

# SCIENTIFIC REPORTS



OPEN

## *De Novo* transcriptome characterization of *Dracaena cambodiana* and analysis of genes involved in flavonoid accumulation during formation of dragon's blood

Received: 19 September 2016

Accepted: 07 November 2016

Published: 06 December 2016

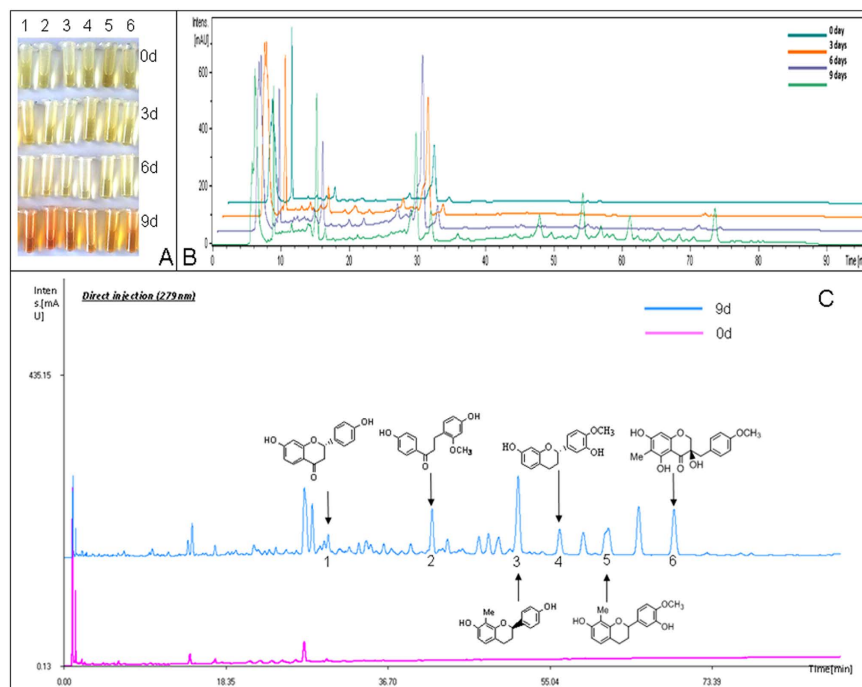
Jia-Hong Zhu\*, Tian-Jun Cao\*, Hao-Fu Dai, Hui-Liang Li, Dong Guo, Wen-Li Mei & Shi-Qing Peng

Dragon's blood is a red resin mainly extracted from *Dracaena* plants, and has been widely used as a traditional medicine in East and Southeast Asia. The major components of dragon's blood are flavonoids. Owing to a lack of *Dracaena* plants genomic information, the flavonoids biosynthesis and regulation in *Dracaena* plants remain unknown. In this study, three cDNA libraries were constructed from the stems of *D. cambodiana* after injecting the inducer. Approximately 266.57 million raw sequencing reads were *de novo* assembled into 198,204 unigenes, of which 34,873 unique sequences were annotated in public protein databases. Many candidate genes involved in flavonoid accumulation were identified. Differential expression analysis identified 20 genes involved in flavonoid biosynthesis, 27 unigenes involved in flavonoid modification and 68 genes involved in flavonoid transport that were up-regulated in the stems of *D. cambodiana* after injecting the inducer, consistent with the accumulation of flavonoids. Furthermore, we have revealed the differential expression of transcripts encoding for transcription factors (MYB, bHLH and WD40) involved in flavonoid metabolism. These *de novo* transcriptome data sets provide insights on pathways and molecular regulation of flavonoid biosynthesis and transport, and improve our understanding of molecular mechanisms of dragon's blood formation in *D. cambodiana*.

Dragon's blood is a red resin excreted by a part of the genus *Dracaena* plants, which has been used as a famous traditional medicine since ancient times<sup>1</sup>. Modern pharmacological studies have found that this resinous medicine has anti-bacterial, anti-spasmodic, anti-inflammatory, analgesic, anti-diabetic, and anti-tumor activities, and is widely applied for the treatment of wounds, leucorrhoea, fractures, diarrhea and piles as well as for intestinal and stomach ulcers<sup>1,2</sup>. In China, *Dracaena cochinchinensis* and *Dracaena cambodiana* were the mainly plant resources of dragon's blood<sup>3</sup>. In nature, *Dracaena* plants grow very slowly, and only trees with 30–50-years-old have the possibility to produce a small amount of dragon's blood. In addition, natural *Dracaena* resources have been destroyed severely owing to overexploitation<sup>4,5</sup>. Our previous studies found an effective inducer (authorized patent: ZL 201310207182.5) could induce the formation of the resin in the stem of *D. Cambodiana*<sup>6</sup>. However, the formation mechanism of dragon's blood is still unknown. Phytochemical studies of dragon's blood revealed that flavonoids are the main chemical constituents, while terpenes, steroids, saponins and phenols have also been identified as constituents<sup>7–11</sup>. Owing to a lack of *Dracaena* plants genomic information, the molecular basis of specific flavonoids biosynthesis and the regulation of formation in *Dracaena* plants remain unknown.

The flavonoids metabolism pathway (especially anthocyanin) has been well characterized in model plants at the genetic, biochemical and molecular levels<sup>12,13</sup>. Given the distribution of flavonoids vary differently among different plants, the molecular mechanisms of their biosynthesis, transport and regulation might be diverse

Key Laboratory of Biology and Genetic Resources of Tropical Crops, Ministry of Agriculture, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, No. 4 Xueyuan Road, Haikou 571101, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to W.-L.M. (email: meiwenli@itbb.org.cn) or S.-Q.P. (email: shqpeng@163.com)



**Figure 1.** Changes of ethanol extracts from the stem of *D. cambodiana* after injecting the inducer. (A) The changes of color of ethanol extracts from the stem of *D. cambodiana* of 0 d, 3 d, 6 d and 9, d after injecting the inducer. (B) HPLC analysis of ethanol extracts from the stem of *D. cambodiana* of 0 d, 3 d, 6 d and 9, d after injecting the inducer. (C) Flavonoids were identified from ethanol extracts from the stem of *D. cambodiana* of 9, d after injecting the inducer.

and complex<sup>14</sup>. Therefore, it is essential to investigate the molecular mechanisms of flavonoids accumulation in *D. cambodiana*. However, the lack of the transcriptomic and genomic data has made it difficult to investigate the mechanisms of flavonoid accumulation and dragon's blood formation in *D. cambodiana*.

Next generation sequencing platforms provide highly efficient tools to identify novel genes associated with biosynthesis of various secondary metabolites in non-model plant species<sup>15</sup>. Specially, it has been widely applied to investigate molecular mechanisms of flavonoids accumulation in plant species such as *Dioscorea alata*<sup>16</sup>, *Prunus avium*<sup>17</sup>, *Olea europaea*<sup>18</sup> and *Gentiana straminea*<sup>19</sup>. Here, we present the transcriptome analysis of *D. cambodiana* using RNA-Seq complemented with HPLC flavonoid profiling to identify potential candidate genes involved in flavonoids accumulation. This study, therefore, may serve as a basis for the future discoveries on pathways and molecular regulation of flavonoid biosynthesis and transport in *D. cambodiana*.

## Results

**Increased flavonoids in the stem of *D. cambodiana* after injecting the inducer.** The color of ethanol extracts from the stem of *D. cambodiana* changed from light yellow to brown after injecting the inducer (Fig. 1A). The compounds of ethanol extracts from the stem of *D. cambodiana* increased 9 d after injecting the inducer (Fig. 1B). We subsequently performed HPLC analysis to detect the flavonoids in the ethanol extract. As shown in Fig. 1C, some flavonoids were identified on the basis of their retention time compared to standards under the same condition (Table S1). The result shows the inducer can induce the formation of flavonoids in the stem of *D. cambodiana* after injecting the inducer.

**Assembling and annotation of *D. cambodiana* transcriptome.** To elucidate the molecular mechanism and to find candidate genes of dragon's blood formation, three cDNA libraries were obtained from the stems of *D. cambodiana* after injecting the inducer in 0 day, 3 days, 6 days respectively (designated 0d, 3d and 6d) and constructed. By sequencing on the platform of Illumina Hiseq 2000, total raw reads of 266.57 M were generated and trimmed to exclude low-quality reads (Table 1). To perform the *de novo* assembly, 217.5 M high-quality reads were further assembled into 198,204 unigenes with an average length of 495 bp, a maximum length of 21,883 bp and an N50 length of 562 bp. After removal of contaminant sequences, a total of 70,122 unigenes longer than 400 bps were selected for further analyses. The length distribution of the unigenes is illustrated in Fig. S1. The sequences ranging from 400 bp to 2,000 bp in length accounted for nearly 92.5% of the total. Up to 3,444 unigenes (4.9%) and 1,842 unigenes (2.6%) were 2,000 bp to 3,000 bp and >3000 bp in length, respectively.

The unique sequences were annotated using BLASTx against the NCBI non-redundant nucleotide database (Nt), the Swiss-Prot protein database, the KEGG database, and the COG database. A total of 34,873 (17.5%) unique sequences could be matched to known genes in the public databases (Table S2). The proportion of the unigenes with BLASTx hits significantly increased for longer unigenes (Fig. S1). The BLASTx searches yielded hits

Trimmed Illumina sequencing data					
		Number of reads	Total bases	N(%)	GC(%)
Raw	0d	87.09 M	8.62 G	0	47.69
	3d	75.87 M	7.51 G	0	48.03
	6d	103.61 M	10.26 G	0	48.64
Filter	0d	71.49 M	7.08 G	0	47.49
	3d	63.13 M	6.25 G	0	47.91
	6d	82.88 M	8.21 G	0	47.91
De novo assembly and annotation		De novo assembly		Annotation	
Total base(MB)		93.63		34.14	
Number of Transcript		198204		34783	
Max Unigene length (bp)		21883		21883	
Mean Unigene length		495		1029	
N50		562		1694	

**Table 1.** Statistical summary of trimmed Illumina sequencing data, *de novo* assembly and annotation.

for 18590 (28.7%) unigenes that were 400 bp to 2,000 bp, 3102 (90.1%) unigenes that were 2,000 bp to 3,000 bp, while 1748 (94.9%) of the unigenes longer than 3,000 bp were annotated. However, there were still a large number of unigenes (163,421, 82.5%) without BLASTX hits. Most of these sequences were short fragments, and some of them might be non-coding RNA sequences or new genes.

### Differentially expressed genes (DEGs) in the stem of *D. cambodiana* after injecting the inducer.

Expression levels of unigenes were determined by aligning the RNA-Seq reads from each library to the assembly. As a result, a total of 148,875 genes were expressed in all three samples, and 2,724, 1,698, 2,155 unigenes were specifically expressed in 0d, 3d and 6d respectively (Fig. 2a). The differences in gene expression patterns were then analyzed for the pairs 0d and 3d, 0d and 6d, 3d and 6d based on the false discovery rate (FDR)  $\leq 0.05$ , and fold change (FC)  $\geq 1$ . A total of 6,986 genes were differentially expressed between the 0d and 3d libraries, among which 4161 were up-regulated and 2,825 were down-regulated. Between the 0d and 6d libraries, 7,106 genes were differentially expressed, including 4,681 up-regulated and 2,425 down-regulated genes. The 3d and 6d libraries showed 6,085 DEGs, including 2451 up-regulated and 3634 down-regulated (Fig. 2b).

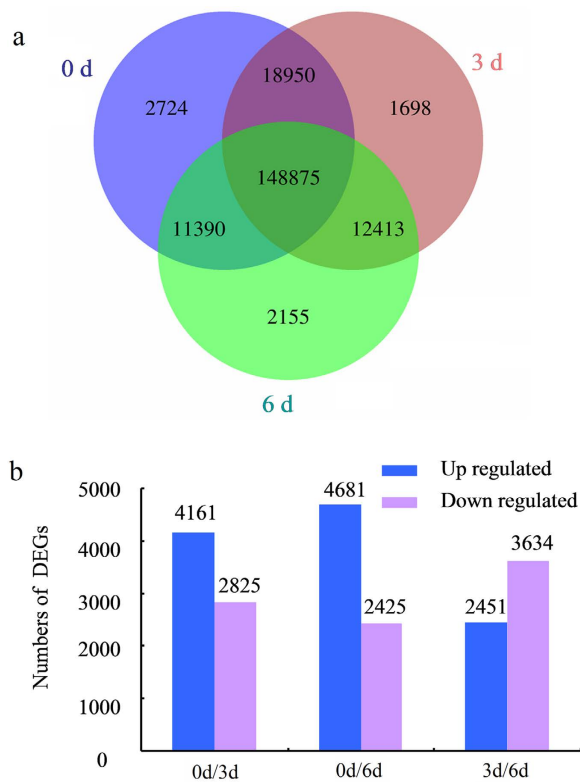
### GO enrichment and KEGG pathway analysis of DEGs.

We did functional categorize to the differential expression unigenes by using gene ontology (GO) terms. The GO terms allow for the definition and standardization of the properties of gene products in any organism. The 3d against 0d libraries, 1,297 different genes were annotated. Of the three main sub ontologies, molecular function was the highly represented, with 1,129 unigenes followed by biological process with 795 unigenes and finally, cellular component with 584 unigenes. Besides this, 1,711 and 1,606 unigenes were annotated in other groups (Fig. 3). GO enrichment analysis show detail information about the changes. GO enrich results with correct p-Value  $< 0.01$  show “molecular function” and “cellular component” between treated and control group. The major sub-categories are list in Table S3, including electron carrier activity (GO: 0009055), extracellular region (GO: 0005576), nucleic acid binding transcription factor activity (GO: 0001071), transporter activity (GO: 0005215), membrane (GO: 0016020), membrane part (GO: 0044425). After further analysis, we found the genes involved in electron carrier activity (GO: 0009055) and extracellular region (GO: 0005576) are mostly the cytochrome P450s which are generally taking an active part in secondary metabolism.

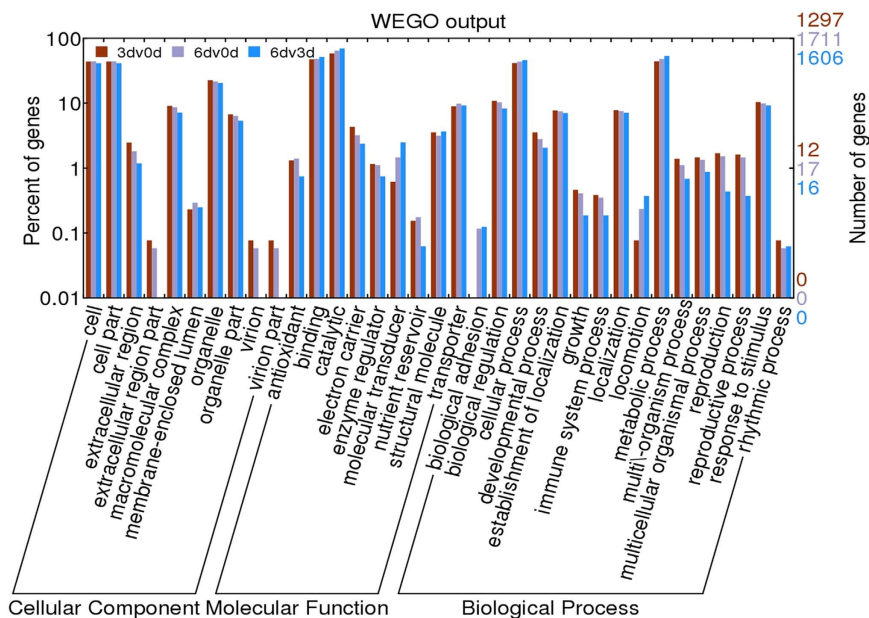
The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways represent collections of manually drawn pathway maps and that are helpful for the understanding of the biological functions and interactions of genes. By mapping Enzyme Commission (EC) numbers to the reference canonical pathways, a total of 998, 1,296, 1,419 distinct genes matched the metabolic pathways in the KEGG pathway database at 3d/0d, 6d/0d, 6d/3d, respectively. Phenylpropanoid biosynthesis, flavonoid biosynthesis and flavone and flavonol biosynthesis were the most prominent classes (Table S4). From the KEGG result, we can find secondary metabolite biosynthesis was animate in the *D. cambodiana*'s transcriptome, where flavonoids, flavone and flavonols, isoquinoline alkaloids, stilbenoid, diarylheptanoid and gingerol biosynthesis were the most relevant pathways by KEGG enrichment (Table S4).

### Candidate genes involved in flavonoid biosynthesis.

The main components of dragon's blood of *D. cambodiana* are flavonoids, especially flavanes, with 7,4'- dihydroxy flavane as basic skeleton. Up to now, the biosynthetic pathway of flavanes has not been reported in other plants. A brief schematic of flavonoid biosynthesis is shown in Fig. 4a, which is modified version from Lepiniec's figure<sup>20</sup> based on analyses of KEGG databases and chemical constituents from dragon's blood of *D. cambodiana*. Flavonoids are synthesized via the phenylpropanoid pathway and are converted from phenylalanine to chalcone by the enzymes phenylalanine ammonia-lyase (EC 4.3.1.24, PAL, 6 unigene), cinnamate 4-hydroxylase (EC 1.14.13.11, C4H, 1 unigene), 4-coumarate CoA ligase (EC 6.2.1.12, 4CL, 18 unigenes) and chalcone synthase (EC 2.3.1.74, CHS, 10 unigene). Chalcone isomerase (EC 5.5.1.6, 6 unigene) catalyses the isomerisation of chalcones into flavanone. Flavanone can be converted either to flavonols through the subsequent action of flavanone 3-hydroxylase (EC 1.14.11.9, F3H, 7 unigenes)

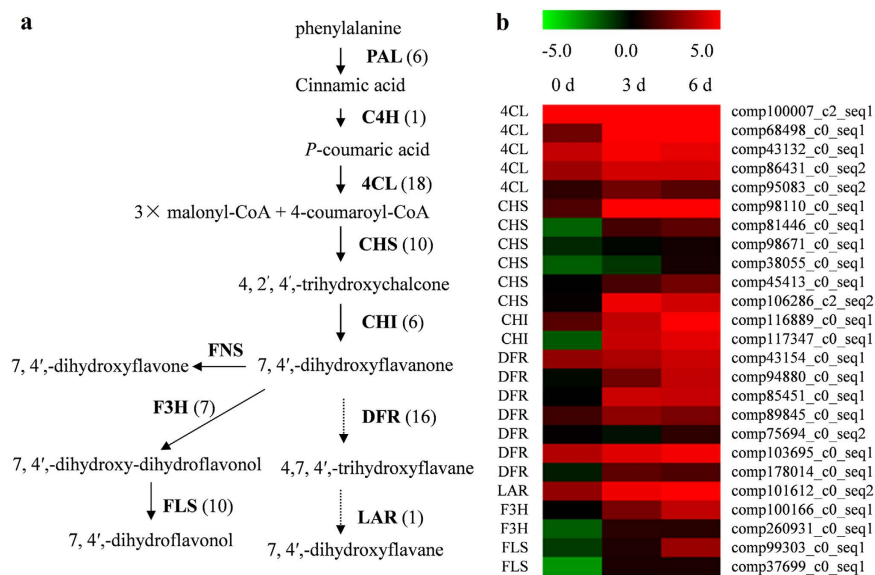


**Figure 2. Analysis of global gene expression among three samples.** (a) The number of common and unique unigenes expressed among three libraries. (b) The number of significantly ( $P$ -value  $\leq 0.05$  and at least two-fold change) up- and down-regulated transcripts between two compared samples.



**Figure 3. Functional classification of differentially expressed genes between each two of three group based on GO classification.** Gene Ontology (GO) terms are summarized in three main categories of biological process, molecular function and cellular component.

and flavonol synthase (EC 1.14.11.23, FLS, 10 unigenes) or to flavane through the action of dihydroflavonol 4-reductase (EC:1.1.1.219, DFR, 16 unigenes) and leucoanthocyanidin reductase (EC:1.17.1.3, LAR, 1 unigene).



**Figure 4. Putative flavonoids biosynthesis pathway.** (a) Proposed pathway for flavonoids biosynthesis. The numbers in brackets following each gene name indicate the number of unigenes annotated to that gene. Enzyme abbreviations: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; LAR, leucoanthocyanidin reductase; FNS, flavone synthase. (b) Expression levels of the candidate unigenes coding key enzyme involved in flavonoids biosynthesis pathways. Green and red colors are used to represent low-to-high expression levels, and color scales correspond to the mean centered  $\log_2$ -transformed FPKM values.

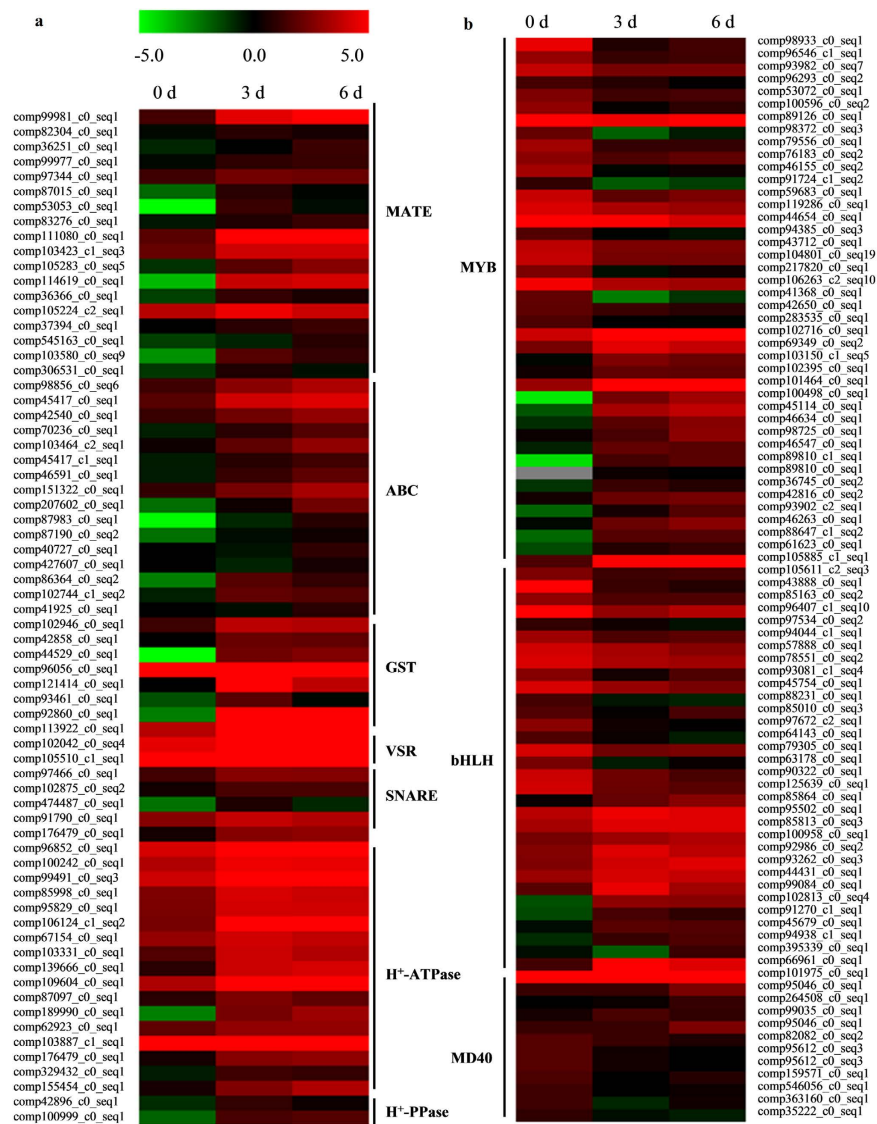
However, no unigene coding for flavone synthase (EC 1.14.11.22, FNS) was detected by RNA-sequencing. The reason may be that FNS genes were short fragments without sequence similarity.

It is known that most genes involved in flavonoid biosynthesis are coordinately expressed, consistent with flavonoid accumulation. We compared the differences in gene expression profile of those flavonoid biosynthesis genes to identify putative genes co-expressed with flavonoid accumulation. Among the above described genes involved in the flavonoid biosynthesis, five 4CL unigenes, six CHS unigenes, two CHI unigenes, seven DFR unigenes, one LAR unigenes, two F3H unigenes and two FLS unigenes were significantly up-regulated after injecting the inducer, consistent with flavonoid accumulation (Fig. 4b). Those genes co-expressed with flavonoid accumulation might play important roles in flavonoid biosynthesis and dragon's blood formation.

The chemical diversity of flavonoids increases enormously by undergoing a variety of modification reactions such as hydroxylation, glycosylation, and/or methylation. Cytochrome P450 (CYP), UDP-glycosyltransferase (UGT) and O-methyltransferase (OMT) have been shown to play roles in the modification of flavonoids<sup>21</sup>. In our study, 11 CYP, 14 UGT and 2 OMT unigenes were found to be significantly up-regulated after injecting the inducer, which may be involved in the modification of flavonoids in *D. Cambodiana* (Table S5).

**Candidate genes involved in flavonoid transport.** Flavonoids are synthesized in the cytosol and transported into the vacuole for storage or to other destinations<sup>22,23</sup>. Multidrug and toxic compound extrusion protein (MATE) transporters, two subfamily of ATP-binding cassette (ABC) transporters (G-type (ABCG) and the multidrug resistance-associated protein (MRP)-type), glutathione S-transferases (GST), vacuolar sorting receptor (VSR) and soluble N-ethylmaleimidesensitive factor attachment protein receptors (SNARE),  $H^+$ -ATPases and  $H^+$ -PPases have been claimed to play roles in sequestration of flavonoids into the vacuole<sup>22,23</sup>. In *D. Cambodiana* transcriptome, 88 unigenes encoding MRP/ABCG, 34 unigenes encoding MATE, 20 unigenes encoding GST, 4 unigenes encoding VSR, 12 unigenes encoding SNARE, 74 unigenes encoding  $H^+$ -ATPases and 13 unigenes encoding  $H^+$ -PPases were found (Table S5). Transcripts expression analysis revealed 13 out of 35 unigenes encoding MRP were highly up-regulated after injecting the inducer; while only 3 out of 53 unigenes encoding ABCG were up-regulated (Fig. 5a). The results indicated that MRP-type ABC transporters might play important roles in flavonoid transport in *D. Cambodiana*. In addition, 18 MATE unigenes, 8 GST unigenes<sup>24</sup>, 2 VSR unigenes, 4 SNARE unigenes, 18  $H^+$ -ATPases and 2  $H^+$ -PPase unigenes were also found to be significantly up-regulated after injecting the inducer (Fig. 5a). These up-regulated unigenes might be involved in transportation of flavonoids from cytosolic synthesis to vacuolar accumulation in *D. Cambodiana*.

**Candidate transcription factors involved in flavonoid biosynthesis and transport.** It is well known that transcription factors (TFs) play an essential role in regulating flavonoid biosynthesis and transport. In most species, flavonoid biosynthesis and transport is controlled by a ternary complex of MYB-bHLH-WD40, which generally regulate expression of many structural genes<sup>25</sup>. Based on our results, a total of 129, 82 and 93 unigenes were respectively predicted to code MYB, bHLH and WD40 proteins in the transcriptome database.



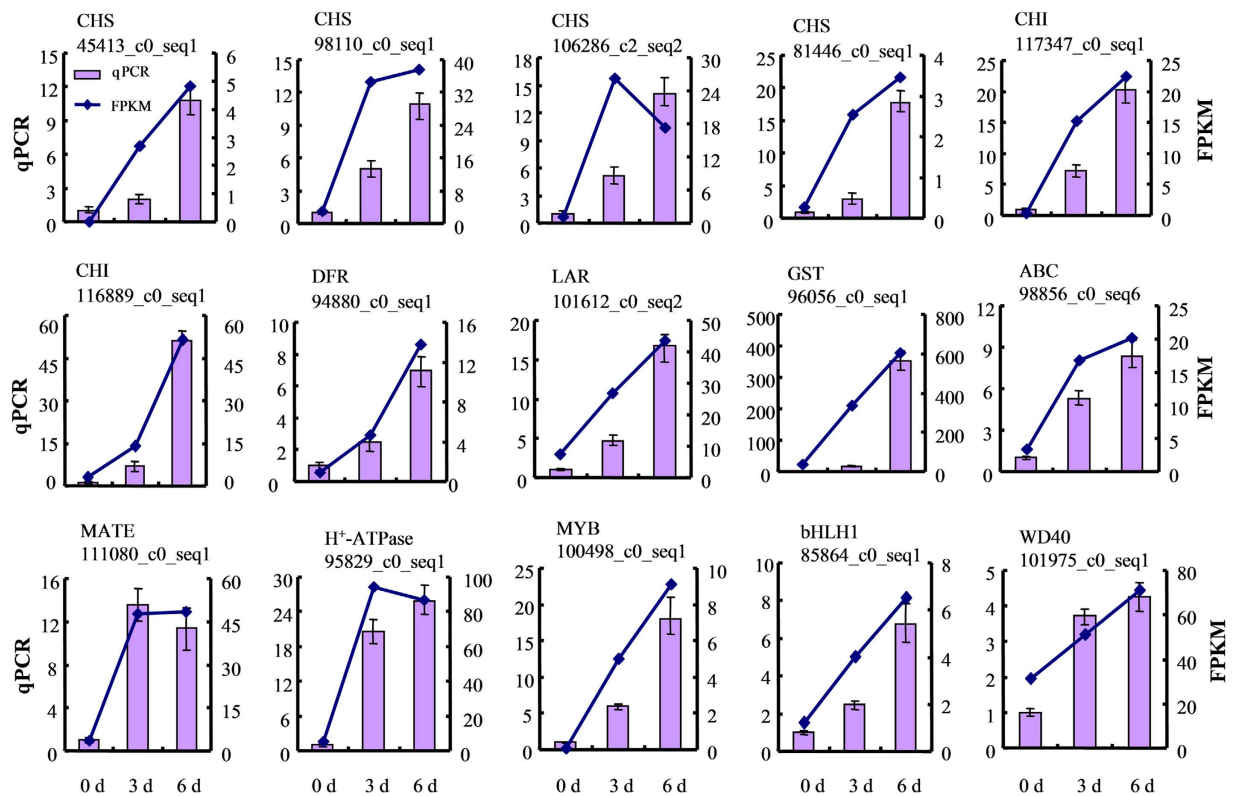
**Figure 5.** Expression levels of candidate unigenes involved flavonoids accumulation. **(a)** Expression levels of the candidate flavonoids transport-related unigenes up-regulated in stems of *D. cambodiana* after injecting the inducer. **(b)** Expression levels of TFs from DEGs involved in flavonoids accumulation. Green and red colors are used to represent low-to-high expression levels, and color scales correspond to the mean centered  $\log_2$ -transformed FPKM values.

Of these genes, the transcriptomic analysis detected 86 TFs that were differentially expressed in the stem of *D. cambodiana* after injecting the inducer, including 41 MYB unigenes (18 up-regulated and 23 down-regulated), 33 bHLH unigenes (15 up-regulated and 18 down-regulated), and 12 WD40 unigenes (5 up-regulated and 7 down-regulated) in the stem of *D. cambodiana* after injecting the inducer (Fig. 5b). These differentially expressed TFs might be involved in regulating flavonoid biosynthesis and transport in *D. cambodiana*.

**Real-time PCR validation of differential expression.** To confirm the unigenes obtained from sequencing and to further analyze the reliability of RNA-seq data in present study, 15 DEGs involved in flavonoid biosynthesis, transport and regulation were chosen for real-time quantitative PCR assay. The expression profiles of these unigenes are shown in Fig. 6. In general, the results showed that all the selected genes revealed similar expression pattern as observed in RNA-seq data. Therefore, our results provide reliable transcriptome and expression profile data for further investigations of key genes involved in flavonoid accumulation in *D. cambodiana*.

## Discussion

*D. cambodiana* is an important medicinal plant used as a source of dragon's blood. The results of the analyses of chemical constituents indicate that the main components of dragon's blood are flavonoids<sup>7–11</sup>. The limited transcriptomic data and genomic data hinder the study of the molecular mechanisms of flavonoids accumulation in *D. cambodiana*. In the present study, we firstly provided the assembly transcriptome sequence for *D. cambodiana*. Based on the sequencing results, 198,204 unigenes were generated, 34,783 of which were similar to known



**Figure 6.** qPCR validations of 15 putative genes involved flavonoids accumulation. The histograms show the qPCR results of 15 unigenes involved in flavonoid biosynthesis, transport and regulation in stems of *D. cambodiana* after injecting the inducer in 0 d, 3 d, 6 d respectively; the line charts show the FPKM values of these unigenes. qPCR results represent the mean(±SD) of three biological replicates.

proteins. All sequences involved in flavonoid biosynthesis, modification, transport and regulation were identified by searching these transcripts against sequence databases using the BLASTx search; and their expression levels were monitored by comparing them in the stems of *D. cambodiana* after injecting the inducer characterized by different flavonoid accumulation.

Flavonoids are synthesized by the phenylpropanoid metabolic pathway catalyzed by PAL, C4H and 4CL, which also serves as a starting point for the production of many other important compounds, such as coumarins, stilbenes, aurones and lignans<sup>26</sup>. In this study, six PAL, one C4H and eighteen 4CL unigenes were identified from the transcriptome data. There were no significant changes in the expression levels of PAL and C4H unigenes in stems after injecting the inducer, while five 4CL unigenes were significantly up-regulated. 4CL genes are divided into three classes, Class I, Class II and Class-4CL like<sup>25</sup>. It has been suggested that members of Class II are closely associated with flavonoid biosynthesis, those in Class I are involved in the biosynthesis of lignin and other phenylpropanoids, while the 4CL-like genes may be associated with other functions<sup>27–29</sup>. Among five 4CL DEGs, one is highly homologous to members of Class I and other four belong to Class-4CL like. In addition, all of the DEGs encoding key enzymes of flavonoid biosynthesis showed significantly up-regulated expression after injecting the inducer, such as CHS (6 unigenes), CHI (2 unigenes), DFR (7 unigenes), LAR (1 unigene), F3H (2 unigene) and FLS (2 unigene) (Fig. 5b), which were consistent with the accumulation of flavonoids in stems. These results indicated that the flavonoid biosynthesis in stems of *D. cambodiana* was activated obviously after injecting the inducer, while the upstream of which, phenylpropanoid metabolic pathway had no significantly change.

The transport of flavonoids from the cytosolic side of the endoplasmic reticulum (ER) to the vacuole may occur through three basic mechanisms: membrane transporter-, glutathione S-transferase (GST)-, or vesicle trafficking-mediated transport<sup>22,23</sup>. The proton gradient between the cytosol and the vacuole by H<sup>+</sup>-ATPases (and H<sup>+</sup>-PPases in the tonoplast) has been proposed to be the main driving force for the transport of some flavonoids<sup>30</sup>. ABCs and MATEs, as important transporters, have been claimed to play a role in sequestration of flavonoids into the vacuole by the membrane transporter-mediated transport system<sup>31–35</sup>. The vesicle trafficking-mediated transport involves flavonoid-containing vesicles releasing their content into the accumulation targets by fusion<sup>36</sup>, which requires VSR and SNARE proteins for budding, targeting, docking, fusion and recycling<sup>23</sup>. In *D. cambodiana* transcriptome, 18 MATE, 2 VSR, 4 SNARE, 18 H<sup>+</sup>-ATPase and 2 H<sup>+</sup>-PPase DEGs were found to be significantly up-regulated after injecting the inducer. Additionally, GSTs could act as flavonoid binding proteins, have been described as participating in vesicle uploading or vacuolar transport<sup>37,38</sup>. We also found that eight out of 20 GST unigenes were up-regulated after injecting the inducer, and expression patterns of three GST DEGs were strongly consistent with dragon's blood accumulation<sup>24</sup>. These results imply that the three distinct transport mechanisms (membrane transporter-, GST-, or vesicle trafficking-mediated transport)

may be all present in *D. cambodiana*; transport of flavonoids may be a multifactorial process, involving different strategies and the contribution of several proteins. These findings provide insights into the research of flavonoid transport mechanisms in *D. cambodiana*. However, in recent years, significant progress in flavonoid transport mechanisms has been focused on anthocyanin, transport of other flavonoids such as flavanes has been limited. Further research is needed to confirm whether there are three similar transport mechanisms in *D. cambodiana* and which mechanism is prevalent in transport of flavanes. Controlled transcription of biosynthetic genes is one major mechanism regulating secondary metabolite production in plant cells<sup>39</sup>. MYB, bHLH and WD40 TFs are well known to be involved in flavonoid metabolism by regulating many structural genes, they can work individually or interact physically to form the MYB-bHLH-WD40 (MBW) complex<sup>25,40</sup>. Moreover, TFs also control the regulation of flavonoid transport. For example, AtTT2 regulates the expression of the MATE transporter gene *TT12* to control the flavonoids transport<sup>41</sup>, and the maize ABC transporter ZmMRP3 involved in anthocyanin transport is regulated by the R (bHLH family) and C1 (R2R3-MYB) TFs<sup>35</sup>; and anthocyanin-related glutathione S-transferase gene *LcGST4* was activated by LcMYB1, a key R2R3-MYB transcription factor that regulates anthocyanin biosynthesis in litchi<sup>35</sup>. Most of TFs are positive regulators in flavonoid biosynthesis, whereas a few of them have been identified as repressors in flavonoid pathway<sup>42</sup>. In our study, we first identified 42 MYB unigenes (19 up-regulated and 23 down-regulated) from the DEGs. Among these unigenes, one unigene (comp103150\_c1\_seq5) was highly homologous to *SbMYB2* in *Scutellaria baicalensis*, which was identified as regulators in the control of flavonoid biosynthesis<sup>43</sup>. Another one (comp102395\_c0\_seq1) was highly homologous to *MdMYB1*, which was also involved in anthocyanin biosynthesis and transport in apple<sup>44</sup>. In addition, 33 bHLH unigenes were also identified from the DEGs. Among these unigenes, comp100958\_c0\_seq1 was highly homologous to members of subfamily 2 of bHLH, which were identified as regulators in flavonoid or anthocyanin metabolism<sup>45</sup>. Comp85864\_c0\_seq1 was highly homologous to *GL3*, which has been reported to positively regulate anthocyanin synthesis<sup>46</sup>. The expression patterns of these 2 candidate bHLH genes were significantly up-regulated after injecting the inducer, consistent with flavonoid accumulation. Finally, 12 candidate WD40 unigenes were found among the DEGs, one of them (comp19899\_c0\_seq1) was expressed at highest abundance (with 10-fold or greater expression than other genes) in stems, the expression patterns of this gene was also positively correlated with flavonoid accumulation in stems. Although MYB, bHLH and WD40 TFs in regulating flavonoid metabolism have been well characterized in plants, the related TFs have not been reported in *D. cambodiana*. Further studies are still needed to determine whether the flavonoid metabolism in *D. cambodiana* is regulated by these candidate genes, and which is independently regulated by a single TF, or controlled by combinations of TF complexes.

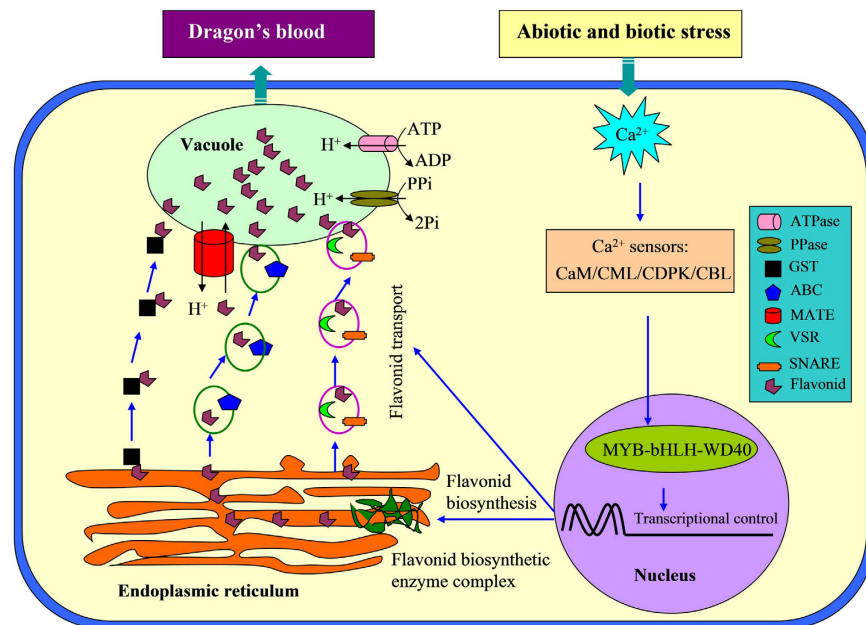
The molecular mechanisms of flavonoid biosynthesis, transport and regulation have been well characterized genetically and biochemically in model plants. Many genes involved in the flavonoid biosynthetic pathway have been cloned and characterized. However, the mechanisms of flavonoid biosynthesis and accumulation in *D. cambodiana* are still unknown. In this work, we performed transcriptome analysis of three samples with different flavonoid accumulation to identify putative genes for flavonoid metabolism in *D. cambodiana*. As a kind of plant secondary metabolite, the formation of dragon's blood was considered as a result of plant defense response to biotic and abiotic responses<sup>4,5</sup>. Calcium is a critical second messenger in the signal transduction pathways of biotic and abiotic stimuli. Calmodulin (CaM), CaM-like proteins (CMLs), calcium-dependent protein kinases (CDPKs) and calcineurin B-like proteins (CBLs) are major Ca<sup>2+</sup> sensors, playing critical roles in Ca<sup>2+</sup>-mediated signaling in plant responses to environmental stresses by binding and regulating downstream effectors<sup>47,48</sup>. In this study, some Ca<sup>2+</sup> sensors (7 CDPKs, 14 CaM/CMLs and 1 CBL) significantly up-regulated were found from the DEGs (Additional file 5), indicating that Ca<sup>2+</sup>-mediated signaling may be involved in the formation of dragon's blood induced by the inducer. Increasing evidence indicates that calcium plays an important role in flavonoid biosynthesis. It has been reported that the calcium can regulate anthocyanin accumulation, possibly by activating flavonoid pathway genes<sup>49–51</sup>. Several components of calcium signaling pathway such as Ca<sup>2+</sup>, calmodulin, and protein kinases have been shown to be involved in regulating the anthocyanin level<sup>51–53</sup>. Recent research suggests that calcium boost anthocyanin accumulation, possibly by acting on a regulatory gene(s) rather than directly activating structural genes<sup>53</sup>. Based on these previous studies and our transcriptome analysis, we proposed a model to explain the mechanism of flavonoid accumulation in *D. cambodiana* (Fig. 7). In this model, different biotic and abiotic stimuli induce transient fluctuations in cytosolic Ca<sup>2+</sup> levels in *D. cambodiana* and cause the calcium signaling, which is decoded by an array of Ca<sup>2+</sup> sensors such as CaM/CMLs, CDPKs and CBLs. These sensors interact with transcriptional activators, resulting in transcriptional activation of genes of flavonoid biosynthesis and transport leading to an increase in flavonoid accumulation. The transport of flavonoids from cytosolic synthesis to vacuolar accumulation may occur through three basic mechanisms: membrane transporter-, GST-, or vesicle trafficking-mediated transport. After reprocessed, the secondary metabolites (the main chemical compositions of dragon's blood) will send out to extracellular in response to a variety of stress. Future studies will focus on verifying this proposed model by identifying putative genes involved in flavonoid accumulation.

In summary, *De novo* characterization of transcriptome of *D. cambodiana* was firstly identified and genes expression was analyzed during formation of dragon's blood on transcriptome level. A total of 198,204 unigenes were identified from the three cDNA libraries, which will contribute significantly to further research of this species and other related species. Many candidate genes involved in flavonoid biosynthesis, modification, transport and regulation were identified, which are worthy of further functional research. The results provide insights on pathways and molecular regulation of flavonoid biosynthesis and transport, and improve our understanding of molecular mechanisms of dragon's blood formation.

## Methods

**Plant materials.** Three-year old *Dracaena cambodiana* Pierre ex Gagnep were planted in Hainan Province of China. The stems were injected with the inducer (37.5 g/L NaCl and 1.25 ml/L acetic acid), and samples were taken from 6 cm above the inject site at 3 days and 6 days after transfusing. The stems cut from healthy trees were





**Figure 7.** Proposed mechanism of flavonoids accumulation.

used to generate material for the 0 day library. All samples were collected, immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  prior to RNA extraction.

**RNA extraction.** Total RNAs from stems of three replicates were extracted using the RNA easy plant Mini Kit (QIAGEN) and treated with DNase I (Thermo) according to manufacturer's instructions. RNA quality was examined using 1% agarose gel and the concentration was determined using Nanodrap (Thermo).

**HPLC analysis.** The dried stems of *D. cambodiana* were refluxed with 95% EtOH. The solvent was evaporated under reduced pressure to yield the 95% EtOH extract. The flavonoids in the extract was chromatographed and detected with an Agilent 1260 Infinity HPLC system equipped with a diode array detector (DAD, G4212B) and a C18 Analytical HPLC Column ( $4.6 \times 100$  mm, 3.5  $\mu\text{m}$ , Agilent) as the stationary phase. The mobile phase consisted of acetonitrile (A) and 0.5% (v/v) formic acid aqueous solution (B), with a gradient elution program, 0–40 min: 10% to 25% A; 40 to 110 min: 25–45% A at a flow rate of  $1.0 \text{ mL} \cdot \text{min}^{-1}$ . The HPLC chromatogram was monitored at 279 nm and the column temperature was set at  $30^{\circ}\text{C}$ . The compounds were determined by comparing to standards isolated in our previous work and their structures were identified by  $^1\text{H NMR}$ ,  $^{13}\text{C NMR}$  and MS.

**cDNA library construction and Illumina sequencing.** Paired-end Illumina mRNA libraries were generated from 4 mg of total RNA following the manufacturer's instructions for mRNAseq sample preparation (Illumina Inc., San Diego, CA). Library quality was assessed with the 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). The samples for transcriptome analysis were prepared using a Truseq<sup>TM</sup> RNA sample prep kit (Illumina) according to the manufacturer's recommendations. Briefly, mRNA was isolated from 0.5 mg of total RNA using oligo (dT) magnetic beads. mRNA was cut into short fragments by adding fragmentation buffer. First-strand cDNA was synthesized using random hexamer-primers, taking these short fragments as templates. RNaseH, buffer, dNTPs, and DNA polymerase I were used to synthesize second-strand cDNA (NEBNext<sup>®</sup> mRNA library prep master mix set for Illumina<sup>®</sup>). Short fragments were purified with Takara's PCR extraction kit (Takara Bio, Inc.). Sequencing adapters were ligated to short fragments and resolved by agarose gel electrophoresis. Proper fragments were selected and purified and subsequently PCR amplified to create the final cDNA library template.

**Analysis of transcriptome assembly.** The transcriptome was sequenced using the Illumina HiSeq<sup>TM</sup> 2000. Four fluorescently labeled nucleotides and a specialized polymerase were used to determine the clusters base by base in parallel. The size of the library was approximately 100 bp and both ends of the library were sequenced. The 100 bp raw pairedend reads were generated on the Illumina sequencing platform. Image deconvolution and quality value calculations were performed using Illumina GA pipeline v1.6. The raw reads were cleaned by removing adaptor sequences, empty reads, and low quality sequences (reads with unknown sequences 'N' or less than 25 bp). The clean reads were assembled into non-redundant transcripts using the Trinity, which has been developed specifically for the de novo assembly of transcriptomes using short reads.

**Annotation of unigenes.** Unigenes were used as query sequences to search against the non-redundant protein (NR) database at NCBI (<http://www.ncbi.nlm.nih.gov>) and the Swiss-Prot protein database (<http://www.ebi>).

ac.uk/uniprot) with *E*-value cut-off of  $1e^{-5}$ . The annotations of the best hits were recorded. Gene Ontology (GO) (<http://www.geneontology.org/>) were further used to category the function of the unigenes by Blast2GO, and the unigenes were assigned to biological functions on the macro levels of “biological process”, “cellular component” and “molecular function”. The Kyoto Encyclopedia of Genes and Genome (KEGG) pathways database (<http://www.genome.jp/kegg/>) were assigned to unigenes by KEGG Automatic Annotation Server (KAAS).

**Differential expression analysis.** Gene expression levels of unigenes in stems were normalized and calculated as fragment per kilobase of exon model per million mapped reads (RPKM) values during the assembly and clustering process. False discovery rate (FDR)  $< 0.001$  and an estimated absolute  $\log_2$  fold-change ( $\log_2$  FC)  $\geq 1$  were used as threshold for identifying differentially expressed genes (DEGs).

**Quantitative real-time RT-PCR.** To verify the RNA-Seq results, quantitative RT-PCR was conducted using SYBR-green (TaKaRa Biotechnology Co., Ltd, Dalian, China) and Stratagene Mx3005P Real Time Thermal Cycler (Agilent, America) with the following program: 95 °C for 30 s, followed by 40 cycles of 95 °C for 10 s, and then annealing at 65 °C–95 °C for 30 s. We used the Primer Premier v. 5.0 and Vector NIT v.11.0 software (Applied Biosystems) to design primers based on the sequences of key genes of interest identified in our library. The *D. cambodiana* actin gene was used as an internal control. The sequences of primers used in this study are provided in Table S6.

## References

- Gupta, D., Bleakley, B. & Gupta, R. K. Dragon's blood: Botany, chemistry and therapeutic uses. *J Ethnopharmacol.* **115**, 361–380 (2008).
- Fan, J. Y. *et al.* A systematic review of the botanical, phytochemical and pharmacological profile of *Dracaena cochinchinensis*, a plant source of the ethnomedicine “Dragon's Blood”. *Molecules* **19**, 10650–10669 (2014).
- Zheng, D. J., Xie, L. S., Wang, Y., Zhang, Z. L. & Zhang, W. Research Advances in Dragon's Blood Plants in China. *Chin. Wild Plant Res.* **28**, 15–20 (2009).
- Wang, X. H. *et al.* Cloning, bioinformatics and the enzyme activity analyses of a phenylalanine ammonia-lyase gene involved in dragon's blood biosynthesis in *Dracaena cambodiana*. *Mol. Biol. Rep.* **40**, 97–107 (2013).
- Wang, X. H., Zhang, C., Yang, L. L. & Gomes-Laranjo, J. Production of dragon's blood in *Dracaena cochinchinensis* plants by inoculation of *Fusarium proliferatum*. *Plant Sci.* **180**, 292–299 (2011).
- Jiang, H. M. *et al.* Antibacterial components from artificially induced Dragon's Blood of *Dracaena cambodiana*. *China. J. Chin. materia medica* **40**, 4002–4006 (2015).
- Zheng, Q. A., Li, H. Z., Zhang, Y. J. & Yang, C. R. Dracaenogenins A and B, new spirostanols from the red resin of *Dracaena cochinchinensis*. *Steroids* **71**, 160–164 (2006).
- Zheng, Q. A. *et al.* Flavonoid oligomers from Chinese dragon's blood, the red resins of *Dracaena cochinchinensis*. *Nat. Prod. Bioprospect.* **2**, 111–116 (2012).
- Luo, Y. *et al.* Cytotoxic and antibacterial flavonoids from dragon's blood of *Dracaena cambodiana*. *Planta Med.* **77**, 2053–2056 (2011).
- Chen, H. Q. *et al.* Two new antimicrobial flavanes from dragon's blood of *Dracaena cambodiana*. *J. Asian. Nat. Prod. Res.* **14**, 436–440 (2012).
- Mei, W. L. *et al.* Two new flavonoids from dragon's blood of *Dracaena cambodiana*. *Bull. Korean Chem. Soc.* **34**, 1791–1794 (2013).
- Pollastri, S. & Tattini, M. Flavonols: old compounds for old roles. *Ann. Bot.* **108**, 1225–1233 (2011).
- Ferreira, M. L. F., Rius, S. P. & Casati, P. Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers Plant Sci.* **3**, 222 (2012)
- Sun, H. *et al.* *De novo*, sequencing and analysis of the cranberry fruit transcriptome to identify putative genes involved in flavonoid biosynthesis, transport and regulation. *BMC Genomics* **16**, 1–18 (2014).
- Ward, J. A., Ponnala, L. & Weber, C. A. Strategies for transcriptome analysis in nonmodel plants. *Am. J. Bot.* **99**, 267–276 (2012).
- Wu, Z. G. *et al.* Transcriptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea alata*, L.) tubers. *BMC Genomics* **16**, 820–829 (2015).
- Wei, H. *et al.* Comparative transcriptome analysis of genes involved in anthocyanin biosynthesis in the red and yellow fruits of sweet cherry (*Prunus avium* L.). *Plos One* **10**, e0121164 (2015).
- Iaria, D. L., Chiappetta, A. & Muzzalupo, I. A *de novo* transcriptomic approach to identify flavonoids and anthocyanins “switch-off” in olive (*Olea europaea* L.) drupes at different stages of maturation. *Front Plant Sci.* **6**, 1246 (2015).
- Zhou, D. W. *et al.* *De novo* sequencing transcriptome of endemic *Gentiana straminea*, (Gentianaceae) to identify genes involved in the biosynthesis of active ingredients. *Gene* **575**, 160–170 (2016).
- Lepiniec, L. *et al.* Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* **57**, 405–430 (2006).
- Saito, K. *et al.* The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol. Biochem.* **72**, 21–34 (2013).
- Petrussa, E. *et al.* Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int. J. Mol. Sci.* **14**, 14950–14973 (2013).
- Zhao, J. Flavonoid transport mechanisms: how to go, and with whom. *Trends Plant. Sci.* **20**, 1360–1385 (2015).
- Zhu, J. H. *et al.* Transcriptome-wide identification and expression analysis of glutathione s-transferase genes involved in flavonoids accumulation in *Dracaena cambodiana*. *Plant Physiol. Biochem.* **104**, 304–311 (2016)
- Hichri I. *et al.* Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* **62**, 2465–2483 (2011).
- Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant.* **3**, 2–20 (2010).
- Zhang, C. H. *et al.* Identification of 4CL genes in desert poplars and their changes in expression in response to salt stress. *Genes* **6**, 901–917 (2015).
- Ehlting, J. *et al.* Three 4-coumarate: coenzyme a ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* **19**, 9–20 (1999).
- Cukovic, D., Ehlting, J., VanZiffle, J. A. & Douglas, C. J. Structure and evolution of 4-coumarate: coenzyme a ligase (4CL) gene families. *Biol. Chem.* **382**, 645–654 (2001).
- Gomez, C. *et al.* Grapevine MATE-type proteins act as vacuolar H<sup>+</sup>-dependent acylated anthocyanin transporters. *Plant Physiol.* **150**, 402–415 (2009).
- Banasia, J. *et al.* A *Medicago truncatula* ABC transporter belonging to subfamily G modulates the level of isoflavonoids. *J. Exp. Bot.* **64**, 1005–1015 (2013).

32. Francisco, R. M. *et al.* ABCC1, an ATP binding cassette protein from grape berry, transports anthocyanidin 3-O-glucosides. *Plant Cell* **25**, 1840–1854 (2013).
33. Klein, M., Martinoia, E., Hoffmann-Thoma, G. & Weissenböck, G. A membrane-potential dependent ABC-like transporter mediates the vacuolar uptake of rye flavone glucuronides: regulation of glucuronide uptake by glutathione and its conjugates. *Plant J.* **21**, 289–304 (2000).
34. Marinova, K. *et al.* The Arabidopsis MATE transporter TT12 acts as a vacuolar flavonoid/H<sup>+</sup>-antiporter active in proanthocyanidin-accumulating cells of the seed coat. *Plant Cell* **19**, 2023–2038 (2007).
35. Goodman, C. D., Casati, P. & Walbot, V. A multidrug resistance-associated protein involved in anthocyanin transport in *Zea mays*. *Plant Cell* **16**, 1812–1826 (2004).
36. Grotewold, E. & Davies, K. M. Trafficking and sequestration of anthocyanins. *Nat. Prod. Commun.* **3**, 1251–1258 (2008).
37. Kitamura, S., Shikazono, N. & Tanaka, A. TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in Arabidopsis. *Plant J.* **37**, 104–114 (2004).
38. Smith, A. P. *et al.* Arabidopsis *AtGSTF2* is regulated by ethylene and auxin, and encodes a glutathione S-transferase that interacts with flavonoids. *Plant J.* **36**, 433–442 (2003).
39. Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S. & Ling, Y. Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim. Biophys. Acta* **1829**, 1236–1247 (2013).
40. Schaart, J. G. *et al.* Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria x ananassa*) fruits. *New Phytol.* **197**, 454–467 (2013).
41. Nesi, N., Jond, C., Debeaujon, I., Caboche, M. & Lepiniec, L. The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell* **13**, 2099–2114 (2011).
42. Hu, B. *et al.* *LcGST4* is an anthocyanin-related glutathione S-transferase gene in Litchi chinensis Sonn. *Plant Cell Rep.* **35**, 831–843 (2016).
43. Yan, J. *et al.* The soybean R2R3 MYB transcription factor GmMYB100 negatively regulates plant flavonoid biosynthesis. *Plant Mol. Biol.* **89**, 35–48 (2015).
44. Yuan, Y. The Scutellaria baicalensis R2R3-MYB transcription factors modulates flavonoid biosynthesis by regulating GA metabolism in transgenic tobacco plants. *PLoS One* **8**, e77275 (2013).
45. Hu, D. G. *et al.* MdMYB1 regulates anthocyanin and malate accumulation by directly facilitating their transport into vacuoles in apples. *Plant Physiol.* **170**, 1315–1330 (2016).
46. Carretero-Paulet, L. *et al.* Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol.* **153**, 1398–412 (2010).
47. Payne, C. T., Zhang, F. & Lloyd, A. M. GL3 encodes a bHLH protein that regulates trichome development in Arabidopsis through interaction with GL1 and TTG1. *Genetics* **156**, 1349–1362 (2000).
48. Virdi, A. S., Singh, S. & Singh, P. Abiotic stress responses in plants: roles of calmodulin-regulated proteins. *Front Plant Sci.* **6**, 809 (2015).
49. Ranty, B. *et al.* Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Front Plant Sci.* **7**, 327 (2016).
50. Gollop, R., Even, S., Colova-Tsolova, V. & Perl, A. Expression of the grape dihydroflavonol reductase gene and analysis of its promoter region. *J. Exp. Bot.* **53**, 1397–1409 (2002).
51. Shin, D. H. *et al.* Calcium dependent sucrose uptake links sugar signaling to anthocyanin biosynthesis in Arabidopsis. *Biochem. Biophys. Res. Commun.* **430**, 634e639 (2013).
52. Vitrac, X. *et al.* Sugar sensing and Ca<sup>2+</sup>-calmodulin requirement in *Vitis vinifera* cells producing anthocyanins. *Phytochemistry* **53**, 659–665 (2000).
53. Wang, C. Q., Zhang, Y. F. & Tao, L. Activity changes of calmodulin and Ca<sup>2+</sup>-ATPase during low-temperature-induced anthocyanin accumulation in *Alternanthera bettzickiana*. *Physiol. Plant.* **124**, 260–266 (2005).
54. Xu, W. X. *et al.* Effect of calcium on strawberry fruit flavonoid pathway gene expression and anthocyanin accumulation. *Plant Physiol. Biochem.* **82**, 289–298 (2014).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.31070321, 31400297), National Nonprofit Institute Research Grant of CATAS-ITBB (1630052016002), and the Special Grant for Modernization of Traditional Chinese Medicine of Hainan Province (2015ZY17).

## Author Contributions

S.Q.P. and W.L.M. conceived the study. J.H.Z., T.J.C., H.L.L., D.G. and H.F.D. performed the experiments and carried out the analysis. J.H.Z., T.J.C., S.Q.P. and W.L.M. designed the experiments and wrote the manuscript. All authors read and approved the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhu, J.-H. *et al.* *De Novo* transcriptome characterization of *Dracaena cambodiana* and analysis of genes involved in flavonoid accumulation during formation of dragon's blood. *Sci. Rep.* **6**, 38315; doi: 10.1038/srep38315 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016