



ARTICLE

Integrative Genome-Wide Analysis of R2R3-MYB Transcription Factors in *Oryza sativa* subsp. *japonica* Reveals Candidate Regulators of Anthocyanin Biosynthesis

N. V. T. Jayaprada¹, Muhammad Hammad², S. Geekiyanage³, Muhammad Shafiq⁴, Heba I. Moamed^{5,*}, Haiam O. Elkatry⁶, Mohamed El Oirdi⁷, Ahmed Mahmoud Ismail⁸, Mohamed M. El-Mogy⁸ and Abdelrahman R. Ahmed^{6,*}

¹Board of Study in Agriculture, Faculty of Graduate Studies, University of Ruhuna, Matara, Sri Lanka

²Department of Horticulture, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan

³Department of Agricultural Biology, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya, Sri Lanka

⁴Department of Horticulture, Faculty of Agricultural Sciences, University of the Punjab, Lahore, Pakistan

⁵Biological and Geological Sciences Department, Faculty of Education, Ain Shams University, Cairo, Egypt

⁶Food and Nutrition Science Department, Agricultural Science and Food, King Faisal University, Al Ahsa, Saudi Arabia

⁷Department of Life Sciences, College of Science, King Faisal University, Al Ahsa, Saudi Arabia

⁸Department of Arid Land Agriculture, College of Agricultural and Food Sciences, King Faisal University, Al Ahsa, Saudi Arabia

*Corresponding Authors: Heba I. Moamed. Email: hebaibrahim79@gmail.com;

Abdelrahman R. Ahmed. Email: arahmed@kfu.edu.sa

Received: 29 December 2025; Accepted: 19 March 2026; Published: 28 April 2026

ABSTRACT: Pigment biofortification in rice, particularly through enhanced anthocyanin accumulation, is increasingly recognized in Asia for its potential to improve nutritional quality and antioxidant capacity while addressing regional health challenges. However, the regulatory network underlying anthocyanin biosynthesis in rice remains incompletely characterized, especially regarding MYB transcription factors that may control pigment accumulation in diverse germplasm. In this study, a genome-wide analysis of MYB transcription factors in *Oryza sativa* subsp. *japonica* identified 109 *OsMYB* genes. Phylogenetic comparison with known anthocyanin regulators from *Arabidopsis thaliana* and *Oryza rufipogon* revealed conserved clades containing potential pigment-related regulators. Integrative analyses including gene structure, conserved motif identification, promoter cis-element profiling, and miRNA target prediction were conducted to explore their regulatory potential. Expression profiling during grain development in black and mixed rice identified 32 *OsMYB* genes as potential candidates associated with anthocyanin accumulation. Promoter analysis revealed enrichment of light-responsive, stress-responsive, and developmental regulatory elements, suggesting coordinated regulation by environmental and developmental signals. Predicted interactions with several Osa-miRNAs further indicate potential post-transcriptional control of these genes. Among the candidates, *OsMYB65* showed strong differential expression in pigmented rice grains and possessed multiple regulatory elements associated with anthocyanin biosynthesis, highlighting it as a promising regulator. Overall, this integrative genome-wide analysis identifies a prioritized set of MYB transcription factors that may regulate anthocyanin biosynthesis in rice, providing valuable targets for functional validation and pigment-oriented crop improvement.

KEYWORDS: R2R3-MYB transcription factors; flavonoid biosynthesis; rice; transcriptional regulation of pigmentation

1 Introduction

Transcription factors (TFs) are regulators of gene expression, acting as molecular switches that modulate the transcriptional system in response to developmental and environmental signals [1]. Among the diverse families of TFs, the *MYB* (myeloblastosis) superfamily represents one of the largest and most functionally diverse classes in plants. Initially discovered in avian myeloblastosis virus, the *MYB* domain is defined by a conserved helix-turn-helix (HTH) DNA-binding motif, typically composed of up to three imperfect repeats (R1, R2, and R3) of ~52 amino acids each [2].

Plant *MYB* genes are classified into four groups based on repeat numbers, namely R1-*MYB*, R2R3-*MYB*, R1R2R3-*MYB* (3R-*MYB*), and deviant *MYBs*. Among these, R2R3-*MYB* TFs are the most abundant and plant-specific, regulating diverse processes such as metabolism, stress tolerance, hormone signaling, and development [3]. Importantly, R2R3-*MYB* TFs regulate phenylpropanoid and flavonoid biosynthesis, including anthocyanins, by activating structural genes such as Chalcone Synthase (*CHS*), Dihydroflavonol 4-Reductase (*DFR*), and Anthocyanidin Synthase (*ANS*) [4,5]. This regulation is mediated through the *MYB*-bHLH-WD40 (MBW) complex. Within this complex, the *MYB* component confers DNA-binding specificity by recognizing conserved cis-regulatory motifs in the promoters of target genes, thereby determining the transcriptional activation of anthocyanin biosynthetic enzymes. The formation and activity of the MBW complex are also influenced by upstream regulatory signals, as *MYB* TFs often bind promoter regions enriched with light- and stress-responsive cis-elements that mediate environmental regulation of pigment biosynthesis. Such regulatory interactions allow plants to modulate anthocyanin accumulation in response to external stimuli, including light intensity and abiotic stresses, which are known to induce flavonoid pathway genes. Moreover, anthocyanin production is frequently restricted to specific tissues or developmental stages, reflecting spatially regulated expression patterns of *MYB* TFs that determine where MBW complexes are assembled and activated. For example, functional characterization of R2R3-*MYB* regulators in *Rosa rugosa* demonstrated that *RrMYB12* and *RrMYB111* directly modulate the transcription of flavonoid biosynthetic genes, leading to coordinated accumulation of flavonols and anthocyanins in pigmented tissues [4]. Therefore, identifying candidate anthocyanin regulators requires integrating phylogenetic relationships with promoter cis-element architecture and tissue-specific expression profiles, enabling a multi-layered evaluation of *MYB* genes that are both evolutionarily related to known regulators and transcriptionally active under relevant regulatory contexts. Flavonoids and anthocyanins not only contribute to pigmentation but also provide photoprotection, pathogen defense, and antioxidative functions [4]. Given their central regulatory roles, R2R3-*MYB* TFs were selected as the primary focus of the present investigation, leading to the identification and characterization of *MYB* TFs of rice, which are potentially involved in anthocyanin biosynthesis.

In *Arabidopsis thaliana*, R2R3-*MYB* TFs such as *AtMYB75* (*PAP1*), *AtMYB90* (*PAP2*), *AtMYB113*, *AtMYB15*, and *AtMYB114* have been well-characterized as key activators of anthocyanin biosynthesis [6]. *AtMYB15* overexpression upregulated genes in the shikimate pathway, a precursor route for phenylpropanoids, possibly enhancing anthocyanin biosynthesis under wounding and stress conditions [7]. These TFs coordinate with bHLH proteins (TT8, GL3) and the WD-repeat protein TTG1 to direct hierarchical activation of biosynthetic genes. Similar regulatory modules have been identified in monocots such as *Zea mays*, with species-specific adaptations and gene duplications that reflect evolutionary divergence [8].

Rice (*Oryza sativa*), the staple food for over half the global population, contains a large number of *MYB* genes, many yet uncharacterized. Genome-wide analyses have identified a large *MYB* TF family in rice, with over 100 *MYB*-like sequences reported and approximately 70–90 belonging to the R2R3-*MYB* subgroup [9–12]. Comparative studies across plant species identified 197 *MYB* genes in rice and 198 in

Arabidopsis, classifying them into multiple subfamilies based on conserved domains and motif composition and suggesting that gene duplication events contributed to both conserved and species-specific expansion patterns [9,10]. Subsequent analyses of 124 rice R2R3-MYB genes revealed diverse expression responses to abiotic stresses, including drought, cold, and salinity, indicating functional diversification within the family [11,12]. Despite these advances in characterizing rice MYB genes, comprehensive phylogenetic and functional analyses of MYBs associated with pigment biosynthesis remain limited. Nevertheless, specific regulators such as *OsMYB3* have been shown to activate anthocyanin biosynthesis by upregulating *OsDFR* and *OsANS* in black rice pericarp, highlighting the regulatory potential of MYB TFs in pigmentation pathways [13]. The deep regulatory network of MYBs influencing pigment biosynthesis and their orthologous relationships with well-studied *Arabidopsis* MYBs need further investigation. Studying pigment production in rice and identifying novel MYB genes is vital for improving grain quality and nutritional value, especially in South Asia, where rice is a staple for over 70% of the population [14]. Pigmented rice varieties rich in anthocyanins and flavonoids offer enhanced antioxidant properties and reduced risk of chronic diseases [15].

Genome-wide analyses enable systematic identification of gene families and provide insights into their evolution, functions, and stress-related roles [9–12]. With a compact 389 Mb genome and high-quality annotation, *O. sativa* serves as a model for monocots and cereal crops. The japonica cultivar “Nipponbare” was the first rice genome to be sequenced [16], and it’s well-assembled. Japonica rice also holds considerable economic and regional importance, being widely cultivated in East Asia and temperate regions, where it contributes significantly to food security. Functionally, japonica varieties exhibit traits such as cold tolerance and unique grain quality, making them valuable for breeding and comparative genomics. Previous studies on *Oryza sativa* MYB TFs have primarily focused on genome-wide identification and basic classification of MYB family members, providing valuable catalogs of these regulators. However, many of these analyses did not systematically integrate evidence such as phylogenetic proximity to known anthocyanin-regulating MYBs, promoter cis-regulatory architecture, and expression patterns in pigmented tissues to prioritize candidates potentially involved in anthocyanin biosynthesis [17]. Consequently, the functional relevance of many MYB genes in pigment regulation remains unclear. To address this gap, the present study applies a comprehensive genome-wide analysis of MYB TFs in *Oryza sativa* subsp. *japonica*, with a particular focus on the R2R3-MYB subgroup. The analysis included structural classification, chromosomal distribution, gene duplication events, conserved motif composition, and phylogenetic relationships. Because the MYB superfamily in rice is large and functionally diverse, such a genome-wide investigation is necessary to systematically identify candidate regulators potentially associated with anthocyanin biosynthesis.

To infer potential functions, phylogenetic relationships were examined through comparative analysis with MYB TFs from *Arabidopsis thaliana*. Candidate pigment-related MYBs were further predicted by integrating multiple lines of evidence, including clustering with known anthocyanin regulators, promoter cis-element analysis, and gene expression profiling. Although several MYB regulators have previously been reported in rice, a unified analysis combining phylogenetic relationships with expression-based candidate prioritization and known anthocyanin activators has not been performed. The integrative framework used in this study therefore provides a systematic approach for identifying candidate MYB regulators involved in flavonoid and anthocyanin biosynthesis. Overall, this work identifies a set of preliminary candidate MYB genes for future functional validation and establishes a foundation for understanding transcriptional regulation of pigment biosynthesis in rice, with potential applications in crop improvement and nutritional enhancement.

2 Materials and Methods

2.1 Database Search and Sequence Retrieval

To identify *MYB* TFs, the conserved *MYB* DNA-binding domain (PF00249) was used as the primary query motif. The amino acid sequence of the *MYB* transcription factor *AtMYB15* (Gene ID: At3g23250) from *Arabidopsis thaliana* was initially retrieved from the UniProt database (<https://www.uniprot.org/>) to confirm the presence of the *MYB*-like DNA-binding motif using the Motif Finder tool available at the KEGG GenomeNet platform (<https://www.genome.jp/tools/motif/>). The identified PF00249 domain was subsequently used to construct a hidden Markov model (HMM) profile for genome-wide searches. The presence of PF00249 in well-characterized anthocyanin regulators, such as *PAP1*, was also confirmed to ensure conservation among functionally validated *MYBs*. Genome-wide identification of *MYB* proteins was performed using the HMMER v3.3 package against protein datasets obtained from Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>) for *Oryza sativa* subsp. *japonica* (cv. Nipponbare) and *Arabidopsis thaliana* and from Ensembl Plants (<https://plants.ensembl.org/>) for *Oryza rufipogon*. Searches were conducted using the PF00249 HMM profile with an E-value cutoff of $1e-5$ to identify candidate *MYB* proteins. Redundant sequences and truncated entries were removed to retain non-redundant protein sets. Candidate sequences were then validated using multiple conserved domain databases, including Pfam and the NCBI Conserved Domain Database (CDD). Sequences lacking a complete *MYB* DNA-binding domain were excluded from further analysis. Using this pipeline, a total of 109, 104, and 108 *MYB* proteins containing the PF00249 domain were identified from *Oryza sativa* subsp. *japonica*, *Arabidopsis thaliana*, and *Oryza rufipogon*, respectively. While initial retrieval included all *MYB*-domain-containing proteins, downstream analyses focused on candidate R2R3-*MYBs*, which are the predominant regulators of anthocyanin biosynthesis.

2.2 Multiple Sequence Alignment and Phylogenetic Analysis

Multiple sequence alignment of *MYB* protein amino acid sequences was performed using MUSCLE v2.1. Poorly aligned regions and positions containing excessive gaps were removed using the partial deletion option implemented in MEGA X to improve alignment quality prior to phylogenetic reconstruction. Phylogenetic analysis was conducted using the Maximum Likelihood (ML) method. The Jones–Taylor–Thornton (JTT) amino acid substitution model was used for tree construction. Rate variation among sites was assumed to follow uniform rates across all positions. *MYB* TFs contain highly conserved R2R3 DNA-binding domains; the alignment primarily consisted of homologous regions with limited sequence length variation after the removal of poorly aligned sites. Under these conditions, assuming uniform substitution rates provides a stable framework for reconstructing higher-level clade relationships within large gene families. The ML heuristic search was performed using the Nearest Neighbor Interchange (NNI) method to explore tree topology. Branch support was evaluated using bootstrap analysis with 1000 replicates, and bootstrap values $\geq 80\%$ were considered to indicate well-supported clades. A total of 321 *MYB* protein sequences containing the PF00249 domain were included in the initial phylogenetic analysis, comprising 109 sequences from *Oryza sativa* subsp. *japonica*, 104 sequences from *Arabidopsis thaliana*, and 108 sequences from *Oryza rufipogon*.

To facilitate functional inference, the phylogenetic dataset was expanded to include previously characterized rice *MYB* TFs and well-established anthocyanin-activating *MYBs* from other plant species. In total, 109 *MYB* proteins from *Oryza sativa* subsp. *japonica* were analyzed together with 20 functionally characterized rice *MYB* TFs and representative anthocyanin-related *MYBs* reported in other plant species. The inclusion of these reference proteins enabled the identification of rice *MYB* genes clustering with

known pigment regulators. Candidate anthocyanin-related MYBs were inferred based on their phylogenetic proximity to experimentally validated anthocyanin activators and supported by bootstrap values $\geq 70\%$. Because the objective of the phylogenetic analysis in this study was to identify clustering patterns with known anthocyanin regulators rather than to estimate evolutionary rates or divergence times, the uniform-rate model was retained.

2.3 Physicochemical Properties and Subcellular Localization Prediction

Information on *Oryza sativa* MYB genes was retrieved from the Phytozome (<https://phytozome-next.jgi.doe.gov/>) and ProtParam databases (<https://web.expasy.org/protparam/>). Phytozome served as the primary source for genomic features, including chromosome number, gene location, and orientation, as well as the lengths of mRNA and corresponding peptides. In contrast, the ProtParam tool was used to predict various physicochemical properties of the MYB proteins, such as molecular weight, isoelectric point (pI), instability index, and Grand Average of Hydropathicity (GRAVY) values (Table S1). Additionally, subcellular localization prediction of the MYB proteins was assessed using the Wolf PSORT server (<https://wolfsort.hgc.jp/>), which analyzes protein sequences to predict potential cellular compartments along with associated confidence scores. This analysis aimed to determine the intracellular distribution of MYB proteins, supporting functional characterization [10].

2.4 Conserved Motifs and Domain Analysis of OsMYB Genes

Domain annotation of MYB proteins from *Oryza sativa* subsp. *japonica* was performed using the NCBI Conserved Domain Database (CDD) (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) to confirm the presence of the MYB DNA-binding domain and associated conserved regions. The identified domains were further verified and visualized using TBtools. To identify conserved sequence motifs within the OsMYB protein sequences, motif discovery was performed using the MEME Suite (version 5.5.9) (<https://web.expasy.org/protparam/>). Protein sequences were analyzed in classic discovery mode with the site distribution set to Zero or One Occurrence Per Sequence (ZOOPS). The maximum number of motifs was limited to 10, and the motif width was restricted to a range of 6–50 amino acids. The significance threshold for motif detection was set to a maximum E-value of 0.05, and a first-order Markov background model generated from the input sequences was used to estimate amino acid frequencies. Motif discovery was carried out using the expectation–maximization algorithm implemented in MEME. The resulting motif sequences and motif logos were extracted to characterize conserved patterns among OsMYB proteins. Motif composition and distribution were further visualized by integrating MEME outputs with phylogenetic tree data using TBtools. Motif 1 and Motif 2 corresponded to the conserved MYB DNA-binding domain (R2/R3 repeats). The PF00249 MYB-like DNA-binding domain of OsMYB was manually scanned for the presence of a bHLH-interaction motif. The sequence of the R3 repeat was examined for the standard pattern [DE]Lx2 [RK]x3Lx6Lx3R, with conserved leucine (L) and arginine/lysine (R/K) residues identified as potential interaction sites.

2.5 Intron-Exon Structure Analysis of OsMYB Genes

The genomic DNA and coding sequences (CDS) of 109 putative OsMYB genes were obtained from the Phytozome database. These sequences were then analyzed using the Gene Structure Display Server 2.0 (GSDS; <http://gsds.cbi.pku.edu.cn/>) to investigate the exon–intron organization within each gene [10].

2.6 Cis-Regulatory Elements of OsMYB Genes

Promoter sequences of *OsMYB* genes from *Oryza sativa* subsp. *japonica* were retrieved from the Phytozome database. For each gene, a 1000 bp region upstream of the translation start site (ATG) was extracted and considered as the putative promoter region. These promoter sequences were analyzed using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html>) to identify all potential cis-regulatory elements present within the sequences. Following the initial identification of cis-acting regulatory elements, the analysis specifically focused on motifs associated with flavonoid and anthocyanin biosynthesis. Particular attention was given to regulatory elements such as MYB-binding sites (MBS), light-responsive elements including ACE, Chs-CMA1a, Chs-CMA2a, and G-box, and hormone- or stress-responsive elements such as ABRE, which are known to influence flavonoid pathway gene expression [18]. Subsequently, the distribution of these anthocyanin-related cis-elements was examined in the promoter regions of the newly identified R2R3 *OsMYB* genes to assess their potential regulatory involvement in pigment biosynthesis. Cis-element distributions and positional patterns were visualized using TBtools. This analysis was conducted as a descriptive survey of regulatory motifs, and no statistical enrichment testing against genomic background sequences was performed.

2.7 The miRNA Analysis for OsMYB Genes

Potential miRNA interactions with *OsMYB* genes from *Oryza sativa* subsp. *japonica* were investigated to explore post-transcriptional regulatory mechanisms. Known rice miRNA sequences were retrieved from the PmiREN database, and coding sequences (CDS) of the *OsMYB* genes were used as query sequences for target prediction. miRNA target prediction was performed using the psRNATarget web server with default scoring parameters. Predicted miRNA–mRNA interactions were filtered based on expectation score and target accessibility criteria provided by the psRNATarget algorithm to reduce false-positive predictions. Following the initial prediction, analysis focused specifically on miRNAs targeting the newly identified R2R3 *OsMYB* TFs, which were prioritized as potential regulators of anthocyanin biosynthesis. Reported functional roles of these miRNAs were obtained from literature surveys, with emphasis on their known or potential involvement in pigment biosynthesis or integration with MBW complex pathways. The predicted miRNA–target interactions were compiled and summarized to support candidate prioritization of R2R3 *OsMYBs* potentially regulated by miRNAs in anthocyanin-related pathways (Table S2).

2.8 Gene Ontology of OsMYB Genes

Gene Ontology (GO) analysis was performed to investigate the functional roles of *OsMYB* genes from *Oryza sativa* subsp. *japonica*. GO annotations associated with molecular function and biological process categories were initially retrieved from the Phytozome database. The identified *OsMYB* gene set (109 proteins) was submitted to ShinyGO v0.741 (<http://bioinformatics.sdstate.edu/go/>) to conduct GO enrichment analysis. During enrichment analysis, the complete annotated gene set of *Oryza sativa* subsp. *japonica* available in ShinyGO was used as background reference. Statistical significance of GO term enrichment was evaluated using a hypergeometric test, and p-values were adjusted for multiple comparisons using the Benjamini–Hochberg false discovery rate (FDR) correction method. GO terms with an adjusted FDR value ≤ 0.05 were considered significantly enriched. To quantify enrichment significance, fold enrichment values were calculated and expressed as $-\log_{10}(\text{FDR})$ scores. These values were used to rank enriched GO categories and visualize the relative significance of each functional group. The enriched GO terms were summarized to identify major functional roles of *OsMYB* TFs related to anthocyanin biosynthesis.

2.9 Protein–Protein Interactions

To further investigate potential protein–protein interactions of candidate *OsMYB* TFs, interaction network analysis was performed using the STRING database (v0.74111). In addition to *OsMYB* proteins, known *bHLH* and WD40 components implicated in anthocyanin biosynthesis in *Oryza sativa* *OsB2*, *OsRc*, and *OsTTG1* were included in the analysis. This allowed the visualization and prediction of interaction networks specifically involving MYB–bHLH–WD40 (MBW) complexes, which are central to the regulation of anthocyanin structural genes. Predicted interactions were examined to identify candidate *OsMYBs* that may associate with these MBW components, providing evidence for their potential regulatory role in anthocyanin biosynthesis.

2.10 Gene Duplication and Synteny Analysis

The analysis based on non-synonymous substitution rate (K_a) and synonymous substitution rate (K_s) values was used to estimate the divergence time of the *MYB* gene family in *Oryza sativa*. The K_a/K_s ratios for each pair of paralogous genes were calculated using the TBtools software. These ratios helped determine the molecular evolutionary rates of gene pairs. The divergence time (T) was calculated using the formula $T = K_s/(2\lambda)$, where the neutral substitution rate (λ) was assumed to be 6.5×10^{-9} substitutions per site per year [19].

Gene duplication events within the *O. sativa* *MYB* gene family were identified using the MCScanX toolkit with default settings, implemented through TBtools, which allowed for the detection of both segmental and tandem duplications. To visualize the genomic relationships of paralogous *MYB* genes, micro-synteny analysis was conducted using the Micro Synteny View function in TBtools. Furthermore, dual syntenic comparisons were performed between *O. sativa* and *A. thaliana* and *O. sativa* and *O. rufipogon* to assess the conservation and divergence of *MYB* genes across species.

2.11 Expression Profiling of *OsMYB* Genes in Rice Pericarps

Transcriptome data (GSE67987) were retrieved from the NCBI Gene Expression Omnibus (GEO) database to examine the expression of *OsMYB* TFs associated with anthocyanin biosynthesis. Expression values of 109 *OsMYB* genes were obtained from pericarp tissues of black rice near-isogenic lines and mixed rice genotypes at 7 and 14 days after heading. Fold-change (FC) values were calculated to assess developmental regulation in black rice [FC (14 vs. 7 DAF, Black)] and genotype-specific differences [FC (Black vs. Mixed, 7 DAF) and FC (Black vs. Mixed, 14 DAF)]. Genes showing FC > 1 across all three conditions were considered as candidate upregulated regulators. Heatmaps based on fold-change values were generated using TBtools to visualize expression patterns across developmental stages and genotypes, where color intensity (white to dark red) represents increasing fold-change levels. Notably, these observations are based solely on fold-change thresholds without statistical testing, and the mixed rice genotype group serves as an approximate comparator rather than a clearly defined control; therefore, these results were interpreted as preliminary signals for candidate gene prioritization rather than definitive evidence of differential expression.

3 Results

3.1 Comprehensive Genome-Wide and Phylogenetic Characterization of *OsMYB* Genes

A phylogenetic analysis using protein sequences from *O. sativa*, *O. rufipogon*, and *A. thaliana* clustered the *MYB* proteins into three major clades (Fig. 1). In Clade I, several putative orthologous groups were identified. *AtMYB70* clustered with *OsMYB18* (bootstrap 1.0000), and *AtMYB59* clustered

with *OsMYB36* (bootstrap 0.9820). The *Arabidopsis* proteins *AtMYB53*, *AtMYB81*, and *AtMYB60* formed a subcluster (bootstrap 0.9920 and 0.9681), which then grouped with *OsMYB94* (bootstrap 0.9401). *AtMYB31* clustered with *OsMYB98* (bootstrap 0.9142), and *AtMYB83* grouped with *OsMYB50* and *OsMYB44*, where the *OsMYB44/OsMYB50* subcluster had bootstraps of 1.0000 and 0.9920, the combined *OsMYB* cluster 0.8563, and the addition of *AtMYB83* 0.8623. Clade II (bootstrap 0.9960) contained the *AtMYB63/OsMYB61* pair. Clade III (bootstrap 0.8683) encompassed multiple putative orthologs, including *AtMYB92/OsMYB75*, *AtMYB37/OsMYB7*, *AtMYB86/OsMYB26*, *AtMYB62/OsMYB12/OsMYB62*, *AtMYB50/AtMYB51/OsMYB1*, *AtMYB102* with *OsMYB35/OrMYB46* (bootstrap 0.8523), *AtMYB39* with *OsMYB9/OrMYB43* (bootstrap 0.9960), and *AtMYB57/AtMYB101* with *OsMYB54/OsMYB87* (bootstrap 0.9960) and *AtMYB55/OsMYB25* (bootstrap 0.9501). Bootstrap values across the tree ranged from 0.85 to 1.00, indicating strong support for these clades and subclades. Phylogenetic analysis further supported the potential role of several newly identified candidate *MYB* genes in anthocyanin-related regulatory networks through downstream analysis. Notably, candidates such as *OsMYB44*, *OsMYB50*, *OsMYB61*, *OsMYB65*, and *OsMYB94* clustered in close proximity to *Arabidopsis* *MYBs*, including *Arabidopsis thaliana* *MYB31*, *MYB83*, *MYB63*, *MYB60*, and *MYB53*, suggesting possible functional relationships. It is important to note that orthology is suggested based on sequence similarity and clustering, and functional conservation should not be inferred without experimental validation. This analysis provides a comprehensive overview of the *OsMYB* gene family and their phylogenetic relationships, highlighting potential evolutionary patterns among rice and *Arabidopsis*. The subcellular localization data indicates distinct patterns of presence of 109 *OsMYB* proteins across various cellular compartments. The majority of *OsMYB* proteins were predominantly localized in the nucleus (Fig. 2).

3.2 Gene Structure Analysis

Gene structure analysis revealed substantial variation in exon–intron organization among the 109 *OsMYB* genes (Fig. 3). The observed exon/intron configurations included 1/0, 2/1, 2/2, 2/3, 3/2, 3/3, 4/0, 6/5, 7/6, and 10/9. Among these, the most common structure was three exons and two introns, followed by two exons and one intron, indicating a general conservation of gene architecture within the family alongside notable diversification.

Focusing on the newly identified candidate genes associated with anthocyanin regulation (*OsMYB5*, *OsMYB6*, *OsMYB7*, *OsMYB8*, *OsMYB9*, *OsMYB15*, *OsMYB24*, *OsMYB25*, *OsMYB32*, *OsMYB41*, *OsMYB43*, *OsMYB44*, *OsMYB45*, *OsMYB47*, *OsMYB50*, *OsMYB59*, *OsMYB60*, *OsMYB61*, *OsMYB62*, *OsMYB65*, *OsMYB68*, *OsMYB69*, *OsMYB70*, *OsMYB71*, *OsMYB81*, *OsMYB86*, *OsMYB89*, *OsMYB94*, *OsMYB98*, *OsMYB101*, *OsMYB104*, and *OsMYB105*), a clear pattern emerged. The majority of these candidates exhibited the 3 exon/2 intron (2/3) or 2 exon/2 intron (2/2) structures, which are characteristic features commonly observed in R2R3-*MYB* genes implicated in flavonoid and anthocyanin biosynthesis in plants. This structural conservation supports their potential functional relevance. In contrast, a few candidate genes displayed more divergent architectures. For example, *OsMYB47* showed the most complex structure with 10 exons and 9 introns, while *OsMYB71* (6/5) also exhibited relatively higher exon–intron numbers. Conversely, simpler structures such as 2/1 (e.g., *OsMYB32*, *OsMYB44*, *OsMYB60*, *OsMYB70*, *OsMYB81*, and *OsMYB89*) and 1/0 (*OsMYB50*) were also observed, indicating structural variability even within candidate regulators. Gene structures alone is insufficient to infer function; therefore, candidate gene identification in this study was based on an integrative approach combining phylogenetic relationships, promoter cis-element analysis, conserved motif detection, and expression profiling.

3.3 Conserved Motif Analysis and Domain Prediction Analysis of OsMYBs

Motif analysis of the *MYB* gene family in *Oryza sativa* confirmed the presence of conserved domains and motifs characteristic of *MYB* TFs (Fig. 4). The identified *MYB* genes possessed several key domains, including *MYB*_DNA-binding, *MYB*_DNA-binding superfamily, *MYB*_DNA-bind_6, the P_C superfamily, and the SANT superfamily. Specifically, Motif 1 (associated with *MYB*_DNA-binding) and Motif 2 (*MYB*_DNA-binding superfamily) were the most frequently occurring domains across the *MYB* gene family suggesting a predominant representation of R2R3-type *MYB* TFs (Fig. 4). These domains, often located toward the N-terminal region, are key components of the R2 and R3 repeats responsible for DNA binding in anthocyanin biosynthesis. For instance, *AtMYB75* from *Arabidopsis* harbor similarly conserved R2R3 motifs and are experimentally validated to activate anthocyanin biosynthesis [20], thus reinforcing the functional relevance of these motifs.

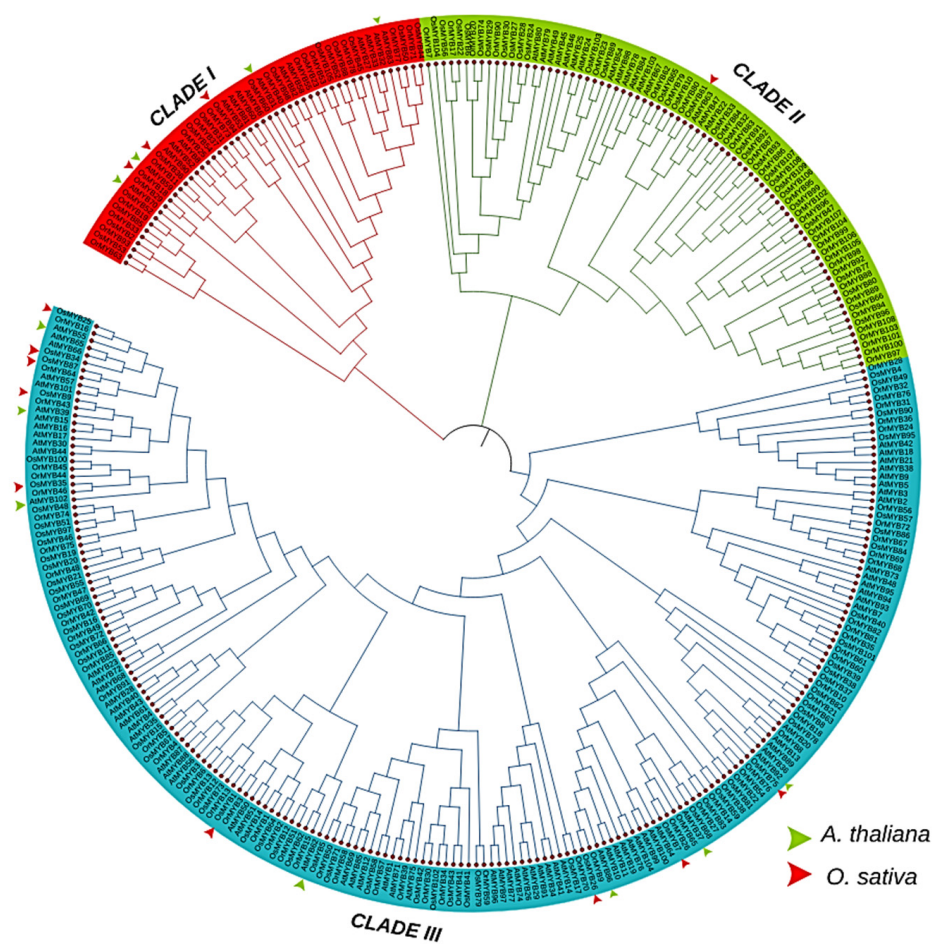


Figure 1: The phylogenetic relationships among *MYB* genes from *O. sativa*, *A. thaliana*, and *O. rufipogon* were investigated using a maximum likelihood (ML) approach with 1000 bootstrap replicates. A total of 109 *MYB* gene sequences were analyzed using MEGA X v12 software. The resulting phylogenetic tree revealed three distinct clades, each visually represented by a unique color. *MYB* genes orthologous from *A. thaliana* and *O. sativa* were clearly distinguished by green and red markers, respectively.

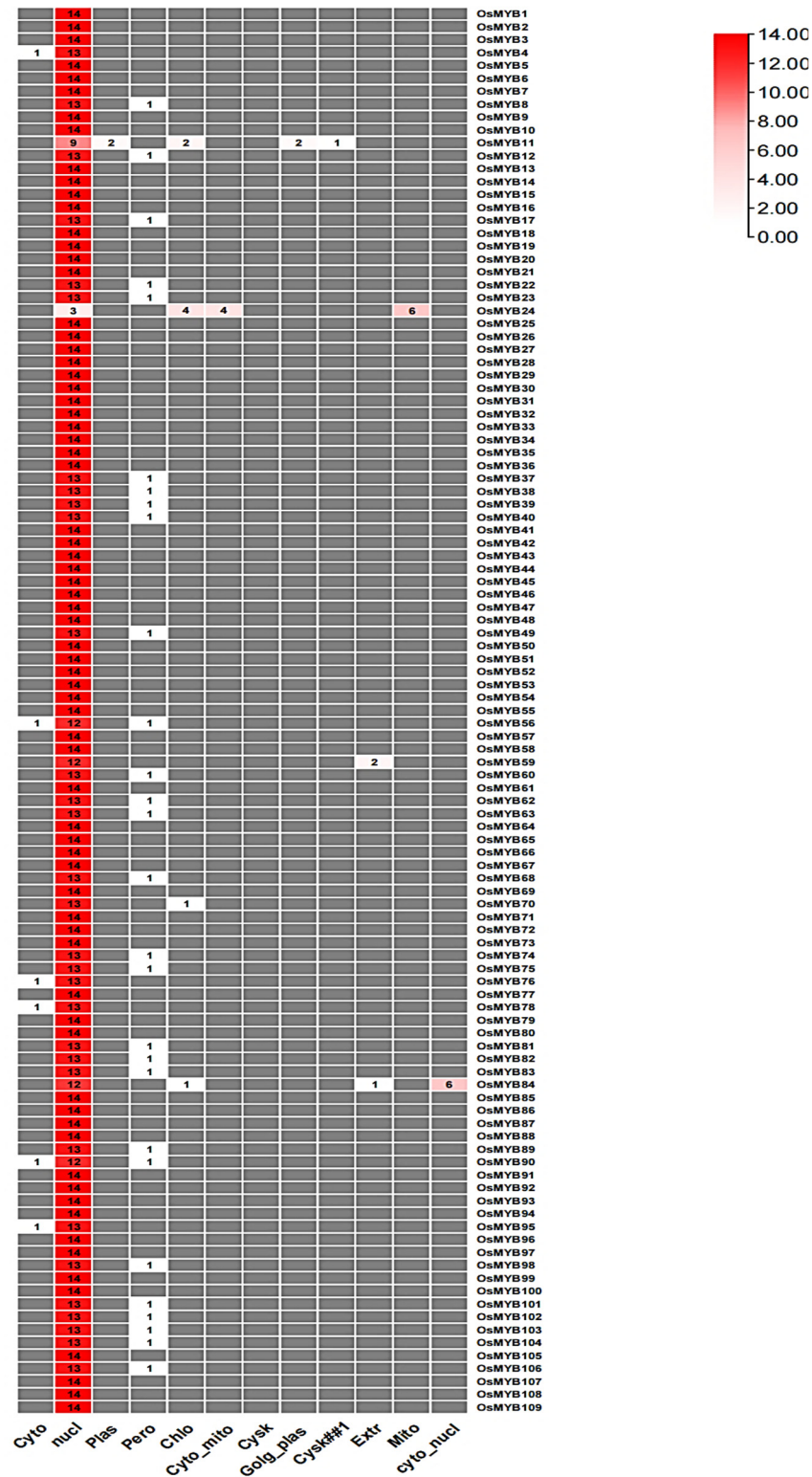


Figure 2: Subcellular localization prediction of *OsMYB* proteins revealed their predominant presence in the nucleus, peroxisomes, and cytoplasm. The majority of proteins were represented by red-colored markers, indicating their primary localization within these cellular compartments.

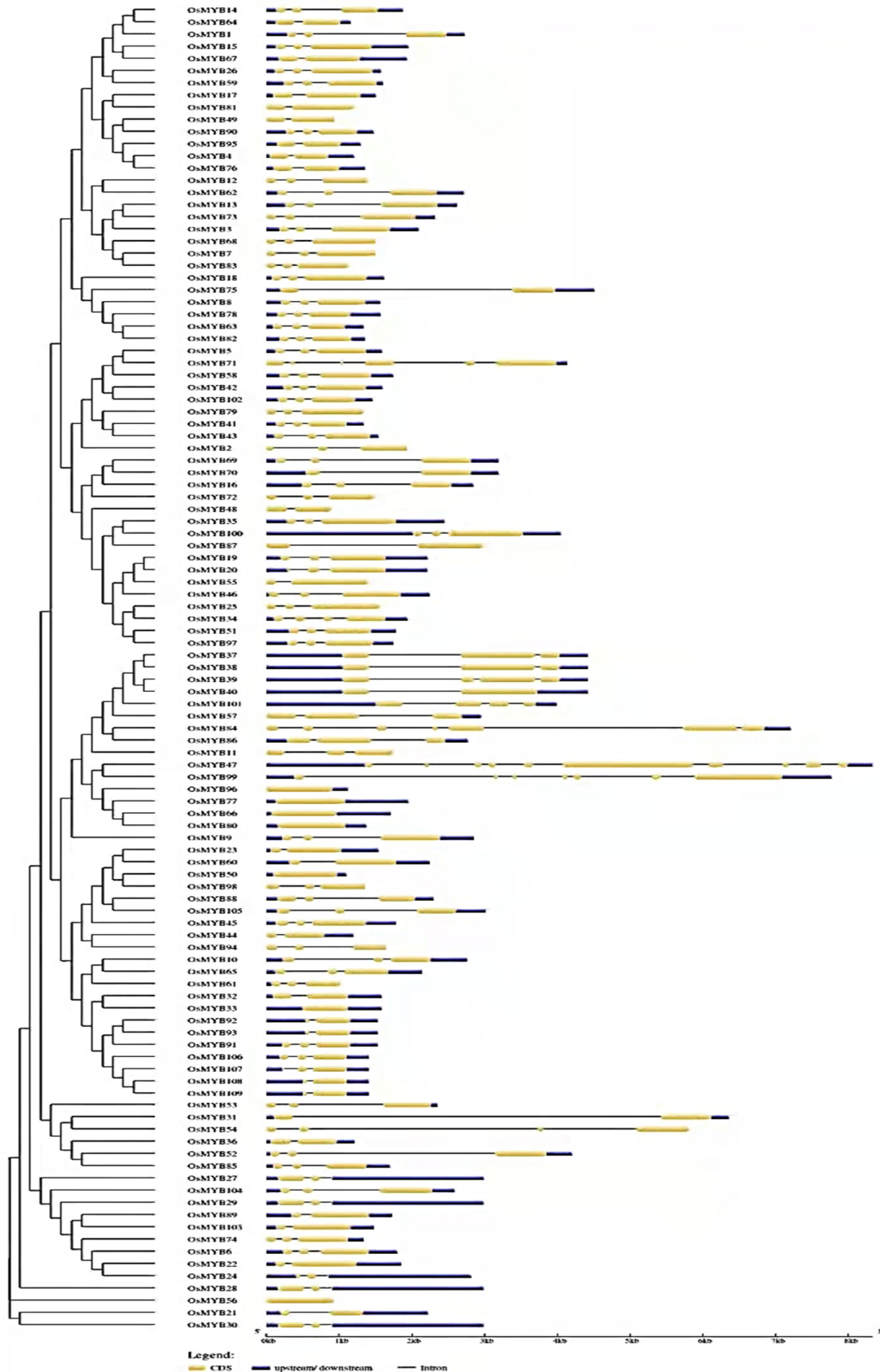


Figure 3: The exon-intron structures were depicted using shapes colored in Yellow and black to represent the exons contained within the genes.

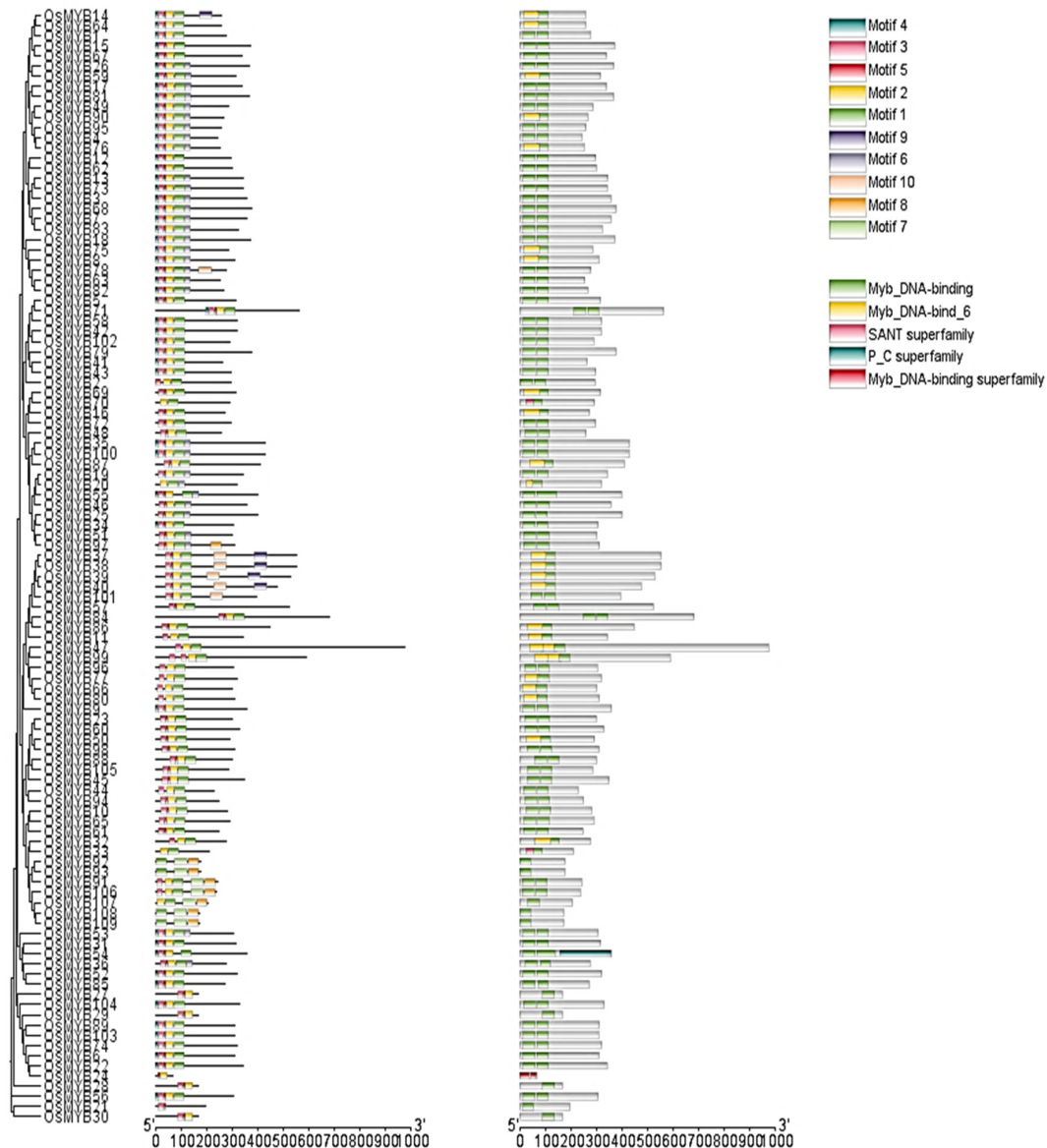


Figure 4: The analysis of *OsMYB* genes involved the identification of conserved domains and motifs, alongside the construction of their phylogenetic relationships. Using MEME version 5.5.2, ten distinct motifs were identified and visualized through a color-coded bar graph. This motif distribution was aligned with a phylogenetic tree, allowing for a comparative view of evolutionary relatedness among the proteins.

In addition to these, five motifs (Motifs 6, 7, 8, 9, and 10) were conserved across several MYB proteins, indicating shared functional or structural roles. Interestingly, *OsMYB70* and *OsMYB33* were the only genes that harbored the SANT superfamily domain, while *OsMYB54* uniquely contained the P_C superfamily domain. The presence of additional domains such as the SANT and P_c superfamily in a few proteins further indicates domain expansion and potential specialization in regulatory roles. Domains unique to these, such as SANT or P_C, possibly point to particular regulatory functions for them. In other TFs, variation in motifs located in the C-terminal regions was found to affect the specificity of secondary metabolic gene targets, including flavonoid biosynthetic genes. Domains in *OsMYBs* that have expanded might indicate possible functional divergence for these to be verified.

3.4 Cis-Element Analysis of OsMYBs

Cis-element analysis of the promoter regions of the 109 *OsMYB* genes identified 105 regulatory elements belonging to several functional categories, including 27 light-responsive elements, 22 stress-responsive elements, 20 development- and metabolism-related elements, 16 hormone-responsive elements, 6 core promoter/general transcription elements, 1 signal transduction/calcium/redox element, and 13 uncharacterized elements (Table 1; Fig. 5). Among these, core promoter elements such as the CAAT-box and TATA-box were the most abundant, consistent with their fundamental role in transcription initiation. Moderate frequencies of regulatory motifs such as AT-TATA-box, ABRE, G-box, MYC, and MYB elements were also detected across multiple *OsMYB* promoters.

Particular attention was given to cis-elements associated with flavonoid and anthocyanin pathway regulation in the promoters of the prioritized candidate *MYB* genes. Several light-responsive motifs, including the MYB-recognition element (MRE), Chs-CMA1a, Chs-CMA2a, G-box, and Box 4, were identified in these promoters [21]. These elements are known to mediate transcriptional activation of flavonoid biosynthetic genes under light-regulated condition. In addition, development-related elements such as AC-I and AC-II, which are associated with tissue-specific expression and flavonoid biosynthesis, were detected in several candidate *OsMYB* promoters. The presence of these regulatory motifs suggests that the prioritized *MYB* genes may participate in the transcriptional regulation of anthocyanin biosynthesis through integration of environmental and developmental signals.

Table 1: Cis-regulatory elements in putative *OsMYB* promoters, which are associated with different plant developmental processes.

SR. No.	Cis-Elements	Function	References
Light-responsive			
1	3-AFI binding site	Involved in light-responsive gene expression regulation in plants	[22]
2	3-AF3 binding site	Functions in light-regulated transcription, binds specific TFs	[23]
3	AE-box	Light-responsive element; interacts with plastid and nuclear genes	[24]
4	ACE	Light-responsive element involved in photoregulated gene expression	[25]
5	AT1-motif	Light-responsive element; contributes to tissue-specific gene regulation	[26]
6	ATC-motif	Light-regulated; enhances transcription in photosynthetic tissues	[27]
7	ATCT-motif	Part of a conserved DNA module involved in light responsiveness	[28]
8	Box 4	Light-responsive element promoting transcription in photosynthetic genes.	[29]
9	Box II	Enhancer-like elements are associated with tissue-specific light response.	[30]
10	G-Box	Recognized by bZIP and bHLH; light/stress regulated. Cis-acting regulatory element involved in light responsiveness	[29]
11	G-box	Light, hormone, and stress response via transcription factor binding. Cis-acting regulatory element involved in light responsiveness	[29]
12	GA-motif	Involved in gibberellin-responsive gene expression.	[31]
13	GATA-motif	Light-regulated expression through GATA transcription factor binding.	[30]
14	GT1-motif	Regulates light-responsive genes and salt-stress inducibility.	[30]
15	GTGGC-motif	Functions in pathogen or abiotic stress response regulation. Part of a light responsive element	[30]
16	Gap-box	Part of a light responsive element	[30]

Table 1: Cont.

SR. No.	Cis-Elements	Function	References
17	H-box	Often linked to pathogen and light-responsive transcription. Root-specific regulatory element	[32]
18	I-box	Participates in light-regulated transcription of photosynthetic genes.	[33]
19	L-box	May relate to light regulation and photomorphogenesis.	[32]
20	LAMP-element	Part of a light-responsive element	[30]
21	MRE	MYB-recognition element involved in light stress response, Flavanoid Biosynthesis	[30]
22	Pc-CMA2c	Part of a light responsive element	[34]
23	Sp1	GC-rich element; binding site for Sp1 TF. Light responsive element	[30]
24	TCCC-motif	May function in light or hormonal regulation.	[30]
25	TCT-motif	Functions in light-regulated transcription activation.	[30]
26	Chs-CMA1a	Light-responsive element in chalcone synthase gene promoters.	[30]
27	Chs-CMA2a	Enhances light-dependent expression of flavonoid genes.	[30]
Hormone-responsive			
28	ABRE	Responsible for ABA (Absciscic acid) signaling	[35]
20	ABRE3a	Variant of ABRE; enhances abscisic acid-mediated stress response	[36]
30	ABRE4	Works with other ABREs to mediate ABA responsiveness	[36]
31	AUXRE	Auxin-responsive element; mediates auxin-regulated gene transcription	[37]
32	AUXRR-core	Core element involved in auxin-responsive gene expression.	[38]
33	CGTCA-motif	Part of ABA or MeJA response pathways.	[39]
34	ERE	Ethylene-responsive element regulating ethylene-mediated gene expression.	[30]
35	GARE-motif	Gibberellin-responsive element for growth-related gene expression.	[30]
36	P-box	Gibberellin-responsive cis-element in growth-related genes.	[30]
37	TATC-box	Gibberellin-responsive motif in growth and development genes.	[30]
38	TGA-element	Auxin-responsive element controlling auxin-related gene expression.	[40]
39	TGAC G-motif	Involved in pathogen and hormonal signaling response. Cis-Acting regulatory element involved in the MeJA-responsiveness	[30]
40	SARE	SA-responsive element in salicylic acid signaling.	[41]
41	TCA	Core of salicylic acid response element.	[42]
42	TCA-element	SA-responsive element controlling pathogen response genes.	[30]
43	as-1	Abiotic stress/Pathogen/stress-inducible element for salicylic acid response.	[43]
Stress-responsive.			
44	ACTCATCCT sequence	Abiotic-related response elements-responsive.	[32]
45	AP-1	Binds Jun/Fos; regulates stress, immunity, and proliferation genes	[44,45]
46	ARE	Anaerobic responsive element; activates genes under low oxygen conditions	[30]
47	AT-rich element	Enhances promoter activity; involved in chromatin remodeling, Binding site of AT-rich DNA binding protein (ATBP-1), Heat-induced expression	[44]
48	DRE core	Essential for drought and cold stress gene induction.	[46]
49	DRE1	Specific drought-responsive element for dehydration-inducible genes.	[46]
50	GC-motif	Involved in anoxic or hypoxic stress response pathways.	[47]
51	LTR	Low temperature-responsive element activating cold-stress genes.	[30]
52	MBS	MYB binding site for drought-inducible gene expression.	[30]

Table 1: Cont.

SR. No.	Cis-Elements	Function	References
53	MBS I	Variant <i>MYB</i> binding site; involved in stress response.	[48]
54	<i>MYB</i>	Core site for <i>MYB</i> TF binding.	[18]
55	<i>MYC</i>	<i>MYC</i> TF binding site; regulates ABA and JA response.	[49]
56	<i>MYB</i> -binding site	Regulates transcription in response to abiotic stress.	[18]
57	<i>Myc</i>	JA/ABA responsive site for <i>MYC</i> family TFs.	[49]
58	STRE	Controls transcription under heat shock and other stresses.	[50]
59	TC-rich repeats	Involved in defense and stress-responsive transcription.	[30]
60	W box	WRKY transcription factor binding site; pathogen defense.	[51]
61	WRE 3	Wound-responsive element for injury-induced gene activation.	[52]
62	WUN_motif	Wound-responsive element activating defense-related genes.	[52]
63	Y-box	Cold-responsive binding site; RNA/DNA binding function.	[53]
64	box S	May regulate stress response and light-sensitive transcription.	-
65	CCGTCC motif	Involved in <i>MYB</i> TF binding for regulation.	
Development/Tissue-specific			
66	A-box	Seed-specific and hormone-responsive gene regulation	[54]
67	AAGAA-motif	Pollen-specific or pathogen-responsive gene expression	[31]
68	AC-I	Xylem specific expression in Cell development, Flavanoid Biosynthesis	[27]
69	AC-II	Xylem specific expression in Cell development, Flavanoid Biosynthesis	[27]
70	CCGTCC-box	Functions in stress-responsive gene expression via <i>MYB</i> proteins. Cis-Acting regulatory element related to meristem specific activation	[30]
71	CAT-box	Associated with meristem expression and tissue-specific activity.	[54]
72	CCAAT-box	Enhances promoter activity, common in housekeeping genes.	[55]
73	CTAG-motif	Involved in endosperm-specific gene expression regulation.	[56]
74	GCN4_motif	Controls endosperm-specific expression of storage protein genes.	[30]
75	HD-zip1	Binding site for HD-ZIP transcription factors in development. In disease resistance. In salt stress response	[57,58]
76	MSA-like	Cell cycle-regulating element active during mitosis.	[59]
77	<i>MYB</i> recognition site	Regulates gene expression during stress or development.	[18]
78	<i>MYB</i> -like sequence	Putative <i>MYB</i> binding site, may regulate transcription.	[18]
79	<i>MYB</i>	Alternative naming for <i>MYB</i> -binding; modulates stress or development.	[18]
80	NON-box	Meristem specific activation	[60]
81	O2-site	Zein gene regulator in maize; Opaque2 binding site.	[30]
82	RY-element	Regulates seed-specific gene expression and dormancy.	[30]
83	circadian	Controls circadian-regulated expression of rhythmic genes.	[61]
84	Plant_Ap-2-like	Related to AP2/ERF TF binding; stress development regulation. Starch biosynthesis	[62]
85	Re2f-1	Cell cycle regulatory element; binds E2F transcription factors.	[63]
Core promoter elements.			
86	CAAT-box	Common cis-acting element in promoter and enhancer regions	[64,65]
87	AT-rich sequence	Functions in DNA-protein interaction and transcription factor binding	[66]
88	AT-TATA-box	Variant of TATA-box; aids in transcription initiation accuracy	-
89	TATA-box	Core promoter element for transcription initiation complex.	[67]
90	OCT	Octamer motif involved in basic promoter activity.	[68]
91	dOCT	Modified octamer motif for tissue-specific transcription.	-
Signal Transduction/Calcium/Redox Elements			
92	CARE	Calcium-responsive element involved in calcium-mediated signal transduction.	[69]
93	Unnamed_1	Uncharacterized element; possible role in transcription regulation.	-
94	Unnamed_1	(same as above)	-

Table 1: *Cont.*

SR. No.	Cis-Elements	Function	References
95	Unnamed_10	Unknown; may be computationally predicted motif.	-
96	Unnamed_12	Unknown; potential regulatory sequence with no confirmed function.	-
97	Unnamed_14	Unidentified motif; role likely under investigation.	-
98	Unnamed_16	Possibly stress or hormone related; function not confirmed.	-
99	Unnamed_2	Uncharacterized; may have context-specific activity.	-
100	Unnamed_2	(same as above)	-
101	Unnamed_4	Unknown; possibly a conserved but unstudied regulatory element.	-
102	Unnamed_6	Likely novel or predicted cis-element; unknown role.	-
103	Unnamed_8	Function unclear; found in some stress-regulated promoters.	-
104	motif I	Broad category; often involved in stress or development.	-
105	NON	Function not well-defined; may act in root-specific expression.	-

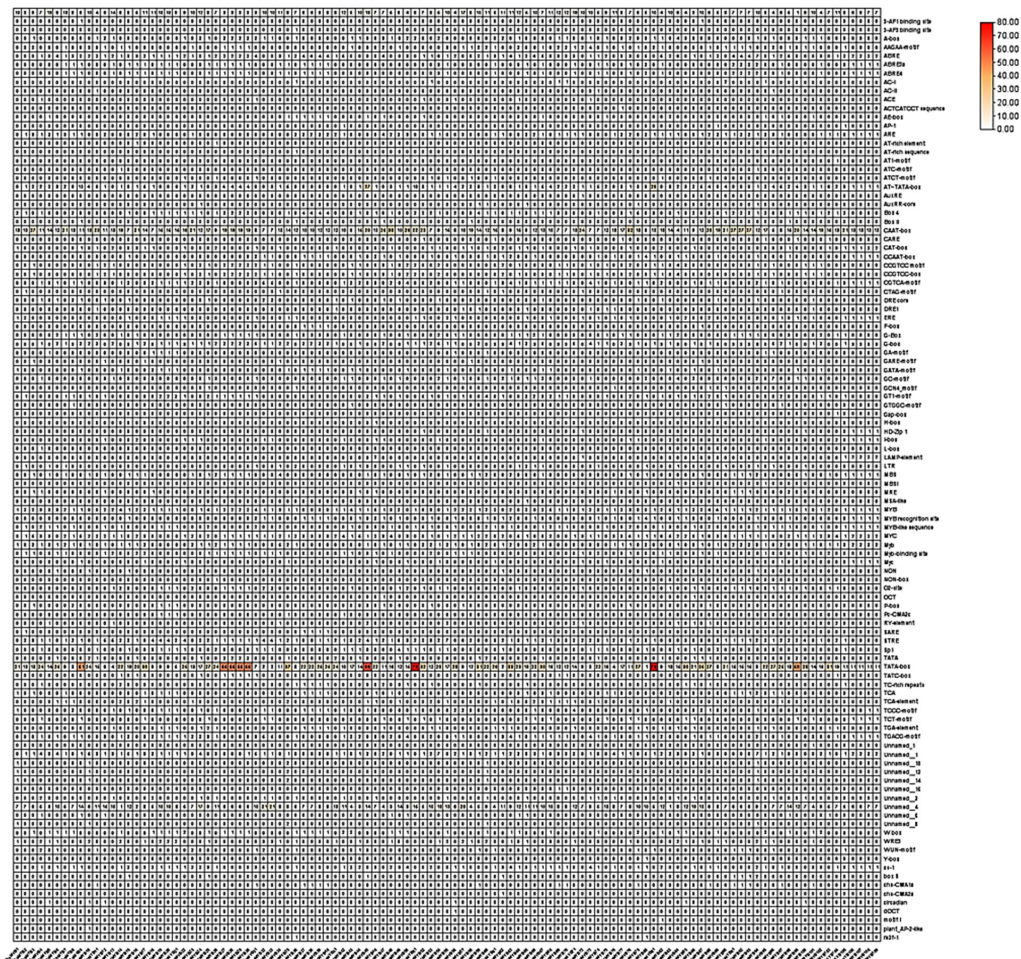


Figure 5: The investigation of cis-regulatory elements within the promoter regions of *O. sativa* MYB genes revealed their potential roles in a wide range of plant developmental and physiological processes. This analysis detailed the presence and frequency of various regulatory elements associated with growth, stress response, hormone signaling, and environmental cues. Furthermore, the spatial distribution of these elements across the promoter regions was assessed, offering insights into their organization.

3.5 Protein–Protein Interaction

To explore potential regulatory roles of candidate OsMYBs in anthocyanin biosynthesis, protein–protein interaction (PPI) networks were analyzed using STRING, including the key MBW components in *O. sativa*: *OsB2*, *OsRc*, and *OsTTG1* (Fig. 6). The resulting network comprised 95 nodes and 682 edges, with an average node degree of 14.4 and an average local clustering coefficient of 0.639, indicating substantial connectivity among the proteins. The expected number of edges based on random chance was 46, and the PPI enrichment *p*-value was $<1.0\text{e}-16$, confirming that the observed interactions are significantly more interconnected than expected by chance.

Within the network, several candidate OsMYBs were predicted to interact directly with MBW components. Specifically, *OsTTG1* interacted with *OsMYB25*, *OsMYB44*, *OsMYB65*, *OsMYB104*, and *OsMYB105*, while *OsRc* was predicted to interact with *OsMYB15* and *OsMYB44*. These predicted interactions suggest that the corresponding OsMYBs may potentially function as components of the MYB–bHLH–WD40 (MBW) complex, which is known to regulate anthocyanin biosynthesis in plants. The network analysis also indicated clusters of OsMYBs with multiple putative interactions, highlighting them as promising candidates for involvement in tissue-specific pigment regulation. However, these interactions are based on computational predictions and should be interpreted with caution. Notably, these OsMYBs were independently identified as candidate regulators through integrative phylogenetic and expression analyses, providing additional support for their possible roles in pigment biosynthesis.

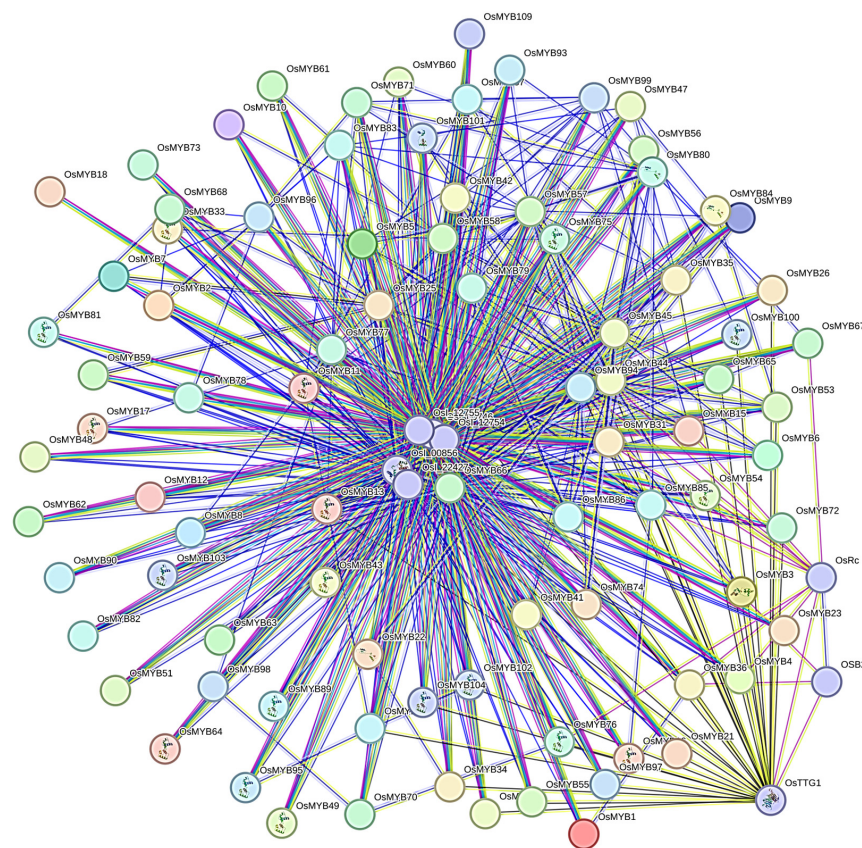


Figure 6: Protein–protein interaction network of OsMYB TFs involved in anthocyanin biosynthesis in *Oryza sativa*. The network illustrates interactions between OsMYBs and the known bHLH and WD40 components, including *OsB2*, *OsRc*, and *OsTTG1*, highlighting connectivity and potential regulatory relationships in the anthocyanin pathway.

3.6 Gene Duplication and Synteny Analysis of *OsMYB*

The expected chromosomal positions indicated that the *OsMYB* genes were widely distributed across all 12 chromosomes of *O. sativa*, with additional loci present on unassembled (ChrUn) and synthetic regions (ChrSy), as shown in Fig. 7. A notable concentration of *OsMYB* genes was observed on chromosome 1 and chromosome 4. Several candidate anthocyanin-associated *OsMYB* genes showed chromosomal clustering patterns. For instance, *OsMYB5*, *OsMYB6*, *OsMYB7*, *OsMYB8*, and *OsMYB9* were all located on chromosome 8, while another group including *OsMYB60*, *OsMYB62*, *OsMYB68*, *OsMYB70*, and *OsMYB71* was located on chromosome 2. Such local clustering of candidate *MYB* genes may reflect historical tandem or segmental duplication events that contributed to the expansion of regulatory *MYB* TFs in rice.

The circular heatmap presents the estimated divergence times of paralogous *O. sativa* *MYB* gene pairs, calculated using synonymous substitution rates (*Ks*) and the molecular clock formula $T = Ks/(2\lambda)$, where λ is set at 6.5×10^{-9} substitutions per site per year, a rate commonly used for monocot plants [19] (Fig. 8). The color gradient from white to deep red represents increasing divergence time in millions of years ago (MYA), with darker shades indicating more ancient duplication events. Several gene pairs, such as *OsMYB34/OsMYB39*, *OsMYB2/OsMYB94*, and *OsMYB15/OsMYB39*, exhibit high *Ks* values and are shaded in deep red, corresponding to estimated divergence times of approximately 400–500 MYA under the assumed substitution rate, likely during ancient whole-genome duplication events. Notably, some of these duplicated genes correspond to candidate anthocyanin regulators identified in this study. For example, both *OsMYB15* and *OsMYB94*, which displayed relatively high *Ks* values in duplication analysis, were also prioritized as potential anthocyanin regulatory candidates based on integrative phylogenetic, promoter, and expression analyses. In contrast, many other gene pairs show much lower *Ks* values, indicating more recent duplications possibly resulting from segmental or tandem duplication events within the rice genome. A *Ka/Ks* ratio 1.0 for paralogous *OsMYB* gene pairs originated over 400–500 MYA indicates that these genes are likely evolving neutrally, with no strong evolutionary pressure to maintain or change their protein-coding sequences. Paralogous *OsMYB* gene pairs with a *Ka/Ks* ratio of 0 are under strong purifying selection, indicating evolutionary pressure to maintain their protein sequence and function. This observation suggests that certain duplication events within the *MYB* family may have contributed to the diversification of TFs potentially involved in pigment-related pathways, although functional validation would be required to confirm these roles [70].

Several *OsMYB* genes in rice exhibit strong syntenic conservation with genes in *Arabidopsis*, indicating evolutionary retention of key regulatory functions (Fig. 9a,b). Both *OsMYB5* and *OsMYB7* were syntenically linked with Chr5 genes, likely arising from segmental duplications and retaining conserved domains seen in *Arabidopsis* *MYBs* [71]. *OsMYB31* and *OsMYB50*, both on Chr1, also showed syntenic conservation, hinting at their descent from ancestral *MYB* genes. *OsMYB50* identified in this study as a candidate regulator of anthocyanin biosynthesis shares conserved genomic context, supporting its potential functional relevance. In Chr3, *OsMYB54* and *OsMYB60* exhibited multiple syntenic links, suggesting stable evolutionary roles. *OsMYB60*, along with *OsMYB62*, *OsMYB68*, and *OsMYB71* on chromosome 2, were among the newly proposed candidate anthocyanin regulators and showed conserved syntenic relationships with regions on chromosomes 3 and 5. These patterns suggest that some of the candidate genes may have retained ancestral regulatory functions following duplication events. Gene duplication and synteny analysis was mainly performed in order to place the *OsMYB* family into an evolutionary context.

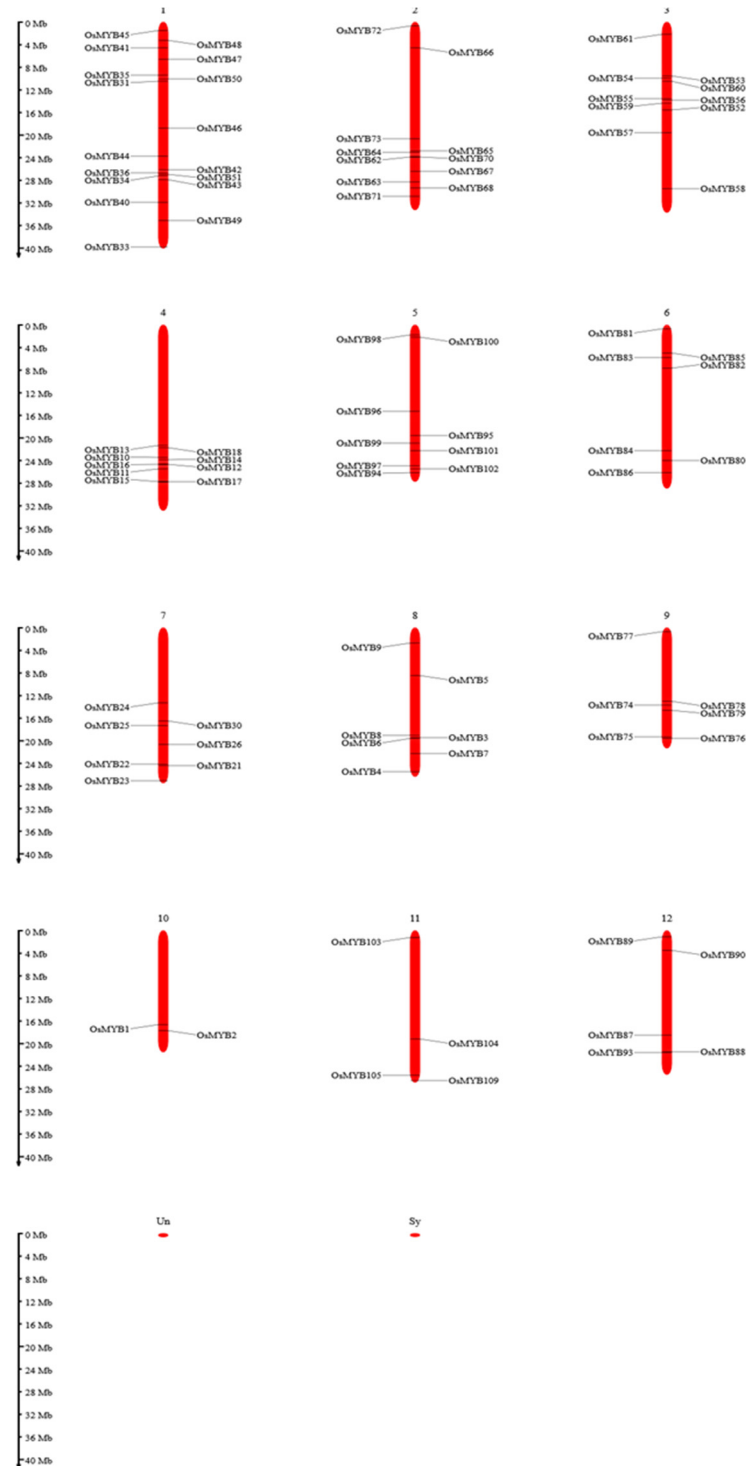


Figure 7: The chromosomal distribution of the *OsMYB* gene family in *O. sativa*. The figure's red bars represent *O. sativa* chromosomes, while the black lines indicate the functional relationships or co-regulation between *OsMYB* genes. This analysis provides insights into the spatial arrangement and potential interactions among *OsMYB* genes within the *O. sativa* genome.

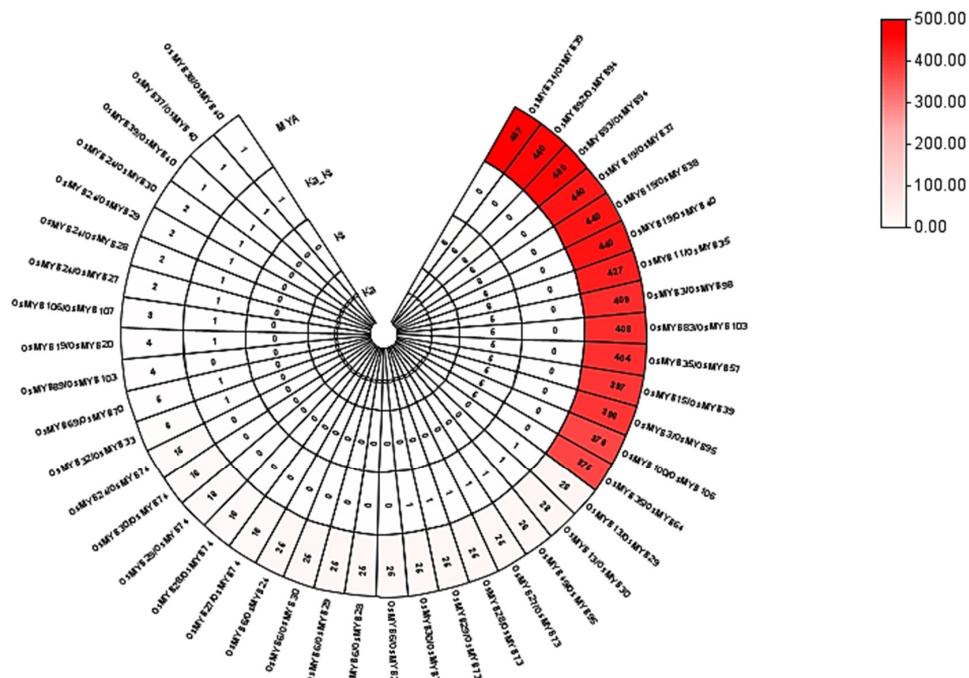


Figure 8: The time of gene duplication for various paralogous *OsMYB* gene pairs in *O. sativa* based on synonymous (Ks) and nonsynonymous (Ka) substitution rates. The analyses were performed using the TBtools software. The Ka/Ks ratio, representing the rate of nonsynonymous (Ka) to synonymous (Ks) substitutions, was calculated to infer the selection pressure acting on these gene pairs.

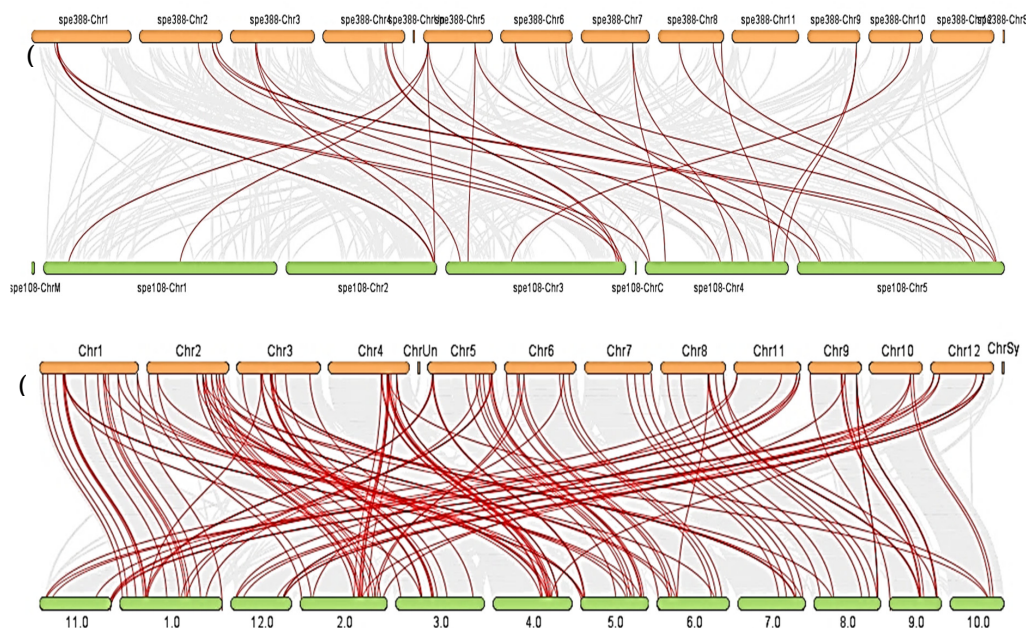


Figure 9: Dual synteny plots were generated to illustrate the genomic collinearity of putative *OsMYB* genes between *O. sativa* and two other plant species. Red lines indicate orthologous gene pairs of *OsMYB* genes across species. Chromosomes of *O. sativa* are shown in orange for clear distinction. (a) Synteny map between *O. sativa* and *A. thaliana*, (b) Synteny map between *O. sativa* and *O. rufipogon*.

3.7 *Osa-miRNA–OsMYB Regulatory Interactions in O. sativa*

Predicted *Osa-miRNA* interactions were analyzed to identify potential post-transcriptional regulators of the newly prioritized R2R3 *OsMYB* genes implicated in anthocyanin biosynthesis (Table S2). Several *miRNA* families were found to target candidate anthocyanin-related *OsMYBs*, suggesting potential roles in modulating their expression. For example, *Osa-miR159a–e* was predicted to target *OsMYBs* associated with reproductive tissue expression, whereas *Osa-miR164a, b, d, and e* were linked to *OsMYBs* expressed in leaves and developing grains [72]. Similarly, *Osa-miR166a, b, d, f, g, and h* were predicted to interact with *OsMYBs* showing tissue-specific expression patterns consistent with anthocyanin accumulation. Several additional *miRNAs*, including *Osa-miR319a/b*, *Osa-miR396a–c*, *Osa-miR408*, *Osa-miR528*, *Osa-miR5505*, and *Osa-miR5809b*, were identified as potential regulators of candidate *OsMYBs* expressed under environmental and developmental cues known to influence pigment accumulation [73–76]. Development-related *miRNAs* such as *Osa-miR171a/h*, *Osa-miR172c*, *Osa-miR444a/d*, and *Osa-miR531b* were also predicted to target *OsMYBs* with spatial or temporal expression patterns relevant to anthocyanin deposition in reproductive tissues and grains.

Overall, these predicted *miRNA*–candidate *OsMYB* interactions highlight specific *miRNA* families that may contribute to post-transcriptional regulation of anthocyanin biosynthesis, either by fine-tuning expression in pigmented tissues or integrating developmental and environmental signals. These interactions are based on computational predictions and literature-based functional annotations, and experimental validation is necessary to confirm their direct regulatory roles in anthocyanin accumulation.

3.8 Gene Ontology

GO enrichment analysis of *Oryza sativa* MYB proteins identified several biological processes associated with transcriptional regulation and metabolic control that are relevant to secondary metabolite biosynthesis. Significant enrichment was observed for regulation of transcription, DNA-templated (3 genes, fold enrichment = 21.7, FDR = 2.0×10^{-3}), regulation of RNA biosynthetic process (3 genes, fold enrichment = 21.7, FDR = 2.0×10^{-3}), and regulation of RNA metabolic process (3 genes, fold enrichment = 21.1, FDR = 2.0×10^{-3}), indicating that a subset of *OsMYB* proteins may function in transcriptional control of downstream metabolic pathways (Fig. 10). Additional enriched categories included regulation of biosynthetic processes (3 genes, fold enrichment = 19.7, FDR = 2.0×10^{-3}) and regulation of metabolic process (3 genes, fold enrichment = 13.9, FDR = 3.3×10^{-3}), which are consistent with regulatory roles in plant secondary metabolism. Since *MYB* TFs are widely recognized regulators of genes involved in the Anthocyanin biosynthesis pathway and the related Flavonoid biosynthesis pathway, these enriched processes suggest that some *OsMYB* proteins may participate in transcriptional networks controlling pigment biosynthesis [77]. In addition, developmental processes such as anther development (1 gene, fold enrichment = 216.2, FDR = 1.2×10^{-2}) and flower development (1 gene, fold enrichment = 34.9, FDR = 4.2×10^{-2}) were also enriched, which may be relevant because anthocyanin pigments commonly accumulate in reproductive tissues. Overall, the enrichment pattern supports a potential regulatory association between *OsMYB* TFs and pathways involved in secondary metabolism and pigment-related biological processes.

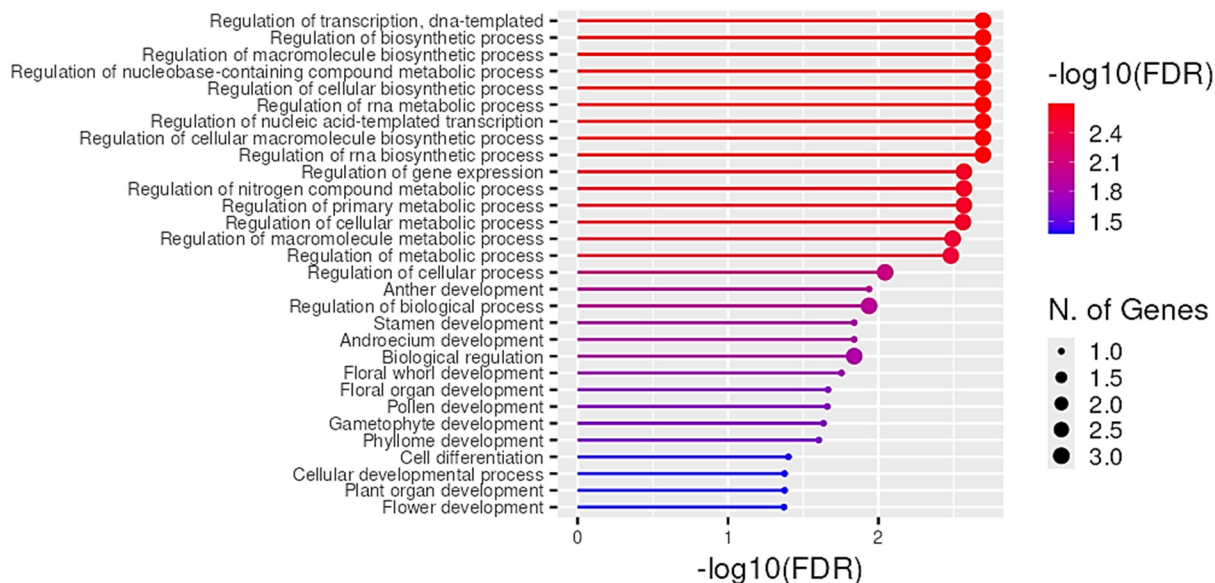


Figure 10: The bar plot illustrates significantly enriched Gene Ontology (GO) categories of 109 OsMYB proteins. This highlights key molecular functions and biological pathways that *MYB* genes are likely involved in to understand the functional roles of these transcription factors in rice. The numerical values of each bar represent the fold enrichment scores expressed as $-\log_{10}$ of the false discovery rate (FDR), indicating the statistical significance of enrichment for each GO term. Higher $\log_{10}(\text{FDR})$ values indicate stronger statistical significance, meaning that the associated GO category is very unlikely to have appeared by random chance.

3.9 Functional Interpretation of OsMYB Genes Based on Arabidopsis Orthologues

The functional interpretation of the *OsMYB* gene family based on Gene Ontology annotations and *A. thaliana* orthologues reveals their diverse regulatory roles in rice (Table S3). Functional interpretation of the candidate *OsMYB* genes using orthologues from *Arabidopsis thaliana* suggests that several of these TFs may be associated with the regulation of flavonoid metabolism, including pathways linked to anthocyanin accumulation. Among the identified candidates, *OsMYB44* shows the strongest association with anthocyanin regulation. Its *Arabidopsis* orthologue, *AtMYB112*, has been reported to positively regulate anthocyanin biosynthesis and enhance tolerance to abiotic stresses such as salinity and drought by activating antioxidant and stress-responsive genes [78]. This functional similarity indicates that *OsMYB44* may contribute to anthocyanin accumulation in rice as part of a broader protective response to environmental stress.

OsMYB45 also represents a potential regulator of flavonoid-derived metabolites. Its orthologue *AtMYB60*-related TFs participate in jasmonate-mediated developmental processes and promote flavonol biosynthesis through activation of the *FLS1* gene [79]. Because flavonols and anthocyanins share common upstream biosynthetic steps within the flavonoid pathway, this relationship suggests that *OsMYB45* may indirectly influence anthocyanin biosynthesis by regulating flavonoid metabolic flux. In addition, several candidate genes including *OsMYB15*, *OsMYB61*, and *OsMYB101* are orthologous to *AtMYB97*-like regulators that control secondary metabolite pathways in *Arabidopsis* [80]. Although these TFs are primarily associated with glucosinolate metabolism and stress-responsive signaling, MYB proteins involved in secondary metabolism frequently exhibit functional diversification across plant species. Therefore, their identification through expression profiling, phylogenetic relationships, and cis-regulatory element analysis suggests that these *OsMYBs* may participate in broader phenylpropanoid or flavonoid metabolic networks that include anthocyanin biosynthesis.

Taken together, orthologue-based functional inference indicates that *OsMYB44* and *OsMYB45* are the most plausible anthocyanin-associated regulators among the candidate genes, while *OsMYB15*, *OsMYB61*, and *OsMYB101* may play supportive or indirect roles in secondary metabolite regulation. These findings are consistent with the phylogenetic placement and expression patterns of these genes, supporting their potential involvement in the transcriptional networks governing anthocyanin accumulation in rice.

3.10 Identification of Novel R2R3 OsMYBs as Putative Anthocyanin Regulators

A phylogenetic analysis was carried out of 109 *OsMYBs* along with the 20 rice *MYB* TFs documented to date [81] and well-characterized anthocyanin-activating *MYBs* from other species (Fig. 11). Based on phylogenetic clustering, several *OsMYBs* were identified as candidates for involvement in pigment biosynthesis due to their orthology with functionally validated anthocyanin regulators from other species. *OsMYB36* clustered with *DcMYB6*, *VlMYBA2*, *CmMYB6*, and *PhAN2*, all known to activate anthocyanin pathways (bootstrap = 0.7924). Similarly, *OsMYB54* showed strong orthology to *SbY1* (*Sorghum bicolor*) (bootstrap = 0.9501), and *OsMYB31* showed strong orthology to *BvMYB12* (*Beta vulgaris*) (bootstrap = 0.8543). The domain analysis also confirms these as R2R3 *OsMYB* TFs. Conversely, *OsMYBs* that were most closely related to the three previously characterized rice pigment regulators were excluded to prioritize novel candidates.

Fold-change analysis of transcriptome data identified several *OsMYB* TFs showing consistent upregulation during pericarp development and higher expression in black rice compared with mixed rice genotypes (Fig. 12). Using the selection criterion of FC > 1 for developmental induction [FC (14 vs. 7 DAF, Black)] and genotype comparisons [FC (Black vs. Mixed, 7 DAF) and FC (Black vs. Mixed, 14 DAF)], a total of 32 *OsMYB* genes were identified as candidate upregulated regulators. Among these, several genes displayed strong expression differences, including *OsMYB8* (FC 9.428, 1.404, and 2.627), *OsMYB24* (FC 1.929, 2.296, and 4.345), *OsMYB45* (FC 3.043, 3.404, and 1.747), *OsMYB60* (FC 3.203, 1.931, and 1.044), *OsMYB61* (FC 1.328, 4.148, and 4.821), *OsMYB65* (FC 1.324, 27.526, and 16.021), and *OsMYB81* (FC 2.440, 7.766, and 2.907), corresponding to FC (14 vs. 7 DAF, Black), FC (Black vs. Mixed, 7 DAF), and FC (Black vs. Mixed, 14 DAF), respectively, indicating strong developmental and genotype-associated upregulation. Notably, *OsMYB65* exhibited the highest genotype-specific expression, with FC values of 27.53 and 16.02 in Black vs. Mixed rice at 7 and 14 DAF, respectively.

The predicted bHLH-interaction motif in these *OsMYBs* was located within the R3 repeat of the PF00249 *MYB*-like DNA-binding domain, which was present in all of these *OsMYBs*. It spans roughly the residues EDEQLR...CRL...HRR, where conserved leucines (L) and arginines (R) were key for interaction. This short motif likely mediates binding to bHLH TFs in rice. Promoter analysis of these potential *OsMYB* anthocyanin candidates revealed several cis-regulatory elements associated with flavonoid and anthocyanin biosynthesis (Table 2). Notably, the *MYB* recognition element (MRE), a key motif involved in *MYB*-mediated regulation of flavonoid pathway genes, was identified in the promoters of *OsMYB44* and *OsMYB98*, suggesting their potential role in regulating structural genes of the anthocyanin biosynthetic pathway. In addition, chalcone synthase-associated motifs, including Chs-CMA1a and Chs-CMA2a, which are known to control light-induced expression of flavonoid biosynthetic genes, were detected in *OsMYB59*, *OsMYB69*, *OsMYB70*, *OsMYB101*, and *OsMYB62*, indicating possible regulation of early anthocyanin pathway enzymes such as chalcone synthase. The presence of AC-I elements, associated with phenylpropanoid metabolism and flavonoid biosynthesis, was observed in *OsMYB69* and *OsMYB70*, further supporting their potential involvement in pigment-related secondary metabolism. Additionally, *MYB* binding motifs and *MYB*-like sequences were widely distributed across promoters of genes such as *OsMYB5*, *OsMYB6*, *OsMYB25*, *OsMYB43*,

and *OsMYB81*, suggesting potential autoregulatory or cross-regulatory interactions within the *MYB* TF network. Together, the occurrence of these anthocyanin-related cis-elements in specific *OsMYB* promoters indicates that these genes may participate in the transcriptional regulation of flavonoid and anthocyanin biosynthesis in rice, serving as preliminary candidate regulators for further experimental validation.

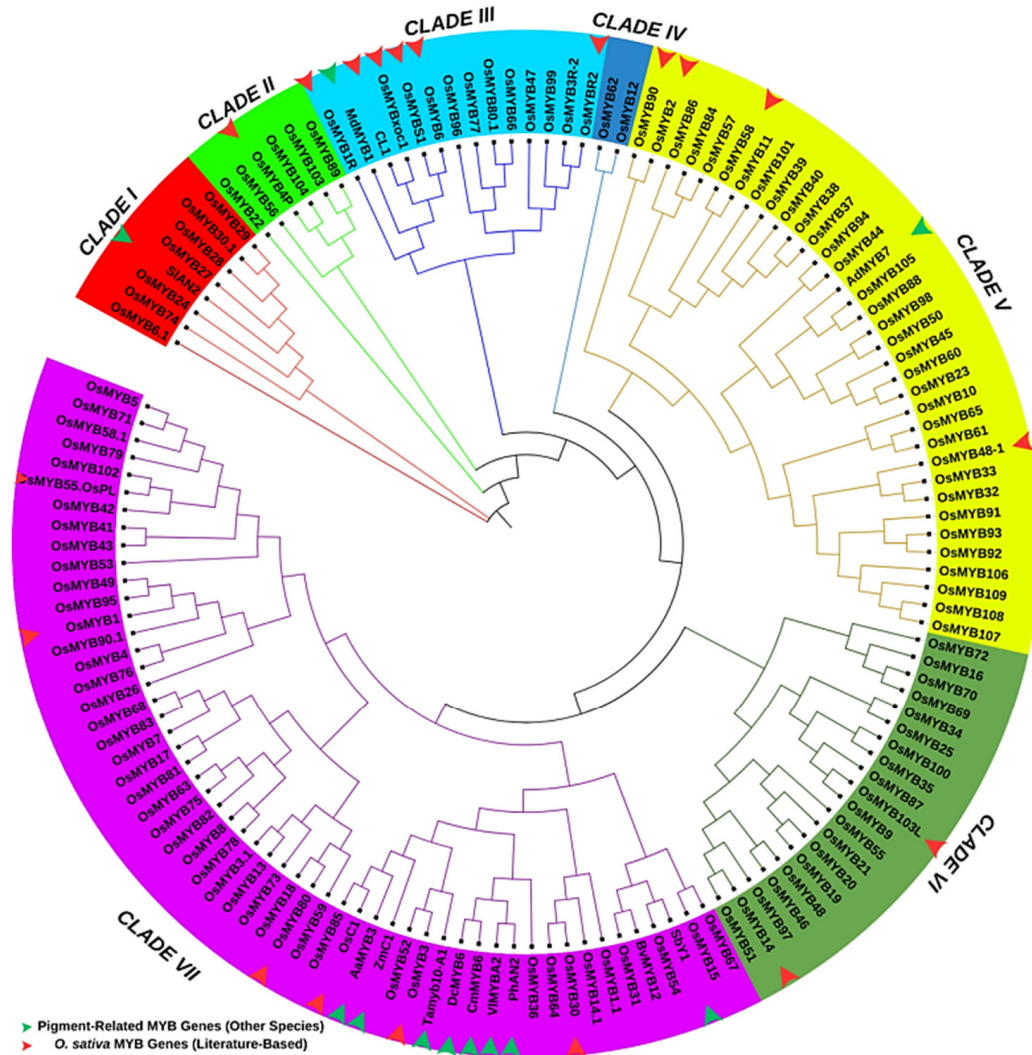


Figure 11: Circular Phylogenetic Tree of 109 *O. sativa* MYB Proteins Integrated with Known Pigment-Related *MYB* Regulators from Rice and Other Plant Species. Red arrows indicate literature-reported *O. sativa* MYB genes involved in pigmentation; green arrows highlight well-characterized anthocyanin-activating *MYB* regulators from other species. Literature-reported *OsMYBs* are labeled with their original gene IDs, while similarly identified gene IDs in our dataset are renamed in the '*OsMYB.1*' format to avoid ID overlapping.

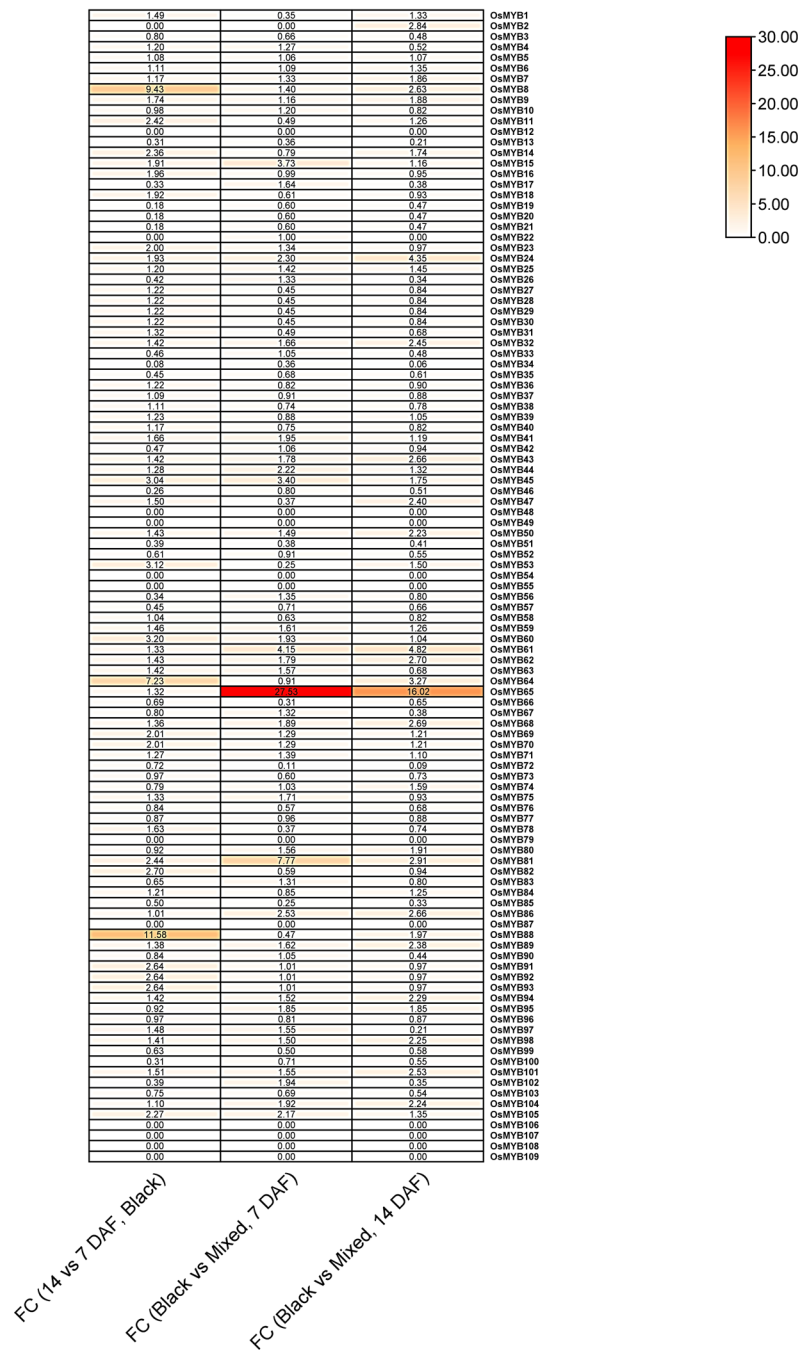


Figure 12: Expression Profiling of *OsMYB* Genes Associated with Anthocyanin Biosynthesis in Developing Rice Pericarps (GSE67987). Heatmaps illustrate fold-change (FC) patterns of 109 *OsMYB* TF genes in pericarp tissues of black rice near-isogenic lines and mixed rice genotypes 7 and 14 days after heading. Fold changes were calculated for developmental regulation in black rice [FC (14 vs. 7 DAF, Black)] and genotype comparisons [FC (Black vs. Mixed, 7 DAF) and FC (Black vs. Mixed, 14 DAF)] using expression data retrieved from the NCBI GEO dataset GSE67987. Color intensity from white to dark red represents increasing fold-change values across developmental stages and genotypic comparisons.

Table 2: Integrated summary of physicochemical characteristics, regulatory elements, and miRNA targeting 32 candidate *OsMYB* genes potentially involved in anthocyanin biosynthesis.

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
						FC (14 vs. 7 DAF, Black)	FC (Black vs. Mixed, 7 DAF)	FC (Black vs. Mixed, 14 DAF)		
<i>OsMYB5</i>	3	316	−0.644	5.89	62.48	1.079	1.057	1.073	AE-box, ARE, AT-TATA-box, CAAT-box, GA-motif, LTR, MBS, <i>MYB</i> , <i>MYB</i> -like sequence, MYC, <i>MYB</i> , RY-element, STRE, TATA-box, W box, WRE3, circadian ABRE, ARE, AT-TATA-box, CAAT-box, CGTCA-motif, DRE core, G-Box, G-box, <i>MYB</i> , <i>MYB</i> -like sequence, <i>MYB</i> , <i>MYB</i> -binding site, NON, TATA, TATA-box, TCA-element, TGACG-motif, Unnamed__1, Unnamed__4, W box, WRE3, WUN-motif, as-1, AAGAA-motif, ABRE, ABRE3a, ABRE4, ARE, AT-TATA-box, CAAT-box, G-Box, G-box, GT1-motif, LTR, MBS, <i>MYB</i> , MYC, <i>MYB</i> , <i>MYB</i> -binding site, STRE, TATA-box, TCA, TCA-element, Unnamed__1, Unnamed__4, W box ABRE, ABRE3a, ABRE4, ARE, Box 4, CAAT-box, CGTCA-motif, DRE core, G-Box, G-box, GT1-motif, GTGGC-motif, <i>MYB</i> , <i>MYB</i> -like sequence, MYC, <i>MYB</i> , P-box, STRE, TATA-box, TCA-element, TCCC-motif, TGACG-motif, Unnamed__4, W box, as-1	Osa-miR531b, Osa-miR5809b
<i>OsMYB6</i>	2	311	−0.592	7.69	50.15	1.107	1.090	1.348		
<i>OsMYB7</i>	2	358	−0.178	6.10	47.90	1.168	1.329	1.855		
<i>OsMYB8</i>	3	310	−0.511	5.41	60.33	9.428	1.404	2.627		Osa-miR444a, Osa-miR444d, Osa-miR444e, Osa-miR444b

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB9</i>	3	360	−0.647	6.13	50.48	1.738	1.157	1.877	3-AF1 binding site, ABRE, ABRE3a, ABRE4, ARE, AT-TATA-box, Box II, CAAT-box, CAT-box, CGTCA-motif, DRE core, G-Box, G-box, GATA-motif, MYB, MYC, Myc, O2-site, STRE, TATA-box, TCT-motif, TGACG-motif, Unnamed__4, Unnamed__6, as-1 AAGAA-motif, ABRE, ABRE3a, ABRE4, Box 4, CAAT-box, CCAAT-box, CGTCA-motif, DRE core, ERE, G-box, MBS, MYB, MYB recognition site, MYB, NON, STRE, Sp1, TATA-box, TCCC-motif, TGACG-motif, Unnamed__1, Unnamed__4, as-1, box S	Osa-miR1430
<i>OsMYB15</i>	3	372	−0.518	4.94	56.25	1.913	3.730	1.160	A-box, AT-TATA-box, AuxRR-core, Box 4, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, G-box, I-box, MYB, MYB, STRE, TATA, TATA-box, TATC-box, TCCC-motif, TGA-element, TGACG-motif, Unnamed__1, Unnamed__2, Unnamed__4, as-1 A-box, ABRE, ABRE3a, ABRE4, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, DRE core, G-box, GC-motif, GCN4_motif, GT1-motif, MBS, MYB, MYB-like sequence, MYC, MYB, MYB-binding site, STRE, TATA-box TCT-motif, TGA-element, TGACG-motif, Unnamed__4 W box, WRE3, as-1, box S	Osa-miR172c
<i>OsMYB24</i>	2	69	−0.441	9.59	46.36	1.929	2.296	4.345		
<i>OsMYB25</i>	3	401	−0.237	6.40	49.84	1.202	1.422	1.448		Osa-miR172c, Osa-miR5495

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB32</i>	2	277	−0.837	6.86	69.20	1.416	1.659	2.453	ABRE, ARE, CAAT-box, CAT-box, DRE core, DRE1, G-box, MYB, MYC, TATA, TATA-box, Unnamed__1, Unnamed__4, WRE3	Osa-miR171h, Osa-miR171g
<i>OsMYB41</i>	3	265	−0.286	5.73	60.22	1.657	1.945	1.186	A-box, AAGAA-motif, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, DRE core, GATA-motif, GTGGC-motif, MYB, MYB-like sequence, MYC, MYB-binding site, O2-site, Sp1, TATA-box, TGACG-motif, Unnamed__2, Unnamed__4, W box, as-1	Osa-miR5809b, Osa-miR172c
<i>OsMYB43</i>	3	299	−0.427	4.88	60.41	1.421	1.784	2.656	A-box. AT-TATA-box, Box 4, CAAT-box, CCAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, G-box, GATA-motif, LTR, MRE, MYB, MYB recognition site, MYC, STRE, Sp1, TATA, TATA-box, TGACG-motif, Unnamed__1, Unnamed__4, WRE3, WUN-motif, as-1,	
<i>OsMYB44</i>	1	229	−0.552	6.32	52.77	1.277	2.225	1.316		
<i>OsMYB45</i>	1	351	−0.486	6.07	58.76	3.043	3.404	1.747	AE-box, AT-TATA-box, CAAT-box, GT1-motif, MBS, MYB, MYC, MYB, P-box, RY-element, STRE, TATA-box, TCA-element, TCCC-motif, Unnamed__1, Unnamed__4, box S	Osa-miR164a, Osa-miR164b, Osa-miR164d, Osa-miR164e, Osa-miR166a, Osa-miR166b, Osa-miR166d, Osa-miR166f, Osa-miR166g, Osa-miR166h, Osa-miR166k, Osa-miR6256

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB47</i>	2	973	−0.593	5.05	54.86	1.502	0.369	2.399	A-box, AAGAA-motif, ABRE, AE-box, ARE, Box II, CAAT-box, CCAAT-box, CCGTCC motif, CCGTCC-box, DRE core, G-Box, G-box, GC-motif, GT1-motif, I-box, MBS, <i>MYB</i> , <i>MYB</i> recognition site, MYC, <i>MYB</i> , OCT, P-box, STRE, TATA-box, TC-rich repeats, TCT-motif, TGA-element, Unnamed__1, Unnamed__4	Osa-miR394, Osa-miR1850, Osa-miR1859, Osa-miR395b, Osa-miR395c, Osa-miR395f
<i>OsMYB50</i>	1	294	−0.589	5.68	56.09	1.431	1.493	2.228	A-box, ABRE, ARE, AT-TATA-box, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, G-Box, G-box, GATA-motif LTR, MBS, MYC, <i>MYB</i> , Myc, O2-site, STRE, Sp1, TATA-box, TC-rich repeats, TCT-motif, TGACG-motif, Unnamed__1, Unnamed__4, Unnamed__6, W box, as-1	
<i>OsMYB59</i>	3	318	−0.269	8.17	51.34	1.461	1.608	1.255	AAGAA-motif, ABRE, AT-TATA-box, AuxRR-core, Box 4, CAAT-box, DRE1, G-box, MBS, <i>MYB</i> , MYC, <i>MYB</i> , TATA, TATA-box, TC-rich repeats, Unnamed__1, Unnamed__4, WUN-motif, chs-CMA1a	

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB60</i>	1	330	−0.613	4.84	59.44	3.203	1.931	1.044	AAGAA-motif, ABRE, AT-TATA-box, AuxRR-core, Box 4, CAAT-box, CAT-box, CCAAT-box, CTAG-motif, GA-motif, GC-moti, GT1-motif, HD-Zip 1, MBS, MYB, MYB recognition site, MYB-like sequence, MYC, MYB, TATA, TATA-box, Unnamed__1, Unnamed__10, Unnamed__12, Unnamed__14, Unnamed__2, Unnamed__4, Unnamed__8, W box, WRE3, AAGAA-motif, ABRE, ABRE3a, ABRE4, ARE, ATCT-motif, AT-TATA-box, Box 4, CAAT-box, CARE, CGTCA-motif, G-box, MYB, MYB-like sequence, Myc, STRE, TATA-box, TCA-element, TGA-element, TGACG-motif, Unnamed__1, Unnamed__4, WUN-motif, as-1	Osa-miR399d
<i>OsMYB61</i>	2	248	−0.879	6.60	72.96	1.328	4.148	4.821	A-box, ABRE, ARE, CAAT-box, CAT-box, CCAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, CTAG-motif, G-box, GC-motif, GT1-motif, MYB recognition site, MYC, MYB, NON, STRE, Sp1, TATA-box, TGA-element, TGACG-motif, Unnamed__1, Unnamed__4, as-1, box S, chs-CMA2a	Osa-miR1861c, Osa-miR399d
<i>OsMYB62</i>	3	301	−0.353	6.96	42.16	1.426	1.794	2.698		

Table 2: Cont.

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB65</i>	2	290	−0.679	5.35	55.35	1.324	27.526	16.021	ABRE, ABRE3a, ABRE4, ARE, AT~TATA-box, AuxRR-core, Box 4, CAAT-box, CAT-box, DRE1, ERE, G-Box, G-box, GC-motif, GT1-motif, LTR, MSA-like, MYB, MYC, STRE, TATA-box, TC-rich repeats, TCA, TCT-motif, Unnamed__1, Unnamed__2, Unnamed__4, Unnamed__6, WRE3, WUN-motif, AT~TATA-box, Box 4, CAAT-box, CGTCA-motif, CGTCA-motif, ERE, GATA-motif, GT1-motif, MYB, MYB-like sequence, MYC, STRE, TATA-box, TCA, TCA-element, TCT-motif, TGACG-motif, Unnamed__4, W box, WRE3, as-1 A-box, AAGAA-motif, ABRE, AC-I, ARE, AT~TATA-box, Box 4, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, G-box, GT1-motif, MYB, MYB-like sequence, STRE, TATA-box, TATC-box, TCA-element, TGACG-motif, Unnamed__1, Unnamed__16, Unnamed__4, as-1, chs-CMA1a,	Osa-miR1430, Osa-miR1846b
<i>OsMYB68</i>	3	378	−0.352	6.25	58.20	1.362	1.894	2.695		
<i>OsMYB69</i>	3	318	−0.731	5.25	68.89	2.007	1.286	1.206		Osa-miR399d, Osa-miR820a, Osa-miR820b, Osa-miR820c

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB70</i>	3	294	−0.513	4.95	66.95	2.007	1.286	1.206	A-box, AAGAA-motif, ABRE, AC-I, ARE, AT-TATA-box, Box 4, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, G-box, GT1-motif, <i>MYB</i> , <i>MYB</i> -like sequence, STRE, TATA-box, TATC-box, TCA-element, TGACG-motif, Unnamed__1, Unnamed__16, Unnamed__4, as-1, chs-CMA1a	<i>Osa-miR399d</i> , <i>Osa-miR820a</i> , <i>Osa-miR820b</i> , <i>Osa-miR820c</i>
<i>OsMYB71</i>	3	564	−0.410	5.59	52.68	1.265	1.395	1.102	AAGAA-motif, ABRE, ABRE3a, ABRE4, ARE, AT-TATA-box, CAAT-box, CAT-box, G-box, MYC, RY-element, STRE, TATA-box, TCA-element, TCCC-motif, TGA-element, Unnamed__1, Unnamed__4, WRE3	<i>Osa-miR5809b</i> ,
<i>OsMYB81</i>	3	368	−0.387	5.62	50.32	2.440	7.766	2.907	A-box, AAGAA-motif, ARE, AT-TATA-box, Box 4, CAAT-box, CCAAT-box, CCGTCC motif, CCGTCC-box, G-box, <i>MYB</i> , <i>MYB</i> recognition site, MYC, RY-element, STRE, TATA-box, Unnamed__1, Unnamed__4, WUN-motif	
<i>OsMYB86</i>	3	448	−0.638	6.41	68.02	1.006	2.526	2.661	ABRE, AT-TATA-box, Box 4, CAAT-box, CAT-box, CGTCA-motif, G-box, GATA-motif, GC-motif, GT1-motif, <i>MYB</i> , <i>MYB</i> -like sequence, O2-site, OCT, RY-element, Sp1, TATA, TATA-box, TC-rich repeats, TCT-motif, TGA-element, TGACG-motif, Unnamed__1, Unnamed__2, Unnamed__4, W box, as-1	<i>Osa-miR159c</i> , <i>Osa-miR159d</i> , <i>Osa-miR159e</i> , <i>Osa-miR159a</i> , <i>Osa-miR159b</i> , <i>Osa-miR2877</i> , <i>Osa-miR820a</i> , <i>Osa-miR820b</i> , <i>Osa-miR820c</i>

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB89</i>	3	312	−0.479	5.59	57.25	1.381	1.619	2.381	ABRE, AP-1, ARE, CAAT-box, CCAAT-box, CGTCA-motif, DRE core, G-box, GATA-motif, <i>MYB</i> , <i>MYB</i> recognition site, MYC, <i>MYB</i> , <i>MYB</i> -binding site, STRE, Sp1, TATA-box, TGACG-motif, Unnamed__1, Unnamed__4, W box, WRE3, as-1 A-box, ABRE, ABRE3a, ABRE4, ARE, ATC-motif, AT~TATA-box, AuxRE, AuxRR-core, Box 4, CAAT-box, CCAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, DRE core, G-Box, G-box, GC-motif, LTR, MBS, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, <i>MYB</i> , NON, STRE, TATA-box, TCA, TCCC-motif, TGA-element, TGACG-motif, Unnamed__1, Unnamed__10, Unnamed__12, Unnamed__14, Unnamed__4, Unnamed__8, WRE3, as-1, box S AAGAA-motif, ABRE, AT~TATA-box, Box 4, Box II, CAAT-box, CAT-box, CCAAT-box, G-Box, G-box, MRE, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, O2-site, STRE, TATA-box, TC-rich repeats, Unnamed__4, Unnamed__6, W box	
<i>OsMYB94</i>	1	250	−0.492	5.97	45.85	1.418	1.524	2.294	DRE core, G-Box, G-box, GC-motif, LTR, MBS, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, <i>MYB</i> , NON, STRE, TATA-box, TCA, TCCC-motif, TGA-element, TGACG-motif, Unnamed__1, Unnamed__10, Unnamed__12, Unnamed__14, Unnamed__4, Unnamed__8, WRE3, as-1, box S AAGAA-motif, ABRE, AT~TATA-box, Box 4, Box II, CAAT-box, CAT-box, CCAAT-box, G-Box, G-box, MRE, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, O2-site, STRE, TATA-box, TC-rich repeats, Unnamed__4, Unnamed__6, W box	Osa-miR399d
<i>OsMYB98</i>	1	311	−0.388	5.65	66.07	1.406	1.504	2.255	DRE core, G-Box, G-box, GC-motif, LTR, MBS, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, <i>MYB</i> , NON, STRE, TATA-box, TCA, TCCC-motif, TGA-element, TGACG-motif, Unnamed__1, Unnamed__10, Unnamed__12, Unnamed__14, Unnamed__4, Unnamed__8, WRE3, as-1, box S AAGAA-motif, ABRE, AT~TATA-box, Box 4, Box II, CAAT-box, CAT-box, CCAAT-box, G-Box, G-box, MRE, <i>MYB</i> , <i>MYB</i> recognition site, <i>MYB</i> -like sequence, MYC, O2-site, STRE, TATA-box, TC-rich repeats, Unnamed__4, Unnamed__6, W box	

Table 2: *Cont.*

Gene	Clade	Amino Acid Length	GRAVY	I.P Value	Instability Index	Expression (FC)			Cis-Elements	miRNA Status
<i>OsMYB101</i>	3	395	−0.786	5.19	47.06	1.510	1.554	2.526	A-box, AAGAA-motif, AE-box, AT-TATA-box, CAAT-box, CCGTCC motif, CCGTCC-box, CGTCA-motif, DRE core, DRE1, ERE, G-box, MYC, MYB, TATA-box, TCT-motif, TGACG-motif, Unnamed__1, Unnamed__4, W box, WRE3, as-1 chs-CMA1a ABRE, ARE, AT1-motif, AT-TATA-box, Box 4, CAAT-box, CGTCA-motif, G-Box, Gap-box, I-box, MYC, MYB, O2-site, RY-element, STRE, TATA-box, TATC-box, TGA-elemen, TGACG-motif, Unnamed__4, as-1 AAGAA-motif, ABRE, AP-1, Box 4, CAAT-box, CAT-box, ERE, G-Box, G-box, GT1-motif, LAMP-element, MBS, MYB, MYB-like sequence, MYC, MYB, STRE, Sp1, TATA-box, TC-rich repeats, Unnamed__1, Unnamed__4, WUN-motif	Osa-miR159c, Osa-miR159d, Osa-miR159e, Osa-miR159a, Osa-miR159b, Osa-miR319a, Osa-miR319b
<i>OsMYB104</i>	2	331	−0.560	6.26	48.71	1.101	1.920	2.239	ABRE, ARE, AT1-motif, AT-TATA-box, Box 4, CAAT-box, CGTCA-motif, G-Box, Gap-box, I-box, MYC, MYB, O2-site, RY-element, STRE, TATA-box, TATC-box, TGA-elemen, TGACG-motif, Unnamed__4, as-1 AAGAA-motif, ABRE, AP-1, Box 4, CAAT-box, CAT-box, ERE, G-Box, G-box, GT1-motif, LAMP-element, MBS, MYB, MYB-like sequence, MYC, MYB, STRE, Sp1, TATA-box, TC-rich repeats, Unnamed__1, Unnamed__4, WUN-motif	Osa-miR2877
<i>OsMYB105</i>	1	286	−0.800	5.60	54.26	2.272	2.171	1.352	ABRE, ARE, AT1-motif, AT-TATA-box, Box 4, CAAT-box, CGTCA-motif, G-Box, Gap-box, I-box, MYC, MYB, O2-site, RY-element, STRE, TATA-box, TATC-box, TGA-elemen, TGACG-motif, Unnamed__4, as-1 AAGAA-motif, ABRE, AP-1, Box 4, CAAT-box, CAT-box, ERE, G-Box, G-box, GT1-motif, LAMP-element, MBS, MYB, MYB-like sequence, MYC, MYB, STRE, Sp1, TATA-box, TC-rich repeats, Unnamed__1, Unnamed__4, WUN-motif	

4 Discussion

This study aimed to identify candidate *MYB* TFs potentially involved in anthocyanin regulation in rice through an integrative genome-wide analysis. A total of 109 *OsMYB* genes were identified in *Oryza sativa* using domain-based screening. Previous studies have reported larger *MYB* families in rice, reflecting differences in search strategies and classification criteria [9–12]. By focusing on *MYB* proteins containing the conserved PF00249 domain associated with anthocyanin regulators such as *AtMYB15* and *PAP1* (*AtMYB75*), our analysis targeted *MYB* members with potential regulatory roles in pigment biosynthesis. This targeted approach is particularly relevant for exploring untapped pigmented rice germplasm in South Asia, including Sri Lankan traditional varieties that exhibit diverse grain pigmentation [82–84]. Identifying *MYB* regulators from such unexploited germplasm would enable targeted breeding of nutritionally superior cultivars, aligning with growing consumer demand for functional foods and dietary diversification [85].

Phylogenetic relationships and conserved motif analysis indicate that several *OsMYBs* cluster with experimentally validated anthocyanin regulators from other plant species [81]. Anthocyanin biosynthesis in plants is typically controlled by the *MYB*–*bHLH*–*WD40* (MBW) regulatory complex, which activates transcription of late flavonoid biosynthetic genes [4]. Several newly identified candidate *OsMYBs* showed predicted interactions with key MBW components, including the *bHLH* proteins *OsB2* and *OsRc*, and the *WD40* protein *OsTTG1*. This suggests that they may function as regulatory hubs with the capacity to modulate MBW complex assembly or stability; such network centrality, when considered alongside their conserved domain architecture and expression patterns, raises the possibility that these factors contribute to the fine-tuning of spatiotemporal anthocyanin biosynthesis rather than acting solely as primary transcriptional activators. The presence of conserved R2R3 DNA-binding motifs and similarity to characterized *MYBs* suggests that several identified *OsMYBs* may participate in similar regulatory modules controlling pigment accumulation in rice grains. Gene structure and domain organization further support the regulatory potential of these TFs. R2R3-*MYB* genes involved in anthocyanin biosynthesis across plant species commonly possess relatively simple exon–intron architectures and conserved DNA-binding domains that enable interaction with promoters of structural genes in the flavonoid pathway [86]. The conservation of these features among several *OsMYBs* identified in this study suggests evolutionary preservation of transcriptional mechanisms controlling flavonoid metabolism.

Promoter analysis suggests that many *OsMYB* genes may integrate environmental and developmental signals relevant to anthocyanin accumulation [87]. Light-responsive elements were particularly abundant, consistent with the well-established role of light signaling in activating anthocyanin biosynthesis [21,30]. The presence of additional stress- and hormone-related cis-elements suggests that pigment accumulation may be coordinated with broader physiological responses, as anthocyanins frequently function as protective antioxidants during environmental stress [35]. However, it should be noted that cis-element predictions are based solely on sequence motifs and do not necessarily indicate functional regulatory activity. Experimental validation such as promoter-reporter essays would therefore be required to confirm the regulatory relevance of these elements [88].

Expression profiling provided an additional layer of evidence for prioritizing candidate regulators. Differential expressions during grain pigmentation enabled the identification of 32 *OsMYB* genes potentially associated with anthocyanin accumulation. Among these, genes such as *OsMYB65* displayed strong expression patterns in pigmented grains, suggesting potential roles in activating flavonoid biosynthetic pathways. The prioritization of candidates based on combined phylogenetic relationships, regulatory motifs, and expression patterns strengthens the likelihood that these TFs participate in pigment regulation. Physicochemical characterization showed that most of these candidate proteins possess negative GRAVY

values, indicating hydrophilic properties typical of TFs functioning in the nucleus [89] (Table S1). Their predicted isoelectric points (pI) varied widely, suggesting structural diversity that may influence DNA-binding affinity and interactions with regulatory complexes controlling secondary metabolism.

Comparative analysis with functionally characterized *MYB* genes highlights the limited number of experimentally confirmed pigment regulators currently known in rice. To date, only a few MYBs including *OsC1*, *OsMYB3/Kala3*, and *OsPL/OsMYB55* have been demonstrated to regulate grain pigmentation [81]. The identification of additional candidate regulators in this study therefore expands the potential regulatory network underlying anthocyanin biosynthesis in rice and provides targets for future functional characterization.

Predicted interactions with several *Osa*-miRNA families further suggest an additional regulatory layer controlling MYB activity. miRNA-mediated regulation has been shown in other plant systems to fine-tune TFs involved in secondary metabolism, including flavonoid biosynthesis [90–93]. These predictions indicate that post-transcriptional control mechanisms may contribute to the spatial or environmental regulation of pigment accumulation, although experimental validation will be necessary to confirm these interactions.

Despite the integrative nature of the analysis, several methodological limitations should be considered. First, promoter cis-element analysis is descriptive and does not demonstrate functional TF binding. Second, the expression dataset used for candidate prioritization represents specific developmental stages and environmental conditions, which may not capture all regulatory contexts in which anthocyanin-related MYBs are active. Some candidate genes may therefore exhibit condition-specific or tissue-specific expression patterns that were not detected in the available data. In addition, the fold-change analysis provides an initial prioritization of candidate *OsMYB* genes that may regulate anthocyanin accumulation in black rice pericarp. However, due to the absence of replicates and statistical testing, these results cannot be interpreted as definitive evidence of differential expression. Future studies with biological replicates and experimental validation are needed to confirm the roles of these candidates in flavonoid and anthocyanin biosynthesis. Finally, phylogenetic similarity alone cannot establish functional equivalence between MYB orthologs across species.

Future studies should therefore focus on functional validation of prioritized candidates. Approaches such as transgenic overexpression, gene knockout using CRISPR/Cas systems, promoter activation assays, and co-expression analysis with structural genes of the flavonoid pathway (e.g., *DFR*, *ANS*, and *UFGT*) would help determine whether these MYBs directly regulate anthocyanin biosynthesis [94]. Validation of predicted miRNA–MYB interactions through degradome sequencing or reporter assays could further clarify post-transcriptional regulatory mechanisms [95].

Overall, this integrative genome-wide analysis identifies a prioritized set of *MYB* TFs potentially involved in anthocyanin regulation in rice. These findings provide a foundation for future functional studies and may support the development of pigment-rich rice varieties with enhanced nutritional and antioxidant properties.

5 Conclusions

This study provides a genome-wide characterization of 109 *OsMYB* TFs in rice, identifying candidate genes potentially involved in anthocyanin biosynthesis and pericarp pigmentation. Phylogenetic analysis, conserved R2R3-MYB domains, and expression profiling highlight genes that may act as key regulators, providing a framework for future functional studies to validate their roles. The analysis of miRNA involved cis-regulatory elements and comparative orthology with *Arabidopsis* and other rice species further supports their potential involvement in pigment accumulation and stress-related pathways. These candidates *OsMYBs*

represent promising targets for future genetic improvement and biofortification efforts, although their functional activity and regulatory impact on anthocyanin biosynthesis require experimental confirmation. Overall, this work lays a foundation for targeted studies to elucidate regulatory networks controlling pigmentation and may ultimately contribute to the development of nutritionally enriched and stress-resilient rice varieties.

Acknowledgement: We thank to Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU254713), for support this research work.

Funding Statement: This work supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU254713).

Author Contributions: Conceptualization, N. V. T. Jayaprada, Muhammad Hammad, S. Geekiyanage, Muhammad Shafiq, Heba I. Moamed, Haiam O. Elkatry, Mohamed El Oirdi, Ahmed Mahmoud Ismail and Abdelrahman R. Ahmed; methodology, N. V. T. Jayaprada, Muhammad Hammad, S. Geekiyanage, Muhammad Shafiq, Heba I. Moamed, Haiam O. Elkatry, Mohamed El Oirdi, Ahmed Mahmoud Ismail and Abdelrahman R. Ahmed; writing—original draft preparation, N. V. T. Jayaprada, Muhammad Hammad, S. Geekiyanage, Muhammad Shafiq, Heba I. Moamed, Haiam O. Elkatry, Mohamed El Oirdi, Ahmed Mahmoud Ismail and Abdelrahman R. Ahmed; writing—review and editing, N. V. T. Jayaprada, Muhammad Hammad, S. Geekiyanage, Muhammad Shafiq, Heba I. Moamed, Haiam O. Elkatry, Mohamed El Oirdi, Ahmed Mahmoud Ismail, Mohamed M. El-Mogy and Abdelrahman R. Ahmed. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The authors confirm that the data supporting the findings of this study are available within the article and its Supplementary Materials.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at <https://www.techscience.com/doi/10.32604/phyton.2026.078307/s1>.

References

1. Cominelli E, Tonelli C. A new role for plant R2R3-MYB transcription factors in cell cycle regulation. *Cell Res.* 2009;19(11):1231–2. [CrossRef].
2. Gonda TJ, Bishop JM. Structure and transcription of the cellular homolog (c-myb) of the avian myeloblastosis virus transforming gene (v-myb). *J Virol.* 1983;46(1):212–20. [CrossRef].
3. Wu Y, Wen J, Xia Y, Zhang L, Du H. Evolution and functional diversification of R2R3-MYB transcription factors in plants. *Hortic Res.* 2022;9:uhac058. [CrossRef]
4. Shi Y, Lu T, Lai S, Li S, Zhang L, Liu R, et al. *Rosa rugosa* R2R3-MYB transcription factors RrMYB12 and RrMYB111 regulate the accumulation of flavonols and anthocyanins. *Front Plant Sci.* 2024;15:1477278. [CrossRef].
5. Jayaprada NVT, Fukai S, Geekiyanage S. Does an R2R3-MYB transcription factor affect *Anthurium spathe* colour variation via regulation of vacuolar pH? *J Natl Sci Found Sri Lanka.* 2024;51(4):629–41. [CrossRef].
6. Gatica-Arias A, Farag MA, Häntzschel KR, Matoušek J, Weber G. The transcription factor AtMYB75/PAP1 regulates the expression of flavonoid biosynthesis genes in transgenic hop (*Humulus lupulus* L.). *Brewingscience.* 2012;65:103–11.
7. Chen Y, Zhang X, Wu W, Chen Z, Gu H, Qu LJ. Overexpression of the wounding-responsive gene AtMYB15 activates the shikimate pathway in *Arabidopsis*. *J Integr Plant Biol.* 2006;48(9):1084–95. [CrossRef].
8. Riaz B, Chen H, Wang J, Du L, Wang K, Ye X. Overexpression of maize ZmC1 and ZmR transcription factors in wheat regulates anthocyanin biosynthesis in a tissue-specific manner. *Int J Mol Sci.* 2019;20(22):5806. [CrossRef].
9. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC. Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genom.* 2012;13(1):544. [CrossRef].

10. Abdullah-Zawawi MR, Ahmad-Nizamuddin NF, Govender N, Harun S, Mohd-Assaad N, Mohamed-Hussein ZA. Comparative genome-wide analysis of WRKY, MADS-box and MYB transcription factor families in *Arabidopsis* and rice. *Sci Rep*. 2021;11(1):19678. [[CrossRef](#)].
11. Kang L, Teng Y, Cen Q, Fang Y, Tian Q, Zhang X, et al. Genome-wide identification of R2R3-MYB transcription factor and expression analysis under abiotic stress in rice. *Plants*. 2022;11(15):1928. [[CrossRef](#)].
12. Zhang HC, Gong YH, Tao T, Lu S, Zhou WY, Xia H, et al. Genome-wide identification of R2R3-MYB transcription factor subfamily genes involved in salt stress in rice (*Oryza sativa* L.). *BMC Genomics*. 2024;25(1):797. [[CrossRef](#)].
13. Zheng J, Wu H, Zhao M, Yang Z, Zhou Z, Guo Y, et al. OsMYB3 is a R2R3-MYB gene responsible for anthocyanin biosynthesis in black rice. *Mol Breed*. 2021;41(8):51. [[CrossRef](#)].
14. Glauber JW, Mamun A. Global rice market: current outlook and future prospects. Washington, DC, USA: International Food Policy Research Institute; 2025.
15. Alam M, Lou G, Abbas W, Osti R, Ahmad A, Bista S, et al. Improving rice grain quality through ecotype breeding for enhancing food and nutritional security in Asia-Pacific Region. *Rice*. 2024;17(1):47. [[CrossRef](#)].
16. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*. 2002;296(5565):92–100. [[CrossRef](#)].
17. Aylward J, Wilson AM, Visagie CM, Spraker J, Barnes I, Buitendag C, et al. IMA Genome—F19: a genome assembly and annotation guide to empower mycologists, including annotated draft genome sequences of *Ceratocystis pirilliformis*, *Diaporthe australafricana*, *Fusarium ophioides*, *Paecilomyces lecythidis*, and *Sporothrix stenoceras*. *IMA Fungus*. 2024;15(1):12. [[CrossRef](#)].
18. Zhu Z, Wang H, Wang Y, Guan S, Wang F, Tang J, et al. Characterization of the *cis* elements in the proximal promoter regions of the anthocyanin pathway genes reveals a common regulatory logic that governs pathway regulation. *J Exp Bot*. 2015;66(13):3775–89. [[CrossRef](#)].
19. Cao J, Shi F. Evolution of the *RALF* gene family in plants: gene duplication and selection patterns. *Evol Bioinform Online*. 2012;8:271–92. [[CrossRef](#)].
20. Czemmel S, Stracke R, Weisshaar B, Cordon N, Harris NN, Walker AR, et al. The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries. *Plant Physiol*. 2009;151(3):1513–30. [[CrossRef](#)].
21. Sheshadri SA, Nishanth MJ, Simon B. Stress-mediated *cis*-element transcription factor interactions interconnecting primary and specialized metabolism in planta. *Front Plant Sci*. 2016;7:1725. [[CrossRef](#)].
22. Qi Y, Feng X, Ding H, Lin D, Lan Y, Zhang Y, et al. Identification and functional analysis of strigolactone pathway genes regulating tillering traits in sugarcane. *Plant Cell Physiol*. 2025;66(2):260–72. [[CrossRef](#)].
23. Aziz RB, Wei J, Wu Q, Song S, Yang H, Chen X, et al. Characterization of main responsive genes reveals their regulatory network attended by multi-biological metabolic pathways in paclobutrazol (PAC)-modulated grape seed development (GSD) at the stone-hardening stage. *Int J Mol Sci*. 2025;26(3):1102. [[CrossRef](#)].
24. Mubeen H, Naseem A, Masood A, Raza S, Naeem N. *Cis*-acting regulatory elements and transcription factors as a key regulator in plant gene expression. *Gene Expr*. 2018;31:275–80.
25. Ezech OS, Yamamoto YY. Combinatorial effects of *Cis*-regulatory elements and functions in plants. *Rev Agric Sci*. 2024;12:79–92. [[CrossRef](#)].
26. Liu J, Xia WR, Hu YP, Yao Y, Fu SP, Duan RJ, et al. Cloning and analysis of McCWINV6 promoter from biofuel plant cassava (*Manihot esculenta* Crantz). *Adv Mater Res*. 2014;986–987:25–9. [[CrossRef](#)].
27. Kaur A, Pati PK, Pati AM, Nagpal AK. *In-silico* analysis of *cis*-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. *PLoS One*. 2017;12(9):e0184523. [[CrossRef](#)].
28. Roy S, Choudhury SR, Singh SK, Das KP. Functional analysis of light-regulated promoter region of AtPolλ gene. *Planta*. 2012;235(2):411–32. [[CrossRef](#)].
29. Gai YP, Zhao YN, Zhao HN, Yuan CZ, Yuan SS, Li S, et al. The latex protein MLX56 from mulberry (*Morus multicaulis*) protects plants against insect pests and pathogens. *Front Plant Sci*. 2017;8:1475. [[CrossRef](#)].
30. Jin Z, Chandrasekaran U, Liu A. Genome-wide analysis of the Dof transcription factors in castor bean (*Ricinus communis* L.). *Genes Genom*. 2014;36(4):527–37. [[CrossRef](#)].
31. Maqsood H, Munir F, Amir R, Gul A. Genome-wide identification, comprehensive characterization of transcription factors, *cis*-regulatory elements, protein homology, and protein interaction network of DREB gene family in *Solanum lycopersicum*. *Front Plant Sci*. 2022;13:1031679. [[CrossRef](#)].

32. Guillermo BV, Marisela RD, Rosalba TR, Reginaldo BS, Martín-Ernesto TH. Physiological function of rhamnogalacturonan lyase genes based in the analysis of *cis*-acting elements located in the promoter region. *Res J Biotechnol.* 2017;12(6):77–108.
33. Saidi A, Hajibarat Z, Hajibarat Z. Identification of responsive genes and analysis of genes with bacterial-inducible *cis*-regulatory elements in the promoter regions in *Oryza sativa* L. *Acta Agric Slov.* 2020;116(1):115–23. [[CrossRef](#)].
34. Khakdan F, Shirazi Z, Ranjbar M. Identification and functional characterization of the CVOMTs and EOMTs genes promoters from *Ocimum basilicum* L. *Plant Cell Tissue Organ Cult PCTOC.* 2022;148(2):387–402. [[CrossRef](#)].
35. Gómez-Porras JL, Riaño-Pachón DM, Dreyer I, Mayer JE, Mueller-Roeber B. Genome-wide analysis of ABA-responsive elements ABRE and CE3 reveals divergent patterns in Arabidopsis and rice. *BMC Genom.* 2007;8:260. [[CrossRef](#)].
36. Buchanan CD, Klein PE, Mullet JE. Phylogenetic analysis of 5'-noncoding regions from the ABA-responsive rab16/17 gene family of sorghum, maize and rice provides insight into the composition, organization and function of *cis*-regulatory modules. *Genetics.* 2004;168(3):1639–54. [[CrossRef](#)].
37. Wu J, Gao F, Li T, Guo H, Zhang L, Fan Y, et al. Genome-wide *cis*-regulatory element based discovery of auxin-responsive genes in higher plant. *Genes.* 2021;13(1):24. [[CrossRef](#)].
38. Chen F, Zhang J, Ha X, Ma H. Genome-wide identification and expression analysis of the Auxin-Response factor (ARF) gene family in *Medicago sativa* under abiotic stress. *BMC Genom.* 2023;24(1):498. [[CrossRef](#)].
39. Liu M, He J, He G, Zhang Y, Zhang M, Wang Y, et al. Genome-wide identification analysis of the ATP-binding cassette transporter family and expression analysis under methyl jasmonate treatment in *Panax ginseng*. *BMC Plant Biol.* 2025;25(1):565. [[CrossRef](#)].
40. Heidari P, Ahmadizadeh M, Najafi-Zarrini H. *In silico* analysis of *Cis*-regulatory elements on co-expressed genes. *J Biol Environ Sci.* 2015;9(25):1–9.
41. Ali S, Ahmad Mir Z, Bhat JA, Tyagi A, Chandrashekar N, Yadav P, et al. Isolation and characterization of systemic acquired resistance marker gene *PR1* and its promoter from *Brassica juncea*. *3 Biotech.* 2017;8(1):10. [[CrossRef](#)].
42. Rahman FU, Ahmad Khan I, Aslam A, Liu R, Sun L, Wu Y, et al. Transcriptome analysis reveals pathogenesis-related gene 1 pathway against salicylic acid treatment in grapevine (*Vitis vinifera* L). *Front Genet.* 2022;13:1033288. [[CrossRef](#)].
43. Redman J, Whitcraft J, Johnson C, Arias J. Abiotic and biotic stress differentially stimulate As-1 element activity in *Arabidopsis*. *Plant Cell Rep.* 2002;21(2):180–5. [[CrossRef](#)].
44. Yi SY, Liu J. Combinatorial interactions of two *cis*-acting elements, AT-rich regions and HSEs, in the expression of tomato Lehs23.8 upon heat and non-heat stresses. *J Plant Biol.* 2009;52(6):560–8. [[CrossRef](#)].
45. Foletta VC, Segal DH, Cohen DR. Transcriptional regulation in the immune system: all roads lead to AP-1. *J Leukoc Biol.* 1998;63(2):139–52. [[CrossRef](#)].
46. Jain A, Joshi G, Chauhan C, Das S. Abiotic stress response in plants: a *Cis*-regulatory perspective. In: Abiotic stress-mediated sensing and signaling in plants: an omics perspective. Singapore: Springer; 2018. p. 183–205.
47. Gasch P, Funderinger M, Müller JT, Lee T, Bailey-Serres J, Mustroph A. Redundant *ERF-VII* transcription factors bind to an evolutionarily conserved *cis*-motif to regulate hypoxia-responsive gene expression in *Arabidopsis*. *Plant Cell.* 2016;28(1):160–80. [[CrossRef](#)].
48. Shukla PS, Agarwal P, Gupta K, Agarwal PK. Molecular characterization of an MYB transcription factor from a succulent halophyte involved in stress tolerance. *AoB Plants.* 2015;7:plv054. [[CrossRef](#)].
49. Li R, Zhu F, Duan D. Function analysis and stress-mediated *cis*-element identification in the promoter region of VqMYB15. *Plant Signal Behav.* 2020;15(7):1773664. [[CrossRef](#)].
50. Martínez-Pastor MT, Marchler G, Schüller C, Marchler-Bauer A, Ruis H, Estruch F. The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.* 1996;15(9):2227–35. [[CrossRef](#)].
51. Xu F, Li L, Zhang W, Cheng H, Sun N, Cheng S, et al. Isolation, characterization, and function analysis of a flavonol synthase gene from *Ginkgo biloba*. *Mol Biol Rep.* 2012;39(3):2285–96. [[CrossRef](#)].
52. Sasaki K, Ito H, Mitsuhara I, Hiraga S, Seo S, Matsui H, et al. A novel wound-responsive *cis*-element, VWRE of the vascular system-specific expression of a tobacco peroxidase gene, tpoxN1. *Plant Mol Biol.* 2006;62(4–5):753–68. [[CrossRef](#)].

53. Al-Fageeh MB, Smales CM. Control and regulation of the cellular responses to cold shock: the responses in yeast and mammalian systems. *Biochem J.* 2006;397(2):247–59. [[CrossRef](#)].
54. Kumar GM, Mamidala P, Podile AR. Regulation of polygalacturonase-inhibitory proteins in plants is highly dependent on stress and light responsive elements. *Plant Omics.* 2009;2(6):238–49.
55. Dolfini D, Zambelli F, Pavesi G, Mantovani R. A perspective of promoter architecture from the CCAAT box. *Cell Cycle.* 2009;8(24):4127–37. [[CrossRef](#)].
56. Mittal S. Genome wide identification of NAC family in ricebean [dissertation]. Wagnaghat, India: Jaypee University of Information Technology; 2024.
57. Li J, Li M, Shen T, Guo Q, Zhang R, Chen Y, et al. Molecular characterization of cassava zinc finger-homeodomain (ZF-HD) transcription factors reveals their role in disease resistance. *Int J Biol Macromol.* 2024;279(Pt 1):134846. [[CrossRef](#)].
58. Zhu Y, Wang D, Yan F, Wang L, Wang Y, Li J, et al. Genome-wide analysis of HD-Zip genes in *Sophora alopecuroides* and their role in salt stress response. *Plant Genome.* 2024;17(3):e20504. [[CrossRef](#)].
59. van der Felden J, Weisser S, Brückner S, Lenz P, Mösch HU. The transcription factors Tec1 and Ste12 interact with coregulators Msa1 and Msa2 to activate adhesion and multicellular development. *Mol Cell Biol.* 2014;34(12):2283–93. [[CrossRef](#)].
60. Bangkim R, Chaudhary K, Panjabi P. Comparative analysis of sequence, chromosomal arrangement and cis-regulatory elements of SQS genes across Brassicaceae. *Int J Ecol Environ Sci.* 2024;50(4):645–57. [[CrossRef](#)].
61. Meireles-Filho ACA, Bardet AF, Yáñez-Cuna JO, Stampfel G, Stark A. Cis-regulatory requirements for tissue-specific programs of the circadian clock. *Curr Biol.* 2014;24(1):1–10. [[CrossRef](#)].
62. Qu J, Xu S, Tian X, Li T, Wang L, Zhong Y, et al. Comparative transcriptomics reveals the difference in early endosperm development between maize with different amylose contents. *PeerJ.* 2019;7:e7528. [[CrossRef](#)].
63. Filyushin MA, Anisimova OK, Shchennikova AV, Kochieva EZ. DREB1 and DREB2 genes in garlic (*Allium sativum* L.): genome-wide identification, characterization, and stress response. *Plants.* 2023;12(13):2538. [[CrossRef](#)].
64. Koul A, Sharma D, Kaul S, Dhar MK. Identification and *in silico* characterization of cis-acting elements of genes involved in carotenoid biosynthesis in tomato. *3 Biotech.* 2019;9(7):287. [[CrossRef](#)].
65. Shokri-Gharelo R, Bandehagh A, Mahmoudi B, Moti-Noparvar P. *In silico* study of cis-acting elements revealing the plastid gene involved in oxidative phosphorylation are responsive to abiotic stresses. *Acta Biol Szeged.* 2017;61(2):179–88.
66. Forde BG. AT-rich elements (ATREs) in the promoter regions of nodulin and other higher plant genes: a novel class of cis-acting regulatory element? *Results Probl Cell Differ.* 1994;20:87–103. [[CrossRef](#)].
67. Lin Z, Wu WS, Liang H, Woo Y, Li WH. The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. *BMC Genom.* 2010;11:581. [[CrossRef](#)].
68. Rajasekaran AK, Zhou Z, Prakash K, Das G, Kreibich G. Functional characterization of the cis-regulatory elements of the rat ribophorin I gene. *Nucleic Acids Res.* 1995;23(3):313–9. [[CrossRef](#)].
69. Griffith EC, Greenberg ME. Fos/AP-1 as an exemplar for the study of neuronal activity-dependent gene regulation. In: *Transcriptional regulation by neuronal activity.* Cham, Switzerland: Springer Nature; 2024. p. 3–50. [[CrossRef](#)].
70. Ebadi M, Bafort Q, Mizrachi E, Audenaert P, Simoens P, Van Montagu M, et al. The duplication of genomes and genetic networks and its potential for evolutionary adaptation and survival during environmental turmoil. *Proc Natl Acad Sci U S A.* 2023;120(41):e2307289120. [[CrossRef](#)].
71. Shukla R, Pokhriyal E, Das S. Complex interplay of tandem, segmental, whole genome duplication, and re-organization drives expansion of SAUR gene family in Brassicaceae. *Biochem Genet.* 2025:1–40. [[CrossRef](#)].
72. Wu B, Zhang X, Zhao J, Zeng B, Cao Z. Identification and analysis of miRNA-mRNA regulatory modules associated with resistance to bacterial leaf streak in rice. *BMC Genom.* 2025;26(1):207. [[CrossRef](#)].
73. Zhang L, Zheng L, Wu J, Liu Y, Liu W, He G, et al. OsCCRL1 is essential for phenylpropanoid metabolism in rice anthers. *Rice.* 2023;16(1):10. [[CrossRef](#)].
74. Kumar V, Srivastava AK, Abdelgawad H. Transcriptional and post-transcriptional regulation of plant growth, development, and stress responses. *J Plant Growth Regul.* 2025;44(4):1317–22. [[CrossRef](#)].
75. Jia Y, Wei K, Qin J, Zhai W, Li Q, Li Y. The roles of microRNAs in the regulation of rice-pathogen interactions. *Plants.* 2025;14(1):136. [[CrossRef](#)].

76. Kumar D, Mulani E, Singh BK, Dutta B, Singh A, Solanke AU, et al. Understanding the role of miRNAs in governing the drought sensitive response of a rice mega variety, Swarna at reproductive stage. *Plant Stress*. 2024;11:100302. [[CrossRef](#)].
77. Shen AQ, Lv MY, Ge YX, Zhou J, Hu ZZ, Ren XQ, et al. Zinc finger-homeodomain transcription factor: a new player in plant growth, stress response, and quality regulation. *Agronomy*. 2025;15(7):1522. [[CrossRef](#)].
78. Wang W, Hu S, Yang J, Zhang C, Zhang T, Wang D, et al. A novel R2R3-MYB transcription factor SbMYB12 positively regulates baicalin biosynthesis in *Scutellaria baicalensis* Georgi. *Int J Mol Sci*. 2022;23(24):15452. [[CrossRef](#)].
79. Wang Y, Zhou H, He Y, Shen X, Lin S, Huang L. MYB transcription factors and their roles in the male reproductive development of flowering plants. *Plant Sci*. 2023;335:111811. [[CrossRef](#)].
80. Chen Z, Lu J, Li X, Jiang D, Li Z. EMBRYONIC FLOWER 1 regulates male reproduction by repressing the jasmonate pathway downstream transcription factor MYB26. *Plant Cell*. 2024;37:koae287. [[CrossRef](#)].
81. Akhter D, Qin R, Nath UK, Eshag J, Jin X, Shi C. A rice gene, OsPL, encoding a MYB family transcription factor confers anthocyanin synthesis, heat stress response and hormonal signaling. *Gene*. 2019;699:62–72. [[CrossRef](#)].
82. Rathnathunga EUU, Padukkage D, Pushpakumari WHDU, Geekiyanage S, Dissanayake N, Senanayake SGJN, et al. Handbook on varietal diversity of Sri Lankan traditional rice (*Oryza sativa*) Volume I. Sri Lanka: National Research Council of Sri Lanka; 2014.
83. Rathnathunga EUU, Senanayake G, Dissanayake N, Seneweera S, Geekiyanage S. Development of a mini core collection from Sri Lankan traditional rice for flowering time variation. *Aust J Crop Sci*. 2016;10(9):1357–67. [[CrossRef](#)].
84. Rezk A, El-Malky M, Mohamed H, El-Beltagi H. Genetic Variability and Phenotypic Correlations Study among Grain Quality Traits and Mineral Elements Concentrations in Colored and Non-Colored Rice. *Phyton*. 2024;93(7):1733. [[CrossRef](#)].
85. Kaya C. Plant metabolic engineering for enhanced nutrition and food security during climate uncertainty. *Food Energy Secur*. 2025;14(1):e70060. [[CrossRef](#)].
86. Koralewski TE, Krutovsky KV. Evolution of exon-intron structure and alternative splicing. *PLoS One*. 2011;6(3):e18055. [[CrossRef](#)].
87. Wu R, Chen B, Jia J, Liu J. Relationship between protein, microRNA expression in extracellular vesicles and rice seed vigor. *Int J Mol Sci*. 2024;25(19):10504. [[CrossRef](#)].
88. Wong DCJ, Schlechter R, Vannozzi A, Höll J, Himmam I, Bogs J, et al. A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. *DNA Res*. 2016;23(5):451–66. [[CrossRef](#)].
89. Maier A, Schrader A, Kokkelink L, Falke C, Welter B, Iniesto E, et al. Light and the E3 ubiquitin ligase COP1/SPA control the protein stability of the MYB transcription factors PAP1 and PAP2 involved in anthocyanin accumulation in *Arabidopsis*. *Plant J*. 2013;74(4):638–51. [[CrossRef](#)].
90. Zhang J, Wang L, Xue M, Gan L, Pei X. microRNA171a regulates plant development and enhances drought stress tolerance. *Environ Exp Bot*. 2024;219:105613. [[CrossRef](#)].
91. Goswami K, Mittal D, Tripathi A, Gautam B, Sopory SK, Sanan-Mishra N. miRNA regulatory networks underlying the root–leaf synergism in salt tolerant pokkali rice. *J Plant Growth Regul*. 2023;42(10):6183–201. [[CrossRef](#)].
92. Azad MF, Dawar P, Esim N, Rock CD. Role of miRNAs in sucrose stress response, reactive oxygen species, and anthocyanin biosynthesis in *Arabidopsis thaliana*. *Front Plant Sci*. 2023;14:1278320. [[CrossRef](#)].
93. Zhang X, Ekwealor JTB, Mishler BD, Silva AT, Yu LA, Jones AK, et al. *Syntrichia ruralis*: emerging model moss genome reveals a conserved and previously unknown regulator of desiccation in flowering plants. *New Phytol*. 2024;243(3):981–96. [[CrossRef](#)].
94. He L, Tang R, Shi X, Wang W, Cao Q, Liu X, et al. Uncovering anthocyanin biosynthesis related microRNAs and their target genes by small RNA and degradome sequencing in tuberous roots of sweetpotato. *BMC Plant Biol*. 2019;19(1):232. [[CrossRef](#)].
95. Rao S, Balyan S, Bansal C, Mathur S. An integrated bioinformatics and functional approach for miRNA validation. In: *Plant gene silencing*. New York, NY, USA: Springer; 2022. p. 253–81. [[CrossRef](#)].