribution pattern of proteins is quite similar, but lots of protein spot intensities varied among the genotypes. Therefore, we identified selected protein spots appearing strikingly different abundance. In this study, we showed that a high degree of electrophoretic heterogeneity of 2 storage proteins, β-conglycinin and glycinin, exists among the genotypes. The accumulation of most of the subunits tested was higher in the parental lines as compared to the 'Agakong'. In addition, we demonstrated that 'Agakong' have significantly higher isoflavone content than its parents,

Table 2. Differently expressed proteins in ‘Agakong’, ‘Eunhakong’, and ‘KLG10084’[1)]

No	Identified protein	Score	Coverage (%)	Accession No	Mw (kDa) /pI	Relative intensity		
						‘Eunhakong’	‘KLG10084’	‘Agakong’
Cell growth/division (5)								
1	Acyl-CoA oxidase	137	12.6	gi|15553480	74/8.3	100	0.1	0.5
2	Seed maturation protein PM36	4,335	30.5	gi|5802242	25/5.3	18.7	61.8	100
3	Seed maturation protein PM31	37.9	16.3	gi|4838149	17/4.6	68.4	2.3	100
4	Dehydrin	434,219	49.2	gi|37495457	23/6.0	100	0.5	59.6
5	Dehydrin	20.4	22.6	gi|37495457	23/6.0	100	68.1	6.2
Metabolism (3)								
6	Cyclin-dependent kinase inhibitor	151	40.0	gi|42362358	22/5.9	64.6	50.3	100
7	Delta 1-pyrroline-5-carboxylate synthetase	1,632	39.2	gi|57283666	30/5.8	100	25.5	54.7
8	Phosphoenolpyruvate carboxylase kinase	937	15.0	gi|23297175	30/6.1	2.1	100	1.3
Defense (4)								
9	Glutathione *S*-transferase GST 12	65.5	26.4	gi|11385439	26/5.9	37.7	35.2	100
10	Glutathione *S*-transferase GST 17	45.1	28.8	gi|11385449	26/6.4	28.1	100	1.2
11	Glutathione *S*-transferase GST 17	22.1	17.0	gi|11385449	26/6.4	100	0.0	52.3
12	Glutathione *S*-transferase GST 7	114	22.2	gi|11385429	26/6.9	51.2	1.0	100
Transporter protein (4)								
13	Lectin precursor (Agglutinin) (SBA)	21,453	42.5	gi|126151	30/5.7	43.2	100	76.6
14	Lectin precursor (Agglutinin) (SBA)	1,545	29.8	gi|126151	30/5.7	55.3	100	89.4
15	Lectin precursor (Agglutinin) (SBA)	47.3	24.2	gi|126151	30/5.7	42.6	37.6	100
16	Lectin precursor (Agglutinin) (SBA)	3,915	30.2	gi|126151	30/5.7	100	95.0	79.0
Storage protein (22)								
17	β-Conglycinin, β chain precursor	1.45×10^6	39.8	gi|121282	50/5.9	82.9	100	81.9
18	β-Conglycinin, β chain precursor	1.13×10^6	36.2	gi|121282	50/5.9	64.1	100	52.8
19	β-Conglycinin, β chain precursor	498,132	43.1	gi|121282	50/5.9	36.2	100	52.7
20	Glycinin G4 precursor	2.95×10^6	53.6	gi|121279	24/4.5	10.7	100	31.5
21	Glycinin G1 precursor	698	13.9	gi|121276	55/5.9	3.8	100	5.3
22	Glycinin G1 precursor	2,058	24.4	gi|121276	55/5.9	4.9	100	8.8
23	Glycinin G2 precursor	66.5	14.4	gi|121277	54/5.5	0.3	100	1.0
24	Glycinin G2 precursor	61.1	14.6	gi|121277	54/5.5	100	1.2	0.2
25	Glycinin G2 precursor	137	19.8	gi|121277	54/5.5	100	38.3	22.4
26	Glycinin G2 precursor	37.7	12.2	gi|121277	54/5.5	100	19.9	8.3
27	Glycinin G3 precursor	51.2	20.8	gi|121278	54/5.7	100	13.2	11.9
28	Glycinin G3 precursor	1,268	19.1	gi|121278	54/5.7	100	20.0	12.7
29	Glycinin G3 precursor	605	16.4	gi|121278	54/5.7	100	40.0	35.6
30	Glycinin G3 subunit	1,677	9.4	gi|121278	54/5.3	100	56.2	50.9
31	Glycinin G3 subunit	18,022	18.0	gi|121278	54/5.3	87.0	100	42.3
32	Glycinin G3 precursor	441	19.3	gi|121278	54/5.7	100	24.6	8.6
33	Glycinin G3 precursor	122	28.1	gi|121278	54/5.7	100	0.9	0.1
34	Glycinin G3 precursor	217	17.0	gi|121278	54/5.7	100	65.4	0.3
35	Glycinin G3 precursor	35.6	12.1	gi|121278	54/5.7	17.7	61.7	100
36	Glycinin G4 precursor	122	14.9	gi|121279	63/5.3	0.4	100	0.8
37	Glycinin G4 precursor	14,790	23.5	gi|121279	63/5.3	28.2	87.1	100
38	Glycinin G4 precursor	620	13.3	gi|121279	63/5.3	35.1	100	62.7
Signal transduction/cell signal (1)								
39	Trypsin inhibitor A precursor	3,464	26.4	gi|125020	24/5.0	22.9	100	94.8
Unknown (2)								
40	Grr1	34.0	14.2	gi|2407790	73/8.0	0.2	100	0.6
41	60S Acidic ribosomal protein P0	76.6	17.8	gi|1710587	34/5.2	100	0.6	54.3

[1)]The MS spectra of protein digests were compared with the NCBInr database using the MS-FIT database-searching program. Protein names and functions have been assigned according to PubMed and Swiss-Prot/TrEMBL. Results are means of 3 independent experiments performed for each condition. The spot numbers are identical to those given in Fig. 2.

'Eunhakong' and 'KLG10084'. Unfortunately, we could not find out protein participated in the fortification of isoflavone accumulation. Nevertheless, our study demonstrated that the comparative studies of the proteins in soybeans would help us to understanding the evolutionary relationship between them.

References

1. Dabrowski KJ, Sosulski FW. Composition of free and hydrolyzable phenolic acids in defatted flours of 10 oilseeds. J. Agr. Food Chem. 32: 128-130 (1984)
2. Lin PY, Lai HM. Bioactive compounds in legumes and their germinated products. J. Agr. Food Chem. 54: 3807-3814 (2006)
3. Choi UK, Jeong YS, Kim MH, Lee NH, Hwang YH. Quality characteristics of *meju* according to germination time of raw soybean (*Glycine max*: 'Hwanggeumkong'). Food Sci. Biotechnol. 16: 386-391 (2007)
4. Mooney BP, Thelen JJ. High-throughput peptide mass fingerprinting of soybean seed proteins: Automated workflow and utility of unigene expressed sequence tag databases for protein identification. Phytochemistry 65: 1733-1744 (2004)
5. Messina MJ, Persky V, Setchell KD, Barnes S. Soy intake and cancer risk: A review of the *in vivo* data. Nutr. Cancer 21: 113-131 (1994)
6. Shon HS, Lee YS, Shin HC, Chung HK. Does soybean isoflavone have adverse effects on human? Korean Soybean Digest 17: 9-19 (2000)
7. Yang SO, Chang PS, Lee HJ. Isoflavone distribution and β-glucosidase activity in *cheonggukjang*, a traditional Korean whole soybean-fermented food. Food Sci. Biotechnol. 15: 96-101 (2006)
8. Lee JD, Jeong YS, Shannon JG, Park SK, Choung MG. Agronomic characteristics of small-seeded RILs derived from 'Eunhakong' (*Glycine max*) ×'KLG10084' (*G soja*). Korean J. Breed. 37: 288-294 (2005)
9. Lee JD, Kwon TH, Cho HY, Hwang YH. Genetic analysis of agronomic characters in interspecific cross in soybean. Agric. Res. Bull. Kyungpook National Universty, Korea 20: 1-8 (2002)
10. Lee JD, Yoon YH, Chung IK, Park SK, Hwang YH. A new G*lycine soja* germplasm accession with green seed-coat color. Breeding Sci. 55: 21-25 (2005)
11. Nam MH, Heo EJ, Kim JY, Kim SI, Kwon KH. Proteomic analysis of the responses of *Panax ginseng* C.A. Meyer leaves to high light; Use of electrospray ionization quadrupole-time of flight mass spectrometry and expressed sequence tag data. Proteomics 3: 2351-2367 (2003)
12. Gianazza E, Eberini I, Arnoldi A, Wait R, Sirtori CR. A proteomic investigation of isolated soy proteins with variable effects in experimental and clinical studies. J. Nutr. 132: 9-14 (2002)
13. Natarajan SS, Xu C, Bae H, Caperna TJ, Garret WM. Characterization of storage proteins in wild (*Glycine soja*) and cultivated (*Glycine max*) soybean seeds using proteomic analysis. J. Agr. Food Chem. 54: 3114-3120 (2006)
14. Wang G, Kuan S, Fransis O, Ware G, Carman AS. A simplified HPLC method for the determination of phytoestrogens in soybean and its processed products. J. Korean Soc. Agric. Chem. Biotechnol. 38: 185-190 (1990)
15. Bradford MM. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein dye binding. Anal. Biochem. 72: 248-254 (1976)
16. Kim KU, Kang TD, Lee JH, Lee IJ, Shin DH. Physio-ecological characteristics of wild soybeans (*Glycine soja*) collected throughout Korea and their response to glyphosate. Korean J. Weed Sci. 23: 153-153 (2003)
17. Zhu D, Hettiarachchy NS, Horax R, Chen P. Isoflavone contents in germinated soybean seeds. Plant Food. Hum. Nutr. 60: 147-151 (2005)
18. Han JS, Hong HD, Kim SR. Changes in isoflavone content and mass balance during soybean processing. Food Sci. Biotechnol. 16: 426-433 (2007)
19. Kim WJ, Lee HY, Won MH, Yoo SH. Germination effect of soybean on its isoflavones and oligosaccharides. Food Sci. Biotechnol. 14: 498-502 (2005)
20. Jang CH, Lim JK, Park CS, Kwon DY, Kim YS, Shin DH, Kim JS. Change of isoflavone content during manufacturing of *cheonggukjang*, a traditional Korean fermented soyfood. Food Sci. Biotechnol. 15: 643-646 (2006)
21. Krishnan HB, Natarajan SS, Mahmoud AA, Bennett JO, Krishnan AH. Assessment of indigenous Nepalese soybean as a potential germplasm resource for improvement of protein in North American cultivars. J. Agr. Food Chem. 54: 5489-5497 (2006)
22. PubMed. A service of the U.S. National Library of Medicine and the National Institute of Health. Available from www.ncbi.nlm.nih.gov/PubMed. Accessed Aug. 1, 2007.
23. Swiss Prot/TrEMBL protein knowledgebase. The expert protein analysis system. Proteomics server of the Swiss Institute of Bioinformatics. Available from http://au.expasy.org/sport. Accessed Aug. 1, 2007.
24. Natarajan SS, Xu C, Caperna TJ, Garrett WM. Comparison of protein solubilization methods suitable for proteomic analysis of soybean seed proteins. Anal. Biochem. 342: 214-220 (2005)
25. Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ. Characterization of the glycinin gene family in soybean. Plant Cell 1: 313-328 (1989)
26. Yaklich R, Helm R, Herman E. Analysis of the distribution of the major soybean allergen in a core collection of *Glycine max* accessions. Crop Sci. 39: 1444-1447 (1999)