Evolutionary Bioinformatics

ORIGINAL RESEARCH

Computational identification of anthocyanin-specific transcription factors using a rice microarray and maximum boundary range algorithm

Chang Kug Kim[1], Shoshi Kikuchi[2], Jang Ho Hahn[1], Soo Chul Park[1], Yong Hwan Kim[1] and Byun Woo Lee[3]

[1]Genomics Division, National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Suwon 441-707, Korea. [2]Division of Genome and Biodiversity Research, National Institute of Agrobiological Sciences (NIAS), Tsukuba, Ibaraki 305-8602, Japan. [3]Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-741, Korea. Corresponding author email: leebw@snu.ac.kr

Abstract: This study identifies 2,617 candidate genes related to anthocyanin biosynthesis in rice using microarray analysis and a newly developed maximum boundary range algorithm. Three seed developmental stages were examined in white cultivar and two black Dissociation insertion mutants. The resultant 235 transcription factor genes found to be associated with anthocyanin were classified into nine groups. It is compared the 235 genes by transcription factor analysis and 593 genes from among clusters of COGs related to anthocyanin functions. Total 32 genes were found to be expressed commonly. Among these, 9 unknown and hypothetical genes were revealed to be expressed at each developmental stage and were verified by RT-PCR. These genes most likely play regulatory roles in either anthocyanin production or metabolism during flavonoid biosynthesis. While these genes require further validation, our results underline the potential usefulness of the newly developed algorithm.

Keywords: black rice, anthocyanin-specific transcription factor, rice microarray, maximum boundary range algorithm

Evolutionary Bioinformatics 2010:6 133–141

doi: 10.4137/EBO.S6077

This article is available from http://www.la-press.com.

Introduction

Anthocyanins are an important group of natural pigments belonging to the flavonoid family. They are widely distributed in nature and responsible for the attractive red, purple, and blue colors of many flowers, fruits, and vegetables.[1] The accumulation of anthocyanins in leaves, shoots, and roots is stimulated by various environmental cues such as temperature, light, and nutrients.[2] Ethylene signaling has been shown to play a negative role in this accumulation in response to sucrose.[3] Many studies have been performed to reveal the functions of anthocyanins.[4–6] Anthocyanin-related transcription factors have been identified that can specify the class of flavonoid pigment produced regardless of positive or negative transcriptional regulation of structural genes.[7–10]

Black rice has high anthocyanin content located in the pericarp layers,[11] and is a good source of fiber, minerals, and several important amino acids.[12] The 135 K rice microarrays were used to identify novel genes that associate consistently with anthocyanin pigmentation in Dissociation (Ds) insertion lines.[13] These Ds insertion lines are one of the most important methods used in breeding rice and in understanding the function of the genome.[14] Several studies have used microarray successfully to analyze gene expression changes in a number of crop species, including black rice.[15,16]

A recent report showed that anthocyanin supplementation in humans improves LDL and HDL levels[17] and can delay cancer development in rodent models of carcinogenesis.[18] In addition, black rice may have anti-atherogenic activity[19] and may improve certain metabolic pathways associated with diets high in fructose.[20] These marked health benefits have been attributed to the antioxidant properties of anthocyanin.[21] These and other recent reports indicating the various potential health benefits of anthocyanin has led to increased interest in the use of this pigment in common food sources.[22] Therefore, the presence of anthocyanin in food could become a major factor considered by consumers when making a purchasing decision and, as such, could ultimately influence human behavior.[23]

In this study, we investigated anthocyanin gene expression in Ds insertion cultivars to gain insight into the causes of pigment production. By analyzing the transcription factors and Clusters of Orthologous Groups (COGs) method with a 135 K rice microarray and a newly developed maximum boundary range algorithm, we identified anthocyanin-specific genes in the black rice.

Methods

Rice

The *Dongjin* white rice cultivar was used in this study along with the black rice cultivars, Ds52306 and Ds92537. Two Ds insertion mutants generated from the *Dongjin* cultivar using the Ac/Ds gene trap system[24] were also studied. These cultivars were useful for comparing gene expression during anthocyanin biosynthesis because they possess similar genetic backgrounds and vastly different anthocyanin content.

Experimental design and statistical analysis

This experiment was designed to assess three factors (ie, *Dongjin* and two Ds cultivars) and three different seed development stages (ie, heading + 7 days, + 14 days, and + 21 days) in triplicate. The samples were harvested from the experimental fields of the National Academy of Agricultural Science (NAAS). The harvested seed samples were frozen in liquid nitrogen and ground to a powder using a mortar and pestle. All experiments were run using only the Cy3 dye to eliminate the dye-swap error value. Sequence, probe, and gene intensity data were collected. Spot intensity was calculated as the median value of the spot compared to the background median value. Gene expression analysis was performed using values produced by a newly developed maximum boundary range algorithm. Transcription factors were defined as those with *P*-values less than 0.05 using the TRANSFAC® 7.0 database (BIOBASE GmbH, Wolfenbuettel, Germany, http://www.gene-regulation.com/) and GeneSpring GX 11 software (Agilent Technologies Inc., Santa Clara, CA, USA, http://www.chem.agilent.com/).

Maximum boundary range algorithm

To classify candidate genes in the microarray data into new categories, we developed a new maximum boundary range algorithm. This new algorithm is a simple transformation procedure based on mechanisms abstracted from modified ratios of gene

expression intensity in the microarray. Therefore, this algorithm performs category identification differently from other existing methods.

Variables X (gene identification), Y (experimental treatment), and Z (intensity) in the microarray experiments are denoted as:

$$(X_i, Y_j, Z_{ij}),\ i = 1, ..., m,\ j = 1, ..., n \qquad (1)$$

where m = number of genes n = number of experimental treatment.

The Z_{ij} represent the functions $F(X_iY_j)$ of to be the values of the X, Y, and Z variables:

$$Z_{ij} = F(X_iY_j),\ i = 1, ..., m,\ j = 1, ..., n \qquad (2)$$

In order to screen the candidate genes, we define the ratio of gene expression intensity (ΔY_i) which is set equals to (Y_{max}/Y_{min}), where Y_{max} = maximum value and Y_{min} = minimum value among Y_j intensity range of X_i.

We assumed the ΔY_i value which is an important factor for evaluating gene selection. If this ratio has a high value, it indicates the great difference among treatments, although the raw intensity expressed a low value to compare the other genes. Table 1 shows the values of candidate genes using the ratio of gene expression intensity (ΔY_i). The other study use a highest intensity value by descending order with the K-nearest Neighbor (KNN) algorithm in the microarray gene expression analysis,[25] according to used the KNN method, candidate gene is "A005". However, this method will select an "A001" gene because ΔY_i has the highest value. To compare the difference between descending order method with the KNN and developed algorithm, we analyze the genes composition between the two groups. When two groups have a similar candidate gene numbers, the difference ratio of gene composition is from 25% to 40%. These ratios have been suggested that new gene (ie, it has a low intensity and big treatment difference) was inserted in the candidate group.

Table 1. The selection of candidate gene by ratio of gene expression intensity (ΔY_i) in the maximum boundary range algorithm.

Gene ID (X_i)	Intensity of each treatment (Y_j)					Ratio (ΔY_i)
A001	116	234	321	567	734	6.33
A002	3476	3516	3724	4762	2783	1.71
A003	12709	8903	16752	13540	13209	1.88
A004	27810	25761	30023	19834	14892	2.02
A005	38765	42361	40231	39782	35672	1.19

Transcription factor analysis

A total of 23,508 upstream sequences from rice were prepared. They were extracted 2 kb upstream of annotated translation starts, which were downloaded from NCBI (http://www.ncbi.nlm.nih.gov/genbank/). Because the rice genome contains 23,508 genes, known and putative motif-binding transcription factors of interest were identified using Match™ and Patch™ in TRANSFAC®.[26] We performed the hypergeometric test for pair-wise correlations between genes that were expressed in the three developmental seed stages and those among all the genes.[27,28]

RNA extraction

Total RNA was extracted from the seeds of one *Dongjin* white rice cultivar and two Ds black rice cultivars. We performed at least three replicates for each treatment. Frozen samples were homogenized with a mortar and pestle in liquid nitrogen. The ground powder was kept frozen in liquid nitrogen until the homogenization procedure was ready to be performed. At that point, 0.5 ml of RLC buffer (Qiagen, Hilden, Germany) was added. The homogenates were vortexed for 10 s, and plant debris was pelleted by centrifugation. RNA was extracted from the supernatant using the RNeasy Kit (Qiagen, Valencia, CA, USA). The RNA samples were further purified using phenol-chloroform-isoamylalcohol (25:24:1) and the RNeasy mini plant kit (Qiagen, Valencia, CA, USA). Total RNA was quantitated by measuring absorbance at 260 nm and 280 nm using the Nanodrop ND-1000-spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, NC, USA).

RT-PCR

RT-PCR was performed on total RNA isolated from the reproductive tract (eg,. uterus, spermathica, and ovaries), 1st–3rd instar larvae, pupa, and the remaining carcass after removal of the reproductive tract and intrauterine offspring. Total RNA (5 μg) from each sample was used to synthesize each pool of cDNA using the SuperScript III First-Strand Synthesis

System (Invitrogen, Carlsbad, CA, USA). For PCR amplification, 1 µl of the resulting cDNA reaction was used as a template. PCR reactions were carried out in a volume of 50 µl volumes using 20 pmol of each primer pair. The PCR program was as follows: 3 min at 94 °C, 25 cycles of 94 °C for 30 s, 60 to 65 °C for 30 s, and 1 min at 72 °C, followed by 5 min at 72 °C. To validate our RT-PCR results, we performed each experiment three times. Actin mRNA was used as a loading control. Primer sequences used for each gene are provided in Table 2.

Microarray experiment

In this study, we used the 3′-TILLING 135 K *Oryza sativa* microarray. Tiling arrays were used to identify specific genomic DNA regions.[29] The 135 K microarray was designed such that four probes could be used to cover a 150-bp region at the 3′ end of the gene. The probe sequences encompassed a region 60 bp upstream of the stop codon and with a 30-bp shift. Finally, the 125,956 probes used to assess the 31,439 genes were deposited at the International Rice Genome Sequencing Project (IRGSP) and the Rice Annotation Project version 2 (RAP2, http://rapdb.dna.affrc.go.jp/).

Results and Discussion

To evaluate the anthocyanin-specific genes in black rice, we performed a five-stage screen. First, the genes associated with signal intensity from the microarray were screened using a newly developed maximum boundary range algorithm. Second, candidate genes from the three seed developmental stages were selected. Third, transcription factors were identified using the hyper-geometric distribution analysis method. Fourth, COGs analysis was performed to classify the genes functionally and to identify orthologous genes. Finally, selected unknown and predicted genes were verified by reverse transcription-polymerase chain reaction (RT-PCR).

Candidate genes associated with anthocyanin biosynthesis

A total of 27 microarray experiments were performed to assess three groups (ie, *Dongjin* and two Ds cultivars) and three seed developmental stages (ie, heading + 7 days, + 14 days, and + 21 days) in triplicate. All signal intensity ratio values were transformed using a new algorithm. Using these transformed values, we screened 5,625 genes to identify those that were up- or downregulated greater than 3.0-fold. The number of upregulated genes across the three seed developmental stages ranged from 1,283 to 2,573, while the number of downregulated genes ranged from 538 and 1,728. A total of 5,625 genes were compared to the International Rice Genome Sequencing Project (IRGSP, http://rgp.dna.affrc.go.jp/E/IRGSP/) database and the rice genome system supported by the National Academy of Agricultural Science (NAAS) website (http://nabic.naas.go.kr/). Comparison with these databases enabled us to eliminate genes that were related to pigments other than anthocyanin. The 2,617 candidate genes were predicted to be potentially related to anthocyanin biosynthesis.

Transcription factors associated with anthocyanin

Transcription factors play important roles in all aspects of the life cycle of higher plants. Studies have suggested that co-expressed gene groups share many putative transcription factors within their promoters.[26,27] We identified nine groups that

Table 2. List of RT-PCR primer sequences used for validation of the nine unknown and hypothetical genes.

Status	Gene identification	Forward primer 5′—3′	Reverse primer 5′—3′
Up-regulation	Os01g0781600	GGC TCC AGT GCG GCA GGA GA	TCA CGC GTC ATG TCC ACG CC
	Os10g0315400	CGC CGT GCG GTT CTG CTT CT	TGG GAC ATT GCG GAT AGG CAG GT
	Os01g0633500	GCC CAC CGT GGA AGG GAT GC	GTG GAT GGT GGC GTC GTG GG
	Os08g0389700	CGTGTGGCCGCCGATGGAAT	ACACCGATGCTGATCCCTAGCA
	Os01g0615050	CCA CAC ATC GCG GTG CAA AAC C	CGA CCC AGA TGC GAA CGC GA
	Os01g0959100	GTG GAA ATG GGG CAC CAC CAC	GTG ATA GGG TAG GTA GCT AAT GG
	Os01g0748150	GCG CAA CCG GTT CCA GGT GA	GAA CGT CCT GGC GTC CTG CC
Down-regulation	Os01g0246400	GCT CGA TCG ATC TCT CTC TGC G	GTA CAC GCA TAC GTA CAT GCA CG
	Os02g0113800	TGC ATT TCA TCG CTG CAA GGT GT	GAA CTC CGG CTC CGG CTT CG

exhibited functionally diverse transcription factor activity involved in anthocyanin-based pigmentation biosynthesis. Of the 2,617 genes, 235 genes (8.9%) were identified as putative transcription factors. These results illustrate the functional diversity of the transcription factor genes and that these factors may be highly activated during anthocyanin production. Table 3 shows the number of transcription factor groups predicted from the 235 candidate genes in the three seed developmental stages.

Based on this analysis, we concluded that most of these transcription factor genes were related to the production of anthocyanins as well as other flavonoid compounds such as flavonols and proanthocyanidins. The MYB and GT transcription factor families showed the highest expression. Both groups accounted for approximately 35% of the predicted transcription factors. The CPRF, TALE, and ERF factors exhibited increased expression throughout seed development, while the MADS and Dof factors were decreased. Figure 1 shows the expression ratio of each specific transcription factor and numbers of predicted candidate genes during three seed developmental stages.

Table 3. Transcription factor groups predicted by the cumulative hyper-geometric test.

TFs	Description	Numbers of genes in tissue[a]		
		7D	14D	21D
MYB	A life cycle activating, Myb-like protein in cell	21	21	35
GT	Nuclear factors related to trans-acting in GT-motif	17	9	14
NAC	Genes related to flavonoid biosynthetic pathway	5	6	9
bHLH	Anthocyanin biosynthesis and signaling	4	4	7
MADS	Determination of floral meristem and organ identity	4	3	3
CPRF	Regulation of transcriptional activity of target genes	8	7	15
TALE	To control meristem formation and maintenance	8	4	9
ERF	Regulation related to growth and development	7	12	15
Dof	Seed storage protein synthesis in endosperm	6	2	3
Etc.	ATHB, AGL, CG-1, ABI, RAV of gene and TFs	27	19	37

Note: [a]In the hyper-geometric test, only *P*-values less than 0.05 were screened.

Involvement of the MYB and GT families, NAC, basic helix-loop-helix (bHLH), and MADS in colored tissues has been reported previously. MYB family members activate flavonoid pigmentation, are involved in specific steps of the anthocyanin pathway, and have evolved specific functions with different biochemical properties.[30–32] GT proteins consist of *trans*-acting factors possessing a GT-motif and include such factors as the nuclear protein of light-responsive in rice.[33–34] The NAC factors play an important role in regulating the expression of flavonoid biosynthesis-related genes,[35] senescence, cell division, and wood formation.[36] The bHLH factors are important regulatory components in transcriptional networks, controlling a diversity of processes from cell proliferation to cell lineage establishment,[37] as well as the determination of red seed color in anthocyanin biosynthesis.[38] Several MADS box genes are expressed preferentially in flowers and cause early flowering when ectopically expressed in rice plants.[39]

The other transcription factor group genes are most likely related to the production of seed storage, growth, and development. The common plant regulatory factor (CPRF) regulates the transcriptional activity of many target genes.[40] The three-amino-acid-loop-extension (TALE) transcription factor has been shown to control meristem formation and/or maintenance, organ morphogenesis, organ position, and several aspects of the reproductive phase.[41] The AP2/ERF family has important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental stimuli.[42] The Dof family has been reported to participate in the regulation of gene expression in processes such as seed storage protein synthesis in the developing endosperm and light regulation of genes involved in carbohydrate metabolism.[43]

Identification of gene expression conservation in orthologs

Conserved genes are importantly used to understand the major function in survival and replication of living organism. Orthology assignment in pigmentation biosynthesis is a critical prerequisite for numerous comparative genomics procedures,[44] and COGs analysis is important to the identification of orthologs.[45] In this study, COGs analysis was performed on the 2,617

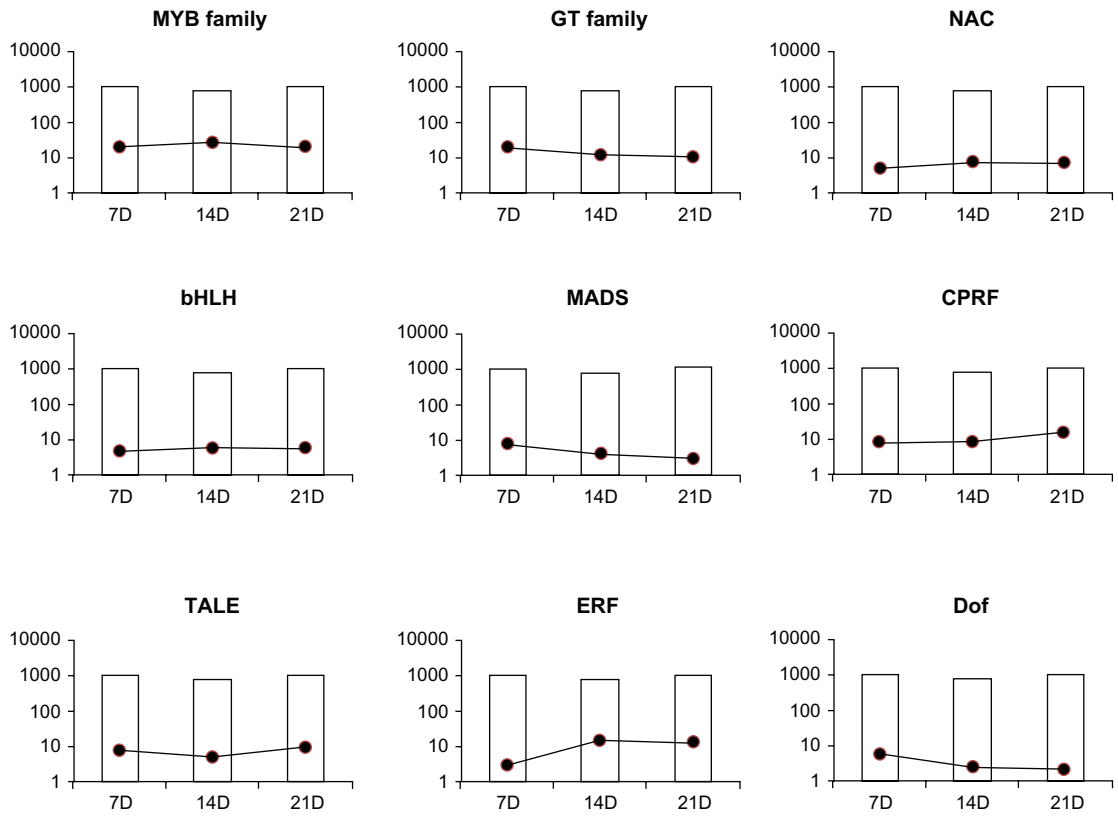

Figure 1. The expression pattern of specific transcription factor groups throughout three seed developmental stages. Line graphs indicate the value of expression ratio which account for specific factor among all transcription factors in the each seed stages. Bar graphs that indicate the number of candidate genes during each seed stages (ie, 7D, 14D, and 21D). 7D, 14D, and 21D represent each seed developmental stage studied, namely heading + 7 days, + 14 days, and + 21 days, respectively.

candidate genes using the NCBI/COGs database (http://www.ncbi.nlm.nih.gov/COG/). The distance value using CLUSTALW multiple alignment program (http://align.genome.jp/) was used as a descriptor of the conservation level of orthologs. Three specific categories were identified based on the presence of highly conserved genes from the distance value analysis, namely the Posttranslational modification, protein turnover, and chaperones cluster, Energy production and conversion cluster and the Carbohydrate transport and metabolism cluster. In addition, upregulated genes were classified by which general function (20.9%) showed the highest ratio in the heading + 14 day stage, and downregulated genes were classified by which general function (21.7%) showed the highest ratio in the heading + 21 days stage. Table 4 shows the results generated from COGs analysis of the 2,617 candidate genes that were up- or downregulated in the three seed developmental stages.

We assumed the clusters of COGs related to anthocyanin functions which three specific clusters have highest conserved genes. In addition, poorly characterized group of COGs were included because multiple candidate genes have unknown function. Total 593 genes were defined from among three specific clusters and poorly characterized group. We compare the 235 genes by transcription factor analysis and 593 genes in the clusters of COGs related to anthocyanin functions. In the common area, 32 genes were found to be expressed at each seed developmental stage. Among these, 23 genes were upregulated and nine genes were downregulated. Therefore, this transcription factor and COGs analysis method using the new algorithm was capable of identifying genes associated with specific biological processes. Finally, 9 unknown and hypothetical genes differed between the *Donjin* and two Ds mutant cultivars (ie, *Donjin/Ds52306* and *Ds92537*).

RT-PCR analysis of selected genes

To validate the selected genes identified by our new algorithm, RT-PCR was performed on the nine

Table 4. Clusters of Orthologous Groups (COGs) analysis of up- and downregulated genes during the three seed developmental stages.

Cluster identification		Upregulated			Downregulated		
		7D[a]	14D	21D	7D	14D	21D
Information storage							
J	Translation, etc	12	9	10	2	1	4
A	RNA processing and modification	17	15	17	4	3	8
K	Transcription	25	18	26	3	4	9
L	Replication, recombination, etc	16	11	12	1	2	4
B	Chromatin structure and dynamics	8	6	8	0	0	2
Cellular processing							
D	Cell cycle control, etc	16	11	19	1	1	5
Y	Nuclear structure	3	2	3	0	1	3
V	Defense mechanisms	4	3	2	1	0	1
T	Signal transduction mechanisms	46	36	46	3	7	14
M	Cell wall/membrane, etc	6	5	5	2	1	3
N	Cell motility	0	0	0	0	0	0
Z	Cytoskeleton	7	4	6	2	1	4
W	Extracellular structure	2	1	1	1	0	2
U	Intracellular trafficking, etc	16	13	20	3	4	9
O	Posttranslational modification, etc	37	25	31	5	6	17
Metabolism							
C	Energy production and conversion	14	11	15	1	1	5
G	Carbohydrate transport, etc	16	14	27	2	2	6
E	Amino acid transport, etc	12	9	9	2	1	2
F	Nucleotide transport and metabolism	5	3	4	1	0	3
H	Coenzyme transport and metabolism	6	4	5	3	1	4
I	Lipid transport and metabolism	15	6	16	2	2	5
P	Inorganic ion transport, etc	9	7	7	1	1	3
Q	Secondary metabolites, etc	11	7	10	5	1	2
Poorly characterized							
R	General function prediction only	76	63	73	15	16	35
S	Function unknown	21	18	19	5	6	11

Note: [a]7D, 14D, and 21D represent each seed developmental stage studied, namely heading + 7 days, + 14 days, and + 21 days, respectively.

unknown and hypothetical genes using the same RNA samples used in the microarray experiments. Total RNA was isolated from reproductive tissues at 7, 14, and 21 DAF (day after heading) rice seed developmental stages. Among the nine genes, seven were upregulated, while two were downregulated (Fig. 2). The RT-PCR results confirmed the expression profiles revealed by the microarray experiments.

The Os01g0781600 and Os01g0748150 genes were highly induced during the early heading stage, suggesting that these genes may play a role in anthocyanin production in the early rice heading stage. Alternatively, the Os01g0615050 gene was not induced until the late heading stage in the two Ds cultivars, indicating that this gene may be critical for the late rice heading stage. Interestingly, the Os01g0633500 gene exhibited vast differences in expression between the *Donjin* and two Ds cultivars at all the seed developmental stages. Because this gene was not expressed in the *Dongjin* cultivar without anthocyanin, it may play a role in anthocyanin biosynthesis. Among the two downregulated genes, Os01g0246400 showed expression in all developmental stages of the *Dongjin* cultivar. We conclude that these genes were related to the inhibition of anthocyanin metabolism or somehow affect the anthocyanin biosynthesis pathway negatively.

Conclusions

We identified putative transcription factor genes related to anthocyanin biosynthesis using the 135 K rice microarray. This microarray is a powerful tool for

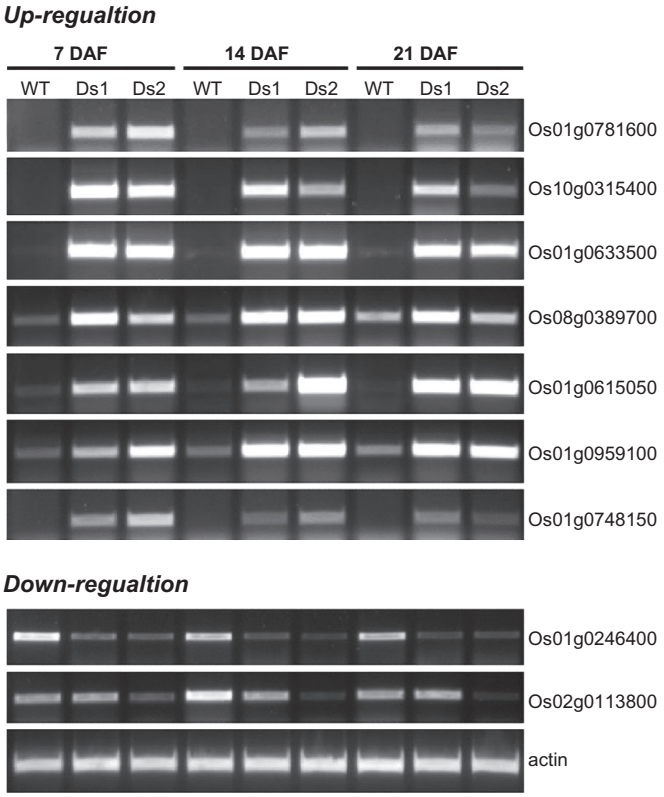

Figure 2. RT-PCR analysis of selected genes identified by microarray analysis. Nine candidate genes identified from the microarray analysis were assessed by RT-PCR in two Ds insertion mutants. Actin was used as a loading control. DAF = Day after heading, WT = white rice, *Dongjin* cultivar, Ds1 = black rice, *Ds52306* mutant, Ds2 = black rice, *Ds92537* mutant.

studying the complex biological responses involved in anthocyanin pigmentation biosynthesis in rice[13]. We performed a five-stage analysis to evaluate the gene expression involved in anthocyanin biosynthesis. First, we transformed the signal intensity of expressed genes using a newly developed maximum boundary range algorithm. We screened 5,625 genes to identify those demonstrating a greater than 3.0-fold up- or downregulation. Second, we selected 2,617 anthocyanin-specific candidate genes. Third, the 235 transcription factor genes were identified using the hyper-geometric test. Fourth, COGs analysis was performed to identify functional classification categories for these genes. Among these, 9 genes were unknown and hypothetical at each seed developmental stage. Finally, nine genes were verified by RT-PCR.

The results from this study suggest that these nine genes play either a regulatory role in the anthocyanin production process or are related to anthocyanin metabolism during flavonoid biosynthesis. While the role of these genes requires further investigation and validation, our results demonstrate the potential and usefulness of this method and algorithm. This computational study provides valuable insight into pigmentation production and will greatly facilitate the future breeding of anthocyanin-rich hybrid rice varieties.

Acknowledgements

This work was supported by the National Academy of Agricultural Science (No. PJ006651 and No. PJ006758).

Disclosure

This manuscript has been read and approved by all authors. This paper is unique and is not under consideration by any other publication and has not been published elsewhere. The authors and peer reviewers of this paper report no conflicts of interest. The authors confirm that they have permission to reproduce any copyrighted material.

References

1. Jackman RL, Yada RY, Tung MA, Speers RA. Anthocyanins as food colorants—a review. *Journal of Food Biochemistry*. 1987;11:201–47.
2. Chalker-Scott L. Environmental significance of anthocyanins in plant stress responses. Photochem. *Photobiol*. 1999;70:1–9.
3. Kwon YR, Oh JE, Noh HN, Hong SW, Bhoo SH, Lee HJ. The ethylene signaling pathway has a negative impact on sucrose-induced anthocyanin accumulation in Arabidopsis. Journal of Plant Research. 2010;online DOI 10.1007/s10265-010-0352-3.
4. Boulton R. The copigmentation of anthocyanins and its role in the color of red wine: a critical review. Am. J. Enol. *Viticult*. 2001;52:67–86.
5. Ovando AC, Hernandez MLP, Hernandez MEP, Rodriguez JA, Vidal CAG. Chemical studies of anthocyanins: A review Food Chem. 2009; 113:859–71.
6. Kim CK, Park SH, Kikuchi S, et al. Genetic analysis of gene expression for pigmentation in Chinese cabbage (Brassica rapa). *BioChip*. 2010;4(2):123–8.
7. Li J, Li X, Guo L, et al. A subgroup of MYB transcription factor genes undergoes conserved alternative splicing in Arabidopsis and rice. *Journal Experimental Botany*. 2006;57:1263–73.
8. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. *Plant J*. 2004;39:366–80.
9. Hartmann U, Sagasser M, Mehrtens F, Stracke R, Weisshaar B. Differential combinatorial interactions of cis-acting elements recognized by R2R3-MYB, BZIP, and BHLH factors control light-responsive and tissue-specific activation of phenylpropanoid biosynthesis genes. Plant Mol. *Biol*. 2005;57:155–71.
10. Reddy AM, Reddy VS, Scheffler BE, Wienand U, Reddy AR. Novel transgenic rice overexpressing anthocyanidin synthase accumulates a mixture of flavonoids leading to an increased antioxidant potential. *Metab Eng*. 2007;9:95–111.
11. Ryu SN, Park SZ, Ho CT. High performance liquid chromatographic determination of anthocyanin pigments in some varieties of black rice. *J. Food Drug Anal*. 1998;6:729–36.
12. Zhang MW, Guo BJ, Peng ZM. Genetic effects on grain characteristics of indica black rice and their uses on indirect selections for some mineral element contents in grains. *Genet Resour Crop*. 2005;52:1121–8.

13. Kim CK, Kikuchi S, Kim YK, et al. Computational identification of seed-specific transcription factors involved in anthocyanin production in black rice. *BioChip Journal*. 2010;4(3):247–55.
14. Zhu QH, Eun MY, Han CD, et al. Transposon insertional mutants: a resource for rice functional genomics. In: Upadhyaya NM, editor. Rice functional genomics challenges progress and prospects. *Springer Press*. 2007;224–62.
15. Nagira Y, Shimamura K, Hirai S, Shimanuki M, Kodama H, Ozeki Y. Identification and characterization of genes induced for anthocyanin synthesis and chlorophyll degradation in regenerated torenia shoots using suppression subtractive hybridization, cDNA microarrays, and RNAi techniques. *J Plant Res*. 2006;119(3):217–30.
16. Kim CK, Kikuchi K, Satoh K, et al. Genetic Analysis of Seed-specific Gene Expression for Pigmentation in Colored Rice. *BioChip*. 2009;3:125–9.
17. Qin Y, Xia M, Ma J, et al. Anthocyanin supplementation improves serum LDL- and HDL-cholesterol concentrations associated with the inhibition of cholesteryl ester transfer protein in dyslipidemic subjects *Am J Clinical Nutrition*. 2009;90(3):485–92.
18. Thomasset S, Teller N, Cai H, et al. Do anthocyanins and anthocyanidins, cancer chemopreventive pigments in the diet, merit development as potential drugs? *Cancer Chemother Pharmacol*. 2009;64(1):201–11.
19. Xia X, Ling W, Ma J, et al. An anthocyanin-rich extract from black rice enhances atherosclerotic plaque stabilization in apolipoprotein E-deficient mice. *Journal Nutr*. 2006;136:2220–5.
20. Guo H, Ling W, Wang Q, Liu C, et al. Effect of anthocyanin-rich extract from black rice (Oryza sativa L. indica) on hyperlipidemia and insulin resistance in fructose-fed rats. *Plant Foods Hum Nutr*. 2007;62:1–6.
21. Wu X, Prior RL. Systematic identification and characterization of anthocyanins by HPLC–ESI–MS/MS in common foods in the United States: fruits and berries. *Journal of Agricultural and Food Chemistry*. 2005;53:2589–99.
22. Zhang Z, Kou X, Fugal K, McLaughlin J. Comparison of HPLC methods for determination of anthocyanins and anthocyanidins in bilberry extracts. *Journal of Agricultural and Food Chemistry*. 2004;52:688–91.
23. Robert Y, Naofumi M. Color enhancing effect of carboxylic acids on anthocyanins. *Food Chemistry*. 2007;105:421–7.
24. Park DS, Park SK, Han SI, et al. Genetic variation through Dissociation (Ds) insertional mutagenesis system for rice in Korea: progress and current status. *Molecular Breeding*. 2009;24:1–15.
25. Sampath D, Henrik B. Classification of Microarrays with kNN: Comparison of Dimensionality Reduction Methods. Intelligent Data Engineering and Automated Learning—IDEAL 2007. *Lecture Notes in Computer Science*. 2007; 4881:800–9.
26. Matys V, Fricke E, Geffers R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*. 2003;31:374–8.
27. Chen L, Zhao H. Gene expression analysis reveals that histone deacetylation sites may serve as partitions of chromatin gene expression domains. *BMC Genomics*. 2005;6:44–53.
28. Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res*. 2003;13:773–80.
29. Taskesen E, Beekman R, de Ridder J, et al. HAT: hypergeometric analysis of tiling-arrays with application to promoter-GeneChip data. *BMC Bioinformatics*. 2010;11:275.
30. Martin C, Paz-Ares J. MYB transcription factors in plants. *Trends Genet*. 1997;13:67–73.
31. Borevitz JO, Xia Y, Blount J, Dixon RA and Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell*. 2000;12:2383–93.
32. Sheng T, Joost K, Maarten K, Sjef S. Sucrose-Specific Induction of Anthocyanin Biosynthesis in Arabidopsis Requires the MYB75/PAP1 Gene1. *Plant Physiology*. 2005;139:1840–52.
33. Reddy VS, Scheffler BE, Wienand U, Wessler SR, Reddy AR. Cloning and characterization of the rice homologue of the maize C1 anthocyanin regulatory gene. *Plant Mol Biology*. 1998;36:497–8.
34. Yujie F, Kabin X, Xin H, Honghong H, Lizhong X. Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. *Mol Genet Genomics*. 2010;283:157–69.
35. Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. Arabidopsis NAC Transcription Factor, ANAC078, Regulates Flavonoid Biosynthesis under High-light. *Plant Cell Physiol*. 2009;50:2210–22.
36. Fang Y, You J, Xie K, Xie W, Xiong L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Mol Genet Genomics*. 2008;280(6):547–63.
37. Toledo-Ortiz G, Huq E, Quail PH. The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell*. 2003;15(8):1749–70.
38. Antonio G, Mingzhe Z, John ML, Alan ML. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant Journal*. 2008;53:814–27.
39. Jeon JS, Jang S, Lee S, et al. Leafy hull sterile1 (lhs1) is a homeotic mutation in a rice MADS box gene affecting rice flower development. *Plant Cell*. 2000;12:871–84.
40. Wellmer F, Schäfer E, Harter K. The DNA binding properties of the parsley bZIP transcription factor CPRF4a are regulated by light. *J Biol Chem*. 2001;276(9):6274–9.
41. Hamant O, Pautot V. Plant development: a TALE story. Comptes Rendus Biologies. 2010;333(4):371–81.
42. Toshitsugu N, Kaoru S, Tatsuhito F, Hideaki S. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiol*. 2006; 140(2):411–32.
43. Lijavetzky D, Carbonero P, Vicente-Carbajosa J. Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *BMC Evolutionary Biology*. 2003;3:17.
44. Dessimoz C, Boeckmann B, Roth AC, Gonnet GH. Detecting non-orthology in the COGs database and other approaches grouping orthologs using genome-specific best hits. *Nucleic Acids Res*. 2006;34:3309–16.
45. Remm M, Storm C, Sonnhammer E. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol*. 2001; 314:1041–52.