Brazilian Journal of Botany (2021) 44:685–695
https://doi.org/10.1007/s40415-021-00720-0

GENETICS & EVOLUTIONARY BIOLOGY - ORIGINAL ARTICLE

Transcriptome analysis of flower color variation in five *Rhododendron* species (Ericaceae)

Zhiliang Li[1,2] · Qiaofeng Yang[3] · Xue Dong[2] · Yu Zhu[2] · Shuang Zhao[2] · Wenying Zhang[1] · Shuzhen Wang[2]

Received: 9 December 2020 / Revised: 26 May 2021 / Accepted: 27 May 2021 / Published online: 25 August 2021

Abstract

Rhododendron genus is famous with important ecology impacts, ornamental values, and high medicine values. To obtain a comprehensive overview of anthocyanin regulatory networks, RNA-seq and de novo assembly of five *Rhododendron* species flower tissues were performed, generating 159,408 unigenes with an average length of 490 bp and an N50 of 552 bp. In particular, 106,766 unigenes could be annotated. Flavonoid biosynthesis was the most abundant KEGG pathway. Genes controlling flower color varied in different species: transcripts involving in carotenoid biosynthesis and isoflavonoid biosynthesis highly expressed in *Rhododendron molle* G. Don possessing yellow color flower; genes involved in flavone and flavonol biosynthesis showed higher abundance in *Rhododendron fortune* Lindl. with light pink flowers; transcripts of unigenes participated in flavonoid biosynthesis exhibited higher levels in *Rhododendron mariesii* Hemsl. with pink color; unigenes involved in anthocyanin biosynthesis showed higher mRNA levels in *Rhododendron simsii* Planch. with red color; genes involved in carotenoid biosynthesis and flavonoid biosynthesis showed higher expression levels in *Rhododendron pulchrum* Sweet with purplish red flowers. The five species were clustered into two main groups: group 1 (*R. fortune* and *R. mariesii*) and group 2 (*R. simsii*, *R. molle* and *R. pulchrum*) based on expression levels of differentially expressed genes (DEGs). In particular, *R. molle* showed a closer relationship with *R. simsii.* This study will provide rich genetic information for further mechanism analysis of flower color variation and genetic improvement of flower color in *Rhododendron* species.

Keywords Anthocyanin · Flavonoid biosynthesis · Genetic improvement · Metabolic network · Rhododendron genus · RNA-seq

1 Introduction

Rhododendron genus is widely distributed in the world due to showy floral displays, medicinal and horticultural value, as well as easy adaptability to various environmental conditions (Scheiber et al. 2000; Popescu and Kopp 2013; Gautam et al. 2020). As the largest genus within Ericaceae family, *Rhododendron* contains over 1000 species and tens of thousands of commercial hybrids, and about 650 species are unique to China, with the juncture of Tibetan, Yunnan, and Sichuan provinces being key distribution center of *Rhododendron* species (Wang et al. 2010, 2018). *Rhododendron* species have been introduced to Europe from southern China and gained extensive popularity, especially in England and the USA (Wang et al. 2018).

Rhododendron species breeding aims at creating novelties with improved flower characteristics, especially flower color (Nishihara and Nakatsuka 2011). *Rhododendron* flower color ranges from purple to carmine red and to red, pink, and

Zhiliang Li and Qiaofeng Yang contributed equally to this work

✉ Wenying Zhang
wyzhang@yangtzeu.edu.cn

✉ Shuzhen Wang
wangshuzhen710@whu.edu.cn

[1] College of Agriculture, Yangtze University, Jingzhou, Hubei 434000, People's Republic of China

[2] Hubei Key Laboratory of Economic Forest Germplasm Improvement and Resources Comprehensive Utilization; Hubei Collaborative Innovation Center for the Characteristic Resources Exploitation of Dabie Mountains; College of Life Sciences, Huanggang Normal University, Huanggang, Hubei 438000, People's Republic of China

[3] College of Food and Bioengineering, Henan University of Animal Husbandry and Economy, Zhengzhou, Henan 450000, People's Republic of China

even white (Keyser et al. 2013). Furthermore, *Rhododendron* flowers can be flecked, which might be caused by transposon activities (Keyser et al. 2013). Though flower color segregation in *Rhododendron* could be predicted by the Mendelian model, the model could not explain pink flowers. Four major phytopigments are responsible for the coloration of plant tissues, including chlorophylls, flavonoids, betalains, and carotenoids (Grotewold 2006). In relation to *Rhododendron* species, anthocyanin and flavonols are the major pigments, and the final color is a combination of both colored pigments (anthocyanins) and colorless flavonols (Mizuta et al. 2009). Moreover, flower color genes and putative QTLs have also been mapped in certain *Rhododendron* species (Dunemann et al. 1999; Daiki et al. 2014). However, anthocyanin biosynthesis has been rarely studied in *Rhododendron* species (Daiki et al. 2014).

RNA sequencing (RNA-seq) shows great potential for identification of genes conferring target traits, analysis of gene expression profile, development of molecular markers, revealing splice variants, mapping transcription start sites, as well as characterization of alternative splicing patterns, which have been successfully used in many species such as pummelo, sweet potato, carrot, and barley (Ozsolak and Milos 2011; Liang et al. 2015; Hill et al. 2016; Sun et al. 2019). Moreover, de novo assembly could benefit studies on non-model species, whose genome information is unavailable (Smith-Unna et al. 2016). However, transcriptome information in *Rhododendron* genus is very limited (Fang et al. 2017; Xing et al. 2017; Choudhary et al. 2018; Xiao et al. 2018).

In this study, Illumina HiSeq 2500 PE125 platform was used for RNA-seq of five *Rhododendron* species with different flower colors, including *Rhododendron fortunei* Lindl. (light pink flowers), *Rhododendron simsii* Planch. (red flowers), *Rhododendron mariesii* Hemsl. (pink flowers), *Rhododendron molle* G. Don (yellow flowers), and *Rhododendron pulchrum* Sweet (purplish red flowers). Based on differential expression genes (DEGs), critical genes involved in flower color formation were clarified, and metabolic networking was constructed. This research will be helpful in elucidating molecular mechanism of color formation and regulation, as well as benefit the genetic improvement of flower color in *Rhododendron* species.

2 Materials and methods

Plant materials – Plant materials –Blooming flowers of *R. fortunei*, *R. mariesii*, , *R. simsii* and *R.molle* were all collected from Taohuachong (Dabie mountains) (116°02′20″–116°10′53″E, 30°57′20″–31°06′10″N, 900–1000 m) (Fig. 1). Furthermore, flower tissue of *R. pulchrum* was sampled from Huanggang Botanical garden (114°55′14.36″E, 30°27′13.73″N, 22–50 m). All samples were immediately frozen by liquid nitrogen and stored at − 80 °C until further use.

RNA isolation and Illumina sequencing – Total RNA was extracted from each sample using TRIzol kit (Takara) according to the manufacturer's instruction. The pair-end cDNA sequencing libraries were constructed by Binggang Biotechnology Co. Ltd. (Wuhan, China). After being treated with DNase I, RNA samples were enriched using magnetic oligo (dT) beads. The enriched RNA was disrupted into short fragments and then was used as template to synthesize second-stand cDNA with random hexamer primers. The double-strand cDNA was purified, and single nucleotide A was added to 3' end. Finally, sequencing adapters were digested to short fragments, and PCR for 16 cycles amplification was performed to construct the cDNA libraries, which were validated by an Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR system. Finally, the cDNA libraries were sequenced using Illumina HiSeq 2500 PE125.

***De novo*assembly and functional annotation –** Raw reads were filtered to generate clean data via a process that include the removal of sequences containing adapters, low-quality sequences with more than 20% of the bases of quality value ≤ 10, and reads with N percentage over 5%. Then, the obtained clean reads were assembled using Trinity software (Dewey and Li 2011). For functional annotation, the unigenes were subjected to BLASTx alignment against the protein non-redundant database of Nr, Protein family (Pfam), Swiss-Prot protein database, Kyoto Encyclopedia of Genes and Genomes (KEGG), euKaryotic Orthologous Groups database (KOG) and were also aligned to the nucleotide sequences database (Nt) by BLASTn. Based on Nr annotation, Gene Ontology (GO) annotation of unigenes according to 'component function,' 'biological process,' and 'cellular component' ontologies was performed using Blast2GO program (Kanehisa and Goto 2000). In particular, BLASTx algorithm, with an E-value threshold of 1E-5, was used in searching Nt, Nr, and Swiss-Prot protein database. The expression levels of unigenes were calculated by mapping clean reads to the assembled transcriptome using FPKM (Fragments Per kb per Million reads) method. Then, the transcriptomes were compared with each other to clarify differentially expressed unigenes. False discovery rate (FDR) was used to calculate the threshold of p value in multiple expression analysis, and FDR ≤ 0.05 was set as the threshold. The DEGs with twofold changes were subjected to GO and KEGG pathway analysis.

Validation of RNA seq data – Validation of RNA-seq data was performed by quantitative real-time PCR (qRT-PCR) and comparative threshold (CT) method (delta-delta CT,

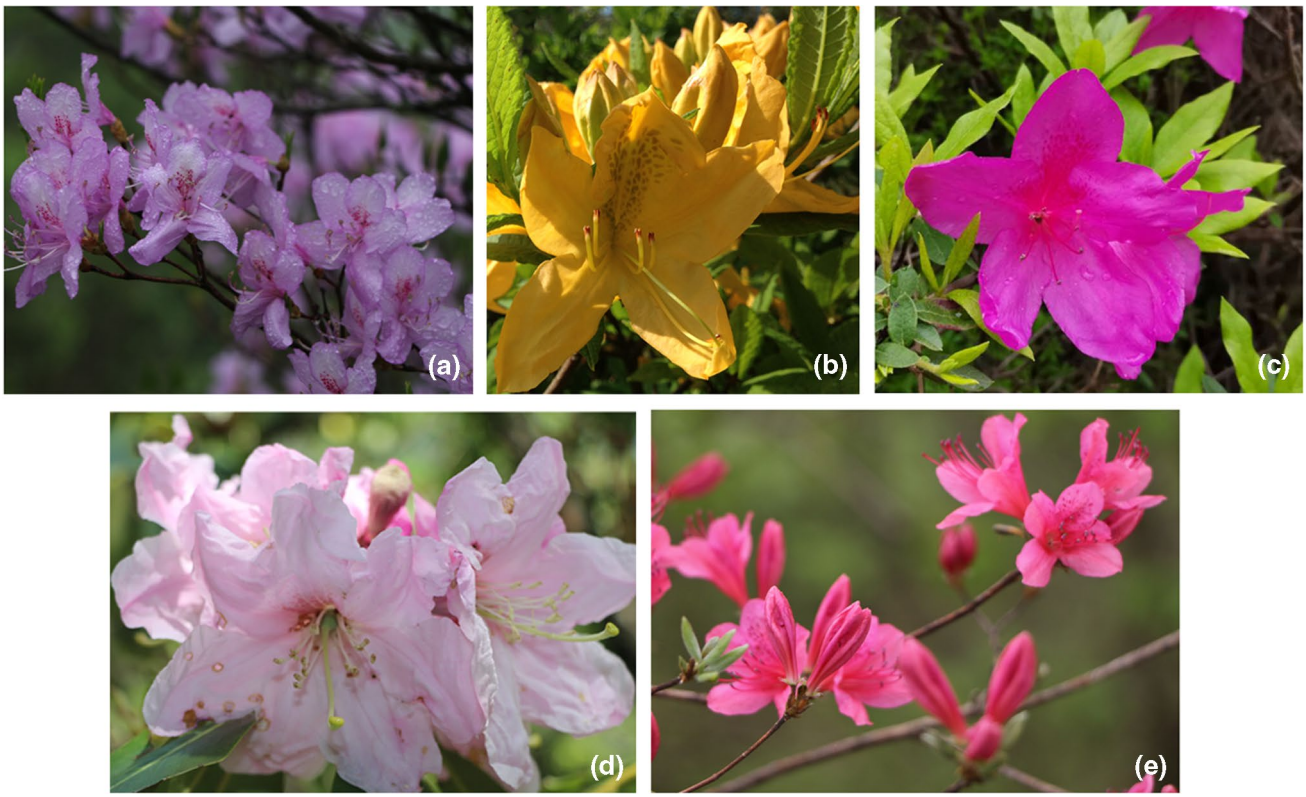

Fig. 1 Information of five *Rhododendron* species. **a** *R. mariesii*; **b** *R. molle*; **c** *R. pulchrum*; **d** *R. fortunei*; **e** *R. simsii*

ΔΔCT) according to Wang et al. (2018). Six genes were randomly selected for real-time PCR amplification in order to confirm the RNA-seq data with primer paris developed by Wang et al. (2018), including *ARR*, *CYCD*, *MYC2*, *SNRNp*, *TIR1*, and *EIN3*, with *EF-1α* gene as the internal control.

3 Results

Illumina sequencing and *de novo* assembly of *Rhododendron* species – Totally, approximately 21,006,137 sequencing reads for *R. fortunei*, 23,207,108 for *R. simsii*, 26,118,281 for *R. mariesii*, 17,411,851 for *R. molle*, and 23,983,039 for *R. pulchrum* were generated (Table 1). Moreover, 20,999,440 clean reads with 5,249,860,000 bp for *R. fortunei*, 23,201,636 clean reads with 5,800,409,000 bp for *R. simsii*, 26,109,041 clean reads with 6,527,260,250 bp for *R. mariesii*, 17,406,941 clean reads with 4,351,735,250 bp for *R. molle*, and 23,977,155 clean reads with 5,994,288,750 bp for *R. pulchrum* were obtained through filtering low-quality sequences, adaptors, and redundant reads. FastQC analysis showed that the QC20 (ratio of high-quality sequences with less than 20% of the bases of quality value ≤ 10) percentage and GC percentage were above 97 and 47%, respectively (Table 1).

High-quality reads were used to assemble transcriptome with Trinity (v2.2.0) software (Grabherr et al. 2011). Totally, 59,887 unigenes with an average length of 882.56 bp and an N50 of 1465 bp were assembled in *R. fortunei*; 92,469 unigenes with an average length of 700.39 bp and an N50 of 1465 bp were identified in *R. simsii*; 81,710 unigenes with an average length of 882.27 bp and an N50 of 1477 bp were obtained in *R. mariesii*; 58,263 unigenes with an average length of 830.26 bp and an N50 of 1372 bp were generated in *R. molle*; 101,021 unigenes with an average length of 1090 bp and an N50 of 674.85 bp were produced in *R. pulchrum* (Table 1). Among these assembled unigenes, 64.5% had lengths greater than 500 bp (Table 1). Furthermore, contigs were pooled and assembled into a non-redundant unigene set to expand the utility of Illumina sequencing data, yielding 159,408 unigenes with an average length of 490 bp and N50 of 552 bp.

Functional annotation of the unigenes – Among the 159,408 unigenes, 106,766 unigenes (66.98%) could be annotated through searching against seven databases. In total, 25,335 unigenes had annotation information in all these seven databases (Table 2, Fig. 2a). Particularly, 55,546 (34.85%), 89,984 (56.45%), and 73,386 (46.04%) unigenes showed high identities to known sequences deposited in Nt, Nr, and

Table 1 Sequence analysis of five *Rhododendron* species

category	*R. fortunei*	*R. simsii*	*R. mariesii*	*R. molle*	*R. pulchrum*
Raw_reads	21,006,137	23,207,108	26,118,281	17,411,851	23,983,039
Clean_reads	20,999,440	23,201,636	26,109,041	17,406,941	23,977,155
Clean_base	5,249,860,000	5,800,409,000	6,527,260,250	4,351,735,250	5,994,288,750
Q20 percentage (%)	98.05	97.95	97.6	97.9	98.2
GC percentage (%)	47.6	47.65	47.8	48.6	47.95
0–300	15,795 (26.37%)	31,674 (34.25%)	21,951 (26.86%)	16,148 (27.72%)	35,846 (35.48%)
300–500	12,845 (21.45%)	22,354 (24.17%)	17,605 (21.55%)	12,923 (22.18%)	25,162 (24.91%)
500–1000	12,269 (20.49%)	17,927(19.39%)	16,936 (20.73%)	12,193 (20.93%)	19,442 (19.25%)
1000–2000	12,824 (21.41%)	14,962(16.18%)	16,812 (20.58%)	11,919 (20.46%)	14,917 (14.77%)
>2000	6,153 (10.27%)	5,551(6.00%)	8,405 (10.29%)	5,079 (8.72%)	5,653 (5.60%)
Unigenes	59,887	92,469	81,710	58,263	101,021
Total length	52,853,689	64,764,610	72,090,665	48,373,333	68,174,524
N50 length	1465	1465	1477	1372	1090
Mean length	882.56	700.39	882.27	830.26	674.85

Table 2 Functional annotation of five *Rhododendron* species

Database	Number of unigenes	Percentage of all unigenes
Nt	55,546	34.85
Nr	89,984	56.45
Swiss-prot	73,386	46.04
KOG	60,040	37.66
KEGG	54,001	33.88
GO	68,131	42.74
Pfam	39,045	24.49
Annotated	106,766	66.98
Total_unigens	159,408	100

Swiss-Prot protein databases, respectively (Table 2). The unigenes with ORFs were searched against Pfam database with a hmmscan threshold of 1E-3, and 39,045 (24.49%) unigenes were similar to known proteins.

Totally, 68,131 unigenes (42.74%) showed significant similarities to known sequences in GO database: 'biological process' accounting for the largest (205,888 unigenes), followed by 'cellular component' (133,669 unigenes) and 'molecular function' (80,565 unigenes) (Fig. 2b). Under 'biological process' category, 'metabolic process' (48,133, 70.64%), 'cellular process' (41,126, 60.36%), and 'single-organism process' (33,683, 49.43%) were main groups. In 'cellular component' category, 30,347 unigenes (44.54%) were assigned both to the most abundant groups of 'cell' and 'cell part,' followed by 'organelle' (22,420, 32.9%). Moreover, 'catalytic activity' (37,474, 55%) and 'binding' (32,320, 47.43%) represented the first and second largest subcategories in 'molecular function' category, respectively.

Furthermore, 60,040 (37.66%) unigenes could be annotated to KOG database and were grouped into 25 functional clusters. Particularly, 'general function prediction' (11,151, 18.5%), 'posttranslational modification, protein turnover, chaperones' (7503, 12.4%), and 'signal transduction mechanisms' (5692, 9.4%) were the first three largest groups (Fig. S1). Moreover, 'energy production and conversion' (4508, 7.5%), 'carbohydrate transport and metabolism' (4377, 7.2%), 'translation, ribosomal structure and biogenesis' (4506, 7.5%), 'intracellular trafficking, secretion, vesicular transport' (3247, 5.4%), 'transcription' (2849, 4.7%), and 'function unknown' (3157, 5.2%) were also the main clusters.

A set of 54,001 (33.88%) unigenes was categorized into 32 pathways (five main categories) by KEGG classification (Fig. 3). The 'metabolism' (40,603, 75.1%) was the largest category containing 12 subcategories, such as 'carbohydrate metabolism' (8751, 16.2%), 'overview' (5552, 10.28%), 'amino acid metabolism' (4847, 8.97%), 'lipid metabolism' (3823, 7.07%), 'energy metabolism' (3448, 6.38%), 'xenobiotics biodegradation and metabolism' (3054, 5.65%), and other subcategories. Furthermore, categories of 'organismal systems,' 'genetic information processing,' 'environment information processing,' and 'cellular processes' accounted for 27.5, 24.6, 17.6, and 13.1%, respectively.

Analysis and functional annotation of DEGs – The five *Rhododendron* species shared 20,509 unigenes in common. Furthermore, a total of 3785, 4382, 5846, 2402, and 12,605 unigenes showed specific expression in *R. fortunei*, *R. simsii*, *R. mariesii*, *R. molle*, and *R. pulchrum*, respectively (Fig. 4a). Moreover, numbers of unigenes shared by *R. fortunei* and *R. simsii*, *R. mariesii* and *R. simsii*, *R. molle* and *R. simsii*, and *R. pulchrum* and *R. simsii* were 28,368, 29,850, 27,408,

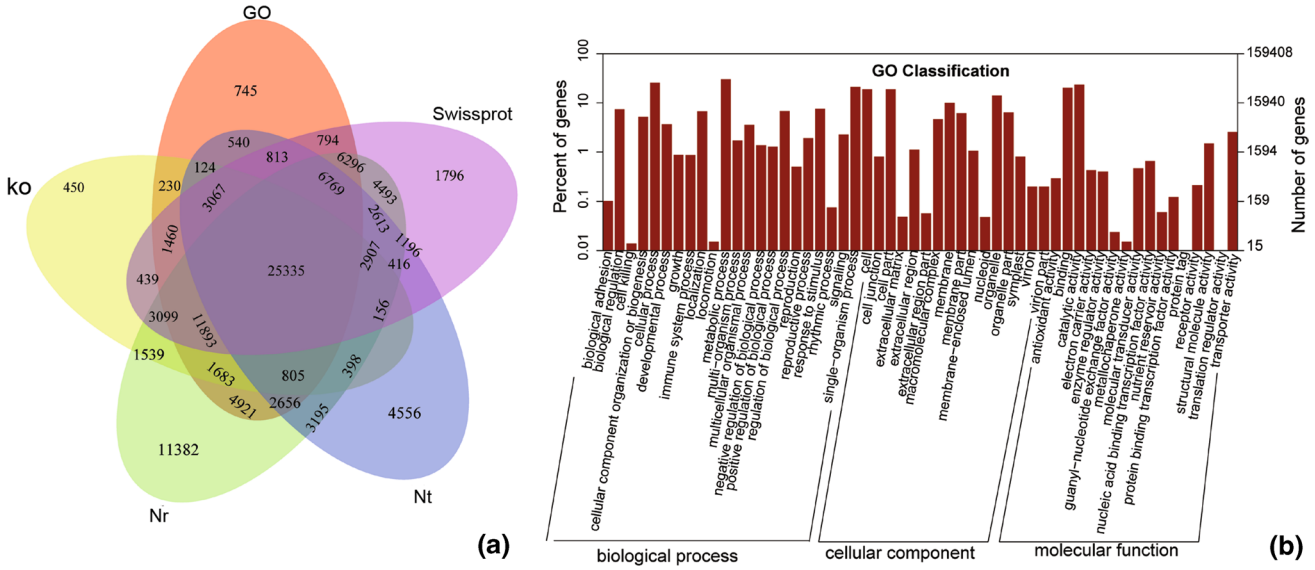

Fig. 2 Veen diagram of the number of unigenes by using BLASTx program against different databases (**a**) and Gene ontology (GO) classification of unigenes (**b**)

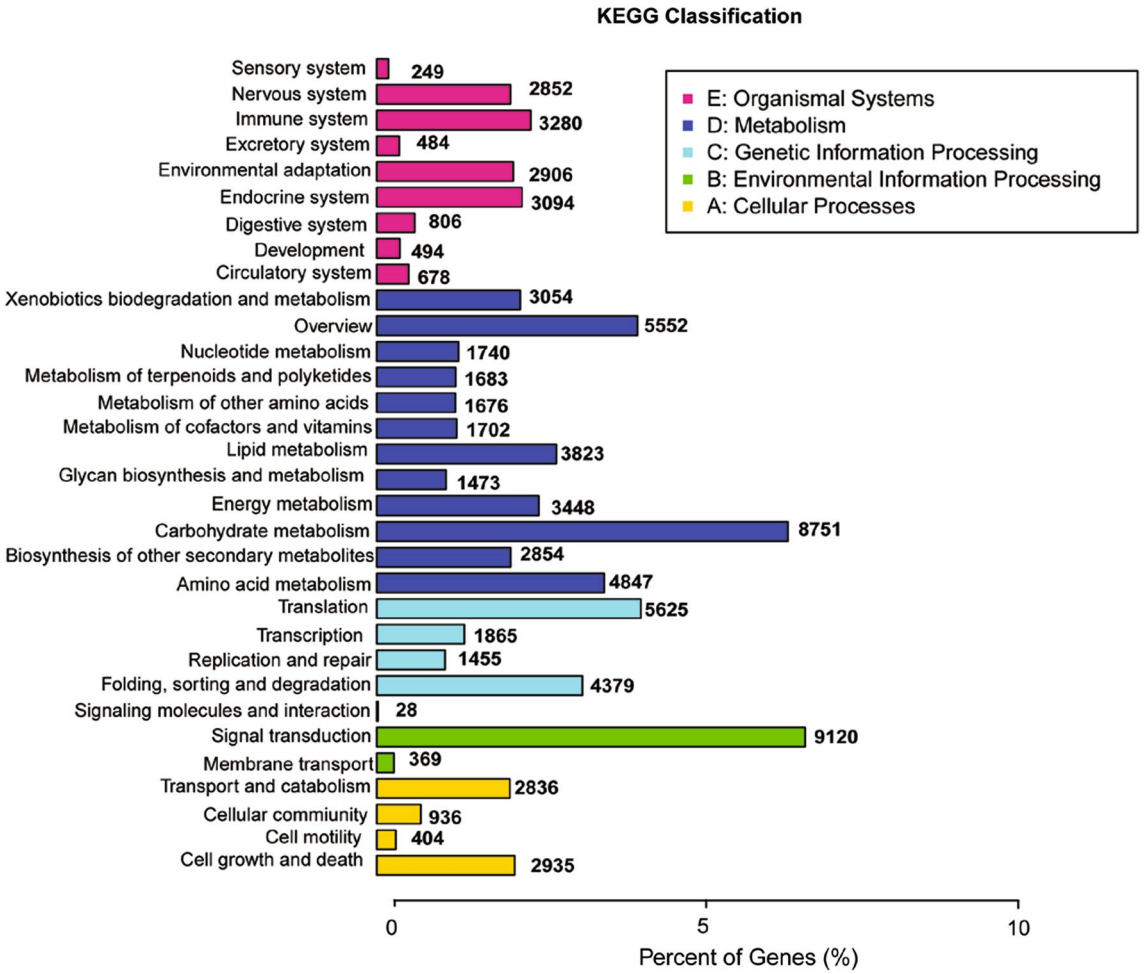

Fig. 3 The KEGG classification of unigenes

and 36,738, respectively. The transcriptomes of *R. fortunei*, *R. mariesii*, *R. molle*, *R. pulchrum* were compared to that of *R. simsii*, yielding 1709, 2,157, 827, and 672 DEGs with an *P* value < 1e-3, FDR < 1e-3 and a | log2 ration | > 2, respectively (Fig. 4b). Compared with *R. simsii*, 1110 down-regulated and 599 up-regulated DEGs were found in *R. fortunei*, 1329 down-regulated and 828 up-regulated DEGs in *R. mariesii*, 471 down-regulated and 356 up-regulated DEGs in *R. molle*, as well as 360 down-regulated and 312 up-regulated DEGs in *R. pulchrum*, respectively (Fig. 4c). By pairwise comparison, 90 common DEGs were found. In addition, seven DEGs were randomly selected for qRT-PCR amplification, and good correlation ($r^2 = 0.8234$) was obtained, confirming the high reliability of these RNA-seq data. Flavonoid biosynthesis was the most abundant KEGG pathway, and genes involved in flavonoid biosynthesis were typical DEGs.

The correlation of these five species was studied by clustering analysis according to expression levels of DEGs. The five species could be divided into two main groups: group 1 (*R. fortune* and *R. mariesii*) and group 2 (*R. simsii*, *R. molle* and *R. pulchrum*) (Fig. 4d). Moreover, group 2 could be further divided into two subgroups: *R. simsii* and *R. pulchrum* were clustered in one subgroup, while *R. molle* was in another subgroup (Fig. 4d). As expected, *R. pulchrum* and *R. simsii* with red flowers were clustered together, *R. fortune* and *R. mariesii* with pink flowers were clustered together, and *R. molle* with yellow flowers was clustered alone.

Among the DEGs between *R. fortunei* and *R. simsii*, 287 were enriched in 30 GO categories. In 'biological process,' the typical DEGs were involved in the response to biotic stimulus (GO: 0,009,607), oxidation–reduction process (GO: 0,055,114), and defense response (GO: 0,006,952) with the numbers of 29, 159, 31, respectively. In 'molecular function,' the most representative categories are heme binding (GO: 0,020,037), monooxygenase activity (GO: 0,004,497), and iron ion binding (GO: 0,005,506) with 51, 30, 49, DEGs, respectively (Fig. S2). Furthermore, chemical carcinogenesis

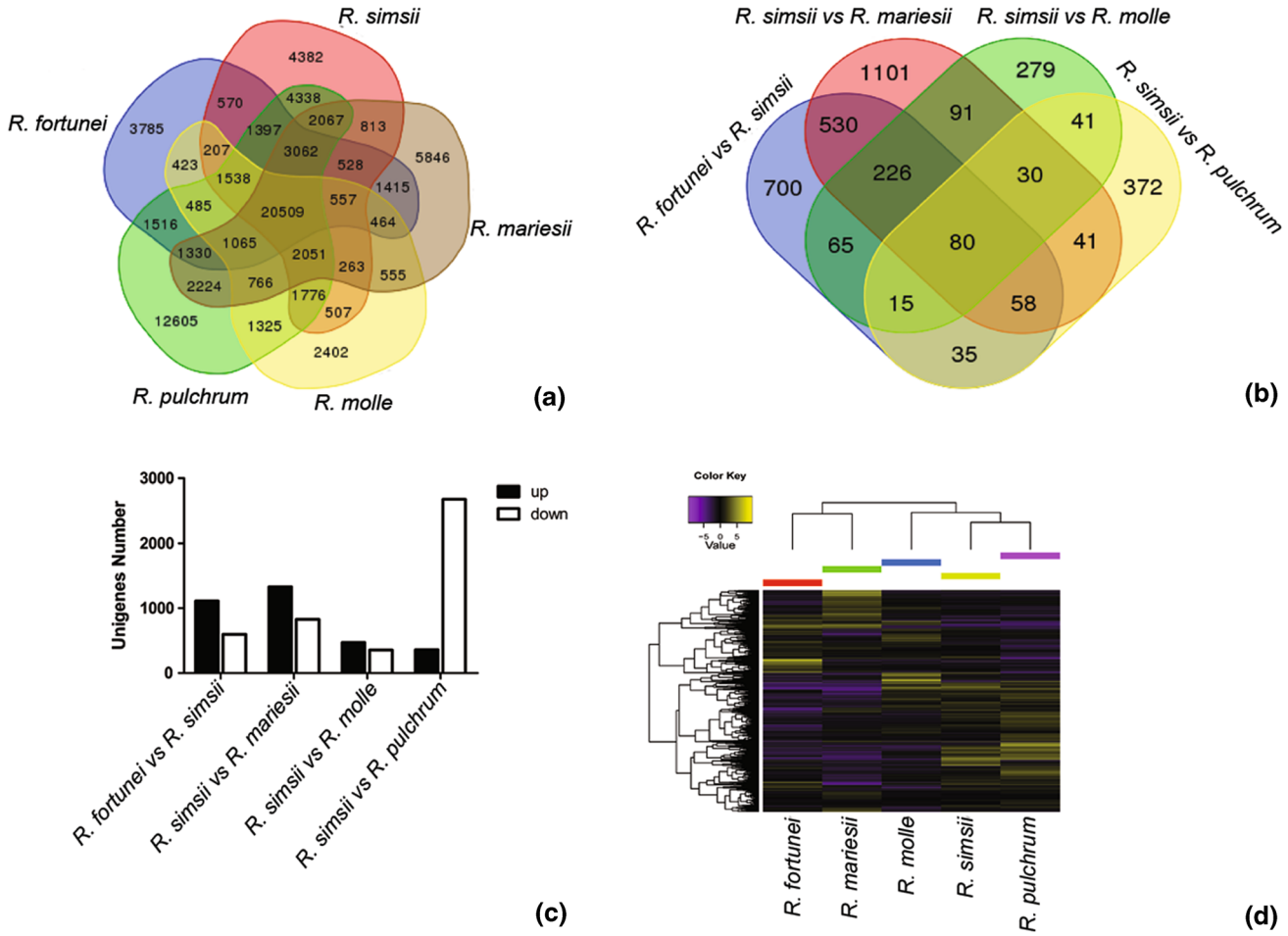

Fig. 4 **a** Veen diagram of the unigenes number of the five samples; **b** Veen diagram of the number of differentially expressed gens (DEGs) obtained through pairwise comparison; **c** Numbers of up-regulated and down-regulated genes based on pairwise comparison; **d** Clustering analysis of DEGs in different species. 0047: *R. fortunei*; 0049: *R. simsii*; 0051: *R. mariesii*; 0053: *R. molle*; 0059: *R. pulchrum*

(ko05204), stibenoid, diarylheptanoid and gingerol biosynthesis (ko00945), as well as drug metabolism-cytochrom P450 (ko00982), were the most KEGG pathway (Fig. S2). In particular, DEGs involved in response to biotic stimulus, defense response, chitin binding, and glutathione transferase activity were all up-regulated in *R. simsii*.

Totally, 438 DEGs between *R. mariesii* and *R. simsii* were enriched in GO categories. The most representative terms were chitin binding (GO: 0,008,061), chitinase activity (GO: 0,004,568), heme binding (GO: 0,020,037), sucrose: proton symporter activity (GO: 0,008,506), and glutathione transferase activity (GO: 0,004,364) with 23, 24, 50, 7, and 13 DGEs, respectively (Fig. S3a). Furthermore, KEGG analysis with DGEs inferred that chemical carcinogenesis (ko05204), drug metabolism-cytochrom P450 (ko00982), metabolism of xenobiotics by cytochrome P450 (ko00980), glutathione metabolism (ko00480), and flavonoid biosynthesis (ko00941) were the most dominant KEGG pathways. In addition, carotenoid biosynthesis (7 DEGs) was also the representative KEGG terms (Fig. S3b). DEGs involved in response to biotic stimulus, chitin binding, chitin catabolic process, cell wall macromolecule, and chitinase activity were all up-regulated in *R. simsii*.

In addition, 172 DEGs between *R. molle* and *R. simsii* were enriched in GO categories. In 'biological process,' 'oxidation–reduction process' (GO: 0,055,114), 'carbohydrate metabolic process' (GO: 0,005,975), 'the response to biotic stimulus' (GO: 0,009,607), and 'defense response' (GO: 0,006,952) were the top four categories with the number of DEGs 70, 23, 12, and 12, respectively. In 'molecular function,' the most representative categories were 'heme binding' (GO: 0,020,037), 'transferase activity, transferring hexosyl groups' (GO: 0,016,758), and 'metallopeptidase activity' (GO: 0,008,237) with 22, 14, 8 DEGs, respectively (Fig. S4a). Most of DEGs involved in response to biotic stimulus (in biological process), extracellular region (in cellular component), metallopeptidase activity (in molecular function) were up-regulated in *R. molle*. Other glycan degradation (ko00511), stibenoid, diarylheptanoid and gingerol biosynthesis (ko00945), and phenylpropanoid biosynthesis (ko00940) were the most enriched KEGG pathways (Fig. S4b).

Furthermore, 176 DEGs between *R. pulchrum* and *R. simsii* were enriched in GO categories. In 'biological process,' oxidation–reduction process (GO: 0055114), defense response (GO: 0006952), the response to biotic stimulus (GO: 0009607) were the top three categories with the number of DEGs 65, 19, 14, respectively. In 'molecular function,' the most representative categories were hydrolase activity, acting on ester bonds (GO: 0016788), glactinol-sucrose galacytosyltransferase activity (GO: 0047274), and pectate lyase activity (GO: 003070) with 16, 15, and 10DEGs, respectively (Fig. S5). DEGs involved in defense response (in biological process), cell wall (in cellular component), and hydrolase activity (in molecular function) were all up-regulated genes in *R. simsii*. Glycan degradation (ko00511), pentose and glucuronate interconversions (ko00040), as well as cutin, suberine, and wax biosynthesis (ko00073), were the most different KEGG pathways.

Identification of genes involved in flower traits – In total, eight terms containing 405 unigenes involved in the synthesis and participation of flower pigments (anthocyanin, carotenoids, and betalains) were found, including flavonoid biosynthesis (GO: 0,009,813), positive regulation of flavonoid biosynthesis (GO: 0,009,963), flavonol biosynthesis process (GO: 0,051,555), anthocyanin-containing compound biosynthesis process (GO: 0,009,718), anthocyanin accumulation in tissue in response to UV light (GO: 0,043,418), regulation of anthocyanin biosynthesis process (GO: 0,031,540), anthocyanin-containing compound metabolic process (GO: 0,046,283), and carotenoid biosynthesis process (GO: 0,016,117). Transcription of DEGs displayed variations in abundance in different species (Fig. 5). Most of genes clustered in Cluster I were involved in carotenoid biosynthesis and isoflavonoid biosynthesis, with higher expression levels in *R. molle* possessing yellow flowers. Within cluster II, unigenes involved in flavone and flavonol biosynthesis showed higher abundance in *R. fortune* with light pink flowers. Within cluster III, transcripts of unigenes participated in flavonoid biosynthesis exhibited higher levels in *R. mariesii* with pink flowers. In cluster IV, unigenes involved in anthocyanin biosynthesis showed higher mRNA levels in *R. simsii* with red flowers. In cluster V, genes involved in carotenoid biosynthesis and flavonoid biosynthesis showed higher expression levels in *R. pulchrum* with purplish red flowers.

In particular, 162 unigenes involved in anthocyanin synthesis pathway were found, mainly containing thirty homologs of *4-coumaryol CoA lignase* (*4CL*), five homologs of *cinnamate 4-hydroxylase* (*C4H*), eight homologs of *anthocyanidin synthase* (*ANS*), fourteen homologs of *chalcone isomerase* (*CHI*), thirty-eight homologs of *chalcone synthase* (*CHS*), thirteen homologs of *dihydroflavonol 4-reductase* (*DFR*), fifteen homologs of *flavonoid 3′,5′-hydroxylase* (*F3′5′H*), six homologs of *flavonoid 3′-hydroxylase* (*F3′H*), ten homologs of *flavonol 3-hydroxylase* (*F3H*), five homologs of *flavonol synthase* (*FLS*), and seventeen homologs of *phenylalanine ammonia lyase* (*PAL*) (Fig. 6). Interestingly, high expression levels of these unigenes were observed in *R. pulchrum*, such as homologous to *F3′H*, *FLS*, and *F3H*. Homologous to *CHI* exhibited higher expression levels in *R. mariesii*. Homologous to *anthocyanidin reductase* (*ANR*) and *leucoanthocyanidin reductase* (*LAR*) were observed higher accumulation in *R. simsii*. *Leucoanthocyanidin dioxygenase* (*LDOX*) exhibited minimum expression in *R. fortune*.

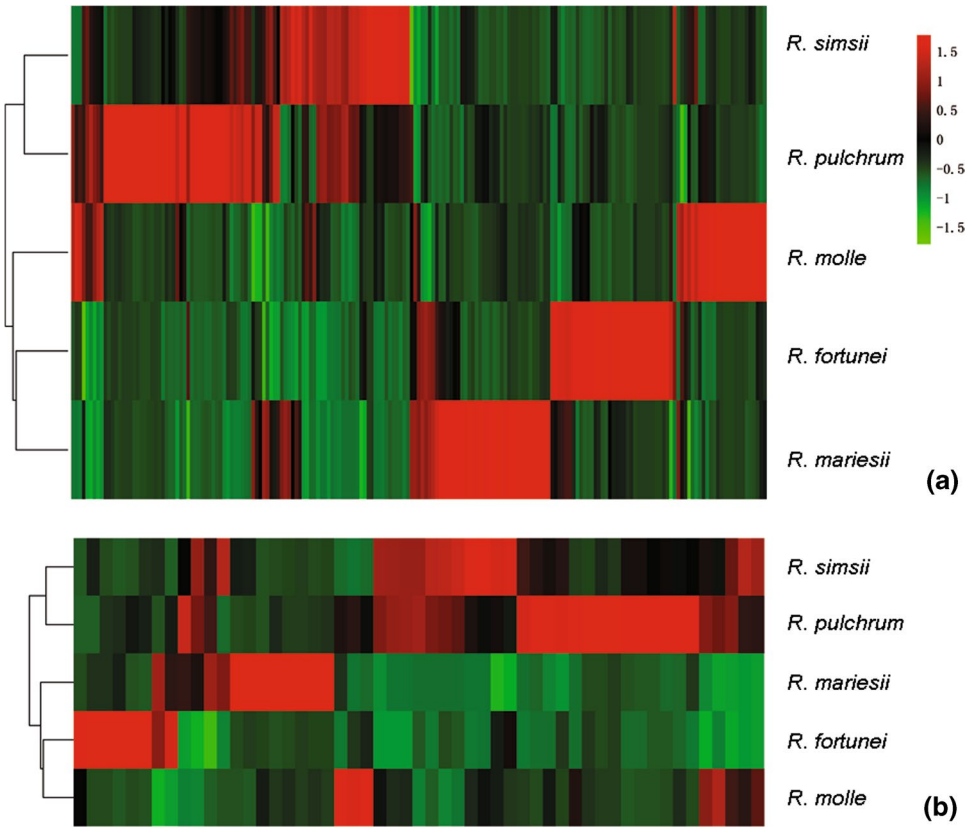

Fig. 5 Heatmaps visualization of differential expressed genes (DEGs) involved in flower. **a** DEGs enriched in flower color related pathways. **b** DEGs likely involved in anthocyanin biosynthesis

Furthermore, 57 unigenes involved in the carotenoid biosynthesis pathway were found: seven homologs of *isopentenyl pyrophosphate isomerase* (*IPI*), five homologs of *phytoene synthase* (*PSY*), two homologs of *phytoene desaturase* (*PDS*), one homolog of *ζ-carotene desaturase* (*ZDS*), three homologs of *lycopene β-cyclase* (*LCYB*), three homologs of *β-ring hydroxylase* (*CHYB*), one homolog of *ε-ring hydroxylase* (*CHYE*), eighteen homologs of *zeaxanthin epoxidase* (*ZEP*), five homologs of *carotenoid isomerase* (*CRTISO*), five homologs of *neoxanthin synthase* (*NSY*), and ten homologs of *9-cis-epoxycarotenoid dioxygenase* (*NCED*) (Fig. 6). Terpenoids are important for fragrance production (Guterman et al. 2002). In total, 92 unigenes were assigned to terpenoids biosynthetic and metabolic pathway, including pentacyclic triterpenoid biosynthetic process (GO: 0,019,745), terpenoid biosynthesis process (GO: 0,016,114), triterpenoid biosynthetic process (GO: 0,016,104), monoterpenoid biosynthesis process (GO: 0,016,099), and tetracyclic triterpenoid biosynthesis process (GO: 0,010,686).

Genes required for photoperiod pathway, light signal transduction, floral meristem development also play important roles in flower development (Irish 2010; Mouradov et al. 2002). In these five *Rhododendron* species, 21 MADS-box genes vital for photoperiod and floral meristem development were found, including three isoforms of *AGL2* (*AGAMOUS LIKE*), one isoform of *AGL8,* one isoform of *AGL15*, one isoform of *AGL16*, and one isoform of *AGL61*. *CONSTANS* (*CO*) is a key regulator of the photoperiod pathway, and eighteen homologs of *CO* were identified. One homolog of *Flowering Locus* (*FT*), the target of *CO*, also existed in these five *Rhododendron* species. Homologous genes of the photoreceptor and circadian clock components were also identified, including twelve homologs of *CRYPTOCHROME* (*CRY*), fifteen homologs of *PHYTOCHROME* (*PHY*), four homologs of *LATE ELONGATED HYPOCOTYL* (*LHY*), ten homologs of *EARLY FLOWERING 4* (*EFL4*), one homolog of *FLAVIN-BINDING KELCH REPEAT* (*FKF1*), and five homologs of *GIGANTEA* (*GI*). One homolog of *APETALA2* (*AP2*), important for floral meristem development, was also searched.

4 Discussion

Flowering is one of the most important events in plant life cycle, therefore, the understanding of genetic mechanisms underlying flowering is vital for plant breeding, especially for horticultural species (Mizuta et al. 2009; Cheon et al. 2011). During the last several decades, molecular mechanisms controlling flowering have been well studied in model plants, particularly in herbaceous plants. As

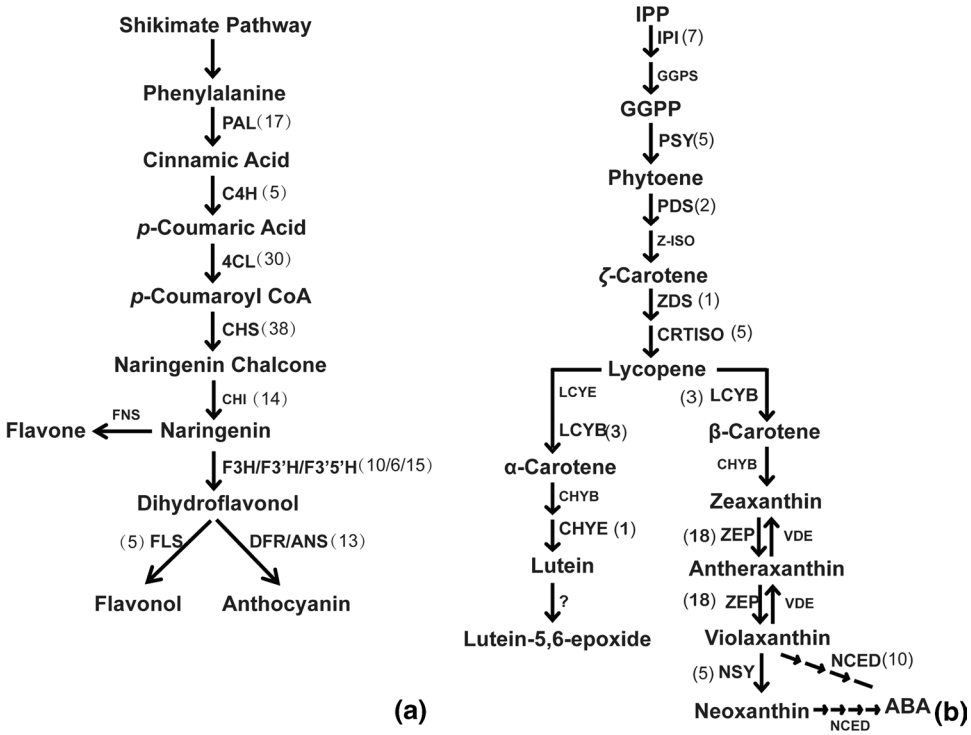

Fig. 6 The unigenes likely involved in flower pigments biosynthesis pathway. **a** Core pathway for anthocyanin biosynthesis pathway; **b** Simplified scheme of the carotenoid biosynthesis pathway

important ornamental woody plant, *Rhododendron* species show large botanical and economic importance. Genes controlling flower initiation and flower color, such as *LEAFY* (*LFY*), *TERMINALFLOWER 1* (*TFL1*), *APETALA 3* (*AP3*), *CHS*, *F3H*, *F3'H*, *FLS* and *CHI* have been isolated and characterized in certain *Rhododendron* species (Nakatsuka et al. 2008; Cheon et al. 2011, 2013; De et al. 2013; Christiaens et al. 2015). Recently, expression levels of genes vital for flower development have been well characterized in *R. pulchrum* (Wang et al. 2018). Anthocyanins and flavonols are the major pigments in various azalea flowers. The main basic groups and derivatives of pigments have been well characterized in *Rhododendron*, such as cyaniding, peonidin, delphinidin, and malvidin (Mizuta et al. 2009). However, genes controlling flower color and other flower traits in different *Rhododedndron* species have not been compared, and no forward and reverse genetics research of *Rhododendron* were yet available due to the lack of *Rhododendron* genomic.

In this study, five *Rhododendron* species possessing different flower colors have been chosen for RNA-seq, hoping for large-scale characterization of genes controlling flowering. In particular, good correlation ($R^2 = 0.8943$) existed between the RNA-seq data and qRT-PCR data, which further confirmed the high reliability of the RNA-seq data obtained (Table S1 and Fig. S6). Among the 159,408 unigenes, 106,766 unigenes (66.98%) could be annotated, which was slightly lower than that in Pummelo (69%) (Liang et al. 2015). The 42.74, 37.66, and 33.88% assembled unigenes in this study were assigned to GO, KOG, and KEGG databases, respectively, indicating that various transcripts are involved in regulating flower traits. Totally, 90 genes encoding key enzymes in pigments biosynthesis were found, containing 71 unigenes in anthocyanin and flavonoid biosynthesis, 12 unigenes in carotenoid biosynthesis, and 7 unigenes in betalains biosynthesis. These genes might be important for anthocyanin synthesis, as 149 known flavonoids, as well as the glycosylated and methylated derivatives, were obtained in blooming flowers of *R. pulchrum* through LC–ESI–MS/MS method, including 38 flavone, 32 flavonol, 14 flavanone, and 7 isoflavone (Wang et al. 2021). Genes of anthocyanin biosynthetic pathway have been divided into early biosynthetic genes and late biosynthetic genes in dicotyledon (Katia and Chiara 2011). Besides *F3'5'H*, the early biosynthetic genes (*CHI*, *F3H*, and *F3'H*) and the late biosynthesis gene *LDOX* also expressed in these five species, and the maximum expression of *F3'5'H* unigenes was in *R. fortune*.

Comparative transcriptome analysis revealed that significant differences existed in these five species. The number of DEGs between *R. simsii* and *R. pulchrum* was minimum (672 DEGs), while the number of DEGs between *R. mariesii* and *R. simsii* was maximum, which was consistent with the significant differences in flower color. *R. molle* with yellow flowers showed a closer relationship with *R. simsii* by clustering analysis. Among these five species, 'oxidation–reduction process' and 'defense response' were the significant GO categories, and 'plant hormone signal transduction' was the typical statistics enriched KEGG pathway, as these genes

might play important roles in flower color diversification. Several lines of previous studies supported this hypothesis. Firstly, antioxidant activity of betalain, one of the main pigments, has been reported (Cai et al. 2003). Secondly, cytochrome P450s, involved in the biosynthesis of anthocyanin pigments, was enriched in KEGG pathway 'metabolism of xenobiotics' (Ayabe and Akashi 2006). Thirdly, ethylene biosynthesis and receptor genes also play vital roles in flower opening (Xue 2008).

Some gene products of DEGs were assigned to flavonoid 3'-monooxygenase and flavonoid 3',5'-hydroxylase, and some were involved in monoterpenoid biosynthesis, which might be involved in anthocyanin pathway in *Petunia* spp. and maize (Marrs et al. 1995; Alfenito et al. 1998). Among DEGs between *R. fortunei* and *R. simsii*, flavonoid biosynthesis (38 DEGs) was the most abundant pathway for the KEGG pathway. Among the identified unigenes involved in flavonoid biosynthesis, one homolog of *F3'H*, three homologs of *F3H*, two homologs of *FLS*, and one homolog of *CHI* were differentially expressed between *R. mariesii* and *R. simsii*. Based on the comparison between *R. molle* and *R. simsii*, one homolog of *F3'5'H*, one homolog of *C4H*, and one homolog of *F3'H* were identified. Among the DEGs between *R. pulchrum* and *R. simsii*, only one homolog of *4CL* was detected. Between *R. pulchrum* and *R. simsii*, five DEGs were enriched for terpenoid backbone biosynthesis in KEGG analysis. All these differences in gene expression profiles might account for flower variation in different *Rhododendron* species.

Co-pigmentation between anthocyanins and flavonols contributes a lot to flower color formation (Alfenito et al. 1998). *F3'5'H* expression level is correlated to petal pigmentation along with flowering progressed (Nakatsuka et al. 2008). The decreased expression of *LDOX* in light pink color flower probably led less dihydroflavonols into the direction of anthocyanin branch, inferring that the expression profile of *LDOX* was correlated with flower color of *Rhododendron*. Previous study in *Paeonia ostill* showed that the late biosynthetic genes were in strong correlation with anthocyanin synthesis (Gao et al. 2016). Lower expression of *ANS* in *R. fortunei* might also probably inhibit the conversion of dihydroflavonols to anthocyanin, resulting in the relatively lower accumulation of anthocyanins in the light pink flower.

In addition to pigments, various factors could also affect flower colors, such as metal ion type and concentration, pH vacuoles, and even shape of surface cells (Yoshida 2009). In this study, genes involved in iron ion binding, copper ion binding, and manganese ion binding, were enriched in GO category 'molecular function' Moreover, DEGs between *R. simsii* and *R. mariesii*, as well as *R. molle* and *R. simsii*, had been assumed to possess beta-primeverosidase activity, which would be involved in aroma formation (Rock and Zeevaart 1991). This study has provided rich genetic information on *Rhododendron* genus, identified lots of genes involved in flower development, as well as obtained DEGs involved in pigments synthesis and fragrance, which would benefit the manipulation of pigments biosynthesis pathways and directional breeding of *Rhododendron* species in the future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40415-021-00720-0.

Acknowledgements This research results reported in this paper are funded by National Natural Science Foundation of China (31500995), fund granted by Hubei Intellectual Property Bureau (2019-1-35), Open fund of Hubei Key Laboratory of Economic Forest Germplasm Improvement and Resources Comprehensive Utilization (201932103 and 201931503), and fund from Assessment and Comprehensive Utilization of Characteristic Biological resources in Dabie Mountains (4022019006).

Author's contribution ZL and QY has performed the RNA-seq and data analysis and prepared the draft manuscript; XD and YZ also did statistical analysis; SZ carried out the qRT-PCR; WZ and SW designed the whole research, as well as wrote the final manuscript.

Availability of data and materials All sequence data generated or supported the findings of this study have been deposited in NCBI (Accession: SRS1967671).

Declaration

Conflicts of interest The authors declare that they have no competing interests.

References

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 10:1135–1149

Ayabe SI, Akashi T (2006) Cytochrome P450s in flavonoid metabolism. Phytochem Rev 5:271–282

Cai Y, Sun M, Corke H (2003) Antioxidant activity of betalains from plants of the Amaranthaceae. J Agr Food Chem 51:2288–2294

Cheon KS, Nakatsuka A, Kobayashi N (2011) Isolation and expression pattern of genes related to flower initiation in the evergreen azalea, *Rhododendron* × *pulchrum* 'Oomurasaki.' Sci Hortic 130:906–912

Cheon KS, Nakatsuka A, Tasaki K (2013) Expression pattern of several flowering-related genes during flower bud formation in *Rhododendron*×*pulchrum* "oomurasaki." J JPN Soc Hortic Sci 82:263–269

Choudhary S, Thakur S, Najar RA, Majeed A, Singh A, Bhardwaj P (2018) Transcriptome characterization and screening of molecular markers in ecologically important Himalayan species (*Rhododendron arboreum*). Genome 61:417–428

Christiaens A, Pauwels E, Gobin B, Van Labeke MC (2015) Flower differentiation of azalea depends on genotype and not on the use of plant growth regulators. Plant Growth Regul 75:245–252

Daiki M, Akira N, Takuya B, Ikuo M, Nobuo K (2014) Pigment composition patterns and expression of anthocyanin biosynthesis genes in *Rhododendron kiusianum*, *R. kaempferi*, and their natural hybrids on Kirishima Mountain mass Japan. J Jpn Soc Hortic Sci 83:156–162

De KE, Desmet L, Van BE, De RJ (2013) How to perform RT-qPCR accurately in plant species? A case study on flower colour gene expression in an azalea (*Rhododendron simsii* hybrids) mapping population. BMC Mol Biol 14:1–15

Dewey CN, Li B (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323–323

Dunemann F, Kahnau R, Stange I (1999) Analysis of complex leaf and flower characters in *Rhododendron* using a molecular linkage map. Theor Appl Genet 98:1146–1155

Fang L, Tong J, Dong YF, Xu DY, Mao J, Zhou Y (2017) De novo RNA sequencing transcriptome of *Rhododendron obtusum* identified the early heat response genes involved in the transcriptional regulation of photosynthesis. PloS One 12:e0186376

Gao L, Yang H, Liu H, Hu Y (2016) Extensive transcriptome changes underlying the flower color intensity variation in *Paeonia ostii*. Front Plant Sci 6:1205

Gautam V, Kohli SK, Kapoor D, Bakshi P, Ahmad P (2020) Stress protective effect of *Rhododendron arboreum* leaves (MEL) on chromium-treated vigna radiata plants. J Plant Growth Regul. https://doi.org/10.1007/s10681-012-0809-7

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fab L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644–652

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annu Rev Plant Biol 57:761–780

Guterman I, Shalit M, Menda N, Piestun D, Dafny-Yelin M, Shalev G, Bar E, Davydov O, Ovadis M, Emanuel M, Wang JH, Adam Z, Pichersky E, Lewinsohn E, Zamir D, Vainstein A, Weiss D (2002) Rose Scent: genomics approach to discovering novel floral fragrance–related genes. Plant Cell 14:2325–2338

Hill CB, Cassin A, Keeble-Gagnère G, Doblin MS, Bacic A, Roessner U (2016) *De novo* transcriptome assembly and analysis of differentially expressed genes of two barley genotypes reveal root-zone-specific responses to salt exposure. Sci Rep 6:31558

Irish VF (2010) The flowering of Arabidopsis flower development. Plant J 61:1014

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Katia P, Chiara T (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci 181:219–229

Keyser ED, Lootens P, Bockstaele EV, Riek JD (2013) Image analysis for QTL mapping of flower colour and leaf characteristics in pot azalea. Euphytica 189:445–460

Liang M, Yang X, Li H, Su S, Yi H, Chai L, Deng X (2015) De novo transcriptome assembly of pummelo and molecular marker development. PLoS One 10:e0120615

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 375:397–400

Mizuta D, Ban T, Miyajima I, Nakatsuka A, Kobayashi N (2009) Comparison of flower color with anthocyanin composition patterns in evergreen azalea. Sci Hortic 122:594–602

Mouradov A, Cermer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. Plant Cell 14:S111–S130

Nakatsuka A, Mizuta D, Kii Y, Miyajima I, Kobayashi N (2008) Isolation and expression analysis of flavonoid biosynthesis genes in evergreen azalea. Sci Horti 118:314–320

Nishihara M, Nakatsuka T (2011) Genetic engineering of flavonoid pigments to modify flower color in floricultural plants. Biotechnol Let 33:433–441

Ozsolak F, Milos PM (2011) Single-molecule direct RNA sequencing without cDNA synthesis. Wires RNA 2:565–570

Popescu R, Kopp B (2013) The genus Rhododendron: an ethnopharmacological and toxicological review. J Ethnopharmacol 147:42–62

Rock CD, Zeevaart JA (1991) The aba mutant of *Arabidopsis thaliana* is impaired in epoxy-carotenoid biosynthesis. Proc Natl Acad Sci USA 88:7496–7499

Scheiber SM, Jarret RL, Robacker CD, Newman M (2000) Genetic relationships within *Rhododendron* L. section Pentanthera G. Don based on sequences of the internal transcribed spacer (ITS) region. Sci Hortic 85:123–135

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S (2016) TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res 26:1134–1144

Sun H, Wang J, Que J, Peng Y, Yu Y, Wang L, Ye H, Huang K, Xue Y, Zhou Y, Ji K (2019) RNA sequencing revealing the role of amp-activated protein kinase signaling in mice myocardial ischemia reperfusion injury. Gene 703:91–101

Wang XQ, Huang Y, Long CL (2010) Cross-amplification and characterization of microsatellite loci for the genus Rhododendron. HortScience 45:1394–1397

Wang S, Li Z, Jin W, Famg Y, Yang Q, Xiang J (2018) Transcriptome analysis and identification of genes associated with flower development in *Rhododendron pulchrum* Sweet (Ericaceae). Gene 679:108–118

Wang S, Huang S, Yang J, Li Z, Zhang M, Fang Y, Yang Q, Jin W (2021) Metabolite profiling of violet, white and pink flowers revealing flavonoids composition patterns in *Rhododendron pulchrum* Sweet. J Biosci 46:3

Xiao Z, Su J, Sun X, Li C, He L, Cheng S, Liu X (2018) De novo transcriptome analysis of *Rhododendron molle* G. Don flowers by Illumina sequencing. Genes Genome 40:591–601

Xing W, Liao J, Cai M, Xia Q, Liu Y, Zeng W, Jin X (2017) *De novo* assembly of transcriptome from *Rhododendron latoucheae* Franch. using Illumina sequencing and development of new EST-SSR markers for genetic diversity analysis in Rhododendron. Tree Genet Genomes 13:53

Xue J (2008) Expression of ethylene biosynthetic and receptor genes in rose floral tissues during ethylene-enhanced flower opening. J Exp Bot 59:2161–2169

Yoshida K (2009) Blue flower color development by anthocyanins: from chemical structure to cell physiology. Nat Prod Rep 26:857–974