Plant Gene 2 (2015) 17–24

Contents lists available at ScienceDirect

Plant Gene

journal homepage: www.elsevier.com/locate/plant-gene

In silico identification, characterization and expression analysis of miRNAs in *Cannabis sativa* L.

Akan Das [a,*], Sumi Chaudhury [a], Mohan C. Kalita [a], Tapan K. Mondal [b]

[a] Department of Bioengineering and Technology, Gauhati University-Institute of Science and Technology, Gopinath Bordoloi Nagar, Guwahati 781014, Assam, India
[b] Division of Genomic Resource, National Bureau of Plant Genetic Resource, Pusa Campus, New Delhi 110012, India

ARTICLE INFO

Article history:
Received 11 February 2015
Received in revised form 13 March 2015
Accepted 23 March 2015
Available online 26 March 2015

Keywords:
Computational
Conserved
Expression
Marijuana
Hemp
Targets

ABSTRACT

Cannabis sativa L. is an annual herb and economically important as a source of fiber, oil, food and for its medicinal and intoxicating properties. MicroRNAs are a class of short (~21 nt), non-coding regulatory RNAs that play a major role in post-transcriptional gene silencing. By in silico analysis of the publically available Transcript Sequence Assemblies (TSA) and Expressed Sequence Tags (ESTs) of *C. sativa*, a total of 18 conserved miRNAs belonging to 9 independent families were identified. To validate the predicted miRNAs, SYBR green based assay of qPCR was applied to detect the tissue-specific (young and mature leaf) expression of 6 putative miRNAs (csa-miR156, csa-miR159a, csa-miR171b, csa-miR172a, csa-miR5021a, csa-miR6034) in *C. sativa*. A total of 80 target genes were also recognized for the newly identified miRNAs, and subsequently assigned to three broad functional categories: biological processes, cellular components and molecular functions as defined for the Arabidopsis proteome. The potential target genes consist of transcription factors (33.75%), transporters (5%), kinase and other enzymes (20%) as well as signaling and other functional proteins (32.50%). The findings in this study on *C. sativa* miRNA precursors, mature miRNAs, and miRNA targets will be helpful for future research on miRNA-mediated gene regulation in this important plant species.

1. Introduction

MicroRNAs (miRNAs) are a class of small (~21 nt) non-coding RNA molecules found in plants, animals, and some viruses, which act on transcriptional and post-transcriptional regulation of gene expression. The precursor sequence of miRNAs (pre-miRNAs) can fold into a stem-loop secondary structure containing the mature miRNA on either of the strand of stem region (B. Zhang et al., 2006). MiRNAs are produced from the precursor molecule to mature miRNAs by several enzymatic reactions inside the nucleus in plants (Rogers and Chen, 2013). Mature miRNAs act as an important fine-tuning regulator in various physiological processes of plants, including leaf, stem and root development (Kidner and Martienssen, 2005), anther development (Millar and Gubler, 2005), floral organ identity and flowering time (Yang et al., 2007), cell signaling (Kidner and Martienssen, 2005), oxidative stress regulation (Sunkar et al., 2006), abiotic and biotic stress responses (Khraiwesh et al., 2012).

Most of the known plant miRNAs are evolutionarily conserved from lower to higher flowering plants (B. Zhang et al., 2006). The conserved nature of plant miRNAs provides the scope of finding homolog sequences of miRNAs in nucleotide resources of different plant species using in silico homology based approach (Das and Mondal, 2010). Most often publically available plant genomic resources, such as Expressed Sequence Tags (ESTs), Genomic Survey Sequences (GSS), Transcript Sequence Assemblies (TSA), High-Throughput Genomic Sequences (HTGS), and non-redundant (NR) nucleotides have been utilized for searching the miRNA homologs. Numbers of conserved miRNAs have been reported in diverse plant species including model and non-model plants by utilizing those genomic resources through homology based in silico approach, such as the identification of 682 miRNAs in 155 diverse plant species (Sunkar and Jagadeeswaran, 2008). In silico identification of miRNAs has also been reported in maize (B. Zhang et al., 2006), tea (Das and Mondal, 2010), cowpea (Lu and Yang, 2010), lotus (Hu et al., 2013), foxtail millet (Fei et al., 2013) and so on. The current trends, however, follow both the methods: genetic and in silico in combination, for the discovery of miRNA and their regulatory functions (Oliveira et al., 2013).

Plant miRNAs have been reported to silence a target gene by binding to a perfect or nearly perfect complementary site. The complementary sequence suggests a powerful method for identifying miRNA target

Abbreviations: ESTs, Expressed Sequence Tags; TSA, transcribed sequence assemblies; MiRNAs, MicroRNAs; DCL, dicer like enzyme; Nt, nucleotide; Pre-miRNAs, precursor microRNAs; BLAST, basic local alignment search tool; GO, gene ontology.

* Corresponding author.
E-mail address: dasakan@gmail.com (A. Das).

http://dx.doi.org/10.1016/j.plgene.2015.03.003

genes by homology analysis using bioinformatics tools (Chorostecki et al., 2012). Most of the known plant miRNA target genes have been predicted on the basis of miRNA complementary site, such as the identification of 115 target genes in maize (B. Zhang et al., 2006), 30 in cowpea (Lu and Yang, 2010), 37 in tea (Das and Mondal, 2010), and 735 in lotus (Hu et al., 2013). Information of known genes and proteins in public domain and gene ontology (GO) categories provide an opportunity to decipher the functions of the predicted target genes (Ashburner et al., 2000). Model plant genome has most often been used for the prediction of target genes for those plants with unavailable genome sequences in public domain (Das and Mondal, 2010).

Marijuana, Bhang or hemp (*Cannabis sativa* L.) is an annual herb and has been cultivated throughout the recorded human history as a source of fiber, oil and protein-rich achenes ("seeds") and for its medicinal and psychoactive properties (Simmonds, 1979; Mikuriya, 1969; Abrams, 1998; Sirikantaramas et al., 2004; Bakel et al., 2011). Great progress has been made on marijuana genomics research in recent years (Sirikantaramas et al., 2004; Bakel et al., 2011; Marks et al., 2009; Divashuk et al., 2014). Recently, draft genome and large scale analysis of transcript sequences of *C. sativa* have also been reported (Bakel et al., 2011). Till date there are 60,029 unplaced genomic scaffold of whole genome shotgun sequences 33,215 TSA, 12,907 ESTs and 37 chloroplast gene sequences and 23 unassembled RNA-Seq datasets of Illumina reads available in sequence databases of NCBI (http://www.ncbi.nlm.nih.gov/) (Bakel et al., 2011; Marks et al., 2009). Besides, there is no report of *C. sativa* miRNA in miRNA registry databases, such as miRBase (http://www.mirbase.org), Plant microRNA Database (PMRD, http://bioinformatics.cau.edu.cn/PMRD/). In this study, we used known miRNAs to systematically search miRNA homologs in available EST and TSA sequences of *C. sativa*. The expression of six mature miRNAs was also validated using qPCR. Target genes of the identified miRNAs were further recognized in Arabidopsis genome, as model dicot plant. The findings in this study on *C. sativa* miRNA precursors, mature miRNAs, and miRNA targets will be helpful for future research on miRNA-mediated gene regulation in this important plant species

2. Materials and methods

2.1. Plant material

Two-month-old seedlings of *C. sativa* L. were collected from the local area of Guwahati, Assam (India). The seedlings were planted on a fertile, neutral to slightly alkaline, well-drained clay loams with moisture retentive subsoil in a medium sized plastic bag, and kept in a shady area within a temperature range of 25–30 °C (Supplementary material, Fig. S1). After commencement of new growth, young (10 day old) and mature (6 week old) leaf samples were immediately frozen in liquid nitrogen and stored at −80 °C until use.

2.2. Reference miRNAs and nucleotide resources of C. sativa

Previously known 7385 miRNAs of diverse plant species were downloaded from miRNA registry database i.e. miRBase (http://www.mirbase.org, Released 20: June, 2013) and clustered by CD-HIT-EST (Li and Godzik, 2006) with c = 1, n = 8, d = 250, and g = 1. From the clustered sequences, 4025 non-redundant miRNAs were selected which in turn used as reference miRNAs for finding the homologs in *C. sativa*. Publically available, 12,907 ESTs and 33,215 TSA sequences of *C.* sativa were downloaded from GenBank, NCBI (http://www.ncbi.nlm.nih.gov/). The low quality sequences were eliminated using Sequencher 5.1 (Gene Code Corporation, USA) and subsequently quality sequences were used to create a local nucleotide sequence database.

2.3. Identification and analysis of conserved miRNAs and their targets

Homology search of reference miRNAs against the local nucleotide sequence database of *C. sativa* was carried out using Standalone BLAST+ 2.2.29 program at an e-value threshold ≤ 0.001 (Altschul et al., 1997). The obtained hits with maximum 3 nt mismatches and without gap were considered for extracting the precursor sequences (pre-miRNA). The pre-miRNA sequences were extracted following the method of a sliding window of about 100 nt in size (moving in increments of approximately 20 nt) from the region ~80 nt upstream of the beginning of the mature miRNA to ~80 nt downstream of the miRNA (Singh and Nagaraju, 2008). The fold-back secondary structures of pre-miRNAs were predicted using Mfold (Zuker, 2003). The following criteria were used for selecting the pre-miRNA structures according to Lu and Yang (2010) as: (1) the sequence could fold into an appropriate stem-loop hairpin secondary structure; (2) predicted mature miRNAs with no more than 3 nt substitutions as compared with the known miRNAs; (3) no more than 6 mismatches are between the predicted mature miRNA sequence and its opposite miRNA* sequence in the secondary structure; (4) mature miRNA hit should be on the stem region of the hairpin structure; (5) no loop or break is in the miRNA* sequences, and (6) predicted secondary structure has higher MFEI and negative MFE. MFEI values are calculated according to B. H. Zhang et al. (2006) following the equation as: MFEI = [(100 × MFE) / Length of RNA / (G + C)]%. For finding the conservation of miRNA of *C. sativa* with other plant species, pre-miRNA sequences of miR172 family from 9 different species, i.e. *Oryza sativa*, *Arabidopsis thaliana*, *Zea mays*, *Sorghum bicolor*, *Gossypium hirsutum*, *Medicago truncatula*, *Brassica rapa*, *Glycine max* and *Solanum lycopersicum* were downloaded from miRBase and aligned by using ClustalX (Thompson et al., 1997). Based on the conservation among the pre-miRNAs, a WebLogo was prepared using a WebLogo online program (Crooks et al., 2004).

The potential target genes of the identified miRNAs were predicted using the plant miRNA target finder program (www.http://plantgrn.noble.org/psRNATarget). The identified mature miRNAs were used as query for finding the complementary sequences in *A. thaliana* unigenes [DFCI gene index (AGI), version 15] using the parameters as: (1) maximum expectation value-3, (2) length of complementary scoring (hspsize)-20, (3) range of central mismatch for translational inhibition 9–11 nt and (4) multiplicity of target site-2. The predicted miRNA/target pairs were obtained and the sequences of the target accession were downloaded. Loci of the predicted target sequences were collected from the TAIR database (www.arabidopsis.org) using the WU-BLAST program (at an e-value threshold of $1e^{-10}$), and used as query for finding the ontology in various functional categories of the target genes on the basis of GOslim categories (www.arabidopsis.org) as: annotations to terms in GOslim category / total annotations to terms in this ontology * 100.

2.4. Expression analysis of the potential C. sativa miRNAs

2.4.1. Isolation of total RNA

Total RNA was isolated from 100 mg of the collected young and mature leaf samples using Trizol reagent (Nitrogen, Carlsbad, CA, USA) following the manufacturer's instruction. The quantity and quality of the isolated RNA were determined using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA) and 2% agarose gel electrophoresis in MOPS [3-(N-morpholino) propanesulfonic acid] buffer, respectively. The isolated RNA was stored at −80 °C until further use.

2.4.2. Polyadenylation and cDNA synthesis

The isolated RNAs (5 μg) were polyadenylated and reverse transcribed at 37 °C for 1 h in 10 μl reaction mixture following the instructions of Mir-X™ miRNA first-strand synthesis kit (Clontech, USA). The reaction mixture contains 1× mRQ buffer and 1.25 μl of mRQ enzyme mix provided with the kit. After 1 h of incubation, the reaction was

terminated by incubating at 85 °C for 5 min and finally the volume was made up to 100 μl by adding deionized water.

2.4.3. Quantitative real-time PCR (qPCR) experiments

Expression analysis of the potential *C. sativa* miRNAs was done following the protocol of Mir-X miRNA SYBR qPCR Kit (Clontech, USA). Briefly, 25 μl PCR reaction mixtures was prepared containing 1× SYBR advantage premix, 1× ROX dye, 0.2 μM each of sense and antisense primers and 2 μl of the first strand cDNA. The entire mature miRNA sequence was used as miRNA specific, 5′ primer (Supplementary material, Table S1) and mRQ, the 3′ primer was used from the kit. The reactions were incubated in a 96 well plate at 95 °C for 2 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 20 s in a Roche 454 qPCR system (Roche, USA). This cycle was followed by a melting curve analysis ranging from 56 to 95 °C, with temperature increasing steps of 0.5 °C every 10 s. Melting curves for each amplicon were observed carefully to confirm the specificity of the primers used. Relative expression levels for each sample were obtained using the 'comparative Ct method' (Schmittgen and Livak, 2008). The threshold cycle (Ct) value obtained after each reaction was normalized to the Ct value of U6 snRNA (U6 snRNA primer was provided with the kit) of which expression was consistent across the conditions. Statistical analyses were conducted using the SAS software of JMP Genomics (SAS Institute, NC, USA). Both biological and technical replicate samples and U6 snRNA reactions were conducted in triplicates.

3. Results and discussion

3.1. Potential C. sativa miRNAs and their in silico analysis

The BLASTn analysis of 4025 unique viridiplantae mature miRNAs in EST and TSA sequences of *C. sativa* produced 166 hits with 0–3 mismatches. The careful evaluation of stem-loop structures at defined criteria for the hit sequences, led us to identify 18 conserved miRNAs in *C. sativa*. Information on predicted *C. sativa* miRNAs, including names, length, sources, and other aspects, were listed in Table 1. The identified putative miRNAs belonged to 9 different families: miR156, miR159, miR166, miR171, miR172, miR1171, miR5021, miR5368, and miR6034 and named according to the rules of Ambros et al. (2003). Out of 18, 12 miRNAs were predicted in TSA and 6 in EST sequences. The number of members ranged from 1 to 7 in each of the miRNA family. MiR172 had the highest 7 members, three families i.e. miR159, miR171 and miR5021 had 2 members in each and the rest of families were with single member. This may indicate the dominant roles of miR172 mediated gene regulation in *C. sativa*. Five different members of miR172 family were reported by Lu and Yang (2010) in cowpea. The familial distribution of miRNAs was also observed in numbers of other plant species, such as *Arabidopsis* (Jones-Rhoades and Bartel, 2004), maize (B. Zhang et al., 2006), Tea (Das and Mondal, 2010), lotus (Hu et al., 2013) and so on.

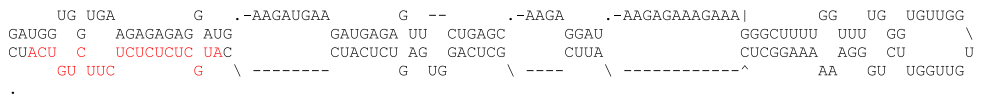
The length of mature and pre-miRNAs was also found diverse in *C. sativa*. The identified mature miRNAs were 19 nt to 23 nt, an average of 21 nt in length (Supplementary material, Fig. S2). The predicted pre-miRNAs varied from 111 nt to 757 nt, an average of 234 nt in length. All the pre-miRNAs form into a stem-loop secondary structure, with the mature miRNAs either on the 5′ or 3′ ends (Fig. 1). The majority of the pre-miRNAs (66.67%) have 100–200 nt and least (22.22%) have 200–350 nt in length. However, there is a few pre-miRNAs (11.11%) above 350 nt in length (Supplementary material, Fig. S3). Our finding is supported by the previous fundamental discovery in *Arabidopsis* (Bologna et al., 2009), where precursor length was found in a range of 50–900 nt. In contrast to animal pre-miRNAs with only 70–90 nt in length (Bologna et al., 2013), a wide range of precursor length in plants has already been reported in a number of diverse plant species, including monocot, such as Maize (B. Zhang et al., 2006) and dicot species, such as Citrus (Wu et al., 2010). It was well demonstarted in *Arabidopsis* that variation in length of pre-miRNAs in plants is due to the multiple RNA recognition patterns during miRNA biogenesis (Bologna et al., 2013). The (A + U) content of stem-loop structures ranged from 43.32% to 72.81%, with an average 63% (Table 1, Supplementary material, Fig. S6). MFE values of stem-loop structures ranged from −195.5 kcal/mol to −44.90 kcal/mol, with an average −80.78 kcal/mol. The highest negative MFE value was −195.5 kcal/mol for csa-miR166 and least −44.90 kcal/mol for miR171a (Table 1). The high negative MFE values of predicted pre-miRNAs in *C. sativa* indicate their stable secondary structures. The MFEI is a criterion for distinguishing a potential miRNA from other types of RNAs such as tRNA (0.64), rRNAs (0.59), and mRNAs (0.62–0.66) (B. H. Zhang et al., 2006). The MFEI of the predicted C. *sativa* pre-miRNAs ranged from 0.67 to 1.37, with an average 1.04 (Table 1), which is above the golden standards. The comparison of pre-miRNAs (156a, 159a, 171b, 172a) and their secondary structures of *C. sativa* with *Arabidopsis* counterparts, revealed that there are no such consistent differences or similarities between them, other than high (A + U) content in *C. sativa* pre-miRNAs (Supplementary Table S2; Fig. S7-A). The *C. sativa* pre-miRNAs are not highly conserved with *Arabidopsis*, although similar members of a family form the same cluster in phylogeny (Supplementary Fig. S7-B-E).

Table 1
Detailed information on the identified potential miRNAs of *Cannabis sativa* L.

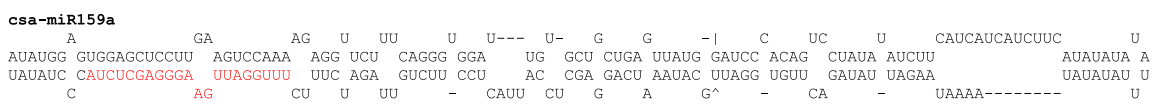
MiRNAs	NS*	AC IDs*	STRAND	SP*	EP*	MM*	PL*	(A + U)	MFE*	MFEI*
csa-miR156	TSA	351623229	+/−	721	702	0	590	60.98	−181.7	0.78
csa-miR159a	EST	347290902	+/+	371	391	0	218	72.81	−81.3	1.37
csa-miR159b	EST	347290902	+/+	371	390	1	218	72.81	−81.3	1.37
csa-miR166	TSA	351625703	+/+	4029	4049	2	757	62.22	−195.5	0.68
csa-miR171a	EST	347293330	+/+	163	182	0	111	69.53	−44.9	1.32
csa-miR171b	EST	347293330	+/+	163	183	0	111	69.53	−44.9	1.32
csa-miR172a	TSA	351612968	+/+	229	249	0	167	61.21	−80.1	1.23
csa-miR172b	TSA	351612968	+/+	227	249	0	167	61.21	−80.1	1.23
csa-miR172c	TSA	351608945	+/+	181	201	0	135	65.56	−48.3	1.03
csa-miR172d	TSA	351608945	+/+	179	201	0	135	65.56	−48.3	1.03
csa-miR172e	TSA	351612968	+/+	106	126	0	167	61.21	−80.1	1.23
csa-miR172f	TSA	351608945	+/+	181	199	0	135	65.56	−48.3	1.03
csa-miR172g	TSA	351612968	+/+	230	249	0	167	61.21	−80.1	1.23
csa-miR1171	TSA	351628069	+/+	3276	3298	2	316	67.72	−69	0.67
csa-miR5021a	TSA	351624730	+/−	65	46	0	311	59.25	−109.8	0.86
csa-miR5021b	EST	347293771	+/−	142	125	1	136	58.62	−51.1	0.9
csa-miR5368	EST	238815235	+/+	228	246	0	170	43.32	−71.1	0.73
csa-miR6034	TSA	351600181	+/+	1735	1755	0	196	61	−58.1	0.76

* (NS = Type of nucleotide sequence, SP = Start position, EP = End position, MM = number of mismatch bases, PL = Precursor length, MFE = Minimal folding free energies, MFEI = Minimal folding free energy index).

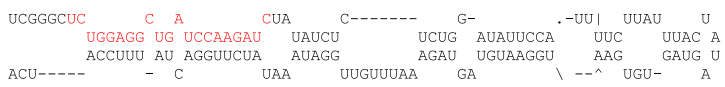
csa-miR156



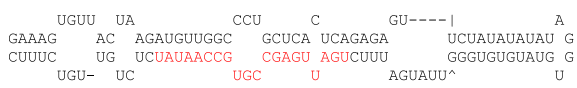
csa-miR159a



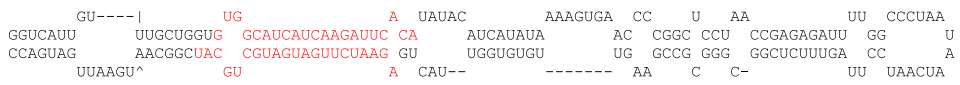
csa-miR166



csa-miR171a

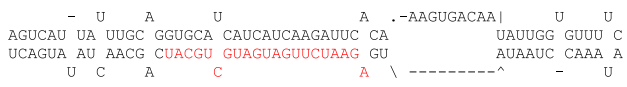


csa-miR172e

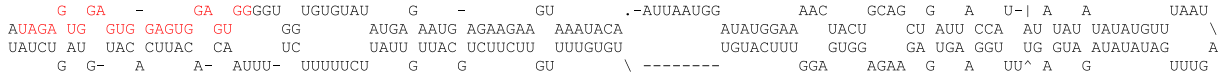


csa-miR172a

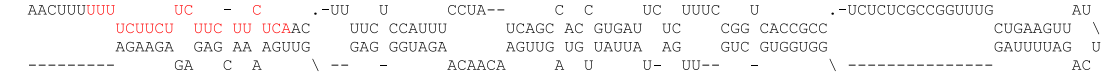
csa-miR172c



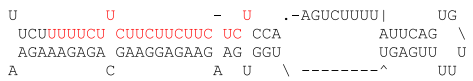
csa-miR1171



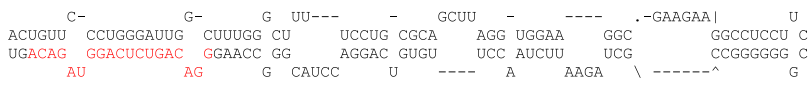
csa-miR5021a



csa-miR5021b



csa-miR5368



csa-miR6034

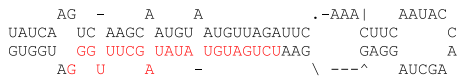


Fig. 1. Predicted hairpin secondary structure of candidate pre-miRNAs. The miRNAs are highlighted (red color) in stem portion.

Our results showed that some precursor sequences encoded more than one miRNA, such as TSA (gi|351612968) encoded csa-miR172a/b/e/g stem-loop structures, which resided at the 229–249, 227–249, 106–126 and 230–249 bp regions, respectively (Table 1). One more TSA (gi|351608945) and two ESTs (gi|347290902, gi|347293330) were found to encode either two or three miRNAs with almost identical sequences with 1–3 nucleotide differences (Supplementary material, Fig. S5). As a result, the 18 pre-miRNAs were actually represented by 11 unique precursor sequences of *C. sativa*. Besides, there were two TSA sequences (gi|351608945, gi|351612968)

which encode different members of the same family, i.e. miR172. Our finding on origination of different miRNAs in the same precursor sequence is supported by earlier reports on Citrus (Wu et al., 2010) and *Arabidopsis* (Zhang et al., 2010). It was demonstrated in *Arabidopsis* that multiple miRNAs could derive from a same miRNA precursor by sequential processing of Dicer or Dicer-like proteins. Their discovery suggested that miRNA-mediated gene regulation might be broader and more complex than that of a previous hypothesis (Zhang et al., 2010).

The conserved nature of plant miRNAs and their precursor sequences in diverse and distantly related species have been well demonstrated in a number of studies (B. Zhang et al., 2006; Hu et al., 2013; Bologna et al., 2013). In the present study, comparison of the pre-miRNA sequences of csa-miR172a from 9 different plant species showed a high degree of conservation of mature miRNAs among them (Fig. 2-A). It was observed that miRNA sequence of miR172a was more conserved than its miRNA* site and pre-miRNA sequence (Fig. 2-B). High degree conservation of csa-miR172a among diverse plant species indicates miR172 family members as possible strong candidates of miRNAs in *C. sativa* and suggested their similar role in post-transcriptional gene regulation and physiological function.

3.2. Potential miRNA target genes and their functional categorization

Here, we recognized 80 potential target genes for the 8 identified miRNA families of *C. sativa* based on the fact of sequence complementarity between miRNAs and their target genes in plants (Supplementary material, Fig. S6, Table S3). There were a maximum of 27 target genes for miR172 family and least 1 for miR166 family, however, no target gene could be predicted for miR5368 family. MiRNA-mediated silencing of target genes may occur either via mRNA degradation or preventing mRNA from being translated (Rogers and Chen, 2013). Out of the 80 potential target genes, the regulation of 62 (77.5%) target genes was due to cleavage and 18 (22.5%) due to translational inhibition.

The identified target genes were assigned to three broad functional categories, i.e., cellular components, biological processes and molecular functions as established for the *Arabidopsis* proteome (Fig. 3). In the biological component category, genes assigned to the 'other cellular processes' category accounted for the largest group (29.73%), followed by 'other metabolic processes' associated genes (17.647%), and genes of 'DNA or RNA metabolism' represented least in the group (0.318%). In the cellular process category, the largest group was 'nucleus' (24.737%) followed by 'other intracellular components' (16.842%) and the 'unknown cellular components' genes were the least in the group (0.526%). In the molecular functions category, the highest percentage was covered by 'transferase activity' (14.554%) followed by 'DNA or RNA binding activity' (13.615%), and the 'receptor binding or activity' were the least in the group (0.939%).

The potential target genes of the 8 identified miRNA families include transcription factors (33.75%), transporters (5%), kinase and other enzymatic proteins (20%), signaling and other functional proteins (32.50%) as well as unknown proteins (8.75%) (Supplementary material, Table S3). MiR156 family found to be targeted by various Squamosa Promoter-Binding protein-like (SPL) genes. The SPLs encode plant-specific transcription factors that play important roles in development, such as phase transition and plant architecture (Huijser and Schmid, 2011). MiR159 is a highly conserved and abundant miRNA in plants which restricts the expression of some MYB transcription factors. The interplay of miR159 and its target MYB is involved in the regulation of vegetative growth, flowering time, anther development, seed shape and germination (Li et al., 2011). In the present study, other target genes found to be regulated by miR159 are leucine-rich repeat transmembrane protein kinase, 1-amino-cyclopropane-1-carboxylate synthase 8 and oligopeptide transporter. Recent research has revealed that miR171 family negatively regulates shoot branching and decreases primary root elongation by targeting GRAS gene family members, such as Scarecrow-like 6-II (SCL6-II), SCL6-III, and SCL6-IV in Arabidopsis (Wang et al., 2010). It is supported by our present analysis where miR171 found to target GRAS family transcription factor. There were two other genes, i.e. ankyrin repeat family protein and acyltransferase family protein reported here to be targeted by miR171. There were a diverse set of genes found to be targeted by miR172. Out of these, there were zinc finger domains containing and basic-leucine zipper (bZIP) transcription factor family protein. It was demonstrated that flowering time regulator, bZIP transcription factor was involved in the process of inducing downstream flower-specific targets such as the MADS-domain proteins APETALA (AP) and FRUITFULL (FUL) (Mathieu et al., 2009). MiR172 serves as a negative regulator of AP2 to specify floral organ identity in *Arabidopsis* (Zhu et al., 2009). However, the present study indicates that miR172 may directly act on

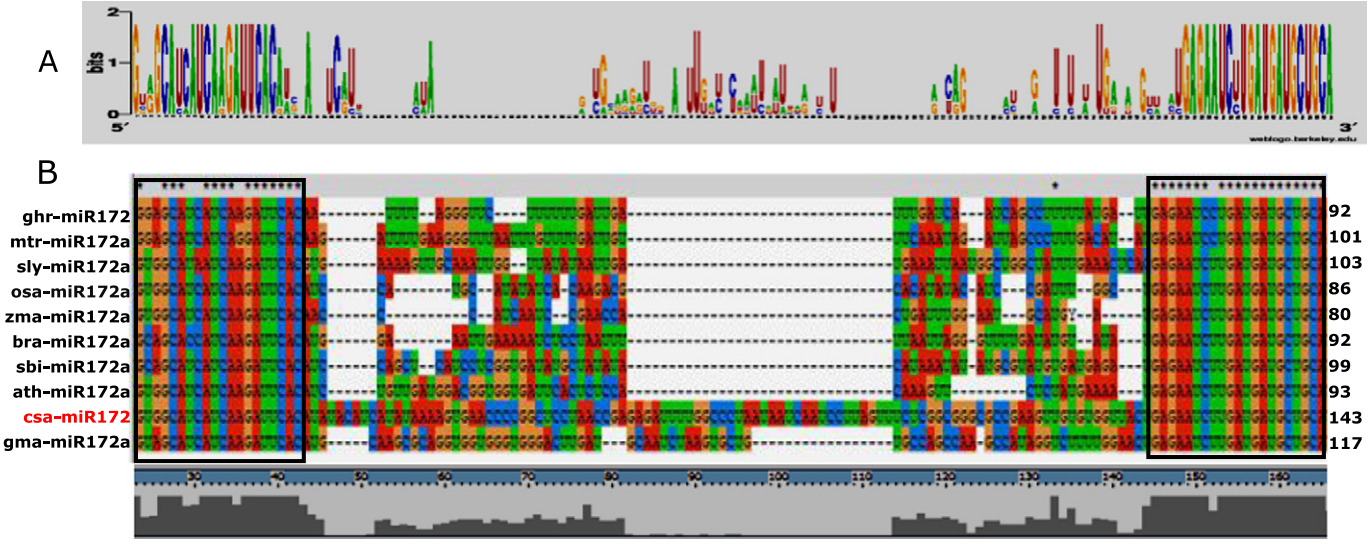


Fig. 2. Conservation of miR172: (A) Sequence logo showing conserved nucleotide composition of pre-miR172 sequences using WebLogo, (B) Alignment of potential precursor csa-miR172 and nine pre-miR172 sequences of diverse plant species [*C. sativa* (csa), *G. hirsutum* (ghr), *M. truncatula* (mtr), *S. lycopersicum* (sly), *Z. mays* (zma), *B. rapa* (bra), *S. bicolor* (sbi), *A. thaliana* (ath), *O. sativa* (osa), *G. max* (gma)].

transcription factors that regulate flowering time. There were two genes, i.e. auxin-independent growth promoter protein-like and calmodulin-dependent protein kinase (CaMK4) found to be targeted by miR1171. The calmodulin-binding protein kinases are likely to be crucial mediators of responses to diverse endogenous and environmental cues in plants (Zhang and Lu, 2003). The miR5021 and miR6034 families also found to target a diverse set of genes in the present study. It includes ribosomal proteins, calcium-binding EF hand family protein, global transcription factor, disease resistance protein, leucine-rich receptor like protein kinase and so on. It indicates their pivotal roles in regulating different biochemical pathways in growth and development, defense and stress resistance in *C. sativa*. The underlining facts may come to light in the future through cloning and characterization of targets and target gene mutant analysis in model plants.

3.3. Expression analysis of potential C. sativa miRNAs

Real time-qPCR based validation of the predicted miRNAs showed differential expression pattern in different tissue types of *C. sativa*. A total of 6 miRNAs i.e., csa-miR156, csa-miR159a, csa-miR171b, csa-miR172a, csa-miR5021a, and csa-miR6034 were validated in young and mature leaf tissues of *C. sativa* (Fig. 4). Compared to the young leaf, three miRNAs, namely csa-miR156, csa-miR5021a and csa-miR6034 were highly expressed in mature leaf. The high expression of csa-miR156 in mature leaf may be for the change of leaf architecture due to aging. The csa-miR171b and csa-miR172a miRNAs were abundant in young leaf. However, there is no available literature information on the expresssion analysis of csa-5021 and csa-6034 in leaf of plants. In the present study, high expression of csa-5021 and csa-6034 in mature

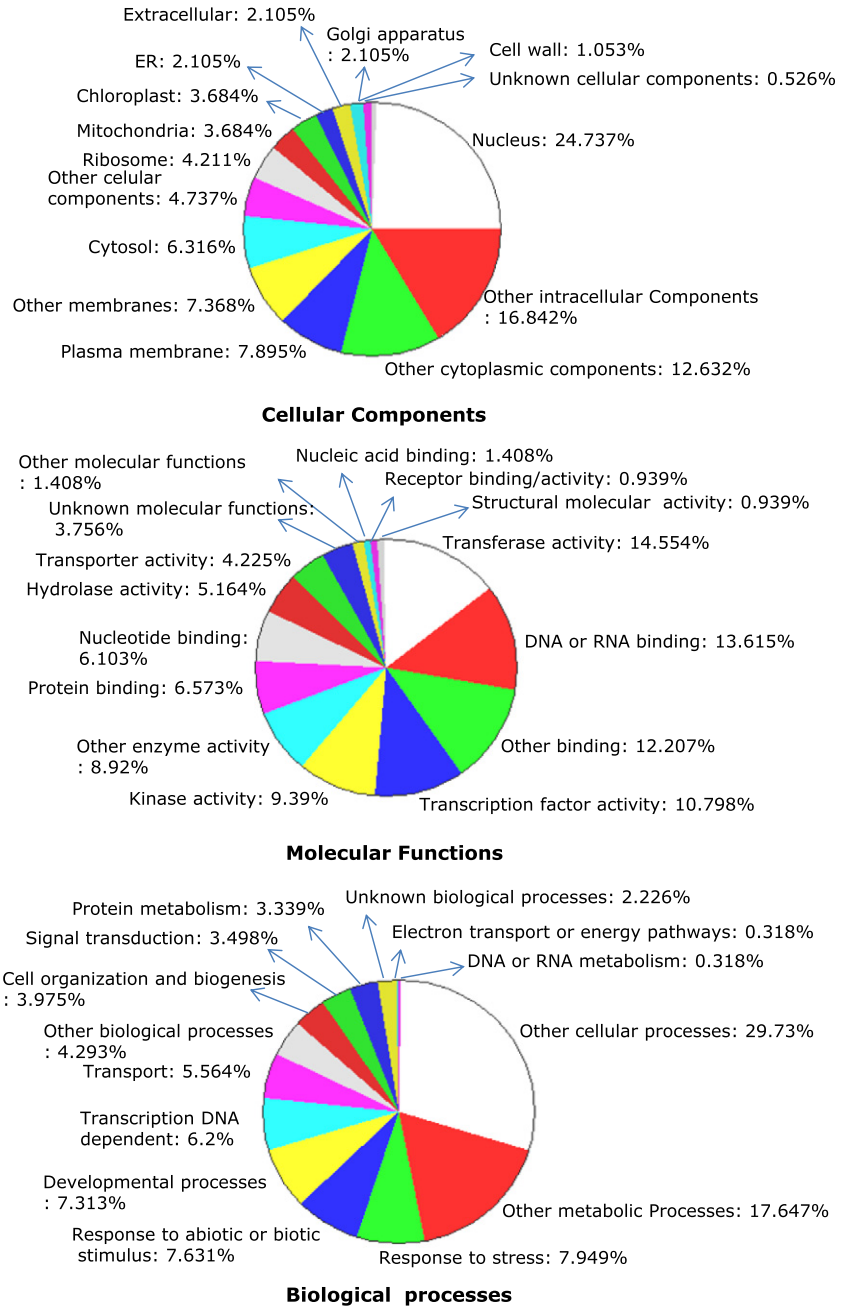


Fig. 3. Gene ontology (GO) analysis of *C. sativa* target genes. The relative frequencies of GO hits for target genes assigned to the GO functional categories: cellular components, biological processes and molecular functions, as defined for the *Arabidopsis* proteome.

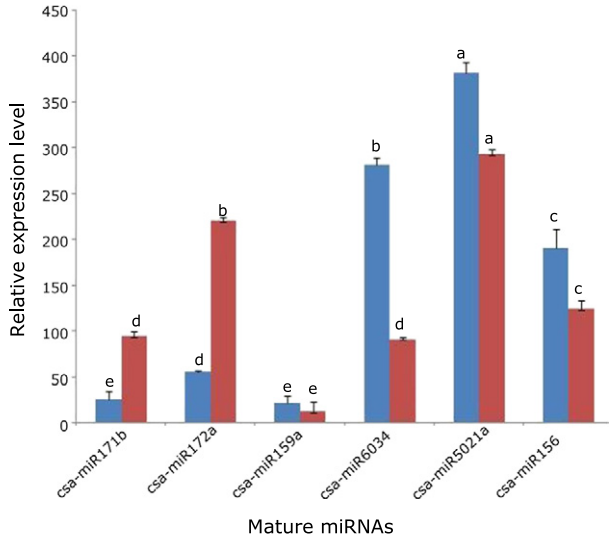


Fig. 4. Relative expression level of miRNAs in mature leaf (blue bar) and young leaf (red bar) of *Cannabis sativa* as revealed by qPCR analysis. Different letters above the bar in between tissue types (mature and young) indicate significant differences in expression (a, b, c, d, e letters indicate the different levels of expression at $<p$ 0.05 in Least Significant Difference analysis using One way analysis of variance).

leaf is reported. The abundance of csa-miR171 in young leaf proved its role in negative regulation of shoot branching during leaf development by targeting transcription factors and csa-miR172 giving a signal of the reproductive growth of the plants. However, the expression level of csa-miR159a was very less in both young and mature leaf tissues but it was comparatively high in mature leaf. It indicates that csa-miR159 may play an important role in flowering and fertility rather than vegetative growth and development in seedlings of *C. sativa*.

4. Conclusion

This study makes it evident that miRNAs play an important role in growth and developmental processes of *C. sativa*. The identified 18 miRNAs of 9 different families targeted 80 potential genes of transcription factors, transporters, kinase and other enzymes as well as signaling and stress responsive proteins. The qPCR based verification of the predicted miRNAs proved the reliability of in silico approach of miRNA identification. The findings in this study on *C. sativa* miRNA precursors, mature miRNAs and miRNA targets will be of immense help in future research on miRNA-mediated gene regulation and physiological function of this economically important plant species.

Acknowledgements

The authors are highly grateful to the Vice-Chancellor of Gauhati University (GU) and Director, GU-Institute of Science and Technology for providing the necessary facilities for the successful completion of this work. The authors are also thankful to Ms Purabi Das for proofreading the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.plgene.2015.03.003.

References

Abrams, D.I., 1998. Medical marijuana: tribulations and trials. J. Psychoactive Drugs 30, 163–169.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
Ambros, V., Bartel, B., Bartel, D.P., et al., 2003. A uniform system for microRNA annotation. RNA 9, 277–279.
Ashburner, M., Ball, C.A., Blake, J.A., et al., 2000. The gene ontology consortium. Nat. Genet. 25, 25–29.
Bakel, H.V., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., Page, J.E., 2011. The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. http://dx.doi.org/10.1186/gb-2011-12-10-r102.
Bologna, N.G., Mateos, J.L., Bresso, E.G., Palatnik, J.F., 2009. A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. EMBO J. 28, 3646–3656.
Bologna, N.G., Schapire, A.L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B.C., Palatnik, J.F., 2013. Multiple RNA recognition patterns during microRNA biogenesis in plants. Genome Res. 23, 1675–1689.
Chorostecki, U., Crosa, V.A., Lodeyro, A.F., Bologna, N.G., Martin, A.P., Carrillo, N., Schommer, C., Palatnik, J.F., 2012. Identification of new microRNA-regulated genes by conserved targeting in plant species. Nucleic Acids Res. 40, 8893–8904.
Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14, 1188–1890.
Das, A., Mondal, T.K., 2010. Computational identification of conserved microRNAs and their targets in tea (*Camellia sinensis*). Am. J. Plant Sci. 2, 77–86.
Divashuk, M.G., Alexandrov, O.S., Razumova, O.V., Kirov, I.V., Karlov, G.I., 2014. Molecular cytogenetic characterization of the dioecious *Cannabis sativa* with an XY chromosome sex determination system. PLoS One. http://dx.doi.org/10.1371/journal.pone.0085118.
Fei, Y., Shaojun, X., Yuwei, L., Xin, Q., Jingjuan, Y., 2013. Genome-wide characterization of microRNA in foxtail millet (*Setaria italica*). BMC Plant Biol. http://dx.doi.org/10.1186/1471-2229-13-212.
Hu, J., Zhang, H., Ding, Y., 2013. Identification of conserved microRNAs and their targets in the model legume *Lotus japonicus*. J. Biotechnol. 164, 520–524.
Huijser, P., Schmid, M., 2011. The control of developmental phase transitions in plants. Development 138, 4117–4129.
Jones-Rhoades, M.W., Bartel, D.P., 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14, 787–799.
Khraiwesh, B., Zhu, J.K., Zhu, J., 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim. Biophys. Acta 1819, 137–148.
Kidner, C.A., Martienssen, R.A., 2005. The developmental role of microRNA in plants. Curr. Opin. Plant Biol. 8, 38–44.
Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659.
Li, Y., Li, C., Ding, G., Jin, Y., 2011. Evolution of MIR159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. BMC Evol. Biol. http://dx.doi.org/10.1186/1471-2148-11-122.
Lu, Y., Yang, X., 2010. Computational identification of novel microRNAs and their targets in Vigna unguiculata. Comp. Funct. Genomics. http://dx.doi.org/10.1155/2010/128297.
Marks, M.D., Tian, L., Wenger, J.P., Omburo, S.N., Soto-Fuentes, W., He, J., Gang, D.R., Weiblen, G.D., Dixon, R.A., 2009. Identification of candidate genes affecting Δ9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J. Exp. Bot. 60, 3715–3726.
Mathieu, J., Yant, L.V., Murdter, F., Kuttner, F., Schmid, M., 2009. Repression of flowering by the miR172 Target SMZ. PLoS Biol. http://dx.doi.org/10.1371/journal.pbio.1000148.
Mikuriya, T.H., 1969. Historical aspects of *Cannabis sativa* in western medicine. New Physician 18, 902–908.
Millar, A.A., Gubler, F., 2005. The *Arabidopsis* GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. Plant Cell 17, 705–721.
Oliveira, J.S., Mendes, N.D., Carocha, V., Graça, C., Paiva, J.A., et al., 2013. A computational approach for microRNA identification in plants: combining genome-based predictions with RNA-seq data. J. Data Min. Genomics Proteomics. http://dx.doi.org/10.4172/2153-0602.1000130.
Rogers, K., Chen, X., 2013. Biogenesis, turnover, and mode of action of plant MICRORNAS. Plant Cell 25, 2383–2399.
Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108.
Simmonds, N.W., 1979. Hemp, *Cannabis sativa* (Moraceae). Evol Crop Plants 15, pp. 203–204.
Singh, J., Nagaraju, J., 2008. In silico prediction and characterization of microRNAs from red flour beetle (*Tribolium castaneum*). Insect Mol. Biol. 17, 427–436.
Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., Taura, F., 2004. The gene controlling marijuana psychoactivity. J. Biol. Chem. 279, 39767–39774.
Sunkar, R., Jagadeeswaran, G., 2008. In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol. http://dx.doi.org/10.1186/1471-2229-8-37.
Sunkar, R., Kapoor, A., Zhu, J.K., 2006. Post-transcriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by down regulation of miR398 and important for oxidative stress tolerance. Plant Cell 18, 2051–2065.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Wang, L., Mai, Y.X., Zhang, Y.C., Luo, Q., et al., 2010. MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in *Arabidopsis*. Mol. Plant 3, 794–806.

Wu, X.M., Liu, M.Y., Xu, Q., Guo, W.W., 2010. Identification and characterization of microRNAs from citrus expressed sequence tags. Tree Genet. Genomes. http://dx.doi.org/10.1007/s11295-010-0319-5.

Yang, T., Xue, L., An, L., 2007. Functional diversity of miRNA in plants. Plant Sci. 172, 423–432.

Zhang, L., Lu, Y.T., 2003. Calmodulin-binding protein kinases in plants. Trends Plant Sci. 8, 123–127.

Zhang, B., Pan, X., Anderson, A.T., 2006a. Identification of 188 conserved maize microRNAs and their targets. FEBS Lett. 580, 3753–3762.

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

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., Zhang, Jin, H., 2010. Multiple distinct small RNAs originate from the same microRNA precursors. Genome Biol. http://dx.doi.org/10.1186/gb-2010-11-8-r81.

Zhu, Q.H., Upadhyaya, N.M., Gubler, F., Helliwell, C.A., 2009. Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). BMC Plant Biol. http://dx.doi.org/10.1186/1471-2229-9-149.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.