Title: Pericarp Pigmentation Correlates with Hormones and Intensifies with Continuation of Bud Sport Generations from 'Red Delicious'

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:**bhch@gsau.edu.cn

Email addresses: 1304784689@qq.com; maojuan-81@163.com; 1947919171@qq.com; 3148123508@qq.com; mazohu@163.com; mmdawuda@yahoo.com; zuocunwu2006@126.com; chu.my@foxmail.com

Running title: Pericarp color correlates with hormones

KEYWORDS apple, 'Red Delicious', bud sport, mutant, pericarp, pigmentation, RNA-seq, hormone

ABSTRACT Bud sport mutants of apple (*Malus domestica* Borkh.) trees with a highly blushed colouring pattern are mainly caused by the accumulation of anthocyanins in the pericarp. Hormones are important factors modulating anthocyanin accumulation. However, a good understanding of the interplay between hormones and anthocyanin synthesis in apples, especially in mutants 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 mutants, including 'Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur',

which were designated G0 to G4, respectively. Pericarp pigmentation gradually proliferated from G0 to G4. The anthocyanin content was higher in the mutants than in 'Red Delicious'. The activation of early phenylpropanoid biosynthesis genes, including *ASP3*, *PAL*, *4CL*, *PER*, *CHS*, *CYP98A* and *F3'H*, was responsible for anthocyanin accumulation in mutants. In addition, IAA and ABA had a positive regulatory effect on the synthesis of anthocyanins, while GA had the reverse effect. The down-regulation of *AACT1*, *HMGS*, *HMGR*, *MVK*, *MVD2*, *IDI1* and *FPPS2* involved in terpenoid biosynthesis influences anthocyanin accumulation by positively regulating transcripts of *AUX1* and *SAUR* that contribute to the synthesis of IAA, *GID2* to GA, *PP2C* and *SnRK2* to ABA. Furthermore, MYB and bHLH members, which are highly correlated (r=0.882–0.980) with anthocyanin content, modulated anthocyanin accumulation by regulating the transcription of structural genes, including *CHS* and *F3'H*, involved in the flavonoid biosynthesis pathway.

INTRODUCTION

Bud sport is a somatic mutation occurring in the shoot cells of perennial fruit trees 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). Although the genetic background of these mutants is nearly identical to that of their parents (Nwafor *et al*. 2014; Otto *et al*. 2014), epigenetic changes have been recognized, such as causing fruit colour alteration in apple (*Malus domestica* Brokh.) (Xu *et al*. 2012; El-Sharkawy *et al*. 2015). Pericarp colour is a key appearance and nutrition quality attribute of apple fruit (Willams and Benkeblia 2018). Anthocyanins are secondary metabolites that contribute to the colours of fruits (Meng *et al*. 2016). Pigmentation in the skin of apple fruit varies among cultivars and is influenced by environmental factors, including temperature (Arrizabalaga *et al*. 2018) and the level of sunlight irradiation (Cominelli *et al*. 2007; Honda and Moriya 2018). Furthermore, hormones are likely to be important factors that modulate light-dependent anthocyanin accumulation (Jeong *et al*. 2004; Carvalho *et al*. 2010; Loreti *et al*. 2010). In summary, exploring the molecular mechanisms of hormones and anthocyanin synthesis in apple

fruit and its bud sport mutants is crucial to research on pigment accumulation and plant somatic mutation.

Previous studies have shown that auxin (IAA), cytokinin (CTK), gibberellins (GA), jasmonate acid (JA), abscisic acid (ABA) and ethylene (ETH) interact in controlling anthocyanin biosynthesis (El-Kereamy *et al*. 2003; Jeong *et al*. 2004; Loreti *et al*. 2010; Liu *et al*. 2014; Ji *et al*. 2014). In addition, the identification and functional characterization of MYB and bHLH transcriptional factors revealed that they play a role in autonomously mediated structural gene transcription. These factors include chalcone isomerase (*CHI*), chalcone synthase (*CHS*), flavonol synthase (*FLS*), leucoanthocyanidin reductase (*LAR*), flavonoid 3'-hydroxylase (*F3'H*) and anthocyanidin reductase (*ANR*), which are involved in the anthocyanin biosynthesis pathway (Deluc *et al*. 2008; Gonzalez *et al*. 2008; Telias *et al*. 2011; Petroni and Tonelli 2011; An *et al*. 2012). MYB proteins are 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 is 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 hormone and anthocyanin synthesis are often applied without a full understanding of the effect 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; Carvalho *et al*. 2010; Loreti *et al*. 2010). Apple cultivar 'Red Delicious' (*M. domestica*) is the most frequently captured sport apple variety that is usually selected on the phenotypic basis of spur type and intense red fruit colour. The cultivar's four continuous generation mutants, namely, 'Starking Red', 'Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee Spur', have been screened. Therefore, these five strains were selected and analysed using a comparative transcriptome combined with physiological and biochemical characteristics to expound the relationships between

hormone and anthocyanin synthesis on apple pericarp pigment accumulation.

MATERIALS AND METHODS

Plant material

'Red Delicious' is the most frequently captured sport apple variety, featuring four continuous generation mutants. 'Starking Red' is a bud sport from 'Red Delicious' and a typical representative of the first generation. The second generation is 'Starkrimson', which is a bud sport from the first-generation 'Starking Red'. The fourth generation, 'Vallee Spur', is a bud sport of the third generation, 'Campbell Redchief', which is bud sport of 'Starkrimson'. Mature apple fruit of these five cultivars range from having red vertical stripes to being completely red (see Figure 1A).

Fruit pericarp samples of 'Red Delicious' and its four continuous 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. Briefly, 20−30 fruits from each of the five strains were sampled at three developmental stages, i.e., 5 August (S1), 25 August (S2), and 14 September (S3) (see Figure 1A). Stages S1, S2 and S3 are equivalent to the pre-veraison, veraison and fruit maturity stages for commercial harvest, respectively. In different experiments, pericarp samples from 6 fruits 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 °C for further analysis of anthocyanin contents, endogenous hormone contents and gene expression profiles (qRT-PCR). In addition, samples from S2 were used for RNA-seq analysis.

Anthocyanin quantification

Lyophilized apple pericarp samples were finely ground, and approximately 500 mg of

119 powdered samples was homogenized in 10 mL of methanol with 1% HCl. The
120 homogenate was transferred to a calibration test tube with a constant volume of 20
121 mL and kept for 20 min at 4 °C with shaking under dark conditions. Then, the
122 samples were filtered through a 0.2 μm polyethersulfone (PES) filter (Krackeler
123 Scientific, Inc., Albany, NY, USA) and analysed using a TU-1900 double beam
124 UV-visible spectrophotometer (Beijing Purkinje General Instrument Co. LTD).
125 Anthocyanin levels were calculated by dividing the absorbance by the coefficient of
126 regression (0.0214) acquired by standard scale measurements.
127
128 **Hormone content measurement**
129 A total of 1.0 g of each lyophilized apple pericarp sample was ground quickly after
130 liquid nitrogen was added and combined with 10 mL of 80% chromatographic pure
131 methanol. Each sample was washed three times with solvent, transferred into a test
132 tube, and stored in a refrigerator at 4 °C overnight in the dark. Then, the samples were
133 centrifuged for 20 min under refrigerated conditions at 4 °C. Supernatant fluid was
134 transferred into a new centrifuge tube. The extract was concentrated, and the methanol
135 was volatilized at 40 °C by rotary evaporation to obtain 2 mL of concentrate. The
136 evaporation bottle wall was then washed continuously with 50% methanol, and the
137 volume was raised to 10 mL with 50% chromatographic pure methanol. The fluid for
138 testing was filtered through a 0.22 μm organic membrane.
139 The determination method was performed with different concentrations of IAA,
140 GA, and ABA, standard samples, which were used to construct a standard curve. The
141 standard samples were purchased from Sigma Company, and the external standard
142 curve and quantitative methods were performed for the measurements. The apparatus
143 used for high-performance liquid chromatography (HPLC) was the LC-20AD system
144 (Shimadzu, Kyoto, Japan) equipped with a Zorbax Eclipse Plus C18 column (4.6 mm
145 × 250 mm × 5.0 μm, Agilent, Palo Alto, CA, USA) and an SPD-20A UV detector. The
146 mobile phase was chromatographic methanol and 0.6% iced acetic acid. The flow
147 velocity was 1.0 mL/min, the wavelength was 254 nm, and the column temperature

148 was 25 °C.

149

150 **RNA extraction**

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

156

157 **Library preparation and sequencing**

158 The 5 triplicate samples (5 varieties at S2) yielded 15 nondirectional cDNA libraries
159 with a total of 68.18 million reads (Table 1), which were prepared from 3.0 μg of total
160 RNA using the NEBNext, Ultra^{TM} RNA Library Prep Kit (NEB, USA). RNA quality
161 and quantity were assessed using a NanoDrop spectrophotometer and an Agilent 2100
162 spectrophotometer (Agilent Technologies, CA, USA). Library fragments were
163 purified with an AMPure XP system (Beckman Coulter, Beverly, USA), and the
164 quality was assessed on an Agilent Bioanalyzer 2100 system. Index-coded samples
165 were clustered on a cBot cluster generation system using the TruSeq PE Cluster Kit
166 v3-cBot-HS (Illumina Inc., San Dego). After cluster generation, library preparations
167 were sequenced on an Illumina HiSeq 2000 platform, in which 125 bp paired-end
168 reads were generated.

169

170 **Analysis of sequencing results: mapping and differential expression**

171 Raw reads were cleaned by removing adapter sequences, reads containing ploy-N,
172 and low-quality sequences ($Q < 20$). Clean reads were aligned onto the apple
173 reference genome (https://iris.angers.inra.fr/gddh13/) (Daccord *et al*. 2017). New
174 transcripts were identified from TopHat alignment results using the Cufflinks v2.1.1
175 reference-based transcript assembly method. An average of 83.95% of reads were
176 mapped for each sample (Table 1). For annotations, all novel genes were searched

against the 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 the 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 the Pfam database using HMMER software, and annotation information of these new genes was obtained.

The DESeq (2010) R package was utilized to detect differentially expressed genes (DEGs). The 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 changes ≥ 2 were used as screening criteria; an absolute value of $\log_2$ (fold change) with reads per kb per million reads (FPKM) ≥ 1.0 was used as a 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 that 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 was considered significantly enriched by DEGs (see Table S1). KOBAS software was used to test the statistical enrichment of different expression genes in KEGG pathways. Pathways with a *Q*-value ≤ 0.05 were defined as genes that displayed significant levels of differential expression (see Table S2).

Common expression pattern clustering analysis of DEGs

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

206 method was K-means clustering or hierarchical clustering.

207

208 **Correlation analysis**

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

214

215 **Quantitative real-time PCR validation of RNA-Seq data**

216 Quantitative reverse transcription RT-PCR analysis DNase-treated RNA (2 μg) was
217 reverse transcribed in a reaction volume of 20 μl using PrimerScript™ RT reagent Kit
218 with gDNA Eraser (Takara, Dalian, China). Gene-specific primers were designed
219 using Primer Express software (Applied Biosystems) (see Table S3). Quantitative
220 reverse transcription PCR (qRT-PCR) assays were performed using 20 ng of cDNA
221 and 300 nM of each primer in a 10 μl reaction with SYBR Green PCR Master Mix
222 (Takara, Dalian, China). Three biological and three technical replicates for each
223 reaction were analysed on a LightCycler® 96 SW 1.1 instrument (Roche). The
224 amplification program consisted of one cycle of 95 °C for 30 s, 40 cycles of 95 °C for
225 5 s, and melting analysis at 60 °C for 34 s, followed by one cycle of 95 °C for 15 s,
226 60 °C for 60 s, and 95 °C for 15 s. Transcript abundance was quantified using
227 standard curves for both target and reference genes, which were generated from serial
228 dilutions of PCR products from corresponding cDNAs. Transcript abundance was
229 normalized to the reference gene *MdGADPH*, which showed high stability across the
230 different apple genotypes and tissues used in this study. Relative gene expression was
231 normalized by comparing with G0 expression and analysed using the comparative
232 $2^{-\Delta\Delta C}$T method (Livak and Schmittgen 2001).

233

234 **Data analysis**

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

Data availability

Supplemental materials available at FigShare. Dataset 1 contains the mean of the normalized expression value per transcript (FPKM fragments per kilobase of mapped reads) of the 5-sample at S2. Dataset 2 contains DEGs identified as commonly up-regulated or down-regulated in each pairwise comparison of 'Red Delicious' and its four generation mutants. Dataset 3 contains gene composition of the six clusters identified using gene expression clustering analysis. Dataset 4 contains gene composition and mean FPKM value of the phenylpropanoid/flavonoid biosynthesis pathway. Dataset 5 contains correlation analysis of anthocyanin content at S2 and MYB transcriptional factors. Dataset 6 contains correlation analysis of anthocyanin content at S2 and bHLH transcriptional factors. Dataset 7 contains gene composition and mean FPKM value of the terpenoid biosynthesis pathway. Dataset 8 contains gene composition and mean FPKM value of the plant hormone signal transduction.

RESULTS

Pericarp pigmentation increased with bud sport generation and maturation

Visual inspection of apple pericarp colour during development revealed that the G2, G3 and G4 strains began colouring at S1, with the most visibly intense red colouring occurring in G3 and G4 (see Figure 1A). Pigmentation of these five strains progressively advanced to much higher levels from S1 to the subsequent stage S3, resulting in red fruit at maturity. In addition, red coloration gradually proliferated from G0 to G4 during the three stages.

Consistent with visual inspection, the analysis of apple pericarp anthocyanin

contents showed that the levels of total anthocyanins in the five strains were initially low and sharply increased with maturation (from S1 to S3) (see Figure 1B). Relative to that at G0, the total anthocyanin level in fruit skin was ~0.78-, ~1.20-, ~5.02- and ~2.85-fold higher in G1, G2, G3 and G4 at stage S1, respectively. At S2, the level in fruit skin was ~1.41-, ~4.65-, ~6.33- and ~4.72-fold higher in G1, G2, G3 and G4, respectively. Finally, at S3, the level in fruit skin was ~2.14-, ~3.41-, ~5.25- and ~4.35-fold higher in G1, G2, G3 and G4, respectively. Briefly, anthocyanin content was lowest in G0 and highest in G3, followed by the contents in G4, G2, and G1. It was concluded that the more intense red colouring pattern in the apple pericarp of bud sport mutants was mainly caused by the accumulation of anthocyanins. In addition, a more blushed colouring pattern was observed with an increase in the number of bud sport generations. However, the third-generation mutant G3 showed greater blushing than did the fourth-generation mutant G4 at S1 and S2, and there was no significant difference at S3.

The contents of IAA and ABA in apple pericarp increased with bud sport generation at veraison, while the content of GA decreased

Hormone levels were also analysed at three time points (S1, S2, and S3). The overall trend of IAA concentrations in apple pericarp first increased and then decreased from S1 to S3 and peaked at S2 (see Figure 1C). However, the GA content decreased from S1 to S3 (see Figure 1D). ABA concentrations of G0, G1 and G2 peaked at S1, whereas those of G3 and G4 peaked at S2 (see Figure 1E). Importantly, the IAA and ABA contents of G3 and G4 in apple pericarp at veraison S2 were considerably higher than those of G2, while G0 and G1 showed considerably lower levels than did G2. Nevertheless, GA concentrations, which displayed a trend opposite that of IAA and ABA, decreased with bud sport generation, that is, from G0 to G4.

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

Triplicate sampling of the pericarp of 'Red Delicious' and its four continuous

294 generation mutants at S2 yielded 15 RNA samples for RNA sequencing (RNA-seq)
295 analysis, and the mapping rate of 20,189,457-26,765,510 clean reads onto the apple
296 reference genome (https://iris.angers.inra.fr/gddh13/) (Daccord *et al*. 2017) ranged
297 from 83.29% to 84.81% (Table 1). The average number of mapped reads ranged from
298 34,093,629 in G2 to 43,153,440.33 in G3. An FDR < 0.01 and fold changes ≥ 2 were
299 used as screening criteria for DEGs; in addition, an FPKM (fragments per kilobase of
300 mapped reads) ≥ 1.0 in at least one of the 5 triplicate samples was considered to be
301 expressed. The mean normalized expression value (FPKM) per transcript of the three
302 biological replicates was calculated for each sample using the geometric
303 normalization method. The resulting dataset comprising 33,192 transcripts was used
304 for subsequent analysis (see Dataset S1).
305
306 **Differentially expressed genes (DEGs) in 'Red Delicious' versus its mutants**
307 **gradually increased with bud sport generation at veraison**
308 To identify DEGs in 'Red Delicious' and its four generation mutants, seven pairwise
309 transcriptome comparisons (i.e., G0 versus G1, G0 versus G2, G0 versus G3, G0
310 versus G4, G1 versus G2, G2 versus G3, and G3 versus G4) were performed at S2
311 (see Figure 2, Figure S1 and Table S1). The total number of DEGs was 3,466,
312 including 1,456 up-regulated DEGs and 2029 down-regulated DEGs (see Figure 2 and
313 Dataset S2). Among them, the number of both up-regulated and down-regulated
314 DEGs increased considerably from the first generation mutant G1 to the fourth
315 generation mutant G4 versus the number observed in G0, and the smallest number
316 was observed in G1 versus that in G2. Moreover, we found that more genes were
317 down-regulated than up-regulated in G1, G2, G3 and G4 relative to G0.
318
319 **Comparative transcriptome enrichment analysis identified key processes**
320 **responsible for anthocyanin accumulation in 'Red Delicious' and its mutants**
321 To understand the major functional categories represented by the DEGs, gene
322 ontology (GO) enrichment analysis was carried out using all reference genes as
323 background via the GOseq R package. GO term enrichment analysis categorized the

annotated sequences into three main categories: biological process, cellular component and molecular function (see Table S2). In the biological process category, three significantly enriched terms, namely, "defence 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 these terms, "1-aminocyclopropane-1-carboxylate biosynthetic process" and "negative regulation of growth" were also enriched in G1 versus G2, G2 versus G3 and G3 versus G4. Furthermore, in addition these two GO terms, three other significantly enriched GO terms, namely, "chitin catabolic process", "regulation of leaf development" and "DNA conformation change", 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 in the seven abovementioned comparison groups: "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". Moreover, "chitinase activity", "sulfur compound binding" and "caffeate O-methyltransferase activity" were only observed 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. The significantly enriched KEGG pathway term "sesquiterpenoid and triterpenoid biosynthesis" was shared in G0 versus G1, G0 versus G2 and G0 versus G3, but not in G0 versus G4 (Table 2). However, the "sesquiterpenoid and triterpenoid biosynthesis" pathway was derived from "terpenoid backbone biosynthesis", which occurred both in G0 versus G3 and

354 G0 versus G4. Furthermore, the number of DEGs belonging to the triterpenoid
355 biosynthesis pathway gradually increased from G0 versus G1 to G0 versus G4. In
356 addition, the "flavonoid biosynthesis" pathway was enriched in G0 versus G2, G0
357 versus G3 and G0 versus G4. Moreover, the "plant hormone signal transduction"
358 pathway was enriched in G0 versus G1, G0 versus G2, G0 versus G3, G0 versus G4,
359 G2 versus G3 and G3 versus G4 but not in G1 versus G2. Thus, the four
360 abovementioned candidate pathways were considered to be heavily involved in
361 anthocyanin accumulation.
362
363 **Functional classification of DEGs in 'Red Delicious' and its four continuous**
364 **generation mutants**
365 To further identify the major functions of DEGs and establish the pericarp pigment
366 transcriptome, clustering analysis was applied to the 3,466 DEGs. These genes were
367 grouped into six expression patterns (see Figure 3 and Dataset S3). Cluster 1
368 contained 561 DEGs whose expression peaked at G0/G1/G2. Cluster 2 contained 336
369 DEGs whose expression peaked at G3/G4. Cluster 3 contained 363 DEGs whose
370 expression peaked at G2. Furthermore, 1049 and 348 DEGs whose expression peaked
371 at G0 were included in clusters 4 and 5, respectively. Cluster 6 contained 809 DEGs
372 whose expression peaked at G4. KEGG analysis was also carried out for DEGs
373 belonging to each pattern with a *P*-value ≤ 0.01. The expression pattern of cluster 2
374 was positively consistent with total anthocyanin content (see Figure 1B), whereas
375 clusters 4 and 5 were negatively aligned. DEGs in cluster 2 were significantly
376 enriched in "plant hormone signal transduction", "flavonoid biosynthesis", "flavone
377 and flavonol biosynthesis", "phenylalanine metabolism", and "phenylpropanoid
378 biosynthesis". Among those, some of the final metabolites of "flavonoid biosynthesis",
379 "flavone and flavonol biosynthesis", "phenylalanine metabolism", and
380 "phenylpropanoid biosynthesis" are anthocyanins (Massonnet *et al*. 2017).
381 Remarkably, the pathway that was co-enriched by clusters 4 and 5 was
382 "sesquiterpenoid and triterpenoid biosynthesis", which was hypothesized to be
383 negatively related to the accumulation of anthocyanins. In addition, the pathway of

384 "terpenoid backbone biosynthesis" was also enriched in cluster 1, and "flavonoid
385 biosynthesis" and "phenylalanine metabolism" were enriched in cluster 3. Overall, the
386 results of the functional classification analysis of the common expression patterns of
387 DEGs combined with KEGG agreed with the results of the aforementioned
388 comparative transcriptome enrichment analysis. Therefore, pathways including
389 "flavonoid/phenylpropanoid biosynthesis", "terpenoid biosynthesis" and "plant
390 hormone signal transduction" were selected for subsequent analysis.
391
392 **Key candidate DEGs responsible for anthocyanin accumulation in 'Red**
393 **Delicious' and its mutants**
394 *Genes involved in phenylpropanoid/flavonoid biosynthesis pathway*
395 Variety-specific trends in the expression of phenylpropanoid/flavonoid biosynthesis
396 pathway genes at S2 were investigated by preparing heat maps (see Figure 4A). We
397 focused on the 28 DEGs involved in this pathway, including 4 from cluster 1, 13 from
398 cluster 2, 10 from cluster 3, and 1 from cluster 5 (Table 3). Differences in the
399 expression pattern of these genes from cluster 2 were found among 'Red Delicious'
400 and its four continuous generation mutants, closely mirroring the differences in total
401 anthocyanin concentration at S2. For example, all DEGs involved in the
402 phenylpropanoid biosynthesis pathway, including one aspartate aminotransferase
403 cytoplasmic (*ASP3*), one Phe ammonia lyase (*PAL*), two beta-glucosidase (*BGLU*),
404 one 4-coumarate-CoA ligase (*4CL*) and three peroxidases (*PER*), were from cluster 2,
405 demonstrating a gradually increasing expression pattern from G0 to G4 (see Figure
406 4A and Dataset S4). Among the DEGs, *ASP3* participates in the synthesis of
407 phenylalanine and phenylpyruvate, which are precursors of anthocyanin synthesis.
408 *BGLU* and *PER* are involved in coumarine and lignin biosynthesis, respectively. *PAL*
409 and *4CL* are phenylpropanoid genes. In addition, down-regulated genes, including
410 two quinate hydroxycinnamoyl transferase (*HCT*) and one caffeoyl-CoA
411 O-methyltransferase (*CCoAOMT*), and up-regulated genes, including three
412 cytochrome P450 98A2-like (*CYP98A*), four *CHS*, one avanone-3-hydroxylase (*F3H*),
413 one dihydroflavonol reductase (*DFR*), and one anthocyanidin synthase (*ANS*), are

414 involved in the synthesis of delphinidin from the phenylpropanoid biosynthesis
415 pathway. Furthermore, the four *CHS* genes described above, one down-regulated and
416 one up-regulated *CHI* gene, two *F3'H* genes, and the aforementioned *F3H*, *DFR* and
417 *ANS* are involved in the synthesis of cyanidin. *F3'H* and other genes involved in
418 pelargonidin synthesis are the same as those involved in cyanidin synthesis. These
419 findings confirm that delphinidin, cyanidin and pelargonidin are the three main
420 substances responsible for the synthesis of anthocyanins via the phenylpropanoid
421 biosynthesis pathway in apple. Moreover, two down-regulated *FLS* genes are
422 involved in flavone and flavonol biosynthesis.
423
424 *Genes involved in myb-like and helix-loop-helix DNA-binding domain transcriptional*
425 *factors*
426 Correlative analysis was carried out between anthocyanin content at S2 and the
427 expression levels of transcription factors that encode myb-like and helix-loop-helix
428 DNA-binding domains in DEGs (see Dataset S5, 6). As a result, 13 MYB and 5
429 bHLH transcription factors were screened. Among them, the expression of MYB
430 members, including *LUX*, *MYB113*, *PCL-like*, *MYB1R1-like*, *MYB6-like*, *MYB308-like*
431 and *MYB5-like*, bHLH members, including *bHLH51* and *bHLH155*, presented a
432 notable positive correlation with anthocyanin content at S2 ($P<0.05$) (see Figure 4B,
433 C). MYB members *LHY-like*, *RVE6*, *GT-3b*, *MYB21*, DNAJC2-like and *MYB4-like*
434 and bHLH members *bHLH62*, *bHLH74* and *bHLH162* showed a remarkably negative
435 correlation with anthocyanin content at veraison S2 ($P<0.05$).
436
437 *Genes involved in terpenoid biosynthesis and plant hormone signal transduction*
438 KEGG analysis showed that terpenoid backbone biosynthesis was over-presented in
439 cluster 1 and that sesquiterpenoid and triterpenoid biosynthesis branching from
440 terpenoid backbone biosynthesis was over-presented in clusters 4 and 5 (see Figure 3),
441 indicating that this pathway may play important roles in anthocyanin accumulation in
442 apple. Moreover, the plant hormone signal transduction pathway, which is modulated
443 by terpenoid biosynthesis, was enriched in cluster 2. A total of 18 DEGs in the

444 terpenoid biosynthesis pathway (see Dataset S7 and Table 4) and 12 DEGs in the
445 plant hormone signal transduction pathway (see Dataset S8 and Table 5) were also
446 investigated by preparing heat maps (see Figure 5). The expression of DEGs involved
447 in terpenoid backbone biosynthesis, including one acetyl-CoA acetyltransferase,
448 cytosolic 1 (*AACT1*), two hydroxymethylglutaryl-CoA synthase-like (*HMGS*), two
449 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1-like (*HMGR*), one mevalonate
450 kinase-like (*MVK*), two diphosphomevalonate decarboxylase MVD2-like (*MVD2*),
451 one isopentenyl-diphosphate Delta-isomerase I (*IDI1*), three farnesyl pyrophosphate
452 synthase 2-like (*FPPS2*) and six squalene monooxygenase-like (*SQMO*), were
453 gradually down-regulated from G0 to G4. Nevertheless, two auxin transporter-like 1
454 (*AUX1*), four auxin-responsive protein *AUX/IAA* and three auxin-responsive protein
455 *SAUR* involved in tryptophan metabolism of auxin biosynthesis; one F-box protein
456 GID2-like (*GID2*) involved in diterpenoid biosynthesis of GA biosynthesis; and one
457 protein phosphatase 2C (*PP2C*) and one sucrose non-fermenting-1-related protein
458 kinase 2 (*SnRK2*) involved in carotenoid biosynthesis of ABA biosynthesis were
459 up-regulated from G0 to G4.

460 To further evaluate the validity of our results, 16 representative DEGs used
461 previously as shown in Figures 4 and 5 were selected for expression level
462 examination by qRT-PCR (see Table S3). The overall trend of relative expression
463 levels at three stages was consistent with that of deep sequencing at S2 (see Figure 6),
464 suggesting that the candidate genes involved in pathways such as
465 phenylpropanoid/flavonoid biosynthesis, terpenoid biosynthesis and plant hormone
466 signal transduction, appended with MYB and bHLH transcriptional factors, were
467 directly correlated with anthocyanin accumulation.

468

469 **DISCUSSION**

470 **The activation of early phenylpropanoid biosynthesis genes was more responsible**
471 **for anthocyanin accumulation in apple pericarp of bud sport mutants**

472 In plants, phenylpropanoid biosynthesis gives rise to a large number of secondary
473 metabolites, including hydroxycinnamic acids, monolignols/lignin, coumarins,

474 benzoic acids, stilbenes, anthocyanins and flavonoids, serving different functions in
475 plant development, reproduction, defence, and protection against biotic/abiotic
476 stresses (Voge 2010; Zhang *et al*. 2013). Differences in the expression pattern of
477 genes involved in phenylpropanoid/flavonoid biosynthesis result in diverse
478 anthocyanin profiles (El-Sharkawy *et al*. 2015; Massonnet *et al*. 2017). Our survey
479 provided a comprehensive profile of the phenylpropanoid/flavonoid biosynthesis
480 pathway in 'Red Delicious' and its four continuous generation mutants. The results
481 showed that all of the early phenylpropanoid biosynthesis pathway genes, including
482 *ASP3*, *PAL*, *4CL*, *BGLU* and *PER*, were aggregated in cluster 2 (see Figure 4A and
483 Table 3), which matched the anthocyanin content (see Figure 1A, B). Other genes in
484 cluster 2 containing *CHS*, *CYP98A* and *F3'H* are involved in the middle steps of the
485 phenylpropanoid pathway, that is, the early steps of the flavonoid biosynthesis
486 pathway. Nevertheless, genes encoding *HCT*, *CCoAOMT*, *CHS*, *CYP98A*, *CHI*, *F3H*,
487 *DFR*, *FLS* and *ANS* were involved in the middle and late steps of the phenylpropanoid
488 biosynthesis pathway and gathered in clusters 1 and 3. Thus, the activation of early
489 phenylpropanoid biosynthesis pathway genes was demonstrated to be most
490 responsible for pigment accumulation in the apple pericarp of bud sport mutants. In
491 addition, *ASP3*, *BGLU* and *PER* were confirmed to be involved in the synthesis of
492 phenylpyruvate, coumarine and lignin, respectively (see Figure 4A). Interestingly, 44
493 stilbene synthase (*STS*) genes involved in stilbene biosynthesis were characterized to
494 influence anthocyanin accumulation during grapevine (*Vitis vinifera*) maturation by
495 Massonnet *et al*. (2017). Nevertheless, these genes do not exist among our 3,466
496 DEGs, possibly because of their variety-specific nature (Zenoni *et al*. 2016).
497
498 **MYB and bHLH modulated anthocyanin accumulation in apple pericarp by**
499 **regulating the transcription of genes involved in the phenylpropanoid/flavonoid**
500 **pathway**
501 MYB and bHLH autonomously mediated the transcription of genes involved in the
502 middle steps of the phenylpropanoid pathway, that is, the early steps of the flavonoid
503 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). The heterologous expression of OjMYB1 in Arabidopsis could enhance anthocyanin content and up-regulate the expression levels of structural gene-related anthocyanin biosynthesis (Feng *et al*. 2017). The red radish (*Raphanus sativus* L.) bHLH transcription factor RsTT8 acts as a positive regulator of anthocyanin biosynthesis (Lim *et al*. 2017). Nevertheless, *AtMYB4*, *AmMYB308*, *FaMYB1*, *ZmMYB31*, *ZmMYB42*, *PhMYB4, VvMYBC2-L1* and *PtrMYB57* have been demonstrated to repress phenylpropanoid synthesis, likely via repression of synthesis genes (Tamagnone *et al*. 1998; Aharoni *et al*. 2001; Colquhoun *et al*. 2011; Huang *et al*. 2014; Wan *et al*. 2017). We corroborated that MYB members, including *LUX*, *MYB113*, *PCL-like*, *MYB1R1-like*, *MYB6-like*, *MYB308-like* and *MYB5-like*, and bHLH members, including *bHLH51* and *bHLH155*, showed a notable positive correlation with anthocyanin content (see Dataset S5, 6 and Figure 4B, C) and were considered to promote anthocyanin synthesis by mediating the transcription of structural genes, *CHS* and *F3'H*, which are involved in the flavonoid pathway. Other MYB members, including *LHY-like*, *RVE6*, *GT-3b*, *MYB21*, *DNAJC2-like* and *MYB4-like*, and bHLH members, including *bHLH62*, *bHLH74* and *bHLH162* showed a remarkably negative correlation with anthocyanin content and were demonstrated to repress anthocyanin synthesis. In addition, HD-Zip I transcription factor MdHB1 was involved in the regulation of anthocyanin accumulation (Lü *et al*. 2014). When MdHB1 is silenced, *MdMYB10*, *MdbHLH3*, and *MdTTG1* are released to activate the expression of *MdDFR* and *MdUFGT* and anthocyanin biosynthesis, resulting in red flesh in apple cv. 'Granny Smith' (Jiang *et al*. 2017). The expression of *F3'5'H*, *DFR* and *ANS* is strongly inhibited by the increase in the expression of *MYBL1*, which is a novel R3 MYB transcription factor classified as an MYB transcriptional repressor (Gates *et al*. 2018). However, a full understanding of the mechanism by which structural genes involved in anthocyanin synthesis are specifically mediated by MYB and bHLH remains elusive clear and requires further investigation.

Terpenoid biosynthesis modulated anthocyanin accumulation by positively regulating plant hormone signal transduction in apple pericarp of bud sport mutants

Hormones are important factors inducing anthocyanin accumulation (Jiang and Fu 2007; Shan 2009; Loreti *et al*. 2010). Carvalho *et al*. (2010) provided evidence 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 an auxin concentration that was too high 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 sesquiterpenoid and triterpenoid biosynthesis along with plant hormone transduction, including tryptophan metabolism for IAA, diterpenoid biosynthesis for GA and carotenoid biosynthesis for ABA, branches from the general terpenoid backbone synthesis pathway and shares the same precursors as glycolysis (see Figure 5). The sesquiterpenoid and triterpenoid biosynthesis pathways comprised six *SQMO* genes in clusters 4 and 5, which were negatively correlated with anthocyanin content. The genes *AACT1*, *HMGS*, *HMGR*, *MVK*, *MVD2*, *IDI1* and *FPPS2*, involved in the early steps of terpenoid backbone synthesis, were in cluster 1, which also reflected the differential accumulation of anthocyanin to some extent. Moreover, *AUX1*, *AUX/IAA* and *SAUR*, associated with tryptophan metabolism for IAA; *GID2*, associated with diterpenoid biosynthesis for GA; and *PP2C* and *SnRK2*, associated with carotenoid biosynthesis for ABA, were in cluster 2, which positively reflected the anthocyanin content. Likewise, the contents of IAA and ABA increased, while GA decreased with maturation and pigment accumulation from G0 to the fourth-generation mutant G4 (see Figure 1C, E). Therefore, the down-regulation of genes involved in terpenoid biosynthesis positively induced the expression of *AUX1*, *AUX/IAA*, *SAUR*, *GID2*, *PP2C* and *SnRK2*, resulting in increased synthesis of IAA and ABA and decreased synthesis of GA, thus modulating anthocyanin accumulation. Loreti *et al*. (2010) suggested the existence of crosstalk between the sucrose and hormone signalling pathways in the regulation of

564 the anthocyanin biosynthetic pathway. Similarly, exogenous ABA promoted fruit
565 ripening by increasing anthocyanin content in sweet cherry (*Prunus avium* L.) cv.
566 Sato Nishiki, and the expression of *PaPP2C3*, *PaPP2C5* and *PaPP2C6* was
567 significantly induced by exogenous ABA (Wang *et al*. 2015). In general, there may be
568 some form of crosstalk between the activation of phenylpropanoid biosynthesis and
569 plant hormone signal transduction in pigment accumulation of apple bud sport
570 mutants.
571

572 **CONCLUSIONS**

573 We investigated the pericarp transcriptome of 'Red Delicious' and its four continuous
574 generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee
575 spur') and identified specific processes that lead to the accumulation of anthocyanin.
576 The results indicate that apple pericarp pigmentation and anthocyanin content were
577 increased in the mutants due to bud sport. Terpenoid biosynthesis influences
578 anthocyanin accumulation by positively regulating the synthesis of IAA and ABA and
579 negatively regulating the synthesis of GA. MYB and bHLH modulate anthocyanin
580 accumulation in apple pericarp by regulating the transcription of genes *CHS* and *F3'H*.
581 *ASP3*, *CYP98A* and *CCoAOMT* are novel anthocyanin-associated genes of apple first
582 reported the present study. This novel set of genes provides not only new insights into
583 anthocyanin biosynthesis but also important clues for more dedicated studies to
584 broaden our knowledge of the anthocyanin pathway in apple.
585

586 **ACKNOWLEDGMENTS**

587 This research was financially supported by the Key Scientific Technology Research
588 Projects of Gansu Province (GPCK2013-2), the Fostering Foundation for the
589 Excellent Ph.D. Dissertation of Gansu Agricultural University (2017002), and the
590 Science and Technology Major Project of Gansu Province (18ZD2NA006).
591 Author contributions: Bai-Hong Chen designed the research. Wen-Fang Li carried out
592 the experiments with the help of Juan Mao, Shi-Jin Yang, Zhi-Gang Guo, Zong-Huan
593 Ma, Cun-Wu Zuo and Ming-Yu Chu. Wen-Fang Li collected the experimental data

and drafted the manuscript. Mohammed Mujitaba Dawuda reviewed the manuscript and part of the data analysis. All authors read and approved the final manuscript.

LITERATURECITED

Aharoni, A., C. H. R. D. Vos, M. Wein, Z. Sun, R. Greco *et al.*, 2001 The strawberry *FaMYB1* transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. Plant J. 28: 319–332.

An, J. P., X. Liu, H. H. Li, C. X. You, and X. F. Wang, 2017 Apple RING E3 ligase MdMIEL1 inhibits anthocyanin accumulation by ubiquitinating and degrading MdMYB1 protein. Plant Cell Physiol. 58(11): 1953–1962.

An, X. H., Y. Tian, K. Q. Chen, X. F. Wang, and Y. J. Hao, 2012 The apple WD40 protein MdTTG1 interacts with bHLH but not MYB proteins to regulate anthocyanin accumulation. J. Plant Physiol. 169: 710–717.

Arrizabalaga, M., F. Morales, M. Oyarzun, S. Delrot, E. Gomès *et al.*, 2018 Tempranillo clones differ in the response of berry sugar and anthocyanin accumulation to elevated temperature. Plant Sci. 267: 74.

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

Benjamini, Y. and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B. 57: 289–300.

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

Colquhoun, T. A., J. Y. Kim, A. E. Wedde, L. A. Levin, K. C. Schmitt *et al.*, 2011 *PhMYB4* fine-tunes the floral volatile signature of *Petunia×hybrida* through *PhC4H*. J. Exp. Bot. 62: 1133–1143.

Daccord, N., J. M. Celton, G. Linsmith, C. Becker, N. Choisne *et al.*, 2017

High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. 49(7): 1099.

Deluc, L., J. Bogs, A. R. Walker, T. Ferrier, A. Decendit *et al*., 2008 The transcription factor VvMYB5b contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. Plant Physiol. 147: 2041–2053.

El-Kereamy, A., C. Chervin, J. P. Roustan, V. Cheyhnier, J. M. Souquet *et al*., 2003 Exogenous ethylene stimulates the long term expression of genes related to anthocyanin biosynthesis in grape berries. Physiol. Plantarum 119: 175–282.

El-sharkawy, I., L. Dong, and K. Xu, 2015 Transcriptome analysis of an apple *(Malus* × *domestica*) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. J. Exp. Bot. 66(22): 7359–7376.

Gomi, K., A. Sasaki, H. Itoh, M. Ueguchi-Tanaka, M. Ashikari *et al*., 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 J. 37(4): 626–634.

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

Honda, C. and S. Moriya, 2018 Anthocyanin biosynthesis in apple fruit. Horticult. J.

Hu, D. G., C. H. Sun, Q.J. Ma, C. X. You, L. Cheng *et al*., 2016 MdMYB1 regulates anthocyanin and malate accumulation by directly facilitating their transport into vacuoles in apples. Plant Physiol. 170(3): 1315–1330.

Huang, Y. F., S. Vialet, J. L. Guiraud, L. Torregrosa, Y. Bertrand *et al*., 2014 A negative MYB regulator of proanthocyanidin accumulation, identified through expression quantitative locus mapping in the grape berry. New Phytol. 201(3): 795–809.

Ito, S., D. Yamagami, M. Umehara, A. Hanada, S. Yoshida *et al*., 2017 Regulation of strigolactone biosynthesis by gibberellin signaling. Plant Physiol. 174(2):

1250–1259.

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

Ji, X. H., Y. T. Wang, R. Zhang, S. J. Wu, M. M. An *et al*., 2015 Effect of auxin, cytokinin and nitrogen on anthocyanin biosynthesis in callus cultures of red-fleshed apple (*Malus sieversii*, f. *niedzwetzkyana*). Plant Cell Tiss. Org. 120(1): 325–337.

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

Jiang, Y., C. Liu, D. Yan, X. Wen, Y. Liu *et al*., 2017 *MdHB1* down-regulation activates anthocyanin biosynthesis in the white-fleshed apple cultivar 'Granny Smith'. J. Exp. Bot. 68(5): 1055.

Lim, S. H., D. H. Kim, J. K. Kim, J.Y. Lee, and S.H. Ha, 2017 A radish basic helix-loop-helix transcription factor, RsTT8 acts a positive regulator for anthocyanin biosynthesis. Front. Plant Sci. 8: 1917.

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

Livak, K. J. and T.D. Schmittgen, 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., G. Povero, G. Novi, C. Solfanelli, A. Alpi *et al*., 2010 Gibberellins, jasmonate and abscisic acid modulate the sucrose-induced expression of anthocyanin biosynthetic genes in Arabidopsis. New Phytol. 179(4): 1004–1016.

Lü, P., C. Zhang, J. Liu, X. Liu, G. Jiang *et al*., 2014 RhHB1 mediates the antagonism of gibberellins to ABA and ethylene during rose (*Rosa hybrida*) petal senescence. Plant J. 78(4): 578–590.

Massari, M. E. and C. Murre, 2000 Helix-loop-helix proteins:regulators of transcription in eucaryotic organisms. J. Mol. Cell Biol. 20: 429–440.

684 Massonnet, M., M. Fasoli, G. B. Tornielli, M. Altieri, M. Sandri *et al*., 2017
685 Ripening transcriptomic program in red and white grapevine varieties correlates
686 with berry skin anthocyanin accumulation. Plant Physiol. 174(4): 2376–2396.
687 Gates, D. J., B. Olson, T. E. Clemente, and S.D. Smith, 2018 A novel R3 MYB
688 transcriptional repressor associated with the loss of floral pigmentation in *Iochroma*.
689 New Phytol. 217(3): 1–11.
690 Meng, R., J. Zhang, L. An, B. Zhang, X. Jiang *et al*., 2016 Expression profiling of
691 several gene families involved in anthocyanin biosynthesis in apple (*Malus*
692 *domestica*, Borkh.) skin during fruit development. J. Plant Growth Regul. 35(2):
693 449–464.
694 Mortazavi, A., B. A. Williams, K. Mccue, L. Schaeffer, and B. Wold, 2008
695 Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:
696 621–628.
697 Petit, R. J. and A. Hampe, 2006 Some evolutionary consequences of being a tree.
698 Annu. Rev. Ecol. Evol. S. 37: 187–214.
699 Petroni, K. and C. Tonelli, 2011 Recent advances on the regulation of anthocyanin
700 synthesis in reproductive organs. Plant Sci. 181: 219–229.
701 Ramsay, N. A. and B. J. Glover, 2005 MYB-bHLH-WD40 protein complex and the
702 evolution of cellular diversity. Trends Plant Sci. 10: 63–70.
703 Shan, X., 2009 Molecular mechanism for jasmonate-induction of anthocyanin
704 accumulation in Arabidopsis. J. Exp. Bot. 60(13): 3849–3860.
705 Stracke, R., H. Ishihara, G. Huep, A. Barsch, F. Mehrtens *et al*., 2007 Differential
706 regulation of closely related R2R3-MYB transcription factors controls flavonol
707 accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J. 50:
708 660–677.
709 Tamagnone, L., A. Merida, A. Parr, S. Mackay, F. A. Culianez-Macia *et al*., 1998
710 The AmMYB308 and AmMYB330 transcription factors from Antirrhinum regulate
711 phenylpropanoid and lignin biosynthesis in transgenic tobacco. Plant Cell 10:
712 135–154.
713 Telias, A., K. Linwang, D. E. Stevenson, J. M. Cooney, R. P. Hellens *et al*., 2011

714 Apple skin patterning is associated with differential expression of *MYB10*. BMC
715 Plant Biol. 11(1): 93.
716 Vasudevan, H. N., P. Mazot, F. He and P. Soriano, 2015 Receptor tyrosine kinases
717 modulate distinct transcriptional programs by differential usage of intracellular
718 pathways. eLife 4: e07186.
719 Vogt, T., 2010 Phenylpropanoid biosynthesis. Mol. Plant 3: 2–20.
720 Wan, S., C. Li, X. Ma, and K. Luo, 2017 PtrMYB57 contributes to the negative
721 regulation of anthocyanin and proanthocyanidin biosynthesis in poplar. Plant Cell
722 Rep. 36(8): 1–14.
723 Wang, Y., P. Chen, L. Sun, Q. Li, S. Dai *et al*., 2015 Transcriptional regulation of
724 *PaPYLs*, *PaPP2Cs* and *PaSnRK2s* during sweet cherry fruit development and in
725 response to abscisic acid and auxin at onset of fruit ripening. Plant Growth Regul.
726 75(2): 455–464.
727 Williams, R. S. and N. Benkeblia, 2018 Biochemical and physiological changes of
728 star apple fruit (*Chrysophyllum cainito*), during different "on plant" maturation and
729 ripening stages. Sci. Hortic. 236: 36–42.
730 Xu, Y., S. Feng, Q. Jiao, C. Liu, W. Zhang *et al*., 2012 Comparison of *MdMYB1*
731 sequences and expression of anthocyanin biosynthetic and regulatory genes
732 between *Malus domestica* Borkh. cultivar 'Ralls' and its blushed sport. Euphytica
733 185: 157–170.
734 Zenoni, S., M. Fasoli, F. Guzzo, S. Dal Santo, A. Amato *et al*., 2016 Disclosing the
735 molecular basis of the postharvest life of berry in different grapevine genotypes.
736 Plant Physiol. 172: 1821–1843.
737 Zhang, Y., E. Butelli, R. De Stefano, H. Schoonbeek, A. Magusin *et al*., 2013
738 Anthocyanins double the shelf life of tomatoes by delaying overripening and
739 reducing susceptibility to gray mold. Curr. Biol. 23: 1094–1100.
740

741 **Figure legends**

742 **Figure 1** A, Close-up views of 'Red Delicious' and its four generation mutants
743 ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur'), named G0 to

744 G4, at three developmental stages (S1–S3) used for anthocyanin quantification,
745 transcriptome profiling and qRT-PCR. B, The changes in total anthocyanin
746 concentrations in the pericarp of the five strains at S1 to S3. Changes in endogenous
747 hormone levels, including IAA (C), GA (D) and ABA (E), in the pericarp of the five
748 strains at S1 to S3. Values are means ± SE. Different lower case letters indicate
749 significant differences among the five strains (P=0.05).
750
751 **Figure 2** Summary of the number of differentially expressed genes (DEGs) identified
752 by RNA-seq analysis in the pericarp tissues of 'Red Delicious' and its four generation
753 mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') at S2,
754 named G0 to G4. The number of total DEGs (A), up-regulated DEGs (B), and
755 down-regulated DEGs (D) are presented by Venn diagrams (FDR < 0.01 and fold
756 change ≥ 2). D, Number of total up-regulated and down-regulated DEGs. The
757 histogram represents the number of commonly down-regulated (blue) and
758 up-regulated (yellow) DEGs.
759
760 **Figure 3** Gene expression profiles and KEGG category distribution of the DEGs in
761 the six common expression clusters composing the pericarp transcriptome at S2.
762 Clusters were derived by coupled clustering analysis of the 3,466 commonly
763 modulated genes. A, Heat map of the overall common expression pattern. B, Each line
764 represents the $\log_2$-transformed average of the mean FPKM values for an individual
765 transcript. Significantly overrepresented KEGG categories are represented by red dots.
766 KEGG category enrichment was computed using the R language, Cluster package,
767 Biobase package, and Q-value package ($P \leq 0.01$).
768
769 **Figure 4** A, Differential expression of genes involved in phenylpropanoid/flavonoid
770 biosynthesis pathway hormone signal transduction pathway in 'Red Delicious' and its
771 four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and
772 'Vallee spur'), named G0 to G4. Differential expression of transcription factors that
773 encode myb-like (B) and helix-loop-helix (C) DNA-binding domains. Heat maps

774 depict the normalized gene expression values, which represent the means ± SD of
775 three biological replicates. Expression values of five libraries are presented as FPKM
776 normalized $\log_{10}$-transformed counts.
777
778 **Figure 5** Differential expression of genes involved in sesquiterpenoid and
779 triterpenoid/terpenoid biosynthesis coupled with hormone signal transduction
780 pathway in 'Red Delicious' and its four generation mutants ('Starking Red',
781 'Starkrimson', 'Campbell Redchief' and 'Vallee spur'), named G0 to G4. Heat maps
782 depict the normalized gene expression values, which represent the means ± SD of
783 three biological replicates. Expression values of five libraries are presented as FPKM
784 normalized $\log_{10}$-transformed counts.
785
786 **Figure 6** Relationships between total anthocyanin contents and transcript levels of the
787 sixteen representative genes from Figures 4 and 5. For each accession, the expression
788 was determined in three developmental stages (S1 to S3) of pericarp tissues. For the
789 qRT-PCR assay, the mean was calculated from three biological replicates, each with
790 three technical replicates (n=9). These replicates were then normalized relative to the
791 expression of *MdGADPH*. The x-axis in each chart is the same and represents
792 different Malus accessions, as indicated by names in the bottom panel, which are
793 arranged in different clusters. The left y-axis represents relative expression levels
794 determined by qRT-PCR.
795
796 **Supporting information**
797 Additional Supporting Information may be found online in the supporting information
798 tab for this article:
799 **Figure S1** Summary of the number of differentially expressed genes (DEGs)
800 identified by RNA-seq analysis in the pericarp tissues of 'Red Delicious' and its four
801 generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee
802 spur') at S2, named G0 to G4. The number of total DEGs (A), up-regulated DEGs (B),
803 and down-regulated DEGs (D) are presented by Venn diagrams (FDR < 0.01 and fold

804 change ≥ 2). D, Number of total up-regulated and down-regulated DEGs. The
805 histogram represents the number of commonly down-regulated (blue) and
806 up-regulated (yellow) DEGs.

807 **Table S1** Summary of RNA-Seq data and mapping metrics

808 **Table S2** Gene ontology (GO) enrichment analyses for DEGs in 'Red Delicious' and
809 its four generation mutants.

810 **Table S3** Sequence of primers used for qRT-PCR analysis.

811

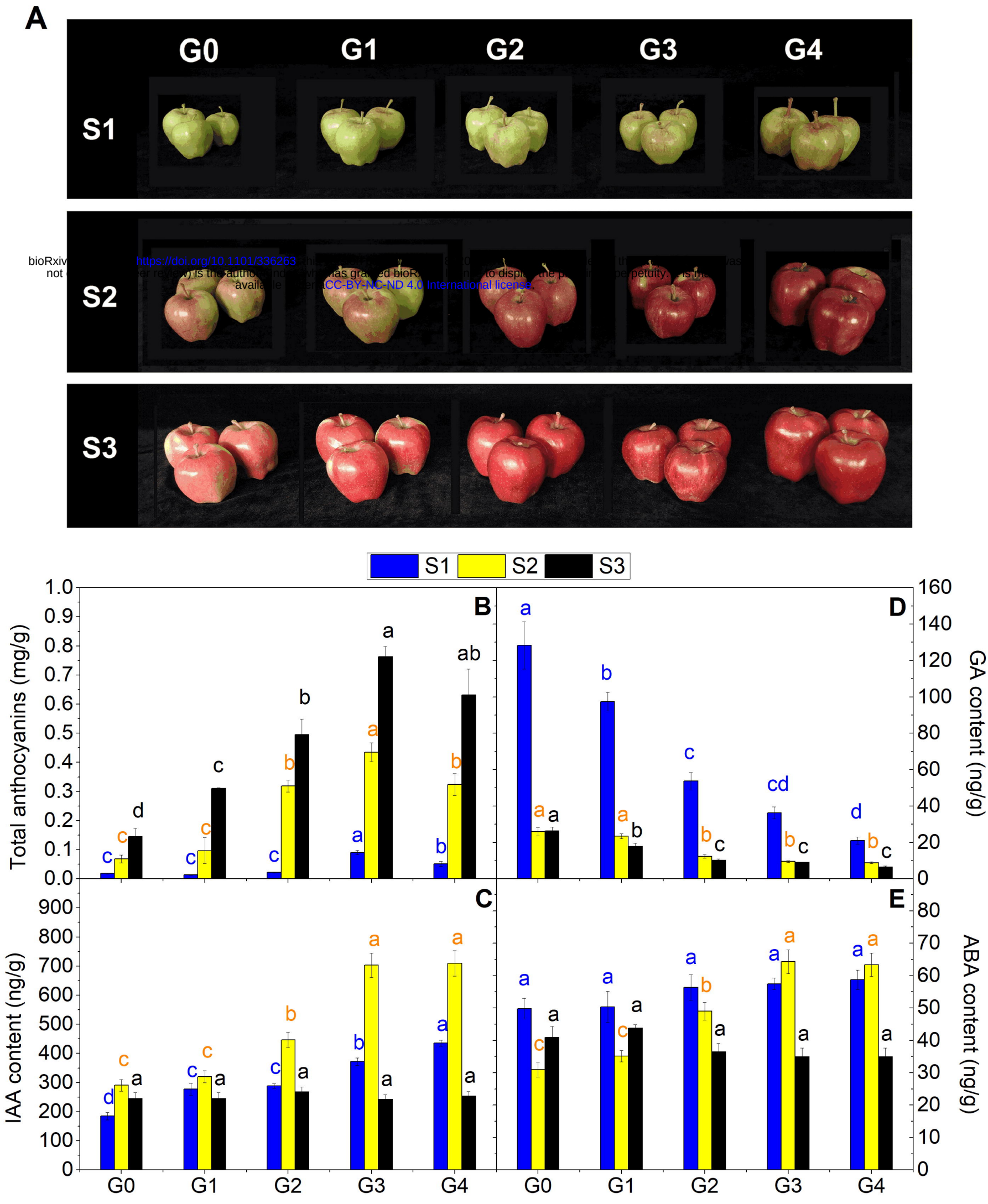

Figure 1. A, Close-up views of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') fruit named as G0 to G4 at three developmental stages (S1-S3) used for anthocyanin quantification, transcriptome profiling and qRT-PCR. B, The changes 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 ($P = 0.05$).

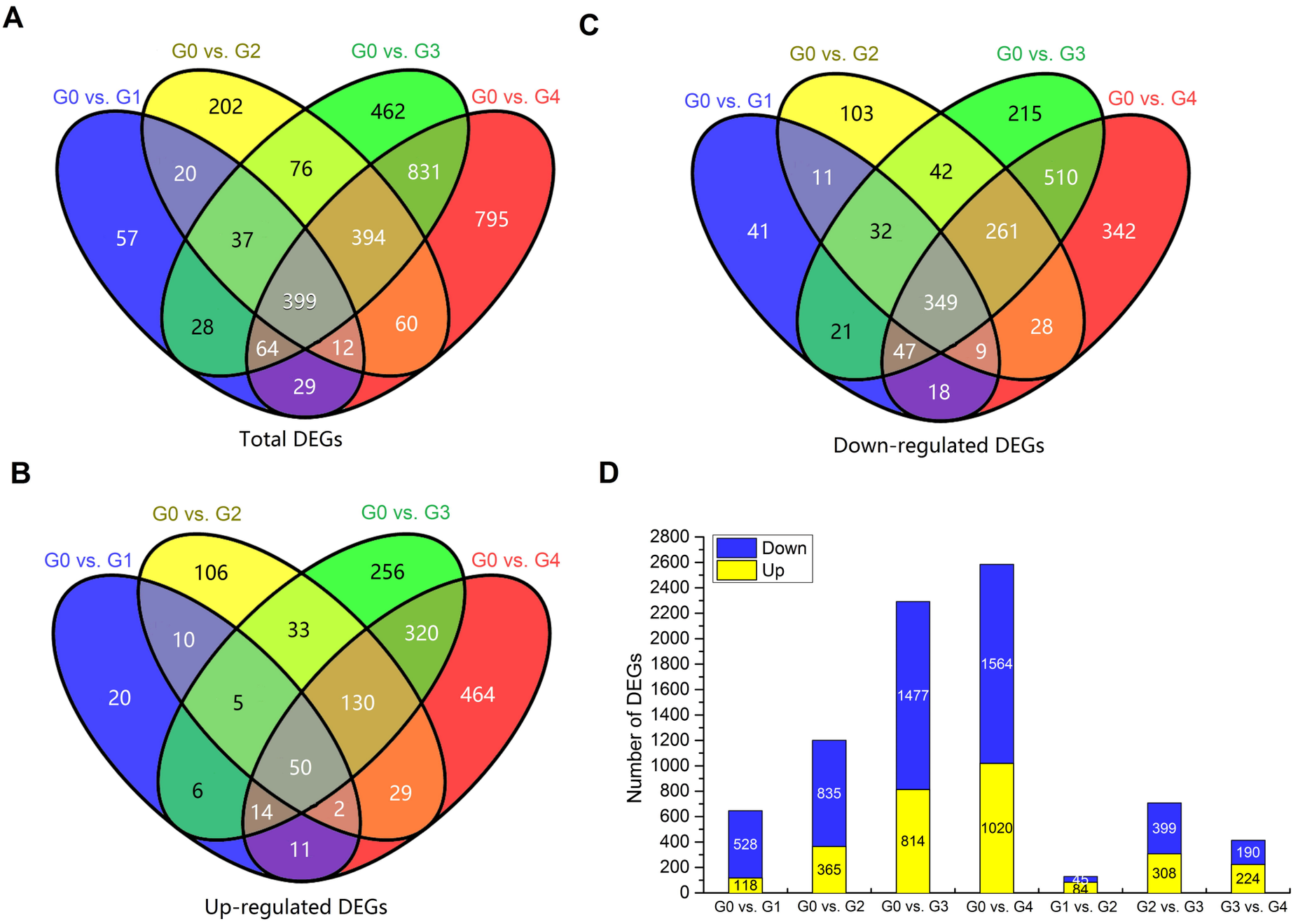

Figure 2. Summary of the number of differentially expressed genes (DEGs) identified by RNA-seq analysis in the pericarp tissues of 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') at S2, named as G0 to G4. Number of total DEGs (A), 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.

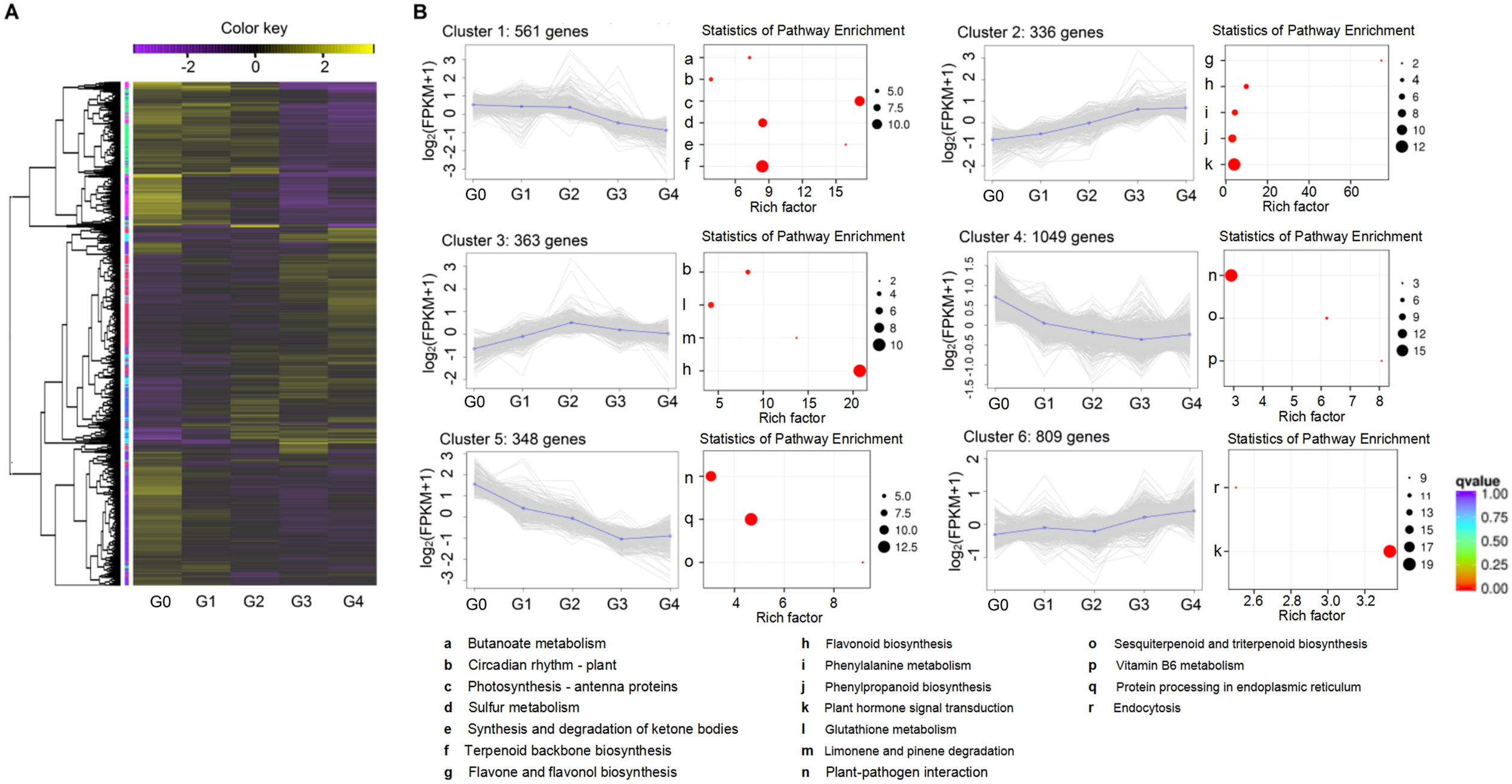

Figure 3. Gene expression profiles and KEGG category distribution of the DEGs in the six common expression clusters composing the pericarp transcriptome at S2. Clusters were derived by coupled clustering analysis of the 3,466 commonly modulated genes. A, Heat map of the overall common expression pattern. 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 ($P \leq 0.01$).

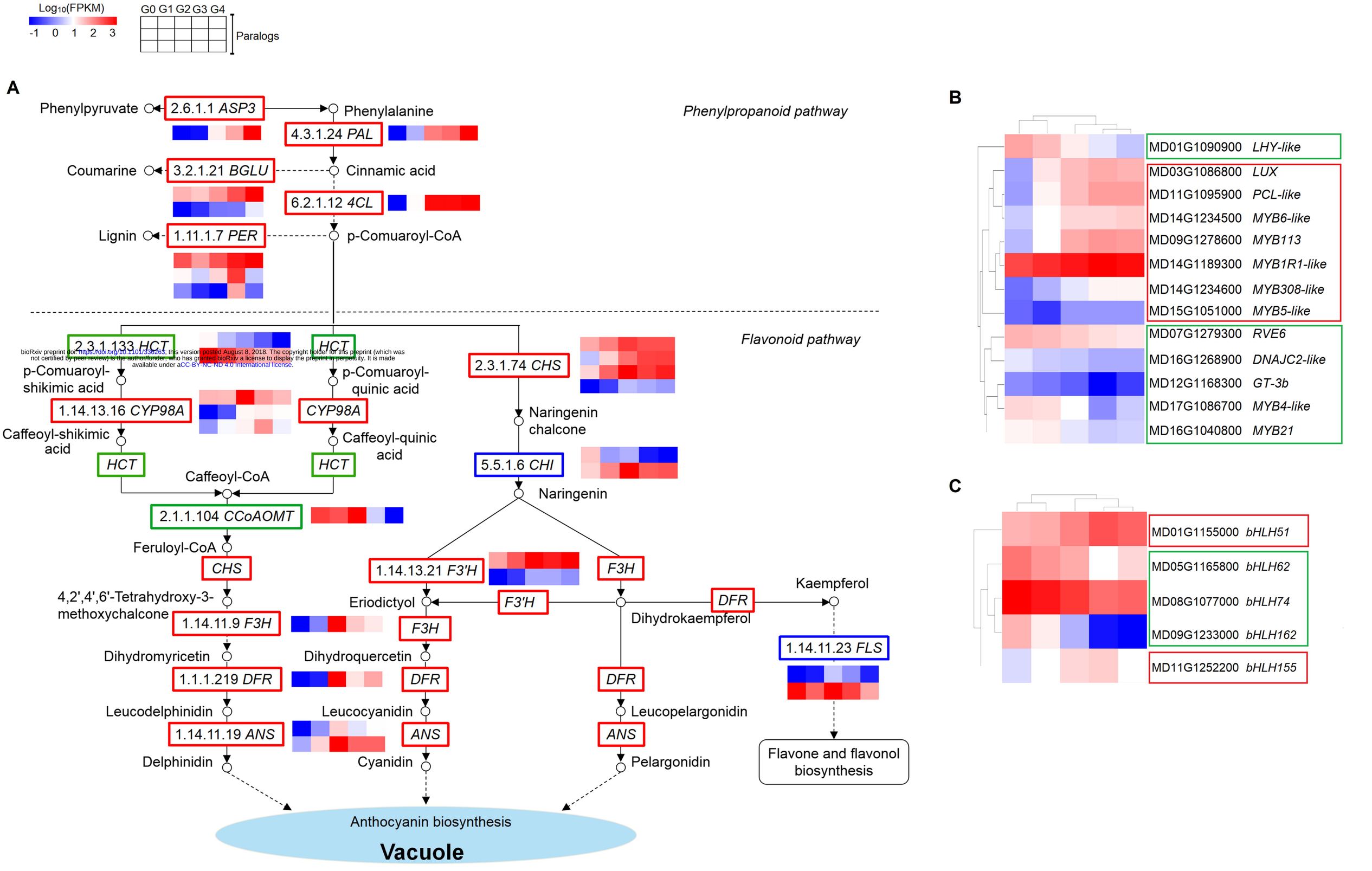

Figure 4. A, Differential expression of genes involved in phenylpropanoid/flavonoid biosynthesis pathway hormone signal transduction pathway in 'Red Delicious' and its four generation mutants ('Starking Red', 'Starkrimson', 'Campbell Redchief' and 'Vallee spur') named as G0 to G4. Differential expression 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.

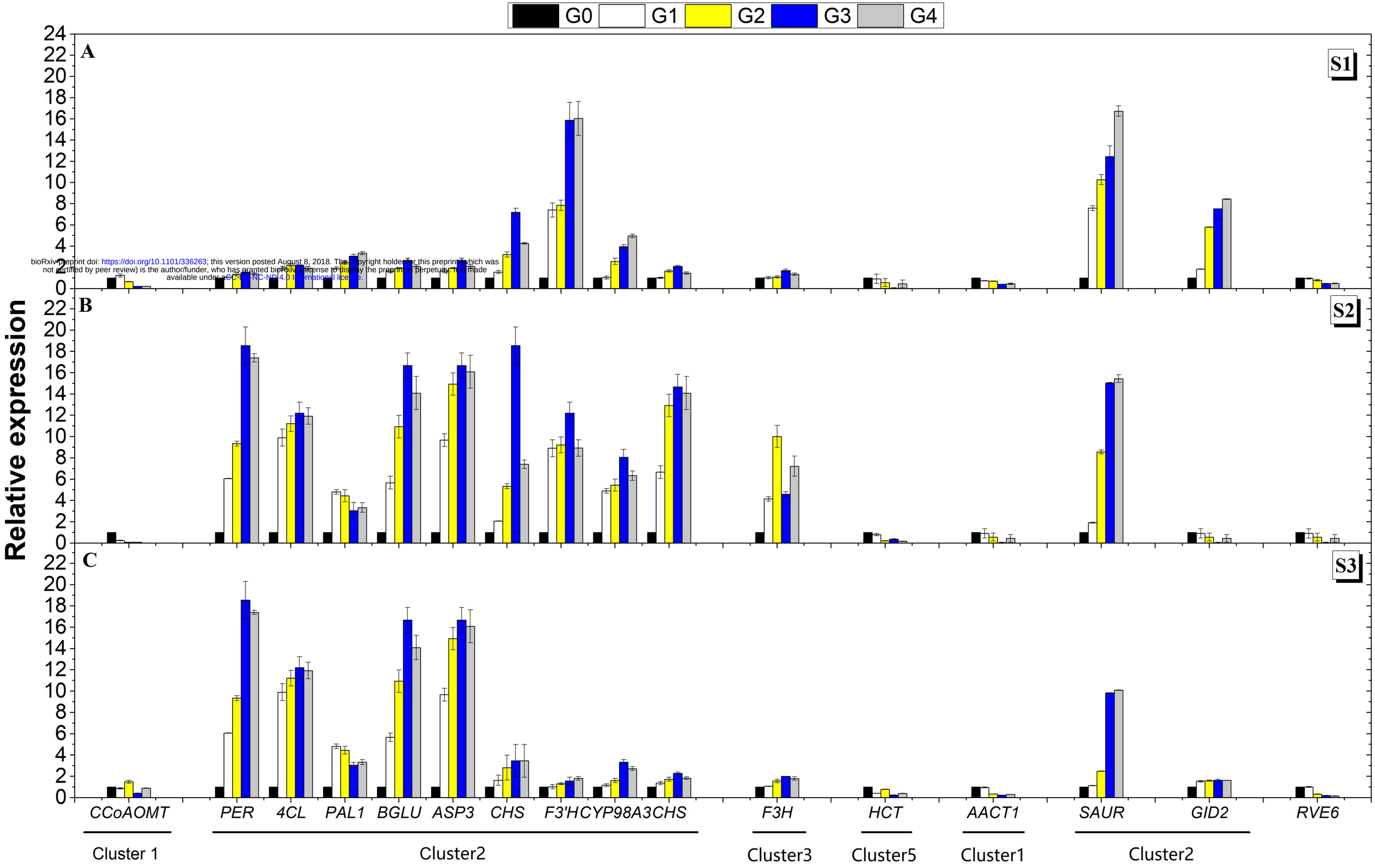

Figure 6. Relationships between total anthocyanin contents and transcript levels of the sixteen representative genes from Fig. 4 and 5. For each accession, the expression was determined in three developmental stages (S1 to S3) of pericarp tissues. For qRT-PCR assay, the mean was calculated from three 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.