

WITHDRAWN: Combined Transcriptomic and Metabolic Analysis Reveals the Potential Mechanism for Fruit Development and Quality Control of *Rubus chingii* Hu

Zhen Chen

jennie1980@163.com

Research article

Keywords: Rubus chingii Hu, transcriptome, metabolome, ellagic acid, kaempferol-3-O-rutinoside

Posted Date: December 2nd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-117873/v1>

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

EDITORIAL NOTE:

The full text of this preprint has been withdrawn by the authors while they make corrections to the work. Therefore, the authors do not wish this work to be cited as a reference. Questions should be directed to the corresponding author.

Abstract

Background: *Rubus chingii* Hu (Chinese raspberry) is an important dual functional food with nutraceutical and pharmaceutical values. Comprehensive understanding of fruit development and bioactive components synthesis and regulation could accelerate genetic analysis and molecular breeding for the unique species.

Results: A combined transcriptomic and metabolic analysis of *R. chingii* fruits from different developmental stages was conducted in this study. A total of 89,188 unigenes was generated and 57,545 unigenes (64.52%) were got annotated. Differential expression genes (DEGs) and differential ions mainly involved in biosynthesis of secondary metabolites. Phenolic acids and flavonol glycosides syntheses were strongly activated at earlier stages, while amino acids, linolenic acid metabolism and anthocyanins synthesis were dominant at later stages. The core genes participated in biosynthesis of ellagic acid (EA) and kaempferol-3-*O*-rutoside (K-3-R) and their corresponding metabolites were elaborately characterized. And some probable MYB and bHLH transcription factors controlling flavonoids synthesis were also identified.

Conclusion: Combined transcriptomic and metabolic analysis initially reveals molecular and chemical mechanism of fruit development of *R. chingii* Hu fruit. The fruit launched the biosynthesis of phenolic acids and flavonols at the very beginning of fruit set and then coordinately accumulated and converted. And it was tightly regulated by expressions of the related genes and transcription factors. The results provide a solid foundation for genetic analysis, functional genes isolation, fruit quality improvement and modifiable breeding.

Background

Rubus chingii Hu (Rosaceae), one of a unique Chinese ‘raspberry’, is a deciduous shrub and widely distributes in the eastern and southern China [1]. It is primitively diploid ($2n = 2x = 14$) and belongs to the same genus of *Rubus* as raspberry and blackberry. Its dried and unripe fruit at green-to-yellow stage, called “Fu-pen-zi”, has been used as a traditional Chinese medicine (TCM) since ancient times to tonify the kidney, reduce urination, and control nocturnal emissions [2–5]. Recently, phytochemical and pharmacological studies have proved that *R. Chingii* enriches polysaccharides, terpenes and phenolic components, especially ellagic acid (EA) and kaempferol-3-*O*-rutoside (K-3-R), for hepatoprotective and anti-oxidative functions [6–8], vasorelaxation [9], anti-osteoporotic [10], antidiabetic [11] and anti-inflammatory, anti-fungal and anti-cancer activities [1, 12–18]. In addition, the leaf of *R. chingii* is historically used as a folk tea due to its astringent and antithrombotic properties [13, 19]. Meanwhile, its ripe fruit is high in vitamins C and PP, mineral K and Mn, and amino acids, as well as SOD activity, supporting its reputation as “superfood” as other *Rubus* fruits [20]. Therefore, *R. chingii* is a typically dual functional food of medicine and edible with prominent health and economic values.

Though the researches of phytochemistry and pharmacology of *R. chingii* have made great progress in recent years, the agronomic trial is still in a neglect situation. Little report related to field management or cultivar screening is available in literature. The quality standards mainly focused on the mixed dried fruits from numerous seedlings. However, synthesis and dynamic changes of the chemical constituents are genetically programmed and highly coordinated during the whole fruit development process, and easily influenced by genotypes, temperature, light, fertilizers, and other field managements [21–23]. Morphological and organoleptic features may not be sufficient for variety selection. The lack of genetic reference for *R. chingii* has distinctly blocked the application of modern breeding. Only when we illuminate the molecular mechanism of fruit development for the unique species, can we catch the law of medicinal or nutritional ingredients accumulations. Then we could find effective methods to regulate fruit quality and occupy the market.

Rubus L. is a large and diverse genus with a total of 900–1000 species attributed to the frequent intraspecific and interspecific hybridization, and is divided into 12 subgenera [24]. Until now, a wide spectrum of wild species and germplasms keep unexploited. The most economically important and popular species of *Rubus* are red raspberry, black raspberry, blackberry, and their hybrids, but their planting areas (50–70°N) are limited due to the strict chilling requirement. Nevertheless, only in recent years has the genomic information of the genus been unveiled. VanBuren et al. [25, 26] revealed the near complete genome of black raspberry with V1 of 243 Mb sequences using next generation sequencing (NGS)-based approach, or modified V3 of 290 Mb sequences using single molecule real-time PacBio sequencing (SMRT) and Hi-C genome scaffolding. In the case of transcriptomic analysis, large-scale sequencing data and genes information of red raspberry (*R. idaeus* cv. Nova) [27], black raspberry (*R. occidentalis*) grew in Korea [28, 29], blackberry (*Rubus spp.* Var. Lochness) [30, 31], black raspberry (*R. occidentalis*) 'Jewel' [25, 26], *R. coreanus* [32] and Himalayan raspberry (*R. ellipticus*) [33] have been identified, and differentially expressed genes during fruit ripening, especially genes involved in anthocyanin accumulations, were generally verified [27, 28, 32]. However, the fruit ripening process is a comprehensive network of metabolites changes relying on related genes expressions and enzymes activities. In this perspective, combined transcriptomics and metabolomics approaches will provide a powerful dissection tool for better comprehension of biochemical, physiological, and organoleptic changes in the reproductive organs. Hyun et al. [28] made first attempt to establish a preliminarily integrated understanding of black raspberry fruit development by combining these two omics analysis and facilitated the fundamental cognition of gene-metabolite relationships. What's more, some transcription factors, such as MYB, bHLH, WRKY and MADS, involving in anthocyanin synthesis were also tested in berries [28, 34]. Nonetheless, different berries had specific biosynthetic routes and distributions of secondary metabolites that regulated by large disparate genetic expression patterns [35, 36]. For instance, mandarin melonberry enriches genistein and its derivatives while Korean black raspberry contains a relatively higher proportion of flavonoid- and anthocyanins-rutinoside forms. Up to now, there's still no genetic, transcriptomic and metabolic information for *R. chingii*. And synthesis of bioactive components in *Rubus* remains unclear.

With the increased demands from consumers for varied, nutritious, healthy foods and locally grown fruits in China under the warmer climate condition (20–40°N), the planting area of *R. chingii* rapidly expanded. Our groups have tracked phenological periods of *R. chingii* for more than 5 years and screened nearly 20 different germplasm resources with special traits, such as thornless, big fruit, sweet fruit, etc, from thousands of seedlings. Among these, L7 was selected in this study with good taste and highly-contained medicinal ingredients. And the gene expressions and metabolites syntheses, especially the phenolic compounds and their regulators, in the process of *R. chingii* fruit development, were analyzed by combined transcriptome and metabolome analysis. The results provide a solid foundation for genetic analysis, functional genes isolation, fruit quality improvement and *Rubus* breeding.

Methods

Sample collection

Rubus chingii Hu is widely distributed in the eastern and southern China, especially in Provinces of Jiangxi, Zhejiang, Anhui, Jiangsu, Fujian and Guangxi. In last decade, numerous wild seedlings from local hills were domesticated and commercially planting in local areas due to its concerned values for human health. More than 12,000 wild and cultivated plantlets were domesticated or bought, and cultivated from 2011, in the farm of Huahan Raspberry Professional Cooperative on a rental basis, which is located at the foot of a hill (28°73'39"N, 121°09'11"E) in Linhai city, Eastern China. Seedlings were all formal identified by taxonomists Pro. Moshun Chen. Actually, there are abundant variations in intra-species of *R. chingii*, but no national variety has been certified yet. We have screened

nearly 20 different germplasm and homogeneous L7 plants with good taste and highly-contained bioactive components were selected in this study.

Fruit set of L7 occurs in late-March and fruits mature in early-May in this area. Eight stages of fruit growth and development were designated according to anthesis, fruit size and color, including small green (SG, 7 days post-anthesis, 7 DPA), medium green (MG, 14 DPA), big green I (BGI, 21 DPA), big green II (BGII, 28 DPA), big green III (BGIII, 35 DPA), green-to-yellow (GY, 42 DPA), yellow-to-orange (YO, 48 DPA) and red (Re, 54 DPA) (Fig. 1). For *R. chingii* cultivation, high-yield period arrives at the third year. So after planting 2 years, a total of about 50 fruits at each stage from different individuals of the same germplasm (L7) were harvested in 2017. Fruit samples were frozen immediately in liquid nitrogen and then kept at -80 °C until use. Three biological repetitions for RNA preparation and six biological repetitions for metabolites extractions were carried out in this study.

Meanwhile, another group of more than 50 fruits at every stage from the same conditions was harvested for measurements of fruit quality, including size, fresh weight, dry weight, and content of total soluble solid, Vc, total phenols, total terpenoids, ellagic acid and kaempferol-3-*O*-rutinoside. Soluble solid of mature fruit was assessed by refractometer PAL-1 (ATAGO, Japan). Vc was measured using molybdenum blue colorimetry [37].

RNA preparation, cDNA library construction and sequencing

Total RNAs of the four representative stages (BGI, GY, YO and Re) fruits was isolated using modified cetyltrimethyl ammonium bromide (CTAB) method [38]. RNA-seq library construction and sequencing was performed by the Beijing Genomics Institute (BGI) (Wuhan, China). mRNA was enriched with Oligo (dT) method and fragmented using divalent cations under elevated temperature. Then first strand cDNA was synthesized by reverse transcriptase with random hexamer primers, and the second strand cDNA was synthesized using DNA polymerase I and RNase H. After purification, the cDNA fragments were connected with adapters and those with suitable size were selected for PCR amplification. Library quality and quantity were assessed by Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System. Finally, the library was sequenced using Illumina (HiSeq X-Ten) platform.

Transcriptome de novo assembly, annotation, and expression analysis

The raw reads were firstly cleaned by removing adaptor-polluted reads, low-quality (with more than 20% Q < 15 bases) and high content of unknown base (N > 5%) reads. Then transcriptome de novo assembly of clean reads was performed using Trinity (v2.0.6) [39]. The harvested transcripts were clustered into final unigenes using Tgicl (v2.0.6) [40]. The unigenes were divided into two types, one type was cluster, which the prefix was CL with the cluster ID behind it (in one cluster, there were several unigenes that the similarity between them was more than 70%), and another type was singleton with the prefix of Unigene.

All assembled transcripts were aligned and annotated using Blastn and Blastx (v2.2.23) [41] or Diamond (v0.8.31) to NCBI nucleotide sequence database (NT), non-redundant protein database (NR) (<http://www.ncbi.nlm.nih.gov/>), EuKaryotic Orthologous Groups (KOG) database (<http://www.ncbi.nlm.nih.gov/pub/COG/KOG>), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database and SwissProt database (<http://ftp.ebi.ac.uk/pub/databases/swissprot>). Blast2GO (v2.5.0) was used to do the Gene Ontology (GO) functional classification [42]. And InterProScan5 (v5.11-51.0) (<http://www.ebi.ac.uk/interpro>) was applied to family and domains annotation [43]. TransDecoder was used to identify the candidate coding area. The longest open reading frame (ORF) was extracted, and the Pfam protein homologous sequence was searched by blast to SwissProt and Hmmscan to predict the coding region. Finally, transcription factor (TF) prediction was performed with getorf (EMBOSS: 6.5.7.0) and hmmsearch (v3.0) [44, 45].

For gene expression analysis, we mapped clean reads to unigenes using Bowtie2 [46] and then calculated the gene expression level by RSEM (v1.2.12) with the formula: Fragments Per Kilobase per Million mapped fragments (FPKM) = $10^6 C / (NL / 10^3)$ [47]. We performed hierarchical clustering analysis with hclust function and PCA analysis with princomp function of R. Software. Differential expression genes (DEGs) were detected with DEseq2 as the request described by Michael et al. [48] Hierarchical clustering for DEGs was carried out using pheatmap. We classified the DEGs based on the GO and KEGG annotation results and official classification, and also mapped functional enrichment using a phyper. Hypergenometric test was adopted, and the terms which false discovery rate (FDR) was no longer than 0.01 were defined as significant enrichment.

Metabolites extraction

Fruit sample (25 mg) from BGI, GY, YO or Re stage was extracted with frozen solution of 800 μ L methanol: water (1:1, v/v) using Tissue-Lyser at 60 Hz for 4 min. Then the mixture was kept at -20 °C for 2 h and centrifuged at 4 °C under 30 000 g for 20 min. 500 μ L supernatant was transferred to a new Ep tube for UPLC-Q-TOF-MS analysis.

UPLC-Q-TOF-MS analysis

All samples were acquired by the LC-MS system followed machine orders at the Beijing Genomics Institute (BGI) (Shenzhen, China). Firstly, all chromatographic separations were performed using an ultra performance liquid chromatography (UPLC) system (Waters, UK). An ACQUITY UPLC BEH C18 column (100 mm $\times$ 2.1 mm, 1.7 μ m, Waters, UK) was used for the reversed phase separation. The column oven was maintained at 50 °C. The mobile phase consisted of solvent A (water + 0.1% formic acid) and solvent B (methanol + 0.1% formic acid) with the flow rate of 0.4 mL/min. Gradient elution conditions were set as follows: 100% phase A from 0 to 2 min, 0-100% phase B from 2 to 11 min, 100% phase B from 11 to 13 min, and 0-100% phase A from 13 to 15 min. The injection volume for each sample was 10 μ L.

Then the eluted metabolites were detected by a high-resolution tandem mass spectrometer Xevo G2-XS QTOF (Waters, UK) operated in both positive and negative ion modes. For positive ion mode, the capillary and sampling cone voltages were set at 1.0 kV and 40.0 V, respectively, while for negative ion mode, they were changed to 3.0 kV and 40.0 V. The mass spectrometry data were acquired in Centroid MSE mode. The TOF mass range was from 50 to 1200 Da with a 0.2 s scan accumulation time. Moreover, for the MS/MS detection, all precursors were fragmented using 20–40 eV, and the scan time was 0.2 s. During the acquisition, the LE signal was monitored to calibrate the mass accuracy. Furthermore, a quality control sample (pool of all samples) was collected after every 10 samples, in order to evaluate the stability of the LC-MS during the whole acquisition.

The raw data were pretreated using metaX to filter low quality ion [49], and the ions of $RSD \leq 30\%$ were kept for next analysis by Progenesis Q1 (v2.2), including baseline correction, peak detection and matching, and retention time alignment. Major metabolites were identified using Progenesis Q1 (v2.2) and KEGG database by comparing both the mass spectra and retention time. The multivariate statistical analysis was performed using metaX, including Principal Component Analysis (PCA) and PLS-DA, and metabolites with a Variable Important for the Projection (VIP) value (≥ 1.0) were selected as suitable products. Finally, metabolites of significantly different expression were screened based on VIP value and Fold change (FC) analysis with $FC \geq 1.2$ or ≤ 0.8333 and $q\text{-value} < 0.05$. Then these differential ions were identified using Progenesis Q1 (version 2.2). Heatmap was generated to visualize metabolite profile and KEGG pathway analysis was performed to help understanding the functions of metabolites.

Real-time quantitative PCR assay

Total RNAs from eight independent stages were isolated respectively using RNAprep Pure Plant Plus Kit (Polysaccharides & Polyphenolics-rich, DP441, Tiangen Biotech, China) according to the manufacturer's instruction. After concentration and purity detections by NanoDrop2000 (Thermo, USA), the RNA was reverse-transcribed to cDNA using PrimeScript™ RT Master Mix Kit (RR036A, TaKaRa, Japan). The quantitative expressions of enzyme and TFs encoding genes involved in synthesis of phenolic acid compounds and flavonoids were determined by qPCR in triplicate with TB Green® *Premix Ex Taq*™ (RR420A, TaKaRa, Japan) using CFX96 real-time PCR detection system (Bio-Rad, USA). The parameter was set as follows: 95 °C for 10 min, followed by 40 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 20 s, and then melting curve was made. Actin was chosen as the control. PCR primers were listed in Table S1. The relative expression level was calculated using $2^{-\Delta\Delta C_t}$ method [13].

Measurement of total phenols and terpenoids

The fully dried fruits were ground to powder and sieved through 60-mesh. Total phenolic (TP) contents were determined by the Folin-Ciocalteu method with a few modifications [19, 31]. 1.0 g of each sample was refluxed and condensed three times with 30 mL 70% (v/v) ethanol (T = 45 °C, t = 30 min). After filtration, the extract was concentrated by rotary evaporation and then diluted with pure deionized water to the final volume of 100 mL. Finally it was thoroughly extracted with petroleum ether. For chromogenic reaction, 1 mL of each extraction was mixed with 1 mL Folin-Ciocalteu solvent in 10 ml volumetric flask with a vortex for 30 s, and then 3 mL of 20% Na₂CO₃ was added to the mixture, following deionized water up to the constant volume. After being kept in dark for 60 min at room temperature, the optical density was detected at wavelength of 765 nm. The standard calibration curve was plotted with 0 ~ 0.08 mg/mL gallic acid.

For total terpenoids (TT) assessment, colorimetric method was applied with 5% vanillin-acetic acid and perchloric acid as the developer [50]. First, 0.1 g fruit powder was extracted by 25 mL 55% ethanol with ultrasonic treatment for 30 min. After filtration, 1 mL extraction was kept at 70 °C water bath for 15 min, and then 0.2 mL 5% vanillin-acetic acid and 0.4 mL perchloric acid were added. The mixture was heated at water bath of 70 °C for 15 min and then 5 mL acetic acid was supplemented. Fifteen minutes later, the absorbance at 480 nm was recorded. Oleanolic acid (0 ~ 0.04 mg/mL) was used as standard substance.

HPLC analysis of ellagic acid and kaempferol-3- O -rutinoside

The principal medical ingredients of *R. chingii* Hu, EA and K-3-R, were analyzed by HPLC in terms of the standards described in Pharmacopoeia of the People's Republic of China [5]. Briefly, for ellagic acid determination, 0.5 g fruit powder was refluxed with 50 mL 70% (v/v) methanol for 60 min and then made up for the weight. After filtration and suitable dilution, sample of 20 µL was injected into HPLC system (LC-20A, Shimadzu, Japan) with C18 column (25 °C) and the mobile phase of acetonitrile – 0.2% phosphate (20:80) at the flow rate of 1 mL/min. The wavelength was set at 254 nm.

For kaempferol-3-*O*-rutinoside measurement, fruit powder of 1 g was refluxed with 50 mL 70% (v/v) methanol for 60 min and the deficient weight was replenished by 70% methanol. After filtration, 25 mL extracting solution was dried by distillation and the residue was re-dissolved with 20 mL water. Then the solution was extracted by 20 ml petroleum ether for three times. After that, the supernatant was extracted by 20 mL water-saturated butanol solvent for another three times. Finally, the butanol was dried by distillation and the residue was re-dissolved with 5 mL methanol. 20 µL of the final extracting solution was analyzed by HPLC with the same mobile phase and flow rate as EA analysis. The wavelength was set at 344 nm and the temperature for C18 column was 30 °C.

Results

Fruit development of *R. chingii*

R. chingii Hu is an erect shrub with biennial canes that sprout from the perennial root system (Fig. 1a). The axillary mixed buds were initiated from summer of the first year of growth, differentiated in autumn, and then followed by dormancy in winter. At early-March of the next year, new leaves and flower buds emerge. One week later seedlings began to blossom and their florescence usually lasted for 20 d. The whole process of fruit development could be divided into eight stages: small green (SG, 7 DPA, March 24), medium green (MG, 14 DPA, March 31), big green I (BGI, 21 DPA, April 7), big green II (BGII, 28 DPA, April 14), big green III (BGIII, 35 DPA, April 21), green-to-yellow (GY, 42 DPA, April 28), yellow-to-orange (YO, 48 DPA, May 4) and red (Re, 54 DPA, May 10) (Fig. 1b). During the two weeks of the big green stages, there were no obvious changes of appearance. The fresh weight, dry weight, vertical diameter and transverse diameter of fruits were increased rapidly at early stages (7 ~ 21 DPA), then remained stable with big green appearance at middle stages (21 ~ 35 DPA), and finally sharp increased at last stages (42 ~ 54 DPA) until full maturity (Fig. 1c and d). The yield of medicinal fresh fruit at GY stage reached about 36.5 Kg per hectare. The content of water and soluble solid of red fruit were 83.11% and 15.8%, respectively. The Vc content of *R. chingii* L7 reached 66.09 mg/100 gFw.

Transcriptome sequencing and de novo assembly

In our project, about 80.42 Gb bases in total were generated on Illumina HiSeq sequencing platform. After removing low-quality reads and trimming adapter sequences, approximately 639.41 million total raw reads with high quality were obtained from all samples of the four representative developmental stages of *R. chingii* fruits (Table S2). Then de novo assembly was carried out with clean reads using Trinity to construct the full-length transcript. After filtering the abundance by Tgicl, a unigene database for *R. chingii* containing 89,188 unigenes was established. The total length, average length, N50, and GC content of these unigenes were 96,361,653 bp, 1,080 bp, 1,916 bp and 41.72%, respectively (Table 1).

Table 1
Quality metrics of unigenes

Sample	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
BG1	39,610	43,866,470	1,107	1,773	1,209	478	43.05
BG2	57,537	51,563,673	896	1,599	912	339	41.65
BG3	38,187	41,638,154	1,090	1,757	1,190	464	43.03
GY1	33,726	38,076,560	1,128	1,795	1,235	497	43.15
GY2	40,506	43,788,492	1,081	1,785	1,199	444	42.6
GY3	41,584	43,303,980	1,041	1,744	1,158	413	42.43
YO1	36,262	38,755,641	1,068	1,721	1,158	451	43.31
YO2	38,656	41,641,351	1,077	1,728	1,173	458	42.89
YO3	39,441	41,500,069	1,052	1,707	1,144	439	43.24
Re1	31,502	34,341,099	1,090	1,729	1,177	475	43.39
Re2	34,734	35,994,711	1,036	1,709	1,113	424	43.15
Re3	33,154	34,287,081	1,034	1,694	1,106	425	43.32
All-Unigene	89,188	96,361,653	1,080	1,916	1,235	413	41.72

Functional annotation of *R. chingii* transcriptome

For functional annotation, the unigenes were first searched using Blastn or Blastx against NCBI database with a cut-off E-value of 10^{-5} . Out of all 89,188 unigenes, 47,957 (53.77%) and 49,755 (55.79%) had BLAST hits to known nucleotides in NT database and proteins in NR database, respectively. Based on the NR database function annotation result, 64.37% annotated unigenes of *R. chingii* were matched with strawberry *Fragaria vesca* subsp. *vesca* proteins (Fig. 2), indicating the conserved evolution between *Rubus* and *Fragaria*. Other related species were *Prunus mume* (7.77%), *Prunus persica* (5.62%), and *Malus domestica* (3.14%), all belonged to the Rosaceae. What's more, all unigenes were annotated by aligning with other 5 functional databases, and finally, 33,500 (SwissProt: 37.56%), 39,051 (KOG: 43.79%), 37,833 (KEGG: 42.24%), 2,837 (GO: 3.18%) and 41,238 (InterPro: 46.24%) unigenes were annotated. The prediction of coding DNA sequences (CDS) showed that the total number, total length, N50, N90, max-length, min-length and GC content were 49,459 bp, 53,839,431 bp, 1,452 bp, 501 bp, 15,330 bp, 297 bp and 45.03%. Taken together, overall 57,545 unigenes (64.52%) were got annotated, and the other 35.48% unigenes might contribute to the special characteristics of *R. chingii*. These results suggest that de novo assembled transcripts covered a wide range of genetic information of *R. chingii*, and provided an invaluable resource for facilitating the identification of novel genes implicated in specific developmental and metabolic processes.

In the case of Eukaryotic Orthologous Groups (KOG) assignment, 25 functional groups were classified. Among these, except for the "general function prediction only", the cluster for "Signal transduction mechanisms" (6,552 unigenes) represented the largest group, followed by "Posttranslational modification, protein turnover, chaperones" (4,772 unigenes), "Transcription" (3,554 unigenes), "Carbohydrate transport and metabolism" (2,705 unigenes) and "RNA

processing and modification” (2,509 unigenes), suggesting the active metabolisms in the procedure of fruit development of *R. chingii* (Figure S1). Further, to identify the precise biochemical pathways, the sequences were mapped against the KEGG pathway. As a result, a total of 37,833 unigenes were involved in different pathways (Figure S2). And most of them were participated in metabolisms, including “Carbohydrate metabolism (3,144 unigenes)”, “Amino acid metabolism (1,719 unigenes)”, “Lipid metabolism (1,533 unigenes)”, “Biosynthesis of other secondary metabolism (1,307 unigenes)”, “Metabolism of terpenoids and polyketides (788 unigenes)”, and so on. All these results were in accordance with the accumulations of fruit nutrients and vigorous secondary metabolic synthesis for medicine constituents in *R. chingii* Hu.

Metabolite profiling of *R. chingii* fruits at different ripening stages

Mass spectrometry based on metabolite profiling of *R. chingii* fruits at four stages was also performed to investigate the changes of metabolic compositions. In total, 8,352 ions were captured in positive ion additive mode (*pos*). After filtering low quality and RSD > 30% ions, 7,611 ions (92.61%) were harvested. The identification numbers at MS and MS2 levels were 4730 and 2999, respectively. Simultaneously, the number of total ions, RSD ≤ 30% ions, MS ions and MS2 ions in negative ion additive mode (*neg*) were 6276, 4683 (76.15%), 2070 and 1130, respectively. Multivariate analyses of Principal Component Analysis (PCA) and partial least squares discriminant analysis (PLS-DA) were performed to support pattern recognition of the metabolic differences in *R. chingii* samples at different stages. In *pos*, the score of PC1 and PC2 were 40.48% and 30.76%, respectively (Fig. 3). Stages were relatively separated from each other, especially the BG and Re stage at PC1 axis. And the same trend was found in *neg*. Reliable PLS-DA models were also established between every two stages with high values of R2 and Q2: 0.998 and 0.991 from GY:BG (*pos*), 0.994 and 0.950 from YO:GY (*pos*), 0.994 and 0.965 from Re:YO (*pos*), and similar in *neg* mode. Then based on the VIP value (≥ 1) by PLS-DA analysis, fold-change (≥ 1.2 or ≤ 0.8333) by univariate analysis and q-value < 0.05, differential ions were screened and identified (Table 2, Figure S3).

Table 2
The number of differential ions (metabolites) during ripening fruits of *R. chingii*

Mode	Group	Diff ion number	Up(MS)	Down(MS)	Up(MS2)	Down(MS2)
<i>pos</i>	GY:BG	2008	852	461	589	284
<i>pos</i>	YO:GY	1716	596	466	414	290
<i>pos</i>	Re :YO	2164	344	984	224	692
<i>neg</i>	GY:BG	1243	409	225	232	125
<i>neg</i>	YO:GY	980	203	257	121	141
<i>neg</i>	Re :YO	1390	210	344	104	197

Differential expression genes and their products during the four stages of *R. chingii* Fruit

Differential expression genes (DEGs) were detected between every two stages with the parameters: Fold Change ≥ 2.0 and Adjusted Pvalue ≤ 0.05. As a result, 8,792 genes were differentially expressed between the green-to-yellow

turning stage (GY) and big green I stage (BGI). Among these genes, expressions of 4,054 genes were upregulated while 4,738 genes were downregulated. Meanwhile 6,439 genes were annotated in KEGG pathway and 134 pathways were involved. The top enrichment pathways were as follows: “Metabolic pathways” with 1,648 (25.59%) DEGs, “Biosynthesis of secondary metabolites” with 1,033 (16.04%) DEGs, “Phenylpropanoid biosynthesis” with 226 (3.51%) DEGs, “Starch and sucrose metabolism” with 164 (2.55%) DEGs, “Flavonoid biosynthesis” with 93 (1.44%) DEGs, and “Phenylalanine metabolism” with 67 (1.04%) DEGs, and so on (Table S3a). Correspondingly, 2008 differential ions (Diff ions) in *pos* or 1243 Diff ions in *neg* were detected between GY and BG stages (Table 2). The top enrichment pathways were consistent with that of gene expressions changes. In *pos*, 415 (20.67%) Diff ions were found in “Metabolic pathways”, followed by 332 (16.53%) Diff ions in “Biosynthesis of secondary metabolites”, 54 (2.69%) Diff ions in “Flavonoid biosynthesis”, 49 (2.44%) Diff ions in “Diterpenoid biosynthesis”, 48 (2.39%) Diff ions in “Tyrosine metabolism”, and 39 (1.94%) Diff ions in “Phenylpropanoid biosynthesis”. The similar trends were found between GY vs YO or YO vs Re stages (Table S3b and c). More visualized results of DEGs in pathways were illustrated by bubble diagrams (Fig. 4a, b, c). Eleven pathways were enriched across the whole fruit development process, mainly focusing on those genes involved in secondary metabolites, such as “Phenylpropanoid biosynthesis”, “Flavonoid biosynthesis”, “Cutin, suberine and wax biosynthesis”, and “Stilbenoid, diarylheptanoid and gingerol biosynthesis”, among others. Interestingly, among these four stages towards maturation, DEGs in pathways of “Phenylalanine, tyrosine and tryptophan biosynthesis”, “Flavone and flavonol biosynthesis”, “Diterpenoid biosynthesis” and “Tropane, piperidine and pyridine alkaloid biosynthesis” were detected specifically in the early stages (Fig. 4a, b, c). And the corresponding products in these pathways might act on plant defenses against biotic and abiotic stresses under wild environments for fruits protection and reproduction, and now exactly contribute to the bioactive roles for human health.³¹ Therefore it confirms that the unripe fruit of *R. chingii* is the perfect resource for medicinal material. In contrast, activities of “Alpha-linolenic acid metabolism”, “Biosynthesis of amino acids”, “C5-branched dibasic acid metabolism”, “Thiamine metabolism”, “Valine, leucine and isoleucine biosynthesis”, as well as “Carotenoid biosynthesis” were significantly enhanced in the later stages (Fig. 4a, b, c), just ready for maturation of the edible, delicious and nutritive red fruits.

X axis represents enrichment factor and Y axis represents pathway name. The color indicates the q value (low: blue, high: white) and the lower q value refer to the more significant enrichment. Point size represents DEG number (the bigger dots is, the larger amount get). Rich Factor represents the value of enrichment factor, which is the quotient of the number of DEGs and total Gene amount. The larger value accounts for the more significant enrichment.

Biosynthesis and regulation of phenolic acids and flavonol derivatives

More than 235 chemical constituents, especially 15 diterpenoids, 15 triterpenoids, 18 flavonoids, 5 coumarins, 9 steroids and 56 organic acids (including phenolic acids, fatty acids, tannins, citric acid, and others) have been identified in *R. chingii* through phytochemical approaches [1]. However, their biosynthetic and metabolic mechanisms remain unclear. Changes of total terpenoids (TT) and total phenols (TP) contents were first determined and the results showed that both of TT and TP contents were gradually decreased with the development of fruits, whereas the accumulations of TT and TP kept at relative stable levels (Table 3). It was coincident with the changes of differential accumulated metabolites and differential expressions of the related core genes in shikimate, phenylpropanoids and flavonoids synthetic pathways, determined by metabolic and transcriptomic analyses. Among these, ellagic acid (EA, a type of phenolic acid) and kaempferol-3-*O*-rutoside (K-3-R, one of the flavonol glycosides) are the two key bioactive components and their extraction and content standards are described in Chinese Pharmacopoeia [5]. So it prompted us to characterize their molecular mechanism of synthesis and regulation.

As shown in Fig. 5a, the EA synthesis is primarily derived from the shikimic acid pathway, which starts with the condensation of erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP) catalyzed by 3-deoxy-7-phosphoheptulonate synthase (DAHPS, EC: 2.5.1.54) [51, 52]. Then with the following catalyses of 3-dehydroquinate synthase (DHQS, EC: 4.2.3.4e) and bifunctional enzyme dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH, or DHD/SDH EC: 4.2.1.10 and E.C. 1.1.1.25), 3-dehydroquinic acid (DHQ) is converted to 3-dehydroshikimate and shikimate. After enolization, 3-dehydroshikimate can be catalyzed by SDH to generate gallic acid (GA). Finally, EA is a dimeric derivative of GA and can be polymerized to form ellagitannins (ETs). In this pathway, seven unigenes encoding DHAPS, DHQS and DQD/SDH were found from *R. chingii* transcriptome library, including two putative *DAHPS* genes, one putative *DHQS* gene and four putative *DHD/SDH* genes (Fig. 5 and Table S4). Intriguingly, the relative expression levels of *DHQS*, *DQD/SDH1*, *DQD/SDH2*, and *DQD/SDH4* reached extremely high at SG stage, decreased sharply at the subsequent stage and then kept at relative stable levels (Fig. 5a and 6), indicating that *R. chingii* triggered the strong expressions of these genes at the very beginning of fruit set. Exceptionally, *DQD/SDH3* expression remained low levels at all stages. Among the four representative stages, *DAHPS1* and *DAHPS2* transcripts were notably more abundant than *DHQS* and *DHD/SDHs* and reached the maximum values at BGI stage, decreased at GY stage, and then had a certain recover at YO stage. Finally, the expression of *DAHPS1* and *DAHPS2* reduced to the lowest levels at Re stage. The length of *DAHPS1* and *DAHPS2* transcripts were 2498 and 2454 bp, respectively, with the deduced proteins of 532 and 539 amino acids (Table S4). The sequence of *R. chingii DAHPS1* had similarities of 89.6% and 89.9% with the *DAHPS1* from *Rosa chinensis* and *F. vesca*, respectively. And the sequence of *R. chingii DAHPS2* exhibited similarities of 89.7% and 87.3% with *DAHPS2* from *R. chinensis* and *F. vesca*. Analogously, the genes information of *DHQS* and *DQD/SDH1-4* were list in Table S4. Correspondingly, UPLC-Q-TOF-MS analysis showed the contents of substrates and products in this pathway, including E4P, DHQ, 3-Dehydroshikimate, GA and EA, were all reached the maximum levels at BGI stage and reduced to the lowest levels at Re stage (Fig. 5a and Table S5). And GA and EA were remarkably enriched, confirming the functional role of *R. chingii* fruits. Furthermore, HPLC results showed that the EA percentage contents gradually declined as ripening progressed, while the EA accumulations in each fruit reached highest at BGI and BGIII stages, and kept relative stable levels from BGIII stage to Re stage (Table 3).

On the other branch, 3-dehydroshikimate can be catalyzed by SDH to generate shikimic acid (SA). Then SA converts to chorismate, the common precursor of phenylalanine (Phe), tryptophan (Trp) and tyrosine (Tyr) (Fig. 5). The flavonoid biosynthetic pathway derives from phenylpropanoid pathway with the initiated compound, Phe, and the products include flavonols, anthocyanins, proanthocyanidins, flavone, flavanone, and isoflavone (Fig. 5b). Dihydrokaempferol and dihydroquercetin are the common precursors for synthesis of flavonols and anthocyanins branches. Seventeen unigenes encoding 9 putative enzymes, including phenylalanine ammonialyase (PAL), cinnamic acid 4-hydroxylase / a cytochrome P450 monooxygenase CYP73A (C4H/CYP73A), 4-coumaric acid: CoA ligase (C4L), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone-3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), flavonol synthase (FLS), and flavonoid 3-O-glucosyltransferase (UGT78D2) / *Rubus* hybrid cultivar 'Arapaho' UDP glucosyltransferase (UGT78H2) were identified to involve in flavonol derivatives synthesis. Accordingly, most of them were initially high expressed and the major corresponding metabolic products were accumulated at earlier stages (Fig. 5b and 6), and then gradually decreased or rebounded with a slight increase at GY or YO stages. On the contrary, expression of another three unigenes for 3 putative enzymes, including dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS) / leucoanthocyanidin dioxygenase (LDOX), and anthocyanidin 5,3-O-glucosyltransferase (UGT88A1), which catalyze dihydroquercetin to anthocyanins, increased as the ripening proceeded and reached the peak at red stage. Two *PAL* genes, two *C4H* genes, three *C4L* genes, one *CHS* gene, two *CHI* genes, one *F3H* gene, and one *UGT78D2/78H2* gene all exhibited similar expression pattern. However,

the expression pattern of *F3'H3-like* was contrary to *F3'H1* and *F3'H2*, the same as *FLS2* and *FLS1*, predicting that the expressions of *F3'H3-like* and *FLS2* might be ready for anthocyanidin synthesis. From the results of UPLC-Q-TOF-MS, we found that the contents of kaempferol and kaempferol-3-glucoside were extremely high at the four stages in comparison with other compounds, whereas kaempferol-3-*O*-glucoside (K-3-R) was not checked out (Fig. 5b and Table S5). That's probably because some groups are easily cleaved in MS/MS analysis. HPLC analysis showed percentage contents of K-3-R were decreased with the development of fruit, but its accumulation in each fruit reached peak at BGII stage, and then gradually decreased (Table 3).

Furthermore, the flavonoids biosynthesis is under stringent regulations [34, 53–55]. Among the four representative stages of *R. chingii* fruits, a total of 2118 transcription factors (TFs) encoding genes were predicted in this study, including 303 *MYB*, 250 *MYB-related*, 95 *bHLH (MYC)*, 111 *NAC*, 74 *WRKY*, and so on (Fig. 7a). Among these, some *MYB* TFs, such as RucC1, RucMYB4, RucMYB5, RucMYB12-1, RucMYB12-2, RucMYB39, RucMYB44, RucMYB46, RucMYB86-1, RucMYB86-2, RucMYB86-3, RucMYB308, RucMYB308-like, RucMYB1R1-1, and some *bHLH* (basic helix-loop-helix) TFs, such as RucbHLH13, RucbHLH63, RucbHLH93, RucbHLH137, as well as WD40-1 and WD40-2, showed the parallel expressions pattern as the core genes of flavonoids synthesis, while other *MYB* TFs, such as RucMYB6, RucMYB44-like, RucODO1 (ODORANT1, *MYB*-TF), RucMYB1R1-2, and other *bHLH* TFs, such as RucbHLH62, RucbHLH77, RucbHLH96 and RucbHLH130, exhibited the opposite expression pattern and mainly regulated red fruit ripening (Fig. 7b and c). The regulations of TFs were also detected by RT-qPCR in another genotype L8 fruits during whole development and showed same tendency (data not shown). Thus, the biosynthesis and conversions of phenolic acids and flavonols may be regulated by the coordinated actions of diverse transcription factors during fruit development, and detailed mechanisms need to be further investigated.

Table 3
Changes of the main medicinal components contents during fruit development in *R. chingii*¹

Developmental Stages (DPA)	Dry weight of single fruit (g)	TT content (%)	TP content (%)	EA content (%)	K-3-R content (%)	TT content per fruit (mg)	TP content per fruit (mg)	EA content per fruit (mg)	K-3-R content per fruit (mg)
14 (MG)	0.109 ± 0.015 e	-	-	0.486 ± 0.004 b	0.0398 ± 0.0022 a	-	-	0.528	0.0432
21 (BGI)	0.194 ± 0.011 d	5.55 ± 0.15 a	2.91 ± 0.20 a	0.536 ± 0.027 a	0.0173 ± 0.0027 d	10.79	5.66	1.042	0.0337
28 (BGII)	0.423 ± 0.048 b	4.24 ± 0.16 b	2.35 ± 0.17 b	0.282 ± 0.016 c	0.0258 ± 0.0028 b	17.95	9.95	1.194	0.1092
35 (BGIII)	0.349 ± 0.030 c	3.37 ± 0.12 c	2.05 ± 0.05 c	0.139 ± 0.003 d	0.0209 ± 0.0020 c	11.75	7.15	0.485	0.0730
42 (GY)	0.450 ± 0.097 b	3.11 ± 0.01 cd	1.22 ± 0.02 d	0.075 ± 0.000 e	0.0096 ± 0.0002 e	13.99	5.49	0.337	0.0433
48 (YO)	0.469 ± 0.088 b	2.86 ± 0.12 d	1.38 ± 0.01 d	0.067 ± 0.001 e	0.0079 ± 0.0002 e	13.41	6.47	0.314	0.0368
54 (Re)	0.766 ± 0.090 a	1.61 ± 0.54 e	0.51 ± 0.01 e	0.056 ± 0.002 e	0.0032 ± 0.0000 f	12.34	3.91	0.429	0.0242
¹ DPA, days post-anthesis; TT, Total terpenoids; TP, Total phenols; EA, Ellagic acid; K-3-R, Kaempferol-3-O-rutinoside.									

Discussion

Recently, consumption as well as planting of the Chinese raspberry, *Rubus chingii* Hu, has rapidly increased for its nutraceutical and pharmaceutical benefits. However, the synthesis and regulation of the bioactive constituents are still in blank. In this study, we performed transcriptomics and metabolomics analyses of *R. chingii* Hu fruits from different development stages. Total of 89,188 unigenes were generated and 57,545 unigenes were got annotated. Blast result showed 64.37% similarity between *R. chingii* and *F. vesca* subsp. *vesca*. The two species belong to the same sub-family Rosoideae of Rosaceae and have the same basal chromosome number of x = 7 [56, 57]. As previous reports, red raspberry *R. idaeus* 'Nova' fruit transcriptome exhibited 68% and 77% similarities with *F. versa* and *R. occidentalis* [27], and 73.5% of blackberry *R. occidentalis* 'Lochness' annotated genes matched similarity with

F. versa [30]. These findings proved the high similarities and certain divergences among the *Rubus* species, and the information of sequence homology and difference will accelerate genetic analysis and molecular breeding of *R. chingii*.

Furthermore, differential expression genes (DEGs) and differential ions (Diff ions) examinations revealed the molecular and chemical mechanism of fruit development in *R. chingii*. And KEGG pathway analysis confirmed the corresponding tendency in metabolites changes and gene expressions. Metabolic pathways, especially secondary metabolites biosynthesis thrived across the whole fruit development process. It was also found in blackberry (*Rubus* sp. Var. Lochness) fruits [30]. What's more, flavone and flavonol biosynthesis and diterpenoid biosynthesis were relatively higher in the early stages while amino acids and linolenic acid metabolism were prominent in the later stages. It is in agreement with the findings in *R. ellipticus* that the concentrations of polyphenolic compounds reached maximum in unripened stages S-1 and S-2, and then decreased significantly as the fruits proceeding towards maturation [58]. The phenols are produced by plants for defensive purposes as "natural pesticides" and then would be oxidized and switch to anthocyanins biosynthesis for fruits maturation [32, 35, 38]. Similarity, in *R. coreanus*, degradation of aromatic compounds, and starch and sucrose metabolism were enriched in the two early stages, while polycyclic aromatic hydrocarbon degradation and bisphenol degradation were detected specially in the later stages, leading to the formation of special aroma and flavor of ripening *Rubus* fruits [32]. Another obvious sign of ripe and soft fruits is the degradation of cell wall components and cuticular/wax layer. So DEGs for cutin, suberin and wax biosynthesis were significantly enriched across *R. coreanus* fruit ripening [32], and this case was also found in *R. chingii* fruit in our study. Altogether, these results will be beneficial for defining the syntheses and conversions of metabolites for fruit qualities and provide the basic data for modifiable filed managements for *R. chingii*.

The ellagitannins (ETs) and ellagic acid (EA) are abundantly existed in berries, such as strawberry, raspberry, blackberry, as well as in tea and some nuts [52, 59]. Their main bioactivities are antioxidant, anti-inflammatory, antiglycation, antidiabetic activity, hepatoprotective and anticancer.^{11,59} A total of 25 ETs were lately identified by HPLC-QTOF-MS/MS in *R. chingii* and 5 ETs were isolated, including monomer and dimer ellagitannins [11]. Nonetheless, the EA was the main component among the ETs according to the highest peak value by HPLC detection and a series of pharmacological tests deeply revealed that EA had the highest activities for ABTS-+ scavenging, anti-glycation than other ETs [59]. The paper suggested that polymerization of EA might lead to the change of action mode. Hence, the decreased EA content at later stages in our study might attribute to the polymerization of EA or the metabolic pathways transfer to the biosynthesis of anthocyanins as the ripening progressed. In addition, Chen et al. [59] found ETs contents in the YF (yellow fruit) group were significantly higher than that in GF (green fruit) group, indicating the fruit's color was closely correlated with the ETs content and composition. However, the "GF" and "YF" meant the color of the dried unripe fruits from different area, not the color of the fresh fruit. So the color may be affect by the drying process. What's more, we found that there were extensive diversities among the *R. chingii* seedlings and the chemical compositions and contents were remarkably varied among the different germplasms (data no shown). Therefore, the fruit development from the homogeneous seedlings was first investigated in this study. The results preferably illuminated the developmental law of *R. chingii* fruit and metabolites synthesis mechanisms.

More specifically, the genes involving in EA synthesis pathway have not been elaborately investigated. In this paper, we identified seven unigenes encoding DHAPS, DHQS and DQD/SDH and assessed their expression patterns during the development of *R. chingii* fruit. As the first committed enzyme in shikimic acid pathway, DAHPS controls carbon flow. *DAHPS* genes in some species, including *Arabidopsis thaliana*, rice (*Oryza sativa*), tomato (*Lycopersicon*

esculentum), potato (*Solanum tuberosum*), and *Gossypium hirsutum* have been isolated, and these species all have at least two putative genes, *DAHPS1* and *DAHPS2* [60]. Expression patterns of these two genes had distinct difference and *AtDAHPS1* performed a strong role in environmental resistances, while tomato two *DAHPS* genes showed different tissue-specific patterns of expression. In this study, expressions of *DAHPS1* and *DAHPS2* in *R. chingii* fruits exhibited the same pattern for fruit development. *DQD/SDH* genes have been functionally characterized in some species and only single *DQD/SDH* gene was identified in *Arabidopsis* with the N-terminal DQD domain (1-316 aa) and C-terminal SDH domain (328–588 aa) [61]. In *Populus trichocarpa*, five putative *DQD/SDH* genes have been found with different activities [51]. In tea (*Camellia sinensis*), *CsDQD/SDH2* and *CsDQD/SDH3* had a similar expression pattern and were synergistic with each other, while *CsDQD/SDH1* exhibited an opposite expression pattern and negatively regulated the other two genes in same family [62]. In our study, four putative *DQD/SDH* genes (encoding 530–534 aa) were identified with similar expression patterns during fruit development. *DQD/SDH1*, *DQD/SDH2* or *DQD/SDH4* may play the dominant roles and their interaction need to be further investigated. Therefore, the identifications of key genes and their expression patterns would be potential to *R. chingii* breeding for higher medicinal quality.

The flavonoid biosynthetic pathway and regulation from transcription factors (TFs) have been widely reported. The flavonols metabolism in *R. chingii* fruits was firstly investigated in this study. The results revealed that *R. chingii* launched flavonols synthesis at the very beginning of fruit set, and then coordinately accumulated and converted. A similar phenomenon also occurred in strawberry receptacle, flavonols (rutin, quercetin glucuronide, kaempferol glucuronide, and others) prominently accumulated at early to middle stages, while significantly declined at red stage [35]. Kaempferol is the most common flavonol and presents in different glycosidic forms. Yu et al. [1] summarized that there were 5 types of kaempferol glucoside in *R. chingii*, including kaempferol-3-*O*- β -D-glucuronic acid methyl ester, kaempferol-7-*O*- α -L-rhamnoside, kaempferol-3-*O*-hexoside, kaempferol-3-glucuronide, and kaempferol-3-*O*- β -D-rutinoside. In this study, kaempferol, kaempferol-3-*O*-glucoside, kaempferol-3-*O*- β -D-glucosylgalactoside, and kaempferol-3-sophorotrioside were identified by UPLC-Q-TOF-MS analysis and kaempferol-3-*O*- β -D-rutinoside was determined by HPLC. Kaempferol and kaempferol-3-*O*-glucoside maintained at high levels within all the fruit developmental stages, while the others mainly accumulated at BG stages. These flavonols accumulated to high levels and provide immature fruits protections against pathogenic attacks [35]. Pharmacological researches have proved that they have strong activities of anti-inflammation, analgesia, hepatic protection and wound healing [63, 64]. Hence, the unripe fruits of *R. chingii* are the perfect resources of flavonols for health care and pharmaceutical applications.

Biosynthesis of flavonols derived from pathways of shikimic acid, phenylpropanoid biosynthesis, and flavonoid biosynthesis. PAL is the first committed enzyme in phenylpropanoid biosynthesis catalyzing the formation of trans-cinnamic acid from Phe. The *PAL* gene family in *A. thaliana* comprises four genes, and *AtPAL1* and *AtPAL2* play the principle roles in flavonoids synthesis [65]. *RiPAL1* in raspberry was associated with early fruit ripening: its expression reached maximum level at Fruits I stage (green and cell expansion), then declined at Fruits II (green but reached mature size) and Fruits III (yellow) stages, and certainly recovered at Fruits IV (ripe) stage [66]. However, expression of *RiPAL2* correlated more with later stages of fruit development. In addition, *RiPAL2* transcripts were 5.6-fold more abundant than those of *RiPAL1*. All these results are strongly consistent with our findings. So we hypothesize that the two *PAL* genes might be controlled by different regulatory mechanism. Blastx results showed that *R. chingii* PAL1 and PAL2 were more than 96% identical with *RiPAL1* (710 amino acids) and *RiPAL2* (730 amino acids) (Table S4), indicating the conservative evolution of PAL in *Rubus*. The information of *C4H*, *C4L*, *CHS*, *CHI*, *F3H*, *F3'H*, *FLS*, *DFR*, and *ANS* genes from *R. chingii* was all clarified in this study (Table S4). What's more, glycosylation is essential for stable accumulation of flavonols. *Arabidopsis* UGT78D2 is one of the flavonoid 3-*O*-

glucosyltransferases (FGTs), belonging to family 1 glycosyltransferases (UGTs), and involves in both of flavonol and anthocyanin glucosylation [65]. And Routaboul et al. [67] has verified that UGT78D2 solely catalyzes the addition of glucose moiety on aglycone kaempferol or quercetin. The sequence of UGT78D2 (Unigene7678_All) in *R. chingii* was identified in this study and the expression pattern was consistent with *PAL* genes. The blast result also showed high similarity (85.1%) between Unigene7678_All and *Rubus* hybrid cultivar 'Arapaho' UDP glucosyltransferase (UGT78H2). Chen et al. [68] observed most transcripts of UGT78H2 were abundant in the immature fruits of blackberry 'Arapaho', and then transcription levels dramatically decreased with fruit maturation. These results confirmed the role of flavonol 3-*O*-glucosyltransferase in kaempferol glucosylation. Subsequently, for synthesis of kaempferol-3-*O*- β -D-rutinoside, rhamnosyltransferase might be needed for the addition of rhamnose moiety. Two clusters of unigenes, CL2419.Contig1/3/4/5_All and Unigene428_All, encoding putative UDP-rhamnose:rhamnosyltransferase 1, were found in transcriptome library, but their expression pattern was on the opposite side of UGT78D2/78H2 and strongly expressed at red stage. So the integrated view of kaempferol-3-*O*- β -D-rutinoside biosynthesis still needs to be further tested.

The structural genes of the enzymes involved in phenolic acids and flavonols formation are predominantly regulated by transcriptional regulators containing MYB or bHLH domains and conserved WD repeats. *Zea mays* ZmC1(COLORED ALERONE1) is the first MYB-type TF found in plants and co-expression *ZmC1* and *ZmLC* (MYC/bHLH -type TF) in tomato fruit clearly enhanced the expression of *PAL*, *F3H*, *F3'H*, *FLS*, *GT* (flavonol-3-glucosyltransferase) and *RT* (flavonol-3-glucoside-rhamnosyltransferase), resulting in significant accumulations of kaempferol (up to 60-fold increase) and its glycosides, e.g. kaempferol-3-*O*-rutinoside [69]. In *A. thaliana*, the flavonols branch is under the transcriptional control of PRODUCTION OF FLAVONOL GLYCOSIED (PFG), including PFG1/MYB12, PFG2/MYB11 and PFG3/MYB111 [53,54,70]. The PFGs specifically recognized the MRE motif (a MYB recognition element) and primarily activated the promoters of *CHS*, *F3H*, and *FLS*, except for *F3'H* or *DFR* promoter, leading to strong accumulations of total quercetins and kaempferols [71]. They also controlled the expressions of GTs, flavonol-3-*O*-rhamnosyltransferase (UGT78D1), flavonoid-3-*O*-GT (UGT78D2), flavonol 7-*O*-rhamnosyltransferase (UGT89C1), and UDP-glycosyltransferase 91A1 (UGT91A1), thereby leading to additive production of three major flavonol glycosides (kaempferol 3-*O*-rhamnoside-7-*O*-rhamnoside, kaempferol 3-*O*-glucoside-7-*O*-rhamnoside, kaempferol 3-*O*-glucoside-rhamnoside - 7-*O*-rhamnoside) [53]. And the effects of the three R2R3-MYBs exhibited tissue and developmental dependence [53]. Further, in tomato, Fernandez-Moreno et al. [54] performed MYB-binding sites (MBSs) searches using FUZZNUC, MEME, and DREME and screened 141 candidate genes as potential direct targets of SIMYB12, including *DAHPS1*, *PAL2*, *4CL3*, *CHS1*, *CHS2*, *F3H*, *F3'H*, among others. In our study, RucC1, RucMYB12-1, and RucMYB12-2 showed initially high expression levels and coincident changes with key genes, such as *DAHPS1*, *DAHPS2*, *PAL1*, *PAL2*, *C4L1 ~ 3*, *CHS*, *CHI*, *FLS1* and *UGT78D2*, in shikimate, phenylpropanoids and flavonols synthetic pathways, confirming the potential influence of the flavonol-specific activators in *R. chingii*. In general, the MYB proteins are characterized by the conserved sequences at N-terminal DNA-binding domains, called MYB repeats (R), and the variable C-terminal region controls the regulatory activity [55]. Hence, we carried out sequence analysis and found R2 to R3 domain of the predicted RucC1 protein has 75.86% similarity with C1 protein in *Amborella trichopoda*, indicating the sequence and functional homology between the C1 proteins. And the deduced protein sequence of RucMYB12-1 was closely related to FaMYB11 in *Fragaria × ananassa* and MdMYB12 × 1 in *Malus domestica*, with 97.87% and 87.34% identities in the R2R3 domain (14 to 61 amino acid and 67 to 112 amino acid), whereas the complete protein sequence showed only 77% and 52% identities. On the other hand, the amino acids of RucMYB12-2 were highly homologous with Fv WER-like (87%) and PaMYB114-like in *Prunus avium*. WEREWOLF (WER), an initial regulator of root hair pattern formation, also controlled the flowering time in *A. thaliana* [72]. MYB114 was identified to involve in anthocyanin biosynthesis [55,

73]. Therefore, RucMYB12-1 and RucMYB12-2 may act in different ways and their functions need to be further verified. What's more, MYB1, MYB4, MYB5 and MYB6 are homologous regulatory genes and they have been found to play positive or repressed regulations in flavonoids biosynthesis in different species [31, 74]. In this study, RucMYB4, RucMYB5 and RucMYB1R1-1 might be positively related to flavonols production, while RucMYB6 and RucMYB1R1-2 (R3-type) were mainly in correlation with the anthocyanins synthesis. In *F. vesca*, MYB39 and MYB86 were down-regulated in yellow fruit, in comparison with red fruit, reflecting their functional involvement in flavonoids biosynthesis [75]. Our results revealed the strong activations of RucMYB86-1, RucMYB86-2 and RucMYB86-3 at BG stages in *R. chingii*. Unexpectedly, the transcripts of MYB44, MYB46, MYB308 and MYB308-like were also abundant in *R. chingii* fruits at BG stage and then changed under the same trend as the expressions of the core genes in flavonols biosynthesis pathway. In addition, the action of MYB was independent or in combination with bHLH and WD40, as the MBW complex [55, 73]. The regulation of AtMYB12 in flavonol synthesis did not depend on bHLH coactivators [71]. However, in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*), MYB12 interacted with bHLH3 and bHLH33 and essentially control proanthocyanidin synthesis, while MYB22 independently activated flavonol pathway by direct combination with FLS [76]. *PsCHS* expression in tree peony could be directly activated by the BMW complex of PsMYB12, bHLH, and WD40 protein [77]. In this study, bHLH13, bHLH63, bHLH93 and bHLH137 might participate in the synthesis of flavonols, whereas bHLH62, bHLH77, bHLH96 and bHLH130 might regulate anthocyanins biosynthesis. bHLH62 have been reported to be related to the anthocyanin biosynthesis in eggplant [78]. However, the roles of the other bHLH-type TFs for fruit development were first found in this paper and their action modes should be further studied. Nonetheless, the regulations of transcription factors were tightly associated with environmental and internal conditions, such as light, temperature, plant hormones, among others. The discoveries of the genes and regulators will provide vital supports to harness maximum nutraceutical and medicinal potential of *R. chingii* via agricultural management strategies and genetic engineering.

In summary, combined transcriptomic and metabolic analysis initially reveals molecular and chemical mechanism of fruit development of *Rubus chingii* Hu fruit. The fruit launched the biosynthesis of phenolic acids and flavonols at the very beginning of fruit set and then coordinately accumulated and converted. And it was tightly regulated by expressions of the related genes and transcription factors. Unveiling the genes and metabolites information will accelerate genetic analysis and molecular breeding for the unique species and then harness maximum nutraceutical and medicinal potential via improved agricultural management.

Abbreviations

EA: ellagic acid; K-3-R: kaempferol-3-O-rutinoside; DPA: days post-anthesis; SG: 7 days post-anthesis; MG: medium green; BGI: big green I; BGII: big green II; BGIII: big green III; GY: green-to-yellow; YO: yellow-to-orange; Re: red; NT: NCBI nucleotide sequence database; NR: non-redundant protein database; KOG: EuKaryotic Orthologous Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; TF: transcription factor; bHLH: basic helix-loop-helix; FPKM: fragments per kilobase per million mapped fragments; DEGs: differential expression genes; UPLC: ultra performance liquid chromatography; PCA: principal component analysis; PLS-DA: partial least squares discriminant analysis; TP: total phenolic; TT: total terpenoids; *pos*: positive ion additive mode; *neg*: negative ion additive mode; E4P: erythrose 4-phosphate; PEP: phosphoenolpyruvate; DAHPS: 3-deoxy-7-phosphoheptulonate synthase; DHQS: 3-dehydroquinase synthase; DQD/SDH: dehydroquinase dehydratase/shikimate dehydrogenase; DHQ: 3-dehydroquinic acid; GA: gallic acid; ETs: ellagitannins; SA: shikimic acid; Phe: phenylalanine; PAL: phenylalanine ammonialyase; C4H: cinnamic acid 4-hydroxylase; CYP73A: a cytochrome P450 monooxygenase; C4L: 4-coumaric acid: CoA ligase; CHS: chalcone synthase; CHI: chalcone isomerase; F3H: flavanone-3-hydroxylase; F3'H: flavonoid 3'-hydroxylase; FLS: flavonol synthase; UGT78D2: flavonoid 3-O-glucosyltransferase; UGT78H2: *Rubus* hybrid cultivar 'Arapaho' UDP

glucosyltransferase; DFR: dihydroflavonol 4-reductase; ANS: anthocyanidin synthase; LDOX: leucoanthocyanidin dioxygenase; UGT88A1: anthocyanidin 5,3-O-glucosyltransferase.

Declarations

Funding

This work was funded by Basic Public Welfare Projects of Zhejiang Province (2017C32082, LGN19C150004), Natural Science Foundation of Zhejiang Province Q18D050003) and Taizhou Science and Technology Project (1801ny10, 1801ny06).

Authors' contributions

ZC designed and performed the experiments, analyzed and discussed the results, wrote the manuscript. JJ designed the experiments and cultivated the plants. LS, revised the manuscript; XL and BQ analyzed the data. XW, XL, HX and JH performed the experiments. All authors have read and approved the manuscript.

Acknowledgments

We thank Youcun Chen of Huahan Raspberry Professional Cooperative for field management. We thank our undergraduate students Jifen Zhang, Ting Feng, Zhenpeng Lin, Yuting Pang, Jianjie Lin, Pihui Li, Yang Lin, Xuefen Wang and Lingjia Yao for helping material collections and experiments performance.

Availability of data and material

The raw sequencing reads of transcriptome data in this study are available in Sequence Read Archive (SRA) database, with the accession number PRJNA671545. All data generated or analyzed during this study are included in this published article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Yu G H, Luo ZQ, Wang WB, Li YH, Zhou YT, Shi YY. *Rubus chingii* Hu: A review of the phytochemistry and pharmacology. *Front. Pharmacol.* 2019;10:799.
2. Tanaka T, Kawamura K, Kitahara T, Kohda H, Tanaka O. *ent*-Labdane-type diterpene glucosides from leaves of *Rubus chingii*. *Phytochemistry*. 1984;23:615-621.
3. Chou WH, Oinaka T, Kanamaru F, Mizutani K, Chen FH, Tanaka O. Diterpene glucosides from leaves of *Rubus chingii* and fruits of *R. suavissimus*, and identification of the source plant of the Chinese folk medicine "Fu-pen-

- zi". Chem Pharm Bull. 1987;35:3021-3024.
4. Hattori M, Kuo KP, Shu YZ, Tezuka Y, Kikuchi T, Nambe T. A triterpene from the fruits of *Rubus chingii*. Phytochemistry. 1988;27:3975-3976.
 5. Chinese Pharmacopoeia Commission. Fupenzi, *Rubi* fructus. Pharmacopoeia of the People's Republic of China (Edition 2015). Beijing: China Medical Science Press; 2015. P. 382.
 6. Yau MH, Che CT, Liang SM, Kong YC, Fong WP. An aqueous extract of *Rubus chingii* fruits protects primary rat hepatocytes against *tert*-butyl hydroperoxide induced oxidative stress. Life Sci. 2002;72: 329-338.
 7. Ding HY. Extracts and constituents of *Rubus chingii* with 1,1-diphenyl-2-picrylhydrazyl (DPPH) free radical scavenging activity. Int J Mol Sci. 2011;12:3941-3949.
 8. Zeng HJ, Liu Z, Wang YP, Yang D, Yang R, Qu LB. Studies on the anti-aging activity of a glycoprotein isolated from Fupenzi (*Rubus chingii* Hu.) and its regulation on klotho gene expression in mice kidney. Int J Biol Macromol. 2018;119:470-476.
 9. Su XH, Duan R, Sun YY, Wen JF, Kang DG, Lee HS, Cho KW, Jin SN. Cardiovascular effects of ethanol extract of *Rubus chingii* Hu (Rosaceae) in rats: an *in vivo* and *in vitro* approach. J Physiol Pharmacol. 2014;65:417-424.
 10. Liang WQ, Xu GJ, Weng D, Gao B, Zheng XF, Qian Y. Anti-osteoporotic components of *Rubus chingii*. Chem Nat Compd+. 2015;51:47-49.
 11. Chen Y, Chen ZQ, Guo QW, Gao XD, Ma QQ, Xue ZH, Ferri N, Zhang M, Chen HX. Identification of ellagitannins in the unripe fruit of *Rubus Chingii* Hu and evaluation of its potential antidiabetic activity. J Agric Food Chem. 2019;67:7025-7039.
 12. Miyasaki Y, Rabenstein JD, Rhea J, Crouch ML, Mocek UM, Kittell PE, Morgan MA, Nichols WS, Van Benschoten MM, Hardy WD, Liu GY. Isolation and characterization of antimicrobial compounds in plant extracts against multidrug-resistant *Acinetobacter baumannii*. PloS One. 2013;8: e61594.
 13. Zhang TT, Lu CL, Jiang JG, Wang M, Wang DM, Zhu W. Bioactivities and extraction optimization of crude polysaccharides from the fruits and leaves of *Rubus chingii* Hu. Carbohydr Polym. 2015;130: 307-315.
 14. Zhang TT, Wang M, Yang L, Jiang JG, Zhao JW, Zhu W. Flavonoid glycosides from *Rubus chingii* Hu fruits display anti-inflammatory activity through suppressing MAPKs activation in macrophages. J Funct Foods. 2015;18:235-243.
 15. Zhang TT, Yang L, Jiang JG. Bioactive comparison of main components from unripe fruits of *Rubus chingii* Hu and identification of the effective component. Food Funct. 2015;6:2205-2214.
 16. Zhong RJ, Guo Q, Zhou GP, Fu HZ, Wan KH. Three new labdane-type diterpene glycosides from fruits of *Rubus chingii* and their cytotoxic activities against five humor cell lines. Fitoterapia. 2015;102: 23-26.
 17. Han B, Chen J, Yu YQ, Cao YB, Jiang YY. Antifungal activity of *Rubus chingii* extract combined with fluconazole against fluconazole-resistant *Candida albicans*. Microbiol Immunol. 2016;60:82-92.
 18. Zhang XY, Li W, Wang J, Li N, Cheng MS, Koike K. Protein tyrosine phosphatase 1B inhibitory activities of ursane-type triterpenes from Chinese raspberry, fruits of *Rubus chingii*. Chin J Nat Med. 2019;17:15-21.
 19. Han N, Gu YH, Ye C, Cao Y, Liu ZH, Yin J. Antithrombotic activity of fractions and components obtained from raspberry leaves (*Rubus chingii*). Food Chem. 2012;132:181-185.
 20. Foster TM, Bassil NV, Dossett M, Worthington ML, Graham J. Genetic and genomic resources for *Rubus* breeding: a roadmap for the future. Hortic Res. 2019;6:116.
 21. McDougall GJ, Martinussen I, Junttila O, Verrall S, Stewart DJ. Assessing the influence of genotype and temperature on polyphenol composition in cloudberry (*Rubus chamaemorus* L.) using a novel mass

- spectrometric method. J Agric Food Chem. 2011;59:10860–10868.
22. Jaakkola, M.; Korpelainen, V.; Hoppula, K.; Virtanena, V. Chemical composition of ripe fruits of *Rubus chamaemorus* L. grown in different habitats. J Sci Food Agric. 2012;92:1324–1330.
 23. Ponder A, Hallmann E. The effects of organic and conventional farm management and harvest time on the polyphenol content in different raspberry cultivars. Food Chem. 2019;301:125295.
 24. Wang Y, Chen Q, Chen T, Tang H, Liu L, Wang XR. Phylogenetic insights into Chinese *Rubus* (Rosaceae) from multiple chloroplast and nuclear DNAs. Front Plant Sci. 2016;7:Article 968.
 25. VanBuren R, Bryant D, Bushakra JM, Vining KJ, Edger PP, Rowley ER, Priest HD, Michael TP, Lyons E, Filichkin SA, Dossett M, Finn CE, Bassil NV, Mockler TC. The genome of black raspberry (*Rubus occidentalis*). Plant J. 2016;87:535–547.
 26. VanBuren R, Wai CM, Colle M, Wang J, Sullivan S, Bushakra JM, Liachko I, Vining KJ, Dossett M, Finn CE, Jibrán R, Chagné D, Childs K, Edger PP, Mockler TC, Bassil NV. A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. GigaScience. 2018;7: 1-9.
 27. Hyun TK, Lee S, Kumar D, Rim Y, Kumar R, Lee SY, Lee CH, Kim JY. RNA-seq analysis of *Rubus idaeus* cv. Nova: transcriptome sequencing and de novo assembly for subsequent functional genomics approaches. Plant Cell Rep. 2014;33:1617–1628.
 28. Hyun TK, Lee S, Rim Y, Kumar R, Han X, Lee SY, Lee CH, Kim JY. De-novo RNA sequencing and metabolite profiling to identify genes involved in anthocyanin biosynthesis in Korean black raspberry (*Rubus coreanus* Miquel). PloS On. 2014;9:e88292.
 29. Lee J, Dossett M, Finn C. Mistaken identity: clarification of *Rubus coreanus* Miquel (Bokbunja). Molecules. 2014;19:10524–10533.
 30. Garcia-Seco D, Zhang Y, Gutierrez-Mañero FJ, Martin C, Ramos-Solano B. RNA-Seq analysis and transcriptome assembly for blackberry (*Rubus* sp. Var. Lochness) fruit. BMC Genomics. 2015;16:5.
 31. Gutierrez E, García-Villaraco A, Lucas JA, Gradillas A, Gutierrez-Mañero FJ, Ramos-Solano B. Transcriptomics, targeted metabolomics and gene expression of blackberry leaves and fruits indicate flavonoid metabolic flux from leaf to red fruit. Front Plant Sci. 2017; 8:Article 472.
 32. Chen Q, Liu XJ, Hu YY, Sun B, Hu YD, Wang XR, Tang HR, Wang Y. Transcriptomic profiling of fruit development in black raspberry *Rubus coreanus*. Int J Genomics. 2018;2018:8084032.
 33. Sharma S, Kaur R, Solanke AK, Dubey H, Tiwari S, Kumar K. Transcriptome sequencing of Himalayan Raspberry (*Rubus ellipticus*) and development of simple sequence repeat markers. 3 Biotech. 2019;9:161.
 34. Thole V, Bassard JE, Ramírez-González R, Trick M, Afshar BG, Breitel D, Hill L, Foito A, Shepherd L, Freitag S, dos Santos CN, Menezes R, Bañados P, Naesby M, Wang L, Sorokin A, Tikhonova O, Shelenga T, Stewart D, Vain P, Martin C. RNA-seq, de novo transcriptome assembly and flavonoid gene analysis in 13 wild and cultivated berry fruit species with high content of phenolics. BMC Genomics. 2019;20:995.
 35. Fait A, Hanhineva K, Beleggia R, Dai N, Rogachev I, Nikiforova VJ, Fernie AR, Aharoni A. Reconfiguration of the achene and receptacle metabolic networks during strawberry fruit development. Plant Physiol. 2008;148:730-750.
 36. Suh DH, Jung ES, Lee GM, Lee CH. Distinguishing six edible berries based on metabolic pathway and bioactivity correlations by non-targeted metabolite profiling. Front Plant Sci. 2018;9:1462.
 37. Zou XY, Niu WQ, Liu JJ, Li Y, Liang BH, Guo LL, Guan YH. Effects of residual mulch film on the growth and fruit quality of tomato (*Lycopersicon esculentum* Mill.). Water Air Soil Pollut. 2017;228: 71.

38. Chen Q, Yu H, Tang H, Wang X. Identification and expression analysis of genes involved in anthocyanin and proanthocyanidin biosynthesis in the fruit of blackberry. *Sci Hortic.* 2012;141:61–68.
39. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644-652.
40. Pertea G, Huang XQ, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee YD, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics.* 2003;19:651-652.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-410.
42. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005; 21:3674-3676.
43. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005;33 (Web Server issue):W116-20.
44. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 2000;16:276-277.
45. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013;41:e121.
46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357-359.
47. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
48. Michael I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
49. Wen B, Mei ZL, Zeng CW, Liu SQ. metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics.* 2017;18:183.
50. Deng B, Fang SZ, Shang XL, Fu XX, Li Y. Influence of provenance and shade on biomass production and triterpenoid accumulation in *Cyclocarya paliurus*. *Agroforest Syst.* 2019;93:483–492.
51. Guo J, Carrington Y, Alber A, Ehrling J. Molecular characterization of quinate and shikimate metabolism in *Populus trichocarpa*. *J Biol Chem.* 2014;289:23846–23858.
52. Lu ZW, Liu YJ, Zhao L, Jiang XL, Li MZ, Wang YS, Xu YJ, Gao LP, Xia T. Effect of low-intensity white light mediated de-etiolation on the biosynthesis of polyphenols in tea seedlings. *Plant Physiol Biochem.* 2014;80:328-336.
53. Stracke R, Jahns O, Keck M, Tohge T, Niehaus K, Fernie AR, Weisshaar B. Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 2010;188:985–1000.
54. Fernandez-Moreno JP, Tzfadia O, Forment J, Presa S, Rogachev I, Meir S, Orzaez D, Aharoni A, Granell A. Characterization of a new pink-fruited tomato mutant results in the identification of a null allele of the SIMYB12 transcription factor. *Plant Physiol.* 2016;171:1821–1836.
55. Mondal SK, Roy S. Genome-wide sequential, evolutionary, organizational and expression analyses of phenylpropanoid biosynthesis associated MYB domain transcription factors in *Arabidopsis*. *J Biomol Struct*

Dyn. 2018;36:1577-1601.

56. Bushakra JM, Stephens MJ, Atmadjaja AN, Lewers KS, Symonds VV, Udall JA, Chagné D, Buck EJ, Gardiner SE. Construction of black (*Rubus occidentalis*) and red (*R. idaeus*) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theor Appl Genet.* 2012;125: 311–327.
57. Potter D, Eriksson T, Evans RC, Oh S, Smedmark JEE, Morgan DR, Kerr M, Robertson KR, Arsenault M, Dickinson TA, Campbell CS. Phylogeny and classification of Rosaceae. *Plant Syst Evol.* 2007;266:5–43.
58. Belwal T, Pandey A, Bhatt ID, Rawal RS, Luo ZS. Trends of polyphenolics and anthocyanins accumulation along ripening stages of wild edible fruits of Indian Himalayan region. *Sci Rep.* 2019;9: 5894 .
59. Chen Y, Xu LL, Wang YJ, Chen ZQ, Zhang M, Panichayupakaranant P, Chen HX. Study on the active polyphenol constituents in differently colored *Rubus Chingii* Hu and the structure-activity relationship of the main ellagitannins and ellagic acid. *LWT-Food Sci Technol.* 2020;121:108967.
60. Yang J, Ji LL, Wang XF, Zhang Y, Wu LZ, Yang YN, Ma ZY. Overexpression of 3-deoxy-7-phosphoheptulonate synthase gene from *Gossypium hirsutum* enhances *Arabidopsis* resistance. *Plant Cell Rep.* 2015;34:1429–1441.
61. Singh SA, Christendat D. Structure of *Arabidopsis* dehydroquinate dehydratase shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway. *Biochemistry.* 2006;45:7787-7796.
62. Sun P, Zhang GY, Xiang P, Lin JK, Lai ZX. Expression and regulation of the shikimic acid pathway gene DHD/SDH in tea plant (*Camellia sinensis*). *Chin J Appl Environ Biol.* 2018;24: 322-327. (in Chinese)
63. Parveen Z, Deng YL, Saeed MK, Dai RJ, Ahamard W, Yu YH. Antiinflammatory and analgesic activities of *Thesium chinense* Turcz extracts and its major flavonoids, kaempferol and kaempferol-3-O-glucoside. *Yakugaku Zasshi.* 2007;127:1275-1279.
64. Wang Y, Tang CY, Zhang H. Hepatoprotective effects of kaempferol 3-O-rutinoside and kaempferol 3-O-glucoside from *Carthamus tinctorius* L. on CCl₄-induced oxidative liver injury in mice. *J Food Drug Anal.* 2015;23:310-317.
65. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, Fernie AR. The flavonoid biosynthetic pathway in *Arabidopsis*: Structural, and genetic diversity. *Plant Physiol Biochem.* 2013;72:21-34.
66. Kumar A, Ellis BE. The phenylalanine ammonia-lyase gene family in raspberry. Structure, expression and evolution. *Plant Physiol.* 2001;127:230-239.
67. Routaboul JM, Dubos C, Bech G, Marquis C, Bidzinski P, Loudet O, Lepiniec L. Metabolite profiling and quantitative genetics of natural variation for flavonoids in *Arabidopsis*. *J Exp Bot.* 2012;63: 3749-3764.
68. Chen Q, Jiang LY, Wang Y, Zhang YT, Wang XR, Tang HR. Characterization of the glycosyltransferase gene *UGT78H2* from blackberry and docking with flavonoid molecules. *Bulletin Botanical Res.* 2015; 35:270-278. (in Chinese)
69. Bovy A, de Vos R, Kemper M, Schijlen E, Almenar Pertejo M, Muir S, Collins G, Robinson S, Verhoeyen M, Hughes S, Santos-Buelga C, van Tunen A. High-flavonol tomatoes resulting from the heterologous expression of the maize transcription factor genes LC and C1. *Plant Cell.* 2002;14: 2509-2526.
70. Zhai R, Zhao YX, Wu M, Yang J, Li XY, Liu HT, Wu T, Liang FF, Yang CQ, Wang ZG, Ma FW, Xu LF. The MYB transcription factor PbMYB12b positively regulates flavonol biosynthesis in pear fruit. *BMC Plant Biol.* 2019;19:85
71. Mehrtens F, Kranz H, Bednarek P, Weisshaar B. The *Arabidopsis* transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol.* 2005;138:1083–1096.

72. Seo E, Yu J, Ryu KH, Lee MM, Lee I. WEREWOLF, a regulator of root hair pattern formation, controls flowering time through the regulation of FT mRNA stability. *Plant Physiol.* 2011;156: 1867–1877.
73. Yao GF, Ming M, Allan AC, Gu C, Li L, Wu X, Wang R, Chang Y, Qi K, Zhang S, Wu J. Map-based cloning of the pear gene MYB114 identifies an interaction with other transcription factors to coordinately regulate fruit anthocyanin biosynthesis. *Plant J.* 2017;92:437–451.
74. Wang XC, Wu J, Guan ML, Zhao CH, Geng P, Zhao Q. *Arabidopsis* MYB4 plays dual roles in flavonoid biosynthesis. *Plant J.* 2020;101:637-652.
75. Zhang YC, Li WJ, Dou YJ, Zhang JX, Jiang GH, Miao LZ, Han GF, Liu YX, Li H, Zhang ZH. Transcript quantification by RNA-Seq reveals differentially expressed genes in the red and yellow fruits of *Fragaria vesca*. *PLoS One.* 2015;10:e0144356.
76. Wang N, Xu HF, Jiang SH, Zhang ZY, Lu NL, Qiu HR, Qu CZ, Wang YC, Wu SJ, Chen XS. MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant J.* 2017;90:276-292.
77. Gu ZY, Zhu J, Hao Q, Yuan YW, Duan YW, Men SQ, Wang QY, Hou QZ, Liu ZA, Shu QY, Wang LS. A Novel R2R3-MYB transcription factor contributes to petal blotch formation by regulating organ-specific expression of *PsCHS* in tree peony (*Paeonia suffruticosa*). *Plant Cell Physiol.* 2019;60: 599-611.
78. Zhang SM, Zhang AD, Wu XX, Zhu ZW, Yang ZF, Zhu YL, Zha DS. Transcriptome analysis revealed expression of genes related to anthocyanin biosynthesis in eggplant (*Solanum melongena* L.) under high-temperature stress. *BMC Plant Biol.* 2019;19:387.

Figures

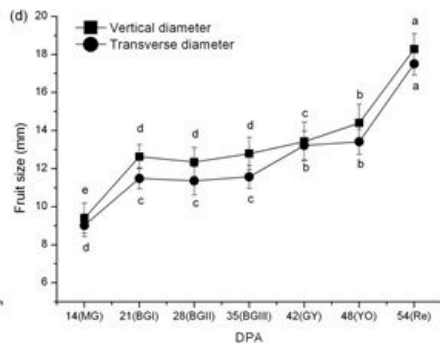
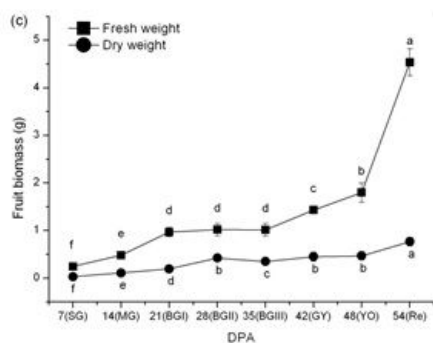
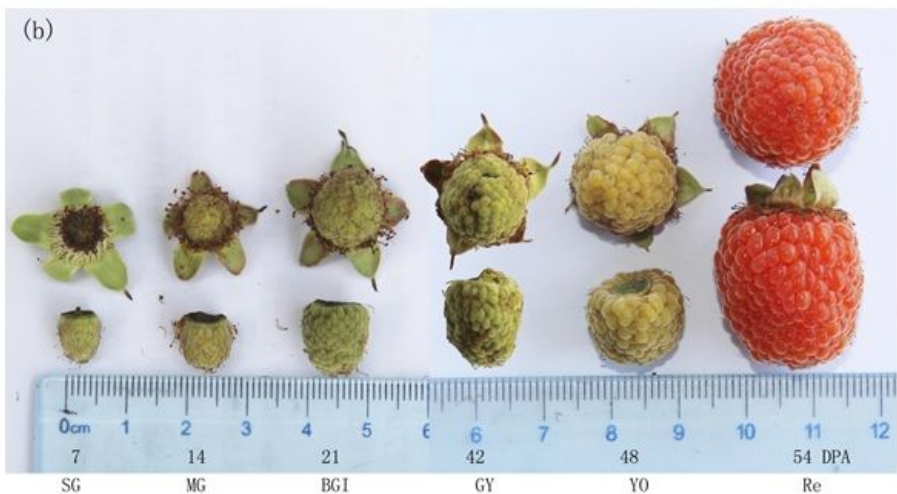


Figure 1

Development of *Rubus chingii* Hu. (a) Leaf, flower, ripe fruit, unripe fruit and dry fruit for medicinal raw material. (b) The six distinguished stages of fruit development: small green (SG), medium green (MG), big green I (BGI), green turning to yellow (GY), yellow turning to orange (YO), and red (Re). (c) Fruit biomass (c) and size (d) records during the developmental process.

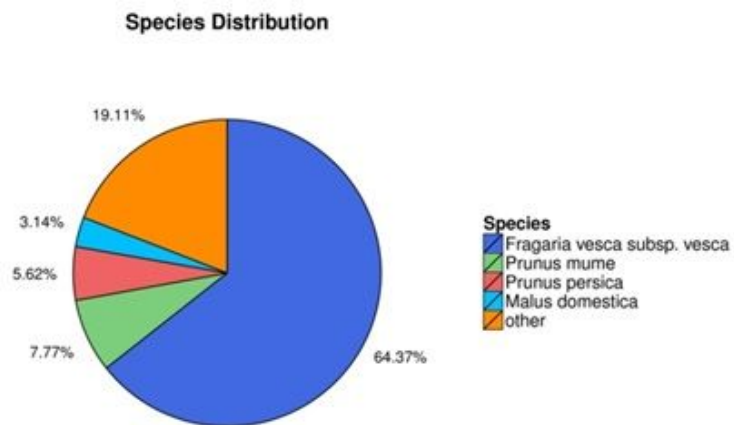


Figure 2

Distribution of NR annotated species.

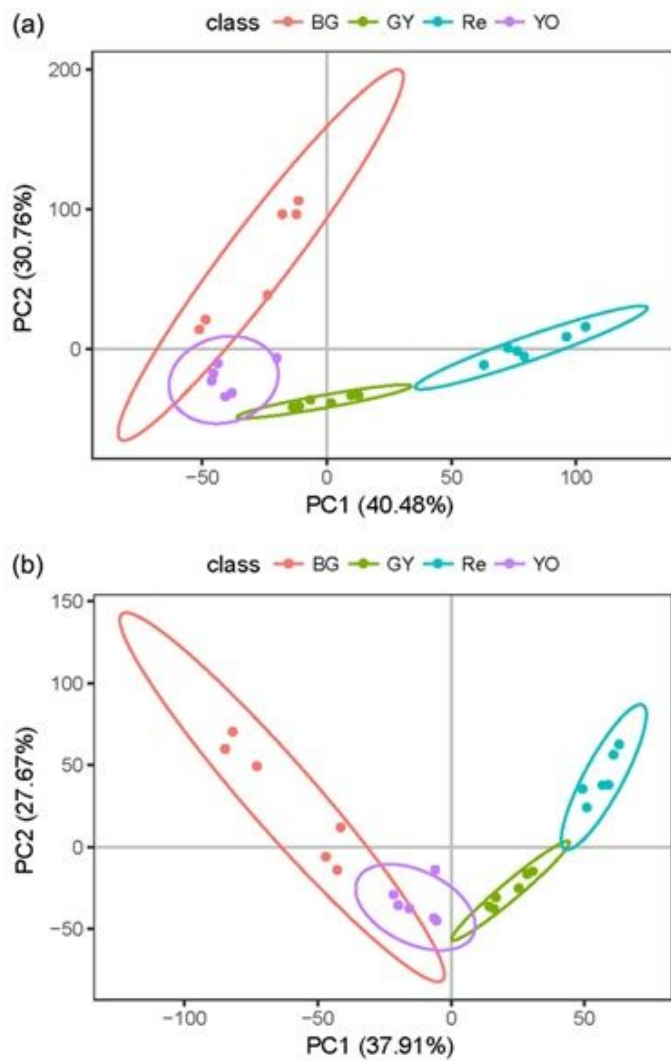


Figure 3

PCA Score plots of metabolite profiles in four different stages of *R. chingii* fruits by UPLC-Q-TOF-MS (a) Positive ion additive mode; (b) negative ion additive mode

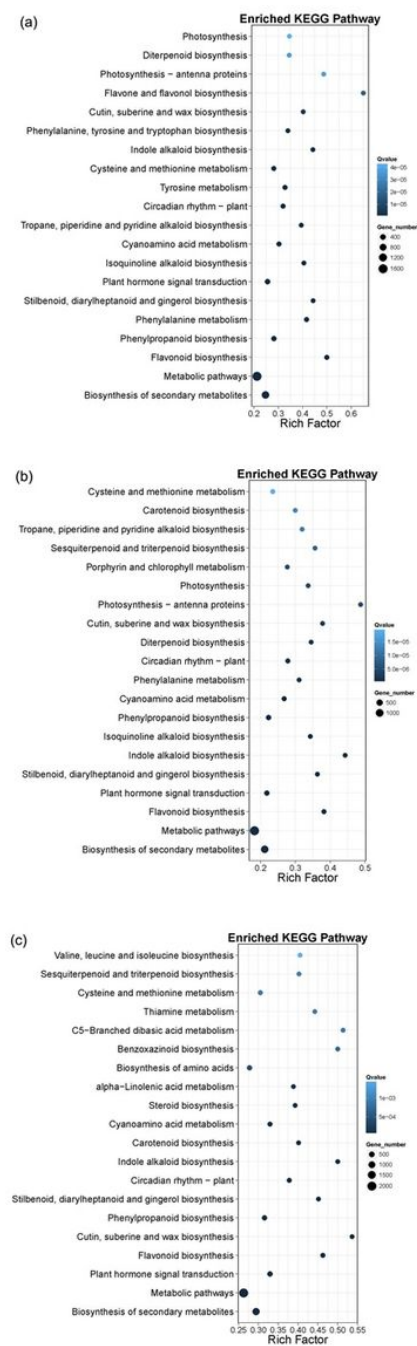


Figure 4

Pathway functional enrichments of differential expression genes between stages (a) green-yellow stage vs big green I stage (GY:BGI); (b) yellow-to-orange stage vs green-to-yellow (YO:GY); (C) red stage vs yellow-to-orange stage (Re:YO).

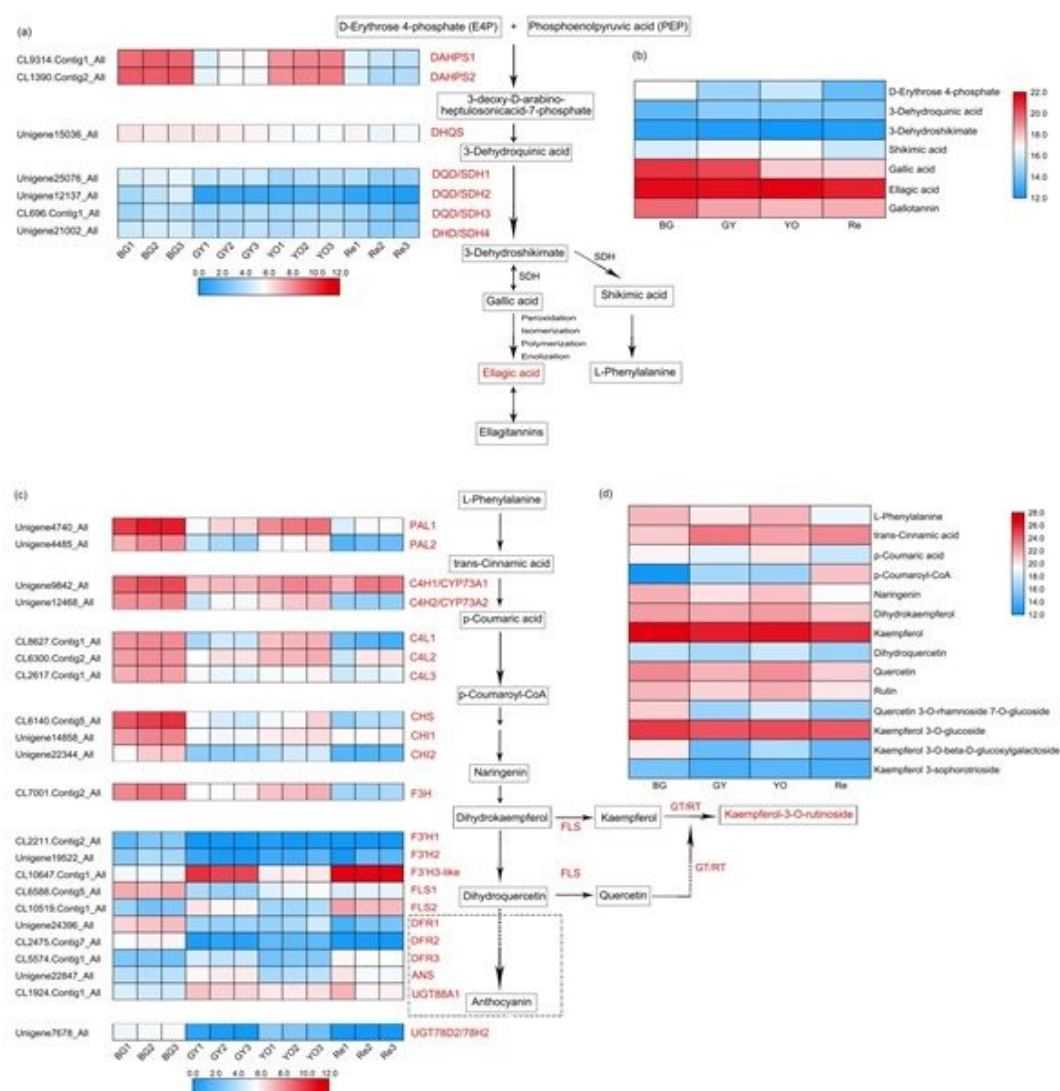


Figure 5

The relative gene expressions in biosynthetic pathway of ellagic acid and kaempferol-3-O-rutinoside by RNA-seq analysis and subsequent metabolites amounts detected by metabolomics of *R. chingii* fruits from four development stages. Enzyme names were abbreviated as follows: DAHPS (3-deoxy-7-phosphoheptulonate synthase), DHQS (3-dehydroquinate synthase), DHD/SDH (or DHQ/SDH, bifunctional 3-dehydroquinate dehydratase / shikimate dehydrogenase), PAL (phenylalanine ammonia-lyase), C4H/CYP73A (cinnamic acid 4-hydroxylase / a cytochrome P450 monooxygenase CYP73A), C4L (4-coumaric acid: CoA ligase), CHS (chalcone synthase), CHI (chalcone isomerase), F3H (flavanone-3-hydroxylase), F3'H (flavonoid 3'-hydroxylase), FLS (flavonol synthase), DFR (dihydroflavonol 4-reductase), ANS (anthocyanidin synthase / LDOX leucoanthocyanidin dioxygenase), UGT88A1 (anthocyanidin 5,3-O-glucosyltransferase), UGT78D2/UGT78H2 (flavonoid 3-O-glucosyltransferase / *Rubus* hybrid cultivar 'Arapaho' UDP glucosyltransferase), GT (Glycosyltransferase), RT (rhamnosyltransferase). The same as follows. (a) Expression heat map of the genes involved in ellagic acid synthesis. (b) Heat map of metabolites in pathway of ellagic acid synthesis. (c) Expression heat map of the genes involved in kaempferol-3-O-rutinoside synthesis. (d) Heat map of metabolites in pathway of kaempferol-3-O-rutinoside synthesis.

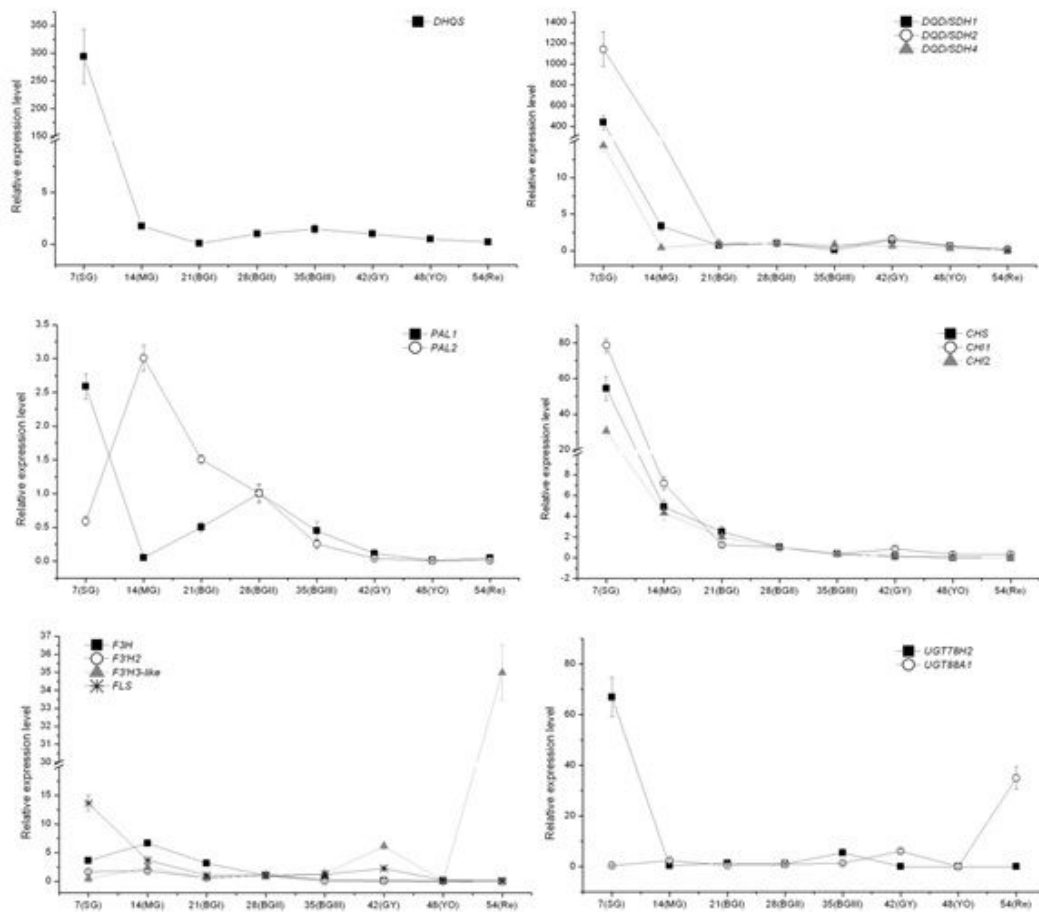


Figure 6

The relative gene expressions in biosynthetic pathway of ellagic acid and kaempferol-3-O-rutinoside by RT-qPCR assesses during the whole stages of fruit development and ripening of *R. chingii* fruits.

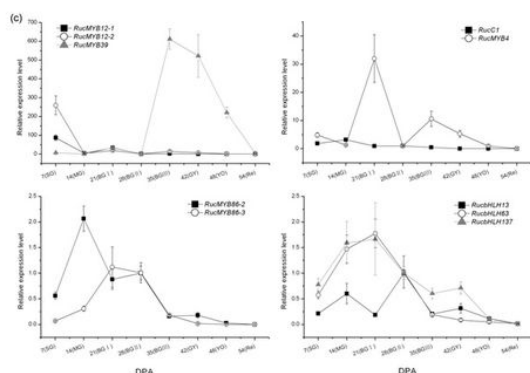
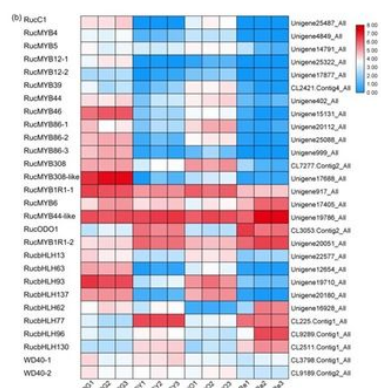
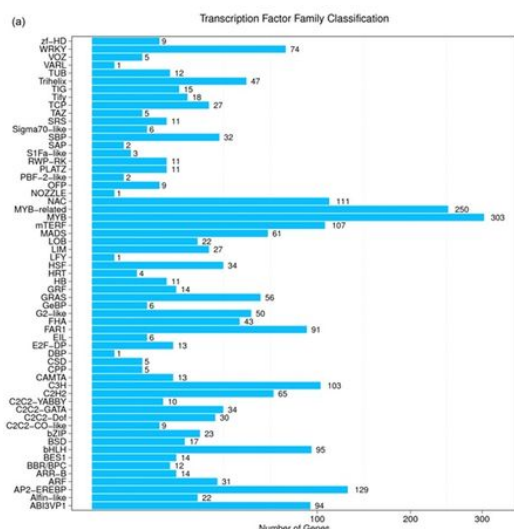


Figure 7

Prediction of transcriptions factors (a) Transcription factor family classification of unigenes. X axis indicates the number of unigenes. Y axis indicates the family of TF. (b) Expression heat map of some MYB, bHLH and WD TFs. (c) qPCR analysis of some MYB and bHLH TFs.

Supplementary Files

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

- [TableS1.pdf](#)

- [TableS2.pdf](#)
- [TableS3abcDEseq2Method.KEGG.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [SIFigures.pdf](#)