Integrated metabolomic and transcriptomic analysis of the anthocyanin regulatory networks in red walnut

Lei Wang
Henan Agricultural University
Lin Li
Henan Agricultural University
Wei Zhao
Henan Agricultural University
Haijun Meng
Henan Agricultural University
Ganggang Zhang
Henan Agricultural University
Wenjiang Wu
Henan Agricultural University
Jiangli Shi
Henan Agricultural University
Guoliang Wu (walnut-wu@126.com)
Henan Agricultural University

Research Article

Keywords: transcriptome, metabolome, anthocyanin, red walnut

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-907324/v1

Abstract

Background

Walnut is one of the most important dry fruit crops worldwide, typically green leaves and yellow-brown or gray-yellow seed coats. A specific walnut type, red walnut ‘RW-1’ with red leaves and seed coats was selected as plant material because of higher anthocyanins contents. Anthocyanins are important colorants with strong antioxidant activity, especially, benefic for human health. However, few studies focused on the molecular mechanism of anthocyanin biosynthesis in walnut.

Results

From the results of Widely Targeted Metabolome and anthocyanidin detection analysis, 395 substances, including 4 procyanidins and 26 anthocyanins, were identified from the red-leaf walnuts of RW-1 natural hybrid progenies (SR) and the green-leaf walnuts of RW-1 natural hybrid progenies (SG). Among these, all the anthocyanins in SR were significantly up-accumulated comparing with SG. Also, delphinidin 3-O-galactoside, cyanidin 3-O-galactoside, delphinidin 3-O-glucoside and cyanidin 3-O-glucoside were identified to the primary components of anthocyanidins because of the higher contents. It was noted that 9 anthocyanins including malvidin 3-O-galactoside, malvidin 3-O-arabinoside, cyanidin 3-O-(6-O-malonyl-beta-D-glucoside), delphinidin 3-O-glucoside, delphinidin 3,5-O-diglucoside (Delphin), peonidin 3-O-(6-O-malonyl-beta-D-glucoside), petunidin 3-O-(6-O-malonyl-beta-D-glucoside), petunidin 3-O-arabinoside and pelargonidin 3-O-(6-O-malonyl-beta-D-glucoside) were detected only in SR walnut. Furthermore, transcriptome analysis demonstrated that the expression of structural genes (*C4H*, *F3H*, *F3’5’H* and *UFGTs*), and four *MYBs* in anthocyanin biosynthetic pathway were significantly higher in SR walnut.

Conclusions

We identified the color formation of SR leaves is due to the accumulation of anthocyanins. And our results obtained the valuable information on the anthocyanin metabolites and candidate genes in anthocyanin biosynthesis, which provided new insights into the anthocyanin biosynthesis in walnuts.

Background

Anthocyanin is one of the most important color-presenting materials in flavonoids, mainly existing in leaves, flowers, fruits of higher plants, formed by the combination of anthocyan and glycosyl [1]. Cyanidin (Cy), delphinidin (Dp), pelargonidin (Pg), peonidin (Pn), petunidin (Pt), and malvidin (Mv) are six common anthocyans, and, common glycosyl includes glucose, galactose, sucrose, etc [2]. Different types and amounts of glycosyls bound to different positions of anthocyans resulted in significant increase of anthocyanins compounds [3]. Anthocyanins takes part in protecting plants from ultraviolet rays damage, attracting insect pollination and resisting to low temperature [4]. As a kind of natural food colorants,

anthocyanins have been widely used in food and cosmetics due to natural nontoxicity and no side effects [5]. Additionally, numerous studies have demonstrated that anthocyanins have many promising benefits for human health, such as the prevention of cardiovascular disease, neuron disease, cancer, diabetes and inflammation [6–10]. Therefore, anthocyanins have risen to fame for high antioxidant capacity, as well as improving human health and extracting natural food pigments.

The anthocyanin biosynthesis pathways have been well characterized in some plants [11]. Generally, phenylalanine is regarded as the initial of anthocyanin biosynthesis, which is composed of a series of enzymatic reactions, including phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), chalcone synthetase (CHS), chalcone isoenzyme (CHI), flavanone 3-hydroxylase (F3H), Flavonoid 3',5'-Hydroxylase (F3’5’H), dihydroflavonol 4-reductase (DFR), anthocyanin synthetase (ANS), UDP glucose: flavonoid-3-O-glucosyltransferase (UFGT) and so on [12, 13].

Meanwhile, the structural genes in anthocyanin biosynthesis are regulated by multiple transcription factors, among which MYB transcription factors have been widely studied. Grape R2R3-MYB TFs VvMYBA1, VvMYBA6 and VvMYBA7 regulated anthocyanin biosynthesis by activating *UFGT* and *3AT* expressions [14]; The overexpression of *MdMYB3* increased the anthocyanin accumulation in tobaccos by activating the gene expressions of *CHS*, *CHI*, *UFGT* and *FLS* [15]; The overexpression of *GhMYB1a* in gerbera and tobacco (*Nicotiana tabacum*) decreased anthocyanin accumulation by upregulating the structural gene expressions of *NtCHS*, *NtF3H*, *NtDFR*, *NtANS*, and *NtUFGT* [16].

It was reported that MYB regulates anthocyanin biosynthesis by either alone or forming MBW complex (MYB-bHLH-WD40) [17, 18]. In MBW complex, bHLH and MYB transcription factors can specifically bind to the promoters of structural genes in anthocyanin biosynthesis pathway, while WD40 protein plays a stable role in MBW complex [17]. The regulatory model of MBW complex included a ternary complex and a binary complex composed of MYB and bHLH. The ternary complex usually directly regulates the transcription of structural genes, but not synthesize intermediate regulators [17]. For example, AcMYB123 and AcbHLH42 in *Actinidia chinensis* cv. Hongyang promote the anthocyanin accumulation by activating the promoters of *AcANS* and *Ac3FGT1* [19]. *PyMYB10* and *PyMYB114* co-transformed with *PybHLH3* induced the promoters activity of *PyDFR*, *PyANS*, *PyUFGT*, *PyGST* and *PyABC transporter* [20]. Therefore, the present study aimed to elucidate the regulation mechanism of main structural genes and transcription factors involving in the anthocyanin accumulation based on red and green walnuts.

Walnut (*Juglans regia* L.) is an ancient fruit tree from the *Juglandaceae* family, ranks the top four nuts in the world due to its higher economic value and its health benefits to humans because of rech in oleic acid, linoleic acid, linolenic acid, protein and many trace elements [21]. In addition, walnuts are also considered to be an important ecological tree. The walnut industry has been included as the core point of industrial poverty alleviation in many areas, and plays an important role in helping targeted poverty alleviation and increasing farmers’ income. According to FAO statistics (http://faostat3.fao.org), China's walnut planting area and yield accounted for 48% and 56% of the world, respectively, both ranking the first in the world in 2019. However, phenotype traits of walnut varieties are close, with green leaves and

yellow-brown or gray-yellow seed coat. Luckily, a red walnut accession ‘RW-1’ was found by our research group, with the red color leaf, pericarp, seed coat and xylem owing to high anthocyanin content [22], and several *bHLH* TFs and *CHS* structural genes were identified in anthocyanin biosynthesis [23, 24]. However, the molecular mechanism of anthocyanin biosynthesis has not yet been clearly elucidated in red walnut.

In recent years, the combination analysis of transcriptome and metabolome has widely used to clarify anthocyanin biosynthesis and accumulation in plants [25–31]. In the current study, the regulatory networks of anthocyanin biosynthesis in walnut were conducted using two ‘RW-1’ natural hybrid accessions, with red and green leaves respectively. The compared analysis of metabolomic and transcriptomics aimed to elucidate the pathway of anthocyanin metabolites and identify the different expressed genes (DEGs) in walnut anthocyanin biosynthesis.

Results

Widely Targeted Metabolome and anthocyanidin detection in red walnut natural hybrid progeny leaves over developmental stages

In order to keep the consistence of genetic background, the RW-1 natural hybrid progenies SG and SR walnuts with different leaves colors were investigated. The SR walnut leaves changed from red to green while the SG walnut leaves displayed green all the time (Fig. 1a). The metabolome of six samples were profiled by Widely Targeted Metabolome and anthocyanidin detection approach. 395 compounds in the walnut leaves were classified into 19 classes (Table S2), including 26 anthocyanins and 4 procyanidins.

Identification of the differentially accumulated metabolites between SR and SG walnut leaves

The PCA chart illustrates the evident separation of the metabolites in the SR and SG, showed that all the biological replicates clustered together, indicating the high reliability of the sequencing data, and the metabolites in SR walnut leaves and SG walnut leaves at different development stages were clearly distinguished (Fig. S2a, b). According to the correlation analysis of population samples, the correlation coefficient of red leaves at different developmental stages was lower, as well as the correlation coefficient between red leaves and green leaves, which suggested the large differences in metabolites from green and red leaves. Meanwhile, the higher correlation of three repetitions indicated that the results in the present study were stable and reliable (Fig. S3).

As Fig. 1b shown, 87, 20 and 36 DAMs among SG-1 vs SR-1、SG-2 vs SR-2、SG-3 vs SR-3 were identified between pair samples, respectively. And 11, 13 and 12 were only were detected only in SR-1, SR-2 and SR-3 comparing with the controls, respectively; and all of them were anthocyanins (Fig. 2). Furthermore, the volcano maps showed that 30 up-accumulated metabolites while 58 down-accumulated in the first stage (Fig. 1c), 13 up-accumulated while 7 down-accumulated in the second stage (Fig. 1d), and 22 up-accumulated while 14 down-accumulated in the third stages (Fig. 1e), respectively. From Fig. 1f-h, the

enrichment analysis of KEGG pathway displayed that the most significant enrichment pathways were anthocyanin biosynthesis, histidine metabolism and pure metabolism in the first period; anthocyanin biosynthesis, isoflavonoid biosynthesis, flavone and flavonol biosynthesis and flavonid biosynthesis in the second period; anthocyanin biosynthesis, pure metabolism and flavonoid biosynthesis in the third period. The above results indicated that the DAMs from the flavonoid biosynthesis pathway were likely to be the key metabolites resulting in the leaves color difference between green and red walnut types.

Components and contents of flavonoids and anthocyanins during walnut leaf development stages

In the work, a total of 88 flavonoid metabolites were identified, including 55 flavonoids, 3 isoflavonoids, 26 anthocyanins and 4 procyanidins.

Among the flavonoid, anthocyanins are the most important colorants in plants [32]. To determine whether the red pigmentation of red walnut is caused by anthocyanins, we analyzed the soluble anthocyanins in green and red leaves using a UPLC-ESI- MS/MS system. A total of 26 anthocyanins were detected in walnut leaves, among them, all the anthocyanins contents were significantly up-accumulated in SR-1 than in SG-1 (Fig. 2). Furthermore, delphinidin 3-O-galactoside, cyanidin 3-O-galactoside, delphinidin 3-O-glucoside and cyanidin 3-O-glucoside were the primary components of anthocyanidins. We also found that 9 anthocyanins compounds existed in only red leaves, including malvidin 3-O-galactoside, malvidin 3-O-arabinoside, cyanidin 3-O-(6-O-malonyl-beta-D-glucoside), delphinidin 3-O-glucoside, delphinidin 3,5-O-diglucoside (Delphin), peonidin 3-O-(6-O-malonyl-beta-D-glucoside), petunidin 3-O-(6-O-malonyl-beta-D-glucoside), petunidin 3-O-arabinoside and pelargonidin 3-O-(6-O-malonyl-beta-D-glucoside).

Moreover, 8 down-accumulated flavonoids contents were found to be significantly different in SG-1 vs SR-1, 2 down-accumulated and 1 up-accumulated in second period, while 2 down-accumulated and 3 up-accumulated in third period (Fig. S4). And 3 procyanidins were also showed up-accumulated in SR-1 than in SG-1, while 2 down-accumulated and 1 up-accumulated in second period, and 3 up-accumulated in third period (Fig. S5). In a word, the difference in anthocyanins compounds accumulation was corresponding to that in the leaf colors.

Transcriptome profiles of green- and red- leaves walnuts

RNA-sequencing (RNA-Seq) was used to profile genome-wide gene expression and transcriptome changes during the leaf development. With three biological replicates, the transcriptome sequencing of the 18 walnut leaf samples yielded a total of 134.01 Gb clean data with 95.22% of bases scoring Q30. Furthermore, 92.23–95.27% of the total clean reads were unique matches with the walnut reference genome [23, 24, 33], 4,131 novel genes were identified, including 3,154 annotated genes.

Differentially expressed genes between green- and red-leaves walnuts

From Fig. 3a, 5,708, 1,587 and 703 DEGs were obtained from SG-1 vs SR-1, SG-2 vs SR-2, and SG-3 vs SR-3, respectively, containing 24 identical genes.

To clear the differences in DEGs between SG and SR walnuts, the volcano maps were performed during the leaf developments of walnut. According to the results of transcriptomic analysis, there were 5,708 DEGs between SG-1 vs SR-1, of which 3,513 were up-regulated and 2,195 were down-regulated. There were 1,587 DEGs between SG-2 vs SR-2, including 810 up-regulated genes and 777 down-regulated genes. There were 703 DEGs between SG-3 vs SR-3, including 364 up-regulated genes and 339 down-regulated genes (Fig. 3b-d). The profiles of the DEGs indicated that the differences in genes expression affected leaves color in walnuts.

Weighted gene correlation network analysis (WGCNA)provides insights into the coloration mechanism in red walnut

To identify potential molecular mechanisms of anthocyanin accumulation in SR, WGCNA was used to identify the characteristic genes of the module and the key genes in the module. By WGCNA analysis, the obtained DEGs were divided into 10 modules according to highly correlated genes (FPKM values) from SR and SG (Fig. 4a). The results highlighted genes in the turquoise and lightcyan modules, with significant difference of FPKM values in SR-1 than SG-1 (Fig. 4b). However, turquoise module exhibited a stronger positive relationship with SR-1 than SG-1, and lightcyan module exhibited negative trend. KEGG enrichment analysis based on the modules revealed that 12 genes might be related to the biosynthesis of anthocyanin of SR walnut leaves in the turquoise module, and only one in lightcyan module (Fig. 4c-d, Table S3).

Analysis of structural genes involved in anthocyanin biosynthesis.

According to the KEGG annotation analysis of all DEGs, 18 DEGs were involved in anthocyanin biosynthesis pathway, of which 17 DEGs were existed in SG-1 vs SR-1. In view of leaves colors difference at Stage 1 on appearance, the combination analysis of transcriptomic and metabolomic data were to explain the anthocyanin biosynthesis pathway in SG and SR walnuts (Fig. 5). It was found that structural genes including *C4H* genes (gene40343 and gene42522), *CHS* genes (gene4994, gene39336, gene35863 and gene32601) [24], *F3H* gene (gene40994), *F3'5'H* gene (gene4387), *ANS* gene (gene1297) and *UFGT* genes (gene28510, gene35146, gene1870, gene24302, gene35144, gene1697, gene36923 and) were all upregulated and only one *UFGT* gene (gene35048) was downregulated in red walnut leaves (Fig. 5).

Identification of TFs Related to Anthocyanin Biosynthesis

As we known, anthocyanin biosynthesis is regulated by MBW (MYB-bHLH-WD40) complex. From Fig. 6a, the number of MYB and bHLH ranked the top two of TFs related to anthocyanin biosynthesis in walnut leaves. Based on the Arabidopsis MYB protein domains, 135 putative walnut MYB protein sequences

were obtained with default parameters using HMMER and BLASTP (Table S4). A phylogenetic tree was constructed using the 135 JrMYBs and other species MYBs related to anthocyanin biosynthesis (Fig. 6b). Nine subfamilies and 25 JrMYBs were obtained from the current study and 4 differentially expressed *MYB* genes in red frames, *JrMYB1b (gene38312)*, *JrMYB6a (gene32351)*, *JrMYB123 (gene9445)*, *JrTT2 (gene39085)* were screened (Fig. 6c). The bHLHs in anthocyanin biosynthesis had been reported by our research group [23].

qPCR analysis of DEGs related to anthocyanin

Based on our above results, 12 differentially expressed structural genes involved in the anthocyanin biosynthetic pathway and 4 *MYB* genes were analyzed using qPCR methods. The results in Fig. 7 showed that eight structural genes *JrC4Hs* (gene40343, gene42522), *JrF3'5'H* (gene4387), *JrUFGTs* (gene35144, gene35146, gene1697, gene24302 and gene1870) and four *MYB* genes (gene38312, gene32351, gene9445, gene39085) were highly expressed in red leaves walnut at the first stage, which indicated that the high expression of genes involved in anthocyanin biosynthesis are related the color in red walnut leaves. It was also further proved that the transcriptome data were accurate and consistent with the expression of genes related to anthocyanin biosynthesis in red walnut.

Discussion

The primary pigments in red skin walnut have been identified as flavonoids particularly anthocyanins [22, 24], and the previous studies on walnut coloration were limited. The current work aimed to explore more comprehensive metabolites involved in the color changes in walnut leaves, and 395 metabolites were obtained from walnut leaves using the widely-targeted metabolomics approach (Table S2). This was the first to present the genome-wide examination of anthocyanins and the gene expression profiles of in walnuts, aiming at providing a more comprehensive landscape of the metabolites involved in the color changes in walnut leaves during development. The transcriptomic and metabolic analysis demonstrated that 17 core genes were identified in the anthocyanin biosynthesis (Fig. 5). These findings provided a theoretical basis for the further study the mechanism of red walnut.

Anthocyanins are secondary metabolites in plants, such as ornamental plants, fruits, vegetables and medical plants, and play various roles in many biological processes including determining fruit quality and flower colors as well as improving resistance, avoiding UV and strong light damage [34, 35]. In addition to the rich nutrition in kernel, walnut is also an important ecological tree. Red walnut RW-1 possess the great ornamental value, nutritional value and economic benefits. A total of 26 anthocyanins were detected in green- and red- leaves walnuts (Fig. 2), and cyanidin 3-O-galactoside, delphinidin 3-O-galactoside, and delphinidin 3-O-glucoside were the main anthocyanin in SR. Importantly, the anthocyanins content was consistent with leaf color.

Interestingly, 6 anthocyanin compounds including cyanidin, delphinidin, malvidin, pelargonidin, peonidin, and petunidin were existed not only in SR but also in SG (Fig. 2). The similar results were also found in

white *Salvia miltiorrhiza* flowers [36]. The color depended on the types and contents of anthocyanins, Co-pigment, chlorophyll, vacuole pH, metal ions [37]. The 6 anthocyanins accumulated lower in green leaves than in red leaves walnuts, suggesting that complete anthocyanin metabolic pathways were also presented in SG. The other factors affecting the color of walnut leaves need to be further studied.

Anthocyanin biosynthesis involves multiple enzymes encoded by early biosynthesis genes (*C4H*, *CHS*, *CHI* and *F3H*) and anthocyanin-specific biosynthesis genes (*F3' H*, *F3'5'H*, *DFR*, *ANS* and *UFGT*) [37]. The results from this study, the expression levels of *C4Hs*, *F3'5'H* and *UFGTs* genes in SR were significantly higher than those in SG (Fig. 7), suggesting that the genes mainly regulated the red coloration in red leaves walnuts. As the leaf development of walnut, these structural gene are down-regulated, and the color was trending to green. As we known, *UFGT* gene was involved in the final steps on the flavonoid biosynthetic pathway (biosynthesis and accumulation of anthocyanins). The expression of *UFGT* genes (gene35144, gene35146, gene1697, gene24302 and gene1870) were higher in SR than in SG at Stage 1, and *UFGT* gene (gene35038) was the opposite. It was worth noting that the 3 *UFGT* genes trend during 3 stages were corresponding to the colour change in walnut leaves. The anthocyanidins are extremely unstable and easy to degrade, therefore glycosylation is important to stabilize. As the last step of anthocyanin biosynthesis pathway, UFGTs were considered as the key enzyme to control anthocyanin biosynthesis in many plants and plays an important role in anthocyanin metabolism [12, 38–40]. The present study first explored the expression profile of *UFGTs* using red walnuts, providing the new sights for study *UFGT* genes function.

Anthocyanin biosynthesis is regulated by MBW (MYB-bHLH-WD40) complex. In MBW complex, bHLH and MYB transcription factors have DNA binding function, which can specifically bind to the promoters of structural genes in anthocyanin biosynthesis pathway, while WD40 protein plays a stable role in MBW complex [17, 41]. *Actinidia chinensis* Planch AcMYB123 and AcbHLH42 promote anthocyanin accumulation in the inner pericarp of *Actinidia chinensis* Planch by activating the promoters of *AcANS* and *AcF3GT1*[19]; PyWRKY26 and PybHLH3 acted together on *PyMYB114* promoter, and co transfection can activate the expression of *PyUFGT* to regulate the accumulation and transportation of anthocyanin in pear [20]. In this study, we have identified five MYBs and four bHLHs based on their expression levels and phylogenetic analysis, how the TFs played roles in regulating the structural genes need further explored. These findings can help to elucidate the molecular mechanism and regulatory networks of anthocyanin biosynthesis in walnut and provide a biological basis for breeding new walnut cultivars.

Conclusions

In this study, the metabolomics and transcriptomics were used reveal the anthocyanin biosynthesis metabolic pathway. A total of 26 anthocyanins in SR were significantly up-accumulated comparing with SG, and the transcriptome analysis demonstrated that the expression of structural genes (*C4H*, *F3H*, *F3'5'H* and *UFGTs*), and four MYBs in anthocyanin biosynthetic pathway were significantly higher in SR walnut. Delphinidin 3-O-galactoside, cyanidin 3-O-galactoside, delphinidin 3-O-glucoside and cyanidin 3-O-glucoside were identified to the primary components of anthocyanidins because of the higher contents.

Our results obtained the valuable information on the anthocyanin metabolites and candidate genes in anthocyanin biosynthesis, which provided new insights into the anthocyanin biosynthesis in walnuts.

Methods

Plant material and growth conditions

The red walnut (*Juglans regia* L. accession RW-1, Germplasm resource number: JUREG4108210002) was introduced from Taihang Mountain, China (Fig. S1). In order to maintain a relatively consistent genetic background, different color phenotypes of natural hybrid plants (SG for green leaf natural hybrid plants, and SR for red leaf natural hybrid plants) grown in the Fruit Tree Experimental Station of Horticulture College, Henan Agricultural University, Zhengzhou, Henan, China. And the sampling permission has been obtained from the public land management agency of Henan Agricultural University. The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1), red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control[24] (Fig. 1a). All samples were immediately frozen in liquid nitrogen and stored at -80°C until RNA and metabolite extraction.

Extraction and quantification of metabolites

The sample Widely Targeted Metabolome and anthocyanidin detection were performed at Wuhan MetWare Biotechnology Co., Ltd. (Wuhan, China) as described previously [28, 42–44].

In brief, 100 mg leaves were crushed into powder and extracted with 0.6 mL of 70% aqueous methanol overnight at 4°C [45]. Following centrifugation at 10,000 g for 10 min, the supernatant was filtered through a microporous membrane (0.22 μm) for subsequent liquid chromatography/tandem mass spectrometry (LC-MS/MS) analysis. Quality control (QC) samples were prepared using mixing sample extracts. The extracts were filtrated and analyzed by Ultra Performance Liquid Chromatography coupled to Electrospray Ionization Tandem Mass Spectrometry (UPLC-ESI-MS/MS). Principal component analysis (PCA) were performed to verify the differences and reliability of metabolites in all the samples. Differentially accumulated metabolites (DAMs) between groups were filtered with VIP value ≥ 1 and absolute Log_2FC (fold change) ≥ 1. Subsequently, the differentially expressed metabolites with the significant enrichment were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.genome.jp/kegg) and Plant Metabolic Network (PMN, https://plantcyc.org/) were used to analyze phenylalanine metabolism pathway, flavonoid metabolism pathway and anthocyanin biosynthesis pathway [45, 46]. We constructed metabolic pathways based on the KEGG database.

RNA extraction, sequencing and transcriptome data analysis

The total RNA from walnut leave was extracted using the Omega Plant RNA Kit (Omega Bio-tek, Norcross, Georgia) according to the manufacturer's instructions. The integrity and quality of total RNA was

examined on 1% agarose gels, and the RNA concentrations were measured by a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). cDNA library was constructed and sequenced on a HiSeq 2,500 platform (Illumina, San Diego, CA, United States) (NCBI accession PRJNA688391) by Biomarker Biotechnology Corp. (Beijing, China), following paired-end reads were produced after cluster generation.

The expression abundance of unigene was represented as FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) value, and the DEGs [false discovery rate (FDR) < 0.05, and | Log_2FC | ≥ 1.5] between red leaves and green leaves were obtained using the DESeq 1.8.3 package [47]. Function annotation was performed using Gene Ontology (GO) and the pathways with significant enrichment were identified based on KEGG pathway employing the cluster Profiler R package [48]. WGCNA (The Weighted Gene Co-Expression Network Analysis) was performed as previously described [49].

Quantitative Real-time (qRT) PCR Assay

The expression of structural genes and transcription factors genes in the anthocyanin biosynthetic pathway were examined by qRT-PCR method. The first-stand cDNA was synthesized using the FastQuant RT Kit (with gDNase) (Tiangen Biotech, Beijing, China), and qRT-PCR was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) on anABI 7500 Real-Time PCR system (Applied Biosystems, Foster City, CA, United States), *Jr18S* (GenBank accession No. XM_019004991.1) as the house-keeping gene[10]. Quantification was evaluated using the $2^{-\Delta\Delta Ct}$ method. All the primers were shown in Table S1.

Statistical analysis

All data were analyzed using SPSS 22.0 software, expressing as the mean ± standard deviation (SD) of three replicates. Significant differences were carried out using a one-sided paired t-test ($^{**}p < 0.01$, $^{*}p < 0.05$) between red leaf and green leaf samples.

Abbreviations

Used
ANS
anthocyanin synthetase
C4H
cinnamate 4-hydroxylase
CHI
chalcone isoenzyme
CHS
chalcone synthetase
Cy
Cyanidin

DAMs
differentially accumulated metabolites
DEGs
the different expressed genes
DFR
dihydroflavonol 4-reductase
Dp
delphinidin
F3H
flavanone 3-hydroxylase
F3’5’H
Flavonoid 3',5'-Hydroxylase
FDR
false discovery rate
FPKM
Fragments Per Kilobase of transcript per Million fragments mapped
GO
Gene Ontology
KEGG
Kyoto Encyclopedia of Genes and Genomes
LC-MS/MS
liquid chromatography/tandem mass spectrometry
Mv
malvidin
PAL
phenylalanine ammonia lyase
PCA
principal component analysis

Declarations

Ethics approval and consent to participate

The collection of wild red walnut were completed under the support of the project ‘Investigation and cataloguing of walnut genetic resources in Henan Province (2016) (Grant ID:GR-2016-06)’, and all the experimental research and field studies on plants comply with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key R&D Program of China (2018YFD1000604) and the Support Plan of Scientific and Technological Innovation Team in Colleges and Universities of Henan Province, China (19IRTSTHN009). The funding organizations provided the financial support to the research projects, but had no role in the design of the study, data collection and analysis, and writing of the manuscript.

Authors' Contributions

GW and JS designed the research. LL and WZ performed experimental works and data analysis, prepared the original draft. LL and LW participated in data analysis; HM, GZ and WW helped review and editing the draft; LW provided support for projects and funds, and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

Accession Codes

We have deposited our raw data in the NCBI Sequence Read Archive (SRA) under accession numbers PRJNA688391 (https://www.ncbi.nlm.nih.gov/).

References

1. Honda C, Moriya S: **Anthocyanin biosynthesis in apple fruit**. *The Horticulture Journal* 2018, **87**(3):305–314.
2. Sasaki N, Nishizaki Y, Ozeki Y, Miyahara T: **The role of acyl-glucose in anthocyanin modifications**. *Molecules* 2014, **19**(11):18747–18766.
3. Shi M-Z, Xie D-Y: **Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana**. *Recent patents on biotechnology* 2014, **8**(1):47–60.

4. Li H, Yang Z, Zeng Q, Wang S, Luo Y, Huang Y, Xin Y, He N: **Abnormal expression of bHLH3 disrupts a flavonoid homeostasis network, causing differences in pigment composition among mulberry fruits**. *Horticulture research* 2020, **7**(1):1–19.
5. Juadjur A, Mohn C, Schantz M, Baum M, Winterhalter P, Richling E: **Fractionation of an anthocyanin-rich bilberry extract and in vitro antioxidative activity testing**. *Food chemistry* 2015, **167**:418–424.
6. Gould K, Davies KM, Winefield C: **Anthocyanins: biosynthesis, functions, and applications**: Springer Science & Business Media; 2008.
7. Kong J-M, Chia L-S, Goh N-K, Chia T-F, Brouillard R: **Analysis and biological activities of anthocyanins**. *Phytochemistry* 2003, **64**(5):923–933.
8. Lev-Yadun S, Gould KS: **Role of anthocyanins in plant defence**. In: *Anthocyanins.* Springer; 2008: 22–28.
9. Blesso CN: **Dietary anthocyanins and human health**. In., vol. 11 (9): Multidisciplinary Digital Publishing Institute; 2019: 2107.
10. Speer H, D'Cunha NM, Alexopoulos NI, McKune AJ, Naumovski N: **Anthocyanins and human health—A focus on oxidative stress, inflammation and disease**. *Antioxidants* 2020, **9**(5):366.
11. Katsumoto Y, Fukuchi-Mizutani M, Fukui Y, Brugliera F, Holton TA, Karan M, Nakamura N, Yonekura-Sakakibara K, Togami J, Pigeaire A: **Engineering of the rose flavonoid biosynthetic pathway successfully generated blue-hued flowers accumulating delphinidin**. *Plant and Cell Physiology* 2007, **48**(11):1589–1600.
12. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S: **A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land**. *The Plant Journal* 2012, **69**(6):1030–1042.
13. Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K *et al*: **Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana**. *The Plant Journal* 2012, **69**(1):154–167.
14. Matus JT, Cavallini E, Loyola R, Höll J, Finezzo L, Dal Santo S, Vialet S, Commisso M, Roman F, Schubert A: **A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus**. *The Plant Journal* 2017, **91**(2):220–236.
15. Vimolmangkang S, Han Y, Wei G, Korban SS: **An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development**. *BMC plant biology* 2013, **13**(1):1–13.
16. Zhong C, Tang Y, Pang B, Li X, Yang Y, Deng J, Feng C, Li L, Ren G, Wang Y: **The R2R3-MYB transcription factor GhMYB1a regulates flavonol and anthocyanin accumulation in Gerbera hybrida**. *Horticulture Research* 2020, **7**(1):1–13.
17. Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V: **Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway**. *Journal of experimental botany* 2011, **62**(8):2465–

2483.

18. Springob K, Nakajima J-i, Yamazaki M, Saito K: **Recent advances in the biosynthesis and accumulation of anthocyanins**. *Natural product reports* 2003, **20**(3):288–303.
19. Wang L, Tang W, Hu Y, Zhang Y, Sun J, Guo X, Lu H, Yang Y, Fang C, Niu X: **A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit Actinidia chinensis cv. Hongyang**. *The Plant Journal* 2019, **99**(2):359–378.
20. Li C, Wu J, Hu K-D, Wei S-W, Sun H-Y, Hu L-Y, Han Z, Yao G-F, Zhang H: **PyWRKY26 and PybHLH3 cotargeted the PyMYB114 promoter to regulate anthocyanin biosynthesis and transport in red-skinned pears**. *Horticulture research* 2020, **7**(1):1–12.
21. Ma Q, Yue J, Song X, Zhou Y, Pei D: **Fruit scientific research in New China in the past 70 years: Walnut**. *Journal of Fruit Science* 2019, **36**(10):1360–1368.
22. Li Y, Luo X, Wu C, Cao S, Zhou Y, Jie B, Cao Y, Meng H, Wu G: **Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in red and green walnut (Juglans regia L.)**. *Molecules* 2018, **23**(1):25.
23. Zhao W, Liu Y, Li L, Meng H, Yang Y, Dong Z, Wang L, Wu G: **Genome-wide identification and characterization of bHLH transcription factors related to anthocyanin biosynthesis in red walnut (Juglans regia L.)**. *Frontiers in genetics* 2021, **12**.
24. Zhao W, Li L, Liu Y, Zhang L, Yang Y, Meng H, Wang L, Wu G: **Identification and bioinformatics analysis of CHS genes in different phenotypic leaves of natural hybrid progenies of red-kernel walnut**. *Journal of Fruit Science* 2021, **38**(02):179–191.
25. Shi Q, Du J, Zhu D, Li X, Li X: **Metabolomic and Transcriptomic Analyses of Anthocyanin Biosynthesis Mechanisms in the Color Mutant Ziziphus jujuba cv. Tailihong**. *Journal of Agricultural and Food Chemistry* 2020, **68**(51):15186–15198.
26. Solhaug EM, Roy R, Chatt EC, Klinkenberg PM, Mohd-Fadzil NA, Hampton M, Nikolau BJ, Carter CJ: **An integrated transcriptomics and metabolomics analysis of the Cucurbita pepo nectary implicates key modules of primary metabolism involved in nectar synthesis and secretion**. *Plant direct* 2019, **3**(2):e00120.
27. Yang B, He S, Liu Y, Liu B, Ju Y, Kang D, Sun X, Fang Y: **Transcriptomics integrated with metabolomics reveals the effect of regulated deficit irrigation on anthocyanin biosynthesis in Cabernet Sauvignon grape berries**. *Food chemistry* 2020, **314**:126170.
28. Zhang Q, Wang L, Liu Z, Zhao Z, Zhao J, Wang Z, Zhou G, Liu P, Liu M: **Transcriptome and metabolome profiling unveil the mechanisms of Ziziphus jujuba Mill. peel coloration**. *Food chemistry* 2020, **312**:125903.
29. Zhang Z, Tian C, Zhang Y, Li C, Li X, Yu Q, Wang S, Wang X, Chen X, Feng S: **Transcriptomic and metabolomic analysis provides insights into anthocyanin and procyanidin accumulation in pear**. *BMC plant biology* 2020, **20**:1–14.
30. Zhuang H, Lou Q, Liu H, Han H, Wang Q, Tang Z, Ma Y, Wang H: **Differential regulation of anthocyanins in green and purple turnips revealed by combined de novo transcriptome and**

metabolome analysis. *International journal of molecular sciences* 2019, **20**(18):4387.

31. Li Y, Chen Q, Xie X, Cai Y, Li J, Feng Y, Zhang Y: **Integrated Metabolomics and Transcriptomics Analyses Reveal the Molecular Mechanisms Underlying the Accumulation of Anthocyanins and Other Flavonoids in Cowpea Pod (Vigna unguiculata L.)**. *Journal of Agricultural and Food Chemistry* 2020, **68**(34):9260–9275.
32. Stintzing FC, Carle R: **Functional properties of anthocyanins and betalains in plants, food, and in human nutrition**. *Trends in food science & technology* 2004, **15**(1):19–38.
33. Zhang J, Zhang W, Ji F, Qiu J, Song X, Bu D, Pan G, Ma Q, Chen J, Huang R: **A high-quality walnut genome assembly reveals extensive gene expression divergences after whole-genome duplication**. *Plant biotechnology journal* 2020, **18**(9):1848–1850.
34. Petroni K, Tonelli C: **Recent advances on the regulation of anthocyanin synthesis in reproductive organs**. *Plant science* 2011, **181**(3):219–229.
35. Naing AH, Kim CK: **Roles of R2R3-MYB transcription factors in transcriptional regulation of anthocyanin biosynthesis in horticultural plants**. *Plant molecular biology* 2018, **98**(1):1–18.
36. Jiang T, Zhang M, Wen C, Xie X, Tian W, Wen S, Lu R, Liu L: **Integrated metabolomic and transcriptomic analysis of the anthocyanin regulatory networks in Salvia miltiorrhiza Bge. flowers**. *BMC Plant Biology* 2020, **20**(1):1–13.
37. Tanaka Y, Sasaki N, Ohmiya A: **Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids**. *The Plant Journal* 2008, **54**(4):733–749.
38. Liu H, Su J, Zhu Y, Yao G, Allan AC, Ampomah-Dwamena C, Shu Q, Lin-Wang K, Zhang S, Wu J: **The involvement of PybZIPa in light-induced anthocyanin accumulation via the activation of PyUFGT through binding to tandem G-boxes in its promoter**. *Horticulture research* 2019, **6**(1):1–13.
39. Koes R, Verweij W, Quattrocchio F: **Flavonoids: a colorful model for the regulation and evolution of biochemical pathways**. *Trends in plant science* 2005, **10**(5):236–242.
40. Montefiori M, Espley RV, Stevenson D, Cooney J, Datson PM, Saiz A, Atkinson RG, Hellens RP, Allan AC: **Identification and characterisation of F3GT1 and F3GGT1, two glycosyltransferases responsible for anthocyanin biosynthesis in red-fleshed kiwifruit (Actinidia chinensis)**. *The Plant Journal* 2011, **65**(1):106–118.
41. Van Nocker S, Ludwig P: **The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function**. *BMC genomics* 2003, **4**(1):1–11.
42. Cao H, Ji Y, Li S, Lu L, Tian M, Yang W, Li H: **Extensive metabolic profiles of leaves and stems from the medicinal plant Dendrobium officinale Kimura et Migo**. *Metabolites* 2019, **9**(10):215.
43. Yuan H, Zeng X, Shi J, Xu Q, Wang Y, Jabu D, Sang Z, Nyima T: **Time-course comparative metabolite profiling under osmotic stress in tolerant and sensitive Tibetan hulless barley**. *BioMed research international* 2018, **2018**.
44. Zhang S, Ying H, Pingcuo G, Wang S, Zhao F, Cui Y, Shi J, Zeng H, Zeng X: **Identification of potential metabolites mediating bird's selective feeding on Prunus mira flowers**. *BioMed research international* 2019, **2019**.

45. Wang Z, Cui Y, Vainstein A, Chen S, Ma H: **Regulation of fig (Ficus carica L.) fruit color: metabolomic and transcriptomic analyses of the flavonoid biosynthetic pathway**. *Frontiers in Plant Science* 2017, **8**:1990.
46. Viant MR, Kurland IJ, Jones MR, Dunn WB: **How close are we to complete annotation of metabolomes?** *Current opinion in chemical biology* 2017, **36**:64–69.
47. Yi D, Zhang H, Lai B, Liu L, Pan X, Ma Z, Wang Y, Xie J, Shi S, Wei Y: **Integrative Analysis of the Coloring Mechanism of Red Longan Pericarp through Metabolome and Transcriptome Analyses**. *Journal of Agricultural and Food Chemistry* 2020.
48. Yu G, Wang L-G, Han Y, He Q-Y: **clusterProfiler: an R package for comparing biological themes among gene clusters**. *Omics: a journal of integrative biology* 2012, **16**(5):284–287.
49. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis**. *BMC bioinformatics* 2008, **9**(1):1–13.

Figures

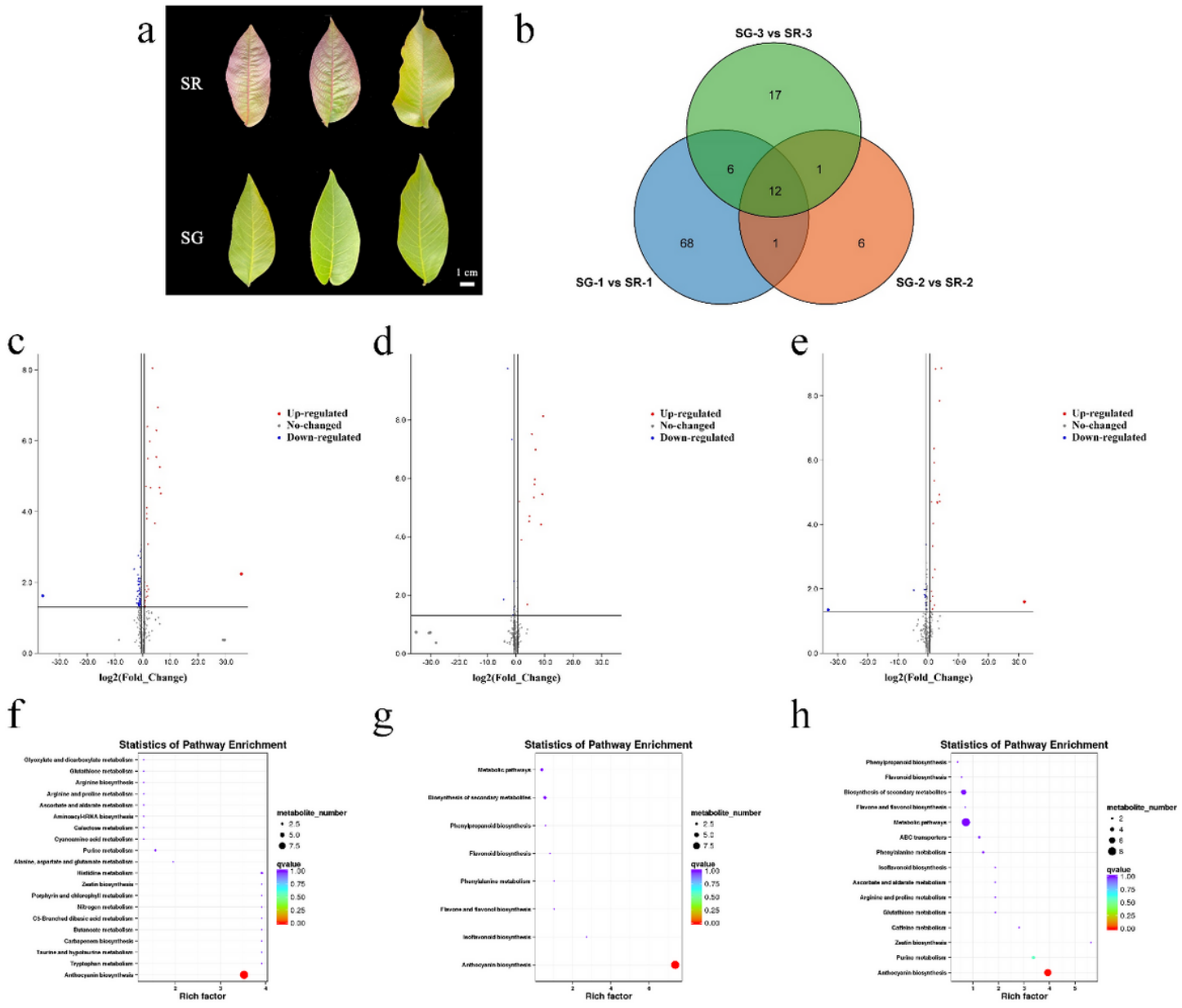

Figure 1

a. Morphological observation of different color leaves in natural hybrid progenies of red walnut; b. Venn diagram depicting the shared and specific metabolites between the three compared groups of leave samples; number of the differentially accumulated metabolites (DAMs) between SG-1 vs SR-1 (c), SG-2 vs SR-2 (d) and SG-3 vs SR-3 (e); KEGG enrichment analysis of the DAMs between SG-1 vs SR-1 (f), SG-2 vs SR-2 (g) and SG-3 vs SR-3 (h). The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1), red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control.

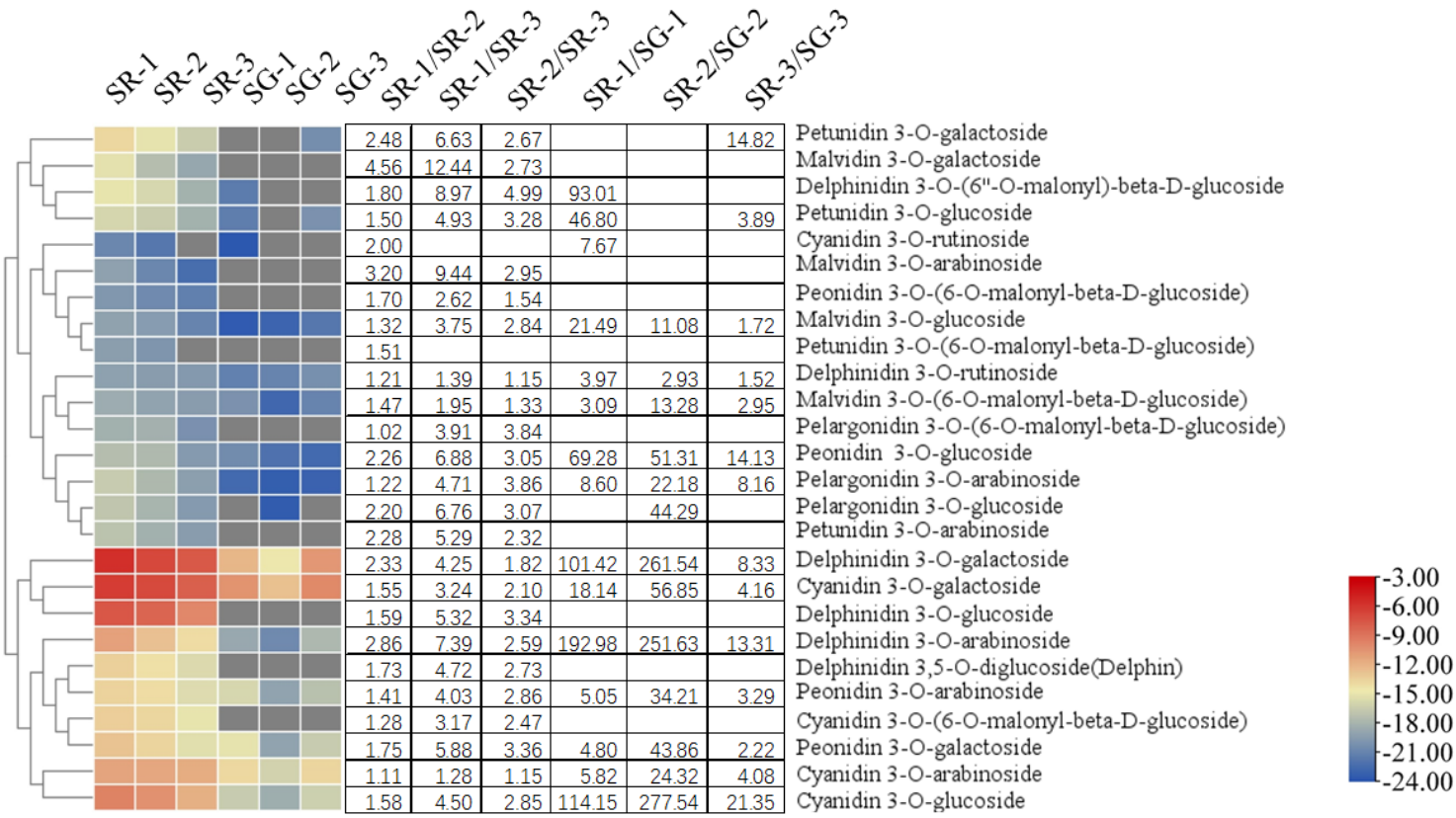

Figure 2

Concentrations of anthocyanins in leaves of SR and SG. Numbers refer to -fold change in anthocyanin contents. SG: Seedling progenies-Green leaves; SR: Seedling progenies-Red leaves; The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1), red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control.

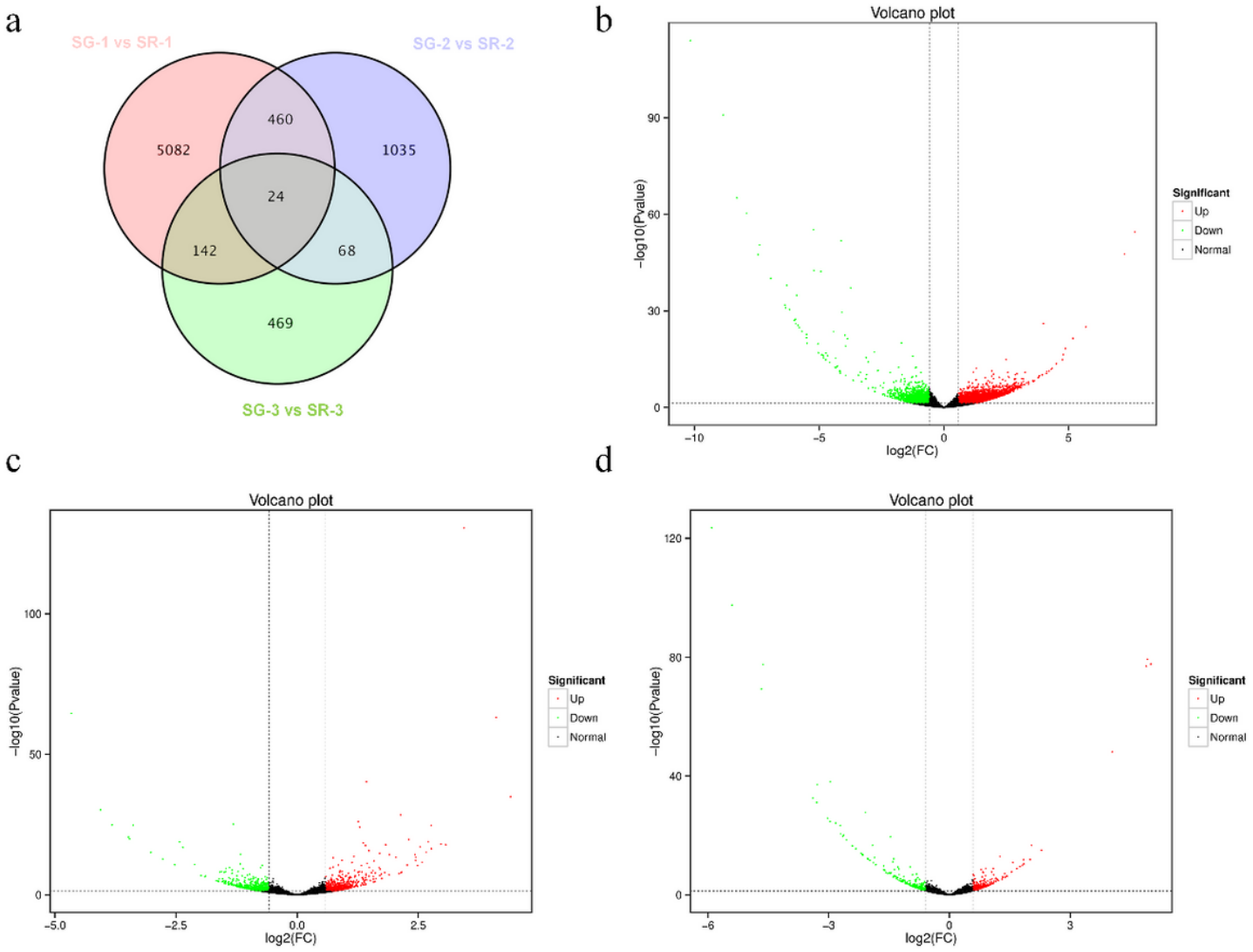

Figure 3

Number of differentially expressed genes (DEGs) identified by RNA-seq analysis. a. Venn diagram representing numbers of DEGs. Volcano plots displaying the up-regulated, down-regulated and no-regulated genes between SG-1 vs SR-1 (b), SG-2 vs SR-2 (c) and SG-3 vs SR-3 (d). SG: Seedling progenies-Green leaves; SR: Seedling progenies-Red leaves; SG-1 and SR-1 refer to leaves collected at the full red period-new shoot growth stage, SG-2 and SR-2 for red and green period-fruit swelling stage, and SG-3 and SR-3 for full green period-early period of fruit ripening, respectively. the green dots indicated the down-regulated expressed genes (fold change, Log2FC ≤ 0.67), the red dots indicate the up-regulated expressed genes (Log2FC ≥ 1.5), and the black dots indicate the non-differentially expressed genes (0.67 < Log2FC < 1.5).

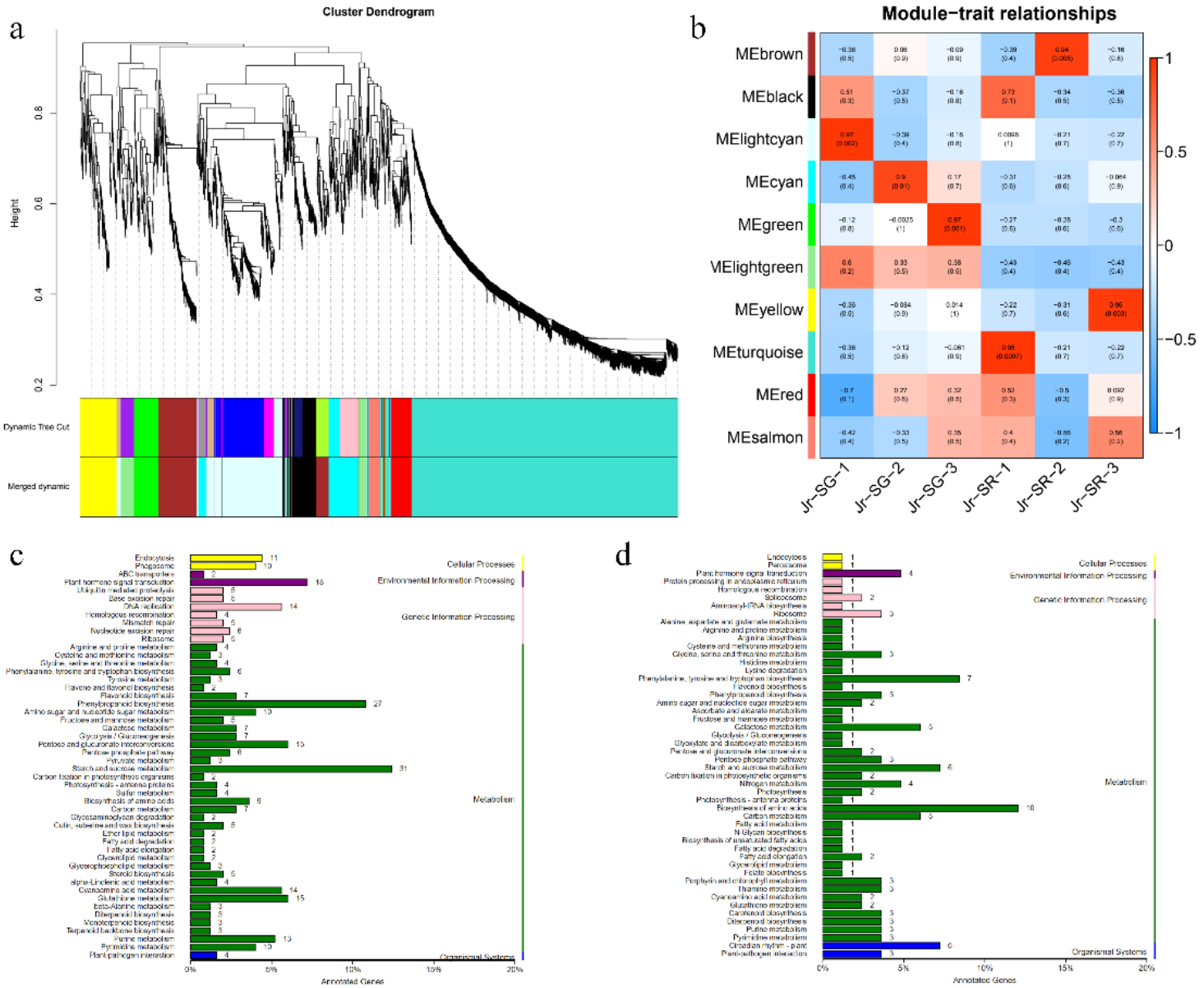

Figure 4

Expression modules constructed by weighted gene correlation network analysis (WGCNA) and associated biological processes of genes in the tuequoise and lightcyan modules. a. Gene dendrogram obtained by hierarchical clustering with the module color indicated by the color of the row underneath. A total of 10 distinct modules were identified. b. Relationships of modules and different samples including leaves in three developmental stages (SR-1,2,3 and SG-1,2,3). Each row in the table corresponds to a module, and each column corresponds to a sample. c and d. Enrichment analyses of GO terms in the tuequoise (c) and lightcyan (d) modules.

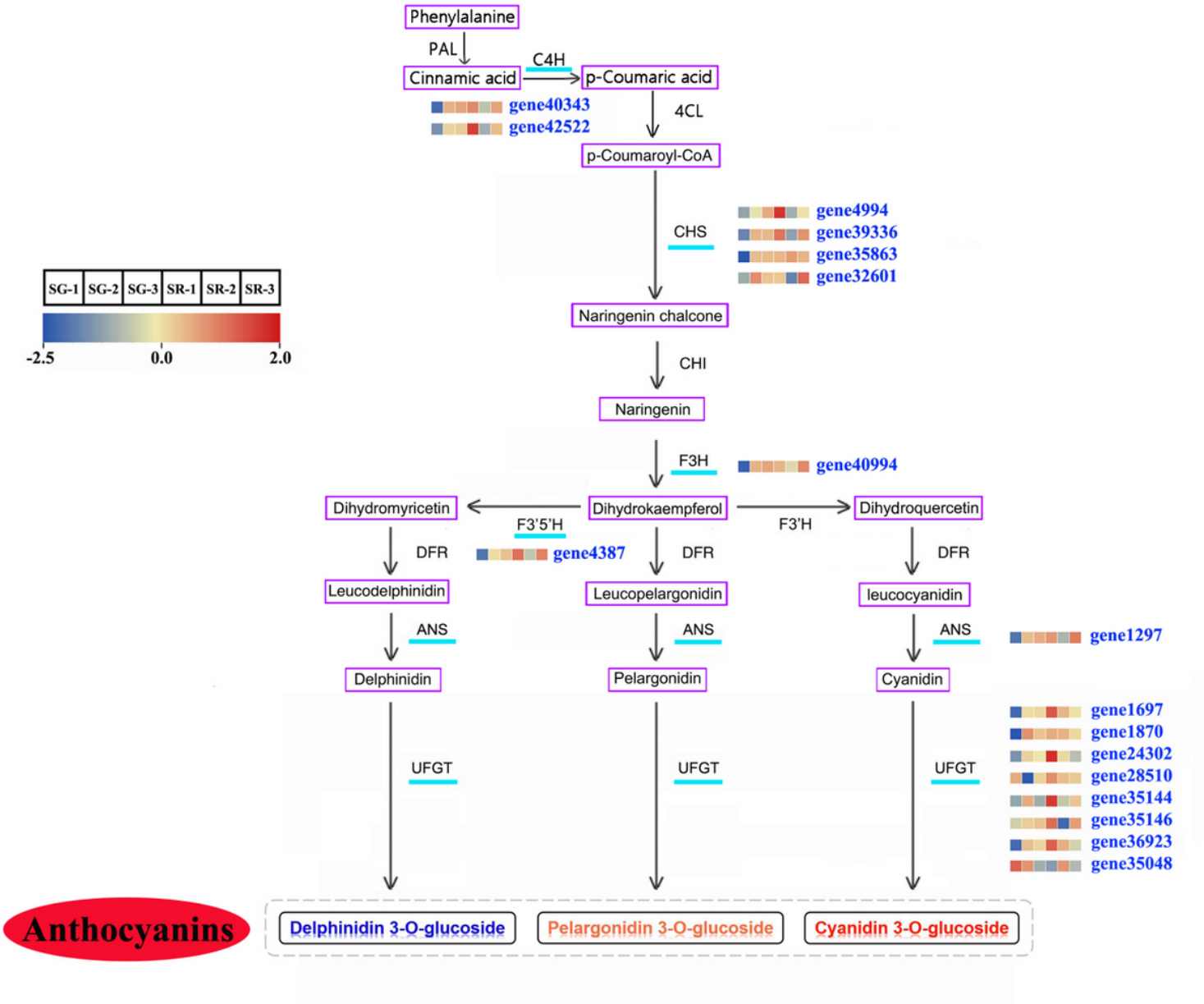

Figure 5

Modulation of anthocyanin biosynthesis pathway genes during walnut leave developmenting. The heatmap represents the expression of corresponding structural genes in SR and SG, and from blue to red in the heatmap indicates the expression levels of structural genes (FPKM value) raning from low to high. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3',5'-hydroxylase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UFGT, UDP-glucose:flavonoid 3-Oglucosyltransferase. SG: Seedling progenies-Green leaves; SR: Seedling progenies-Red leaves; The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1), red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control.

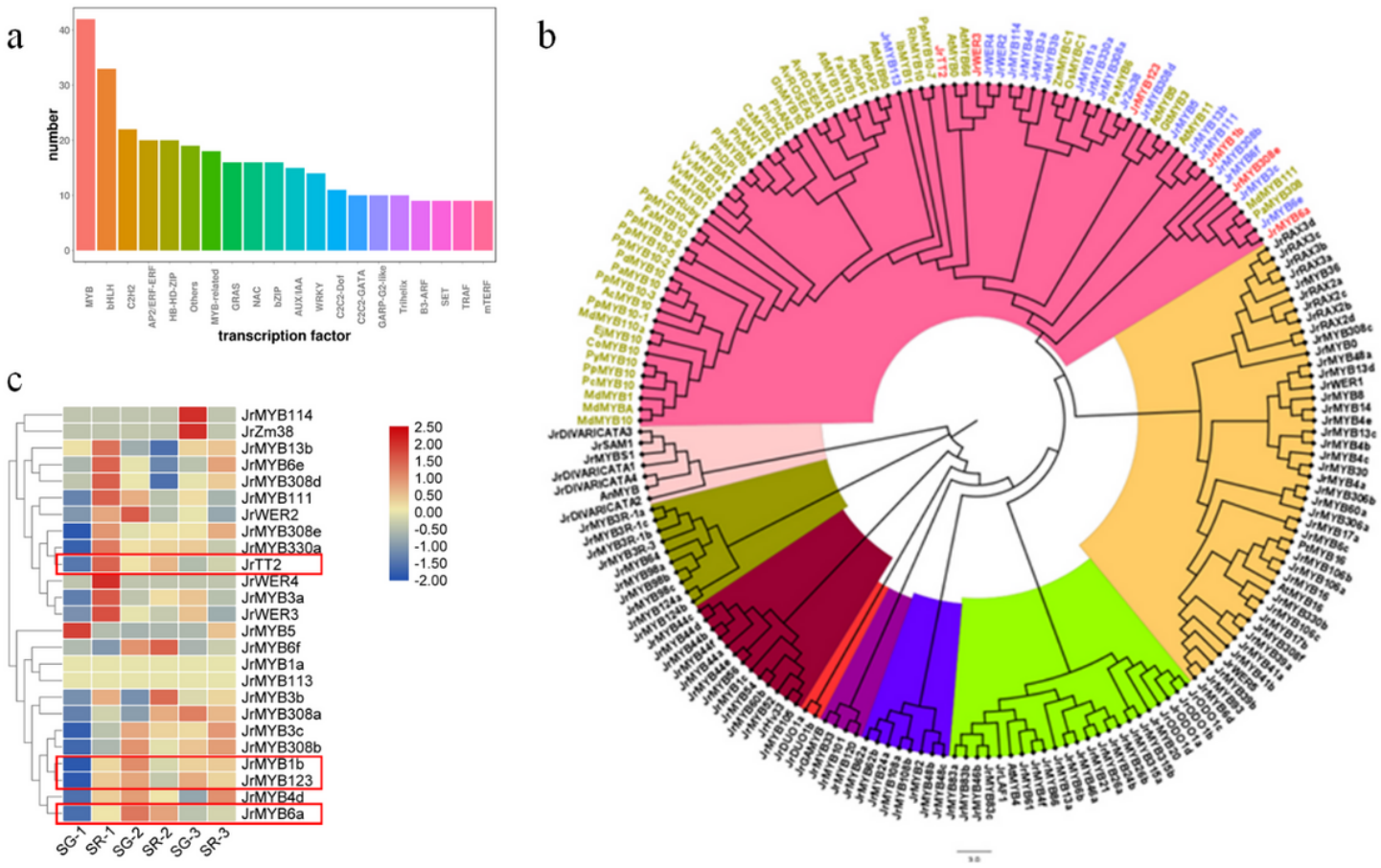

Figure 6

Analysis of transcription factors related to anthocyanin biosynthesis. a. Number of different expression transcription factors in three period. b. Phylogenetic relationship of MYB protein between walnut, Arabidopsis thaliana, and MYBs related to anthocyanin biosynthesis in other species. c. Gene expression pattern of 25 JrMYBs highly homologous to anthocyanin related MYBs of other species in different phenotypic (red leaves and green leaves) of red walnut natural hybrid progeny by RNA-seq. The genes in red frames were the differentially expressed MYB genes. Leaves of the different phenotypes (red leaves and green leaves) of red walnut natural hybrid progeny were collected at three stages. SG: Seedling progenies-Green leaves; SR: Seedling progenies-Red leaves; The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1), red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control.

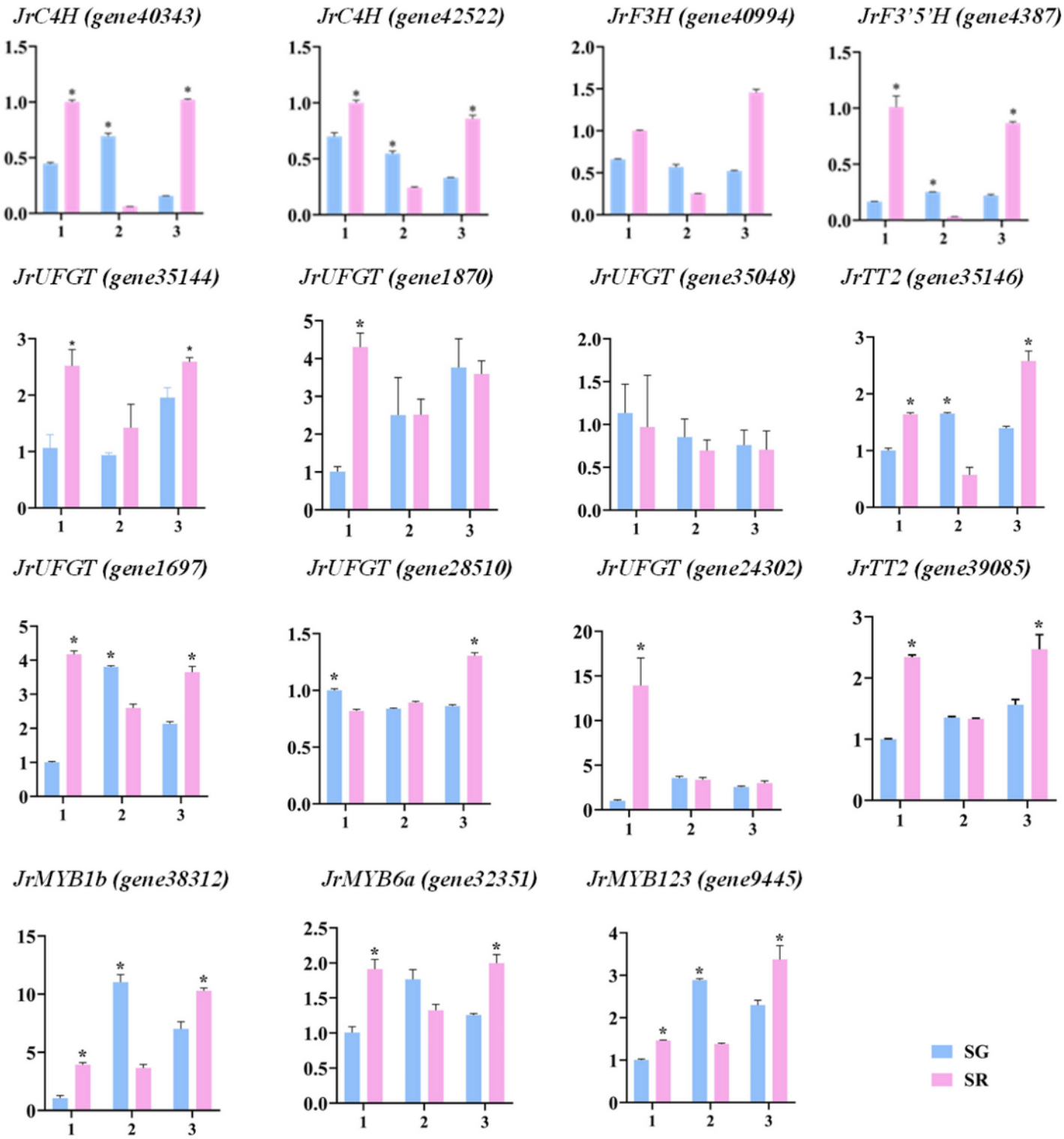

Figure 7

Expression of the structural genes and MYB genes in the different phenotypes (red leaves and green leaves) of the red walnut natural hybrid progeny. The relative expressions were detected by qRT-PCR. Leaves of the different phenotypes (red leaves and green leaves) of red walnut natural hybrid progeny were collected at three stages. SG, Seedling progenies-Green leaves; SR, Seedling progenies-Red leaves. The leaves were collected according to the color changing of SR walnut leaves, as full red period (SR-1),

red-green period (SR-2), and full green period (SR-3), the leaves of SG walnut in the same period were collected as the control. Significant differences were determined using a one-sided paired t-test (*p< 0.05). Expression values (±SE) of three replicates were normalized using Jr18s as the internal control.

Supplementary Files

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

- SupplementalFiguresS1S5.docx
- SupplementalTablesS1S4.docx