

# Candidate genes associated with red colour formation revealed by comparative genomic variant analysis of red- and green-skinned fruits of Japanese apricot (*Prunus mume*)

Xiaopeng Ni<sup>1,2</sup>, Song Xue<sup>1</sup>, Shahid Iqbal<sup>1</sup>, Wanxu Wang<sup>1</sup>, Zhaojun Ni<sup>1</sup>, Muhammad Khalil-ur-Rehman<sup>1</sup> and Zhihong Gao<sup>1</sup>

<sup>1</sup>Laboratory of Fruit Tree Biotechnology, College of Horticulture, Nanjing Agricultural University, Nanjing, China

<sup>2</sup>Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Nanjing, China

## ABSTRACT

The red-skinned fruit of Japanese apricot (*Prunus mume* Sieb. et Zucc) appeals to customers due to its eye-catching pigmentation, while the mechanism related to its colour formation is still unclear. In this study, genome re-sequencing of six Japanese apricot cultivars was carried out with approximately 92.2 Gb of clean bases using next-generation sequencing. A total of 32,004 unigenes were assembled with an average of 83.1% coverage rate relative to reference genome. A wide range of genetic variation was detected, including 7,387,057 single nucleotide polymorphisms, 456,222 insertions or deletions and 129,061 structural variations in all genomes. Comparative sequencing data revealed that 13 candidate genes were involved in biosynthesis of anthocyanin. Significantly higher expression patterns were observed in genes encoding three anthocyanin synthesis structural genes (*4CL*, *F3H* and *UFGT*), five transcription factors (MYB–bHLH–WD40 complexes and NAC) and five anthocyanin accumulation related genes (*GST1*, *RT1*, *UGT85A2*, *ABC* and *MATE* transporters) in red-skinned than in green-skinned Japanese apricots using reverse transcription-quantitative polymerase chain reaction. Eight main kinds of anthocyanin s were detected by UPLC/MS, and cyanidin 3-glucoside was identified as the major anthocyanin (124.2 mg/kg) in red-skinned cultivars. The activity of UDP-glucose flavonoid-3-O-glycosyltransferase enzyme determined by UPLC was significantly higher in all red-skinned cultivars, suggesting that it is the potential vital regulatory gene for biosynthesis of anthocyanin in Japanese apricot.

Submitted 26 October 2017

Accepted 24 March 2018

Published 4 May 2018

Corresponding author

Zhihong Gao,  
gaozhihong@njau.edu.cn

Academic editor

Scott Edwards

Additional Information and  
Declarations can be found on  
page 24

DOI 10.7717/peerj.4625

© Copyright  
2018 Ni et al.

Distributed under  
Creative Commons CC-BY 4.0

**OPEN ACCESS**

**Subjects** Genomics, Molecular Biology, Plant Science

**Keywords** *Prunus mume*, Sequencing, Anthocyanin, UFGT, Gene expression

## INTRODUCTION

Japanese apricot (*Prunus mume* Sieb. et Zucc), an attractive fruit tree, originated from Southwest China and has been extensively cultivated in all of East Asia and Japan. China, being the origin of Japanese apricot, is rich in good quality germplasm with an approximately 190 fruiting cultivars (Chu, 1999). The fruit is usually processed into many

value-added products, containing salted-fruit, drinks and juice, is well advised to high nutritional and medicinal value ([Chen, 1989](#)). Fruit colour is an significant factor for determining fruit quality, which increases its value and attracts consumer's attention ([Gao, Han & Zhang, 2003](#)). The whole genome of Japanese apricot was sequenced in 2012 ([Zhang et al., 2012](#)), laying the foundation for detailed study of important traits related to pigmentation, flowering time, dormancy and other commercially relevant aspects. Moreover, genome-based tools can also be advanced to develop breeding efficiency and carry out other functional genomic research.

The emergence of next-generation sequencing (NGS) technologies had a tremendous impact on genomic research over the past decade. NGS technologies have become a major tool for acquiring data related to genetic variation inside a variant and recognizing exclusive genotypic molecular markers. Sequencing technologies have been applied to discover and quantify genomic variants in many species of plants and enabled researchers to illustrate genomic alterations among specific genotypes ([Zhang et al., 2014](#)). Sequencing-based bulked segregation analysis (seq-BSA) was used to map resistance genes by re-sequencing in pigeon pea and has been proven an effective way to carry out genomics-assisted breeding. Whole genome re-sequencing has also been helpful in identifying the particular location of equally crossover and non-crossover recombination measures in an Avian pedigree ([Smeds et al., 2016](#)). Whole genome re-sequencing has made genomic research easier and more accurate.

Anthocyanins are the main water-soluble pigments localized in vacuoles belonging to parent class of flavonoids synthesized via the phenylpropanoid pathway and accumulated in many flowers, fruits, leaves, stems, roots and other plant tissues ([Mol, Grotewold & Koes, 1998](#); [Dixon & Steele, 1999](#); [Escribano-Bailón, Santos-Buelga & Rivas-Gonzalo, 2004](#); [Espley et al., 2007](#); [Welch, Wu & Simon, 2008](#); [Jimenez-Garcia et al., 2013](#)). The accumulation of anthocyanins contributes to diverse characteristic: red, purple or blue hues, reliant on pH of the vacuole, which not only attracts consumers' attention but also contributes to fruit quality and its medicinal importance ([Rao & Rao, 2007](#); [Singh et al., 2008](#); [Wang & Stoner, 2008](#); [He & Giusti, 2010](#); [Batra & Sharma, 2013](#)).

The anthocyanin biosynthesis pathway has been widely reported in most fruits: apple ([Kim et al., 2003](#)), grapes ([Boss, Davies & Robinson, 1996a](#)), peach ([Tsuda et al., 2004](#)), pear ([Steyn et al., 2004](#)) and plum ([Usenik, Štampar & Veberič, 2009](#)). Primary structural pathway genes encoding enzymes and transcription factors (TFs), which regulate structural gene transcription, are two main elements involved in regulation of anthocyanin synthesis ([Koes, Verweij & Quattrocchio, 2005](#)). The specific flavonoid pathway involving several structural genes and the phenylpropanoid pathway was the first pathway discovered, in which cinnamic acid 4-hydroxylase (*C4H*), phenylalanine ammonia-lyase (*PAL*) and 4-coumarate-CoA ligase (*4CL*) converted phenylalanine to coumarate-CoA. We found some structural genes in the flavonoid pathway which named chalcone isomerase (*CHI*), chalcone synthase (*CHS*), dihydroflavonol 4-reductase (*DFR*), flavanone 3-hydroxylase (*F3H*), flavonol 3'-hydroxylase (*F3' H*), flavonol synthase (*FLS*), leucoanthocyanidin dioxygenase (*LDOX*) and UDP-glucose flavonoid-3-O-glycosyltransferase (*UFGT*). Expression of structural genes in pathways could be

regulated directly by TFs whose functions have been demonstrated in many plants. An intricate of MYB TF, basic helix-loop-helix (bHLH) TFs and WD-repeat proteins in MYB–bHLH–WD40 (MBW) complex combined with promoters of structural genes in anthocyanin biosynthesis (*Gonzalez et al., 2008*). In apple and pear, expression of the *MYB10* gene has been shown to control red apple fruit flesh and foliage anthocyanin accumulation to the exclusion of other non-red pigments (*Takos et al., 2006; Espley et al., 2007*). The biosynthesis of anthocyanins in lotus (*Nelumbo Adans*) was regulated by MBW complex (*NnMYB5*, *NnbHLH1* and *NnTTG1*); the study showed that *NnMYB5* was a transcriptional activator of anthocyanin synthesis that interacted with *NnbHLH1* and *NnTTG1* (*Sun et al., 2016*). TTG1, a WD40 protein, participated in anthocyanin synthesis had also been characterized from fruit species including grapevine (*Kobayashi, Goto-Yamamoto & Hirochika, 2004*) and strawberry (*Schaart et al., 2013*).

In Japanese apricot, fruit colour is ascertained by the accumulation of anthocyanins, most abundant flavonoids. Cyanidin 3-glucoside remained the major anthocyanin in several diversities of red-skinned cultivars. In this present study, red-skinned cultivars were found to be more attractive to consumers than their green counterparts due to their eye-catching appeal and therefore were used for more findings of anthocyanin biosynthesis.

## MATERIALS AND METHODS

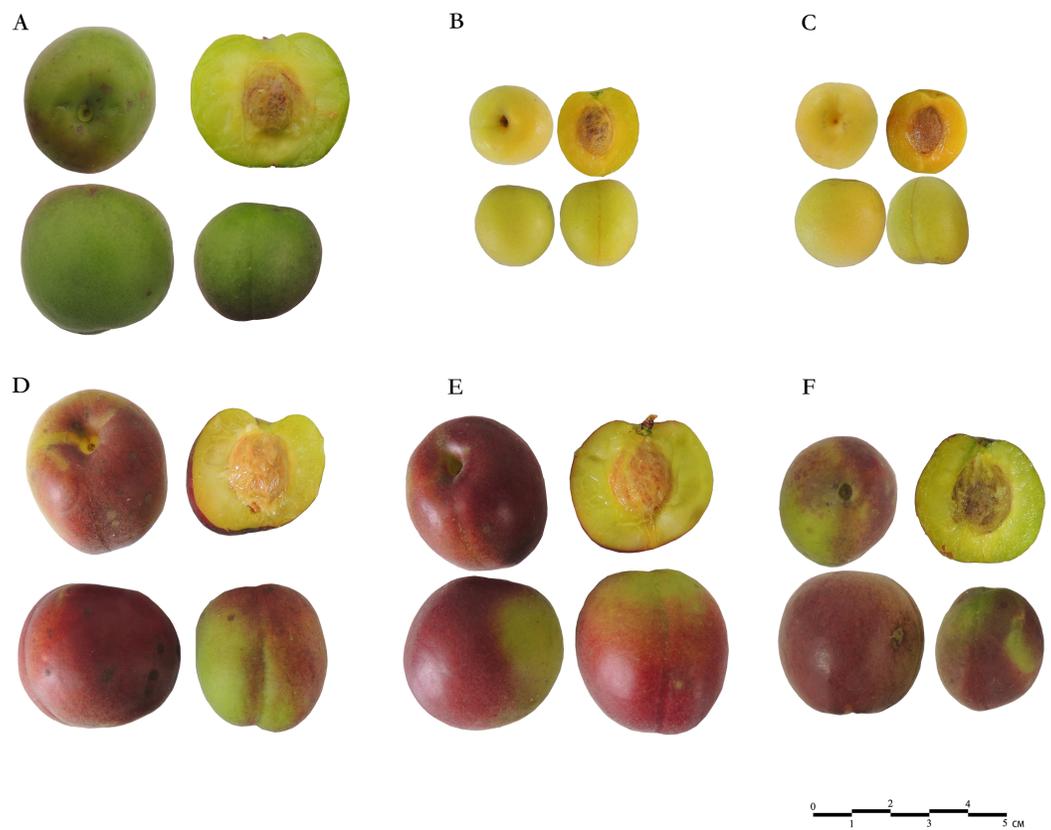
### Plant materials and DNA extraction

Fruits of six different cultivars of Japanese apricot were collected from National Field Genebank for *P. mume* and Waxberry at the Jiangpu Agricultural Research Station, Nanjing Agricultural University, Nanjing, P.R. China. Based on fruit skin colour, these cultivars were classified into two groups: green- and red-skinned cultivars. Green-skinned cultivars included ‘QingjiaNo.2’ (QJM), ‘Yanglao’ (YLM) and ‘Shinano koume’ (SKM), and red-skinned cultivars included ‘Ruantiao hongmei’ (RHM), ‘Xiaoye zhugan’ (XZM) and ‘Zhonghong’ (ZHM). Red-skinned cultivars can be easily distinguished from green-skinned by their specific phenotypic traits. DNA samples were extracted from young leaves of all cultivars using CTAB method following the manufacturer’s protocol to perform genome re-sequencing. Fruits were harvested in triplicate for re-sequencing on 28th May and 2nd June.

### Exploration of physiological indicators and phenotypic traits

Physiological indicators were measured at the same ripening time (*Fig. 1*). The vertical height, flank diameter and width were measured by vernier calliper (GuangLu, Guilin, China). Single fruit weight was measured with an electronic balance (METTLER TOLEDO, Zurich, Switzerland). Fruits were peeled and pulp was squeezed in a cheesecloth, and filtered, homogenized juice was used for determination of soluble solids contents using a digital refractometer (ATAGO, Tokyo, Japan).

The colour of the fruits was observed and calculated with a Minolta CR-400 Chroma Meter (Konica Minolta Sensing, Inc., Osaka, Japan) using an illumination of D75 and at measuring angle of 10° after calibration with an ordinary white plate



**Figure 1** Fruits of six Japanese apricot cultivars used in this study. The top row are three green-skinned cultivars: (A) 'QJM,' (B) 'SKM,' (C) 'YLM'; the lower row are three red-skinned cultivars: (D) 'RHM,' (E) 'XZM,' (F) 'ZHM.' The photographs are provided by Xiaopeng Ni.

Full-size  DOI: [10.7717/peerj.4625/fig-1](https://doi.org/10.7717/peerj.4625/fig-1)

( $Y = 94.00$ ,  $x = 0.3158$ ,  $y = 0.3322$ ). Three records of  $L^*$  represents lightness,  $a^*$  and  $b^*$  represent red–green and yellow–blue chromaticity which coordinates were noted for individually Japanese apricot sunny surface. Each sample was chosen from 10 replicates by changing the position of sunny surface at maturity stage. Chroma and hue angle were calculated as  $\text{Chroma} = (a^{*2} + b^{*2})^{1/2}$  and  $\text{Hue} = \tan^{-1}(b^*/a^*)$  (Oms-Oliu, Soliva-Fortuny & Martín-Belloso, 2008).

### Construction of libraries and Illumina sequencing

To construct a whole genome shotgun sequencing library for whole genome re-sequencing, high-quality genomic DNA was extracted and randomly interrupted by ultrasonic shearing at intervals of 150–800 bp (pair-end library). DNA fragments of required length were isolated by electrophoresis, and then Klenow DNA polymerase, T4 DNA polymerase and T4 PNK were utilized to convert sticky ends to blunt ends. Amplified DNA fragments were finally sequenced using an Illumina HiSeq™ 2500 platform (1gene, Zhejiang, China). The sequences were aligned to the Japanese apricot reference genome (<https://doi.org/10.6084/m9.figshare.6197219.v1>) using the SOAP2 (Short Oligonucleotide Alignment Program 2) software (<http://soap.genomics.org.cn/soapaligner>).

html), and SOAP2 was matched to the single-end and paired-end fragment. According to alignment results, it was possible to calculate the depth and coverage of sequencing with respect to reference genome. Statistical analyses were carried out by means of SPSS 18.0 software (SPSS Inc., Chicago, IL, USA).

### Single nucleotide polymorphisms detection

Single nucleotide polymorphisms (SNPs) were recognized using SOAPsnp (<http://soap.genomics.org.cn/soapsnp.html>) to explain the effects of substitutions (including synonymous SNPs and non-synonymous SNPs). According to the comparison results, a Bayesian model was used to determine the probability of each possible genotype based on actual observed data, while considering data characteristics, sequencing quality and experimental factors. The genotype of specific locus of sequenced individual was selected as the maximum of probability and a quality value reflecting genotype was assigned on the basis of that probability, then consensus sequence was obtained. Based on this consensus sequence, reference sequence exists, polymorphism screening and filtering was done. The SAM tools function 'mpileup' was used to identify unprocessed SNPs population by means of reads through a mapping value  $\geq 20$ . Using the SAM tools database 'vcfutils,' SNPs extracted using the beyond procedure were initially filtered to yield sequencing depths between 35 and 91. Raw SNPs sites were additionally filtered based on subsequent criteria: copy number  $\leq 2$ , and a least of 5 bp apart, with the exclusion of insignificant allele frequencies ( $MAF \geq 0.05$ ) where SNPs were engaged when distance between SNPs was  $< 5$  bp. To check the SNPs calling accuracy of SAMtools, fragments were casually selected and amplified using corresponded primers, and subsequent PCR products exposed to Sanger sequencing for concordance rates. Both HTML and text production files were created from SOAPsnp; output included the chromosome (Chr), position, reference base  $\leftrightarrow$  sample base, SNPs status, strand, annotation type, feature type, codon phase (phase of the codon mutation), codon mutate (the codon before mutation  $\leftrightarrow$  the codon after mutation), amino acid mutate (amino acids before mutation  $\leftrightarrow$  amino acids after mutation), synonymous mutations, non-synonymous mutations, start position, end position, gene id and its functions. Using these annotation files, SNPs could correspond to a homologous functional element region, such as coding sequence (CDS) region, exon region, UTR region and others. This information could help us to select desired SNPs.

### Insertions or deletions detection

SOAPindel software was designed for all paired-end sequencing data, wherein SOAPindel comparison allowed open gaps, and short insertions or deletion (InDel) detections was performed. During short InDels testing, all reads met paired-end requirements, and only one end gaps contained the alignment. In this process, the length of gap was set, not more than 5 bp. In the detection of InDels, at least three pairs of reads were required to define an InDel. For InDels annotation, analysis of regional information and impact of InDels on CDS region, such as in a frameshift mutation, was possible.

### Structural variations detection

The SOAPsv (<http://soap.genomics.org.cn/SOAPsv.html>) program was used to detect structural variations by whole genome de novo assembly system requirements. Structural variations (SVs) is one of the important variations between individuals of the same species. In contrast to paired-end, if there is a structural difference an individual sequence and the reference sequence, a comparison cannot be made. The abnormal condition of this double end alignment can be used to detect SVs. At present, the main types of SVs can be detected, such as insertion, deletion, duplication (tandem duplication and dispersed duplication), inversion (upstream and downstream) and complex. Sometimes, two types of SVs can occur simultaneously.

### DNA level functional annotation and screening of variant genes

Gene functions were annotated by BLAST alignment with Japanese apricot reference genes (*Zhang et al., 2012*) (<http://www.ncbi.nlm.nih.gov/genome/?term=Prunus+mume>) and *Arabidopsis* proteins (*Huala et al., 2001*) (<http://www.arabidopsis.org/>). In order to analyse the gene's biological function, we mapped all the different genes to NR (non-redundant sequence databases) (*Holm & Sander, 1998*) (<http://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins/>), Swiss-Prot protein database (Swiss-Prot) (*Bairoch & Apweiler, 2000*) ([http://web.expasy.org/docs/swiss-prot\\_guideline.html](http://web.expasy.org/docs/swiss-prot_guideline.html)), GO (Gene Ontology) (*Ashburner et al., 2000*) (<http://geneontology.org/page/go-database>), COG (Clusters of Orthologous Groups) (*Tatusov, Koonin & Lipman, 1997*) (<https://www.ncbi.nlm.nih.gov/COG/>) and KEGG (Kyoto Encyclopedia of Genes and Genomes) (*Kanehisa et al., 2004*) (<http://www.genome.jp/kegg/>). These annotation databases were used for comparisons with homologous genes.

Explanation of data based on homologous genes was utilized to monitor for unique genes associated with colour formation. Non-synonymous SNPs, InDels or SVs were examined on the basis of green-skinned ('QJM,' 'YLM' and 'SKM') and red-skinned ('RHM,' 'XZM' and 'ZHM') cultivars phenotypes.

### Reverse transcription-quantitative polymerase chain reaction

Reverse transcription-quantitative polymerase chain reaction (qRT-PCR) was carried out to identify the comparative alteration in expression of genes identified in re-sequencing analysis. Total RNA samples were extracted from peels using CTAB extraction method and used for expression analysis (*Tong et al., 2009*).

SuperScript II RT (Invitrogen) was used to synthesize the first-stand cDNA from Individual sample total RNA with an Oligo primer. qRT-PCR reactions were performed in 96-wells plates using a 7300 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) using SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA). Beacon designer 7.0 program (Premier Biosoft International, Palo Alto, CA, USA) was used to design gene-specific primers. Each qRT-PCR reaction (20  $\mu$ L) confined 8.6  $\mu$ L ddH<sub>2</sub>O, 0.2  $\mu$ L (100 nM) of each primer, 10  $\mu$ L SYBR Green II Master Mix and 1  $\mu$ L of diluted cDNA. To standardize the entire quantity of cDNA in individually reaction, a housekeeping gene, *RPII* (sense: 5'-TGAAGCATACACCTATGATGATGAAG-3',

**Table 1** Sequences of gene associated the biosynthesis of anthocyanin primers used in qRT-PCR analysis of different genes.

Functions	Gene	Primer name	Sequence 5'-3'	Gene ID	Size (bp)
Structural genes	<i>4-CL</i>	4-CL.F	CGCCAAAATCCTTCAACCC	103339831	298
		4-CL.R	ACGATTGAGAAGAAGCAGAC		
	<i>F3H</i>	F3H.F	CCAAGTGGCAACTGCTGAAAC	103326444	413
		F3H.R	CCGTGATCTCCCTTGCTCAG		
	<i>UFGT</i>	UFGT.F	ATTAAATGCCTTTGAATTGGTA	103328988	297
		UFGT.R	CAAAATCAGAACCACCGAAT		
Transcription factors	<i>MYB114</i>	MYB114.F	GTAGACTAAGGTGGTTGAA	103342473	139
		MYB114.R	TGATGGTTAGGGAAGAGA		
	<i>MYB29</i>	MYB29.F	GCAGCTTGACCATCCTCCATA	103334039	307
		MYB29.R	ATGGTGGGTTGTTGGGTACAT		
	WD40	WD40.F	TCCGGCCTTCTCCACCGA	103323546	300
		WD40.R	ACTGCAAAGTCCCAATAGCGAAG		
	bHLH30	BHLH30.F	CTCTTCGACCCTTTTACGC	103324936	178
		BHLH30.R	GATTCTCAAAGCCGCCAA		
	NAC	NAC.F	CCCCAAAACAGCAAAGAACGAG	103323499	667
		NAC.R	TCCCGAAAGAGCCCAACCTCA		
Regulative genes	<i>GST1</i>	GST1.F	GAAGATCCCAGTGCTCGTTC	103323824	154
		GST1.R	ATCGTCAGCGAATTTAACCC		
	<i>RT1</i>	RT1.F	CTGCCTTGCTTGTATGACCA	103335106	109
		RT1.R	CCAATGATTTCTGCTAAACGA		
	<i>UGT85A2</i>	UGT85A2.F	TTCCAGGAATGAACGGCATC	103323808	161
		UGT85A2.R	AACATCTTGCTCCAATGCATC		
	<i>ABCG12</i>	ABCG12.F	TGGCAGAAGTGTAAATCACC	103326861	121
		ABCG12.R	TAGAGCCATATTTGCTTGTCC		
	<i>MATE1</i>	MATE1.F	ATCTATATGAGTCTTCGAGCA	103326834	101
		MATE1.R	TTGAGTAAACCAGCGTCCA		

**Note:**

*4-CL*, 4-coumarate-CoA ligase-like 7; *F3H*, flavanone 3-hydroxylase; *UFGT*, UDP-glucose flavonoid 3-O-glucosyltransferase 3-like; *MYB114*, transcription factor MYB114; *MYB29*, transcription factor MYB29-like; *WD40*, WD repeat-containing protein 75; *bHLH30*, transcription factor bHLH30-like; *NAC*, NAC transcription factor 29; *GST1*, probable glutathione S-transferase; *RT1*, UDP-rhamnose: rhamnosyltransferase 1; *UGT85A2*, UDP-glycosyltransferase 85A2-like; *ABCG12*, ABC transporter G family member 12-like; *MATE1*, MATE efflux family protein 1.

antisense: 5'-CTTTGACAGCACCAGTAGATTCC-3') was used as a reference gene in control reactions (*Tong et al., 2009*). All primers that were used in reactions are listed in [Table 1](#). The expression levels of relative genes were measured using the  $2^{-\Delta\Delta C_t}$  method (*Ramakers et al., 2003*).

### Extraction and analysis of anthocyanin

Extraction of anthocyanins was carried out conferring to previously reported methods with some alterations (*Pomar, Novo & Masa, 2005*). Peels (2 g) were detached from pulp to obtain juice with 2% formic acid in methanol (10 mL) for 24 h in dark. The process was reiterated four times. The extracts attained were assorted and centrifuged at 12,000 rpm for 25 min, and then the supernatant was evaporated to dryness at 40 °C and reconstituted in 2% formic acid in methanol (5 mL). Individual sample was accomplished for third

times, and the extracts were used for anthocyanin pigmentation identification. Cyanidin 3-glucoside (Cy-3-glc) (Sigma Chemical, St. Louis, MO, USA) was used as a standardized, dissolved in acidified MeOH (2% formic acid) to obtain a concentration of 1 mg/mL. The UPLC system with the Xevo<sup>®</sup> G2-XS QToF Mass Spectrometer was used for separation according to the method described by *Hosseiniyan, Li & Beta (2008)*.

### UFGT enzyme activity

Pericarp (1 g) was pulverized with a mortar and pestle in liquid N<sub>2</sub>. Reaction substrate was prepared following the method (*Wu et al., 2017*) with some modifications. Reaction tubes including enzyme were incubated at 25 °C for 50 min, and stopped through accumulation of 20% trichloroacetic acid in methanol (75 µL).

An ultra-performance liquid chromatograph (UPLC) was fortified with a opposite segment column ACQUITY UPLC HSS C18 (1.8 µm particle sizes, 100 mm × 2.1 mm I.D.) (Waters, Milford, MA, USA) for separation. The UPLC technique was carried out with the method described by *Wu et al. (2017)*. UFGT were enumerated using quercetin-3-galactoside (Sigma Chemical, St. Louis, MO, USA) as an ordinary. Lastly, UFGT action was characterized by mg quercetin-3-gal · g<sup>-1</sup> FW.

### Statistical analysis

Analysis of variance (ANOVA) was carried out to compare cultivar mean values using IBM SPSS Statistics 18 (SPSS Inc., Chicago, IL, USA). The least significant difference test was employed to determine variances between means at a 5% significance level. GraphPad Prism version 6.0 (GraphPad Software, San Diego, CA, USA) was used for graph plotting.

## RESULTS

### Variation in phenotypic traits of six Japanese apricot cultivars

Phenotypic observations showed a significant difference between red- and green-skinned cultivars due to red pigmentation of fruit skin (*Fig. 1*). 'XZM' showed highest red pigmentation from all cultivars followed by 'RHM' and 'ZHM,' whereas green-skinned cultivars 'SKM' and 'YLM' had trace amounts or no colour pigmentation. Phenotypic indicators revealed no significant differences in flank diameter, vertical diameter and width among all red-skinned, whereas the values for 'QJM' and other two green-fleshed cultivars, 'SKM' and 'YLM,' were significantly lower than those of all other cultivars (*Table 2*). The 'QJM' cultivar had the highest fruit weights, while other two green-skinned cultivars were approximately 1/8 to 1/7 the weight of 'QJM.' The recorded fruit weights of three green-skinned cultivars 'QJM,' 'SKM' and 'YLM' were 36.89, 5.27 and 4.36 g, respectively. The vertical diameter of all red-skinned were more than all green-skinned cultivars, for 'width and flank diameter,' 'QJM' and all red-skinned cultivars were close in value and had no significant difference except two green-skinned cultivars, 'SKM' and 'YLM.' The weights of red-skinned cultivars 'RHM,' 'XZM' and 'ZHM' were recorded as 26.68, 24.22 and 30.17 g, respectively (*Table 2*). The highest soluble solids were found in 'QJM' (9.16%), closely matched by 'SKM' (9.12%), whereas the concentration in other

**Table 2** Skin colour indexes of six Japanese apricot cultivars.

Parameters	Green-skinned			Red-skinned		
	'QJM'	'SKM'	'YLM'	'RHM'	'XZM'	'ZHM'
Vertical diameter (mm)	33.04 ± 0.83 <sup>c</sup>	22.31 ± 0.38 <sup>d</sup>	20.11 ± 0.38 <sup>e</sup>	34.37 ± 0.49 <sup>bc</sup>	35.21 ± 0.49 <sup>b</sup>	37.59 ± 1.03 <sup>a</sup>
Width (mm)	35.95 ± 0.85 <sup>b</sup>	20.61 ± 0.32 <sup>c</sup>	19.58 ± 0.35 <sup>c</sup>	38.42 ± 0.55 <sup>a</sup>	35.82 ± 0.55 <sup>b</sup>	37.79 ± 1.08 <sup>a</sup>
Flank diameter (mm)	33.84 ± 1.04 <sup>ab</sup>	19.76 ± 0.32 <sup>c</sup>	18.43 ± 0.37 <sup>c</sup>	33.79 ± 0.48 <sup>ab</sup>	32.13 ± 0.52 <sup>b</sup>	35.06 ± 1.27 <sup>a</sup>
Per Fruit Weight (g)	36.89 ± 2.04 <sup>ab</sup>	5.27 ± 0.21 <sup>d</sup>	4.37 ± 0.25 <sup>d</sup>	26.68 ± 0.79 <sup>bc</sup>	24.22 ± 1.03 <sup>c</sup>	30.17 ± 3.06 <sup>b</sup>
Soluble Solids (%)	9.16 ± 0.41 <sup>a</sup>	9.12 ± 0.33 <sup>a</sup>	8.22 ± 0.22 <sup>b</sup>	7.09 ± 0.14 <sup>c</sup>	5.79 ± 0.21 <sup>d</sup>	8.83 ± 0.64 <sup>b</sup>
$L^*$	53.76 ± 0.41 <sup>b</sup>	59.95 ± 0.61 <sup>a</sup>	60.41 ± 0.59 <sup>a</sup>	30.49 ± 0.65 <sup>d</sup>	31.52 ± 0.34 <sup>d</sup>	33.12 ± 0.33 <sup>c</sup>
$a^*$	-18.83 ± 0.23 <sup>d</sup>	-9.77 ± 1.07 <sup>c</sup>	-17.01 ± 0.79 <sup>d</sup>	15.93 ± 0.75 <sup>a</sup>	16.58 ± 0.32 <sup>a</sup>	12.62 ± 0.49 <sup>b</sup>
$b^*$	34.26 ± 0.31 <sup>c</sup>	43.91 ± 0.73 <sup>a</sup>	40.52 ± 0.84 <sup>b</sup>	11.93 ± 0.59 <sup>dc</sup>	11.18 ± 0.37 <sup>c</sup>	13.37 ± 0.43 <sup>d</sup>
C	39.1 ± 0.37 <sup>b</sup>	45.35 ± 0.76 <sup>a</sup>	44.25 ± 0.64 <sup>a</sup>	20.08 ± 0.81 <sup>c</sup>	20.04 ± 0.42 <sup>c</sup>	18.53 ± 0.47 <sup>c</sup>
h	118.78 ± 0.15 <sup>a</sup>	101.87 ± 1.56 <sup>c</sup>	113.15 ± 1.17 <sup>b</sup>	37.13 ± 1.48 <sup>e</sup>	33.82 ± 0.69 <sup>f</sup>	46.84 ± 1.42 <sup>d</sup>

**Note:**

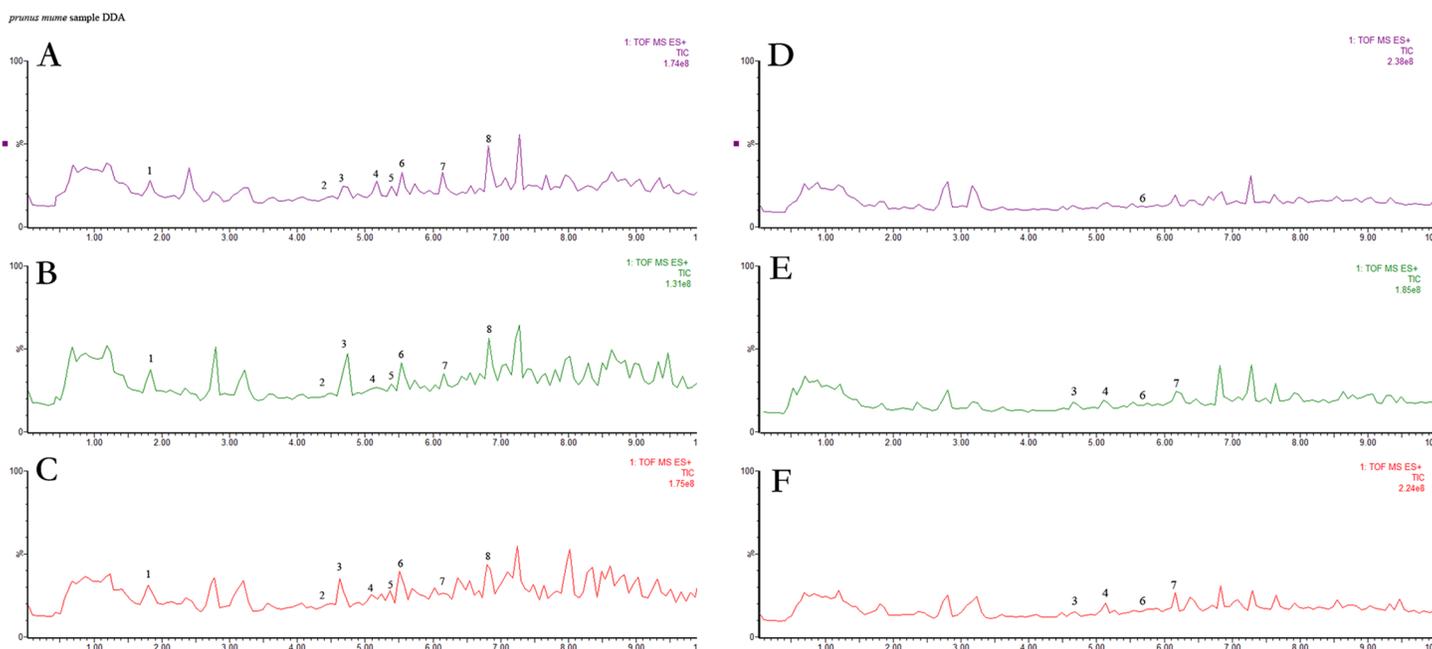
$L^*$  means lightness,  $a^*$  means red–green chromaticity,  $b^*$  means yellow–blue chromaticity. Each sample chose from 10 replications by changing the position of the sunny surface at maturity stage. C means Chroma, and h means hue angle, they were calculated being Chroma =  $(a^{*2} + b^{*2})^{1/2}$  and Hue =  $\tan^{-1}(b^*/a^*)$ , respectively. Values represent as a mean ± standard deviation,  $n = 3$ . Different letters in rows indicate significant differences among mean values of treatments ( $p < 0.005$ ).

cultivars ranged between 5.79% and 8.83%. The soluble solids contents in red-skinned cultivars were recorded lower than those in green-skinned cultivars (Table 2).

CIELab parameters were also measured for these cultivars (Table 2). Significant differences appeared between red- and green-skinned cultivars. 'XZM' exhibited highest  $a^*$  (16.58) value and lowest  $b^*$  (11.18) value, representing the highest amount of red pigmentation among the studied cultivars, followed by 'RHM' and 'ZHM' with  $a^*$  values of 15.93 and 12.62, respectively, and  $b^*$  values of 11.93 and 13.37, respectively. 'QJM' showed the lowest  $a^*$  (-18.83) value and a higher  $b^*$  (34.26) value, revealing the highest amount of green pigmentation, whereas 'YLM' revealed the second lowest  $a^*$  (-17.01) value above 'SKM' (-9.77) and a lower  $b^*$  (40.52) than 'SKM' (43.91), indicating higher amount of green pigmentation but a lower amount of yellow pigmentation. Moreover, lightness ( $L^*$ ), hue angle ( $H^*$ ) and Chroma ( $C^*$ ) varied significantly between red- and green-skinned cultivars, all three parameters were higher in green-skinned than in red-skinned cultivars.

### Anthocyanin identified in Japanese apricot by UPLC/MS

Anthocyanins are the key pigments responsible for colour in many flowers and fruits, accumulating only during plant ripening (Lancaster et al., 1994; Salvatierra et al., 2010; Rahim, Busatto & Trainotti, 2014; Wei et al., 2015; Wang et al., 2015). Anthocyanins coordinate to the flavonoid class of compounds and structurally comprise an anthocyanidin aglycone bound to one or more sugar moieties (Jaakola, 2013). An entire of eight anthocyanin compounds were recognized on the basis of reference cyanidin-3-*O*-glucosides and mass spectrometry (MS) (Mazza, Cacace & Kay, 2004) (Fig. 2). UPLC in combination with MS provided a sophisticated technique for anthocyanin analysis. Eight anthocyanin compounds were present in all red varieties studied with differences in retention time (Rt), including cyanidin-3-*O*-glucoside (peak 1), cyanidin-3-*O*-(6-*O*-malonyl- $\beta$ -D-glucoside)



**Figure 2** UPLC chromatograms of anthocyanin extracts captured at 520 nm for anthocyanin of six cultivars. (A) ‘ZHM,’ (B) ‘XZM,’ (C) ‘RHM,’ (D) ‘YLM,’ (E) ‘SKM,’ (F) ‘QJM.’ Peak identification is shown in Table 3. [Full-size](#) DOI: 10.7717/peerj.4625/fig-2

(peak 2), cyanidin-3,5-*O*-diglucoside (peak 3), delphinidin-3-*O*-arabinoside (peak 4), cyanidin-3-*O*-(6''-acetyl-galactoside) (peak 5), delphinidin-3-*O*-glucoside (peak 6), delphinidin-3-*O*-(6''-acetyl-galactoside) (peak 7) and peonidin-3-*O*-glucoside (peak 8) (Fig. 2), whereas few green-skinned cultivars contained all the anthocyanins. Three anthocyanin compounds: cyanidin-3,5-*O*-diglucoside, delphinidin-3-*O*-arabinoside and delphinidin-3-*O*-(6''-acetyl-galactoside) were only detected in ‘QJM’ and ‘SKM’ and not the ‘YLM’ cultivar. Delphinidin-3-*O*-(6''-acetyl-galactoside) was detected in all three green-skinned cultivars (Table 3). Cyanidin-3-*O*-glucoside (Cy-3-glu) content was the highest among all anthocyanin compounds, followed by cyanidin-3-*O*-(6-*O*-malonyl-β-D-glucoside) (Cy-3-mal-glu) and delphinidin-3-*O*-arabinoside (Dp-3-ara). Peonidin-3-*O*-Glucoside is the lowest from all anthocyanins. ‘ZYM’ had the highest content of ‘Cy-3-glu’ (124.2 mg/kg), followed by ‘RHM’ (35.2 mg/kg) and ‘ZHM’ (32.3 mg/kg) (Table 3).

### Re-sequencing of six cultivars

Genomic variants amongst the six genotypes incorporated SNPs, InDels and SVs. Six DNA libraries were built in order to perform whole genome re-sequencing. In total, 92.2 gigabits (Gb) of crude sequence data, about 15.3 Gb (clean data) per individual remained after filtering the paired-end sequences of individual cultivar (Table 4), that translated to an average of 49.1× depth of coverage per genotype. The reference genome was available from NCBI Sequence Read Archive (SRA) for alignment (BioProject PRJNA371370, accessions SRR5241550–SRR5241555), and full assembly and analysis data are available at <https://doi.org/10.6084/m9.figshare.6197219.v1>. Every individual sample read sequence was compared to the reference genome by means of SOAP2

**Table 3** Anthocyanins identification in all six Japanese apricot cultivars by UPLC/MS.

Peak No.	Anthocyanins	Retention time (min)	[M] <sup>+</sup> (m/z)	Frag. MS2 (m/z)	Content (mg/kg)					
					QJM	SKM	YLM	RHM	ZYM	ZHM
1	Cyanidin-3-O-glucoside	1.80	449	287	ND	ND	ND	35.2 ± 0.6 <sup>b</sup>	124.2 ± 0.1 <sup>a</sup>	32.3 ± 0.6 <sup>b</sup>
2	Cyanidin-3-O-(6-O-malonyl-β-D-glucoside)	4.51	535	287	ND	ND	ND	52.8 ± 0.3 <sup>b</sup>	99.2 ± 0.7 <sup>a</sup>	56.2 ± 0.3 <sup>b</sup>
3	Cyanidin-3,5-O-diglucoside	4.69	611	287	4.8 ± 0.5 <sup>d</sup>	8.0 ± 0.3 <sup>c</sup>	ND	48.8 ± 0.4 <sup>a</sup>	24.3 ± 0.3 <sup>b</sup>	12.8 ± 0.5 <sup>c</sup>
4	Delphinidin-3-O-arabinoside	5.19	465	303	13.6 ± 0.8 <sup>c</sup>	0.8 ± 0.6 <sup>d</sup>	ND	72.8 ± 0.3 <sup>a</sup>	49.6 ± 0.6 <sup>b</sup>	88.2 ± 1.0 <sup>a</sup>
5	Cyanidin-3-O-(6"-acetyl-galactoside)	5.39	491	287	ND	ND	ND	19.3 ± 0.1 <sup>a</sup>	9.2 ± 0.8 <sup>b</sup>	3.7 ± 0.1 <sup>c</sup>
6	Delphinidin-3-O-glucoside	5.63	465	303	2.6 ± 0.3 <sup>d</sup>	6.4 ± 0.8 <sup>c</sup>	5.6 ± 0.3 <sup>c</sup>	17.5 ± 0.5 <sup>a</sup>	14.1 ± 0.4 <sup>b</sup>	16.6 ± 0.3 <sup>a</sup>
7	Delphinidin-3-O-(6"-acetyl-galactoside)	6.18	507	303	1.6 ± 0.2 <sup>c</sup>	3.2 ± 0.4 <sup>c</sup>	ND	10.7 ± 0.6 <sup>a</sup>	10.4 ± 0.2 <sup>a</sup>	5.7 ± 0.4 <sup>b</sup>
8	Peonidin-3-O-Glucoside	6.99	463	301	ND	ND	ND	0.8 ± 1.0 <sup>b</sup>	0.3 ± 0.5 <sup>c</sup>	3.2 ± 0.4 <sup>a</sup>

**Notes:**  
Values represent as a mean ± standard deviation,  $n = 3$ . Different letters in rows indicate significant differences among mean values of treatments ( $p < 0.005$ ). ND, no detected.

**Table 4** Summary of genome re-sequencing and variations for all six Japanese apricot cultivars.

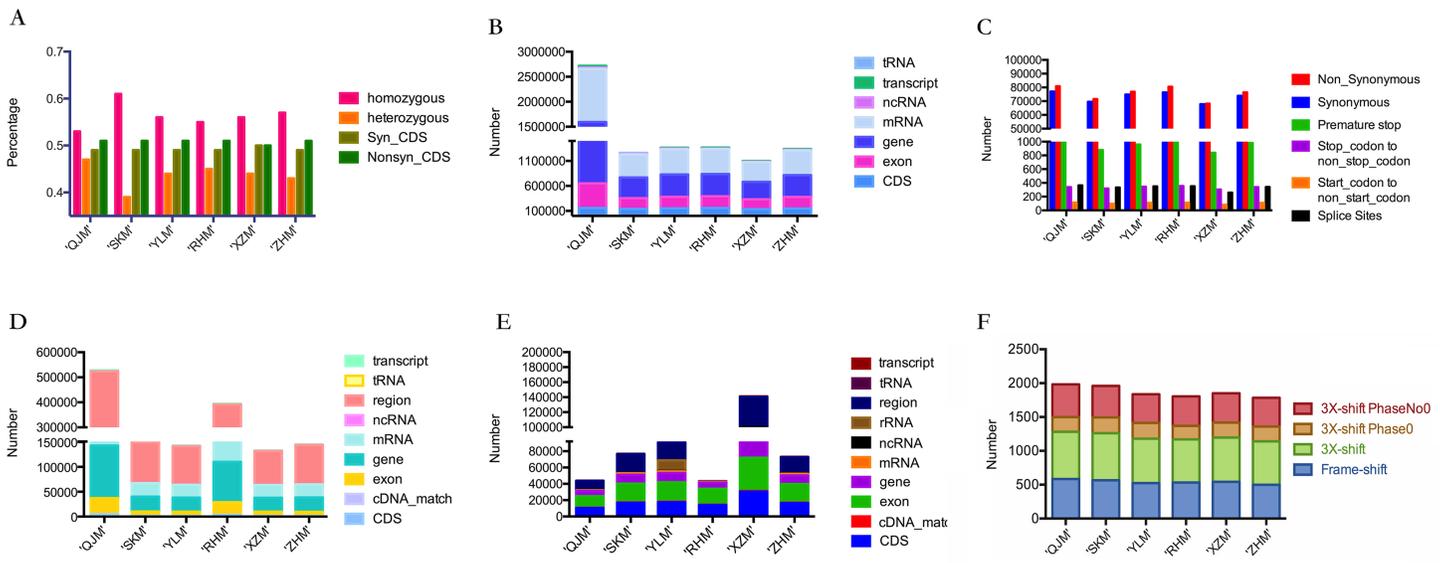
Classification	Genotype	Clean bases (Gb) <sup>a</sup>	Coverage rate (%) <sup>b</sup>	Sequencing depth <sup>c</sup>	Mapping bases rate (%) <sup>d</sup>	SNP (M) <sup>e</sup>	InDel (K) <sup>f</sup>	SV (K) <sup>g</sup>
Green-skinned	'QJM'	8.9	85.3	32.4	78.9	1.4	77.4	40.1
	'SKM'	20.5	83.0	64.0	67.9	1.2	82.1	22.0
	'YLM'	21.9	82.8	69.2	68.6	1.3	77.4	12.9
Red-skinned	'RHM'	8.6	84.5	31.3	79.3	1.2	72.2	19.9
	'XZM'	11.9	79.9	33.7	61.3	0.9	68.1	11.4
	'ZHM'	20.4	83.0	63.8	68.1	1.3	79.2	22.9
	Total	92.2	83.1	49.1	70.7	7.4	456.2	129.1

**Notes:**  
M and K mean million and thousand, respectively. Three total terms of last row (coverage rate, sequencing depth and mapping bases) was the average of data, others total terms were the summation of data.  
<sup>a</sup> Correction or removal of erroneous (dirty) data caused by contradictions, disparities, keying mistakes, missing bits, etc.  
<sup>b</sup> Percent of target genome size in the reference genome size.  
<sup>c</sup> Percent of clean data in the reference genome size.  
<sup>d</sup> Percent of bases mapping in the reference genome size.  
<sup>e</sup> Single nucleotide polymorphism.  
<sup>f</sup> Insert and deletion.  
<sup>g</sup> Structural variant.

(Short Oligonucleotide Alignment Program 2). The coverage rate ranged from 79.9% to 85.3%. Of all cultivars studied, 'QJM' had the highest coverage rate (85.3%), while 'XZM' (79.9%) had the lowest, and that of the other four cultivars was approximately 83%. The base mapping rate ranged from 61.3% for 'XZM' to 79.3% for 'RHM,' and 70.7% was mapped to reference sequence on average, including 55.6% that mapped uniquely.

### SNPs detection

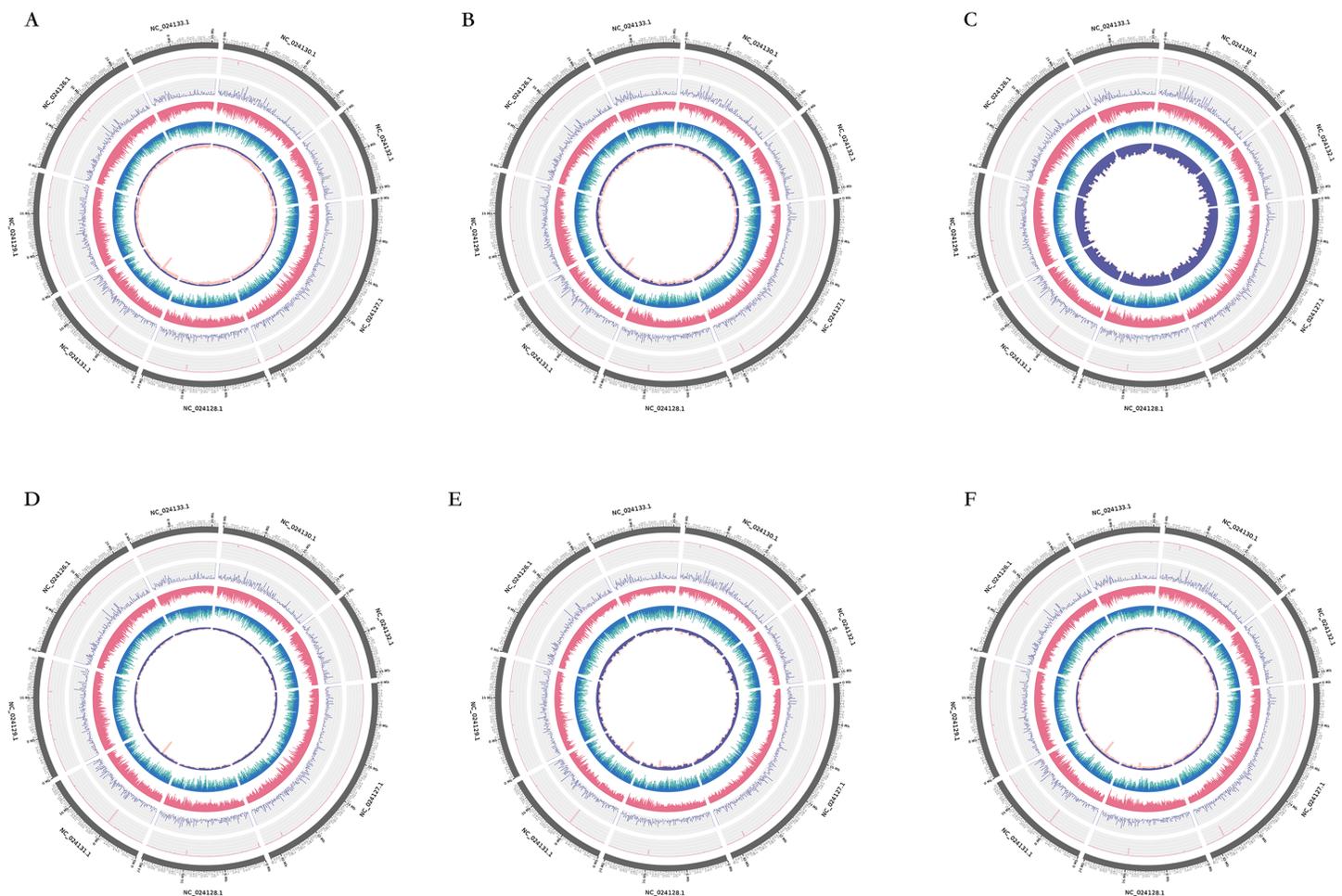
SOAPSnp (<http://soap.genomics.org.cn/soapsnp.html>) software was used to detect SNPs and considered the main variant among these six genotypes. On the basis of same



**Figure 3** Statistics of SNPs, InDel and SVs. (A) Total number of variants, type and zygosity of variants in each genotype. Syn\_CDS means synonymous SNPs in the CDS region, nonsyn\_CDS means non-synonymous SNPs in the CDS region; (B) the distribution of SNPs in different regions among all six cultivars; (C) summary of large effect SNPs in all cultivars of Japanese apricot; (D) the distribution of InDels in different regions among all six cultivars; (E) the distribution of SVs in different regions among all six cultivars; (F) frameshift mutation occurred among all Japanese apricot cultivars. Frame-shift means frameshift mutation in the CDS region, 3X-shift means 3X-shift mutation in CDS region, 3X-shift Phase0 means 3X-shift mutation in CDS region (phase 0), 3X-shift PhaseNo0 means 3X-shift mutation in CDS region (phase No 0).

Full-size [DOI: 10.7717/peerj.4625/fig-3](https://doi.org/10.7717/peerj.4625/fig-3)

sequence information (i.e. after comparing all SNPs of the six samples), filtering out the polymorphic loci detected between reference sequence and genotype by low quality and depth, high reliability of SNP data was attained. A total of 1,387,518 SNPs were detected in ‘QJM,’ of which 53.09% were homozygous and 46.91% were heterozygous (Fig. 3A). Approximately 48.90% were synonymous SNPs in CDS regions (Syn\_CDS) and 51.10% were non-synonymous SNPs in CDS regions (Nonsyn\_CDS). A total of 1,240,360 SNPs were detected in ‘SKM,’ of which 60.54% were homozygous and 39.46% were heterozygous. Approximately 49.41% were Syn\_CDS and 50.59% were Nonsyn\_CDS. A total of 1,331,341 SNPs were identified in ‘YLM,’ of which 56.42% were homozygous and 43.58% were heterozygous. Approximately 49.46% were Syn\_CDS, and 50.54% were Nonsyn\_CDS. A total of 1,231,625 SNPs were detected in ‘RHM,’ of which 54.98% were homozygous and 45.02% were heterozygous. Approximately 48.78% were Syn\_CDS and 51.22% were Nonsyn\_CDS. A total of 894,302 SNPs were detected in ‘XZM,’ of which 56.24% were homozygous and 43.76% were heterozygous. Approximately 49.97% of which were Syn\_CDS and 50.03% were Nonsyn\_CDS. A total of 1,301,911 SNPs were detected in ‘ZHM,’ of which 56.65% were homozygous and 43.35% were heterozygous. Approximately 49.29% were Syn\_CDS and 50.71% were Nonsyn\_CDS (Fig. 3A). The distribution of SNPs corresponding to functional element regions (such as CDS, exons, UTR, etc.) revealed that the mRNA region (46.15%) occupied the majority of SNPs among all functional regions, followed by the gene region (41.05%). Exons accounted for third most enriched region, and considered to be an important CDS region, the other



**Figure 4** Diagrammatic representation of structural variation in the sequence data obtained from the re-sequencing of six Japanese apricot cultivars. (A) 'QJM,' (B) 'SKM,' (C) 'YLM,' (D) 'RHM,' (E) 'XZM,' (F) 'ZHM.' There are six circles inside and out, and the eight chromosomes are portrayed along the perimeter of the outer circle. The depth of re-sequencing and coverage are represented in the second and third circle, respectively, every window's size is 100 kb. The fourth circle represents the distribution of SNP density (red); the quantity of InDel which contains an insertion (green) or deletion (blue) represents in the fifth circle. The innermost circle indicates the account of SV which contains an insertion (pink) or deletion (purple). [Full-size !\[\]\(f5a508cc6d05e5d06b117ced927b1acd\_img.jpg\) DOI: 10.7717/peerj.4625/fig-4](https://doi.org/10.7717/peerj.4625/fig-4)

three regions (ncRNA, transcript and tRNA) had a small part in SNPs variation. The distribution of SNPs in different regions was not significantly different among the six cultivars, except for SNPs in exon, gene and mRNA regions of 'QJM' and SNPs in transcript and tRNA regions of 'SKM' (Fig. 3B). The distribution of SNPs types was similar among eight chromosomes in all six cultivars (Table S1; Fig. 4). 'QJM' exhibited the highest sequence variation of all six cultivars with respect to the *P. mume* reference genome, while 'SKM' showed the lowest sequence variation from all cultivars. The genome-wide mutation rate for 'QJM' was one change per 64 bases versus one change per 110 bases for 'SKM.' Among eight chromosomes that contain the reference genome, uppermost quantity of SNPs was perceived in chr 2 in all six genotypes. Particularly, chr 7 in both 'QJM' and 'SKM' showed the lowest level of mutations. Chr 8 in the other four cultivars exhibited the lowest level of mutations, respectively.

Single nucleotide polymorphisms in CDS region are worthy of attention based on explanation of the *P. mume* reference genome. SNPs that change the translation start–stop site or variable splice sites would cause phenotypic change. Approximately 10,397 (0.14%) of identified SNPs were considered large-effect SNPs consisting of premature stop (5,787 SNPs), stop-codon to non-stop codon (1,997 SNPs), start-codon to non-start codon (616 SNPs) and splice site (1,997 SNPs). The cultivar ‘QJM’ demonstrated the highest number of large-effect SNPs (1,882 SNPs), ‘RHM’ (1,878 SNPs) showed no significant difference from ‘QJM,’ followed by ‘ZHM’ (1,768 SNPs), ‘YLM’ (1,762 SNPs) and ‘SKM’ (1,626 SNPs), whereas ‘XZM’ revealed the lowest number of large-effect SNPs (Fig. 3C).

### Short InDels detection

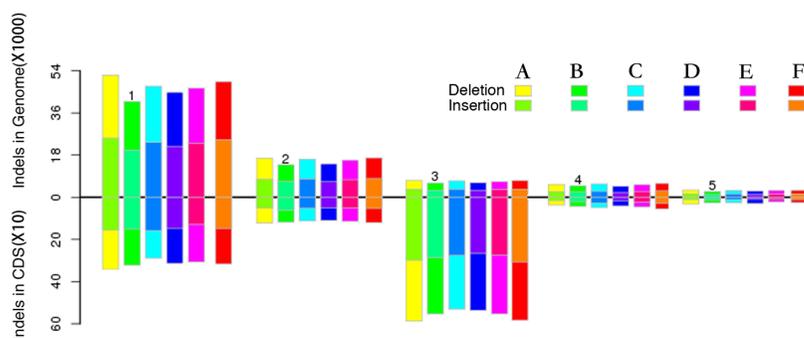
SOAPindel software was used to detect short InDels, in accordance with the statistics of SNPs (Table 4). A total of 1,493,577 InDels were acquired in all six cultivars, and majority of variants were found in region (53.1%), followed by the gene region (20.1%) and mRNA region (18.5%). Approximately 1/3 of all InDels were fixed in ‘QJM’ (527,815), giving ‘QJM’ the highest number of InDels variations. ‘XZM’ contained the lowest number of InDels in all six cultivars (132,876), and remaining four cultivars: ‘SKM’ (150,691), ‘YLM’ (142,748), ‘XZM’ (132,876) and ‘ZHM’ (145,182) are shown (Fig. 3D).

The distribution of InDels was not significantly different from that of SNPs among eight chromosomes in all six cultivars. ‘SKM’ (82,072) showed highest number of InDels of all six genotypes, followed by ‘ZHM’ (79,160), ‘QJM’ (77,390) and ‘YLM’ (77,381), while ‘XZM’ contained the lowest number of InDels of all cultivars. The percentages of insertions and deletions among eight chromosomes in all six cultivars were similar: insertions ranged from 47.1% to 49.5%, slightly less than deletions (50.5% to 52.9%). Chr 2 had the most abundant variation of all eight chromosomes, accounting for 20% of all variation (Table S2).

Frameshift mutations are genetic mutations affected by insertions and deletions of multiple nucleotides in a DNA sequence. Some InDels were projected to be frameshift mutations, but the lengths of InDels in coding regions were more probable from 1 to 5 (the length of a codon), than InDels in the rest of the genome (Fig. 5). In this study, two types of mutation were detected; the most frequent frameshift was a 3X-shift mutation in CDSs (3,987), including phase0 (1,320) and phaseNo0 (2,667), followed by frameshift mutations in CDSs, which were the second highest number of mutations (3,245), whereas ‘QJM’ had the highest number of two-shift mutations (Fig. 3F).

### SVs detection

Structure variation was important between individuals of the same species, and the types detected could be insertions and deletions, repetitions, inversions and translocations. Because of their small contribution to variation, types of variation other than insertions and deletions were classified as others. The distribution of SVs in different regions showed (Fig. 3E) no significant difference between SNPs and InDels, and this region was the highest amount of structure variation among all regions. ‘XZM’ had the most variation in CDSs, the regions of most interest to us, ‘QJM’ cultivar has the least variable while other



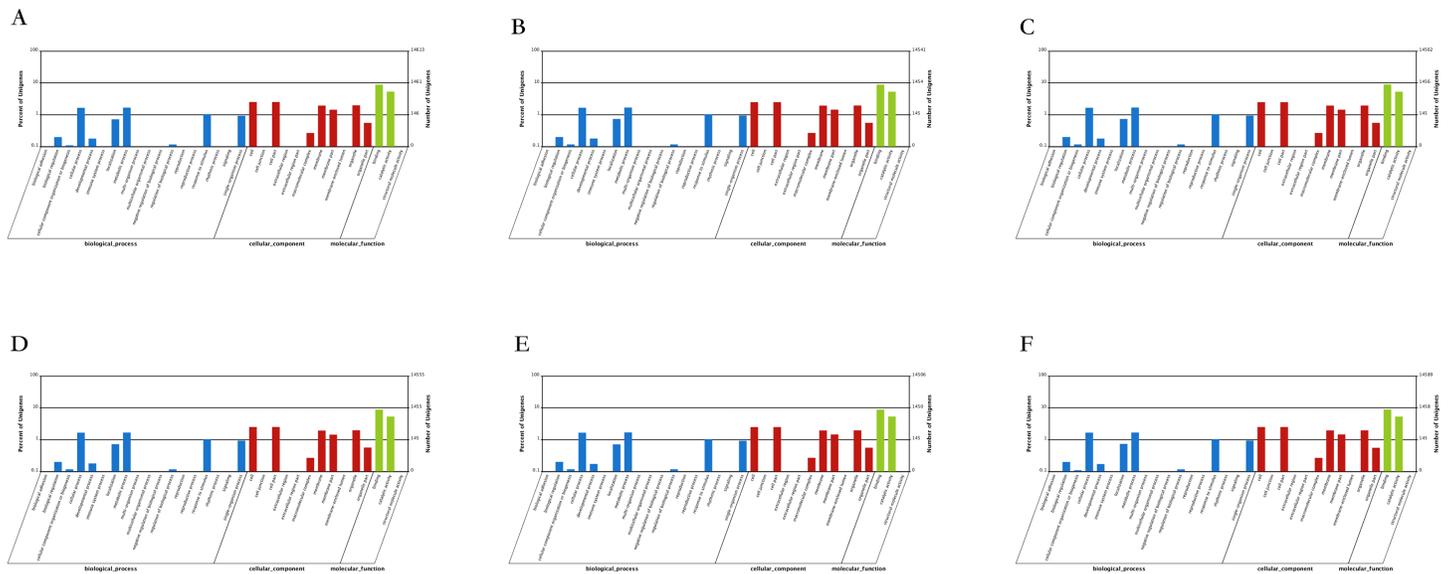
**Figure 5** Different InDel size distributions and frameshift mutations between the whole genome and CDS region. (A) ‘QJM,’ (B) ‘SKM,’ (C) ‘YLM,’ (D) ‘RHM,’ (E) ‘XZM,’ (F) ‘ZHM.’ The  $y$ -axis indicates the number of InDels in the CDS and genome, the  $x$ -axis indicates the different sizes of InDel distributions and frameshift mutations. [Full-size !\[\]\(ebfe6d37ad86655679811e032f633da4\_img.jpg\) DOI: 10.7717/peerj.4625/fig-5](https://doi.org/10.7717/peerj.4625/fig-5)

four cultivars were not significantly different. The distribution of SVs was not significantly different from that of InDels and SNPs among eight chromosomes in all six cultivars (Table S3). The highest variation was observed on chr 2 (16.6%–18.4%), and distribution of other chromosomes was similar. The percentage of insertions and deletions was significantly different among all cultivars between green- and red-skinned: approximately 60% of SVs was insertion in ‘QJM,’ while ‘YLM’ had the lowest number of insertions of all cultivars (7%). A total of 40,055 SVs were detected in ‘QJM,’ in which insertions accounted for 60% (the highest amount), and deletions accounted for 43.76% (the lowest amount) of the identified SVs. The opposite trend was observed in ‘YLM’: it had the lowest number of insertions and the highest number of deletions, whereas the distributions of other types of SVs differed little, one to another.

### Functional annotation of databases

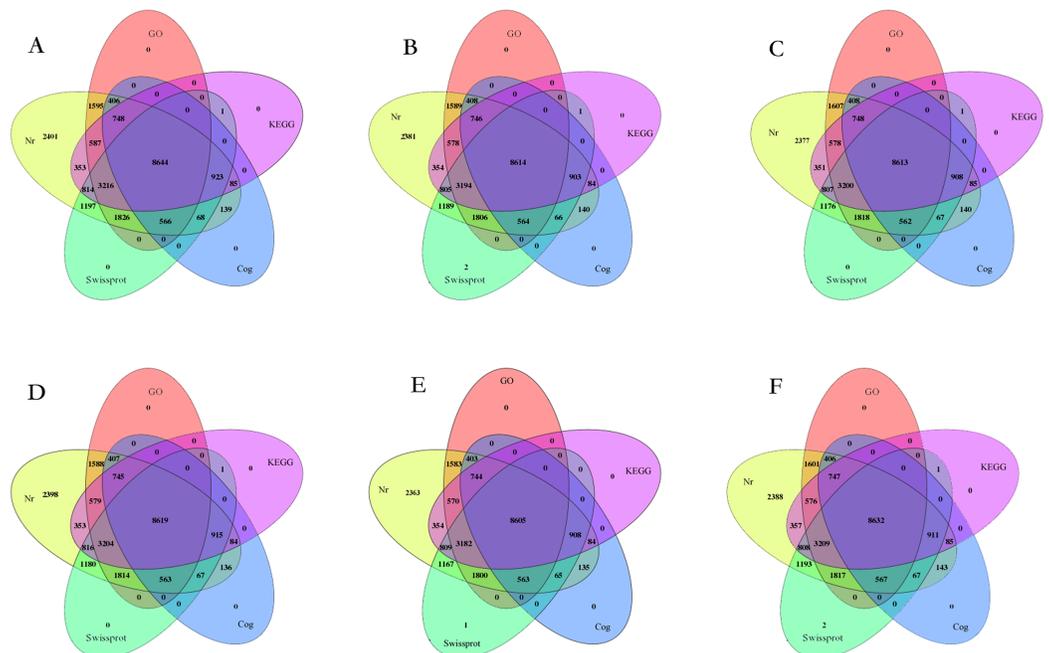
A change in function of a gene was usually because of the existence of non-synonymous (NS) mutation, frameshift (F) and SVs; in turn, expression of related protein was influenced. All tested and functionally explained genes were categorized into GO (Fig. 6). Approximately 4,736 genes per cultivar were separated into three groups: molecular function, cellular components and biological process. Among all categories, the highest number of genes (2,060 on average) were recorded in molecular function, while the lowest number of genes were observed in cellular components (1,657) and biological process categories (1,020), respectively. GO enrichment classification recommended that the genes from molecular function, cellular component and biological process groupings could be classified into three, 11 and 18 groups, respectively.

Genes explained by BLASTx, with a threshold of  $10^{-5}$ , were noted in six public databases comprising the GO, NCBI non-redundant (NR), nucleotide sequence (NT), Swiss-Prot, Clusters of Orthologous Groups of Proteins (COG) and KEGG databases. A total of 23,698 genes in ‘QJM,’ 23,554 genes in ‘SKM,’ 23,572 genes in ‘YLM,’ 23,597 genes in ‘RHM,’ 23,458 genes in ‘XZM’ and 23,640 genes in ‘ZHM’ were annotated successfully (Fig. 7). All the genes were noted in the NR database, and the distributions of genes annotated in six databases were not significantly different. At least 8,605 genes (‘XZM’)



**Figure 6** GO classification for differential unigenes between the six Japanese apricot cultivars. The results are the percentage of each GO category of genes in groups of cellular component, molecular function and biological process. The left  $y$ -axis indicates the percentage of genes in a category, and the right  $y$ -axis indicates the number of genes in a category. The  $x$ -axis indicates the three ontologies, respectively. A corrected ( $p \leq 0.05$ ) was chosen as the threshold value. The GO term ( $p \leq 0.05$ ) is defined as a differentially expressed gene-enriched GO term.

Full-size DOI: 10.7717/peerj.4625/fig-6



**Figure 7** Homology analysis of the six Japanese apricot cultivars relative to reference genome. (A) 'QJM,' (B) 'SKM,' (C) 'YLM,' (D) 'RHM,' (E) 'XZM,' (F) 'ZHM.' A total of six public databases included NR—Non-redundant sequence databases, SwissProt—Swiss-Prot protein database, GO—Gene Ontology, COG—Clusters of Orthologous Groups and KEGG—Kyoto Encyclopedia of Genes and Genomes.

Full-size DOI: 10.7717/peerj.4625/fig-7

**Table 5** Candidate genes participated in the biosynthesis of anthocyanin.

	SNP	InDel	SV
All positions	181331	3318	51427
All unigenes	19774	1937	13850
Compare positions	3143	60	337
Compare unigenes	1350	32	122

were annotated in all six databases, while ‘QJM’ had the most genes annotated in all databases. The other four cultivars showed 8,614 genes in ‘SKM,’ 8,613 genes in ‘YLM,’ 8,619 genes in ‘RHM’ and 8,632 genes in ‘ZHM.’

Overall, CDS is a region of DNA that is translated to form protein, and detection of sequence variation occurring in the CDS region becomes extremely important for gene function analysis. Through genome-wide comparative analysis of six re-sequence data sets in the CDS region, a total of 181,331 positions of SNPs, 3,318 positions of InDels and 51,427 SVs detected 19,774, 1,937 and 13,850 unigenes, respectively. The significant difference between green- and red-skinned cultivars is the colour pigmentation due to existence of anthocyanin, in order to obtain the same sequence variations occurring in all green-skinned cultivars, while red-skinned cultivars have either no variation or synonymous mutations only associated with the biosynthesis and accumulation of anthocyanin. After comparison found that a sum of 3,143 SNPs (1,350 unigenes), 60 InDels (32 unigenes) and 337 SVs (122 unigenes) differed between green- and red-skinned cultivars (Table 5), including 25 unigenes containing two sequence variations at same time, it was a simple task to find the important variations which leads to pigmentation changes: there are four pathways, namely phenylpropanoid biosynthesis (31 unigenes, ko00940), flavonoid biosynthesis (15 unigenes, ko00941), flavone and flavonol biosynthesis (seven unigenes, ko00944), isoflavonoid biosynthesis (two unigenes, ko00943), that were found to be related to biosynthesis of anthocyanin according to previous research results. TFs (26 unigenes), as a series of regulatory genes that control structural gene transcription, were detected, including genes encoding MYB (four unigenes), bHLH (two unigenes), WD40-repeat protein (one unigenes), NAC (one unigenes), etc. Twenty-one unigenes may be related to accumulation of anthocyanin.

### Genes validation by qRT-PCR at transcriptional level

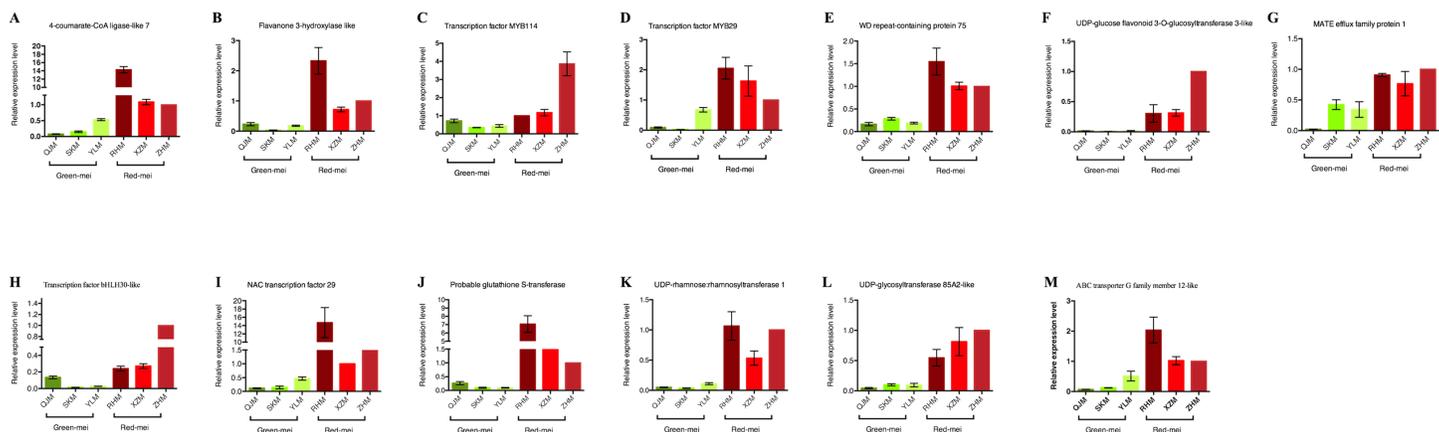
Fruit colour is determined by the accumulation and biosynthesis of anthocyanins, mediated by multiple enzymes in the phenylpropanoid and flavonoid pathways, regulatory genes and TFs. According to functional annotation results, 13 candidate homologous genes were revealed related to colour pigmentation, the translated proteins were detected as NS, InDels or SVs and 13 candidate genes participated in the biosynthesis of anthocyanin (Table 6). The expression patterns of genes encoding three anthocyanin synthesis structural genes (*4CL*, *F3H* and *UFGT*), five TFs (MBW complexes and NAC) and five related regulated genes (*GST1*, *RT1*, *UGT85A2*, *ABC* and *MATE* transports) were definite through quantitative reverse transcription-polymerase chain reaction (qRT-PCR), expression pattern of these candidate genes were not significantly different,

**Table 6** Candidate genes associated with the anthocyanin synthesis.

Functions	Description	Gene	Type red-skinned	Type green-skinned
Structural genes	4-Coumarate-CoA ligase-like 7	4-CL	S <sup>a</sup>	NS <sup>b</sup> /InDel <sup>c</sup> /SV <sup>d</sup>
	Flavanone 3-hydroxylase	F3H	S	NS
	UDP-glucose flavonoid 3-O-glucosyltransferase 3-like	UFGT	S	NS
Transcription factors	Transcription factor MYB114	MYB114	S	NS
	Transcription factor MYB29	MYB29	NS	S
	WD repeat-containing protein 75	WD40	S	S/NS
	Transcription factor bHLH30-like	bHLH30	NS	S
	NAC transcription factor 29	NAC	NS	S
	Related genes	Probable glutathione S-transferase	GST1	S
UDP-rhamnose: rhamnosyltransferase 1		RT1	NS	S
UDP-glycosyltransferase 85A2-like		UGT85A2	NS/SV	S
ABC transporter G family member 12-like		ABCG12	S	NS
MATE efflux family protein 1		MATE1	S	NS

**Notes:**

- <sup>a</sup> There are synonymous SNPs in genes.  
<sup>b</sup> There are non-synonymous SNPs in genes.  
<sup>c</sup> There are insert or deletion in genes.  
<sup>d</sup> There are structure variation in genes.

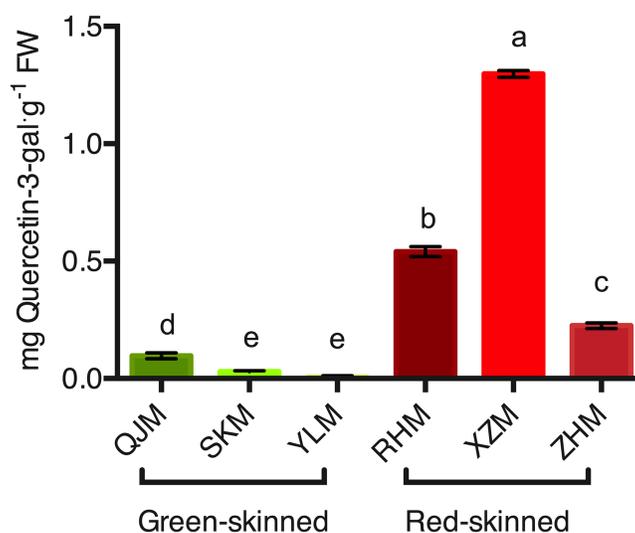
**Figure 8** Expression analysis of the anthocyanin biosynthetic genes in all of Japanese apricot cultivars using qRT-PCR. Relative levels of gene expression were calculated using RP II. The vertical bars represent the standard deviation (SD) of three biological replicates.

Full-size DOI: 10.7717/peerj.4625/fig-8

and transcription levels of genes in red-skinned cultivars were significantly higher than those of the genes in green-skinned cultivars (Fig. 8).

### UFGT enzyme analysis

UDP-glucose flavonoid-3-O-glycosyltransferase is assumed to be an important enzyme for anthocyanin biosynthesis in many plants. In this study, UFGT activity was identified in all cultivars. While the lowest activity was  $0.01 \text{ mg g}^{-1}$  in 'YLM' and  $0.03 \text{ mg/g}$  in 'SKM', compared with green-skinned cultivars, red-skinned cultivars showed significantly higher UFGT activity. 'XZM' cultivars had the highest activity ( $1.3 \text{ mg/g}$ ) of all cultivars, followed by 'RHM' ( $0.5 \text{ mg/g}$ ) and 'ZHM' ( $0.2 \text{ mg/g}$ ) (Fig. 9).



**Figure 9** Assay of UFGT activity between the six Japanese apricot cultivars. Values represent mean  $\pm$  standard deviation,  $n = 3$ . Different letters in rows indicate significant differences among mean values of treatments ( $p < 0.005$ ). [Full-size !\[\]\(2154c9f569edab9d15ee6b3ae0aa8398\_img.jpg\) DOI: 10.7717/peerj.4625/fig-9](https://doi.org/10.7717/peerj.4625/fig-9)

## DISCUSSION

In the present study, we mainly focused on understanding the colouring mechanism in Japanese apricot. Different experiments were conducted for this study. Firstly, we used UPLC/MS method to identify the main anthocyanin components, we identified a component ‘cyanidin-3-glucoside’ a major anthocyanin in Japanese apricot fruit, consistent with previous studies conducted in flowers (*Changling, Weiming & Junyu, 2004*). However, the molecular mechanisms of anthocyanin biosynthesis in Japanese apricot are not yet clear. After that, we performed genome re-sequencing to investigate the molecular mechanism of synthesis and accumulation of anthocyanin, comparing gene variations between red- and green-skinned cultivars and confirmed using quantitative qRT-PCR to find the expression pattern of those candidate genes, we found that three anthocyanin synthesis structural genes (*4CL*, *F3H* and *UFGT*), five TFs (MBW complexes and NAC) and five genes (*GST1*, *RT1*, *UGT85A2*, *ABC* and *MATE* transports) were possibly involved in the anthocyanin metabolic pathway of Japanese apricot. In addition, developing SNP makers in a wider range of Japanese apricot varieties by using re-sequencing data.

### Genes encoding enzymes in the anthocyanin biosynthetic pathway

Multiple enzymes participated in anthocyanin biosynthesis in phenylpropanoid and flavonoid pathways. In this study, three candidate enzymes were discussed (*Fig. 8*). *4CL* was the final enzyme of the general phenylpropanoid pathway, converting phenylalanine to 4-coumaroyl-CoA, a product that resulted in a range of flavonoid compounds, such as pro-anthocyanidins and anthocyanins (*Boss, Davies & Robinson, 1996b*). Ozeki (*Ozeki & Komamine, 1985*) studied the regulated relationship between 2,4-D and anthocyanin synthesis, and when no anthocyanin synthesis occurred, the activities of *4CL* increased one day after transfer due to transfer effect but consequently decreased and remained at

low levels. When the activities of 4CL increased, and reached up to maximum, anthocyanin was synthesized most rapidly. Christie (*Christie, Alfenito & Walbot, 1994*) found anthocyanin accumulation in all tested lines with 4CL-homologous transcripts increased at least threefold over levels in unstressed plants (no anthocyanin accumulation). These observations agree well with those qRT-PCR results mentioned above.

The primary phases of the flavonoid pathway were from 4-coumaroyl CoA through chalcone and naringenin to dihydroflavonol, in which F3H was one of three main enzymes (other two were CHS, CHI converted naringenin flavanone to dihydroflavonols). *F3H* gene expression seems to be fundamental in regulation at bifurcation of anthocyanin and flavonol branches, catalyses the stereospecific hydroxylation of (2S)-flavanones at 3-position of C-ring to (2R, 3R)-dihydroflavonols (*Sparvoli et al., 1994*). Zuker (*Zuker et al., 2002*), working on antisense suppression to block the expression of a flower colour gene encoding *F3H* of carnation (*Dianthus caryophyllus* L.), found that transgenic plants showed flower colour alterations ranging from attenuation to complete loss of their orange/reddish colour. *F3H* expression was explored to clarify the molecular mechanism of red colouration in apple; results indicated that the *F3H* gene was co-ordinately expressed during fruit development, and its levels of expression was positively related to the degree of anthocyanin concentration (*Honda et al., 2002*). Castellarin (*Castellarin et al., 2007*) performed dehydration to magnify the expression of *F3H* gene, which described well the increase of anthocyanin content in grapevines (*Vitis vinifera* L.) during experimental seasons. In radish (*Raphanus sativus*) varieties, the *F3H* gene was remarkably expressed through accumulation of sucrose in red radish; in contrast, the expression of *F3H* was powerfully suppressed in the white variety despite the accumulation of sucrose, showing that *F3H* activated the biosynthesis of anthocyanins via regulation of TFs (*Hara et al., 2004*).

UDP-glucose flavonoid-3-O-glycosyltransferase, an enzyme participated in the late step in anthocyanin biosynthesis, transfers the glucosyl moiety from UDP-glucose to the 3-hydroxyl group of acceptor molecules in a glucosyltransferase catalytic reaction. The role of UFGT in anthocyanidins were glucosylated by UFGT during red-skinned fruit ripening. Dissimilarities in colour strength make contribution to the expression differences of structural genes. High expression of UFGT has been detected in red skin of grapes, while in white grapes, almost extremely low expression was found (*Boss, Davies & Robinson, 1996b*). The lack of anthocyanins detected in white Malay apple fruits was due to the absence of detectable levels of UFGT transcripts (*Kotepong, Ketsa & van Doorn, 2011*). Takos (*Takos et al., 2006*) confirmed that the transcript levels of genes encoding UFGT in multiple anthocyanin pathways was much higher in red-skinned apples than in others. The transcript expression pattern of gene encoding UFGT reach the fastest accumulation speed at both primary and late developmental stage was confirmed in other fruits like peach (*Tsuda et al., 2004*), bilberry (*Jaakola et al., 2002*).

In the current study, transcript levels of 4CL, F3H and UFGT, consistent to the flavanol and anthocyanin levels, were greater in red-skinned fruits than in green-skinned cultivars (*Fig. 8*); up-regulated structural genes in red-skinned fruits indicated three genes contributed in the accumulation of anthocyanin that led to red pigmentation in plants.

## Regulation of TFs through anthocyanin accumulation

The anthocyanin biosynthetic pathway has been comprehensively studied in several species such as tobacco (*Aharoni et al., 2001*), *Arabidopsis* (*Tohge et al., 2005*), grapevine (*Deluc et al., 2006*), apple (*Espley et al., 2007*), tomato (*Butelli et al., 2008*), mangosteen (*Palapol et al., 2009*) and Chinese bayberry (*Niu et al., 2010*); the structural genes encoding pathway enzymes have been mentioned above, and other regulatory genes were TFs which could control structural gene transcription (Fig. 8).

The anthocyanin pathway genes are regulated by interaction of DNA binding R2R3-MYB TFs and MYC-like bHLH and WD40-repeat proteins (MBW complexes) in plants (*Chagné et al., 2007*). Two MYB genes, *MYB29* and *MYB114*, were found in this present study to be pertinent in regulation of anthocyanin synthesis (*Li et al., 2014; Schwinn et al., 2016; Sun et al., 2016*) (Fig. 8). Schwinn (*Schwinn et al., 2016*) reported that *MYB29* was associated with bulb colour (*Allium cepa* L., Alliioideae, Asparagales): as an R2R3-MYB TF that regulated the flavonoid pathway, it was demonstrated that *MYB29* belonged to sub-group that carried out biosynthesis of flavonoids. Overexpression of *MYB5* or *MYB114* strongly induces proanthocyanidin (PA) accumulation in *Medicago truncatula* hairy roots, and both *MYB5* and *MYB114*, mutants of *M. truncatula*, which exhibit darker seed coat colour than wild-type plants, were revealed to physically interact and synergistically activate the promoters of anthocyanidin reductase and leucoanthocyanidin reductase (*Liu, Jun & Dixon, 2014*). Gonzalez (*Gonzalez et al., 2008*) reported that overexpression of *MYB114* resulted in enhanced pigment production in a TTG1- and bHLH-dependent manner. Li (*Li et al., 2014*) found that JA promotion of anthocyanin accumulation under far-red light is dependent on the phytochrome A signalling pathway; at the same time, knockdown expression of *MYB114* significantly reduces methyl jasmonic acid promotion of anthocyanin accumulation.

As a regulatory complex comprising another candidate TF, bHLH30-like (bHLH30), and WD repeat-containing protein 75 (WD40), the expression arrangements of these two candidate genes, *bHLH30* and *WD40*, in six cultivars were the same as those of MYB family genes (*MYB5* and *MYB114*), which are upregulated throughout fruit ripening in red-skinned associated with green-skinned cultivars. Lc (leaf color) as the first plant bHLH protein involved in maize tissue-specific anthocyanin pigmentation by mutant analyses (*Dias et al., 2003*). Zhou (*Zhou, Shi & Xie, 2012*) reported that nitrogen regulated the synthesis of anthocyanin in red cells operates through the mechanism by which the expression levels of genes encoding equally major constituents of *TTG1-GL3/TT8-PAP1* complex and negative regulators was effected by nitrogen, which was a complex of bHLH and WD40; this may be same as in our present study. All three regulatory complex genes were highly upregulated in red-skinned relative to green-skinned cultivars.

Like MBW regulatory complex, the TF NAC was also found and related to red colour formation. The NAC family are plant-particular TFs and have 106 and 149 members pretend in *Arabidopsis* and rice, correspondingly (*Gong et al., 2004; Xiong et al., 2005*). Morishita (*Morishita et al., 2009*) found that NAC078 protein regulated the expression of genes related to the biosynthesis of flavonoids and the level of anthocyanins were

expressively increased in NAC078-containing plants and reduced in NAC078-knockout plants of *Arabidopsis*. In orange fruit, a gene encoding TF NAC domain protein was induced the biosynthesis of anthocyanins in blood oranges but not in common oranges. The same phenomena occurred in raspberry in a previous study: a candidate gene encoding NAC (*CUC2-like*) was accountable for the accumulation of cyanidin and pelargonidin pigments (*Kassim et al., 2009*). Our study showed that NAC TF 29 was highly upregulated in red-skinned relative to green-skinned cultivars, which suggested that it is a potential TF contributing to the regulation of red colour formation in Japanese apricot.

Among these candidate TFs, the functions of only two MYB TFs (MYB29 and MYB114) have been studied in *Arabidopsis* and other plants, whereas reports on TF function are rare in Japanese apricot (*Fig. 8*). The other three candidate genes have not been stated to be intricate in the biosynthesis of anthocyanins. Advance studies are desired to regulate whether changes in the transcription of these candidate genes are correlated to regulation of anthocyanin metabolism. Moreover, in Japanese apricot, the relationship between regulatory complex (MBW) and anthocyanin biosynthesis leftovers is unclear. Subsequently, this matter should also be explored.

### Related genes contributed to accumulation of anthocyanin

There were several candidate genes participating in biosynthesis or the accumulation of anthocyanin. However, these genes were not the same as structural genes, which could directly impact the anthocyanin synthesis pathway, and neither were they like TFs, which could control structural gene transcription; they may be related to regulation of the anthocyanin biosynthesis pathway.

In this study, seven candidate genes were identified that might be associated to the biosynthesis of anthocyanin (*Fig. 8*). Glutathione S-transferase (*GST1*) probably have the same function as *Bz2* gene in maize (*Marrs et al., 1995*), which encodes a protein with GST activity conjugated with glutathione and the inhibitor vanadate in plant tonoplast and has been shown to inhibit the accumulation of anthocyanins in vacuole. Alfenito (*Alfenito et al., 1998*) revealed that *An9*, which encoded a kind of glutathione S-transferase which transported glutathionated cyanidin 3-glucoside (C3G) to the vacuole, regulated transcriptional activator of the anthocyanin pathway. In *Vitis vinifera* L., two other GST genes (*VvGST1* and *VvGST4*) showed different induction patterns, but all their transcriptional profiling showed that they induced intensively different accumulation of anthocyanin from that of *Bz2* and *An9* in maize and petunia, correspondingly (*Conn et al., 2008*). TT19 was demonstrated in *Arabidopsis* that works as a carrier for uptake of anthocyanins or proanthocyanidin precursors into the vacuole and may protect cyanidin from degradation during transport process (*Kitamura, Shikazono & Tanaka, 2004*). GST in strawberry was also mentioned as one of the list of differentially expressed gene related to biosynthesis anthocyanin (*Gu et al., 2015*).

In *petunia hybrida*, expression of UDP-rhamnose/rhamnosyltransferase (Rt) gene was induced by intense light, and transcription of UDP-rhamnose: anthocyanidin-3-glucoside rhamnosyltransferase (3RT) played a key role in promoting the accumulation of anthocyanin in epidermal cells of the petal (*Morita et al., 2005*). Regarding secondary

metabolism in Norton grape (*Ali et al., 2011*), the transcriptional profile of UDP-rhamnose/rhamnosyltransferase was proposed to be associated with the anthocyanin pathway. The RT gene downregulated the activity of *PAL* when it was low; however, *PAL* was a structural gene in the anthocyanin pathway, indicating that RT gene influenced the biosynthesis of anthocyanin (*Given, Venis & Grierson, 1988*).

The transcriptional expression of UDP-glycosyltransferase 85A2-like (UGT85A2) was highly upregulated in red-skinned comparative to green-skinned cultivars, indicating that it may be associated with the biosynthesis of anthocyanin. In *Crocus* species, transcription of UDP-glycosyltransferase 703B1 was found only in stigmas and petals where anthocyanin accumulates (*Nørbæk et al., 2002; Moraga et al., 2009*). A relationship between *Arabidopsis thaliana* mutant UGT72B1 and decreased accumulation of anthocyanin in floral stems was confirmed, and an association between UGT86A1 with the anthocyanin biosynthesis pathway in strawberry was also mentioned (*Gu et al., 2015*). The ABC and MATE transporter families are two important groups of proteins whose members participate in an extensive variety of transport processes. The upregulation of ABC transporters demonstrated that ABC transporters take part in transport of anthocyanin into vacuoles of *Arabidopsis* (*Lu, Li & Rea, 1997*). In yeast, two multidrug resistance-associated protein-type ABC transporters were found to facilitate vacuolar uptake of anthocyanin–glutathione conjugates (*Rea, 1999*), and same results were reported in maize (*Debeaujon et al., 2001*) and tea (*Li, 2016*). The MATE transporters TT12 of *Arabidopsis* work as a precursors that active in cells synthesizing proanthocyanidins, thus inducing vacuolar accumulation of proanthocyanidins in the seed (*Gomez et al., 2009*).

All related genes intricate in regulation of the biosynthesis of anthocyanin were studied in this experiment and showed same expression pattern as the structural genes and TFs, also representing that these genes may be associated to anthocyanin accumulation.

### Gene expression encoding UFGT enzyme

As the gene encoding enzyme accountable for preceding step in biosynthesis of anthocyanin, the *UFGT* gene played an important role in preminent substances of anthocyanins and flavonols in red pericarps and their effective absence in green pericarps. In this study, gene expression of *UFGT* was significantly upregulated in red pericarps, while almost no one was detected in green pericarps (*Fig. 9*). At the same time, the enzyme activity was detected among all Japanese apricot cultivars, whether in coloured skin where anthocyanin was accumulated or in green pericarps where no colour pigmentation was present, but the activity of *UFGT* was expressively greater in red pericarps than in green pericarps (*Fig. 9*). Studies of other plant species have shown that the expression of *UFGT* gene was acute for anthocyanin biosynthesis. For example, the expression of *StUFGT*, which induced corresponding by gibberellic acid and sucrose, was detected in roots, leaves and stems, suggested that *UFGT* was related to the accumulation of anthocyanin in potato (*Hu et al., 2011*). In grape berry, *UFGT* gene expression was only identified in red-skinned spots, and in Kyoho it was detected by northern blot analysis, indicating that the *UFGT* gene is a main regulator of anthocyanin biosynthesis (*Kobayashi et al., 2001*). Unlike other

genes of the anthocyanin pathway, expression of UFGT was expected to mirror the anthocyanin content, suggesting that fruit colouration was strongly influenced by UFGT expression in litchi (*Zhao et al., 2012*). In conclusion, the *UFGT* gene was the most important gene correlated with the biosynthesis and accumulation of anthocyanins, contributing to red pigmentation of Japanese apricot.

## CONCLUSION

Empiric genome re-sequencing on an Illumina platform was a realistic means to scientifically investigate the universal genomic variations associated with the accumulation of anthocyanins in Japanese apricot fruits. We investigated that the genes participating in developments of anthocyanin biosynthesis, TFs and related genes were only upregulated in red-skinned cultivars, confirming that these genes may have the function of regulating anthocyanin biosynthesis. Generally, colourful fruits such as those of the ornamental Japanese apricot are very attractive to consumers. Considering the molecular mechanisms underlying their formation is imperative, predominantly for understanding the transcriptional regulatory networks of dissimilar coloured fruits and alterations in accumulation and distribution of different compounds in coloured cultivars. This data can provide valuable information and will subsidize to the selection of plant germplasm resources for marker-assisted selection as well as for developing the nutritional and medicinal importance of Japanese apricot in the future.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the Six Talent Fund of Jiangsu Province (No. NY068) and the project of crop germplasm collection and evaluation (2016NWB032). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Six Talent Fund of Jiangsu Province: NY068.

Project of crop germplasm collection and evaluation: 2016NWB032.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Xiaopeng Ni conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools, prepared figures and/or tables.
- Song Xue analyzed the data.
- Shahid Iqbal authored or reviewed drafts of the paper.
- Wanxu Wang performed the experiments.
- Zhaojun Ni analyzed the data.

- Muhammad Khalil-ur-Rehman authored or reviewed drafts of the paper.
- Zhihong Gao conceived and designed the experiments, approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The raw sequencing data in this paper are uploaded to the GenBank SRA database. The BioProject is [PRJNA371370](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA371370). The data for the ‘Xiaoyezhugan’ cultivar have been deposited at SRA under accession [SRR5241555](https://www.ncbi.nlm.nih.gov/sra/SRR5241555). The data for ‘Zhonghong’ cultivar have been deposited at SRA under accession [SRR5241554](https://www.ncbi.nlm.nih.gov/sra/SRR5241554). The data for the ‘Shinanokoume’ cultivar have been deposited at SRA under accession [SRR5241553](https://www.ncbi.nlm.nih.gov/sra/SRR5241553). The data for the ‘Yanglao’ cultivar have been deposited at SRA under accession [SRR5241552](https://www.ncbi.nlm.nih.gov/sra/SRR5241552). The data for the ‘Ruantiaohongmei’ cultivar have been deposited at SRA under accession [SRR5241550](https://www.ncbi.nlm.nih.gov/sra/SRR5241550). The data for the ‘QingjiaNo.2’ cultivar have been deposited at SRA under accession [SRR5241549](https://www.ncbi.nlm.nih.gov/sra/SRR5241549).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.4625#supplemental-information>.

## REFERENCES

- Aharoni A, De Vos CH, Wein M, Sun Z, Greco R, Kroon A, Mol JN, O’Connell AP. 2001. The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant Journal* 28(3):319–332 DOI [10.1046/j.1365-313x.2001.01154.x](https://doi.org/10.1046/j.1365-313x.2001.01154.x).
- Alfenito MR, Souer E, Goodman CD, Buell R, Mol J, Koes R, Walbot V. 1998. Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases. *Plant Cell* 10(7):1135–1149 DOI [10.1105/tpc.10.7.1135](https://doi.org/10.1105/tpc.10.7.1135).
- Ali MB, Howard S, Chen S, Wang Y, Yu O, Kovacs LG, Qiu W. 2011. Berry skin development in Norton grape: distinct patterns of transcriptional regulation and flavonoid biosynthesis. *BMC Plant Biology* 11(1):7 DOI [10.1186/1471-2229-11-7](https://doi.org/10.1186/1471-2229-11-7).
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* 25:25–29 DOI [10.1038/75556](https://doi.org/10.1038/75556).
- Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* 28(1):45–48 DOI [10.1093/nar/24.1.21](https://doi.org/10.1093/nar/24.1.21).
- Batra P, Sharma AK. 2013. Anti-cancer potential of flavonoids: recent trends and future perspectives. *3 Biotech* 3(6):439–459 DOI [10.1007/s13205-013-0117-5](https://doi.org/10.1007/s13205-013-0117-5).
- Boss PK, Davies C, Robinson SP. 1996a. Analysis of the expression of anthocyanin pathway genes in developing *Vitis vinifera* L. cv Shiraz Grape Berries and the implications for pathway regulation. *Plant Physiology* 111(4):1059–1066 DOI [10.1104/pp.111.4.1059](https://doi.org/10.1104/pp.111.4.1059).
- Boss PK, Davies C, Robinson SP. 1996b. Expression of anthocyanin biosynthesis pathway genes in red and white grapes. *Plant Molecular Biology* 32(3):565–569 DOI [10.1007/bf00019111](https://doi.org/10.1007/bf00019111).
- Butelli E, Titta L, Giorgio M, Mock H-P, Matros A, Peterek S, Schijlen EGWM, Hall RD, Bovy AG, Luo J, Martin C. 2008. Enrichment of tomato fruit with health-promoting

anthocyanins by expression of select transcription factors. *Nature Biotechnology* **26**(11):1301–1308 DOI [10.1038/nbt.1506](https://doi.org/10.1038/nbt.1506).

- Castellarin SD, Pfeiefer A, Sivilottip P, Degan M, Peterlunger E, Digaspero G. 2007.** Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant, Cell & Environment* **30**(11):1381–1399 DOI [10.1111/j.1365-3040.2007.01716.x](https://doi.org/10.1111/j.1365-3040.2007.01716.x).
- Chagné D, Carlisle CM, Blond C, Volz RK, Whitworth CJ, Oraguzie NC, Crowhurst RN, Allan AC, Espley RV, Hellens RP, Gardiner SE. 2007.** Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple. *BMC Genomics* **8**(1):212 DOI [10.1186/1471-2164-8-212](https://doi.org/10.1186/1471-2164-8-212).
- Changling Z, Weiming G, Junyu C. 2004.** Preliminary study on the categories and contents of the flower color pigments of *Prunus mume* sieb. et Zucc. *Journal of Beijing Forestry University* **26**:68–73.
- Chen J. 1989.** *Chinese Mei flower cultivars*. Beijing: China Forestry Publishing House.
- Christie PJ, Alfenito MR, Walbot V. 1994.** Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta* **194**(4):541–549 DOI [10.1007/bf00714468](https://doi.org/10.1007/bf00714468).
- Chu MY. 1999.** *China Fruit Records-Mei*. Beijing: China Forestry.
- Conn S, Curtin C, Bézier A, Franco C, Zhang W. 2008.** Purification, molecular cloning, and characterization of glutathione S-transferases (GSTs) from pigmented *Vitis vinifera* L. cell suspension cultures as putative anthocyanin transport proteins. *Journal of Experimental Botany* **59**(13):3621–3634 DOI [10.1093/jxb/ern217](https://doi.org/10.1093/jxb/ern217).
- Debeaujon I, Peeters AJ, Léon-Kloosterziel KM, Koornneef M. 2001.** The TRANSPARENT TESTA12 gene of Arabidopsis encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* **13**(4):853–871 DOI [10.1105/tpc.13.4.853](https://doi.org/10.1105/tpc.13.4.853).
- Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde J-P, Mérillon J-M, Hamdi S. 2006.** Characterization of a grapevine R2R3-MYB transcription factor that regulates the phenylpropanoid pathway. *Plant Physiology* **140**(2):499–511 DOI [10.1104/pp.105.067231](https://doi.org/10.1104/pp.105.067231).
- Dias AP, Braun EL, McMullen MD, Grotewold E. 2003.** Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiology* **131**(2):610–620 DOI [10.1104/pp.012047](https://doi.org/10.1104/pp.012047).
- Dixon RA, Steele CL. 1999.** Flavonoids and isoflavonoids—a gold mine for metabolic engineering. *Trends in Plant Science* **4**(10):394–400 DOI [10.1016/S1360-1385\(99\)01471-5](https://doi.org/10.1016/S1360-1385(99)01471-5).
- Escribano-Bailón MT, Santos-Buelga C, Rivas-Gonzalo JC. 2004.** Anthocyanins in cereals. *Journal of Chromatography A* **1054**(1–2):129–141 DOI [10.1016/j.chroma.2004.08.152](https://doi.org/10.1016/j.chroma.2004.08.152).
- Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty Amma S, Allan AC. 2007.** Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant Journal* **49**(3):414–427 DOI [10.1111/j.1365-313X.2006.02964.x](https://doi.org/10.1111/j.1365-313X.2006.02964.x).
- Gao Z, Han Z, Zhang Z. 2003.** Core collection and molecular marker of Japanese apricot (*Prunus mume* Sieb. et Zucc.) [D]. China Agricultural University Doctorate Degree dissertation.
- Given NK, Venis MA, Grierson D. 1988.** Phenylalanine ammonia-lyase activity and anthocyanin synthesis in ripening strawberry fruit. *Journal of Plant Physiology* **133**(1):25–30 DOI [10.1016/S0176-1617\(88\)80079-8](https://doi.org/10.1016/S0176-1617(88)80079-8).
- Gomez C, Terrier N, Torregrosa L, Vialet S, Fournier-Level A, Verries C, Souquet J-M, Mazauric J-P, Klein M, Cheynier V. 2009.** Grapevine MATE-type proteins act as vacuolar H<sup>+</sup>-dependent acylated anthocyanin transporters. *Plant Physiology* **150**(1):402–415 DOI [10.1104/pp.109.135624](https://doi.org/10.1104/pp.109.135624).

- Gong W, Shen Y-P, Ma L-G, Pan Y, Du Y-L, Wang D-H, Yang J-Y, Hu L-D, Liu X-F, Dong C-X, Ma L, Chen Y-H, Yang X-Y, Gao Y, Zhu D, Tan X, Mu J-Y, Zhang D-B, Liu Y-L, Dinesh-Kumar SP, Li Y, Wang X-P, Gu H-Y, Qu L-J, Bai S-N, Lu Y-T, Li J-Y, Zhao J-D, Zuo J, Huang H, Deng XW, Zhu Y-X. 2004. Genome-wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. *Plant Physiology* 135(2):773–782 DOI 10.1104/pp.104.042176.
- 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(5):814–827 DOI 10.1111/j.1365-3113X.2007.03373.x.
- Gu X, Chen Y, Gao Z, Qiao Y, Wang X. 2015. Transcription factors and anthocyanin genes related to low-temperature tolerance in rd29A:Rdreb1BI transgenic strawberry. *Plant Physiology and Biochemistry* 89:31–43 DOI 10.1016/j.plaphy.2015.02.004.
- Hara M, Oki K, Hoshino K, Kuboi T. 2004. Effects of sucrose on anthocyanin production in hypocotyl of two radish (*Raphanus sativus*) varieties. *Plant Biotechnology* 21(5):401–405 DOI 10.5511/plantbiotechnology.21.401.
- He J, Giusti MM. 2010. Anthocyanins: natural colorants with health-promoting properties. *Annual Review of Food Science and Technology* 1(1):163–187 DOI 10.1146/annurev.food.080708.100754.
- Holm L, Sander C. 1998. Removing near-neighbour redundancy from large protein sequence collections. *Bioinformatics* 14(5):423–429 DOI 10.1093/bioinformatics/14.5.423.
- Honda C, Kotoda N, Wada M, Kondo S, Kobayashi S, Soejima J, Zhang Z, Tsuda T, Moriguchi T. 2002. Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiology and Biochemistry* 40(11):955–962 DOI 10.1016/s0981-9428(02)01454-7.
- Hosseinian FS, Li W, Beta T. 2008. Measurement of anthocyanins and other phytochemicals in purple wheat. *Food Chemistry* 109(4):916–924 DOI 10.1016/j.foodchem.2007.12.083.
- Hu C, Gong Y, Jin S, Zhu Q. 2011. Molecular analysis of a UDP-glucose: flavonoid 3-O-glucosyltransferase (UFGT) gene from purple potato (*Solanum tuberosum*). *Molecular Biology Reports* 38(1):561–567 DOI 10.1007/s11033-010-0141-z.
- Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, Reiser L, LaFond F, Hanley D, Kiphart D, Zhuang M, Huang W, Mueller LA, Bhattacharyya D, Bhaya D, Sobral BW, Beavis W, Meinke DW, Town CD, Somerville C, Rhee SY. 2001. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Research* 29(1):102–105 DOI 10.1093/nar/29.1.102.
- Jaakola L. 2013. New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends in Plant Science* 18(9):477–483 DOI 10.1016/j.tplants.2013.06.003.
- Jaakola L, Määttä K, Pirttilä AM, Törrönen R, Kärenlampi S, Hohtola A. 2002. Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiology* 130(2):729–739 DOI 10.1104/pp.006957.
- Jimenez-Garcia SN, Guevara-Gonzalez RG, Miranda-Lopez R, Feregrino-Perez AA, Torres-Pacheco I, Vazquez-Cruz MA. 2013. Functional properties and quality characteristics of bioactive compounds in berries: biochemistry, biotechnology, and genomics. *Food Research International* 54(1):1195–1207 DOI 10.1016/j.foodres.2012.11.004.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Research* 32(90001):D277–D280 DOI 10.1093/nar/gkh063.

- Kassim A, Poette J, Paterson A, Zait D, McCallum S, Woodhead M, Smith K, Hackett C, Graham J. 2009.** Environmental and seasonal influences on red raspberry anthocyanin antioxidant contents and identification of quantitative traits loci (QTL). *Molecular Nutrition & Food Research* **53**(5):625–634 DOI [10.1002/mnfr.200800174](https://doi.org/10.1002/mnfr.200800174).
- Kim S-H, Lee J-R, Hong S-T, Yoo Y-K, An G, Kim S-R. 2003.** Molecular cloning and analysis of anthocyanin biosynthesis genes preferentially expressed in apple skin. *Plant Science* **165**(2):403–413 DOI [10.1016/S0168-9452\(03\)00201-2](https://doi.org/10.1016/S0168-9452(03)00201-2).
- Kitamura S, Shikazono N, Tanaka A. 2004.** TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant Journal* **37**(1):104–114 DOI [10.1046/j.1365-313X.2003.01943.x](https://doi.org/10.1046/j.1365-313X.2003.01943.x).
- Kobayashi S, Goto-Yamamoto N, Hirochika H. 2004.** Retrotransposon-induced mutations in grape skin color. *Science* **304**(5673):982 DOI [10.1126/science.1095011](https://doi.org/10.1126/science.1095011).
- Kobayashi S, Ishimaru M, Ding CK, Yakushiji H, Goto N. 2001.** Comparison of UDP-glucose: flavonoid 3-O-glucosyltransferase (UFGT) gene sequences between white grapes (*Vitis vinifera*) and their sports with red skin. *Plant Science* **160**:543–550 DOI [10.1016/S0168-9452\(00\)00425-8](https://doi.org/10.1016/S0168-9452(00)00425-8).
- Koes R, Verweij W, Quattrocchio F. 2005.** Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in Plant Science* **10**(5):236–242 DOI [10.1016/j.tplants.2005.03.002](https://doi.org/10.1016/j.tplants.2005.03.002).
- Kotepong P, Ketsa S, van Doorn WG. 2011.** A white mutant of Malay apple fruit (*Syzygium malaccense*) lacks transcript expression and activity for the last enzyme of anthocyanin synthesis, and the normal expression of a MYB transcription factor. *Functional Plant Biology* **38**(1):75–86 DOI [10.1071/fp10164](https://doi.org/10.1071/fp10164).
- Lancaster JE, Grant JE, Lister CE, Taylor MC. 1994.** Skin color in apples—influence of copigmentation and plastid pigments on shade and darkness of red color in five genotypes. *Journal of the American Society for Horticultural Science* **119**:63–69.
- Li J. 2016.** Transcriptome analysis reveals the accumulation mechanism of anthocyanins in “Zijuan” tea (*Camellia sinensis* var. *assamica* (Masters) Kitamura) leaves. *Plant Growth Regulation* **81**(1):51–61 DOI [10.1007/s10725-016-0183-x](https://doi.org/10.1007/s10725-016-0183-x).
- Li T, Jia K-P, Lian H-L, Yang X, Li L, Yang H-Q. 2014.** Jasmonic acid enhancement of anthocyanin accumulation is dependent on phytochrome A signaling pathway under far-red light in Arabidopsis. *Biochemical and Biophysical Research Communications* **454**(1):78–83 DOI [10.1016/j.bbrc.2014.10.059](https://doi.org/10.1016/j.bbrc.2014.10.059).
- Liu C, Jun JH, Dixon RA. 2014.** MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiology* **165**(4):1424–1439 DOI [10.1104/pp.114.241877](https://doi.org/10.1104/pp.114.241877).
- Lu YP, Li ZS, Rea PA. 1997.** AtMRP1 gene of Arabidopsis encodes a glutathione S-conjugate pump: isolation and functional definition of a plant ATP-binding cassette transporter gene. *Proceedings of the National Academy of Sciences of the United States of America* **94**(15):8243–8248 DOI [10.1073/pnas.94.15.8243](https://doi.org/10.1073/pnas.94.15.8243).
- Marrs KA, Alfenito MR, Lloyd AM, Walbot V. 1995.** A glutathione S-transferase involved in vacuolar transfer encoded by the maize gene Bronze-2. *Nature* **375**(6530):397–400 DOI [10.1038/375397a0](https://doi.org/10.1038/375397a0).
- Mazza G, Cacace JE, Kay CD. 2004.** Methods of analysis for anthocyanins in plants and biological fluids. *Journal of AOAC International* **87**:129–145.
- Mol J, Grotewold E, Koes R. 1998.** How genes paint flowers and seeds. *Trends in Plant Science* **3**(6):212–217 DOI [10.1016/S1360-1385\(98\)01242-4](https://doi.org/10.1016/S1360-1385(98)01242-4).

- Moraga AR, Mozos AT, Ahrazem O, Gómez-Gómez L. 2009. Cloning and characterization of a glucosyltransferase from *Crocus sativus* stigmas involved in flavonoid glucosylation. *BMC Plant Biology* 9(1):109 DOI 10.1186/1471-2229-9-109.
- Morishita T, Kojima Y, Maruta T, Nishizawa-Yokoi A, Yabuta Y, Shigeoka S. 2009. Arabidopsis NAC transcription factor, ANAC078, regulates flavonoid biosynthesis under high-light. *Plant and Cell Physiology* 50(12):2210–2222 DOI 10.1093/pcp/pcp159.
- Morita Y, Hoshino A, Kikuchi Y, Okuhara H, Ono E, Tanaka Y, Fukui Y, Saito N, Nitasaka E, Noguchi H, Iida S. 2005. Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose: anthocyanidin 3-O-glucoside-2"-O-glucosyltransferase, due to 4-bp insertions in the gene. *Plant Journal* 42(3):353–363 DOI 10.1111/j.1365-313X.2005.02383.x.
- Niu S-S, Xu C-J, Zhang W-S, Zhang B, Li X, Lin-Wang K, Ferguson IB, Allan AC, Chen K-S. 2010. Coordinated regulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*) fruit by a R2R3 MYB transcription factor. *Planta* 231(4):887–899 DOI 10.1007/s00425-009-1095-z.
- Nørbæk R, Brandt K, Nielsen JK, Ørgaard M, Jacobsen N. 2002. Flower pigment composition of *Crocus* species and cultivars used for a chemotaxonomic investigation. *Biochemical Systematics and Ecology* 30(8):763–791 DOI 10.1016/S0305-1978(02)00020-0.
- Oms-Oliu G, Soliva-Fortuny R, Martín-Belloso O. 2008. Using polysaccharide-based edible coatings to enhance quality and antioxidant properties of fresh-cut melon. *LWT—Food Science and Technology* 41(10):1862–1870 DOI 10.1016/j.lwt.2008.01.007.
- Ozeki Y, Komamine A. 1985. Changes in activities of enzymes involved in general phenylpropanoid metabolism during the induction and reduction of anthocyanin synthesis in a carrot suspension culture as regulated by 2,4-D. *Plant and Cell Physiology* 26(5):903–911 DOI 10.1093/oxfordjournals.pcp.a076985.
- Palapol Y, Ketsa S, Lin-Wang K, Ferguson IB, Allan AC. 2009. A MYB transcription factor regulates anthocyanin biosynthesis in mangosteen (*Garcinia mangostana* L.) fruit during ripening. *Planta* 229(6):1323–1334 DOI 10.1007/s00425-009-0917-3.
- Pomar F, Novo M, Masa A. 2005. Varietal differences among the anthocyanin profiles of 50 red table grape cultivars studied by high performance liquid chromatography. *Journal of Chromatography A* 1094(1–2):34–41 DOI 10.1016/j.chroma.2005.07.096.
- Rahim MA, Busatto N, Trainotti L. 2014. Regulation of anthocyanin biosynthesis in peach fruits. *Planta* 240(5):913–929 DOI 10.1007/s00425-014-2078-2.
- Ramakers C, Ruijter JM, Deprez RHL, Moorman AF. 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* 339(1):62–66 DOI 10.1016/s0304-3940(02)01423-4.
- Rao AV, Rao LG. 2007. Carotenoids and human health. *Pharmacological Research* 55:207–216 DOI 10.1016/j.phrs.2007.01.012.
- Rea PA. 1999. MRP subfamily ABC transporters from plants and yeast. *Journal of Experimental Botany* 50(90001):895–913 DOI 10.1093/jxb/50.Special\_Issue.895.
- Salvatierra A, Pimentel P, Moya-Leon MA, Caligari PDS, Herrera R. 2010. Comparison of transcriptional profiles of flavonoid genes and anthocyanin contents during fruit development of two botanical forms of *Fragaria chiloensis* ssp. *chiloensis*. *Phytochemistry* 71(16):1839–1847 DOI 10.1016/j.phytochem.2010.08.005.
- Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AMML, de Vos RCH, Jonker HH, Xu W, Routaboul J-M, Lepiniec L, Bovy AG. 2013. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin

- biosynthesis in strawberry (*Fragaria* × *ananassa*) fruits. *New Phytologist* **197**(2):454–467 DOI [10.1111/nph.12017](https://doi.org/10.1111/nph.12017).
- Schwinn K, Ngo H, Kenel F, Brummell D, Albert N, McCallum J, Pither-Joyce M, Crowhurst R, Eady C, Davies K. 2016. The onion (*Allium cepa* L.) R2R3-MYB gene MYB1 regulates anthocyanin biosynthesis. *Frontiers in Plant Science* **7**:1865 DOI [10.3389/fpls.2016.01865](https://doi.org/10.3389/fpls.2016.01865).
- Singh M, Arseneault M, Sanderson T, Murthy V, Ramassamy C. 2008. Challenges for research on polyphenols from foods in Alzheimer's disease: bioavailability, metabolism, and cellular and molecular mechanisms. *Journal of Agricultural and Food Chemistry* **56**(13):4855–4873 DOI [10.1021/jf0735073](https://doi.org/10.1021/jf0735073).
- Smeds L, Mugal CF, Qvarnström A, Ellegren H. 2016. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an Avian pedigree. *PLOS Genetics* **12**(5):e1006044 DOI [10.1371/journal.pgen.1006044](https://doi.org/10.1371/journal.pgen.1006044).
- Sparvoli F, Martin C, Scienza A, Gavazzi G, Tonelli C. 1994. Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (*Vitis vinifera* L.). *Plant Molecular Biology* **24**(5):743–755 DOI [10.1007/bf00029856](https://doi.org/10.1007/bf00029856).
- Steyn WJ, Holcroft DM, Wand SJE, Jacobs G. 2004. Regulation of pear color development in relation to activity of flavonoid enzymes. *Journal of the American Society for Horticultural Science* **129**:6–12.
- Sun S-S, Gugger PF, Wang Q-F, Chen J-M. 2016. Identification of a R2R3-MYB gene regulating anthocyanin biosynthesis and relationships between its variation and flower color difference in lotus (*Nelumbo* Adans.). *PeerJ* **4**:e2369 DOI [10.7717/peerj.2369](https://doi.org/10.7717/peerj.2369).
- Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR. 2006. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiology* **142**(3):1216–1232 DOI [10.1104/pp.106.088104](https://doi.org/10.1104/pp.106.088104).
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* **278**(5338):631–637 DOI [10.1126/science.278.5338.631](https://doi.org/10.1126/science.278.5338.631).
- Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima JI, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M. 2005. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *Plant Journal* **42**(2):218–235 DOI [10.1111/j.1365-313x.2005.02371.x](https://doi.org/10.1111/j.1365-313x.2005.02371.x).
- Tong Z, Gao Z, Wang F, Zhou J, Zhang Z. 2009. Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Molecular Biology* **10**(1):71 DOI [10.1186/1471-2199-10-71](https://doi.org/10.1186/1471-2199-10-71).
- Tsuda T, Yamaguchi M, Honda C, Moriguchi T. 2004. Expression of anthocyanin biosynthesis genes in the skin of peach and nectarine fruit. *Journal of the American Society for Horticultural Science* **129**:857–862.
- Usenik V, Štampar F, Veberič R. 2009. Anthocyanins and fruit colour in plums (*Prunus domestica* L.) during ripening. *Food Chemistry* **114**(2):529–534 DOI [10.1016/j.foodchem.2008.09.083](https://doi.org/10.1016/j.foodchem.2008.09.083).
- Wang Y-R, Lu Y-F, Hao S-X, Zhang M-L, Zhang J, Tian J, Yao Y-C. 2015. Different coloration patterns between the red- and white-fleshed fruits of malus crabapples. *Scientia Horticulturae* **194**:26–33 DOI [10.1016/j.scienta.2015.07.041](https://doi.org/10.1016/j.scienta.2015.07.041).
- Wang L-S, Stoner GD. 2008. Anthocyanins and their role in cancer prevention. *Cancer Letters* **269**(2):281–290 DOI [10.1016/j.canlet.2008.05.020](https://doi.org/10.1016/j.canlet.2008.05.020).
- Wei H, Chen X, Zong X, Shu H, Gao D, Liu Q. 2015. Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). *PLOS ONE* **10**(3):e0121164–20 DOI [10.1371/journal.pone.0121164](https://doi.org/10.1371/journal.pone.0121164).

- Welch CR, Wu Q, Simon JE. 2008.** Recent advances in anthocyanin analysis and characterization. *Current Analytical Chemistry* **4**(2):75–101 DOI [10.2174/157341108784587795](https://doi.org/10.2174/157341108784587795).
- Wu X, Gong Q, Ni X, Zhou Y, Gao Z. 2017.** UFGT: the key enzyme associated with the petals variegation in Japanese apricot. *Frontiers in Plant Science* **8**:67–14 DOI [10.3389/fpls.2017.00108](https://doi.org/10.3389/fpls.2017.00108).
- Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M. 2005.** Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. *Plant Molecular Biology* **59**(1):191–203 DOI [10.1007/s11103-005-6503-6](https://doi.org/10.1007/s11103-005-6503-6).
- Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G. 2012.** The genome of *Prunus mume*. *Nature Communications* **3**:1318.
- Zhang S, Chen W, Xin L, Gao Z, Hou Y, Yu X, Zhang Z, Qu S. 2014.** Genomic variants of genes associated with three horticultural traits in apple revealed by genome re-sequencing. *Horticulture Research* **1**(1):14045 DOI [10.1038/hortres.2014.45](https://doi.org/10.1038/hortres.2014.45).
- Zhao ZC, Hu GB, Hu FC, Wang HC, Yang ZY, Lai B. 2012.** The UDP glucose: flavonoid-3-O-glucosyltransferase (*UFGT*) gene regulates anthocyanin biosynthesis in litchi (*Litchi chinensis* Sonn.) during fruit coloration. *Molecular Biology Reports* **39**(6):6409–6415 DOI [10.1007/s11033-011-1303-3](https://doi.org/10.1007/s11033-011-1303-3).
- Zhou L-L, Shi M-Z, Xie D-Y. 2012.** Regulation of anthocyanin biosynthesis by nitrogen in TTG1–GL3/TT8–PAP1-programmed red cells of *Arabidopsis thaliana*. *Planta* **236**(3):825–837 DOI [10.1007/s00425-012-1674-2](https://doi.org/10.1007/s00425-012-1674-2).
- Zuker A, Tzfira T, Ben-Meir H, Ovadis M, Shklarman E, Itzhaki H, Forkmann G, Martens S, Neta-Sharir I, Weiss D, Vainstein A. 2002.** Modification of flower color and fragrance by antisense suppression of the flavanone 3-hydroxylase gene. *Molecular Breeding* **9**:33–41 DOI [10.1023/A:1019204531262](https://doi.org/10.1023/A:1019204531262).