

Metabolomics combined with transcriptomics reveals the accumulation mechanism of the fruit pulp colour of *Baccaurea ramiflora* Lour

Jianjian Huang

Hanshan Normal University

Jie Chen

Guangdong Ocean University

Hui Zhu

Hanshan Normal University

Yuzhong Zheng

Hanshan Normal University

Jiaqi Zheng

Hanshan Normal University

Xueying Wen

Hanshan Normal University

Suying Cai

Hanshan Normal University

Yingchun Zhu

Hanshan Normal University

Qinghan Wu

Hanshan Normal University

Zikai Chen

Hanshan Normal University

Fengnian Wu

`fengnian.wu@hstc.edu.cn`

Hanshan Normal University

Article

Keywords: *Baccaurea ramiflora* Lour., metabolome, transcriptome, anthocyanin, pulp colour

Posted Date: March 27th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3964227/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Baccaurea ramiflora Lour., a wild fruit tree with edible, ornamental, and medicinal qualities. The mechanism behind the color accumulation in its fruit pulp, which can be either pink or milky-white, remains unclear. This study investigates the metabolome and transcriptome of two *B. ramiflora* pulp types—LR (milky-white at maturity) and BR (pink at maturity)—to elucidate their coloration processes. We identified 35 flavonoids, including nine involved in the anthocyanin synthesis pathway, confirming cyanidin as the pivotal pigment for the pink pulp coloration. An examination of the flavonoid and anthocyanin biosynthetic pathways in *B. ramiflora* pulp uncovered 38 differentially expressed genes associated with structural genes. The genes *F3'5'H* and *UFGT* exhibited high expression levels in the first two developmental stages of BR, significantly more than in LR, and were almost non-existent in later stages, signifying their crucial role in the differential color accumulation between BR and LR pulps. Additionally, the expression levels of *CHI* and *FLS*, early-stage structural genes in the anthocyanin synthesis pathway, correlated with the concentrations of naringenin and quercetin, indicating their importance in the anthocyanin biosynthesis pathway of *B. ramiflora* pulp. These discoveries provide new insights that could facilitate the breeding of *B. ramiflora* varieties with diverse pulp colors.

Introduction

Baccaurea ramiflora Lour., an evergreen arboreal species from the *Baccaurea* genus, is nestled within the Phyllanthaceae family, under the order of Malpighiales (APG IV system). This versatile plant, alternatively known as the Burmese grape in English, Mafai in Thai, Latkan in Hindi, and Lotka in Bengali, flourishes predominantly in South China and southern Yunnan and extends its reach across Southeast Asia and South Asia^{1,2}. With its economic potency and developmental prospects, *B. ramiflora* is a revered wild resource. It is distinguished by a pleasing arboreal silhouette and robust adaptability. Its fruit, reminiscent of a grape, offers a harmonious blend of sweetness and tartness, exudes a delightful aroma, and possesses a tender texture. The fruit can be savored fresh or can be processed into fruit juice, preserved fruit, jam, and fruit wine. The wood it provides is suitable for crafting furniture and joinery. The plant's roots, stems, leaves, and fruits find medicinal use in alleviating asthma and cough, detoxifying the body, relieving itching, and producing antitumour effects^{3,4}.

In the contemporary era, plant-based foods are gaining increasing popularity. Owing to their rich nutrient content, they play an instrumental role in providing a balanced diet for human health⁵. The fruit of *B. ramiflora*, with its delightful balance of sweetness and sourness, has garnered considerable attention, and its nutritional value has been validated by researchers⁶. The fruit's edible rate stands at 49.2%, and its composition includes significant percentages of water, fat, starch, fibre, vitamin C, titratable acid, and total sugar^{7–10}. The fruit pulp, while low in fat, ash, and protein, is abundant in dietary fibre, ascorbic acid (vitamin C), and mineral elements such as calcium, magnesium, and phosphorus¹⁰. As a species of wild fruit tree with potential for development, *B. ramiflora* can cater to the escalating demand for plant-based food. A 2018 survey revealed that the price of fresh *B. ramiflora* fruit in major supermarkets in the

Honghe Hani and Yi Autonomous Prefecture of Yunnan Province was approximately RMB 10/kg, and the market price in southern Guangxi ranged from RMB 7–10/kg. Given that an adult tree can yield 100–150 kg per plant, cultivating *B. ramiflora* can yield substantial economic benefits for farmers¹¹.

The hue of a fruit significantly influences consumer preferences¹². The myriad fruit pulp colours, including shades of red, pink, red-purple, purple, and blue, are primarily attributed to the differential accumulation of flavonoids and anthocyanins. These flavonoids, ubiquitous in plants, are secondary metabolites synthesized via the phenylalanine pathway^{13,14}. They serve a plethora of functions, including defense against pathogens, prevention of UV damage, antioxidation, regulation of auxin transport and allelopathy, and enhancement of plant resistance in adversity (biological or abiotic stress)^{15,16}, thereby bolstering the plant's immunity and pathogen protection. Anthocyanins, a specific class of flavonoids, have been extensively studied. They are water-soluble pigments and include six common types: pelargonin, cyanin, petunin, malvin, peonin, and delphinin¹⁷. Anthocyanins, like other flavonoids, can prevent cardiovascular and cerebrovascular diseases¹⁸, diabetes¹⁹, cancer, and heart disease^{20–22}. Consequently, the colour qualities and value imparted by flavonoids in fruits have been progressively recognized, and they are particularly favoured by consumers and can be used as indicators in fruit quality research.

In recent years, the accumulation mechanism of pigments in various plant tissues has drawn increasing attention from researchers. Numerous studies, both within China and internationally, have been conducted on the anthocyanin biosynthetic pathway, revealing that this pathway is a downstream segment of the flavonoid metabolic pathway. This biosynthetic pathway is highly conserved across plants^{23,24}. Both anthocyanins and flavonoids are synthesized via the phenylalanine pathway and primarily undergo three stages of metabolism: the initial stage of phenylalanine cleavage, the intermediate stage of flavonoid metabolism, and the final stage of anthocyanin synthesis^{25,26}. During these stages, various compounds are synthesized, catalyzed by a sequence of enzymes, leading to the production of diverse compounds, including flavonols and anthocyanins, which are ultimately stored in vacuoles.

While the anthocyanin biosynthesis pathways of tropical fruits such as lychee²⁷ and longan²⁸ have been studied, there is a dearth of reports on the colouring mechanism of *B. ramiflora* fruit pulp. In this study, we performed metabolome and transcriptome analyses on the pulps of two *B. ramiflora* cultivars, LR (milk-white pulp) and BR (pink pulp), throughout their development to elucidate the molecular mechanism of anthocyanin accumulation in the pulp.

Results

Flavonoids in the pulp of *B. ramiflora*.

Flavonoids and phenylpropanes are the precursors of anthocyanin synthesis and are related to the synthesis of anthocyanins. Flavonoids have a variety of stress-protective effects on plants and have

health-promoting effects on the human body²⁹. A branch of the flavonoid biosynthetic pathway is involved in the production and regulation of anthocyanins, procyanidins, and flavonols²⁹. The metabolomic analysis of the fruit pulp of *B. ramiflora* showed that there were 35 types of flavonoids (Table S2), and no carotene was identified. The flavonoids were mainly rhusflavanone, procyanidin B1, and (+)-catechin. Nine metabolites of flavonoids and anthocyanin biosynthetic pathways were identified in the fruit pulp of *B. ramiflora*: naringenin chalcone, naringenin, and eriodictyol, taxifolin/dihydroquercetin, kaempferol, quercetin, (+)-catechin, (-)-epicatechin, and procyanidin B1.

Dynamic changes in anthocyanin pathway metabolites during LR and BR pulp development.

The changes in the 9 flavonoids in the anthocyanin synthesis pathway during the developmental stages are shown in Fig. 2. Overall, they showed a decline with time. Compared with BR, the anthocyanin synthesis pathways of LR had a greater rate of decline during the period from 52 DAF to 73 DAF, which may have been associated with the more abundant anthocyanins in the pulp of BR. The accumulation of the contents of various substances mainly occurred in the first two stages, indicating that the anthocyanins in the fruit pulp of *B. ramiflora* were mainly synthesized in the early stage.

Differential metabolites in LR and BR pulp during development stages. Naringin chalcone showed a significant decrease only in LR4 vs. LR3 ($p = 0.028$, FC -2.437, VIP 1.045). Dihydroquercetin significantly decreased in BR5 vs. BR4 ($p = 0.020$, FC -1.338, VIP 1.209), LR3 vs. BR3 ($p = 0.011$, FC -1.346, VIP 1.282), and LR4 vs. BR4 ($p = 0.006$, FC -2.070, VIP 1.661). Dihydroquercetin is an essential precursor for the synthesis of cyanidin. We speculated that the increase in dihydroquercetin content in BR was an important reason for the pink colour of the pulp. Kaempferol significantly decreased in BR5 vs. BR4 ($p = 0.021$, FC -1.339, VIP 1.256), LR3 vs. BR3 ($p = 0.011$, FC -1.378, VIP 1.335), and LR4 vs. BR4 ($p = 0.004$, FC -2.070, VIP 1.661). Procyanidin B1 showed a decreasing trend during the entire fruit development process, and we speculated that the gradual decrease in the bitterness and astringency of the fruit pulp was related to its decrease. (+)-Catechin is an upstream substance for the synthesis of procyanidin B1, and it exhibited a decreasing trend during development, which is consistent with the decrease in the synthesis of procyanidin B1, (+)-catechin only showed a significant decrease in LR4 vs. LR3 ($p = 0.017$, FC -2.140, VIP 1.002). Procyanidin B1 and (+)-catechin gradually decreased, which may be related to the synthesis of anthocyanidins or the reduction of upstream substrate synthesis.

Transcriptome sequencing and evaluation.

The RNA at different developmental stages of the fruit pulp was sequenced using PE150 on the Illumina NovaSeq platform. The filtered data are shown in Table S3. The clean reads after processing ranged from 22,698,981 to 27,405,036; the base data after processing ranged from 6.80 Gb to 8.21 Gb, which met the requirements we set. The GC content was in the range of 43.65% – 46.99%. Q20 was in the range of 96.86% – 97.65%, and Q30 was between 91.56% and 92.98%, both greater than 88%, indicating that the RNA sequencing data in this study had good quality and met our requirements. The comparison of the 30 RNA-seq clean reads of *B. ramiflora* with the *B. ramiflora* genome showed that the total number of

reads in each sample was between 20,335,899 and 33,541,270, and the alignment rate reached 88.97% – 92.79%, further indicating that the quality of the transcriptome sequencing data was satisfactory.

The Pearson correlation coefficient is an important indicator of the reliability of gene expression experiments done on various samples and of the rationality of sample selection. Based on the fragments per kilobase per million expression values of all genes in each sample, the intragroup and intergroup correlation coefficients of the samples were calculated, and a heatmap was plotted to display the sample duplication within the group and the sample difference between the groups. The heatmap of the square of Pearson's correlation coefficient (R^2) of the 30 samples of fruit pulp is shown in Fig. 3. The R^2 is more closer to 1, the more similar the expression patterns are between the samples. In general, the requirement of biological replication is $R^2 > 0.8$. The R^2 between samples in this study was between 0.807 and 0.997. The three biological samples had good repeatability and met the requirement of the correlation coefficient.

PCA is often used to assess intergroup differences and intragroup sample reproducibility. PCA performs dimensionality reduction of gene expression as well as principal component calculation through a linear algebraic calculation method. Ideally, the same group of samples should be clustered together. The PCA of the three biological replicates at each stage is shown in Fig. 4. The samples at the first developmental stage of LR and BR were closely clustered, indicating that the reproducibility of the samples was good; the samples at the second and third stages are more scattered, indicating that the samples at these two developmental stages are quite different, but the overall clustering is one type; the clustering of samples from the last two developmental stage indicates that the biological reproducibility is good. The samples from the five stages were divided into three categories: 30 DAF was the young fruit stage, 52 DAF and 73 DAF were the growth stage, and 93 DAF and 112 DAF were the ripening stage. PCA indicated that the fruit started to undergo ripening from 73 DAF to 93 DAF, consistent with the metabolomic analysis.

Structural genes in the anthocyanin biosynthetic pathway in fruit pulp of *B. ramiflora*.

The biosynthetic pathways of flavonoids in plants have been relatively clear and are relatively conserved among different species^{14,30}. In this study, after obtaining the DEGs, based on the GO enrichment and KEGG enrichment analysis of DEGs at different developmental stages, the structural genes in the biosynthetic pathways of flavonoids and anthocyanins in the fruit pulp of *B. ramiflora* and DEGs at different developmental stages were distinguished (Table S1), and a pathway map containing the expression heatmap of various structural genes in the anthocyanin biosynthesis pathway in the fruit pulp was constructed (Fig. 5). A total of 43 structural genes related to the flavonoid biosynthetic pathway were found, and differential expression of 39 genes was found (Table S4). The early structural genes of anthocyanin synthesis included *PAL* (ctg2456.g21765, ctg836.g07135), *C4H* (ctg733.g06345), *C4L* (ctg3279.g28008, ctg2580.g23157, ctg836.g07174, ctg3058.g26217 and ctg3058.g26218), *CHS* (ctg655.g05350, ctg3105.g26705), *CHI* (ctg965.g08335, ctg502.g04473), *F3H* (ctg1825.g16638, ctg1147.g09820), *F3'5'H* (ctg3090.g26553, ctg1305.g11051, ctg2135.g19559, ctg2839.g24678 and ctg17.g00099), *F3'H* (ctg2495.g22288, ctg287.g02877, ctg2135.g19557), *DFR*

(ctg1578.g14126), and *FLS* (ctg1560.g13893, ctg2313.g20660, ctg2657.g23512). The structural genes in the middle and late stages of anthocyanin synthesis included *LAR* (ctg2548.g22735, ctg1760.g15871), *LDOX* (leucoanthocyanidin dioxygenase) (ctg438.g03741, ctg3056.g26143), and *UFGT/3GT* (ctg2661.g23571, ctg1496.g13357, ctg1652.g14649, ctg1652.g14646, ctg1306.g11055, ctg1210.g10432, ctg2440.g21453, ctg2170.g19838, ctg3000.g25733, ctg2661.g23551, ctg2055.g18534), and *ANR* (ctg1329.g11603, ctg1g2699.g23958). In the transcriptome analysis, the *ANS* gene was not annotated, but the two homologous *LDOX* sequences were annotated. Cyanidin is produced from leucoanthocyanidin via *ANS* or an iron-dependent 2-gluconate dioxygenase (also known as *ANS*³¹). The *ANS* and *LDOX* genes have the same biological function. In addition, the *ANR* gene that acts in the procyanidin pathway was discovered as a DEG here, and its role in the metabolome is of great significance to further understand the procyanidin pathway.

DEGs in the anthocyanin synthesis pathway.

In the DEGs of LR1 vs. BR1, *C4L* (ctg2580.g23157 and ctg836.g07174) was downregulated; *CHI* (ctg965.g08335) was upregulated; ctg1825.g16638 of *F3H* was upregulated, and ctg1147.g09820 was downregulated; and 3 *F3'5'H* genes, ctg1305.g11051, ctg2839.g24678 and ctg17.g00099 were downregulated; ctg287.g02877 of *F3'H* was upregulated; ctg1578.g14126 of *DFR* was upregulated; ctg1760.g15871 of *LAR* was upregulated; and ctg1306.g11055 and ctg1210.g10432 of *UFGT/3GT* were downregulated.

In the DEGs of LR2 vs. BR2, ctg836.g07174 was downregulated in *C4L*; ctg1147.g09820 of *F3H* was downregulated; ctg2495.g22288 of *F3'H* was downregulated; and ctg2661.g23571, ctg1652.g14646, ctg1306.g11055, ctg2661.g23, and ctg2055.g18534 of *UFGT/3GT* were all downregulated.

In the DEGs of LR3 vs. BR3, ctg836.g07174 of *C4L* was downregulated; ctg965.g08335 and ctg502.g04473 of *CHI* were all upregulated; and ctg1305.g11051, ctg2839.g24678 and ctg17.g00099 of *F3'5'H* were all upregulated; ctg1578.g14126 of *DFR* was upregulated; ctg1560.g13893 of *FLS* was upregulated; ctg1210.g10432 of *UFGT/3GT* was upregulated.

Among the DEGs of LR4 vs. BR4, only ctg1147.g09820 of *F3H* was downregulated.

In the DEGs of LR5 vs. BR5, ctg2456.g21765 of *PAL* was downregulated; ctg2580.g23157 and ctg836.g07174 of *C4L* were downregulated; ctg1147.g09820 of *F3H* was downregulated; and ctg1305.g11051 and ctg2135.g19559 of *F3'5'H* were upregulated; ctg2495.g22288 of *F3'H* was downregulated; for *UFGT/3GT*, ctg1496.g13357 and ctg1652.g14649 were both upregulated, and ctg1306.g11055 was downregulated.

Transcription factors associated with the anthocyanin biosynthesis pathway. Structural genes involved in anthocyanin biosynthesis are transcriptionally regulated by the MBW (MYB-bHLH-WD40) complex^{30,32}. The transcription factors that were related to the regulation of flavonoid and anthocyanin biosynthetic pathways in this study included *MYB* (ctg100.g00957-0F, ctg1410.g12339-1F, ctg1554.g13859-0F,

ctg1784.g16146-2F, ctg2470.g22007-0F, ctg2470.g22008-0F, ctg2470.g22009-0F, ctg842.g07305-0F, ctg911.g07979-0F, ctg1317.g11116-0F), and *MYB-related* (ctg1226.g10502-0F, ctg2353.g20962-0F). We speculate that these 12 MYB/MYB-related transcription factors are involved in the regulation of anthocyanin synthesis in the pulp of *B. ramiflora*.

qRT-PCR gene expression analysis.

To verify the sequencing results of the transcriptome of the *B. ramiflora* pulp transcriptome, 8 DEGs related to anthocyanin synthesis pathways (*CHS*: ctg655.g05350; *CHI*: ctg965.g08335; *F3'5'H*: ctg2839.g24678 and ctg2135.g19559; *UFGT*: ctg1923.g16986; *UFGT89B2*: ctg2170.g19838; and *UFGT88F3*: ctg2661.g23571) were selected for qRT-PCR analysis. As shown in Fig. 6, the expression levels of the DEGs in qRT-PCR were similar to those of the RNA-seq data, indicating that the RNA-seq data were reliable. This ensured the DEGs analysis and enrichment analysis of the transcriptome data in the later stage.

Discussion

Anthocyanins, pivotal metabolites within the flavonoid biosynthetic network, play a crucial role in dictating the chromaticity of fruits. These compounds, as secondary metabolites within the flavonoid schema, are integral to the nuanced palette of pigmentation³³. The inherent nature of the flavonoid precursor is a determinant for the resultant anthocyanin derivatives. The enzymatic conversion of dihydroflavonols to their respective forms—dihydrokaempferol, dihydroquercetin, and dihydromyricetin—is facilitated by *F3'5'H*. Subsequently, leucoanthocyanidins are synthesized through the action of *DFR*; *ANS* catalyzes the formation of anthocyanins; and, ultimately, *UFGT* mediates the synthesis of anthocyanins³⁴.

In the developmental trajectory of *B. ramiflora*, the flavonoid and anthocyanin biosynthetic pathway metabolites—naringin chalcone, naringenin, eriodictyol, dihydroquercetin, kaempferol, and quercetin—exhibit heightened concentrations during the initial two stages of maturation. A marked decline in these compounds commences at 73 DAF during the tertiary stage of development, suggesting that the biosynthesis of flavonoids and anthocyanins in *B. ramiflora*'s fruit pulp predominantly occurs in the early phases of fruit maturation. In the anthocyanin synthesis pathway, leucocyanidins undergo transformation into catechin/flavan-3-ol via the enzymatic activity of *LAR*, and anthocyanins are converted into epicatechin/epiflavan-3-ol by *ANR*, while procyanidins are synthesized through the polymerization of catechin or epicatechin. The concentrations of procyanidin B1, (+)-catechin, and (-)-epicatechin were notably elevated during the first two stages of development and began to diminish precipitously at 73 DAF, indicative of a substantial conversion of anthocyanins to epicatechin and a consequent significant reduction in anthocyanin accumulation.

Cyanidins bestow upon an array of fruits, vegetables, and edible flowers their characteristic pink or red hues, as observed in strawberries, red currants, potatoes, chicories, and onions³⁵. The absence of

specific anthocyanin components in the non-targeted metabolome analysis may be attributed to the low anthocyanin content within the pulp, elucidating why the pink pigmentation in *B. ramiflora* pulp is localized to the endocarp rather than the mesocarp. Dihydroquercetin serves as a precursor in cyanidin synthesis. The profusion of procyanidin B1, (+)-catechin, and (-)-epicatechin—each a branch of the cyanidin synthesis pathway—alongside the absence of other branches and related metabolites, suggests that the pink hue of the endocarp is determined by cyanin.

Integrating metabolome analysis, the expression profiles of structural genes implicated in the anthocyanin biosynthesis pathway—*CHI* (ctg502.g04473), *FLS* (ctg1560.g13893), and *LAR* (ctg1760.g15871)—correlate with the levels of their downstream metabolites: naringenin, quercetin, and (-)-epicatechin, respectively. We postulate that these genes are instrumental in the pigmentation of *B. ramiflora*'s fruit pulp.

Flavonoids predominantly accumulate during the initial two stages of fruit development. Contrasting with LR, the BR variant exhibits a pink pulp, indicative of an upsurge in the expression of numerous structural genes related to anthocyanin synthesis early in BR's development, culminating in a greater accumulation of flavonoids and anthocyanins. In the comparative analysis of BR versus LR, genes such as *F3'5'H* (ctg1305.g11051, ctg2839.g24678, and ctg17.g00099) and *UFGT* (ctg1210.g10432) are upregulated during the first developmental stage. The expression of these genes in BR is pronounced during the initial two stages and diminishes to negligible levels in the subsequent stages, underscoring their pivotal role in the differentiation of pulp coloration between BR and LR. The upregulated differentially expressed genes (*DEGs*) in the early stages include the early-stage gene *C4L* (ctg836.g07174), the mid-stage gene *F3H* (ctg1147.g09820), and the late-stage gene *UFGT* (ctg1306.g11055), denoting their involvement in flavonoid and anthocyanin synthesis. Notably, the expression of *F3H* (ctg1147.g09820) throughout BR's developmental period is consistently higher than in LR, further supporting *F3H*'s role as a key gene in the anthocyanin synthesis pathway and as a determinant of the color variation between BR and LR.

LDOX and *ANS* are functionally analogous, both facilitating the conversion of leucoanthocyanins to anthocyanins^{31–36}. Although the *ANS* gene was not delineated in the GO enrichment or KEGG pathway enrichment analyses of the anthocyanin biosynthetic pathway in *B. ramiflora*'s pulp, two *LDOX* transcripts (ctg438.g03741 and ctg3056.g26143) with functional similarity to *ANS* were annotated, intimating that *B. ramiflora* may have evolved distinct gene expression peculiarities. The pronounced expression of *ANR* in the early stages, surpassing that of *UFGT*, enables it to more efficiently catalyze the synthesis of (-)-epicatechin from anthocyanin substrates, thereby precipitating a swift reduction in anthocyanin content. This provides a plausible explanation for the non-detection of corresponding anthocyanins in the non-targeted metabolome analysis. The comparative expression analysis between LR1 and BR1 reveals heightened expression of *ANR* (ctg2699.g23958) and a significant decrease in *UFGT* (ctg1306.g11055 and ctg1210.g10432) expression, elucidating the lower anthocyanin metabolism in LR, which coherently accounts for the pink pigmentation of BR fruits.

In the anthocyanin synthesis pathway, leucocyanidin is transformed into catechin/flavan-3-ol by *LAR*, and anthocyanin is converted into epicatechin/epicatechin-3-ol by *ANR*, with procyanidin being synthesized from the polymerization of catechin or epicatechin. The elevated expression of *LAR* (ctg2548.g22735) and *ANR* (ctg2699.g23958) is instrumental to procyanidin synthesis.

Elucidating the molecular regulatory mechanisms underpinning the pigment accumulation variance in *B. ramiflora* pulp across different developmental stages is pivotal for the selection and breeding of superior cultivars. A profound comprehension of these biochemical pathways can expedite the exploitation of *B. ramiflora*, fostering the selection, cultivation, and dissemination of its varieties, and providing scientific substantiation for the conservation, development, and utilization of *B. ramiflora*'s germplasm diversity.

Materials and Methods

Plant material.

Accessions of *B. ramiflora* were obtained from the Fangcheng District in Fangchenggang city, located in the Guangxi region (N 21°42'33", E 108°6'29", alt 20 m). This collection includes two indigenous strains, which are germplasm of significant value that have been carefully curated and conserved by the local villagers for generations and are now widely propagated throughout the area. Authentication of the samples was conducted by the Plant Research Laboratory at Hanshan Normal University. All germplasm accessions were acquired in strict accordance with national and international guidelines for genetic resource conservation. The collection process was conducted under the guidance and with the authorization of Hanshan Normal University, ensuring that all authors adhered to the pertinent local and national regulations.

The two local strains of *B. ramiflora* were tested: BR, whose fruit peel is milk-white at maturity and whose pulp (endocarp) is pink, and LR, whose fruit peel is light green at maturity and whose pulp (endocarp) is milk-white. The taste of BR is relatively sweet compared with that of LR. The trees were more than 10 years old, were growing healthily, and were in the full fruiting period. This study encompassed the determination of fruit samples from two *B. ramiflora* pulp types (BR and LR). For each type, three trees exemplifying robust health and uniform development were selected as the source for sampling. Sampling was conducted at five distinct maturation stages, specifically at 30, 52, 73, 93, and 112 days post-anthesis. At each maturation stage, four fruits exhibiting consistent maturity and good condition were harvested from each tree, ensuring uniformity and comparability of the samples. This protocol yielded 12 samples per cultivar at each maturation stage, culminating in a total of 120 samples for the entire study. After field collection, the samples were immediately dipped in liquid nitrogen and stored in a -80°C freezer for later use, subsequently utilized for metabolomic and transcriptomic sequencing analyses. The voucher specimen of the fruits was deposited in the Plant Research Laboratory of Hanshan Normal University, Chaozhou, Guangdong, China (No: BR01-BR60, LR01-LR60).

Identification of flavonoids by UPLC–MS/MS.

Ultraperformance liquid chromatography–tandem mass spectrometry (UPLC-MS/MS) technology was done for nontargeted metabolomic analysis of the fruit pulp of *B. ramiflora*, and the metabolites were statistically analysed, including metabolite classification annotation and functional annotation, using the metaX package of R software³⁷, which was independently developed by BerryGenomics. Principal component analysis (PCA) was used to reduce the dimensionality of the multivariate raw data to analyse the intragroup and intergroup similarities and differences and the outliers (whether there were abnormal samples). The partial least squares–discriminant analysis (PLS-DA) model³⁸ was used to calculate the variable importance of projection (VIP) of the two principal components. VIP can measure the effect strength and explanatory power of the expression pattern of each metabolite on the classification of samples and can help with the screening of metabolic markers. The data were first subjected to log₂ transformation, and then the PLS-DA model was established. The scaling method was Par; seven interactive verifications were performed; and 200 response permutation tests were performed to determine the quality of the PLS-DA model. Differential metabolites were screened by considering the fold change (FC) obtained by univariate analysis and the results of Student's t test. Both PCA and FC were subjected to log₂ treatment, and the screening criteria were $p < 0.05$, $VIP \geq 1$, and $FC \geq 1.2$ or $FC \leq 0.83$ ⁶.

RNA library construction and sequencing.

Total RNA was extracted using the RNA extraction kit by BerryGenomics (RS-122-2001 (30 samples) (Set A: 12 indices)). RNA-seq library construction and sequencing included total RNA sample extraction, sample detection, RNA library construction, library quality inspection, and sequencing. Pulp mRNA was enriched by magnetic oligo(dT) beads; fragmentation buffer was added to break the mRNA into short fragments of 200 nt; and these mRNAs were used as the template for reverse transcription into first-strand cDNA, primed by random hexamers. Next, buffer, dNTPs, RNase H, and DNA polymerase I were added to replicate and synthesize the second cDNA strand. The double-stranded cDNA was then subjected to end repair with addition of base A at the 3' end. After purification with the QIAQuick PCR kit and elution with EB buffer, the DNA was end-repaired and ligated with sequencing adapters. Then the library was subjected to fragment size selection by agarose gel electrophoresis, and PCR amplification was performed to enrich the cDNA.

In this study, a total of 30 libraries of *B. ramiflora* pulp were constructed, 15 each for BR and LR, which were three biological replicate libraries for each of 30 days after full flowering (DAF), 52 DAF, 73 DAF, 93 DAF, and 112 DAF. After library construction was completed, the nucleic acid was preliminarily quantified using the Qubit 2.0 fluorescent reagent, and then the insert size of the library was detected in the Agilent 2100 analyser. If the insert size was in line with expectations, the effective concentration of the library was accurately quantified using the real-time fluorescence quantitative PCR method to ensure the quality of the library for sequencing the transcriptome of the fruit pulp. After the library was qualified, PE150 sequencing was performed on the Illumina NovaSeq platform. The library construction and sequencing were entrusted to BerryGenomics (Beijing, China).

The cDNA library was sequenced on the Illumina NovaSeq high-throughput sequencing platform, and many high-quality reads, which are called raw data, were produced. The raw data were filtered by FASTQC. For paired-end sequencing, the raw data of each sequencing sample included two FASTQ files, which contained the reads measured at both ends of all cDNA fragments. The raw data were subjected to rigorous quality assessment, and high-quality clean reads were obtained for subsequent data analysis. Using the whole genome sequenced from the third generation of *B. ramiflora* obtained in this study as the reference genome, bioinformatic analysis was performed, which mainly focussed on the gene expression, differentially expressed genes (DEGs), and transcription factors in the tissues of various developmental stages of the pulp and the Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis of DEGs.

The raw data obtained from the sequencing contained some reads with adapters, duplicates, and low quality, which would affect the subsequent analysis and comparison. First, we filtered the raw data. The screening criteria were as follows: the paired reads with > 3 single-end sequencing reads were excluded; paired reads with quality values below 5 and base ratios $\geq 20\%$ in single-end sequencing reads were excluded; when removing the adapter sequence, the adapter sequence must match at least 8 bp. Therefore, clean and effective high-quality clean reads can be obtained for subsequent statistical analysis. The quality assessment was based on the quality values Q20 and Q30, which measure the accuracy. Generally, the higher the Q20 and Q30 (the basic requirement is $\geq 85\%$), the higher the quality of the clean reads after sequencing.

Analysis of DEGs.

The DEGs in each sample were analysed using EdgeR software³⁹, and the *p* value and adjusted *p* value (padj) value of the up- or downregulation were calculated. padj is the corrected p-value, and the smaller its value, the more significant the gene expression difference. The screening criteria were as follows: padj < 0.05 and $|\log_2FC| > 1$; if the number of DEGs was too small, the parameters were adjusted to $p < 0.05$ and $|\log_2FC| > 1$ for screening. The input data for gene differential expression analysis were the read count data obtained from the quantitative analysis of genes, which was mainly divided into three parts: first, the read count data were standardized; then, the hypothesis test on the probability (p value) was calculated according to the model; and finally, the multiple hypothesis testing correction was performed to obtain the padj value (false detection rate).

The corrected DEGs were input into enrichment analysis. topGO software was used for the GO enrichment analysis of the DEGs⁴⁰, and KOBAS software was used for the KEGG pathway enrichment analysis. A *p* value of the GO and KEGG pathways < 0.05 indicated significant enrichment, which was used to analyse the relationship between DEGs and the related biological functions. The enrichment pathways can provide a better understanding of the signalling pathways of genes involved in metabolism and their specific biological functions.

Transcription factor analysis.

Transcription factors regulate the binding of RNA polymerase to DNA templates. Through the analysis of transcription factors, we can comprehensively understand the transcription process of genes. Transcription factor prediction was analysed using iTAK software. According to the Pfam file of the transcription factor family, Pfam was annotated, and transcription factors were identified by Hmmscan⁴¹.

qRT-PCR validation.

To verify the accuracy of RNA-seq in each sample of fruit pulp, 8 DEGs (including the structural genes of the flavonoid biosynthesis pathway, the structural genes of the glucose metabolism pathway, and the transcription factors) were selected for fluorescence real-time quantitative PCR analysis. The first-strand cDNA was synthesized by the reverse transcriptase Servicebio®RT First Strand cDNA Synthesis Kit (Servicebio G3330). The cDNA obtained by reverse transcription was amplified by PCR. The internal reference gene was the GAPDH gene. The primers for the DEGs were designed using Primer 6.0 (Table S1).

Statement on Permissions and Appropriate Licenses for Plant Specimen Collection

This study adheres to the relevant institutional, national, and international guidelines and legislation governing experimental research and field studies on plants, including the collection of plant materials. The authors, Jianjian Huang and Jie Chen, acquired accessions of *B. ramiflora* from the Fangcheng District in Fangchenggang City, located within the Guangxi region (N 21°42'33", E 108°6'29", altitude 20 meters). This collection comprises two native strains, representing valuable germplasm that has been meticulously maintained and conserved by local villagers for generations and is now extensively propagated across the region. Authentication of the samples was conducted by Jianjian Huang and Qinhan Wu in the Plant Research Laboratory at Hanshan Normal University. The collection license has been obtained and is uploaded as supplementary material to the journal system. Voucher specimens are securely stored in the Plant Research Laboratory at Hanshan Normal University (Nos: BR01-BR60, LR01-LR60).

Conclusions

Cyanin, the anthocyanin responsible for imparting the resplendent pink hue to the fruit pulp of *B. ramiflora*, is synthesized through a complex pathway involving a cadre of pivotal genes. Within this biosynthetic labyrinth, *CHI* (ctg502.g04473), *FLS* (ctg1560.g13893), *C4L* (ctg836.g07174), *F3H* (ctg1147.g09820), coupled with the *F3'5'H* isoforms (ctg1305.g11051, ctg2839.g24678, and ctg17.g00099), and *UFGT* (ctg1210.g10432 and ctg1306.g11055) emerge as quintessential for the orchestration of anthocyanin assembly within the pulpy matrix of *B. ramiflora*. Notably, the *F3'5'H* ensemble (ctg1305.g11051, ctg2839.g24678, and ctg17.g00099) and *UFGT* (ctg1210.g10432) represent the genetic cornerstone distinguishing the chromatic divergence between the BR and LR variants.

Concurrently, *LAR* (ctg1760.g15871) and *ANR* (ctg2699.g23958) are recognized as the master regulators in the synthesis of procyanidins.

The intricate molecular regulatory mechanisms dictating pigment accretion within the pulp of *B. ramiflora*, discernible across its developmental continuum, are of paramount significance in the discernment and cultivation of superior cultivars. An enriched understanding of these biochemical pathways holds the promise of enhancing the valorization of *B. ramiflora*, expediting the selection, domestication, and propagation of its varieties, and furnishing scientific underpinnings for the conservation, advancement, and judicious exploitation of the *B. ramiflora* germplasm spectrum.

Declarations

Data Availability

Genomic and transcriptome data has been uploaded to NCBI and will not be published until the paper is published. Please reference PRJNA845134 (Genome) and PRJNA848621 (Transcriptome) in your publication. This manuscript includes all data generated or analyzed during this study. Other necessary data of this study are available with the corresponding author on reasonable request.

Acknowledgments

We would like to thank editor and reviewers who revised this manuscript providing insightful comments. This research was supported by the projects of Guangdong Provincial Key Laboratory of Functional Substances in Medicinal Edible Resources and Healthcare Products (2021B1212040015), Forestry Science and Technology Innovation Project of Guangdong province (2018KJCX023), and the program for scientific research start-up funds of Guangdong Ocean University (060302052305).

Author Contributions

Conceptualization, J.J.H. and J.C.; methodology, J.J.H.; software, J.Q.Z., X.Y.W., S.Y.C., Y.C.Z.; validation, Y.Z.Z., Z.K.C. and Q.H.W.; formal analysis, J.J.H.; investigation, J.Q.Z.; resources, X.Y.W.; data curation, S.Y.C.; writing—original draft, J.J.H., F.N.W.; writing—review and editing, F.N.W.; visualization, Z.K.C.; supervision, J.C., F.N.W.; project administration, J.J.H., H.Z., Y.Z.Z., F.N.W.; funding acquisition, H.Z., Y.Z.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Supplementary Materials

The online version contains supplementary material available at....

References

1. Abrahams, S. *et al.* The *Arabidopsis* TDS4 gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *T. Plant J.* 35, 624–636 (2003).
2. Alexa, A., Rahnenfuhrer, J. TopGO: enrichment analysis for Gene Ontology, R package version 2.36.0 (2019).
3. Barker, M., Rayens, W. Partial least squares for discrimination. *J. Chemometr.* 17, 166–173 (2003).
4. Basumatary, A. *et al.* Bamboo shoots as a nutritive boon for Northeast India: an overview. *3 Biotech.* 7, 169 (2017).
5. Baudry, A. *et al.* TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380 (2004).
6. Cassidy, A. *et al.* High anthocyanin intake is associated with a reduced risk of myocardial infarction in young and middle-aged women. *Circulation* 127, 188–196 (2013).
7. Chen, J. *et al.* Identification of key taste components in *Baccaurea ramiflora* Lour. fruit using non-targeted metabolomics. *Food Sci. Hum. Well.* 12, 94–101 (2023).
8. Goyal, A. K., Middha, S. K., Talambedu, U. *Baccaurea ramiflora* Lour.: a comprehensive review from traditional usage to pharmacological evidence. *Adv. Tradit. Med.* 22, 231–249 (2020).
9. Goyal, A. K., Mishra, T., Sen, A. Antioxidant profiling of Latkan (*Baccaurea ramiflora* Lour.) wine. *Indian J. Biotechnol.* 12, 137–139 (2013).
10. He, J., Giusti, M. Anthocyanins: natural colorants with health-promoting properties. *Annu. Rev. Food Sci. Technol.* 1, 163–187 (2010).
11. Hossain, M. F., Islam, M. A., Akhtar, S., Anwar, M. Nutritional value and medicinal uses of minor fruits: Burmese Grape (*Baccaurea ramiflora* Lour.). *Int J Nutr Food Sci.* 6, 211–214 (2017).
12. Hu, J. X., Xiao, C. F., Zheng, L. L. Wild fruit tree - *Baccaurea ramiflora*. *S. China Fruits.* 32, 49 (2003).
13. Jaakola, L. New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci.* 18, 477–483 (2013).
14. Jun, J. H., Xiao, X. R., Rao, X. L., Dixon, R. A. Proanthocyanidin subunit composition determined by functionally diverged dioxygenases. *Nat. Plants.* 4, 1034–1043 (2018).
15. 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.* 9, e86293 (2014).
16. Li, B. T., Michael, G. G. Flora of China. Beijing: Science Press 11, 216–217 (2008).
17. Li, W. X. *et al.* Determination of soluble saccharide contents of *Baccaurea ramiflora* Lour. using anthrone colorimetric method. *Chin. Hortic. Abstr.* 12, 7–8 (2015).
18. Liu, C. C. *et al.* The bZip transcription factor HY5 mediates CRY1a-induced anthocyanin biosynthesis in tomato. *Plant Cell Environ.* 41, 1762–1775 (2018).

19. Liu, Y., Li, D., Zhang, Y. H., Sun, R. F., Xia, M. Anthocyanin increases adiponectin secretion and protects against diabetes-related endothelial dysfunction. *Am. J. Physiol-End. M.* 306, 975–988 (2014).
20. Liu, Y. *et al.* Anthocyanin biosynthesis and degradation mechanisms in *Solanaceous* vegetables: A Review. *Front Chem.* 6, 52 (2018).
21. Lloyd, A. *et al.* Advances in the MYB–bHLH–WD Repeat (MBW) pigment regulatory model: addition of a WRKY Factor and Co-option of an anthocyanin MYB for betalain regulation. *Plant Cell Physiol.* 58, 1431–1441 (2017).
22. Nesa, M. L. *et al.* Screening of *Baccaurea ramiflora* (Lour.) extracts for cytotoxic, analgesic, antiinflammatory, neuropharmacological and antidiarrheal activities. *BMC Complem. Altern. M.* 18, 35 (2018).
23. Pandey, Y. *et al.* Nutritional compositions of *Baccaurea sapida* and *Eleaocarpus sikkimnesis* of Sikkim Himalaya. *Int. J. Curr. Microbiol. App. Sci.* 7, 2101–2106 (2018).
24. Paulino, P. R. *et al.* PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res.* 38, D822–827 (2010).
25. Pojer, E., Mattivi, F., Johnson, D., Stockley, C. S. The case for anthocyanin consumption to promote human health: a review. *Compr Rev Food Sci. F.* 12, 483–508 (2013).
26. Qian, M. J. *et al.* Isolation and expression analysis of anthocyanin biosynthesis genes from the red Chinese sand pear, *Pyrus pyrifolia Nakai cv. Mantianhong*, in response to methyl jasmonate treatment and UV-B/VIS conditions. *Plant Mol. Biol. Rep.* 32, 428–437 (2013).
27. Qin, Y. *et al.* Cyanidin-3-O-glucoside ameliorates diabetic nephropathy through regulation of glutathione pool. *Biomed. Pharmacother* 103, 1223–1230 (2018).
28. Ramsay, N. A., Glover, B. J. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci.* 10, 63–70 (2005).
29. Wang, H. The control of red color by a family of MYB transcription factors in octoploid strawberry (*Fragaria × ananassa*) fruits. *Plant Biotechnol. J.* 18, 1169–1184 (2020).
30. Wang, L. S., Stoner, G. D. Anthocyanins and their role in cancer prevention. *Cancer Lett.* 269, 281–290 (2008).
31. Wen, B., Mei, Z. L., Zeng, C. W., Liu, S. Q. MetaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics.* 18, 183 (2017).
32. Yang, Z. Q. *et al.* Development prospect and countermeasure of *Baccaurea ramiflora* in Guangxi, Chin. *Hortic. Abstr.* 30, 59 (2014).
33. Yi, D.B. Integrative analysis of the coloring mechanism of red longan pericarp through metabolome and transcriptome analyses. *J. Agr. Food Chem.* 69, 1806–1815 (2021).
34. Yuan, Y. W., Rebocho, A. B., Sagawa, J. M., Jr, H. D. B. Competition between anthocyanin and flavonol biosynthesis produces spatial pattern variation of floral pigments between *Mimulus* species. *Proc. Natl. Acad. Sci.* 113, 2448–2453 (2016).

35. Zhang, Y. J. *et al.* Metabolic and molecular analysis of nonuniform anthocyanin pigmentation in tomato fruit under high light. *Hortic. Res.* 6, 56 (2019).
36. Zhang, Y., Butelli, E., Martin, C. Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19, 81–90 (2014).
37. Shen, N. *et al.* Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* 383: 132531 (2022).
38. Smyth, G. K. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139 (2010).
39. Telias, A., Bradeen, J. M., Luby, J. J., Hoover, E. E., Allan, A. C. Regulation of anthocyanin accumulation in apple peel. *Hortic. Rev.* 38, 357–391 (2011).
40. Tena, N., Martín, J., Asuero, A. G. State of the art of anthocyanins: Antioxidant activity, sources, bioavailability, and therapeutic effect in human health. *Antioxidants* 9, 451 (2020).
41. Tsuda, T. Dietary anthocyanin-rich plants: Biochemical basis and recent progress in health benefits studies. *Mol. Nutr. Food Res.* 56, 159–170 (2012).

Figures



Figure 1

Five fruit profiles of the maturation process of *B. ramiflora* in LR and BR. From left to right: 30 d, 52 d, 73 d, 93 d and 112 d after full bloom.

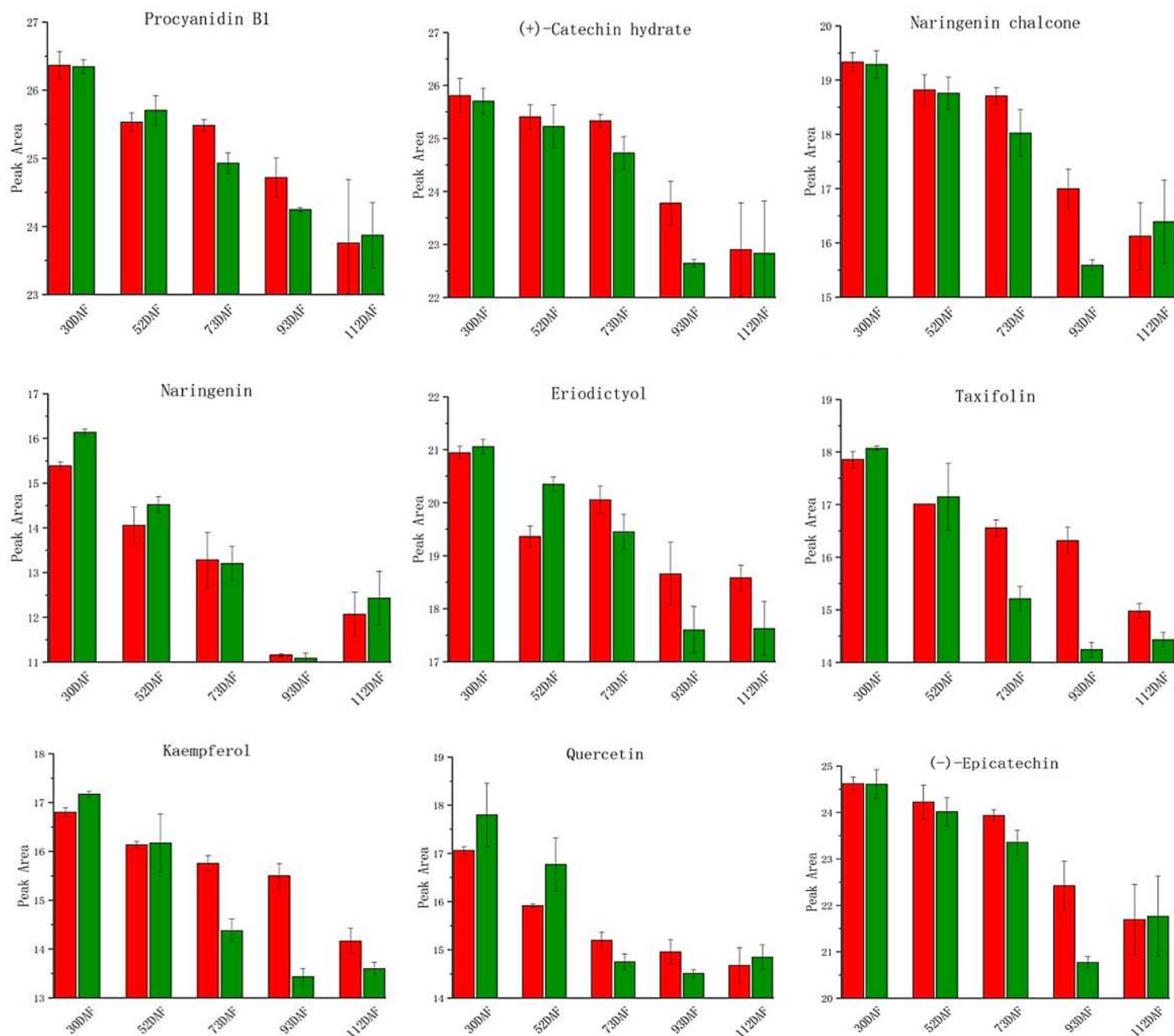


Figure 2

Nine flavonoids content of different developmental stages in *B. ramiflora* fruit. In the figure, the red bar graph represents LR (milky-white at maturity), while the green bar graph indicates BR (pink at maturity).

Pearson correlation coefficient

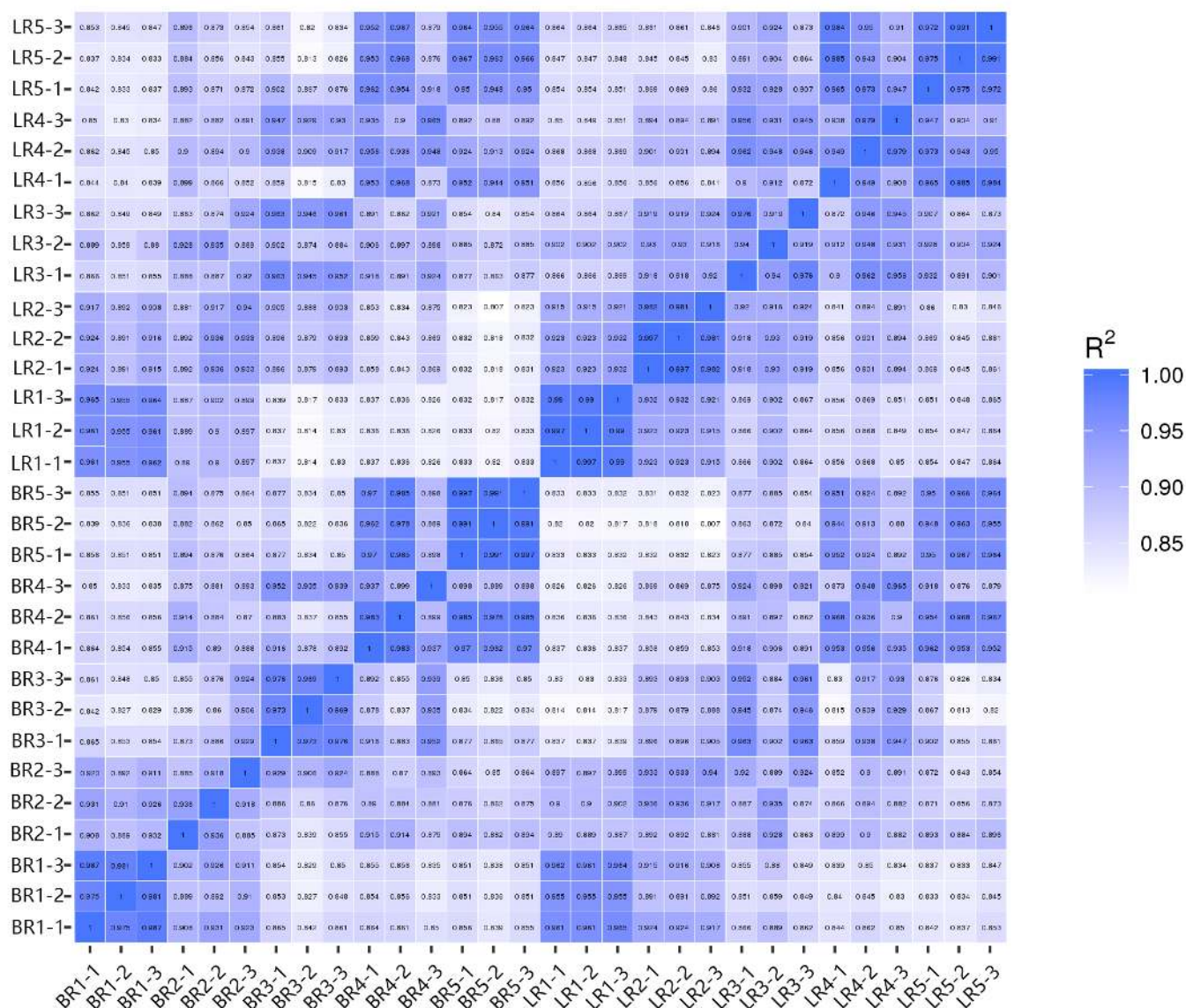


Figure 3

Heatmap of Pearson correlation analysis of gene expression among 30 pulp samples

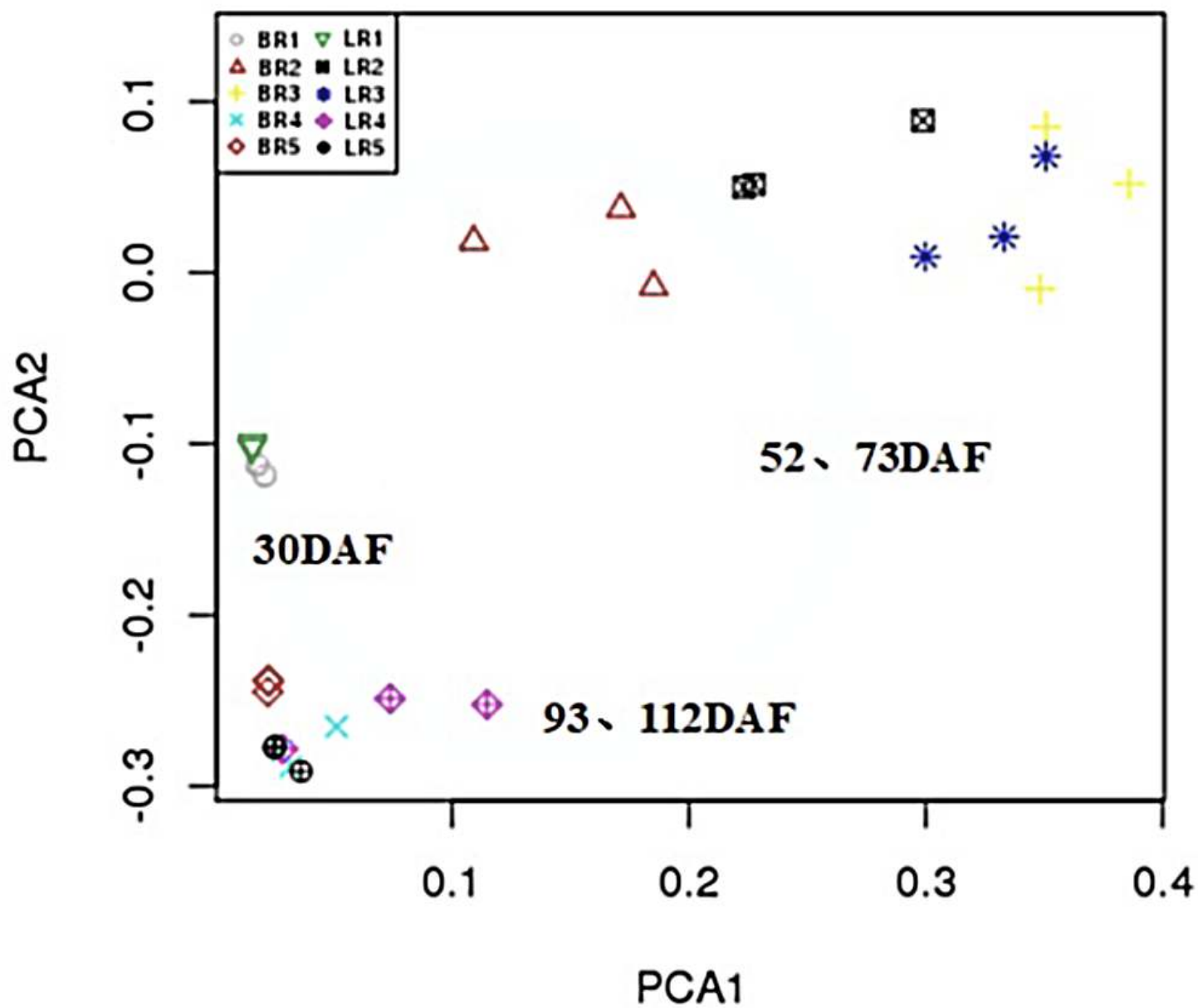


Figure 4

Principal component analysis (PCA) of 30 pulp samples (DAF: days after full flowering).

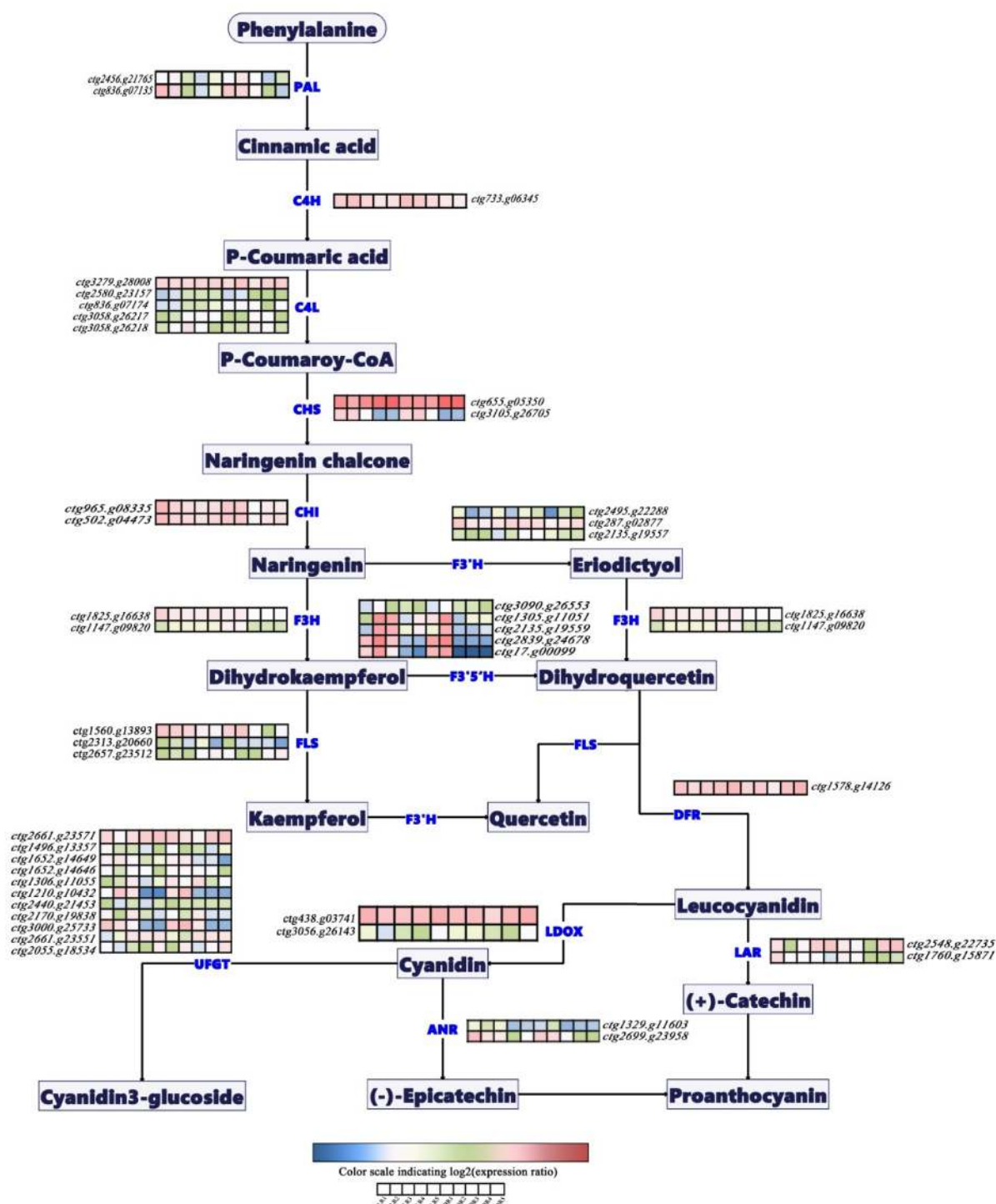


Figure 5

The expression levels of structural genes related to the anthocyanin biosynthetic pathway during pulp development of *B. ramiflora* fruit. The heatmap represents the expression levels of the listed structural genes in the two types of *B. ramiflora*. The colour changes from blue to red indicate that the gene expression levels are from low to high, and the same applies below.

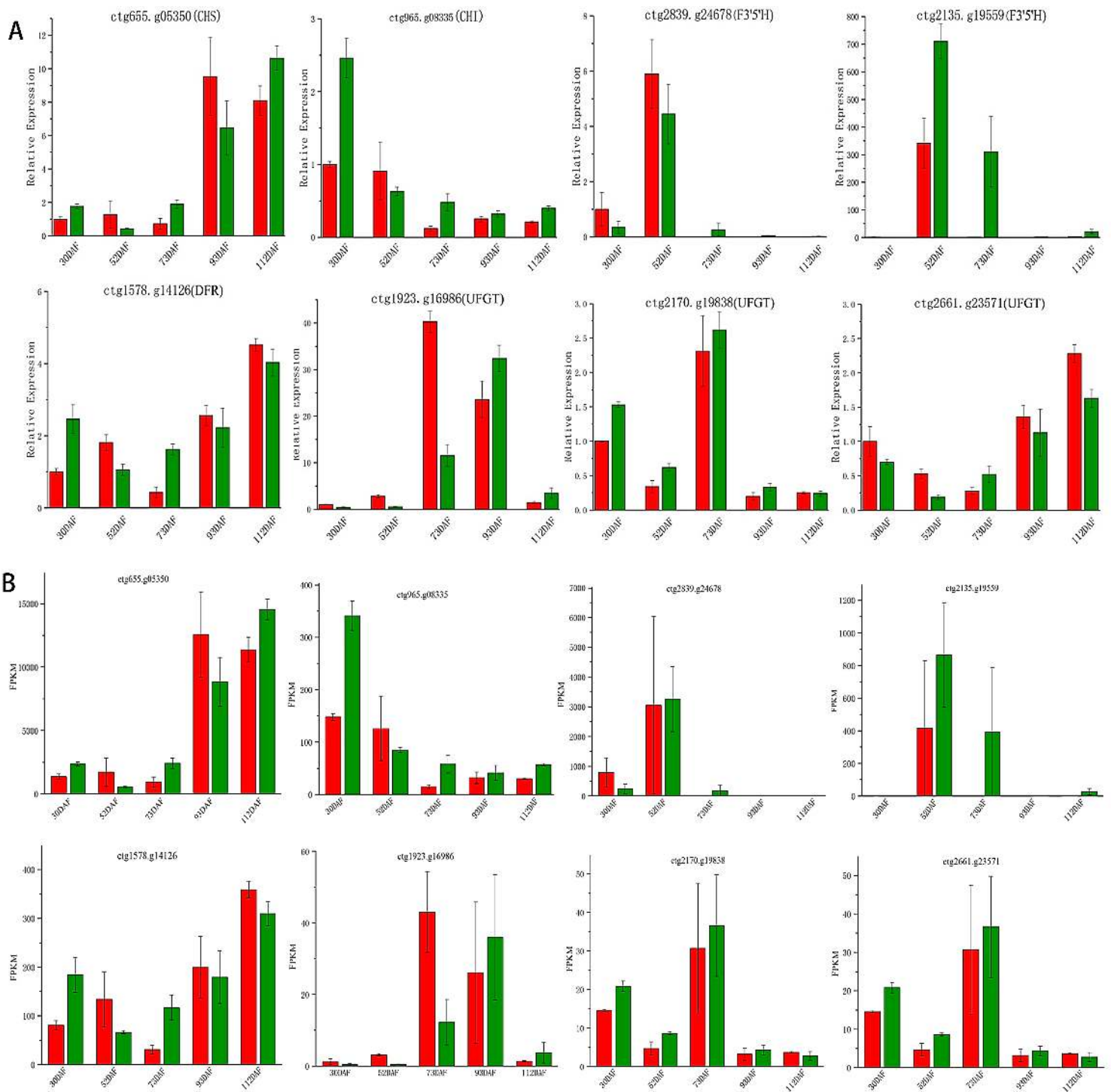


Figure 6

Expression analysis of 8 DEGs related to the anthocyanin biosynthesis pathway: (A) qRT-PCR validation analysis (B) RNA-seq expression quantity. In the figure, the red bar graph represents LR (milky-white at maturity), while the green bar graph indicates BR (pink at maturity).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTableS1andS3.pdf](#)
- [SupplementaryTableS2.xlsx](#)
- [SupplementaryTableS4.xlsx](#)