

RESEARCH ARTICLE

Open Access



Metabolome and transcriptome analyses of the molecular mechanisms of flower color mutation in tobacco

Fangchan Jiao^{1,2}, Lu Zhao^{1,2}, Xingfu Wu^{1,2}, Zhongbang Song^{1,2*} and Yongping Li^{1,2*}

Abstract

Background: Anthocyanins determinate the flower color of many plants. Tobacco is a model plant for studying the molecular regulation of flower coloration. We investigated the mechanism underlying flower coloration in tobacco by profiling flavonoid metabolites, expression of anthocyanin biosynthetic structural genes and their regulator genes in the pink-flowered tobacco cultivar Yunyan 87 and white-flowered Yunyan 87 mutant.

Result: Significant down-accumulation of anthocyanins, including cyanidin 3-O-glucoside, cyanin, cyanidin 3-O-rutinoside, pelargonidin 3-O-beta-D-glucoside, cyanidin O-syringic acid, pelargonin, and pelargonidin 3-O-malonylhexoside (\log_2 fold change < -10), endowed the flower color mutation in Yunyan 87 mutant. Transcriptome analysis showed that the coordinately down-regulated anthocyanin biosynthetic genes including chalcone isomerase, naringenin 3-dioxygenase, dihydroflavonol 4-reductase and UDP-glucose:flavonoid 3-O-glucosyltransferase played critical roles in suppressing the formation of the aforesaid anthocyanins. Several genes encoding MYB and bHLH transcription factors were also found down-regulated, and probably the reason for the suppression of structural genes.

Conclusion: This is the first study of tobacco flower coloration combining metabolome and transcriptome analyses, and the results shed a light on the systematic regulation mechanisms of flower coloration in tobacco. The obtained information will aid in developing strategies to modify flower color through genetic transformation.

Keywords: Tobacco, Flower coloration, Anthocyanin, Mutation, Gene expression

Background

Flower color is a key trait for ornamental plants, and flower coloration has been one of the hotspots in biological studies [1, 2]. Anthocyanins, carotenoids and betalains are the main pigments in plants [3]. Anthocyanins confer all orange, pink, red, purple, blue and blue-black flower colors [4, 5]. Carotenoids confer yellow color while betalains existing exclusively in Caryophyllales, are responsible for the yellow, orange, red and purple colors. In previous studies, anthocyanin contents were regarded as

the major factors endowing the flower color as pink or light red in transgenic tobacco [6–10]. The anthocyanin biosynthetic pathways have been characterized extensively in higher plants such as *Arabidopsis thaliana* [11], *Medicago truncatula* [12], *Fragaria ananasa* [13], *Vaccinium myrtillus* [14], *Trifolium repens* [15], *Antirrhinum majus* [16], *Vitis vinifera* [17, 18], *Daucus carota* [19], and *Zea mays* [20]. The conserved structural genes including chalcone synthase (*CHS*), chalcone isomerase (*CHI*), flavanone 3-hydroxylase (*F3H*), dihydroflavonol reductase (*DFR*), anthocyanidin synthase (*ANS*), and UDP-glucose:flavonoid 3-o-glucosyltransferase (*UFGT*) involved in anthocyanin biosynthetic pathway have been identified [3, 5, 21], but the regulatory mechanisms vary across plant species [22].

* Correspondence: zbsoon@vip.163.com; 1303046997@qq.com

¹Yunnan Academy of Tobacco Agricultural Sciences, Kunming 650021, Yunnan, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

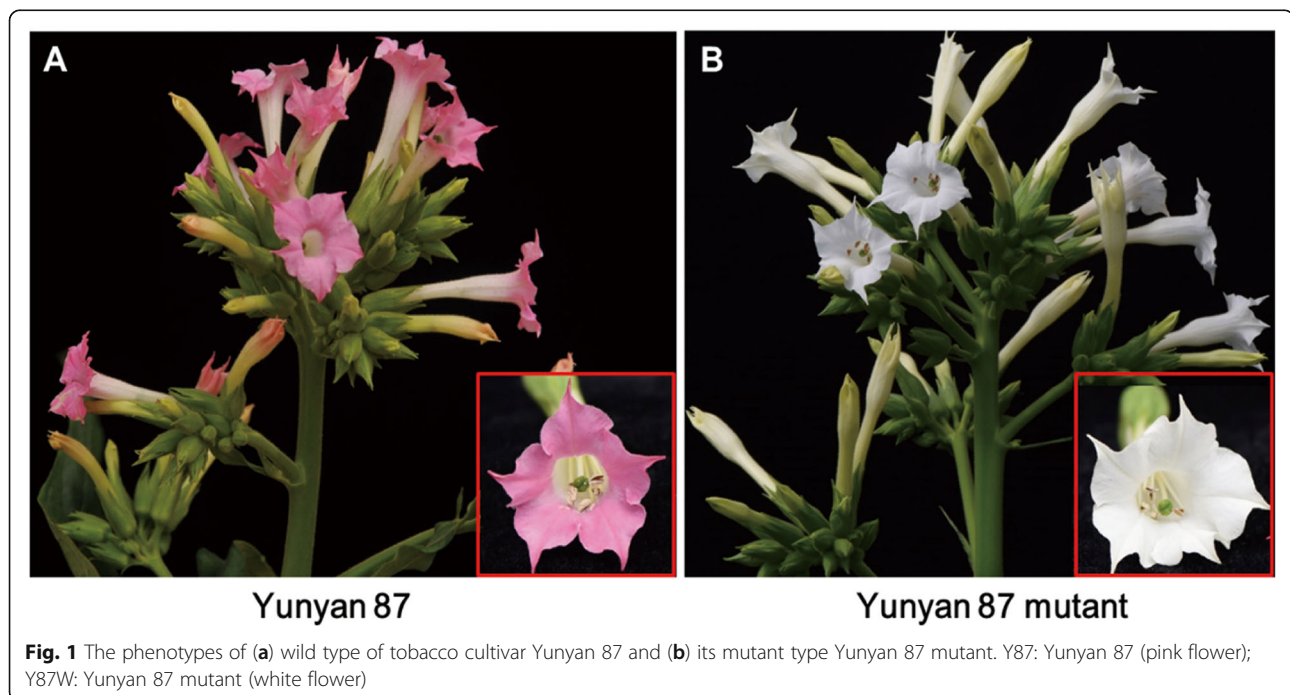
The anthocyanidin are glycosylated into anthocyanin by *UFGT* [23]. Flavonoid production has been limited by the inhibition of one step in the biosynthetic pathway in flowers of several plant species, including chrysanthemum, cyclamen, tobacco, gerbera, carnation, gentian, rose, lisianthus, petunia, and torenia [24, 25]. In addition, the anthocyanin biosynthesis is regulated by key transcription factors belonging to the families of *R2R3 MYB*, *bHLH* and *WD40*, interacting together to form a *MYB-bHLH-WD40* (MBW) complex [8, 26–29]. MYB transcription factors are able to induce anthocyanin accumulation in plants [26, 30–35]. While in some cases, some *MYBs* require co-expression of specific *bHLHs* to efficiently induce the anthocyanin biosynthesis [36, 37]. Tobacco (*Nicotiana tabacum*) is an important model plant for studying the molecular mechanisms of flower coloration. Unfortunately, the previous studies of flower coloration in tobacco are limited and mainly focused on the control of one of the core anthocyanin biosynthetic genes or the regulatory factors such as MYB transcription factors [7–10, 24, 25, 36]. Hence, a clear insight into the molecular mechanism underlying the tobacco flower coloration is still lacking. Recently, metabolomics and transcriptomics have facilitated identification of molecules and key genes involved color formation in plants [38–42]. Herein, we profiled the changes in metabolome and transcriptome between pink and white flower petals of tobacco in order to pinpoint key genes controlling anthocyanin composition in the flowers of Yunyan 87 and Yunyan 87 mutant.

Results

Metabolome profiling of the petal samples of pink-flowered and white-flowered tobacco

Fresh flowers with the pink and/or white petals were collected from tobacco cultivar Yunyan 87 and Yunyan 87 mutant, respectively (Fig. 1a, b). Flower samples were analyzed for metabolite concentrations between pink flowers of Yunyan 87 (Y87) and white flowers of Yunyan 87 mutant (Y87W).

Since anthocyanin content is determinant for flower color change in tobacco [6–10], we profiled the metabolome of the petal samples of Yunyan 87 and Yunyan 87 mutant using the flavonoid-targeted metabolomics approach [42]. Each sample had three replicates, so in total six samples were analyzed. We detected in the tobacco flower a total of 215 compounds, which could be grouped into nine classes including flavone, flavonol, flavone C-glycosides, anthocyanins, flavanone, isoflavone, catechin derivatives, proanthocyanidins, flavonolignan (Fig. 2a; Table S1). A hierarchical heatmap clustering analysis of the petal samples of Yunyan 87 and Yunyan 87 mutant was performed based on the metabolite quantification. All of the biological replicates were clustered together which indicated that the quality of the generated metabolome data was high (Fig. 2b). Interestingly, the pink flower samples (Y87) and the white flower samples (Y87W) were clearly separated. The result indicated that distinct flavonoid profiles in Y87 and Y87W samples. In consistence, the metabolites were also clustered



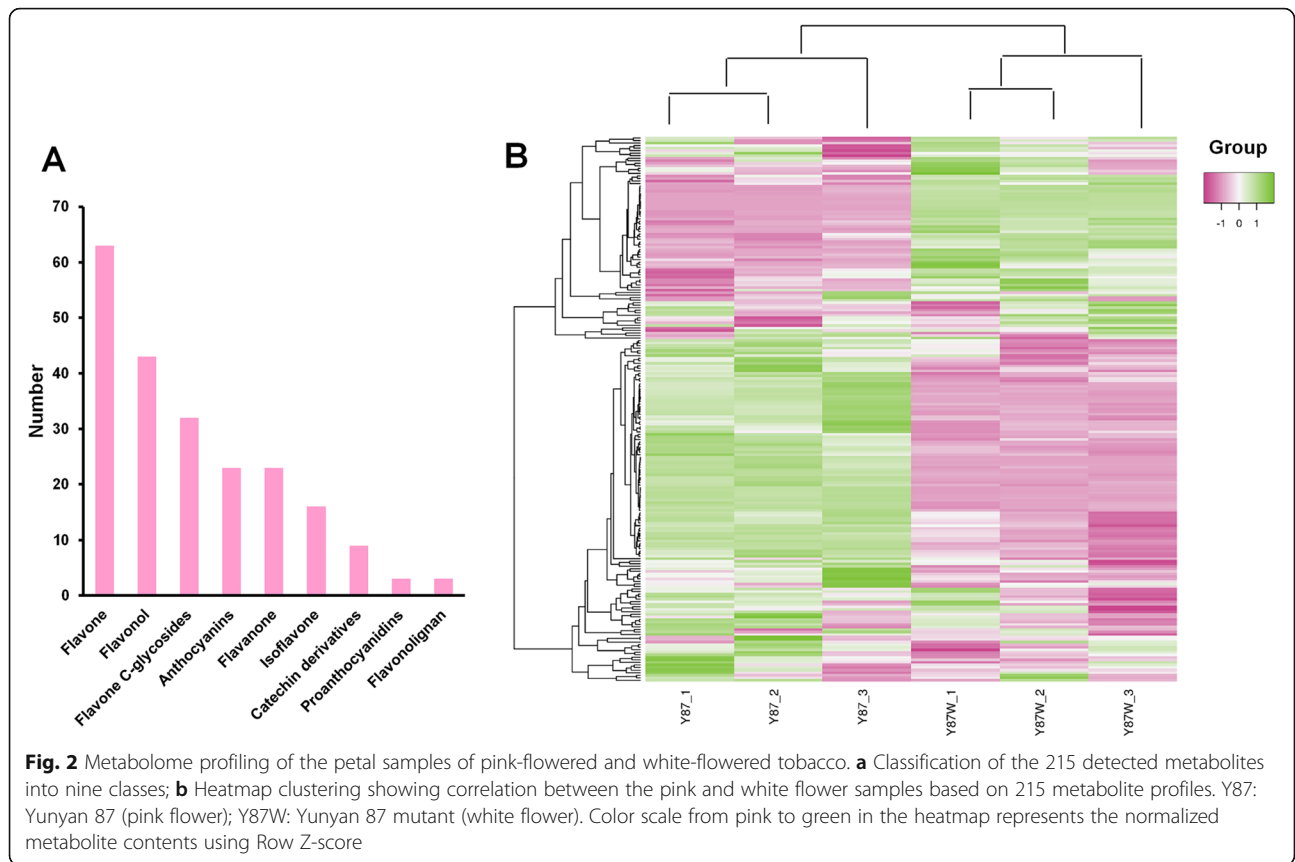


Fig. 2 Metabolome profiling of the petal samples of pink-flowered and white-flowered tobacco. **a** Classification of the 215 detected metabolites into nine classes; **b** Heatmap clustering showing correlation between the pink and white flower samples based on 215 metabolite profiles. Y87: Yunyan 87 (pink flower); Y87W: Yunyan 87 mutant (white flower). Color scale from pink to green in the heatmap represents the normalized metabolite contents using Row Z-score

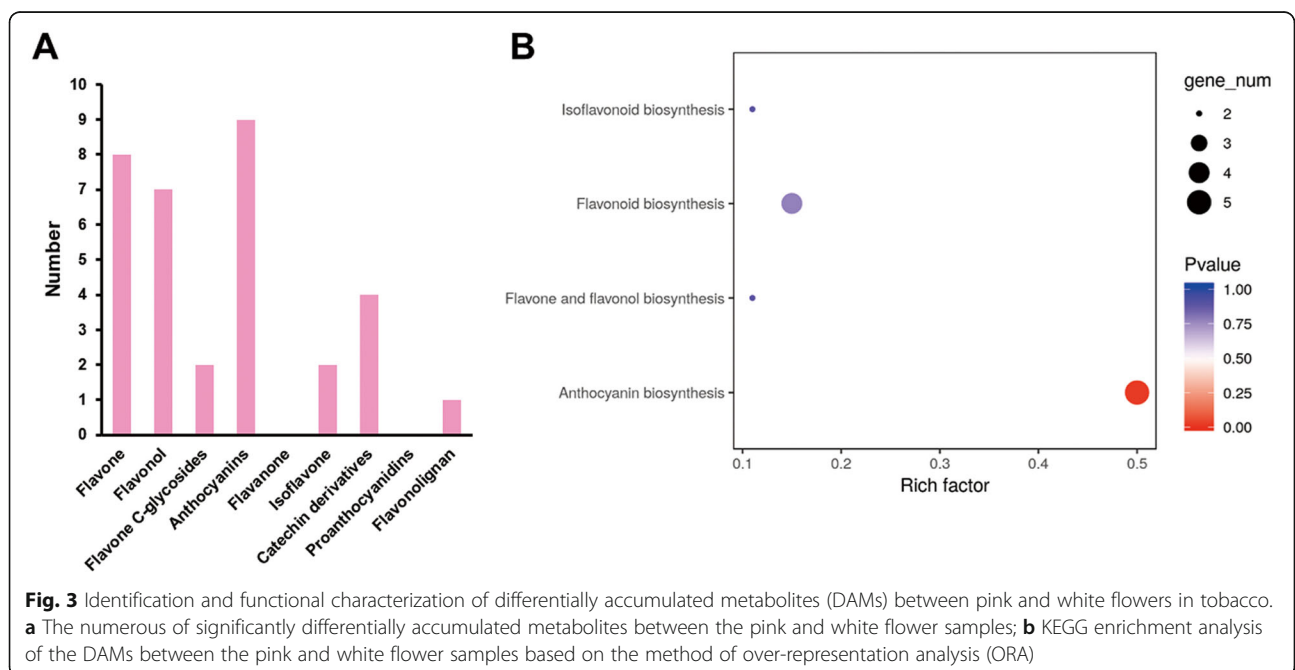


Fig. 3 Identification and functional characterization of differentially accumulated metabolites (DAMs) between pink and white flowers in tobacco. **a** The numerous of significantly differentially accumulated metabolites between the pink and white flower samples; **b** KEGG enrichment analysis of the DAMs between the pink and white flower samples based on the method of over-representation analysis (ORA)

into two main groups demonstrating the opposite accumulation patterns between the pink-colored and white-colored flower samples (Fig. 2b).

Identification of the differentially accumulated metabolites between pink and white flowers in tobacco

The differentially accumulated metabolites (DAMs) in petal samples between Yunyan 87 and Yunyan 87 mutant (Y87 vs. Y87W) were determined based on the variable importance in projection (VIP) ≥ 1 and fold change ≥ 2 or fold change ≤ 0.5 [43]. There were 33 DAMs showing significantly different accumulation between the compared samples (Fig. 3a; Table S2). The most enriched KEGG term among the DAMs detected for the compared samples was anthocyanin biosynthesis, based on the method of over-representation analysis (ORA) (Fig. 3b).

Interestingly, all the differentially accumulated anthocyanins were down-accumulated in white flower samples, including the extremely significantly down-accumulated (\log_2 fold change < -10) anthocyanins (cyanidin 3-O-glucoside, cyanin, cyanidin 3-O-rutinoside, pelargonidin 3-O-beta-D-glucoside, cyanidin O-syringic acid, pelargonin, and pelargonidin 3-O-malonylhexoside) (Table 1). This result indicates that the anthocyanin branch was blocked in the white mutant phenotype.

Analysis of differentially expressed genes between pink and white-colored tobacco flowers using RNA-sequencing

We synthesized six cDNA libraries from flowers collected from Yunyan 87 and Yunyan 87 mutant plants and generated transcriptome RNA-sequencing (RNA-seq) data in Yunyan 87 and its mutant phenotype Yunyan 87 mutant. The RNA-seq yielded a total of 46.99 Gb clean data with 91.82% of bases scoring Q30 and above (Table 2).

Of the total clean reads, 83.04 to 85.01% were mapped in proper pairs with the *Nicotiana tabacum* reference genome [44] (Table S3). Principal component analysis (PCA) of the samples based on fragments per kilobase of

exon model per million reads mapped (FPKM) clearly separated the two flower sample types, implying that the differentially accumulated metabolites between the two phenotypes are regulated by differential expressed genes (Fig. 4a).

To identify the differentially expressed genes (DEGs) between Y87 and Y87W, the raw standardized count of assembled unigenes were imputed in DESeq and edgeR packages. Only DEGs commonly detected by both packages were used in this study. After comparison, we obtained a total of 3715 DEGs, including 3045 up-regulated and 670 down-regulated genes in the tobacco flowers (Fig. 4b). To further validate the results of the DEGs, eight differentially expressed genes were selected (Table S4) and their expression levels from flowers of the Y87 (pink) and Y87W (white) were analyzed using qRT-PCR. The strongly correlation between the qRT-PCR results and the RNA-seq data ($R^2 = 0.8994$, Figure S1) confirmed the RNA-seq data obtained in this study are highly reliable.

Major transcription factors were differentially regulated between pink and white flowers in tobacco

Previous study demonstrated that *MYB* alone, co-expression of *MYB* and *bHLH*, and formation of the *MYB-bHLH-WD40* complex were all sufficient to induce anthocyanin accumulation in plants [8, 29–31, 37]. Given the down-regulated anthocyanin accumulation in white flowers (Table 1), we focused on the most significantly down-regulated transcription factors *MYB*, *bHLH* and *WD40* (\log_2 fold change < -4). Expression fold change of these TFs showed that three *MYB* genes (*MYB113*, *MYB3*, and *MYB PHL11*) and two *bHLH* genes (*bHLH162*; *gene14770* and *gene27472*) were strikingly down-regulated in white flowers of Yunyan 87 mutant (Table 3; Table S5). In other words, the aforesaid *MYB* genes and *bHLH* genes were significantly up-regulated in pink flower of Yunyan 87. The results indicated that inhibition of the *MYB* genes (*MYB113*,

Table 1 Differentially accumulated anthocyanins between pink and white flowers in tobacco. Values represent the ion intensity (concentration) of the metabolites

| Metabolite code | Metabolite name | Y87 (ion intensity) | Y87W (ion intensity) | Log ₂ fold change | VIP |
|-----------------|---|---------------------|----------------------|------------------------------|------|
| pmb0550 | Cyanidin 3-O-glucoside (Kuromanin) | 4,060,000 | 9.00 | -18.8 | 3.93 |
| pme1777 | Cyanidin 3,5-O-diglucoside (Cyanin) | 2,780,000 | 9.00 | -18.2 | 3.87 |
| pme1773 | Cyanidin 3-O-rutinoside (Keracyanin) | 2,720,000 | 9.00 | -18.2 | 3.87 |
| pme3392 | Pelargonidin 3-O-beta-D-glucoside (Callistephin chloride) | 1,510,000 | 9.00 | -17.4 | 3.78 |
| pmb2957 | Cyanidin O-syringic acid | 879,000 | 9.00 | -16.6 | 3.69 |
| pme1793 | Pelargonin | 66,400 | 9.00 | -12.8 | 3.25 |
| pmb0554 | Pelargonidin 3-O-malonylhexoside | 48,300 | 9.00 | -12.4 | 3.19 |
| pme1397 | Pelargonidin | 393,000 | 110,000 | -1.84 | 1.23 |
| pmb0558 | Delphinidin O-malonyl-malonylhexoside | 44,100 | 16,900 | -1.38 | 1.06 |

Table 2 Overview of the transcriptome sequencing dataset and quality check

| Samples | clean_reads | clean_bases | error_rate | Q20 | Q30 | GC (%) |
|---------|-------------|---------------|------------|--------|--------|--------|
| Y87_1 | 50,822,512 | 7,623,376,800 | 0.35 | 97.145 | 92.6 | 42.99 |
| Y87_2 | 46,783,110 | 7,017,466,500 | 0.37 | 96.955 | 91.975 | 42.855 |
| Y87_3 | 51,680,598 | 7,752,089,700 | 0.345 | 97.225 | 92.6 | 42.875 |
| Y87W_1 | 54,330,202 | 8,149,530,300 | 0.355 | 97.135 | 92.31 | 42.945 |
| Y87W_2 | 56,821,124 | 8,523,168,600 | 0.375 | 96.875 | 91.825 | 42.97 |
| Y87W_3 | 52,801,194 | 7,920,179,100 | 0.365 | 96.955 | 92.035 | 42.985 |

MYB3, and *MYB PHL11*) and *bHLH* genes (*bHLH162*; *gene14770* and *gene27472*) may contribute to manipulating the changes in flower color from pink to white in tobacco cultivar Yunyan 87.

Modulation of anthocyanin biosynthesis and accumulation in tobacco flowers

Flower color is mainly determined by the composition and concentrations of the plant’s pigments [1]. In this study, four key genes including *CHI* (*gene66881*), *F3H* (*gene9170*), *DFRs* (*gene59526*, *gene61321*, *gene64820*), and *UFGTs* (*gene13307*, *gene43584*) involved in the anthocyanin biosynthesis were significantly down-regulated, which is consistent with the low anthocyanin contents in Yunyan 87 mutant (Fig. 5a, b). Surprisingly, the enzyme directly involved in anthocyanin biosynthesis, (anthocyanidin

synthase; *ANS*) showed a higher level in Yunyan 87 mutant (Fig. 5b), which contradicts the down-accumulation of anthocyanins. In addition, our study demonstrated that the *MYB* genes (*MYB113*, *MYB3*, and *MYB PHL11*) and *bHLH* genes (*bHLH162*) were significantly down-regulated in white flowers of Yunyan 87 mutant. Taken together, our results suggest that the coordinately down-regulations of *CHI* (*gene66881*), *F3H* (*gene9170*), *DFRs* (*gene59526*, *gene61321*, *gene64820*), and *UFGTs* (*gene13307*, *gene43584*) plays critical roles in suppressing the formation of the anthocyanins including cyanidin 3-O-glucoside, cyanin, cyanidin 3-O-rutinoside, pelargonidin 3-O-beta-D-glucoside, cyanidin O-syringic acid, pelargonin, and pelargonidin 3-O-malonylhexoside (log₂ fold change < - 10), endowing the changes in flower color from pink to white in tobacco.

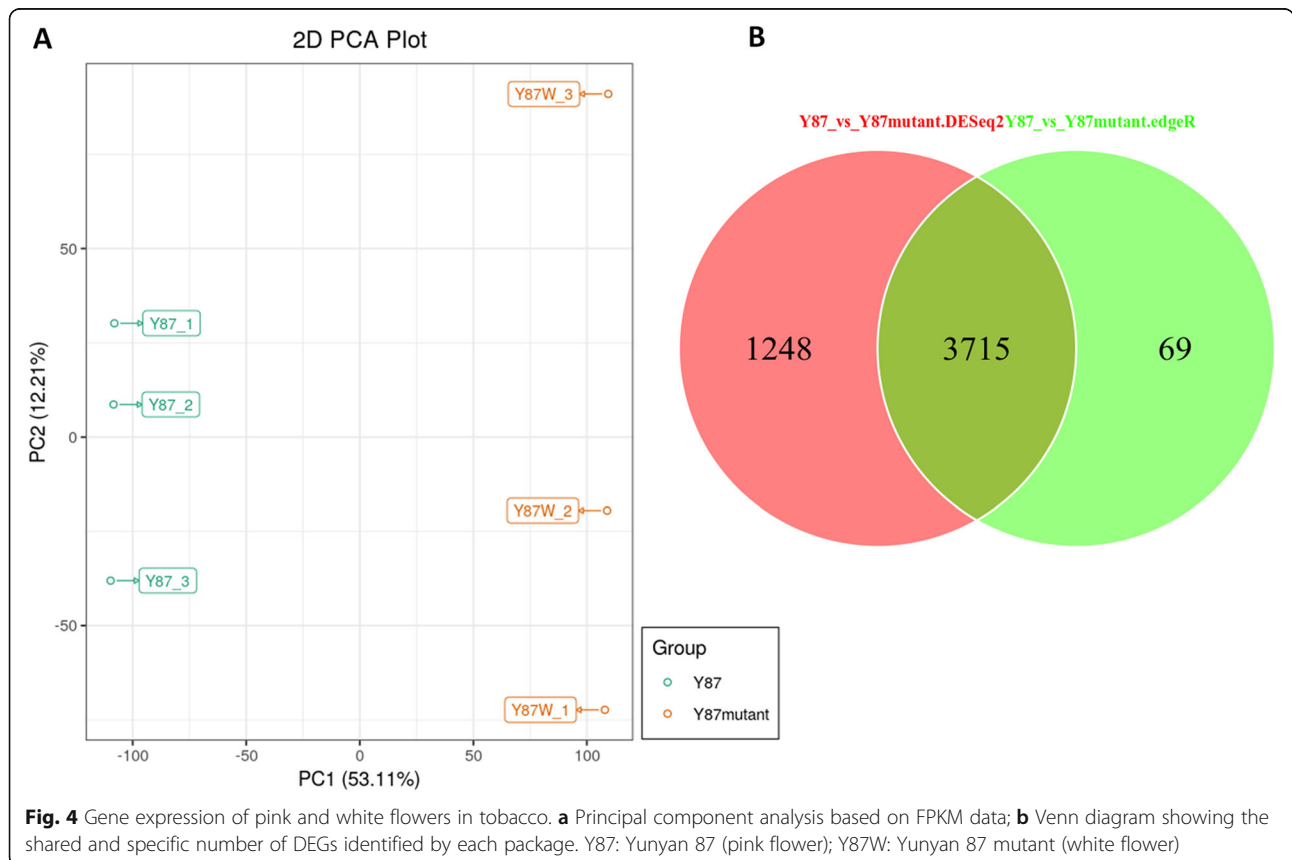


Fig. 4 Gene expression of pink and white flowers in tobacco. **a** Principal component analysis based on FPKM data; **b** Venn diagram showing the shared and specific number of DEGs identified by each package. Y87: Yunyan 87 (pink flower); Y87W: Yunyan 87 mutant (white flower)

Table 3 Key transcription factors involved in the flower color change from pink to white in tobacco

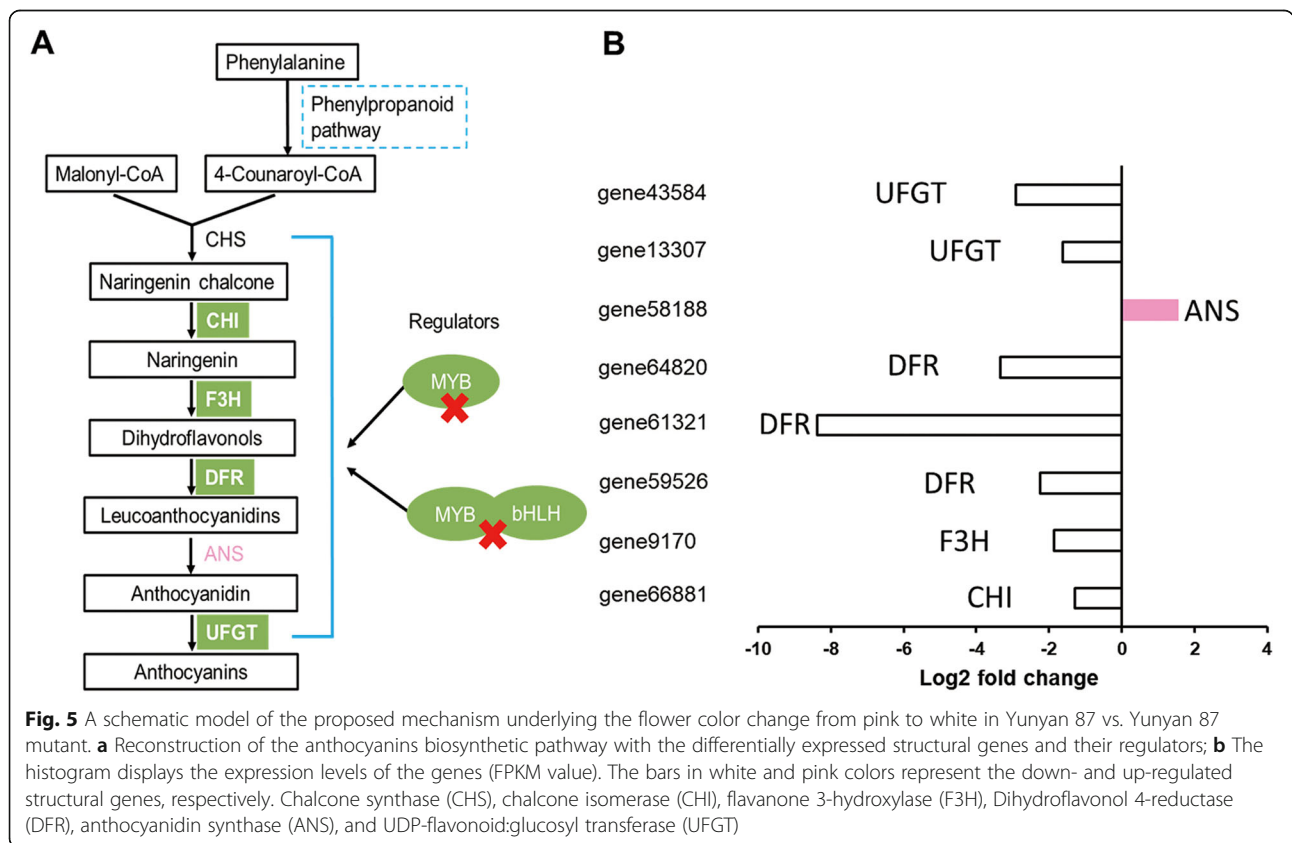
| Transcription factors | Gene ID | Log ₂ fold change | Gene description |
|-----------------------|-----------|------------------------------|------------------|
| MYB | gene54221 | -9.277 | MYB113 |
| | gene14314 | -6.073 | MYB3 |
| | gene31475 | -4.868 | MYB PHL11 |
| bHLH | gene14770 | -5.740 | bHLH162 |
| | gene27472 | -4.071 | bHLH162 |

Discussion

The flower color is an important trait in many plants and flower coloration has been one of the hotspots in biological studies [1, 2, 44]. Although color formation is well studied in plants, there are species-specific peculiarities of pigment regulation. For example, some plants only harbor betalains while other harbor anthocyanins as pigments. Moreover, the numbers of structural genes involved in the biosynthetic pathways of these pigments varied considerably across species. Finally, various mechanisms could lead to different colorations such as competition between pathways, mutations in structural genes or regulatory genes, activity of microRNAs, etc. Therefore, studying color formation in a species or new genetic materials can reveal specific genes and specific

mechanism of regulation, which will generate new knowledge in this field. In this study, we observed two distinct flower phenotypes in tobacco cultivar Yunyan 87 and Yunyan 87 mutant. The wild phenotype (Yunyan 87) exhibits pink-colored petals, while its mutant phenotype (Yunyan 87 mutant) presents white-colored petals (Fig. 1a, b). Therefore, the tobacco cultivar Yunyan 87 and its mutant12 are desired materials to investigate the systematic molecular mechanisms of flower color mutation in tobacco.

Flavonoids are the major molecules involved in plant pigmentation [37]. There are mainly six groups of flavonoids in plant tissues, including anthocyanins, flavan-3-ols (catechins and proanthocyanidins), flavanonols, flavonols, flavones and phenolic acid [38]. The previous studies on tobacco coloration process were limited to very few flavonoids metabolites such as anthocyanins, flavonols, etc. [7, 45, 46]. In this study, using the flavonoid-targeted metabolomics approach, we investigated the changes in 215 metabolites between the pink flowers of Yunyan 87 and white flowers of Yunyan 87 mutant (Fig. 2a; Table S2), aiming at providing a more comprehensive landscape of the metabolites involved in the flower coloration in tobacco. Analysis of the differentially accumulated metabolites between the two



phenotypes revealed that the significantly decreased anthocyanins are the main metabolites modulating the color change of tobacco flowers from pink to white. Anthocyanins comprise a class of water-soluble pigments in plants which contribute to the color of flowers, fruits, stems and leaves [47, 48]. Previous studies highlighted that anthocyanin contents were regarded as the major factors endowing the flower color as pink or light red in transgenic tobacco [6–10]. However, they did not identify the specific anthocyanins responsible for the color alteration of tobacco petals from pink to white. In this study, we found that cyanidin 3-O-glucoside, cyanin, cyanidin 3-O-rutinoside, pelargonidin 3-O-beta-D-glucoside, cyanidin O-syringic acid, pelargonin, and pelargonidin 3-O-malonylhexoside were all extremely significantly down-accumulated (\log_2 fold change < -10) in white-flowered tobacco in comparison with the pink-flowered tobacco and may be regarded as the major anthocyanins contributing to the color change from pink to white in tobacco.

By manipulating the structural genes of anthocyanin biosynthetic pathway or their regulatory genes, flower colors of several species have been modified [3, 6–10]. We revealed that most of the structural genes (*CHI*, *F3H*, *DFR*, and *UFGT*) involved in the anthocyanin biosynthetic pathway were significantly down-regulated in the white flowers, with the exception of *ANS* gene (Fig. 5b). The up-regulation of *ANS* gene might be the result of feedback regulation [49]. *DFR* plays a key role in the formation of anthocyanins, directly determines the development of pink or white color in tobacco flowers [50]. Interestingly, our study reveals that *DFR* genes (*gene61321*, *gene64820*, and *gene 59,526*) are the most significantly down-regulated genes, which seem to be one of the important candidate genes in determining the flower color in tobacco cultivar Yunyan87. The *UFGT* genes (*gene43584* and *gene13307*) were strictly inhibited in white flowers. It has been reported that the down-regulation of *McDFR* in apple fruit reduced the expression levels of some structural genes (*F3H*, *F3'H*, *DFR*, *ANS* and *UFGT*), while the *CHS* and *CHI* genes were up-regulated [51]. It indicates that the altered expression of *DFR* also affected the expression of other anthocyanin biosynthetic genes. In this study, it is likely that the significant down-regulation of *DFR* genes led to the concomitant decreases in the expression levels of *CHI*, *F3H*, *DFR*, and *UFGT* genes.

MYB alone, co-expression of MYB and bHLH, and formation of the MYB-bHLH-WD40 complex were all sufficient to induce anthocyanin accumulation in plants [8, 22, 29–31, 49]. In tobacco, the regulatory effects of MYB transcription factors (such as *PamMybA.1*, *PamMybA.3* and *PamMybA.5*) resulted in different accumulation patterns of anthocyanins [8]. Ectopic expression of

maize bHLH transcription factor Lc enhanced anthocyanin concentrations in many plants including tobacco [7]. In this study, three MYB genes (*MYB113*, *MYB3*, and *MYB PHL11*) and two bHLH genes (*bHLH162*; *gene14770* and *gene27472*) were strikingly down-regulated in white flowers of Yunyan 87 mutant (Table 3; Table S5). The down-regulated MYB and bHLH transcription factors will not activate the expression of structural genes involved in anthocyanin biosynthetic pathway, leading to decreased anthocyanin concentrations. This indicates that the altered flower color in Yunyan 87 mutant might be attributed to the inhibition of MYB genes (*MYB113*, *MYB3*, and *MYB PHL11*) and bHLH genes (*bHLH162*; *gene14770* and *gene27472*). However, further studies are required to identify the spontaneously mutated gene(s) in Yunyan 87 mutant, and evaluate how the mutations can cause the reductions in the expression levels of these transcription factors and then the structural genes. This will decipher the systematic regulation mechanisms of the flower coloration in tobacco and benefit the genetic engineering of flower modification in other plants.

Conclusion

We combined metabolome and transcriptome data to decipher the molecular mechanisms underlying flower coloration in tobacco. By comparing pink and white colored flowers, our data showed that strong down-regulation of anthocyanin biosynthetic structural genes correlated with the significant reduction of anthocyanins in the white flower as compared to the pink samples. Several transcription factors mainly MYB and bHLH were predicted to regulated the anthocyanin biosynthetic structural genes. Collectively, this study offers candidate genes for functional characterization and for manipulation of flower color in tobacco.

Methods

Plant materials

The Yunyan 87 mutant, discovered in the field, is a natural mutant of the pink-flowered tobacco cultivar Yunyan 87. The plant materials are available at Yunnan Academy of Tobacco Agricultural Sciences, China. The formal identification of the plant materials was undertaken by the corresponding author of this article (Professor Yongping Li). No voucher specimen of this material has been deposited in a publicly available herbarium. The plants of tobacco Yunyan 87 and Yunyan 87 mutant were grown under controlled conditions at Yanhe, Yuxi, China. The day and night growth temperatures were 28 °C and 25 °C, respectively. During flowering stages, the flowers showed two distinct petal colorations in Yunyan 87 (wild phenotype, pink) and Yunyan 87 mutant (mutant phenotype, white) (Fig. 1). The fresh

flowers with pink petals from tobacco cultivar Yunyan 87 and white petals from Yunyan 87 mutant were harvested and named as Y87 and Y87W, respectively. Each Y87 or Y87W sample, consisted of 20 flowers, was frozen in liquid nitrogen and stored at -80°C until further use. Three biological replicates (20 individual flowers/replicate) were analyzed for Y87 and Y87W.

Metabolic profiling

The sample preparation and metabolite analysis were performed according to the methods as previously described by Yuan et al. [43]. The flower samples collected from Y87 and Y87W were crushed to powder and subjected to LC-MS analysis [52].

Quality control (QC) analysis was conducted before the data analysis. PLS-DA analysis was applied to calculate the corresponding variable importance in projection (VIP) value. When the $\text{VIP} \geq 1$, and fold change ≥ 2 or fold change ≤ 0.5 , the metabolites were considered as differentially changed metabolites. Heatmap clustering analysis was performed in the R software (www.r-project.org).

RNA extraction, library preparation, and sequencing

For transcriptome sequencing, six libraries representing the collected petal samples of Y87 and Y87W (three replicates of each) were constructed. Total RNAs were extracted using TRIzol reagent (TaKaRa, China). RNA contamination and RNA integrity number (RIN) were monitored on 1% agarose gels and the Agilent 2100 Bioanalyzer system (Agilent Technologies, CA, USA), respectively. A total amount of $3\ \mu\text{g}$ RNA per sample was used as input material for construction of pair-end (PE) sequencing libraries. Following manufacturer's instructions, the libraries were generated using NEBNext[®] UltraTM RNA Library Prep Kit for Illumina[®] (NEB, USA), and then added index codes to attribute sequences in each sample. According to the manufacturer's recommendations of TruSeq PE Cluster Kit v3-cBot-HS (Illumina), the clustering of the index-coded samples was performed on a cBot Cluster Generation System. Following the libraries were sequenced by paired-end sequencing on Illumina HiSeq platform.

Assembled transcriptome data of Illumina HiSeq sequencing

The raw paired-end reads were cleaned through removing adaptor sequences, poly-N, and low quality sequences. The FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to trim the adaptor sequences and low quality sequences (i.e., the percentage of bases of quality value ≤ 5 exceeds 50% in the read). Meanwhile, short sequences (< 50 bp) were also removed using a custom Perl program. The clean data with high quality were applied to the downstream analyses.

Reference genome and gene model annotation files were downloaded from ftp://ftp.sgn.cornell.edu/genomes/Nicotiana_tabacum/edwards_et_al_2017/ [53]. Index of the reference genome was built and paired-end clean reads were aligned to the reference genome using Hisat2 [54]. The transcriptome data of pink-flowered Yunyan 87 and white-flowered Yunyan 87 mutant have been deposited to national center for biotechnology information (NCBI) sequence read archive (SRA) under accession number PRJNA590063 (<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA590063>)

Differential expression analysis

The reads numbers mapped to each gene were counted using featureCounts v1.5.0-p3 [55]. Then, calculating the expected number of fragments per kilobase of exon model per million reads mapped (FPKM) of each gene based on the length of each gene and reads count mapped to the gene. DEGs of two groups of colored samples were identified using the DESeq R package (v1.18.0) [56] and edgeR package (v 3.24.3). The threshold p -value in multiple tests to judge the significance of gene expression difference was based on false discovery rate (FDR) method. When $\text{FDR} \leq 0.05$ and FPKM values showing at least 2-fold difference among samples, the gene was considered as significant DEG [57]. DEGs commonly detected by both packages were used in this study.

Validation of gene expression using qRT-PCR

RNA-seq samples (in triplicate) were used for quantitative real time-PCR (qRT-PCR) to test the dependability of the transcriptome results following descriptions of Dossa et al. [58]. Primer pairs of eight selected genes were designed using the Primer Premier 5.0 [59] (Table S4). Data are presented as relative transcript level based on the $2^{-\Delta\Delta\text{Ct}}$ method [60]. We estimated the Pearson correlation coefficient between the gene expression profiles in the qRT-PCR and RNA-seq in R version 3.6.3.

GO and KEGG enrichment analysis of DEGs

The GOseq R package [61] was used to analysis of Gene Ontology (GO) enrichment with DEGs. The adjusted P -value of significantly substantiated GO terms was less than 0.05. The KEGG pathways enriched with DEGs ($\text{FDR} < 0.05$) were detected using KOBAS 2.0 software [62] based on the method of over-representation analysis (ORA).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-07028-5>.

Additional file 1: Table S1. List, characteristics and concentration (ion intensity) of the metabolites detected in pink-colored and white-colored tobacco flowers.

Additional file 2: Table S2. List, characteristics and concentration (ion intensity) of the differentially accumulated metabolites between pink-colored and white-colored tobacco flowers.

Additional file 3: Table S3. Overview of the transcriptome sequencing dataset and mapping statistics in tobacco flowers collected from Yunyan 87 (Y87) and Yunyan 87 mutant (Y87W).

Additional file 4: Table S4. The primer sequences of genes used for quantitative real-time PCR.

Additional file 5: Table S5. Expression fold change of up- and down-regulated transcription factors MYB, bHLH and WD40.

Additional file 6: Figure S1. qRT-PCR results of 8 selected genes (left) and correlation between transcriptome data and real time PCR results (right).

Abbreviations

CHI: Chalcone isomerase; *F3H*: Naringenin 3-dioxygenase; *DFR*: Dihydroflavonol 4-reductase; *UGFT*: UDP-glucose:flavonoid 3-O-glucosyltransferase; *CHS*: Chalcone synthase; *ANS*: Anthocyanidin synthase; *MBW*: *MYB-bHLH-WD40*; *LIT*: Linear ion trap; *QQQ*: Triple quadrupole; *Q TRAP*: Quadrupole-linear ion trap mass spectrometer; *MRM*: Multiple reaction monitoring; *DP*: De-clustering potential; *CE*: Collision energy; *VIP*: Variable importance in projection; *PLS-DA*: Partial least squares-discriminant analysis; *DAMs*: Differentially accumulated metabolites; *DEGs*: Differentially expressed genes

Acknowledgements

Not applicable.

Authors' contributions

Conception and designed the study: Y L, Z S. Drafted the manuscript: F J, Z S. Performed the experiments: F J, L Z, X W. Analyzed the data: F J, Z S. All authors read and approved the final version of this manuscript.

Funding

This work was supported by the Yunnan Academy of Tobacco Agricultural Sciences (Grant No. YNTC-2017YN05). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

RNA-seq data is available at the SRA database in National Center of Biotechnology Information with the accession number PRJNA590063.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Yunnan Academy of Tobacco Agricultural Sciences, Kunming 650021, Yunnan, China. ²National Center for Tobacco Gene Engineering, Kunming 650021, Yunnan, China.

Received: 6 May 2020 Accepted: 27 August 2020

Published online: 07 September 2020

References

- Mol J, Grotewold E, Koes R. How genes paint flowers and seeds. *Trends Plant Sci.* 1998;3:212–7.
- Ben Meir H, Zuker A, Weiss D, Vainstein A. In: Vainstein A, editor. *In Breeding for Ornamentals: Classical and Molecular Approaches*. Springer; 2002. p. 253–72.
- Grotewold E. The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol.* 2006;57:761–80. <https://doi.org/10.1146/annurev.arplant.57.032905.105248>.
- Tanaka Y, Sasaki N, Ohmiya A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 2008;54:733–49. <https://doi.org/10.1111/j.1365-3113X.2008.03447.x>.
- Davies KM, Schwinn KE. Molecular Biology and Biotechnology of Flower Pigments. In: *Plant Developmental Biology-Biotechnological Perspectives*; 2010. p. 161–87.
- Kumar V, Yadav SK. Overexpression of *CsANR* increased flavan-3-ols and decreased anthocyanins in transgenic tobacco. *Mol Biotechnol.* 2013;54:426–35. <https://doi.org/10.1007/s12033-012-9580-1>.
- Huang ZA, Zhao T, Wang N, Zheng SS. Ectopic expression of *Lc* differentially regulated anthocyanin biosynthesis in the floral parts of tobacco (*Nicotiana glauca* L.) plants. *Botanical Stud.* 2016;57:24. <https://doi.org/10.1186/s40529-016-0138-6>.
- Dasgupta K, Thilmony R, Stover E, Oliveira ML, Thomson J. Novel R2R3-MYB transcription factors from *Prunus americana* regulate differential patterns of anthocyanin accumulation in tobacco and citrus. *GM Crops Food.* 2017;8:85–105. <https://doi.org/10.1080/21645698.2016.1267897>.
- Nishihara M, Nakatsuka T, Yamamura S. Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. *FEBS Lett.* 2005;579:6074–8. <https://doi.org/10.1016/j.febslet.2005.09.073>.
- Yang CC, Yu JJ, Zhao Q, Zhu DY, Ao GM. Influence of maize *Lc* regulatory gene on flower colour of transgenic tobacco and petunia. *J Agric Biotechnol.* 2007;15:85–9.
- Lepiniec L, et al. Genetics and biochemistry of seed flavonoids. *Ann Rev Plant Biology.* 2006;57:405–30. <https://doi.org/10.1146/annurev.arplant.57.032905.105252>.
- Pang Y, Peel GJ, Wright E, Wang Z, Dixon RA. Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. *Plant Physiology.* 2007;145:601–15. <https://doi.org/10.1104/pp.107.107326>.
- Aharoni A, et al. The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J.* 2001;28:319–32. <https://doi.org/10.1046/j.1365-3113x.2001.01154.x>.
- Jaakola L, et al. Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiology.* 2002;130:729–39. <https://doi.org/10.1104/pp.006957>.
- Abeynayake SW, et al. Biosynthesis of proanthocyanidins in white clover flowers: cross talk within the flavonoid pathway. *Plant Physiology.* 2012;158:666–78. <https://doi.org/10.1104/pp.111.189258>.
- Schwinn K, et al. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell.* 2006;18:831–51. <https://doi.org/10.1105/tpc.105.039255>.
- Gargouri M, et al. Structure and epimerase activity of anthocyanidin reductase from *Vitis vinifera*. *Acta Crystallogr Sect D.* 2009;65:989–1000. <https://doi.org/10.1107/S0907444909025013>.
- Terrier N, et al. Ectopic expression of VvMybPA2 promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiology.* 2009;149:1028–41. <https://doi.org/10.1104/pp.108.131862>.
- Xu ZS, Feng K, Que F, Wang F, Xiong AS. A MYB transcription factor, DcMYB6, is involved in regulating anthocyanin biosynthesis in purple carrot taproots. *Sci Rep.* 2017;7:45324. <https://doi.org/10.1038/srep45324>.
- Morohashi K, et al. A genome-wide regulatory framework identifies maize *pericarp color1* controlled genes. *Plant Cell.* 2012;24:2745–64. <https://doi.org/10.1105/tpc.112.098004>.
- Tian J, et al. Characteristics of dihydroflavonol 4-reductase gene promoters from different leaf colored *Malus crabapple* cultivars. *Horticulture Research.* 2017;4:17070. <https://doi.org/10.1038/hortres.2017.70>.
- Quattrocchio F, Wing JF, van der Woude K, Mol JN, Koes R. Analysis of bHLH and MYB domain proteins: species-specific regulatory differences are caused by divergent evolution of target anthocyanin genes. *Plant J.* 1998;13:475–88. <https://doi.org/10.1046/j.1365-3113x.1998.00046.x>.
- Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science.* 2003;299:396–9. <https://doi.org/10.1126/science.1078540>.
- Davies KM, Schwinn KE. In: Andersen ØM, Markham KR, editors. *In Flavonoids. Chemistry, Biochemistry, and Applications*. CRC Press/Taylor & Francis; 2006. p. 143–218.
- To KY, & Wang C. K. *Molecular breeding of flower color*. Vol. 1 300–310 (Global Science Books), (2006).

26. Allan AC, Hellens RP, Laing WA. MYB transcription factors that colour our fruit. *Trends Plant Sci.* 2008;13:99–102. <https://doi.org/10.1016/j.tplants.2007.11.012>.
27. Ramsay NA, Glover BJ. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 2005;10:63–70. <https://doi.org/10.1016/j.tplants.2004.12.011>.
28. Butelli E, et al. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol.* 2008;26:1301–8. <https://doi.org/10.1038/nbt.1506>.
29. Xie Y, Tan H, Ma Z, Huang J. DELLA proteins promote anthocyanin biosynthesis via sequestering MYBL2 and JAZ suppressors of the MYB/bHLH/WD40 complex in *Arabidopsis thaliana*. *Mol Plant.* 2016;9:711–21. <https://doi.org/10.1016/j.molp.2016.01.014>.
30. Zhou H, et al. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* 2015;82:105–21. <https://doi.org/10.1111/tpj.12792>.
31. Boase MR, et al. Failure to launch: the self-regulating *Md-MYB10_{RG}* gene from apple is active in flowers but not leaves of *Petunia*. *Plant Cell Rep.* 2015;34:1817–23. <https://doi.org/10.1007/s00299-015-1827-4>.
32. Hichri I, et al. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Botany.* 2011;62:2465–83. <https://doi.org/10.1093/jxb/erq442>.
33. Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell.* 2000;12:2383–94. <https://doi.org/10.1105/tpc.12.12.2383>.
34. Dubos C, et al. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 2010;15:573–81. <https://doi.org/10.1016/j.tplants.2010.06.005>.
35. Qiu J, et al. Arabidopsis AtPAP1 transcription factor induces anthocyanin production in transgenic *Taraxacum brevicorniculatum*. *Plant Cell Rep.* 2014;33:669–80. <https://doi.org/10.1007/s00299-014-1585-8>.
36. Huang YJ, et al. Differential activation of anthocyanin biosynthesis in *Arabidopsis* and tobacco over-expressing an R2R3 MYB from Chinese bayberry. *Plant Cell Tissue Organ Cult.* 2013;113:491–9.
37. Lai B, et al. *LcMYB1* is a key determinant of differential anthocyanin accumulation among genotypes, tissues, developmental phases and ABA and light stimuli in *Litchi chinensis*. *PLoS One.* 2014;9:e86293. <https://doi.org/10.1371/journal.pone.0086293>.
38. Lou Q, et al. Transcriptome sequencing and metabolite analysis reveals the role of delphinidin metabolism in flower colour in grape hyacinth. *J Exp Bot.* 2014;65:3157–64. <https://doi.org/10.1093/jxb/eru168>.
39. Matus JT. Transcriptomic and metabolomic networks in the grape berry illustrate that it takes more than flavonoids to fight against ultraviolet radiation. *Front Plant Sci.* 2016;7:1337. <https://doi.org/10.3389/fpls.2016.01337>.
40. Wang Z, Cui Y, Vainstein A, Chen S, Ma H. Regulation of fig (*Ficus carica* L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway. *Front Plant Sci.* 2017;8:1990. <https://doi.org/10.3389/fpls.2017.01990>.
41. Wang Z, et al. Transcriptome analysis reveals candidate genes related to color fading of 'Red Bartlett' (*Pyrus communis* L.). *Front Plant Sci.* 2017;8:455. <https://doi.org/10.3389/fpls.2017.00455>.
42. Li Y, et al. Combined analysis of the fruit metabolome and transcriptome reveals candidate genes involved in flavonoid biosynthesis in *Actinidia arguta*. *Int J Mol Sci.* 2018;19. <https://doi.org/10.3390/ijms19051471>.
43. Yuan H, et al. Time-course comparative metabolite profiling under osmotic stress in tolerant and sensitive Tibetan hullless barley. *Biomed Res Int.* 2018;2018:9415409. <https://doi.org/10.1155/2018/9415409>.
44. Sobel JM, Streisfeld MA. Flower color as a model system for studies of plant evo-devo. *Front Plant Sci.* 2013;4:321. <https://doi.org/10.3389/fpls.2013.00321>.
45. Wang X, Bennetzen JL. Current status and prospects for the study of Nicotiana genomics, genetics, and nicotine biosynthesis genes. *Mol Gen Genomics.* 2015;290:11–21. <https://doi.org/10.1007/s00438-015-0989-7>.
46. Jia HF, et al. Changes in flavonol content and transcript levels of genes in the flavonoid pathway in tobacco under phosphorus deficiency. *Plant Growth Regul.* 2015;76:225–31.
47. Andersen M, Jordheim M. *Anthocyanins*. Chichester: Wiley; 2010. p. 1–12.
48. Gould K, Davies K, Winefield C. *Anthocyanins: Biosynthesis, Functions, and Application*. Springer; 2009.
49. Deng J, et al. Proteomic and epigenetic analyses of Lotus (*Nelumbo nucifera*) petals between red and white cultivars. *Plant Cell Physiol.* 2015;56:1546–55. <https://doi.org/10.1093/pcp/pcv077>.
50. Luo P, et al. Disequilibrium of Flavonol Synthase and Dihydroflavonol-4-Reductase Expression Associated Tightly to White vs. Red Color Flower Formation in Plants. *Front Plant Sci.* 2015;6:1257. <https://doi.org/10.3389/fpls.2015.01257>.
51. Tian J, et al. The balance of expression of Dihydroflavonol 4-reductase and Flavonol synthase regulates flavonoid biosynthesis and red foliage coloration in crabapples. *Sci Rep.* 2015;5:12228. <https://doi.org/10.1038/srep12228>.
52. Chen W, et al. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant.* 2013;6:1769–80. <https://doi.org/10.1093/mp/sst080>.
53. Edwards KD, et al. A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics.* 2017;18:448. <https://doi.org/10.1186/s12864-017-3791-6>.
54. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12:357–60. <https://doi.org/10.1038/nmeth.3317>.
55. Liao Y, Smyth GK, Shi W. Feature counts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
56. Anders S, Huber W. Differential Expression of RNA-Seq Data at the gene level—the DESeq package: European Molecular Biology Laboratory (EMBL); 2012.
57. Li G, et al. De novo assembly and characterization of the spleen transcriptome of common carp (*Cyprinus carpio*) using Illumina paired-end sequencing. *Fish and Shellfish Immunology.* 2015;44:420–9. <https://doi.org/10.1016/j.fsi.2015.03.014>.
58. Dossa K, Mmadi MA, Zhou R, Zhang T, Su R, Zhang Y, Wang L, You J, Zhang X. Depicting the core transcriptome modulating multiple abiotic stresses responses in sesame (*Sesamum indicum* L.). *Int J Mol Sci.* 2019;20(16):3930.
59. Lalitha S. Primer premier 5. Biotech Software and Internet Report. 2000;1:270–2.
60. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{−(Delta Delta C(T))} method. *Methods.* 2001;25:402–8. <https://doi.org/10.1006/meth.2001.1262>.
61. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11:R14. <https://doi.org/10.1186/gb-2010-11-2-r14>.
62. Xie C, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39:W316–22. <https://doi.org/10.1093/nar/gkr483>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

