Comparative Transcriptome Analysis of Alpinia oxyphylla reveals Tissue-specific Expression of Flavonoid Biosynthesis Genes

Bingmiao Gao (gaobingmiao@hainmc.edu.cn)

Research article

Keywords: Alpinia oxyphylla, transcriptome analysis, differentially expressed genes, secondary, metabolites, flavonoidbiosynthesis

Posted Date: December 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-115139/v1

Article

Comparative Transcriptome Analysis of *Alpinia oxyphylla* reveals Tissue-specific Expression of Flavonoid Biosynthesis Genes

Lin Yuan[1], Kun Pan[1], Yonghui Li[1], Bo Yi[2]*, Bingmiao Gao[1]*

1 Key Laboratory of Tropical Translational Medicine of Ministry of Education, Hainan Key Laboratory for Research and Development of Tropical Herbs, Hainan Medical University, Haikou 571199, China;
2 Department of Pharmacy, 928th Hospital of PLA Joint Logistics Support Force, Haikou 571159, China.

* Correspondence: gaobingmiao@hainmc.edu.cn (B.G.); Tel.: +86-0898-6689-3826 (B.G.); boyicn@126.com (B.Y.); Tel:+86-898-66968410

Abstract

Background: *Alpinia oxyphylla* is an important edible and medicinal herb, and its dried fruits are widely used in traditional herbal medicine. Flavonoids are one of the main chemical compounds in *A. oxyphylla*; however, the genetic and molecular mechanisms of flavonoid biosynthesis are not well understood.

Methods: We performed transcriptome analysis in the fruit, root, and leaf tissues of *A. oxyphylla* to delineate tissue-specific gene expression and metabolic pathways in this medicinal plant.

Results： In all, 8.85, 10.10, 8.68, 6.89, and 8.51 Gb clean data were obtained for early-, middle-, and late-stage fruits, leaves, and roots, respectively. Furthermore, 50,401 unigenes were grouped into functional categories based on four databases, namely Nr (47,745 unigenes), Uniprot (49,685 unigenes), KOG (20,153 unigenes), and KEGG (27,285 unigenes). A total of 3,110 differentially expressed genes and five distinct clusters with similar expression patterns were obtained, in which 27 unigenes encoded 13 key enzymes (such as *CHS, CHI, F3H, FLS, ANS*) associated with flavonoid biosynthesis.

Conclusion： The tissue-specific expression of the genes corresponds to accumulation of flavonoids in these tissues.These results provide insights into the molecular mechanism of flavonoid biosynthesis in *A. oxyphylla* and application of genetically engineered varieties of *A. oxyphylla*.

Keywords: *Alpinia oxyphylla*; transcriptome analysis; differentially expressed genes; secondary metabolites; flavonoid biosynthesis

Background

Alpinia oxyphylla, a member of the Zingiberaceae family, is an important plant species for traditional Chinese medicine, which originates in the Hainan Province and is widely cultivated in southern China [1]. The dried fruits of *A. oxyphylla* are regarded as a valuable drug that has a long clinical history as a well-known constituent of the four southern Chinese medicines in China [2, 3]. The fruits of *A. oxyphylla* are widely used in the treatment of ulcerations, gastralgia, diarrhea, dementia, diabetes, and Alzheimer's disease [4-9]. Numerous studies have reported that the fruits of *A. oxyphylla* are rich in flavonoids, diarylheptanoids, terpenoids, volatile oils, and steroids and their glycosides [10-13]. Among these compounds, flavonoids and terpenoids are the main active ingredients of *A. oxyphylla* fruits, which have been found to exert various pharmacological activities [13].

Usually, there are variations in the distribution of secondary metabolites in different tissues of higher plants [14,15]. The concentration of chemical constituents was comparable in roots and leaves of *A. oxyphylla*, but was significantly higher in fruits [16]. Therefore, the secondary metabolites accumulate preferentially in fruits, instead of roots or leaves. In addition, the content of chemical compounds in the fruits of *A. oxyphylla* harvested at different times indicates that the 45-day harvested fruit had the highest content of chemicals [16,17]. The metabolic processes and regulatory mechanisms

of these chemical compounds in different tissues and fruits at different stages have not yet been elucidated.

The transcriptome is a complete set of RNA transcripts in a cell at a specific developmental stage, and provides information on gene expression and regulation related to a variety of cellular processes including secondary metabolite biosynthesis [18,19]. With the development of next-generation sequencing, RNA sequencing is an effective method for investigating the metabolic pathways influenced by active ingredients and associated gene expression in different tissues or samples, such as flavonoid biosynthesis in *Ampelopsis megalophylla* [20], terpenoids metabolism in ginseng roots [21] and polysaccharide and alkaloid content in *Dendrobium* [22]. To date, there are no studies on the genetic modification of *A. oxyphylla* either toward increased production of secondary metabolites or biomass accumulation. Therefore, it is important to explore the whole genome transcriptome of *A. oxyphylla* to identify candidate genes contributing to metabolic processes and regulatory mechanisms.

In this study, fruits at three different growth stages, roots, and leaves of *A. oxyphylla* were collected and transcriptome sequencing was performed by Illumina 4000 sequencing technology. We attempted to predict accumulation of compounds and metabolites among different tissues and stages of the fruits through this analysis. The assembled unigenes were annotated by searches against public databases such as NCBI non-redundant protein (Nr), Universal Protein (Uniport), EuKaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) classifications. After assembly and annotation, the differentially expressed genes were identified in specific tissues and stages of fruits. We also identified the pathway of flavonoid biosynthesis and analyzed the expression of associated genes. Therefore, the results of this study may serve as a significant resource for developing genetically engineered varieties of *A. oxyphylla* with improved quality and yield.

Results

De novo assembly

The three tissue samples (fruits of different developmental stages, leaves, and roots) of *A. Oxyphylla* were sequenced using Illumina HiSeq 4000 which generated approximately 29.50, 33.67, 28.93, 22.98, and 27.84 million pair-end short reads with a length of 150 bp for early-fruits, middle-fruits, late-fruits, leaves, and roots, respectively. After filtering out low-quality reads and adapters, we obtained 8.85, 10.10, 8.68, 6.89, and 8.51 Gb clean data for each sample, and the clean data ratio were estimated to be 99.84%, 99.85%, 99.84%, 99.80%, and 99.86%, respectively (Table 1). The raw reads were deposited in the NCBI Sequence Read Archive (BioProject accession PRJNA559252, SRA accession SRR9937427, SRR9937428, SRR9937429, SRR9937430, SRR9937426). *De novo* assembly of the short reads generated 262,114 contigs and 140,126 unigenes for the whole transcriptome, and N50 was calculated to be 1567 bp and 1073 bp and the mean lengths were 916 bp and 658 bp. The average GC content of contigs and unigenes for the *A. oxyphylla* transcriptome were 43.76% and 43.78%, respectively (Table 1).

Table 1. Sequencing statistics and assembly summary for the fruits, leaves, and roots of *A. oxyphylla*.

Samples	Fruits			Leaves	Roots
	Early	Middle	Late		
			Raw data		
Total Reads	29496176	33671483	28927107	22975241	27836177
Total length (bp)	8848852800	10101444900	8678132100	6892572300	8350853100
Read length (bp)	150	150	150	150	150
			Clean data		
Total Reads	29448034	33622040	28882070	22928184	27796543
Total length (bp)	8834410200	10086612000	8664621000	6878455200	8338962900
Clean data ratio	99.84%	99.85%	99.84%	99.80%	99.86%
			Contigs		
Total Number			262114		
Total Length (bp)			240350061		
Mean Length (bp)			916		

N50 (bp)	1567
N70 (bp)	939
N90 (bp)	352
GC Content	43.76%
	Unigenes
Total Number	140126
Total Length (bp)	92262411
Mean Length (bp)	658
N50 (bp)	1073
N70 (bp)	507
N90 (bp)	263
GC Content	43.78%

Functional annotation and classification

To investigate the function of unigenes, annotation was performed based on four databases. A total of 50,401 unigenes were grouped into the databases, Nr (47,745 unigenes), Uniprot (49,685 unigenes), KOG (20,153 unigenes), and KEGG (27,285 unigenes), respectively, while an additional 89,725 unigenes were not found in these databases. A detailed comparison of the unigenes annotated by four different databases are illustrated in Fig. 1.

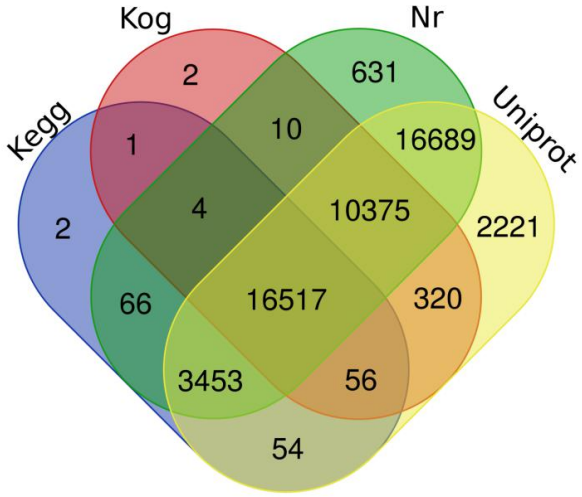

Figure 1. Venn diagram describing the unigenes annotated by four different databases. The integration of unique similarity search results against the NCBI non-redundant protein (Nr), Universal Protein (Uniport), EuKaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases.

GO analysis illustrated that 218,989 unigenes of *A. oxyphylla* were annotated into three categories: molecular function (43,557), cellular component (61,441), and biological process (39,881), respectively (Fig. S1). The binding (19,730) and catalytic activity (17,452) functional groups were the most prominent molecular functions (Fig.S1, MF). A total of 20,153 unigenes of *A. oxyphylla* were further annotated and grouped into 25 molecular families in KOG database (Fig. S2). These molecular families were grouped into four categories: information storage and processing (5575), cellular processes and signaling (7377), metabolism (6180), and poorly characterized (5803). For KEGG analysis, 29,211 unigenes of *A. oxyphylla* had significant matches in the database and were assigned to five primary categories: cellular processes (3324), environmental information processing (2571), genetic information processing (5073), metabolism (13,599), and organismal systems (4644) (Fig. S3). A majority of unigenes were assigned to metabolism, and global and overview maps had the highest number of annotated unigenes (5005).

Differential gene expression analysis

There were 35,278 differentially expressed genes (DEGs) identified between the leaf vs fruit sample, including 15,063 up-regulated and 20,215 down-regulated DEGs (Fig. 2a). A total of 34,846 DEGs were identified between root vs. fruit sample, including 14,807 up-regulated and 20,039 down-regulated DEGs (Fig. 2b). There were 19,776 DEGs between root vs. leaf sample, out of which 8797 were up-regulated and 10,979 were down-regulated (Fig. 2c). Using a Venn diagram, we compared the data sets from the three comparison groups (leaf vs. fruit, root vs. fruit, and root vs. leaf). In this comparison, 19,266 DEGs were identified as common (Fig. 2d) to all three groups. A total of 16,213 DEGs were identified in both "leaf vs. fruit" and "root vs. fruit" comparisons; 19,266 DEGs were identified in both "leaf vs. fruit" and "root vs. leaf" comparisons; while 19,266 DEGs were identified in both "root vs. fruit" and "root vs. leaf" comparisons.

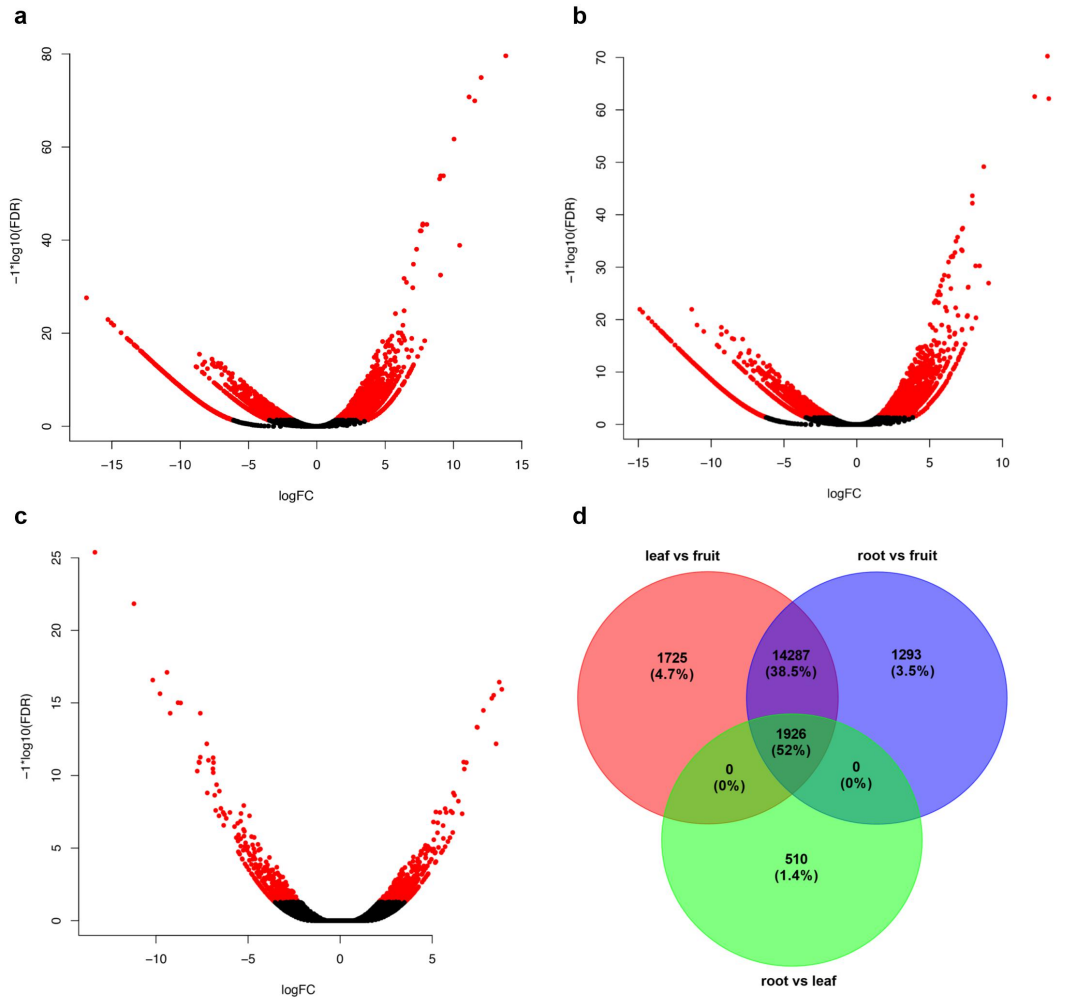

Figure 2. Volcano plots of the differentially expressed genes (DEGs) in the comparison group of (a) leaf vs. fruit, (b) root vs fruit, and (c) root vs. leaf. Venn diagram of DEGs in three different comparisons groups represented by three circles. The overlapping parts of the circles represent the number of DEGs in common in the comparison groups.

Cluster and KEGG enrichment analysis of DEGs

To investigate the expression trends of DEGs in different tissues, we performed a cluster analysis using normalized expression values from each individual replicate of five different samples of *A. oxyphylla*. As a result, a total of 3,110 DEGs and five distinct clusters with similar expression patterns were obtained, containing 606, 807, 954, 725, and 18 genes, respectively (Fig 3a). As shown in Fig. 3b, the expression level of cluster I (606) and cluster IV (725) genes in fruits of *A. oxyphylla* were higher than in roots and leaves， and the expression levels of cluster II (807), cluster III (954), and cluster V

(18) in fruits were lower than in roots and leaves. The secondary metabolites in fruits are higher than roots and leaves, for instance, flavonoids in fruits are 1000 times higher than roots and leaves [16]. Therefore, the DEGs related to secondary metabolite biosynthesis should be in cluster I and cluster IV. Signal pathway analysis of DEGs in the five clusters showed that cluster I contains DEGs involved in flavonoid biosynthesis, isoquinoline alkaloid biosynthesis, and biosynthesis of secondary metabolites (Fig. 4).

Through further comparative analysis, there were 35 and 44 DEGs related to secondary metabolites in root vs fruit and leaf vs fruit, repetively (Table 2). These DEGs were mainly distributed in phenylpropanoid, flavonoid and isoquinoline alkaloid biosynthesis pathways. For phenylpropanoid biosynthesis pathways, 14 DEGs were up-regulated and 3 DEGs were down-regulated in root vs fruit, and 19 DEGs were up-regulated, 5 DEGs were down regulated in leaf vs fruit. It is noteworthy that all the 8 DEGs mapped to flavonoids biosynthesis, and they were both up-regulated in leaf vs fruit (Table 2 and Fig. S4). In addition, 2 DEGs were up-regulated in anthocyanin biosynthesis, 3 DEGs were down-regulated in diarylheptanoid and gingerol biosynthesis, 1 DEGs were up-regulated and 2 DEGs were down-regulated in sesquiterpenoid and triterpenoid biosynthesis. In conclusion, phenylpropanoid, flavonoids and isoquinoline alkaloid biosynthesis related DEGs were significantly up-regulated, while diarylheptanoid, gingerol, sesquiterpenoid, triterpenoid and carotenoid biosynthesis related DEGs were down-regulated in fruits compared with roots and leaves.

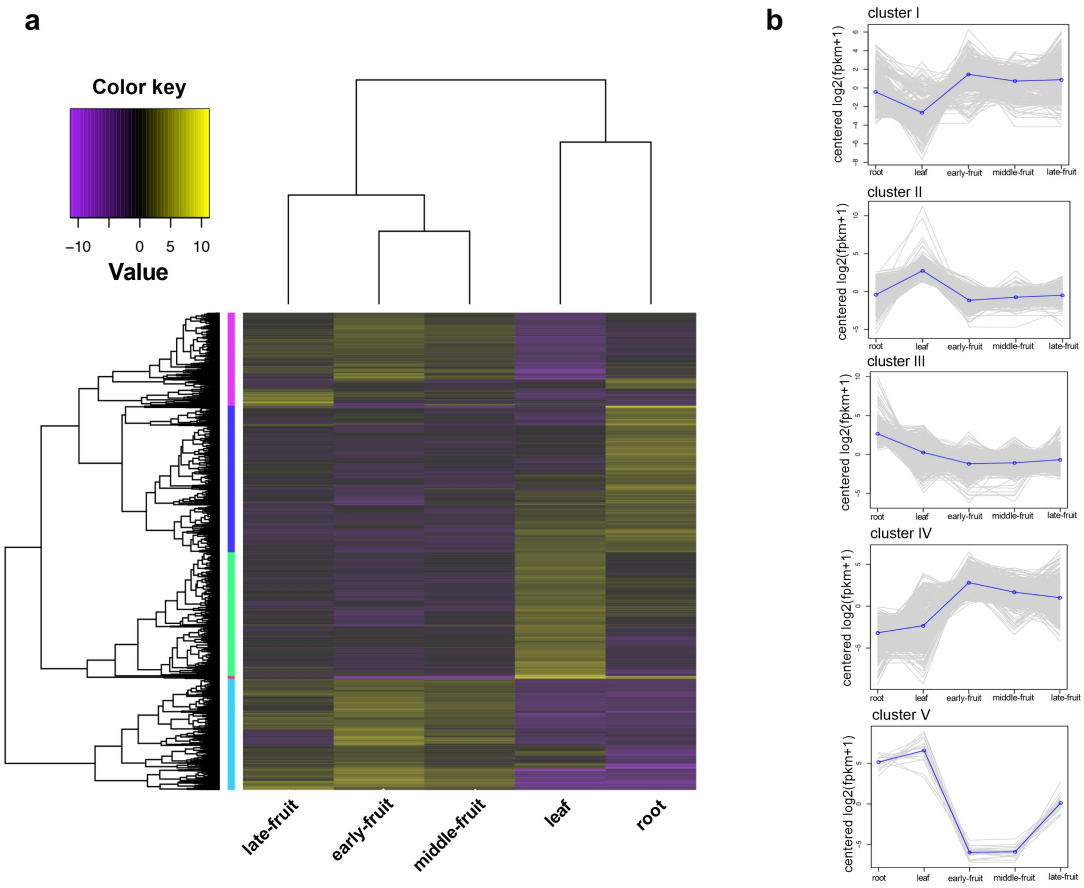

Figure 3. Cluster analysis of DEGs (a) Heat-map showing the expression of DEGs using RNA-seq data derived from mean value of three replicates of each sample based on log 2 (FPKM) values. Color code indicates expression levels. Similarity between samples and unigenes with hierarchical clustering is shown above and on the left of the heatmap, respectively. (b) Cluster analysis of all DEGs. The y-axis in each graph represents the mean-centered log2 (FPKM+1) value. Expression of a single gene is plotted in gray, while the mean expression of the genes in each cluster is plotted in blue.

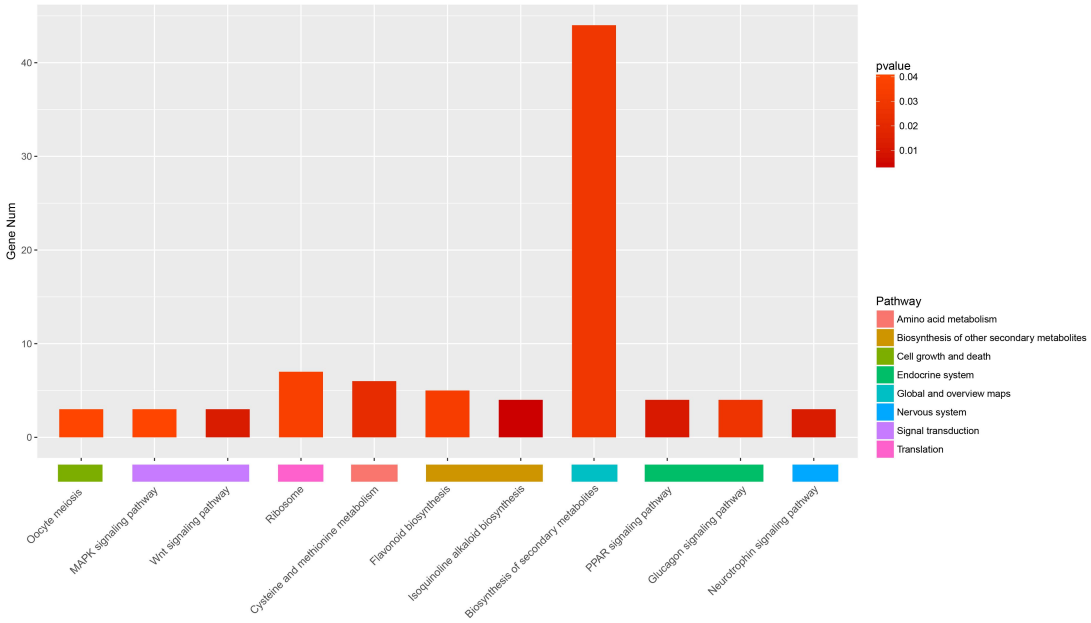

Figure 4. Distribution map of DEGs in cluster I signaling pathway.

Table 2 Comparative analysis of gene expression regulation of secondary metabolites biosynthesis in fruits, roots and leaves

Group	ROOT	Second	mapID	Description	DEGs	up-gene in Fruit	down-gene in fruit
root vs fruit	metabolism	biosynthesis of other secondary metabolites	map00940	phenylpropanoid biosynthesis	35	14	3
			map00942	anthocyanin biosynthesis		2	0
			map00945	stilbenoid, diarylheptanoid and gingerol biosynthesis		0	3
		metabolism of terpenoids and polyketides	map00909	sesquiterpenoid and triterpenoid biosynthesis		1	2
leaf vs fruit	metabolism	biosynthesis of other secondary metabolites	map00940	phenylpropanoid biosynthesis	44	19	5
			map00941	flavonoid biosynthesis		8	0
			map00950	isoquinoline alkaloid biosynthesis		7	0
		metabolism of terpenoids and polyketides	map00906	carotenoid biosynthesis		0	5

Candidate genes associated with flavonoid biosynthesis

Flavonoids are one of the main chemical compounds found in *A. oxyphylla* and are important for evaluating its quality [17]. To understand the regulation of flavonoid biosynthesis in *A. oxyphylla*, key regulatory genes involved in the pathways for phenylpropanoid and flavonoid biosynthesis were identified in this study. Twenty-seven unigenes encoding 13 key enzymes observed in this study were mostly associated with biosynthesis of flavonoids. Furthermore, results of the microarray analysis of

tissue-specific transcriptomes demonstrated that the majority of genes encoding enzymes in the biosynthesis of flavonoids were expressed preferentially in the fruit of *A. oxyphylla* (Fig. 5a). In particular, 9 DEGs, including CHS, CHI, F3H, FLS, ANS, DFR, and ANR unigenes, were significantly up-regulated in fruits, whereas expression of 11 DEGs including F3'5'H, HCT, CCoAMT, 4CL and PAL, were highly up-regulated in roots. However, the flavonoid biosynthesis associated genes exhibited low expression levels in leaves, particularly *4CL* and *FLS* displayed an expression value of 0 (Table S1). In previous studies, flavonoids are found in high concentrations in fruits, followed by roots, and are found in the lowest concentrations in leaves [16]. Expression analysis of flavonoid biosynthesis genes in the present study also showed a similar trend. The putative flavonoid synthesis pathway is shown in Fig. 5b. Flavonoids are synthesized via the phenylpropanoid pathway and are converted from phenylalanine to chalcone by the enzymes PAL, C4H, 4CL, and CHS. CHI catalyzes the isomerization of chalcones into flavanone. Flavanone can be converted either to flavonols through the subsequent action of F3H and FLS, or to flavone through the action of DFR and LAR. However, no unigene coding for flavone synthase (FNS) was detected in the transcriptome analysis. A similar situation has been reported in the transcriptome sequencing of other plants such as *Sophora japonica*, which may be attributed to the fact that FNS genes are short fragments without sequence similarity [23].

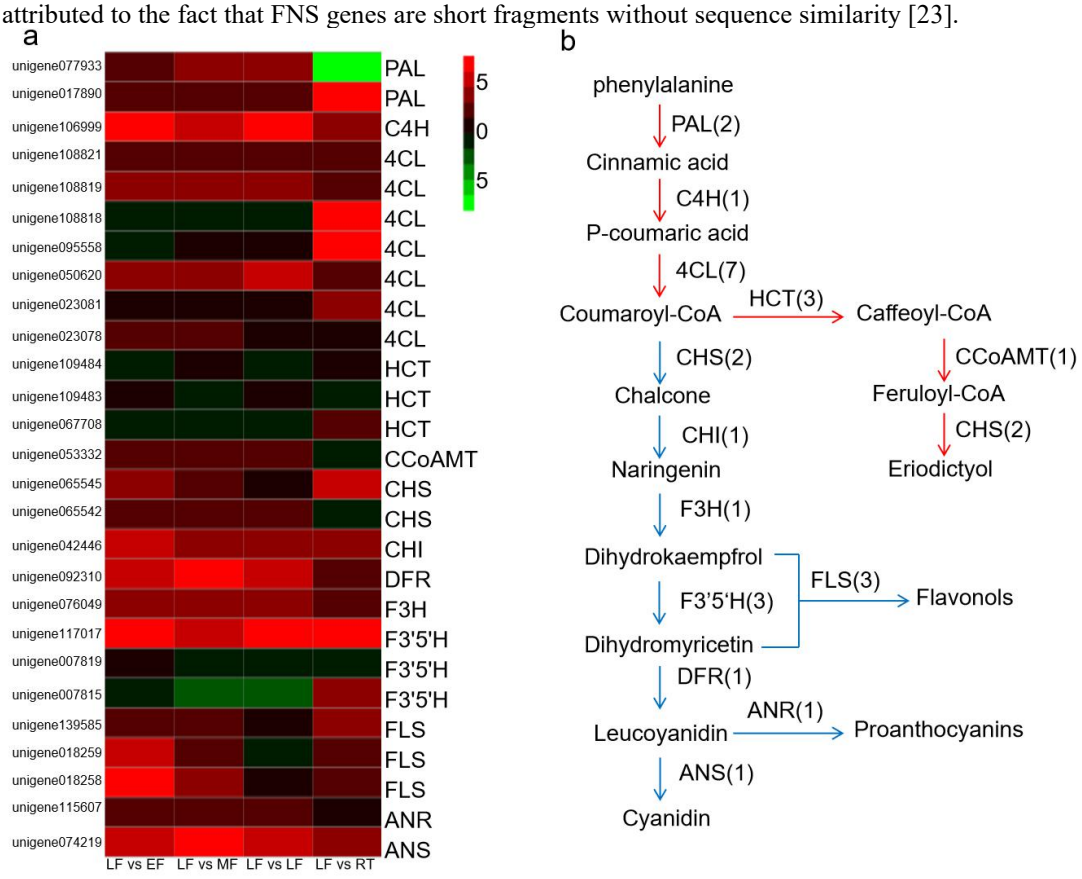

Figure 5. Putative flavonoid biosynthesis pathway in *A. oxyphylla*. (a) Expression level of candidate *A. oxyphylla* unigenes coding for key enzymes involved in flavonoid biosynthesis pathways. Green and red colors are used to represent low-to-high expression levels (mean centered $\log_2$-transformed FPKM values). (b) Pathway for flavonoid biosynthesis. The numbers in brackets following each gene name indicate the number of *A. oxyphylla* unigenes corresponding to that gene. Enzyme abbreviations are as follows: PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; CHS, chalcone synthase; CCoAMT, Caffeoyl Co-A transferase; 4CL, 4-coumarate-CoA ligase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'5'H, flavonoid-3', 5'-hydroxylase; DFR, dihydroflavonol-4-reductase; flavonoid-3-O-glucosyltransferase; ANR, anthocyanidin reductase; ANS, anthocyanidin synthase.

Discussion

There are about 250 species of *Alpinia* plants distributed in tropical Asia [24]. The roots and fruits of *Alpinia* plants are often used for medicinal applications [2, 25]. The capsular fruit of *A. oxyphylla* has been used as a medicinal constituent or health supplement for centuries as one of the four famous southern Chinese medicines [2, 3]. Studies in natural product chemistry reveal that the capsular fruit, root, and leaf contain flavonoids, sesquiterpenes, diarylheptanoids, essential oils, glycosides, and steroids [14,16]. The main chemical components of *A. oxyphylla* flavonoids comprise of tectochrysin, izalpinin, chrysin, and kaempferide, of which tectochrysin is the second most abundant flavonoid concentrated in fruits [11]. Therefore, flavonoids are one of the most important active chemical components in *A. oxyphylla* and are important for evaluating its quality. However, the molecular mechanism of tissue-specific flavonoid biosynthesis and accumulation in *A. oxyphylla* remains largely unexplored.

In this study, we collected three tissue samples (fruits of different developmental stages, leaves, and roots) of *A. oxyphylla* and performed a comparative transcriptome analysis, with a particular focus on flavonoid biosynthesis genes. The characterized flavonoids, including tectochrysin, izalpinin, chrysin, and kaempferide, are found in greatest concentrations in fruits, followed by roots, and are found in the lowest concentrations in leaves [16]. To analyze if the gene expression of biosynthetic genes also follow this pattern, high-throughput transcriptome sequencing technology was employed. Indeed, transcriptional analysis showed that a large number of transcripts exhibited a tissue-specific expression. The number of DEGs in the 'leaf vs. fruit' and 'root vs. fruit' comparison groups was higher than that in the 'root vs. leaf' comparison group. These results suggest that the medicinal properties and associated biological processes are concentrated in the fruits of *A. oxyphylla*. To investigate the trends of DEGs in gene expression, we performed a cluster analysis using normalized expression values from each individual replicate of five different samples of *A. oxyphylla*. A total of 3110 DEGs were divided into five distinct clusters according to their expression patterns. Further analysis showed that only the cluster I of DEGs were related to flavonoid biosynthesis, isoquinoline alkaloid biosynthesis and biosynthesis of secondary metabolites, and the expression level in fruits was significantly higher than that in leaves and roots. The expression level of flavonoid related genes was consistent with that of chemical components. The enriched KEGG pathways results showed that all the DEGs related to flavonoid biosynthesis were up-regulated, and most of the DEGs involved in phenylpropanoid biosynthesis were also up-regulated, but the DEGs related to stilbenoid, diarylheptanoid and gingerol biosynthesis were down-regulated in fruits, indicating that flavonoids were the main secondary metabolites.

The biosynthesis of flavonoids has been reported in many other medicinal plants such as *Astragalus membranaceus* var. mongholicus, *Apocynum venetum,* and *Eucommia ulmoides*, and phenylpropanoid biosynthesis is the common core pathway for the synthesis of flavonoids [26-28]. The first step in flavonoid biosynthesis is regulated by enzymes (PAL, C4H, and 4CL) in the phenylpropanoid pathway. The substrate 4-coumaroyl-CoA is converted into chalcone by CHS in the first rate-limiting step of flavonoid biosynthesis [29]. Next, different flavonoid subgroups are synthesized through modification of the molecular backbone, which is controlled by flavonoid, flavone and flavonol biosynthesis enzymes such as HCT, CCoAMT, CHS, CHI, F3H, F3′,5′H, DFR, ANR, and ANS [28-31]. In this study, homologous unigenes and the expression levels of these genes were investigated in samples of different tissues from *A. oxyphylla*.

Interestingly, DEGs encoding *CHS, CHI, F3H, FLS, ANS, DFR* and *ANR* were highly expressed in the samples from fruits than the other two tissues, and DEGs encoding *PAL, 4CL, HCT, CCoAMT,* and *F3'5'H* were highly expressed in the samples from roots than the other two tissues. It is noteworthy that *PAL* and *4CL* display high expression in roots, but the flavonoids are not concentrated in the root [16]. It is speculated that in the initial stages of flavonoid synthesis, phenylpropanoid biosynthesis pathway initiates synthesis of substrates in the root, part of which is converted into eriodictyol by HCT, CCoAMT, and F3'5'H, and the rest is transported to the fruit, where it is modified and processed by CHS, CHI, F3H, FLS, ANS, DFR, and ANR to form flavonoids, flavones, and flavonols (Fig. 5). Therefore, it reasonable to primarily utilize fruits of *A. oxyphylla* as components of traditional medicine, rather than the root as done in species such as *A. officinarum*. These results provide insights into the molecular processes of flavonoid biosynthesis in *A. oxyphylla* and offer a

significant resource for the application of genetic engineering to develop varieties of *A. oxyphylla* with improved quality.

Methods

Plant Material

A. oxyphylla were collected from cultivated fields in Baisha County, Hainan Province, China (N.109.437569, E.19.19680). The fruits were sampled at the following three developmental stages: early-fruit (15 days), middle-fruit (30 days) and late-fruit (45 days). Fresh *A. oxyphylla* fruits were obtained from the three plants simultaneously during each phase. Then, the materials of same phase were mixed for further experiments. After harvesting the fruit, the leaves and roots were obtained from the same plant. All the samples of *A. oxyphylla* were immediately frozen in liquid nitrogen and stored at -80℃ prior to processing.

RNA sequencing and De Novo Assembly

The total RNA was extracted from different plant tissues using the RNAprep Pure Plant Kit (Tiangen, Beijing, China) as per the standard protocol [32]. The RNA concentration and quantity were assessed using the Nanodrop 2000 spectrometer (Thermo Fisher Scientific, Wilmington, DE, USA) and Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). A Stranded Total RNA Library Prep Kit (Illumina, Inc., San Diego, AR, USA) was used for cDNA library construction and normalization. The cDNA library was sequenced using Illumina HiSeq 4000 as per standard protocol. Raw reads were filtered by removing the adapter and low-quality sequences to produce high-quality clean reads and the reads were assembled to generate unigene libraries. Trinity software (v.2.8.5, the Broad Institute, Cambridge, MA, USA) was used to assemble the clean data into unigenes according to a basic group quality score of more than Q30 [33].

Functional Annotation

Function annotation of the assembled unigenes were obtained from public databases NCBI non-redundant protein (Nr) (http://www.ncbi.nlm.nih.gov), Universal Protein (Uniport) (https://www.uniprot.org/), EuKaryotic Orthologous Groups (KOG) (ftp://ftp.ncbi.nih.gov/pub/COG/KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) classifications (http://www.genome.jp/kegg/).

Analysis of DEGs

Unigene expression level was calculated using the fragments per kilobase of transcript per million mapped (FPKM) method. The DEGs were screened using the edgeR package with the threshold set as described previously [34]. GO and KEGG enrichment analysis of the identified DEGs was performed using the GOAtools version 0.5.9 (https://github.com/tanghaibao/Goatools) and KOBAS version 2.0.12 with default settings, respectively. The corrected p-value for identifying significant differences in expression was calculated and adjusted by the hypergeometric Fisher exact test. GO terms with a corrected p-value < 0.05 were considered to be significantly enriched. Next, we employed the same method for KEGG pathway functional enrichment analysis of DEGs.

Supplementary Materials: The following are available online at www.xx/xxx/s1, Figure S1: GO classification of assembled unigenes of *A. Oxyphylla*., Figure S2: KOG classification of assembled unigenes of *A. Oxyphylla.*, Figure S3: KEGG functional classification of assembled unigenes of *A. Oxyphylla*., Figure S4: Diferentially expressed genes (DEGs) involved in the flavonoids biosynthesis pathways in leaf vs fruit, Table S1: Expression level of candidate *A. oxyphylla* unigenes coding for key enzymes involved in flavonoid biosynthesis pathways.

Author Contributions: L.Y. and B.G. performed the experiments, data analysis, and the writing of the manuscript; K.P. and Y.L. prepared the sample and the part of data analysis; B.G. and B.Y. made revisions to the final manuscript. All authors have read and approved the final manuscript.

Funding: This work is supported by National Natural Science Foundation of China (No. 81560611) and Hainan Provincial Keypoint Research and Invention Program (ZDYF2018138).

Acknowledgments: We would like to thank Editage (www.editage.cn) for English language editing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gao, B.; Yuan, L.; Tang, T.; Hou, J.; Pan, K.; Wei, N. The complete chloroplast genome sequence of *Alpinia oxyphylla* Miq. and comparison analysis within the Zingiberaceae family. *PLoS One* **2019**, 14, e0218817.
2. Wang, H.; Liu, X.; Wen, M.; Pan, K.; Zou, M.; Lu, C.; Liu, S.; Wang, W. Analysis of the genetic diversity of natural populations of *Alpinia oxyphylla* Miquel using inter-simple sequence sepeat markers. *Crop Science* **2012**, 52,1767-1775.
3. Sharifi-Rad, M.; Varoni, E.M.; Salehi, B.; Sharifi-Rad, J.; Matthews, K.R.; Ayatollahi, S.A.; Kobarfard, F.; Ibrahim, S.A.; Mnayer, D.; Zakaria, Z.A.; Sharifi-Rad, M.; Yousaf, Z.; Iriti, M.; Basile, A.; Rigano, D. Plants of the genus Zingiber as a source of bioactive phytochemicals: from tradition to pharmacy. *Molecules* **2017**, 22, E2145.
4. Zhang, J.Q.; Wang, S.; Li, Y.H.; Xu, P.; Chen, F.; Tan, Y.F.; Duan, J.A. Anti-diarrheal constituents of *Alpinia oxyphylla. Fitoterapial*. **2013**, 89, 149-156.
5. Zhang, Q.; Zheng, Y.; Hu, X.; Hu, X.; Lv, W.; Lv, D.; Chen, J.; Wu, M.; Song, Q.; Shentu, J. Ethnopharmacological uses, phytochemistry, biological activities, and therapeutic applications of *Alpinia oxyphylla* Miquel: A review. *J Ethnopharmacol*. **2018**, 224,149-168.
6. Duan, L.H.; Li, M.; Wang, C.B.; Wang, Q.M.; Liu, Q.Q.; Shang, W.F.;Shen, Y.J.; Lin, Z.H.; Sun, T.Y.; Wu, Z.Z.; Li, Y.H.; Wang, Y.L.; Luo, X. Protective effects of organic extracts of *Alpinia oxyphylla* against hydrogen peroxide-induced cytotoxicity in PC12 cells. *Neural Regen Res*. **2020**, 15, 682-689.
7. Huang, K.K.; Lin, M.N.; Hsu, Y.L.; Lu, I.H.; Pan, I.H.; Yang, J.L. *Alpinia oxyphylla* fruit extract ameliorates experimental autoimmune encephalomyelitis through the regulation of Th1/Th17 Cells. *Evid Based Complement Alternat Med*. **2019**, 6797030.
8. Xie, Y.; Xiao, M.; Ni, Y.; Jiang, S.; Feng, G.; Sang, S.; Du, G. *Alpinia oxyphylla* Miq. extract prevents diabetes in mice by modulating gut Microbiota. *J Diabetes Res*. **2018**, 4230590.
9. Xu, J.; Wang, F.; Guo, J.; Xu, C.; Cao, Y.; Fang, Z.; Wang, Q. Pharmacological mechanisms underlying the neuroprotective effects of *Alpinia oxyphylla* Miq. on Alzheimer's disease. *Int J Mol Sci*. **2020**, 21, 2071.
10. Sun, Z.; Kong, X.; Zuo, L.; Kang, J.; Hou, L.; Zhang, X. Rapid extraction and determination of 25 bioactive constituents in *Alpinia oxyphylla* using microwave extraction with ultra high performance liquid chromatography with tandem mass spectrometry. *J Sep Sci*. **2016**, 39, 603-610.
11. Li, Y.H.; Chen, F.; Wang, J.F.; Wang, Y.; Zhang, J.Q.; Guo, T. Analysis of nine compounds from *Alpinia oxyphylla* fruit at different harvest time using UFLC-MS/MS and an extraction method optimized by orthogonal design. *Chem Cent J*. **2013**, 7, 134.

12. Niu, Q.; Gao, Y.; Liu, P. Optimization of microwave-assisted extraction, antioxidant capacity, and characterization of total flavonoids from the leaves of *Alpinia oxyphylla* Miq. *Prep Biochem Biotechnol*. **2020**, 50, 82-90.
13. He, B.; Xu, F.; Yan, T.; Xiao, F.; Wu, B.; Wang, Y.; Bi, K.; Jia, Y. Tectochrysin from *Alpinia Oxyphylla* Miq. alleviates Abeta1-42 induced learning and memory impairments in mice. *Eur J Pharmacol*. **2019**, 842, 365-372.
14. Williams, C.A.; Goldstone, F.; Greenham, J. Flavonoids, cinnamic acids and coumarins from the different tissues and medicinal preparations of Taraxacum officinale. *Phytochemistry* **1996**, 42, 121-127.
15. Weng, Z.; Zeng, F.; Zhu, Z.; Qian, D.; Guo, S.; Wang, H.; Duan, J.A. Comparative analysis of sixteen flavonoids from different parts of *Sophora flavescens* Ait. by ultra high-performance liquid chromatography-tandem mass spectrometry. *J Pharm Biomed Anal*. **2018**, 156, 214-220.
16. Li H.; Tan Y.; Wang Y.; Wei N.; Li Y.; Zhang J. Chemical constituents of flavones part from the stems and leaves of *Alpinia oxyphylla* Miq. *Nat Prod Res Dev.* **2014**, 26:1038-1042.
17. Li, Y.H.; Chen, F.; Wang, J.F.; Wang, Y.; Zhang, J.Q.; Guo, T. Analysis of nine compounds from *Alpinia oxyphylla* fruit at different harvest time using UFLC-MS/MS and an extraction method optimized by orthogonal design. *Chem Cent J*. **2013**, 7, 134.
18. Sun, L.; Yu, D.; Wu, Z.; Wang, C.; Yu, L.; Wei, A.; Wang, D. Comparative transcriptome analysis and expression of genes reveal the biosynthesis and accumulation patterns of key flavonoids in different varieties of *Zanthoxylum bungeanum* leaves. *J Agric Food Chem*. **2019**, 67, 13258-13268.
19. Liang, W.; Ni, L.; Carballar-Lejarazu, R.; Zou, X.; Sun, W.; Wu, L.;Yuan, X.; Mao, Y.;Huang, W.; Zou, S. Comparative transcriptome among *Euscaphis konishii* Hayata tissues and analysis of genes involved in flavonoid biosynthesis and accumulation. *BMC Genomics* **2019**, 20, 24.
20. Yang, M.; Zhou, P.N.; Gui, C.; Da, G.Z.; Gong, L.; Zhang, X.Q. Comparative transcriptome analysis of *Ampelopsis megalophylla* for identifying genes involved in flavonoid biosynthesis and accumulation during different seasons. *Molecules* **2019**, 24, 1267.
21. Fan, H.; Li, K.; Yao, F.; Sun, L.W.; Liu, Y.J. Comparative transcriptome analyses on terpenoids metabolism in field- and mountain-cultivated ginseng roots. *Bmc Plant Biol*. **2019**, 19, 82.
22. Yang, F.; Wei, N.N.; Gao, R.; Piao, X. C.; Lian, M.L. Effect of several medium factors on polysaccharide and alkaloid accumulation in protocorm-like bodies of *Dendrobium candidum* during bioreactor culture. *Acta Physiol Plant*. **2015**, 37, 94.
23. Zhang, F.S.; Wang, Q.Y.; Pu, Y.J.; Chen, T.Y.; Qin, X.M.; Gao, J. Identification of genes involved in flavonoid biosynthesis in *Sophora japonica* through transcriptome sequencing. *Chem Biodivers*. **2017**, 14.
24. Joshi, R. K.; Mohanty, S.; Kar, B.; Nayak, S. Assessment of genetic diversity in Zingiberaceae through nucleotide binding site-based motif-directed profiling. *Biochem Genet.* **2012**, 50(7-8), 642-656.
25. Basri, A.M.; Taha, H.; Ahmad, N. A review on the pharmacological activities and phytochemicals of *Alpinia officinarum* (Galangal) extracts derived from bioassay-guided fractionation and isolation. *Pharmacogn Rev*. **2017**, 11, 43-56.
26. Liang, J.; Li, W.; Jia, X.; Zhang, Y.; Zhao, J. Transcriptome sequencing and characterization of *Astragalus membranaceus* var. mongholicus root reveals key genes involved in flavonoids biosynthesis. Genes Genomics. **2020**, doi:10.1007/s13258-020-00953-5.

27. Gao, G.; Chen, P.; Chen, J.; Chen, K.; Wang, X.; Abubakar, A.S.; Liu N.; Yu, C.; Zhu, A. Genomic survey, transcriptome, and metabolome analysis of *Apocynum venetum* and *Apocynum hendersonii* to reveal major flavonoid biosynthesis pathways. Metabolites, **2019**, 9(12):296.
28. Li, L.; Liu, M.; Shi, K.; Yu, Z.; Zhou, Y.; Fan, R.; Shi, Q. Dynamic changes in metabolite accumulation and the transcriptome during leaf growth and development in *Eucommia ulmoides*. Int. J. Mol. Sci. **2019**, 20(16):4030.
29. Zhu, J. H.; Cao, T.J.; Dai, H.F.; Li, H.L.; Guo, D.; Mei, W.L.; Peng, S.Q. De Novo transcriptome characterization of *Dracaena cambodiana* and analysis of genes involved in flavonoid accumulation during formation of dragon's blood. *Sci Rep*. **2016**, 6, 38315.
30. Hamamouch, N.; Winkel, B.S.J.; Li, C.; Davis, E.L. Modulation of arabidopsis flavonol biosynthesis genes by cyst and root-knot nematodes. *Plants (Basel)*. **2020**, 9, 253.
31. Zuk, M.; Szperlik, J.; Hnitecka, A.; Szopa, J. Temporal biosynthesis of flavone constituents in flax growth stages. *Plant Physiol Biochem*. **2019**, 142, 234-245.
32. Wu, X.; Chen, Y.; Hou, J.; Gao, B. Comparing study for isolating of total RNA in fruits of southern medicine Alpinia oxyphylla Miquel. *Lishizhen Medicine and Materia Medica Research* **2018**, 29, 766-768.
33. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.;Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; Palma, F.D.; Birren, B.W.; 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.
34. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, 26, 139-140.

Appendix A

Figure S1. GO classification of assembled unigenes of *A. Oxyphylla*. A total of 218,989 unigenes were annotated into three categories: molecular function (MF, blue), cellular component (CC, green) and biological process (BP, red).

Figure S2. KOG classification of assembled unigenes of *A. Oxyphylla.* [J] Translation, ribosomal structure and biogenesis; [A] RNA processing and modification; [K] Transcription; [L] Replication; recombination and repair; [B] Chromatin structure and dynamics; [D] Cell cycle control, cell division, chromosome partitioning; [Y] Nuclear structure; [V] Defense mechanisms; [T] Signal transduction mechanisms; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [Z] Cytoskeleton; [W] Extracellular structures; [U] Intracellular trafficking, secretion and vesicular transport; [O] Posttranslational modification, protein turnover, chaperones; [C] Energy production and conversion; [G] Carbohydrate transport and metabolism; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [P] Inorganic ion transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism; [R] General function prediction only; [S] Function unknown.

Figure S3. KEGG functional classification of assembled unigenes of *A. Oxyphylla*. The unigenes were divided into five primary categories: (A) cellular processes, (B) environmental information processing, (C) genetic information processing, (D) metabolism, and (E) organismal systems. The x-axis represents the number of unigenes, whereas the y-axis represents the functional categories.

Figure S4. Diferentially expressed genes (DEGs) involved in the flavonoids biosynthesis pathways in leaf vs fruit. The red columns indicate genes expressed at a signifcantly high level in the fruit of *A. Oxyphylla*. This color-coded map of DEGs corresponds to map00941 in the KEGG database.

Table S1. Expression level of candidate *A. oxyphylla* unigenes coding for key enzymes involved in flavonoid biosynthesis pathways.

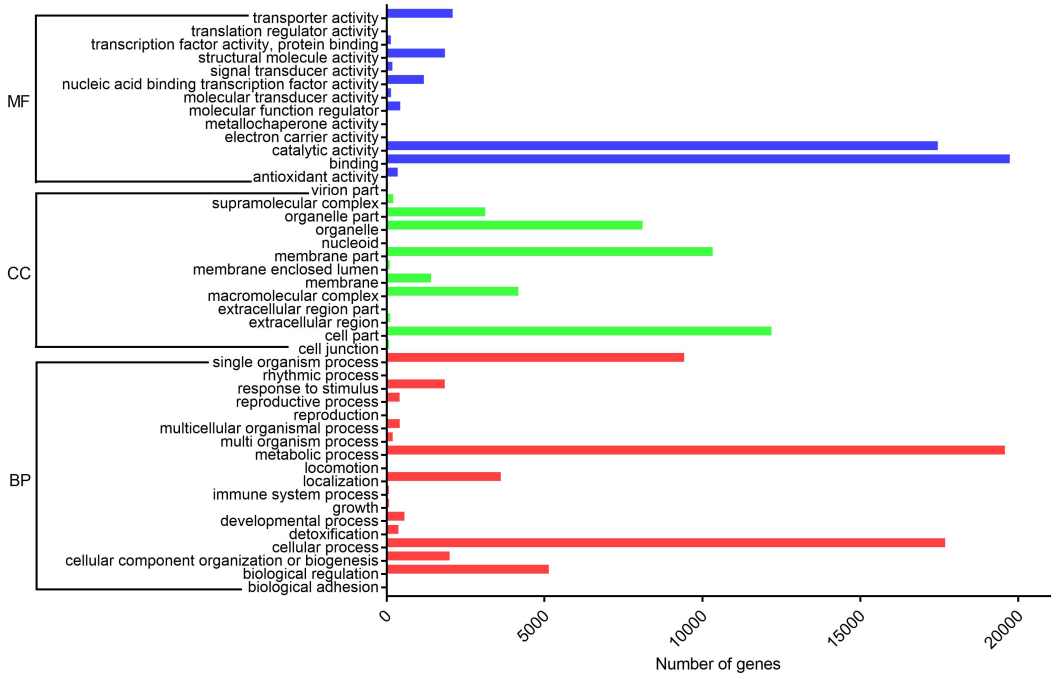

Figure S1

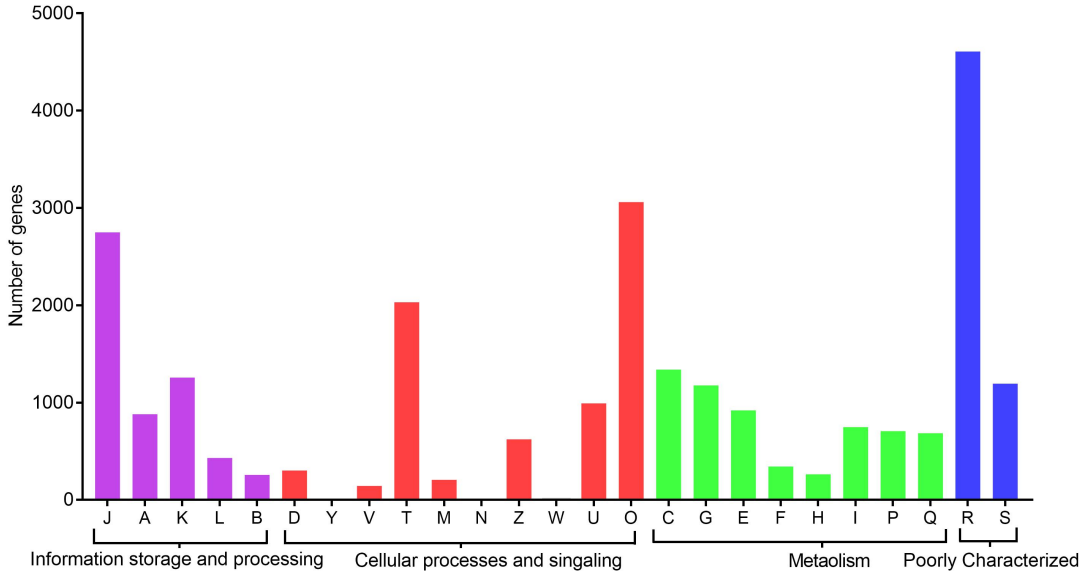

Figure S2

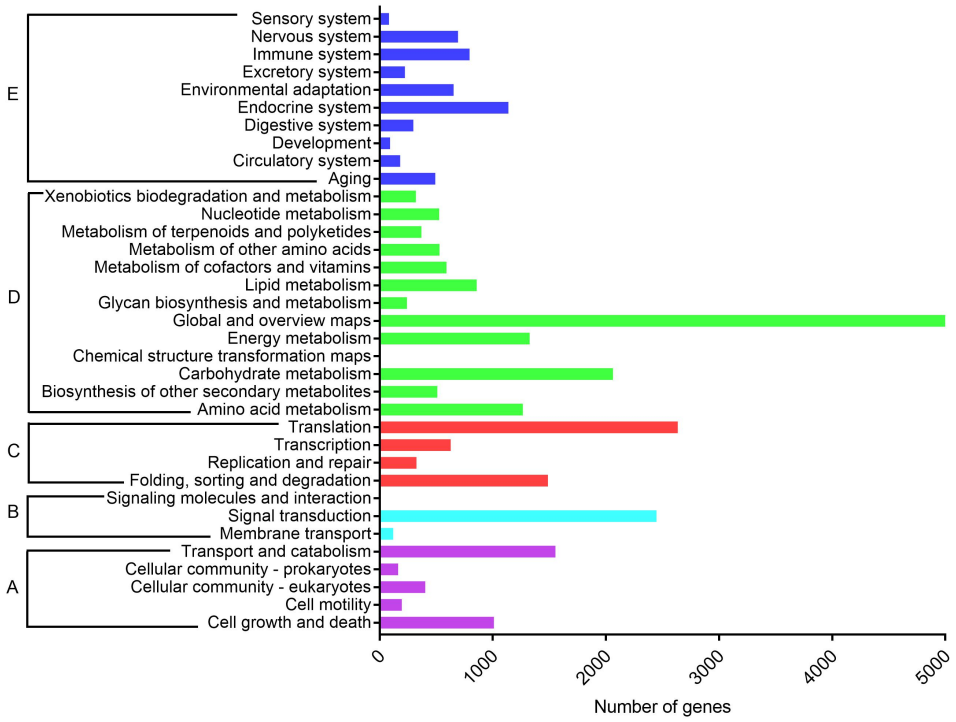

Figure S3

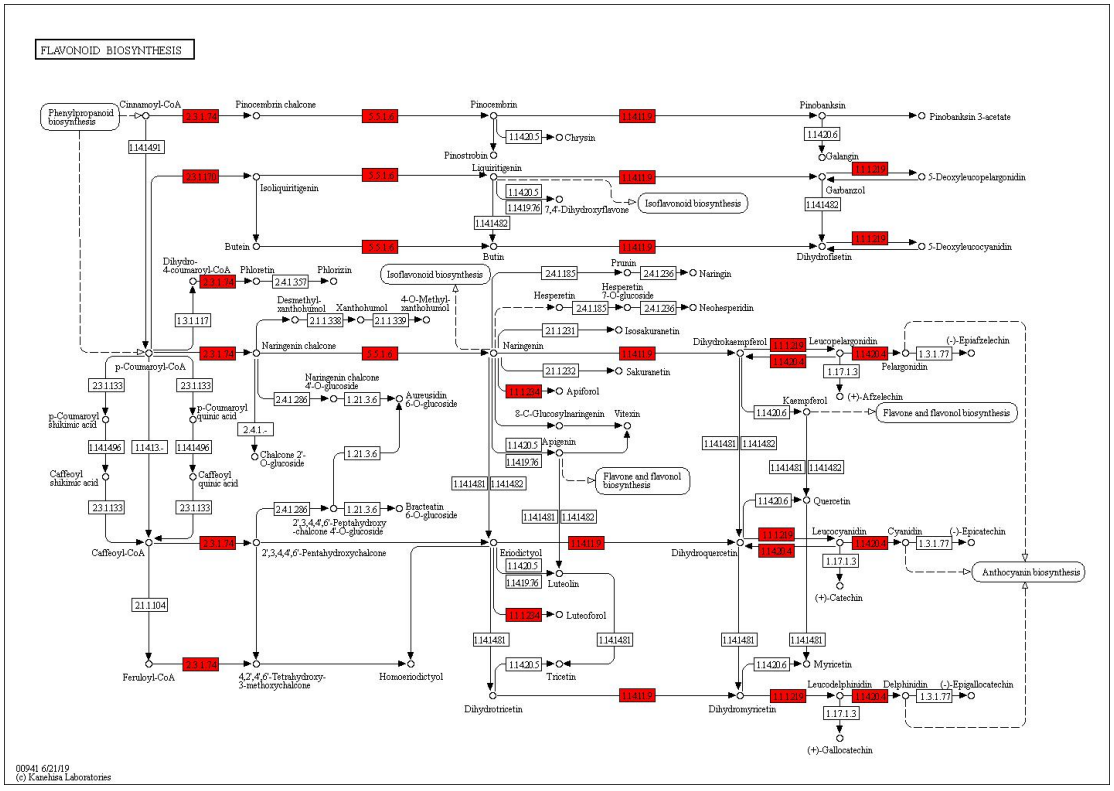

Figure S4

Table S1.

Name	Gene ID	FPKM

		early-fruit	middle-fruit	late-fruit	root	leaf
FLS	unigene139585	458.62	179.04	82.35	0.09	59.11
F3’5’H	unigene117017	27.31	12.91	14.99	49.87	0.12
ANR	unigene115607	1020.66	2113.57	1768.39	498.78	337.65
HCT	unigene109484	4.55	6.74	2.05	3.14	8.03
HCT	unigene109483	68.88	53.75	71.79	210.44	70.62
4CL	unigene108821	43.81	37.85	42.6	128.7	6.12
4CL	unigene108819	98.89	150.82	118.44	39.41	5.2
4CL	unigene108818	2.7	2.03	1.94	68.18	7.12
C4H	unigene106999	51.68	34.58	131.28	60.8	0.37
4CL	unigene095558	10.83	25.16	19.73	7.38	23.21
DFR	unigene092310	307.69	440.88	231.95	73.62	2.67
PAL	unigene077933	12.25	22.33	18.37	12.17	1.38
F3H	unigene076049	4797.64	6785.61	4773.57	2370.02	427.85
ANS	unigene074219	911.45	1216.78	774.03	348.66	9.27
HCT	unigene067708	53.92	45.31	53.43	26.37	75.53
CHS	unigene065545	43.39	13.51	8.57	46.91	3.58
CHS	unigene065542	1158.04	870.33	562.51	81.74	160.18
CCoAMT	unigene053332	355.08	147.74	245.95	591.43	53.1
4CL	unigene050620	1.34	1.89	5.85	0	0
CHI	unigene042446	1129.52	588.88	300.05	235.97	17.68
4CL	unigene023081	33.21	34.13	35.64	96.7	14.42
4CL	unigene023078	28.25	35.02	21.21	35.23	8.07
FLS	unigene018259	15.21	0.97	0.1	23.87	0.19
FLS	unigene018258	25.59	1.04	0	38.39	0
PAL	unigene017890	165.85	211.45	167.33	409.88	56.49
F3’5’H	unigene007819	7.61	1.53	1.58	41.96	3.93
F3’5’H	unigene007815	116.68	12.14	11.78	219.54	153.46

434

Figures

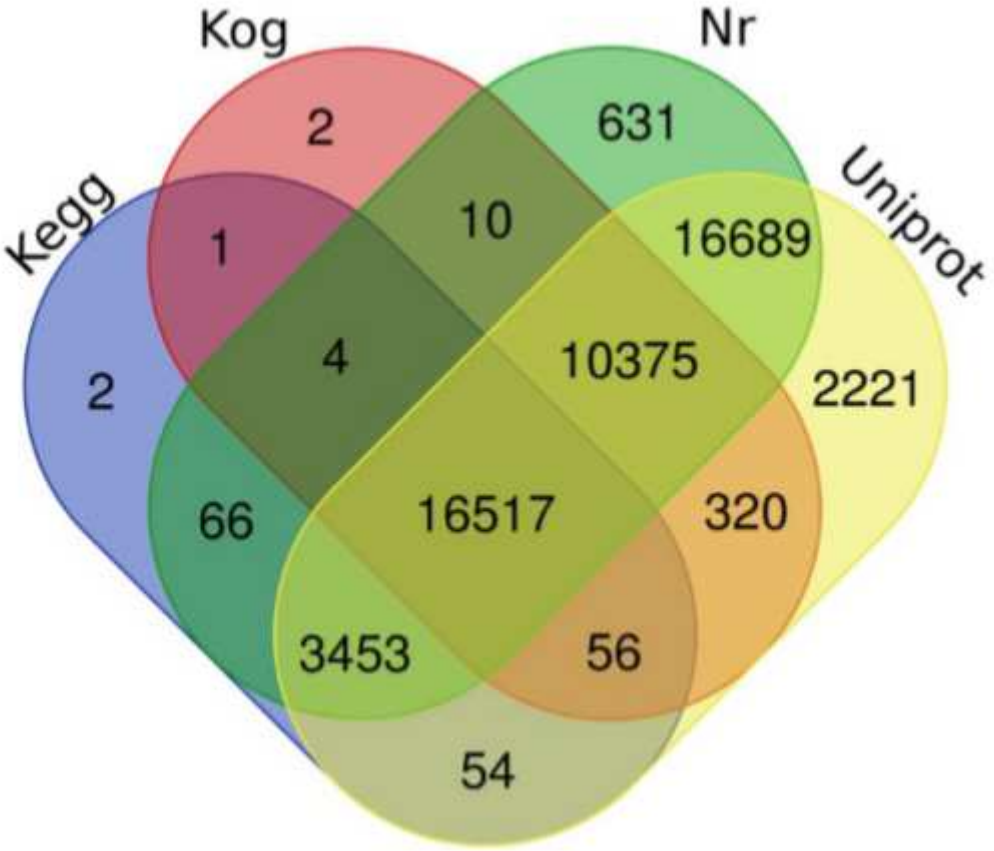

Figure 1

Venn diagram describing the unigenes annotated by four different databases. The integration of unique similarity search results against the NCBI non-redundant protein (Nr), Universal Protein (Uniport), EuKaryotic Orthologous Groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG)databases.

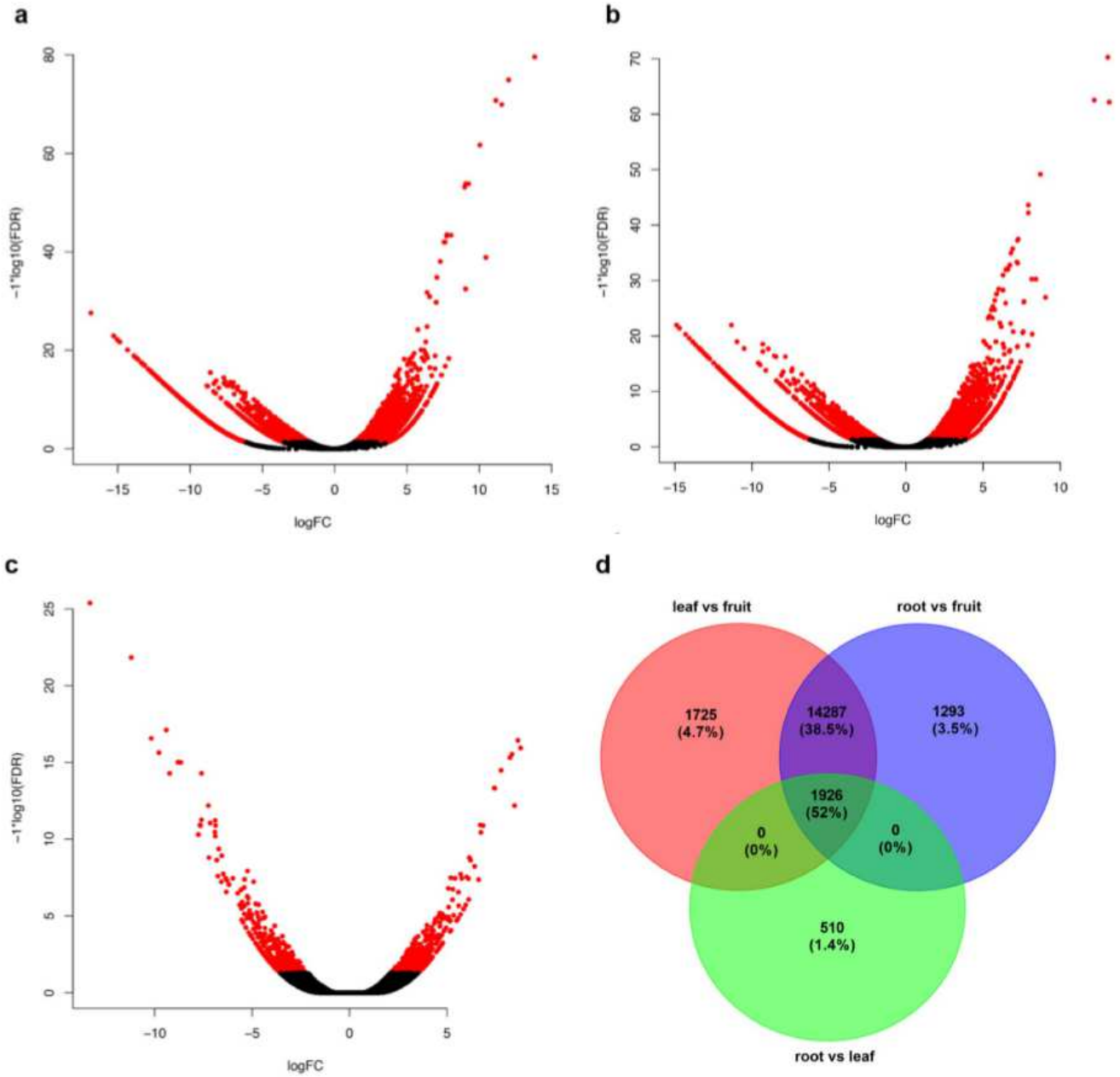

Figure 2

Volcano plots of the differentially expressed genes (DEGs) in the comparison group of (a) leaf vs. fruit, (b) root vs fruit, and (c) root vs. leaf. Venn diagram of DEGs in three different comparisons groups represented by three circles. The overlapping parts of the circles represent the numberofDEGsincommoninthecomparisongroups.

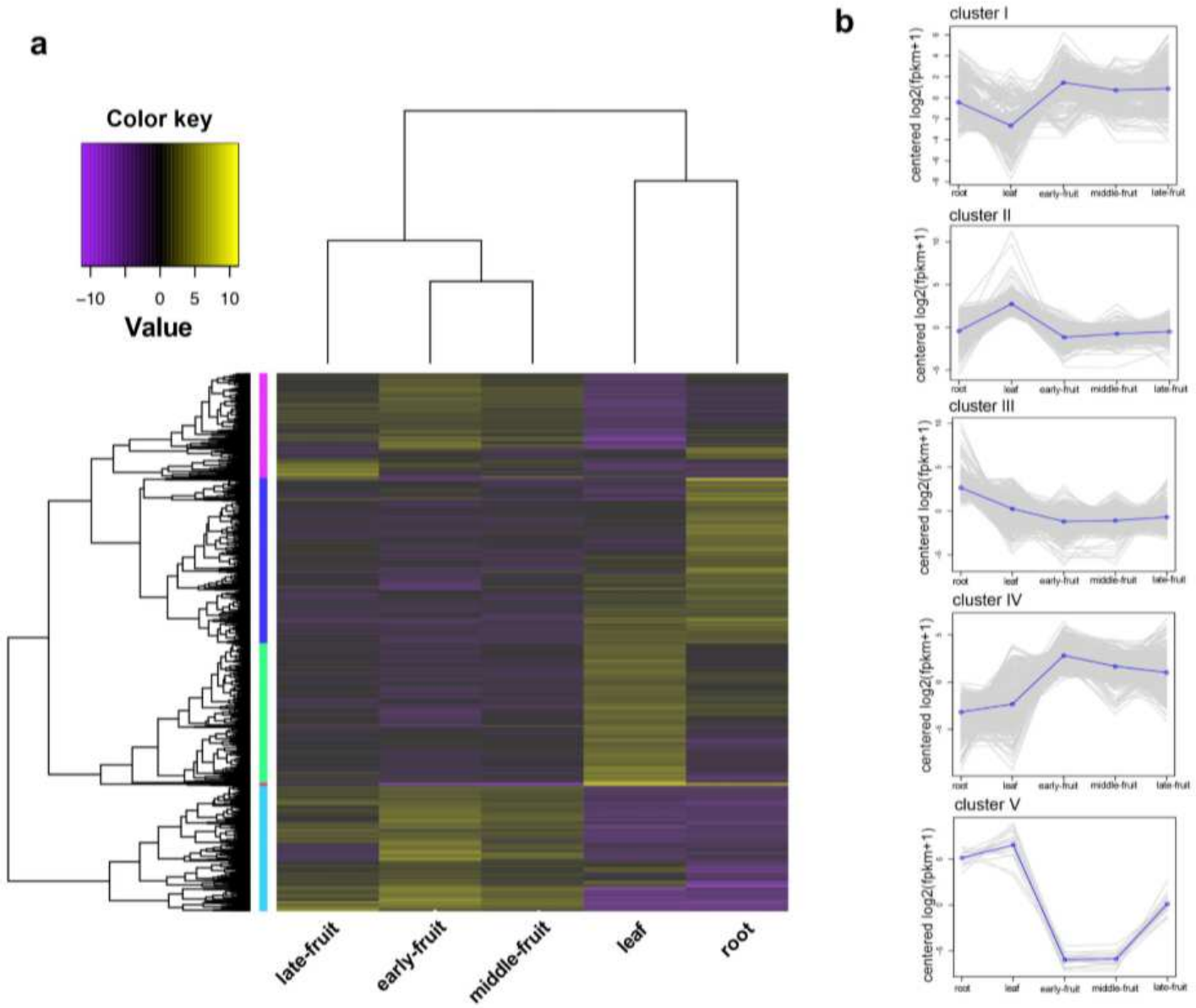

Figure 3

Cluster analysis of DEGs (a) Heat-map showing the expression of DEGs using RNA-seq data derived from mean value of three replicates of each sample based on log 2 (FPKM) values. Color code indicates expression levels. Similarity between samples and unigenes with hierarchical clustering is shown above and on the left of the heatmap, respectively. (b) Cluster analysis of all DEGs. The y-axis in each graph represents the mean-centered log2 (FPKM+1) value. Expression of a single gene1 isplottedingray,whilethemeanexpressionofthegenesineachclusterisplottedinblue.

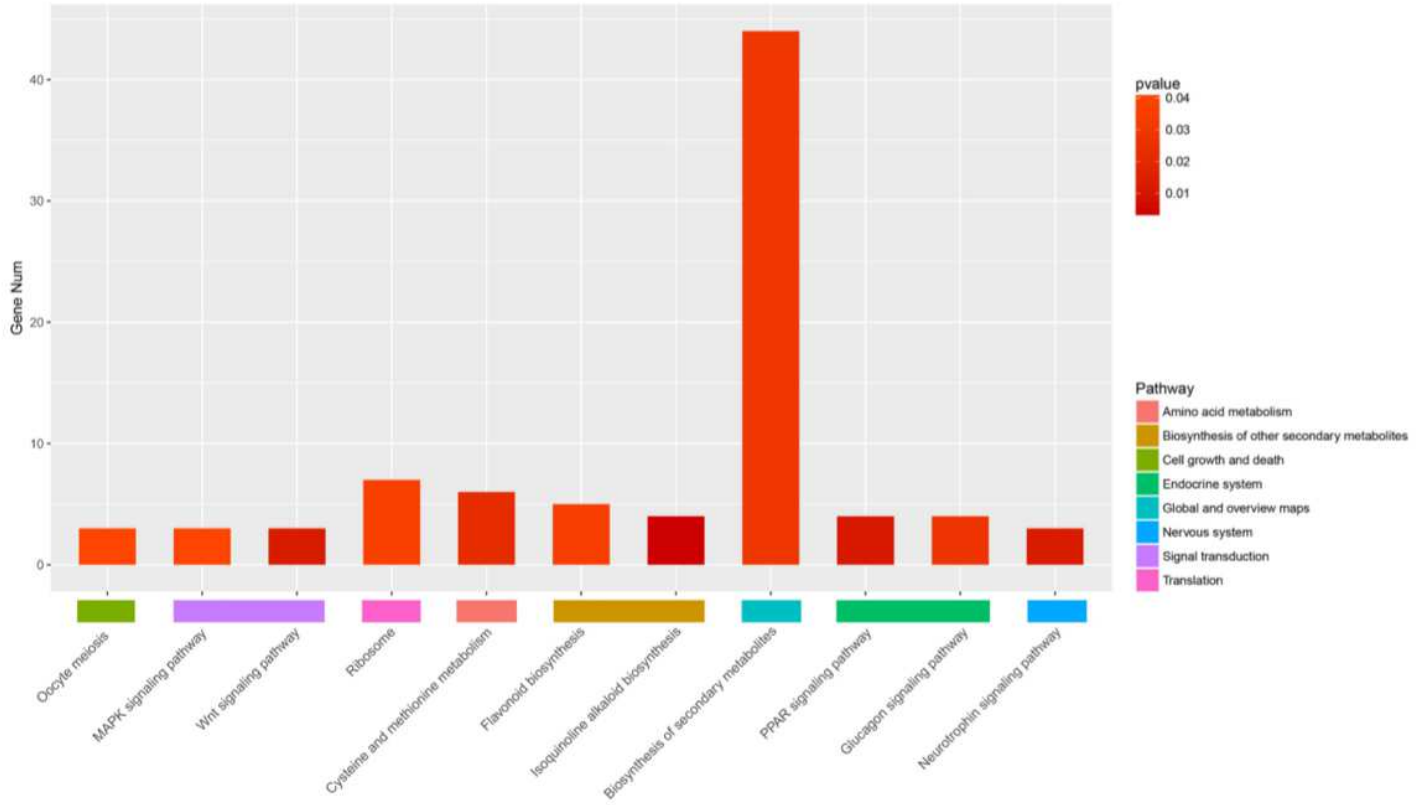

Figure 4

Distribution map of DEGs in cluster I signaling pathway

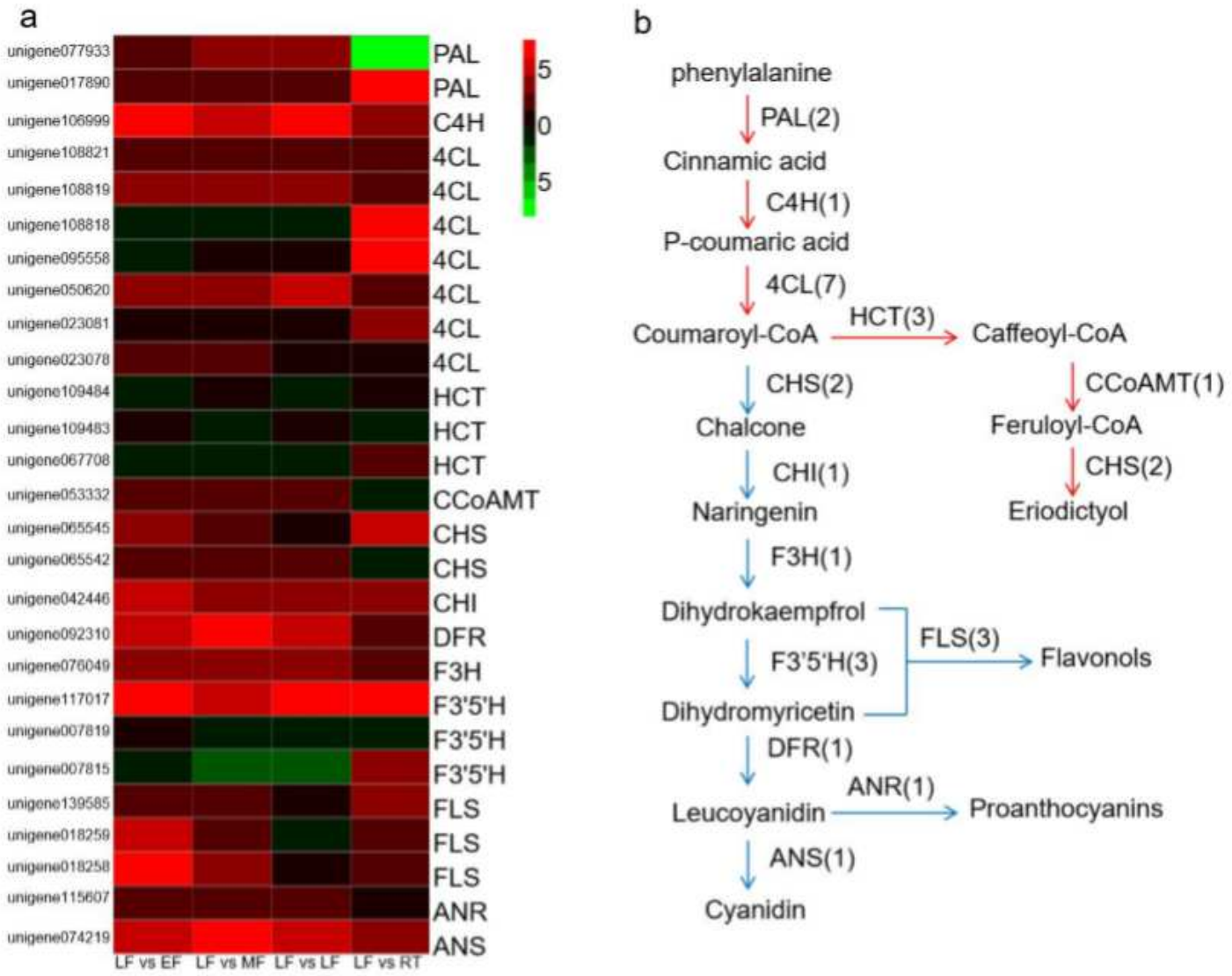

Figure 5

Putative flavonoidbiosynthesispathway inA. oxyphylla. (a)Expression level ofcandidateA. oxyphylla unigenes coding for key enzymes involved in flavonoid biosynthesis pathways. Green and red colors are used to represent low-to-high expression levels (mean centered log2-transformed FPKM values). (b) Pathway for flavonoid biosynthesis. The numbers in brackets following each gene name indicate the number of A. oxyphylla unigenes corresponding to that gene. Enzyme abbreviations are as follows: PAL, phenylalanineammonia-lyase; C4H, cinnamate 4-hydroxylase; CHS, chalcone synthase;CCoAMT, Caffeoyl Co-A transferase; 4CL, 4-coumarate-CoA ligase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3’5’H, flavonoid-3’, 5’-hydroxylase; DFR, dihydroflavonol-4-reductase; flavonoid-3-O-glucosyltransferase; ANR,anthocyanidinreductase;ANS,anthocyanidinsynthase.