Gene Reports 5 (2016) 62–69

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Profiling and annotation of microRNAs and their putative target genes in chilli (*Capsicum annuum* L.) using ESTs

Muhammad Din *, Muhammad Younas Khan Barozai, Iftekhar Ahmed Baloch

Department of Botany, University of Balochistan, Sariab Road, Quetta, Balochistan, Pakistan

ARTICLE INFO

Article history:
Received 16 June 2016
Accepted 30 August 2016
Available online 31 August 2016

Keywords:
miRNAs
Conserved nature
Post-transcriptional gene regulation
Comperative genomics

ABSTRACT

MicroRNAs (miRNAs) are short, endogenous and non-protein coding RNAs that are 18–26 nucleotides (nt) in length. The main function of the miRNAs is to negatively control the protein coding sequences at post-transcriptional level. By nature, majority of the miRNAs are evolutionarily conserved. This conserved nature provides the logic to predict the new miRNAs orthologus in a number of plant species. In plants, miRNAs are involved in almost all biological processes from growth and development to biotic and abiotic stresses with metabolism and cell signaling. In the current research, various computational tools were used for the profiling and characterization of new conserved miRNAs and their targets in chilli (*Capsicum annuum* L.). Consequently, a total of 88 miRNAs (belonging to 81 miRNA families) were discovered from the mining of 118,572 express sequence tags (ESTs). In this study, two miRNA genes were also found as pre-miRNA clusters (can-mir8019 and can-mir8036). In addition, nine miRNAs are randomly selected and experimentally validated through RT-PCR. Furthermore, a total of 409 protein targets were also predicted for these newly profiled miRNAs. These predicted protein targets were classified as: transcription factor, stress related, disease related, growth and development, hypothetical protein, transporters, signaling pathways, metabolism and structural protein. All of these newly identified miRNAs were reported in chilli for the first time. These results will serve as reference data to improve the regulation, management, and modification of this economically important plant at the molecular level. This will also help us to improve the chilli plant for production and biotic and abiotic stress tolerance in the near future.

1. Introduction

Chilli is the universal and important spice (Abrar et al., 2009). The mature fruit of chilli coloration is mostly due to the pigments of Capsanthin and Capsorubin, belong to carotenoids. These carotenoids are involved in the vitamin A activities which are useful for health (Ittah et al., 1993).

MicroRNAs (miRNAs) are a newly class of small, endogenous and non-coding RNA sequences ranging from 18 to 26 nucleotides (nt) long that control gene expression negatively at post-transcriptional level (Bartel, 2004). They carry out their regulatory functions either by degradation or by translation inhibition of targeted mRNAs (Filipowicz et al., 2008). The miRNAs are transcribed from miRNA genes by Polymerase II enzymes in plants as a longer primary transcripts (pri-miRNAs). The pri-miRNAs are then processed by Dicer-like 1 enzyme (DCL1) to form precursor miRNAs (pre-miRNAs) with stem-loop secondary structures. The pre-miRNAs are further chopped to a miRNAs duplex (miRANA:miRNA*) and this process is also catalysed by DCL1 (Bartel, 2004). The duplex is then transported into the cytoplasm by HASTY; orthologue of exportin 5 in plants (Park et al., 2002). In the cytoplasm, miRNAs are become unfold by helicase into mature miRNAs (Bartel, 2004). These mature miRNAs are then incorporated with Argonaute (AGO) proteins to form RNA-induced silencing complex (RISC) (Hammond et al., 2000), where they guided the RISC complex to negatively regulate targeted gene expression at post-transcriptional level (Bartel, 2004). Plant miRNAs play pivotal role in numerous biological processes viz; growth (Chen, 2003) organogenesis (Kidner and

Abbreviations: ath, *Arabidopsis thaliana*; ABC, ATP-binding cassette; bp, Base pair; BLAST, Basic local alignment search tool; can, *Capsicum annuum*; cDNA, Complementary DNA; dbEST, Database of EST; DCL1, Dicer-like Enzyme1; ESTs, Expressed Sequence Tags; FAD, Fatty acid desaturase; GC %, GC percentage; GO, Gene Ontology; GA, Gibberellin; gma, *Glycine* max; HSP, Heat shock factor; miR, Mature miRNA; MS, Mature sequence; MSA, Mature sequence arm; ML, Mature sequence length; Mtr, *Medicago truncatula*; MRNA, Messenger RNA; MIR, MicroRNA gene; miRNAs, MicroRNAs; 3′, 3 prime; 5′, 5 prime; MFE, Minimum Free Energy; MYB, Myeloblastosis; NCBI, National Center for Biotechnology Information; NBS-LRR, Nucleotide-binding site-leucine rich repeat; nt, Nucleotides; NM, Number of mismatches; miRNA*, Opposite strand of miRNA; OE, Organ of expression; Pi, Orthophosphate; psRNATarget, Plant small RNA target; *ptc*, *Populus trichocarpa*; PL, Precursor miRNA length; pre-miRNAs, Precursor of miRNAs; pri-miRNAs, Primary transcripts of mature miRNAs; refseq_rna, Reference mRNA sequence; RISC, RNA induced silencing complex; RT-PCR, Reverse Transcription-Polymerase Chain Reaction; sly, *Solanum lycopersicum*; sme, *Solanum melongena*; SE, Source EST; SL, Strand location; TFs, Transcription factors; *zma*, *Zea mays*.

* Corresponding author.
E-mail addresses: deen.2006@yahoo.com, drmdin75@gmail.com (M. Din).

http://dx.doi.org/10.1016/j.genrep.2016.08.010
2452-0144/

Martienssen, 2005), transgene suppression (Allen et al., 2005) signaling pathway (Yoshikawa et al., 2005), environmental stresses (Lu et al., 2012), diseases (Adai et al., 2005), stress responses (biotic and abiotic) (Sunkar and Zhu, 2004) and defense against the invading viruses (Bennasser et al., 2004) and phase transition from vegetative growth to reproductive growth (Chen, 2003).

The majority miRNAs from different plant species are observed with conserved nature, this suggests highly converged evolutionary behaviour for these tiny molecules. This property of convergence becomes a logical approach to identify new miRNAs in other plant species. Many researchers, based on this conserved nature of miRNAs, have identified a huge number of novel miRNAs using comparative genomic approaches in a wide range of plant species, including apricot (Baloch et al., 2015) tomato (Din and Barozai, 2014), potato (Din et al., 2014), Phaseolus (Barozai et al., 2013b), tobacco (Frazier et al., 2010), switchgrass (Xie et al., 2010), *Glycine* max (Zhang et al., 2008), cotton species (Zhang et al., 2007; Barozai et al., 2008), *Zea mays* (Zhang et al., 2006a) and *Arabidopsis thaliana* (Wang et al., 2004). These reports strongly suggest that comparative genomic strategies are valid, highly efficient, convenient, and economical-friendly methods to profile new potential miRNAs in species whose complete genomic data is not available. By the employment of bioinformatics approaches, there thousands of miRNAs have been identified both in plants and animals, submitted in miRBase (Release 21: June 2014) (Griffiths-Jones, 2004).

In spite of great importance of chilli, no miRNA is reported in miRBase. In this research, a total of 88 new miRNAs (from 81 miRNA families) in chilli using expressed sequence tags (ESTs) through the bioinformatics tools were identified and characterized. In this study, two miRNA genes were found as pre-miRNA clusters (can-mir8019 and can-mir8036). These newly identified miRNAs were also validated through their protein targets. A total of 409 protein targets were also predicted for these newly profiled miRNAs. These predicted protein targets were classified as; transcription factor, stress related, disease related, growth and development, hypothetical protein, transporters, signaling pathways, metabolism and structural protein. Furthermore, to confirm the expression of the stem loop precursor miRNAs, nine miRNAs are randomly selected and experimentally validated through RT-PCR.

2. Materials and methods

In order to use different bioinformatics algorithms to profile and characterize the new miRNA families in chilli, the detail steps are as follows:

2.1. Reference miRNAs fetching

Total 8749 known monocot and dicot plant precursor and mature miRNA sequences were downloaded as reference miRNAs from miRBase: miRNA Registry Database (Version Release 21: June 2014) (Griffiths-Jones, 2004). These reference miRNAs were subjected to profile the potential miRNAs from chilli (*Capsicum annuum* L.) expressed sequence tags (ESTs) using the same approach described by Barozai et al. (2008) with little modification as sketched in Fig. S1.

2.2. ESTs from EST database (dbEST)

The EST database (dbEST) is a sum of single-read cDNA sequences from Genbank. The aim of this database is to annotate genes, evaluate gene expression and to provide potential variation among the genes (http://www.ncbi.nlm.nih.gov/nucest/). The dbEST (database of EST) comprises the total entries of 74,186,692 ESTs from various plants and animals species (dbEST release 130101, 01 January 2013) at (http://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/). Chilli was observed with 1,18,572 ESTs in dbEST. Chilli ESTs were mined by applying comparative genomics approach with various bioinformatics tools using 8749 reference miRNAs for profiling and characterization of new miRNAs in chilli.

2.3. Profiling initial potential miRNAs

In this important step, the known precursor and mature plant miRNAs sequences (8749) retrieved from miRBase were used as reference miRNAs. These reference miRNAs were subjected to Basic Local Alignment Search Tool (BLAST) for alignment against publicly available chilli (1,18,572) ESTs from the dbEST, using BLASTn program (Altschul et al., 1990). Adjusted BLAST parameter settings were as follows: expect values were set at 1000; low complexity was chosen as the sequence filter, database (others. Chilli) program selection (somewhat similar sequence) and all other parameters were used as default. The FASTA formats of all the initial potential candidates' sequences having 0–4 mismatches with the reference miRNAs were saved.

2.4. Creation of single tone EST

After the profiling of initial potential miRNAs, the next step is going to create the single tone EST per miRNA. In this crucial step, The ESTs created from the same mRNA were found by BLAST against the chilli ESTs Database, using BLASTn program (Altschul et al., 1990). Adjusted BLAST parameter settings were as follows: expect values were set at 1000; low complexity was chosen as the sequence filter, database (others. Chilli) program selection (somewhat similar sequence) and all other parameters were used as default. The repeated ESTs for the same miRNAs were discarded and single tone EST for each miRNA was created and saved for downstream analysis.

2.5. Removal of protein coding sequences through BLASTx

The validation of the newly identified potential miRNAs as a non-coding RNA, is very significant criterion (Barozai et al., 2014). To assure these newly predicted miRNAs in chilli as a non-protein coding, the initial candidate sequences were subjected through BLAST against the protein database at NCBI (http:/www.ncbi.nlm.nih.gov) using blastx with default parameters (Altschul et al., 1997). All the protein coding sequences were discarded. Only potential candidate miRNA sequences which showing non-significant similarity with protein sequences, at protein database, were carried out for the next step of profiling and characterization of newly identified miRNAs in chilli.

2.6. Generation of stem-loop secondary structures

In aforesaid steps the initial potential candidates of chilli miRNAs fulfilled the criteria of single tone sequences, having 0–4 mismatches with the reference miRNAs and non-protein coding nature, were subjected to secondary structures generation. For this purpose, publicly available Zuker folding algorithm publically available at (http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi), known as MFOLD (version 3.6) (Zuker, 2003), was applied to predict the secondary structures for these potential candidate miRNAs. The MFOLD parameters were adjusted as RNA sequence (linear), folding temperature (37 °C), ionic condition (1 M NaCl with no divalent ions), percent sub-optimality number (5); maximum interior/bulges loop size (30), and all others with defaults values. The lowest free energy structures were selected for manual inspection, as described by Barozai et al. (2008). The threshold values used to select a miRNA were same as described by Zhang et al. (2006b). The stem regions of the stem-loop structures were checked and confirmed for the mature sequences with either at least 16 or equal to the reference miRNAs base pairing involved in Watson-Crick or G/U base pairing between the mature miRNA and the opposite strand (miRNA*).