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