Agric. Chem. Biotechnol. **48**(3), 127-132 (2005)

Article

Classification and Expression Profiling of Putative R2R3 MYB Genes in Rice

Bong Gyu Kim, Jae Hyung Ko, Shin Young Min and Joong-Hoon Ahn*

Department of Molecular Biotechnology, Bio/Molecular Informatics Center, Konkuk University, Seoul 143-701, Korea

Received July 11, 2005; Accepted August 18, 2005

MYB genes, comprising group of related genes found in animal, plant, and fungal genomes, encode common DNA-binding domains composed of one to four repeat motifs. MYB genes containing two repeats (R2R3) constitute largest MYB gene family in plants. R2R3 MYB genes play important roles in regulation of secondary metabolism, control of cell shape, disease resistance, and hormone response. Eight-four R2R3 MYB genes were retrieved from rice genome for functional characterization of MYB genes. Analysis of MYB domains revealed each MYB domain contains three α-helices with regularly spaced tryptophan residues. R2R3 MYB genes were divided into four subfamilies based on phylogenic analysis result. Real-time PCR analysis of 34 MYB genes revealed 12 MYB genes were highly expressed in seeds than in leaves, whereas 4 genes were highly expressed in leaves.

Key words: *MYB transcription factors,* Oryza sativa*, real-time PCR*

Organisms have delicate systems to control complicated biochemical processes. One of the control points for the regulation of gene expressions is at the transcription level. Various protein involved in various proteins, known as transcription factors, including bHLH, bZIP, WERKY, and MYB are composed of domains serving as classification units.[1)]

MYB gene originating from v-Myb oncogene of the myeloblatosis virus[2)] became a classification unit of the transcription factors. Members of MYB gene family were identified in diverse plants and animals with animals containing relatively few MYB genes compared to the plants.[3)] MYB proteins are defined by a highly conserved DNA-binding domain (termed the MYB domain), with each domain composed of one to four helix-turn-helix motifs, which exist as tandem repeats (referred to as R0R1R2R3) in a single MYB protein. Each repeat is composed of about 50 amino acids with regularly spaced tryptophan residues, and forms three α-helices. The first two helices of each MYB genes form a helix-turn-helix motif with the third α-helix involved in the DNA binding.[4)] Major MYB proteins found in plants consists of two MYB domains (called R2R3 type). Even though biological functions of most MYB proteins are still under investigation, some of them are known to be involved in the regulation of secondary metabolism, the control of cell shape, disease resistance, and hormone response.[1,3,5-8)] MYB proteins may also play crucial roles in generating morphological and physiological variations by controlling the developmental processes.[9-12)]

More than 100 R2R3 MYB genes have been identified from the *Arabidopsis* genome,[1,6)] with some of their biological functions characterized. Eight-five R2R3 genes have so far been found in rice[13)] with more MYB genes expected to be discovered. Completion of the rice genome sequencing will enable the entire complement of MYB genes to be identified and described. Therefore, much experimental work is required to determine the specific biological function of each gene, and as an initial step, MYB genes from rice genome were retrieved and classified to explore their biological functions. Expression patterns of the 34 selected MYB genes from seeds and leaves were also analyzed.

Materials and Methods

Database search. Multiple databases searches were performed to identify the members of R2R3 MYB genes Basic Local Alignment Search Tool (BLAST) available on the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov) and TIGR from The Institute for Genomic Research (http://www.tigr.org) were used to search the published sequence of the entire rice genome. As a query sequence, we used the conserved domain of R2R3 MYB family. Duplicated R2R3 MYB genes were sorted out and excluded.

Sequence pileups and phylogeny construction. The sequences were aligned using the program cluster W with the default parameters (http://prodes.toulouse.inra.fr/multalin.multalin.html) and further adjusted visually. The BoxShade program (http://www.ch.embnet.org/softwareBox_doc.html) was used to highlight conserved and similar amino acids. Phylogenetic tree was constructed with Genebee (http://www.genebee.msu.su/services/phtree_reduced.html).

RNA isolation and reverse transcription. Total RNA from stem and seed of *Oryza sativa* japonica cv. Heukjinju

*Corresponding author
Tel: 82-2-450-3764; Fax: 82-2-343706196
E-mail: jhahn@konkuk.ac.kr

Table 1. List of primers used for real-time PCR

Gene	Forward Primer	Reverse Primer	Gene	Forward Primer	Reverse Primer
KKAMYB 1	agatggaggcccttctgtgc	ccagtccatgtcgagcaggt	KKAMYB 23	gctcatcttccgattcgttc	cacctcctgcatccagtctt
KKAMYB 2	cgtgttgtgcagcttgaact	gtacgaagcttccccacctt	KKAMYB 28	gctttctggagggagtttga	catcttgaagaggtggagca
KKAMYB 3	tccctttggagaccaacat	aacgttccatggagcaagtc	KKAMYB 29	gtggccttttggaagagttg	tacatggcgtaccgacagaa
KKAMYB 4	cataccaagcagcgacgac	cacgaattggactgcatcgt	KKAMYB 31	acgtccgagtcgtcgctg	tggtcgtcctccatgctc
KKAMYB 5	aacgatgacatgtggggctc	tacaggttgcccaagaagaa	KKAMYB 32	ccaccagatgcaggatcagt	cgtcgtcgtagtacagccc
KKAMYB 6	gcgacagctccagcaacag	gaacatgtccatgtgcggc	KKAMYB 35	ggtcaagaaccagagaccca	tctctgcaacggcagaagta
KKAMYB 10	acgtcgtcgagctcattacc	gtcgtcggtgtcaagaaacg	KKAMYB 43	ccctgttgttgtttgttgtt	accccacatgacccaaatac
KKAMYB 11	tgccgattggctataagctc	gtcaccgtcgtcgtcttcct	KKAMYB 44	ctgttcccggaattctgcat	gacaagctcggatcgagagg
KKAMYB 12	atgcagttcatggtggcttc	gagatcgcctaactgctcgt	KKAMYB 53	tccagtgaatggaaaggagg	gggaaagagcagagggttct
KKAMYB 13	ccaccactactcggagacga	aggaggaggaagaggccgac	KKAMYB 81	ggacgactcctgctcctgc	gttgtccacgagaacaccg
KKAMYB 15	gacagcttctggtcggagac	cagtagtccatgtcgtcggc	KKAMYB 85	cggtgatggacttcatggac	aatccagtgaggttgcatgg
KKAMYB 16	gtgccccgacctcaacct	agccgaggtggtagcagag	KKAMYB 89	tgggatgaagaagaagcaca	ggtcccataaaaaggaggga
KKAMYB 17	atatcggagcaggcttcaca	cacctggcttctttgctcat	KKAMYB 90	gatgacgaacccgttggact	ggtcgtcgtggaagaactcc
KKAMYB 18	actgctgcttcgacgttcat	catcagctgctgcttgctct	KKAMYB 93	cacggtctcgagtagcaaca	tcgtcgtcgtcgtcgtagag
KKAMYB 19	ggaggtgaccgggatgatg	tgagccagaactccatgtcg	KKAMYB 96	ggatagaccccgtcacgca	caagtccaagggcagctggt
KKAMYB 20	actgctgcttcgacgttcat	catcagctgctgcttgctct	KKAMYB 102	atgactatgaggcgaagccg	gcacttgcactccttctgga
KKAMYB 21	agtcgatgtacctgatgggg	gatcctctggtacgcgttga	OsMYB1	cttcgtcatctcgcggac	tgatgtcgtggatgctgg

Table 2. Rice R2R3MYB used in this study and their accession numbers

KKAMYB	Accession number	KKAMYB	Accession number	KKAMYB	Accession number
KKAMYB 01	AAL84631	KKAMYB 33	BAB39921	KKAMYB 110	AAL84630
KKAMYB 02	CAD40990	KKAMYB 34	BAB92511	KKAMYB 111	CAE05473
KKAMYB 03	BAB39972	KKAMYB 36	BAD08950	KKAMYB 112	AAP92750
KKAMYB 04	AAO62334	KKAMYB 40	NP_917132	KKAMYB 113	BAD07916
KKAMYB 05	BAB20661	KKAMYB 44	NP_914560	KKAMYB 117	AAL78372
KKAMYB 06	BAB89216	KKAMYB 45	CAA65525	KKAMYB 120	CAE04573
KKAMYB 07	CAC85051	KKAMYB 46	NP_914382	KKAMYB 121	BAD04025
KKAMYB 08	BAC07040	KKAMYB 47	AAL84627	KKAMYB 122	BAD04029
KKAMYB 09	BAC20674	KKAMYB 52	AAF34434	KKAMYB 124	BAC79723
KKAMYB 10	CAA75509	KKAMYB 55	NP_916704	KKAMYB 125	AAL84626
KKAMYB 11	BAA23339	KKAMYB 58	XP_470731	KKAMYB 126	NP_916974
KKAMYB 12	BAA23340	KKAMYB 62	CAC19439	KKAMYB 128	BAD10148
KKAMYB 13	BAA23338	KKAMYB 75	NP_909096	KKAMYB 129	BAD04027
KKAMYB 14	CAA72218	KKAMYB 85	NP_916576	KKAMYB 132	AAM08125
KKAMYB 15	CAA72217	KKAMYB 87	NP_918017	KKAMYB 133	AAL78373
KKAMYB 16	BAA23337	KKAMYB 88	NP_917110	KKAMYB 134	BAD04026
KKAMYB 17	CAD44619	KKAMYB 89	BAC64999	KKAMYB 137	BAD05679
KKAMYB 18	CAA72187	KKAMYB 90	BAD09322	KKAMYB 139	CAE04731
KKAMYB 19	CAA72185	KKAMYB 93	BAC84030	KKAMYB 141	CAD41609
KKAMYB 21	CAA72186	KKAMYB 95	BAD10751	KKAMYB 142	NP_922180
KKAMYB 23	AAK08983	KKAMYB 96	AAL84624	KKAMYB 143	NP_912968
KKAMYB 24	BAB89293	KKAMYB 98	AAR06367	KKAMYB 144	BAD10634
KKAMYB 25	NP_910296	KKAMYB 102	AAL84628	KKAMYB 145	BAD08736
KKAMYB 28	BAA23341	KKAMYB 103	CAE03051	KKAMYB 150	AAM47303
KKAMYB 29	CAA67000	KKAMYB 104	CAC85050	KKAMYB 151	NP_914401
KKAMYB 30	BAC22341	KKAMYB 106	CAE04569	KKAMYB 154	BAB61618
KKAMYB 31	BAB67851	KKAMYB 108	BAD079322	KKAMYB 160	BAC84082
KKAMYB 32	BAC21351	KKAMYB 109	CAE05465	OsmybS1	AAN63152

was isolated using the Qiagen plant total RNA isolation kit (Qiagen, Gaitherburg, MD, USA). The extracts were treated with Rnase-free Dnase I (Roche, Indianapolis, IN, USA) to eliminate the residual DNA present in the preparation. First strand cDNA was synthesized using Ominiscript transcriptase (Qiagen) with oligo $(dT)_{15}$.

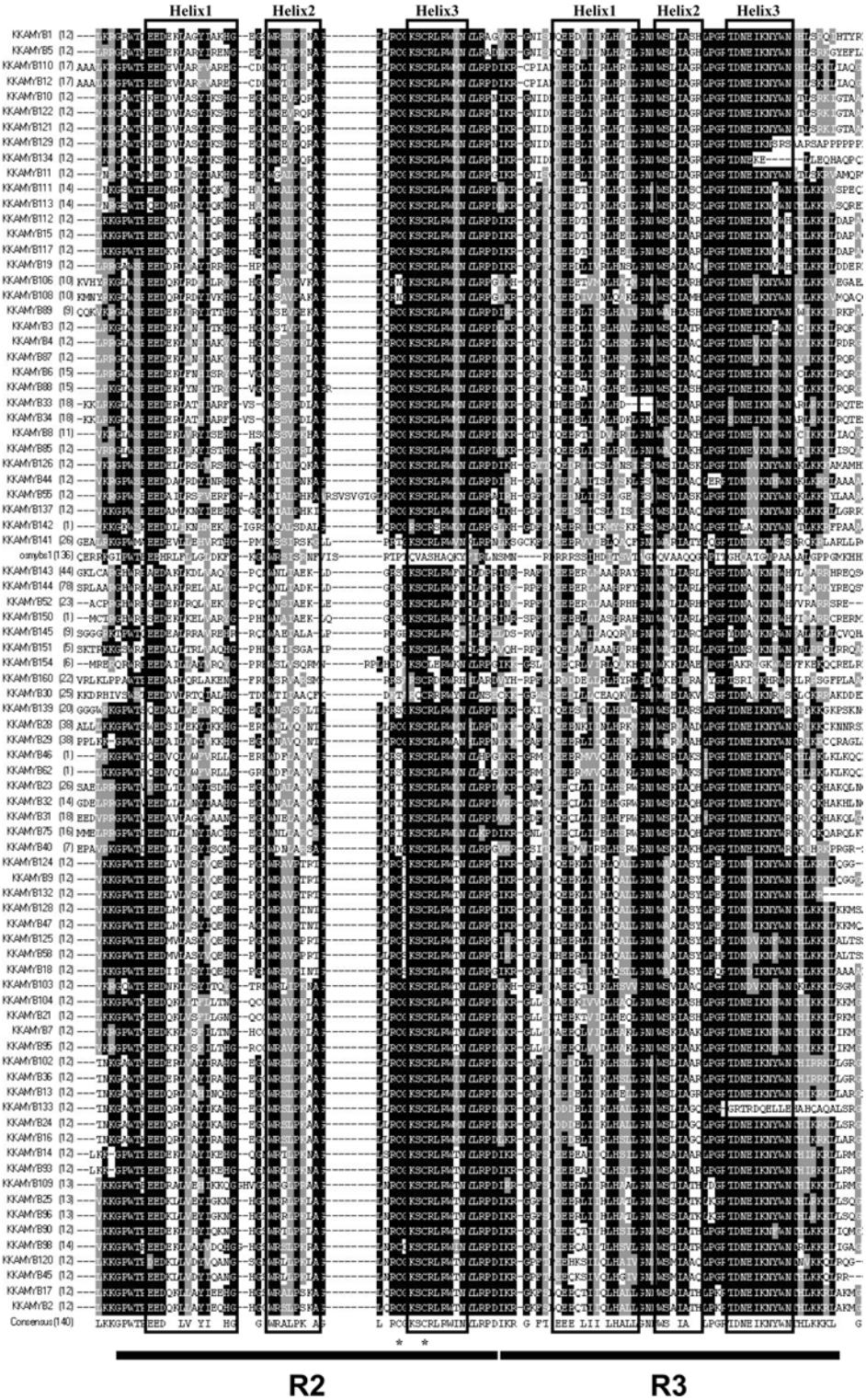

Fig. 1. A amino acid sequence alignments of DNA-binding domain from R2R3 MYB genes from rice. Well-conserved cystein residues are marked as *.

Real time-PCR. cDNA was amplified using SYBR-GreenR PCR Kit containing Hotstart DNA polymerase (Qiagen, Germany) on a Rotor-Gene RG-3000A themocycler (Corbett Research, Sydney, Austrailia). Primers used are listed in Table 1. The PCR was performed as follows: incubation at 95°C for 15 min to activate the hot start Taq DNA polymerase, followed by 45 cycles of 10 s at 95°C, 15 s at 60°C, and 20 s at 72°C. Specificity of the PCR amplification was checked throughheat dissociation kinetics. Specificity of each primer set was also checked by sequencing the PCR products. The results obtained on different samples were standardized to the constitutive actin gene expression level.

Results and discussion

Database search and Phylogeny Construction. MYB consensus sequence LKKGPWTPEEDEKLINYILKHGEGN WRSLPKKAGLKRCGKSCRLRW[1)] was used as the query sequence to search the rice genome and 163 MYB homologues were found. Elimination of the overlapped clones resulted in 104 MYB genes, with 1, 12, and 91 showing three, one, and two copies of MYB domain, respectively. By eliminating those showing incomplete MYB domains or those with the location of MYB domains shifted to C-terminal, total 84 R2R3 genes were chosen (Table 2).

Alignments of 84 rice R2R3 MYB proteins were carried out. The N-terminal portion at which MYB domains are located is well-conserved, while the C-terminal portion is diverse. Only the alignment of MTB domain is shown in Fig 1. As with the other MYB proteins, each MYB domain consists of about 50 amino acids and contains three α-helices with regularly spaced tryptophan residues. Among the three α-helices of each MYB domain, the third helix, known to contact directly with a target DNA, showed more similarity than the others. Two cystein residues (marked as stars in Fig 1) were well-conserved. In particular, the second cystein located at the third α-helix of the first MYB domain was conserved in all the R2R3 MYBs. These two cysteins are known to form a disulfide bond under non-reducing conditions, which prevents R2R3 MYB domains from binding to the DNA,[14)] suggested as one of the MYB transcription factor regulation mechanism.

Eight four rice MYB proteins were divided into four major groups based on the phylogenetic analysis of each MYB domain (Fig. 2). Arabidopsis MYB showed only three groups.[15)]

Analysis of tissue-specific expression using real-time PCR. R2R3 MYB genes are known to mediate various metabolic events. We were interested in those involved in seed-specific R2R3 MYB. To this end, expression patterns of 34 R2R3 genes whose full length cDNA were already cloned were analyzed by real-time PCR. The primers of each gene were designed, and the specificity of the primers was verified by sequencing the PCR product. To isolate the total RNA, leaves were harvested upon maturation of the seeds. Templates were prepared through reverse transcription with RNAs from seeds and leaves. Relative expression profiles of the 34 R2R3 MYB genes in leaves and seeds were examined. Figure 3 illustrates the relative transcript levels of the 34 R2R3 MYB genes standardized to the constitutive actin gene expression level. Due to high variations in the relative transcript levels, a logarithmic scale was chosen to present the results.

In leaves, expression levels of *KKAMYB2, 4, 6*, and *43* were much lower than those of others. Furthermore whether they were transcriptionally active in leaves could not be determined with certainty. *KKAMYB6* also showed low expression in seeds. On the other hand, *KKAMYB53* showed the highest expression levels in both leaves and seeds showing about 72

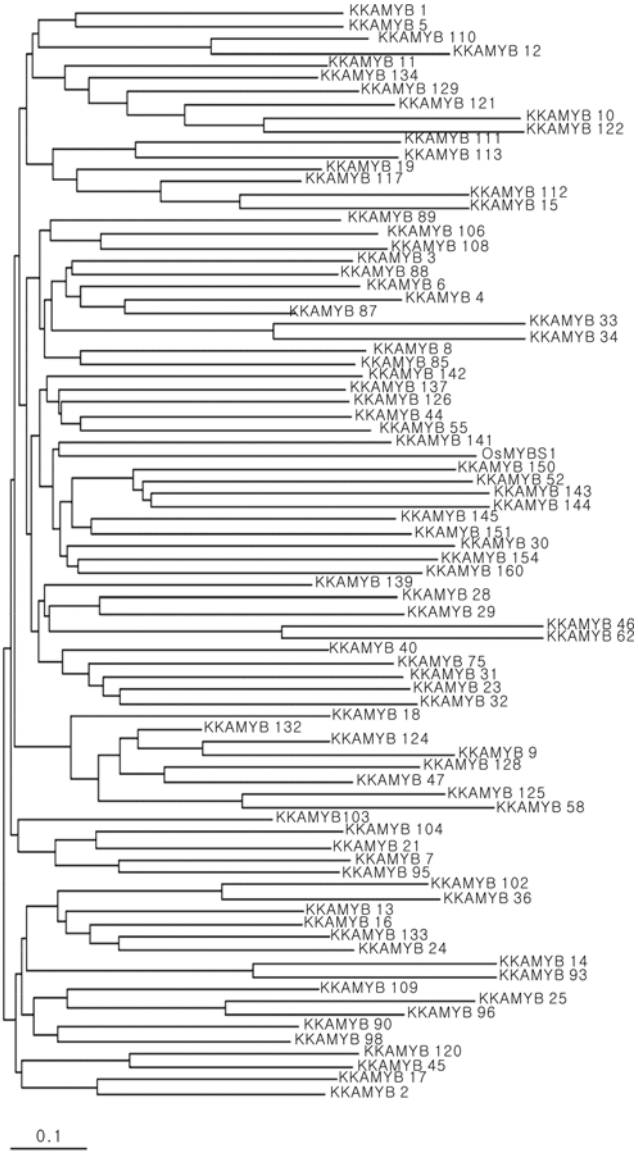

Fig. 2. Phylogenetic relationship among the DNA-binding domains of rice R2R3 MYB genes.

% of actin in both leaves and seeds. *KKAMYB5, 16, 23, 44, 81*, and *93* exhibited relatively higher expressions in leaves than in seeds, raging from 10 to 31 % of actin.

Seeds displayed different expression patterns. As mentioned above, *KKAMY53* showed the highest expression in seeds. *KKAMYB12, 13, 15, 16, 23, and 44* also showed the higher expressions in seeds, ranging from 15 to 84% of actin, among which , only *KKAMYB 16* showed high expression both in seeds and but also in leaves. In addition, *KKAMYB2, 4, 10,* and *85* displayed higher expressions in seeds compared to leaves, even though the overall expression levels of them in both tissues were relatively lower. *KKAMYB 28* and *89* were expressed 45 and 73 times higher in leaves than in seeds, respectively. Expression of *KKAMYB 81* was rarely detected in seeds (0.07% actin), whereas reached about 31 % of actin in leaves, indication of leaf-specific expression. Comparison

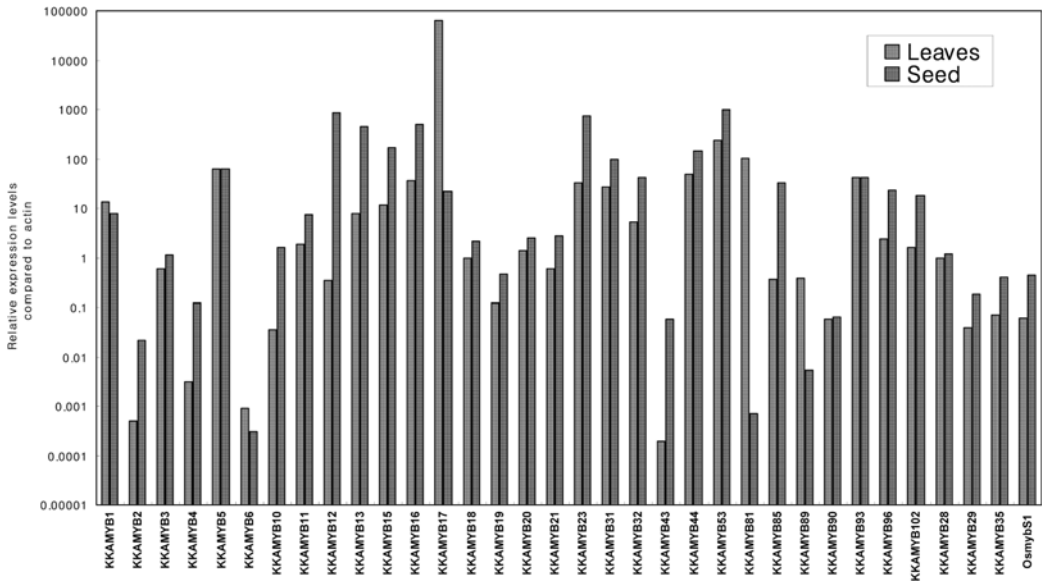

Fig. 3. Relative expression profiles of 34 rice MYB gene in seeds and leaves.

of expressions in seeds and leave showed 12 MYB genes were expressed much higher in seeds than in leaves, whereas 4 genes were expressed higher in leaves than in seeds (Table 3).

BLAST analysis of genes showing higher expression in either of the tissues revealed high sequence similarity with other MYB genes onlyat the N-terminal region, where MYB domain is located. *KKAMYB10* showed 45 times higher expression in seeds than in leavse. The gene product of KKAMYB10 binds to the myb-responsive elements in the dihydroflavonol reductase (*Dfr*) and the anthocyaninidin synthase (*Ans*) of rice.[16] Because the rice cultivar used in this experiment accumulates high anthocyanin in the seeds, MYB gene that activates genes in the anthocyanin pathway is likely to be expressed at high level. *KKAMYB16* was expressed about 14 times higher in seeds than in leaves, which agreed with the previous study by RT-PCR.[17] In addition, the expression of *Osmyb4* that corresponds to *KKAMYB12* was rarely detected in leave by RT-PCR while its expression in seeds was higher,[17] which also agrees with our real- time PCR result. Expression levels of other Osmybs studied by RT-PCR were compared with the results of real-time PCR. RT-PCR results showed that *Osmyb2*, and *5* were expressed almost at the same levels in seeds and leaves, whereas real time PCR results revealed that *Osmyb2* (*KKAMYB 13*) showed higher expression in seed and *Osmyb5* (*KKAMYB 28*) in leaves. The discrepancy could be attributed to several reasons; the first is that RT-PCR result was not quantified. RT-PCR was set to a fixed a number of cycles which sometimes resulted in the saturation of amplification. Thus, even though there was a difference in the expression level, they could not be detected. The second is that different RNA sampling times were different. Some genes are expressed at a certain period of development. Therefore, different sampling time may result in different expression level. Third, expression of genes sometimes depends on cultivars. For example, a certain rice cultivar produced more anthocyanin than others and thus would show different gene expression patterns in anthocyanin biosynthesis including transcription factors.

Table 3. Relative expression levels of thirty-five rice R2R3 MYB genes

Gene	Seeds/Leaves	Gene	Seeds/Leaves
KKAMYB1	0.5793	KKAMYB23	23.0442
KKAMYB2	42.6000	KKAMYB31	3.5970
KKAMYB3	1.9033	KKAMYB32	7.9305
KKAMYB4	37.8125	KKAMYB43	285.0000
KKAMYB5	0.9876	KKAMYB44	3.0957
KKAMYB6	0.3333	KKAMYB53	4.1951
KKAMYB10	45.0276	KKAMYB81	0.0000
KKAMYB11	4.0070	KKAMYB85	89.5522
KKAMYB12	2412.5072	KKAMYB89	0.0136
KKAMYB13	57.7742	KKAMYB90	1.0774
KKAMYB15	14.7080	KKAMYB93	0.9713
KKAMYB16	13.8167	KKAMYB96	9.5293
KKAMYB17	0.0003	KKAMYB102	11.2798
KKAMYB18	2.1888	KKAMYB28	0.0000
KKAMYB19	3.7521	KKAMYB29	4.5000
KKAMYB20	1.7582	KKAMYB35	5.8571
KKAMYB21	4.5349	OsmybS1	7.5000

In rice, only a few MYB genes have been functionally analyzed with mutants. Mutation in MYB (OsGAMYB), a regulator of gibberellin-dependent α-amylase expression, impaired α-amylase expression in aleurone and flower development.[18] OsGAMYB is a typical R2R3 MYB gene. And, mutation in anther indehiscence1 (aid1), which encodes a single MYB domain protein, showed defects in anther development in rice.[19] Because there are various available mutant lines which were made by either transposon-tagging system or T-DNA insertion system, the functional analyses of

individual MYB genes would be feasible.

Acknowledgments

This work was supported by a grant of Biogreen 21 Program, Rural Development Administration, Republic of Korea.

References

1. Stracke, R., Werber, M. and Weisshaar, B. (2001) The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447-456.
2. Klempnauer, K, H., Gonda, T. J. and Bishop, J. M. (1982) Nucleotide sequence of the retroviral leukemia gene v-*myb* and its cellular progenitor c-*myb*: the architecture of a transduced oncogene. *Cell*, **31**, 453-463.
3. Martin, C. and Paz-Ares, J. (1997) MYB transcription factors in plants. *Trends Genet.* **13**, 67-73.
4. Rabinowicz, P. D., Braun, E. L., Wolfe, A. D., Bowen, B. and Grotewold, E. (1999) Maize R2R3 Myb genes: sequence analysis reveals amplification in the higher plants. *Genetics* **153**, 427-444.
5. Urao, T., Yamaguchi-Shinozaki, K,, Urao, S. and Shinozaki, K. (1993) An *Arabidopsis* myb homolog is induced by dehydration stress and its gene product binds to the conserved MYB crcogintion sequence. *Plant Cell* **5**, 1592-1539.
6. Kranz, H. D., Denejamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R. C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., *et al.* (1998) Towards functional characterization of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant J.* **16**, 263-276.
7. Braun, E. L. and Grotewold, E, (1999) Newly discovered plant c-myb-like genes revrite the evolution of the plant myb gene family. *Plant Physiol.* **121**, 21-24.
8. Jin, H. and Martin, C. (1999) Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.* **41**, 577-585
9. Doebley, J. and Lukens, L. (1998) Transcriptional regulators and the evolution of plant from. *Plant Cell* **10**, 1075-1082.
10. Stern, D. (1998) A role of *ultrabithorax* in morphological differences between *Drosophila* species. *Nature* **396**, 463-466.
11. Cubas, P., Vincent, C. and Coen, E. (1999) An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401**, 157-161.
12. Arthur, W, (2002) The emerging conceptual framework of evolutionary developmental biology. *Nature* **415**, 757-764.
13. Jiang, C., Cu, X. and Perteson, T. (2004) Identification of conserved structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sative* L. ssp. Indica. *Genome Biology* **5**, R46.1-R46.11.
14. Heine, G. F., Hernandez, J. M. and Grotewold E. (2004) Two cysteines in plant R2R3MYB domaind participate in REDOX-dependent DNA binding. *J. Biol. Chem.* **279**, 37878-37885.
15. Romero, I., Fuertes, A., Benito, M. J., Malpica, J. M., Leyva, A. and Paz-Ares, J. (1998) More than 80 R2R3-MYB regulatory gnnes in the genome of *Arabidopsis thaliana*. *Plant J.* **14**, 273-284.
16. Ithal, N. and Reddy, A. R. Rice flavonoid pathway genes, *OsDfr* and *OsAns*, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, *OsC1-MYB*. *Plant Sci.* **166**, 1505-1513.
17. Suzuki, A., Suzuki, T., Tanabe, F., Toki, S., Washida, H., Wu, C.-Y. and Takaiwa, F. (1997) Cloning and expression of five myb-related genes from rice seed. *Gene* **198**, 393-398.
18. Kaneko, M., Inukai, Y., Ueguchi-Tanaka, M., Itoh, H., Izawa, T., Kobayashi, Y., Hattori, T., Miyao, A., Hirochika, H., Ashikari, M. and Matsuoka, M. (2004) Loss-of-function mutations of the rice *GAMYB* gene impair á-amylase expression in aleurone and flower development. *Plant Cell* **16**, 33-44.
19. Zhu, Q.-H., Ramm K., Shivakkumar R., Dennis E. S. and Upadhyaya N. M. (2004) The anther indehiscence1 gene encoding a single MYB domain protein in involved in anther development in rice. *Plant Physiol*. **135**, 1514-1525.