Comparative analysis of the root and leaf transcriptomes in *Chelidonium majus* L.

Helen Pourmazaheri¹,²☯, Aboozar Soorni³☯, Bahram Baghban Kohnerouz¹*, Nafiseh Khosravi Dehaghi²*, Enayatollah Kalantar⁴, Mansoor Omidi⁵, Mohammad Reza Naghavi¹*  

¹ Department of Plant Breeding and Biotechnology, College of Agriculture, University of Tabriz, Tabriz, Islamic Republic of Iran, ² Department of Pharmacognosy, Faculty of Pharmacy, Alborz University of Medical Sciences, Karaj, Islamic Republic of Iran, ³ Department of Biotechnology, College of Agriculture, Isfahan University of Technology, Isfahan, Iran, ⁴ Department of Microbiology and Immunology, Faculty of Medicine, Alborz University of Medical Science, Karaj, Islamic Republic of Iran, ⁵ Agronomy and Plant Breeding Department, Agricultural & Natural Resources College, University of Tehran, Karaj, Islamic Republic of Iran  

☯ These authors contributed equally to this work.  
* bahramrouz@yahoo.com (BBK); khosravi@abzums.ac.ir (NKD); mnaghavi@ut.ac.ir (MRN)

Abstract

*Chelidonium majus* is a traditional medicinal plant, which commonly known as a rich resource for the major benzylisoquinoline alkaloids (BIAs), including morphine, sanguinarine, and berberine. To understand the biosynthesis of *C. majus* BIAs, we performed de novo transcriptome sequencing of its leaf and root tissues using Illumina technology. Following comprehensive evaluation of de novo transcriptome assemblies produced with five programs including Trinity, Bridger, BinPacker, IDBA-tran, and Velvet/Oases using a series of k-mer sizes (from 25 to 91), BinPacker was found to produce the best assembly using a k-mer of 25. This study reports the results of differential gene expression (DGE), functional annotation, gene ontology (GO) analysis, classification of transcription factor (TF) families, and SSR and miRNA discovery. Our DGE analysis identified 6,028 transcripts that were up-regulated in the leaf, and 4,722 transcripts that were up-regulated in the root. Further investigations showed that most of the genes involved in the BIA biosynthetic pathway are significantly expressed in the root compared to the leaf. GO analysis showed that the predominant GO domain is “cellular component”, while TF analysis found bHLH to be the most highly represented TF family. Our study further identified 10 SSRs, out of a total of 39,841, that showed linkage to five unigenes encoding enzymes in the BIA pathway, and 10 conserved miRNAs that were previously not detected in this plant. The comprehensive transcriptome information presented herein provides a foundation for further explorations on study of the molecular mechanisms of BIA synthesis in *C. majus*.

Introduction

*Chelidonium majus* L. is an herbaceous medicinal plant belonging to the botanical family Papaveraceae. *C. majus* is widely distributed in Europe and Western Asia and also as an
introduced species in Northern America. The species is commonly known as celandine, greater celandine, celandine poppy, elon-wort, felon-wort, rock poppy, swallow-wort, and tetter-wort. *C. majus* is highly toxic due to the presence of various secondary metabolites in the roots and stems, but is used in both traditional and modern medicines [1].

Pharmacological properties ascribed to *C. majus* include anti-viral [2], anti-bacterial [3], anti-fungal [4], anti-protozoal and radioprotective [5], anti-inflammatory [6], anti-alzheimer [7], anti-cancer [8], hepatoprotective [9], and natriuretic and antidiuretic effects [10]. The diverse array of secondary metabolites present in *C. majus* is responsible for its therapeutic properties. Alkaloids are the most common group of secondary metabolites present in *C. majus*. Chelidonine, berberine, sanguinarine, coptisine, chelerythrine, and protopine, are among the various alkaloids synthesized by *C. majus* [11]. Flavonoids, saponins, vitamins (e.g. vitamin A and C), mineral elements, sterols, and acids and their derivatives [12] are other secondary metabolites present in *C. majus* as well.

Transcriptome sequencing can be effectively utilized to identify and characterize pathways associated with the biosynthesis of secondary metabolites in plants [13–15], and enables the exploration of gene sequence and expression levels in an organism that lacks genomic resources [16–17]. A *de novo* transcriptome assembly, coupled with a liquid chromatography-electrospray ionization-tandem mass spectrometry (LC–ESI-MS/MS) proteomic approach, has been previously performed for *C. majus* to examine its protein composition, which showed novel defense-related proteins characteristic of its latex [18]. Also, Hagel et al. (2015) established an essential resource for the elucidation of benzylisoquinoline alkaloids (BIA) metabolism from the transcriptomes of 20 BIA-accumulating plants, but the structural diversity of the alkaloids and their biosynthetic pathways are not well studied in *C. majus*.

Considering the benefits of RNA sequencing technology, we used the root and leaf tissues as the basic materials to generate RNA-seq reads using Illumina HiSeq 2000 to obtain a better understanding about genes involved in the BIA biosynthesis pathway. We also mined the assembly to identify expressed sequence tag simple sequence repeats (EST-SSRs) and miRNAs that have not yet been characterized in *C. majus*.

**Materials and methods**

**Plant materials, RNA extraction, and nucleotide sequencing**

The *Chelidonium majus* tissues used in this study were collected from a high producer of chelidonine (Voucher number: IBRCP1006619), Mahmudabad-Amol, Mazandaran, Iran (Longitude coordinates: 52 17' 0.9", Latitude coordinates: 36 35' 15.1") [19]. To collect samples, plants were grown in the greenhouse facilities (28°C day/20°C night under natural light conditions) of the Iranian Biological Resources Center (IBRC) in Alborz, Iran. The root and leaf tissues were harvested from the plants, washed thoroughly with sterile water, frozen in liquid nitrogen and immediately stored at -80°C. Total RNA from the harvested plant materials was extracted using TRIzol® Reagent according to the manufacturer’s instructions (Invitrogen, USA). RNA samples were sent to the Beijing Genomic Institute (BGI) for transcriptome sequencing. Libraries were constructed using illumina TruSeq RNA sample preparation kit, while sequencing was performed with the Illumina HiSeq 2000 platform to generate paired-end (2×150 base) reads.

**De Novo transcriptome assembly**

We obtained a draft transcriptome from the raw RNA sequencing data using five popular assembly programs including (1) Trinity v. 2.4.0 [16], (2) Velvet v.1.2.10 and Oases v.0.2.09 [20–21], (3) IDBA-tran [22], (4) Bridger [23], and (5) BinPacker [24]. Trinity was used with a
fixed k-mer size of 25 as suggested by the authors. Oases-Velvet and IDBA-tran were used with a series of k-mer sizes from 25 to 91 and with an increment of 2. For BinPacker and Brigger we used two k-mer sizes (25 and 27). All tools were run with default settings and only assembled transcripts longer than 200 bp were retained. Subsequently, the most basic metrics for transcriptome assemblies including contig number, length distribution, assembly size, percentage of reads that could be mapped back to the transcriptome assembly (RMBT), and N50 were assessed and compared for all assemblies.

**Gene expression levels and transcript annotation**

The RNA-seq by Expectation Maximization (RSEM) package was used to estimate gene expression levels based on the mapping of RNA-seq reads to the assembled transcriptome [25]. To estimate the individual transcript abundances, the RNA-seq reads had first to be aligned to the transcriptome assembly. After indexing the reference transcriptome, separately, fastq files from the individual libraries of each sample were mapped to the final transcript set using script align_and_estimate_abundance.pl. The program Bowtie was used to generate alignments for each sample. Combining the read counts from all samples into a matrix was performed using script abundance_estimates_to_matrix.pl [16, 25]. Finally, identification of differentially expressed genes was carried out using run_DE_analysis.pl, which involves the Bioconductor package EdgeR in the R statistical environment [26–27]. Transcripts with very low read counts were filtered out across all libraries. Gene expression values were measured in FPKM (fragments per kilobase of transcript per million reads mapped) [26,28] and were used to make pairwise comparisons. Clustering analysis was performed on the differentially expressed genes, with FDR and the logFC cutoff defined by the–P 1e-3 -C 2 parameters using analyze_diff_expr.pl script.

Functional annotation of the *de novo* transcriptome was conducted using TransDecoder v2.0.1 to predict open reading frames (ORFs) at least 100 amino acids long, and the Trinotate pipeline v3.0.2 (http://trinotate.github.io/) was used to annotate the predicted ORFs using the following programs: BLASTX v2.2.29 and BLASTP v2.2.29 to search against Swissprot-UniProt database [29], Hmmer v.3.1b2 to identify protein domains (PFAM) [30–31], SignalP v.4.1 to predict the presence of signal peptides [32], Tmhmm v.2.0c for prediction of transmembrane helices in proteins [33], and Rnammer v.1.2 to predict ribosomal RNA [30]. All results from the bioinformatics analyses performed above were imported into a Trinotate SQLite database. To obtain Gene Ontology (GO) annotations, we used the Trinotate-integrated UniProtKB GO annotations and WEGO software [34] for GO functional classification.

**TF identification and EST-SSR analysis**

Homology searches against PlantTFDB using BLASTx with a cut-off *E*-value of 1e−5 were performed in order to identify transcription factors [35]. The assembled sequences were scanned to identify single sequence repeats (EST-SSRs) using the MicroSatellite Identification Tool (MISA, http://gpcr.inpk-gatersleben.de/misa/) [36]. For this purpose, a FASTA file containing all of the assembled sequences was used as the input file in MISA Perl script to screen for EST-SSRs with motifs of 1 to 6 nucleotides and a minimum repeat number of 10, 6, 5, 5, 5, and 5, respectively. PCR primers were designed using Primer3 [37]. The parameters for designing primers were as follows: PCR product size range of 100 to 300 bp; primer length of 18–25 nucleotides; annealing temperature between 55 and 62°C with 57°C as the optimum melting temperature.
In silico miRNA identification

To identify potential miRNAs in *C. majus*, transcripts and previously-known plant miRNAs from the miRBase database [38] were initially clustered using CD-HIT-EST [39] with the following parameters: c = 1, n = 10, d = 0, and M = 16000. The clustered sequences were then aligned against non-redundant miRNAs using BLASTn v 2.2.30 [29]. The obtained hits with alignment length \( > 20 \), e-value threshold \( \leq 0.001 \), and without mismatches and gaps were considered for extracting the precursor sequences (pre-miRNA). A sliding window of about 400 nt from the region 200 nt upstream of the beginning of the mature miRNA to 200 nt downstream of the miRNA from the filtered sequences was then used as a query in BLASTX searches against the NCBI non-redundant protein database to remove protein coding sequences. The secondary structures of the retained sequences were predicted using the web server mfold (Zuker, 2003). Only sequences with the following criteria were considered as potential miRNA precursors: (1) \( > 20 \) nt mature miRNA sequence within one arm of the hairpin (2) with higher negative minimal free energies and higher MFEIs [40], (3) no more than six mismatches with the opposite miRNA, and (4) no loop or break in miRNA sequences. In the last step, we used the web tool psRNA-target (http://bioinfo3.noble.org/psRNATarget/) to predict the potential miRNA targets.

Orthogroup identification

We used OrthoFinder [41] with the default parameters, aligned sequences with MAFFT v 7.271 [42] and built trees with FastTreeMP v 2.1.8 [43], to identify conserved orthogroups for eight species, including *Argemone mexicana*, *Papaver bracteatum*, *Eschscholzia californica*, *Glaucium flavum*, *Stylophorum diphyllum*, *Sanguinaria canadensis*, and *Corydalis cheilanthifolia* published previously along with *C. majus*. The corresponding transcriptome assemblies for seven species were downloaded from www.phytometasyn.ca [44]. The predicted protein sequences were obtained using TransDecoder v2.0.1 (http://transdecoder.sourceforge.net/). The rooted species tree was drawn using Dendroscope v 3.5.9 [45].

Results and discussion

Short-read sequencing and de novo transcriptome assembly

A total of 188.98 million clean PE RNA-seq reads of 150 bp in length with quality scores of \( > Q20 \) were obtained after sequencing root and leaf tissues on Illumina HiSeq 2000™ platform. Subsequently all 188.98 million of the high quality were used for de novo assembly using different packages.

The primary assembly statistics showed variable patterns of performance with the different tools; for example, Trinity produced the largest number of contigs with the highest number of bps, followed closely by Bridger (Table 1). The number of predicted transcripts is strongly affected by the k-mer size [46]. With Velvet/Oases, the number of predicted transcripts dropped from 325,276 with k-mer 25 to 94,116 with k-mer 91, similarly to previously reported results [46–48]. However, using IDBA-tran, the number of contigs generally increased with increasing k-mer size (168,305 contigs with k-mer 25, and 210,145 with k-mer 91). Some previous studies have indicated that N50, a metric commonly used in genome assembly, is not suitable for transcriptome assembly, because longer N50 values may indicate a high level of chimerism [6,49], although it has also been observed that larger N50s can reflect a higher quality assembly [50–51]. BinPacker gave the largest N50 compared to Trinity and Bridger. With increased k-mer size, the N50 increased for all Velvet/Oases and IDBA-tran k-mer assemblies. The total assembly length showed a similar trend N50 for IDBA-tran, while for Velvet/Oases
with increasing k-mer size, the number of bps increased up to k = 45, at which point the number of bps declined. Across all assembly strategies performed using the different programs, Trinity, Bridger, and BinPacker consistently produced similar percentages of paired-end reads that mapped back to the relative assembly, ranging from 90.22 to 93.78%. Assemblies produced by Velvet/Oases had the lowest percentage of mapped reads (>70%). BinPacker was faster, compared to Trinity and Bridger. These conflicting patterns show that the outputs of the assembly programs can be quite variable.

Based on the assembly statistics, the assembly generated by BinPacker with k-mer 25, which had the highest N50 value (1,585 bp), average transcript length, and RMBT percentage, whilst keeping fewer number of contigs (232,701) and larger total assembly size (216.24 Mbp) as long as possible was selected for downstream analysis.

### Identification of differentially expressed genes (DEGs)

The identification of DEGs was performed by estimating individual transcript abundance by mapping the cleaned reads back to the assembled transcripts with RSEM, and their expression levels were represented as FPKM values. More than 93% of trimmed reads in the four libraries could be mapped to the transcriptome assembly successfully, which indicates the quality of the de novo transcriptome assembly. Digital abundance analysis identified 10,750 unique transcripts as being significantly different between leaves and roots with two biological replicates where the criteria for FDR was set to 0.001 and fold-change was set to 2(2) or 4-fold; 6,028 transcripts were up-regulated in leaf and 4,722 transcripts in root. The fold-change ranged from 2 to 14.

In order to identify the active pathways represented in the leaf and root transcriptome of *C. majus*, the DEG sequences were used as queries in searches against the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database. A total of 3,379 transcripts (31.43%), 1,828 transcripts being up-regulated in the leaf and 1,507 transcripts up-regulated in the root, were assigned to 354 pathways. These canonical pathways were classified into six categories (Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, and Human Diseases) and 43 sub-categories. Among the pathways, “metabolic pathways” (with 347 transcripts), “biosynthesis of secondary metabolites” (192 transcripts), and “ribosome” (101 transcripts) were the most abundant. Since *C. majus* produces a major group of secondary metabolites (especially alkaloids), it was necessary to identify the most active genes involved in the metabolic pathways.

### Genes related to alkaloid biosynthesis pathways

The principal pathway for metabolism of morphinans (codeine and morphine), protoberberines (berberine) and benzophenanthridines (sanguinarine) starts with the formation of (S)-

| Tools          | Trinity | Bridger | BinPacker |
|---------------|---------|---------|-----------|
| Kitmer Size   | 25      | 25      | 27        |
| Number of contigs | 392,555 | 336,021 | 334,664   |
| Total size (Mb)  | 220.67  | 207.94  | 204.34    |
| Maximum length (bp) | 16,238  | 37,322  | 25,856    |
| Minimum length (bp) | 224     | 201     | 201       |
| Average length (bp) | 562.15  | 618.84  | 610.58    |
| N50 length (bp)   | 674     | 1,013   | 993       |

https://doi.org/10.1371/journal.pone.0215165.t001
reticuline. Berberine and sanguinarine are found simultaneously in only a few species [52]. The biosynthesis of (S)-reticuline begins with the conversion of tyrosine to dopamine and 4-hydroxyphenylacetaldehyde (4HPAA) [53–54]. Tyrosine/dopa decarboxylase (TYDC), which yields tyramine or dopamine, has been isolated from a range of plant species [55–56]. TYDC gene was found to be strongly expressed in root (Fig 1), displaying different expression patterns in different organs. Previous studies have shown that TYDCs are regulated by multiple factors and are differentially expressed in response to elicitor treatments [57–59].

Norcoclaunine synthase (NCS) yields (S)-norcoclaurine through the condensation of 4-HPAA and dopamine. The NCS gene sequence was initially isolated from meadow rue (Thalictrum flavum) [54] and then from opium poppy [60]. In our study, the expression of NCS was significantly higher in root than in leaves, which is consistent with the results of previous studies, although NCS-specific mRNA has been detected in flower buds and germinating seeds [61–62]. Coclaurine N-methyltransferase (CNMT), which is expressed in roots, stems, flower buds, and at lower levels in leaves [63] is an N-methyltransferase which converts (S)-Coclaurine to (S)-N-methylcoclaurine [64]. In this study, gene-specific transcripts of CNMT were detected in both tissues. (S)-N-methylcoclaurine 3′-hydroxylase (CYP80B1) is a P450 hydroxylase [65]. Three transcripts related to CYP80B1 showed high levels of expression in rootManu-
script, which is consistent with previously published results [66]. We detected three, two, and four genes in the sanguinarine, berberine, and morphine pathways, respectively. In the sanguinarine pathway, the gene for tetrahydroprotoberberine cis-N-methyltrans-
ferase (TNMT), which converts (S)-stylopine to (S)-cis-N-methylstylopine, showed higher expression levels in leave as compared to root, but the methylstylopine hydroxylase (MSH) and protopine 6-hydroxylase (P6H) genes had higher expression levels in root. MSH and P6H both belong to the P450 enzyme family [67]. Most previous studies showed that TNMT, MSH, and P6H are highly expressed in root, with the lowest expression levels detected in leave, fruits, or bulb initi-
ation [68]. However, Liscombe and Facchini (2007) measured the highest levels of TNMT activity in the stem and leaf tissues of opium poppy, with lower levels in roots and flower buds, which is consistent with our results [69]. In the berberine pathway, genes for (S)-scoulerine-9-O-methyltransferase (SMT) [70] and (S)-canadine synthase (CYP719A1) [71] were both up-regulated in root, but of four genes detected in the morphine pathway, three, including those for salutaridine reductase (SalR), salutaridinol 7-O-acetyltransferase (SalAT), and codeine O-demethylase (CODM), were up-regulated in root while transcription of the codeinone reduc-
tase (COR) gene was up-regulated in leaf. COR appears twice in the codeine and morphine pathway; (1) it catalyzes the NADPH-dependent reduction of codeinone to codeine [72], and (2) it is involved in the conversion of morphinone to morphine [73].

Functional annotation and GO classification

Gene annotation is one of the most important parts of transcriptome analysis, because it enables us to interpret the content of transcriptome assembly. A total of 97,275 (41.8%) and 196,640 sequences (84.5%) gave significant hits against the Swiss-Prot database using BLASTx and BLASTp searches, respectively. Furthermore, 14,894 unique Pfam protein motifs were assigned and 8,805 transcripts were predicted to encode proteins with signal peptides. Of the transcripts that returned BLASTx hits, 110,757 were associated with a total of 853,310 Gene Ontology (GO) terms. Of these annotated transcripts, 5,609 had only a single GO term. Fig 2 summarizes the percentage of genes belonging to the top 10 categories in the "biological process", “cellular component”, and “molecular function” GO domains. Among the three main domains, “cellular component” was the most highly represented, and within this category most of the genes belonged to the “cell” class, followed by the “cell part” and “intracellular”
Fig 1. Schematic representation of the benzylisoquinoline alkaloid biosynthesis pathway and expression of the genes for pathway enzymes in *Chelidonium majus*.

https://doi.org/10.1371/journal.pone.0215165.g001
classes. In the case of the “biological process” domain, the most abundant categories were “cellular process” and “metabolic process”, and for the “molecular function” domain, the predominant categories were “binding” and “catalytic activity”. The GO term abundance results are similar to those from a large number of transcriptome studies that have been reported for other non-model and medicinal plants, such as saffron [74], gardenia [75], safflower [76], and chrysanthemum [77]; however, compared to a previous study on *C. majus* [18], the distribution of genes in the three main ontologies was different. The most noticeable difference was observed in the distribution of genes in “molecular function” GO domain. Possible reasons for the discrepancies between our study and that of Nawrot et al. (2016) could include variations in the structure of the cDNA libraries and/or the number of sequences used to retrieve GO terms [18].

Fig 2. GO classification of genes expressed in *Chelidonium majus*. The bar chart shows the percentage of genes (Y-axis) belonging to the top 10 categories (X-axis) in the “cellular component”, “biological process”, and “molecular function” GO domains.
https://doi.org/10.1371/journal.pone.0215165.g002
Identification and analysis of transcription factor genes

Transcription factors (TFs) play multiple key roles in plants by controlling the synthesis of bioactive components, especially secondary metabolism and regulation of gene expression through DNA-binding and cis-acting elements [78–79]. Here, a total of 69,971 putative TF encoding transcripts were identified and further classified into 64 different families in the *C. majus* transcriptome. Among the transcription factor families, bHLH was the most highly represented, with 7,736 transcripts (11.06%), followed by NAC (4,992; 7.13%), MYB-related (4,545; 6.50%), ERF (4,003; 5.72%), C2H2 (3,300; 4.72%), and WRKY (3,077; 4.40%) (Fig 3).

Previous studies have demonstrated that the bHLH TFs could play major roles not only in the developmental processes including control of cell proliferation [80] and formation of trichome and light signal transduction [81], but also in the regulation of the expression of many genes which participate in the biosynthesis of plant secondary metabolites such as flavonoids and alkaloids [82]. In addition to bHLH, other TF families such as WRKY, MYB, and C2H2 are involved in secondary metabolism pathways. Two transcription factors, CjWRKY1, a WRKY-type TF [83] and CjbHLH1, a basic helix-loop-helix TF [84] have been identified in the alkaloid pathway to independently regulate berberine biosynthesis. CjWRKY1 is the first transcription factor which has been characterized to play a positive role in berberine synthesis in *Coptis japonica* [83]. CjbHLH1 is a non-MYC2-type bHLH TF, and two homologs, EcbHLH1-1 and EcbHLH1-2, that are associated with the regulation of sanguinarine synthesis, have been identified in the California poppy, *Eschscholzia californica* [85]. The ERF subfamily, which belongs to the AP2/ERF family, have only a single AP2/ERF domain, and are known to be involved in dehydration or ethylene responses [86]. ERF189 and ERF221/ORC1 in *N. tabacum* and ORCA2 and ORCA3 in *C. roseus* are members of the AP2/ERF TF family that have been identified as being involved in alkaloid biosynthesis [87]. MYB transcription factors control diverse biological processes such as the regulation of primary/secondary metabolism and hormone syntheses [88–90], whereas NAC family members participate in regulating plant growth and developmental processes [91–93].

EST-SSR frequency and distribution

EST-SSRs have been extensively used in the study of genetic variation, evolutionary relationships, linkage mapping, and genotyping due to their abundance, high polymorphic information content, good reproducibility, and relative ease of use. At present, there are no studies addressing the genetic diversity and classification of *C. majus* germplasm resources based on EST-SSR markers, because have not been identified so far. In this study, for the first time, large-scale transcriptome sequencing was used to identify expressed sequence tag simple sequence repeats (EST-SSR) markers. To develop new markers for *C. majus*, all of the 232,701 transcripts generated by BinPacker were screened to find potential microsatellite motifs using the MISA search tool. Due to both sequencing and assembly errors, mononucleotide repeats may not be reliable, so we excluded them from further analyses. A total of 39,841 EST-SSRs (2–6 nt) were identified in 45,277 (19.45%) transcripts (Fig 4), and 15,293 sequences were found to contain more than one EST-SSR motif. The dinucleotide repeat motifs were the most abundant (21,887 or 54.94%), followed by trinucleotide repeats (17,180 or 43.12%), and only 540 (1.38%), 64 (0.16%), and 160 (0.40%) of the identified EST-SSRs harbored predominately tetra-, penta-, and hexanucleotide repeat motifs respectively. Within EST-SSR data sets, dinucleotide repeat frequencies are usually higher than trinucleotide repeat frequencies. This is supported by studies on medicinal plants such as *Andrographis paniculata* [94], *Ginkgo biloba* L. [95], *Gleditsia sinensis* [96], *Crocus sativus* [74], *Boea clarkeana* [97], *Phyllanthus amarus* [98], and *Cinnamomum longepaniculatum* [99]. However, the tri-nucleotide repeats are more
frequent than di-nucleotide repeats in some other medicinal plants such as *Mucuna pruriens* [100] and *Epimedium sagittatum* [101]. These distribution frequencies vary with respect to the different plant species, the employed datasets, and the tools and standards used for EST-SSR searches and identification.

Among the dinucleotide repeat motifs, we found that AG/CT was the most common (48.05%) in *C. majus*, and this is the case for plants in general [102]. The presence of CT repeat sequences in 5'UTRs is probably related to reverse transcription and has a significant role in gene regulation. Of the trinucleotide repeats, AAG/CTT was the most frequent motif (12.89%)
in *C. majus*, followed by ACC/GGT (6.94%) (Fig 5). The (AAG/CTT)n repeats and their complements are the most common tri-nucleotide repeat motifs in plants [103]. We succeeded in identifying several novel EST-SSRs which were linked to unigenes that putatively encode enzymes involved in morphine and sanguinarine biosynthesis. Finally we designed high-quality primers to amplify these potential EST-SSR loci (Table 2). Our findings will enrich the molecular marker resources and help spearhead molecular genetic research on *C. majus*.

**Discovery of miRNAs**

A high stringency filtering approach on BLAST results identified a total of 104 potential miRNAs belonging to 108 sequences that were retained for secondary structure analysis. After filtering based on secondary structure, nine folded miRNA precursors were predicted from nine different families for the first time in *C. majus* (Table 3). In this study, the identified precursors...
had high MFEI values (0.71–0.83) with an average of 0.76, which is higher than that of rRNAs (0.59), tRNAs (0.64) or mRNAs (0.62–0.66) [40].

Most mature miRNAs are evolutionarily conserved between species within the plant kingdom, some of which have a large number of potential targets. Of these, miR319 regulates transcription factors belonging to the TCP family which regulate plant developmental processes such as leaf morphogenesis in Arabidopsis [104]. miR396 is necessary for normal development in Arabidopsis, and regulates the Growth-Regulating Factor (GRF) family of transcription factors. GRFs are known to control cell proliferation in Arabidopsis leaves [105]. miR159 has a

---

**Fig 5.** Expressed sequence tag simple sequence repeats (SSRs) identified in the Chelidonium majus transcriptome. The y-axis indicates frequencies of the 10 most abundant SSRs motifs. The x-axis indicates 10 groups of SSRs motifs.

https://doi.org/10.1371/journal.pone.0215165.g005

---
very similar sequence to miR319 but regulates different genes [106]. miR828 appears to target transcription factor genes for DNA binding domain-containing proteins such as CONSTANS-like 5 related cluster protein and zinc finger protein-B box [107]. Most studies have shown that the miR171 family negatively regulates (decreases) primary root elongation and shoot branching by targeting GRAS gene family members [108]. Auxin Response Factors (ARFs), proteins that play important roles in plant growth and development, have been reported to be targets of the miR167 family in *Oryza sativa* [109]. miR169 is mostly expressed in the roots and regulates CCAAT motif-binding transcription factors [107].

### Construction of orthogroups across multiple species of Papaveracea

To facilitate comparative studies and to demonstrate the utility of transcriptome assemblies for phylogenetic analysis, candidate coding regions generated by TransDecoder from transcriptome assemblies of seven species were compared with potential proteins based on ORF

### Table 2. Identification of SSR motifs in putative morphine and sanguinarine biosynthesis genes.

| ID                  | SSR        | Forward Primer (5′-3′)                      | Tm (°C) | size | Reverse Primer (5′-3′)                      | Tm (°C) | size | PS (bp)** |
|---------------------|------------|---------------------------------------------|---------|------|---------------------------------------------|---------|------|-----------|
| S-norcoclaurine(NCS)| (AAG) 5    | AAGAAACTCGAGCAAGAGAGA                      | 60.134  | 20   | TCTTGTAGGCTGACAGTTCTCTCC                  | 59.936  | 23   | 157       |
| S-norcoclaurine(NCS)| (TA)8      | CTCGAGGGCGGCTCAAGAT                        | 59.962  | 20   | CGTGGTAGTAGAGATACCAATTAA                 | 57.965  | 25   | 214       |
| Salutaridinol 7-O-acetyltransferase(SalAT)| (TCA) 5    | CATCAGTGCCTGGTGTGTC                      | 60.005  | 20   | TGAGAATGGGGTGGTAA                     | 60.162  | 20   | 189       |
| Salutaridinol 7-O-acetyltransferase(SalAT)| (ACA) 5    | GCAGTTGCGCTTGAATATGA                      | 59.983  | 20   | GAAGACGAGCAATGACCA                     | 59.637  | 20   | 227       |
| (S)-tetrahydroprotoiberine(TNMT)| (CT)8      | CTTCCTCCCATACCCACACTA                     | 59.92   | 20   | CTCCTCCCATAGTACCTAA                  | 60.029  | 20   | 256       |
| (S)-tetrahydroprotoiberine(TNMT)| (GAC) 5    | CGACGGAGAGTGGGTTGTT                       | 60.073  | 20   | CCAGACGTTGAGTACCTG                    | 60.028  | 20   | 238       |
| (S)-tetrahydroprotoiberine(TNMT)| (GGT) 5    | CCGGACTACAAGCGCTGAT                      | 59.989  | 20   | GGTTTTCTCCTGCGGAGTA                | 60.192  | 20   | 118       |
| (S)-tetrahydroprotoiberine(TNMT)| (AT)7      | CACAATTAGGCACACTAA                      | 59.395  | 20   | GCCTTGATAGATATCTGA               | 59.799  | 20   | 203       |
| Methyltetrahydroprotoiberine 14-monooxygenase (MSH) | (AT)6 | ACACAAACACAAAGCGACAAAGC | 60.154  | 20   | GGAGTGGAAGCGGACAT                   | 59.973  | 20   | 219       |
| NADPH-dependent codeinone reductase (COR) | (AG)8 | TTTCCATTAGGCAACATCC | 59.762  | 20   | TTGCGATCTCCCTACCTGAG                | 60.21   | 20   | 226       |

** PS: PRODUCT1 size.

https://doi.org/10.1371/journal.pone.0215165.t002

### Table 3. High-probability miRNAs proposed for *Chelidonium majus*.

| miRNA family | miRNA ID  | Mature sequence     | Mismatch | ΔG = -kcal/mol | MFEI  |
|--------------|-----------|---------------------|----------|----------------|-------|
| mir477       | mes-miR477h | ACUCUCCUCAAGGCCUCAG  | 4        | 82.6           | 0.77  |
| mir319       | ath-miR319a | UUGGACAGUAGGGAGUCUCU | 2        | 116.2          | 0.83  |
| mir396       | cca-miR396c | UUCAAGAAGCCUGGGGAAAA | 1        | 117.3          | 0.77  |
| mir159       | pde-miR159 | UUGUGUGAAGGGAGCUCCA  | 4        | 116.5          | 0.74  |
| mir828       | vvi-miR828a | UCUUGUCUCAUUAGAGIUAUCCA | 3       | 102            | 0.71  |
| mir171       | ctr-miR171 | UUGACGGCGCCGCAAUAUCC  | 1        | 125.80         | 0.73  |
| mir167       | ath-miR167d | UGAAGGAGGAAAGAUGAGG   | 2        | 105.60         | 0.71  |
| mir4376      | sly-miR4376 | ACCAGGAGAGAGAUGAGCUCCA | 5      | 109.6          | 0.81  |
| mir169       | gma-miR169d | UGACCAAGAGAUGAGUGCCGU  | 4        | 98.1           | 0.77  |

MFEI: Minimal folding free energy index.

https://doi.org/10.1371/journal.pone.0215165.t003
predictions in the C. majus transcriptome using OrthoFinder. The number of shared orthogroups between each pair of species ranged from 10,925 (between Eschscholzia californica and Papaver bracteatum) to 15,498 (between C. majus and Argemone mexicana). A total of 8,483 orthogroups were identified among all species present, and there were 59 single-gene orthogroups in our species comparison. The family Papaveraceae is divided into four subfamilies based on critical details of the morphological traits [110]. In this study, with the exception of Corydalis cheilanthifolia, which belongs to the Fumarioideae subfamily, all other species belong to the Papaveroideae subfamily. The species tree strongly supports genetic relationship between C. majus and S. diphyllum (Fig 6). This tree suggests that C. majus and S. diphyllum are the most divergent from C. cheilanthifolia in the Fumarioideae subfamily.

Conclusions

In the current study, we generated and characterized a fully annotated and deep-sequencing transcriptome assembly for leaves and root tissues of C. majus. This represents an important initial resource that will enable further studies on the molecular mechanisms of bioactive alkaloids biosynthesis, as well as for studies of the molecular genetics and functional genomics of this important medicinal plant. Based on transcriptome assembly metrics, BinPacker was found to be the best among all the assemblers used in this study. Generally, our analysis revealed that most of the genes involved in the sanguinarine, berberine, and morphine pathways are broadly expressed in root. We observed that relatively few of these genes are up-regulated in leaves. Our results also showed that the most frequent transcription factor families
represented here are involved in regulating secondary metabolism pathways, especially those for alkaloid biosynthesis. Development of a large number of EST-SSR markers and the design of high-quality PCR primers for potential EST-SSR loci amplification in the *C. majus* transcriptome will be useful for evaluating genetic diversity and also in marker-assisted breeding in *C. majus*. Furthermore, our computational methods enabled the identification of a set of potential miRNAs which were previously unknown for this plant.

**Acknowledgments**

We thank David Zaitlin (Kentucky Tobacco Research and Development Center, University of Kentucky, Lexington.) for critical reading of the manuscript and language editing.

**Author Contributions**

Data curation: Aboozar Soorni.

Formal analysis: Aboozar Soorni.

Investigation: Helen Pourmazaheri.

Methodology: Aboozar Soorni.

Project administration: Bahram Baghban Kohnerouz, Nafiseh Khosravi Dehaghi, Mohammad Reza Naghavi.

Resources: Helen Pourmazaheri.

Software: Aboozar Soorni.

Visualization: Aboozar Soorni.

Writing – original draft: Aboozar Soorni.

Writing – review & editing: Bahram Baghban Kohnerouz, Nafiseh Khosravi Dehaghi, Enayatollah Kalantar, Mansoor Omidi, Mohammad Reza Naghavi.

**References**

1. Monavari S. H, Shahrabad M. S, Keyvani H, Bokharaei-Salim F. Evaluation of In Vitro Antiviral Activity of *Chelidonium majus* L. against Herpes Simplex Virus Type1. Afr J Microbiol Res. 2012; 6:4360–4364.

2. Gerencer M, Turecek P.L, Kistner O, Mitterer A, Savidis-Dacho H, Barrett N.P. In Vitro and In Vivo Antiretroviral Activity of the Substance Purified from the Aqueous Extract of *Chelidonium majus* L. Antiviral Res. 2006; 72:153–156. [PMID: 16647765]

3. Miao F, Yang X.J, Zhou L, Hu H.J, Zheng F, Sun X.D, et al. Structural Modification of Sanguinarine and Chelerythrine and Their Antibacterial Activity. Nat Prod Res. 2011; 25:863–875. [PMID: 21491327]

4. Hou Z, Yang R, Zhang C, Zhu L.F, Miao F, Yang X.J, et al. 2(Substituted Phenyl)-3,4-Dihydroisouquinolin2-iums as Novel Antifungal Lead Compounds: Biological Evaluation and Structure-Activity Relationships. Mol. 2013; 18:10413–10424.

5. Kim D.S, Kim S.J, Kim M.C, Jeon Y.D, Um J.Y, Hong S.H. The Therapeutic Effect of Chelidonic Acid on Ulcerative Colitis. Biol Pharm Bull. 2012; 35:666–671. [PMID: 22687399]

6. Park J.E, Cuong T.D, Hung T.M, Lee I, Na Min Kyun, Kim J.C, et al. Alkaloids from *Chelidonium majus* and Their Inhibitory Effects on LPS Induced NO Production in RAW264.7 Cells. Bioorg Med Chem Lett. 2011; 21:69606863.

7. Cahlikova L, Opletal L, Kurfurst M, Macakova K, Kulhankova A, Hostalkova A. Acetylcholinesterase and Butyrylcholinesterase Inhibitory Compounds from *Chelidonium majus* (Papaveraceae). Nat Prod Commun. 2010; 5:1751–1754. [PMID: 21213973]

8. Moussa S.Z, El-Meadawy S.A, Ahmed H.A, Refat M. Efficacy of *chelidonium majus* and Propolis against Cytotoxicity Induced by Chlorhexidine in Rats. J Biochem Mol Biol. 2007; 25:42–68.
9. Biswas S.J, Bhattacharjee N, KhudaBukhsh A.R. Efficacy of a Plant Extract (Chelidonium majus L.) in Combating Induced Hepatocarcinogenesis in Mice. Food Chem Toxicol. 2008; 46:1474–1487. https://doi.org/10.1016/j.fct.2007.12.009 PMID: 18215450

10. Koriem K.M, Arbid M.S, Asaad G.F. Chelidonium majus Leaves Methanol Extract and Its Chelidonine Alkaloid Ingredient Reduce Cadmium-Induced Nephrotoxicity in Rats. J Nat Med. 2013; 67:159–167. https://doi.org/10.1007/s11412-016-0667-6 PMID: 22484604

11. Gilsa M, Gaman L, Panait E, Stoian I, Atanasiu V. Chelidonium majus: an Integrative Review: Traditional Knowledge versus Modern Findings. Forsch Komplementmed. 2010; 17:241–248. https://doi.org/10.1159/000321397 PMID: 20980763

12. Kopytko Y.F, Dargaeva T.D, Sokolskaya T.A, Groditskaya E.I, Kopnin A.A. New Methods for the Quality Control of a Homeopathic Matrix Tincture of Greater Celandine. Pharm Chem J. 2005; 39:603609.

13. Rama Reddy NR, Mehta RH, Soni PH, Makasana J, Gajbhiye NA, Ponunchamy M, et al. Next Generation Sequencing and Transcriptome Analysis Predicts Biosynthetic Pathway of Sennosides from Senna (Cassia angustifolia Vahl.), a Non-Model Plant with Potent Laxative Properties. PLOS ONE. 2015; https://doi.org/10.1371/journal.pone.0129422 PMID: 26098898

14. Hagel JM, Morris JS, Lee EJ, Desgagné-Penix I, Bross CD, Chang L, et al. Transcriptome analysis of 20 taxonomically related benzylisoquinolinealkaloid-producing plants. BMC Plant Biol. 2015; 15:227. https://doi.org/10.1186/s12870-015-0596-0 PMID: 26384972

15. Xu ZC, Peters RJ, Weirather J, Luo HM, Liao BS, Zhang X, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. Plant J. 2015; 82:951–961. https://doi.org/10.1111/tpj.12865 PMID: 25912611

16. 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. Nat Biotechnol. 2011; 29: 644–652. https://doi.org/10.1038/nbt.1883 PMID: 21572440

17. Garg R, Jain M. Transcriptome Analyses in Legumes: A Resource for Functional Genomics. The Plant Genome. 2013; https://doi.org/10.3835/plantgenome2013.04.0011

18. Nawrot R, Barylski J, Lippmann R, Altschmied L, Mock HP. Combination of transcriptomic and proteomic approaches helps to unravel the protein composition of Chelidonium majus L. milky sap. Planta. 2016; 244:1055–1064. https://doi.org/10.1007/s00425-016-2566-7 PMID: 27401454

19. Pourmazaheri H, Baghbani Kohnerouz B, Khosravi Dehaghi N, Naghavi M.R, Kalantar E, Mohammadkhani E, et al. High-Content Analysis of Chelidonine and Berberine from Iranian Chelidonium majus L. Ecotypes in Different Ontogenetical Stages Using Various Methods of Extraction. J Agr Sci Tech. 2017; 19:1381–1391.

20. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. https://doi.org/10.1101/gr.074492.107 PMID: 18349386

21. Schulz M.H, Zerbino D.R, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. 2012; 28:1086–1092. https://doi.org/10.1093/bioinformatics/bts094 PMID: 22368243

22. Peng Y, Leung H.C.M, Yiu SM, Lv M.J, Zhu XG, Chin F.Y.L. IDBA-tran: a more robust de novo transcriptome assembly from RNA-seq data. PLOS Comput Biol. 2016; 12:e1004772. https://doi.org/10.1371/journal.pcbi.1004772 PMID: 26894997

23. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040

24. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data Genome Biol. 2010; 11:R25. https://doi.org/10.1186/gb-2010-11-3-r25 PMID: 20196867

25. R Development Core Team. R: A Language and Environment for Statistical Computing. In: The R Foundation for Statistical Computing. Vienna, Austria. 2011. http://www.R-project.org.

26. Trapnell C, Williams B.A, Pertea G, Mortazavi A, Kwan G, van Baren M.J, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. https://doi.org/10.1038/nbt.1621 PMID: 20436464
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990; 215:403–410.

30. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011; 39:W28–W37. https://doi.org/10.1093/nar/gkr367 PMID: 21593126

31. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families database. Nucleic Acids Res. 2012; 40:D290–D301. https://doi.org/10.1093/nar/gkr1065 PMID: 22127870

32. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 2011; 8:785–786. https://doi.org/10.1038/nmeth.1701 PMID: 21959131

33. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol. 2001; 305:567–580. https://doi.org/10.1006/jmbi.2002.1031 PMID: 12589540

34. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, et al. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 2006; 34:W293–7. https://doi.org/10.1093/nar/gkl031 PMID: 16845012

35. Jin J, Zhang H, Kong L, Gao G, Luo J. Plant TFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Res. 2014; 42:D1182–7. https://doi.org/10.1093/nar/gkt1016 PMID: 24174544

36. Thiel T, Michalew E, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003; 106:411–22. https://doi.org/10.1007/s00122-002-1031-0 PMID: 12589540

37. Rozen S, Skaletsky H. Primer3 on the WWW for General Users and for Biologist Programmers. In: Misener S, Krawetz S.A, editors. Bioinformatics Methods and Protocols. Berlin, Germany: Springer; 1999. p. 365–386.

38. Griffiths-Jones S, Grocock R J, van Dongen S, Bateman A, Enright A. MiRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006; 34:D140–D144. https://doi.org/10.1093/nar/gkj112 PMID: 16381832

39. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; 22:1658–1659. https://doi.org/10.1093/bioinformatics/btt158 PMID: 16731699

40. Zhang B.H, Pan X.P, Cox S.B, Cobb G.P, Anderson T.A. Evidence that miRNAs are different from other RNAs. Cell Mol Life Sci. 2006; 63:246–254.

41. Emms D.M, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015; 16:157. https://doi.org/10.1186/s13059-015-0721-2 PMID: 26243257

42. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30:3059–3066. PMID: 12136088

43. Price M, Dehal P, Arkin A. Fast Tree2: approximately maximum-likelihood trees for large alignments. PLOS ONE. 2010; https://doi.org/10.1371/journal.pone.0009490 PMID: 20224823

44. Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJ, Chakrabarty R, et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J Biotechnol. 2013; 166:122–34. https://doi.org/10.1016/j.jbiotec.2013.04.004 PMID: 23602801

45. Huson DH, Scornavacca C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Syst Biol. 2012; 61:1061–1067. https://doi.org/10.1093/sysbio/sys062 PMID: 22780991

46. Wang S, Gribskov M. Comprehensive evaluation of de novo transcriptome assembly programs and their effects on differential gene expression analysis. Bioinformatics. 2016; 33:327–333.

47. Haznedaroglu BZ, Reeves D, Rismani-Yazdi H, Peccia J. Optimization of de novo transcriptome assembly from high-throughput short read sequencing data improves functional annotation for non-model organisms. BMC Bioinformatics. 2012; 13:170. https://doi.org/10.1186/1471-2105-13-170 PMID: 22808927

48. Moreton J, Dunham SP, Emes RD. A consensus approach to vertebrate de novo transcriptome assembly from RNA-seq data: assembly of the duck (Anas platyrhynchos) transcriptome. Front Genet. 2014; https://doi.org/10.3389/fgene.2014.00190 PMID: 25009566

49. Smith-Unnas R, Boursnell C, Patro R, Hibberd J.M, Kelly S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 2016; 26:1134–1144. https://doi.org/10.1101/gr.196469.115 PMID: 27222236

50. Schliesky S, Gowik U, Weber A, Braeutigam A. RNA-Seq assembly--Are we there yet?. Front Plant Sci. 2012; 3:220. https://doi.org/10.3389/fpls.2012.00220 PMID: 23056003
51. Neil S, Emrich S. Assessing de novo transcriptome assembly metrics for consistency and utility. BMC Genomics. 2013; 14:465. https://doi.org/10.1186/1471-2164-14-465 PMID: 23837739
52. Gu Y, Qian D, Duan J, Wang Z, Guo J, Tang Y, et al. Simultaneous determination of seven main alkaloids of Chelidonium majus L. by ultra-performance LC with photodiode-array detection. J Sep Sci. 2010; 33:1004–1009. https://doi.org/10.1002/jssc.200900690 PMID: 20183823
53. Samanani N, Facchini PJ. Purification and characterization of norcoclaurine synthase. The first committed enzyme in benzylisoquinoline alkaloid biosynthesis in plants. J Biol Chem. 2002; 277:33878–33883. https://doi.org/10.1074/jbc.M203051200 PMID: 12107162
54. Samanani N, Liscombe DK, Facchini PJ. Molecular cloning and characterization of norcoclaurine synthase, the first committed enzyme in benzylisoquinoline alkaloid biosynthesis. Plant J. 2004; 40:302–313. https://doi.org/10.1111/j.1365-313X.2004.02210.x PMID: 15447655
55. Trezzini GF, Horrichs A, Sommssich IE. Isolation of putative defense-related genes from Chelidonium majus and expression in fungal elicitor-treated cells. Plant Mol Biol. 1993; 21:385–389.
56. Facchini P, De Luca V. Differential and tissue-specific expression of a gene family for tyrosine/dopa decarboxylase in opium poppy. J Biol Chem. 1994; 269:26684–26690. PMID: 7929401
57. Facchini P, Park SU. Developmental and inducible accumulation of gene transcripts involved in alkaloid biosynthesis in Opium Poppy. Phytochemistry. 2003; 64:177–186.
58. Park SU, Johnson AG, Penzes-Yost C, Facchini PJ. Analysis of promoters from tyrosine/dihydroxyphenylalanine decarboxylase and berberine bridge enzyme genes involved in benzylisoquinoline alkaloid biosynthesis in Opium Poppy. Plant Mol Biol. 1999; 40:121–31. PMID: 10394951
59. Gurkok T, Ozsuner E, Parmaksiz I, Özcan S, Turkas M, İpek A, et al. Functional Characterization of 4’OMT and 7OMT Genes in BIA Biosynthesis. Front Plant Sci. 2016; 7:98. https://doi.org/10.3389/fpls.2016.00098 PMID: 26909086
60. Liscombe DK, MacLeod BP, Loukanina N, Nandi OI, Facchini PJ. Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. Phytochemistry. 2005; 66:2501–2520. PMID: 16342378
61. Samanani N, Facchini PJ. Isolation and partial characterization of norcoclaurine synthase, the first committed step in benzylisoquinoline alkaloid biosynthesis, from opium poppy. Planta. 2001; 213:898–906. PMID: 11722126
62. Lee EJ, Facchini P. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. Plant Cell. 2010; 22:3489–503. https://doi.org/10.1105/tpc.110.077958 PMID: 21037103
63. Samanani N, Alcantara J, Bourgault R, Zulak K.G, Facchini P.J. The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy. Plant J. 2006; 47:547–563. https://doi.org/10.1111/j.1365-313X.2006.02801.x PMID: 16813579
64. Choi K.B, Morisige T, Shitani N, Yazaki K, Sato F. Molecular cloning and characterization of coclaurine N-methyltransferase from cultured cells of Coptis japonica. J Biol Chem. 2002; 277:830–835. https://doi.org/10.1074/jbc.M106405200 PMID: 11682473
65. Pauli H.H, Kutchan T.M. Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcoclaurine 3’-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. Plant J. 1998; 13:793–801. PMID: 9681018
66. Huang FC, Kutchan TM. Distribution of morphinan and benzophenanthidine alkaloid gene transcript accumulation in Papaver somniferum. Phytochemistry. 2000; 53:555–64. PMID: 10724180
67. Vrba J, Vrublova E, Modriansky M, Ultrichova J, Protopine and allocrypto pine increase mRNA levels of cytochromes P450 1A in human hepatocytes and HepG2 cells independently of AhR. Toxicol Lett. 2011; 203:135–141. https://doi.org/10.1016/j.toxlet.2011.03.015 PMID: 21419197
68. Zeng J, Liu Y, Liu W, Liu X, Liu F, Huang P, et al. Integration of Transcriptome, Proteome and Metabolism Data Reveals the Alkaloids Biosynthesis in Macleaya cordata and Macleaya microcarpa. PLOS ONE. 2013; https://doi.org/10.1371/journal.pone.0053409 PMID: 23326424
69. Liscombe DK, Facchini PJ. Molecular Cloning and Characterization of Tetrahydroprotoberberinecis-N-Methyltransferase, an Enzyme Involved in Alkaloid Biosynthesis in Opium Poppy. J Biol Chem. 2007; 282:14741–14751. https://doi.org/10.1074/jbc.M611908200 PMID: 17389594
70. Takeshita N, Fujisawa H, Mimura H, Fitchen J.H, Yamada Y, Sato F. Molecular cloning and characterization of S-adenosyl-L-methionine:scoloterine-9-O-methyltransferase from cultured cells of Coptis japonica. Plant Cell Physiol. 1995; 36:29–36. PMID: 7719631
71. Ikezawa N, Tanaka M, Nagayoshi M, Shinkyo R, Sakaki T, Inouye K, et al. Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured Coptis japonica cells. J Biol Chem. 2003; 278:38557–38565. https://doi.org/10.1074/jbc.M302470200 PMID: 12732624
91. Larsson E, Sundström JF, Sitbon F, von Arnold S. Expression of PaNAC01, a Picea abies CUP-Chen X, Facchini PJ. Short-chain dehydrogenase/reductase catalyzing the final step of noscapine biosynthesis. Sci Rep. 2016; https://doi.org/10.1038/srep22456 PMID: 26936416

89. Gomez-Gomez L, Trapero-Mozos A, Gomez MD, Rubio-Moraga A, Ahrazem O. Identification and characterization of a novel MYB transcription factor with potential roles in flavonoid and lignin biosynthesis. Plant Mol. Biol. 2013; 79:509–515. https://doi.org/10.1007/s11103-012-0110-2 PMID: 23094314

87. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant stress responses. Biochim Biophys Acta. 2012; 1819:86–96. https://doi.org/10.1016/j.bbagrm.2011.08.004 PMID: 21867785

85. Luliu H, Xiao Y, Pei S, Wen T, Shangqin H. The first illumina-based de novo transcriptome sequencing of the shoot apical meristem and formation of separate cotyledons. Annal Bot. 2012; 110:923–931. https://doi.org/10.1093/pcp/pcc137 PMID: 23098306

83. Kato N, Dubouzet E, Kokabu Y, Yoshida S, Taniguchi Y, Dubouzet JG, et al. Identification of a WRKY marker discovery. PLOS ONE. 2013; https://doi.org/10.1371/journal.pone.0062293 PMID: 23626799

81. Khanna R, Huq E, Kikis EA, Al-Sady B, Lanzatella C, Quail PH. A novel molecular recognition motif SHAPED COTYLEDON orthologue, is regulated by polar auxin transport and associated with differentiation of the shoot apical meristem and formation of separated cotyledons. Annal Bot. 2012; 110:923–934.

79. Chou J, Feng Q, Li X, Li T, Sun L, Wu Q, et al. The first illumina-based de novo transcriptome sequencing of the shoot apical meristem and formation of separate cotyledons. Annal Bot. 2012; 110:923–934.

77. Wang H, Jiang J, Chen S, Qi X, Peng H, Li P, et al. Next-generation sequencing of the Chrysanthemum nankense (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. PLOS ONE. 2013; https://doi.org/10.1371/journal.pone.0062293 PMID: 23626799

75. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol. 2004; 14:283–291. https://doi.org/10.1016/j.sbi.2004.06.004 PMID: 15193307

73. Hosseini B, Shahriari-Ahmadi F, Hashemi H, Marashi M.H, Mohseniazar M, Farokhzad A, et al. Transient Expression of cor Gene in Papaver somniferum. BiolImpacts. 2011; 1:229–235. PMID: 23678433

72. Lenz R, Zenk M.H. Purification and properties of codeinone reductase (NADPH) from Papaver somniferum cell cultures and differentiated plants. Eur J. 1995; 233:132–9.

70. Jain M, Srivastava PL, Verma M, Ghangal R, Garg R, Kikis EA, Al-Sady B, Lanzatella C, Quail PH. A novel molecular recognition motif SHAPED COTYLEDON orthologue, is regulated by polar auxin transport and associated with differentiation of the shoot apical meristem and formation of separated cotyledons. Annal Bot. 2012; 110:923–34.
93. Christiansen MW, Gregersen PL. Members of the barley NAC transcription factor gene family show differential co-regulation with senescence-associated genes during senescence of flag leaves. J Exp Bot. 2014; 65:4009–4022. https://doi.org/10.1093/jxb/eru046 PMID: 24567495

94. Cherukupalli N, Divate M, Mitappelli SR, Khareedu VR, Vudem DR. De novo Assembly of Leaf Transcriptome in the Medicinal Plant Andrographispaniculata. Front Plant Sci. 2016; 7:1203. https://doi.org/10.3389/fpls.2016.01203 PMID: 27582746

95. Han SM, Wu ZJ, Jin Y, Yang WN, Shi HZ. RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (Ginkgo biloba L.). Tree Genet Genomes. 2015; 11:37.

96. Han S, Wu Z, Wang X, Huang K, Jin Y, Yang W, et al. De novo assembly and characterization of Gleditsia sinensis transcriptome and subsequent gene identification and SSR mining. Genet Mol Res. 2015; https://doi.org/10.4238/gmr.15017740 PMID: 26909943

97. Wang Y, Liu K, Bi D, Zhou S, Shao J. Characterization of the transcriptome and EST-SSR development in Boea clarkeana, a desiccation-tolerant plant endemic to China. PeerJ. 2017; https://doi.org/10.7717/peerj.3422 PMID: 28630801

98. Bose Mazumdar A, Chattopadhyay S. Sequencing, De novo Assembly, Functional Annotation and Analysis of Phyllanthus amarus Leaf Transcriptome Using the Illumina Platform. Front Plant Sci. 2016; 6:1199. https://doi.org/10.3389/fpls.2015.01199 PMID: 26858723

99. Yan K, Wei Q, Feng R, Zhou W, Chen F. Transcriptome analysis of Cinnamomum longepaniculatum by high-throughput sequencing. Electron J Biotechnol. 2017; 28:58–66.

100. 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; 18:409. https://doi.org/10.1186/s12864-017-3780-9 PMID: 28545396

101. Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe BA, Wang Y. Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, Epimedium sagittatum (Sieb. Eizucc.) Maxim. BMC Genomics. 2010; 11:94–104. https://doi.org/10.1186/1471-2164-11-94 PMID: 20141623

102. Wei WL, Qi XQ, Wang LH, Zhang YX, Hua W, Li DH, et al. Characterization of the sesame (Sesamum indicum L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics. 2011; https://doi.org/10.1186/1471-2164-12-451 PMID: 21929789

103. Liu Y, Zhang P, Song M, Hou J, Qing M, Wang W, et al. Transcriptome Analysis and Development of SSR Molecular Markers in Glycyrrhiza uralensis Fisch. PLOS ONE. 2015; https://doi.org/10.1371/journal.pone.0143017 PMID: 26571372

104. Schommer C, Bresso EG, Spinelli SV, Palatnik JF. Role of MicroRNA miR319 in plant development. In: Sunukar R, editor. MicroRNAs in Plant Development and Stress Responses. Berlin, Germany: Springer; 2012. p. 29–47.

105. Debernardi JM, Rodriguez RE, Meccia MA, Palatnik JF. Functional Specialization of the Plant miR396 Regulatory Network through Distinct MicroRNA–Target Interactions. PLOS Genet. 2012; https://doi.org/10.1371/journal.pgen.1002419 PMID: 22242012

106. Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, Rodriguez R, et al. Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. Dev Cell. 2007; 13:115–125. https://doi.org/10.1016/j.devcel.2007.04.012 PMID: 17609114

107. Prabu G, Mandal A. Computational identification of miRNAs and their target genes from expressed sequence tags of tea (Camellia sinensis). Genom Proteom Bioinform. 2010; 8:113–21.

108. Branscheid A, Devers EA, May P, Krajinski F. Distribution pattern of small RNA and degradome reads provides information on miRNA gene structure and regulation. Plant Signal Behav. 2011; 6:1609–1611. https://doi.org/10.4161/psb.6.10.17305 PMID: 21957499

109. Liu H, Jia SH, Shen DF, Liu J, Li J, Zhao H, et al. Four AUXIN RESPONSE FACTOR genes downregulated by microRNA167 are associated with growth and development in Oryza sativa. Funct Plant Biol. 2012; 39:736–744.

110. Brezinaova B, Macak M, Effimova J. The morphological diversity of selected traits of world collection of poppy genotypes (genus papaver). J Cent Eur Agr. 2009; 10:183–190.