Title: Pericarp Pigmentation in 'Red Delicious' and Its Four Generation Mutants Correlates with Hormones

Running Title: Hormones modulate anthocyanin accumulation

Wen-Fang Li[1], Juan Mao[1], Shi-Jin Yang[1], Zhi-Gang Guo[1], Zong-Huan Ma[1], Mohammed Mujitaba Dawuda[1,2], Cun-Wu Zuo[1], Ming-Yu Chu[1], Bai-Hong Chen[1]*

[1] College of Horticulture, Gansu Agricultural University, Lanzhou 730070, PR China

[2] Department of Horticulture, FoA, University for Development Studies, P. O. Box TL 1882, Tamale, Ghana

***Corresponding author:** Baihong Chen, Gansu Agricultural University, College of Horticulture

e-mail: bhch@gsau.edu.cn

Abstract

Anthocyanins are fulfil many physiological and ecological functions, but a full understanding of the interplay between hormones and anthocyanin synthesis in apple at the molecular level remains elusive. Here, physiological and comparative transcriptome approaches were used to reveal the molecular basis of pericarp pigmentation in 'Red Delicious' and its four generation mutants including 'Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur'. It was found that the pericarp pigmentation was gradually proliferated from G0 to G4. The activation of early phenylpropanoid biosynthesis pathway genes including *ASP3*, *PAL*, *4CL*, *PER*, *CHS*, *CYP98A* and *F3'H* were more responsible for anthocyanin accumulation. In addition, the down-regulation of *AACT1*, *HMGS*, *HMGR*, *MVK*, *MVD2*, *IDI1* and *FPPS2* involved in terpenoid biosynthesis influence anthocyanin accumulation by positively regulating the expression of *AUX1* and *SAUR* that contributed to the

synthesis of IAA, *GID2* to GA, *PP2C* and *SnRK2* to ABA. MYB and bHLH members which highly correlated (0.882<|r|=0.980) with anthocyanin contents modulated anthocyanin accumulation in apple pericarp by regulating the transcription of structural genes including *CHS* and *F3'H* involved in flavonoid pathway. In particular, MYB members including *LUX*, *MYB113*, *PCL-like*, *MYB1R1-like*, *MYB6-like*, *MYB308-like* and *MYB5-like*, bHLH members including *bHLH51* and *bHLH155* notable positive correlation with the anthocyanin content, were confirmed to promote anthocyanin synthesis. Other MYB members of *LHY-like*, *RVE6*, *GT-3b*, *MYB21*, *DNAJC2-like* and *MYB4-like*, bHLH members of *bHLH62*, *bHLH74* and *bHLH162* had a remarkably negative correlation with the anthocyanin content, were demonstrated to repress anthocyanin synthesis.

Keywords: Apple, 'Red Delicious', somatic mutation, anthocyanin, RNA-seq, hormone.

Abbreviations: *4CL*, 4-coumarate-CoA ligase; *AACT1*, acetyl-CoA acetyltransferase, cytosolic 1; ABA, abscisic acid; *ANR*, anthocyanidin reductase; *ANS*, anthocyanidin synthase; *ASP3*, aspartate aminotransferase cytoplasmic; *BGLU*, beta-glucosidase; *CCoAOMT*, caffeoyl-CoA O-methyltransferase; *CHI*, chalcone isomerase; *CHS*, chalcone synthase; CTK, cytokinin; *CYP98A*, cytochrome P450 98A2-like; DEGs, differentially expressed genes; *DFR*, dihydroflavonol reductase; ETH, ethylene; *F3H*, avanone-3-hydroxylase; *F3'H*, flavonoid 3'-hydroxylase; *FLS*, flavonol synthase; FPKM, fragments per kilobase of mapped reads; *FPPS2*, farnesyl pyrophosphate synthase 2-like; GA, gibberellins; *HCT*, quinate hydroxycinnamoyl transferase; *HMGR*, 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1-like; *HMGS*, hydroxymethylglutaryl-CoA synthase-like; IAA, auxin; *IDI1*, isopentenyl-diphosphate Delta-isomerase I; JA, jasmonate acid; *LAR*, leucoanthocyanidin reductase; *MVD2*, diphosphomevalonate decarboxylase MVD2-like; *MVK*, mevalonate kinase-like; *PAL*, Phe ammonia lyase; *PER*, peroxidase; *SQMO*, squalene monooxygenase-like.

Introduction

Bud sport is a somatic mutation occurring in shoot cells of perennial fruits, and an important source of discovery of new cultivars or strains that are superior to the parent (Petit and Hampe, 2006; El-Sharkawy *et al*., 2015). The genetic background of these mutants albeit almost identical to their parents (Nwafor *et al*., 2014; Otto *et al*., 2014),

epigenetic changes still have been recognized, like causing fruit colour alteration in apple (*Malus domestica* Brokh.) (Xu *et al*., 2012; El-Sharkawy *et al*., 2015). Pericarp color is a key appearance and nutrition quality attribute of apple fruit (Willams and Benkeblia, 2018). Anthocyanins are secondary metabolites that contribute to the colors of fruits (Meng *et al*., 2016). Hormones are likely to be important factors modulating the light-dependent anthocyanin accumulation (Jeong *et al*., 2004; Loreti *et al*., 2008; Carvalho *et al*., 2010). In summary, exploring the molecular mechanisms of hormones and anthocyanin synthesisin apple fruit is crucial for further research on pigment accumulation and plant somatic mutation.

The basal level of anthocyanins can be modulated by various internal and external stimuli, such as temperature (Arrizabalaga *et al*., 2018), light (Cominelli *et al*., 2007), carbohydrates and plant hormones (Loreti *et al*., 2008). Previous studies have shown that auxin (IAA), cytokinin (CTK), gibberellins (GA), jasmonate acid (JA), abscisic acid (ABA) and ethylene (ETH) displayed the interaction in controlling anthocyanin biosynthesis (El-Kereamy *et al*., 2003; Jeong *et al*., 2004; Loreti *et al*., 2008; Liu *et al*., 2014; Ji *et al*., 2014). In addition, the identification and functional characterization of MYB and bHLH transcriptional factors revealed that they have the role of autonomously mediated structural gene transcription, such as chalcone isomerase (*CHI*), chalcone synthase (*CHS*), flavonol synthase (*FLS*), leucoanthocyanidin reductase (*LAR*), flavonoid 3'-hydroxylase (*F3'H*) and anthocyanidin reductase (*ANR*) involved in anthocyanin biosynthesis pathway (Deluc *et al*., 2008; Gonzalez *et al*., 2008; Telias *et al*., 2011; Petroni and Tonelli, 2011; An *et al*., 2012). We know that MYB proteins characterized by two imperfect repeats of the DNA-binding motifs R2 and R3 (Ramsay and Glover, 2005), and bHLH proteins are characterized by the basic helix-loop-helix domain which responsible for sequence-specific DNA binding (Massari and Murre, 2000).

The publication of the apple reference genome (Daccord *et al*., 2017) and the development of new tools for transcriptomics have facilitated recent advances in the genome-wide analysis of dynamic gene expression during pericarp development (El-Sharkawy *et al*., 2015; Massonnet *et al*., 2017). The strategies of hormones and

anthocyanin synthesis often are applied without a full understanding of the impact at the molecular level, with the exception of a few studies that have correlated biochemical and physiological outcomes with transcriptomic changes (Jeong *et al*., 2004; Loreti *et al*., 2008; Carvalho *et al*., 2010). Apple cultivar 'Red Delicious' (*M. domestica*) is the most frequently captured sport apple variety that usually selected on the phenotypic basis of spur-type and more intense red fruit color. Its four continuous generation mutants, including 'Starking Red', 'Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee Spur' haven been screened out so far. Therefore, these five strains were selected and analyzed using comparative transcriptome combined with physiological and biochemical characteristics to expound the relationships between hormones and anthocyanin synthesis on apple pericarp pigment accumulation.

Materials and methods

Fruit Material

'Red Delicious' is the most frequently captured sport apple variety, containing four generation mutants. 'Starking Red' is bud sport from 'Red Delicious' and a typical representative of the first generation. The second generation contains 'Starkrimson', which is bud sport from the first generation 'Starking Red'. The fourth generation 'Vallee Spur' is bud sport from the third generation 'Campbell Redchief',which is bud sport from 'Starkrimson'. Mature apple fruit of these five cultivars are often from vertical strips red to full red (Fig. 1A).

Fruit pericarp samples of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') were named G0 to G4 and collected in 2017 from 12−year−old trees grown in apple demonstration gardens at Tianshui, China, respectively. Briefly, 20−30 fruit from each of the five strains were sampled at three developmental stages, i.e. 5 August (S1) , 25 August (S2), and 14 September (S3), respectively (Fig. 1A). Stage S1, S2 and Stage S3, is equivalent to the prior to veraison, veraison and fruit maturity stage for commercial harvest,

respectively. In different experiments, pericarp samples from 6 fruit per replicate with three independent biological replicates were collected. All samples were collected at the same time of day (9–10 $_{AM}$), immediately frozen in liquid nitrogen and stored at −80 ^{o}C for further analysis of anthocyanin contents, endogenous hormones contents and gene expression profiles (qRT-PCR). In addition, samples from S2 were used for RNA-seq.

Anthocyanin Quantification

Lyophilized apple pericarp samples were finely ground, and about 500 mg of powdered samples were homogenized in 10 ml of methanol with 1% HCl. The homogenate was transferred to calibration test tube with constant volume to 20 mL, and kept for 20 min at 4°C with shaking in dark conditions. Then the samples were filtered through a 0.2 μm polyethersulfone (PES) filter (Krackeler Scientific, Inc., Albany, NY, USA) and analysed using a TU-1900 double beam UV-visible spectrophotometer (Beijing Purkinje General Instrument Co. LTD). Anthocyanin levels were calculated by dividing the absorbance by the coefficient of regression (0.0214) acquired by standard scale measurements.

Hormone content measurement

A total of 1.0 g of each lyophilized apple pericarp samples were ground quickly after liquid nitrogen was added and combined with 10 mL of 80% chromatographic pure methanol. Each sample was washed three times with solvent, transferred into a test tube, and stored in a refrigerator at 4 °C overnight in the dark. Then, the samples were centrifuged for 20 min under refrigerated conditions at 4 °C. Supernatant fluid was transferred into a new centrifuge tube. The extract was concentrated, and the methanol was volatilized under 40 °C by rotary evaporation to obtain 2 mL of concentrate. The evaporation bottle wall was then washed continuously with 50% methanol, and the volume was raised to 10 mL with 50% chromatographic pure methanol. The fluid for testing was filtered through 0.22 μm organic membrane.

The extract for IAA, GA, and ABA determination was injected into an Agilent 1100 series LC-MS apparatus (Agilent Technologies, Waldbronn, Germany) equipped with Extend-C_{18} (4.6 mm × 250 mm, 5 μm, Agilent, Germany). The mobile phase was chromatographic methanol and 0.6% iced acetic acid. The flow velocity was 1.0 mL/min, the wavelength was 254 nm, and the column temperature was 25 °C.

RNA Extraction

Total RNA was extracted from approximately 200 mg of lyophilized apple pericarp samples ground in liquid nitrogen using the RNase-Free DNase Set (Qiagen, Valencia, CA, USA) then cleaned up with the RNeasy Mini Kit (Qiagen). RNA quality and quantity were determined using a Pultton P200 Micro Volume Spectrophotometer (Pultton Technology Limited).

Library Preparation and Sequencing

The 5 triplicate samples (5 varities at S2) yielded 15 nondirectional cDNA libraries with a total of 68.18 million reads (Supplementary Table S1), which were prepared from 3.0 μg of total RNA using NEBNext Ultra^{TM} RNA Library Prep Kit (NEB, USA). RNA quality and quantity were assessed using a NanoDrop spectrophotometer and an Agilent 2100 spectrophotometer (Agilent Technologies, CA, USA). Library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA) and the quality was assessed on Agilent Bioanalyzer 2100 system. Clustering of index-coded samples was performed on a cBot cluster generation system using TruSeq PE Cluster Kit v3-cBot-HS (Illumia Inc., San Dego). After cluster generation, library preparations were sequenced on an Illumina Hiseq 2000 platform, where 125 bp paired-end reads were generated.

Analysis of Sequencing Results: Mapping and Differential Expression

Raw reads were cleaned by removing adapter sequences, reads containing ploy-N, and low-quality sequences (Q < 20). Clean reads were aligned onto the apple

reference genome (https://iris.angers.inra.fr/gddh13/) (Daccord *et al*., 2017). New transcripts were identified from TopHat alignment results using Cufflinks v2.1.1 reference based transcript assembly method. An average of 83.95% of reads was mapped for each sample (Supplemental Table S1). For annotations, all novel genes were searched against NCBI non-redundant protein sequence database (Nr), Swiss-Prot, Gene Ontology (GO) database, Cluster of Orthologous Groups of proteins (COG), protein family (Pfam), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database using BLASTx with 10^{-5} as E-value cut-off point; sequences with the highest similarities were retrieved. After amino acid sequences of new genes were predicted, they were searched against Pfam database using HMMER software, then, annotation information of these new genes were obtained.

DESeq (2010) R package was utilized to detect differentially expressed genes (DEGs). False discovery rate (FDR) was used to identify the *P*-value threshold in multiple tests (Benjamini and Hochberg, 1995). An FDR < 0.01 and fold change ≥ 2 were used as screening criteria; absolute value of $\log_2$ (fold change) with reads per kb per million reads (FPKM) ≥ 1.0 was used as threshold to determine significant DEGs (Mortazavi *et al*., 2008).

Functional Analysis of Differentially Expressed Genes (DEGs)

Functional enrichment analysis including GO and KEGG was performed to identify DEGs, which were significantly enriched in GO terms or metabolic pathways. GO enrichment analysis of DEGs was implemented by the GOseq R package, in which gene length bias was corrected. GO coupled with *KS*<0.01 were considered as significantly enriched by DEGs (Supplemental Table S2). KOBAS software was used to test statistical enrichment of different expression genes in KEGG pathways. Pathways with *Q*-value ≤ 0.05 were defined as genes that display significant levels of differential expression (Supplemental Table S3).

Common Expression Pattern Clustering Analysis of DEGs

The different expression patterns of DEGs among the five strains were analyzed using R language, Cluster package, Biobase package, and *Q*-value package. The DEGs with the common expression trend were divided into a data set, which was expressed as a model map. The distance measure used was Euclidean distance, clustering method is K-means clustering or hierarchical clustering.

Correlation Analysis

A correlation matrix was prepared using SPSS statistical software and Pearson's correlation coefficient as the statistical metric. The analysis was performed using the content of anthocyanin at S2 and the FPKM average of each candidate DEGs. Correlation values were converted into distance coefficients to define the height scale of the dendrogram.

Quantitative Real-time PCR Validation of RNA-Seq Data

Quantitative reverse transcription RT-PCR analysis DNase-treated RNA (2 μg) was reverse transcribed in a reaction of 20 μl using PrimerScript^{TM} RT reagent Kit with gDNA Eraser (Takara, Dalian, China). Gene-specific primers were designed using Primer Express software (Applied Biosystems) (Supplementary Table S8). Quantitative reverse transcription PCR (qRT-PCR) assays were performed using 20 ng of cDNA and 300 nM of each primer in a 10 μl reaction with SYBR Green PCR Master Mix (Takara, Dalian, China). Three biological and three technical replicates for each reaction were analysed on a LightCycler® 96 SW 1.1 instrument (Roche). Amplification program consisted of one cycle of 95 °C for 30 s, 40 cycles of 95 °C for 5 s, and melting analysis at 60 °C for 34 s, followed by one cycle of 95 °C for 15 s, 60 °C for 60 s, and 95 °C for 15 s. Transcript abundance was quantified using standard curves for both target and reference genes, which were generated from serial dilutions of PCR products from corresponding cDNAs. Transcript abundance was normalized to the reference gene *MdGADPH* that showed high stability across the different apple genotypes and tissues used in this study. Relative gene expression was

normalized by comparing with G0 expression and analyzed using comparative $2^{-\Delta\Delta C}T$ Method (Livak and Schmittgen, 2001).

Data Analysis

Data on the content of anthocyanin and the relative expression level of specific genes were analyzed by the ANOVA and treatment means were separated by Duncan Multiple Range Test at $P < 0.05$ with the aid of SPSS statistical software. For correlation analysis, Pearson correlation coefficient (r) was calculated and a two-tailed test was carried out.

Results

Anthocyanin profile during development in 'Red Delicious' and its four generation mutants

Fruit pericarp were sampled in triplicate at three developmental stages, prior to veraison (S1), veraison (S2), and harvest (S3). 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') were named G0 to G4. Visual inspection of apple pericarp colour during development revealed that G2, G3 and G4 strains began coloring at S1, with visibly more intense red colour in G3 and G4 (Fig. 1A). Pigmentation of these five strains that progressively advanced at much higher levels from S1 to the subsequent stage S3, resulting in red coloured fruit at maturity. In addition, red coloration was gradually proliferated from G0 to G4 at the three stage, respectively.

Consistent with the visual inspection, the analysis of apple pericarp anthocyanin contents showed that the levels of total anthocyanin in the five strains were initially low and sharply increased along with maturation progression (from S1 to S3) (Fig. 1B). Relative to G0, the total anthocyanin level was ~0.78-, ~1.20-, ~5.02- and ~2.85-fold higher in G1, G2, G3 and G4 fruit skin at the S1 stage, respectively. At S2, it was ~1.41-, ~4.65-, ~6.33- and ~4.72-fold higher in G1, G2, G3 and G4 fruit skin, respectively. Finally at S3, it was ~2.14-, ~3.41-, ~5.25- and ~4.35-fold higher in G1,

G2, G3 and G4 fruit skin, respectively. Briefly, anthocyanin content from the three stages was the lowest in G0, and highest in G3, followed by G4, G2, and G1.

Endogenous hormone content during development in 'Red Delicious' and its four generation mutants

The hormone levels also were analyzed at three time points (S1, S2,and S3). The overall trend of IAA concentrations in apple pericarp increased firstly and then decreased from S1 to S3 and peaked at S2 (Fig. 1C). However, GA content were decreasing from S1 to S3 (Fig. 1D). ABA concentrations of G0, G1 and G2 peaked at S1, whereas G3 and G4 peaked at S2 (Fig. 1E). Significantly, the IAA and ABA contents of G3 and G4 in apple pericarp at S2 were considerably higher than G2, but G0 and G1 showed considerably lower levels than G2. Remarkably, GA concentrations, which displayed a contrary trend compared with IAA and ABA, were decreased from G0 to G4 at S2.

Transcriptomic profiling of 'Red Delicious' and its four generation mutants pericarp

The triplicate sampling of 'Red Delicious' and its four generation mutants pericarp at S2 yielded 15 RNA samples for RNA sequencing (RNA-seq) analysis, and the mapping rate of 20,189,457-26,765,510 clean reads onto the apple reference genome (https://iris.angers.inra.fr/gddh13/) (Daccord *et al*., 2017) was ranged from 83.29% to 84.81% (Supplemental Table S1). The average mapped reads ranged from 34,093,629 in G2 to 43,153,440.33 in G3. An FDR < 0.01 and fold change ≥ 2 were used as screening criteria of DEGs, combined with a FPKM ≥ 1.0 in at least one of the 5 triplicate samples were considered to be expressed. The mean normalized expression value (fragments per kilobase of mapped reads, FPKM) per transcript of the three biological replicates was calculated for each sample using the geometric normalization method. The resulting data set comprising 33,192 transcripts was used for subsequent analysis (Supplemental Data Set S1).

Identification of differentially expressed genes (DEGs) in 'Red Delicious' and its four generation mutants

In order to identify DEGs in 'Red Delicious' and its four generation mutants, seven pairwise transcriptome comparisons (i.e. G0 versus G1, G0 versus G2, G0 versus G3, G0 versus G4, G1 versus G2, G2 versus G3, and G3 versus G4) were made at S2 (Fig. 2, Supplemental Fig. S1, Supplemental Table S2). The total number of DEGs was 3,466, including 1,456 up-regulated DEGs and 2029 down-regulated DEGs (Fig. 2, Supplemental Data Set S2). Among them, the number of both up-regulated and down-regulated DEGs was considerably increased from G1 to G4 versus G0, and the least in G1 versus G2. Moreover, we found that more genes were down-regulated rather than up-regulated in G1, G2, G3 and G4 relative to 'Red Delicious'.

Comparative transcriptome enrichment analysis identified key processes responsible for anthocyanin accumulation that received in 'Red Delicious' and its four generation mutants

To understand the major functional categories represented by the DEGs, Gene Ontology (GO) enrichment analysis was carried out using all reference genes as background, and implemented by the GOseq R package. GO term enrichment analysis categorized the annotated sequences into three main categories: biological process, cellular component and molecular function (Supplemental Table S3). In the biological process category, three significantly enriched terms including "defense response to fungus", "1-aminocyclopropane-1-carboxylate biosynthetic process" and "negative regulation of growth" were shared in G0 versus G1, G0 versus G2, G0 versus G3 and G0 versus G4, among them, "1-aminocyclopropane-1-carboxylate biosynthetic process" and "negative regulation of growth" also were enriched in G1 versus G2, G2 versus G3 and G3 versus G4. Furthermore, except for the two GO terms, other three significantly enriched GO terms including "chitin catabolic process", "regulation of leaf development" and "DNA conformation change" also were shared in G0 versus G1, G1 versus G2, G2 versus G3 and G3 versus G4. The cellular component category was further classified into "elongator holoenzyme complex", which was enriched in

G0 versus G1, G0 versus G2, G0 versus G3 and G0 versus G4, whereas, “elongator holoenzyme complex”, “mitochondrial intermembrane space” and “U12-type spliceosomal complex” were shared in G0 versus G1, G1 versus G2, G2 versus G3 and G3 versus G4. In the molecular function category, the DEGs were further classified into nine terms including “ADP binding”, “L-iditol 2-dehydrogenase activity”, “acyl-CoA hydrolase activity”, “3-beta-hydroxy-delta5-steroid dehydrogenase activity”, “O-methyltransferase activity”, “1-aminocyclopropane-1-carboxylate synthase activity”, “naringenin-chalcone synthase activity”, “catechol oxidase activity” and “trans-cinnamate 4-monooxygenase activity” in the above seven comparison groups. Moreover, “chitinase activity”, “sulfur compound binding” and “caffeate O-methyltransferase activity” were only in G0 versus G1, G1 versus G2, G2 versus G3 and G3 versus G4.

To further systematically understand the molecular interactions among the DEGs, we performed KEGG analysis using KOBAS software. Significantly enriched KEGG pathway “sesquiterpenoid and triterpenoid biosynthesis” was shared in G0 versus G1, G0 versus G2 and G0 versus G3, but not in G0 versus G4 (Supplemental Table S4). However, “sesquiterpenoid and triterpenoid biosynthesis” pathway comes from “terpenoid backbone biosynthesis”, which was both in G0 versus G3 and G0 versus G4. Furthermore, the number of DEGs belonging to the triterpenoid biosynthesis pathway were gradually increased from G0 versus G1 to G0 versus G4. In addition, “flavonoid biosynthesis” pathway was enriched in G0 versus G2, G0 versus G3 and G0 versus G4. Moreover, “plant hormone signal transduction” pathway was enriched in G0 versus G1, G0 versus G2, G0 versus G3, G0 versus G4, G2 versus G3 and G3 versus G4, but not in G1 versus G2. Thus, the above four candidate pathways were considered to be remarkably involved in anthocyanin accumulation.

Functional classification of DEGs in ‘Red Delicious’ and its four generation mutants

In order to further identify the major functions of DEGs and establish the pericarp pigment transcriptome, clustering analysis was applied to the 3,466 DEGs. These

genes grouped into six expression patterns (Fig. 3, Supplemental Data Set S3). Cluster 1 contained 561 DEGs whose expression peaked at G0/G1/G2. Cluster 2 contained 336 DEGs whose expression peaked at G3/G4. Cluster 3 contained 363 DEGs whose expression peaked at G2. Further, 1049 and 348 DEGs whose expression peaked at G0 were included in Cluster 4 and 5, respectively. Cluster 6 contained 809 DEGs whose expression peaked at G4. KEGG analysis also was carried out for DEGs belonging to each pattern with *P*-value ≤ 0.01. The expression pattern of cluster 2 was positively consistent with total anthocyanin content (Fig. 1B), whereas cluster 4 and 5 were negatively aligned. DEGs in cluster 2 were significantly enriched in "plant hormone signal transduction", "flavonoid biosynthesis", "flavone and flavonol biosynthesis", "phenylalanine metabolism", and "phenylpropanoid biosynthesis". Among them, part of the final metabolites of "flavonoid biosynthesis", "flavone and flavonol biosynthesis", "phenylalanine metabolism", and "phenylpropanoid biosynthesis" are anthocyanin (Massonnet *et al*., 2017). Remarkably, the pathway that were co-enriched by cluster 4 and 5 was "sesquiterpenoid and triterpenoid biosynthesis", which was hypothesized to be negatively related to the accumulation of anthocyanins. In addition, the pathway of "terpenoid backbone biosynthesis" also was enriched in cluster 1, "flavonoid biosynthesis" and "phenylalanine metabolism" were enriched in cluster 3. Overall, the results of the functional classification analysis of the common expression patterns of DEGs combined with KEGG were in accordance to the results of the comparative transcriptome enrichment analysis aforementioned. Therefore, pathways including flavonoid/phenylpropanoid biosynthesis, terpenoid biosynthesis and plant hormone signal transduction were selected for subsequent analysis.

Key candidate DEGs responsible for anthocyanin accumulation that received in 'Red Delicious' and its four generation mutants

Genes involved in phenylpropanoid/flavonoid biosynthesis pathway

Variety-specific trends in the expression of phenylpropanoid/flavonoid biosynthesis pathway genes at S2 were investigated by preparing heat maps (Fig. 4A). We focused

on the 28 DEGs involved in this pathway, among which 4 from cluster 1, 13 from cluster 2, 10 from cluster 3, and 1 from cluster 5 (Supplemental Table S5). Differences in the expression pattern of these genes from cluster 2 were found among 'Red Delicious' and its four generation mutants, strictly mirroring the differences in the total anthocyanin concentration at S2. For example, all of the DEGs involved in phenylpropanoid biosynthesis pathway, including one aspartate aminotransferase cytoplasmic (*ASP3*), one Phe ammonia lyase (*PAL*), two beta-glucosidase (*BGLU*), one 4-coumarate-CoA ligase (*4CL*) and three peroxidase (*PER*) were from cluster 2, which showing gradually increased expression pattern from G0 to G4 (Fig. 4A; Supplemental Data Set S4). Among them, *ASP3* participates in the synthesis between phenylalanine and phenylpyruvate, which is the precursor of anthocyanin synthesis. *BGLU* and *PER* involved in coumarine and lignin biosynthesis, respectively. *PAL* and *4CL* are the phenylpropanoid genes. In addition, down-regulated genes including two quinate hydroxycinnamoyl transferase (*HCT*) plus one caffeoyl-CoA O-methyltransferase (*CCoAOMT*), and up-regulated genes including three cytochrome P450 98A2-like (*CYP98A*), four *CHS*, one avanone-3-hydroxylase (*F3H*), one dihydroflavonol reductase (*DFR*), and one anthocyanidin synthase (*ANS*) were involved in the synthesis of delphinidin from phenylpropanoid biosynthesis pathway. Further, four *CHS* described above, one down-regulated and another up-regulated *CHI* genes, two *F3'H*, along with aforementioned *F3H*, *DFR* and *ANS w*ere involved in the synthesis of cyanidin. Removal of *F3'H*, other genes involved in pelargonidin synthesis were the same as those in cyanidin synthesis. These data confirming that delphinidin, cyanidin and pelargonidin are the three main substances responsible for the synthesis of anthocyanin from phenylpropanoid biosynthesis pathway in apple. Moreover, two down-regulated *FLS* genes involved in flavone and flavonol biosynthesis.

Genes involved in myb-like and helix-loop-helix DNA-binding domain transcriptional factors

Correlative analysis was carried out between anthocyanin content at S2 and the

expression levels of transcription factors that encode myb-like and helix-loop-helix DNA-binding domain in DEGs (Supplemental Data Set S5, 6). As a result, 13 MYB and 5 bHLH transcription factors were screened. Among them, the expression of MYB members including *LUX*, *MYB113*, *PCL-like*, *MYB1R1-like*, *MYB6-like*, *MYB308-like* and *MYB5-like*, bHLH members including *bHLH51* and *bHLH155* presented a notable positive correlation with the anthocyanin content at S2 ($P<0.05$) (Fig. 4B, C). Other MYB members of *LHY-like*, *RVE6*, *GT-3b*, *MYB21*, DNAJC2-like and *MYB4-like*, bHLH members of *bHLH62*, *bHLH74* and *bHLH162* had a remarkably negative correlation with the anthocyanin content at S2 ($P<0.05$).

Genes Involved in Terpenoid Biosynthesis and Plant Hormone Signal Transduction

KEGG analysis showed terpenoid backbone biosynthesis was over-presented in cluster 1, sesquiterpenoid and triterpenoid biosynthesis branching from the terpenoid backbone biosynthesis was over-presented in cluster 4, and 5 (Fig. 3), indicating that this pathway may play important roles in anthocyanin accumulation of apple. Meanwhile, plant hormone signal transduction pathway, which was modulated by terpenoid biosynthesis, was enriched in cluster 2. A total of 18 DEGs in terpenoid biosynthesis pathway (Supplemental Data Set S7; Supplemental Table S6) and 12 DEGs in plant hormone signal transduction pathway (Supplemental Data Set S8; Supplemental Table S7) also were investigated by preparing heat maps (Fig. 5). The expression of DEGs involved in terpenoid backbone biosynthesis, including one acetyl-CoA acetyltransferase, cytosolic 1 (*AACT1*), two hydroxymethylglutaryl-CoA synthase-like (*HMGS*), two 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1-like (*HMGR*), one mevalonate kinase-like (*MVK*), two diphosphomevalonate decarboxylase MVD2-like (*MVD2*), one isopentenyl-diphosphate Delta-isomerase I (*IDI1*), three farnesyl pyrophosphate synthase 2-like (*FPPS2*) and six squalene monooxygenase-like (*SQMO*) were gradually down-regulated from G0 to G4. Nevertheless, two auxin transporter-like 1 (*AUX1*), four auxin-responsive protein *AUX/IAA* and three auxin-responsive protein *SAUR* involved in tryptophan metabolism of auxin biosynthesis, coupled with one F-box protein GID2-like (*GID2*)

involved in diterpenoid biosynthesis of GA biosynthesis, as well as one protein phosphatase 2C (*PP2C*) and one sucrose non-fermenting-1-related protein kinase 2 (*SnRK2*) involved in carotenoid biosynthesis of ABA biosynthesis, were up-regulated from G0 to G4.

To further evaluate the validity of our results, 16 representative DEGs used previously in Fig. 4 and 5 were selected for expression levels examination by qRT-PCR (Supplemental Table S8). The overall trend of relative expression levels at three stage were consistent with that of deep sequencing at S2 (Fig. 6), suggesting that the candidate genes involved in pathways including phenylpropanoid/flavonoid biosynthesis, terpenoid biosynthesis and plant hormone signal transduction, appended with MYB and bHLH transcriptional factors were directly correlated with anthocyanin accumulation.

Discussion

The activation of early phenylpropanoid biosynthesis pathway genes were more responsible for anthocyanin accumulation

In plants, phenylpropanoid biosynthesis gives rise to a large number of secondary metabolites, including hydroxycinnamic acids, monolignols/lignin, coumarins, benzoicacids, stilbenes, anthocyanins and flavonoids, serving different functions in plant development, reproduction, defence, and protection against biotic/abiotic stresses (Voge, 2010; Zhang et al., 2013). Differences in the expression pattern of genes involved in phenylpropanoid/flavonoid biosynthesis resulting in diverse anthocyanin profiles (El-Sharkawy *et al*., 2015; Massonnet *et al*., 2017). Our survey provided a comprehensive profile of the phenylpropanoid/flavonoid biosynthesis pathway in ‘Red Delicious’ and its four generation mutants, and showed that all of the early phenylpropanoid biosynthesis pathway genes including *ASP3*, *PAL*, *4CL*, *BGLU* and *PER* were aggregated in cluster 2 (Fig. 4A; Supplemental Table S5), which matched with anthocyanin content (Fig. 1A, B). Other genes in cluster 2 containing *CHS*, *CYP98A* and *F3’H* involved in the middle steps of phenylpropanoid pathway, that is the early steps of flavonoid biosynthesis pathway. Nevertheless, genes

encoding *HCT*, *CCoAOMT*, *CHS*, *CYP98A*, *CHI*, *F3H*, *DFR*, *FLS* and *ANS*, were involved in the middle and late steps of phenylpropanoid biosynthesis pathway, and gathered in cluster 1 and 3. Thus, we demonstrated that the activation of early phenylpropanoid biosynthesis pathway genes were more responsible for pigment accumulation in apple pericarp. In addition, aforementioned *ASP3*, *BGLU* and *PER* were confirmed to be involved in the synthesis of phenylpyruvate, coumarine and lignin, respectively (Fig. 4A). Interestingly, 44 stilbene synthase (*STS*) genes involved in stilbene biosynthesis were characterized to influence anthocyanin accumulation during grapevine (*Vitis vinifera*) maturation by Massonnet *et al* (2017), nevertheless, it does not exist in our 3,466 DEGs. This may be due to its varietal-specific (Zenoni *et al*., 2016).

MYB and bHLH modulated anthocyanin accumulation in apple pericarp by regulating the transcription of genes involved in phenylpropanoid/flavonoid pathway

MYB and bHLH autonomously mediated the transcription of genes involved in the middle steps of phenylpropanoid pathway, that is the early steps of the flavonoid biosynthetic pathway (*CHS*, *CHI*, *F3H*, *F3'H* and *FLS*) which leads to the production of colourless dihydroflavonol compounds (Baudry et al., 2004; Stracke et al., 2007; Petroni and Tonelli, 2011; Hu *et al*., 2016; An *et al*., 2017). *AtMYB4*, *AmMYB308*, *FaMYB1*, *ZmMYB31*, *ZmMYB42*, *PhMYB4* and *VvMYBC2-L1* were demonstrated to repress phenylpropanoid synthesis, probably via repression of the expression of synthesis genes (Tamagnone *et al*., 1998; Aharoni *et al*., 2001; Colquhoun *et al*., 2011; Huang *et al*., 2014). We corroborated that MYB members including *LUX*, *MYB113*, *PCL-like*, *MYB1R1-like*, *MYB6-like*, *MYB308-like* and *MYB5-like*, bHLH members including *bHLH51* and *bHLH155* notable positive correlation with the anthocyanin content (Supplemental Data Set S5, 6; Fig. 4B, C), were confirmed to promote anthocyanin synthesis by mediating the transcription of structural genes, *CHS* and *F3'H* involved in flavonoid pathway. Other MYB members of *LHY-like*, *RVE6*, *GT-3b*, *MYB21*, *DNAJC2-like* and *MYB4-like*, bHLH members of *bHLH62*, *bHLH74* and

bHLH162 had a remarkably negative correlation with the anthocyanin content, were demonstrated to repress anthocyanin synthesis. However, the mechanism of which genes involved in anthocyanin synthesis specifically mediated by MYB and bHLH is still not clear, and requires further investigate.

Terpenoid biosynthesis modulated anthocyanin accumulation by positively regulating plant hormone signal transduction during maturation

Hormones are important factors inducing anthocyanin accumulation (Jiang and Fu, 2007; Loreti *et al*., 2008; Shan 2009). Carvalho *et al*. (2010) have provided evidences that anthocyanin accumulation was promoted by exogenous ABA and CTK, and inhibited by GA in tomato hypocotyls. Co-treatment of IAA and CTK significantly enhanced the cytokinin-induced increase in anthocyanin levels but too high auxin concentration strongly inhibited anthocyanin synthesis even in the presence of cytokinin in callus cultures of red-fleshed apple (*M. sieversii* f.*niedzwetzkyana*) as shown by Ji *et al*. (2014). Our results showed that the sesquiterpenoid and triterpenoid biosynthesis along with plant hormone transduction, including tryptophan metabolism for IAA, diterpenoid biosynthesis for GA and carotenoid biosynthesis for ABA, branching from the general terpenoid backbone synthesis pathway and sharing the same precursors of glycolysis (Fig. 5). Among them, sesquiterpenoid and triterpenoid biosynthesis pathway comprised with six *SQMO* was in cluster 4 and 5, which negatively correlated with anthocyanin content. Genes of *AACT1*, *HMGS*, *HMGR*, *MVK*, *MVD2*, *IDI1* and *FPPS2* in the early steps of terpenoid backbone synthesis were in cluster 1 also reflected the differential accumulation of anthocyanin to some extent. Moreover, *AUX1*, *AUX/IAA*, *SAUR*, *GID2*, *PP2C* and *SnRK2* belonging to plant hormone transduction pathway were in cluster 2, which positively reflected the anthocyanin content. Likewise, the contents of IAA and ABA were increased, while GA were decreased with the maturation and pigment accumulation (Fig. 1C, E). Therefore, we corroborated that there may be some form of crosstalk between the activation of phenylpropanoid biosynthesis and plant hormone signal transduction in apple pigment accumulation.

Conclusions

We have investigated the pericarp transcriptome of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') and also have identified specific processes that lead to the accumulation of anthocyanin. The results emphasized that the terpenoid biosynthesis influence anthocyanin accumulation by positively regulating the synthesis of IAA and ABA, while negatively regulating the synthesis of GA. MYB and bHLH modulated anthocyanin accumulation in apple pericarp by regulating the transcription of genes, *CHS* and *F3'H*. *ASP3*, *CYP98A* and *CCoAOMT* are novel anthocyanin-associated genes first reported in apple of the present study. This novel set of genes provides not only new insights into the anthocyanin biosynthesis, but also important clues for more dedicated studies to broaden our knowledge of the anthocyanin pathway in apple.

Acknowledgments

This research was financially supported by the Key Scientific Technology Research Projects of Gansu Province (GPCK2013-2), and the Fostering Foundation for the Excellent Ph.D. Dissertation of Gansu Agricultural University (2017002).

References

Aharoni A, De Vos CHR, Wein M, Sun Z, Greco R, Kroon A, Mol JNM, O'Connell AP. 2001. The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant Journal **28,** 319–332.

An JP, Liu X, Li HH, You CX, Wang XF, Hao YJ. 2017. Apple RING E3 ligase MdMIEL1 inhibits anthocyanin accumulation by ubiquitinating and degrading MdMYB1 protein. Plant Cell Physiology **58(11),** 1953–1962.

An XH, Tian Y, Chen KQ, Wang XF, Hao YJ. 2012. The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin

accumulation. Journal of Plant Physiology **169:** 710–717.

Arrizabalaga M, Morales F, Oyarzun M, Delrot S, Gomès E, Irigoyen JJ, Hilbert G, Pascual I. 2018. Tempranillo clones differ in the response of berry sugar and anthocyanin accumulation to elevated temperature. Plant Science **267,** 74.

Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. 2004. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. Plant Journal **39,** 366–380.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society **57,** 289–300.

Carvalho RF, Vera Q, Lep P. 2010. Hormonal modulation of photomorphogenesis-controlled anthocyanin accumulation in tomato (*Solanum lycopersicum* L. cv Micro-Tom) hypocotyls: physiological and genetic studies. Plant Science **178(3),** 258–264.

Colquhoun TA, Kim JY, Wedde AE, Levin LA, Schmitt KC, Schuurink RC, Clark DG. 2011. *PhMYB4* fine-tunes the floral volatile signature of *Petunia×hybrida* through *PhC4H*. Journal of Experiment Botany **62,**1133–1143.

Daccord N, Celton J M, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, *et al*. 2017. High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetic **49(7),** 1099.

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, Robinson SP, Barrieu F. 2008. The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiology **147,** 2041–2053.

El-Kereamy A, Chervin C, Roustan JP, Cheyhnier V, Souquet JM, Moutounet M, Raynal J, Ford C, Latché A, Pech JC, Bouzayen M. 2003. Exogenous ethylene stimulates the long term expression of genes related to anthocyanin biosynthesis in grape berries. Physiologia Plantarum **119,** 175–282.

El-sharkawy I, Dong L, Xu K. 2015. Transcriptome analysis of an apple *(Malus* ×

domestica) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. Journal of Experiment Botany **66(22),** 7359–7376.

Gomi K, Sasaki A, Itoh H, Ueguchi-Tanaka M, Ashikari M, Kitano H, Matsuoka M. 2004. GID2, an F-box subunit of the SCF E3 complex, specifically interacts with phosphorylated SLR1 protein and regulates the gibberellin-dependent degradation of SLR1 in rice. Plant Journal **37(4),** 626–634.

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. Plant Journal **53,** 814–827.

Hu DG, Sun CH, Ma QJ, You CX, Cheng L, Hao YJ. 2016. MdMYB1 regulates anthocyanin and malate accumulation by directly facilitating their transport into vacuoles in apples. Plant Physiology **170(3),** 1315–1330.

Huang YF, Vialet S, Guiraud JL, Torregrosa L, Bertrand Y, Cheynier V, This P, Terrier N. 2014. A negative MYB regulator of proanthocyanidin accumulation, identified through expression quantitative locus mapping in the grape berry. New Phytologist **201(3),** 795–809.

Ito S, Yamagami D, Umehara M, Hanada A, Yoshida S, Sasaki Y, *et al*. 2017. Regulation of strigolactone biosynthesis by gibberellin signaling. Plant Physiology **174(2),** 1250–1259.

Jeong ST, Goto-Yamamoto N, Kobayashi S, Esaka M. 2004. Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skins. Plant Science **167(2),** 247–252.

Ji XH, Wang YT, Zhang R, Wu SJ, An MM, Li M, Wang CZ, Chen XL, Zhang YM, Chen XS. 2015. Effect of auxin, cytokinin and nitrogen on anthocyanin biosynthesis in callus cultures of red-fleshed apple (*Malus sieversii*, f. *niedzwetzkyana*). Plant Cell Tissue and Organ Culture **120(1),** 325–337.

Jiang C, Fu X. 2007. Phosphate starvation root architecture and anthocyanin accumulation responses are modulated by the gibberellin-DELLA signaling pathway in Arabidopsis. Plant Physiology **145(4),** 1460–70.

Liu Z, Shi MZ, Xie DY. 2014. Regulation of anthocyanin biosynthesis in *Arabidopsis thaliana* red pap1-D cells metabolically programmed by auxins. Planta **239(4),** 765.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. Methods **25,** 402–408.

Loreti E, Povero G, Novi G, Solfanelli C, Alpi A, Perata P. 2008. Gibberellins, jasmonate and abscisic acid modulate the sucrose-induced expression of anthocyanin biosynthetic genes in Arabidopsis. New Phytologist **179(4),** 1004–1016.

Massari ME, Murre C. 2000. Helix-loop-helix proteins:regulators of transcription in eucaryotic organisms. Journal of Molecular Cell Biology **20,** 429–440.

Massonnet M, Fasoli M, Tornielli GB, Altieri M, Sandri M, Zuccolotto P, Paci P, Gardiman M, Zenoni S, Pezzott M. 2017. Ripening transcriptomic program in red and white grapevine varieties correlates with berry skin anthocyanin accumulation. Plant Physiology **174(4),** 2376–2396.

Meng R, Zhang J, An L, Zhang B, Jiang X, Yang Y, Zhao Z. 2016. Expression profiling of several gene families involved in anthocyanin biosynthesis in apple (*Malus domestica*, Borkh.) skin during fruit development. Journal of Plant Growth Regulation **35(2),** 449-464.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods **5,** 621–628.

Petit RJ, Hampe A. 2006. Some evolutionary consequences of being a tree. Annual Review of Ecology Evolution and Systematics **37,** 187–214.

Petroni K, Tonelli C. 2011. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Science **181,** 219–229.

Ramsay NA, Glover BJ. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends in Plant Science **10,** 63–70.

Shan X. 2009. Molecular mechanism for jasmonate-induction of anthocyanin accumulation in Arabidopsis. Journal of Experiment Botany **60(13),** 3849–3860.

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. 2007. Differential regulation of closely related R2R3-MYB transcription factors

controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant Journal **50,** 660–677.

Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martin C. 1998. The AmMYB308 and AmMYB330 transcription factors from Antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. Plant Cell **10,** 135–154.

Telias A, Linwang K, Stevenson D E, Cooney JM, Hellens RP, Allan AC, Hoover EE, Bradeen JM. 2011. Apple skin patterning is associated with differential expression of MYB10. BMC Plant Biology **11(1),** 93.

Vasudevan HN, Mazot P, He F, Soriano P. 2015. Receptor tyrosine kinases modulate distinct transcriptional programs by differential usage of intracellular pathways. eLife **4,** e07186.

Vogt T. 2010. Phenylpropanoid biosynthesis. Molecular Plant **3,** 2-20.

Williams RS, Benkeblia N. 2018. Biochemical and physiological changes of star apple fruit (*Chrysophyllum cainito*), during different "on plant" maturation and ripening stages. Scientia Horticulturae **236,** 36–42.

Xu Y, Feng S, Jiao Q, Liu C, Zhang W, Chen W, Chen X. 2012. Comparison of MdMYB1 sequences and expression of anthocyanin biosynthetic and regulatory genes between *Malus domestica* Borkh. cultivar 'Ralls' and its blushed sport. Euphytica **185,** 157–170.

Zenoni S, Fasoli M, Guzzo F, Dal Santo S, Amato A, Anesi A, Commisso M, Herderich M, Ceoldo S, Avesani L *et al.* 2016. Disclosing the molecular basis of the postharvest life of berry in different grapevine genotypes. Plant Physiology **172,** 1821–1843.

Zhang Y, Butelli E, De Stefano R, Schoonbeek H, Magusin A, Pagliarani C, Wellner N, Hill L, Orzaez D, Granell A. 2013. Anthocyanins double the shelf life of tomatoes by delaying overripening and reducing susceptibility to gray mold. Current Biology **23,** 1094–1100.

Figure legends

Fig. 1. A, Close-up views of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') fruit named as G0 to G4 at three developmental stages (S1–S3) used for anthocyanin quantification, transcriptome profiling and qRT-PCR. B, The changes of total anthocyanin concentrations in the pericarp of the five strains at S1 to S3. The changes of endogenous hormone levels including IAA (C), GA (D) and ABA (E) in the pericarp of the five strains at S1 to S3. Values are means ± SE. Different lower case letters indicate the significant differences among the five strains ($P = 0.05$).

Fig. 2. Summary of the number of differentially expressed genes (DEGs) identified by RNA-seq analysis in the pericarp tissues of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') at S2, named as G0 to G4. Number of total DEGs (A), up-regulated DEGs (B), down-regulated DEGs (D) were presented by Venn diagrams (FDR < 0.01 and fold change ≥ 2). D, Number of total up-regulated and down-regulated DEGs. The histogram represents the number of commonly down-regulated (blue) and up-regulated (yellow) DEGs.

Fig. 3. Gene expression profiles and KEGG category distribution of the DEGs in the six common expression clusters composing the pericarp transcriptome at S2. Clusters were derived by coupled clustering analysis of the 3,466 commonly modulated genes. A, Heat map of the overall common expression pattern. B, Each single line represents the $\log_2$-transformed average of the mean FPKM values for an individual transcript. Significantly overrepresented KEGG categories are represented by red dot. KEGG category enrichment was computed using R language, Cluster package, Biobase package, and *Q*-value package ($P \leq 0.01$).

Fig. 4. A, Differential expression of genes involved in phenylpropanoid/flavonoid biosynthesis pathway hormone signal transduction pathway in 'Red Delicious' and its

four generation mutants (‘Starking Red’, ‘Starkrimson’, ‘Campbell Redchief’ and ‘Vallee spur’) named as G0 to G4. Differential expression of transcription factors that encode myb-like (B) and helix-loop-helix (C) DNA-binding domain. Heatmaps depict the normalized gene expression values, which represent means ±SD of three biological replicates. Expression values of five libraries are presented as FPKM normalized $\log_{10}$-transformed counts.

Fig. 5. Differential expression of genes involved in sesquiterpenoid and triterpenoid/terpenoid biosynthesis coupled with hormone signal transduction pathway in ‘Red Delicious’ and its four generation mutants (‘Starking Red’, ‘Starkrimson’, ‘Campbell Redchief’ and ‘Vallee spur’) named as G0 to G4. Heatmaps depict the normalized gene expression values, which represent means ±SD of three biological replicates. Expression values of five libraries are presented as FPKM normalized $\log_{10}$-transformed counts.

Fig. 6. Relationships between total anthocyanin contents and transcript levels of the sixteen representative genes from Fig. 4 and 5. For each accession, the expression was determined in three developmental stages (S1 to S3) of pericarp tissues. For qRT-PCR assay, the mean was calculated from three biological replicates each with three technical replicates (n=9). These were then normalized relative to the expression of *MdGADPH*. The x-axis in each chart is the same and represents different Malus accessions as indicated by names at the bottom panel, which are arranged in different clusters. The left y-axis represents relative expression levels determined by qRT-PCR.

Supplemental Fig. 1. Summary of the number of differentially expressed genes (DEGs) identified by RNA-seq analysis in the pericarp tissues of ‘Red Delicious’ and its four generation mutants (‘Starking Red’, ‘Starkrimson’, ‘Campbell Redchief’ and ‘Vallee spur’) at S2, named as G0 to G4. Number of total DEGs (A), up-regulated DEGs (B), down-regulated DEGs (D) were presented by Venn diagrams (FDR < 0.01 and fold change ≥ 2). D, Number total up-regulated and down-regulated DEGs. The

histogram represents the number of commonly down-regulated (blue) and up-regulated (yellow) DEGs.

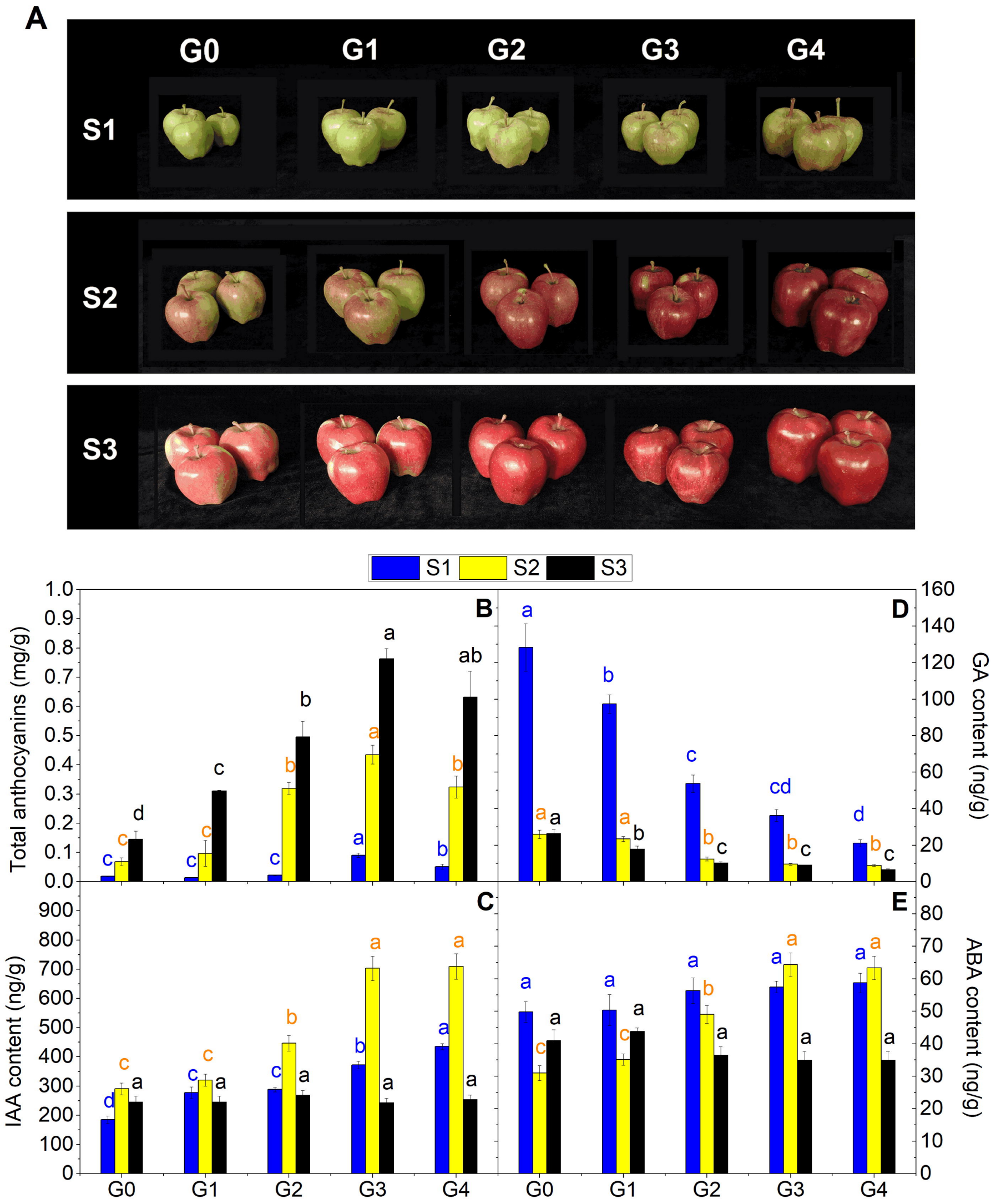

Figure 1. A, Close-up views of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') fruit named as G0 to G4 at three developmental stages (S1-S3) used for anthocyanin quantification, transcriptome profiling and qRT-PCR. B, The changes of total anthocyanin concentrations in the pericarp of the five strains at S1 to S3. The changes of endogenous hormone levels including IAA (C), GA (D) and ABA (E) in the pericarp of the five strains at S1 to S3. Values are means ± SE. Different lower case letters indicate the significant differences among the five strains (P = 0.05).

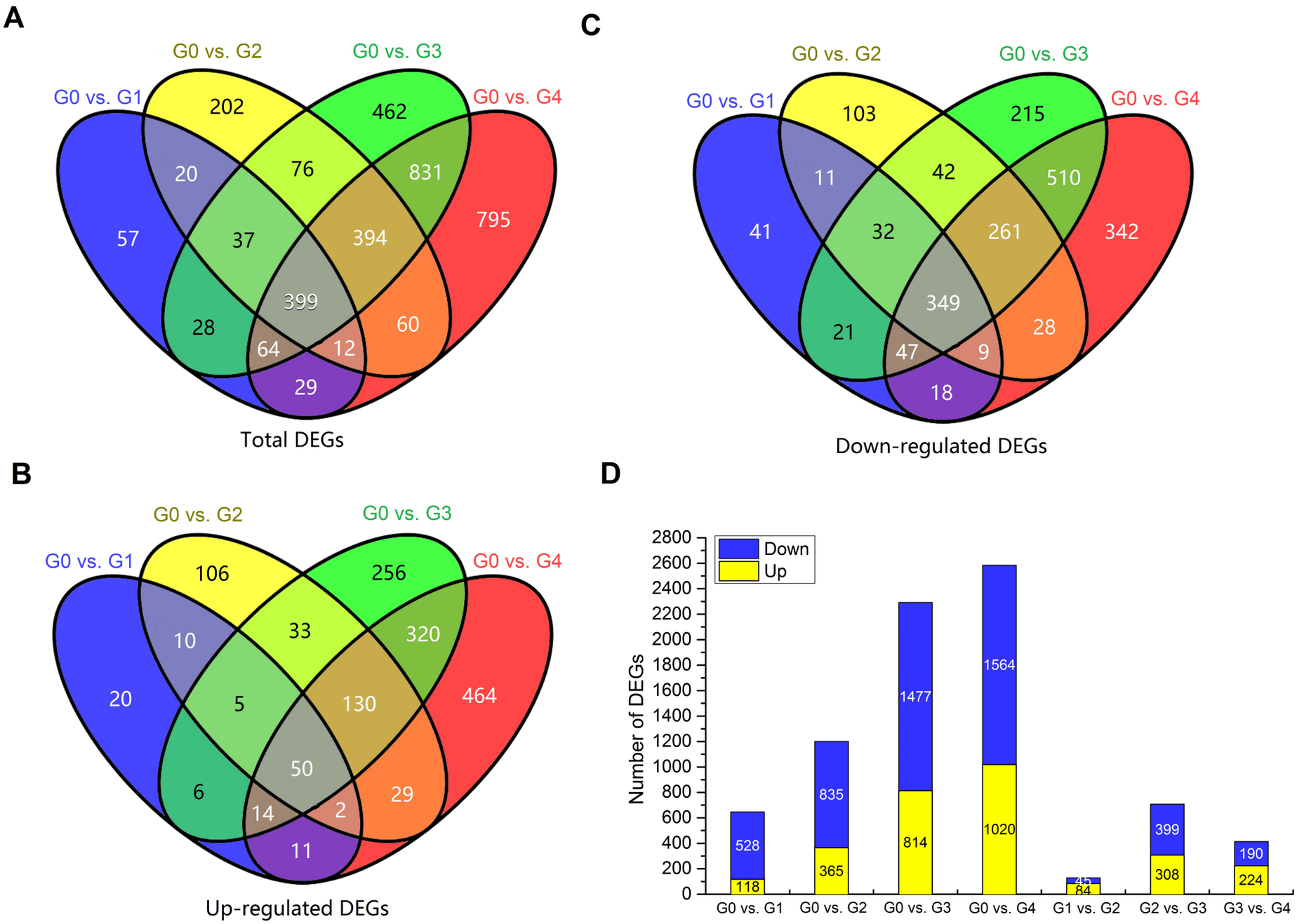

Figure 2. Summary of the number of differentially expressed genes (DEGs) identified by RNA-seq analysis in the pericarp tissues of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') at S2, named as G0 to G4. Number of total DEGs (A), up-regulated DEGs (B), down-regulated DEGs (D) were presented by Venn diagrams (FDR < 0.01 and fold change ≥ 2). D, Number of total up-regulated and down-regulated DEGs. The histogram represents the number of commonly down-regulated (blue) and up-regulated (yellow) DEGs.

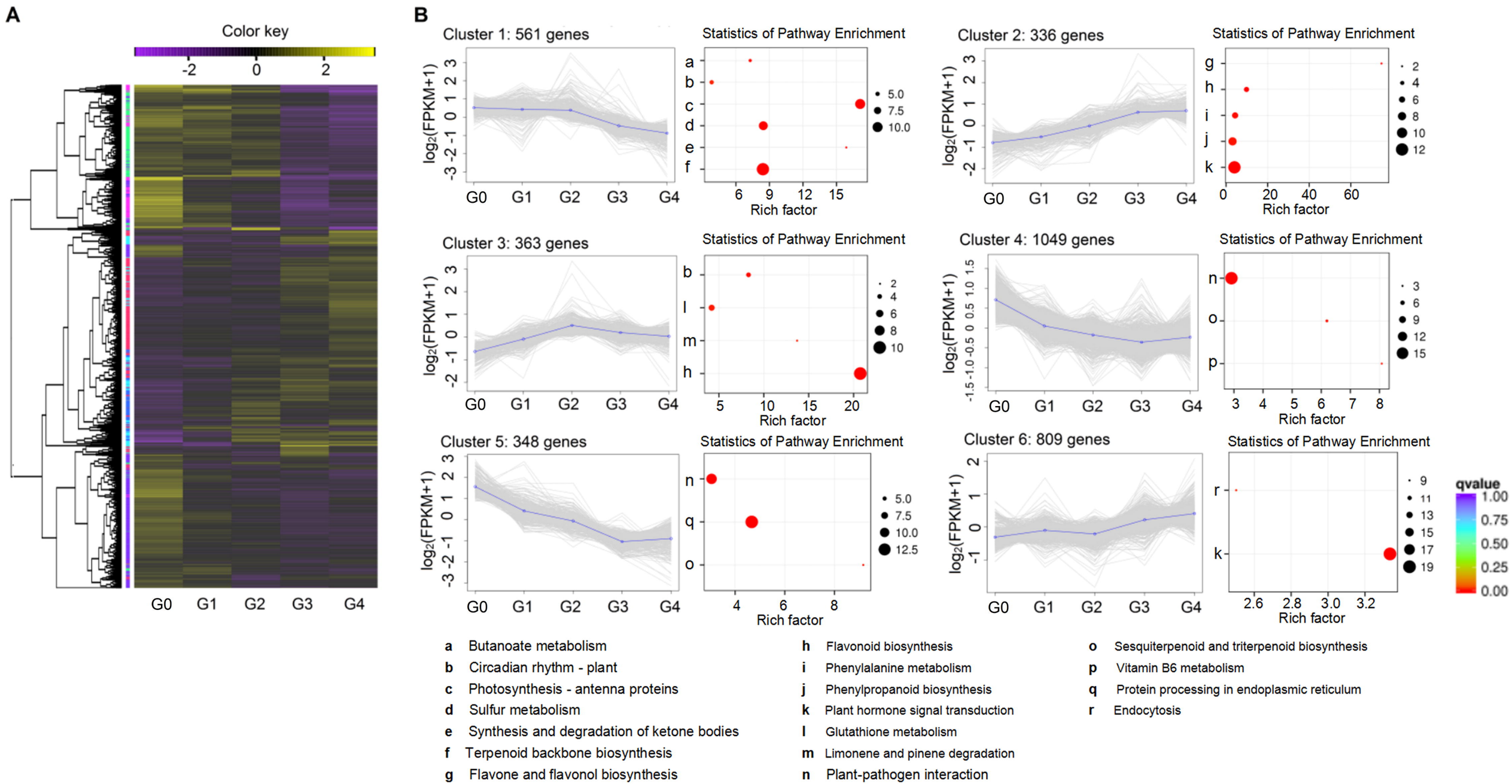

Figure 3. Gene expression profiles and KEGG category distribution of the DEGs in the six common expression clusters composing the pericarp transcriptome at S2. Clusters were derived by coupled clustering analysis of the 3,466 commonly modulated genes. A, Heat map of the overall common expression pattern. B, Each single line represents the $\log_2$-transformed average of the mean FPKM values for an individual transcript. Significantly overrepresented KEGG categories are represented by red dot. KEGG category enrichment was computed using R language, Cluster package, Biobase package, and *Q*-value package ($P \leq 0.01$).

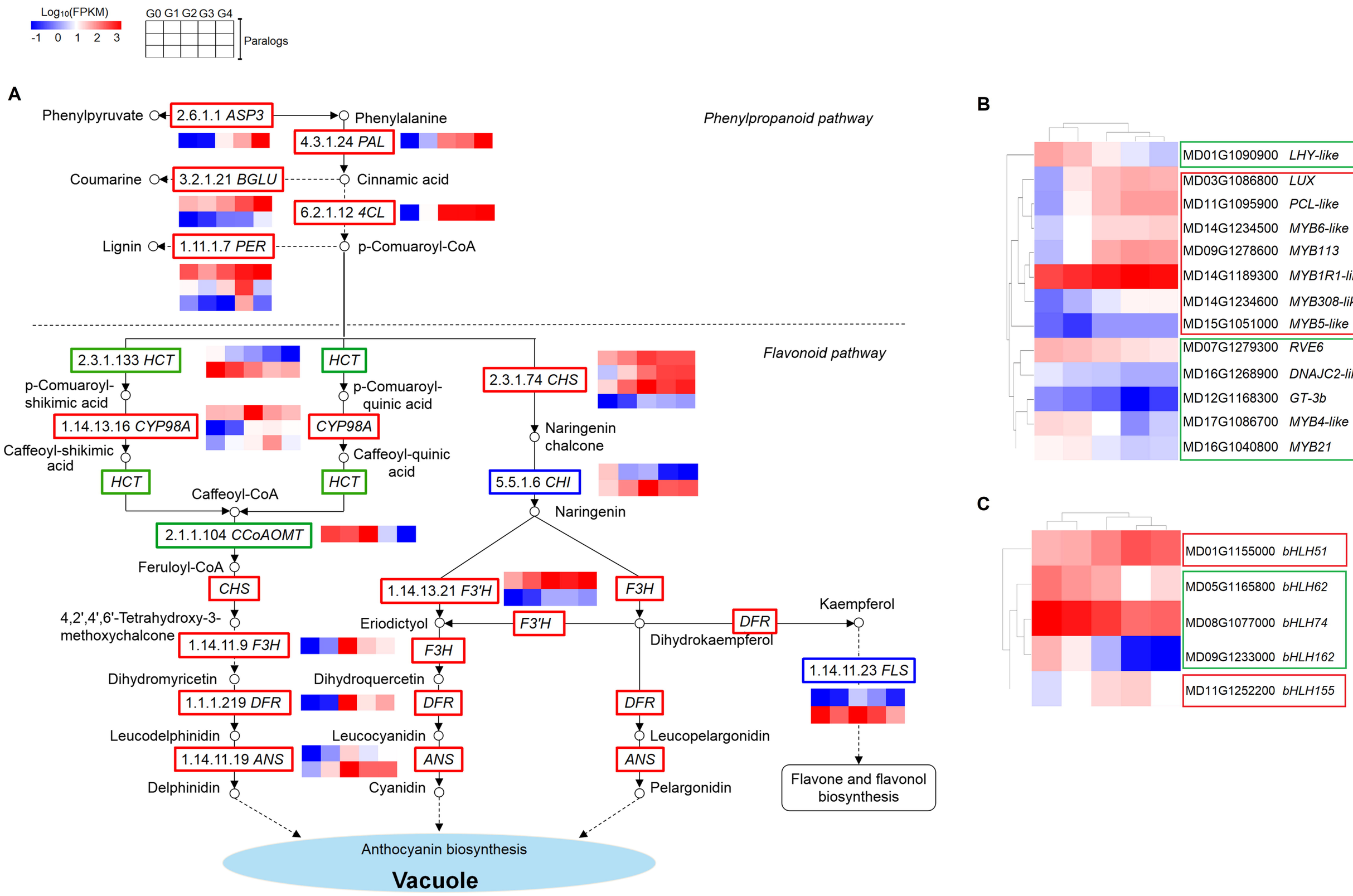

Figure 4. A, Differential expression of genes involved in phenylpropanoid/flavonoid biosynthesis pathway hormone signal transduction pathway in 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') named as G0 to G4. Differential expression of transcription factors that encode myb-like (B) and helix-loop-helix (C) DNA-binding domain. Heatmaps depict the normalized gene expression values, which represent means ±SD of three biological replicates. Expression values of five libraries are presented as FPKM normalized $\log_{10}$-transformed counts.

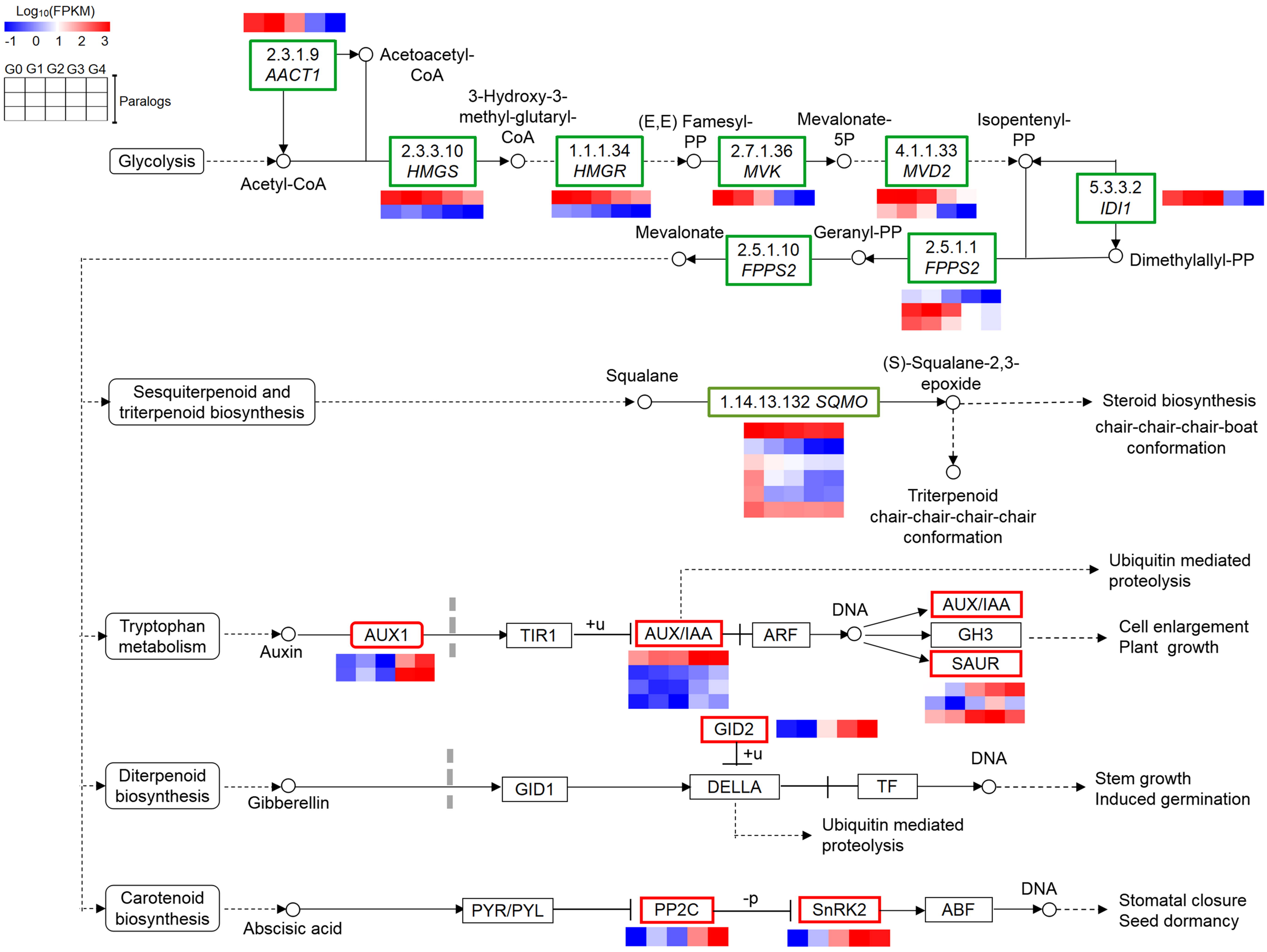

Figure 5. Differential expression of genes involved in sesquiterpenoid and triterpenoid/terpenoid biosynthesis coupled with hormone signal transduction pathway in ‘Red Delicious’ and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') named as G0 to G4. Heatmaps depict the normalized gene expression values, which represent means ±SD of three biological replicates. Expression values of five libraries are presented as FPKM normalized log_{10}-transformed counts.

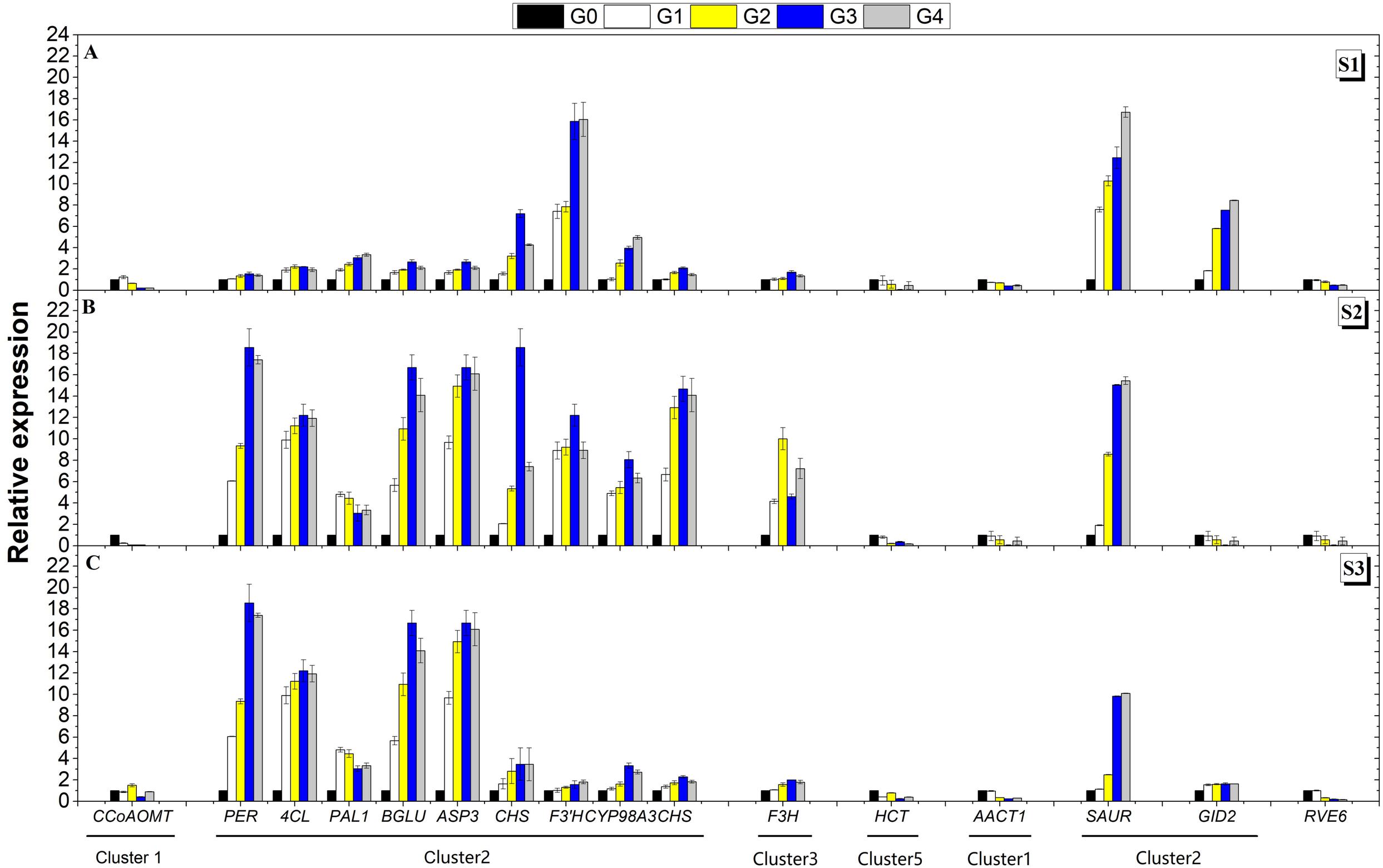

Figure 6. Relationships between total anthocyanin contents and transcript levels of the sixteen representative genes from Fig. 4 and 5. For each accession, the expression was determined in three developmental stages (S1 to S3) of pericarp tissues. For qRT-PCR assay, the mean was calculated from three biological replicates each with three technical replicates (*n*=9). These were then normalized relative to the expression of *MdGADPH*. The x-axis in each chart is the same and represents different Malus accessions as indicated by names at the bottom panel, which are arranged in different clusters. The left y-axis represents relative expression levels determined by qRT-PCR.