Tree Physiology 41, 1046–1064
doi:10.1093/treephys/tpaa152

Research paper

Metabolic profiling and gene expression analysis provides insights into flavonoid and anthocyanin metabolism in poplar

Yuru Tian[1,2], Qianqian Li[1,2], Shupei Rao[1,2], Aike Wang[3,4], Hechen Zhang[5], Liangsheng Wang[6], Yue Li[1,2] and Jinhuan Chen[1,2,7]

[1]Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Biological Sciences and Technology, Beijing Forestry University, No. 35 Qinghua East Road, Haidian District, Beijing 100083, China; [2]National Engineering Laboratory for Tree Breeding, Beijing Forestry University, No. 35 Qinghua East Road, Haidian District, Beijing 100083, China; [3]Yucheng Institute of Agricultural Sciences, Shangqiu, Henan 476000, China; [4]Shangqiu Zhongxing Seedling Planting Co., Ltd, Shangqiu, Henan 476000, China; [5]Henan Academy of Agricultural Sciences, Horticultural Research Institute, Zhengzhou, Henan 450002, China; [6]Key Laboratory of Plant Resources and Beijing Botanical Garden, Chinese Academy of Sciences, Institute of Botany, No.20 Nanxincun, Haidian District, Beijing 100093, China; [7]Corresponding author (chenjh@bjfu.edu.cn)

Received March 28, 2020; accepted October 29, 2020; handling Editor Jörg-Peter Schnitzler

Poplar, a woody perennial model, is a common and widespread tree genus. We cultivated two red leaf poplar varieties from bud mutation of *Populus sp.* Linn. '2025' (also known as Zhonglin 2025, L2025 for shot): *Populus deltoides* varieties with bright red leaves (LHY) and completely red leaves (QHY). After measuring total contents of flavonoid, anthocyanin, chlorophyll and carotenoid metabolites, a liquid chromatography-electrospray ionization-tandem mass spectrometry system was used for the relative quantification of widely targeted metabolites in leaves of three poplar varieties. A total of 210 flavonoid metabolites (89 flavones, 40 flavonols, 25 flavanones, 18 anthocyanins, 16 isoflavones, 7 dihydroflavonols, 7 chalcones, 5 proanthocyanidins and 3 other flavonoid metabolites) were identified. Compared with L2025, 48 and 8 flavonoids were more and less abundant, respectively, in LHY, whereas 51 and 9 flavonoids were more and less abundant in QHY, respectively. On the basis of a comprehensive analysis of the metabolic network, gene expression levels were analyzed by deep sequencing to screen for potential reference genes for the red leaves. Most phenylpropanoid biosynthesis pathway-involved genes were differentially expressed among the examined varieties. Gene expression analysis also revealed several potential anthocyanin biosynthesis regulators including three *MYB* genes. The study results provide new insights into poplar flavonoid metabolites and represent the theoretical basis for future studies on leaf coloration in this model tree species.

Keywords: anthocyanin, flavonoids, gene expression, leaf coloration, metabolic profiling, poplar.

Introduction

Flavonoids, which form a large family of polyphenolic secondary metabolites with low molecular weight, are widespread throughout the plant kingdom, ranging from mosses to angiosperms (Koes et al. 1994). Structurally, they comprise two aromatic rings (A and B) linked by a central three-carbon chain (C6–C3–C6). Flavonoids are classified as flavanones, dihydroflavonols, isoflavones, flavones, flavanols and anthocyanins depending on the modifications to the heterocycle, with three carbon atoms at the center. In addition to the coloration of flowers, fruits and leaves (e.g., red, blue and purple), which is aesthetically pleasing, flavonoids have diverse functions. For example, the floral pigmentation attracts pollinators and seed dispersers. Flavonoids protect plants from ultraviolet light and pathogenic microorganisms, and also function as signaling molecules during plant–microbe interactions (Dooner et al. 1991, Koes et al. 1994, Dixon and Paiva 1995).

Researchers have focused on flavonoids mainly because they are crucial for the color and flavor of fruits, vegetables and nuts, which form an integral part of the human diet (Parr and Bolwell 2000). The secondary metabolite composition greatly influences the quality and potential health benefits of food

products. Among the secondary metabolites, research has confirmed that flavonoids protect against oxidative stress, coronary heart disease, certain cancers and other age-related diseases (Chung et al. 2001, Ross and Kasum 2002). Accordingly, there has been increasing interest in flavonoids among researchers. Many structural and regulatory genes related to flavonoids have been isolated in various plant species, including *Zea mays* L., *Petunia hybrida* and *Arabidopsis*.

Anthocyanins, which are a type of flavonoid, are synthesized in the cytoplasm from phenylalanine via the phenylpropanoid and flavonoid biosynthetic pathways, after which they are transported to vacuoles (Dooner et al. 1991, Jeong et al. 2004). Anthocyanins are derived from six common anthocyanidins, namely cyanidin, delphinidin, pelargonidin, peonidin, petunidin and malvidin, which are differentiated according to the hydroxyl pattern or methoxy substitutions of the aromatic B ring. Delphinidin, peonidin and malvidin produce purple and dark colors, whereas the derivatives of cyanidin and pelargonidin are the main bright red pigments (Jaakola 2013, K.Cho et al. 2016). Besides the coloration of fruits, flowers, seeds and vegetables (e.g. pink, red, purple and blue), anthocyanins have other important roles, such as minimizing the damage due to cold, drought and UV irradiation (Bridle and Timberlake 1997), protecting plants from infections by viruses, bacteria and fungi (Lev-Yadun and Gould 2008) and promoting pollination and seed dispersal by attracting insects and animals (Hatier and Gould 2009). Anthocyanins have recently been designated as food materials with physiological functions because of their biological activities (e.g., antioxidant and antiaging activities), which may decrease the risk of cancers and diabetes (Hagiwara et al. 2002, Lila 2004). Anthocyanin biosynthesis has been extensively studied in many species, especially in plants with colorful flowers and food crops, such as sweet potato (*Ipomoea batatas*) (A.M.Wang et al. 2018), maize (*Z. mays*) (Hanson et al. 1996) and tomato (*Solanum lycopersicum*) (Kerckhoffs et al. 1997). In poplar, several anthocyanins have been identified by researchers, including cyanidin, cyanidin 3-*O*-glucoside and delphinidin 3-*O*-sambubioside, delphinidin 3-*O*-glucoside (Alcalde-Eon et al. 2016).

The genus *Populus* is a critical genetic resource that includes approximately 30 species, which are widely used for producing timber, pulp and paper, as well as for being a source of renewable energy (Taylor 2002). Poplar is a model system for characterizing tree-specific processes such as dormancy, secondary wood formation, and responses to biotic and abiotic stresses. *Populus deltoides* is a biologically and economically important species, which is often used in breeding new varieties, because of its rapid growth, good morphology and ease of vegetative propagation. A number of natural and artificially cultivated varieties of *P. deltoides*, which has horticultural traits including leaf color variation, are commercially available (Gao et al. 2009). In 1972, China introduced a number of modern clones from the Italian Poplar Research Institute, including *P. deltoides* cv. I-69/55 (poplar I-69), which is often used as the parent to generate fast-growing poplars (Wang et al. 1980). *Populus* sp. Linn. '2025' (known as Zhonglin 2025 in China, L2025 for short) is a variety bred from poplar I-69 as the female parent and *P. deltoides* as the male parent (Tian 2003). The genotypes we studied here are bud mutations of green leaves *Populus* sp. Linn. '2025' (wild-type), which has bright red leaves (described as LHY) and completely red leaves (described as QHY).

At present, most studies on flavonoids in poplars have focused on the expression of individual genes to regulate the metabolism of flavonoids (Ma et al. 2018, L.J.Wang et al. 2019), and a small body of research has explored the differences in flavonoids in different poplars (Alcalde-Eon et al. 2016). A series of poplar lines exhibiting intense anthocyanin-based pigmentation has been created, thereby enabling researchers to study flavonoids, including anthocyanins, in this model tree species. Key research questions to address include the following: What are the differences between these three varieties at the transcriptome and metabolome level? Which metabolic pathways and candidate genes contribute to leaf coloration in poplar? To answer these questions, the flavonoid and anthocyanin metabolites were compared among the three varieties, and the corresponding biosynthetic and regulatory genes were screened and analyzed by RNA-sequencing and a quantitative real-time polymerase chain reaction (qRT-PCR). Our results provide new insights into the flavonoid metabolism and represent the theoretical basis for future studies of metabolites biosynthesis and leaf coloration in this model tree species.

Material and methods

Plant materials

The poplar variety 'Zhonglin 2025' (L2025) is an officially released clone, which was provided by the Chinese Academy of Forestry (Beijing, China). QHY and LHY are bud-mutant varieties derived from L2025. For each variety, five seedlings were planted in pots (15 l capacity) filled with commercial growth medium comprising perlite, vermiculite and peat in the spring of 2017. The soil moisture content in the pots was maintained at 70% measured by a TRIME-PICO 64/32 TDR portable soil moisture meter (IMKO, Ettlingen, Germany). The growing conditions in the pots were consistent with the climatic conditions of Shangqiu (city). Shangqiu (city), Henan (province) is a region in China that experiences a warm temperate continental monsoon climate and the latitude and longitude is 115.65 and 34.44, respectively. The annual average temperature is 14 °C, with 726.5 mm annual precipitation and a 271-day frost-free period. The average daylight hours in the four seasons are 620.8, 599.8, 407.7 and 376.7 h, and the average temperatures are 16.1 °C, 26.8 °C, 15.3 °C and 3.6 °C, respectively.

In September 2018, a total of 12 seedlings including four seedlings from each variety with uniformly growth rate were chosen. From 9:00 a.m. to 9:30 a.m., the second to fifth of the fully expanded mature leaves from seedlings of the same variety were harvested, evenly mixed and immediately frozen with liquid nitrogen. The samples were then stored at −80 °C and used for content determination, metabolites detection, RNA-sequencing and qRT-PCR analysis.

Leaf color measurement

Leaf color was measured in accordance with the method recommended by the International Commission on Illumination. The NF333 spectrophotometer (Nippon Denshoku Industries Co. Ltd, Japan) was used to determine the $L^*a^*b^*$ values. The chroma [$C^* = (a^{*2} + b^{*2})^{1/2}$] and the chroma angle [$h = \arctan b^*/a^*$] were calculated based on the a^* and b^* values. Leaf color analysis was completed with five leaves per variety, and all measurements involved the central part of the leaves.

Flavonoid, anthocyanin, chlorophyll and carotenoid quantification

To quantify the total flavonoid content, samples were dried to a constant weight at 55–60 °C, crushed by an ultrasonicator (Biosafer 650-92, China) and filtered through a 40-mesh sieve before being weighed. Subsequently, 2 ml of 60% ethanol was added to approximately 0.02 g of sample, and the resulting mixtures were oscillated at 60 °C for 2 h. After centrifugation at 2000*g* for 10 min, 108 µl of supernatant was retained and transferred to a 2-ml centrifuge tube to be tested, taking 108 µl of 60% ethanol into another 2-ml centrifuge tube as a control. Thereafter, 6 µl of 0.05 g ml^{-1} $NaNO_2$ was added to the two centrifuge tubes separately, mixed well and left to stand for 6 min before 6 µl of 0.1 g ml^{-1} $AlCl_3$ was added. The mixture was allowed to stand for another 6 min and added with 80 µl of 1 mol l^{-1} NaOH. After standing for 15 min, the absorbance of the mixture was measured at 510 nm using a spectrophotometer Jasco UV-530 (Japan). The results obtained from the test solution were denoted as A1, and the results obtained from ethanol were denoted as A2. The flavonoid levels were calculated by using the following equation: amount (mg g^{-1} fresh weight) = (A1 − A2 − 0.0007)/0.02 (Chlopicka et al. 2012).

Anthocyanin was extracted according to a previously described pH differential method (Dong et al. 2019). The protocol was described as follows: frozen samples were crushed into a powder in liquid nitrogen. About 100 mg of powder was used and extracted separately with 2 ml of buffer A (100 mM KCl and 300 mM HCl, pH 1.0) and 2 ml of buffer B (200 mM sodium acetate and 120 mM HCl, pH 4.5). The mixtures were centrifuged at 10,000*g* for 15 min at 4 °C. The supernatants were collected for absorbance measurement at 510 nm using a spectrophotometer. The total anthocyanin levels were calculated by using the following equation: amount (mg g^{-1} fresh weight) = (A1 − A2) × 484.8 × 1000/24,825. A1 represents the absorbance at 510 nm of the supernatants collected from buffer A solution, while A2 represents that from buffer B solution. The value 484.8 represents the molecular mass of cyanidin-3-glucoside, while 24,825 reflects its molar absorptivity at 510 nm.

Chlorophyll and carotenoids were extracted from freeze-dried leaves with 95% (v/v) ethanol, after which the chlorophyll content was measured as previously described (Lichtenthaler 1987). The absorbance of the extracts was determined at 470, 645 and 663 nm using a spectrophotometer. The total flavonoid, anthocyanin, chlorophyll and carotenoid contents for each sample were determined with three biological replicates, each with three technical replicates.

Sample preparation and extraction

Freeze-dried leaf samples were crushed with the MM 400 mixer mill (Retsch) and zirconia beads for 1.5 min at 30 Hz. The crushed samples were weighed, after which 100 mg of powder was mixed with 1.0 ml of 70% (v/v) aqueous methanol for overnight extraction at 4 °C. After centrifuging the solutions at 10,000*g* for 10 min, the extracts were added to a CNWBOND Carbon-GCB SPE Cartridge (250 mg, 3 ml) and then filtered through the SCAA-104 membrane (0.22 µm pore size; ANPEL, Shanghai, China) for subsequent liquid chromatography-mass spectrometry (LC–MS) analysis.

High-performance liquid chromatography and tandem mass spectrometry

For each variety, three biological replicates were independently and randomly analyzed to minimize bias. Sample extracts were analyzed with an high-performance liquid chromatography (HPLC)-electrospray ionization (ESI)-tandem mass spectrometry (MS/MS) system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS/MS, Applied Biosystems 6500 Q TRAP), with the following conditions: HPLC: column, Waters ACQUITY UPLC HSS T3 C18 (1.8 µm, 2.1 mm × 100 mm); solvents, solvent A (water and 0.04% acetic acid) and solvent B (acetonitrile and 0.04% acetic acid); gradient program, 100:0 $V_{(A)}/V_{(B)}$ at 0 min, 5:95 $V_{(A)}/V_{(B)}$ at 11.0 min, 5:95 $V_{(A)}/V_{(B)}$ at 12.0 min, 95:5 $V_{(A)}/V_{(B)}$ at 12.1 min and 95:5 $V_{(A)}/V_{(B)}$ at 15.0 min; flow rate, 0.40 ml min^{-1}; temperature, 40 °C; and injection volume, 2 µl. The eluate was analyzed with an ESI-triple quadrupole (QQQ)-linear ion trap (LIT; Q TRAP) mass spectrometer. The LIT and QQQ scans were acquired with the API 6500 Q TRAP LC–MS/MS system equipped with an ESI Turbo Ion-Spray interface, operating in a positive ion mode and controlled by Analyst 1.6.3 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature, 500 °C; ion spray voltage, 5500 V; curtain gas, 25.0 p.s.i.; and the collision gas was high. In the QQQ, each ion pair was scanned for detection based on the

optimized decompression potential and collision energy (CE). To produce maximal signals, CE and declustering potential (DP) were optimized for individual MRM transitions. A specific set of MRM transitions was monitored for each period on the basis of the metabolites eluted within the period (see Figure S1 available as Supplementary data at *Tree Physiology* Online; Chen et al. 2013). Changes in flavonoids were identified and quantified using molecular formula-based mass accuracy and specific features of their MS^2 spectra.

Qualitative and quantitative analysis of metabolites

Metabolites detected by the LC-ESI-MS/MS system were identified based on a search of accurate masses of significant peak features against the online MWDB (MetWare database from MetWare Biotechnology Co., Ltd, Wuhan) (Chen et al. 2013, Zhu et al. 2018, Shen et al. 2019). The MWDB was based on the MS^2 spectral tag (MS2T) library, which was constructed by Metware Biotechnology Co., Ltd Metabolites in the MS2T library were annotated by matching the fragmentation pattern (delivered by ESI-Q TRAP-MS/MS), combined with the retention time and the accurate *m/z* value (delivered by ESI-QqTOF-MS/MS). Isotope signals, repeated signals of K^+, Na^+ and NH_4^+, and repeated signals of fragment ions of substances with a relatively high molecular weight were removed in the analysis (Chen et al. 2013). To produce maximal signals, CE and DP were optimized for each precursor–product ion (Q1–Q3) transition. In this mode, the quadrupole first screened the precursor ions of the target substance and eliminated the ions of other molecular weight substances to prevent preliminary interference. The precursor ions were ionized in the collision chamber and then fragmented. The fragment ions were filtered through the QQQ to obtain a fragment ion with the required characteristic and eliminate the interference of non-target ions. On the basis of the retention time and peak shape information of the metabolites, the mass spectrum peaks detected in different samples of each metabolite were manually corrected to ensure the quantitative accuracy. The peak area of each chromatographic peak represents the phase content of the corresponding metabolites, and the quantitative analysis integration results of all samples were obtained (Fraga et al. 2010). Thus, our data can be used for relative comparisons across L2025, LHY and QHY.

Sample quality control analysis

Quality control (QC) samples (mix) were prepared from a mixture of sample extracts and used to analyze the reproducibility of samples under the same processing methods. The quality control samples were injected after every 10 experimental samples throughout the analytical run to assess the repeatability and reliability of the data. The repeatability of metabolite extraction and detection (technical duplication) was assessed by overlapping the total ions chromatography (TIC) of mass spectrometry analysis of different QC samples. The stability of the instrument provides an important guarantee for the repeatability and reliability of the data. Only the RNAs with integrity number (RIN) >7.0, 260/230 ratio between 2.1 and 2.5 and 260/280 ratio between 1.8 and 2.0, were chosen for subsequent sequencing.

Transcriptome assembly and analysis

Total RNA was isolated and purified with the CTAB method (Chang et al. 1993). The integrity, purity and concentration of purified RNA were assessed with the Agilent 2100 Bioanalyzer and the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). The mRNA extracted from the total RNA in the samples was isolated using Oligo dT. Libraries were generated and purified using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs Inc., Ipswich, MA, USA) and AMPure XP Beads (Beckman Coulter, Inc., Indianapolis, IN, USA), using the fragmented mRNA as the template, following the manufacturer's recommendations. The concentration, integrity and quantification of the library were determined by using a Qubit™ Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA, USA) and a Qsep100 DNA Analyzer (KAPA Biosystems), respectively. The denatured libraries were subjected to high-throughput parallel sequencing of both ends of the library using an Illumina HiSeq X™ Ten System sequencing platform. The quality of the raw data was evaluated using FasQC v0.10.1 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) with default settings. The clean data were separated using Cutadapt v1.9 (http://cutadapt.readthedocs.org/), and the quality threshold was set to Q30, which removed the sequencing adapters and the primer sequence from the raw data to filter out low-quality data. In this study, de novo transcriptome assembly was performed according to Grabherr (Grabherr et al. 2011).

The transcript level was quantified using RSEM software, and the length of the transcript in the sample was normalized to fragments per kilobase of exon per million reads mapped (FPKM) values (Peng et al. 2019). The false discovery rate was used to adjust the *P*-values of differentially expressed genes (DEGs). Genes with an expression-level change of $\log_2 > 2$ and an adjusted $P < 0.05$ were considered DEGs and further annotated based on Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The enrichment of specific KEGG pathways among the DEGs was assessed with Fisher's exact test.

Integrative analysis of metabolome and transcriptome

Pearson correlation coefficients were calculated to integrate the metabolome and transcriptome data. Specifically, we calculated the mean metabolome and transcriptome data for all biological replicates of each variety. The fold changes in each genotype were then calculated for the metabolome and transcriptome data

and compared with the corresponding data for the control poplar variety. Finally, the correlation coefficients were calculated based on the $\log_2$ fold change of each metabolite and transcript. Correlations with a coefficient $R > 0.9$ were selected. Metabolome and transcriptome relationships were visualized with Cytoscape (version 2.8.2).

Validation of DEGs through quantitative real-time PCR

Total RNA extracted from the leaves of three poplar varieties was used for the reverse transcription with the FastQuant RT Kit with DNase (TianGen Biotech Co., Ltd, China) to synthesize the first-strand cDNA. A qRT-PCR assay was performed with an optical 96-well reaction plate, the ABI PRISM 7500 Real-time PCR system (Applied Biosystems) and SuperReal PreMix Plus SYBR Green (TianGen Biotech Co., Ltd). Each reaction contained 12.5 µl of SYBR Premix ExTaq, 0.5 µl of ROX Reference Dye, 2.0 µl of cDNA and 1.0 µl of gene-specific primers in a final volume of 25 µl. The PCR program was as follows: 95 °C for 10 s and then 45 cycles at 95 °C for 5 s and 60 °C for 40 s. The primers were listed in Table S1 available as Supplementary data at *Tree Physiology* Online. The qRT-PCR data were analyzed according to the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Statistical analysis

The metabolome and transcriptome analyses were completed with three biological replicates. Hierarchical clustering analysis (HCA), principal component analysis (PCA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were completed with R (http://www.r-project.org/). The criteria of fold change ≥ 2 or ≤ 0.5 and a variable importance in projection (VIP) score >1 was used to identify differentially accumulated metabolites among the analyzed poplar varieties (A.M.Wang et al. 2018). The content of all differentially accumulated flavonoid/anthocyanins or gene expression was analyzed by hierarchical cluster analysis with MultiExperiment Viewer (MEV) 4.9.0 software. Computerized algorithms were used by maximum difference normalization to homogenize the metabolites or transcript expression levels in MeV (Howe et al. 2011). The metabolites with the same or similar expression were clustered. Data were analyzed with the SPSS 23.0 program (IBM Corporation, Armonk, NY, USA). Values were compared with the one-way ANOVA and Duncan's multiple range tests to determine the significance of any differences ($P < 0.05$). The figures presented herein were drawn with the OriginPro 2016 program (OriginLab Corporation, Northampton, MA, USA) and Adobe Illustrator CC.

Results

Color characteristics of three poplar varieties

The QHY and LHY bud mutant varieties originated from the L2025 poplar variety, which was cultivated by Zhongxing Ltd

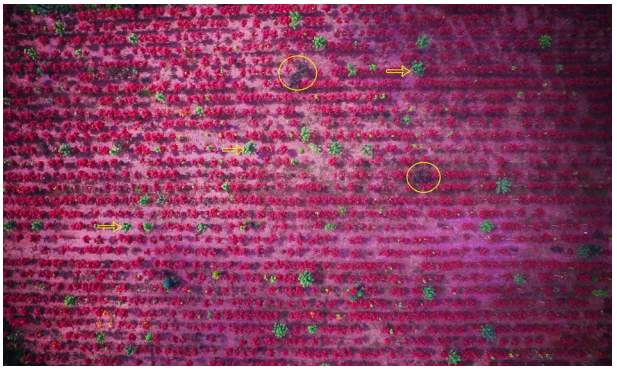

Figure 1. L2025 forest with QHY (cycle indicated) and L2025 (arrowhead indicated) sporadically distributed. The forest landscape was photographed at 16:00h at Shangqiu City, Henan Province.

From April (when the buds germinated) to late September, all QHY leaves were purple–red, after which the leaves turned bright red. By contrast, the LHY leaves were bright red from April to September. The LHY landscape forest was photographed at 16:00h on July 28 at Shangqiu, Henan. The L2025 forest, in which QHY trees were sporadically distributed, was also photographed (Figure 1). Considerable differences in leaf color were obvious among the three poplar varieties (Figure 2A). Hue angles determined with the NF333 spectrophotometer indicated the three poplar varieties could be divided into three clusters. Relatively similar hue values were observed for QHY and LHY. The intensity of red hues was greater for the QHY and LHY leaves than for the L2025 leaves. The LHY leaves were intensely red, followed by the QHY leaves (Figure 2B). The L^* value, which indicates brightness, was highest for L2025 leaves (Figure 2C).

Total flavonoid and anthocyanin contents increased in red poplar varieties

The abundance and diversity of pigments primarily determine plant colors. Thus, the flavonoid, anthocyanin and chlorophyll contents were quantified and compared among the three poplar varieties. The flavonoid content of the control (i.e., green) leaves was only 58 mg g^{-1} dry weight (DW). Comparatively, the flavonoid contents of the LHY and QHY leaves were 105 and 165 mg g^{-1} DW, respectively. The total flavonoid content of the red poplar leaves was 1.6- to 3-fold higher (Figure 3A). Consistent with the intense pigmentation observed in these leaves, anthocyanin accumulated substantially more in the QHY and LHY leaves than in the L2025 leaves (Figure 3B). Compared with the flavonoid content of the QHY and LHY leaves, which was two times higher than that of L2025 leaves, the anthocyanin level in the QHY and LHY leaves was about 10 times higher than that of the L2025 leaves. These results suggested that the reddish leaf color is closely related to anthocyanin accumulation. Interestingly, although the leaf colors varied, the anthocyanin content of the LHY and QHY leaves did not differ. The chlorophyll and carotenoid contents were lower in the LHY and QHY leaves

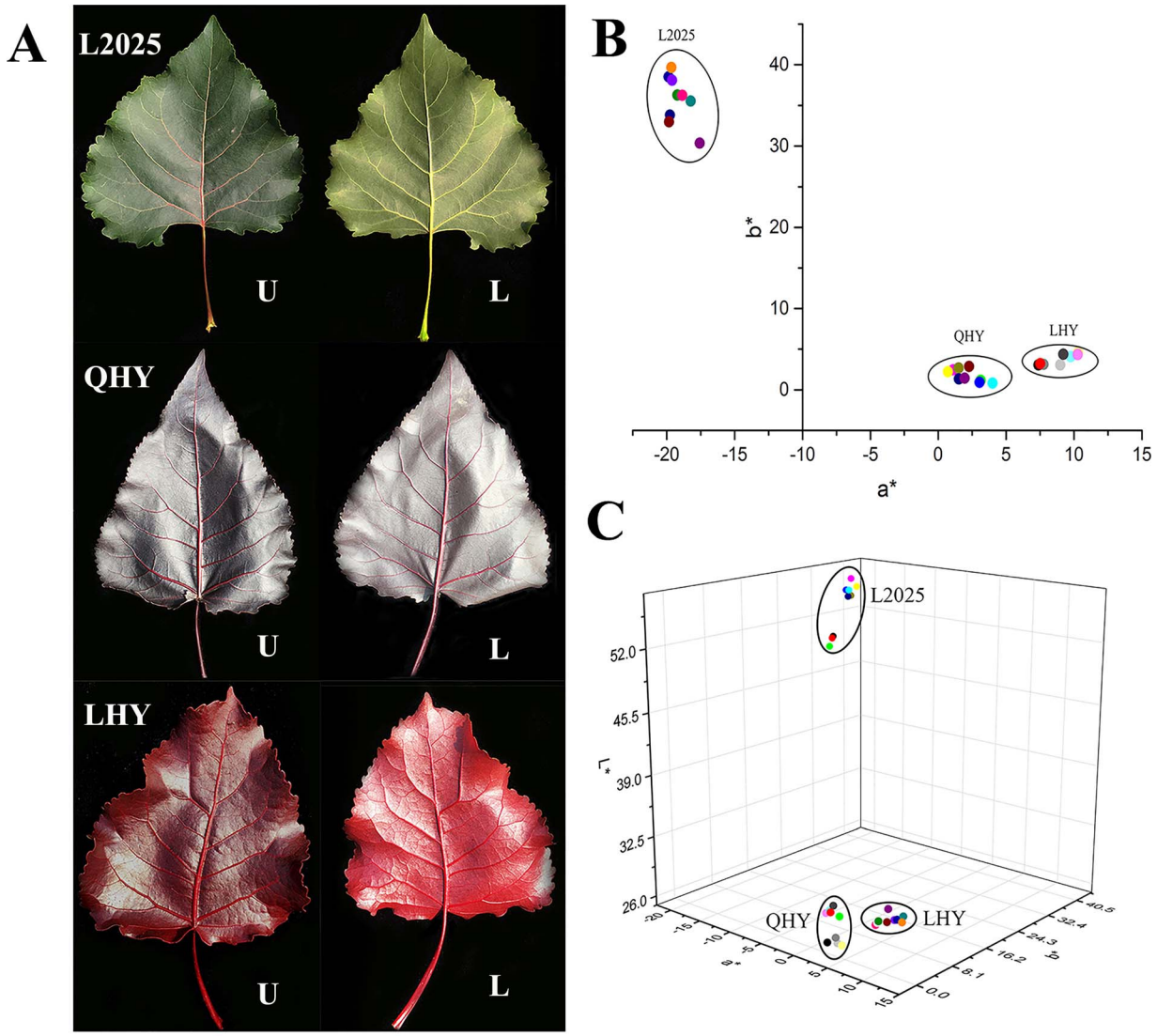

Figure 2. Morphology of green and red poplar leaves. (A) Digital image of the expanded leaves from three poplar varieties, namely, L2025, QHY and LHY. (B and C) Chroma meter measurement. (B) and (C) represent two- and three-dimensional graphics, respectively. The value of a^* represents green–red axis, b^* represents the yellow–blue axis and L^* represents brightness. Nine biological replicates were included for each variety.

than in the L2025 leaves (Figure 3C). The chlorophyll content of LHY leaves was >70% lower than that of L2025 leaves. However, the phenomenon of chlorophyll reduction was not observed for QHY leaves. The carotenoid contents of QHY and LHY leaves were 32.1 and 46.4% lower than that of L2025 leaves, respectively (Figure 3D).

Transcriptomic analysis of three poplars

To evaluate the gene expression profile in three poplars, nine cDNA libraries were constructed and sequenced. A total of 393.848 M raw reads were produced. After filtering these raw reads, we obtained a total of 364.541 M clean reads (see Table S2 available as Supplementary data at *Tree Physiology* Online). Clean reads were assembled with de novo, and transcripts with 42.09% GC and unigenes with 41.89% GC were obtained, respectively (Table 1 and see Figures S2 and S3 available as Supplementary data at *Tree Physiology* Online). To annotate the transcriptome with putative functions, we searched the assembled unigenes against public databases (see Figures S4 and S5 available as Supplementary data at *Tree Physiology* Online).

Metabolic profiling of flavonoids in poplar

To ensure that the observed differences in the flavonoid contents among the three analyzed varieties were accurate, we used a newly developed HPLC-MS/MS platform for a widely targeted metabolomics examination (Chen et al. 2013, Peng et al. 2017, Zhu et al. 2018). On the basis of databases developed by MetWare Co. and by a multivariate statistical analysis, we

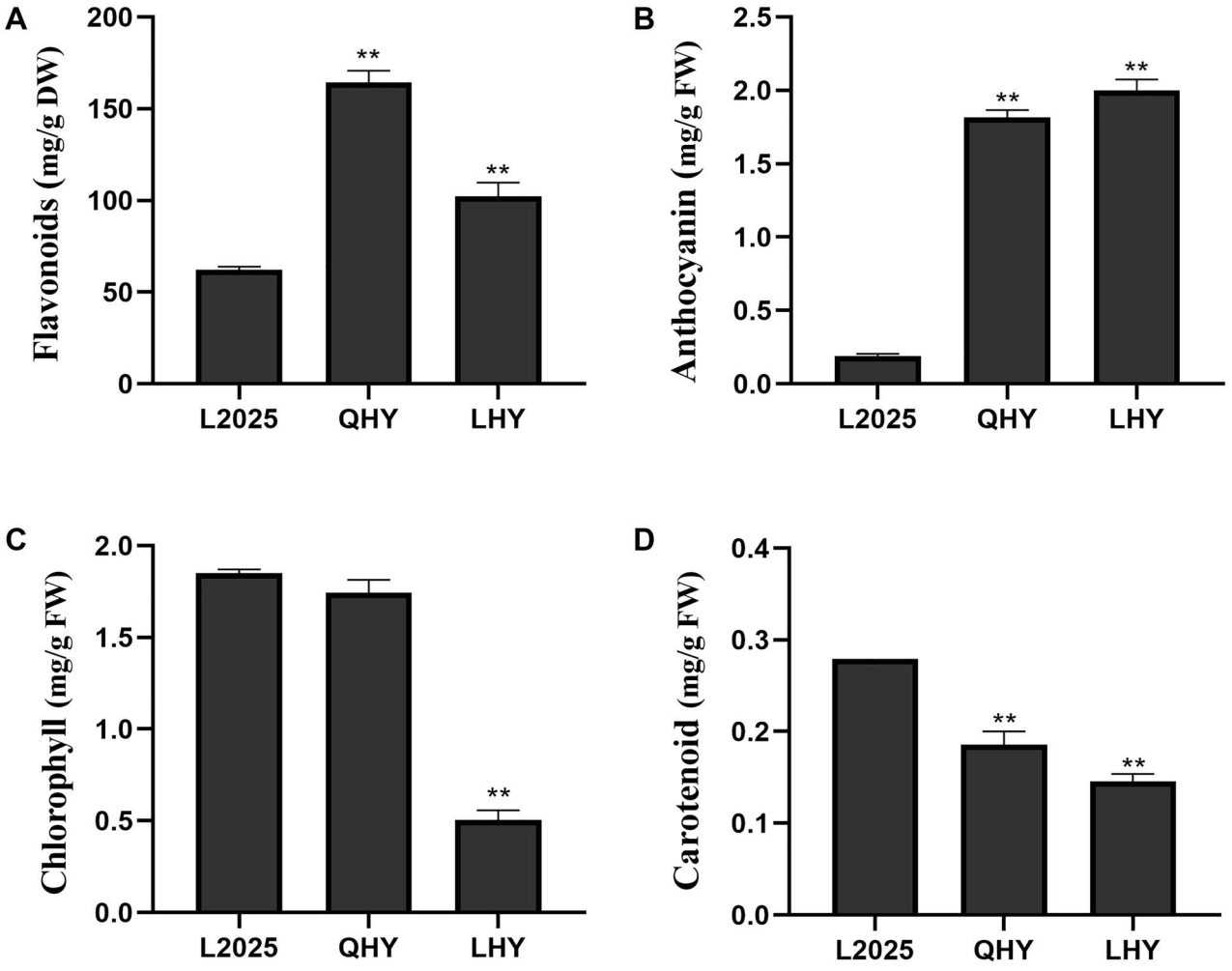

Figure 3. Contents of flavonoids, anthocyanin, chlorophyll and carotenoid in different poplars. (A) Total flavonoids content in the leaves of three poplar varieties. (B) Total anthocyanin content in the leaves of three poplar varieties. (C) Total chlorophyll content in the leaves of three poplar varieties. (D) Carotenoid content in the leaves of three poplar varieties. Error bars indicate standard errors (SEs). Asterisks above the column indicate significant differences of correlation at 0.01 level ($P < 0.01$).

Table 1. Statistical results of de novo transcriptome assembly.

	All (≥200 bp)	≥500 bp	≥1000 bp	N50	GC (%)	Total length	Max length	Min length	Average length
Transcript	71,959	39,911	22,998	1405	42.09	65,117,675	11,073	224	904
Unigene	53,016	23,729	11,794	1115	41.89	39,223,846	11,073	224	739

detected 210 flavonoid metabolites (see Table S3 available as Supplementary data at *Tree Physiology* Online), including 89 flavones, 40 flavonols, 25 flavanones, 18 anthocyanins, 16 isoflavones, 7 flavanonols, 7 chalcones and 5 proanthocyanidins. Among these flavonoids, the following 15 were present in the red poplar varieties (QHY and LHY), but not in L2025: cyanidin 3-*O*-glucoside, luteolin *O*-sinapoylhexoside, malvidin 3-*O*-galactoside, malvidin 3-*O*-glucoside, phloretin, delphinidin 3-*O*-glucoside, cyanidin 3-*O*-rutinoside, cyanidin 3,5-*O*-diglucoside, pelargonin, butein, quercetin 7-*O*-β-D-glucuronide, rotenone, 7,4′-dihydroxyflavone, delphinidin 3,5-*O*-glucoside and diosmin (Table 2). Only tricin 4′-*O*-syringic acid was present in L2025, but not in LHY and QHY. A total of 108 flavonoid metabolites could be assigned to a known flavonoid biosynthesis pathway (see Figure S6 available as Supplementary data at *Tree Physiology* Online), including 24 flavonols, 40 flavones, 20 flavanones, 13 anthocyanins and 11 isoflavones.

Four common anthocyanins comprising glycosylated cyanidin, delphinidin, peonidin or pelargonidin were identified. Additionally, we detected five kinds of anthocyanidins (peonidin, pelargonidin, cyanidin, malvidin and delphinidin) and 14 kinds of anthocyanins (peonidin *O*-hexoside, peonidin 3-*O*-glucoside,

Table 2. A metabolites specifically accumulated in QHY and LHY and L2025.

Compounds	Class	L2025	QHY	LHY
Cyanidin 3-*O*-glucoside (Kuromanin)	Anthocyanins	×	◯	◯
Luteolin *O*-sinapoylhexoside	Flavone	×	◯	◯
Malvidin 3-*O*-galactoside	Anthocyanins	×	◯	◯
Malvidin 3-*O*-glucoside (Oenin)	Anthocyanins	×	◯	◯
Phloretin	Flavanone	×	◯	◯
Delphinidin 3-*O*-glucoside (Mirtillin)	Anthocyanins	×	◯	◯
Pelargonin	Anthocyanins	×	◯	◯
Quercetin 7-*O*-β-D-Glucuronide	Flavonol	×	◯	◯
Rotenone	Isoflavone	×	◯	◯
7,4′-Dihydroxyflavone	Flavone	×	◯	◯
Delphin chloride	Anthocyanins	×	◯	◯
Diosmin	Flavonoid	×	◯	◯
Tricin 4′-*O*-syringic acid	Flavone	◯	×	×
Chrysoeriol *O*-homovanillic acid	Flavone	◯	◯	×
Tricin *O*-glucuronic acid	Flavone	◯	×	◯
Procyanidin A1	Proanthocyanidins	◯	◯	×
7-*O*-Methyleriodictyol	Flavanone	◯	◯	×
Glabridin	Flavonoid	◯	◯	×
Chrysoeriol *O*-glucuronic acid-*O*-hexoside	Flavone	◯	×	◯
Hesperetin *C*-hexosyl-*O*-hexosyl-*O*-hexoside	Flavone	◯	◯	×

'×' means the metabolites were not detected; '◯' means the metabolites were detected.

peonidin 3,5-diglucoside, cyanidin 3-*O*-glucoside, cyanidin *O*-syringic acid, cyanidin *O*-diacetyl-hexoside-*O*-glyceric acid, cyanidin 3-*O*-rutinoside, cyanidin 3,5-*O*-diglucoside, pelargonidin 3,5-*O*-glucoside, pelargonidin 3-*O*-glucoside, malvidin 3-*O*-galactoside, malvidin 3-*O*-glucoside, delphinidin 3-*O*-glucoside and delphinidin 3,5-*O*-glucoside).

We detected 15 kinds of *C*-glycosyl flavonoids, including luteolin glycosides, hesperetin glycosides and apigenin glycosides. The 15 *C*-glycosyl flavonoids were classified as 12 *C*-hexosides and 3 *C*-glucosides (luteolin 6-*C*-glucoside, apigenin *C*-glucoside and apigenin 6,8-*C*-diglucoside; Table 3). Most of the *C*-glycosyl flavonoids kurtosis values were higher for QHY and LHY than for L2025, implying that the red poplar leaves were rich in stable flavonoids.

In the current study, the flavonoids were identified using molecular formula-based mass accuracy and specific features of their MS^2 spectra. Here, we selected cyanidin 3-*O*-glucoside and apigenin *C*-glucoside from the identified substances above and attached their MS^2 spectra in Figure S7A and B available as Supplementary data at *Tree Physiology* Online, respectively. In addition, 18 polyphenol metabolites were detected, which are not further discussed here (see Table S4 available as Supplementary data at *Tree Physiology* Online).

Overview of the differences in the flavonoid metabolite profiles

The HCA and PCA were performed to clarify the differences in the metabolite profiles among the three analyzed varieties. The HCA resulted in seven clusters based on the relative differences in the accumulation in the different varieties (Figure 4A and see Table S5 available as Supplementary data at *Tree Physiology* Online). The results show that the curves of TIC during metabolite detection overlapped, that is, the retention time and peak intensity were consistent, indicating that the mass spectrometry had better signal stability when the same sample was detected at different times (Figure 4B). The flavonoids in clusters 4 and 5 were obviously more abundant in LHY and QHY than in L2025, whereas the flavonoid levels in cluster 6 were higher in L2025 and LHY than in QHY. Thus, the red poplar varieties were clearly distinguished from L2025, but were similar to each other regarding the flavonoid clusters.

We carried out PCA analysis on the 210 metabolites. The PCA clearly separated the two cultivars and quality control (QC) samples (mix), suggesting that the observed differences in the flavonoid profiles were correlated with leaf color (Figure 4C). In addition, the correlation coefficients between samples were >0.5, and the correlation coefficient for the same variety was >0.9, indicating that the results were reliable and repeatable (Figure 4D).

In this study, the OPLS-DA model involved pair-wise comparisons of the flavonoid metabolite contents of the samples to evaluate the differences between L2025 and QHY (R^2X = 0.887, R^2Y = 1 and Q^2 = 0.994; Figure 4E), between L2025 and LHY (R^2X = 0.922, R^2Y = 1 and Q^2 = 0.999; Figure 4F) and between LHY and QHY (R^2X = 0.769, R^2Y = 0.999 and Q^2 = 0.957; Figure 4G). The Q^2 values of all comparison groups exceeded 0.9, demonstrating that these models were stable and reliable and the differences in the flavonoid metabolites could be screened further.

Table 3. The *C*-glycosyl flavonoids in three poplar varieties.

C-glycosyl flavonoids	Fold change		*P*-values	
	LHY/L2025	QHY/L2025	$P_{LHY/L2025}$	$P_{QHY/L2025}$
Naringenin *C*-hexoside	3.15*	3.10*	0.00	0.00
Luteolin *C*-hexoside	4.39*	2.82*	0.00	0.00
Luteolin 8-*C*-hexosyl-*O*-hexoside	6.84*	7.59*	0.00	0.00
Luteolin 6-*C*-glucoside	3.64*	2.48*	0.00	0.01
Hesperetin *C*-hexosyl-*O*-hexosyl-*O*-hexoside	0.00*	0.62	0.00	0.10
Eriodictyol *C*-hexoside	1.00	1.18	1.00	0.84
Eriodictiol *C*-hexosyl-*O*-hexoside	2.53*	2.88*	0.02	0.00
Eriodictiol 6-*C*-hexoside 8-*C*-hexoside-*O*-hexoside	4.61*	5.86*	0.00	0.01
Chrysoeriol *C*-hexosyl-*O*-rhamnoside	2.92	2.95*	0.05	0.01
Chrysoeriol 8-*C*-hexoside	0.67	0.49*	0.10	0.04
Chrysin *C*-hexoside	0.75	1.19	0.53	0.69
Apigenin *C*-glucoside	3.98*	3.17*	0.00	0.00
Acacetin *C*-hexoside	0.38	1.19	0.07	0.72
'Apigenin 6,8-*C*-diglucoside'	0.83	0.76	0.09	0.06
Naringenin *C*-hexoside	3.15*	3.10*	0.00	0.00

*A statistical difference at $P < 0.05$ level between the mean of the two samples.

Differential metabolite abundance among three poplar varieties

To clarify the metabolic differences among QHY, LHY and L2025, we screened for differentially abundant metabolites among the 210 annotated flavonoids according to their fold changes and the VIP scores. The screening results were illustrated with volcano plots (Figure 5A–C) and Venn diagrams (Figure 5D). A total of 79 differentially abundant flavonoids were detected among the three poplar varieties, of which 60 were differentially accumulated (51 increased and 9 decreased) between QHY and L2025 (Figure 5A). Additionally, 56 flavonoids were differentially accumulated (48 increased and 8 decreased) between LHY and L2025 (Figure 5B). Moreover, 47 differentially abundant flavonoids were common between QHY vs L2025 and LHY vs L2025 comparisons. These flavonoids may be important metabolites influencing poplar leaf colors. Moreover, the 29 flavonoids detected as differentially accumulated between QHY and LHY (Figure 5C and D) were considered to be responsible for the variability in the red coloration of these two mutant varieties.

To reveal the flavonoid metabolite accumulation patterns in the control and red mutant poplar varieties, pair-wise comparisons were completed between LHY/QHY and L2025. The HCA of the flavonoids that were differentially accumulated among the three poplar varieties revealed only two clusters for QHY vs L2025 (Figure 6A) and LHY vs L2025 (Figure 6B). Cluster 1 represented metabolites that were more abundant in L2025 than in the red poplar mutants (Figure 6A and B). Cluster 2 included the metabolites that were less abundant in L2025 than in the red poplar varieties. The cluster 2 metabolites may be crucial for determining the pigmentation and coloration of the red poplar varieties. However, the clusters for the 29 differentially accumulated flavonoids between the two mutant poplar varieties were complex (Figure 6C), implying that distinguishing between LHY and QHY may be difficult.

Differences in the metabolic pathways and anthocyanin derivatives between green and red poplars

The KEGG classification results indicated that the differentially abundant flavonoid metabolites between L2025 and QHY were mainly related to isoflavone biosynthesis, flavone and flavonol biosynthesis, anthocyanin biosynthesis and others (Figure 7A and B). The *P*-value of the anthocyanin biosynthesis pathway was the smallest one among the four pathways. Similar results of classification were obtained in comparison between L2025 and LHY (Figure 7C and D). These pathways, especially the anthocyanin biosynthesis pathway, may be responsible for the observed variability in leaf coloration in red poplars.

All anthocyanins accumulated more in QHY and LHY than in L2025. Six kinds of anthocyanidins were exclusive to the red poplar varieties (Table 4). A phylogenetic tree constructed according to the anthocyanin accumulations revealed that almost all anthocyanins had a similar accumulation pattern in the three varieties (Figure 8). Nearly all of the anthocyanins were considerably more abundant in QHY and LHY than in L2025, especially peonidin 3,5-di-*O*-glucoside and pelargonidin-based anthocyanins. Cyanidin 3-*O*-glucoside and delphinidin 3,5-*O*-glucoside were exclusive to LHY and QHY (Figure 8 and Table 4). Another predominant anthocyanin in QHY and LHY was pelargonidin 3,5-*O*-glucoside, which likely contributes to the red coloration. Our results indicated that the anthocyanin composition differed between green and red

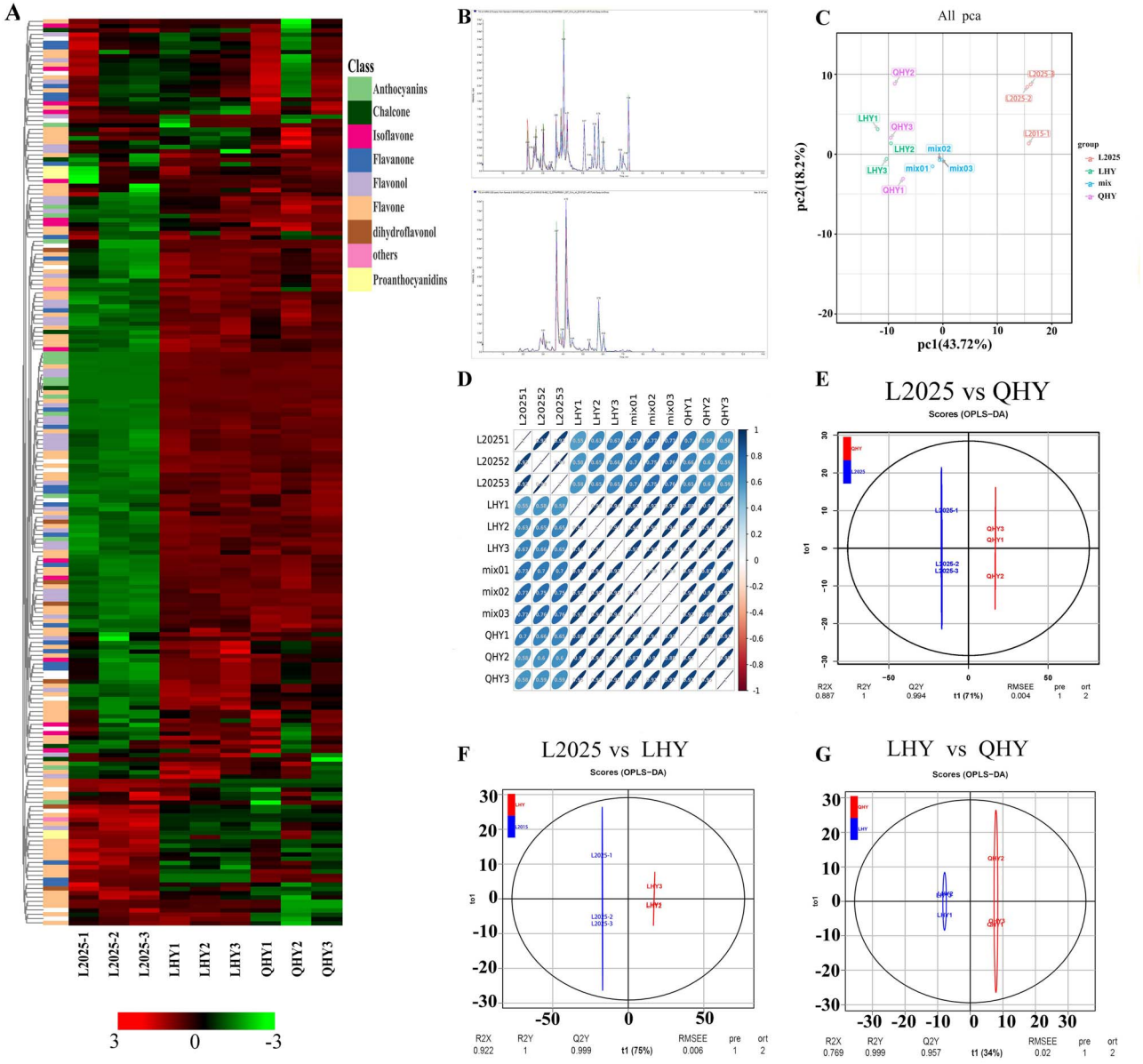

Figure 4. Heat map visualization, TIC, correlation coefficient, PCA and PLS-DA of the relative differences in flavonoids in different poplar varieties. (A) Clustering heat map of all flavonoid metabolites. The high and low abundances of metabolites were indicated by red and green, respectively. (B) Overlay of QC sample mass detection total ions chromatography plot (TIC plot). (C) Score plots for PCA showed high cohesion within groups and good separation among three poplar varieties. The mix sample indicated quality control. (D) Correlation coefficient graph of L2025, LHY, QHY and mix. (E-G) PLS-DA model plots and loading plots for L2025, QHY and LHY.

poplars. Additionally, the anthocyanin accumulation in QHY and LHY was responsible for the observed development of red leaves. Notably, there were no obvious differences in the anthocyanin types and contents between QHY and LHY.

Flavonoid and anthocyanin pathway genes are highly expressed in red poplars

To identify the DEGs related to leaf coloration, we performed an RNA-sequencing analysis of the same leaf samples used for metabolic profiling. The expression levels of 7274 and 4004 genes were up-regulated in QHY and LHY relative to the L2025 expression levels. Additionally, the expression levels of 5256 and 6191 genes were down-regulated in QHY and LHY compared with the L2025 expression levels (see Figure S8 available as Supplementary data at *Tree Physiology* Online). We further examined the key candidate genes involved in the phenylpropanoid biosynthesis pathway, which were responsible for flavonoid and anthocyanin biosynthesis. An analysis of the

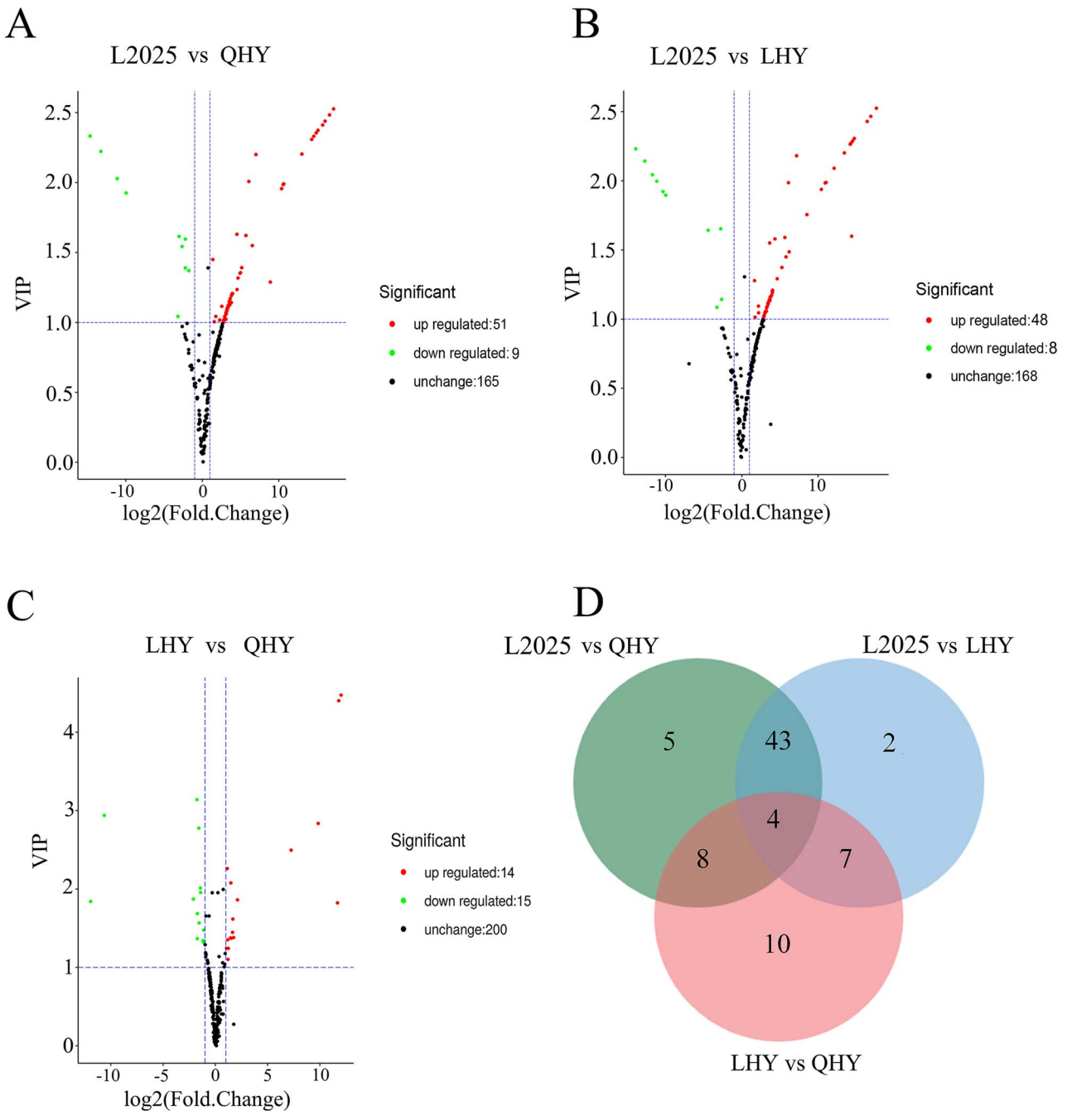

Figure 5. Metabolites from QHY and LHY compared with L2025. (A–C) The volcano plot shows the differential metabolite expression levels between (A) L2025 and QHY, (B) L2025 and LHY and (C) QHY and LHY. Green and red dots represent down-regulated and up-regulated differentially accumulated metabolites, respectively. Black dots represented non-differentially expressed metabolites. (D) Venn diagram shows the overlapping and cultivar-specific differential metabolites from QHY, LHY and L2025.

transcriptomic data uncovered six *PAL* genes and six *C4H* genes, of which five *PAL* genes and five *C4H* genes exhibited up-regulated expression in the red poplar varieties. Two chalcone synthase (*CHS*) genes, one chalcone isomerase (*CHI*) gene, two dihydroflavonol reductase (*DFR*) genes, one anthocyanidin synthase (*ANS*) gene, and two *LAR* genes were identified in the sequencing libraries, and all were more highly expressed in the red poplar varieties than in the control poplar variety. By contrast, the expression levels of two *4CL* genes were markedly down-regulated in the red varieties. Furthermore, complex expression patterns were observed for the genes encoding flavanone 3′-hydroxylase (F3′H) and flavonol synthase (FLS) in the three analyzed poplar varieties. Regarding the five *F3′H* genes detected in the libraries, three and two exhibited down- and up-regulated expression, respectively, in the red poplar varieties compared with the green poplar variety. Of the five

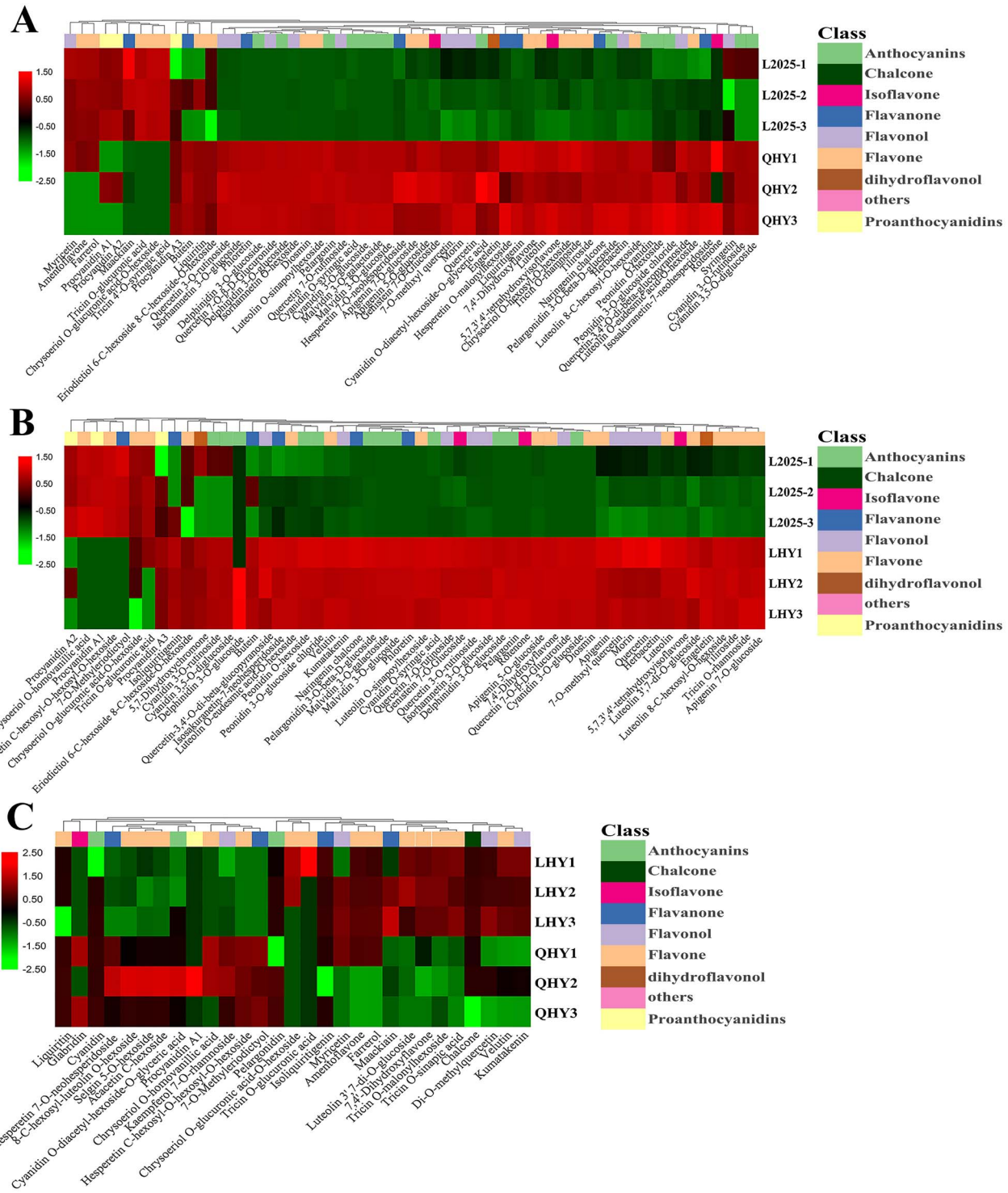

Figure 6. Metabolites with significant difference in three poplar varieties. Significantly different metabolites of green and red poplars were assessed by variable important in projection (VIP) and fold change. Each different colored rectangle on the bottom side of the heatmap represents the kinds of metabolite. The high and low abundances of metabolites were indicated by red and green, respectively. (A) The difference metabolites between L2025 and QHY. (B) The difference metabolites between L2025 and LHY. (C) The difference metabolites between LHY and QHY.

FLS genes identified in the libraries, two numbers exhibited up-regulated expression, two showed down-regulated expression in the red poplar varieties relative to the control variety, and the fifth gene only exhibited up-regulated expression in QHY. The qPCR results were inconsistent with the RNA-seq results (see Figure S9 available as Supplementary data at *Tree Physiology* Online), which also indicated that most of the genes involved in the phenylpropanoid biosynthesis pathway were up-regulated in

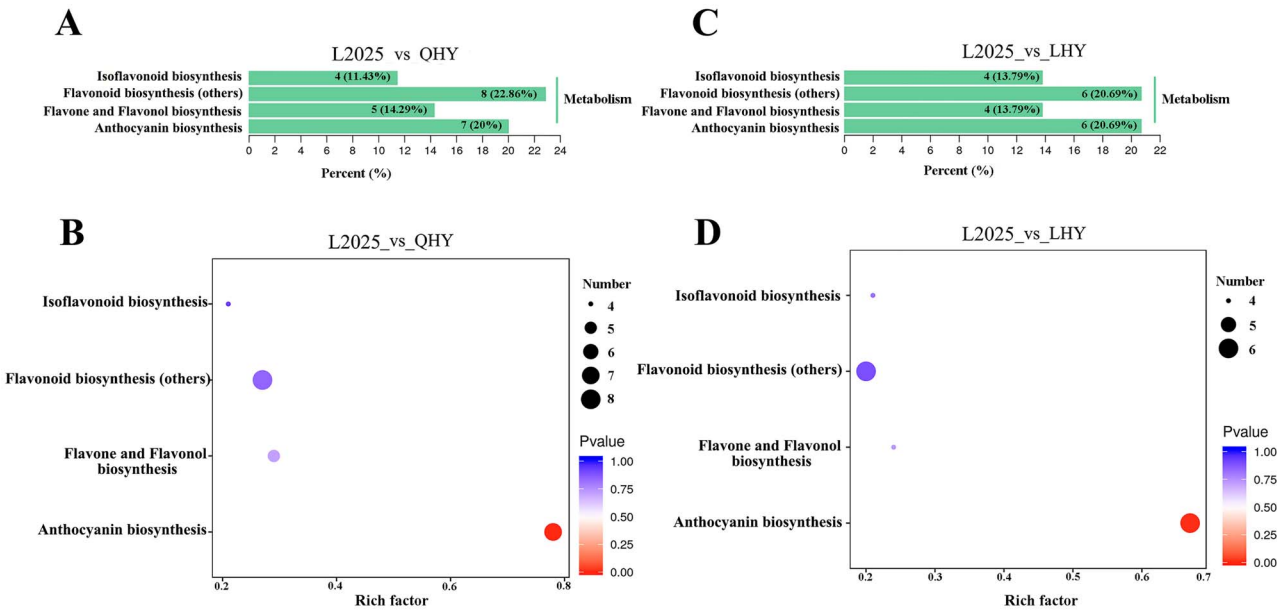

Figure 7. The KEGG classification of differential metabolites between groups L2025 and QHY (A and B), and L2025 and LHY (C and D).

Table 4. Anthocyanins and anthocyanidins detected in L2025, QHY and LHY.

Anthocyanidins	Anthocyanins	L2025	QHY	LHY
Peonidin		292,333	353,000*	548,333*
	Peonidin *O*-hexoside	1,560,000	18,966,667**	18,400,000**
	Peonidin 3-*O*-glucoside	1,490,000	20,713,333**	19,800,000**
	Peonidin 3, 5-diglucoside	271,667	111,767**	124,333**
Pelargonidin		76,133	140,500*	40,433
	Pelargonidin 3,5-*O*-glucoside	0	14,600**	17,367**
	Pelargonidin 3-*O*-glucoside	507,667	16,266,667**	19,333,333**
Cyanidin		73,167	2,236,667**	1,290,003
	Cyanidin 3-*O*-glucoside	0	1,326,667**	1,836,667**
	Cyanidin *O*-syringic acid	361,667	9,186,667**	9,010,000**
	Cyanidin *O*-diacetyl-hexoside-*O*-glyceric acid	17,700	415,000	123,900
	Cyanidin 3-*O*-rutinoside	66,400	1,496,667**	1,536,667**
	Cyanidin 3,5-*O*-diglucoside	338,000	14,500,000**	16,200,000**
Delphinidin		144,000	480,333**	464,000**
	Delphinidin 3-*O*-glucoside	0	325,000**	290,000**
	Delphinidin 3,5-*O*-glucoside	0	13,700**	12,317**
Malvidin		0	0	0
	Malvidin 3-*O*-galactoside	0	613,667**	786,333**
	Malvidin 3-*O*-glucoside	0	919,333**	1,089,333**

The number indicates the areas of the peaks obtained for each compound in the MRM analysis. These numbers could only be compared for the same metabolite in different samples.
*$P < 0.05$.
**$P < 0.01$.

the red poplar. Notably, differences were observed between QHY and LHY regarding the extent of the up-regulated expression compared with the L2025 expression levels. Some of the anthocyanin biosynthesis genes were more highly expressed in LHY than in QHY (Figure 9), which may explain why the anthocyanin levels were higher in LHY than in QHY (Figure 3).

Correlation between the metabolites and transcripts in three poplar varieties

We analyzed the correlation between the metabolite and transcript data. A 9-quadrant graph based on association analysis indicated that the metabolites in the third and seventh quadrants had the same transcript and metabolite accumulation

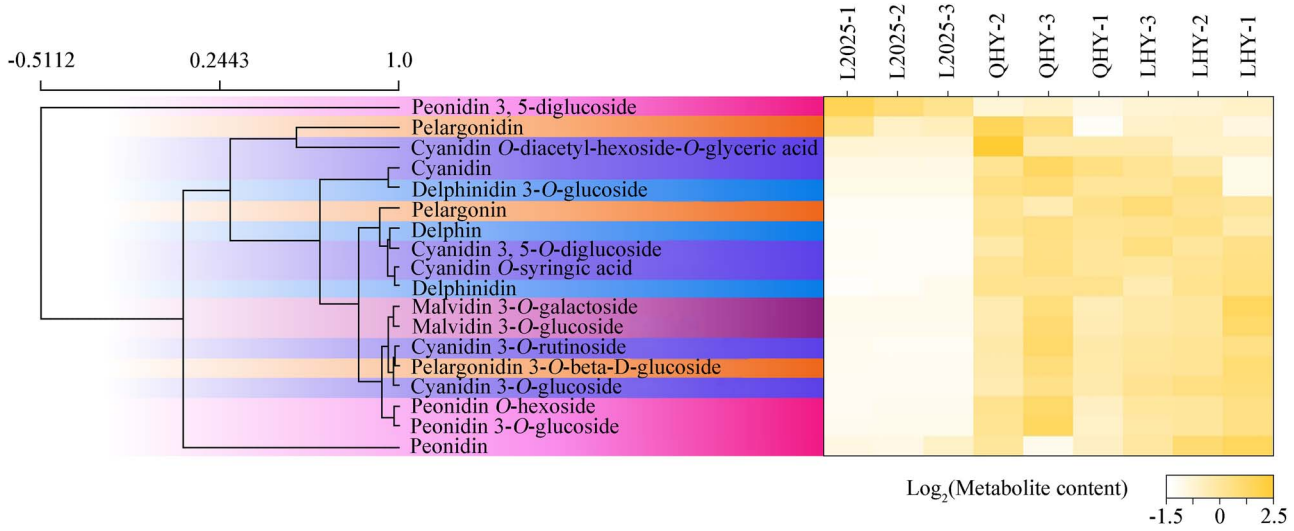

Figure 8. Heatmap and expression pattern cluster of all anthocyanin metabolites detected in L2025, LHY and QHY. The cluster on the left indicate the consistency of the metabolite expression pattern, based on all the anthocyanins detected, and the different background colors represent the color of metabolites. The rectangles with different colors on the right side of the heatmap indicate the content of metabolites. A rectangle's color close to yellow indicates high metabolite content. The metabolite content data were normalized by maximum difference normalization.

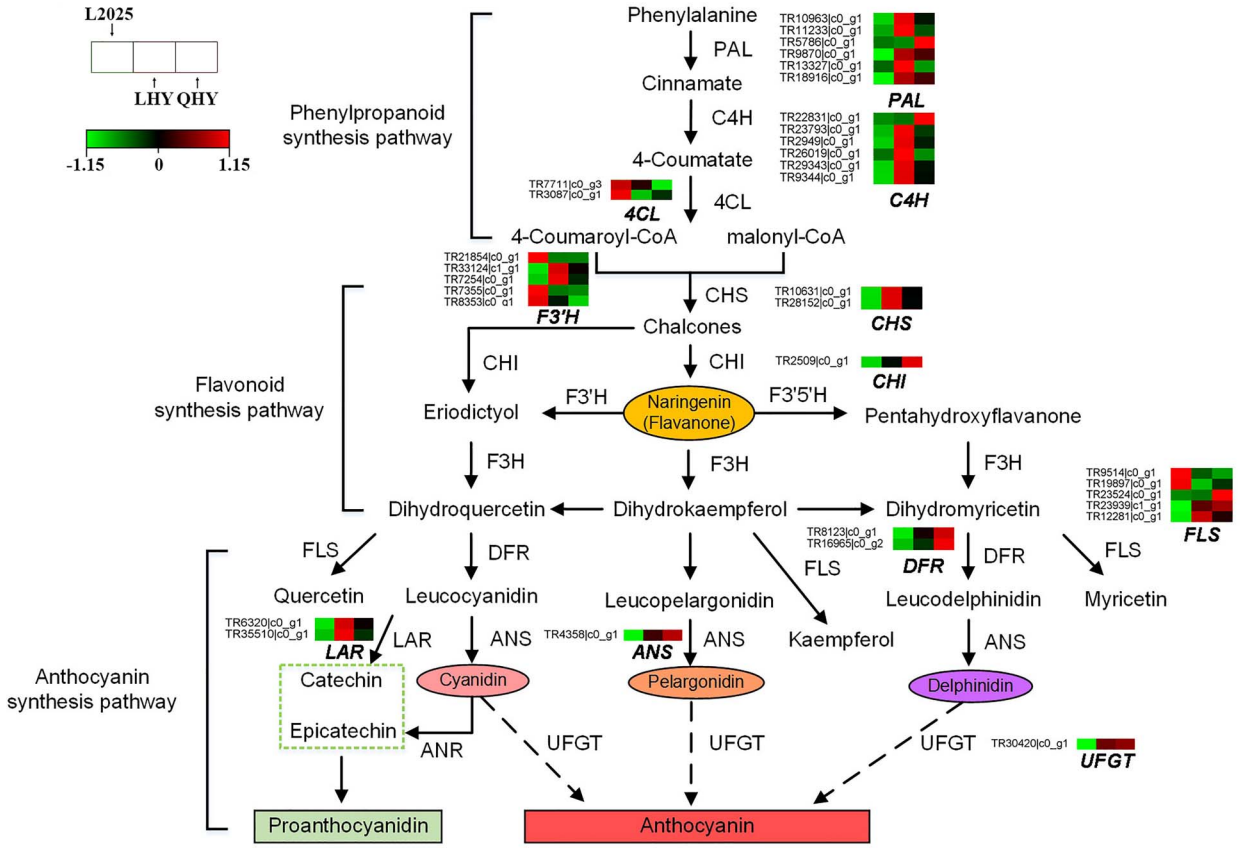

Figure 9. Phenylpropane metabolic pathway illustrates enhanced flavonoid biosynthesis in poplar. The flavonoid biosynthesis pathways have a common precursor. The expression profile of anthocyanin/flavonoid pathway genes was indicated. Red and green represent the up-regulated and down-regulated genes, respectively. PAL, Phe ammonia lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumaroyl-CoA synthase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavonoid 3-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol-4-reductase; ANS, anthocyanidin synthase; LAR, leucoanthocyanidin reductase; ANR, anthocyanidin reductase; UFGT, The UDP-glucose: Flavonoid-3-*O*-glucosyltransferase.

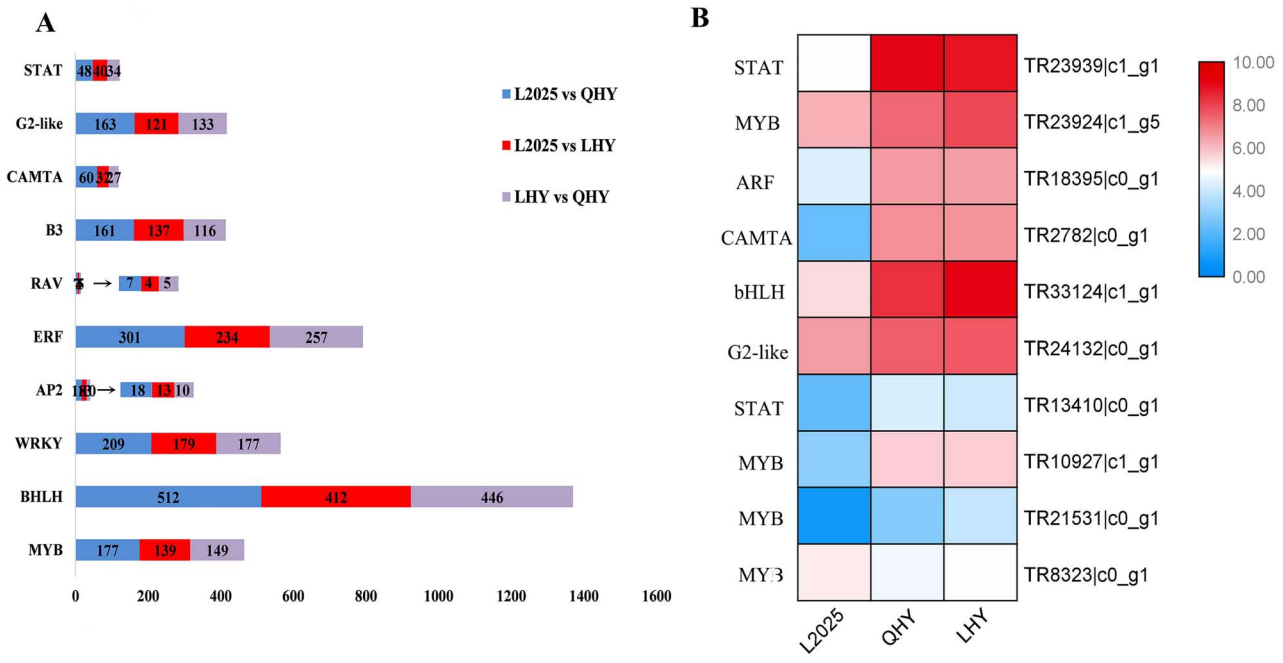

Figure 10. Main TFs differentially expressed among three poplar varieties. (A) The numbers of the main TFs family that were differentially expressed among three poplar varieties (the abscissa refers to all the genes of a single transcription factor such as MYB TF in the pairwise comparison of three poplars and the ordinate refers to the name of transcription factor all TFs). (B) Expression pattern of 10 TFs.

trends, whereas the metabolites in the first and ninth quadrants had the opposite transcript and metabolite accumulation patterns (see Figure S10 available as Supplementary data at *Tree Physiology* Online). Moreover, in the fifth quadrant, the transcripts and metabolites were relatively unchanged. In the third quadrant for L2025 vs QHY and L2025 vs LHY comparisons, methyl quercetin-*O*-hexoside, naringenin-*C*-hexoside, isorhamnetin *O*-hexoside, apigenin *C*-glucoside, cyanidin 3-*O*-glucoside, engeletin, chrysoeriol *O*-hexosyl-*O*-hexoside and 4,2′,4′,6′-tetrahydroxychalcone, along with their related transcripts, were identified as up-regulated and positively related. Given that the metabolites in the third quadrant and their related transcripts were detected at higher levels in the mutant poplar varieties, these transcripts or metabolites may be important for flavonoid production or coloration in red poplars.

Expression analysis of MYB transcription factors

Transcription factors (TFs) are the major regulators of gene expression profiles. Therefore, we investigated the TF families that were differentially expressed among the three varieties. Of the detected TFs, many belonged to the MYB, bHLH, WRKY and ERF families (Figure 10A). Consequently, we compared the expression of *MYB* genes in three poplar varieties. Of the R2R3-MYB members, TR23924|c1_g5, TR10927|c1_g1 and TR21531|c0_g1 exhibited obvious up-regulation in the red poplar varieties. Additionally, the expression level of TR8323|c0_g1 was very low in QHY and LHY, but relatively higher in L2025. The three MYBs may be crucial for regulating the expression of structural genes involved in poplar leaf coloration (Figure 10B).

Discussion

What is responsible for the diversity in poplar colors?

The flavonoids involved in color formation are mainly anthocyanins, which give flowers, fruits and other plant tissues their characteristic red, purple and blue hues (Gould et al. 2008, Noda et al. 2013). To date, several papers have focused on the anthocyanin composition of *Populus* (Alcalde-Eon et al. 2016), whereas only the anthocyanin composition in leaves of *Populus tremula* and *Populus tremuloides* Michx have been reported (Bendz and Haglund 1968, Chang et al. 1989). Previous reports indicated that red-hued flowers contain pelargonidin- and cyanidin-based anthocyanins (Jaakola 2013, J.S.Cho et al. 2016). In the current study, these two anthocyanins were highly abundant in red poplar leaves. Pelargonin was exclusive to the red poplar leaves. Additionally, the cyanidin 3-*O*-rutinoside, cyanidin 3-*O*-glucoside, cyanidin 3,5-*O*-diglucoside and cyanidin *O*-syringic acid contents were significantly higher in the red poplars than in the control variety, and this finding was in accordance with the colors of these poplar leaves. A previous study has indicated that cyanidin 3-*O*-glucoside is the major pigment in both autumn leaves and anthocyanin-colored anthers of several poplar species (Alcalde-Eon et al. 2016). Here, we show that it was only present in red poplars, making it a valuable color-related

chemotaxonomic marker for poplar. Two delphinidin-based anthocyanins (delphinidin 3,5-*O*-glucoside and delphinidin 3-*O*-glucoside) were only detected in red poplar leaves, which was consistent with the results of an earlier investigation which indicated that delphinidin-based anthocyanins potentially responsible for blue hues may also contribute to red coloration (Li et al. 2012). Among the anthocyanins identified in this study, delphinidin 3-*O*-glucoside, cyanidin 3-*O*-glucoside and cyanidin 3-*O*-rutinoside have been identified in poplars (Alcalde-Eon et al. 2016), and cyanidin *O*-syringic acid was reported in the study of *Camellia sinensis* for the first time (Rothenberg et al. 2019). Although the anthocyanin types and contents were significantly different between the green and red poplars, no significant differences were found in the anthocyanin derivatives between the two red poplar varieties. The genes involved in the flavonoid pathway may be classified as structural genes encoding enzymes directly involved in flavonoid production or regulatory genes that control the expression of the structural genes. On the basis of our analysis on flavonoid pathway gene expression, most of the early (e.g., *PAL* and *C4H*), intermediate (e.g., *CHS* and *CHI*) and late (e.g., *ANS* and *DFR*) biosynthesis genes were more highly expressed in the red poplar varieties than in the green (control) variety. By contrast, the expression level of the early biosynthesis gene *4CL* was down-regulated in the red poplar varieties, as were the *F3′H* and *FLS* expression levels. These genes potentially represent the full pathway for the production of flavonoids and anthocyanins from phenylalanine. Their up-regulated expression in QHY and LHY leaves was in accordance with their presumed roles in the biosynthesis of anthocyanins and other flavonoids. Flavonoid biosynthesis is regulated by three types of TFs, namely R2R3-MYB factors, basic helix–loop–helix (bHLH) proteins and WD40 repeat proteins, which interact to form the MYB–bHLH–WD40 TF complex (Koes et al. 2005, Ramsay and Glover 2005, J.S.Cho et al. 2016, H.H. Wang et al. 2019, L.J.Wang et al. 2019). In this complex, MYB is the main component that determines the production of specific anthocyanins (H.H.Wang et al. 2019). As indicated in our transcript data, we identified several potential key TFs from these three families that may regulate flavonoid and anthocyanin production in poplar. Among them, three R2R3-MYB TFs altered their transcript abundance in all red poplars and likely played crucial roles in leaf coloration (Figure 10). We further found that the *MYB118* gene, which has been proved to promote anthocyanin biosynthesis (H.H.Wang et al. 2019), was unregulated in red leaf poplars. This result indicated the important roles of those MYBs. The function of two other MYBs on flavonoid or anthocyanin biosynthesis will be investigated further via gene transformation methods.

Insight into the potential utility of red poplars in animal feeds

Flavonoids, which play important roles in protecting human health, can fight cancer by removing free radicals. Similarly, they have a certain positive effect on the immune system, cardiovascular system and secretory system of humans and animals. Among them, flavonols, chalcones, isoflavones and other substances have been linked to anticancer effects and have shown great potential as cytotoxic anti-cancer agents that promote apoptosis in cancer cells (Yuan et al. 2014, 2016, Abotaleb et al. 2018). Flavanols are present in high amounts in plants that are of widespread consumption, i.e., tea (*C. sinensis*), grapes and cocoa; meanwhile the distribution of isoflavones is limited to soybeans, legumes and some microbes (Abotaleb et al. 2018, Perez-Vizcaino and Fraga 2018). Our results demonstrated that the leaves of red poplar varieties had relatively higher flavonoid content than the green leaves of common poplar variety, including in quercetin class, rhamnetin class, isorhamnetin class and kaempferol class (Table 5). Syringetin induces human osteoblast differentiation and may be useful for treating cancer (Bando et al. 2017). Quercetins may strengthen blood vessels, enhance the use of vitamin C and increase collagen production. It may also decrease cholesterol levels and blood pressure as well as preventing blood clots (Nishimura et al. 2016, Zhang et al. 2017). Some epidemiological studies have found a positive association between the consumption of foods containing kaempferol and a reduced risk of developing several disorders such as cancer and cardiovascular diseases (Calderón-Montaño et al. 2011). Isorhamnetin is one of the most important active ingredients in the leaves of *Ginkgo biloba* L., which possesses extensive pharmacological activities. Isorhamnetin has the effects of cardiovascular and cerebrovascular protection, anti-tumor, anti-inflammatory, anti-oxidation, organ protection and prevention of obesity (Gong et al. 2020). Many of the substances described above showed an upward trend in accumulation in red leaf poplar varieties. Thus, besides anthocyanins, other flavonoids are also important inresearch on directions for metabolic pathways in red poplar. Whether the flavonoid active substances in red poplar are as high as those in *Hippophae rhamnoides* L. or *G. biloba*, and whether the content of these substances will differ in varying growth cycles of red poplar, remain unknown. These are the scientific questions that we and more researchers should focus on in the future. Given that poplar leaf is a common feedstuff for animals, particularly for sheep, goats and rabbits, the relatively flavonoid-rich red poplar varieties may be more useful as animal feed than the common poplar (Ayers et al. 1996), even though the widely targeted method is impossible to know the quantitative relevance of each compound or each flavonoid class in the samples.

A colorful model for studying tree flavonoids

Populus spp., commonly named as poplars, aspens and cottonwoods, are distributed in the northern hemisphere (J.S.Cho et al. 2016). Poplar is a model for studying tree-specific perennial growth and development. Previously, 15 flavonoids

Table 5. Flavonol and isoflavone that accumulated in three poplar varieties.

Category	Fold change		*P*-values	
	LHY/L2025	QHY/L2025	$P_{LHY/L2025}$	$P_{QHY/L2025}$
Flavonol				
Quercetin	4.01*	3.90*	0.01	0.00
Quercetin-3,4′-*O*-di-beta-glucopyranoside	3.13*	2.94*	0.00	0.00
Quercetin 7-*O*-rutinoside	2.86*	3.09*	0.00	0.00
Quercetin 3-*O*-rutinoside	2.91*	2.80*	0.00	0.00
Kumatakenin	2.94*		0.00	
Rhamnetin (7-*O*-methyl quercetin)	4.03*	3.65*	0.00	0.00
Morin	3.91*	3.81*	0.02	0.00
Herbacetin	3.21*	3.11*	0.01	0.00
Isorhamnetin 3-*O*-glucoside	3.29*	3.21*	0.00	0.00
Isoflavone				
Orobol (5,7,3′,4′-tetrahydroxyisoflavone)	4.040*	3.36*	0.00	0.01
Genistein 7-*O*-Glucoside	3.99*	3.96*	0.00	0.00

*A statistical difference at $P < 0.05$ level between the mean of the two samples.

present in apical tissues of two full-sib poplar families were analyzed via HPLC. The results indicated that metabolite profiling combined with quantitative trait locus analysis can be used to identify loci that control metabolite abundance in trees (Morreel et al. 2006). Recently, by applying a widely targeted liquid LC–MS method, flavonoid was profiled in 14 plant species spanning the pre-angiosperm, monocot and eudicot clades, *Populus deltoides* was included. The flavonoid profiles visualized by hierarchical cluster analysis indicated that *Populus* has a different flavonoid composition from monocots and herbaceous dicots. Compared with flavonoid biosynthesis pathways, which have been extensively elucidated in *Arabidopsis*, petunia, maize and tomato (Povero et al. 2011, Zhang et al. 2019), poplar flavonoid biosynthesis needs further studies. On the basis of the flavonoid metabolic pathways that have been identified in other plants, the multi-enzyme complexes that catalyze sequential reactions in flavonoid pathways have been fully understood. Phenylalanine, which is a direct precursor for the biosynthesis of anthocyanins and other flavonoids, is converted by phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H) and 4-coumarate-CoA ligase (4CL) into 4-coumaroyl-CoA and malonyl-CoA (N.Wang et al. 2018). The next key reaction is the conversion of 4-coumaroyl-CoA to dihydroflavonol, which can be converted to anthocyanins, flavonols and other flavonoids via the activities of various enzymes. The expression of genes encoding these enzymes is important for the coloration of plant tissues. Therefore, a model has been proposed for the enzyme reactions starting with phenylalanine and PAL (Figure 9). Metabolome and transcriptome studies have revealed that flavonoids are abundant in medicinal plants, crops and flowers. Comparatively, there is relatively little available data regarding flavonoid- and anthocyanin-producing trees. A previous report indicated that poplar contains eight major classes of flavonoids, namely chalcones, dihydrochalcones, flavanones, flavones, dihydroflavonols, flavonols, anthocyanins and proanthocyanidins (Tsai et al. 2006). However, flavonoid composition and contents vary among poplar species and clones (Greenaway et al. 1991, 1992, Donaldson et al. 2006). Analyses of the red mutants of the L2025 poplar variety enabled us to clarify the steps of the flavonoid and anthocyanin pathways because color changes are related to flavonoid and anthocyanin metabolism and the expression of related genes. We observed considerable variability in the flavonoid metabolites in the three analyzed poplar varieties. The flavonoid content was much higher in the red poplar varieties than in the control poplar variety with green leaves. The 210 flavonoid metabolites detected by HPLC-ESI-MS/MS were more than what has been reported for most plants, and were not limited to trees. Thus, we propose that the red poplar varieties may be useful as models for studying flavonoids in trees.

Conclusions

Poplar is recognized as a model for studying perennial growth and development. To our knowledge, the differences in metabolic profiles on the basis of leaf coloration have not been investigated yet. This study focused on the specific metabolic diversity of flavonoids in three poplar varieties with different leaf colors to elucidate the factors responsible for leaf coloration. Using a novel widely targeted metabolomics method, which has been established and widely applied in plants in recent years, a total of 210 flavonoids with various modifications were detected and annotated. Most of these flavonoids had high accumulation levels in red poplars. Gene expression analysis revealed that the phenylpropanoid biosynthesis genes displayed higher levels in red poplars than in green poplars. Several potential anthocyanin biosynthesis regulators especially *MYBs* were overrepresented in red poplars. On the basis of a comprehensive analysis

relating flavonoids compounds to gene expression profiles, the mechanism of coloration was discussed. These findings provide new insights into poplar flavonoid metabolites as well as improving our understanding of the mechanisms accounting for leaf coloration in trees.

Supplementary data

Supplementary data for this article are available at *Tree Physiology* Online.

Acknowledgments

We appreciate the help from Xiangjun Cheng of Shangqiu Zhongxing Seedling Planting Co., Ltd, the help from Jiang Feng and Lei Lei from the Metware Co., Ltd and the advice of Baoqing Zhu from Beijing Forestry University.

Conflicts of interest

The authors declare that they have no conflict of interest.

Authors' contributions

J.H.C. designed and supervised the study. Q.Q.L. and R.Y.T. performed the experiments; A.K.W. and H.C.Z. cultivated and collected the poplars; J.H.C., R.Y.T. and S.P.R. analyzed the data, J.H.C. drafted the manuscript. Y.L. and L.S.W. revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by National Natural Science Foundation of China (31670610, 31370597)and the Fundamental Research Funds for the Central University (2016ZCQ05).

References

Alcalde-Eon C, García-Estévez I, Rivas-Gonzalo JC, Rodríguez de la Cruz D, Escribano-Bailón MT (2016) Anthocyanins of the anthers as chemotaxonomic markers in the genus *Populus L*. differentiation between *Populus nigra*, *Populus alba* and *Populus tremula*. Phytochemistry 128:35–49.

Ayers AC, Barrett RP, Cheeke PR (1996) Feeding value of tree leaves (hybrid poplar and black locust) evaluated with sheep, goats and rabbits. Anim Feed Sci Tech 57:51–62.

Bando SI, Hatano O, Takemori H, Kubota N, Ohnishi K (2017) Potentiality of syringetin for preferential radiosensitization to cancer cells. Int J Radiat Biol 93:286–294.

Bridle P, Timberlake CF (1997) Anthocyanins as natural food colours—selected aspects. Food Chem 58:103–109.

Calderón-Montaño JM, Burgos-Morón E, Pérez-Guerrero C, López-Lázaro M (2011) A review on the dietary flavonoid kaempferol. Mini Rev Med Chem 11:298–344.

Chang SJ, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116.

Chen W, Gong L, Guo ZL, Wang WS, Zhang HY, Liu XQ, Yu SB, Xiong LZ, Luo J (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. Mol Plant 6:1769–1780.

Chlopicka J, Pasko P, Gorinstein S, Jedryas A, Zagrodzki P (2012) Total phenolic and total flavonoid content, antioxidant activity and sensory evaluation of pseudocereal breads. LWT- Food Sci Technol 46:548–555.

Cho JS, Nguyen VP, Jeon HW et al. (2016) Overexpression of PtrMYB119, a R2R3-MYB transcription factor from *Populus trichocarpa*, promotes anthocyanin production in hybrid poplar. Tree Physiol 36:1162–1176.

Cho K, Cho KS, Sohn HB, Ha IJ, Hong SY, Lee H, Kim YM, Nam MH (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. J Exp Bot 67:1519–1533.

Chung SY, Janelle ML, Huang MT, Newmark HL (2001) Inhibition of carcinogenesis by dietary polyphenolic compounds. Carcinogenesis 21:381–406.

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7:1085–1097.

Donaldson JR, Stevens MT, Barnhill HR, Lindroth RL (2006) Age-related shifts in leaf chemistry of clonal aspen (*Populus tremuloides*). J Chem Ecol 32:1415–1429.

Dong T, Han R, Yu J, Zhu M, Zhang Y, Gong Y, Li Z (2019) Anthocyanins accumulation and molecular analysis of correlated genes by metabolome and transcriptome in green and purple asparaguses (*Asparagus officinalis,* L.). Food Chem 271:18–28.

Dooner HK, Robbins TP, Jorgensen RA (1991) Genetic and developmental control of anthocyanin biosynthesis. Annu Rev Genet 25:173–199.

Fraga CG, Clowers BH, Moore RJ, Zink EM (2010) Signature-discovery approach for sample matching of a nerve-agent precursor using liquid chromatography-mass spectrometry, XCMS, and chemometrics. Anal Chem 82:4165–4173.

Gao J, Zhang Y, Wang C, Zhang SG, Qi LW, Song WQ (2009) AFLP fingerprinting of *Populus deltoides* and *Populus* × *canadensis* elite accessions. New For 37:333–344.

Gong G, Guan YY, Zhang ZL, Rahman K, Wang SJ, Zhou S, Luan X, Zhang H (2020) Isorhamnetin: a review of pharmacological effects. Biomed Pharmacother 128:110301.

Gould K, Davies K, Winefield C (eds) (2008) Anthocyanins: biosynthesis, functions, and applications. Springer Verlag, New York, NY, pp 1–20.

Grabherr MG, Haas BJ, Yassour M et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

Greenaway W, Gümüsdere I, Whatley FR (1991) Analysis of phenolics of bud exudate of *Populus euphratica* by GC-MS. Phytochemistry 30:1883–1885.

Greenaway W, English S, Whatley FR (1992) Relationships of *Populus* × *acuminata* and *Populus* × *generosa* with their parental species examined by gas chromatography mass spectrometry of bud exudates. Can J Bot 70:212–221.

Hagiwara A, Yoshino H, Ichihara T et al. (2002) Prevention by natural food anthocyanins, purple sweet potato color and red cabbage color, of 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP)-associated colorectal carcinogenesis in rats. J Toxicol Sci 27: 57–68.

Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF (1996) Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics 143:1395.

Hatier JHB, Gould KS (eds) (2009) Anthocyanin function in vegetative organs. Springer, New York, NY.
Howe EA, Sinha R, Schlauch D, Quackenbush J (2011) RNA-Seq analysis in MeV. Bioinformatics 27:3209–3210.
Jaakola L (2013) New insights into the regulation of anthocyanin biosynthesis in fruits. Trends Plant Sci 18:477–483.
Jeong ST, Goto-Yamamoto N, Kobayashi S, Esaka M (2004) Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skins. Plant Sci 167:247–252.
Kerckhoffs LHJ, Schreuder MEL, Tuinen AV, Koornneef M, Kendrick RE (1997) Phytochrome control of anthocyanin biosynthesis in tomato seedlings: analysis using photomorphogenic mutants. Photochem Photobiol 65:374–381.
Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242.
Koes RE, Quattrocchio F, Mol J (1994) The flavonoid biosynthetic pathway in plants: function and evolution. Bioessays 16:123–132.
Lev-Yadun S, Gould KS (eds) (2008) Role of anthocyanins in plant defence. Springer, New York, NY.
Li X, Thwe AA, Park NI, Suzuki T, Kim SJ, Park SU (2012) Accumulation of phenylpropanoids and correlated gene expression during the development of tartary buckwheat sprouts. J Agr Food Chem 60:5629–5635.
Lichtenthaler HK (1987) Chlorophyll and carotenoids: pigments of photosynthetic biomembranes. Meth Enzymol 148:331–382.
Lila MA (2004) Anthocyanins and human health: an in vitro investigative approach. J Biomed Biotechnol 5:306–313.
Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408.
Ma D, Reichelt M, Yoshida K, Gershenzon J, Constabel CP (2018) Two R2R3-MYB proteins are broad repressors of flavonoid and phenylpropanoid metabolism in poplar. Plant J 96:949–965.
Morreel K, Goeminne G, Storme V et al. (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. Plant J 47:224–237.
Nishimura M, Ohkawara T, Sato Y, Satoh H, Suzuki T, Ishiguro K, Noda T, Morishita T, Nishihira J (2016) Effectiveness of rutin-rich Tartary buckwheat (*Fagopyrum tataricum* Gaertn.) 'Manten-Kirari' in body weight reduction related to its antioxidant properties: a randomised, double-blind, placebo-controlled study. J Funct Foods 26:460–469.
Noda N, Aida R, Kishimoto S, Ishiguro K, Fukuchi-Mizutani M, Tanaka Y, Ohmiya A (2013) Genetic engineering of novel bluer-colored chrysanthemums produced by accumulation of delphinidin-based anthocyanins. Plant Cell Physiol 54:1684–1695.
Parr AJ, Bolwell GP (2000) Phenols in the plant and in man. The potential for possible nutritional enhancement of the diet by modifying the phenols content or profile. J Sci Food Agr 80:985–1012.
Peng M, Shahzad R, Gul A et al. (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. Nat Commun 8:1975.
Peng XJ, Liu H, Chen PL et al. (2019) A chromosome-scale genome assembly of paper mulberry (*Broussonetia papyrifera*) provides new insights into its forage and papermaking usage. Mol Plant 12:661–677.
Perez-Vizcaino F, Fraga CG (2018) Research trends in flavonoids and health. Arch Biochem Biophys 646:107–112.
Povero G, Gonzali S, Bassolino L, Mazzucato A, Perata P (2011) Transcriptional analysis in high-anthocyanin tomatoes reveals synergistic effect of Aft and atv genes. J Plant Physiol 168:270–279.
Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends Plant Sci 10:63–70.
Ross JA, Kasum CM (2002) Dietary flavonoids: bioavailability, metabolic effects, and safety. Annu Rev Nutr 22:19–34.
Rothenberg DO, Yang HJ, Chen MB, Zhang WT, Zhang LY (2019) Metabolome and transcriptome sequencing analysis reveals anthocyanin metabolism in pink flowers of anthocyanin-rich tea (*Camellia sinensis*). Molecules 24:1064.
Shen JZ, Zhang YD, Zhou L et al. (2019) Transcriptomic and metabolomic profiling of *Camellia sinensis* L. cv. 'Suchazao' exposed to temperature stresses reveals modification in protein synthesis and photosynthetic and anthocyanin biosynthetic pathways. Tree Physiol 39:1583–1599.
Taylor G (2002) Populus: Arabidopsis for forestry. Do we need a model tree? Ann Bot 90:681–689.
Tian SH (2003) Zhonglin 2025 (in Chinese). Rural Newspaper 5:29. doi: 10.19433/j.cnki.1006-9119.2003.05.028.
Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytol 172:47–62.
Wang AM, Li RS, Ren L, Gao XL, Zhang YG, Ma ZM, Ma DF, Luo YH (2018) A comparative metabolomics study of flavonoids in sweet potato with different flesh colors (*Ipomoea batatas* (L.) Lam). Food Chem 260:124–134.
Wang HH, Wang XQ, Song WM, Bao Y, Jin YL, Jiang CM, Wang CT, Li B, Zhang HX (2019) PdMYB118, isolated from a red leaf mutant of *Populus deltoids*, is a new transcription factor regulating anthocyanin biosynthesis in poplar. Plant Cell Rep 38:927–936.
Wang LJ, Lu WX, Ran LY, Dou LW, Yao S, Hu J, Fan D, Li CF, Luo KM (2019) R2R3-MYB transcription factor MYB6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. Plant J 99:733–751.
Wang MM, Huang MS, Tan Y, Zhou BL, Jiang ZL (1980) A comparative study of five modern clones of Aigeiros (in Chinese). J Nanjing Forest Univ 23:1–9.
Wang N, Jiang SH, Zhang ZY, Fang HC, Xu HF, Wang YC, Chen XS (2018) *Malus sieversii*: the origin, flavonoid synthesis mechanism, and breeding of red-skinned and red-fleshed apples. Hortic Res 5:70.
Yuan L, Wang J, Wu WQ, Liu Q, Liu XB (2016) Effect of isoorientin on intracellular antioxidant defence mechanisms in hepatoma and liver cell lines. Biomed Pharmacother 81:356–362.
Yuan L, Wei SP, Wang J, Liu XB (2014) Isoorientin induces apoptosis and autophagy simultaneously by reactive oxygen species (ROS)-related p53, PI3K/Akt, JNK, and p38 signaling pathways in HepG2 cancer cells. J Agr Food Chem 62:5390–5400.
Zhang HC, Koes R, Shang HQ et al. (2019) Identification and functional analysis of three new anthocyanin R2R3-MYB genes in Petunia. Plant Direct 1:e00114.
Zhang LJ, Li XX, Ma B et al. (2017) The tartary buckwheat genome provides insights into rutin biosynthesis and abiotic stress tolerance. Mol Plant 10:1224–1237.
Zhu GT, Wang SC, Huang ZJ et al. (2018) Rewiring of the fruit metabolome in tomato breeding. Cell 172:249–261.e12.