*plants*

MDPI

*Article*

# The Characterization of R2R3-MYB Genes in Water Lily *Nymphaea colorata* Reveals the Involvement of NcMYB25 in Regulating Anthocyanin Synthesis

**Qi Liu [1,2], Shujuan Li [1,*], Tuanjie Li [3], Qian Wei [1] and Yan Zhang [1]**

[1] Xi'an Botanical Garden of Shaanxi Province, Institute of Botany of Shaanxi Province, Shaanxi Engineering Research Centre for Conservation and Utilization of Botanical Resources, Xi'an 710061, China; bio_67@163.com (Q.L.)
[2] College of Life and Environmental Sciences, Minzu University of China, Beijing 100081, China
[3] Longcaoping Forestry Bureau of Shaanxi Province, Hanzhong 723400, China
* Correspondence: lishujuanxbg@163.com; Tel.: +86-029-61109924

**Abstract:** *Nymphaea colorata*, valued for its diverse flower colors and attractive shapes, is a popular ornamental aquatic plant. Anthocyanins provide color to flowers, and their biosynthesis is regulated by the R2R3-MYB transcription factor. In this study, we identified and analyzed the R2R3-MYB genes in *N. colorata*, focusing on their structure, evolution, expression patterns, regulatory mechanisms, and biological functions. We also investigated the role of the *NcMYB25* gene in anthocyanin biosynthesis. There were 59 R2R3-MYB genes in *N. colorata*, distributed across 14 chromosomes. Among these, 14 genes were involved in segmental duplications and 6 in tandem duplications. Multiple R2R3-MYB transcription factors appeared to play a role in biological processes in *N. colorata*, including NcMYB48 in flavonoid synthesis, NcMYB33 in lignin synthesis, NcMYB23 in cold stress response, and NcMYB54 in osmotic stress response. Additionally, we identified 92 miRNAs in *N. colorata*, with 43 interacting with 35 R2R3-MYB genes. The NcMYB25 protein is localized in the nucleus and possesses transcriptional activation activity. Overexpression of the *NcMYB25* gene in an apple pericarp resulted in anthocyanin accumulation. These findings provide insight into the evolutionary trajectory of the R2R3-MYB genes in *N. colorata* and highlight the regulatory function of the *NcMYB25* gene in anthocyanin biosynthesis.

**Keywords:** *Nymphaea colorata*; R2R3-MYB transcription factors; anthocyanins

check for updates

**Citation:** Liu, Q.; Li, S.; Li, T.; Wei, Q.; Zhang, Y. The Characterization of R2R3-MYB Genes in Water Lily *Nymphaea colorata* Reveals the Involvement of NcMYB25 in Regulating Anthocyanin Synthesis. *Plants* **2024**, *13*, 2990. https://doi.org/10.3390/plants13212990

Academic Editor: Konstantin V. Kiselev

Received: 19 September 2024
Revised: 22 October 2024
Accepted: 24 October 2024
Published: 26 October 2024

**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

The term "water lily" refers to plants in the genus *Nymphaea* within the family Nymphaeaceae. Ecologically, water lilies are divided into tropical and hardy varieties, with the former mainly found in tropical regions and the latter found in subtropical and temperate regions [1]. Currently, there are over 50 species and 1000 horticultural cultivars of water lilies [2,3]. Due to their diverse flower colors, aroma, and shapes, water lilies are often used in garden water features, indoor floral decorations, and floral arrangements. Water lilies have extensive root systems and require substantial amounts of nitrogen and phosphorus during growth. They also have a strong ability to absorb heavy metals like mercury, chromium, and manganese from water, making them valuable for aquatic ecological restoration [4]. Moreover, as one of the basal angiosperms, water lilies are important for studying early plant evolution [5]. *Nymphaea colorata* is renowned for its blue flowers. As a significant ornamental aquatic plant, breeding new varieties with different flower colors is a current focus in cultivating water lily varieties. The high-quality genome of *N. colorata* has recently been published, greatly facilitating the analysis of evolution, gene functions, and molecular regulatory networks in this species [6].

In plants, including *N. colorata*, flower coloration is often closely related to the types and content of anthocyanins in the petals. Anthocyanins are important components of plant flavonoids and are typically found in multiple plant tissues, giving them vibrant colors [7]. The color of flowers not only contributes to their ornamental value but plays a significant role in pollen dispersal and UV radiation tolerance [7,8]. Key enzymes in the anthocyanin biosynthesis pathway have been identified and analyzed in various plants. The synthesis of anthocyanins starts with the reaction between malonyl-CoA and coumaroyl-CoA catalyzed by chalcone synthase (CHS). Naringenin chalcone is then converted into anthocyanidins through the subsequent actions of chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavonoid 3′,5′-hydroxylase (F3′5′H), dihydroflavonol 4-reductase (DFR), and anthocyanin synthase (ANS) [9,10]. Finally, anthocyanidins are converted into stable and soluble anthocyanin through the action of the flavonoid-3-O-glucosyltransferase (UFGT) [11,12]. Genes affecting anthocyanin synthesis in plants are divided into two categories. One category includes the genes involved directly in the anthocyanin biosynthesis pathway, such as DFR and ANS. Overexpression of the *HvDFR* gene from *Hosta ventricosa* increased the accumulation of anthocyanins in transgenic tobacco [13], and overexpression of the *TaDFR-1* gene in wheat led to a significant accumulation of anthocyanin in transgenic *Arabidopsis* [14]. The other category includes the transcription factors that regulate the expression of key genes in the anthocyanin biosynthesis pathway, including R2R3-MYB transcription factors. These factors regulate anthocyanin synthesis by activating or inhibiting the transcription of genes related to the anthocyanin biosynthesis pathway [15].

MYB proteins are among the largest transcription factor families in plants and are involved in various biological processes [16]. MYB proteins are characterized by a highly conserved DNA-binding domain known as the MYB domain, which typically consists of no more than four repeats (R) of about 52 amino acids each, forming three alpha-helices. The second and third helices of each R form a helix-turn-helix (HTH) structure with three regularly spaced tryptophan (or hydrophobic) residues, creating a hydrophobic core in the 3D HTH structure [17]. MYB proteins are categorized into different classes based on the number and type of adjacent R: 1R-MYB, R2R3-MYB, 3R-MYB, and 4R-MYB. Among these, R2R3-MYB proteins are the most numerous and involved in the widest range of biological processes, including development, metabolism, and responses to biotic and abiotic stresses [18]. One hypothesis is that R2R3-MYB genes evolved from ancestral R1R2R3-MYB genes by losing the R1 sequence during evolution, leading to the formation of the R2R3-MYB gene family [18,19]. Many R2R3-MYB transcription factors have been characterized, participating in the regulation of primary and secondary metabolism, growth, development, and responses to biotic and abiotic stresses. Studies have shown that several R2R3-MYB transcription factors regulate flavonoid biosynthesis, such as AtMYB11, AtMYB12, and AtMYB111, which are involved in the synthesis of flavonols [20], and AtMYB75, AtMYB90, AtMYB113, and AtMYB114, which regulate anthocyanin synthesis [21].

Flower color is an eye-catching phenotype and a significant source of both ornamental and commercial value for plants [22]. The biosynthesis of anthocyanins, the basis of flower coloration, is typically regulated by multiple R2R3-MYB transcription factors. Some transcription factors regulating anthocyanin synthesis have been identified. In *Arabidopsis*, AtMYB75, AtMYB90, AtMYB113, and AtMYB114 act as positive regulators in the anthocyanin biosynthesis pathway [21,23]. Transient expression of the *MdMYB10* gene enhanced the accumulation of anthocyanin in transgenic tobacco and apples [15]. In a mature pear pericarp, the expression level of *PyMYB10* gene was positively correlated with *CHS*, CHI, *DFR*, and *ANS* genes [24]. Overexpression of the *MdMYB3* gene in tobacco activated the transcription of *CHS*, *CHI*, and *UFGT* genes, leading to the accumulation of anthocyanin in flowers [25]. Furthermore, some studies have shown that R2R3-MYB transcription factors influence the accumulation of anthocyanins by binding to the promoters of target genes and regulating their expression. The LvMYB1 transcription factor activated the transcription of the *LvANS* gene by binding to its promoter, thereby promoting the accumulation of anthocyanins [26]. With the publication of the high-quality genome of *N. colorata*, some

key enzymes in the anthocyanin biosynthesis pathway have been identified. However, the R2R3-MYB transcription factor, an important regulator in anthocyanin synthesis, its quantity, types, and how it affects anthocyanin synthesis remains unclear.

In this study, we first identified the R2R3-MYB gene family members in *N. colorata* based on high-quality genomic data and then analyzed their structures, evolution, expression patterns, and regulatory networks. Finally, we focused on the biological functions of the NcMYB25 transcription factor in anthocyanin biosynthesis and explored the molecular mechanisms underlying flower color formation in *N. colorata*.

## 2. Results

### *2.1. Identification and Physicochemical Properties of R2R3-MYB Proteins in N. colorata*

A total of 59 R2R3-MYB genes were identified in the *N. colorata* genome, distributed across 14 chromosomes. Chromosome 1 harbored the highest number, with 8 genes, followed by chromosomes 2 and 10 with 7 genes each, while chromosomes 6 and 14 contained only 1 gene each (Figure 1). The amino acid number of *N. colorata* R2R3-MYB proteins ranged from 142 to 513, with molecular weights from 16.39 to 55.06 kDa. The NcMYB4 protein possessed the largest molecular weight, and the NcMYB21 protein possessed the smallest. Based on theoretical isoelectric points (pI), the R2R3-MYB proteins were categorized into acidic (32 proteins) and basic (27 proteins). All R2R3-MYB proteins in *N. colorata* were hydrophilic (Table S1).

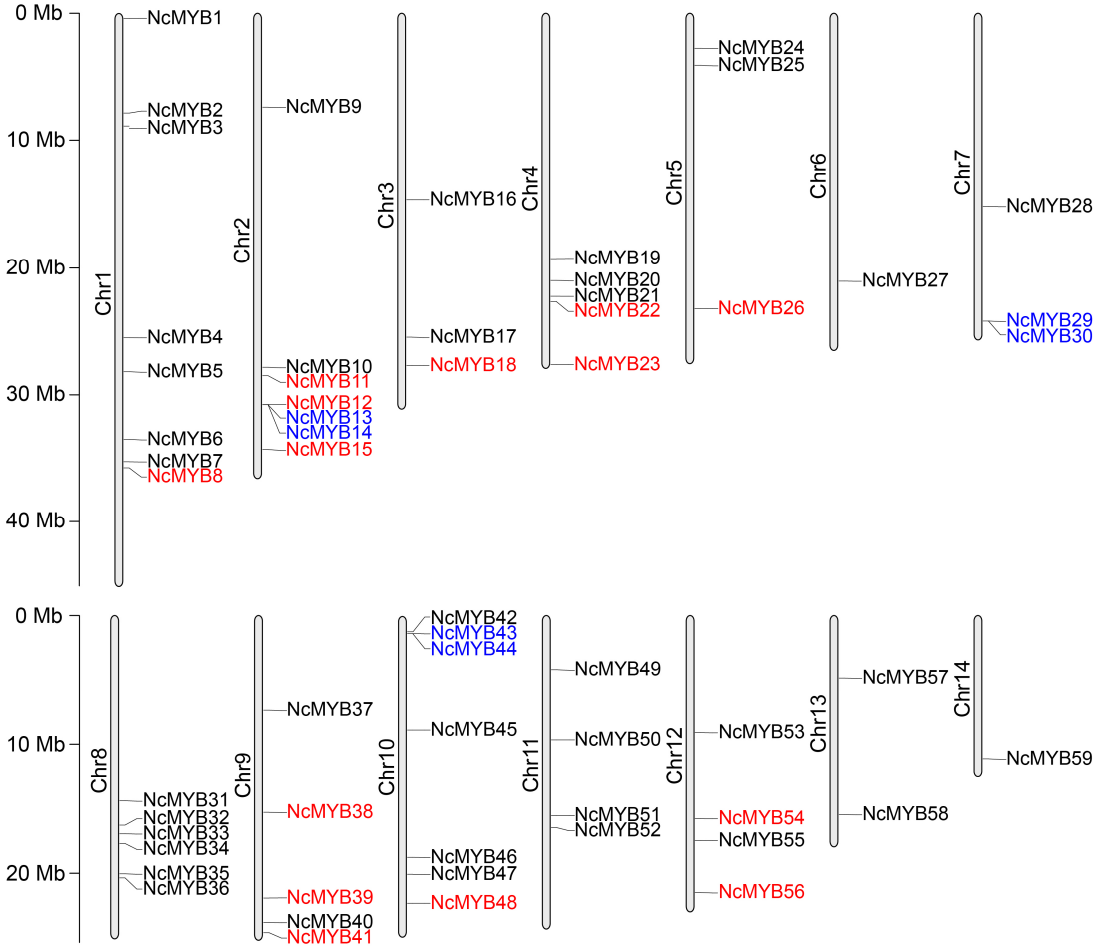



**Figure 1.** Distribution of R2R3-MYB genes on the chromosomes in *N. colorata*. The scale on the left side of the image represented the length of the chromosomes. Genes labeled in red indicate segmental duplication genes, while genes labeled in blue indicate tandem duplication genes.

### 2.2. Structural Analysis of R2R3-MYB in *N. colorata*

To analyze the structure of the R2R3-MYB gene, exon–intron analysis was performed. The R2R3-MYB genes in *N. colorata* showed generally fewer introns, with *NcMYB58* gene possessing the most at four introns. Forty R2R3-MYB genes contained two introns, and *NcMYB32* and *NcMYB47* genes possessed no intron (Figure 2A). To analyze the protein structure of R2R3-MYB in *N. colorata*, a MEME analysis was performed. Motif2, motif3, and motif5 were located within the R2 domain, while motif1 and motif4 were located in the R3 domain. Motif3 was present in all R2R3-MYB proteins in *N. colorata*, motif1 in 56 members, motif2 in 54, motif4 in 36, and motif5 in 39. Other motifs were found in only a few proteins (Figures 2B and S1).

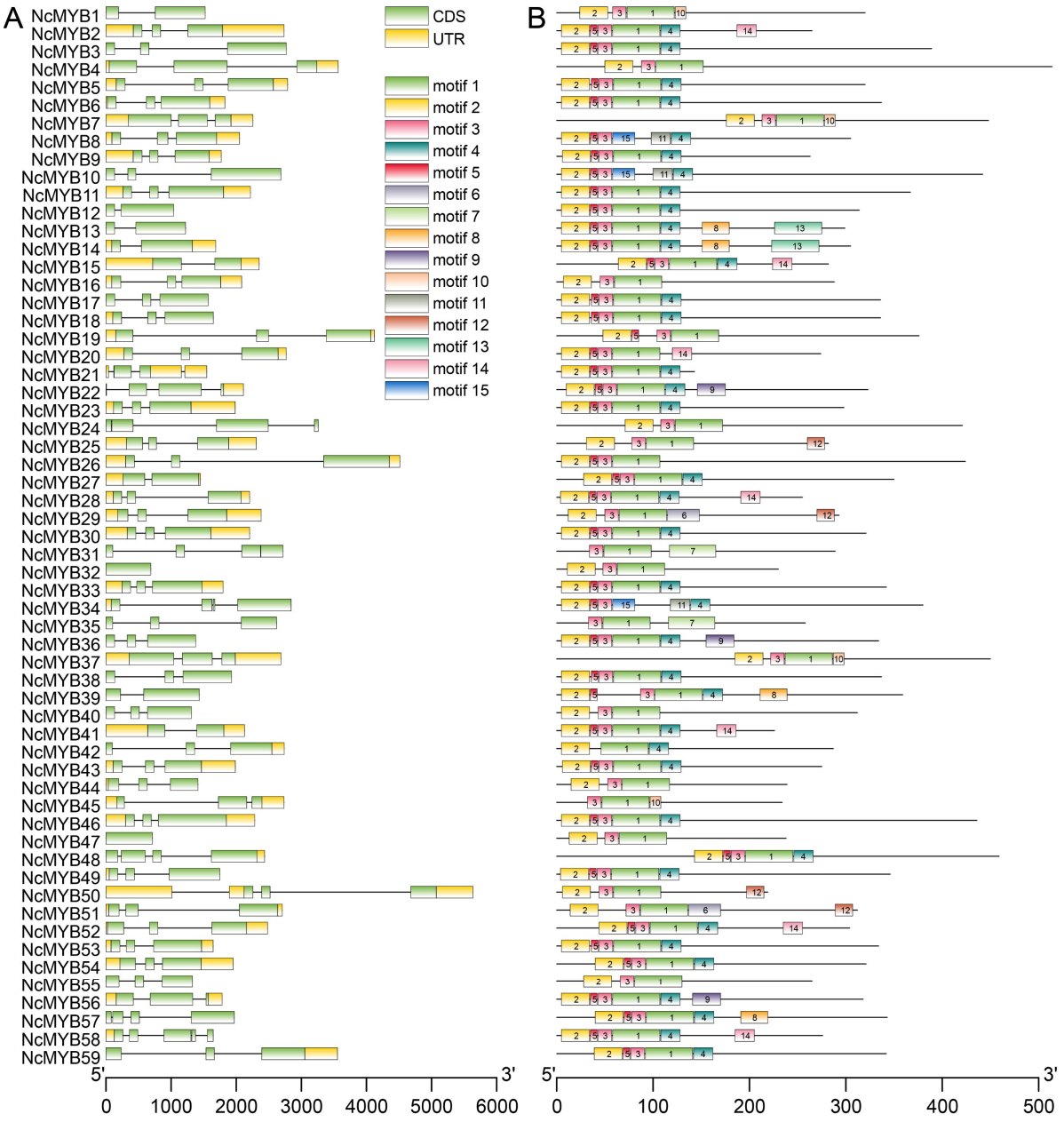



**Figure 2.** Structural analysis of R2R3-MYB in *N. colorata*. (**A**) Gene structure of R2R3-MYB genes. The scale bar at the bottom represented gene length. Yellow rectangles indicate untranslated regions (UTR), while green rectangles represent coding sequence (CDS). (**B**) The conserved motif of R2R3-MYB transcription factors. The scale bar at the bottom represents the number of amino acids. Each rectangle represents a motif, with detailed sequences provided in Figure S1.

### 2.3. Multiple Sequence Alignment and Phylogenetic Analysis

To investigate the phylogeny of R2R3-MYB proteins in *N. colorata*, multiple sequence alignments and a phylogenetic tree were constructed. The conserved R2 and R3 domains were present in all proteins in *N. colorata* (Figure S2). A phylogenetic tree was constructed using all R2R3-MYB protein sequences in *N. colorata* and partial R2R3-MYB protein sequences in *Arabidopsis*. The majority of R2R3-MYB proteins in *N. colorata* were correctly classified, with the highest numbers in groups S4 and S14 (Figure 3). There were no members in groups S10 and S12 in *N. colorata*. Using an alternative classification rule, the R2R3-MYB proteins in *N. colorata* were divided into 10 groups, with the most members in groups VIII-D and VIII-E-e2 and the fewest in group VII. Proteins in groups VIII-E-e1, III, IV, V, and VII clustered together, suggesting closer evolutionary relationships and potential functional similarities (Figure 3).

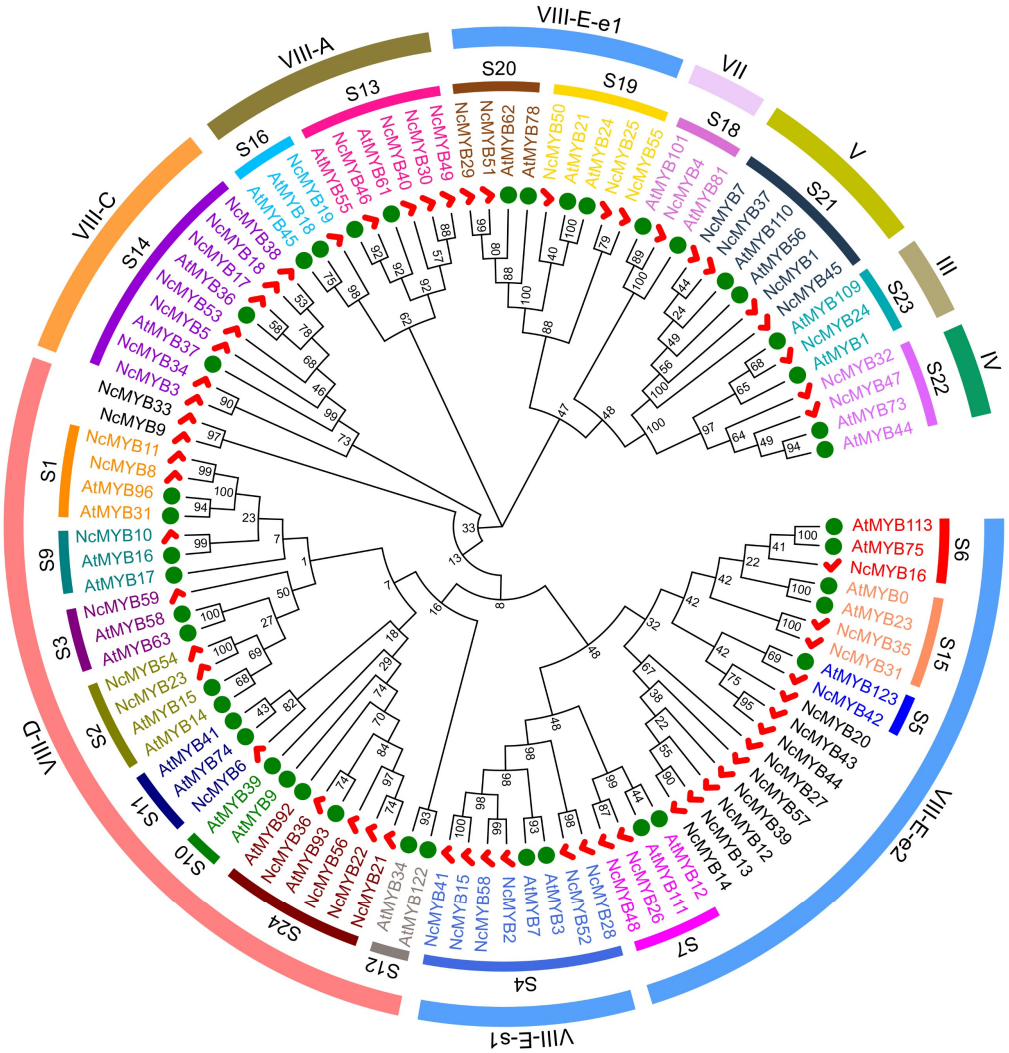



**Figure 3.** Phylogenetic analysis of R2R3-MYB proteins in *N. colorata*. The R2R3-MYB proteins of *N. colorata* were classified based on two classification schemes used for R2R3-MYB proteins in *Arabidopsis*. Green circles represent R2R3-MYB proteins in *Arabidopsis*, while red checks represent R2R3-MYB proteins in *N. colorata*.

### 2.4. Evolutionary Analysis of R2R3-MYB Genes in N. colorata

To analyze the evolution of the R2R3-MYB gene family, a collinearity analysis was performed. The collinearity analysis revealed that 14 genes underwent segmental duplication and 6 genes underwent tandem duplication (Figures 1 and 4A). The gene duplication event between *NcMYB8* and *NcMYB11* genes likely derived from a recent whole genome

duplication (WGD) event, while duplications between *NcMYB23* and *NcMYB54* genes likely occurred during an earlier WGD event (Figure S3 and Table S2).

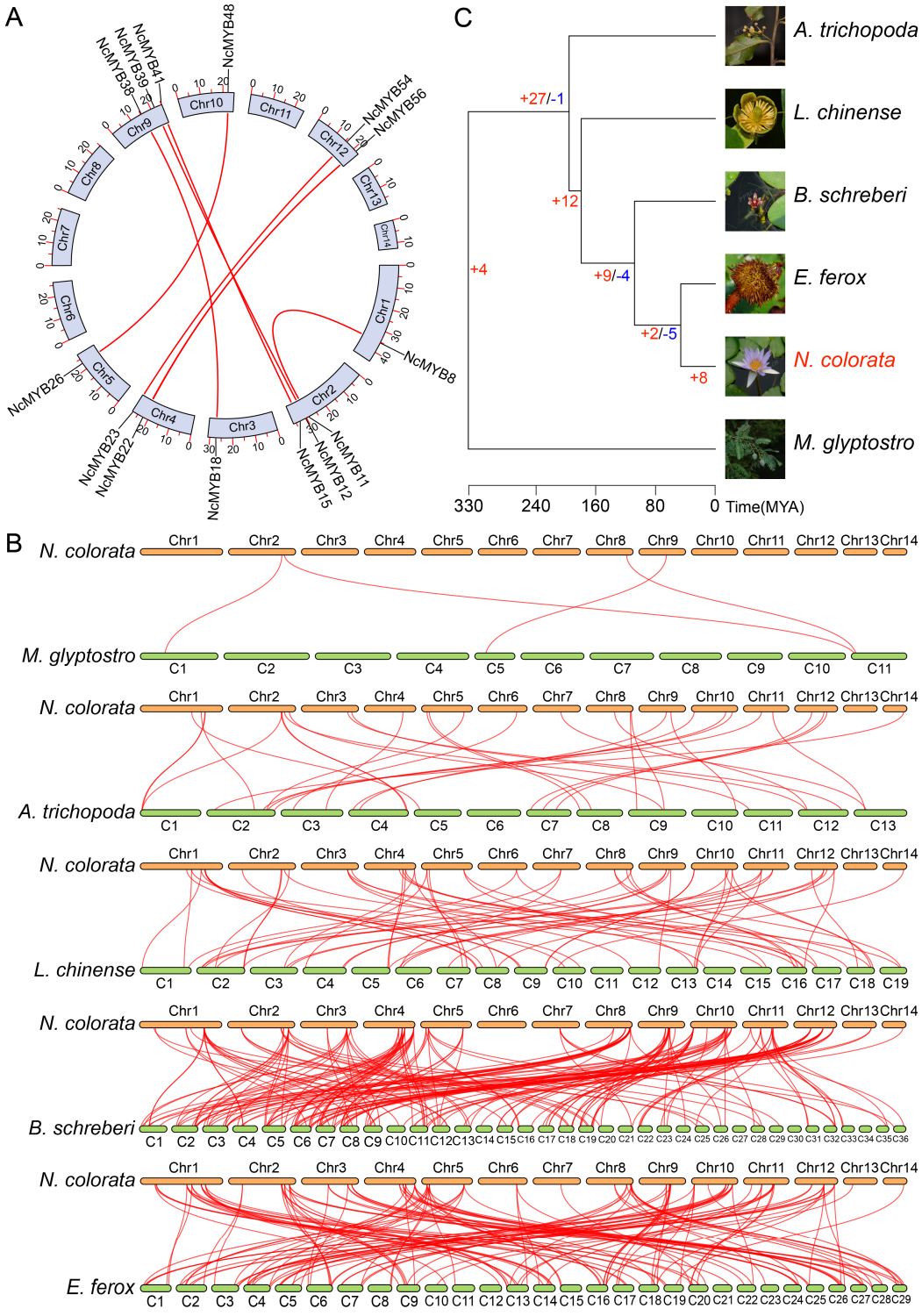



**Figure 4.** Evolutionary analysis of R2R3-MYB genes in *N. colorata*. (**A**) Segmental duplication genes of R2R3-MYB genes in *N. colorata*. (**B**) Collinearity analysis between *N. colorata* and five other species. (**C**) Evolutionary history of R2R3-MYB genes in *N. colorata*. Species trees and divergence times between species were obtained from the TimeTree website. Red-labeled numbers indicate newly evolved genes, while blue numbers indicate lost genes. Species illustrations were from the Plant Photo Bank of China (PPBC).

To analyze the evolutionary history of the R2R3-MYB gene family in *N. colorata*, a collinearity analysis was executed between *N. colorata* and *M. glyptostroboides*, *A. trichopoda*, *L. chinense*, *B. schreberi*, and *E. ferox*, respectively. The number of homolog genes was correlated with the phylogenetic relationships and the chromosome numbers of the genomes analyzed (Figure 4B,C).

### *2.5. Cis-Acting Elements Analysis of R2R3-MYB Genes in N. colorata*

To explore their potential biological functions, the *cis*-acting elements in R2R3-MYB gene promoters were analyzed. The promoters of R2R3-MYB genes in *N. colorata* contained the *cis*-acting elements responsive to hormone signals and environmental stress. ABA-responsive *cis*-acting elements were found in 49 R2R3-MYB gene promoters, auxin-responsive *cis*-acting elements in 30, gibberellin-responsive *cis*-acting elements in 37, MeJA-responsive *cis*-acting elements in 46, and SA-responsive *cis*-acting elements in 29 (Figure 5). Most promoters of R2R3-MYB genes in *N. colorata*, except those of *NcMYB50* and *NcMYB57* genes, possessed more than one type of hormone-responsive element. Light-responsive *cis*-acting elements were present in all promoters. Additionally, 37 gene promoters possessed drought stress-responsive elements, and 23 genes possessed cold stress-responsive elements. Fourteen gene promoters possessed *cis*-acting elements responsive to both cold and drought stress, indicating possible involvement in the responses to drought and cold stress (Figure 5).

| Gene name | abscisic acid responsiveness | auxin responsiveness | gibberellin responsiveness | MeJA responsiveness | salicylic acid responsiveness | anaerobic induction | drought responsiveness | defense and stress responsiveness | light responsiveness | low-temperature responsiveness |
|---|---|---|---|---|---|---|---|---|---|---|
| NcMYB1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 6 | |
| NcMYB2 | | | 2 | | 1 | | 2 | | 3 | 1 |
| NcMYB3 | 2 | 1 | 1 | 4 | 1 | 1 | | 1 | 4 | |
| NcMYB4 | 2 | | 2 | 2 | | | | 1 | 4 | 1 |
| NcMYB5 | | | 1 | 8 | 3 | | 1 | | 3 | |
| NcMYB6 | 12 | | 2 | 6 | | 2 | 1 | | 12 | |
| NcMYB7 | | 1 | 3 | | | 2 | 1 | 1 | 6 | 1 |
| NcMYB8 | 2 | 1 | 2 | 4 | 1 | 2 | 1 | | 2 | 2 |
| NcMYB9 | 4 | | 4 | 2 | 1 | 2 | | | 5 | |
| NcMYB10 | | 2 | 2 | 4 | 1 | 11 | 1 | 1 | 3 | |
| NcMYB11 | 3 | 2 | 1 | 8 | | 2 | | | 10 | |
| NcMYB12 | 4 | 2 | 1 | 2 | 2 | 3 | 2 | 2 | 6 | 1 |
| NcMYB13 | 1 | | 1 | 2 | | 2 | 1 | | 1 | |
| NcMYB14 | 5 | | | 8 | | 2 | | 1 | 6 | |
| NcMYB15 | 2 | 1 | | | | 3 | 2 | | 2 | 1 |
| NcMYB16 | 8 | | 2 | 2 | 2 | 1 | | 1 | 9 | |
| NcMYB17 | 2 | | 1 | 4 | | 3 | | | 5 | |
| NcMYB18 | 5 | | 3 | | | 2 | 2 | | 7 | 1 |
| NcMYB19 | 1 | | 2 | 6 | 1 | 2 | 2 | | 1 | 1 |
| NcMYB20 | 8 | 2 | | 8 | | 1 | 4 | | 8 | 1 |
| NcMYB21 | 2 | 1 | | 6 | | 4 | 1 | 1 | 4 | 1 |
| NcMYB22 | 1 | | 1 | 4 | 1 | 5 | 1 | | 4 | |
| NcMYB23 | 4 | | | 6 | 1 | 6 | | | 7 | |
| NcMYB24 | 3 | | | 6 | 2 | 3 | 1 | | 12 | |
| NcMYB25 | 1 | 1 | 2 | 2 | 1 | 1 | | | 2 | 2 |
| NcMYB26 | 3 | 2 | 3 | 4 | | 2 | | | 7 | |
| NcMYB27 | 3 | | 3 | | 1 | 1 | 2 | 1 | 3 | |
| NcMYB28 | 3 | | 1 | 8 | | 3 | 1 | 1 | 6 | |
| NcMYB29 | 5 | | | 4 | 3 | 2 | 3 | 2 | 4 | 1 |
| NcMYB30 | 2 | | 2 | 2 | | | 1 | | 3 | |

| Gene name | abscisic acid responsiveness | auxin responsiveness | gibberellin responsiveness | MeJA responsiveness | salicylic acid responsiveness | anaerobic induction | drought responsiveness | defense and stress responsiveness | light responsiveness | low-temperature responsiveness |
|---|---|---|---|---|---|---|---|---|---|---|
| NcMYB31 | | 2 | | 10 | 1 | 2 | 3 | 1 | 4 | |
| NcMYB32 | | 3 | 1 | 2 | | 1 | | | 6 | 1 |
| NcMYB33 | 6 | | | 2 | | 4 | 1 | | 10 | 2 |
| NcMYB34 | 7 | 1 | 1 | 4 | 1 | | | 1 | 10 | 1 |
| NcMYB35 | 5 | 2 | | 2 | | 2 | | 2 | 9 | 2 |
| NcMYB36 | 3 | | 1 | 4 | | 1 | 1 | | 4 | |
| NcMYB37 | 1 | 1 | 1 | 4 | 2 | 1 | 1 | | 3 | |
| NcMYB38 | 5 | 1 | | 4 | | 1 | | 1 | 8 | |
| NcMYB39 | 1 | 1 | | 4 | | 3 | 1 | | 5 | |
| NcMYB40 | 7 | | 2 | 2 | | 1 | 2 | | 7 | |
| NcMYB41 | | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 2 |
| NcMYB42 | 3 | | | 4 | | 4 | | | 7 | |
| NcMYB43 | 3 | 1 | | | | 1 | 2 | 1 | 4 | |
| NcMYB44 | 7 | 2 | | 6 | | | | | 10 | |
| NcMYB45 | 5 | 2 | | | 4 | 1 | 2 | 1 | 8 | |
| NcMYB46 | 1 | | 1 | | 2 | 2 | | | 4 | |
| NcMYB47 | 1 | 1 | 3 | | | 3 | 1 | 1 | 6 | |
| NcMYB48 | 3 | 2 | 3 | 6 | 3 | 3 | 3 | | 3 | |
| NcMYB49 | 2 | | | 6 | 1 | | | | 2 | 1 |
| NcMYB50 | 1 | | | | | 1 | 1 | | 2 | |
| NcMYB51 | 1 | 1 | 1 | | | 4 | | | 3 | 2 |
| NcMYB52 | 3 | | | 8 | 1 | | 1 | | 7 | |
| NcMYB53 | | 1 | 2 | 2 | | 4 | | | 1 | |
| NcMYB54 | 5 | | 3 | 8 | 2 | 4 | | | 5 | 1 |
| NcMYB55 | 10 | 1 | 1 | 2 | 1 | 1 | 1 | | 10 | 2 |
| NcMYB56 | 1 | 1 | | | 2 | 4 | | | 5 | 1 |
| NcMYB57 | 6 | | | | | 1 | 1 | 1 | 9 | 1 |
| NcMYB58 | | 2 | 2 | 2 | | 4 | 1 | | 5 | |
| NcMYB59 | | | | 4 | 1 | 2 | 1 | | 6 | |

**Figure 5.** The *cis*-acting elements of the R2R3-MYB gene promoter. Purple labels indicate hormone-related *cis*-acting elements, red labels denote environmental stress-related *cis*-acting elements, and numbers represent the quantity of *cis*-acting elements. The colored boxes in the figure represent the number of *cis*-acting elements, with the gradient from green to orange indicating an increasing number.

### 2.6. Expression Patterns Analysis of R2R3-MYB Genes in N. colorata

An expression analysis revealed significant variation in R2R3-MYB genes in *N. colorata* across different tissues. Notably, several genes exhibited tissue-specific expression, suggesting specialized functions. *NcMYB50* and *NcMYB54* genes showed high expression in stamens, sepals, and petals, indicating a role in flower organ development. *NcMYB10* and *NcMYB41* genes were expressed in multiple tissues, with higher expression in reproductive organs and leaves, suggesting roles in both vegetative and reproductive growth. *NcMYB17* and *NcMYB57* genes exhibited strong root-specific expression, indicating roles in root biology (Figure 6A). In *Nymphaea White colorata*, the *NcMYB11* gene was highly expressed in sepals and petals. The *NcMYB50* gene was highly expressed in sepals, petals, and carpels, especially in petals, suggesting its role in flower development. *NcMYB10*, *NcMYB39*, and *NcMYB54* genes were notably expressed in carpels, implying specific roles in carpel development or function. *NcMYB2*, *NcMYB23*, and *NcMYB41* genes were expressed in multiple tissues, indicating involvement in various physiological functions (Figure 6B).

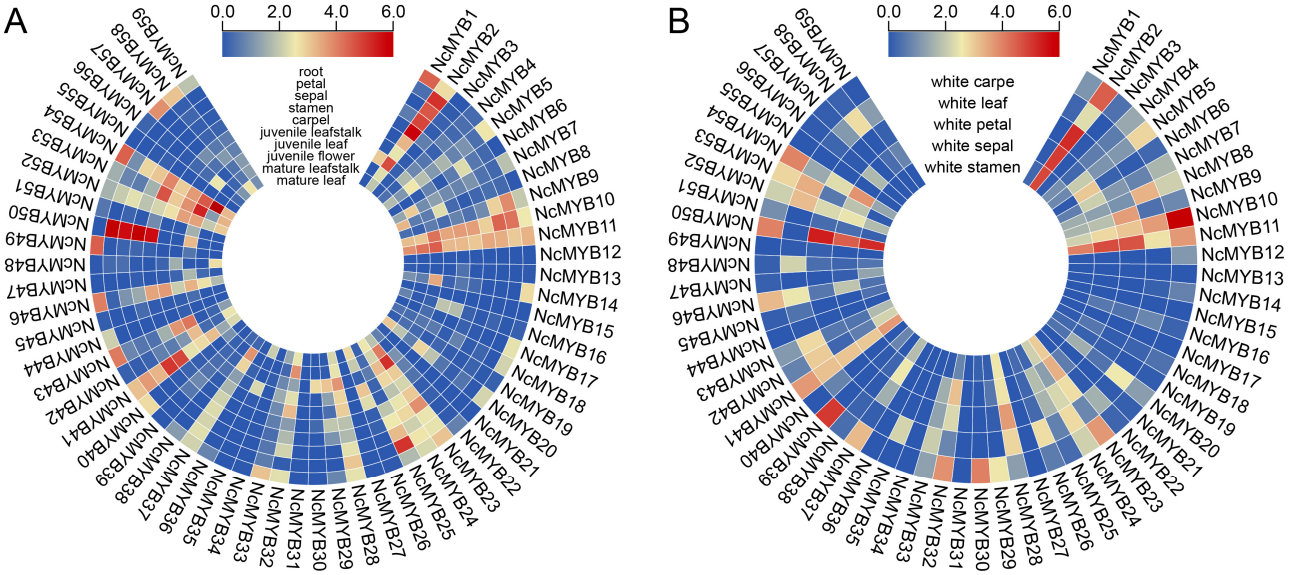



**Figure 6.** Expression pattern of R2R3-MYB genes in different tissues based on transcriptome data. (**A**) Expression patterns of R2R3-MYB genes in *N. colorata* petals. (**B**) Expression patterns of R2R3-MYB genes in *White colorata* petals.

To further validate the expression patterns, a qRT-PCR analysis of R2R3-MYB genes in roots, stems, leaves, petioles, and petals was performed. The qRT-PCR results showed higher expression levels of most R2R3-MYB genes in petioles and stems, and lower expression in leaves and petals. In petiole tissues, *NcMYB26* and *NcMYB43* genes were highly expressed, while *NcMYB2* and *NcMYB58* genes were low. In stems and petals, *NcMYB50* and *NcMYB53* genes showed the highest expression levels, with *NcMYB53* genes highest in leaves (Figure 7).

### 2.7. PPI Analysis of R2R3-MYB Proteins in N. colorata

To elucidate the complex biological processes involving R2R3-MYB proteins in *N. colorata*, a PPI analysis was conducted. The PPI network included 56 nodes and 374 edges, involving various complex biological processes (Figure S4). Two core modules, Cluster 1 and Cluster 2, were identified using the MCODE plugin in Cytoscape v3.8.2 software. Cluster 1, with 19 nodes and 113 edges, was associated with flavonoid biosynthesis (Figure 8A). MYB114, TT8, MYB75, and LDOX regulated the synthesis of flavonoids and anthocyanins, while F3H and FLS were key enzymes in the flavonoid biosynthesis pathway. Additionally, the MYB32, MYB52, and MYB86 transcription factors regulated flavonoid synthesis (Figure 8A). Cluster 2, with 12 nodes and 52 edges, was related to plant growth, secondary

metabolism, and stress responses (Figure 8B). MYB103, MYB43, and MYB58 participate in the regulation of cell wall components and lignin synthesis, while LBD15, LBD30, and BLH6 are involved in root, shoot, and flower organ formation. NAC30, NAC43, and NAC66 are associated with responses to drought and cold stress (Figure 8B). These findings suggest that R2R3-MYB proteins in *N. colorata* might regulate flavonoid biosynthesis, development regulation, and stress response.

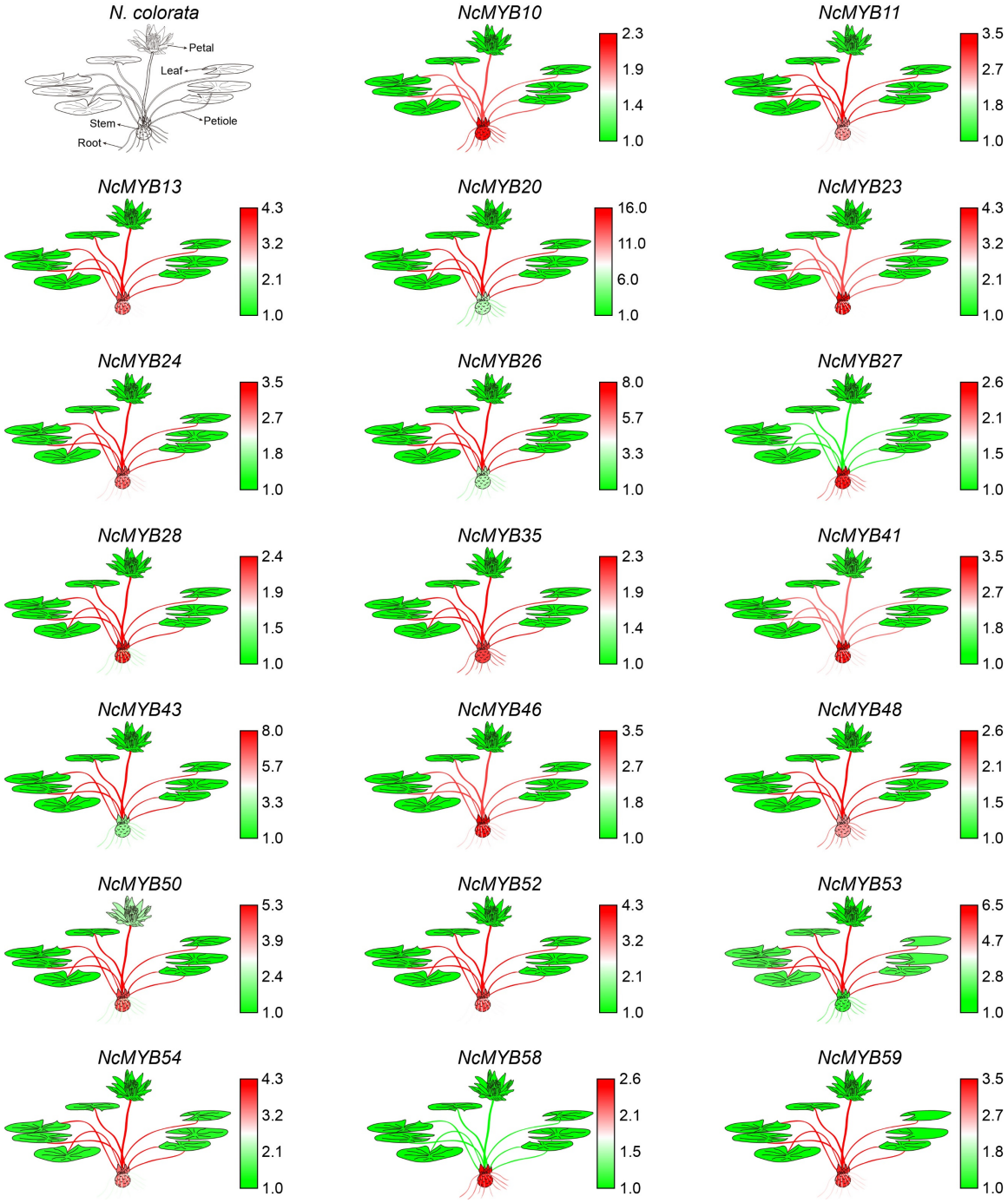



**Figure 7.** Expression levels of R2R3-MYB genes in *N. colorata* based on qRT-PCR analysis.

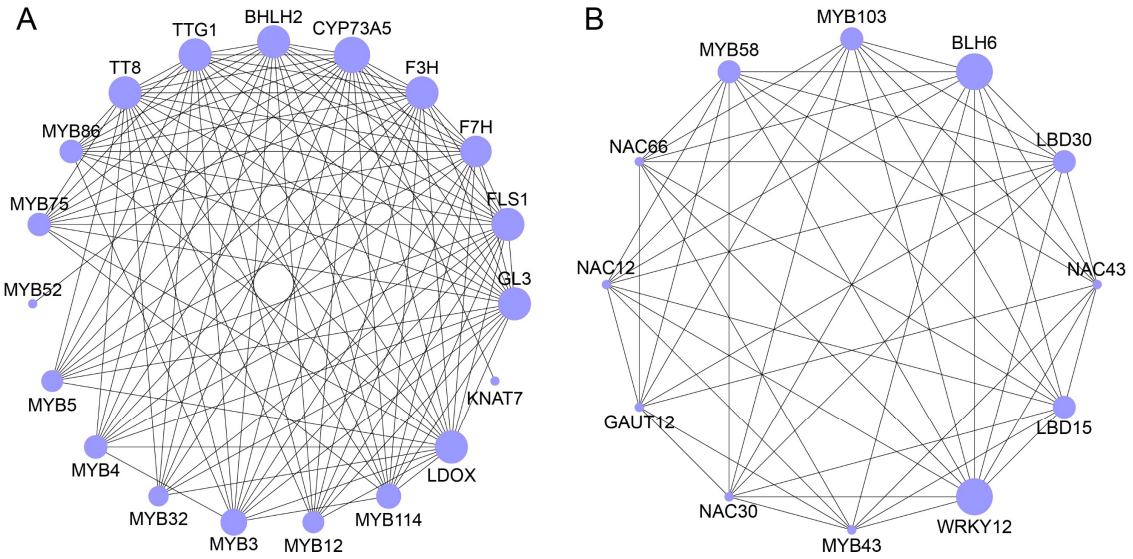



**Figure 8.** Core network of PPI analysis for R2R3-MYB in *N. colorata*. (**A**) Cluster 1 module. (**B**) Cluster 2 module.

### *2.8. GO Functional Analysis of R2R3-MYB Proteins in N. colorata*

To analyze the functions of R2R3-MYB proteins in *N. colorata*, a GO functional annotation analysis was performed. A total of 27 R2R3-MYB proteins in *N. colorata* were annotated in the GO database, encompassing multiple terms across cellular component, biological process, and molecular function categories (Figure 9 and Table S3). All R2R3-MYB proteins in *N. colorata* are located in the nucleus and possess DNA-binding activity. The R2R3-MYB proteins in *N. colorata* likely participate in various biological processes, including stress responses, growth regulation, and secondary metabolism. *NcMYB23*, *NcMYB48*, *NcMYB54*, and *NcMYB59* genes may respond to cold, freezing, osmotic, and salt stresses. In growth regulation, four genes affect root growth, and three genes influence flower development (Figure 9). For secondary metabolism, NcMYB48 regulate flavonoid synthesis, NcMYB33 regulate lignin synthesis, and NcMYB15, NcMYB18, NcMYB41, and NcMYB58 are involved in cinnamic acid synthesis (Figure 9).

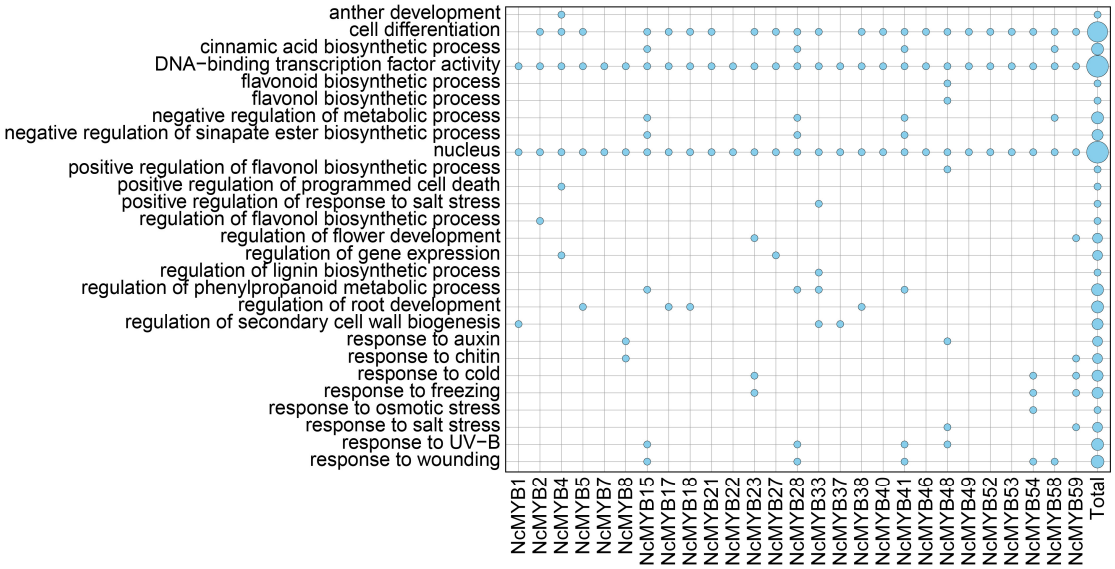



**Figure 9.** GO functional analysis of R2R3-MYB proteins in *N. colorata*. The *x*-axis represents R2R3-MYB proteins, while the *y*-axis represents GO terms. Larger circles indicate a greater number of proteins.

### *2.9. Regulation of R2R3-MYB Gene Expression by miRNAs in N. colorata*

To investigate the regulatory mechanisms of the R2R3-MYB gene in *N. colorata*, miRNAs were identified in *N. colorata*, and their regulatory relationships with R2R3-MYB genes were analyzed. A total of 92 miRNAs were identified in *N. colorata*, including 72 conserved miRNAs and 20 non-conserved miRNAs (Figure 10A and Table S4). All conserved miRNAs belong to 22 families, with the highest number of members in the MIR165, MIR157, and MIR167 families, possessing 10, 8, and 6 members, respectively (Figure 10A and Table S4). The precursor sequences of all miRNAs exhibit typical stem–loop structure (Figure 10B).

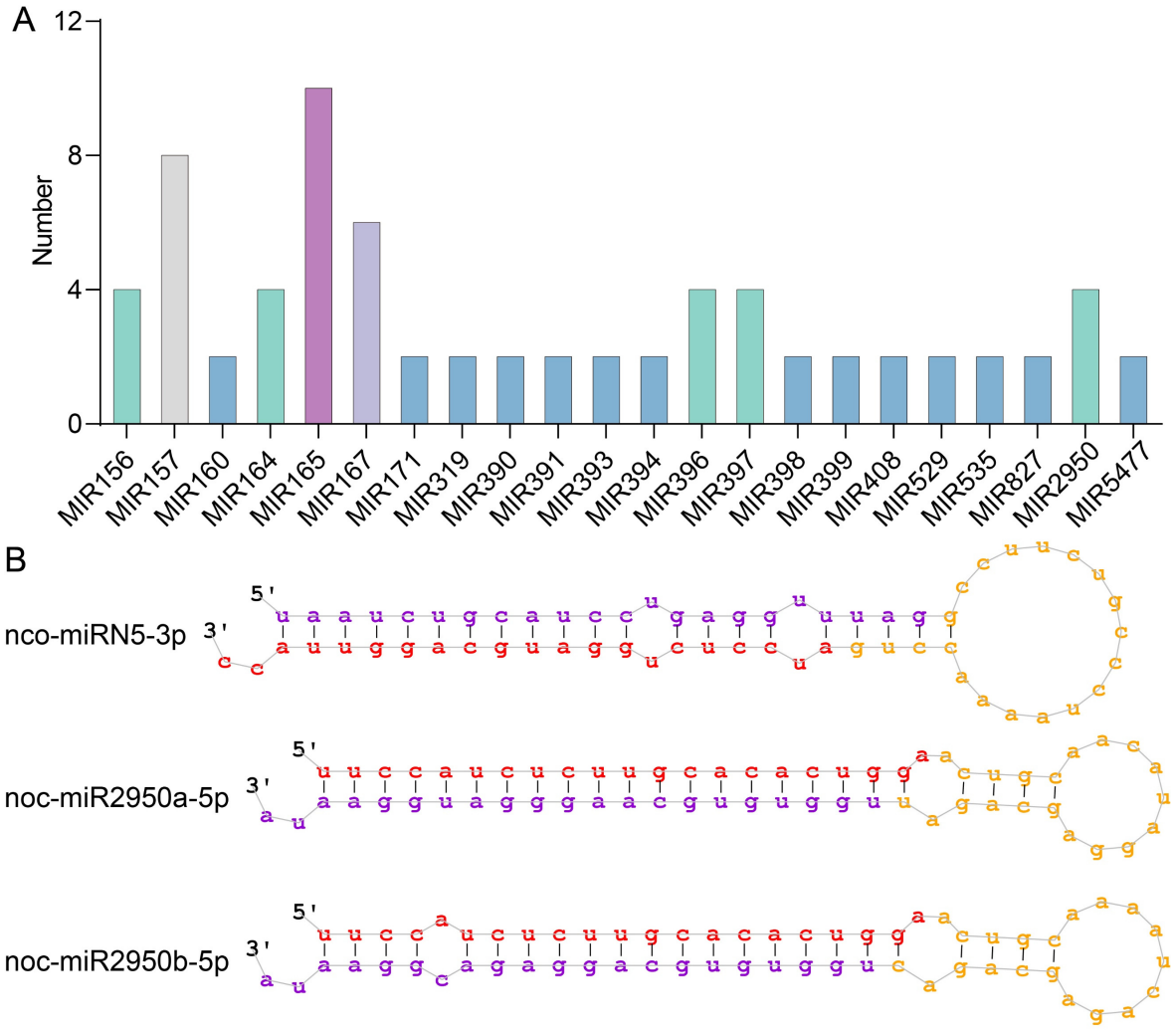



**Figure 10.** Identification of miRNAs in *N. colorata*. (**A**) Statistics of identified miRNA family members. (**B**) Representative stem–loop structures of identified miRNA precursors. Red is the mature miRNA sequence, purple is the miRNA-star sequence, and yellow is the non-miRNA and loop sequences in the stem.

To investigate the regulation of R2R3-MYB gene expression by miRNAs, the potential miRNAs that regulate R2R3-MYB gene expression were predicted. A total of 84 regulatory relationships were found between 43 miRNAs and 35 R2R3-MYB genes. Specifically, 20 R2R3-MYB genes may be regulated by a single miRNA each, two R2R3-MYB genes could be regulated by two miRNAs, and 13 R2R3-MYB genes could be potentially regulated by at least three miRNAs. Notably, both *NcMYB2* and *NcMYB33* genes could be regulated by seven different miRNAs. Among these miRNAs, 26 targeted a single R2R3-MYB gene each, 7 targeted two genes, and nco-miR165b-5p targeted the most genes, with a total of 7 (Figure 11).

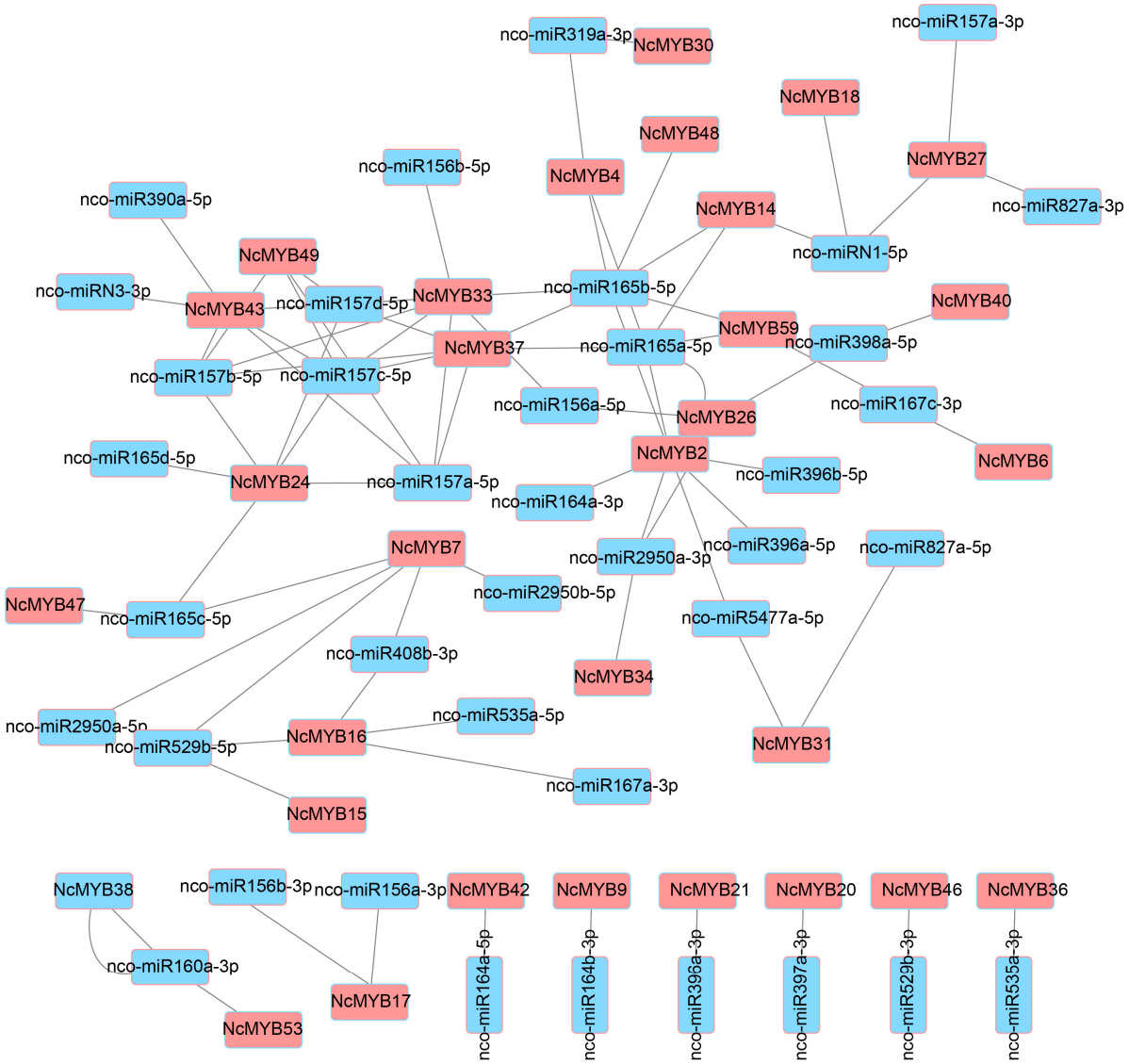



**Figure 11.** Regulation of R2R3-MYB gene expression miRNA in *N. colorata*. The blue-filled rectangle with a red border represents the miRNA in *N. colorata*, and the red-filled rectangle with a blue border represents the R2R3-MYB gene in *N. colorata*. Gray lines indicate potential regulatory relationships between the miRNAs and R2R3-MYB genes.

### *2.10. NcMYB25 May Be Involved in the Regulation of Anthocyanin Synthesis in N. colorata*

To identify the R2R3-MYB transcription factors potentially regulating anthocyanin synthesis in *N. colorata*, the expression levels of R2R3-MYB genes in the petals of *N. colorata* and *White colorata* were compared (Figure 12A,B). Compared to *White colorata*, 12 R2R3-MYB genes exhibited higher expression levels in *N. colorata*, with *NcMYB10*, *NcMYB25*, and *NcMYB50* showing the highest expression. Among the R2R3-MYB genes analyzed, five showed differential expression fold changes greater than 2 between *N. colorata* and *White colorata*. Notably, the *NcMYB25* gene exhibited the highest fold change, reaching 21.73 (Figure 12B). Based on gene expression levels and fold changes, the *NcMYB25* gene was selected for further analysis to investigate its role in anthocyanin synthesis.

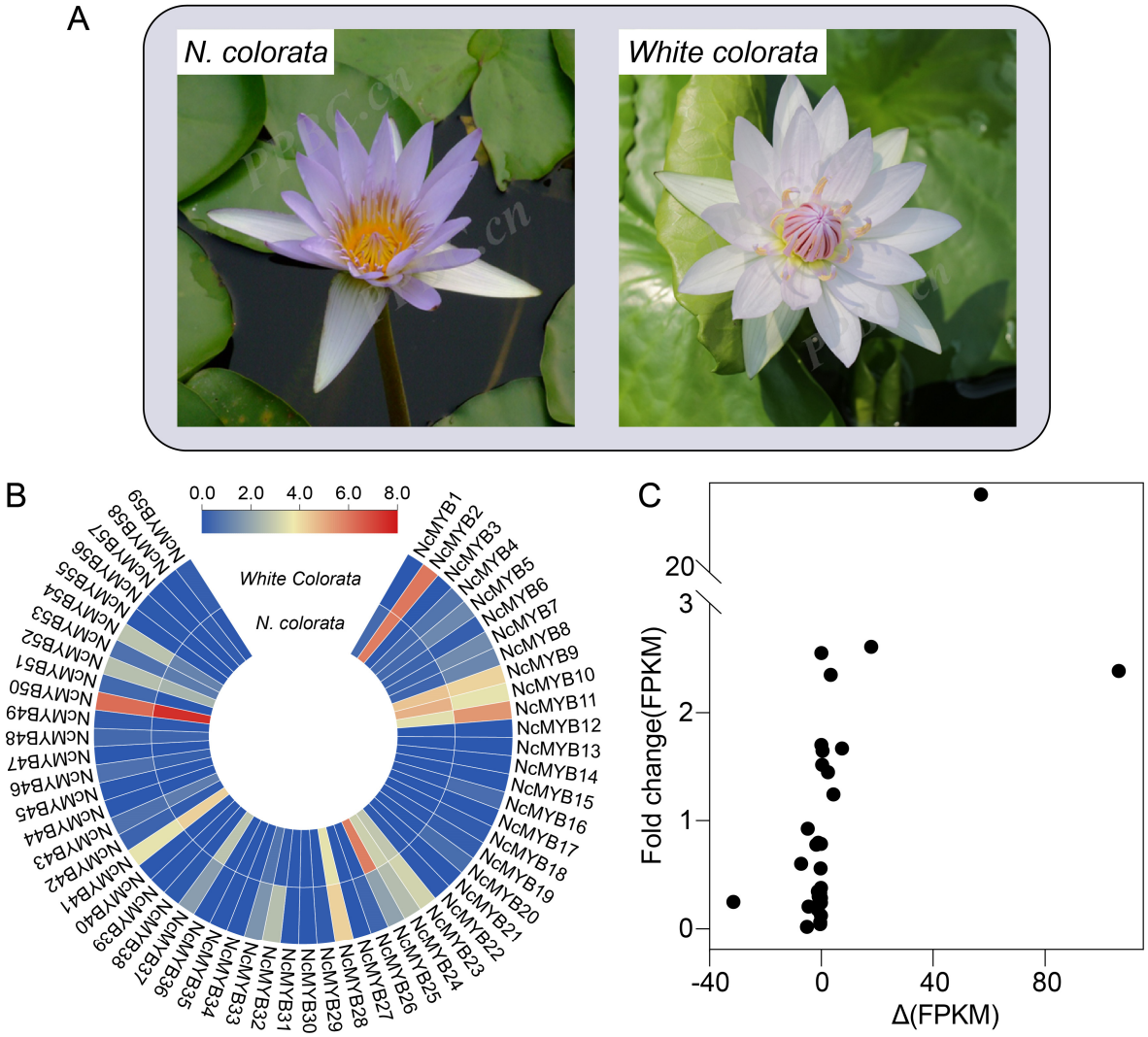



**Figure 12.** The expression levels of the R2R3-MYB gene in *N. colorata* and *White colorata* in petals. (**A**) Representative images of *N. colorata* and *White colorata*. Images were from PCBB. (**B**) Expression levels of R2R3-MYB genes in *N. colorata* and *White colorata* in petals. (**C**) Scatter plot of the expression levels of R2R3-MYB gene. The expression level of the MYB genes in *N. colorata* was denoted as FPKM (*N. colorata*). Similarly, the expression level of the MYB genes in *White colorata* was denoted as FPKM (*White colorata*). The *x*-axis shows the difference in the expression level of the R2R3-MYB gene between *N. colorata* and *White colorata*, and the difference was calculated as ΔFPKM = FPKM (*N. colorata*) – FPKM (*White colorata*). The *y*-axis shows the fold change of the expression level of R2R3-MYB gene between *N. colorata* and *White colorata*. The fold change was calculated using the formula: Fold Change (FPKM) = FPKM (*N. colorata*)/FPKM (*White colorata*).

### *2.11. Localization of NcMYB25 Protein in the Nucleus*

To verify whether the NcMYB25 protein was localized in the nucleus, a subcellular localization analysis was conducted using the tobacco expression system (Figure 13A). In tobacco leaves transformed with the *35S::GFP* vector, green fluorescence was observed around the cell nucleus and cell membrane. By contrast, tobacco leaves transformed with the *35S::NcMYB25-GFP* vector showed green fluorescence exclusively in the nuclear region, indicating that the NcMYB25 protein was localized in the nucleus (Figure 13B).

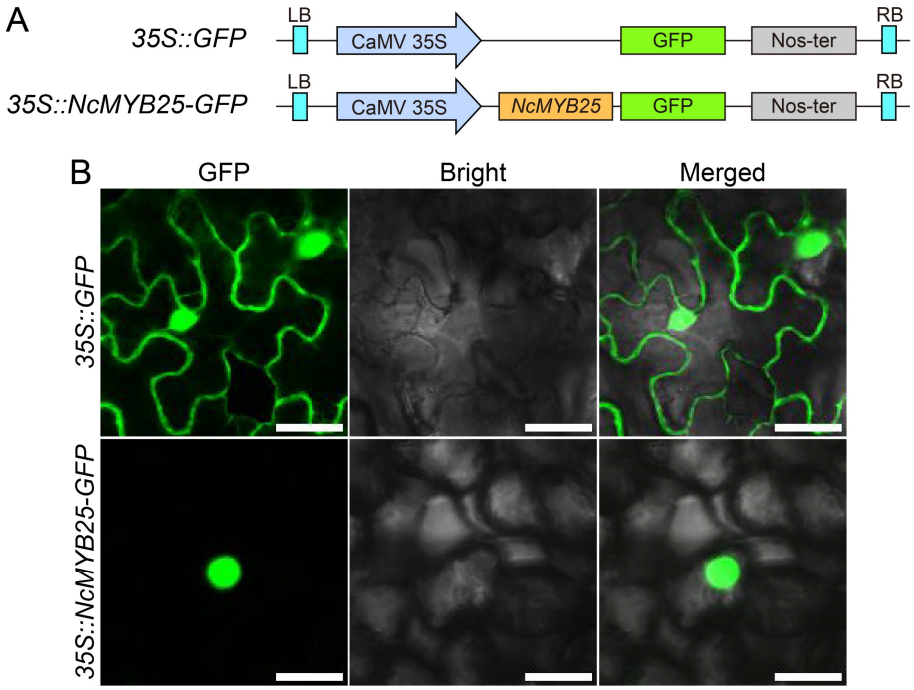



**Figure 13.** Subcellular localization analysis of the NcMYB25 protein. (**A**) Schematic diagrams of the *35S::GFP* and *35S::NcMYB25-GFP* vectors used for subcellular localization. (**B**) Fluorescence signals of the GFP protein and the NcMYB25-GFP fusion protein in tobacco leaf cells observed under a laser confocal microscope. Scale bar = 25 μm.

### *2.12. Transcriptional Activation Activity of NcMYB25 Transcription Factors*

To analyze the transcriptional activation activity of the NcMYB25 transcription factor, we conducted an assay using the *pGBKT7* vector and the AH109 yeast system (Figure 14A). Yeast cells expressing the *NcMYB25* gene and positive control yeast cells grew normally on media lacking tryptophan and media lacking both tryptophan and histidine, while negative control yeast cells only grew normally on media lacking tryptophan, indicating that the NcMYB25 transcription factor possesses transcriptional activation activity (Figure 14B).

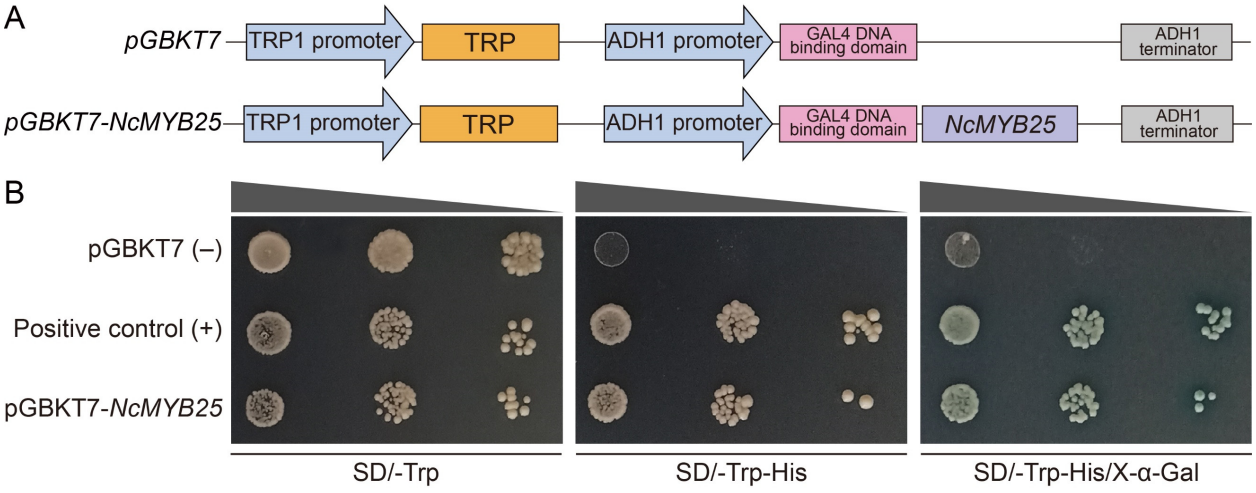



**Figure 14.** Transcriptional activation activity analysis of the NcMYB25 transcription factor. (**A**) Schematic diagrams of the *pGBKT7* and *pGBKT7-NcMYB25* vectors used in the transcriptional activation activity assay. (**B**) Representative images of the transcriptional activation activity analysis results, with pGBKT7 as the negative control.

### *2.13. Overexpression of NcMYB25 Gene Enhanced the Accumulation of Anthocyanins in Apple Pericarp*

To study the effect of the NcMYB25 transcription factor on anthocyanin synthesis, the *NcMYB25* gene was transiently expressed in an apple pericarp, and anthocyanin accumulation was measured (Figure 15A). Compared to that of the control, the apple pericarp expressing the *NcMYB25* gene showed distinct red coloration (Figure 15B). Compared to the control, the anthocyanin content in the NcMYB25 group was significantly higher (Figure 15C). These results showed that overexpression of *NcMYB25* gene increased the accumulation of anthocyanin in the transgenic apple pericarp, suggesting that the NcMYB25 transcription factor played a positive regulatory role in anthocyanin synthesis.

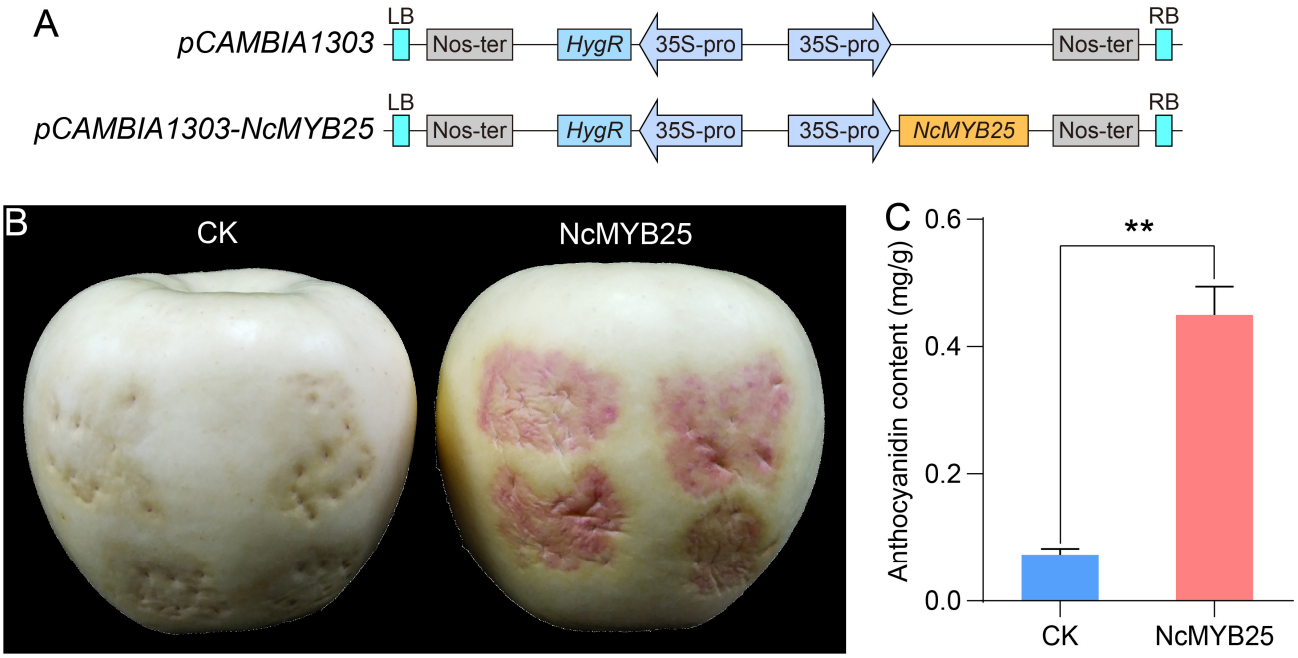



**Figure 15.** Overexpression of *NcMYB25* gene enhanced the accumulation of anthocyanins in an apple pericarp. (**A**) Schematic diagrams of the *pCAMBIA1303* and *pCAMBIA1303-NcMYB25* vectors used in the apple pericarp transient transformation assay. (**B**) The color change in the apple pericarp after *Agrobacterium* injection. (**C**) Anthocyanin content. "**" indicated significance at $p < 0.01$.

## 3. Discussion

Water lily plays an increasingly important role in water gardens, water purification, and ecological civilization construction. *N. colorata* is a typical tropical water lily with striking flowers and high ornamental value [27]. Flower color has long been a focal point in water lily breeding and research. In the formation of anthocyanins, the fundamental compounds of flower color, R2R3-MYB transcription factors act as crucial regulatory factors [10,28]. Numerous R2R3-MYB transcription factors have been identified and analyzed across various species, revealing substantial variation in their numbers among different species. In the *Marchantia polymorpha* genome, there were 29 R2R3-MYB genes; in *Salvinia cucullata*, 41 genes; in *Gnetum montanum*, 37 genes; in *Oryza sativa*, 113 genes; in *Vitis vinifera*, 130 genes; and in *Glycine max*, 298 genes [16]. The differences in the number of R2R3-MYB genes among species relate to various factors, such as the evolutionary stage of the species, WGD events, and the quality of genome sequencing and annotation [16]. Generally, higher plants have more R2R3-MYB genes compared to lower plants, and species that have undergone more WGD events tend to have a higher number of R2R3-MYB genes. In this study, 59 R2R3-MYB genes were identified in the *N. colorata*. The water lily belongs to one of the early differentiation groups of angiosperms, and the relatively small number of R2R3-MYB genes in *N. colorata* may reflect its evolutionary status [6]. On the other hand,

the relatively stringent criteria used in gene identification in this study may also contribute to the lower number of identified R2R3-MYB genes in *N. colorata*.

Based on phylogenetic and evolutionary analysis results, the evolutionary history of R2R3-MYB genes in *N. colorata* was inferred. *M. glyptostroboides*, as a representative of gymnosperms, contained four orthologous genes of the R2R3-MYB genes in *N. colorata*. It is speculated that *NcMYB11*, *NcMYB12*, *NcMYB36*, and *NcMYB39* genes may be the oldest members of the R2R3-MYB gene family in *N. colorata*. *A. trichopoda* is one of the earliest diverging angiosperms with no phylogenetic relationship to most other angiosperms. The genome of *A. trichopoda* contained 30 orthologous genes of the R2R3-MYB genes in *N. colorata*, suggesting that 27 orthologous genes of the R2R3-MYB genes in *N. colorata* had newly emerged during the evolution from gymnosperms to angiosperms, and the NcMYB25 gene may have formed during this historical stage. When the most recent common ancestor (MRCA) of *A. trichopoda* and *L. chinense* diverged into *L. chinense*, 12 orthologous genes of the R2R3-MYB genes in *N. colorata* had newly emerged. When the MRCA of *L. chinense* and *B. schreberi* diverged into *B. schreberi*, 9 orthologous genes of the R2R3-MYB genes in *N. colorata* had newly emerged. *B. schreberi*, as a unique group among aquatic plants, provides valuable insights into the origins, differentiation, and evolutionary history of aquatic plant groups. During the evolutionary stage when the MRCA of *B. schreberi* and *E. ferox* diverged into *E. ferox*, the orthologous genes of the *NcMYB44* and *NcMYB57* genes were newly formed. Additionally, when the MRCA of *E. ferox* and *N. colorata* diverged into *N. colorata*, eight R2R3-MYB genes in *N. colorata* were newly emerged (Table S5).

The classification of the R2R3-MYB transcription factor remains contentious, primarily due to lineage-specific gene loss or expansion [16,29,30]. Currently, the two classifications of R2R3-MYB proteins in *Arabidopsis* are often used as references for classifying R2R3-MYB proteins in other plants [16,18]. R2R3-MYB transcription factors are involved in regulating various biological processes in plants, such as growth and development, metabolic processes, and responses to environmental stresses. Additionally, the biological processes involving plant R2R3-MYB transcription factors show a clear correlation with the classification of R2R3-MYB proteins. In *Arabidopsis*, several R2R3-MYB transcription factors were involved in the regulation of flavonoid biosynthesis. The S7 group members, such as AtMYB11, AtMYB12, and AtMYB111, were involved in regulating flavonols biosynthesis [20]. The S6 group members, such as AtMYB75, AtMYB90, AtMYB113, and AtMYB114, have been confirmed to regulate anthocyanin biosynthesis [21]. AtMYB16, a transcription factor from the S9 group, was involved in the formation of petal epidermal cells, while AtMYB17 was involved in flower development and seed germination [31,32]. Similarly, NcMYB10, as a member of the S9 group, is speculated to play a role in flower development in *N. colorata*. Many R2R3-MYB transcription factors are also involved in plant responses to biotic and abiotic stresses in plants. AtMYB96, a transcription factor from the S1 group, regulated the responses to drought stress and disease resistance via the ABA signaling pathway [33,34]. AtMYB15 (S2 group) was involved in responses to cold stress [35], AtMYB2 (S20 group) regulated the expression of genes responsive to salt and drought stress [36], and AtMYB41 and AtMYB102 (S11 group) was involved in plant defense response [37–39]. It is speculated that members of the S2 and S20 groups in *N. colorata*, such as NcMYB23, NcMYB54, NcMYB62, and NcMYB78, may be involved in the response to abiotic stresses, while NcMYB6, NcMYB8, and NcMYB11 may be involved in responses to biotic stresses.

Several studies have shown that the expression of R2R3-MYB genes is regulated by various internal and external factors, with external factors primarily including biotic and abiotic stresses. In *Arabidopsis*, *AtMYB2* and *AtMYB96* genes were significantly upregulated in response to drought stress [33,36], and *AtMYB102* gene showed high responses to salt stress and wounding [39]. Many R2R3-MYB genes are regulated by miRNAs. For instance, miR159 targeted *AtMYB33*, *AtMYB35*, *AtMYB65*, and *AtMYB101* genes [40]; while miR858 targeted and silenced the expression of *AtMYB11*, *AtMYB13*, and *AtMYB20* genes [41]. Plant hormones are also intrinsic factors affecting the expression of R2R3-MYB genes. The *RhMYB108* gene was significantly upregulated after treatment with exogenous ethylene and

JA [42]. The *TcMYB29a* gene from *Taxus chinensis* was strongly suppressed after treatment with exogenous methyl jasmonate (MeJA) [43]. The *CmMYB2* gene was upregulated after ABA treatment [44]. In this study, a *cis*-acting element analysis showed that all R2R3-MYB genes contained hormone response elements, with the highest number of ABA and MeJA response elements, suggesting that the expression of most R2R3-MYB genes in *N. colorata* may be regulated by ABA or MeJA signaling pathways.

Anthocyanin biosynthesis in plants is a part of the phenylpropanoid pathway [45]. The initial step in anthocyanin biosynthesis involves the CHS catalyzing the reaction between malonyl-CoA and coumaroyl-CoA to form naringenin chalcone. Naringenin chalcone is then converted to dihydromyricetin through the action of CHI, F3′H, and F3′5′H. DFR, a key enzyme in the anthocyanin biosynthetic pathway, catalyzes the conversion of dihydromyricetin into the colorless anthocyanin precursor leucodelphinidin, which is then converted to anthocyanins by ANS, marking the final step in anthocyanin synthesis [9,10]. Although UFGT does not directly participate in anthocyanin synthesis, it catalyzes the conversion of unstable anthocyanins into more soluble and stable anthocyanidins stored in cells [11]. Anthocyanin synthesis is primarily influenced by two types of genes: those involved in the biosynthetic pathway and regulatory genes, such as R2R3-MYB, bHLH, and WD40 transcription factors [46]. Many key genes in the anthocyanin biosynthesis pathway are regulated by R2R3-MYB transcription factors. The MdMYB3 transcription factor in apples activated the expression of *CHS* and *CHI* genes [25]. DFR and ANS are additional key enzymes in the anthocyanin biosynthesis pathway. The LvMYB1 transcription factor promoted anthocyanin synthesis by activating *LvANS* gene expression through binding to its promoter [26]. The PdMYB118 transcription factor from *Populus deltoids* regulated anthocyanin biosynthesis by binding to MYB-binding sites on the promoters of *DRF2* and *ANS1* genes [47]. In this study, we revealed the positive regulatory role of the NcMYB25 transcription factor in anthocyanin synthesis.

## 4. Materials and Methods

### *4.1. Identification, Properties, and Structural Analysis of R2R3-MYB in N. colorata*

The genomic data of *N. colorata* were downloaded from the Genome Warehouse in BIG Data Center (accession number: GWHAAYW00000000) [6]. Genomic data for *Metasequoia glyptostroboides*, *Amborella trichopoda*, *Liriodendron chinense*, *Brasenia schreberi*, and *Euryale ferox* were downloaded from the NCBI database. MYB transcription factors were identified using HMMER v3.4 and the PF00249 model [48,49]. R2R3-MYB transcription factors were manually selected based on previously described sequence features of R2R3-MYB proteins [18]. The distribution of R2R3-MYB genes on chromosomes was visualized using Tbtools-II software [50]. Protein properties were predicted using the ProtParam tool from the Expasy database. The structure of R2R3-MYB proteins was analyzed using the MEME suite and visualized using Tbtools-II software [50].

### *4.2. Phylogenetic and Evolutionary Analysis*

Multiple sequence alignments of R2R3-MYB proteins were performed using the MUSCLE algorithm [51], and the results were visualized using Jalview v2.11.4.0 software [52]. Phylogenetic trees were constructed using MEGA-X software [53]. R2R3-MYB transcription factors were classified based on the classification criteria of the R2R3-MYB factors in *Arabidopsis* [16,18]. Phylogenetic trees were visualized using the Evolview website [54]. Collinearity analysis was conducted using MCScanX software [55]. Ka, Ks, and Ka/Ks ratios were calculated using KaKs_Calculator 3.0 software [56]. Species trees and divergence times were obtained from the TimeTree website [57].

### *4.3. Expression Pattern Analysis*

The region 2000 bp upstream of the gene start codon was used as a promoter sequence. *Cis*-acting elements in promoters were analyzed using the PlantCARE database [58]. To analyze expression patterns, 15 RNA-Seq datasets were downloaded from the Genome Se-

quence Archive (Accession Numbers: CRR058573~CRR058582 and CRR059175~CRR059179). The data from CRR058573~CRR058582 correspond to transcriptomic data from different tissues of *N. colorata* (mature leaf, mature leaf stalk, juvenile flower, juvenile leaf, juvenile leaf stalk, carpel, stamen, sepal, petal, and root), while the data from CRR059175~CRR059179 correspond to transcriptomic data from different tissues of the *White colorata* (stamen, sepal, petal, leaf, and carpel). Transcriptome data were analyzed on the Galaxy website, and gene expression levels were calculated using the featureCounts tool [59].

### *4.4. Total RNA Extraction and qRT-PCR Analysis*

Fresh roots, stems, leaves, petioles, and petals of *N. colorata* were collected for RNA extraction. Total RNA was extracted using the TRIpure Total RNA Rapid Extraction Kit (Huayueyang Biotechnology (Beijing) Co., Ltd., Beijing, China). Reverse transcription was performed using the FastQuant RT Kit (With gDNase) (Tiangen Biotech (Beijing) Co., Ltd., Beijing, China) with 1 μg of total RNA. A 10-fold diluted cDNA solution was used as the template for qRT-PCR analysis. NcActin gene was used as the housekeeping gene in the qRT-PCR analysis. Primers for qRT-PCR were designed using Primer Premier 6 software; the primer sequences used for qRT-PCR are listed in Table S6. The qRT-PCR analysis was performed with FastFire qPCR PreMix (SYBR Green) (Tiangen Biotech (Beijing) Co., Ltd., Beijing, China) on QuantStudio 3 real-time PCR system (Thermo Fisher Scientific Inc., Waltham, MA, USA). The qRT-PCR reaction system (20 μL) included 10 μL PreMix, 2 μL of template, 0.6 μL each of forward and reverse primers, and 6.8 μL RNase-Free dd$H_2O$. The qRT-PCR cycling conditions: initial denaturation at 95 °C for 1 min; 40 cycles of denaturation at 95 °C for 5 s, and annealing/extension at 60 °C for 15 s. Each experiment was conducted in triplicate. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [60].

### *4.5. Analysis of Possible Functions of R2R3-MYB Proteins in N. colorata*

Functional analysis of R2R3-MYB proteins in *N. colorata* was performed using protein–protein interaction (PPI) and gene ontology (GO) functional analysis. PPI analysis was conducted using the STRING database [61]. GO functional annotation was performed using eggNOG-mapper v2 software [62].

### *4.6. Identification of miRNAs and Prediction of Target Genes of miRNAs*

To identify miRNAs in *N. colorata*, a total of 7 ncRNA-Seq datasets were downloaded from the Genome Sequence Archive (SRX8123259, SRX8123358~SRX8123363). The miRDeep2 v2.0.1.2. software was used for miRNA identification [63]. The target genes of the miRNA were predicted using psRNATarget software (2017 Update) with default parameters [64].

### *4.7. Subcellular Localization Analysis Assay*

Subcellular localization was analyzed using the tobacco transient transformation system [65,66]. The coding sequence (CDS) of the *NcMYB25* gene was cloned into the *pCAMBIA1305-GFP* vector to create the *pCAMBIA1305-NcMYB25-GFP* construct. This construct, along with the *pCAMBIA1305-GFP* vector, was transformed into *Agrobacterium tumefaciens* GV3101 separately. The *Agrobacterium* solution carrying the constructs was injected into the leaves of tobacco seedlings, and the plants were incubated under weak light for one day, then under normal conditions for 1–2 days. The fluorescence signal of GFP proteins and NcMYB25-GFP fusion proteins in tobacco leaves was observed using a laser scanning confocal microscope, with an excitation wavelength of 488 nm and an emission wavelength of 500–530 nm (Leica, Mannheim, Germany). Primers used for subcellular localization analysis are listed in Table S7.

### *4.8. Transcriptional Activation Analysis Assay*

The transcriptional activation activity of the NcMYB25 transcription factor was analyzed using a yeast expression system [67]. The CDS of the *NcMYB25* gene was cloned into

the *pGBKT7* vector to create the *pGBKT7-NcMYB25* construct. This construct and *pGBKT7* vector were transformed into *Saccharomyces cerevisiae* AH109 separately, and growth was tested on SD/−Trp and SD/−His/−Trp media. X-α-Gal (20 μg/mL) was added to SD/−His/−Trp media as an indicator of transcriptional activation activity. Primers used for transcriptional activation analysis assay are listed in Table S8.

### *4.9. Apple Pericarp Transient Transformation Assay*

The apple pericarp transient transformation assay was performed following previously described methods [49]. The CDS of *NcMYB25* gene was cloned into the *pCAMBIA1303* vector to create the *pCAMBIA1303-NcMYB25* construct. This construct, along with the *pCAMBIA1303* vector, was transformed into *A. tumefaciens* GV3101 separately and injected into apple pericarps. The injected apples were placed in a dark incubator for 24 h, followed by 5 days in a lighted incubator. The apples inoculated with Agrobacterium containing the *pCAMBIA1303-NcMYB25* vector and those inoculated with Agrobacterium containing the *pCAMBIA1303* vector were maintained under identical environmental conditions. After this period, the color of the apple pericarp at the injection site was observed, and the anthocyanin concentration in the apple pericarp at the injection site was measured. The anthocyanin content in the pericarp was measured using the Anthocyanin Content Assay Kit (Biosharp Biotechnology Co., Ltd., Hefei, China). Primers used for apple pericarp transient transformation assay are listed in Table S9.

### *4.10. Statistical Analysis*

A Student's *t*-test was employed to determine the statistical significance of differences between two sample means, with "*" indicating significance at $p < 0.05$ and "**" indicating significance at $p < 0.01$. The statistical analysis was performed using Microsoft Office Excel 2019.

## 5. Conclusions

In this study, 59 R2R3-MYB genes were identified in *N. colorata*, distributed across 14 chromosomes. Segmental duplication and tandem duplication were the primary driving forces of the R2R3-MYB gene family formation in *N. colorata*. The R2R3-MYB transcription factors in *N. colorata* were likely involved in regulating flavonoid biosynthesis, plant growth and development, and responses to abiotic stress. We identified 92 miRNAs in *N. colorata*, with 43 interacting with 35 R2R3-MYB genes. The NcMYB25 transcription factor, which localized in the nucleus and possessed transcriptional activation activity, positively regulated anthocyanin synthesis. This study characterized the features of R2R3-MYB genes in *N. colorata* and the biological function of the *NcMYB25* gene in anthocyanin synthesis, contributing to the understanding of molecular mechanisms of flower color formation and providing data support for molecular breeding of water lily varieties.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants13212990/s1, Figure S1: Sequence conservation analysis of motifs in R2R3-MYB proteins of *N. colorata*; Figure S2: Multiple sequence alignment of R2R3-MYB proteins in *N. colorata*; Figure S3: Density curve of Ks in the *N. colorata* genome; Figure S4: PPI analysis of R2R3-MYB in *N. colorata*; Table S1: Physicochemical properties analysis of R2R3-MYB protein in *N. colorata*; Table S2: Analysis of evolutionary selective pressure of R2R3-MYB gene in *N. colorata*; Table S3: GO annotation of transcription factor R2R3-MYB in *N. colorata*; Table S4: The identification of miRNA in *N. colorata*; Table S5: Evolution history of R2R3-MYB gene in *N. colorata*; Table S6: Primers sequence for qRT-PCR assay; Table S7: Primers sequence for subcellular localization assay; Table S8: Primers sequence for transcription activation activity assay; Table S9: Primer sequences for transient expression assay in apple pericarps.

**Author Contributions:** Conceptualization, Q.L. and S.L.; methodology, Q.L.; software, Q.L.; validation, Q.L., S.L., T.L., Q.W. and Y.Z.; formal analysis, T.L., Q.W. and Y.Z.; investigation, Q.L.; resources, T.L.; data curation, Q.L.; writing—original draft preparation, Q.L.; writing—review and editing, Q.L. and S.L.; visualization, Q.L.; supervision, S.L.; project administration, Q.L. and S.L.; funding acquisition, S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Key Research and Development Program of the Shaanxi Provincial Science and Technology Department, grant number 2024NC-YBXM-070 and the Key Cultivation Research Project of the Shaanxi Academy of Sciences, 2022K-03.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data supporting this study are available within the paper and within the Supplementary Materials published online.

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

## References

1. Qiu, P.; Liu, T.; Xu, Y.; Ye, C.; Zhang, R.; Wang, Y.; Jin, Q. Multi-omic dissection of the cold resistance traits of white water lily. *Hortic. Res.* **2024**, *11*, uhae093. [CrossRef] [PubMed]
2. Huang, X.; Yang, M.; Guo, J.; Liu, J.; Chu, G.; Xu, Y. Genome-Wide Survey and Analysis of Microsatellites in Waterlily, and Potential for Polymorphic Marker Development. *Genes* **2022**, *13*, 1782. [CrossRef] [PubMed]
3. Yoo, M.-J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Divergence times and historical biogeography of *Nymphaeales*. *Syst. Bot.* **2005**, *30*, 693–704. [CrossRef]
4. Li, C.; Wang, M.; Luo, X. Uptake of uranium from aqueous solution by *Nymphaea tetragona* Georgi: The effect of the accompanying heavy metals. *Appl. Radiat. Isot.* **2019**, *150*, 157–163. [CrossRef] [PubMed]
5. Sun, C.; Chen, F.; Teng, N.; Xu, Y.; Dai, Z. Comparative analysis of the complete chloroplast genome of seven *Nymphaea* species. *Aquat. Bot.* **2021**, *170*, 103353. [CrossRef]
6. Zhang, L.; Chen, F.; Zhang, X.; Li, Z.; Zhao, Y.; Lohaus, R.; Chang, X.; Dong, W.; Ho, S.Y.W.; Liu, X.; et al. The water lily genome and the early evolution of flowering plants. *Nature* **2020**, *577*, 79–84. [CrossRef]
7. Hu, X.; Liang, Z.; Sun, T.; Huang, L.; Wang, Y.; Chan, Z.; Xiang, L. The R2R3-MYB transcriptional repressor TgMYB4 negatively regulates anthocyanin biosynthesis in tulips (*Tulipa gesneriana* L.). *Int. J. Mol. Sci.* **2024**, *25*, 563. [CrossRef]
8. Zeng, S.; Lin, S.; Wang, Z.; Zong, Y.; Wang, Y. The health-promoting anthocyanin petanin in *Lycium ruthenicum* fruit: A promising natural colorant. *Crit. Rev. Food Sci. Nutr.* **2024**, *64*, 10484–10497. [CrossRef]
9. Liu, W.; Feng, Y.; Yu, S.; Fan, Z.; Li, X.; Li, J.; Yin, H. The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* **2021**, *22*, 12824. [CrossRef]
10. Yan, H.; Pei, X.; Zhang, H.; Li, X.; Zhang, X.; Zhao, M.; Chiang, V.L.; Sederoff, R.R.; Zhao, X. MYB-mediated regulation of anthocyanin biosynthesis. *Int. J. Mol. Sci.* **2021**, *22*, 3103. [CrossRef]
11. Zhao, Z.C.; Hu, G.B.; Hu, F.C.; Wang, H.C.; Yang, Z.Y.; Lai, B. The UDP glucose: Flavonoid-3-O-glucosyltransferase (*UFGT*) gene regulates anthocyanin biosynthesis in litchi (*Litchi chinesis* Sonn.) during fruit coloration. *Mol. Biol. Rep.* **2012**, *39*, 6409–6415. [CrossRef] [PubMed]
12. Campanella, J.J.; Smalley, J.V.; Dempsey, M.E. A phylogenetic examination of the primary anthocyanin production pathway of the plantae. *Bot. Stud.* **2014**, *55*, 10. [CrossRef] [PubMed]
13. Qin, S.; Liu, Y.; Cui, B.; Cheng, J.; Liu, S.; Liu, H. Isolation and functional diversification of dihydroflavonol 4-Reductase gene *HvDFR* from *Hosta ventricosa* indicate its role in driving anthocyanin accumulation. *Plant Signal. Behav.* **2022**, *17*, 2010389. [CrossRef] [PubMed]
14. Shin, D.H.; Choi, M.-G.; Kang, C.-S.; Park, C.-S.; Choi, S.-B.; Park, Y.-I. Overexpressing the wheat dihydroflavonol 4-reductase gene *TaDFR* increases anthocyanin accumulation in an *Arabidopsis dfr* mutant. *Genes Genom.* **2016**, *38*, 333–340. [CrossRef]
15. Espley, R.V.; Hellens, R.P.; Putterill, J.; Stevenson, D.E.; Kutty-Amma, S.; Allan, A.C. Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J.* **2007**, *49*, 414–427. [CrossRef]
16. Jiang, C.K.; Rao, G.Y. Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. *Plant Physiol.* **2020**, *183*, 637–655. [CrossRef]
17. Ogata, K.; Kanei-Ishii, C.; Sasaki, M.; Hatanaka, H.; Nagadoi, A.; Enari, M.; Nakamura, H.; Nishimura, Y.; Ishii, S.; Sarai, A. The cavity in the hydrophobic core of MYB DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.* **1996**, *3*, 178–187. [CrossRef]
18. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2010**, *15*, 573–581. [CrossRef]
19. Rosinski, J.A.; Atchley, W.R. Molecular evolution of the MYB family of transcription factors: Evidence for polyphyletic origin. *J. Mol. Evol.* **1998**, *46*, 74–83. [CrossRef]
20. Stracke, R.; Ishihara, H.; Huep, G.; Barsch, A.; Mehrtens, F.; Niehaus, K.; Weisshaar, B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* **2007**, *50*, 660–677. [CrossRef]
21. Gonzalez, A.; Zhao, M.; Leavitt, J.M.; Lloyd, A.M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* **2008**, *53*, 814–827. [CrossRef] [PubMed]

22. Zhang, X.; Xu, Z.; Yu, X.; Zhao, L.; Zhao, M.; Han, X.; Qi, S. Identification of two novel R2R3-MYB transcription factors, *PsMYB114L* and *PsMYB12L*, related to anthocyanin biosynthesis in *Paeonia suffruticosa*. *Int. J. Mol. Sci.* **2019**, *20*, 1055. [CrossRef] [PubMed]
23. Borevitz, J.O.; Xia, Y.; Blount, J.; Dixon, R.A.; Lamb, C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **2000**, *12*, 2383–2394. [CrossRef] [PubMed]
24. Feng, S.; Wang, Y.; Yang, S.; Xu, Y.; Chen, X. Anthocyanin biosynthesis in pears is regulated by a R2R3-MYB transcription factor PyMYB10. *Planta* **2010**, *232*, 245–255. [CrossRef] [PubMed]
25. Vimolmangkang, S.; Han, Y.; Wei, G.; Korban, S.S. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* **2013**, *13*, 176. [CrossRef]
26. Yin, X.; Zhang, Y.; Zhang, L.; Wang, B.; Zhao, Y.; Irfan, M.; Chen, L.; Feng, Y. Regulation of MYB transcription factors of anthocyanin synthesis in lily flowers. *Front. Plant Sci.* **2021**, *12*, 761668. [CrossRef]
27. Tang, H.; Zhang, L.; Chen, F.; Zhang, X.; Chen, F.; Ma, H.; Van de Peer, Y. *Nymphaea colorata* (Blue-Petal Water Lily). *Trends Genet.* **2020**, *36*, 718–719. [CrossRef]
28. Li, C.; Yu, W.; Xu, J.; Lu, X.; Liu, Y. Anthocyanin biosynthesis induced by MYB transcription factors in plants. *Int. J. Mol. Sci.* **2022**, *23*, 11701. [CrossRef]
29. Wilkins, O.; Nahal, H.; Foong, J.; Provart, N.J.; Campbell, M.M. Expansion and diversification of the Populus R2R3-MYB family of transcription factors. *Plant Physiol.* **2009**, *149*, 981–993. [CrossRef]
30. Soler, M.; Camargo, E.L.; Carocha, V.; Cassan-Wang, H.; San Clemente, H.; Savelli, B.; Hefer, C.A.; Paiva, J.A.; Myburg, A.A.; Grima-Pettenati, J. The Eucalyptus grandis R2R3-MYB transcription factor family: Evidence for woody growth-related evolution and function. *New Phytol.* **2015**, *206*, 1364–1377. [CrossRef]
31. Baumann, K.; Perez-Rodriguez, M.; Bradley, D.; Venail, J.; Bailey, P.; Jin, H.; Koes, R.; Roberts, K.; Martin, C. Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development* **2007**, *134*, 1691–1701. [CrossRef] [PubMed]
32. Zhang, Y.; Cao, G.; Qu, L.J.; Gu, H. Characterization of *Arabidopsis* MYB transcription factor gene *AtMYB17* and its possible regulation by LEAFY and AGL15. *J. Genet. Genomics* **2009**, *36*, 99–107. [CrossRef] [PubMed]
33. Seo, P.J.; Xiang, F.; Qiao, M.; Park, J.Y.; Lee, Y.N.; Kim, S.G.; Lee, Y.H.; Park, W.J.; Park, C.M. The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiol.* **2009**, *151*, 275–289. [CrossRef] [PubMed]
34. Seo, P.J.; Park, C.M. MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in *Arabidopsis*. *New Phytol.* **2010**, *186*, 471–483. [CrossRef]
35. Agarwal, M.; Hao, Y.; Kapoor, A.; Dong, C.H.; Fujii, H.; Zheng, X.; Zhu, J.K. A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance. *J. Biol. Chem.* **2006**, *281*, 37636–37645. [CrossRef]
36. Abe, H.; Urao, T.; Ito, T.; Seki, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **2003**, *15*, 63–78. [CrossRef]
37. Lippold, F.; Sanchez, D.H.; Musialak, M.; Schlereth, A.; Scheible, W.R.; Hincha, D.K.; Udvardi, M.K. AtMyb41 regulates transcriptional and metabolic responses to osmotic stress in *Arabidopsis*. *Plant Physiol.* **2009**, *149*, 1761–1772. [CrossRef]
38. Cominelli, E.; Sala, T.; Calvi, D.; Gusmaroli, G.; Tonelli, C. Over-expression of the *Arabidopsis AtMYB41* gene alters cell expansion and leaf surface permeability. *Plant J.* **2008**, *53*, 53–64. [CrossRef]
39. De Vos, M.; Denekamp, M.; Dicke, M.; Vuylsteke, M.; Van Loon, L.; Smeekens, S.C.; Pieterse, C.M. The *Arabidopsis thaliana* transcription factor AtMYB102 functions in defense against the insect herbivore *Pieris rapae*. *Plant Signal. Behav.* **2006**, *1*, 305–311. [CrossRef]
40. Allen, R.S.; Li, J.; Stahle, M.I.; Dubroué, A.; Gubler, F.; Millar, A.A. Genetic analysis reveals functional redundancy and the major target genes of the *Arabidopsis* miR159 family. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16371–16376. [CrossRef]
41. Addo-Quaye, C.; Eshoo, T.W.; Bartel, D.P.; Axtell, M.J. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Curr. Biol.* **2008**, *18*, 758–762. [CrossRef] [PubMed]
42. Zhang, S.; Zhao, Q.; Zeng, D.; Xu, J.; Zhou, H.; Wang, F.; Ma, N.; Li, Y. RhMYB108, an R2R3-MYB transcription factor, is involved in ethylene- and JA-induced petal senescence in rose plants. *Hortic. Res.* **2019**, *6*, 131. [CrossRef] [PubMed]
43. Cao, X.; Xu, L.; Li, L.; Wan, W.; Jiang, J. TcMYB29a, an ABA-responsive R2R3-MYB transcriptional factor, upregulates taxol biosynthesis in *Taxus chinensis*. *Front. Plant. Sci.* **2022**, *13*, 804593. [CrossRef] [PubMed]
44. Shan, H.; Chen, S.; Jiang, J.; Chen, F.; Chen, Y.; Gu, C.; Li, P.; Song, A.; Zhu, X.; Gao, H.; et al. Heterologous expression of the chrysanthemum R2R3-MYB transcription factor CmMYB2 enhances drought and salinity tolerance, increases hypersensitivity to ABA and delays flowering in *Arabidopsis thaliana*. *Mol. Biotechnol.* **2012**, *51*, 160–173. [CrossRef]
45. Shen, N.; Wang, T.; Gan, Q.; Liu, S.; Wang, L.; Jin, B. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* **2022**, *383*, 132531. [CrossRef]
46. Wang, X.C.; Wu, J.; Guan, M.L.; Zhao, C.H.; Geng, P.; Zhao, Q. *Arabidopsis* MYB4 plays dual roles in flavonoid biosynthesis. *Plant J.* **2020**, *101*, 637–652. [CrossRef]
47. Wang, H.; Wang, X.; Song, W.; Bao, Y.; Jin, Y.; Jiang, C.; Wang, C.; Li, B.; Zhang, H. PdMYB118, isolated from a red leaf mutant of *Populus deltoids*, is a new transcription factor regulating anthocyanin biosynthesis in poplar. *Plant Cell Rep.* **2019**, *38*, 927–936. [CrossRef]

48. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [CrossRef]
49. Sumbur, B.; Gao, F.; Liu, Q.; Feng, D.; Bing, J.; Dorjee, T.; Li, X.; Sun, H.; Zhou, Y. The characterization of R2R3-MYB genes in *Ammopiptanthus nanus* uncovers that the miR858-AnaMYB87 module mediates the accumulation of anthocyanin under osmotic stress. *Biomolecules* **2023**, *13*, 1721. [CrossRef]
50. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant.* **2023**, *16*, 1733–1742. [CrossRef]
51. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf.* **2004**, *5*, 113. [CrossRef] [PubMed]
52. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]
53. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
54. Subramanian, B.; Gao, S.; Lercher, M.J.; Hu, S.; Chen, W.H. Evolview v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **2019**, *47*, W270–W275. [CrossRef] [PubMed]
55. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]
56. Zhang, Z. KaKs_Calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genom. Proteom. Bioinf.* **2022**, *20*, 536–540. [CrossRef]
57. Kumar, S.; Suleski, M.; Craig, J.M.; Kasprowicz, A.E.; Sanderford, M.; Li, M.; Stecher, G.; Hedges, S.B. TimeTree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* **2022**, *39*, msac174. [CrossRef]
58. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]
59. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef]
60. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [CrossRef]
61. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P. STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2015**, *43*, D447–D452. [CrossRef] [PubMed]
62. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [CrossRef] [PubMed]
63. Mackowiak, S.D. Identification of novel and known miRNAs in deep-sequencing data with miRDeep2. *Curr. Protoc. Bioinformatics* **2011**, *36*, 12.10.1–12.10.15. [CrossRef] [PubMed]
64. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [CrossRef]
65. Zhang, J.; He, S. Tobacco system for studying protein colocalization and interactions. *Methods Mol. Biol.* **2021**, *2297*, 167–174.
66. Ma, L.; Lukasik, E.; Gawehns, F.; Takken, F.L. The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. *Methods Mol. Biol.* **2012**, *835*, 61–74.
67. Liu, Q.; Zheng, L.; Wang, Y.; Zhou, Y.; Gao, F. AmDHN4, a winter accumulated SKn-type dehydrin from *Ammopiptanthus mongolicus*, and regulated by AmWRKY45, enhances the tolerance of Arabidopsis to low temperature and osmotic stress. *Int. J. Biol. Macromol.* **2024**, *266*, 131020. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.