

RESEARCH ARTICLE

Characterization of the *Rosa roxburghii* Tratt transcriptome and analysis of *MYB* genes

Xiaolong Huang¹✉, Huiqing Yan²✉*, Lisheng Zhai¹, Zhengting Yang¹, Yin Yi¹*

1 Key Laboratory of State Forestry Administration on Biodiversity Conservation in Karst Mountainous Areas of Southwestern China /Key Laboratory of Plant Physiology and Developmental Regulation/ School of Life Sciences, Guizhou Normal University, Guiyang, China, **2** School of Life Sciences, Guizhou Normal University, Guiyang, China

✉ These authors contributed equally to this work.

* yanhuiqing@gznu.edu.cn (HY); yiyin@gznu.edu.cn (YY)



OPEN ACCESS

Citation: Huang X, Yan H, Zhai L, Yang Z, Yi Y (2019) Characterization of the *Rosa roxburghii* Tratt transcriptome and analysis of *MYB* genes. PLoS ONE 14(3): e0203014. <https://doi.org/10.1371/journal.pone.0203014>

Editor: David L. Remington, University of North Carolina at Greensboro, UNITED STATES

Received: August 9, 2018

Accepted: February 16, 2019

Published: March 12, 2019

Copyright: © 2019 Huang 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 relevant data are within the manuscript, its Supporting Information files, or at the NCBI Gene Expression Omnibus (GEO) database under the accession number GSE122014.

Funding: This work was supported by grants from the National Natural Science Foundation of China (Grant No. 31660554, 31600214 and 31660046), Science Foundation of Guizhou Provinces (Qiankehe J zi (2015) 2117). Guizhou Science and technology program (Qiankehe Platform for talents [2017]5726-50).

Abstract

Rosa roxburghii Tratt (Rosaceae) has a fruit that is flavorful, economically valuable, and highly nutritious, providing numerous health benefits. Myeloblastosis (*MYB*) proteins play key roles in the development and fruit quality of *R. roxburghii*. However, there is little available genomic and transcriptomic information for *R. roxburghii*. Here, a normalized cDNA library was constructed from five tissues, including the stem, leaf, flower, young fruit, and mature fruit, using the Illumina HiSeq 3000 platform. *De novo* assembly was performed, and 470.66 million clean reads were obtained. In total, 63,727 unigenes, with an average GC content of 42.08%, were discovered, 60,406 of which were annotated. In addition, 9,354 unigenes were assigned to Gene Ontology categories, and 20,202 unigenes were assigned to 25 Eukaryotic Ortholog Groups. Additionally, 19,508 unigenes were classified into 140 pathways of the Kyoto Encyclopedia of Genes and Genomes database. Based on the transcriptome, 163 unigenes associated with *MYB*s were detected. Among these genes, 75 genes were significantly expressed in the various tissues, including 10 R1 *MYB*, 42 R2R3 *MYB*, one R1R2R3 *MYB*, three R4 *MYB* and 19 atypical *MYB*-like proteins. The expression levels of the 12 *MYB* genes randomly selected for quantitative real-time PCR analysis corroborated the RNA sequencing results. A total of 37,545 microsatellites were detected, with an average expressed sequence tag-simple sequence repeat frequency of 0.59 (37,545/63,727). This transcriptome data improves our understanding of the role of *MYB* in *R. roxburghii* and will be valuable for identifying genes of interest.

Introduction

Rosa roxburghii Tratt (Rosaceae), commonly known as chestnut rose, is a fruit crop and deciduous horticultural shrub that is mainly distributed in southwestern China. It has several nutritional and functional components, including polysaccharides, flavonoids, triterpenes, and also exhibits superoxide dismutase activity [1, 2]. Some pharmacological researches indicate that it could be used for the inhibition of the metastasis and invasion of ovarian cancer cells and

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

senescence-retardation [3, 4]. In addition, its fruits exhibit radio-protective, anti-tumor, anti-mutagenic, and genoprotective activities [5–7]. *Rosa roxburghii* is also a traditional Chinese medicinal plant that is further processed into fruit juice, preserves, and fruit wine [8].

Myeloblastosis (MYB) proteins have been reported to significantly influence plant development and fruit quality [9, 10]. Among the different transcription factor families, MYBs form the largest and most functionally diverse superfamily and are involved in regulating cell activities and plant development [11]. The N-terminus of a MYB domain is composed of adjacent tandem repeats [10]. The repeat encodes 50–53 amino acid residues and contains helices forming a helix-turn-helix domain that interact with the major grooves of specific DNA sequences [12]. MYB superfamily members are classified into several subfamilies based on the number of conserved domains (R), including R1 MYB (with one R), R2R3 MYB (with two Rs), R1R2R3 MYB (with three Rs), R4 MYB (with four Rs), and atypical MYB-like proteins [13]. Many MYB superfamily proteins and their functions have been determined in different species. Due to the conserved domains, the R2R3 MYB type is predominant in plants [14]. MYBs are involved in regulating plant growth, development, and stress resistance, including the anthocyanin biosynthetic pathway, trichome initiation and development, flavonoid or phenylpropanoid metabolism, secondary wall biosynthesis, sugar signaling and responses to abiotic or biotic stress [15, 16].

The genes of the *R. roxburghii* MYB superfamily have not been comprehensively characterized. Thus, for species that lack reference genomes, high-throughput Illumina sequencing can be used to generate new transcripts, determine gene expression levels, and produce an accurate transcriptome profile [17]. Assembled unigenes with different database annotations can be assayed to evaluate genetic characteristics and metabolic pathways. Recently, the fruits of *R. roxburghii* at three different developmental stages were analyzed using Illumina sequencing [18]. The author detected several MADS-box genes related to fruit development, as well as NAD genes. Several transcripts related to ascorbate biosynthesis were also identified. However, the sequencing depth was insufficient to represent the entire transcriptome. In another study, genomic survey sequencing for the genetic background of *R. roxburghii* was also performed using next-generation sequencing technology by HiSeq 2500 sequencing. The estimated genome size was 480.97 Mb based on the ratio of K-mer number to peak depth, and the findings mainly illustrated the genes related to the biosynthesis of ascorbic acid [19].

To elucidate the transcriptomic profiles of different tissues in *R. roxburghii*, the leaf, stem, flower, young fruit and mature fruit were subjected to RNA sequencing (RNA-Seq) analysis. The Illumina platform was used to construct a cDNA library using these five tissues to obtain transcriptome information. MYBs, which were significantly expressed in the various tissues, were identified. Expressed sequence tag-simple sequence repeats (EST-SSRs) were used to assess the genetic diversity of *R. roxburghii*. These results provide a valuable resource for functional gene analyses, particularly that of MYB genes.

Materials and methods

Biological materials

Seedlings of wild *Rosa roxburghii* were provided by Hongying Zhou (Guizhou Botanical Garden, Guizhou, China) and planted outside in pots (one pot per seedling) under natural conditions at Guizhou Normal University, Guizhou province, China (N 26°42.408'; E 106°67.353'). Different tissues were collected from these plants exhibiting healthy growth, including the leaves, stems, flowers at the full-blossom stage, young fruit (50 days after flowering, YF) and mature fruit (120 days after flowering, MF). The former three samples were collected at the same time in May, 2017, while the latter two samples were respectively obtained in July and

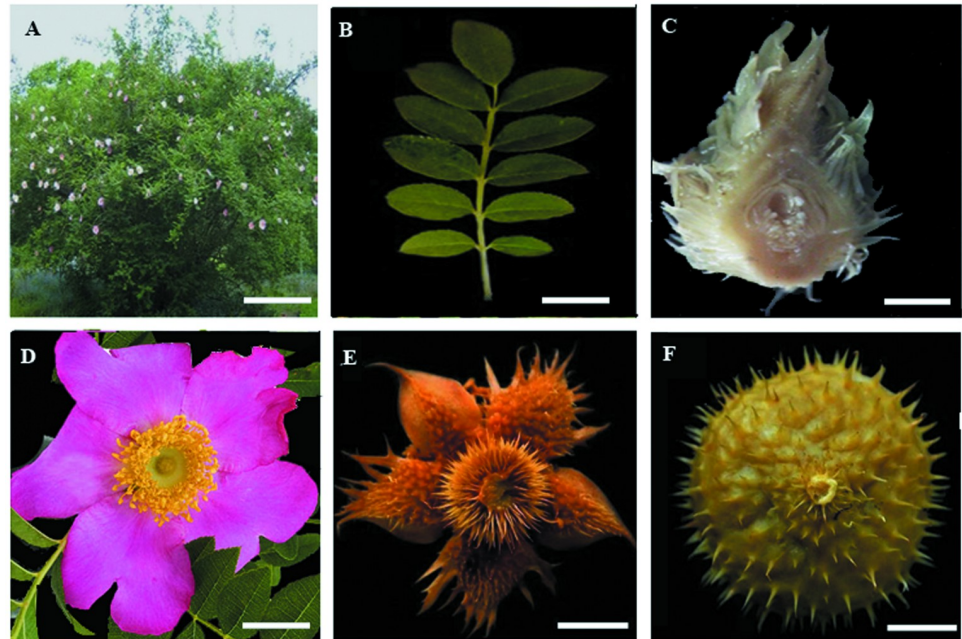


Fig 1. Photographs of the *Rosa roxburghii*. (A) plant, bar = 50 cm; (B) leaf and stem, bar = 0.5 cm; (C) flower bud, bar = 1mm; (D) flower, bar = 1 cm; (E) young fruit (50 days after flowering, DAF), bar = 1 cm; (F) mature fruit (120 DAF), bar = 1 cm.

<https://doi.org/10.1371/journal.pone.0203014.g001>

September (Fig 1A–1F). Three biological replicates were sampled for each tissue, resulting in a total of 15 samples. These materials were immediately frozen in liquid nitrogen following collection and then mechanically ground into a fine powder and stored at -80°C until further analysis.

RNA extraction and sequencing

A total of 0.1 g of frozen leaf, stem, flower, young fruit and mature fruit tissue were weighed. RNA was isolated using the Trizol method (Takara, Japan) with the addition of RNAiso-mate (Takara, Japan) according to the manufacturer's guidelines. RNA quality was determined with a Nanodrop spectrophotometer (Wilmington, USA) and Agilent 2100 bioanalyzer (Santa Clara, USA). The total RNA was treated with RNase-free DNase I to eliminate DNA contamination. The enrichment of mRNA was performed with Oligo (dT)-attached magnetic beads, and fragmentation was performed using divalent cations under high temperature in NEBNext First Strand Synthesis Reaction Buffer (5 \times) to randomly interrupt mRNA. The first-strand cDNA chain was synthesized by random hexamers and M-MuLV Reverse Transcriptase (RNase H) with the mRNA template. Subsequently, the second-strand cDNA chain was synthesized by adding buffer, dNTPs, RNase H and DNA polymerase I. The the double-stranded cDNA was purified using AMPure XP beads. Following end-pair and single nucleotide A (adenine) addition for the purified cDNA, adapters were used to distinguish the different samples by selection with AMPure XP beads for PCR amplification, following which the cDNA library was obtained. Strand-specific cDNA libraries were constructed for sequencing using an Illumina HiSeq 3000 (Illumina, San Diego, USA) at Huazhong Agricultural University (Wuhan, China) to generate 150-bp paired-end reads.

De novo assembly and functional annotation

The clean reads were trimmed by filtering out adaptor-only reads, reads with more than 5% unknown N bases, and low-quality reads (reads containing more than 50% bases with a Q-value ≤ 10). *De novo* assembly was performed using the Trinity assembly program with default settings based on the de Bruijn graph algorithm [20]. Clean reads were first broken into shorter fragments (K-mers) and then assembled into longer fragments named contigs. The overlapping reads were mapped back to the corresponding contigs. Based on the paired-end reads, different contigs from the same transcripts were identified, and the distances among these contigs were calculated. These contigs were further assembled using Trinity to obtain sequences that were no longer defined as unigenes [20]. In addition, the assembled sequences that were less than 200 bp in size were removed. The assembled genome was used as the genomic reference. The raw sequence reads have been archived at the NCBI Gene Expression Omnibus (GEO) database with the accession number (GEO No. GSE122014).

To determine the function of the unigenes, BLASTx alignment with an E-value $\leq 10^{-5}$ was performed with different databases, including KOG (Eukaryotic Ortholog Groups, <http://www.ncbi.nlm.nih.gov/KOG/>) [21], Nr (NCBI non-redundant protein database, <http://www.ncbi.nlm.nih.gov/>), KEGG (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/>) [22], Gene Ontology (GO, <http://www.geneontology.org/>) and Swiss-Prot (<http://www.expasy.ch/sprot>) [23, 24]. Based on the Nr annotations, the program Blast2GO [25] was applied to acquire the GO terms of the unigenes. After the GO annotations were obtained, WEGO software was used to perform GO functional classification and identify the distribution of gene functions in *R. roxburghii* at the macro level [26]. The best-aligned results were used to identify the sequence direction of the unigenes. If different databases conflicted, the results were prioritized in the following order: Nr, Swiss-Prot, KEGG, COG and GO [27]. When transcripts did not align to any of the databases, EST scans (<http://myhits.isb-sib.ch/cgi-bin/estscan>) was conducted to decide the sequence direction.

Homology with other model *Rosa* species was also assessed. Nucleic acid sequences from strawberry (*Fragaria X ananassa*) (*Fragaria ananassa* GDR RefTrans V1), apple (*Malus x domestica*) (*Malus x domestica* GDR RefTrans V1) and cherry (*Prunus avium*) (*Prunus avium* GDR RefTrans V1) and *R. roxburghii* sequences were aligned using BLAST in the Uni-prot database. This was done in order to compare the four species based on exactly the same search parameters and database type.

Identification and conserved motif analysis of *Rosa roxburghii* superfamily

Bowtie2 [28] and RSEM (RNA-Seq by Expectation-Maximization) [29] were used to map clean reads to the assembled transcriptome and evaluate the abundance of each transcript, respectively. To compare the expression levels of transcripts in the different samples, FPKM (fragments per kilobase of transcript per million mapped reads) was used to normalize the gene expression [30] according to the formula: $FPKM = (10^6 \times C \times 10^3) / NL$. C represents the number of reads uniquely aligned to a certain unigene, N indicates the total number of reads uniquely aligned to all unigenes, and L is the base number of this unigene.

All genes related to MYBs were selected and analyzed based on Nr annotation and then described. The expression level of all MYB in the various tissues were assessed. The MYBs were predicted using Open Reading Frame (ORF) in TransDecoder (v2.1.0) (<http://transdecoder.github.io/>) with default parameters to obtain the protein. The longest ORF was used to identify the transcription sequence. Conserved motifs shared by MYB proteins, which were significantly expressed using a threshold value of absolute \log_2 FC (fold change) ≥ 1 with an FDR (false discovery rate) ≤ 0.01 , were analyzed using Multiple Em for Motif Elicitation

(MEME Version 5.0.2, <http://meme-suite.org/tools/meme>) online tool by uploading the amino acid sequences of the MYB superfamily members. The following parameter settings were applied: R1-MYB proteins, with one R; R2R3-MYB proteins, with two Rs; R1R2R3-MYB proteins, with three Rs; R4-MYB proteins, with four Rs. Others belonging to atypical MYB families were determined [31].

Quantitative real-time PCR (qRT-PCR) assays

Twelve cDNAs encoding MYB transcription factors, all of which have potential roles in the regulation plant development, were selected for qRT-PCR validation. Primers were designed (S1 Table) with primer premier 6. Total RNAs were isolated from leaf, stem, flower, young fruit (YF) and mature fruit (MF) using the TRIzol, followed by purification with an RNA purification kit with added RNAiso-mate. The RNA was then reverse-transcribed using an RT-PCR Kit (TaKaRa, Japan) with an oligo dT-adaptor primer according to the manufacturer's protocol. The volume of each reaction was 20 μ L prepared with 10 μ L of SYBR Green Master mix (Toyobo, Osaka, Japan), 1 μ L of each primer pair, and 1 μ L of cDNA template, and 7 μ L sterilized ddH₂O. The qRT-PCR reactions were performed in a Roche LightCycler480 machine with β -actin as an endogenous control. Three biological replicates were tested. Amplification was performed for 95°C for 2 min, followed by 40 cycles at 95°C for 15 s, annealing at 58°C for 30 s, and 72°C for 30 s. The expression levels relative to the control were estimated by calculating $\Delta\Delta C_t$ and were subsequently analyzed using $2^{-\Delta\Delta C_t}$ method.

Microsatellite detection

The program MISA (<http://pgrc.ipksgatersleben.de/misa/>) [32] was used to detect microsatellite repeat motifs for each unigene in order to determine the distributions of microsatellites (also known as SSRs) and to develop new markers in the transcriptome of *R. roxburghii*. The number of core repeat motifs in mononucleotide, di-nucleotides, tri-nucleotide tetra-nucleotide, penta-nucleotide and hexa-nucleotides was counted.

Results

Illumina sequencing and sequence assembly

A total of 470.66 million reads were obtained for all five tissues and all repetitions. After trimming and quality filtration of the raw data, each tissue was represented by an average of 31.38 million reads (S2 Table). The total reads per biological condition are indicated in S2 Table. For each sample, at least 86.21% of the reads could be mapped uniquely to contigs assembled using Trinity software. A total of 212,534 transcripts were obtained by assembling the clean reads using Trinity, with an average guanine plus cytosine (GC) content of 42.08%, an average length of 1437.69 bp and an N50 length of 2,085 bp. A total of 63,727 unigenes were assembled and all longer than 200 bp, with average and N50 lengths of 995 bp and 1,895 bp, respectively (Table 1). Of the 63,727 unigenes, 78.03% (49,727) were longer than 600 bp and 56.07%

Table 1. Summary of the trinity assembly for *Rosa roxburghii*.

Length range (bp)	Transcript	Unigene
Total Number	212,534	63,727
Total length	305,558,508	85,677,327
N50 length	2,085	1,895
Median Length	1,077	995

<https://doi.org/10.1371/journal.pone.0203014.t001>

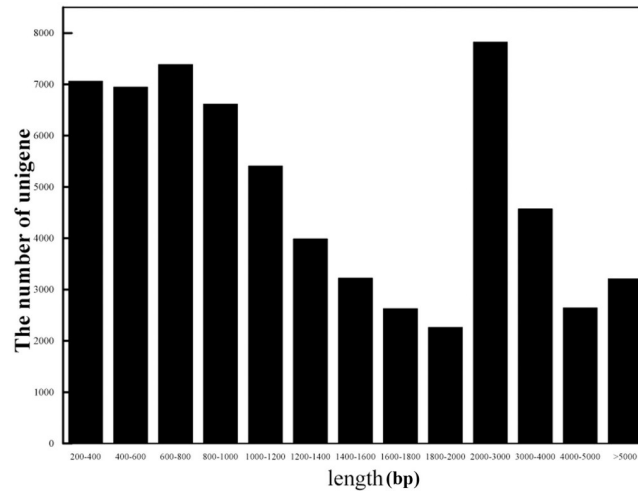


Fig 2. Length distribution of unigenes assembly for *Rosa roxburghii* Tratt.

<https://doi.org/10.1371/journal.pone.0203014.g002>

(35,732) were longer than 1 kb (Fig 2). In addition, most unigenes (60,901) were less than 5,200 bp (95.57%) (S3 Table and Table 1).

Functional annotation of unigenes

Using a BLASTX-based algorithm, 60,406 out of the 63,727 unigenes could be mapped to the GO, KOG, KEGG, Nr, and Swiss-Prot databases (S4 Table). Swiss-prot and Nr contained the highest number of homologous unigenes (55,151 and 55,118, respectively). In total, 1,948 unigenes could be annotated to all databases (Fig 3A). Comparative sequence alignment identified differences and similarities among the various *Rosa* species. Nucleic acid sequences from *Rosa* species, including strawberry, apple, and cherry, were aligned using a BLAST algorithm-based search of the Uniprot database. We found that strawberry had the highest number of similar genes, followed by cherry and then apple, with the numbers of unigenes matching those of *R. roxburghii* being 30,577 (47.98%), 20,105 (31.55%), and 17,677 (27.74%), respectively (Fig 3B).

There are three GO categories: biological process, cellular component, and molecular function (S5 Table). The Category “biological process” consisted of 20 functional groups, with the

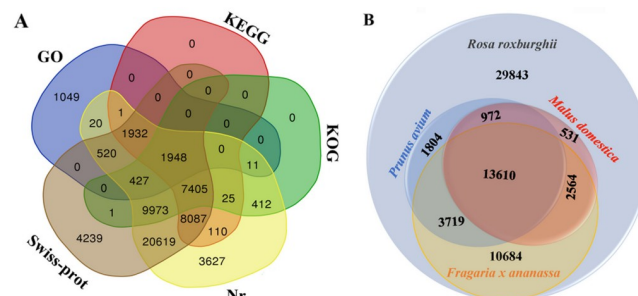


Fig 3. (A) Venn diagram of the number of unigenes annotated in five different databases. The number in the circles represents the number of unigenes annotated by single or multiple databases; **(B)** homology to strawberry (*Fragaria x ananassa*), apple (*Malus domestica*) and cherry (*Prunus avium*).

<https://doi.org/10.1371/journal.pone.0203014.g003>

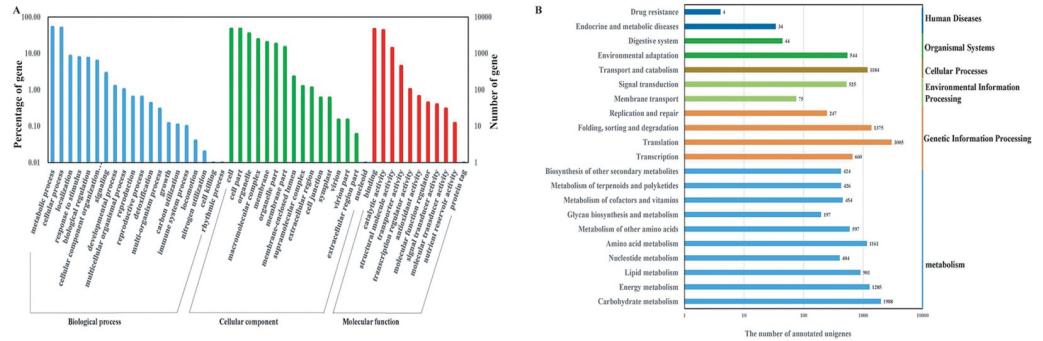


Fig 4. (A) GO functional classification of the unigenes. Blue indicates biological process, green indicates cellular process, and red represents molecular function. The axis labels are “Percentage of genes” (on the left) and “Number of genes” (on the right). (B) KEGG pathway analysis of the unigenes. The x-axis indicates the number of unigenes in a specific category. The left y-axis indicates the clustered functional groups, and the right y-axis indicates the specific category of the genes in the main category.

<https://doi.org/10.1371/journal.pone.0203014.g004>

major groups being metabolic process (56.56%) and cellular process (54.02%), followed by localization (9.10%) and response to stimulus (8.38%). For the category “cellular part”, 16 groups were predicted, with cell (49.89%), cell part (49.80%), and organelle (37.24%) constituting the three major groups. For “molecular function”, binding (49.02%) and catalytic activity (46.01%) were the dominant groups, followed by structural molecule activity (14.89%) (Fig 4).

A total of 20,202 unigenes were identified using the KOG database (S1 Fig) and annotated to 25 functional categories. General function prediction (46.37%) was the largest group, followed by signal transduction mechanisms (24.26%), posttranslational modification, protein turnover, and chaperones (23.49%), and translation, ribosomal structure, and biogenesis (19.80%). The numbers of unigenes assigned to transcription (11.81%), carbohydrate transport and metabolism (11.70%), energy production and conversion (11.67%), and intracellular trafficking, secretion, and vesicular transport (11.29%) were almost identical. In addition, lipid transport and metabolism represented 10.29%. However, there were still 1,939 unigenes with unknown functions. KOG classifications revealed the potential biological functions and provided an insight into the chemical reactions involved in the molecular processes in *R. roxburghii*.

A total of 19,508 annotated unigenes were assayed to determine the biological pathways represented in *R. roxburghii*. Briefly, these unigenes matched 140 KEGG pathways, as summarized in S6 Table. Translation in genetic information processing (3,005) was the dominant pathway, followed by carbohydrate metabolism (1,988), folding, sorting and degradation (1,375), energy metabolism (1,285), transport and catabolism (1,184), amino acid metabolism (1,161), and lipid metabolism (901) (Fig 4). The KEGG pathways can provide new insights into the biological properties of *R. roxburghii* and contribute to the prediction of the higher-level complexity of cellular processes and organismal behavior.

Genes involved with MYB transcriptional factors in five different tissues

According to the functional database annotations, 163 MYBs were identified. Descriptions of the MYBs are listed in S7 Table, and 61 MYBs in *R. roxburghii* could be annotated with *Fragaria vesca*. MYBs regulate secondary metabolism and gene expression and are involved in environmental stress responses. The expression levels of 159 putative MYB genes in five tissues are indicated in a heatmap in Fig 5, produced in the R statistical environment (version 3.1.3), while four other genes (*DN141656_c0_g3*, *DN143182_c2_g1*, *DN145149_c1_g6* and

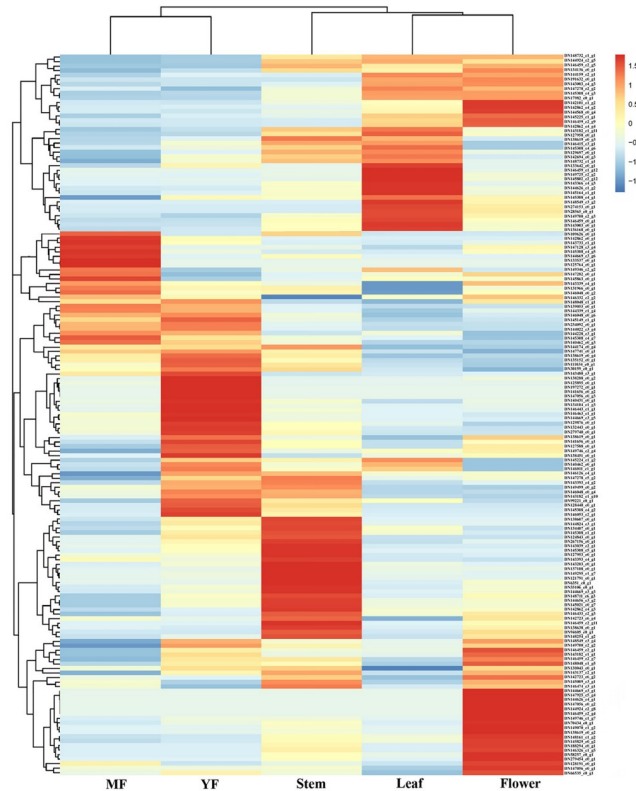


Fig 5. Heatmap and dendrogram indicating the expression levels of all MYBs using R software (version 3.1.3).

<https://doi.org/10.1371/journal.pone.0203014.g005>

DN142101_c1_g3) were not expressed in any tissues (S7 Table). Various MYBs exhibited differential expression among five tissues. Based on an expression value of \log_2 FC (fold change) ≥ 1 between any two tissues, there were 75 differentially expressed MYBs in the various tissues (S8 Table).

To investigate the features of homologous domains and each repeat MYB domain, the online MEME was used to search for the conserved motifs shared by these proteins by uploading the amino acid sequences. In total, 75 candidate MYBs were analyzed and are indicated in Fig 6. Ten R1 MYB, 42 R2R3 MYB, one R1R2R3 MYB, and three R4 MYBs were identified, while the remaining 19 proteins belonged to atypical MYB-like families (S8 Table). The sequences of the different conserved motifs are illustrated in S2 Fig

Verification of RNA-seq results by qRT-PCR

qRT-PCR was conducted to validate the identification of the MYB genes obtained by RNA-seq analysis in the five different tissues. Using β -actin as the internal control, 12 genes related to MYB transcription factors were randomly selected. Validation results showed that the change trends of the 12 genes were nearly consistent with the gene expression patterns identified by RNA-seq (Fig 7), thereby confirming the reliability and accuracy of the RNA-seq analysis.

Microsatellite analysis and SSR distribution

Using MISA software, 63,727 unigenes with a total length of 109,644,660 bp were screened for microsatellite determination. A total of 37,545 potential EST-SSRs were identified. The

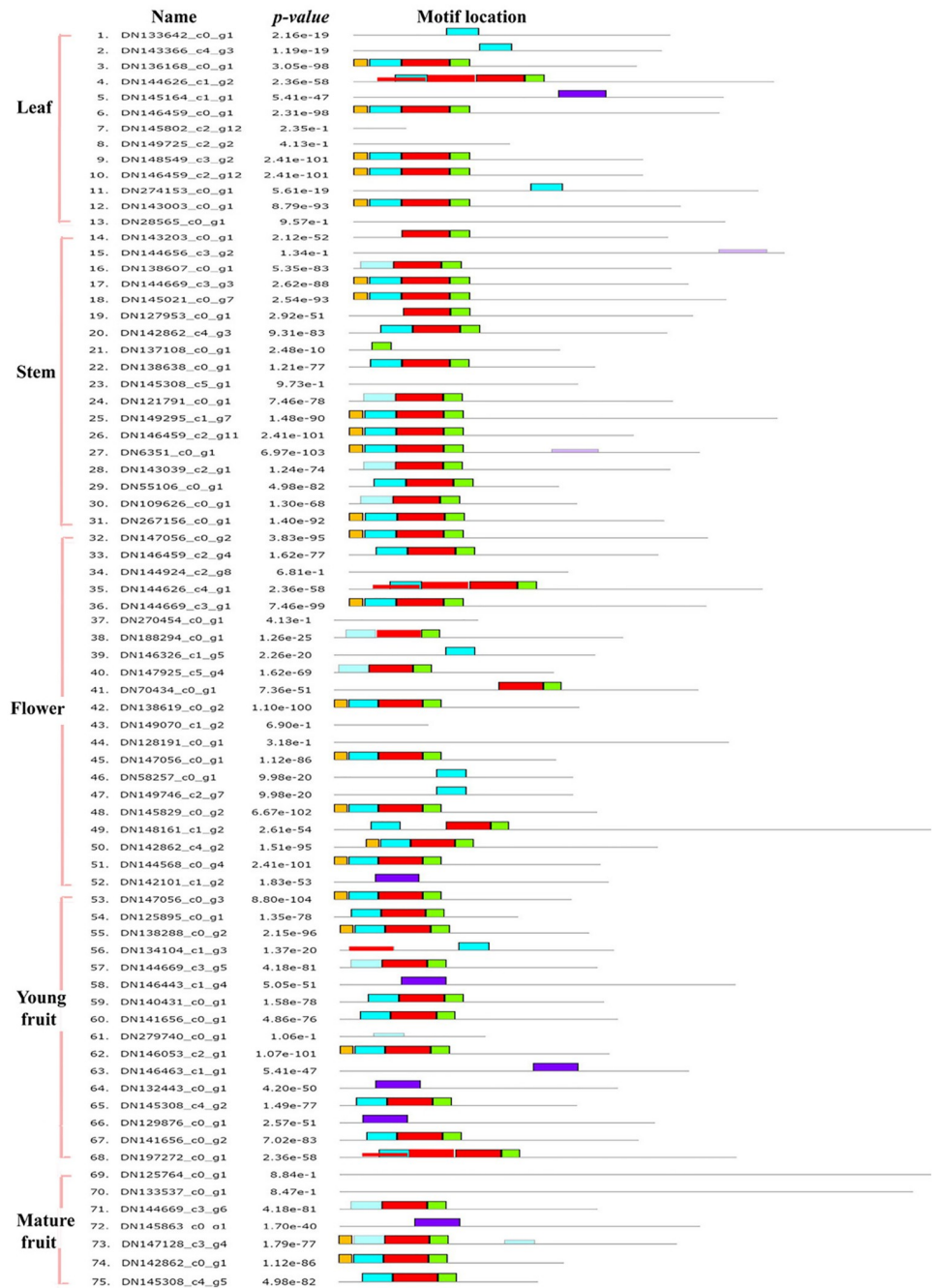


Fig 6. Seventy-five MYB motifs that were significantly expressed in the different tissues.

<https://doi.org/10.1371/journal.pone.0203014.g006>

average distribution of the SSRs was calculated to be 1:2,920 bp (37,545/109,644,660), and the average frequency of an EST-SSR was 0.59 (37,545/63,727). In total, 20,321 unigenes contained one type of SSR (54.12%), and 8,757 contained more than one type of SSR (23.32%; S9 Table). In total, 5,275 (14.05%) SSRs were present in a compound formation. Transcriptome types of the SSRs, from single nucleotide to hexanucleotide, were abundant.

Among the identified 37,545 SSRs, repeats with mononucleotide motifs were the most abundant (19,589, 52.17%), followed by di-nucleotides (11,504, 30.64%) (Table 2). The most

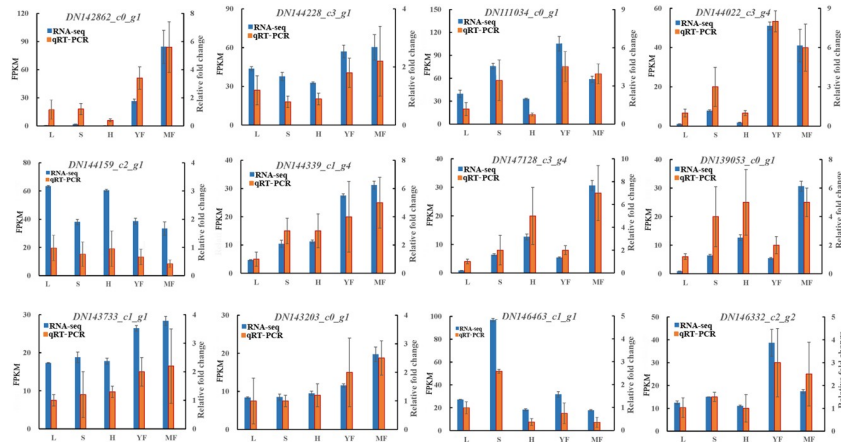


Fig 7. The relative expression levels of 12 randomly selected MYB by RNA-Seq and qRT-PCR. The horizontal axis indicates different tissues, including leaf (L), stem (S), flower (F), young fruit (YF) and mature fruit (MF). The left and right vertical axis individually indicate FPKM using RNA-seq and relative expression level using q RT-PCR. The error bars represent the standard deviation derived from each individual sample in triplicate.

<https://doi.org/10.1371/journal.pone.0203014.g007>

abundant motif was A or T (18,855, 50.22%), followed by AG or CT (8,055, 21.45%) (Table 2). Among SSRs with tri-, tetra-, and penta-nucleotides, the most abundant types were AAG/CTT (2,189, 5.83%), AAAT/ATTT (141, 0.38%), and AAAAT/ATTTT (19, 0.05%), respectively. The hexa-motifs AAGGAG/CCTTCT and ACCTCC/AGGTGG were the most abundant types and were equally present (7, 0.02%) (Table 2). The repeat positions of the SSR types were analyzed and ranged from 5 to 121. Most SSR types were repeated more than 15 times [19.01% (7,136)], while 17.10% (6,422) were repeated 10 times (Table 3). With the exception of mononucleotides, the repeat numbers for most SSRs ranged from 5 to 12 (9,612, 75.9%), with only a small percentage being repeated more than 15 times (1,177, 9.3%).

Discussion

The limited available genomic information for *R. roxburghii* has constrained previous genetic studies. The generation of RNA-Seq libraries using short-read Illumina sequencing technology is considered as an effective approach for assessing transcriptional expression in tissues of interest in species lacking reference genomes. In the present study, the clean reads of five tissues were used for library construction, and 470,657,040 clean reads were obtained with 150-bp paired-end reads. This is greater than in a previous transcriptome study of *R. roxburghii* [18]. Similar numbers of unigenes were assigned to GO categories as a previous study, while 19,508 unigenes were assigned to 140 KEGG pathways compared with the 7,480

Table 2. Summary of EST-SSRs identified from the transcriptome of *Rosa roxburghii*.

SSR Type	Number	Percentage	Dominant motif	Number of dominant motif	Percentage of dominant motif in all EST-SSR
Mononucleotide	19,589	52.17%	A/T	18,855	50.22%
Dinucleotide	11,504	30.64%	AG/CT	8,055	21.45%
Trinucleotide	5,879	15.66%	AAG/CTT	2,189	5.83%
Tetranucleotide	373	0.99%	AAAT/ATTT	141	0.38%
Pentanucleotide	93	0.25%	AAGAG/CTCTT	19	0.05%
Hexanucleotide	107	0.28%	AAGGAG/CCTTCT	7	0.02%
			ACCTCC/AGGTGG	7	0.02%

<https://doi.org/10.1371/journal.pone.0203014.t002>

Table 3. Summary of different repeat times for SSRs isolated from the transcriptome of *Rosa roxburghii*.

SSR Type	5	6	7	8	9	10	11	12	13	14	15	>15	Total
Mononucleotide	0	0	0	0	0	5448	2864	2019	1380	1165	905	5808	19589
Dinucleotide	0	2441	1765	1368	1158	883	985	858	216	259	263	1308	11504
Trinucleotide	2965	1385	727	461	91	86	69	40	20	8	7	20	5879
Tetranucleotide	242	95	15	7	4	5	0	3	2	0	0	0	373
Pentanucleotide	67	17	8	1	0	0	0	0	0	0	0	0	93
Hexnucleotide	74	22	7	1	1	0	1	1	0	0	0	0	107
Total	3348	3960	2522	1838	1254	6422	3919	2921	1618	1432	1175	7136	37545
Percentage	8.92%	10.55%	6.72%	4.90%	3.34%	17.10%	10.44%	7.78%	4.31%	3.81%	3.13%	19.01%	100%

<https://doi.org/10.1371/journal.pone.0203014.t003>

unigenes assigned to 124 pathways in the previous study [18]. This is because more tissues were analyzed and more unigenes were annotated with a greater number of pathways in the present study. The GC content of *R. roxburghii* was calculated to be 42.05%, which is higher than the 38.63% detected in an earlier next-generation sequencing study [19]. A total of 63,727 unigenes were predicted with an average length of 1437.69 bp in our study, while 22,721 genes with an average length of 2311.52 bp were detected in the aforementioned study [19]. The greater volume of data in the present study increased the coverage depth and accuracy, allowing for the identification of a large number of genes involved in different metabolic pathways. Although five tissues with three repetitions each were used to construct cDNA libraries, this does not cover the entire transcriptome of *R. roxburghii*, as some transcripts, not expressed in these tissues, may be missed.

Based on the MYB expression patterns, 75 MYBs were significantly expressed in the various tissues. Among these MYBs, 42 R2R3 MYBs were identified, proving to be the dominant type. The MYBs presented similar patterns and conserved motifs, suggesting that their conserved features play similar roles in group-specific functions. For example, DN141656_c0_g3, which could constitute part of the MYB-bHLH-WD40 complex, is involved in plant trichome development [33]. Some R2R3 MYBs are related to secondary metabolism and responses to abiotic and biotic stresses [34–35]. Several R2R3 MYBs, such as DN111034_c0_g1 annotated as MYB12 and DN146459_c2_g7 annotated as MYB108, were analyzed for their possible involvement in flavonoid or phenylpropanoid biosynthesis and accumulation, as well as their regulation under different stress and hormone treatments [36–38]. MYBs are also reported to be implicated in sugar signaling, fruit-skin coloration, and anthocyanin metabolism [39–41]. The majority of the R2R3 MYBs detected in the present study are associated with plant growth and fruit development and could be further studied to improve fruit quality and stress resistance in *R. roxburghii*. Based on the similar functions of homologous gene with other *Rosa* species, the results of this study will aid functional studies of the MYB genes involved in *R. roxburghii* plant development.

A total of 37,545 microsatellites with different repeat types were detected from 63,727 unigenes, indicating that each unigene, on average, contained 0.59 SSR. The SSR locus density was 1:2,920 bp, compared with 1:4.00 kb in a previous study [18]. Various criteria and parameters for SSR detection, as well as the diversity of genomic structures and compositions can influence SSR density [42]. While a positive and significant association between genome size and SSRs motif length exists in gymnosperms, SSR density and length contribute less to genome size diversity in angiosperms [43]. Errors in sequencing and assembly mistakes that resulted in mononucleotide SSRs were relatively low [44]. Except mononucleotides, the most common SSR motif was dinucleotide repeats in the transcriptome, which was similar to

previously reported results. Here, AC/GT was the common type. In conclusion, detected EST-SSRs (37,545) are more closely associated with functional genes than genomic SSRs [45]. Therefore EST-SSRs could provide valuable information for genetic and genomic analyses [46].

In summary, a deep RNA-seq analysis was conducted on five tissues, and a total of 470.66 million reads were generated. In total, 63,727 unigenes were obtained using Trinity, of which nearly 94.79% (60,406) were successfully annotated. The results of this study have provided increased scope for the detection of genes involved in various metabolic pathways and have further elucidated the potential roles of *MYBs* in the different tissues in *R. roxburghii*. The detected microsatellites allow for the identification of genetic linkage mapping construction and maker-assisted selection.

Supporting information

S1 Table. Primers for real-time PCR.

(XLSX)

S2 Table. Summary of read mapping in leaf, flower, stem, young fruit and mature fruit with three repetitions.

(XLSX)

S3 Table. Analysis of size distribution of unigene for *Rosa roxburghii* Tratt.

(XLSX)

S4 Table. The number of unigenes annotated according to the NCBI non-redundant (Nr), Swiss-Prot, KOG, Gene Ontology (GO) and KEGG database in *Rosa roxburghii* Tratt.

(XLSX)

S5 Table. The number of unigenes annotated with GO and classified with three categories.

(XLSX)

S6 Table. The number of unigenes involved in KEGG pathways.

(XLSX)

S7 Table. Analysis of genes related to MYB transcription factor in leaf, stem, flower, young fruit and mature fruit of *Rosa roxburghii* Tratt. FPKM means fragments per kilobase of transcript per million mapped reads.

(XLSX)

S8 Table. The tissues where *MYB* exhibited significantly differential expression based on a threshold value of absolute \log_2 FC (fold change) ≥ 1 , and types of MYB superfamily were listed.

(XLSX)

S9 Table. Summary of different type EST-SSRs identified from the transcriptome.

(XLSX)

S1 Fig. Histogram of Eukaryotic Ortholog Groups (KOG) classification of assembled unigenes.

(TIF)

S2 Fig. Logo sequences of the conserved motifs obtained using the MEME by uploading the amino acid sequences of seventy-five MYB. The overall height of each individual stack represents the conservation of the sequence at that position. The Arabic numerals under the colored capital letters represent the position of each residue and the width of the motif. Each

color of the English letters indicates a different type of amino acid residue. (TIF)

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (Grant No. 31660554, 31600214 and 31660046), Science Foundation of Guizhou Provinces (Qiankehe J zi (2015) 2117). Guizhou Science and technology program (Qiankehe Platform for talents [2017]5726–50).

We would like to thank LetPub (www.letpub.com) for providing linguistic assistance during the preparation of this manuscript.

Author Contributions

Formal analysis: Lisheng Zhai, Zhengting Yang.

Funding acquisition: Xiaolong Huang, Zhengting Yang.

Investigation: Lisheng Zhai.

Methodology: Lisheng Zhai, Yin Yi.

Resources: Yin Yi.

Supervision: Xiaolong Huang.

Visualization: Xiaolong Huang.

Writing – original draft: Xiaolong Huang.

Writing – review & editing: Huiqing Yan.

References

1. Chen G, Kan J. Characterization of a novel polysaccharide isolated from *Rosa roxburghii* Tratt fruit and assessment of its antioxidant in vitro and in vivo. *Int J Biol Macromol*. 2018; 107: 166–174. <https://doi.org/10.1016/j.ijbiomac.2017.08.160> PMID: 28866014.
2. Xu P, Liu X, Xiong X, Zhang W, Cai X, Qiu P, et al. Flavonoids of *Rosa roxburghii* Tratt Exhibit Anti-Apoptosis Properties by Regulating PARP-1/AIF. *J Cell Biochem*. 2017; 118(11): 3943–3952. <https://doi.org/10.1002/jcb.26049> PMID: 28398610.
3. Chen Y, Liu ZJ, Liu J, Liu LK, Zhang ES, Li WL. Inhibition of metastasis and invasion of ovarian cancer cells by crude polysaccharides from *rosa roxburghii* tratt in vitro. *Asian Pac. J. Cancer Prev*. 2014; 15(23):10351–10354. PMID: 25556474
4. Ma YX, Zhu Y, Wang CF, Wang ZS, Chen SY, Shen MH, et al. The aging retarding effect of ‘Long-Life CiLi’. *Mechanisms of Ageing & Development*. 1997; 96(1–3):171–180.
5. Xu P, Zhang WB, Cai XH, Lu DD, He XY, Qiu PY, et al. Flavonoids of *Rosa roxburghii* Tratt act as radio-protectors. *Asian Pac J Cancer Prev*. 2014; 15(19): 8171–8175. PMID: 25339001.
6. Xu SJ, Zhang F, Wang LJ, Hao MH, Yang XJ, Li NN, et al. Flavonoids of *Rosa roxburghii* Tratt offers protection against radiation induced apoptosis and inflammation in mouse thymus. *Apoptosis*. 2018; 23(9–10): 470–483. <https://doi.org/10.1007/s10495-018-1466-7> PMID: 29995207.
7. Chen G, Kan J. Ultrasound-assisted extraction, characterization, and antioxidant activity in vitro and in vivo of polysaccharides from Chestnut rose (*Rosa roxburghii* tratt) fruit. *J Food Sci Technol*. 2018; 55(3): 1083–1092. <https://doi.org/10.1007/s13197-017-3023-8> PMID: 29487451
8. Liu MH, Zhang Q, Zhang YH, Lu XY, Fu WM, He JY. Chemical Analysis of Dietary Constituents in *Rosa roxburghii* and *Rosa sterilis* Fruits. *Molecules*. 2016; 21(9): 1204–1224. <https://doi.org/10.3390/molecules21091204> PMID: 27618004.
9. Li G, Chen D, Tang X, Liu Y. Heterologous expression of kiwifruit (*Actinidia chinensis*) GOLDEN2-LIKE homolog elevates chloroplast level and nutritional quality in tomato (*Solanum lycopersicum*). *Planta*. 2018; 247(6): 1351–1362. <https://doi.org/10.1007/s00425-018-2853-6> PMID: 29520458.

10. Zhang C, Ma R, Xu J, Yan J, Guo L, Song J, et al. Genome-wide identification and classification of MYB superfamily genes in peach. *PLoS One*. 2018; 13(6): e0199192–e0199217. <https://doi.org/10.1371/journal.pone.0199192> PMID: 29927971
11. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science*. 2000; 290(5499):2105–2110. PMID: 11118137
12. Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF. Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins. *Plant Journal*. 2004; 40(1): 22–34. <https://doi.org/10.1111/j.1365-313X.2004.02183.x> PMID: 15361138.
13. Mmadi MA, Dossa K, Wang L, Zhou R, Wang Y, Cisse N, et al. Functional Characterization of the Versatile MYB Gene Family Uncovered Their Important Roles in Plant Development and Responses to Drought and Waterlogging in Sesame. *Genes (Basel)*. 2017; 8(12): 362–379.
14. Hajiebrahimi A, Owji H, Hemmati S. Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome*. 2017; 60(10): 797–814. <https://doi.org/10.1139/gen-2017-0059> PMID: 28732175.
15. Stracke R, Ishihara H, Hupé G, Barsch A, Mehrtens F, Niehaus K, et al. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J*. 2007; 50(4): 660–677. <https://doi.org/10.1111/j.1365-313X.2007.03078.x> PMID: 17419845
16. Feng K, Xu ZS, Que F, Liu JX, Wang F, Xiong AS. An R2R3-MYB transcription factor, OjMYB1, functions in anthocyanin biosynthesis in *Oenothera javanica*. *Planta*. 2018; 247(2): 301–315. <https://doi.org/10.1007/s00425-017-2783-8> PMID: 28965159.
17. Zhu Y, Chen L, Zhang C, Hao P, Jing X, Li X. Global transcriptome analysis reveals extensive gene remodeling, alternative splicing and differential transcription profiles in non-seed vascular plant *Selaginella moellendorffii*. *BMC Genomics*. 2017; 18 (Suppl 1):1042–1056. <https://doi.org/10.1186/s12864-016-3266-1> PMID: 28198676
18. Yan X, Zhang X, Lu M, He Y, An H. De novo sequencing analysis of the *Rosa roxburghii* fruit transcriptome reveals putative ascorbate biosynthetic genes and EST-SSR markers. *Gene*. 2015; 561(1): 54–62. <https://doi.org/10.1016/j.gene.2015.02.054> PMID: 25701597.
19. Lu M, An H, Li L. Genome Survey Sequencing for the Characterization of the Genetic Background of *Rosa roxburghii* Tratt and Leaf Ascorbate Metabolism Genes. *PLoS One*. 2016; 11(2):e0147530–0147546. <https://doi.org/10.1371/journal.pone.0147530> PMID: 26849133
20. 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):644–652. <https://doi.org/10.1038/nbt.1883> PMID: 21572440
21. Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, et al. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol*. 2004; 5(2): R7–R34. <https://doi.org/10.1186/gb-2004-5-2-r7> PMID: 14759257
22. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 2004; 32: D277–280. <https://doi.org/10.1093/nar/gkh063> PMID: 14681412
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25(1): 25–29. <https://doi.org/10.1038/75556> PMID: 10802651
24. UniProt Consortium T. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2018; 46(5): 2699. <https://doi.org/10.1093/nar/gky092> PMID: 29425356
25. 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):3674–3676. <https://doi.org/10.1093/bioinformatics/bti610> PMID: 16081474.
26. Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, et al. WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res*. 2018; 46(W1):W71–W75. <https://doi.org/10.1093/nar/gky400> PMID: 29788377
27. Zhang J, Schmidt CJ, Lamont SJ. Transcriptome analysis reveals potential mechanisms underlying differential heart development in fast- and slow-growing broilers under heat stress. *BMC Genomics*. 2017; 18(1): 295–309. <https://doi.org/10.1186/s12864-017-3675-9> PMID: 28407751
28. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012; 9(4):357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
29. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *Bmc Bioinformatics*. 2011; 12(1):323.

30. T C, W BA, P G, MA, K G, vB MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*. 2010; 28(5):511–515. <https://doi.org/10.1038/nbt.1621> PMID: 20436464
31. Li X, Xue C, Li J, Qiao X, Li L, Yu L, et al. Genome-Wide Identification, Evolution and Functional Divergence of MYB Transcription Factors in Chinese White Pear (*Pyrus bretschneideri*). *Plant Cell Physiol*. 2016; 57(4):824–847. <https://doi.org/10.1093/pcp/pcw029> PMID: 26872835.
32. MISA: MicroSatellite identification tool <http://pgrc.ipk-gatersleben.de/misa/>
33. Pattanaik S, Patra B, Singh SK, Yuan L. An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. *Front Plant Sci*. 2014; 5:259–266. <https://doi.org/10.3389/fpls.2014.00259> PMID: 25018756
34. Wei Q, Luo Q, Wang R, Zhang F, He Y, Zhang Y, et al. A Wheat R2R3-type MYB Transcription Factor TaODORANT1 Positively Regulates Drought and Salt Stress Responses in Transgenic Tobacco Plants. *Front Plant Sci*. 2017; 8: 1374–1388. <https://doi.org/10.3389/fpls.2017.01374> PMID: 28848578
35. Brendolise C, Espley RV, Lin-Wang K, Laing W, Peng Y, McGhie T, et al. Multiple Copies of a Simple MYB-Binding Site Confers Trans-regulation by Specific Flavonoid-Related R2R3 MYBs in Diverse Species. *Front Plant Sci*. 2017; 8: 1864–1868. <https://doi.org/10.3389/fpls.2017.01864> PMID: 29163590
36. Mehrtens F, Kranz H, Bednarek P, Weissshaar B. The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. *Plant Physiol*. 2005; 138(2):1083–1096. <https://doi.org/10.1104/pp.104.058032> PMID: 15923334
37. Wang N, Xu H, Jiang S, Zhang Z, Lu N, Qiu H, et al. MYB12 and MYB22 play essential roles in proanthocyanidin and flavonol synthesis in red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant J*. 2017; 90(2):276–292. <https://doi.org/10.1111/tpj.13487> PMID: 28107780.
38. Cui F, Brosche M, Sipari N, Tang S, Overmyer K. Regulation of ABA dependent wound induced spreading cell death by MYB108. *New Phytol*. 2013; 200(3):634–640. <https://doi.org/10.1111/nph.12456> PMID: 23952703.
39. Chen YS, Chao YC, Tseng TW, Huang CK, Lo PC, Lu CA. Two MYB-related transcription factors play opposite roles in sugar signaling in Arabidopsis. *Plant Mol Biol*. 2017; 93(3): 299–311. <https://doi.org/10.1007/s11103-016-0562-8> PMID: 27866313.
40. Tuan PA, Bai S, Yaegaki H, Tamura T, Hihara S, Moriguchi T, et al. The crucial role of PpMYB10.1 in anthocyanin accumulation in peach and relationships between its allelic type and skin color phenotype. *BMC Plant Biol*. 2015; 15: 280–293. <https://doi.org/10.1186/s12870-015-0664-5> PMID: 26582106
41. Medina-Puche L, Cumplido-Laso G, Amil-Ruiz F, Hoffmann T, Ring L, Rodriguez-Franco A, et al. MYB10 plays a major role in the regulation of flavonoid/phenylpropanoid metabolism during ripening of *Fragaria x ananassa* fruits. *J Exp Bot*. 2014; 65(2): 401–417. <https://doi.org/10.1093/jxb/ert377> PMID: 24277278.
42. Zheng X, Pan C, Diao Y, You Y, Yang C, Hu Z. Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (*Araceae*). *BMC Genomics*. 2013; 14: 490–500. <https://doi.org/10.1186/1471-2164-14-490> PMID: 23870214
43. Ranade SS, Lin YC, Zuccolo A, Van dP Y, García-Gil MR. Comparative in silico analysis of EST-SSRs in angiosperm and gymnosperm tree genera. *BMC Plant Biology*. 2014; 14(1):1–10.
44. Tong C, Zhang C, Zhang R, Zhao K. Transcriptome profiling analysis of naked carp (*Gymnocypris przewalskii*) provides insights into the immune-related genes in highland fish. *Fish Shellfish Immunol*. 2015; 46(2): 366–377. <https://doi.org/10.1016/j.fsi.2015.06.025> PMID: 26117731
45. Jain N, Patil GB, Bhargava P, Nadgouda RS. In silico mining of EST-SSRs in *Jatropha curcas* L. towards assessing genetic polymorphism and marker development for selection of high oil yielding clones. *American Journal of Plant Sciences*. 2014; 5(11):1521–1541.
46. Sathyanarayana N, Pittala RK, Tripathi PK, Chopra R, Singh HR, Belamkar V, et al. Transcriptomic resources for the medicinal legume *Mucuna pruriens*: *de novo* transcriptome assembly, annotation, identification and validation of EST-SSR markers. *BMC genomics*. 2017; 18(1):409. <https://doi.org/10.1186/s12864-017-3780-9> PMID: 28545396