

RESEARCH ARTICLE

# High-Throughput Sequencing and *De Novo* Assembly of Red and Green Forms of the *Perilla frutescens* var. *crispa* Transcriptome

Atsushi Fukushima<sup>1\*</sup>, Michimi Nakamura<sup>2</sup>, Hideyuki Suzuki<sup>3</sup>, Kazuki Saito<sup>1,2</sup>, Mami Yamazaki<sup>2\*</sup>

**1** RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa, 230–0045, Japan, **2** Graduate School of Pharmaceutical Sciences, Chiba University, Chiba-shi, Chiba, 263–8522, Japan, **3** Kazusa DNA Research Institute, Kazusa-Kamatari 2-6-7, Kisarazu, Chiba, 292–0818, Japan

\* [atsushi.fukushima@riken.jp](mailto:atsushi.fukushima@riken.jp) (AF); [mamiy@faculty.chiba-u.jp](mailto:mamiy@faculty.chiba-u.jp) (MY)



**OPEN ACCESS**

**Citation:** Fukushima A, Nakamura M, Suzuki H, Saito K, Yamazaki M (2015) High-Throughput Sequencing and *De Novo* Assembly of Red and Green Forms of the *Perilla frutescens* var. *crispa* Transcriptome. PLoS ONE 10(6): e0129154. doi:10.1371/journal.pone.0129154

**Academic Editor:** Xianlong Zhang, National Key Laboratory of Crop Genetic Improvement, CHINA

**Received:** February 22, 2015

**Accepted:** May 5, 2015

**Published:** June 12, 2015

**Copyright:** © 2015 Fukushima et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All FASTQ files are available from the DNA Data Bank of Japan DRA database. Accession number: DRA003003.

**Funding:** This work was supported, in part, by Health Labour Sciences Research Grant from The Ministry of Health Labour and Welfare, and JSPS KAKENHI Grant Number 24390030. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

*Perilla frutescens* var. *crispa* (Labiatae) has two chemo-varietal forms, i.e. red and green forms of perilla, that differ in the production of anthocyanins. To facilitate molecular biological and biochemical studies in perilla-specialized metabolism we used Illumina RNA-sequencing technology in our comprehensive comparison of the transcriptome map of the leaves of red and green forms of perilla. Sequencing generated over 1.2 billion short reads with an average length of 101 nt. *De novo* transcriptome assembly yielded 47,788 and 47,840 unigenes in the red and green forms of perilla plants, respectively. Comparison of the assembled unigenes and existing perilla cDNA sequences showed highly reliable alignment. All unigenes were annotated with gene ontology (GO) and Enzyme Commission numbers and entered into the Kyoto Encyclopedia of Genes and Genomes. We identified 68 differentially expressed genes (DEGs) in red and green forms of perilla. GO enrichment analysis of the DEGs showed that genes involved in the anthocyanin metabolic process were enriched. Differential expression analysis revealed that the transcript level of anthocyanin biosynthetic unigenes encoding flavonoid 3'-hydroxylase, dihydroflavonol 4-reductase, and anthocyanidin synthase was significantly higher in red perilla, while the transcript level of unigenes encoding limonene synthase was significantly higher in green perilla. Our data serve as a basis for future research on perilla bio-engineering and provide a shortcut for the characterization of new functional genes in *P. frutescens*.

## Introduction

Plants can produce a diverse range of secondary metabolites that are beneficial for human health, food, and medicines. When exposed to environmental changes such as drought and water-deficiency, plants can respond to these stresses by producing soluble phenolics, mainly flavonoids (for example, see [1]) and lignins [2]. One group of flavonoids is an anthocyanin pigment [3]. Genes involved in the central flavonoid biosynthetic pathway (see review: [4]),

their modification reactions, and their transcriptional regulation have been characterized by the combinatorial approach of transcriptomic and metabolomic profiles with a reverse genetic technique in *Arabidopsis* [5–8] and other plants [9, 10]. A more detailed genetic, transcriptomic, and metabolomic characterization of pigment plants will lead to a better understanding of the transcriptional regulatory and metabolic systems for anthocyanin biosynthesis.

*Perilla frutescens* var. *crispa* (Labiatae) is a medicinal plant common in Southeast Asia. Among its two chemo-varietal forms, red and green forms of perilla, only red perilla ('Aka-jiso' in Japanese) can produce anthocyanins, mainly malonylshisonin [11, 12]. The differential display of mRNA [13] from red and green forms of perilla plants was used for the characterization of genes associated with regulation of the expression of biosynthetic genes [14], for example, the Myb-like gene [15] and the Myc-like gene [16]. Other anthocyanin-related genes have been identified [17–20] and a normalized cDNA library from whole young perilla was constructed and 4,582 uni-expressed sequence tags (uniESTs) were identified [21]. As early methods such as the mRNA differential display, differential hybridization, and serial analysis of gene expression (SAGE) can only monitor a small coverage of the transcript profile, the establishment of fundamental molecular and genetic resources/tools such as DNA microarray- and EST databases remains far from complete in perilla plants.

Recent advances in high-throughput RNA-sequencing technologies (RNA-seq) allow the monitoring of genome-wide transcription, i.e. a complete set of transcripts of an organism (see reviews, [22] and [23]). RNA-seq is applicable to both model organisms with reference genome sequences and to non-model species without an existing reference genome, including crops, trees, and vegetables [24, 25]. It can also detect novel transcribed regions in a genome, small/micro RNAs, and novel alternative splicing patterns. The Medicinal Plant Genomics Resource (MPGR) consortium (<http://medicinalplantgenomics.msu.edu/>) provides RNA-seq data for 14 medicinal plants including *Catharanthus roseus*; transcriptome data from 23 different tissues in *C. roseus* are available [26]. RNA-seq technology is helpful for a better understanding of the perilla-specialized metabolism and its regulation.

Using RNA-seq technology, we analyzed and here described the whole transcriptome map of red and green forms of perilla leaves. We generated over 1.2 billion bases of high-quality short reads using an Illumina sequencer and now demonstrate the suitability of our sequencing for *de novo* transcriptome assembly and the functional annotation of unigenes in perilla leaves. We compared transcript levels in red and green forms of perilla, especially the biosynthetic pathways of anthocyanin and perillyl alcohol. Our findings serve as a basis for future studies on perilla bio-engineering and provide a shortcut to the discovery of new functional genes in *P. frutescens*.

## Results and Discussion

### Sample preparation and Illumina sequencing

For the comprehensive characterization of red and green forms of the perilla transcriptome, total RNA samples were isolated from leaves. Using a bioanalyzer we performed DNase treatment and confirmed RNA integrity. Then, the samples were mixed equally. Total RNA was utilized in the mRNA preparation, fragmentation, and cDNA synthesis. After the removal of adaptor sequences and low-quality and ambiguous reads, Illumina sequencing yielded 1,214,546,008 and 1,240,000,000 clean reads from the mRNA pool isolated from *Perilla frutescens* var. *crispa* f. *purpurea* (red perilla) (Table 1) and *P. frutescens* var. *crispa* f. *viridis* (green perilla), respectively (S1 Table). The short reads showed mean quality scores 36.2% in red- and 36.3% in green perilla, indicating that our RNA sequencing was adequate for *de novo* assembly.

**Table 1. Summary of the sequence assembly after Illumina sequencing in red perilla.**

	Raw reads	Contigs	Unigenes
Total length (bp)	1,214,546,008	44,923,850	41,869,105
Number of contigs	12,025,208	54,500	47,788
Average length (bp)	101	824	876
Median length (bp)	101	515	580
Max length (bp)	101	12,157	12,157
Min length (bp)	101	201	201
N50 (bp)	101	1,312	1,349

doi:10.1371/journal.pone.0129154.t001

### De novo transcriptome assembly of red and green forms of perilla

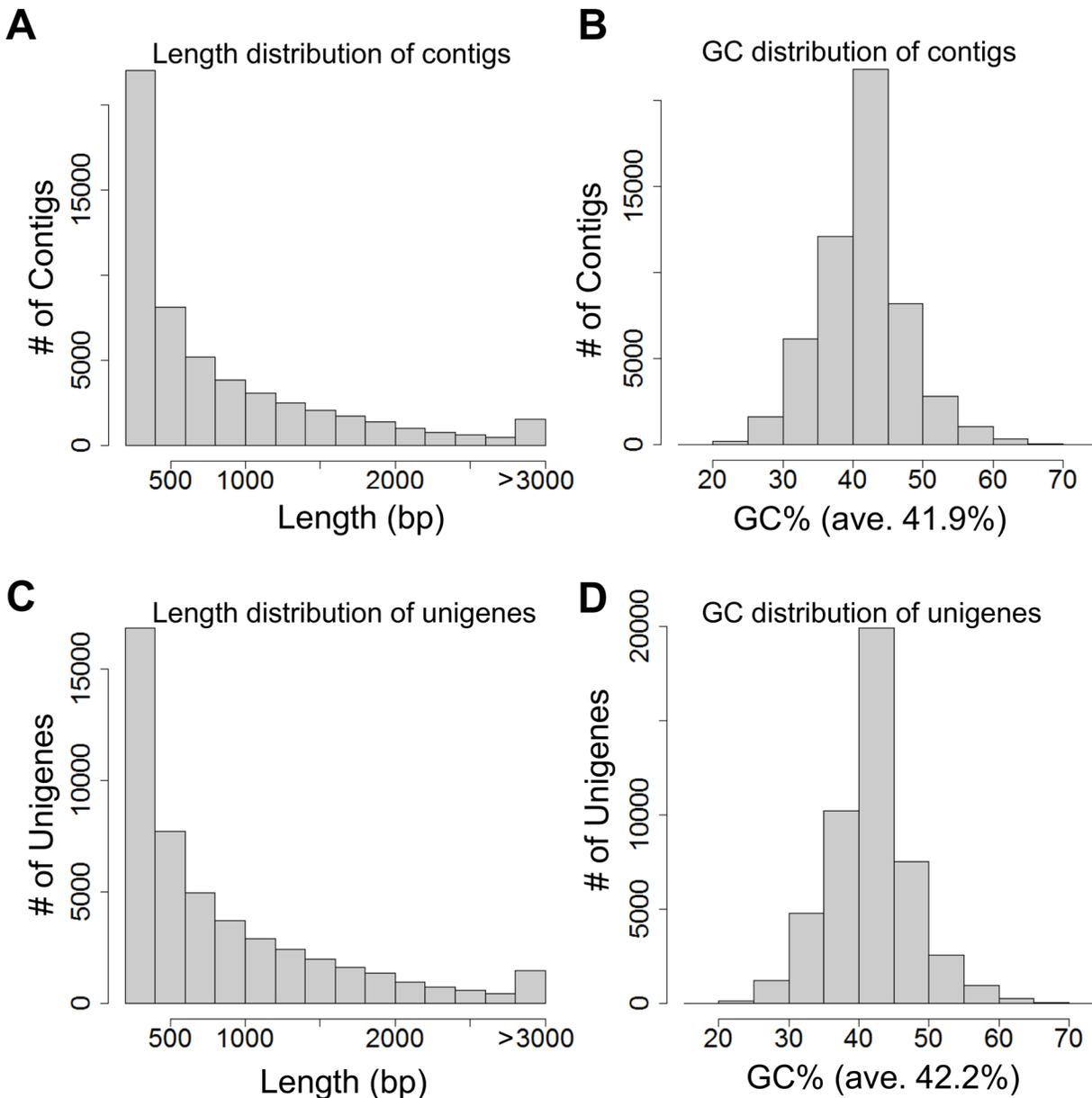
Using the Trinity program [27], all clean reads of red perilla were assembled *de novo* into 54,500 contigs with an average length of 824 base pairs (bp) and an N50 of 1,312 bp (S1 File). In green perilla we obtained 54,445 contigs with an average length of 844 bp and an N50 of 1,368 bp. The length and GC% distribution for all contigs for red and green forms of perilla are shown in Fig 1A and 1B, respectively, and in S1 Fig To estimate expression abundance we used Bowtie [28] and RSEM [29] for the contigs. We obtained 47,788 unigenes with an average length of 876 bp and an N50 of 1,349 bp in red perilla (Table 1) and 47,840 unigenes with an average length of 897 bp and an N50 of 1,399 bp in green perilla (S1 Table). The length and GC content distribution of all assembled unigenes in red and green perilla are shown in Fig 1C and 1D, respectively, and in S1 Fig To provide a general overview of the unigenes we calculated basic statistics. The results showed that in red perilla, 21,174 unigenes were shorter than 500 bp and 1,152 unigenes were longer than 3,000 bp; in green perilla 21,186 unigenes were shorter than 500 bp and 1,257 were longer than 3,000 bp. In 3,909 unigenes of red- and in 3,810 unigenes of green perilla the GC content exceeded 50%.

### Comparison of assembled unigenes and perilla sequences deposited in GenBank

To assess the quality of the assembled unigenes we used all *P. frutescens* cDNA sequences available as of December 2014 from NCBI GenBank [30] which contains all 5,911 *P. frutescens* sequences (5,538 ESTs and 373 nucleotides). Of the 5,911 cDNA sequences in GenBank, 4,252 (71.9%) could be matched with red perilla unigenes using a cutoff E-value of  $10^{-10}$  (BLASTn). Of 47,788 unigenes, 3,957 (8.3%) matched with 5,911 perilla cDNA sequences, the others were unmatched. There were significant similarities with previously characterized perilla genes encoding glutathione S-transferase (AB362191.1: 98.6% identity and E-value = 0.0) [31], anthocyanin 5-O-glucosyltransferase (AB013596.1: 97.6% identity and E-value = 0.0) [20], anthocyanin 5-O-glucoside-6"-O-malonyltransferase (AF405204.1: 98.9% identity and E-value = 0.0) [32], a WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis (AB059642.1: 98.8% identity and E-value = 0.0) [33], and cytochrome P450 reductase (GQ120439.1: 99.0% identity and E-value = 0.0) [34]. These results indicate that our assembled unigenes have a wide coverage with known perilla cDNA sequences because many unigenes had not been sequenced or had not been assembled correctly.

### Functional annotation and classification of perilla unigenes

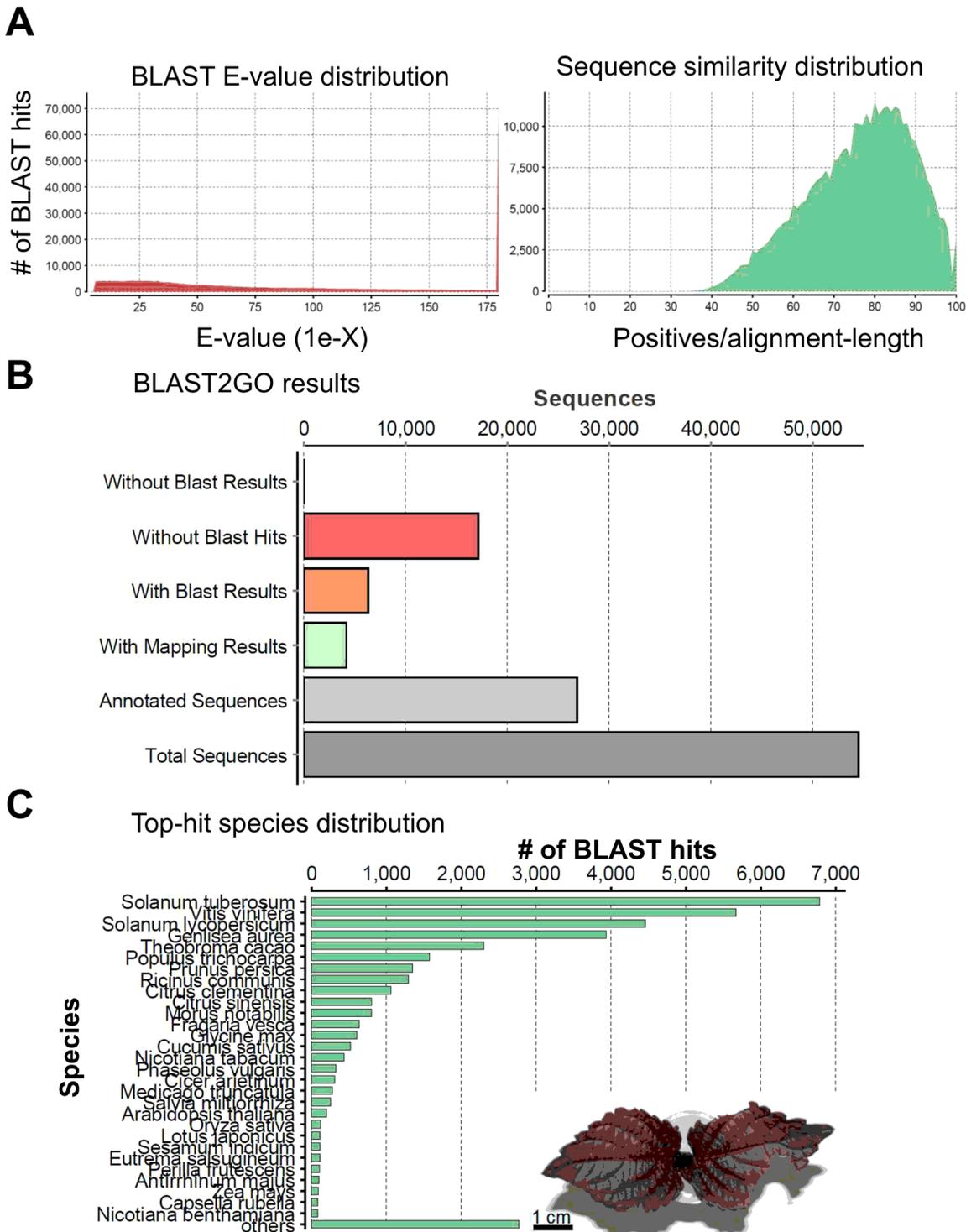
Next we validated and annotated the assembled unigenes. Our homology search against an NCBI non-redundant (NR) protein database (<http://www.ncbi.nlm.nih.gov>; formatted on



**Fig 1. Overview of the *de novo* transcriptome assembly in *Perilla frutescens purpurea* (red perilla).** (A and B) Length and GC distribution of contigs assembled from high-quality clean reads by the Trinity program [27]. (C and D) Length and GC distribution of unigenes generated from further contig assembly.

doi:10.1371/journal.pone.0129154.g001

April 7, 2014) was based on the BLASTx program [35] for all unigenes using a cutoff E-value  $< 10^{-5}$ , and the best aligning results were selected to annotate the unigenes. As a result, 81.7% of the aligned sequences in red perilla exhibited significant homology with entries in the NR database (E-value  $< 1E^{-5}$ ) (left panel in Fig 2A). The annotation results for green perilla are shown in S2 Fig. Based on the BLAST similarity distribution, 11,630 sequences in red perilla exhibited alignment identities greater than 80% (right panel in Fig 2A). To obtain gene ontology (GO) [36] for the unigenes we used the Blast2GO program v 2.7.1 [37]; it can also assign an Enzyme Commission (EC) number and Kyoto Encyclopedia of Genes and Genomes (KEGG)



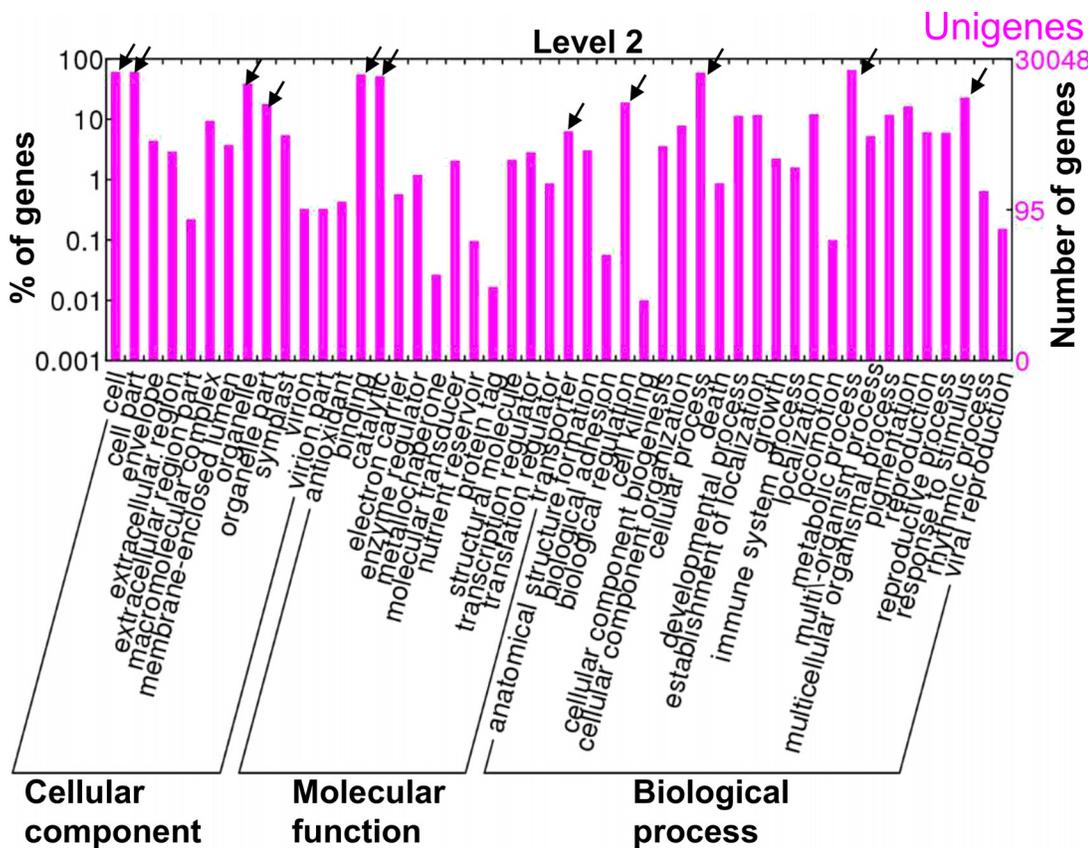
**Fig 2. Characterization of the assembled unigenes based on a non-redundant (NR) protein database search in red perilla.** (A) (Left panel): E-value distribution of BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ . (Right panel): Similarity score distribution of the top BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ . (B) Bar chart of the data distribution from BLAST2GO [37]. (C) Species distribution of the top BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ .

doi:10.1371/journal.pone.0129154.g002

[38] information based on the BLAST results. The annotation results for red perilla are presented as a bar chart of the data distribution from BLAST2GO (Fig 2B).

A large number of the hits matched the sequences of *Solanum tuberosum* (18.2%), *Vitis vinifera* (15.2%), and *Solanum lycopersicum* (12.0%); other hits were identified within the reference protein databases of *Genlisea aurea* (10.6%), *Theobroma cacao* (6.2%), *Populus trichocarpa* (4.2%), *Prunus persica* (3.6%), *Ricinus communis* (3.5%), *Citrus clementina* (2.8%), and *Citrus sinensis* (2.2%) (Fig 2C). The species distribution of the top BLAST hits for the assembled unigenes from red and green forms of perilla was quite similar (S2 Fig). Although there were many unigenes with no BLAST hits, they may be uncharacterized genes that were not represented in the annotated protein databases or assembled sequences too short to produce hits.

We used the BGI WEGO program [39] to perform GO functional classification of all unigenes and of the distribution of gene functions of the species (Fig 3). WEGO can map all of the annotated unigenes to GO terms and identify the number of unigenes involved in each GO term. All 30,048 unigenes in red perilla were categorized by three main GO terms: cellular component, molecular function, and biological process. Within the cellular component, most unigenes were assigned to “cell” and “cell parts”, followed by “organelle” and “organelle part”. Within the molecular function category, the great majority of unigenes was associated with the terms “binding”, “catalytic”, and “transporter”. Within the biological process group, the great majority of unigenes was related to the terms “cellular process”, “metabolic process”,



**Fig 3. Gene ontology assignments for all assembled unigenes in red perilla.** The results are summarized in terms of three functional categories: cellular component, molecular function, and biological process. 30,048 unigenes were categorized by GO terms. The GO terms were visualized using WEGO (<http://wego.genomics.org.cn>) [39].

doi:10.1371/journal.pone.0129154.g003

“biological regulation”, and “response to stimulus” (for green perilla, see [S3 Fig](#)). All unigenes with functional annotations are presented in [S2](#) and [S3](#) Tables.

### Functional classification by KEGG pathways

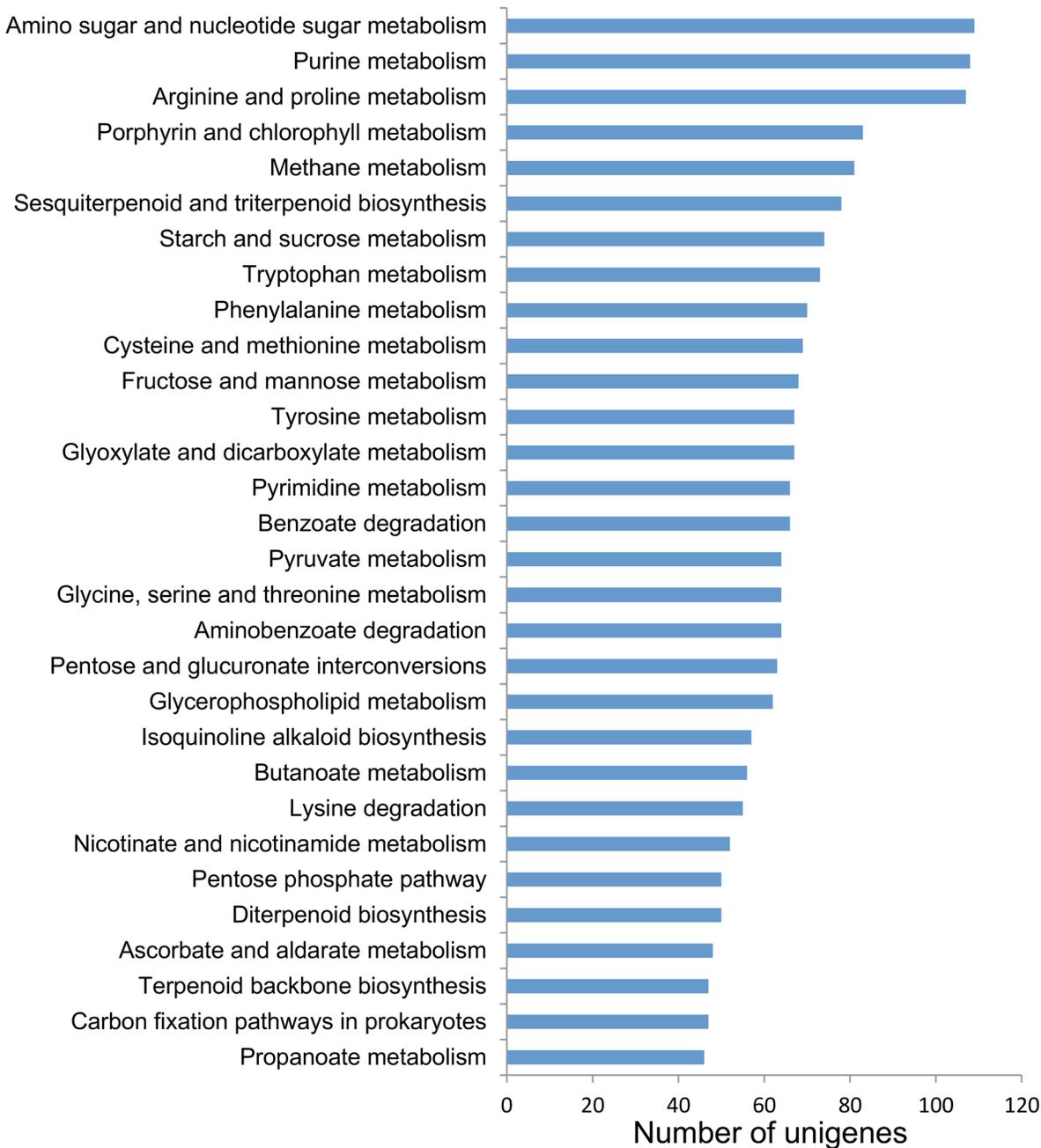
The KEGG pathway database [38] stores the molecular network interactions of cellular components. Pathway-based annotation helps to further understand the biological functions of unigenes. The annotated unigenes in red perilla were grouped into 139 KEGG pathways. [Fig 4](#) represents the top 30 pathways in our enrichment analysis of assembled unigenes in red perilla. The top 3 ranking pathways were amino sugar and nucleotide sugar metabolism (109 unigenes), purine metabolism (108 unigenes), and arginine and proline metabolism (107 unigenes) ([Fig 4](#) and [S4 Table](#)). The top 3 ranking pathways were identical for red and green perilla. The number of unigenes associated with the anthocyanin biosynthetic pathway in red and green forms of perilla was 25 and 19, respectively (see also [S4 Fig](#)).

### Identification of simple sequence repeats (SSRs)

SSRs, or microsatellites, are ubiquitous repetitive DNA sequences in eukaryotic genomes (see reviews, [40] and [41]). They are important markers for determining functional genetic variation including paternity determination, genetic diversity assessment, population genetics studies, and for the development of a genetic map. To identify SSRs we searched all unigenes in the red and green forms of perilla with MISA [42]. We detected a total of 15,156 SSRs in 12,024 transcripts of red perilla ([Table 2](#) and [S5 Table](#)). All SSRs can be classified by the number of repeat units. Di-nucleotide SSRs represented the largest fraction (46.5%) of SSRs identified, followed by mono-nucleotide (36.1%) and tri-nucleotide (17.2%) SSRs. Although only a small fraction of tetra- (120), penta- (12), and hexa-nucleotide (18) SSRs were identified in red perilla transcripts, their number was significant. Our identified SSRs of *Perilla* species may provide potential genetic markers for population genetics and comparative genomics research to enhance our understanding of the genetic control of adaptive traits.

### Identification of differentially expressed genes (DEGs) in different forms of perilla plants

We identified 68 differentially-expressed genes [false discovery rate (FDR) < 0.05] using the TCC package [43], which is for comparing raw tag count data with a robust normalization method. In [S6 Table](#), we also listed uniquely expressed unigenes in the different forms of perilla plants. The tables feature 22,359 and 22,187 unigenes in the red and green form, respectively. We then performed GO enrichment analysis using a hypergeometric test implemented in BiNGO [44]. This yielded significantly enriched GO functional categories in DEGs compared to the genomic background ([S5 Fig](#)). GO functional categories with an FDR < 0.05 were defined as significantly over-represented in DEGs. The top 5 enriched GO terms were “anthocyanin metabolic process (FDR = 1.6E-05)”, “flavonoid metabolic process (FDR = 1.6E-05)”, “phenylpropanoid metabolic process (FDR = 4.6E-05)”, “flavonoid biosynthetic process (FDR = 6.7E-05)”, and “phenylpropanoid biosynthetic process (FDR = 6.7E-05)”. Our identification of DEGs suggests that mainly the biological processes of anthocyanin biosynthesis, but not those of other metabolites, are different in red and green forms of perilla as has been reported previously [12]. The leaves of the red form contain many anthocyanin pigments such as malonylshisonin, shisonin, cis-isomers of malonylshisonin, and peonidin 3-*O*-malonylglucoside-5-*O*-*p*-coumarylglucoside. Among them, malonylshisonin was the main anthocyanin, representing approximately 70% of the total anthocyanins in red perilla leaves. In contrast, green perilla leaves did not accumulate these anthocyanins [12].



**Fig 4. Pathway enrichment analysis of assembled unigenes in red perilla.** Annotated unigenes were grouped into 139 KEGG pathways. The top 30 pathways containing unigenes are displayed.

doi:10.1371/journal.pone.0129154.g004

### Exploring gene expressions associated with the biosynthetic pathway of anthocyanins

To detect more genes belonging to the relevant biosynthesis pathway of anthocyanin in the transcriptome sequences we carried out a comparative inspection of transcriptome data from red and green forms of perilla plants involved in the biosynthesis of phenylpropanoid and flavonoid skeletons (Fig 5). In this pathway, red and green forms of perilla differed in that red

**Table 2. Statistics of SSRs detected in red perilla.**

<b>Results of SSR searches</b>	
Total number of sequences examined:	54,500
Total size of examined sequences (bp):	44,923,850
Total number of identified SSRs:	15,156
Number of SSR containing sequences:	12,024
Number of sequences containing more than 1 SSR:	2,460
Number of SSRs present in compound formation:	1,135
<i>Distribution to different repeat type classes</i>	
Mono-nucleotide	5,352
Di-nucleotide	7,050
Tri-nucleotide	2,604
Tetra-nucleotide	120
Penta-nucleotide	12
Hexa-nucleotide	18

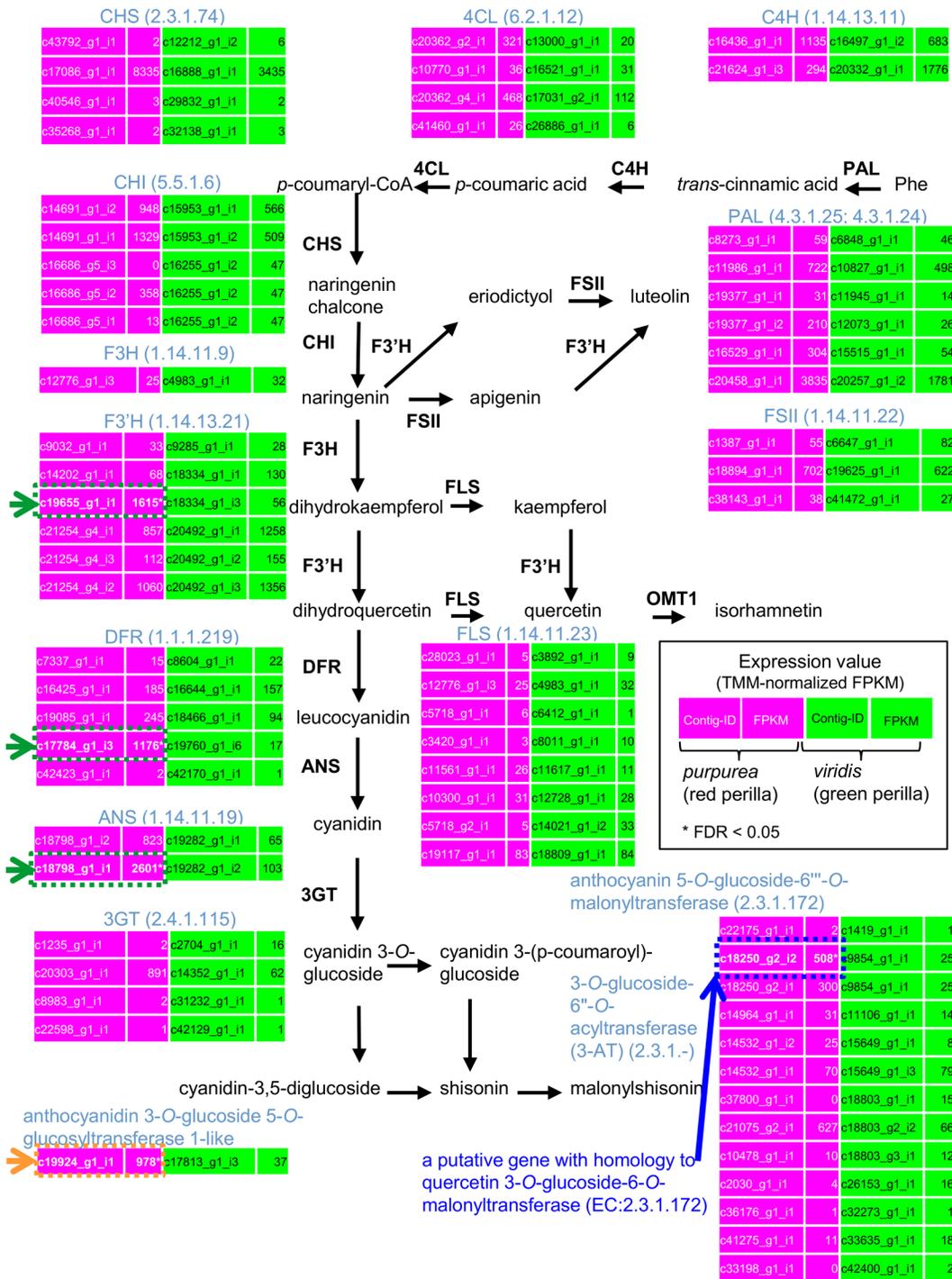
doi:10.1371/journal.pone.0129154.t002

plants manifested higher expression levels of genes encoding flavanone 3'-hydroxylase (F3'H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS) than did green plants (FDR < 0.05) (green arrows in Fig 5). Because F3'H plays a crucial role in pigment biosynthesis in Arabidopsis, mutant plants lacking F3'H (called *tt7*) produce pale brown seeds due to reduced levels of brown pigment [45]. DFR catalyzes the first committed reaction leading to anthocyanin and proanthocyanidin [46]. ANS, also called leucoanthocyanidin dioxygenase (LDOX), can catalyze the formation of cyanidin from leucoanthocyanidin with oxygen and 2-oxo-glutaric acid as co-substrates [18, 47].

In Arabidopsis and maize, a set of transcription factors including MYB [48, 49], basic helix-loop-helix (bHLH) [50], and WD40 [33, 51] plays a central role in the regulation of anthocyanin genes. Tohge et al. [5] suggest that MYB75/PAP1 (PRODUCTION OF ANTHOCYANIN PIGMENT 1) and its homolog MYB90/PAP2 specifically induce the expression of genes associated with the biosynthesis of anthocyanin in Arabidopsis, including DFR and ANS/LDOX. Genes encoding a putative gene with homology to quercetin 3-O-glucoside-6-O-malonyltransferase (EC:2.3.1.172) (unigene ID: 'c18250\_g2\_i2') (blue arrow in Fig 5) and anthocyanidin 3-O-glucoside 5-O-glucosyltransferase 1-like (unigene ID: 'c19924\_g1\_i1') (orange arrow in Fig 5) were significantly expressed in red perilla. We also compared the expression levels of previously identified genes including transcription factors and enzymes in red and green forms of perilla (S6 Fig). In red perilla, there was a significant up-regulation (FDR < 0.05) of genes encoding F3G1 (AB103172), bHLH transcription factors associated with the regulation of the flavonoid pathway [52, 53], and glutathione S-transferase (AB362191.1) [31].

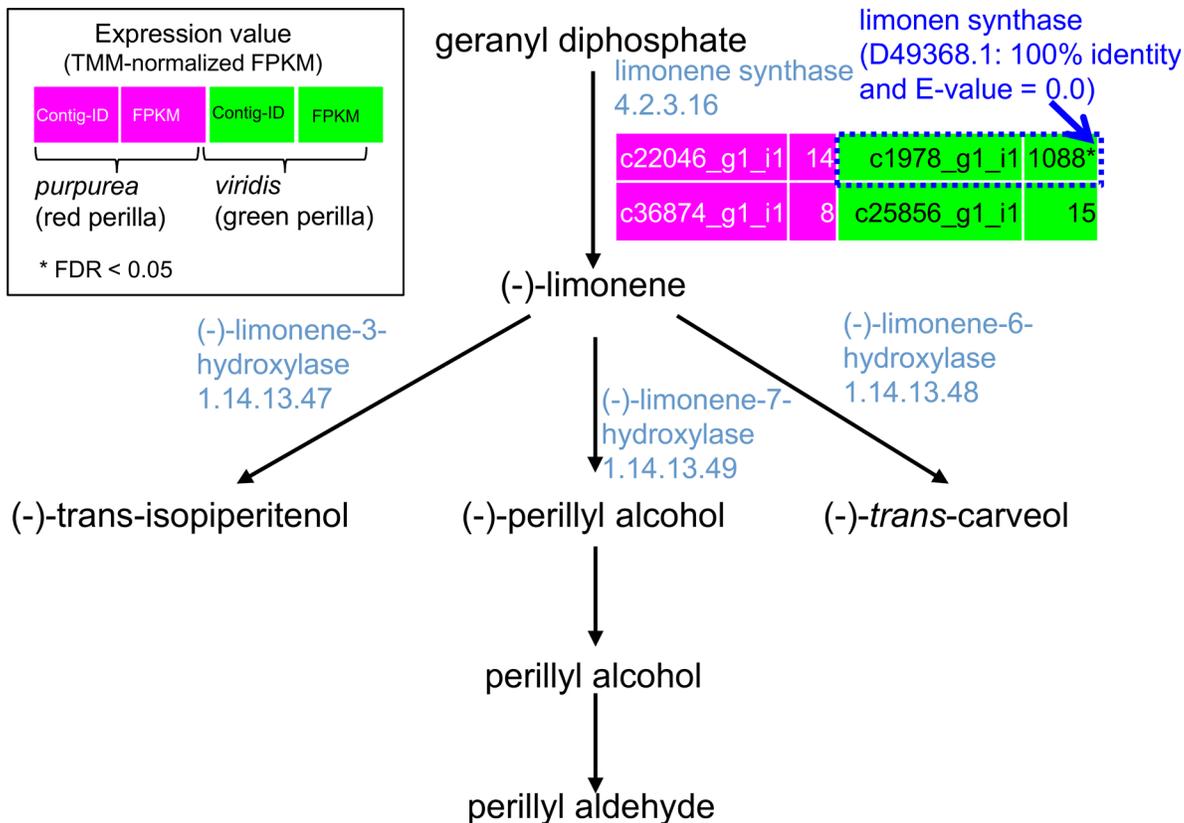
## Gene expressions involved in the biosynthetic pathway of monoterpenes in red and green forms of perilla

Monoterpenes produced by plants play crucial roles in their defense against insects and microbes [54, 55]. Perillyl alcohol, a cyclic monoterpene, is secreted by numerous plant species including lavender, mints, and perilla. The flavor of the perilla herb is characterized by perillaldehyde, an index compound for quality control of the perilla herb in the Japanese Pharmacopoeia (JP). Fig 6 is a comparative representation of transcriptome data related to the biosynthetic pathway of monoterpenes in red and green forms of perilla. Unigene encoding limonene synthase (D49368.1: 100% identity and E-value = 0.0) showed significantly greater



**Fig 5. Comparative representation of transcriptome data in red and green forms of perilla plants.** Biosynthetic pathways and the expression of unigenes involved in the biosynthesis of phenylpropanoid and flavonoid skeletons are shown. The expression levels (TMM-normalized FPKM values) of unigenes encoding the enzymes of each step are displayed. Homologous genes in red and green perilla represent the reciprocal best-hit BLAST results. Asterisks identify the false discovery rate, (FDR) < 0.05, with the TCC package [43]. TMM, Trimmed Mean of M values [63]; PAL, phenylalanine ammonia-lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumaric acid: CoA ligase; ACC, acetyl-CoA carboxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; OMT1, O-methyltransferase 1; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; 3GT, UDP-glucose: anthocyanidin 3-O-glucosyltransferase.

doi:10.1371/journal.pone.0129154.g005



**Fig 6. Comparative representation of transcriptome data in red and green perilla plants.** Biosynthetic pathways and the expression of unigenes involved in the biosynthesis of perillyl alcohol are displayed. The expression levels (TMM-*FPKM* values) of unigenes encoding the enzymes of each step are shown. Homologous genes in red and green perilla show the reciprocal best-hit BLAST results. Asterisks indicate the false discovery rate, (FDR) < 0.05, obtained with the TCC package [43]. TMM, Trimmed Mean of M values [63].

doi:10.1371/journal.pone.0129154.g006

up-regulation in green than red perilla. It produces limonene from geranyl diphosphate and is an important step in the biosynthesis of perillyl alcohol [56]. One of the major essential oil components of green perilla is limonene. A study that applied gas chromatography—flame ionization detection (GC—FID) showed that the chemical composition of limonene from the aerial parts of red and green forms of perilla was 1.1% and 12.6%, respectively [57]. We think that our results at least partially support the chemical composition of the essential oils of red and green forms of perilla.

In another recent work, Tong et al. [58] analyzed the differences between red and green forms of *P. frutescens* var. *crispa* to identify candidate genes involved in leaf color. More studies are needed for a better understanding of the complex regulation of the biosynthetic pathway(s) of anthocyanin and perillyl alcohol in perilla plants and its physiological significance. Although the expression patterns at the protein level must be further investigated, our data and those of others [58], are a basis for future studies on perilla bioengineering and may help to develop an approach for the characterization of new functional genes in *Perilla* species.

## Conclusion

Our study represents comprehensive transcriptome resource for perilla plants that feature two varietal forms of anthocyanin accumulation (red and green forms). Our datasets are an integrated genomic resources for molecular cloning and for identifying genes of interest in perilla.

Given the incomplete knowledge on the molecular control mechanism(s) of the biosynthetic pathways associated with anthocyanin and monoterpenes, our transcriptome analysis provides useful information regarding the specialized metabolism of perilla plants.

## Materials and Methods

### Plant materials, RNA isolation, and cDNA synthesis

The plants of *Perilla frutescens* var. *crispa* f. *purpurea* (red perilla) and *P. frutescens* var. *crispa* f. *viridis* (green perilla), were grown in the experimental gardens of the Center of Medicinal Resources, Graduate School of Medical and Pharmaceutical Sciences, Chiba University, Chiba, Japan. *P. frutescens* var. *crispa* is not an endangered or protected species. Fresh leaves were collected from healthy plants in May 2012. The leaves were frozen by liquid nitrogen and subsequently powdered using a Multi Beads Shocker (Yasui Kikai, Japan). Total RNA was extracted from powdered red and green leaves of *P. frutescens* var. *crispa* with the RNeasy Mini Kit (Qiagen, USA), cleaned by ethanol precipitation, and processed using an Illumina TruSeq Prep Kit v2 according to the manufacturer's protocol (Illumina, San Diego, CA, USA). We used unreplicated data for each form of perilla (i.e., one sample per form).

### Illumina sequencing

cDNA libraries were sequenced on an Illumina HiSeq 1000 sequencer (Illumina Inc., San Diego, CA, USA) and 100-bp paired-end (PE) reads were produced. After the removal of adaptor sequences and ambiguous and low-quality reads, Illumina sequencing resulted in 1,214,546,008 and 1,240,000,000 clean reads from the mRNA pool isolated from red perilla and green perilla, respectively. All raw read sequences are available at the DDBJ Sequence Read Archive [59] under accession number DRA003003.

### Data pre-processing, filtering, and *de novo* transcriptome assembly

For transcriptome assembly we filtered the raw reads and removed adapter sequences, non-coding RNA, low-quality reads with ambiguous 'N' bases, and raw reads with an average length less than 20 bases. The Trinity program [27] was used for *de novo* transcriptome assembly, it combines read sequences of a certain overlap length to form longer fragments without 'N' gaps (contigs). We then processed these contigs for read alignment and abundance estimation with Bowtie [28] and RSEM [29]. To calculate unigene expression we used the Fragments Per Kilobase of exon per Million mapped fragments (FPKM) method. In the calculation of gene expression it can exclude sequencing discrepancies and the influence of different gene lengths. The number of unigenes was 47,788 in red- and 47,840 in green perilla at a threshold more than FPKM = 1. The length and GC% distribution of all assembled unigenes are shown in Fig 1C and 1D and in S1 Fig. To calculate the GC content and basic statistics values used custom Ruby/Bioruby script [60], the R/Bioconductor package "ShortRead" [61], and "Biostrings" [62].

### Functional annotation and classification of unigenes

We performed a homology search against the NCBI NR protein database (<http://www.ncbi.nlm.nih.gov>, formatted on April 7, 2014) based on the BLASTx program [35] for all unigenes using a cutoff E-value < 10<sup>-5</sup>. The best aligning results were selected to annotate the unigenes. For their further annotation we used the Blast2GO program v 2.7.1 [37] to assign GO terms, an EC number, and KEGG [38] information according to the BLAST results. For visualization of the GO functional classification of all unigenes and the distribution of the gene functions in the

different species we used the BGI WEGO program [39]. The microsatellite identification tool (MISA) (<http://pgrc.ipk-gatersleben.de/misa/>) [42] with its default parameters was applied to identify microsatellites in the unigenes.

### Identification of differentially expressed genes (DEGs)

Differential gene expression analysis between red and green forms of perilla was with the TCC package [43]. Briefly, the main algorithm to identify DEGs with the TCC package is based on the combination of TMM normalization [63] and a DEG identification method [e.g., edgeR [64] and DESeq [65]]. In the DEG identification step, we used a negative binomial test implemented in DESeq [65]. BiNGO [44], a tool to calculate the over-representation of DEGs, was used to analyze significantly over-represented GO categories.

### Global BLAST search against currently available *Perilla frutescens* sequences

To identify putative orthologous genes in red and green forms of perilla plants, all 5,911 *P. frutescens* var. *crispa* sequences (5,538 ESTs and 373 nucleotides) were downloaded from NCBI GenBank [30] and then submitted to reciprocal best-hit BLASTn searches against unigenes; the cutoff E-value was  $< 10^{-10}$ .

## Supporting Information

**S1 Fig. Overview of our *de novo* transcriptome assembly in *Perilla frutescens* var. *crispa* f. *viridis* (green perilla).** (A and B) Length and GC distribution of the contigs assembled from high-quality clean reads by the Trinity program [27].

(C and D) Length and GC distribution of the unigenes generated from further contig assembly. (PPTX)

**S2 Fig. Characterization of the assembled unigenes based on a non-redundant (NR) protein database search in green perilla.**

1. (Left panel) E-value distribution of BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ . (Right panel) Similarity score distribution of the top BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ .
2. Bar chart of the data distribution from BLAST2GO [37].
3. Species distribution of the top BLAST hits for the assembled unigenes with a cutoff of E-value  $< 10^{-5}$ . (PPTX)

**S3 Fig. Gene ontology (GO) assignments for all assembled unigenes in green perilla.** Results summarized in three functional categories: cellular component, molecular function, and biological process. 29,813 unigenes were categorized by GO terms. The GO terms were visualized using WEGO (<http://wego.genomics.org.cn>) [39].

(PPTX)

**S4 Fig. Pathway enrichment analysis of assembled unigenes in green perilla.** Annotated unigenes were grouped into 137 KEGG pathways. The top 30 pathways containing unigenes are displayed.

(PPTX)

**S5 Fig. Significantly enriched GO functional categories in 68 differentially expressed genes in red and green forms of perilla.** The colored nodes indicate the significantly over-represented GO categories. The pseudo-colored bar represents the significance [false discovery rate (FDR)]. (PPTX)

**S6 Fig. Comparative representation of transcriptome data in red and green perilla plants.** Previously identified genes encoding transcription factor(s) and enzymes, and the expression of unigenes are shown. The expression levels (TMM-normalized FPKM values) are displayed. Homologous genes in red and green perilla indicate the reciprocal best-hit BLAST results. Asterisks represent the false discovery rate (FDR) < 0.05 obtained with the TCC package [43]. TMM, Trimmed Mean of M values [63]. (PPTX)

**S1 File. Unigene sequences of *Perilla frutescens* var. *crispa* f. *purpurea* (red perilla) in FASTA format.** (FASTA)

**S2 File. Unigene sequences of *Perilla frutescens* var. *crispa* f. *viridis* (green perilla) in FASTA format.** (FASTA)

**S1 Table. Summary of the sequence assembly after Illumina sequencing in green perilla.** (XLSX)

**S2 Table. List of all annotated unigenes in red perilla.** (XLSX)

**S3 Table. List of all annotated unigenes in green perilla.** (XLSX)

**S4 Table. Functional classification of red (A) and green (B) forms of perilla unigenes by KEGG pathways.** (XLSX)

**S5 Table. Statistics of simple sequence repeats detected in green perilla.** (XLSX)

**S6 Table. Differential expression analysis and identification of uniquely expressed genes in different forms of perilla.**

1. Detailed lists of differentially expressed genes (DEGs) up- or down-regulated between red and green forms of perilla.
2. Uniquely expressed genes in red perilla.
3. Uniquely expressed genes in green perilla. (XLSX)

## Acknowledgments

We thank Ms. Ursula Petralia for editorial assistance and Dr. Tetsuya Sakurai and Mr. Yutaka Yamada (RIKEN Center for Sustainable Resource Science) for computational assistance. We also thank Mr. Tsutomu Hosouchi (Kazusa DNA Research Institute) for technical support in Illumina sequencing.

## Author Contributions

Conceived and designed the experiments: AF HS KS MY. Performed the experiments: MN HS MY. Analyzed the data: AF HS MY. Contributed reagents/materials/analysis tools: MN HS MY. Wrote the paper: AF KS MY.

## References

1. Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. *Current opinion in plant biology*. 2002; 5(3):218–23. PMID: [11960739](#).
2. Moura JC, Bonine CA, de Oliveira Fernandes Viana J, Dornelas MC, Mazzafera P. Abiotic and biotic stresses and changes in the lignin content and composition in plants. *Journal of integrative plant biology*. 2010; 52(4):360–76. doi: [10.1111/j.1744-7909.2010.00892.x](#) PMID: [20377698](#).
3. Koes R, Verweij W, Quattrocchio F. Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends in plant science*. 2005; 10(5):236–42. doi: [10.1016/j.tplants.2005.03.002](#) PMID: [15882656](#).
4. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, et al. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant physiology and biochemistry: PPB / Societe francaise de physiologie vegetale*. 2013; 72:21–34. doi: [10.1016/j.plaphy.2013.02.001](#) PMID: [23473981](#).
5. Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, et al. Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. *The Plant journal: for cell and molecular biology*. 2005; 42(2):218–35. doi: [10.1111/j.1365-313X.2005.02371.x](#) PMID: [15807784](#).
6. Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, et al. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *The Plant cell*. 2008; 20(8):2160–76. doi: [10.1105/tpc.108.058040](#) PMID: [18757557](#); PubMed Central PMCID: PMC2553606.
7. Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, et al. Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana. *The Plant journal: for cell and molecular biology*. 2012; 69(1):154–67. doi: [10.1111/j.1365-313X.2011.04779.x](#) PMID: [21899608](#); PubMed Central PMCID: PMC3507004.
8. Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, et al. A flavonoid 3-O-glucoside:2"-O-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. *The Plant journal: for cell and molecular biology*. 2014; 79(5):769–82. doi: [10.1111/tbj.12580](#) PMID: [24916675](#).
9. Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, et al. Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. *The Plant journal: for cell and molecular biology*. 2012; 70(4):624–36. doi: [10.1111/j.1365-313X.2012.04903.x](#) PMID: [22229385](#).
10. Gong L, Chen W, Gao Y, Liu X, Zhang H, Xu C, et al. Genetic analysis of the metabolome exemplified using a rice population. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110(50):20320–5. doi: [10.1073/pnas.1319681110](#) PMID: [24259710](#); PubMed Central PMCID: PMC3864304.
11. Kondo T, Tamura H, Yoshida K, Goto T. Structure of malonylshisonin, a genuine pigment in purple leaves of *Perilla ocimoides* L. var. *crispa* Benth. *Agric Biol Chem*. 1989; 53(3):797–800.
12. Yamazaki M, Nakajima J, Yamanashi M, Sugiyama M, Makita Y, Springob K, et al. Metabolomics and differential gene expression in anthocyanin chemo-varietal forms of *Perilla frutescens*. *Phytochemistry*. 2003; 62(6):987–95. PMID: [12590125](#).
13. Yamazaki M, Saito K. Differential display analysis of gene expression in plants. *Cellular and molecular life sciences: CMLS*. 2002; 59(8):1246–55. PMID: [12363028](#).
14. Saito K, Yamazaki M. Biochemistry and molecular biology of the late-stage of biosynthesis of anthocyanin: lessons from *Perilla frutescens* as a model plant. *New Phytol*. 2002; 155(1):9–23. doi: [10.1046/j.1469-8137.2002.00440.x](#)
15. Gong ZZ, Yamazaki M, Saito K. A light-inducible Myb-like gene that is specifically expressed in red *Perilla frutescens* and presumably acts as a determining factor of the anthocyanin forma. *Molecular & general genetics: MGG*. 1999; 262(1):65–72. PMID: [10503537](#).
16. Gong ZZ, Yamagishi E, Yamazaki M, Saito K. A constitutively expressed Myc-like gene involved in anthocyanin biosynthesis from *Perilla frutescens*: molecular characterization, heterologous expression in

- transgenic plants and transactivation in yeast cells. *Plant molecular biology*. 1999; 41(1):33–44. PMID: [10561066](#).
17. Kitada C, Gong Z, Tanaka Y, Yamazaki M, Saito K. Differential expression of two cytochrome P450s involved in the biosynthesis of flavones and anthocyanins in chemo-varietal forms of *Perilla frutescens*. *Plant & cell physiology*. 2001; 42(12):1338–44. PMID: [11773526](#).
  18. Saito K, Kobayashi M, Gong Z, Tanaka Y, Yamazaki M. Direct evidence for anthocyanidin synthase as a 2-oxoglutarate-dependent oxygenase: molecular cloning and functional expression of cDNA from a red forma of *Perilla frutescens*. *The Plant journal: for cell and molecular biology*. 1999; 17(2):181–9. PMID: [10074715](#).
  19. Gong Z, Yamazaki M, Sugiyama M, Tanaka Y, Saito K. Cloning and molecular analysis of structural genes involved in anthocyanin biosynthesis and expressed in a forma-specific manner in *Perilla frutescens*. *Plant molecular biology*. 1997; 35(6):915–27. PMID: [9426610](#).
  20. Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, et al. Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. *The Journal of biological chemistry*. 1999; 274(11):7405–11. PMID: [10066805](#).
  21. Lee SC, Lee J, Kim NH, Park JY, Kim HU, Lee HO, et al. Analysis of expressed sequence tags from a normalized cDNA library of perilla (*Perilla frutescens*). *J Plant Biol*. 2014; 57(5):312–20.
  22. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews Genetics*. 2009; 10(1):57–63. doi: [10.1038/nrg2484](#) PMID: [19015660](#); PubMed Central PMCID: PMC2949280.
  23. Garber M, Grabherr MG, Guttman M, Trapnell C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods*. 2011; 8(6):469–77. doi: [10.1038/nmeth.1613](#) PMID: [21623353](#).
  24. Gongora-Castillo E, Buell CR. Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural product reports*. 2013; 30(4):490–500. doi: [10.1039/c3np20099j](#) PMID: [23377493](#).
  25. Fukushima A, Kusano M. A network perspective on nitrogen metabolism from model to crop plants using integrated 'omics' approaches. *Journal of experimental botany*. 2014; 65(19):5619–30. doi: [10.1093/jxb/eru322](#) PMID: [25129130](#).
  26. Gongora-Castillo E, Childs KL, Fedewa G, Hamilton JP, Liscombe DK, Magallanes-Lundback M, et al. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS one*. 2012; 7(12):e52506. doi: [10.1371/journal.pone.0052506](#) PMID: [23300689](#); PubMed Central PMCID: PMC3530497.
  27. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*. 2011; 29(7):644–52. doi: [10.1038/nbt.1883](#) PMID: [21572440](#); PubMed Central PMCID: PMC3571712.
  28. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*. 2009; 10(3):R25. doi: [10.1186/gb-2009-10-3-r25](#) PMID: [19261174](#); PubMed Central PMCID: PMC2690996.
  29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics*. 2011; 12:323. doi: [10.1186/1471-2105-12-323](#) PMID: [21816040](#); PubMed Central PMCID: PMC3163565.
  30. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic acids research*. 2014. doi: [10.1093/nar/gku1216](#) PMID: [25414350](#).
  31. Yamazaki M, Shibata M, Nishiyama Y, Springob K, Kitayama M, Shimada N, et al. Differential gene expression profiles of red and green forms of *Perilla frutescens* leading to comprehensive identification of anthocyanin biosynthetic genes. *The FEBS journal*. 2008; 275(13):3494–502. doi: [10.1111/j.1742-4658.2008.06496.x](#) PMID: [18513325](#).
  32. Suzuki H, Nakayama T, Nishino T. Proposed mechanism and functional amino acid residues of malonyl-CoA:anthocyanin 5-O-glucoside-6"-O-malonyltransferase from flowers of *Salvia splendens*, a member of the versatile plant acyltransferase family. *Biochemistry*. 2003; 42(6):1764–71. doi: [10.1021/bi020618g](#) PMID: [12578391](#).
  33. Sompornpailin K, Makita Y, Yamazaki M, Saito K. A WD-repeat-containing putative regulatory protein in anthocyanin biosynthesis in *Perilla frutescens*. *Plant molecular biology*. 2002; 50(3):485–95. PMID: [12369624](#).
  34. Mau CJ, Karp F, Ito M, Honda G, Croteau RB. A candidate cDNA clone for (-)-limonene-7-hydroxylase from *Perilla frutescens*. *Phytochemistry*. 2010; 71(4):373–9. doi: [10.1016/j.phytochem.2009.12.002](#) PMID: [20079506](#).

35. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990; 215(3):403–10. doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/).
36. The Gene Ontology C. Gene Ontology Consortium: going forward. *Nucleic acids research*. 2014. doi: [10.1093/nar/gku1179](https://doi.org/10.1093/nar/gku1179) PMID: [25428369](https://pubmed.ncbi.nlm.nih.gov/25428369/).
37. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21(18): 3674–6. doi: [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610) PMID: [16081474](https://pubmed.ncbi.nlm.nih.gov/16081474/).
38. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic acids research*. 2014; 42(Database issue): D199–205. doi: [10.1093/nar/gkt1076](https://doi.org/10.1093/nar/gkt1076) PMID: [24214961](https://pubmed.ncbi.nlm.nih.gov/24214961/); PubMed Central PMCID: [PMC3965122](https://pubmed.ncbi.nlm.nih.gov/PMC3965122/).
39. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. *Nucleic acids research*. 2006; 34(Web Server issue):W293–7. doi: [10.1093/nar/gkl031](https://doi.org/10.1093/nar/gkl031) PMID: [16845012](https://pubmed.ncbi.nlm.nih.gov/16845012/); PubMed Central PMCID: [PMC1538768](https://pubmed.ncbi.nlm.nih.gov/PMC1538768/).
40. Hancock JM. Simple sequences in a "minimal" genome. *Nature genetics*. 1996; 14(1):14–5. doi: [10.1038/ng0996-14](https://doi.org/10.1038/ng0996-14) PMID: [8782812](https://pubmed.ncbi.nlm.nih.gov/8782812/).
41. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends in biotechnology*. 2005; 23(1):48–55. doi: [10.1016/j.tibtech.2004.11.005](https://doi.org/10.1016/j.tibtech.2004.11.005) PMID: [15629858](https://pubmed.ncbi.nlm.nih.gov/15629858/).
42. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *TAG Theoretical and applied genetics Theoretische und angewandte Genetik*. 2003; 106(3):41–22. doi: [10.1007/s00122-002-1031-0](https://doi.org/10.1007/s00122-002-1031-0) PMID: [12589540](https://pubmed.ncbi.nlm.nih.gov/12589540/).
43. Sun J, Nishiyama T, Shimizu K, Kadota K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC bioinformatics*. 2013; 14:219. doi: [10.1186/1471-2105-14-219](https://doi.org/10.1186/1471-2105-14-219) PMID: [23837715](https://pubmed.ncbi.nlm.nih.gov/23837715/); PubMed Central PMCID: [PMC3716788](https://pubmed.ncbi.nlm.nih.gov/PMC3716788/).
44. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005; 21(16):3448–9. doi: [10.1093/bioinformatics/bti551](https://doi.org/10.1093/bioinformatics/bti551) PMID: [15972284](https://pubmed.ncbi.nlm.nih.gov/15972284/).
45. Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B. Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. *Biological chemistry*. 2000; 381(8):749–53. doi: [10.1515/BC.2000.095](https://doi.org/10.1515/BC.2000.095) PMID: [11030432](https://pubmed.ncbi.nlm.nih.gov/11030432/).
46. Shirley BW, Hanley S, Goodman HM. Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *The Plant cell*. 1992; 4(3):333–47. doi: [10.1105/tpc.4.3.333](https://doi.org/10.1105/tpc.4.3.333) PMID: [1354004](https://pubmed.ncbi.nlm.nih.gov/1354004/); PubMed Central PMCID: [PMC160133](https://pubmed.ncbi.nlm.nih.gov/PMC160133/).
47. Devic M, Guillemot J, Debeaujon I, Bechtold N, Bensaude E, Koornneef M, et al. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *The Plant journal: for cell and molecular biology*. 1999; 19(4):387–98. PMID: [10504561](https://pubmed.ncbi.nlm.nih.gov/10504561/).
48. Grotewold E, Drummond BJ, Bowen B, Peterson T. The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell*. 1994; 76(3):543–53. PMID: [8313474](https://pubmed.ncbi.nlm.nih.gov/8313474/).
49. Procissi A, Dolfini S, Ronchi A, Tonelli C. Light-Dependent Spatial and Temporal Expression of Pigment Regulatory Genes in Developing Maize Seeds. *The Plant cell*. 1997; 9(9):1547–57. doi: [10.1105/tpc.9.9.1547](https://doi.org/10.1105/tpc.9.9.1547) PMID: [12237395](https://pubmed.ncbi.nlm.nih.gov/12237395/); PubMed Central PMCID: [PMC157032](https://pubmed.ncbi.nlm.nih.gov/PMC157032/).
50. Shin J, Park E, Choi G. PIF3 regulates anthocyanin biosynthesis in an HY5-dependent manner with both factors directly binding anthocyanin biosynthetic gene promoters in *Arabidopsis*. *The Plant journal: for cell and molecular biology*. 2007; 49(6):981–94. doi: [10.1111/j.1365-313X.2006.03021.x](https://doi.org/10.1111/j.1365-313X.2006.03021.x) PMID: [17319847](https://pubmed.ncbi.nlm.nih.gov/17319847/).
51. Carey CC, Strahle JT, Selinger DA, Chandler VL. Mutations in the pale aleurone color1 regulatory gene of the *Zea mays* anthocyanin pathway have distinct phenotypes relative to the functionally similar TRANSPARENT TESTA GLABRA1 gene in *Arabidopsis thaliana*. *The Plant cell*. 2004; 16(2):450–64. doi: [10.1105/tpc.018796](https://doi.org/10.1105/tpc.018796) PMID: [14742877](https://pubmed.ncbi.nlm.nih.gov/14742877/); PubMed Central PMCID: [PMC341916](https://pubmed.ncbi.nlm.nih.gov/PMC341916/).
52. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant journal: for cell and molecular biology*. 2011; 66(1):94–116. doi: [10.1111/j.1365-313X.2010.04459.x](https://doi.org/10.1111/j.1365-313X.2010.04459.x) PMID: [21443626](https://pubmed.ncbi.nlm.nih.gov/21443626/).
53. Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of experimental botany*. 2011; 62(8):2465–83. doi: [10.1093/jxb/erq442](https://doi.org/10.1093/jxb/erq442) PMID: [21278228](https://pubmed.ncbi.nlm.nih.gov/21278228/).
54. Loza-Tavera H. Monoterpenes in essential oils. Biosynthesis and properties. *Advances in experimental medicine and biology*. 1999; 464:49–62. PMID: [10335385](https://pubmed.ncbi.nlm.nih.gov/10335385/).

55. Degenhardt J, Kollner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*. 2009; 70(15–16):1621–37. doi: [10.1016/j.phytochem.2009.07.030](https://doi.org/10.1016/j.phytochem.2009.07.030) PMID: [19793600](https://pubmed.ncbi.nlm.nih.gov/19793600/).
56. Yuba A, Yazaki K, Tabata M, Honda G, Croteau R. cDNA cloning, characterization, and functional expression of 4S-(-)-limonene synthase from *Perilla frutescens*. *Archives of biochemistry and biophysics*. 1996; 332(2):280–7. doi: [10.1006/abbi.1996.0343](https://doi.org/10.1006/abbi.1996.0343) PMID: [8806736](https://pubmed.ncbi.nlm.nih.gov/8806736/).
57. Tabanca N, Demircib B, Alia A, Alia Z, Blythec EK, Khan IA. Essential oils of green and red *Perilla frutescens* as potential sources of compounds for mosquito management. *Ind Crop Prod*. 2015; 65:36–44. doi: [10.1016/j.indcrop.2014.11.043](https://doi.org/10.1016/j.indcrop.2014.11.043)
58. Tong W, Kwon SJ, Lee J, Choi IY, Park YJ, Choi SH, et al. Gene set by de novo assembly of *Perilla* species and expression profiling between *P. frutescens* (L.) var. *frutescens* and var. *crispa*. *Gene*. 2015; 559(2):155–63. doi: [10.1016/j.gene.2015.01.028](https://doi.org/10.1016/j.gene.2015.01.028) PMID: [25597767](https://pubmed.ncbi.nlm.nih.gov/25597767/).
59. Kosuge T, Mashima J, Kodama Y, Fujisawa T, Kaminuma E, Ogasawara O, et al. DDBJ progress report: a new submission system for leading to a correct annotation. *Nucleic acids research*. 2014; 42(Database issue):D44–9. doi: [10.1093/nar/gkt1066](https://doi.org/10.1093/nar/gkt1066) PMID: [24194602](https://pubmed.ncbi.nlm.nih.gov/24194602/); PubMed Central PMCID: PMC3964987.
60. Goto N, Prins P, Nakao M, Bonnal R, Aerts J, Katayama T. BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics*. 2010; 26(20):2617–9. doi: [10.1093/bioinformatics/btq475](https://doi.org/10.1093/bioinformatics/btq475) PMID: [20739307](https://pubmed.ncbi.nlm.nih.gov/20739307/); PubMed Central PMCID: PMC2951089.
61. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*. 2009; 25(19):2607–8. doi: [10.1093/bioinformatics/btp450](https://doi.org/10.1093/bioinformatics/btp450) PMID: [19654119](https://pubmed.ncbi.nlm.nih.gov/19654119/); PubMed Central PMCID: PMC2752612.
62. Pages H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2321. 2009.
63. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*. 2010; 11(3):R25. doi: [10.1186/gb-2010-11-3-r25](https://doi.org/10.1186/gb-2010-11-3-r25) PMID: [20196867](https://pubmed.ncbi.nlm.nih.gov/20196867/); PubMed Central PMCID: PMC2864565.
64. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26(1):139–40. doi: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) PMID: [19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/); PubMed Central PMCID: PMC2796818.
65. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biology*. 2010; 11(10):R106. doi: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106) PMID: [20979621](https://pubmed.ncbi.nlm.nih.gov/20979621/); PubMed Central PMCID: PMC3218662.