Acta Pharmaceutica Sinica B 2025;15(4):2253−2272

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

A chromosome-level *Dendrobium moniliforme* genome assembly reveals the regulatory mechanisms of flavonoid and carotenoid biosynthesis pathways

Jiapeng Yang[a,b,†], Qiqian Xue[a,b,†], Chao Li[a,b], Yingying Jin[a,b], Qingyun Xue[a,b], Wei Liu[a,b], Zhitao Niu[a,b,*], Xiaoyu Ding[a,b,*]

[a]*College of Life Sciences, Nanjing Normal University, Nanjing 210023, China*
[b]*Jiangsu Provincial Engineering Research Center for Technical Industrialization of Dendrobiums, Nanjing 210023, China*

Received 29 December 2024; received in revised form 24 January 2025; accepted 16 February 2025

KEY WORDS

Dendrobium moniliforme;
Genome;
R2R3-MYB;
Flavonoid;
Carotenoid;
UPLC−MS/MS;
Yeast one-hybrid;
Dual-luciferase

Abstract *Dendrobium moniliforme* (*D. moniliforme*) is a traditional medicinal herb widely cultivated in Asia. Flavonoids, one of the largest groups of secondary metabolites in plants, are significant medicinal components in *Dendrobium* species. Several subgroups of R2R3-MYB proteins have been validated to directly regulate flavonoid biosynthesis. Using PacBio sequencing technology, we assembled a high-quality chromosome-level *D. moniliforme* genome with a total length of 1.20 Gb and a contig N50 of 3.97 Mb. The BUSCO assessment of genome annotation was 91.4%. By integrating the genome and transcriptome, we identified biosynthesis pathway enzyme genes related to flavonoids, polysaccharides, carotenoids, and alkaloids. A total of 90 R2R3-MYBs were identified in *D. moniliforme* and classified into 21 subgroups. Studies on the functions of R2R3-MYB transcription factors revealed that R2R3-MYB in SG6 can up-regulate flavonoid biosynthesis. Various validation experiments, including subcellular localization, transient overexpression, UPLC−MS/MS, HPLC, yeast one-hybrid, and dual-luciferase assays, demonstrated that DMYB69 directly up-regulates the expression of enzyme genes involved in flavonoid biosynthesis, increasing the content of flavonoids such as anthocyanin, flavone, and flavonol. Additionally, DMYB44 was shown to directly up-regulate the expression of carotenoid biosynthesis enzyme genes, thereby increasing carotenoid content. This study provides an essential genome resource and theoretical basis for molecular breeding research in *D. moniliforme*.

*Corresponding authors.
E-mail addresses: niuzhitaonj@163.com (Zhitao Niu), dingxynj@263.net (Xiaoyu Ding).
†These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.03.005
2211-3835 © 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

© 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Flavonoids are one of the largest secondary metabolites in plants, comprising over ten thousand identified compounds[1,2]. They can be broadly categorized into subgroups such as flavones, flavonols, anthocyanins, proanthocyanidins, and isoflavones[3-5]. Flavonoids play crucial roles in plants, including providing color and fragrance to flowers[6], attracting pollinators for seed dispersal[7], supporting seedling growth and development[8], and protecting plants from abiotic stress[9]. Additionally, flavonoids possess various therapeutic properties, including anti-bacterial, anti-herpetic, anti-oxidant, anti-inflammatory, anti-ulcer, and anti-diabetic activities[10]. The biosynthesis of flavonoids is a complex process originating from the phenylpropanoid metabolic pathway[11-13]. The expression of flavonoid biosynthesis enzyme genes is regulated by the MBW complex, comprising R2R3-MYB transcription factors (TFs), basic helix-loop-helix (bHLH) TFs, and WD-repeat proteins[14-16].

Carotenoids are the largest class of natural pigments, with more than 750 different compounds found in various organisms, including algae and higher plants[17]. In the green tissues of plants, carotenoids are located in chloroplasts, where they serve as auxiliary molecules to photosynthetic pigments and function as antioxidants to prevent photo-oxidative damage[18]. In non-green tissues, carotenoids contribute various colors to the chromoplasts of flowers and fruits, attracting pollinators and seed dispersers[19].

The MYB family is one of the largest transcription factor families in plants, playing a crucial role in regulating phytochemical biosynthesis pathways and responding to both abiotic and biotic stress[20]. The first plant MYB gene, *COLORED1*, identified in maize, regulates flavonoid biosynthesis[21]. The most prominent feature of the MYB TF family is the tandem arrangement of 1−4 incomplete repeats of MYB domains, each containing approximately 50−55 amino acid residues at the N-terminal region of the protein sequence[22,23]. Based on the number of repeats, MYB TFs can be classified into four types: 1R-MYB, 2R-MYB (R2R3-MYB), 3R-MYB (R1R2R3-MYB), and 4R-MYB. Among these, R2R3-MYB proteins are the most abundant in plants, characterized by two repeats at the N-terminus and typically a transcriptional activation function at the C-terminus[24,25]. The functions of R2R3-MYB proteins are diverse, including roles in flavonoid biosynthesis[26-28], carotenoid biosynthesis[29-31], photomorphogenesis[32], epidermal cell development[33], auxin biosynthesis[34,35], response to abiotic stress[36-38], and fruit ripening[39-41].

Dendrobium moniliforme (L.) Sw. (*D. moniliforme*) (Fig. 1A), distinguished by its thin stems and small flowers, is a medicinal crop widely cultivated throughout Asia. As a representative herb in traditional Chinese medicine, it is often processed into finished herbal materials[42]. The primary active constituents of *D. moniliforme*, which endow it with significant economic value due to its ornamental and medicinal properties[43], include polysaccharides, alkaloids, flavonoids, bibenzyls, phenanthrenes, glycosides, and several trace mineral elements[44]. Research on flavonoid biosynthesis in *Dendrobium* plants continues to garner attention, particularly studies focusing on the regulation of flavonoid biosynthesis by R2R3-MYB proteins. For instance, in *Dendrobium hybrids*, transient expression of *DhMYB2* results in purple pigment spots on white petal tissues[45]. Using the dsRNA method, silencing the *DhMYB22* and *DhMYB60* genes in *D. officinale* floral organs leads to relatively low expression levels of *F3′H* and *DFR*[46]. In *Phalaenopsis*-type Dendrobiums, yeast one-hybrid assays demonstrate that DhMYB2 directly binds to the promoters of *DhF3′H1*, *DhF3′5′H2*, *DhANS*, and *DhGT4*. Dual-luciferase reporter gene analysis shows that DhMYB2 can activate these enzyme genes by 1.5−2.5 fold[47]. Transient overexpression of the *DoMYB5* gene results in an increase in total anthocyanin content in *D. officinale* leaves on Days 10[48]. Carotenoids also significantly contribute to the ornamental and medicinal values of *Dendrobium* plants, but research in this area remains limited[49,50].

We present the sequencing, assembly, and annotation of a chromosome-level genome of *D. moniliforme* using PacBio sequencing technology, along with comparative genomic analyses. By integrating genome and transcriptome data, we identified biosynthesis pathway enzyme genes related to medicinal components. Specifically, we pinpointed candidate R2R3-MYB proteins that likely regulate flavonoid and carotenoid biosynthesis in the *D. moniliforme* genome and validated their transient expression in *D. moniliforme* leaves. Furthermore, we conducted comprehensive TF functional validation using assays such as ultra performance liquid chromatography-mass spectrometry/mass spectrometry (UPLC−MS/MS), high performance liquid chromatography (HPLC), yeast one-hybrid, and dual-luciferase assays. This study provides a high-quality *D. moniliforme* genome and *in vivo* TF functional validation, offering a crucial molecular foundation and theoretical reference for future molecular breeding research on medicinal components.

2. Materials and methods

2.1. Genome sequencing and assembly

D. moniliforme (voucher specimen: Yang202301) used for genome sequencing was collected from Huoshan, Anhui (116.32°E, 31.38°N) and preserved in the Herbarium of the College of Life Sciences, Nanjing Normal University (Nanjing, China). Genomic DNA was extracted from young leaves using an improved cetyltrimethylammonium bromide (CTAB) method. Genome sequencing was performed using both PacBio long-read sequencing and Illumina high-throughput sequencing technologies (Benagen LTD). The next-generation sequencing data were filtered to obtain 446.67 million clean reads (∼67.0 Gb) (Supporting Information Table S1). The filtering criteria were: (1) removal of reads with an N base content exceeding 5%; (2) removal of low-quality reads with a base call quality score of 50% or less (quality value≤5); (3) removal of adapter-contaminated reads; and (4) removal of duplicate sequences caused by PCR amplification. PacBio long-read sequencing was conducted on the PacBio Sequel II platform (Menlo Park, CA, USA), generating a total of ∼158.38 Gb of sequence data, with

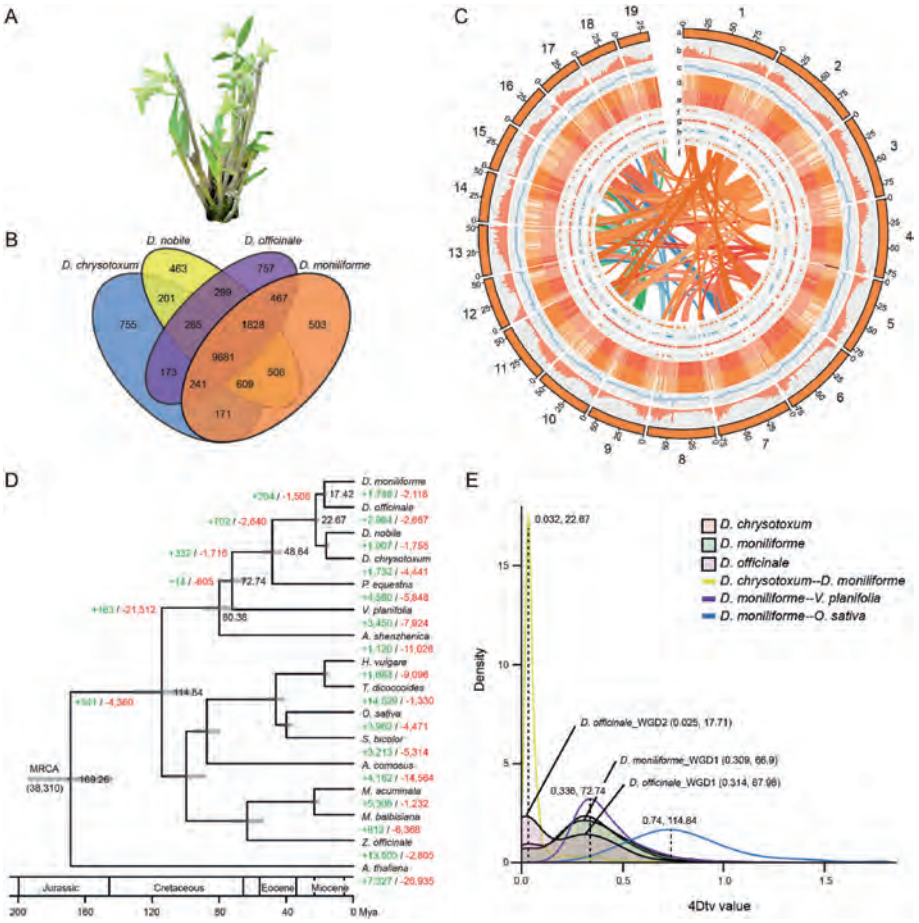

Figure 1 Genome annotation and comparative genomics results of *D. moniliforme*. (A) Images of *D. moniliforme*. (B) Venn diagram analysis of gene families in *Dendrobium*. (C) Basic information regarding the genome annotations of *D. moniliforme*: a. Chromosome length; b. Number of protein-coding genes; c. GC content; d. Number of TEs; e. Number of LTRs; f. Number of miRNAs; g. Number of snRNAs; h. Number of rRNAs; i. Number of tRNAs; j. Intraspecific collinearity. (D) Phylogenetic analysis and assessment of gene family expansion and contraction based on whole-genome data. (E) Analysis of WGD events using the 4DTv method. 4DTv, transversion of four-fold degenerate sites; LTR, long terminal repeat; miRNA, micro RNA; MRCA, most recent common ancestor; Mya, million years ago; rRNA, ribosomal RNA; snRNA, small nuclear RNA; TE, transposable element; tRNA, transfer RNA; WGD, whole-genome duplication.

a subread length N50 value of 24,454 bp (Supporting Information Table S2). The genome size was estimated to be approximately 1.24 Gb using the GCE (version 1.0.2) (https://github.com/fanagislab/GCE) (Supporting Information Table S3).

Next-generation sequencing data were assembled using next-Denovo software (version 2.4.0). Based on PacBio sequencing data, the spliced results were corrected twice using Racon software (version 1.4.13). Two rounds of assembly correction were then performed with Pilon (version 1.23)[51] to obtain the final assembly, using default parameters. The Hi-C library[52] was constructed on the Illumina platform with paired-end 150 reads. The raw Hi-C data were filtered to remove adapters and low-quality reads, resulting in high-quality clean data. The HICUP software (version 0.8.0) was used to align the clean data with the initially assembled genome, obtaining unique paired-end reads that were mapped to the genome. The effective Hi-C data were then used to anchor the initially assembled contigs onto chromosomes. The genome assembly was evaluated based on next-generation mapping rate, coverage, and Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis. The coverage was 99.16%, with an average depth of 53.74 × . The accuracy of the genome assembly was assessed using the BUSCO software (version 4.1.4)[53]. The BUSCO database contained relatively few conserved gene products related to the *Dendrobium* orchids.

2.2. *Genome annotation*

RepeatMasker software (version 4.0.9) was employed to identify repetitive elements, based on the RepBase repository[54]. RepeatModeler software (version 1.0.11) was then used to reconstruct the *de novo* database according to sequence characteristics, followed by alignment using RepeatMasker software (version 4.0.9)[55]. Subsequently, TEprotein (version 4.0.9) was utilized to predict repetitive elements[56]. Finally, all repeat prediction results were consolidated, redundancy was eliminated, and the final genomic repeat element set was obtained.

The annotation of high-quality protein-coding genes was performed by integrating homology-based, *de novo*, and transcriptome data predictions. For homology-based predictions, Exonerate software (version 2.2.073)[57] was used to map protein-coding gene sets from five plant genomes (*Dendrobium catenatum*, GCA_001605985.2; *Phalaenopsis equestris*, GCA_001263595.1;

Asparagus officinalis, GCA_001876935.1; *Ananas comosus*, GCA_001540865.1; *Aegilops tauschii*, GCA_000347335.2) onto the *D. moniliforme* genome assembly. *De novo* predictions were conducted using Augustus (version 3.3.2)[58], Genscan (version 1.0), and GlimmerHMM (version 3.0.4)[59]. Transcriptomes from the roots, stems, leaves, and flowers of *D. moniliforme* were obtained to assist with gene annotation. RNA-seq data were mapped to the genome sequence using StringTie software (version 2.1.1)[60].

Gene function annotation was primarily performed using two strategies: sequence similarity search and motif similarity search. For sequence similarity search, protein sequences encoded by genes were compared with existing protein databases such as UniProt, NR, and Kyoto Encyclopedia of Genes and Genomes (KEGG) to obtain functional and metabolic pathway information. KEGG annotations were linked to KEGG orthology and pathways using KOBAS (version 3.0) (https://github.com/xmao/kobas). For motif similarity search, InterPro secondary databases were queried using InterProScan (version 5.33). Additionally, conserved sequences, motifs, and structural domains of proteins were identified using HMMscan (version 3.1, e-value 1e-5)[61]. The completeness and quality of gene models were then assessed using BUSCO (version 4.1.2).

Four types of non-coding RNAs (ncRNAs) were annotated: transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), micro RNAs (miRNAs), and small nuclear RNAs (snRNAs). The tRNA sequences were annotated using tRNAscan-SE (version 1.23)[62]. RNAmmer (version 1.2) was utilized for rRNA prediction. The genome was searched for miRNA and snRNA sequences using INFERNAL (version 1.1.2), based on the Rfam RNA database (http://rfam.xfam.org/).

2.3. Gene family analysis

Protein sequences from the *D. moniliforme* genome (this study) and fifteen other genomes, including *Dendrobium officinale* Kimura et Migo (GCA_019514585.1), *Dendrobium chrysotoxum* Lindl. (GCA_019925795.1), *Dendrobium nobile* Lindl. (GCA_022539455.1), *Apostasia shenzhenica* Z.J. Liu & L.J. Chen (GCA_002786265.1), *Phalaenopsis equestris* (Schauer) Rchb. (GCA_001263595.1), *Vanilla planifolia* Andrews (GCA_016413895.1), *Oryza sativa* L. (GCA_001433935.1), *Ananas comosus* (L.) Merr. (GCA_001540865.1), *Hordeum vulgare* L. (GCA_904849725.1), *Sorghum bicolor* (L.) Moench (GCA_000003195.3), *Triticum dicoccoides* (Asch. & Graebn.) Schweinf. (GCA_002162155.3), *Musa acuminata* Colla (GCA_000313855.2), *Musa balbisiana* Colla (GCA_004837865.1), *Zingiber officinale* Roscoe (GCA_018446385.1), and *Arabidopsis thaliana* (L.) Heynh. (GCA_000001735.2), were utilized for gene family clustering through OrthoFinder (version 2.5.4)[63].

2.4. Phylogenomic analysis

MAFFT (version 7.450)[64] was utilized to align the nucleotide sequences of single-copy genes within orthogroups. Well-aligned single-copy genes were retained for the generation of a super-alignment matrix. A phylogenetic tree was constructed using the Maximum Likelihood (ML) method and 1000 bootstrap replicates with RAxML (version 1.3), employing *A. thaliana* as the outgroup. Divergence times were estimated using BEAST2 based on the super-alignment matrix[65]. The following constraints were applied for time calibrations: (1) The root age was set at 170 million years ago (Mya) with a prior distribution of normal (mean: 170, sd: 8)[66]. (2) Based on the oldest known fossil of monocotyledons, the earliest divergence time among monocotyledons was established at 115 Mya with a prior distribution of normal (mean: 115, sd: 3.05)[67]. Three independent Markov Chain Monte Carlo (MCMC) analyses were conducted to assess convergence, each comprising 10,000,000 generations. The results from the three independent runs were combined using LogCombiner, discarding the first 10% of data as burn-in.

By utilizing CAFÉ version 4.2[68], the expansion and contraction of gene families were analyzed by comparing the family sizes of the most recent common ancestor (MRCA). Gene families with more than 100 gene copies from one or several species were excluded from the analysis. The *P*-values for each lineage were calculated using conditional likelihood, and families exhibiting a *P*-value of less than 0.05 were deemed significantly expanded or contracted. Additionally, gene families from four *Dendrobium* genomes, including *D. moniliforme*, *D. chrysotoxum*, *D. officinale*, and *D. nobile*, were extracted for Venn diagram analysis to identify species-specific gene families within the genus *Dendrobium*.

To investigate whole-genome duplication (WGD) in the *Dendrobium* genus, an integrated analysis was performed utilizing wgd[69] and MCScanX[70], along with the calculation of the transversion of four-fold degenerate sites (4DTv). Protein sequences from *D. moniliforme*, *D. chrysotoxum*, *D. officinale*, *V. planifolia*, and *O. sativa* were employed to conduct BLASTP searches (e-value $<$ 1e-5) both within and between the genomes to identify collinear blocks. This was followed by searches for collinear blocks using MCScanX software based on gene positions and BLAST results. Homologous genes within these collinear blocks were aligned using MAFFT for multiple sequence alignment, from which coding DNA sequence (CDS) alignments were generated based on the protein alignment results. Finally, 4DTv values were computed from these alignments, and a scatter plot of the values for all gene pairs was created to identify the inferred WGD events and the divergence between each pair of species.

2.5. Transcriptome sequencing and analysis

Three-year-old potted seedlings were used as samples and treated with a 10 mmol/L methyl jasmonate (MeJA) solution every 12 h. After 2 days, leaf samples were frozen in liquid nitrogen for transcriptome sequencing. The mRNA from the three replicate leaf samples was purified using oligonucleotide-attached magnetic beads to remove rRNA and tRNA, yielding a total RNA amount of 2 μg per sample. Quality control verification was performed on the samples using an Agilent Technologies 2100 bioanalyzer (Sunnyvale, CA, USA). The final constructed library was amplified using phi29 to generate DNA nanoballs, which were subsequently loaded onto a nanoarray. Paired-end reads of 150 bp were obtained on the T7 platform at Wuhan Benagen Technology Co., Ltd.

The raw data were initially processed using FastQC (version 0.11.9) to eliminate low-quality sequences[71]. Subsequently, the clean reads were aligned to the *D. moniliforme* genome with HISAT. The gene expression levels for each sample were calculated using the RSEM package, expressed as fragments per kilobase of transcript per million fragments mapped (FPKM)[72]. The gene expression levels of stress samples were compared to those of control samples to identify differentially expressed genes (DEGs). Gene Ontology (GO) and KEGG enrichment analyses were performed using clusterProfiler (version 3.8)[73].

2.6. Identification and phylogenetic analyses of R2R3-MYB proteins

The Hidden Markov Model (HMM) profile of the MYB DNA binding domain was accessed from the Pfam database under accession number PF00249. The HMMER (version 3.0)[74], utilizing default parameters, was employed to perform an HMM search within the *D. moniliforme* genome to identify MYB proteins, using this profile as a query. R2R3-MYBs from *A. thaliana* were obtained from prior research, with corresponding sequences sourced from the TAIR database. All sequences were retrieved by executing a BLASTP analysis against *A. thaliana* R2R3-MYB protein dataset using default settings. To further validate the reliability of the candidate R2R3-MYB proteins, the R2R3-MYB sequences were submitted to Pfam, InterPro, and SMART to confirm the completeness of the R2 and R3 MYB repeat sequences and the integrity of the MYB domains. Proteins with incomplete conserved MYB domains were excluded, while the remaining proteins were considered to belong to the R2R3-MYB family in *D. moniliforme*. The theoretical isoelectric point (pI), molecular mass, hydrophilic mean coefficient, instability coefficient, and aliphatic coefficient of the R2R3-MYB proteins were calculated using the Expasy online tool (https://web.expasy.org/protparam/). Amino acid sequences were aligned using Clustal X 2.0, and an ML phylogenetic tree with 1000 bootstrap replicates was constructed using MEGA X. Based on the classification of R2R3-MYB proteins in *A. thaliana*, all R2R3-MYB proteins were categorized into subgroups.

2.7. Subcellular localization assay

The CDSs of *DMYB69*, *DMYB44*, and *DMYB7* were amplified using HiFi-enzyme PCR with primers that lacked stop codons (Supporting Information Table S4). The PCR products of *DMYB69*, *DMYB44*, and *DMYB7* were ligated into the pCambia1300-35S-EGFP plasmid, where they were fused to a green fluorescent protein (GFP) and driven by the CaMV 35S promoter for protein expression. Following sequence confirmation, the recombinant plasmids were designated as 35S:DMYB69::GFP, 35S:DMYB44::GFP, 35S:DMYB7::GFP, and 35S:GFP (control), and transformed into *Agrobacterium* strain GV3101. The bacterial suspensions (OD_{600} = 0.6−1.0) were then infiltrated into *Nicotiana benthamiana* leaves and maintained in the greenhouse for 72 h. Subsequently, the transformed tobacco leaves were examined under a confocal laser scanning microscope (Nikon Corporation, NIKON A1 Ti-E-A1R, Tokyo, Japan).

2.8. Overexpression assays

The primers utilized for plasmid construction are detailed in Table S4. The CDSs of *DMYB69*, *DMYB44*, and *DMYB7* were inserted into the binary vector pCAMBIA1301-35SN under the control of the CaMV 35S promoter to construct overexpression vectors. These vectors were amplified in *Escherichia coli* strain DH5α and subsequently transformed into *Agrobacterium* GV3101 using an electric shock method. *Agrobacterium* without the transfected plasmid served as the control. The *Agrobacterium* suspension was prepared and adjusted to an optical density at 600 nm (OD_{600}) of 0.4. The suspension was injected into the abaxial surface of the leaves, and the samples were kept in the dark for 12 h. Gene expression levels were assessed by qPCR on Day 3 following the transformation.

2.9. Quantitative real-time PCR assay

Total RNA was extracted from leaf tissue using the FastPure Plant Total RNA Isolation Kit (Vazyme, Nanjing, China) following the instructions. The concentration and quality of each RNA sample were confirmed using a Nanodrop spectrophotometer. First-strand cDNA for qPCR was synthesized with HiScript III RT SuperMix for qPCR (Vazyme, Nanjing, China). qPCR was performed on the LightCycler® 480 II Real-Time PCR Detection System (Roche, Basel, Switzerland) utilizing Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China). The qPCR primers are listed in Table S4. The cycling conditions included a pre-incubation step at 95 °C for 5 min, 40 cycles of amplification step at 95 °C for 10 s and 60 °C for 30 s, and a melting step at 95 °C for 15 s, 60 °C for 60 s, and 95 °C for 15 s. The expression levels of *GAPDH* (F primer: TTCGG AAGGA TTGGA AGGCT TGTAG; R primer: GAGAT GATAA CCTTC TTGGC ACCGC) served as an internal control. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method, with all quantification experiments conducted in triplicate.

2.10. UPLC−MS/MS assay of flavonoid

The content of flavonoid compounds was determined using the UPLC−MS/MS method. The data acquisition system primarily consisted of ultra high performance liquid chromatography (UPLC) and a high-resolution mass spectrometer (Q Exactive, Thermo, Waltham, MA, USA). The Q Exactive high-resolution mass spectrometry detection system (Thermo) was employed for data collection. The raw spectra obtained from the UPLC−Q Exactive/MS analysis were pre-processed using TraceFinder software, which included baseline filtering, peak identification, peak matching, retention time correction, and peak alignment, resulting in a data matrix that contained retention time, mass-to-charge ratio, and peak intensity. The total ion chromatogram represents a spectrum generated by summing the intensities of all ions in the mass spectrum at each time point and plotting them continuously. The horizontal axis corresponds to retention time, while the vertical axis represents ion flow intensity. A standard curve was generated using TraceFinder software, and the absolute quantification of the compounds was performed using the external standard method.

2.11. Yeast one-hybrid assay

The yeast one-hybrid assays were used to detect the binding of DMYB69 to the promoters of *DmCHS1* and *DmF3H1*, as well as the interaction between DMYB44 and the promoters of *DmBCH2* and *DmPSY*. The protein-coding sequences of *DMYB69* and *DMYB44* were cloned as prey into the pB42AD vector at the EcoRI-XhoI sites. Promoter fragments from *DmCHS1* (−1481 to −1091 bp), *DmF3H1* (−690 to −456 bp), *DmBCH2* (−1785 to −1509 bp), and *DmPSY* (−1830 to −1527 bp) were amplified and inserted into the pLacZi vector to serve as bait. The pBait-LacZi plasmids and pB42AD-prey plasmids were co-transformed into the EJ48 yeast strain, using X-gal chromogenesis to indicate protein binding to DNA. Blank plasmids were used to show self-activation, and pLacZi-pB42AD was used as a blank control group.

2.12. HPLC assay of carotenoid

The content of carotenoid compounds was determined by HPLC method. The chromatographic system used a Thermo DGLC dual

ternary ultra high performance liquid chromatography system (Waltham, MA, USA) with a DAD detector. The liquid phase used a YMC Carotenoid S-3 μm column, the injection volume was 2 μL, the column temperature was set to 40 °C, and the detection wavelength was set to 450 nm. Chromeleon7 software was used for preprocessing the raw data, including baseline filtering, peak identification, peak matching, retention time correction, and peak alignment, resulting in a data matrix containing retention time, mass-to-charge ratio, and peak intensity. The chromatogram was defined as the graph showing the distribution of the detection signal of the separated components over time, with the horizontal axis being the retention time and the vertical axis being the response value. Standard curves were fitted using Chromeleon7 software, and absolute quantification of the compounds was performed using the external standard method.

2.13. Dual-luciferase assays

The CDSs of *DMYB69* and *DMYB44* were amplified and connected to the pGreenII62-SK vector to generate effectors, respectively. The promoter fragments of *DmCHS1*, *DmF3H1*, *DmBCH2*, and *DmPSY* were cloned into pGreenII0800-LUC vector to generate reporters, respectively. The constructed plasmids of effectors and reporters were transformed into *Agrobacterium tumefaciens* strain GV3101 and cultured in the culture medium. The *A. tumefaciens* cells carrying the effector and reporter plasmids were mixed in a 1:1 ratio (*v*/*v*) and resuspended in an infiltration buffer. After culturing for 3 days, the infiltrated leaves were brushed with Beetle Luciferin Potassium Salt (Yeasen Biotechnology, shanghai, China), and the luciferase (LUC) signal was detected with BERTHOLD NightSHADE LB 985 (Berthold, Bad Wildbad, Germany). We validated the DmCHS1, DmF3H1, DmBCH2, and DmPSY promoter activity with dual-luciferase reporter assays involving multiple combinations. LUC and *Renilla* luciferase (REN) were detected using the Dual-Luciferase Reporter Assay System kit (Vazyme, Nanjing, China).

3. Results

3.1. Genome sequencing and assembly

The genome size, heterozygosity, and repeat content were evaluated using *k*-mer distribution analysis. A 19 *k*-mer frequency peak appeared at a depth of 34.78 (Supporting Information Fig. S1A, Table S3). The estimated genome size of *D. moniliforme* was 1.24 Gb, with heterozygosity and repeat content at 2.03% and 58.58%, respectively. For genome assembly, the PacBio Sequel II platform yielded 158.38 Gb of PacBio subreads with a subread length N50 of 24,454 bp (Table S2). Using PacBio data and next-generation sequencing, 500 contigs were assembled and anchored to 19 pseudo-chromosomes with 128.85 Gb of Hi-C clean reads (anchoring rate 100%; Supporting Information Tables S5 and S6). Juicebox software was employed to create a relationship map between interaction intensity and the location of contigs (Fig. S1B). The results indicated that interaction intensity at diagonal positions was higher than at non-diagonal positions in each group, reflecting high interaction intensity between adjacent sequences, consistent with the principle of Hi-C auxiliary assembly. A total of 19 pseudo-chromosomes were obtained, with a total length of 1.20 Gb, representing 96.77% of the estimated *D. moniliforme* genome, and a contig N50 of 3.97 Mb, with a GC content of 35.60% (Table 1). The quality and integrity of the assembled *D. moniliforme* genome were assessed by comparison with next-generation sequencing data and BUSCO analysis. The reads covered 99.16% of the genome sequence, with an average Illumina coverage depth of 53.74 × . A total of 1614 plant-specific orthologs were identified, of which 1409 were complete genes (Supporting Information Table S7). These results indicated that the genome sequencing and assembly were of high quality.

3.2. Genome annotation

The annotation results indicated that 71.59% of the *D. moniliforme* genome consisted of repeat sequences, amounting to 862.83 Mb (Supporting Information Table S8). The predominant type of repeat sequences was long terminal repeat (LTR) retrotransposons, which accounted for 39.19% of the total genome (Fig. 1C). Long interspersed nuclear elements (LINEs) made up 7.86% of the genome, while DNA transposons constituted 5.38%. A total of 27,231 high-quality protein-coding genes were annotated, with an average of 5.46 exons per gene (Supporting Information Table S9). The average lengths of genes and CDSs were 14,924 and 1122 bp respectively. Among all annotated genes, 90.94% and 90.67% had functional annotations in the NR and UniProt databases, respectively (Supporting Information Table S10). Functional annotations revealed that 74.41% of the genes were categorized by GO terms, and 33.61% of the genes were annotated to KEGG pathways. Additionally, 2952 ncRNA genes were identified (Supporting Information Table S11), including 105 miRNAs, 1412 snRNAs, 1019 rRNAs, and 416 tRNAs. The *D. moniliforme* genome included 195 collinear blocks (Fig. 1C), containing a total of 2360 gene pairs. Assessment tests showed that the annotation contained 91.4% of BUSCO genes (Table S7), suggesting a high degree of accuracy in the annotated genome.

3.3. Comparative genomic analyses

Gene families were identified using OrthoFinder, and comparative genomic analyses were performed to elucidate evolutionary relationships. The results revealed that the *D. moniliforme* genome contained 14,006 gene families, which included the 25,015 annotated protein-coding genes. Among these 14,006 gene families, 503 were specific to the *D. moniliforme* genome when compared to the genomes of *D. chrysotoxum*, *D. officinale*, and *D. nobile* (Fig. 1B). These 503 gene families encompassed 1229 genes annotated in databases. Functional GO enrichment analysis showed enrichment in ATP binding, integral component of

Table 1 Statistics of *D. moniliforme* genome assembly results.

Species	*D. moniliforme*
Genome size	1.20 Gb
Contig number	500
Contig N50	3.97 Mb
Contig N90	1.25 Mb
Max contig length	21.06 Mb
Min contig length	37,844 b
Scaffold N50	—
Hi-C anchored rate	100%
GC content	35.60%

membrane, and nucleic acid binding (Supporting Information Fig. S2A). KEGG functional analysis indicated enrichment in translation, carbohydrate metabolism, and lipid metabolism. These findings suggested that there were no significant differences among these four *Dendrobium* species in terms of functional genes related to secondary metabolites.

Phylogenetic analysis was conducted by screening single-copy genes with high homology and the results indicated a close relationship between *D. moniliforme* and *D. officinale*, which diverged approximately 17.42 Mya (Fig. 1D). The origin of *Dendrobium* orchids was estimated to have occurred around 22.67 Mya. Using CAFE v4.2, we analyzed the differences between the ancestral species and *D. moniliforme*, identifying 1748 expanded gene families and 2118 contracted gene families. The most significant functional categories included metal ion transport, nucleosome, integral component of membrane, oxidoreductase activity, and exonuclease activity (Fig. S2B). KEGG enrichment analysis suggested that rapidly evolving genes were associated with carbohydrate metabolism, lipid metabolism, and other secondary metabolites (Fig. S2C). These findings indicated that the differences among the four *Dendrobium* species were concentrated on growth-related functional genes, which may be linked to their significant morphological differences.

A comparative analysis of 4DTv values suggested that *D. moniliforme*, *D. officinale*, and *D. chrysotoxum* underwent two rounds of WGD events (Fig. 1E). The self-comparisons peaked at 0.309 for *D. moniliforme*, 0.314 for *D. chrysotoxum*, and 0.314 for *D. officinale*, indicating the first shared WGD event. A self-comparison peak value of 0.025 for *D. officinale* indicated the second shared WGD event. The *D. moniliforme*−*O. sativa* comparison peak was 0.74, corresponding to a divergence time of 114.84 Mya. The *D. moniliforme*−*V. planifolia* comparison peak was 0.336, associated with a divergence time of 72.74 Mya. The *D. chrysotoxum*−*D. moniliforme* comparison peak was 0.032, indicating a divergence time of 22.67 Mya. Comparisons of the 4DTv values suggested that the first shared WGD event occurred at approximately 66.9 Mya, predating the divergence time of *Dendrobium* (22.67 Mya) and close to the speciation time of *V. planifolia* (72.74 Mya). The second shared WGD event took place approximately 17.71 Mya. These results indicated that *Dendrobium* species have jointly experienced two WGD events. During their divergence, the types of gene families have not undergone significant changes; however, recent gene families related to morphological differences have shown varying degrees of expansion and contraction.

3.4. Transcriptome analysis

To further investigate the biosynthesis pathways of medicinal components, we obtained transcriptome data following MeJA treatment. The raw data were filtered using fastp to obtain valid data for subsequent analysis. The statistical results of the transcriptome data are presented in Supporting Information Table S12. The filtered transcriptome sequences were aligned to the reference genome (Supporting Information Table S13), with alignment rates for the six samples ranging from 93.54% to 95.85%. Correlation analysis between samples is shown in Supporting Information Fig. S3. The correlation between replicates (0.926−0.965) was significantly higher than that between different treatments (0.727−0.815). DEGs were identified between the treatment and control groups. Compared to the control group, 3673 genes were up-regulated and 4206 genes were down-regulated in the treatment group (Supporting Information Fig. S4A and S4B). GO and KEGG enrichment analyses were conducted on the DEGs. GO annotations were categorized into biological processes, cellular components, and molecular functions (Fig. S4C). In the cellular component category, the most enriched genes were related to membranes. In the molecular function category, the most enriched genes were related to catalytic activity. KEGG annotation results indicated that protein-coding genes were primarily associated with metabolic pathways, secondary metabolite biosynthesis, plant−pathogen interaction, and plant hormone signal transduction (Fig. S4D).

3.5. Identification of enzyme genes related to the medicinal component biosynthesis pathways

In the transcriptome data, genes related to the flavonoid biosynthesis pathway, such as *phenylalanine ammonia-lyase* (*PAL*), *4-hydroxylase* (*C4H*), *4-coumarate CoA ligase (4CL), chalcone synthase* (*CHS*), *chalcone isomerase* (*CHI*), *flavonone 3-hydroxylase* (*F3H*), *flavonol synthase* (*FLS*), *F3′H*, *F3′5′H*, *dihydroflavonol reductase* (*DFR*), and *anthocyanin synthase* (*ANS*), were identified (Fig. 2A, Supporting Information Table S14). After MeJA treatment, the expression levels of genes such as *PAL* (*XZSH0192080*, *XZSH0040260*), *C4H* (*XZSH0034140*), *4CL* (*XZSH0046700*, *XZSH0051590*), *F3′5′H* (*XZSH0034990*), and *ANS* (*XZSH0107140*) were significantly up-regulated, whereas the expression levels of genes such as *CHS* (*XZSH0270730*) and *CHI* (*XZSH0143280*) were significantly down-regulated. Genes related to the polysaccharide biosynthesis pathway, including *AKR1B*, *SORD*, *SLDH*, *HK*, *PMM*, *GMPP*, *GMDS*, *ALDO*, *TSTA3*, *TPI*, *DAK*, and *FUK*, were also identified (Fig. 2B). Following MeJA treatment, genes such as *AKR1B* (*XZSH0217300*) and *ALDO* (*XZSH0057240*, *XZSH0176770*) showed significantly up-regulation, while genes such as *SORD* (*XZSH0087920*), *HK* (*XZSH0196060*), *PMM* (*XZSH0001260*), *ALDO* (*XZSH0145910*), and *FUK* (*XZSH0193370*) exhibited significant down-regulation. Additionally, genes involved in the carotenoid biosynthesis pathway, such as *phytoene synthase* (*PSY*), *phytoene desaturase* (*PDS*), *zeta-carotene desaturase* (*ZDS*), *lycopene β-cyclase* (*LCY-β*), and *BCH2*, were identified (Fig. 2C). The expression level of the *PSY* (*XZSH0151520*) gene was significantly up-regulated.

Genes associated with the alkaloid biosynthesis pathway, including *DHS*, *DHQS*, *DHQ*, *SK*, *EPSPS*, *CS2*, *AS*, *DXS*, *ispG*, *ACAT*, *HMGCS*, *HMGCR*, *MVK*, *mvaK2*, *MVD*, *STR1*, and *FDPS*, were identified (Fig. 3). Following MeJA treatment, the expression levels of genes such as *DHS* (*XZSH0094000*), *DHQS* (*XZSH0206000*), *SK* (*XZSH0243080*), *EPSPS* (*XZSH0196730*), *CS2* (*XZSH0081100*), *DXS* (*XZSH0212430*), and *ACAT* (*XZSH0198690*) were significantly up-regulated, while the expression levels of genes such as *DXS* (*XZSH0065310*) and *HMGCR* (*XZSH0070500*) were significantly down-regulated. Comparative transcriptome analysis revealed that multiple genes involved in the biosynthesis pathways of flavonoids, polysaccharides, and alkaloids could be activated by MeJA.

3.6. Characteristics and classification of R2R3-MYB proteins

To identify the R2R3-MYB proteins in the *D. moniliforme* genome, an HMM search was performed using the MYB binding domain HMM profile, and local BLASTP queries against the genome were conducted using the R2R3-MYB proteins from *A. thaliana*. The R2R3-MYB proteins were validated using the Pfam and NCBI CDD databases, confirming 90 typical *D.*

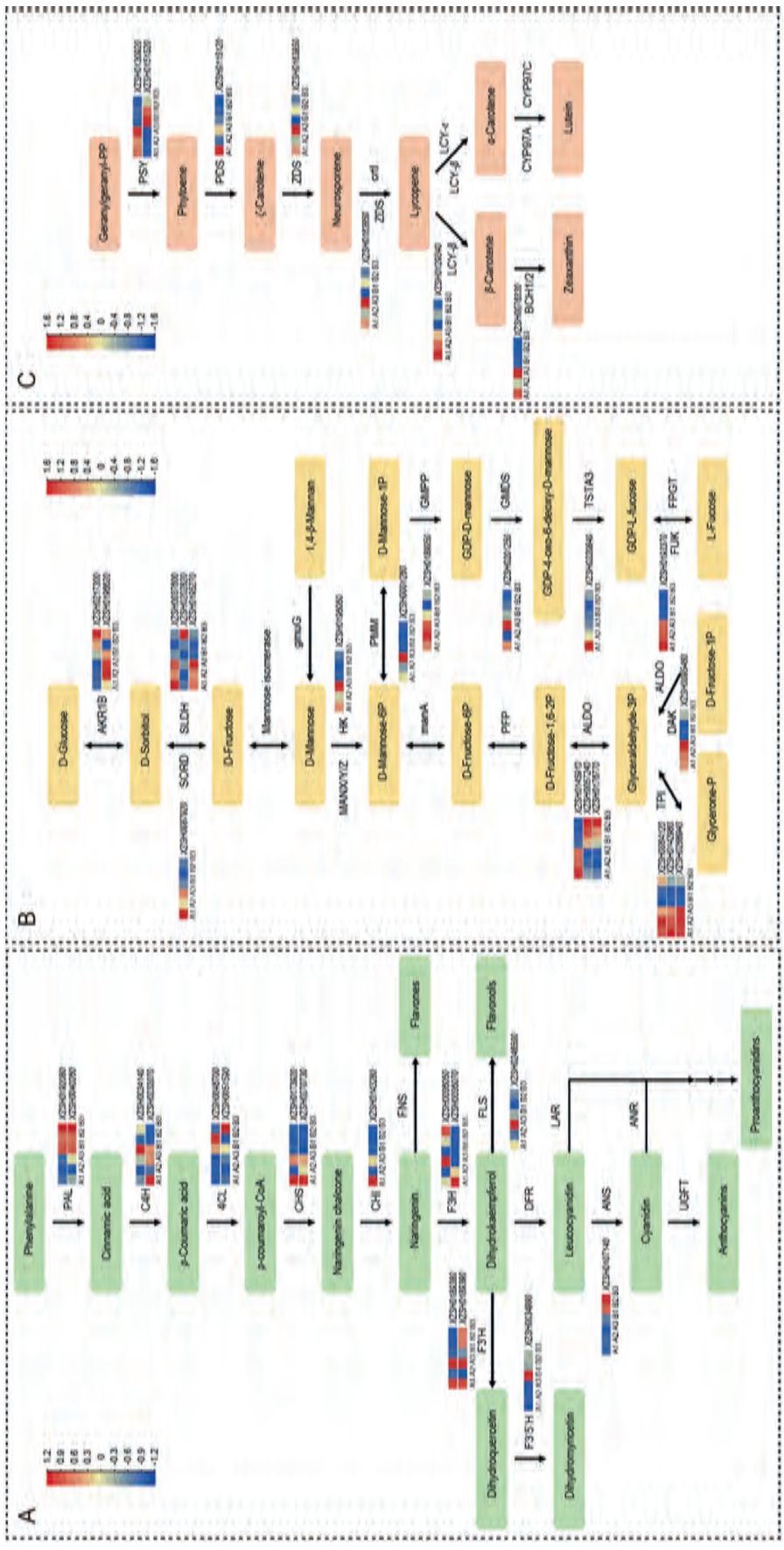

Figure 2 Biosynthesis pathway maps of medicinal components. (A) Flavonoid biosynthesis pathway; (B) Polysaccharide biosynthesis pathway; (C) Carotenoid biosynthesis pathway. The enzymes genes identified in the figure are annotated, and the heat map shows their relative expression levels in the transcriptome.

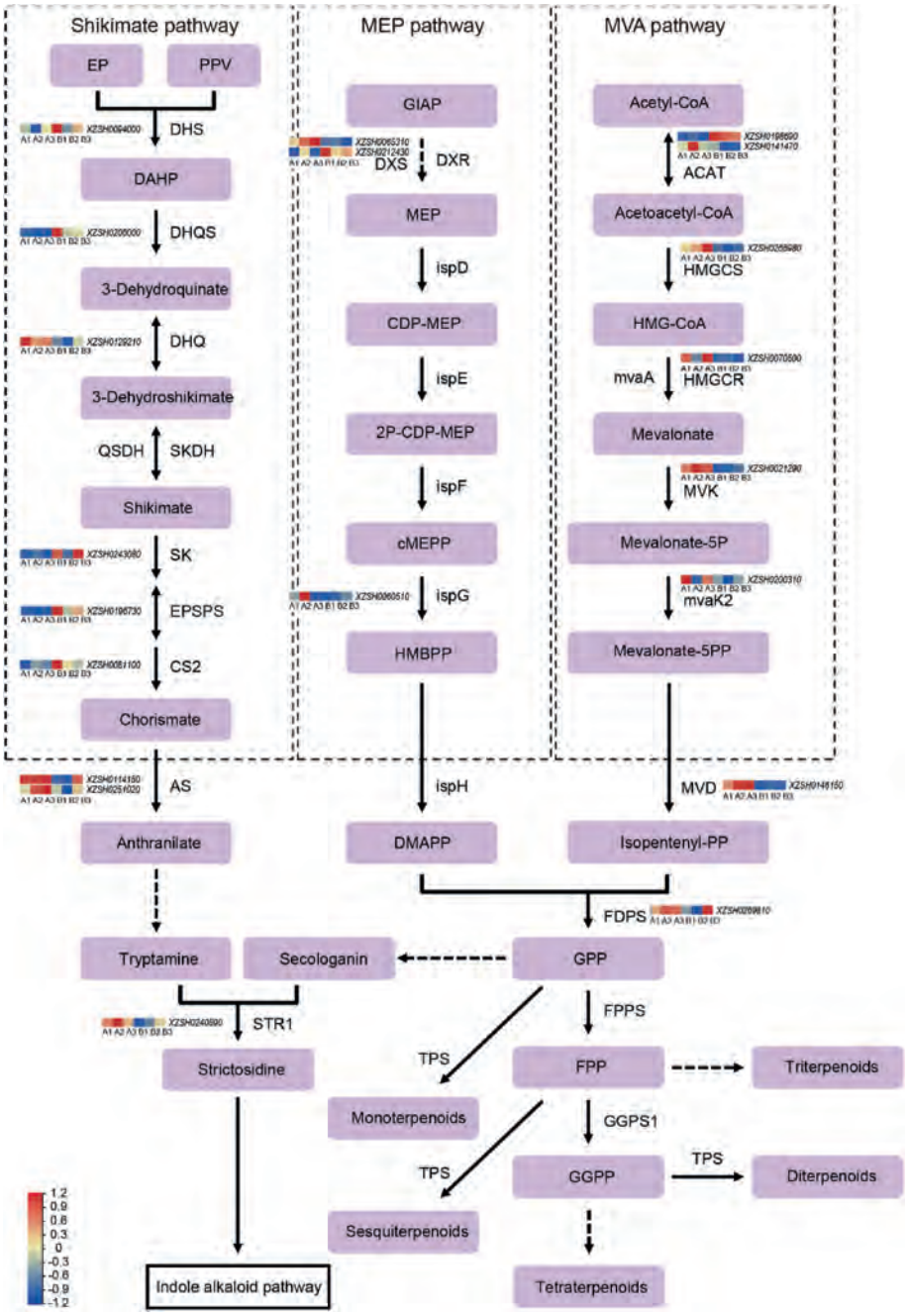

Figure 3 Biosynthesis pathway map of alkaloids and terpenoids. The enzymes genes identified in the figure are annotated, and the heat map shows their relative expression levels in the transcriptome. MEP, methylerythritol phosphate; MVA, mevalonate.

moniliforme R2R3-MYB proteins, named DMYB1-DMYB90 (Supporting Information Fig. S5, Supporting Information Table S15). Basic information revealed that the lengths of the R2R3-MYB proteins ranged from 164 to 586 amino acids (aa), with molecular weights ranging from 19.08 to 64.64 kDa and pI values between 4.9 and 10.38. The predicted grand average of hydropathicity (GRAVY) values of the R2R3-MYB proteins ranged from −1.063 to −0.285, indicating that all *D. moniliforme* R2R3-MYB proteins were hydrophilic. Instability indexes ranged from 33.19 to 81.09, and aliphatic indexes from 54.09 to 89.16. The transcriptome analysis revealed that following MeJA treatment, the expression levels of seven R2R3-MYB genes were up-regulated, namely *DMYB7*, *DMYB23*, *DMYB25*, *DMYB29*, *DMYB44*, *DMYB61*, and *DMYB85*, which were distributed across five different subgroups. Conversely, the expression levels of 16 R2R3-MYB genes were down-regulated, including *DMYB4*, *DMYB8*, *DMYB11*, *DMYB27*, *DMYB30*, *DMYB34*, *DMYB35*, *DMYB37*, *DMYB38*, *DMYB39*, *DMYB49*, *DMYB57*, *DMYB72*, *DMYB75*, *DMYB82*, and *DMYB84*, which were distributed across nine subgroups. Specifically, four genes were in subgroup 1 (SG1), three in SG13, and three in SG21.

To investigate the evolutionary relationship between *D. moniliforme* and *A. thaliana* R2R3-MYB proteins, the predicted 90 *D. moniliforme* R2R3-MYB proteins were aligned with 113

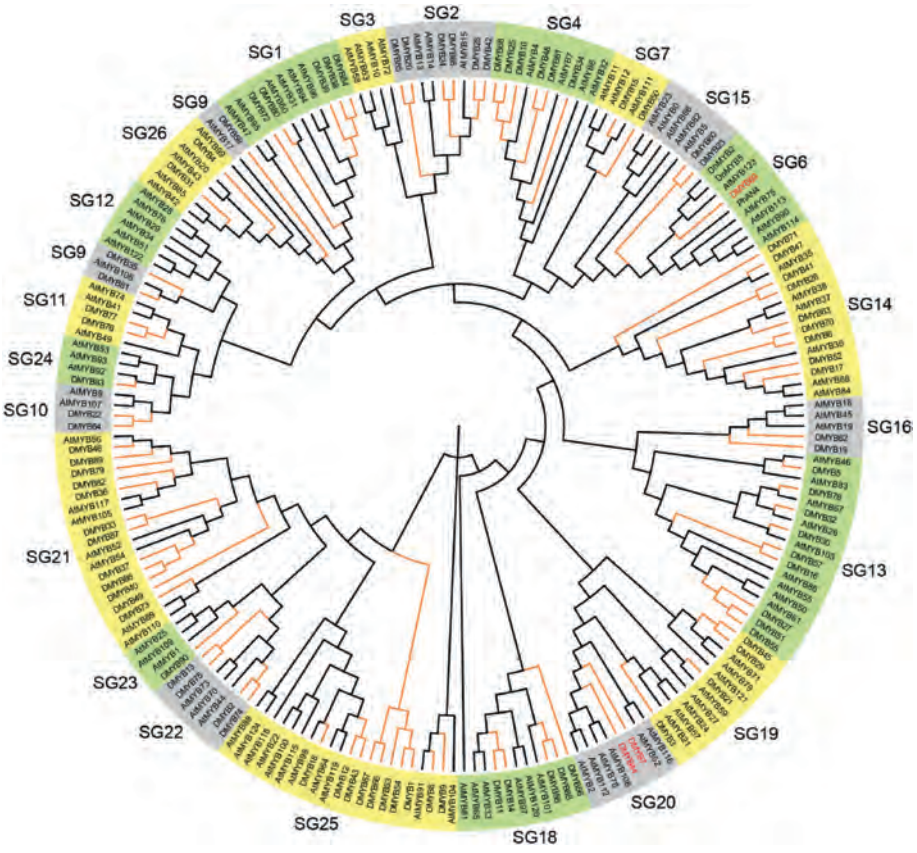

Figure 4 Phylogenetic analysis and classification results of R2R3-MYB proteins. SG, subgroup.

Arabidopsis R2R3-MYB proteins, DhMYB2[47], DoMYB5[48], and PhAN4[75], and an ML phylogenetic tree was constructed to infer their evolutionary relationships. Following the *Arabidopsis* classification[76,77], all R2R3-MYB proteins were categorized into 23 subgroups (Fig. 4). In the phylogenetic tree, SG21, SG22, SG23, and SG25 clustered together, consistent with the reference classification results[76]. Next were SG18, SG20, SG19, SG13, and SG16. Consistent with the reference, the remaining subgroups were primarily divided into two branches, the one of which included SG1, SG9, SG11, SG24, and SG10. Overall, the phylogenetic tree largely aligned with the structure of the reference, though some minor differences were observed.

3.7. Expansion and contraction of the R2R3-MYB genes

To investigate the evolutionary history of R2R3-MYB genes, we identified a total of 889 R2R3-MYB proteins in *O. sativa*, *A. shenzhenica*, *V. planifolia*, *P. equestris*, *D. chrysotoxum*, *D. nobile*, *D. officinale*, and *D. moniliforme*. The length of R2R3-MYB proteins in *D. chrysotoxum* ranged from 160 to 852 aa, with 98.9% (93/94) of the proteins having sequence lengths between 160 and 718 aa. In *D. nobile*, the protein lengths ranged from 161 to 1133 aa, with 99% (108/109) of the proteins measuring between 161 and 527 aa. For *D. officinale*, the lengths ranged from 160 to 1075 aa, with 98.8% (84/85) of the proteins spanning between 160 and 689 aa. Subsequently, the R2R3-MYB proteins were grouped and quantified in the aforementioned seven species based on their phylogenetic tree, informed by the ML tree and the grouping of *A. thaliana* (Supporting Information Table S16). The number of SG4+ R2R3-MYB proteins varied significantly among the species; for instance, *O. sativa* contained 37, while *D. nobile* and *D. moniliforme* had none. We speculated that R2R3-MYB proteins underwent multiple gene duplication events, with SG4+ genes being retained after functional degeneration. The SG23 had a small number of R2R3-MYB proteins, with only 1−2 copies in Orchidaceae and none in *D. chrysotoxum*. SG16 R2R3-MYB proteins were retained in *Dendrobium* and *Phalaenopsis* but degenerated in other Orchidaceae. Genes in SG6 and SG12 generally showed degeneration in Orchidaceae. Other SGs maintained relatively conserved numbers within Orchidaceae (Fig. 5B, Table S16).

3.8. R2R3-MYB gene clusters

To evaluate the conservation of R2R3-MYB proteins in *Dendrobium*, collinearity analysis was performed using the genomes of *O. sativa*, *V. planifolia*, *D. chrysotoxum*, *D. nobile*, *D. officinale*, and *D. moniliforme*. The renamed chromosome information is provided in Supporting Information Table S17. The analysis revealed seven collinear blocks between *O. sativa* and *V. planifolia* (Fig. 5C), containing 55 gene pairs. Between *V. planifolia* and *D. chrysotoxum*, 549 collinear blocks were identified, comprising 5414 gene pairs. There were 95 collinear blocks between *D. chrysotoxum* and *D. nobile*, containing 14,071 gene pairs. Between *D. nobile* and *D. officinale*, 124 collinear blocks were found, containing 16,588 gene pairs. Lastly, 156 collinear blocks were observed between *D. officinale* and *D. moniliforme*, containing 16,740 gene pairs. The R2R3-MYB gene clusters in

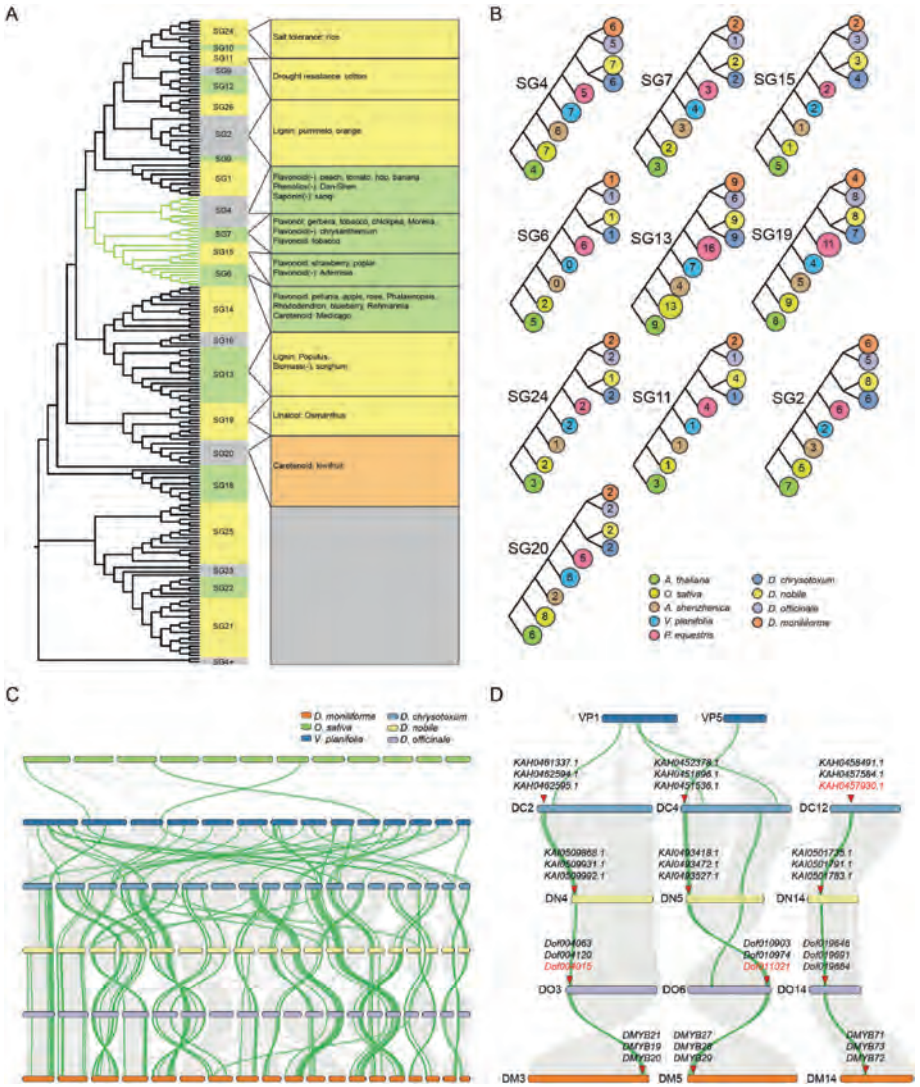

Figure 5 Results of the comparative analysis of R2R3-MYB proteins in the Orchidaceae family. (A) Validated functions of R2R3-MYBs across different subgroups. (B) Variation in the number of R2R3-MYB subgroups in Orchidaceae, including SG4, SG7, SG15, SG6, SG13, SG19, SG24, SG11, SG2, SG20. (C) Collinearity analysis among Orchidaceae genomes, with R2R3-MYB genes indicated by green lines. (D) Analysis of R2R3-MYB gene clusters. SG, subgroup.

different species were identified, revealing three shared gene clusters within the *Dendrobium* genus (Fig. 5D, Supporting Information Table S18). The first gene cluster included genes from SG19, SG2, and SG16; the second cluster comprised genes from SG13, SG2, and SG19; and the third cluster contained genes from SG14, SG21, and SG1.

3.9. Identification and overexpression assay of DMYB69

Functional verification studies of R2R3-MYB TFs, which are positively associated with flavonoid biosynthesis, are primarily focused on SG6 (Fig. 5A) and involve various species. DMYB69, a member of SG6, was identified as a candidate TF for regulating flavonoid biosynthesis in subsequent experiments. *DMYB69* was located between 49,904,827 and 49,907,645 bp on chromosome DM13, comprising three exons and two introns (Fig. 6A). To determine the subcellular localization of the DMYB69 protein, a GFP reporter construct was created to express the DMYB69-GFP fusion protein, with the same vector expressing only GFP used as a control. The fusion construct and the GFP-only control were then transformed into the lower epidermal cells of tobacco. As expected, DMYB69 specifically localized to the nucleus (Fig. 6B), consistent with its hypothesized role as a TF.

An overexpression plasmid for *DMYB69* was constructed and transformed into *D. moniliforme* leaves, using *Agrobacterium* without the overexpression plasmids as a blank control. The qPCR results indicated that the relative expression level of *DMYB69* was 9.28-fold higher, significantly up-regulated on Day 3 (*t*-test, $P < 0.05$) (Fig. 6C). In the flavonoid biosynthesis pathway, the relative expression levels were as follows: *DmANS* at 0.92-fold, *DmCHS1* at 6.71-fold, *DmCHS2* at 1.17-fold, *DmDFR* at 1.82-fold, *DmF3′5′H* at 1.54-fold. The relative expression levels of *DmF3′H1*, *DmF3′H2*, *DmF3H1*, and *DmF3H2* were 1.04-, 2.9-, 2.34-, and 0.93-fold, respectively. Both *DmCHS1* (*t*-test, $P < 0.05$) and *DmF3H1* (*t*-test, $P < 0.01$) were significantly

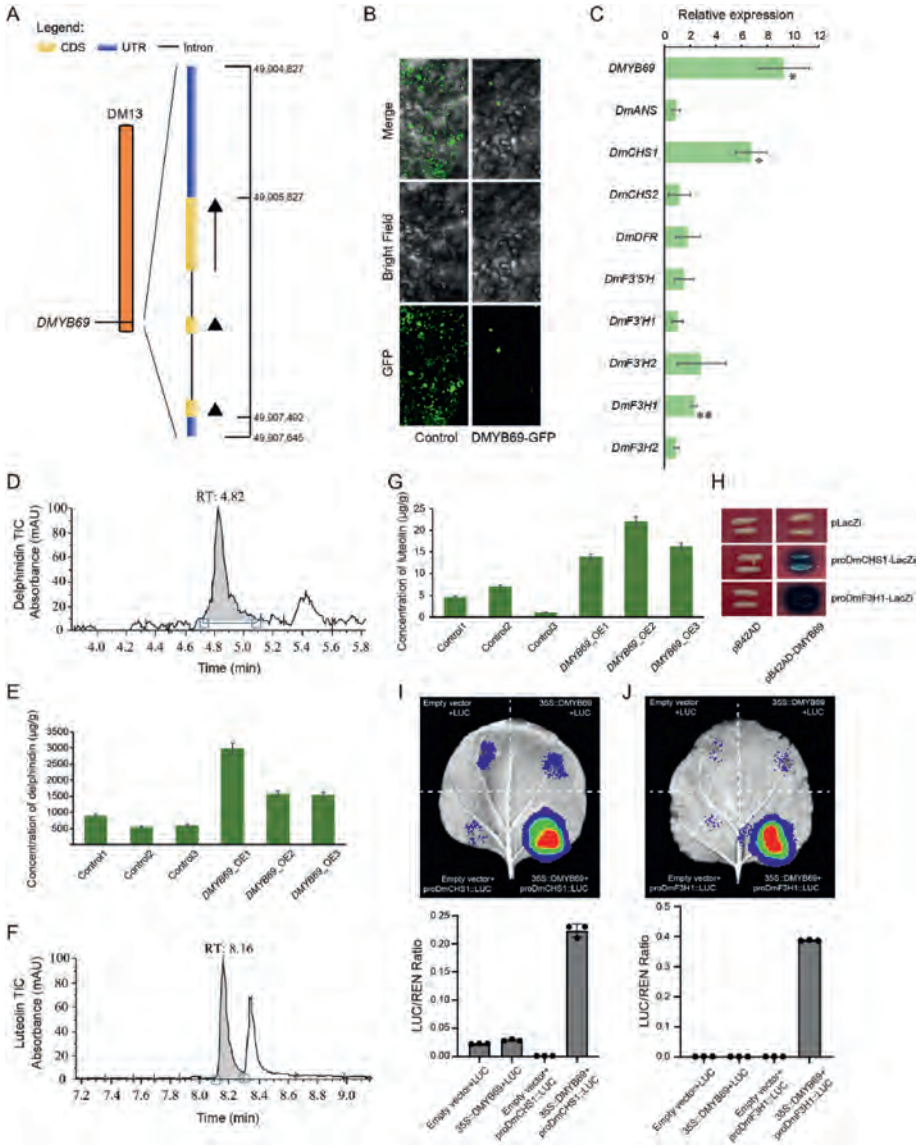

Figure 6 Function validation results of DMYB69. (A) Chromosomal localization results of *DMYB69*. (B) Subcellular localization results of DMYB69. (C) Expression levels of *DMYB69* and enzyme genes in overexpression experiments. (D) UPLC−MS/MS TIC graphs of delphinidin. (E) Absolute content of delphinidin detected by UPLC−MS/MS assays. (F) UPLC−MS/MS TIC graphs of luteolin. (G) Absolute content of luteolin detected by UPLC−MS/MS assays. (H) Results of DMYB69 yeast one-hybrid assays. (I) LUC experiment demonstrating the binding of DMYB69 to proDmCHS1. (J) LUC experiment demonstrating the binding of DMYB69 to proDmF3H1. CDS, coding DNA sequence; GFP, green fluorescent protein; LUC, luciferase; OE, overexpression; REN, *Renilla* luciferase.

up-regulated. These results suggested that DMYB69 functioned as a TF to up-regulate the expression of *DmCHS1* and *DmF3H1*.

3.10. Flavonoid contents detected by UPLC−MS/MS

UPLC−MS/MS assays were conducted to quantify flavonoids, including anthocyanins (delphinidin, malvidin, peonidin, cyanidin, and cyanidin-3-galactoside), flavones (luteolin), and flavonols (rutin and quercetin) (Supporting Information Tables S19 and S20), in the leaves of *DMYB69*-overexpression (*DMYB69*-OE) *D. moniliforme* and the blank control group. The results indicated that the delphinidin content in the *DMYB69*-OE line (2038.1 μg/g) was 3-fold that of the blank control group (675.6 μg/g), demonstrating a significant increase (Fig. 6D and E). The luteolin content in the *DMYB69*-OE line (17.36 μg/g) was 4.1-fold that of the blank control group (4.24 μg/g), showing a significant increase (Fig. 6F and G). Significant increases were also observed in the cyanidin-3-galactoside and rutin contents (Supporting Information Fig. S6). Conversely, the malvidin content in the *DMYB69*-OE line (1.28 μg/g) was only 0.35-fold that of the blank control group (3.69 μg/g), indicating a significant decrease. The quercetin content varied greatly between the overexpression group and the blank control group, with no significant difference observed. These results suggested that the overexpression of *DMYB69* can up-regulate the contents of four flavonoids in *D. moniliforme*: delphinidin, cyanidin-3-galactoside, luteolin, and rutin.

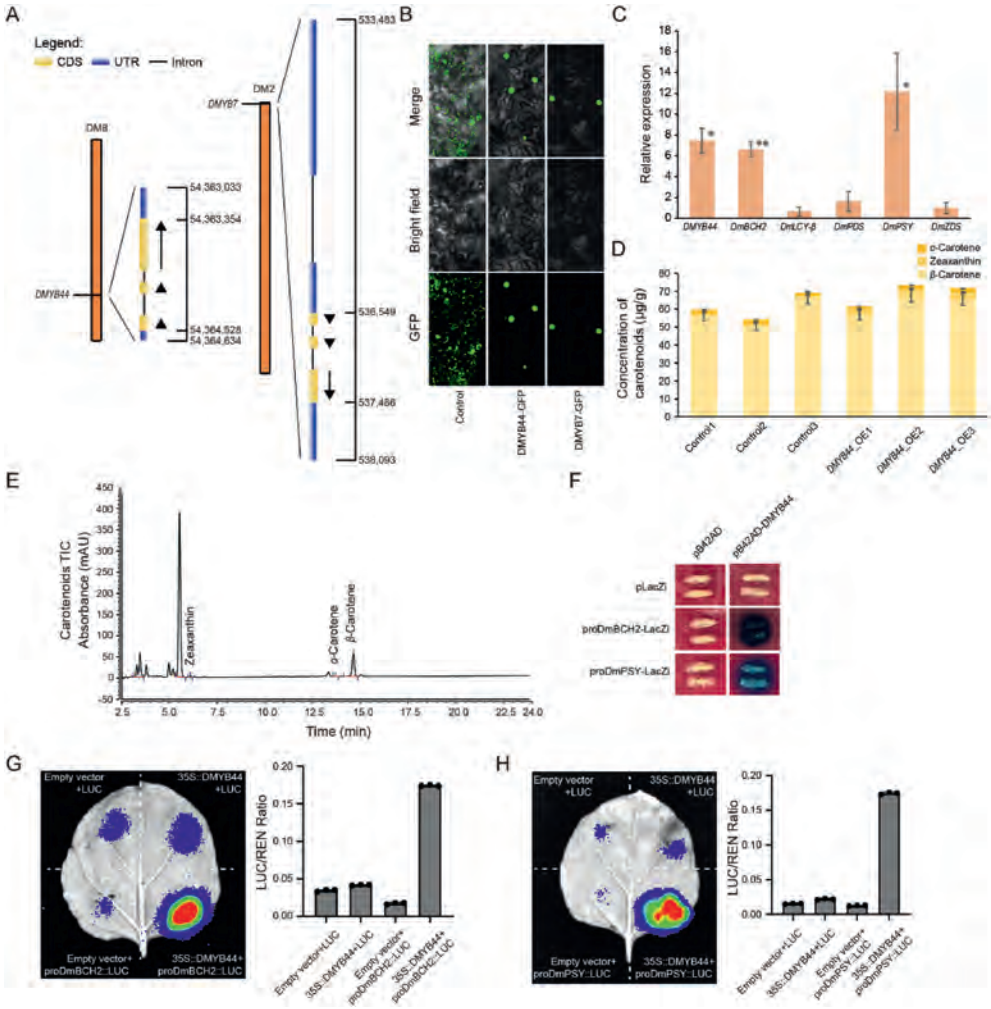

Figure 7 Function validation results of DMYB44. (A) Chromosomal localization results of *DMYB44* and *DMYB7*. (B) Subcellular localization results of DMYB44 and DMYB7. (C) Expression levels of *DMYB44* and enzyme genes in overexpression experiments. (D) Absolute content of carotenoid compounds detected by HPLC assays. (E) HPLC TIC graph of carotenoid standard substances, including zeaxanthin, α-carotene, and β-carotene. (F) Results of DMYB44 yeast one-hybrid assays. (G) LUC experiment demonstrating the binding of DMYB44 to proDmBCH2. (H) LUC experiment demonstrating the binding of DMYB44 to proDmPSY. CDS, coding DNA sequence; GFP, green fluorescent protein; LUC, luciferase; OE, overexpression; REN, *Renilla* luciferase.

3.11. Yeast one-hybrid and dual-luciferase assays of DMYB69

To verify the direct transcriptional activation of DMYB69 on *DmCHS1* and *DmF3H1*, we conducted yeast one-hybrid and dual-luciferase assays. MYB binding sites were identified in the 2000 bp upstream regions of both *DmCHS1* and *DmF3H1*. The results revealed 11 non-redundant candidate MYB binding sites in the upstream promoter region of *DmCHS1* (Supporting Information Fig. S7A) and six non-redundant candidate MYB binding sites in the upstream promoter region of *DmF3H1*. Yeast one-hybrid assays confirmed the interaction of DMYB69 with the upstream promoters of *DmCHS1* and *DmF3H1*. The blank control group showed no color change, and there was no self-activation of the pB42AD-DMYB69, proDmCHS1-LacZi, or proDmF3H1-LacZi plasmids. The experimental group demonstrated that DMYB69 directly binds to the upstream promoter regions of *DmCHS1* and *DmF3H1* (Fig. 6H). Specifically, the MYB binding region of *DmCHS1* ranged from −1481 to −1091 bp, and the MYB binding region of *DmF3H1* ranged from −690 to −456 bp (Fig. S7B).

The promoters of *DmCHS1* and *DmF3H1* were fused into the pGreenII0800-LUC vector to generate reporter plasmids. The effector plasmid 62SK-DMYB69 was constructed by inserting the *DMYB69* CDS into the binary vector pGreenII62-SK. Dual-luciferase activity was measured in tobacco leaves. We observed that, when treated with empty plasmids, the *in vivo* imagers showed no luminescence in tobacco leaves, and the LUC/REN ratios were relatively low. Co-transformation of 62SK-DMYB69 with either proDmCHS1-0800-LUC or proDmF3H1-0800-LUC resulted in higher LUC/REN ratios, stronger luminescence signals, and increased LUC activities in tobacco leaves (Fig. 6I and J). Both yeast one-hybrid and dual-luciferase assays demonstrated that DMYB69 can directly bind to the promoters of *DmCHS1* and *DmF3H1*, thereby activating their transcription.

3.12. Identification and overexpression assay of DMYB44

The R2R3-MYB TFs in SG20 have been validated to positively regulate carotenoid biosynthesis in kiwifruit[30]. DMYB44 and

DMYB7, identified as candidate TFs for carotenoid biosynthesis regulation and subsequently undergoing functional verification, belonged to SG20. *DMYB44* was located on chromosome DM8 between 54,363,033 and 54,364,634 bp, containing three exons and two introns (Fig. 7A). *DMYB7* was located on chromosome DM2 between 533,483 and 538,093 bp, also comprising three exons and two introns. To determine the subcellular localization of the DMYB44 and DMYB7 proteins, GFP reporter constructs expressing DMYB44-GFP and DMYB7-GFP fusion proteins were created, with a vector expressing only GFP serving as a control. These constructs and the GFP-only control were then transformed into tobacco lower epidermal cells. As expected, DMYB44 and DMYB7 specifically localized to the nucleus (Fig. 7B), consistent with their proposed role as TFs.

The effects of DMYB44 and DMYB7 on the expression of enzyme genes in the carotenoid biosynthesis pathways were investigated. Overexpression plasmids for the DMYB44 and DMYB7 TFs were constructed and introduced into the leaves of *D. moniliforme* using *Agrobacterium*; *Agrobacterium* without the overexpression plasmids served as a blank control. The qPCR results revealed that the relative expression levels of *DMYB44* were significantly up-regulated by 7.45-fold (*t*-test, $P < 0.05$) (Fig. 7C). In the carotenoid biosynthesis pathway, the relative expression levels of *DmBCH2* increased by 6.65-fold, *DmLCY-β* decreased to 0.68-fold, *DmPDS* increased by 1.64-fold, *DmPSY* increased by 12.16-fold, and *DmZDS* showed no significant change at 1.01-fold. Both *DmBCH2* (*t*-test, $P < 0.01$) and *DmPSY* (*t*-test, $P < 0.05$) were significantly up-regulated. When *DMYB7* was overexpressed, no significant changes were observed in the expression of the enzyme genes in the carotenoid biosynthesis pathway. Compared to other SG20 TFs, DMYB7 had a 16 aa deletion in the conserved region of the MYB domain (Fig. S7C), which may impact its original function. These findings indicated that overexpression of *DMYB44* can significantly up-regulate the expression of *DmBCH2* and *DmPSY*.

3.13. Carotenoid contents detected by HPLC

HPLC assays were employed to quantify carotenoids, including α-carotene, zeaxanthin, and β-carotene, in the leaves of the *DMYB44*-OE lines and the control group. The results indicated a significant increase in zeaxanthin content in the *DMYB44*-OE line (3.84 μg/g), which was 2.49 times higher than that of the control group (1.54 μg/g) (Fig. 7D and E).

3.14. Yeast one-hybrid and dual-luciferase assays of DMYB44

To validate the direct transcriptional activation of DMYB44 on *DmBCH2* and *DmPSY*, we performed yeast one-hybrid and dual-luciferase assays. We examined all MYB binding sites within the 2000 bp upstream regions of *DmBCH2* and *DmPSY*. The results identified eight distinct candidate MYB binding sites in the upstream promoter region of *DmBCH2* (Fig. S7D) and six unique MYB binding sites in the upstream promoter region of *DmPSY*. The yeast one-hybrid assays confirmed the interaction between DMYB44 and the upstream promoters of *DmBCH2* and *DmPSY*. The negative control group showed no color change, indicating no self-activation of the pB42AD-DMYB44, proDmBCH2-LacZi, and proDmPSY-LacZi plasmids. In contrast, the experimental group demonstrated that DMYB44 could directly bind to the upstream promoter regions of *DmBCH2* and *DmPSY* (Fig. 7F). The MYB binding site in the DmBCH2 promoter was located between −1785 and −1509 bp, while the MYB binding site in the DmPSY promoter was between −1830 and −1527 bp (Fig. S7E).

The dual-luciferase assay results revealed that treatment with empty plasmids resulted in no luminescence in tobacco leaves, and the LUC/REN ratios were relatively low. However, co-transformation with 62sk-DMYB44 and either proDmBCH2-0800-LUC or proDmPSY-0800-LUC resulted in higher LUC/REN ratios and significantly stronger luminescence signals and elevated LUC activity in tobacco leaves (Fig. 7G and H). Both yeast one-hybrid and dual-luciferase assays confirmed that DMYB44 can directly bind to the promoters of *DmBCH2* and *DmPSY*, thereby activating their transcription.

Table 2 Statistics of *Dendrobium* genome annotation results.

Species	Genome size	Gene number	Contig N50	BUSCO annotation (%)
D. moniliforme (this study)	1.20 Gb	27,231	3.97 Mb	91.4
D. officinale	1.23 Gb	27,631	1.44 Mb	93.5
D. chrysotoxum	1.37 Gb	30,044	1.54 Mb	87.5
D. nobile	1.19 Gb	29,476	1.54 Mb	93.43

BUSCO, Benchmarking Universal Single-copy Ortholog.

4. Discussion

4.1. The high-quality chromosome-level genome of D. moniliforme enriches the knowledge of orchids

Since the publication of the *Phalaenopsis aphrodite* genome in 2018[78], nine chromosomal-level orchid genomes have been published, including those of *D. huoshanense*[79], *D. chrysotoxum*[80], *Cymbidium sinense*[81], *Gastrodia elata*[82], *Dendrobium officinale*[83], *D. nobile*[84], *Vanilla planifolia*[85], *Platanthera guangdongensis*, and *Platanthera zijinensis*[86]. In this study, we obtained a high-quality genome of *D. moniliforme*, with a total genome size of 1.20 Gb and a contig N50 of 3.97 Mb, further enriching the genome resources of orchids (Table 1). The BUSCO assessment of genome annotation was 91.4%. The evolutionary relationships among orchid species have garnered significant interest from evolutionary biologists and botanists[87,88]. This research suggests that orchids, characterized by unique floral morphologies and high ornamental value, originated between 80.38 and 114.84 Mya (Fig. 1D). By constructing a phylogenetic tree, we inferred the relationships of *D. moniliforme* with six other orchid species: (i) the seven orchid species form a monophyletic group, and (ii) *D. moniliforme* and *D. officinale* are sister clades. Previously, four chromosome-level genomes have been published for *Dendrobium* species (*D. officinale*, *D. chrysotoxum*, *D. nobile*, and *D. huoshanense*), among which the assemblies of *D. officinale* (contig N50: 1.44 Mb), *D. chrysotoxum* (contig N50: 1.54 Mb), and *D. nobile* (contig N50: 1.54 Mb) genomes are of high quality (Table 2). The *D. moniliforme* genome assembled in this study showed further improvement in quality, with a contig N50 of 3.97 Mb.

There is substantial evidence suggesting that WGD events and transposable elements (TEs) amplification are primary drivers of genome expansion[89,90]. In this study, 4DTv values indicate that the genus *Dendrobium* underwent two rounds of WGD events, occurring approximately 17.71 and 66.9 Mya (Fig. 1E). TEs,

particularly LTR retrotransposons, constitute a significant portion of angiosperm genomes[91]. Numerous studies have reported that LTR retrotransposons can alter the expression of adjacent genes, thereby influencing the phenotype of the species[92,93]. For instance, in *Brassica oleracea*, the insertion of a retrotransposon into the *BoCYP704B1* gene, which is involved in a distinct developmental process, suppresses sporophyte formation and results in male sterility[94]. TEs comprise 71.59% of the *D. moniliforme* genome, closely resembling the 76.77% observed in the *D. officinale* genome[83]. Notably, LTR elements constitute a large proportion of TEs, accounting for 54.74%, while other TEs represent only 45%. There are 1877 protein-coding genes that contain LTR sequences within their exons. Additionally, eight miRNAs, 30 rRNAs, 297 snRNAs, and 55 tRNAs contain LTR sequences, suggesting that TEs play a crucial role in the functional gene evolution of *D. moniliforme*.

4.2. Combining genome and transcriptome to explore biosynthesis pathway enzyme genes of medicinal components

D. moniliforme contains various medicinal components, including polysaccharides, alkaloids, flavonoids, bibenzyls, and several trace mineral elements[95,96]. Recent studies have indicated that these components possess antioxidant, anti-inflammatory, and therapeutic properties for osteoporosis[97]. *In vitro*, *D. moniliforme* demonstrates antioxidant effects, while *in vivo*, it exhibits renoprotective effects in high-fat diet-treated mice by lowering serum glucose levels, total cholesterol concentration, and renal lipid accumulation[95]. Numerous investigations have confirmed that the phytohormone MeJA enhances the biosynthesis of secondary metabolites in medicinal plants[98]. In *Dendrobium* species, MeJA induces alkaloid biosynthesis by up-regulating the expression of alkaloid synthesis genes[99,100]. This study investigates enzyme genes related to the biosynthesis pathways of medicinal components through transcriptome data treated with MeJA.

The flavonoid biosynthesis pathway is a complex process that originates in the phenylalanine metabolic pathway[27,101-103]. Initially, PAL, C4H, and 4CL catalyze the decomposition of phenylalanine to form 4-coumaroyl-CoA, the initial substrate for flavonoid biosynthesis (Fig. 2A). CHS then catalyzes the production of chalcones from 4-coumaroyl-CoA[104]. Subsequently, CHI converts chalcones to naringenin. FNS catalyzes the conversion of naringenin into various flavones. F3H catalyzes the production of dihydrokaempferol[105], which is further converted into various flavonols by FLS. Anthocyanins are further methylated, glycosylated, and acylated through the continuous catalysis of DFR and ANS, resulting in various anthocyanins[106,107]. Most of the genes encoding enzymes involved in flavonoid biosynthesis were identified through similarity comparison. After MeJA treatment, the expression levels of genes such as *PAL*, *C4H*, *4CL*, *F3′5′H*, and *ANS* were up-regulated, while the expression levels of *CHS* and *CHI* were down-regulated.

Polysaccharides are natural macromolecules composed of multiple monosaccharide units and are structurally diverse[108]. The primary components of *Dendrobium* polysaccharides include glucose and mannose, with additional constituents such as arabinose, galacturonic acid, xylose, rhamnose, and galactose[109,110]. Extensive pharmacological research has demonstrated that rhizome polysaccharides exhibit various functions, including antioxidant, anti-herpes, antibacterial, anticancer, hypoglycemic, and hypolipidemic activities, as well as lung protection and antiviral effects[111]. Previous studies have also indicated that polysaccharides can act as compatible solutes, enhancing plant tolerance to abiotic stresses such as drought and salt stress[112]. Through comparative analysis, most of the polysaccharide biosynthesis enzyme genes were identified. Following MeJA treatment, the expression levels of certain genes, such as *AKR1B* and *ALDO*, were up-regulated, while the expression levels of *SORD*, *HK*, *PMM*, and *FUK* were down-regulated.

The initial step in the carotenoid biosynthesis pathway is catalyzed by the enzyme PSY, producing the colorless carotenoid 15-*cis*-phytoene (Fig. 2C)[30]. This 15-*cis*-phytoene is desaturated by PDS to yield 9,15,9′-tri-*cis*-ζ-carotene, which is then converted to ζ-carotene by ζ-carotene isomerase (ZISO) and/or photoisomerization[113]. Fig. 2C illustrates the two consecutive desaturation reactions mediated by ZDS, which synthesize neurosporene and ultimately prolycopene[114]. Subsequently, LCY-β forms β-rings at both ends of the lycopene molecule to produce beta-carotene, while it collaborates with lycopene ε-cyclase (LCY-ε) to form α-carotene. Carotenoid β-hydroxylase (BCH1/2) hydroxylates β-carotene to produce zeaxanthin, and CYP97A and CYP97C hydroxylate α-carotene to produce lutein[115]. After MeJA treatment, only *PSY* expression level was up-regulated.

Early research on the pharmacological effects of *Dendrobium* revealed that *Dendrobium* alkaloids enhance human immunity and benefit the stomach[116]. The alkaloid biosynthesis pathway involves the mevalonate (MVA), methylerythritol phosphate (MEP), and Shikimate pathways (Fig. 3). Products from the MVA and MEP pathways are catalyzed by FDPS to produce GPP, which is subsequently catalyzed by various enzymes to form different types of terpene alkaloids. GPP is further catalyzed by multiple intermediate enzymes to form secologanin, which, along with tryptamine from the Shikimate pathway, is catalyzed by STR1 to form strictosidine, thereby initiating the biosynthesis of indole alkaloids[117]. Through similarity comparison, most of the alkaloid biosynthesis enzyme genes were identified. After MeJA treatment, the expression levels of several genes, such as *DHS*, *DHQS*, *SK*, *EPSPS*, *CS2*, *DXS*, and *ACAT*, were up-regulated. The genome and transcriptome were analyzed to identify enzyme genes in the biosynthesis pathways of flavonoids, polysaccharides, carotenoids, and alkaloids, some of which can be up-regulated by MeJA activation.

4.3. SG6 is a conserved R2R3-MYB subgroup that up-regulates flavonoid biosynthesis

MYB TFs, which are widely distributed in plants, play a significant role in regulating various biological processes[118]. These MYBs can function as transcription activators or repressors to control the transcription of downstream genes. Among the different types of plant MYB proteins, R2R3-MYB proteins are the most predominant[25]. These R2R3-MYB proteins are involved in regulating a variety of plant-specific biological processes, including primary and secondary metabolism, hormone responses, environmental stress responses, disease resistance, and leaf morphogenesis[119,120].

We have summarized the functional studies of R2R3-MYB TFs and identified that their functions, upon complete functional verification, are mainly grouped into SG24, SG11, SG2, SG4, SG7, SG15, SG6, SG13, SG19, and SG20 (Fig. 5A). The functions of SG24 and SG11 are associated with abiotic stress[121,122]. SG2 and SG13 are linked to lignin synthesis[123-126]. The R2R3-MYB TFs in SG19 have been verified to be involved in species-specific linalool synthesis[127]. Those in SG20 have been

confirmed to positively regulate carotenoid biosynthesis in kiwifruit[30].

R2R3-MYB TFs related to flavonoid biosynthesis are concentrated in SG4, SG7, SG15, and SG6. Specifically, the R2R3-MYB TFs in SG4 have been shown to negatively regulate flavonoid compounds in peach, tomato, hop, and banana[128-131]. Additionally, these TFs in SG4 are negatively correlated with the synthesis of phenolics and saponin[77,132]. The R2R3-MYB TFs in SG7 and SG15 have varying regulatory effects on flavonoid biosynthesis across different species[101,133-135]. Moreover, the TFs in SG7 positively regulate the biosynthesis of flavonols in gerbera, tobacco, chickpea, and *Morella*[11,12,136,137]. Functional verification studies indicate that R2R3-MYB TFs positively related to flavonoid biosynthesis are primarily concentrated in SG6, involving species such as petunia, apple, rose, *Phalaenopsis*, *Rhododendron*, blueberry, and *Rehmannia*[26,28,75,138-141].

The R2R3-MYB TFs associated with the regulation of flavonoid biosynthesis are grouped into a single clade, which includes SG4, SG7, SG15, and SG6. Within *Dendrobium*, the SG4 branch possesses a relatively abundant number of R2R3-MYB TFs (Fig. 5B), potentially indicating a regulatory function for secondary metabolites beyond flavonoids. SG7 and SG15 consistently maintain one to four copies within *Dendrobium*, with functions related to flavonoid compounds, though their performance varies across different species. SG6, containing only one copy within *Dendrobium*, likely contributes to its stable positive regulation of flavonoid biosynthesis. These findings suggest that SG6 is a conserved R2R3-MYB subgroup with a role in positively regulating flavonoid biosynthesis, whereas SG4, SG7, and SG15 within the same clade may represent the results of functional divergence.

4.4. DMYB69 directly up-regulates flavonoid biosynthesis in D. moniliforme

DMYB69, belonging to the SG6 and homologous to PhAN4, promotes flavonoid accumulation by up-regulating the expression of flavonoid biosynthesis genes in the *Petunia hybrid*. DMYB69 is localized in the nucleus (Fig. 6B), indicating its role as a TF. Overexpression of *DMYB69* in *D. moniliforme* leaves significantly up-regulated the expression levels of *DmCHS1* and *DmF3H1* (Fig. 6C). UPLC−MS/MS assays revealed a significant increase in flavonoid content, including anthocyanins (delphinidin, cyanidin-3-galactoside), flavones (luteolin), and flavonols (rutin). Notably, DMYB69 had the most pronounced effect on delphinidin levels. Promoter analysis of *DmCHS1* and *DmF3H1* showed a significant number of MYB binding sites. Yeast one-hybrid and dual-luciferase assays demonstrated that DMYB69 directly binds to the upstream promoter fragments of *DmCHS1* and *DmF3H1*. Our findings align with previous studies; for instance, overexpression of *PhAN4* increased anthocyanin accumulation by upregulating genes such as *CHSA*, *CHSI*, *F3H*, *F3′H*, *F3′5′H*, *DFR*, and *ANS*. CRISPR/Cas9-mediated mutation of *PhAN4* resulted in the absence of *Petunia hybrid* corolla tube venation[75]. In apples, MdMYB90-like regulated anthocyanin biosynthesis both directly, by activating anthocyanin biosynthesis genes, and indirectly, by activating other TFs involved in anthocyanin biosynthesis[26]. Overexpression of *RcMYB1* significantly promoted anthocyanin accumulation in white rose petals[28]. Collectively, our results demonstrate that the DMYB69 TF directly up-regulates flavonoid biosynthesis in *D. moniliforme*.

4.5. DMYB44 directly up-regulates the carotenoid biosynthesis

R2R3-MYB proteins involved in carotenoid regulation are present in SG20. In SG20, AdMYB7 modulates carotenoid and chlorophyll (Chl) pigment accumulation in tissues by activating the transcription of metabolic pathway genes in kiwifruit[30]. DMYB44, homologous to AdMYB7, also modulates carotenoid and Chl pigment accumulation in kiwifruit. DMYB44 is localized in the nucleus (Fig. 7B), indicating its function as a TF. Transient overexpression of *DMYB44* in *D. moniliforme* leaves significantly up-regulated the expression levels of *DmBCH2* and *DmPSY* (Fig. 7C). HPLC assay results indicated a significant increase in zeaxanthin content. Analysis of the promoter regions of *DmBCH2* and *DmPSY* revealed numerous MYB binding sites. Yeast one-hybrid and dual-luciferase assays demonstrated that DMYB44 directly binds to upstream promoter fragments of *DmBCH2* and *DmPSY*. These results collectively demonstrate that the DMYB44 TF can directly up-regulate carotenoid biosynthesis.

5. Conclusions

This study presents a high-quality chromosome-level genome of *D. moniliforme*, with a genome size of 1.2 Gb and a contig N50 of 3.97 Mb. The genome annotation was evaluated at 91.4% using BUSCO. Comparative genomic analysis indicated that orchids originated between 80.38 and 114.84 Mya, and *Dendrobium* species diverged between 22.67 and 48.64 Mya. *Dendrobium* species experienced two WGD events at approximately 17.71 and 66.9 Mya. These findings provide valuable insights for future research on the evolutionary history of Orchidaceae and *Dendrobium* species. By integrating genome and transcriptome data, we identified biosynthesis pathway enzyme genes associated with flavonoids, polysaccharides, carotenoids, and alkaloids. R2R3-MYBs were identified in *D. moniliforme*, and 90 R2R3-MYBs were classified into 21 subgroups. Reviewing studies on R2R3-MYB transcription factors, we found that R2R3-MYB in SG6 can up-regulate flavonoid biosynthesis. The quantity of most SGs was found to be relatively conserved among orchids. Transient expression experiments in leaves and UPLC−MS/MS assays confirmed that DMYB69 can directly up-regulate flavonoid biosynthesis enzyme genes and increase flavonoid content. Similarly, DMYB44 was verified to directly up-regulate carotenoid biosynthesis enzyme genes and increase carotenoid content through transient expression experiments and HPLC assays. The direct regulatory functions of DMYB69 and DMYB44 were further validated using yeast one-hybrid and dual-luciferase assays. This study provides a crucial molecular foundation and theoretical reference for future molecular breeding research.

Acknowledgments

Our work was funded by the National Natural Science Foundation of China (Grant No. 32070353 and 32470384), Forestry Science and Technology Innovation and Promotion Project of Jiangsu Province (LYKJ[2021]12, China), Agricultural Science and Technology Independent Innovation Fund Project of Jiangsu Province (CX(22)3147, China), Graduate Research and Practice Innovation Project of Jiangsu Province (181200003023127, China).

Author contributions

Jiapeng Yang, Zhitao Niu, and Xiaoyu Ding designed the study. Jiapeng Yang and Yingying Jin performed the experiments. Qingyun Xue, Wei Liu, Zhitao Niu, and Xiaoyu Ding were responsible for preparing materials. Jiapeng Yang, Qiqian Xue, Chao Li, and Yingying Jin analyzed the data. Jiapeng Yang wrote the manuscript, which was revised by Zhitao Niu and Xiaoyu Ding. All authors approved the final version of the manuscript.

Conflict of interest

The authors have no conflicts of interest to declare.

Data availability

Genome data assembled in this study have been deposited in NCBI (BioProject: PRJNA1121055). Raw data of transcriptome data have been deposited in NCBI (BioProject: PRJNA1180727).

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2025.03.005.

References

1. Xu K, Ren X, Wang J, Zhang Q, Fu X, Zhang PC. Clinical development and informatics analysis of natural and semi-synthetic flavonoid drugs: a critical review. *J Adv Res* 2024;**63**:269−84.
2. Mutha RE, Tatiya AU, Surana SJ. Flavonoids as natural phenolic compounds and their role in therapeutics: an overview. *Futur J Pharm Sci* 2021;**7**:25.
3. Naik J, Misra P, Trivedi PK, Pandey A. Molecular components associated with the regulation of flavonoid biosynthesis. *Plant Sci* 2022;**317**:111196.
4. Tohge T, de Souza LP, Fernie AR. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J Exp Bot* 2017;**68**:4013−28.
5. Gao XR, Zhang H, Li X, Bai YW, Peng K, Wang Z, et al. The B-box transcription factor IbBBX29 regulates leaf development and flavonoid biosynthesis in sweet potato. *Plant Physiol* 2023;**191**:496−514.
6. Tan BC, Tan SK, Wong SM, Ata N, Rahman NA, Khalid N. Distribution of flavonoids and cyclohexenyl chalcone derivatives in conventional propagated and *in vitro*-derived field-grown *Boesenbergia rotunda* (L.) Mansf. *Evid-based Compl Alt* 2015;**2015**:451870.
7. Dudek B, Warskulat AC, Schneider B. The occurrence of flavonoids and related compounds in flower sections of *Papaver nudicaule*. *Plants-Basel* 2016;**5**:28.
8. Dong NQ, Lin HX. Contribution of phenylpropanoid metabolism to plant development and plant−environment interactions. *J Integr Plant Biol* 2021;**63**:180−209.
9. Chen S, Wu F, Li Y, Qian Y, Pan X, Li F, et al. NtMYB4 and NtCHS1 are critical factors in the regulation of flavonoid biosynthesis and are involved in salinity responsiveness. *Front Plant Sci* 2019;**10**:178.
10. Karlson CKS, Mohd Noor SN, Khalid N, Tan BC. CRISPRi-mediated down-regulation of the *Cinnamate-4-Hydroxylase* (*C4H*) gene enhances the flavonoid biosynthesis in *Nicotiana tabacum*. *Biology-Basel* 2022;**11**:1127.
11. Saxena S, Pal L, Naik J, Singh Y, Verma PK, Chattopadhyay D, et al. The R2R3-MYB-SG7 transcription factor CaMYB39 orchestrates surface phenylpropanoid metabolism and pathogen resistance in chickpea. *New Phytol* 2023;**238**:798−816.
12. Cao Y, Zhang R, Xing M, Ren C, Li J, Qian J, et al. Synergistic actions of three MYB transcription factors underpins the high accumulation of myricetin in *Morella rubra*. *Plant J* 2023;**115**:577−94.
13. Li D, Ye G, Li J, Lai Z, Ruan S, Qi Q, et al. High light triggers flavonoid and polysaccharide synthesis through *DoHY5*-dependent signaling in *Dendrobium officinale*. *Plant J* 2023;**115**:1114−33.
14. 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**:1513−30.
15. Zhang H, Koes R, Shang H, Fu Z, Wang L, Dong X, et al. Identification and functional analysis of three new anthocyanin R2R3-MYB genes in *Petunia*. *Plant Direct* 2019;**3**:e00114.
16. Shao D, Zhu QH, Liang Q, Wang X, Li Y, Sun Y, et al. Transcriptome analysis reveals differences in anthocyanin accumulation in cotton (*Gossypium hirsutum* L.) induced by red and blue light. *Front Plant Sci* 2022;**13**:788828.
17. Fraser PD, Bramley PM. The biosynthesis and nutritional uses of carotenoids. *Prog Lipid Res* 2004;**43**:228−65.
18. Demmig-Adams B, Adams 3rd WW. Antioxidants in photosynthesis and human nutrition. *Science* 2002;**298**:2149−53.
19. Howitt CA, Pogson BJ. Carotenoid accumulation and function in seeds and non-green tissues. *Plant Cel Environ* 2006;**29**:435−45.
20. An JP, Wang XF, Zhang XW, Xu HF, Bi SQ, You CX, et al. An apple MYB transcription factor regulates cold tolerance and anthocyanin accumulation and undergoes MIEL1-mediated degradation. *Plant Biotechnol J* 2020;**18**:337−53.
21. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H. The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *Embo J* 1987;**6**:3553−8.
22. Jia L, Clegg MT, Jiang T. Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies *indica* and *japonica* genomes. *Plant Physiol* 2004;**134**:575−85.
23. Kranz H, Scholz K, Weisshaar B. *C-MYB* oncogene-like genes encoding three MYB repeats occur in all major plant lineages. *Plant J* 2000;**21**:231−5.
24. Krieg J, Oelgeschläger M, Janknecht R, Lüscher B. High affinity DNA binding of native full length c-Myb and differential proteolytic sensitivity of its N- and C-terminal domains. *Oncogene* 1995;**10**:2221−8.
25. Du H, Feng BR, Yang SS, Huang YB, Tang YX. The R2R3-MYB transcription factor gene family in maize. *PLoS One* 2012;**7**:e37463.
26. Sun C, Wang C, Zhang W, Liu S, Wang W, Yu X, et al. The R2R3-type MYB transcription factor MdMYB90-like is responsible for the enhanced skin color of an apple bud sport mutant. *Hortic Res* 2021;**8**:156.
27. Su L, Lv A, Wen W, Fan N, Li J, Gao L, et al. *MsMYB741* is involved in alfalfa resistance to aluminum stress by regulating flavonoid biosynthesis. *Plant J* 2022;**112**:756−71.
28. He G, Zhang R, Jiang S, Wang H, Ming F. The MYB transcription factor RcMYB1 plays a central role in rose anthocyanin biosynthesis. *Hortic Res* 2023;**10**:uhad080.
29. Meng Y, Wang Z, Wang Y, Wang C, Zhu B, Liu H, et al. The MYB activator WHITE PETAL1 associates with MtTT8 and MtWD40-1 to regulate carotenoid-derived flower pigmentation in *Medicago truncatula*. *Plant Cell* 2019;**31**:2751−67.
30. Ampomah-Dwamena C, Thrimawithana AH, Dejnoprat S, Lewis D, Espley RV, Allan AC. A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol* 2019;**221**:309−25.
31. Yang J, Huang J, Wu X, Xu Y, Gu Z, Chen Y, et al. NtMYB1 and NtNCED1/2 control abscisic acid biosynthesis and tepal senescence in *Chinese narcissus* (*Narcissus tazetta*). *J Exp Bot* 2023;**74**:6505−21.
32. Yan Y, Li C, Dong X, Li H, Zhang D, Zhou Y, et al. MYB30 is a key negative regulator of *Arabidopsis* photomorphogenic development

that promotes PIF4 and PIF5 protein accumulation in the light. *Plant Cell* 2020;**32**:2196−215.
33. Lu HC, Lam SH, Zhang D, Hsiao YY, Li BJ, Niu SC, et al. R2R3-MYB genes coordinate conical cell development and cuticular wax biosynthesis in *Phalaenopsis aphrodite*. *Plant Physiol* 2022;**188**: 318−31.
34. Zhao Y, Xing L, Wang X, Hou YJ, Gao J, Wang P, et al. The ABA receptor PYL8 promotes lateral root growth by enhancing MYB77-dependent transcription of auxin-responsive genes. *Sci Signal* 2014; **7**:ra53.
35. Cai Y, Liu Y, Fan Y, Li X, Yang M, Xu D, et al. MYB112 connects light and circadian clock signals to promote hypocotyl elongation in *Arabidopsis*. *Plant Cell* 2023;**35**:3485−503.
36. An JP, Li R, Qu FJ, You CX, Wang XF, Hao YJ. R2R3-MYB transcription factor MdMYB23 is involved in the cold tolerance and proanthocyanidin accumulation in apple. *Plant J* 2018;**96**: 562−77.
37. Zhang P, Wang R, Yang X, Ju Q, Li W, Lü S, et al. The R2R3-MYB transcription factor AtMYB49 modulates salt tolerance in *Arabidopsis* by modulating the cuticle formation and antioxidant defence. *Plant Cel Environ* 2020;**43**:1925−43.
38. Chen C, Zhang Z, Lei YY, Chen WJ, Zhang ZH, Li XM, et al. *MdMYB44-like* positively regulates salt and drought tolerance *via* the MdPYL8-MdPP2CA module in apple. *Plant J* 2024;**118**:24−41.
39. Zhu F, Luo T, Liu C, Wang Y, Yang H, Yang W, et al. An R2R3-MYB transcription factor represses the transformation of α- and β-branch carotenoids by negatively regulating expression of *CrBCH2* and *CrNCED5* in flavedo of *Citrus reticulate*. *New Phytol* 2017;**216**: 178−92.
40. Cao H, Chen J, Yue M, Xu C, Jian W, Liu Y, et al. Tomato transcriptional repressor MYB70 directly regulates ethylene-dependent fruit ripening. *Plant J* 2020;**104**:1568−81.
41. Zhang L, Xu Y, Li Y, Zheng S, Zhao Z, Chen M, et al. Transcription factor CsMYB77 negatively regulates fruit ripening and fruit size in citrus. *Plant Physiol* 2024;**194**:867−83.
42. Wu X, Yuan J, Luo A, Chen Y, Fan Y. Drought stress and re-watering increase secondary metabolites and enzyme activity in *Dendrobium moniliforme*. *Ind Crop Prod* 2016;**94**:385−93.
43. Pant B, Paudel M, Chand MB, Wagner SH. *Treasure troves of orchids in Central Nepal*. Kirtipur, Nepal: Central Department of Botany, Tribhuvan University; 2016.
44. Zhao JR, Wang YH, Jin Y, Jiang C, Zhan ZL. Research advances in chemical constituents and pharmacological activities of *Dendrobium* plants. *Zhongguo Zhong yao za zhi* 2022;**47**:2358−72.
45. Li C, Qiu J, Ding L, Huang M, Huang S, Yang G, et al. Anthocyanin biosynthesis regulation of *DhMYB2* and *DhbHLH1* in *Dendrobium* hybrids petals. *Plant Physiol Bioch* 2017;**112**:335−45.
46. Khairul-Anuar MA, Mazumdar P, Othman RY, Harikrishna JA. *DhMYB22* and *DhMYB60* regulate pigment intensity and floral organ shape in *Dendrobium* hybrid. *Ann Bot* 2022;**130**:579−94.
47. Wang Y, Yin H, Long Z, Zhu W, Yin J, Song X, et al. DhMYB2 and DhbHLH1 regulates anthocyanin accumulation *via* activation of late biosynthesis genes in *Phalaenopsis*-type *Dendrobium*. *Front Plant Sci* 2022;**13**:1046134.
48. Yang K, Hou Y, Wu M, Pan Q, Xie Y, Zhang Y, et al. DoMYB5 and DobHLH24, transcription factors involved in regulating anthocyanin accumulation in *Dendrobium officinale*. *Int J Mol Sci* 2023; **24**:7552.
49. Hieber AD, Mudalige-Jayawickrama RG, Kuehnle AR. Color genes in the orchid *Oncidium Gower* Ramsey: identification, expression, and potential genetic instability in an interspecific cross. *Planta* 2006;**223**:521−31.
50. Wang Y, Xu J, Liu A. Identification of the carotenoid cleavage dioxygenase genes and functional analysis reveal *DoCCD1* is potentially involved in beta-ionone formation in *Dendrobium officinale*. *Front Plant Sci* 2022;**13**:967819.
51. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;**9**: e112963.
52. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 2017;**356**: 92−5.
53. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**: 3210−2.
54. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna-Uk* 2015;**6**:11.
55. Price AL, Jones NC, Pevzner PA. *De novo* identification of repeat families in large genomes. *Bioinformatics* 2005;**21**:i351−8.
56. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res* 2018;**46**:e126.
57. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. *Bmc Bioinformatics* 2005;**6**:31.
58. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 2008;**24**:637−44.
59. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 2007;**23**:673−9.
60. Shumate A, Wong B, Pertea G, Pertea M. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Comput Biol* 2022;**18**:e1009730.
61. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res* 2018;**46**:W200−4.
62. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol* 2019;**1962**:1−14.
63. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;**20**:238.
64. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772−80.
65. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. Beast 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2019;**15**:e1006650.
66. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007;**449**:463−7.
67. Ramírez SR, Gravendeel B, Singer RB, Marshall CR, Pierce NE. Dating the origin of the Orchidaceae from a fossil orchid with its pollinator. *Nature* 2007;**448**:1042−5.
68. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol* 2013;**30**:1987−97.
69. Zwaenepoel A, Van de Peer Y. wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 2019;**35**:2153−5.
70. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012;**40**:e49.
71. Davis EM, Sun Y, Liu Y, Kolekar P, Shao Y, Szlachta K, et al. SequencErr: measuring and suppressing sequencer errors in next-generation sequencing data. *Genome Biol* 2021;**22**:37.
72. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics* 2011;**12**:323.
73. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012;**16**: 284−7.
74. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 2011;**39**:W29−37.
75. Zhang B, Xu X, Huang R, Yang S, Li M, Guo Y. CRISPR/Cas9-mediated targeted mutation reveals a role for *AN4* rather than *DPL*

in regulating venation formation in the corolla tube of *Petunia hybrida*. *Hortic Res* 2021;**8**:116.
76. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 2010;**15**:573−81.
77. Man J, Shi Y, Huang Y, Zhang X, Wang X, Liu S, et al. PnMYB4 negatively modulates saponin biosynthesis in *Panax notoginseng* through interplay with PnMYB1. *Hortic Res* 2023;**10**:uhad134.
78. Chao YT, Chen WC, Chen CY, Ho HY, Yeh CH, Kuo YT, et al. Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol J* 2018;**16**:2027−41.
79. Han B, Jing Y, Dai J, Zheng T, Gu F, Zhao Q, et al. A chromosome-level genome assembly of *Dendrobium huoshanense* using long reads and Hi-C data. *Genome Biol Evol* 2020;**12**:2486−90.
80. Zhang Y, Zhang GQ, Zhang D, Liu XD, Xu XY, Sun WH, et al. Chromosome-scale assembly of the *Dendrobium chrysotoxum* genome enhances the understanding of orchid evolution. *Hortic Res* 2021;**8**:183.
81. Yang FX, Gao J, Wei YL, Ren R, Zhang GQ, Lu CQ, et al. The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol J* 2021;**19**:2501−16.
82. Xu Y, Lei Y, Su Z, Zhao M, Zhang J, Shen G, et al. A chromosome-scale *Gastrodia elata* genome and large-scale comparative genomic analysis indicate convergent evolution by gene loss in mycoheterotrophic and parasitic plants. *Plant J* 2021;**108**:1609−23.
83. Niu Z, Zhu F, Fan Y, Li C, Zhang B, Zhu S, et al. The chromosome-level reference genome assembly for *Dendrobium officinale* and its utility of functional genomics research and molecular breeding study. *Acta Pharm Sin B* 2021;**11**:2080−92.
84. Xu Q, Niu SC, Li KL, Zheng PJ, Zhang XJ, Jia Y, et al. Chromosome-scale assembly of the *Dendrobium nobile* genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of *Dendrobium*. *Front Plant Sci* 2022;**13**:844622.
85. Piet Q, Droc G, Marande W, Sarah G, Bocs S, Klopp C, et al. A chromosome-level, haplotype-phased *Vanilla planifolia* genome highlights the challenge of partial endoreplication for accurate whole-genome assembly. *Plant Commun* 2022;**3**:100330.
86. Li MH, Liu KW, Li Z, Lu HC, Ye QL, Zhang D, et al. Genomes of leafy and leafless *Platanthera* orchids illuminate the evolution of mycoheterotrophy. *Nat Plants* 2022;**8**:373−88.
87. Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, et al. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *P Roy Soc B-biol Sci* 2015;**282**:20151553.
88. Huang G, Song L, Du X, Huang X, Wei F. Evolutionary genomics of camouflage innovation in the orchid mantis. *Nat Commun* 2023;**14**:4821.
89. Bennetzen JL. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 2002;**115**:29−36.
90. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 2009;**10**:725−32.
91. Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 2009;**102**:533−41.
92. Berg JA, Appiano M, Santillán Martínez M, Hermans FW, Vriezen WH, Visser RG, et al. A transposable element insertion in the susceptibility gene *CsaMLO8* results in hypocotyl resistance to powdery mildew in cucumber. *Bmc Plant Biol* 2015;**15**:243.
93. Chen J, Zhao H, Zheng X, Liang K, Guo Y, Sun X. Recent amplification of Osr4 LTR-retrotransposon caused rice *D1* gene mutation and dwarf phenotype. *Plant Divers* 2017;**39**:73−9.
94. Ji JL, Yang LM, Fang ZY, Zhuang M, Zhang YY, Lv HH, et al. Recessive male sterility in cabbage (*Brassica oleracea* var. *capitata*) caused by loss of function of *BoCYP704B1* due to the insertion of a LTR-retrotransposon. *Theor Appl Genet* 2017;**130**:1441−51.
95. Lee W, Eom DW, Jung Y, Yamabe N, Lee S, Jeon Y, et al. *Dendrobium moniliforme* attenuates high-fat diet-induced renal damage in mice through the regulation of lipid-induced oxidative stress. *Am J Chin Med* 2012;**40**:1217−28.
96. Zhao N, Yang G, Zhang Y, Chen L, Chen Y. A new 9,10-dihydrophenanthrene from *Dendrobium moniliforme*. *Nat Prod Res* 2016;**30**:174−9.
97. Baek JM, Kim JY, Ahn SJ, Cheon YH, Yang M, Oh J, et al. *Dendrobium moniliforme* exerts inhibitory effects on both receptor activator of nuclear factor kappa-b ligand-mediated osteoclast differentiation *in vitro* and lipopolysaccharide-induced bone erosion *in vivo*. *Molecules* 2016;**21**:295.
98. Akhgari A, Laakso I, Maaheimo H, Choi YH, Seppänen-Laakso T, Oksman-Caldentey KM, et al. Methyljasmonate elicitation increases terpenoid indole alkaloid accumulation in *Rhazya stricta* hairy root cultures. *Plants-Basel* 2019;**8**:534.
99. Jiao C, Song C, Zheng S, Zhu Y, Jin Q, Cai Y, et al. Metabolic profiling of *Dendrobium officinale* in response to precursors and methyl jasmonate. *Int J Mol Sci* 2018;**19**:728.
100. Adejobi OI, Guan J, Yang L, Hu JM, Yu A, Muraguri S, et al. Transcriptomic analyses shed light on critical genes associated with bibenzyl biosynthesis in *Dendrobium officinale*. *Plants-Basel* 2021;**10**:633.
101. Zhou LJ, Geng Z, Wang Y, Wang Y, Liu S, Chen C, et al. A novel transcription factor CmMYB012 inhibits flavone and anthocyanin biosynthesis in response to high temperatures in chrysanthemum. *Hortic Res* 2021;**8**:248.
102. Ma D, Constabel CP. MYB Repressors as regulators of phenylpropanoid metabolism in plants. *Trends Plant Sci* 2019;**24**:275−89.
103. Wang F, Ji G, Xu Z, Feng B, Zhou Q, Fan X, et al. Metabolomics and transcriptomics provide insights into anthocyanin biosynthesis in the developing grains of purple wheat (*Triticum aestivum* L.). *J Agric Food Chem* 2021;**69**:11171−84.
104. Feder A, Burger J, Gao S, Lewinsohn E, Katzir N, Schaffer AA, et al. A kelch domain-containing F-box coding gene negatively regulates flavonoid accumulation in muskmelon. *Plant Physiol* 2015;**169**:1714−26.
105. Gholami A, De Geyter N, Pollier J, Goormachtig S, Goossens A. Natural product biosynthesis in *Medicago* species. *Nat Prod Rep* 2014;**31**:356−80.
106. Falcone Ferreyra ML, Rius SP, Casati P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front Plant Sci* 2012;**3**:222.
107. Zhang A, Zheng J, Chen X, Shi X, Wang H, Fu Q. Comprehensive analysis of transcriptome and metabolome reveals the flavonoid metabolic pathway is associated with fruit peel coloration of melon. *Molecules* 2021;**26**:2830.
108. Yuan Y, Zhang J, Kallman J, Liu X, Meng M, Lin J. Polysaccharide biosynthetic pathway profiling and putative gene mining of *Dendrobium moniliforme* using RNA-Seq in different tissues. *Bmc Plant Biol* 2019;**19**:521.
109. Luo QL, Tang ZH, Zhang XF, Zhong YH, Yao SZ, Wang LS, et al. Chemical properties and antioxidant activity of a water-soluble polysaccharide from *Dendrobium officinale*. *Int J Biol Macromol* 2016;**89**:219−27.
110. Li XL, Xiao JJ, Zha XQ, Pan LH, Asghar MN, Luo JP. Structural identification and sulfated modification of an antiglycation *Dendrobium huoshanense* polysaccharide. *Carbohyd Polym* 2014;**106**:247−54.
111. Chen L, Xu S, Liu Y, Zu Y, Zhang F, Du L, et al. Identification of key gene networks controlling polysaccharide accumulation in different tissues of *Polygonatum cyrtonema* Hua by integrating metabolic phenotypes and gene expression profiles. *Front Plant Sci* 2022;**13**:1012231.
112. Yu Z, Zhang G, Teixeira da Silva JA, Li M, Zhao C, He C, et al. Genome-wide identification and analysis of DNA methyltransferase and demethylase gene families in *Dendrobium officinale* reveal their potential functions in polysaccharide accumulation. *Bmc Plant Biol* 2021;**21**:21.

113. Chen Y, Li F, Wurtzel ET. Isolation and characterization of the *Z-ISO* gene encoding a missing component of carotenoid biosynthesis in plants. *Plant Physiol* 2010;**153**:66−79.
114. Breitenbach J, Sandmann G. Zeta-Carotene cis isomers as products and substrates in the plant poly-cis carotenoid biosynthetic pathway to lycopene. *Planta* 2005;**220**:785−93.
115. Flores-Ortiz C, Alvarez LM, Undurraga A, Arias D, Durán F, Wegener G, et al. Differential role of the two ζ-carotene desaturase paralogs in carrot (*Daucus carota*): *ZDS1* is a functional gene essential for plant development and carotenoid synthesis. *Plant Sci* 2020;**291**:110327.
116. Ng TB, Liu J, Wong JH, Ye X, Wing Sze SC, Tong Y, et al. Review of research on *Dendrobium*, a prized folk medicine. *Appl Microbiol Biot* 2012;**93**:1795−803.
117. Wang Y, Tong Y, Adejobi OI, Wang Y, Liu A. Research advances in multi-omics on the traditional Chinese herb *Dendrobium officinale*. *Front Plant Sci* 2022;**12**:808228.
118. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Pla* 2013;**19**:307−21.
119. Millard PS, Kragelund BB, Burow M. R2R3 MYB transcription factors - functions outside the DNA-binding domain. *Trends Plant Sci* 2019;**24**:934−46.
120. Jiang CK, Rao GY. Insights into the diversification and evolution of R2R3-MYB transcription factors in plants. *Plant Physiol* 2020;**183**: 637−55.
121. Yu Y, Guo DD, Min DH, Cao T, Ning L, Jiang QY, et al. Foxtail millet MYB-like transcription factor SiMYB16 confers salt tolerance in transgenic rice by regulating phenylpropane pathway. *Plant Physiol Bioch* 2023;**195**:310−21.
122. Liu R, Shen Y, Wang M, Liu R, Cui Z, Li P, et al. *GhMYB102* promotes drought resistance by regulating drought-responsive genes and ABA biosynthesis in cotton (*Gossypium hirsutum* L.). *Plant Sci* 2023;**329**:111608.
123. Shi M, Liu X, Zhang H, He Z, Yang H, Chen J, et al. The IAA- and ABA-responsive transcription factor CgMYB58 upregulates lignin biosynthesis and triggers juice sac granulation in pummelo. *Hortic Res* 2020;**7**:139.
124. Song F, Li Z, Wang C, Jiang Y, Wang Z, He L, et al. CsMYB15 positively regulates *Cs4CL2*-mediated lignin biosynthesis during juice sac granulation in navel orange. *Front Plant Sci* 2023;**14**:1223820.
125. Jiang PF, Lin XY, Bian XY, Zeng QY, Liu YJ. Ectopic expression of *Populus MYB10* promotes secondary cell wall thickening and inhibits anthocyanin accumulation. *Plant Physiol Bioch* 2022;**172**:24−32.
126. Zheng H, Gao Y, Sui Y, Dang Y, Wu F, Wang X, et al. R2R3 MYB transcription factor SbMYBHv33 negatively regulates sorghum biomass accumulation and salt tolerance. *Theor Appl Genet* 2023;**136**:5.
127. Lan Y, Zhang K, Wang L, Liang X, Liu H, Zhang X, et al. The R2R3-MYB transcription factor OfMYB21 positively regulates linalool biosynthesis in *Osmanthus fragrans* flowers. *Int J Biol Macromol* 2023;**249**:126099.
128. Zhou H, Lin-Wang K, Wang F, Espley RV, Ren F, Zhao J, et al. Activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation. *New Phytol* 2019;**221**:1919−34.
129. Zhang L, Duan Z, Ma S, Sun S, Sun M, Xiao Y, et al. SlMYB7, an AtMYB4-Like R2R3-MYB Transcription factor, inhibits anthocyanin accumulation in *Solanum lycopersicum* fruits. *J Agric Food Chem* 2023;**71**:18758−68.
130. Mishra AK, Kocábek T, Nath VS, Khan A, Matoušek J, Hazzouri KM, et al. The multifaceted roles of R2R3 transcription factor HlMYB7 in the regulation of flavonoid and bitter acids biosynthesis, development and biotic stress tolerance in hop (*Humulus lupulus* L.). *Plant Physiol Bioch* 2023;**197**:107636.
131. Deng GM, Zhang S, Yang QS, Gao HJ, Sheng O, Bi FC, et al. MaMYB4, an R2R3-MYB repressor transcription factor, negatively regulates the biosynthesis of anthocyanin in banana. *Front Plant Sci* 2021;**11**:600704.
132. Tian Q, Han L, Zhu X, Zhang C, Li Y, Xue X, et al. *SmMYB4* is a R2R3-MYB transcriptional repressor regulating the biosynthesis of phenolic acids and tanshinones in *Salvia miltiorrhiza*. *Metabolites* 2022;**12**:968.
133. Zhao L, Song Z, Wang B, Gao Y, Shi J, Sui X, et al. R2R3-MYB transcription factor NtMYB330 regulates proanthocyanidin biosynthesis and seed germination in tobacco (*Nicotiana tabacum* L.). *Front Plant Sci* 2022;**12**:819247.
134. Jiang L, Yue M, Liu Y, Zhang N, Lin Y, Zhang Y, et al. A novel R2R3-MYB transcription factor FaMYB5 positively regulates anthocyanin and proanthocyanidin biosynthesis in cultivated strawberries (*Fragaria* × *ananassa*). *Plant Biotechnol J* 2023;**21**:1140−58.
135. Lv Z, Li J, Qiu S, Qi F, Su H, Bu Q, et al. The transcription factors TLR1 and TLR2 negatively regulate trichome density and artemisinin levels in *Artemisia annua*. *J Integr Plant Biol* 2022;**64**:1212−28.
136. Zhong C, Tang Y, Pang B, Li X, Yang Y, Deng J, et al. The R2R3-MYB transcription factor GhMYB1a regulates flavonol and anthocyanin accumulation in *Gerbera hybrida*. *Hortic Res* 2020;**7**:78.
137. Song Z, Zhao L, Ma W, Peng Z, Shi J, Pan F, et al. Ethylene inhibits ABA-induced stomatal closure *via* regulating NtMYB184-mediated flavonol biosynthesis in tobacco. *J Exp Bot* 2023;**74**:6735−48.
138. Wang R, Mao C, Ming F. PeMYB4L interacts with PeMYC4 to regulate anthocyanin biosynthesis in *Phalaenopsis* orchid. *Plant Sci* 2022;**324**:111423.
139. Wang D, Liu G, Yang J, Shi G, Niu Z, Liu H, et al. Integrated metabolomics and transcriptomics reveal molecular mechanisms of corolla coloration in *Rhododendron dauricum* L. *Plant Physiol Bioch* 2024;**207**:108438.
140. Tang Q, Chi FM, Liu HD, Zhang HJ, Song Y. Single-molecule real-time and illumina sequencing to analyze transcriptional regulation of flavonoid synthesis in blueberry. *Front Plant Sci* 2021;**12**:754325.
141. Zuo X, Miao C, Li M, Gu L, Yang X, Song C, et al. Purple *Rehmannia*: investigation of the activation of R2R3-MYB transcription factors involved in anthocyanin biosynthesis. *Physiol Plantarum* 2023;**175**:e13920.