Metabolomic and transcriptomic analyses reveal the regulation of pigmentation in the purple variety of *Dendrobium officinale*

Xinqiao Zhan1,2*, Jufeng Qi1, Bin Zhou3 & Bizeng Mao1*

We performed an integrated analysis of the transcriptome and metabolome from purple (Pr) and normal cultivated varieties (CK) of *Dendrobium officinale* to gain insights into the regulatory networks associated with phenylpropanoid metabolism and to identify the key regulatory genes of pigmentation. Metabolite and transcript profiling were conducted by ultra-performance liquid chromatography electrospray tandem mass spectrometry (UPLC-ESI-MS/MS) and RNA sequencing. Pr had more flavonoids in the stem than did CK. Metabolome analyses showed that 148 differential metabolites are involved in the biosynthesis of phenylpropanoids, amino acids, purines, and organic acids. Among them, the delphinidin and quercetin derivatives were significantly higher in Pr. A total of 4927 differentially expressed genes (DEGs) were significantly enriched \((p \leq 0.01)\) in 50 Gene Ontology (GO) terms. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses revealed significantly enriched phenylpropanoid biosynthesis and phytohormone signal transduction in Pr versus CK. The expression levels of flavanone 3-hydroxylase (F3H) and leucoanthocyanidin dioxygenase (LDOX) affected the flux of dihydroflavonol, which led to a color change in Pr. Moreover, DEG enrichment and metabolite analyses reflected flavonoid accumulation in Pr related to brassinosteroid (BR) and auxin metabolism. The results of this study elucidate phenylpropanoid biosynthesis in *D. officinale*.

*Dendrobium* is the second largest genus in Orchidaceae, and contain more than 1500 species of high ornamental and medicinal value, distributed across Asia and Oceania. *Dendrobium officinale* is a popular health product in China because its active components and health-promoting effects1,2. Most of the active medicinal components of *Dendrobium* are polysaccharides3,4, with immunomodulatory and hepatoprotective activities5. *Dendrobium* also contains complex secondary metabolites, including alkaloids, phenols, terpenes, coumarins, and flavonoids6,7. In recent years, transcriptome and metabolomics technologies have been used to investigate the biosynthesis of secondary metabolites in several *Dendrobium* species. Eight flavonoid glycosides were identified by ultra-high performance liquid-chromatography–mass spectrometry in *D. catenatum* from three different locations (Zhejiang, Guangxi and Guangdong) in China. Vicenin, 2″-O-glucopyranosylvitexin, and schafoseide were present in all the *D. catenatum* samples. Violanthin and 6,8-di-C-α-l-arabinosylapigenin were only present in Guangdong samples, while the vicenin II content was remarkably high in the Zhejiang samples8. The polysaccharide and alkaloid contents of *D. huoshanense* are associated with mevalonate (MVA) and 2-C-methyl-D-erythritol 4-phosphate (MEP) pathways8. Moreover, studies of stress resistance in *Dendrobium* have shown that the major CIPK gene transcript is affected by the biological clock under abiotic stress and is also associated with signal transduction and energy metabolism9. However, no studies have investigated the regulation of pigmentation in *Dendrobium*.

Flavonoids are a class of widely distributed secondary metabolites that play an important role in biotic and abiotic stresses10,11. Flavonoids are synthesized through phenylpropanoid metabolism, transforming phenyllalnine into naringenin, which then enters the anthocyanin, flavone, and flavonol biosynthesis pathways12. Anthocyanins and their derivatives have been identified in numerous flowers, fruits, leaves, stems, and seeds. More than 600 naturally occurring anthocyanins are derived from three major anthocyanins: pelargonidin, cyanidin, and delphinidin. These pigments contribute to the colors orange to red, red to magenta, and magenta to purple, respectively13. Anthocyanin biosynthesis is a branch of general flavonoid metabolism for which biosynthetic
enzymes were identified that has been reviewed extensively. Two dihydroflavonol, dihydrokaempferol and dihydroquercetin, serve as major precursors that can be catalyzed by leucoanthocyanidins by dihydroflavonol 4-reductase (DER). Subsequently, anthocyanidin synthase (ANS) is responsible for the formation of anthocyanidins from the colorless leucoanthocyanidins. Although the biosynthetic pathway of anthocyanins has been identified, the precise relationship between anthocyanin metabolism and the regulation of color change remains unresolved. This is largely because of the diversity of anthocyanin species and differences between plant varieties.

Flavonoid biosynthesis is influenced by external factors such as UV-B light, pathogens, temperature, wounds, and phytohormones. Flavonoids are synthesized in the cytosol and transported to the vacuole for storage. The vacuolar sequestration of flavonoids involves vesicle trafficking, membrane transporters, and glutathione S-transferase (GST). There is an increasing body of evidence indicating that flavonoids accumulation are tightly related with phytohormone signals in plant growth. Certain flavonoids, such as kaempferol, quercetin, and apigenin, have been shown to inhibit auxin transport and improve localized auxin level in plants. In the transparent testa (tt4) mutant, auxin transport from the shoot to the root was found to be enhanced compared to that in the wild-type. Ethylene has a repressed function similar to that of auxin and modulates the flavonoids in response to gravity. In addition, ethylene-induced flavonol accumulation in guard cells moderates ABA-mediated stomatal closure. ABA signaling regulates the flavonoid biosynthesis by transcription factors HYS and MYB. MYB2 and MYB111 can be activated by HYS, and then mediate flavonol synthase expression. ABA accumulation during fruit ripening stimulates the MYB-bHLH-WD40 complex transcript levels to upregulate the expression of anthocyanin biosynthesis genes. Recent researches reveal that flavonoid may regulate the ABA-signaling network by detoxifying reactive oxygen species and enhancing ABA biosynthesis. The flavonoid pathway transcription factor TT8 directly regulates jasmonic acid (JA) and brassinosteroid (BR) biosynthesis. In Arabidopsis, JAZ interacts with the MYB-bHLH-WD40 complex to regulate JA-mediated anthocyanin biosynthesis. BR enhanced expression of BEE1 factor is an early response BR signaling component required for the full BR response. BEE1 is also involved in cold stress responses by regulating anthocyanin accumulation. Furthermore, BEE1 is regulated by other phytohormones, such as ABA, which represses the transcription of BEE1, and auxins and ethylene, which enhance BE1 expression. Therefore, the crosstalk between flavonoids and phytohormone metabolism play an important role in plant development.

The combination of genetic studies and multiple-omics data represents an efficient approach to decipher the regulatory networks of the active medicinal components in *Dendrobium*. However, this method is hampered by the lack of cultivars with different phenotype traits. In this study, metabolomic and transcriptome analyses were combined to investigate changes in the phenylpropanoid metabolism (and the related genes) of a novel purple variety of *D. officinale*. We explain the complex metabolic networks underlying biochemical traits in the purple variety (i.e., pigmentation) and provide new insights into the development of active medicinal components in *Dendrobium*.

**Results**

**Pigment accumulation in the *D. officinale* purple variety (Pr) stem.** Pigment variation in the purple variety (Pr) stem differed from that in the normal cultivated variety (CK) stem (Fig. 1a–c). Red pigments were observed in the Pr methanol HCl (hydrochloric acid in methanol) extracts of the stem (Fig.1d). We also measured the pigment content of methanol extracts and found that the flavonoid and anthocyanin contents were 28.8% and 203.1% higher, respectively, in Pr than in CK (Fig. 1e,f). The extracts were analyzed by thin-layer chromatography (TLC) and exhibited chromatographic bands corresponding to the flavonoids rutin, quercetin, and myricetin (Figure S1). These results indicate that Pr accumulated more flavonoids in the stem than did CK.

**Nontargeted metabolomic analysis and overall metabolite identification.** To further explore differences in the metabolite content of CK and Pr, the total extracts were subjected to ultrahigh-pressure liquid chromatography triple quadrupole mass spectrometry (UPLC-TQMS) for nontargeted metabolomics. The repeatability of CK and Pr extracts was judged by an overlapping analysis of the total ion current (TIC) in the quality control (QC) samples in positive and negative modes (Figure S2). A principal component analysis (PCA) was used to analyze the contribution rate of the first two components, which were 75.75% and 80.37% in positive and negative modes, respectively. The three period materials were clearly separated, and each formed a cluster in positive and negative modes (Fig. 2a).

After quality validation, 464 metabolites were identified in CK and Pr from the MS2 spectral data (Table S2). Superclass information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to classify the metabolites, and the first 20 pathways are presented in Figure S3. Most of the metabolite classifications were connected to phenylpropanoid metabolism, and the largest proportions were benzenoids (9%), lipids (30%), nucleotides (5%), organic acids (11%), organoheterocyclic compounds (9%), and phenylpropanoids (8%) (Fig. 2b).

**Identification of the differentially accumulated metabolites.** Differentially accumulated metabolites (DAMs) were defined by a fold change ≥2 or ≤0.5, and a variable importance in project (VIP) ≥ 1 between Pr and CK (p < 0.05). A total of 148 DAMs were identified (Table S3), and volcano plots were generated to show that 44 of 148 (29.7%) were up-regulated and 104 of 148 (70.3%) were down-regulated (Fig. 3a). A hierarchical cluster analysis (HCA) was also performed to assess the DAMs and QC sample accumulation patterns (Figure S4). Partial least-squares discriminant analysis (PLS-DA) was used to observe differences in the metabolite compositions of the two materials, and each formed a cluster (Fig. 3b). Based on the KEGG classifications, we created a HCA heat map to investigate the metabolites involved in phenylpropanoid, amino acid, purine, and organic acid metabolism in Pr and CK. The majority of the flavonoids were accumulated at significantly higher
levels in Pr, while most of the amino acids, organic acids, and purines accumulated at significantly lower levels in Pr (Fig. 3c).

Transcriptome data analysis. To analyze the gene expression patterns of the *D. officinale* purple variety, the raw reads from two transcript libraries were qualified, and the adapters were removed, yielding 38.48 GB of sequence data (21.03 GB from CK and 17.45 GB from Pr). The GC percentage ranged from 46 to 47%. The percentage of clean reads ranged from 88.89 to 95.40%. The probability of incorrect base calling was used to evaluate the sequencing quality according to the value of Q30. High Q30 proportions (97%) indicated high-quality RNA sequence (RNA-seq) data (Table S1). PCA revealed that Pr and CK presented different gene expression patterns. Pr clustered away from CK in PCA1, which explains 64.23% of the variation (Figure S5a). Moreover, the HAC showed that the Pr and CK gene expression profiles of the three independent biological replicates clustered together (Figure S5b).

Identification and functional enrichment analysis of the DEGs. The transcript abundances of each gene from the CK and Pr data were analyzed using the fragments per kilobase of transcript per million mapped reads (FPKM) method. A false discovery rate (FDR) ≤ 0.05 and fold change (FC) > 1 were used as cutoffs to identify differentially expressed genes (DEGs). In total, 1573 genes were up-regulated and 3354 genes were down-regulated in Pr vs. CK, respectively (Fig. 4a).

In our study, the DEGs were evaluated using GO and KEGG pathway analyses. The number of DGEs in Pr vs. CK are summarized in three main GO categories (Fig. 4b, Table S4). In detail, 18 GO terms were significantly enriched (*p* < 0.005) in the biological process category in Pr vs. CK. Three of these GO terms had the largest DEGs, including “transmembrane receptor protein tyrosine kinase signaling pathway” (3 up-regulated and 59 down-regulated), “cell wall organization” (13 up-regulated and 58 down-regulated), and “protein phosphorylation” (29 up-regulated and 132 down-regulated). In the cellular components category, eight GO terms were significantly enriched (*p* < 0.005) in the Pr vs. CK comparison. In particular, 631 DEGs were enriched in the “plasma membrane” GO term, with 165 up-regulated and 466 down-regulated. A total of 633 DEGs were enriched in the “integral component of membrane” GO term, with 178 up-regulated and 455 down-regulated genes. In the molecular functions category, 10 GO terms were significantly enriched (*p* < 0.005) in Pr vs. CK. Among them, “protein serine/threonine kinase activity” and “glycosyl groups transferase activity” respectively enriched 172 DEGs (35 up-regulated and 137 down-regulated) and 80 DEGs (15 up-regulated and 65 down-regulated). GO terms associated with lipid and wax metabolism were also observed in the DEGs (Figure S6, Table S4). Four GO terms were significantly enriched (*p* < 0.005) in Pr vs. CK, including “lipid catabolic process” (*p* = 8.794E−04), “wax biosynthetic process” (*p* = 1.874E−04), “lipid transport” (*p* = 1.210E−03), and “lipid binding” (*p* = 1.716E−05). Five GO terms were significantly enriched (*p* < 0.05) in Pr vs. CK, including “very
long-chain fatty acid metabolic process" \((p = 5.470E−03)\), "galactolipid metabolic process" \((p = 1.867E−02)\), "fatty acid biosynthetic process" \((p = 4.087E−02)\), "very long-chain fatty acid biosynthetic process" \((p = 4.440E−02)\), and "long-chain fatty acid-CoA ligase activity" \((p = 4.440E−02)\).

To further explore the biological functions of the DEGs, an enrichment analysis based on the KEGG database was performed. Among the 130 enriched KEGG pathways (Table S5), the strongest significance \((p < 0.05)\) was observed for "plant hormone signal transduction" \((p = 4.252E−06)\), "flavonoid biosynthesis" \((p = 2.641E−05)\), "other types of O-glycan biosynthesis" \((p = 4.647E−05)\), "stilbenoid, diarylheptanoid, and gingerol biosynthesis" \((p = 2.110E−04)\), "plant-pathogen interaction" \((p = 2.322E−04)\), "phenylpropanoid biosynthesis" \((p = 1.778E−03)\), "phenylalanine metabolism" \((p = 3.518E−02)\), and "isoflavonoid biosynthesis" \((p = 4.660E−02)\) (Fig. 5).

DEGs expression profiles associated with flavonoid biosynthesis. Flavonoids are synthesized through the phenylpropanoid metabolism pathway, and many enzymes participate in the catalyzing steps. In this study, the transcript levels of 56 flavonoid biosynthesis-related structural genes were analyzed (Fig. 6a). Two putative PAL (Dca018706 and Dca009106) and two C4H (Dca007625 and Dca004773) genes were identified, and all were down-regulated in Pr compared to CK. Five 4CL (Dca0001588, Dca010921, Dca000325, Dca021161, and Dca012359) genes were identified, among which the expression levels of Dca021161 and Dca012359 increased.

![Figure 2. Overall analysis of metabolomics data. (a) PCA analysis of the two samples (Pr, green; CK, red) and quality control samples (QC, blue). The x-axis represents the first principal component and the y-axis represents the second principal. (b) Component analysis of the identified metabolites.](https://example.com/figure2.jpg)
8-fold and 2.2-fold in Pr vs. CK, respectively. Chalcone synthase (CHS) plays an important role in flavonoid biosynthesis; two putative CHS genes (Dca003406 and Dca016502) were identified and Dca003406 was up-regulated by 16.2-fold in Pr compared to in CK. One CHI (Dca006141) gene was identified and showed no change in Pr vs. CK. Expression of most of the five F3H genes were significantly increased in Pr vs. CK, including 14.8-fold Dca026362, 3.7-fold Dca020695, and 12.5-fold Dca023004. Expression of one F3′H (Dca008783) gene and one F3R (Dca002396) were also identified, and Dca020665 was up-regulated 4.7-fold in Pr vs. CK.

Usually, flavonoids are glycosylated in plant cells, thereby increasing their solubility and facilitating their transport. Fifteen UGT genes were identified, and three genes showed more than tenfold up-regulation in Pr vs. CK, including 40-fold Dca014179, 14.8-fold Dca028143, and 13.4-fold Dca006464. Two 5GT (Dca024061 and Dca009740) and one 3MAT (Dca017013) gene showed more than 11-fold upregulation in Pr compared to CK, except for Dca009740 (Fig. 6a). Anthocyanins are synthesized on the cytosolic surface of the endoplasmic reticulum (ER) and transported to the vacuole. Eleven GST genes and six MATE genes were identified, and most of these genes were up-regulated in Pr vs. CK, such as 15.8-fold Dca026334, 14.7-fold Dca028214, 8.3-fold Dca019469, and 4.6-fold Dca024549 (Fig. 6a). Moreover, 15 anthocyanin biosynthesis-related genes were chosen for expression validation by qRT-PCR (Fig. 6b, Figure S7). All of these 15 selected genes showed similar expression patterns in the RNA-seq data, which indicated that the RNA-seq data were reliable.

**Integrated metabolomic and transcriptomic analyses in phenylpropanoid metabolism.** We integrated metabolomic and transcriptomic data to analyze the phenylpropanoid metabolism pathway in *D. officinale* (Fig. 7a). The phenylpropanoid pathway initiates from the aromatic amino acid phenylalanine, which decreased 4.3-fold in Pr vs. CK. Cinnamic acid decreased 3.1-fold in Pr vs. CK, while two PAL were down-regulated by 18.7-fold and 17.3-fold, respectively. 4-Coumarate decreased 4.2-fold in Pr vs. CK, while two C4H were down-regulated by 14.7-fold and 6.5-fold, respectively. Two CHS were up-regulated by 16.2-fold and 1.1-fold, respectively; nevertheless, chalcone decreased 1.2-fold in Pr vs. CK. Naringenin decreased 5.8-fold in Pr vs. CK, which was sustained by a 1.1-fold decrease in its mRNA. The 4CL family plays key roles in lignin metabolism and participates in sinapoyl-CoA, feruloyl-CoA, and caffeoyl-CoA biosynthesis. We found that sinapic acid,
conifer alcohol, and sinapoyl alcohol increased 3.53-fold, 9.65-fold, and 6.56-fold, respectively. One COMT gene (caffeic acid 3-O-methyltransferase) increased 3.80-fold in Pr compared to CK. Lignan is also an important active compound in Chinese herbs, which shares a common upstream pathway with lignins and branches after the synthesis of coniferyl alcohol. These results suggest that the medicinal value of *D. officinale* requires further exploration.

As with lignin metabolism, increased levels of flavone and flavanol (i.e., dihydrokaempferol, dihydroquercetin, eriodictyol, quercetin, quercetin glucoside, and rutin) were associated with significant increases in mRNA (Fig. 7a). For example, dihydroquercetin was estimated to increase 19-fold in Pr compared to that in CK, which was underpinned by a fivefold increase in its mRNA. Substantial changes in Pr were also observed for compounds connected to anthocyanin metabolism (Fig. 7a). Anthocyanin metabolism plays an important role in plant coloration; cyanidin and delphinidin contribute to red to magenta and magenta to purple colors, respectively. In Pr the total delphinidin content increased 45 times compared to CK, while one LDOX (*Dca020665*) was up-regulated 4.7-fold and the other (*Dca026251*) showed no obvious change, suggesting that the pigmentation of Pr was due to delphinidin (Fig. 7a,b).

For a systematic overview of the associations between the DAMs and DEGs, we compared the log2 fold changes between Pr and CK in mRNA and metabolites in Fig. 7c (the data details are shown in Table S8, Table S9). The metabolites and their correlated mRNA shown in red dots and green dots were both more abundant in Pr. For instance, three up-regulated genes (*Dca026362, Dca020695, and Dca023004*) of *F3H* were correlated with accumulation of dihydrokaempferol, dihydroquercetin, quercetin and eriodictyol. Down-regulated *PAL* (*Dca018706*) was correlated with the reduction of phenylalanine and cinnamic acid. The mRNA shown in orange

**Figure 4.** Identification of the DEGs between Pr and CK. (a) Volcano plot showing the differential genes in Pr vs. CK. (b) The Go classification of the DEGs in Pr vs. CK.
dots, were more abundant than metabolites in Pr, while the metabolites, shown in blue dots, were more abundant than mRNA in Pr. For instance, two 4CLL (Dca021161, Dca012359) were up-regulated and correlated with the reduction of 4-coumarate. The other two 4CL (Dca010921 and Dca000325) were down-regulated and correlated with the accumulation of caffeyl alcohol, coniferyl alcohol, sinapyl alcohol and sinapic acid.

Hormone signal transduction pathway associated with flavonoid biosynthesis. Based on the transcriptomes, 106 significant DEGs were enriched in "Plant hormone signal transduction" (ko04075), most of
which are involved in the auxin, cytokinin, ethylene, JA, ABA, GA, SA, and BR signaling pathways (Figure S8, Table S6). To further explore plant hormone metabolism associated with flavonoid biosynthesis, 16 related GO terms were identified, and two pathways were significantly enriched ($p < 0.05$) in Pr compared to in CK (Fig. 8a), including “response to auxin” (GO:0009733), “positive regulation of auxin-mediated signaling pathway” (GO:0010929), and “response to brassinosteroid” (GO:0009741). For the GO term “response to auxin”, we analyzed the expression of 21 auxin metabolism- and signaling pathway-related genes and found that most of the genes were down-regulated in Pr (Fig. 8b). In “response to brassinosteroid”, genes encoding AP2/ERF and B3 domain-containing protein and ornithine aminotransferase (Dca002334 and Dca002616) were significantly up-regulated in Pr, while two other genes (Dca017429 and Dca019769) were down-regulated in Pr (Fig. 7b). The endogenous brassinolide content was higher in Pr than in CK (Fig. 8c), and the auxin precursor (tryptophan) content was reduced in Pr (Fig. 8d). Together, these results suggest that BR and auxin metabolism affect flavonoid accumulation in Pr.

**Identifications of TFs families in Pr vs. CK.** Several TFs have been reported to play important roles in flavonoid biosynthesis. In our study, 477 putative TF encoding genes belonging to 16 major TF families were analyzed in *D. officinale* (Table S10). A large number of TFs were included in the MYB family (90 genes), bHLH family (85 genes), AP2/ERF family (78 genes) and WRKY family (54 genes) (Table S11). Furthermore, DEGs analysis revealed that most of the TFs were significantly down-regulated in Pr (Fig. 9). For example, the AP2/ERF family, Dca024190 and Dca007398, decreased 118-fold and 141-fold in Pr vs. CK, respectively. The MYB family (Dca004957) was greatly decreased by 434-fold in Pr vs. CK. Three bHLH family (Dca024215, Dca011717, and Dca021380) were down-regulated by more than 100-fold in Pr vs. CK. However, some up-regulated TFs showed especially strong responses in Pr (Fig. 9). For example, the AP2/ERF family, Dca016529 and Dca016530 were
more than 20-fold up-regulated in Pr vs. CK. The MYB-related transcription factor RADIALIS (Dca027426) was greatly increased by 119-fold in Pr vs. CK.

Discussion

Polysaccharides and alkaloids are common active constituents of *Dendrobium* species\(^ {31-33} \). The biosynthesis of polysaccharides and alkaloids is strongly associated with plant primary and secondary metabolism, and plants contain three major groups of secondary metabolites: phenolics, terpenoids, and alkaloids. Phenylpropanoids are the first class of phenolics that can be divided into several groups, such as flavonoids, anthocyanins, and lignin\(^ {34} \). However, few studies have qualitatively or quantitatively studied phenylpropanoids in *Dendrobium*. Eight flavonoid glycosides were identified in *D. catenatum* from three locations. Total flavonoid content in Guangxi samples were the highest, with a content of 3.87 μg/mg, while the total flavonoid content of the samples from the Guangdong and Zhejiang provinces were lower, at 2.40 and 2.85 μg/mg, respectively\(^ {8} \). In this study, we identified a Pr of *D. officinale* with higher flavonoid and anthocyanin content in stems compared to that in CK (Fig. 1).
Nontargeted metabolomics revealed 148 different metabolites involved in the biosynthesis of phenylpropanoids, amino acids, purines, and organic acids (Fig. 2). Phenylpropanoid biosynthesis is a complex process, and the initial three steps of the pathway, catalyzed by PAL, C4H, and 4CL, are necessary and provide the basis for all subsequent branches. This suggests that the PAL, C4H, and 4CL genes play important roles in flavonoid enrichment. The pal1 pal2 double mutants were deficient of tannin pigments in the seed coat. However, two isoforms of PAL and C4H were identified in our transcript data, and both were down-regulated in Pr compared to CK.
and glucose in the mature stage, suggesting that the purple variety of *D. officinale*, a 65–75% relative humidity. Stem tissues were collected from three independent plants for transcriptomic analysis.

Nontargeted metabolomic and analysis. The collected samples were thawed on ice, and metabolites were extracted with 50% methanol buffer. Briefly, 20 μL of sample was extracted with 120 μL of precooled 50% methanol, vortexed for 1 min, and incubated at room temperature (24 ± 2 °C) for 10 min. The samples were

Materials and methods

Plant materials and culture conditions. The stem samples were harvested from two-year-old cultivated *D. officinale*, including the purple variety and normal variety, in August 2019 from a growth chamber of Zhejiang University, Hangzhou, China. The growth conditions were set at 25 ± 2 °C with a light/dark cycle of 12/12 h and a 65–75% relative humidity. Stem tissues were collected from three independent plants for transcriptomic analysis (three biological replicates). The stem tissues were collected from six independent plants for metabolomic analysis (six biological replicates). All samples were immediately frozen in liquid nitrogen and stored at −80 °C.

Nontargeted metabolomic and analysis. The collected samples were thawed on ice, and metabolites were extracted with 50% methanol buffer. Briefly, 20 μL of sample was extracted with 120 μL of precooled 50% methanol, vortexed for 1 min, and incubated at room temperature (24 ± 2 °C) for 10 min. The samples were

(Int. 6). It could be that the late biosynthetic genes play a more important role than the early biosynthetic genes in phenylpropanoid metabolism in mature plants. For instance, the overexpression of *SQD2.2*, which encodes a glycosyltransferase in rice, conferred plants with enhanced flavonol levels at the mature stage. Flavonol synthase (FLS) and dihydroflavonol 4-reductase (DFR) have strong competition for common dihydroquercetin substrates that affect flower color in grape hyacinth.

Integrated transcriptomic and metabolomic analyses demonstrated the gene-to-metabolite networks and enabled the mechanisms involved in Pr pigmentation to be deciphered (Fig. 7a). The flow of individual metabolites with mRNA, including *F3H*, *DFR*, and *LDOX*, suggests that delphinidin is the predominant anthocyanin in the purple variety. The biosynthesis of anthocyanin via the phenylpropanoid pathway is controlled by two types of genes: structural genes and regulatory genes. Structural genes can be divided into early biosynthetic genes (such as *PAL* and *CHI*) and late biosynthetic genes (such as *LDOX* and *UGT*). In our study, the expression pattern of late biosynthetic genes was more complex than that of late biosynthetic genes. For example, *PAL* and *CHI* correlated with their metabolites and were down-regulated in Pr vs. CK. In contrast, one *LDOX* (*Dca020665*) was up-regulated 4.7-fold and the other (*Dca026251*) showed no obvious change in Pr vs. CK, and their related products, delphinidin and cyanidin, increased by 45.5-fold and 1.16-fold, respectively, in Pr vs. CK (Fig. 7a). Regulatory genes including MYB, bHLH and WD-repeat proteins, constitute a ternary complex predominantly responsible for orchestrating regulation of anthocyanin synthesis. Our transcriptomic data identified 24 MYB genes and 32 bHLH genes as DEGs in Pr vs. CK, and most of them were significantly down-regulated (Fig. 9).

This could be due to metabolic feedback or the preference of regulatory genes. In Arabidopsis, the MYB-bHLH-WD40 complex predominantly regulates late biosynthetic genes over early biosynthetic genes. The expression of other late biosynthetic genes, such as *DFR* and *LDOX*, is nearly off or is undetectable in *ttg1* and bHLH anthocyanin mutants. Moreover, anthocyanin accumulation is closely related to subcellular transport, including vesicle trafficking, membrane transporters and GSTs. GSTs are also associated with ER and vacuole membranes that produce high levels of anthocyanins. In Arabidopsis, the Golgi-localized membrane protein can transport proanthocyanidin to the central vacuole via vesicle trafficking. In our study, a large number of DEGs were classified into eight GO terms related to "cellular component", such as "plasma membrane" and "trans-Golgi network" (Fig. 4b). This finding suggests that membrane transport is likely to be an important regulator of anthocyanin accumulation in Pr.

The final flower color is determined by three major anthocyanins: pelargonidin, cyanidin, and delphinidin. Cyanidin contributes to red to magenta colors, and delphinidin contributes to magenta to purple colors. Anthocyanins are widely present in plants and are responsible for the purple coloration of plant stems and leaves. Several studies have reported that sugars (e.g., glucose, sucrose, maltose, and turanose) induce anthocyanin biosynthesis in Arabidopsis hypocotyls and leaves. The stems of cultivated *D. officinale* are rich in mannose and glucose in the mature stage, suggesting that the purple variety of *D. officinale* is a good model for studying the regulatory mechanisms of secondary metabolite accumulation. Dihydrokaempferol and dihydroquercetin provided substrates for quercetin derivative biosynthesis that were significantly up-regulated in Pr, especially rutin (Fig. 7a). Rutin is a potent antioxidant that inhibits sorbitol, reactive oxygen species, advanced glycation end-product precursors, and inflammatory cytokines. Therefore, the active medicinal components of *D. officinale* require further exploration for their pharmacological potential.

Auxin is an essential hormone that regulates almost every aspect of plant growth and development. Normally, auxin is transported from synthesis sites in the apices to the distal parts of the plant. Indole-3-acetic acid (IAA) is the primary natural auxin that plays a role in gravity-induced flavonoid synthesis in the root tips. IAA is synthesized via tryptophan (Trp)-dependent and Trp-independent pathways. Our data showed that the Trp content was reduced in Pr (Fig. 8d) and that most of the auxin-related genes were down-regulated in Pr (Fig. 8b). In Arabidopsis *tt* mutants, the flavonol-overproducing mutants exhibited reduced auxin transport, whereas the nonflavonoids-accumulating mutant increased auxin transport. These data suggest that high flavonoid accumulation is negatively correlated with auxin metabolism in Pr. Brassinolide, the most potent BR, is produced by several cytochrome P450 monooxygenases from campesterol. Similar to animal hormones, acetyl-CoA enters the mevalonic acid pathway to make mevalonate. Some studies have indicated that exogenous BR could stimulate the biosynthesis of flavonoids in leaves. Moreover, brassinolide signaling has been implicated in PIN2 sorting and intracellular distribution to delimit root gravitropism. Our transcriptomic data identified the "response to brassinosteroid" and "lipid and wax metabolism" GO terms, and the endogenous brassinolide contents were increased in Pr compared to CK (Fig. 8c, Figure S6). Therefore, brassinolide stimulated increased flavonoid concentrations in Pr. These results suggest that brassinolide signaling and auxin signaling crosstalk play important roles in the regulatory network of phenylpropanoid metabolism in *D. officinale*. 

Vol.:(0123456789)
stored at – 80 °C prior to the LC–MS analysis. In addition, pooled QC samples were prepared by combining 10 µL of each extraction mixture.

For metabolite separation, we used the method previously described by Dunn et al.44. First, all chromographic separations were performed using an ultra-performance liquid chromatography (UPLC) system (SCIEX, UK). An ACQUITY UPLC T3 column (100 mm × 2.1 mm, 1.8 µm, Waters, UK) was used for the reversed phase separation. The column oven was maintained at 35 °C. The flow rate was 0.4 mL/min and the mobile phase consisted of solvent A (water, 0.1% formic acid) and solvent B (Acetonitrile, 0.1% formic acid). Gradient elution conditions were set as follows: 0–0.5 min, 5% B; 0.5–7 min, 5% to 100% B; 7–8 min, 100% B; 8–8.1 min, 100% to 5% B; 8.1–10 min, 5% B. The injection volume for each sample was 4 µL.

A high-resolution tandem mass spectrometer TrippleTOF 5600plus (SCIEX, UK) was used to detect metabolites eluted from the column34. The Q-TOF was operated in both positive and negative ion modes. The curtain gas was set 30 PSI, ion source gas1 was set 60 PSI, ion source gas2 was set 60 PSI, and ion source temperature was 650 °C. For the positive ion mode, the ions pray voltage floating was set at 5000 V. For the negative ion mode, the ions pray voltage floating was set at 4500 V. The mass spectrometry data were acquired in IDA mode. The TOF mass ranged from 60 to 1200 Da. The survey scans were acquired in 150 ms and as many as 12 product ion scans were collected if a threshold of 100 counts per second (counts/s) was exceeded with a 1 + charge state. The total cycle time was fixed to 0.56 s. Four time bins were summed for each scan at a pulser frequency value of 11 kHz by monitoring the 40 GHz multichannel TDC detector with four-anode/channel detection. Dynamic exclusion was performed for 4 s. During the acquisition, the mass accuracy was calibrated for every sample. The online KEGG database was used to annotate the metabolites by matching the exact molecular mass data (m/z) of samples with those from the database. If the mass difference between the observed and the database value was less than 10 ppm, the metabolite would be annotated and the molecular formula of metabolites would further be identified and validated by the isotopic distribution measurements. We used an in-house fragment spectrum library of metabolites to validate metabolite identification. Student’s t-tests were conducted to detect differences in metabolite concentrations between the two phenotypes. The p-value was adjusted for multiple tests using the FDR (Benjamini–Hochberg).

Flavonoid and anthocyanin measurement. Flavonoids were washed twice with 70% ethanol to remove chlorophyll, and the remaining sediment was suspended in methanol (1:10, v/v) and incubated overnight in the dark at 4 °C to extract the flavonoids. The methanol extract was used to measure the concentration of total flavonoids, which was determined using the AlCl3 method as described previously52. Absorbance at 510 nm for flavonoids was determined, and rutin was used as the standard.

Anthocyanin content of the stem was determined using the protocol described by Mita et al.46. The stems (100 mg) were extracted for 1 day at 4 °C in 1 mL of 1% (v/v) hydrochloric acid in methanol. The mixture was centrifuged at 13,000 rpm for 15 min and the absorbance of the supernatant was measured at 530 nm and 650 nm.

Thin layer chromatography of all extracts was performed on TLC precoated silica gel 60 GF254 plate using hexane ethyl acetate formic acid (7:3:0.1) as solvent system. TLC plates were detected under UV light at 254 nm and 366 nm. TLC fingerprints were determined using the protocol outlined by Sithisarn et al.57.

Transcriptomic analysis. Stem samples from three independent biological replicates for both CK and Pr were harvested and frozen in liquid nitrogen. Total RNA was extracted using Trizol reagent (Invitrogen, CA, USA) following the manufacturer’s procedure. Total RNA quantity and purity were analyzed with Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, CA, USA) with RIN number > 7.0. Approximately 10 µg of total RNA representing a specific adipose type was subjected to isolation of (A) mRNA with poly-T oligo attached magnetic beads (Invitrogen). Following purification, the mRNA was fragmented into small pieces using divalent cations at elevated temperatures. Then, the cleaved RNA fragments were reverse-transcribed to create the final cDNA library in accordance with the protocol for the mRNASeq sample preparation kit (Illumina, San Diego, USA). The average insert size for the paired-end libraries was 300 bp (± 50 bp). Paired-end sequencing was performed on an IlluminaHiseq4000 (Illumina Inc., San Diego, CA, USA) by the Lianchuan Biotechnology Company (Hangzhou, China). The RNA-seq data were submitted to the BIG Data Center of the Chinese Academy of Sciences (https://bigd.big.ac.cn) with accession number CRA002691.

Differentially expressed genes analysis. The raw data were filtered using Trimmomatic (https://www.usadellab.org/cms/?page=trimmomatic) and then were assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) to obtain clean reads. Clean reads were aligned to the reference genome of D. catenatum2 using the HISAT program. Subsequently, the expression levels of unigene were calculated as FPKM (fragments per kilo bases of exons for per million mapped reads) using the software package StringTie. The DEGs were further characterized and estimated using the edgeR package according to the results from StringTie. FDRs ≤ 0.05, log2 fold-change (FC) > 1 or log2 FC < − 1 and with statistical significance (p value < 0.05) were used as threshold for determining the DEGs between Pr and CK. The DEGs were also subjected to Gene Ontology (GO) enrichment analysis using GOSep software and KEGG pathway enrichment analysis by KOBASE 2.0.35.
analyses were performed using a SYBRGreen qPCR kit (TransGenBiotech) with a MyiQ system (Bio-Rad) as described previously\(^\text{63}\). The primers are listed in Table S7.

**Statistical analysis.** The data are displayed as the mean ± standard deviation (SD). Statistical analyses were performed using SPSS 17.0 software (SPSS Inc., Chicago, IL USA), and an ANOVA was applied to compare the differences between the two groups. Data were also treated by hierarchical clustering using the R package FactoMineR (R version 3.6.0, https://cran.r-project.org/bin/windows/base/old/3.6.0/).

Received: 6 May 2020; Accepted: 6 October 2020
Published online: 19 October 2020

**References**

1. He, C. et al. A genome-wide identification of the WRKY family genes and a survey of potential WRKY target genes in *Dendrobium officinale*. *Sci. Rep.* 7, 9200 (2017).
2. Zhang, G. Q. et al. The Apostasia genome and the evolution of orchids. *Nature* 495, 379–383 (2013).
3. He, C. et al. Cytotoxic localization of polysaccharides in *Dendrobium officinale* and the involvement of DoCSLA6 in the synthesis of mannose polysaccharides. *Front. Plant Sci.* 8, 173 (2017).
4. He, C. et al. Structure and bioactivity of the polysaccharides in medicinal plant *Dendrobium huoshanense*. *Bioorg. Med. Chem.* 16, 6054–6068 (2008).
5. Ng, T. B. et al. Review of research on *Dendrobium*, a prized folk medicine. *Appl. Microbiol. Biotechnol.* 93, 1795–1803 (2012).
6. Yuan, Y. et al. Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC Genomics* 19, 978 (2018).
7. Jin, Q. et al. Metabolic analysis of medicinal *Dendrobium officinale* and *Dendrobium huoshanense* during different growth years. *PLoS ONE* 11, e0146607 (2016).
8. Lei, Z. et al. Transcriptome analysis reveals genes involved in flavonoid biosynthesis and accumulation in *Dendrobium catenatum* from different locations. *Sci. Rep.* 8, 6373 (2018).
9. Wang, X., Zou, L.-H., Zheng, B.-Q. & Wang, Y. Circadian regulation of alternative splicing of drought-associated CIPK genes in *Dendrobium catenatum* (Orchidaceae). *Int. J. Mol. Sci.* 20, 688 (2019).
10. Zhan, X. et al. Rice sulfooquinovosyltransferase SQD21 mediates flavonoid glycosylation and enhances tolerance to osmotic stress. *Plant Cell Environ.* 42, 2215–2230 (2019).
11. Nakabayashi, R. et al. Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. *Plant J.* 77, 367–379 (2014).
12. Falcone Ferreyra, M. L., Rius, S. P. & Casati, P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3, 222 (2012).
13. Cho, K. et al. Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. *J. Exp. Bot.* 67, 1519–1533 (2016).
14. Saito, K. et al. The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34 (2013).
15. Kusano, M. et al. Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of *Arabidopsis* to UV-B light. *Plant J.* 67, 354–369 (2011).
16. Dixon, R. A. & Paiva, N. L. Stress-induced polyphenol metabolism. *Plant Cell* 7, 1085–1097 (1995).
17. Brunetti, C., Sebastiani, F. & Tattini, M. Review: ABA, flavonoids, and the evolution of land plants. *Trends Plant Sci.* 20, 1495–1497 (2015).
18. Zhu, J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant Sci.* 20, 576–585 (2015).
19. Peer, W. A. & Murphy, A. S. Flavonoids and auxin transport: modulators or regulators?. *Trends Plant Sci.* 12, 556–563 (2007).
20. Buer, C. S. & Muday, G. K. The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of *Arabidopsis* roots to gravity and light. *Plant Cell* 16, 1191–1205 (2004).
21. Buer, C. S., Sukumar, P. & Muday, G. K. Ethylene modulates flavonoid accumulation and gravitropic responses in roots of *Arabidopsis*. *Plant Physiol.* 140, 1384–1396 (2014).
22. Watts, J. M., Hechler, P. J. & Muday, G. K. Ethylene-induced flavonoid accumulation in guard cells suppresses reactive oxygen species and moderates stomatal aperture. *Plant Physiol.* 164, 1707–1717 (2014).
23. Lu, G. et al. ABA mediates development-dependent anthocyanin biosynthesis and fruit coloration in *Lycium plants*. *BMC Plant Biol.* 19, 317 (2019).
24. Choudhury, F. K., Rivero, R. M., Blumwald, E. & Mittler, R. Reactive oxygen species, abiotic stress and stress combination. *Plant J.* 90, 856–867 (2017).
25. Bai, A. et al. Coordinate regulation of metabolite glycosylation and stress hormone biosynthesis by TT8 in *Arabidopsis*. *Plant Physiol.* 171, 2499 (2016).
26. Qi, T. et al. The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MBF complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell* 23, 1795–1814 (2011).
27. Petridis, A., Döll, S., Nichelmann, L., Bilger, W. & Mock, H. P. Rice sulfoquinovosyltransferase SQD21 mediates flavonoid glycosylation and enhances tolerance to osmotic stress. *Bioorg. Med. Chem.* 24, 978 (2016).
28. Friedrichsen, D. M. et al. Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* 162, 1445–1456 (2002).
29. Brito, A. F. & Zang, Y. A review of lignan metabolism, milk enterolactone concentration, and antioxidant status of dairy cows fed flaxseed. *Molecules* https://doi.org/10.3390/molecules24010041 (2018).
30. Pan, Y. et al. Composition analysis and antioxidant activity of polysaccharide from *Dendrobium denneanum*. *Int. J. Biol. Macromol.* 45, 169–173 (2009).
31. Sheng, C. et al. Identification and analysis of genes associated with the synthesis of bioactive constituents in *Dendrobium officinale* using RNA-Seq. *Sci. Rep.* 7, 187 (2017).
32. Lu, J. et al. High-density genetic map construction and stem total polysaccharide content-related QTL exploration for Chinese endemic *Dendrobium* (Orchidaceae). *Front. Plant Sci.* 9, 398 (2018).
33. Vogt, T. Polyphenol biosynthesis. *Mol. Plant* 3, 2–20 (2010).
The authors declare no competing interests.

Province Agricultural Chinese Medicine New Variety Breeding Major Science and Technology Special Project

The experiments. X.Z. analyzed data. X.Z. and B.M. wrote and revised the manuscript.

We thank Lianchuan Biotechnology Co. Ltd for transcriptomic and metabolomic detection.

Acknowledgements

We thank Dr. Shen Qingwen (Huazhong Agricultural University) who was consulted for the study and read the article. We thank Lianchuan Biotechnology Co. Ltd for transcriptomic and metabolomic detection.

Author contributions

X.Z. and B.M. conceived and designed the experiments. J.Q. and B.Z. collected samples. X.Z. and J.Q. performed the experiments. X.Z. analyzed data. X.Z. and B.M. wrote and revised the manuscript.

Funding

This work was financially supported by the Zhejiang Provincial Postdoctoral Research Program (No. 219756), the Key Research and Development Projects of Zhejiang Province in 2018 (No. 2018C02034), and Zhejiang Province Agricultural Chinese Medicine New Variety Breeding Major Science and Technology Special Project (No. 2016C02058-3).

Competing interests

The authors declare no competing interests.
