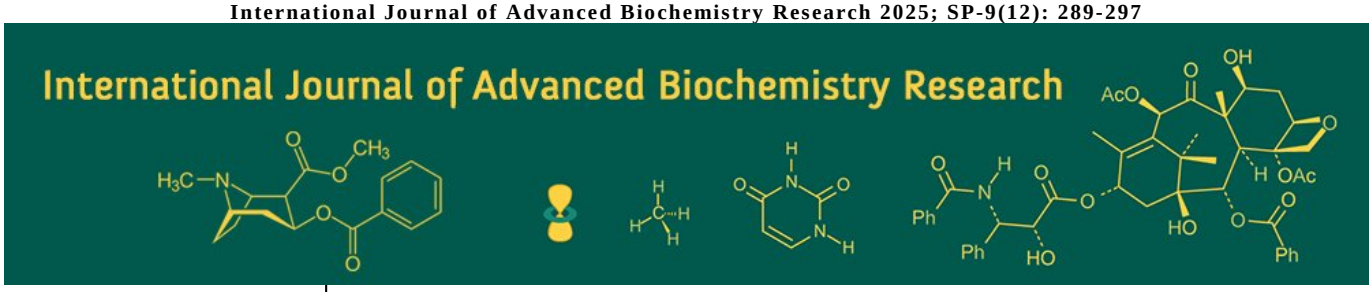


ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2025): 5.29
IJABR 2025; SP-9(12): 289-297
www.biochemjournal.com
Received: 15-10-2025
Accepted: 19-11-2025

Dhananjay Kumar
Faculty of Agricultural Science and Technology, Assam down town University, Guwahati, Assam, India

Hrishikesh Ojah
Department of Plant Breeding and Genetics, Sarat Chandra Singha College of Agriculture (SCSCA), Dhubri, Assam, India

Khanin Pathak
Department of Agricultural Biochemistry, Sarat Chandra Sinha College of Agriculture (SCSCA), Dhubri, Assam, India

Subhankar Saha
Department of Horticulture, Sarat Chandra Sinha College of Agriculture (SCSCA), Dhubri, Assam, India

Prasanta Kumar Das
(1) Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam, India
(2) Department of Agricultural Biotechnology, Sarat Chandra Sinha College of Agriculture (SCSCA), Dhubri, Assam, India

Corresponding Author:
Prasanta Kumar Das
(1) Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat, Assam, India
(2) Department of Biotechnology, Sarat Chandra Sinha College of Agriculture (SCSCA), Dhubri, Assam, India

In silico network analysis of structural and regulatory genes governing anthocyanin accumulation in black rice pericarp

Dhananjay Kumar, Hrishikesh Ojah, Khanin Pathak, Subhankar Saha and Prasanta Kumar Das

DOI: https://www.doi.org/10.33545/26174693.2025.v9.i12Sd.6518

Abstract
The *Chakhao poreitin*, a variety of black rice exclusively found in the northeastern state of Manipur in India and its black colour is imparted by anthocyanin, a primary pigment that is responsible for the presence of black, red and blue colouration in various grains, flowers, fruits and vegetables. To identify the genes responsible for anthocyanin enrichment in *Chakhao poreitin*, an *in-silico* transcriptome study between white rice and black rice was performed to identify the differentially expressed genes in black rice. The genes were then subjected to COG and GO analysis, which resulted in a total of 135 genes related to anthocyanin production in some way or the other. Pathway enrichment and mapping to the existing flavonoid pathway to identify the underlying genes resulted in 24 biosynthetic genes being present in the pathway. These genes were mapped to the chromosomes of the rice genome to identify their genetic map positions. Phylogenetic reconstruction of these genes together with their homologous genes in *A. thaliana* was carried out using the maximum-likelihood method. A protein-protein interaction network between the expected regulatory genes (belonging to either MYB, WD40 or bHLH domains) and the biosynthetic genes was constructed and also mapped to a repository for regulatory element information for plants to identify the genes that regulate the biosynthetic genes. From the results, it can be predicted that the early biosynthetic genes (EBGs) are mainly regulated by R2R3-MYB genes, namely, MYB55 and MYBS3 whereas the late biosynthetic genes (LBGs) are regulated by the WD40 gene, TTG1 and the helix-loop-helix gene, Rc. Thus, we have highlighted the possible interaction of differentially expressed regulators and structural genes encoding potential biomarkers for the pigmentation trait in Black rice.

Keywords: Black rice, anthocyanin biosynthesis, differential gene expression, flavonoid pathway, regulatory genes, biomarkers

Introduction
Oryza sativa L., seeds of the Poaceae family, commonly known as Asian rice or simply rice, could be considered as one of the most important grains since it is consumed by half of the world population [1] and especially in Southeast Asia and in the Indian subcontinent. Black rice is a type of the rice species *Oryza sativa* L., which is glutinous, packed with high level of nutrients and mainly cultivated in Asia. Black rice is also known as purple rice, forbidden rice, heaven rice, imperial rice, king's rice and prized rice. Many people assume this rice as a panacea of many culinary diseases because of its high nutritive value and curative effect. There are more than 200 types of black rice varieties in the world [2]. The black scented rice (Chakhao) of Manipur, a North-eastern State of India has their importance as scented and is dark purple colour which is used for the community feast as well as ceremonial purposes as a delicacy. These are one of the high rated dishes serve as desserts, flakes, bread, cakes, beverages and a special snack "Utong Chak" prepared within bamboo sticks. The black scented rice (Chakhao) of Manipur has been used by the traditional medical practitioners. They are sold in the local markets at a premium rate. The black scented rice cultivars of Manipur are poor yielders, which are found only in this state of India and little is known about it throughout the Indian region [3]. The pericarp (outer part) of kernel of this rice colour is black due to a pigment known as anthocyanin, an antioxidant [2]. Anthocyanins, a member of flavonoids are the primary pigments in grains, flowers, fruits and vegetables

showing blue, purple and red pigmentations. It is estimated that there are more than 400 naturally occurring anthocyanins [4]. The structural genes in their biosynthesis are placed into the general phenylpropanoid pathway, the early flavonoid pathway and the late anthocyanin-specific pathway [5]. It was also reported that black-colored rice is rich in anthocyanin and its main anthocyanin pigments are composed of cyanidin and peonicin [6]. The first committed genes of the flavonoid biosynthesis pathway (i.e., CHS, CHI, F3H and F3'H) have been named early biosynthetic genes (EBGs), and by extension the downstream genes of the pathway are usually named late biosynthetic genes (LBGs) [7]. It has also been reported that the genes on the Myb Domain, bHLH domain and WD40 repeats domain commonly known as the MBW complex play a pivotal role in the regulation of anthocyanin production [7]. MYB and basic helix-loop-helix (bHLH) TFs are found in all eukaryotes and are among the largest families of plant TFs [8, 9, 10, 11].TG1 is a WDR (WD repeat) protein characterized by the pleiotropic phenotypes of the corresponding mutants that are affected in flavonoid biosynthesis and the fate of various epidermal cells [12] Here, we perform an in silico study to investigate the structural genes of the anthocyanin biosynthesis via expression studies to gain an insight into its mechanism and also to determine the regulatory genes of the MBW complex for the early biosynthetic genes as well as the late biosynthetic genes that may prove to regulate the expression of these structural/biosynthetic genes. The results of this study may prove beneficial for improved and selective breeding and cultivation of Chakhao poreitin with required traits and characteristics.

Materials and Methods

RNA Isolation and cDNA library construction

Total RNA was isolated from the seeds of the black rice variety 'Chakhao poreitin' using Trizol Reagent, according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). The RNA was then treated with RNase-free DNase I (Takara, Otsu, Shiga, Japan) at 37 °C for 30 min to remove residual DNA. The RNA quality was verified using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and confirmed using RNase-free agarose gel electrophoresis. The concentration of the total RNA was further quantified using Nano-Drop 2000 (Thermo Fisher Scientific, Wilmington, Delaware USA). Equal amounts of total RNA from the rice sample were quickly frozen in liquid nitrogen for storage at-80 °C until further use.

A cDNA library of pooled RNA was obtained using a TruSeq RNA Sample preparation kit (Illumina, USA), according to the manufacturer's instructions. Poly-T oligo-attached magnetic beads (Illumina Inc., San Diego USA) were used to isolate poly-A mRNA from total RNA.

First-strand cDNA was synthesized from the fragmented mRNA using random hexamer prim-ers and reverse transcriptase (Invitrogen, USA). The single-end cDNA library was prepared according to Illumina's recommended protocol.

Illumina Sequencing, Quality control and Assembly

Reference based transcriptome was performed ona total of four rice samples which are two biological replicates each from both black rice as well as white rice [13]. From a total of 83.28 million Illumina HiSeq (150x2) reads 77.25 million high quality adapter free reads were utilized in analysis. An average of ~91.43% of processed reads aligned to the reference genome. Average of 31650 transcripts are expressed across all samples. The raw data generated was checked for the quality using FastQC 1 and pre processed, which includes removing the adapter sequences and removing the low quality bases (<q30). Pre-processing of the data is done with Cutadapt.

HISAT, a fast and sensitive spliced alignment program was used to align and assemble the high quality data to the reference genome with the default parameters. Reads were classified into aligned reads (which align to the reference genome) and unaligned reads.

Differential Gene expression

The protocol by Trapnell *et al.* [14] was followed for the identification of differential genes using FPKM values and log2fold changes as a measure of expression. Cufflinks was used to estimate and calculate transcript abundance. It results in normalized read count in the form of FPKM values. Cuffdiff was used to calculate the differentially expressed transcripts and categorize them into Up, Down and Neutrally regulated based on the log2fold change values.

Functional Annotation and Pathway Analysis

The software Blast2Go [15] was used for the functional annotation of the DEGs according to the three Gene Ontology terms i.e, Biological Process, Molecular Function and Cellular Component using the blastx suite using an E-value cutoff of 1E-0.5. Furthermore, the COG analysis was also used which helped us identify the genes associated with anthocyanin which was carried out using the PLAZA 4.0 database server [16]. The pathway analysis was performed using the KEGG mapper [17] with an aim to reveal the biosynthetic genes of the flavonoid biosynthetic pathway which we intend to further narrow down to exculsively the genes belonging to the anthocyanin biosynthetic pathway.

Localization and Phylogenetic analysis of anthocyanin related biosynthetic genes

Our next goal to localize the regions to which these genes belong led us to mapping their chromosomal locations which was achieved using the FSTVAL software [18]. To identify the phylogenetic relationships among the targeted genes as well as with their well classified homologs in *Arabidopsis thaliana*, a maximum likelihood phylogenetic analysis using MEGA7 software was performed [19] to reveal the evolutionary relationship among the targeted genes as well as with their homologs in Arabidopsis thaliana.

Regulatory network analysis

Next we wanted to idnetify the relation between transcription factors and the biosynthetic genes so as to gain an insight on how these biosynthetic genes are regulated. PlantRegMap [20] (http://plantregmap.cbi.pku.edu.cn/network.php) was used for transcription factor enrichment as well as finding interactions between various transcription factors and the biosynthetic genes which were then carried out using network reconstruction further, a protein-protein interaction network was also constructed using string database (https://string-db.org/) [21] to identify other possible interactions as well. The reconstructed networks were then visualised using Cytsoscape software (http://www.cytoscape.org) [22].

Results

In order to identify the genes responsible for anthocyanin biosynthesis in black rice, rna-seq analysis was performed on

the black and white rice samples expressing different levels of anthocyanins on the Illumina HiSeq (150X2) paired-end sequencing platform. The RNAseq analysis identified a total of 5,184 differentially expressed genes, which were either up-regulated or down-regulated. All these genes were functionally annotated using Blast2go software according to their Gene Ontology (Biological Process, BP), Molecular Function (MF) and Cellular Component (CC)) and Pathways by performing blastn function on all these differentially expressed genes.

COG and GO analyses

To identify which genes are responsible for anthocyanin pigmentation in black rice, we performed a Cluster of Orthologous Group (COG) analysis and Gene Ontology (GO) annotation in which the orthologs from different monocots were identified according to the GO terms related to all anthocyanin processes, including metabolic, biosynthetic and regulation of anthocyanin accumulation. After performing the COG and GO analyses, we were successfully able to identify a total of 135 genes related to anthocyanin of which 88 were found to be upregulated and 55 were downregulated. This shows that the number of upregulated genes related to anthocyanin is larger than that of the down-regulated genes.

Pathway enrichment and KEGG pathway mapping

The 135 genes were subjected to pathway enrichment and mapped to the flavonoid-biosynthetic pathway map 00941 of Kyoto Encyclopedia of Genes and Genomes (KEGG) of which 24 genes were successfully mapped onto the pathway (Fig.1). The genes in the catabolic pathway were found to be downregulated and those in the anabolic pathway were seen to be upregulated. It can also be noticed from the RNASeq data that among the genes that were mapped onto the pathway, the transcript expression of Chalcone Synthase (CHS), Anthocyanidin Synthase (ANS) and Flavone 3-hydroxylase 1(F3H1) was significantly upregulated, whereas expression of Anthocyanidin Reductase (ANR) and Quinate 3-hydroxylase(C3H1) was significantly downregulated. This shows that the genes ANS, CHS and F3H1 play a significant role in the activation of the pathway. Genes that are upregulated are depicted in red, those that are downregulated are shown in blue, and the uncoloured boxes indicate that none of our genes were mapped there.

Localization of anthocyanin related biosynthetic genes

To visualize the exact location of our genes on the rice chromosome the sequences of the 24 biosynthetic genes were mapped to one of the three sequence types i.e., exon, intron or intergenic represented in different colours (Fig.2.1). The results showed that 6 genes (26%) were found in the exonic region, 7 genes (30%) in the intronic region and 5(22%) genes in both the 5' upstream region as well as the 3' downstream region (Fig.2.II). The table below shows the mapping matrix of the 24 genes of which Chalcone Synthase (CHS), Chalcone Isomerase (CHI), C3H and Flavone Synthase (FLS) are the early flavonoid biosynthesis genes whereas, Anthocyanidin Synthase (ANS), Leucoanthocyanidin reductase (LAR), Anthocyanidin Reductase (ANR) and Dihydroflavonol 4-reductase (DFR) are the late anthocyanin specific biosynthetic genes. In contrast, the rest of the genes are associated with flavonoid biosynthesis [23, 24].

Phylogenetic analysis of the biosynthetic genes

To identify the phylogenetic relationships among the 24 genes, a maximum likelihood phylogenetic analysis was performed using MEGA7 software [19]. The resulting tree illustrates that these genes can be divided broadly into two subgroups. Group I primarily contains Cytochrome P450 enzymes (CYPs) and a few other enzymes functioning earlier in the flavonoid pathway (Fig3.I). Dominant Members are: CYP75B4, CYP75B3, CYP75A11, belonging to the flavonoid 3′/5′ hydroxylase group involved in anthocyanin and flavonoid structural modifications; C4H (Cinnamate-4-hydroxylase), A core enzyme early in the phenylpropanoid pathway and; C3H, CFI, CoA1, CHI (chalcone isomerase), LAR (leucoanthocyanidin reductase). These genes are associated with early to mid-stage flavonoid biosynthesis Group I represents upstream or branching enzymes involved in the modification of flavonoid backbones**.** The clustering suggests that these enzymes share evolutionary ancestry**,** likely due to related functions ns in hydroxylation, ring modifications, and early-pathway reactions.Group II contains genes viz., CHS, F3H, DFR, ANR, involved later in the flavonoid and proanthocyanidin (PA) biosynthesis pathway. Group II represents downstream flavonoid/anthocyanin/PA-related enzymes. Their clustering suggests functional similarity and shared evolutionary history among enzymes performing reduction, oxidation, and ring-closure steps.The separation indicates functional divergence**,** likely driven by: substrate specialization, pathway branching points and plant species diversification.

As part of the study, we were also interested in finding the homologous genes in *Arabidopsis thaliana* as a comparison with well-characterised homologs may reveal functional insights (Fig3.II). Thus, the 24 genes were subjected to a homology search of which 12 homologs were identified in *A.thaliana*. The phylogenetic tree thus formed can be divided into two main sub-clusters and each sub-cluster is further divided into two other subclusters. Clusters I(red) and II (green) consists of anthocyanin related genes while, cluster III (blue) and IV (yellow) consist of flavonoid genes. It is also observed that the clusters II and III consists of all upregulated genes. Lastly, clusters I and IV consists of mostly the downregulated genes. Arabidopsis and rice genes cluster together within each functional group, showing shared ancestry and conserved biochemical function. CYP gene family is the most phylogenetically distinct and they form a large, separate cluster because of broad functions, conserved P450 structural motifs and older evolutionary origin.

Regulatory network analysis

It has been known that the anthocyanin biosynthesis is regulated by the genes lying on three major domains i.e., MYBR2R3, WD40 repeats and bHLH domains collectively known as the MBW complex. Previous studies have proven that this complex plays a major regulatory process for the anthocyanin biosynthesis [25]. To elucidate this case, we identified all differentially expressed genes from our study that belongs to either one of these three domains. We then made a combined list (Refer to Table.2 for the interacting node's MSU IDs with their corresponding gene names) containing both the biosynthetic genes and the genes belonging to the MBW complex. This combined list of genes was then mapped to PlantRegNetwork [20] (http://plantregmap.cbi.pku.edu.cn/network.php), a database that shows regulatory information on several genes for plants

via network reconstruction. From this database we were able to identify that the early biosynthetic genes of the anthocyanin pathway such as CHS, F3'H, F3H1, FLS are mainly regulated by R2R3-Myb domain proteins namely MYBS3 and MYB55 (Fig4.I). This network suggests that these genes work together as part of a shared biological pathway**.** Hub genes may be potential targets for genetic modification **or** markers for functional studies**.** Directional pathways imply regulatory hierarchies, where some genes act as upstream regulators.

A protein-protein interaction was also constructed using the string database (https://string-db.org/) [21] in order to understand all the possible interactions between both the biosynthetic as well as regulatory genes that play a role in the anthocyanin pathway (Fig. 4. II). As expected, we were able to identify the interaction between TTG1, a WD40 protein interacting with Rc, a bHLH protein which in turn regulates the late anthocyanin biosynthetic genes, ANS, ANR, DFR, LAR and UGT. This interaction was visualized by using the Cytoscape software (http://www.cytoscape.org/) [22].

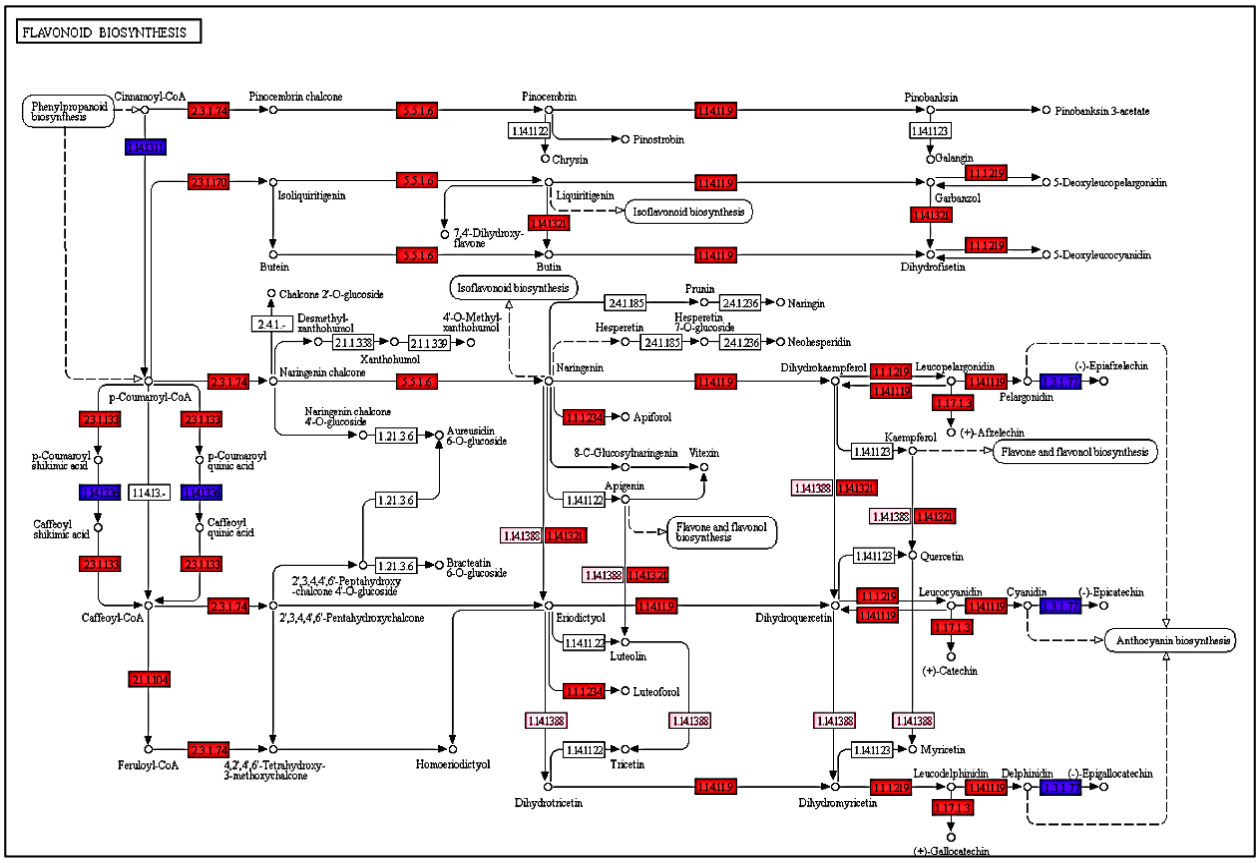
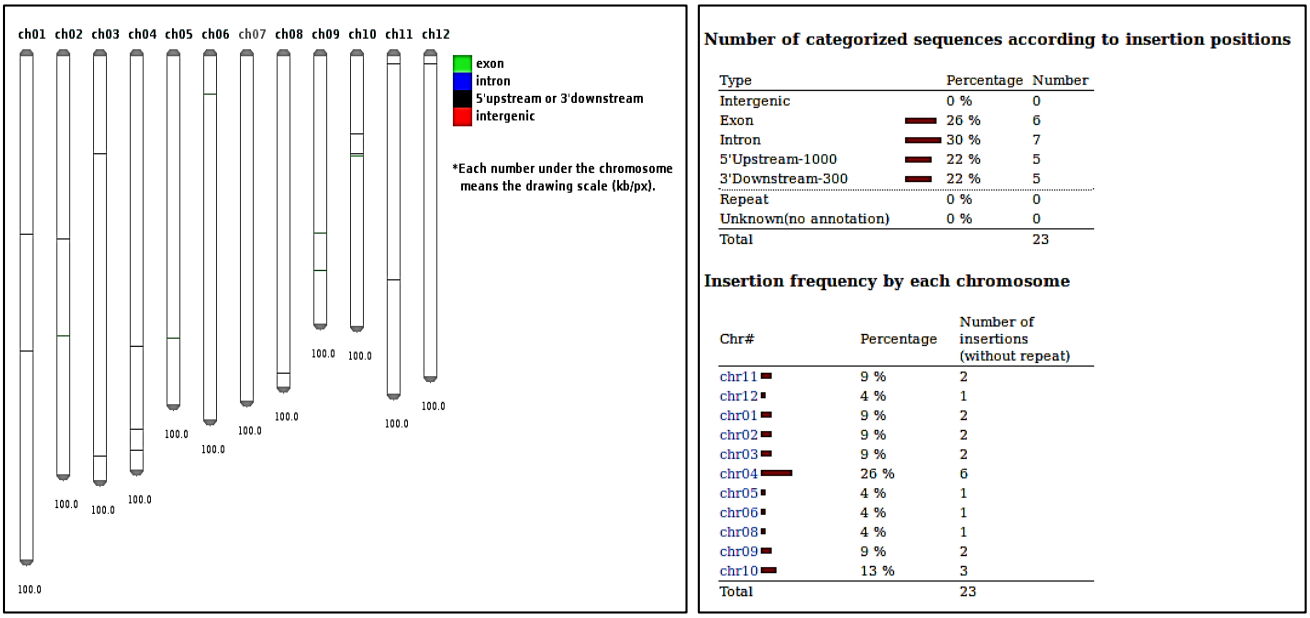


Fig 1: Expression of flavonoid-biosynthetic pathway genes in black rice. Differentially expressed genes from the data were mapped to the flavonoid-biosynthetic pathway (KEGG, map00941). Upregulated genes are shown in red. Downregulated genes are shown in blue. Neutrally regulated and genes not mapped are shown in pink and white respectively.



Number of categorized sequences according to insertion positions

Type	Percentage	Number
Intergenic	0 %	0
Exon	26 %	6
Intron	30 %	7
5'Upstream-1000	22 %	5
3'Downstream-300	22 %	5
Repeat	0 %	0
Unknown(no annotation)	0 %	0
Total		23

Insertion frequency by each chromosome

Chr#	Percentage	Number of insertions (without repeat)
chr11	9 %	2
chr12	4 %	1
chr01	9 %	2
chr02	9 %	2
chr03	9 %	2
chr04	26 %	6
chr05	4 %	1
chr06	4 %	1
chr08	4 %	1
chr09	9 %	2
chr10	13 %	3
Total		23

Fig 2: I Chromosomal locations of the biosynthetic genes. II: Number of categorised sequences according to insertion.

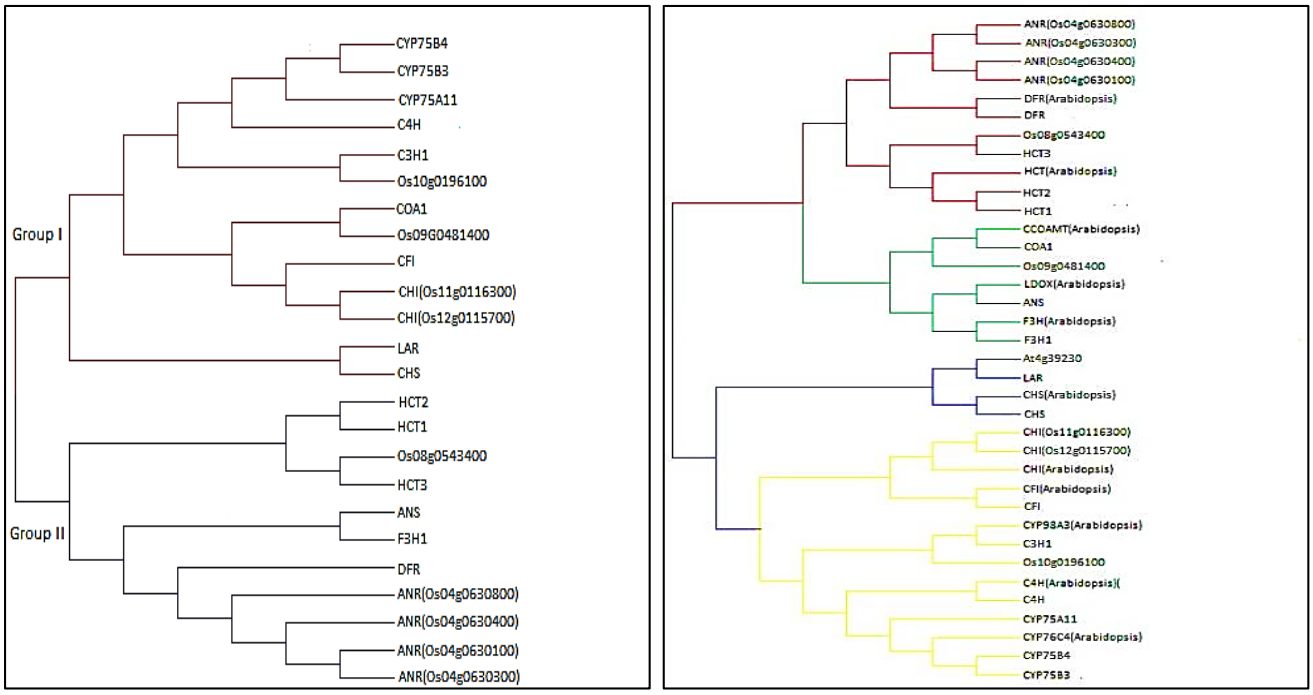


Fig 3 (I): Group I primarily contains phylogenetic clustering 24 genes encloding Cytochrome P450 enzymes (CYPs) and a few other enzymes functioning earlier and downstream flavonoid/PA regulated enzymes in the flavonoid pathway. (II): *Oryza sativa* Phylogenetic relationship among the 24 flavanoid biosynthetic genes along with their homologs in A. Thaliana.

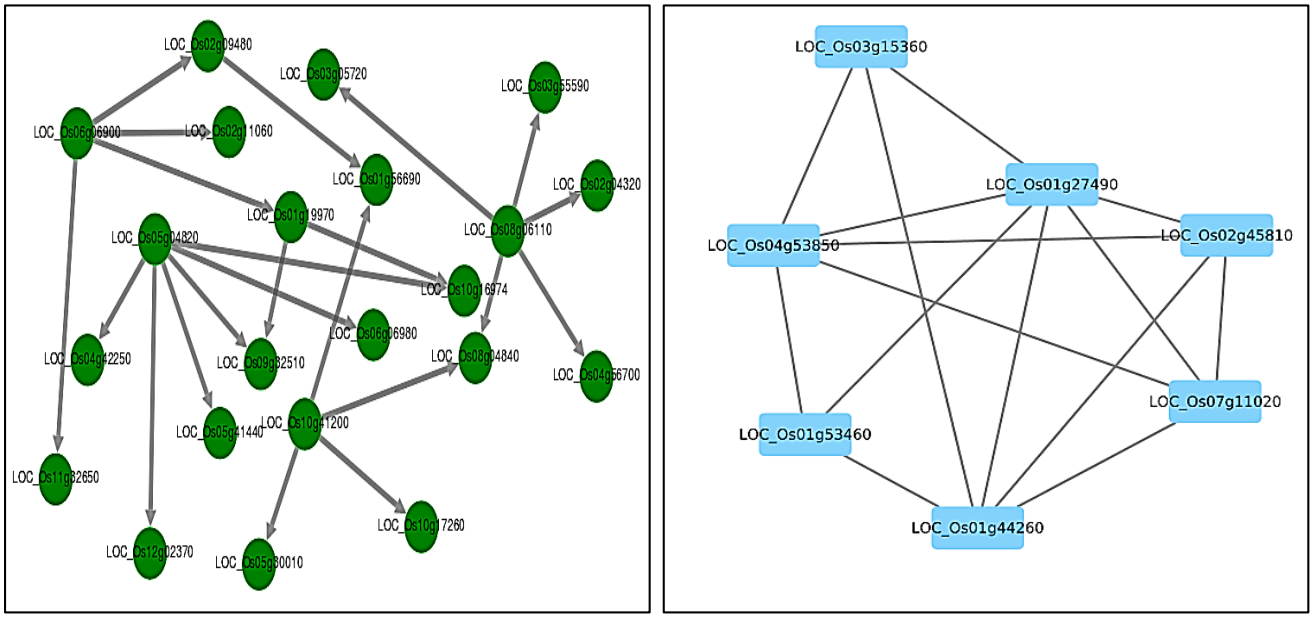


Fig 4: (I) Regulatory network between biosynthetic genes and genes of the MBW complex.(II) Interaction between TTG1, a WD40 protein interacting with Rc, a bHLH protein

Table 1: Mapping matrix of anthocyanin biosynthetic genes.

Query	Chromo-some	matching %	*	query		chromosome		e-value	score	type, gene id, description
				start	end	start	end			
196100	chr10	100.00	1024, 0, 0	1	1024	6715605	6716628	0.0	2030.0	**5'Upstream-1000** **Os10t0196100-00 upstream 0.001kb** Similar to Cytochrome P450 98A1.
481400	chr09	100.00	773, 0, 0	316	1088	18475347	18476119	0.0	1467.0	**Exon** **Os09t0481400-01** Similar to Caffeoyl-CoA O-methyltransferase (EC 2.1.1.104) (Trans-caffeoyl-CoA 3-O-methyltransferase) (CCoAMT) (CCoAOMT).
543400	chr08	100.00	2054, 0, 0	588	2641	27202857	27204910	0.0	3929.0	**Intron** **Os08t0543400-02** Transferase family protein.
ANR(630100)	chr04	100.00	1551, 0, 0	1082	2632	32050143	32051693	0.0	2783.0	**Intron** **Os04t0630100-01** NAD(P)-binding domain containing protein.
ANR(630300)	chr04	100.00	1644, 0, 0	1188	2831	32060746	32062389	0.0	3176.0	**Intron** **Os04t0630300-01** NAD(P)-binding domain containing protein.
ANR(630400)	chr04	100.00	1976, 0, 0	1	1976	32063755	32065730	0.0	3763.0	**5'Upstream-1000** **Os04t0630400-01 upstream 0.001kb** NAD(P)-binding domain containing protein.
ANR(630800)	chr04	100.00	2661, 0, 0	1	2661	32085611	32088271	0.0	4924.0	**5'Upstream-1000** **Os04t0630800-01 upstream 0.001kb** Similar to Anthocyanidin reductase.
ANS	chr01	100.00	1590, 0, 0	1	1590	15346351	15347940	0.0	3104.0	**3'Downstream-300** **Os01t0372500-00 downstream 0.001kb** Similar to Leucoanthocyanidin dioxygenase (EC 1.14.11.19) (LDOX) (Leucocyanidin oxygenase) (Leucoanthocyanidin hydroxylase).
C3H1	chr05	100.00	4044, 0, 0	12	4055	24271100	24275143	0.0	7737.0	**Exon** **Os05t0494000-01** Similar to Cytochrome P450 98A1 (EC 1.14.-.-).
C4H	chr02	100.00	1525, 0, 0	1	1525	15717044	15718568	0.0	3023.0	**5'Upstream-1000** **Os02t0466900-00 upstream 0.001kb** Hypothetical protein.
CFI	chr03	100.00	1131, 0, 0	1	1131	34394507	34395637	0.0	2242.0	**3'Downstream-300** **Os03t0819600-01 downstream 0.001kb** Chalcone isomerase (EC 5.5.1.6).
CHI(115700)	chr12	100.00	2138, 0, 0	1	2138	766440	768577	0.0	4197.0	**5'Upstream-1000** **Os12t0115801-00 upstream 0.359kb** Non-protein coding transcript.
CHI(116300)	chr11	100.00	3242, 0, 0	1	3242	729513	732754	0.0	6320.0	**3'Downstream-300** **Os11t0116300-01 downstream 0.001kb** Chalcone isomerase domain containing protein.
CHS	chr11	100.00	1358, 0, 0	1571	2928	19279166	19280523	0.0	2288.0	**Intron** **Os11t0530600-01** Similar to Chalcone synthase C2 (EC 2.3.1.74) (Naringenin-chalcone synthase C2).
COA1	chr06	100.00	1187, 0, 0	153	1339	3314265	3315451	0.0	2294.0	**Exon** **Os06t0165800-01** Similar to Caffeoyl-CoA 3-O-methyltransferase (Fragment).
CYP75A11	chr10	100.00	4806, 0, 0	4129	8934	8498375	8503180	0.0	9206.0	**Intron** **Os10t0317900-01** Chrysoeriol 5'-Hydroxylase, Flavonoid B-ring hydroxylase, Tricin biosynthesis
CYP75B3	chr10	100.00	1791, 0, 0	166	1956	8679475	8681265	0.0	3550.0	**Exon** **Os10t0320100-01** Similar to Flavonoid 3'-monooxygenase (EC 1.14.13.21) (Flavonoid 3'-hydroxylase) (Cytochrome P450 75B2).
DFR	chr01	100.00	1088, 0, 0	878	1965	25383590	25384677	0.0	2074.0	**Intron** **Os01t0633500-00** Similar to Dihydroflavonol reductase.
F3H1	chr04	100.00	2797, 0, 0	1	2797	33809615	33812411	0.0	5206.0	**3'Downstream-300** **Os04t0662600-00 downstream 0.001kb** Similar to Naringenin,2-oxoglutarate 3-dioxygenase.
HCT1	chr04	100.00	5367, 0, 0	255	5621	24994163	24999529	0.0	10270.0	**Intron** **Os04t0500700-01** Similar to Hydroxyanthranilate hydroxycinnamoyltransferase 3.
HCT2	chr02	100.00	2496, 0, 0	438	2933	24081140	24083635	0.0	4573.0	**Exon** **Os02t0611800-01** Similar to Hydroxyanthranilate hydroxycinnamoyltransferase 3.
HCT3	chr09	100.00	2569, 0, 0	1072	3640	15253081	15255649	0.0	4950.0	**Exon** **Os09t0422000-01** Transferase family protein.
LAR	chr03	100.00	1488, 0, 0	1	1488	8406808	8408295	0.0	2950.0	**3'Downstream-300** **Os03t0259400-00 downstream 0.001kb** Similar to Leucoanthocyanidin reductase (EC 1.17.1.3) (Leucocyanidin reductase).

Table 2: The list of MSU IDs and their corresponding gene names used in the regulatory network reconstruction.

MSU ID	Gene Name
LOC_Os01g19970	Os01g0305900
LOC_Os01g56690	Os01g0773800
LOC_Os02g04320	WD40-36
LOC_Os02g09480	Os02g0187700
LOC_Os02g11060	WD40-38
LOC_Os03g05720	WD40-63
LOC_Os03g55590	Os03g0764600
LOC_Os04g42250	HCT1
LOC_Os04g56700	F3H1
LOC_Os05g04820	MYB55
LOC_Os05g30010	Os05g0363500
LOC_Os05g41440	C3H1
LOC_Os06g06900	bHLH108
LOC_Os06g06980	COA1
LOC_Os08g04840	Os08g144000
LOC_Os08g06110	CCA1
LOC_Os09g32510	bHLH092
LOC_Os10g16974	CYP75B4
LOC_Os10g17260	F3'H
LOC_Os10g41200	MYBS3
LOC_Os11g32650	CHS
LOC_Os12g02370	CHI
LOC_Os03g15360	LAR
LOC_Os04g53850	ANR
LOC_Os01g53460	UGT
LOC_Os01g44260	DFR
LOC_Os01g27490	ANS
LOC_Os07g11020	RC
LOC_Os01g27490	TTG1

Discussion

KEGG pathway enrichment (map00941) mapped 24 of the 135 differentially expressed genes to the flavonoid biosynthesis pathway, revealing a strong transcriptional bias toward biosynthesis. Anabolic genes were predominantly upregulated, while catabolic genes were downregulated. Notably, CHS, F3H1, and ANS showed significant upregulation, indicating activation of both the entry and late steps of anthocyanin biosynthesis [26-28]. In contrast, ANR was downregulated, suggesting reduced metabolic flux toward proanthocyanidin formation and preferential channelling toward anthocyanin accumulation [29]. Downregulation of C3H1 further reflects regulation of upstream phenylpropanoid flux [30]. Overall, these results demonstrate coordinated transcriptional reprogramming of the flavonoid pathway to favour anthocyanin biosynthesis.

Phylogenetic analysis of the 24 flavonoid and anthocyanin biosynthetic genes resolved two major functional-evolutionary groups consistent with pathway architecture. Group I comprised primarily Cytochrome P450 enzymes and early to mid-pathway genes (C4H, C3H, CFI, CoA1, CHI and LAR), representing upstream and backbone-modifying steps of flavonoid biosynthesis. The tight clustering of CYP75B4, CYP75B3 and CYP75A11 highlights strong conservation of flavonoid 3′/5′-hydroxylation, a critical determinant of flavonoid structural diversity and pigment composition [31-32]. Group II contained downstream structural genes (CHS, F3H, DFR and ANR) that catalyze committed reactions toward anthocyanin and proanthocyanidin formation, reflecting their close functional coordination in controlling metabolic flux [33-34].

Comparative phylogenetic analysis with *Arabidopsis thaliana* identified 12 homologs and resolved four sub-clusters distinguishing anthocyanin-and flavonoid-related genes. The exclusive clustering of upregulated genes in two subgroups indicates modular transcriptional regulation of pigment biosynthesis [36]. The conserved co-clustering of rice and Arabidopsis genes demonstrates deep evolutionary conservation of flavonoid pathway organization across angiosperms [35-37]. Overall, these results indicate that anthocyanin accumulation in black rice is governed by evolutionarily conserved structural modules under coordinated transcriptional control.

Anthocyanin biosynthesis in rice is tightly regulated by the conserved MYB-bHLH-WD40 (MBW) transcriptional complex, which integrates developmental and environmental signals to control pigment accumulation [38-39]. In the present study, regulatory network analysis of differentially expressed genes confirmed the involvement of MYBR2R3, bHLH and WD40 domain proteins in coordinating anthocyanin biosynthesis in black rice. PlantRegNetwork-based reconstruction revealed that early biosynthetic genes, including CHS, F3′H, F3H1 and FLS, are predominantly regulated by R2R3-MYB transcription factors MYBS3 and MYB55, consistent with the known role of MYB regulators in initiating flavonoid pathway flux in rice [40-41].

Protein-protein interaction analysis further supported the integrity of the MBW complex by identifying the interaction between TTG1 (WD40) and Rc (bHLH), a key regulatory module controlling late biosynthetic genes such as ANS, ANR, DFR, LAR and UGT [42-44]. This MBW-mediated regulation of late structural genes is a defining feature of anthocyanin accumulation in pigmented rice. The identification of hub regulators within this network highlights promising candidate genes for molecular breeding and

metabolic engineering aimed at enhancing anthocyanin content in rice grain.

Conclusion

RNA Seq analysis for the identification of differentially expressed genes was used in this study as it has been proven to be one of the best techniques for transcriptomics studies. The COG analysis helped us narrow down the number of DEGs that are restricted only to anthocyanin related genes in some way or the other. The pathway analysis was performed using the KEGG mapper revealed the biosynthetic genes of the flavonoid biosynthesis pathway, which we were able to further narrow down to exclusively the genes belonging to the anthocyanin biosynthetic pathway in the further steps of our study. The results of the pathway analysis suggest that the genes ANS, CHS and F3H1 are key players in the activation of the anthocyanin biosynthetic pathway.

Our next goal to localize the regions to which these genes belong led us Mapping their chromosomal locations showed that 6 genes (26%) were found in the exonic region, 7 genes (30%) in the intronic region and 5(22%) genes in both the 5' upstream region as well as the 3' downstream region. Finally, after the identification of the biosynthetic genes, we were also interested in finding out the genes that regulate them; therefore, a regulatory network was constructed consisting of the biosynthetic genes and the genes from the MBW complex. Our results showed that early anthocyanin biosynthetic genes such as CHI, C3H1 and FLS were largely regulated by genes belonging to the MYBR2R3 domain namely MYB55 and MYBS3 whereas the late biosynthetic genes i.e., ANR, ANS, LAR and DFR were regulated as a complex between the bHLH and WD40 repeats domain.

Acknowledgments

We would like to acknowledge the DBT-NECAB Centre for providing the necessary facilities and support for this work.

References

1. Watson RR, Preedy VR, Zibadi S. Wheat and Rice in Disease Prevention and Health: Benefits, risks and mechanisms of whole grains in health promotion. Academic Press; 2014. p. 409-430.
2. Kushwaha UKS. Black Rice Research, History and Development. 1st ed. Springer International Publishing; 2016. p. 21-47.
3. Asem ID, Imotomba RK, Mazumder PB, Laishram JM. Anthocyanin content in the black scented rice (Chakhao): its impact on human health and plant defense. Symbiosis. 2015;66:47-54.
4. Kong JM, Chia LS, Goh NK, Chia TF, Brouillard R. Analysis and biological activities of anthocyanins. Phytochemistry. 2003;64:923-933.
5. Shi MZ, Xie DY. Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana. Recent Pat Biotechnol. 2014;8:47-60.
6. Abdel-Aal ESM, Young JC, Rabalski I. Anthocyanin composition in black, blue, pink, purple, and red cereal grains. J Agric Food Chem. 2006;54:4696-4704.
7. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015;20:1360-1385.
8. Dubos C, *et al.* MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15:573-581.
9. Heim MA, *et al.* The basic helix-loop-helix transcription factor family in plants. Mol Biol Evol. 2003;20:735-747.
10. Toledo-Ortiz G, *et al.* The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell. 2003;15:1749-1770.
11. Feller A, *et al.* Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J. 2011;66:94-116.
12. Li Q, *et al.* Genome-wide analysis of WD-repeat proteins in cucumber and Arabidopsis. Mol Genet Genomics. 2014;289:103-124.
13. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57-63.
14. Trapnell C, *et al.* Differential gene and transcript expression analysis of RNAseq with TopHat and Cufflinks. Nat Protoc. 2012;7:562-578.
15. Conesa A, *et al.* Blast2GO: a universal tool for functional genomics. Bioinformatics. 2005;21:3674-3676.
16. Van Bel M, *et al.* PLAZA 4.0: an integrative resource for plant genomics. Nucleic Acids Res. 2018;46:D1190-D1196.
17. Kanehisa M, *et al.* KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44:D457-D462.
18. Kim JS, *et al.* FSTVAL: a web tool to validate bulk flanking sequence tags. Plant Methods. 2012;8:24-31.
19. Oh JH, *et al.* Computational identification of anthocyanin-related genes using transcriptome data from black rice. Cell Dev Biol. 2015;4:1-10.
20. Jin J, *et al.* PlantTFDB 4.0: transcription databases in plants. Nucleic Acids Res. 2017;45:D1040-D1045.
21. Jensen LJ, *et al.* STRING 8-protein networks. Nucleic Acids Res. 2009;37:D412-D416.
22. Shannon P, *et al.* Cytoscape: integrated biomolecular interaction networks. Genome Res. 2003;13:2498-2504.
23. Fujiwara H, *et al.* Anthocyanin 5-aromatic acyltransferase from Gentiana triflora. Eur J Biochem. 1997;249:45-51.
24. Nelson DR, *et al.* Comparative genomics of rice and Arabidopsis cytochrome P450 genes. Plant Physiol. 2004;135:756-772.
25. Xu W, Dubos C, Lepiniec L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. Trends Plant Sci. 2015;20:1360-1385. (duplicate removed if used above)
26. Winkel-Shirley B. Flavonoid biosynthesis. Plant Physiol. 2001;126:485-493.
27. Holton TA, Cornish EC. Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell. 1995;7:1071-1083.
28. Petroni K, Tonelli C. Regulation of anthocyanin synthesis in reproductive organs. Plant Sci. 2011;181:219-229.
29. Xie DY, *et al.* Anthocyanidin reductase in proanthocyanidin biosynthesis. Arch Biochem Biophys. 2004;422:91-102.
30. Schoch G, *et al.* Cinnamate-3-hydroxylase. J Biol Chem. 2001;276:36566-36574.
31. Falginella L, *et al.* Flavonoid 3’,5’-hydroxylases in grapevine. Plant Sci. 2010;178:456-464.
32. Seitz C, *et al.* Evolution of flavonoid 3’,5’-hydroxylase. Plant J. 2006;49:949-963.

33. Holton TA, Cornish EC. Genetics and biochemistry of anthocyanin biosynthesis. Plant Cell. 1995;7:1071-1083. (duplicate)
34. Lepiniec L, *et al.* Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 2006;57:405-430.
35. Xu W, Dubos C, Lepiniec L. MYB-bHLH-WD40 complexes. Trends Plant Sci. 2015;20:552-561.
36. Nelson D, Werck-Reichhart D. A P450-centric view of plant evolution. Plant J. 2011;66:194-211.
37. Deng Y, *et al.* Flavonoid biosynthesis genes in plants. BMC Plant Biol. 2018;18:114-130.
38. Sweeney MT, *et al.* Rc encodes bHLH conditioning red pericarp in rice. Plant Cell. 2006;18:283-294.
39. Furukawa T, *et al.* Rc and Rd genes in rice pericarp. Plant Cell. 2007;19:3579-3594.
40. Sun X, *et al.* C-S-A gene system regulates hull pigmentation in rice. Plant Physiol. 2018;178:924-936.
41. Peng Y, *et al.* Overexpression of R2R3-MYB increases flavonoids. BMC Plant Biol. 2016;16:39-50.
42. Ramsay NA, Glover BJ. MYB-bHLH-WD40 and evolution of cellular diversity. Annu Rev Plant Biol. 2005;56:509-531.
43. Wang E, *et al.* Control of rice grain filling and yield by domestication gene. Plant Cell. 2011;23:3283-3299.
44. Ma D, Constabel CP, Song J. Comparative analysis of flavonoid biosynthesis genes. Plant Biotechnol J. 2018;16:1258-1271.