A Genome-wide Approach for Functional Analysis Using Rice Mutant

Won Cheol Yim*, Dong Sub Kim**, Jun-Cheol Moon***, Cheol Seong Jang****, and Byung-Moo Lee*[†]

*Department of Plant Biotechnology, Dongguk University, Seoul, Korea
**Radiation Research Center for Bio-technology, Advanced Radiation Technology Institute, Korea Atomic Energy Research Institute, Jeongeup, Korea
***Institute of Life Science & Natural Resources, Korea University, Seoul, Korea
****Plant Genomics Lab., Division of Bio-resources Technology, Kangwon National University, Chunchon, Korea

ABSTRACT Rapid extension of genomic database leads to the remarkable advance of functional genomics. This study proposes a novel methodology of functional analysis using 5-methyltrytophan (5 MT) mutant together with their 2-DE analysis and public microarray database. A total of 24 proteins was changed in 5 MT mutant and four remarkably different expressed proteins were identified. Among them, three spots were converted to Affymetrix probe. A total of 155 microarray samples from Gene Expression Omnibus (GEO) in NCBI was retrieved and followed by constructing gene co-expression networks over a broad range of biological issues through Self-Organising Tree Algorithm. Three co-expressing gene clusters were retrieved and each functional categorization with differential expression pattern was exhibited from 5 MT resistance mutant rice. It was indicated new co-expression networks in the mutant. This study suggests that on investigating possibility which correspond 2-DE to microarray database with their full potential.

Keywords : rice (*Oryza sativa*), 5 MT, 2-DE, genome-wide, mutant, proteomic, GEO, clustering analysis

Rice (*Oryza sativa* L.) is the most important crop in eastern Asia. The agronomic importance of rice shared evolutionary history and syntenic relationship with the other cereal species. (Gale and Devos, 1998). Rice is also a monocot model plant for biological research because of its small genome size (Devos and Gale, 2000). It has led to the establishment of a high-quality finished genome sequence (International Rice Genome Sequencing Project, 2005). Therefore, now that the initial progress of sequencing has been reached, it is important to unravel the function, regulation, locus-locus interaction, and characterization of uncovered gene.

A far vast number of rice cultivars are widely grown, and field of genetic and molecular focuses is being in active. It conceivably consists of about 430 million base pair (Sasaki, 1998), and about 30,000 genes can be expressed in the tissue. The protein handling work is developing for total protein in convenience. Consequently, sequencing of a protein separated by two-dimensional electrophoresis became possible with their molecular weight database. The rapid developments of bioinformatics over last year, the combined cross analysis have been available. Moreover, the advent of microarray databases has made it possible to analysis the expression levels of thousands of gene in parallel. In their remarkable stream of bioinformatics, some novel method can be possible with cross analysis. In addition, a new generation of web-based tools based on quantitative information about genes contained in large microarray databases may be possible to discover novel crosstalk networks in diverse physiological pathways in plants (Zimmermann *et al.*, 2005).

Over last decade, plant tissue culture methods are developed rapidly, induced mutation lines in rice are increasing availability for research. An example, somaclonal variation induced by the process of plant cell culture, has also proven effective at generating a variety of kinds of genetic variation (Philips *et al.*, 1994). These mutant lines can be elucidated for particular biochemical processes that are difficult or impossible to study in wild type plants (Larkin and Scowcroft, 1981; Schaeffer *et al.*, 1984). In earlier, Kim *et al.* (2004) investigated 5-methyltryptophan (5 MT)-resistance rice mutant lines using in vitro mutagenesis with gamma

[†]Corresponding author: (Fax) +82-2-2269-2770
 (E-mail) bmlee@dongguk.edu <Received August 23, 2009>

rays and a continuous selection process that verified their resistance to 5 MT.

A total of 155 microarray samples from Gene Expression Omnibus (GEO) in NCBI was downloaded and followed by constructing gene co-expression networks over a broad range of biological issues through Self-Organising Tree Algorithm (SOTA). The dataset of co-expressing genes allows us to discover some degree of gene regulatory networks in diverse plant physiological issues (Yim and Jang, 2007). Therefore, this study has been proposed a novel methodology of functional analysis using mutants together with their two-dimensional analysis and public microarray database. The focus of the this study was to determine the extent of the involvement of gene-expression in rice mutant through transcriptome and proteomic cross-analysis

MATERIALS AND METHODS

Plant growth

Two genotypes were used: Mutant 7-6-21 resistance to 5 MT (Kim *et al.*, 2004) and wild-type rice (*O. sativa* L. Japonica cv. DonganByo). Rice seedlings were grown in pots and maintained in a growth chamber with 16 h light at 28℃ and 8 h dark at 25℃. The seedlings were harvested at three weeks after seeding. The leaf blades were immediately frozen in liquid nitrogen and stored at -80℃ until sample preparation for proteomic and genomic analysis.

Protein extraction

The extraction was performed according to method of Kamo *et al.* (2005) with some modifications. A portion (1 g) of samples was ground in liquid nitrogen and suspended in 10% w/v TCA in acetone with 0.07% w/v DTT at -20℃ for 1 h, followed by centrifugation for 15 min at 18,000 g. The pellets were washed with ice-cold acetone containing 0.07% DTT, incubated at -20℃ for 1 h and centrifuged again at 4℃. The pellets were lyophilized after three times washed. The pellet powders were then solubilized in lysis buffer (8.5 M urea, 4% w/v CHAPS, 5% w/v ampholines pH 3-10, 1% w/v DTT and 2% 2-mercaptoethanol). The protein concentration was determined by the Bradford method (Kruger, 2002) using a commercial dye reagent (Bio-Rad, Hercules, CA, USA).

Two−Dimensional Electrophoresis

Proteins (200 μg) were separated by 2D-PAGE. The first dimension by isoelectric focusing (IEF) was performed in individual immobilized pH gradient (IPG) strips (pH 4-10 and 17 cm in length, Bio-Rad, Hercules, CA, USA). The second dimension was separated by SDS-PAGE. IPG strips were equilibrated for 12 h with rehydration buffer containing 7 M urea, 2% w/v CHAPS (Bio-Rad, Hercules, CA, USA), 5% w/v ampholines pH 3-10 (Bio-Rad, Hercules, CA, USA), and 1% w/v DTT.

IEF was performed at 20℃ using PROTEAN IEF Cell (Bio-Rad, Hercules, CA, USA). Electrophoresis carried out at 100 V for 1 h, followed 1000 V for 12 h, and 2000 V for 1 h. After IEF, strips were incubated for 15 min in equilibration buffer (0.375 M Tris-HCl pH 8.8, 6 M Urea, 2% SDS, 20% glycerol and 130 mM DTT) and transferred to alkylation buffer (0.375 M Tris-HCl pH 8.8, 6 M Urea, 2% SDS, 20% glycerol and 135 mM iodoacetamide).

SDS-PAGE in the second dimension was carried out according to the method of O'Farrell *et al.* (1975), using 12% polyacrylamide gel. The 2-DE gels were stained by silver staining method of Gromova and Celis (2004). The pI and relative molecular mass of each protein were determined by using data of Rice Proteome Database (http://gene64. dna.affrc.go.jp/RPD/). To visualize gels and related data obtained through the use of Melanie software (version 3.08, Genbio, Geneva, Switzerland).

MALDI−TOF MS analysis

Peptide mass fingerprints of differentially expressed proteins in 2-DE were analyzed using Voyager-DE STR Biospectrometry Workstation MALDI TOF/MS (Applied Biosystems, Framingham, MA, USA). The machine was run in reflector/ time-lag-focusing mode controlled by Voyager Control Software. Four protein spots which displayed more than two-fold variation of expression between wild type and mutant were excised from the stained gels and digested according to Shevchenko *et al.* (2000) using modified sequencing grade porcine trypsin (Promega, Medison, WI, USA). Gel pieces were washed with 50% Acetonitrile (ACN) reagent (Wako, Osaka, Japan) and dried under vacuum. Washed pieces rehydrated with trypsin (10 ng/μl, sequencing grade, Promega) solution by incubation for overnight at 37℃. The pro-

teolytic reaction was terminated by addition of 5 μl of 0.5% trifluoroacetic acid (Aldrich, Poole, Dorset, UK). Tryptic peptides were recovered by combining the aqueous phase from several extractions of gel pieces with 50% aqueous ACN. After concentration, the peptide mixture was desalted using C18ZipTips (Millipore Billerica, MA, USA), and peptides were eluted in 1-5 μl of ACN. An aliquot of this solution was mixed with an equal volume of a saturated solution of α-cyano-4-hydroxycinnamic acid (CHCA) in 50% aqueous ACN. 1 μl of the mixture was spotted onto a target plate. Peptide were selected in the mass range between 500 and 3000 Da. Proteins were identified by database search against the NCBI nr database (ftp://ftp.ncbi.nih.gov/ blast/db/) by using MASCOT (Matrix Science, London, UK). The results were converted to microarray probe by using NetAffx (http://www.affymetrix.com) probe-to-gene annotations.

Rice expression data

A set of 155 gene expression omnibus samples (GSM) of Affymetrix GeneChip rice genome array (accession no. GPL2025) was downloaded from the Gene Expression Omnibus (GEO) database in NCBI, as of July 1 2007. Affymetrix GeneChip spot IDs matched to the rice nsLTP genes were retrieved via TIGR Genome Bowser (http://www.tigr.org/ tigr-scripts/osa1_web/gbrowse/rice/). The expression data of each of the nsLTP genes were gathered manually from the 155 GSM data of the rice genome array. The expression values of each plate were subsequently normalized by dividing each individual signal by the average of the plate. For the clustering analysis of individual rice nsLTP genes and GSM arrays based on the Affymetix GeneChip expression data, the average linkage hierarchical clustering of the data with the uncentered Pearson correlation was conducted using the C clustering library source code (de Hoon et al., 2004) through 64-bit arithmetic on an IBM p690 Unix platform at KISTI (Korea Institute of Science and Technology Information, Dajeon, Korea).

In order to find clusters of co-expressing genes with each of the gene (identified protein CDS) above the dataset, analysis using the SOTA array program (http://gepas.bioinfo.cipf.es/ cgi-bin/sotarray) with default parameters with the exception of complete linkage using correlation distance and linear correlation coefficiency was conducted (Herrero et al., 2001).

The expression profile of co-expressing genes in each cluster was visualized using SigmaPlot version 9.0 (Systat Software, Richmond, CA, USA).

RESULTS

Proteomic analysis between wild type and 5MT mutant

Proteomic analysis was performed to test whether the protein expression was different in wild type and 5 MT mutant leaves when phenotype equalized. Independently, two 2-DE gels from the proteins of seedling of wild type and 5 MT mutant plants were carried out. Quantitative image analysis, Melanie software revealed that 24 protein spots were changed their intensities (±1.5 fold over the WT spot). Out of 24 proteins changed in 5 MT mutant, 10 protein spots showed increased expression, whereas 14 protein spot showed reduced expression. These changed spots were verified by visually checking the gel image under high-resolution (600 dpi). The four remarkably different protein spots (±3 fold over the WT spot) were excised and analyzed by MALDI-TOF MS (Fig. 1). All 4 proteins resolved from the samples, were identified with accurate intensity pick. To gain protein information about these pick, MASCOT web base program (http://www.matrixscience. com/) was used. And these results were subjected to BLASTP (ftp://ftp.ncbi.nih.gov/BLAST) to investigate their homologies with other proteins in NCBI nr database (ftp://ftp.ncbi.nih. gov/blast/db/). The corresponding homologues with the highest
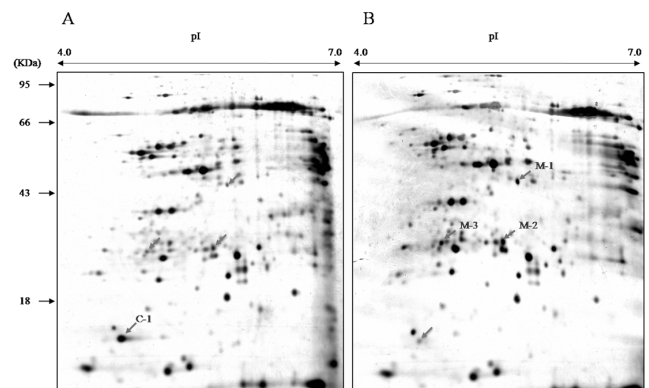
Fig. 1. Representative 2-DE gel patterns of leaf blade proteins. (A) 2-DE gel of wild type. (B) 2-DE gel of 5 MT mutant leaf protein. Red/blue arrow heads indicate more than three fold up/down regulated spots.

Table 1. The identification of differentially expressed protein spots with MALDI-TOF analysis

Spot Name	Expect value	PM[†]	SC[‡]	Protein name	Species	Acc. no[§]
C-1	1.46	22	33	hypothetical protein OsI_025208	*Oryza sativa*	EAZ03976
M-1	0.46	9	26	hypothetical protein OsI_020000	*Oryza sativa*	EAY98767
M-2	0.5	20	62	Rieske region domain containing protein	*Oryza sativa*	NP_001067589
M-3	0.72	15	45	BED zinc finger family protein	*Oryza sativa*	ABF99849

[†]PM, number of peptide matched.
[‡]SC, sequence coverage by MASCOT using MALDI-TOF MS.
[§]Acc. no., accession number in NCBI database.

homology are shown in Table 1. The two proteins up-regulated in 5 MT mutant were unambiguously identified, others might have no similar annotated proteins.

Clustering of co–expressing genes

The clustering of co-expressing genes were performed with a total of 155 rice GSM samples of Affymetrix GeneChip rice genome array using the SOTA array program with variability threshold of 90%, resulting in a total of 1660 clusters. To evaluate the distribution of expression pattern among genes in this dataset, a total of 20,000 gene pairs was selected randomly followed by measuring the degree of similarity using Pearson correlation coefficient (Yim and Jang, 2007). Ninety nine percent of the r values obtained from random gene pairs had $r < 0.55$ with an average r value of 0.07, providing a model for determining not-randomized clustering of the expression profiles. In order to verify that the clustering is performed well, one cluster with 85 co-expressing genes was selected and measured the correlation coefficient among genes. The r values among the co-expressing genes distributed from -0.08 to 0.98 with an average r value of 0.66. It appears that genes included in each cluster were related to common functional role.

Functional categorization with expression profiles

The co-expression of genes across a given cluster is most commonly explicable through some common functional role. Retrieving co-expressing genes with identified spots was conducted using the SOTA array dataset. However, current Affymetrix *O. Sativa* GeneChip was not supported full matching probes with whole rice genome. For this reason, the identified spot M-1 did not match complete probe.

In order to verify that the clustering was performed well
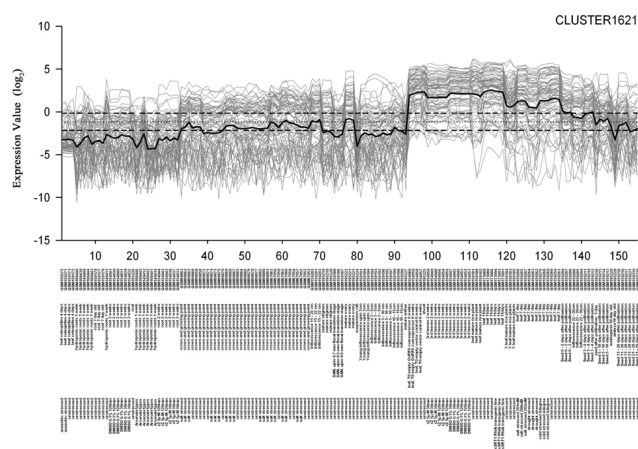
Fig. 2. Expression profile of numbers of genes that might be associated with the Calvin cycle. Thick lines and thin lines indicate average and individual gene expression profile, respectively. Dot and dash line represented average and expression value ± 2 fold values.

in function, two genes encoding phosphoribulokinase and sedoheptulose 1,7-bisphosphatase in the Calvin cycle were selected and then a cluster included both genes was detected. Interestingly, cluster 1621 with a set of 77 spots encoded some Calvin cycle enzymes, e.g. ribulose bisphosphate carboxylase small chain A, ribulose bisphosphate carboxylase/ oxygenase activase, glyceraldehyde-3-phosphate dehydogenase B, aldolase, fructose-bisphosphate aldolase, phosphoribulokinase and sedoheptulose 1,7-bisphosphatase. It revealed the notion that co-expressing genes grouping into a cluster appear to evidence common functional role. In addition, the co-expression genes to be grouped into the cluster 1621 appeared to be expressed more abundantly in leaves than in other tissues, reflecting some of the features of the Calvin cycle (Fig. 2).

The gene expression profiles of each of genes from identified spots were visualized. Three microarray probes
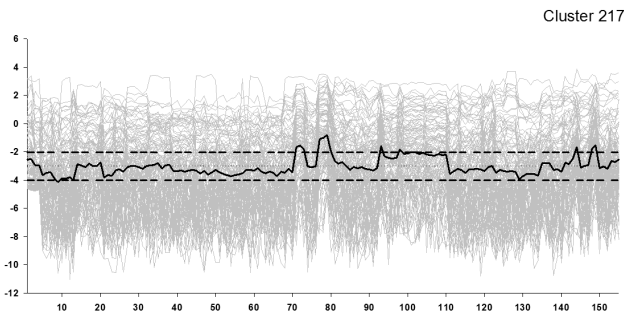
Fig. 3. Expression profile of cluster 217 that associated with up regulated protein spot in 5 MT. For descriptions in detail see the Fig. 2 legend.
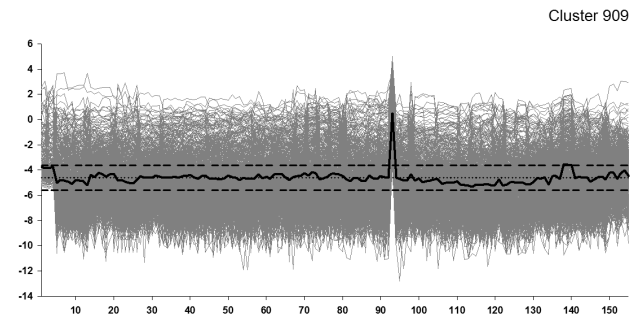
Fig. 5. Expression profile of cluster 909 that associated with up regulated protein spot in 5 MT. For descriptions in detail see the Fig. 2 legend.
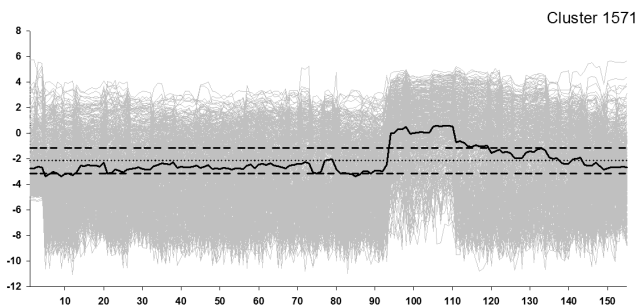
Fig. 4. Expression profile of cluster 1571 that associated with up regulated protein spot in 5 MT. For descriptions in detail see the Fig. 2 legend.

corresponding to the spots were founded. Three co-expressing gene clusters (217: M-2, 1571: M-3 and 909: C-1) were paralleled each functional categorization. All cluster showed high or low peaks with two-folds of the average expression values, indicating the different expression patterns of co-expression genes in each cluster by specific tissue and/or stress-response. Cluster 217 indicated high expressions profiling in mature stigmas and ovary tissue (Fig. 3). Tissue specific expression pattern was represented in this cluster. It may be appear that tryptophan biosynthesis was related to pathway for pollination biology. Moreover cluster 217 was included 3 glucosyl related spots (Os.26829.2.S1_s_at, Os.52346.1.S1_at and Os.54198.1.S1_at). The indole glucosinolates was believed to be derived from tryptophan (Haughn *et al.*, 1991). The diversion of tryptophan away from the indole glucosinolates biosynthetic pathway effectively redirected, quantitatively converted into tryptamine production. It is consistent with the hypothesis that tryptophan is a

precursor to the indole glucosinolates (Radwanski *et al.*, 1995). Cluster 217 was included candidate QTL related to drought stress (Yim and Jang, 2007). The probe (Os.4379. 1S1_at) corresponding to the TRAB1 gene, transcription factor responsible for ABA regulation (Hobo *et al.*, 1999), located on chromosome 1. This was included into cluster 217 together with 120 co-expressing genes, including two different transcription factors, Myb and transcription initiation factor IIB. It was predictable evidence that 5 MT mutant related drought stress resistance. Co-expressing genes in cluster 1571 appeared to be expressed in a common physiological function with 430 genes. For instance, 23 Leucine Rich Repeat family, 13 Cytochrome P450 family, 12 oxygen related, 11 ATP related and 10 Zinc finger proteins were included. The cluster showed mature leaf (over 2 weeks) specific expression pattern (Fig. 4). Cluster 909 was highly expressed in mature anther and two fold higher than average expression in inflorescence (Fig. 5). The expression levels of cluster were stable in broad dataset except for the tissues. The functional categories of these transcripts are protein synthesis and cellular communication and transcription. In comparison of co-expression genes related to cluster 217 and cluster 909, their components were similar. However, genes included in cluster 217 were more related to stress inducible gene (e.g. HAD family protein, BTB/POZ domain, acyltransferase family protein and AUX/IAA family protein).

DISCUSSION

In this study the utility of 2-DE and Microarray database

analysis approach for uncover gene discovery in rice mutant has been demonstrated. After inducing mutations, mutated regions were tried to identify. Several methods were employed to selection theses mutants for crop improvement. However, the methods including structural genomics, transcriptome, proteomics and metabolomics with plant physiology and plant conventional breeding, were unable to identify whole mutation region. Especially, the rapid development of bioinformatics tool may allow helping the previously mentioned methods. For this purpose, it was designed a new mutational analysis strategy to identify mutated genes.

Discovery of gene regulatory networks by using bioinformatics tools allows not only to gain a broad-scale view of how genes, proteins, and small molecules interact to regulate complex cellular processes (Zeng *et al*., 2007) but also to extend further to practice whole genome selection for the breeders to design the superior genotypes *in silico* (Varshney *et al*., 2005). This approach is likely to be a good example of gene network discovery by using a model plant. The criterion utilized in this report showed to be somewhat higher than those of other reports. For example, Blanc and Wolfe (2004) consider that any gene pairs with $r \geq 0.52$ at α=95% could be significantly co-regulated in *Arabidopsis*. In microorganism study, expression divergence of yeast duplicates as determined gene pairs with $r < 0.5$ (Gu *et al*., 2002). Therefore the variability threshold of 90%, representing an average r value of 0.66 evidenced that co-expressing genes involved in one cluster are possible to have a common functional role, thereby providing worthy of further study.

The 2-DE and Microarray analysis have been widely used as high-throughput screening tools for functional genomics. By using proteome analysis of differential protein expression and microarray analysis of differential gene expression, we find some overlaps in the result. This approach in present study was confirmed interaction of gene to protein. In the next logical step understands the way the function and pathway. Although not fully available annotated, the rice genome sequencing is now available. The most challenging task is to assign the function of each and every gene of a genome, thereby defining functional genomics as the next step for whole mutation analysis.

In plant, the tryptophan produce pathway are related to polypeptides and metabolized into auxin, glucosinolates, phytoalexins, alkaloids, and other indolic compounds (Radwanski and Last, 1995). These kind metabolites play diverse roles in plant biological processes, including development, plant-pathogen interactions, herbivory and pollination biology. The profile plots (Fig. 3, 5) showed pollination related and regulated differently. It is one of particular solution to understand how pathway is regulated and what mutated gene is for biosynthetic for tryptophan. The other cluster profile (Fig. 4), including drought related QTL gene and up-regulated protein spot, would be indicated 5MT resistance mutant with chance for tolerant / in tolerant about drought stress. It would not be a possibility but definitely related.

This study proposed a novel method for mutant screening method of based on *in silico*. This approach was tried an experiment of the mutant screening using the microarray database, and successfully discriminated against a 5 MT resistance mutant line. At practical to analyze as many rice mutants in our dataset as possible is desired. Thus it is necessary to conduct a standard microarray cluster dataset to process functional role. This cross analysis based on bioinformatics, microarray, and proteomics have important implications for understanding the functional role and pathway of mutant. In the same way, it has a wide possibility to apply to transformation plants and other breeding programs for crop improvement.

ACKNOWLEDGEMENT

REFERENCES

Blanc, G. and K. H. Wolfe. 2004. Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16 : 1679-1691.

de Hoon, M. J., S. Imoto, J. Nolan, and S. Miyano. 2004. Open source clustering software. Bioinformatics 20 : 1453-1454.

Devos, K. M. and M. D. Gale. 2000. Genome Relationships: The Grass Model in Current Research. The Plant Cell 12 : 637-646.

Gale, M. D. and K. M. Devos. 1998. Comparative genetics in

the grasses. Proc. Natl Acad. Sci 95 : 1971-1974.

Gromova, I. and J. E. Celis. 2004. Protein detection in gels by silver staining: A procedure compatible with mass-spectrometry. Cell Biology. A Laboratory Handbook.

Gu, Z., D. Nicolae, H. H. Lu, and W. H. Li. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends in Genetics 18 : 609-613.

Haughn, G. W., L. Davin, M. Giblin, and E. W. Underhill. 1991. Biochemical Genetics of Plant Secondary Metabolites in Arabidopsis thaliana: The Glucosinolates. Plant Physiology 97 : 217-226.

Herrero, J., A. Valencia, and J. Dopazo. 2001. A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics 17 : 126-136.

Hobo, T., Y. Kowyama, and T. Hattori. 1999. A bZIP factor, TRAB1, interacts with VP1 and mediates abscisic acid-induced transcription. Proc. Natl Acad. Sci 96(26) : 15348-15353.

W. C. Yim. and Jang, C. S. 2007. Discovery of gene regulatory networks via *in silico* analysis and their application in abiotic stress responses. Korean J. Breed 37(4) : 464-472.

Kamo, M., T. Kawakami, N. Miyatake, and A. Tsugita. 1995. Separation and characterization of Arabidopsis thaliana proteins by two-dimensional gel electrophoresis. Electrophoresis 16 : 423-430.

Kim, D. S., I. S. Lee, C. S. Jang, D. Y. Hyun, Y. W. Seo, and Y. I. Lee. 2004. Selection of 5-methyltryptophan resistant rice mutants from irradiated calli derived from embryos. Euphytica 135 : 9-19.

Kruger, N. J. 2002. The Bradford Method for Protein Quantitation. The Protein Protocols Handbook. Humana Press, Inc. 15-20.

Larkin, P. J. and W. R. Scowcroft. 1981. Somaclonal variation -a novel source of variability from cell cultures for plant improvement. Theor. App. Genet. 60 : 197-214.

O'Farrell, P. H. 1975. High resolution two-dimensional electro-phoresis of proteins. J Biol Chem 250 : 4007-4021.

Phillips, R. L., S. M. Kaeppler, and P. Olhoft. 1994. Genetic Instability of Plant Tissue Cultures: Breakdown of Normal Controls. Proc. Natl Acad. Sci 91 : 5222-5226.

Project, I. R. G. S. 2005. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. BMC Biol 3 : 20.

Radwanski, E. R., J. Zhao, and R. L. Last. 1995. Arabidopsis thaliana tryptophan synthase alpha: Gene cloning, expression, and subunit interaction. Mol. Genet. Genome. 248 : 657-667.

Sasaki, T. 1998. The rice genome project in Japan. Proc. Natl Acad. Sci 95(5) : 2027-2028.

Schaeffer, G. W., F. T. Sharpe, and P. B. Cregan. 1984. Variation for improved protein and yield from rice anther culture. Theor. App. Genet. 67 : 383-389.

Varshney, R. K., A. Graner, and M. E. Sorrells. 2005. Genomics-assisted breeding for crop improvement. Trends Plant Sci. 10 : 621-630.

Zeng, F., X. Zhang, L. Cheng, L. Hu, L. Zhu, J. Cao, and X. Guo. 2007. A draft gene regulatory network for cellular totipotency reprogramming during plant somatic embryogenesis. Genomics 90 : 620-628.

Zimmermann, P., L. Hennig, and W. Gruissem. 2005. Gene-expression analysis and network discovery using Genevesti-gator. Trends Plant Sci. 10 : 407-409.