Gene 552 (2014) 277–282

Contents lists available at ScienceDirect

Gene

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

Short communication

In-silico identification of miRNAs and their regulating target functions in *Ocimum basilicum*

Noopur Singh *, Ashok Sharma

Biotechnology Division, CSIR—Central Institute of Medicinal and Aromatic Plants, P.O. CIMAP, Lucknow 226015, UP, India

ARTICLE INFO

Article history:
Received 27 March 2014
Received in revised form 29 August 2014
Accepted 18 September 2014
Available online 23 September 2014

Keywords:
miRNA
miRNA-targets
Secondary metabolic pathway
Plant development
Stress
Ocimum basilicum

ABSTRACT

microRNA is known to play an important role in growth and development of the plants and also in environmental stress. *Ocimum basilicum* (Basil) is a well known herb for its medicinal properties. In this study, we used in-silico approaches to identify miRNAs and their targets regulating different functions in *O. basilicum* using EST approach. Additionally, functional annotation, gene ontology and pathway analysis of identified target transcripts were also done. Seven miRNA families were identified. Meaningful regulations of target transcript by identified miRNAs were computationally evaluated. Four miRNA families have been reported by us for the first time from the Lamiaceae. Our results further confirmed that uracil was the predominant base in the first positions of identified mature miRNA sequence, while adenine and uracil were predominant in pre-miRNA sequences. Phylogenetic analysis was carried out to determine the relation between *O. basilicum* and other plant pre-miRNAs. Thirteen potential targets were evaluated for 4 miRNA families. Majority of the identified target transcripts regulated by miRNAs showed response to stress. miRNA 5021 was also indicated for playing an important role in the amino acid metabolism and co-factor metabolism in this plant. To the best of our knowledge this is the first in silico study describing miRNAs and their regulation in different metabolic pathways of *O. basilicum*.

1. Introduction

Diverse group of regulatory events is controlled by non-coding RNAs. microRNAs (miRNAs) are 21–25-nucleotide small non-coding RNAs controlling the gene expression at the posttranscriptional level through transcriptional cleavage or translational repression. The presence of terpenoids, alkaloids, and flavonoids in the *Ocimum basilicum* supports for its enormous pharmaceutical potential and medicinal use. MEP/terpenoid and shikimate/phenylpropanoid pathways point to multiple levels of metabolic control in sweet basil glandular trichomes. No work has been reported on identifying microRNAs (miRNAs) and their targets in this medicinal plant. With concern to medicinal properties a deep understanding of the regulatory system governed by the miRNAs is required in this important medicinal plant.

Numerous reports are available on identification of miRNAs in plants through computational and experimental approaches (Griffiths-Jones et al., 2006; Zhang et al., 2006a). High cost and lower expression of the gene in the experimental techniques of the miRNA identification promote the computational screening of potential miRNAs. It has also been successful for the discovery of conserved miRNAs as well as new miRNAs (Jones-Rhoades and Bartel, 2004; Zhang et al., 2006a). Genome sequence and expressed sequence tags (ESTs) are the key elements used in these approaches (Zhang et al., 2005, 2006a; Xie et al., 2007). Since majority of known mature miRNAs are conserved across plants species, it is also logical to perform the bioinformatics search for putative miRNAs (J.F. Wang et al., 2004; X.J. Wang et al., 2004). A series of computational programs have been successfully developed for prediction of miRNA gene in *Arabidopsis*, rice and other plant species (Bonnet et al., 2004a; Jones-Rhoades and Bartel, 2004; J.F. Wang et al., 2004; X.J. Wang et al., 2004; Adai et al., 2005). Majority of bioinformatics tool available for miRNA prediction requires genome sequence or next generation sequencing data of plants (Wu et al., 2011; Yang and Li, 2011; Xie et al., 2012; Jha and Shankar, 2013). Deficiency of the genome sequences works as a restrictive aspect to predict miRNAs and targets in medicinal plants. However, availability of EST sequences can also be utilized to identify miRNA and their target in medicinal plants. No data on genomic sequence or next generation sequencing is yet available for *O. basilicum*. As ESTs are available, this study was initiated with the EST data set. ESTs have also been used earlier for miRNA identification in several previous studies (Zhang et al., 2005; Unver et al., 2010; Barozai et al., 2012; Catalano et al., 2012). Comprehensive plant miRNA target analysis is still limited due to limited ability of the available tools in predicting translational inhibition and integrating transcriptome data. To overcome this problem, different types of tool at each step of this study were used with the aim to identify possible miRNA mediated regulation in metabolites pathways.

Abbreviations: miRNA, microRNA; MFE, minimal folding free energy; MFEI, minimal folding free energy index.

* Corresponding author.
E-mail address: singh.rajpoot.noopur@gmail.com (N. Singh).

http://dx.doi.org/10.1016/j.gene.2014.09.040
0378-1119/

2. Materials and methods

2.1. Data retrieval and software employed

A set of reported 23,260 ESTs of *O. basilicum* were downloaded from the NCBI (www.ncbi.nlm.nih.gov/nucest). To remove the redundancy from the ESTs sequences CAP3 (Huang and Madan, 1999) was used. miRNA and their target prediction was performed by using a comprehensive tool C-mii version 1.11 (Numnark et al., 2012) from 3436 contigs. Nine sequences were excluded according to the criteria of miRNA identification module (length exceeding limit 3000 bps) of C-mii. Online web sever psRNATarget (Dai and Zhao, 2011) and Target-align (Xie and Zhang, 2010) were used to evaluate the prediction results of C-mii. Blast2GO (Conesa et al., 2005; Conesa and Gotz, 2008) was used for annotation of identified target transcripts.

2.2. miRNA and its target prediction

miRNA identification module of C-mii used in this study uses homology search based approach. Primary miRNA folding and precursor miRNA folding, sub-modules were used with their default parameters. To process the predicted results, stability of the secondary structure of the pre-miRNAs needs to be evaluated. A contig was considered as miRNA candidate if it fits in the following criteria; 1) length of predicted miRNAs should be in range of 19–25 nucleotides; 2) more than three substitutions were not allowed for the predicted mature miRNAs against with a known miRNAs; 3) one arm localization of the mature miRNA within stem–loop structure; 4) maximum 6 mismatches were allowed between miRNA sequence and miRNA* sequence of stem–loop structure; 5) A + U content should be high compared to G + C content; and 6) high negative minimal folding free energy (MFE) and high MFE index (MFEI) value of the secondary structure.

Predicted miRNAs were then used for target search against all contig sequences of *O. basilicum*. Target identification module of C-mii is based on perfect or near-perfect complementarities of plant miRNA and its target. Target scanning was performed to search complementary site of predicted miRNAs for all contigs. The following criteria were set for the prediction of miRNA-target genes: 1) maximum three mismatches between predicted mRNAs and target gene; 2) no mismatches were allowed for 10th and 11th positions of complementary site; 3) MFE of miRNA and target duplex should be negative; and 4) no more than 4 GU pair was allowed in the complimentary alignment.

2.3. Validation

Least number of allowed mismatches for the alignment of putative miRNAs and known miRNAs is the foremost feature to get accurate results. In-silico validation was done only for miRNAs having targets in predicted results. User submitted small RNAs and user submitted transcript section of psRNATaget were used in this study. The same alignment approach is also used by Target-align. During validation of predicted result by Target-align 3 mismatches were allowed from base 1 to 9. The following criteria were taken into account for the validation of miRNA and its targets. 1) Only 3 non-consecutive or consecutive mismatches were allowed; all aligned pairs having more than 3 mismatches by both validating tool were filtered; 2) The sequences having aligned section were considered as final results; any predicted miRNA or transcripts showing no alignment by both tools were ignored; 3) to improve the accuracy of the results only cleavage type of inhibition were considered by psRNATarget; and 4) significance of the predicted results with at least one validation tool was considered for the study.

Evaluated target transcripts were selected for functional annotation. BlastX was performed against plant/*Arabidopsis thaliana* protein database. GO and pathway analyses were performed with its analyzed default parameters, which also evaluates C-mii annotated results.

Note — miRNA* complementary strand of the guide mRNA strand.

2.4. Phylogenetic analysis

To determine the phylogeny of the identified miRNAs of *O. basilicum*, pre-miRNA sequence of all available plants was obtained from miRBase (version 20, June 2013) (Griffiths-Jones et al., 2008). Sequences were aligned using MUSCLE and neighbor-joining (NJ) phylogenies were generated by MEGA 6.0 (Tamura et al., 2013) with its default parameters.

3. Results and discussion

3.1. miRNA prediction

After carefully considering the homology and secondary structure prediction results, 9 sequences were selected as miRNA candidates. These predicted sequence belonged to 7 different families (Table 1). Evolutionary conserved miRNA families such as miR160 and miR164 were identified in this study. The distribution of identified miRNAs in the investigation was not uniform. Majority of the miRNA families identified had only one member. In the case of miR160 three members were identified from the same transcript (Table 1), only one miRNA candidate was considered for the study. The secondary structures of predicted miRNAs were showed in Supplementary data 6.

3.1.1. Characterization of predicted pre-miRNAs

3.1.1.1. Length variation. Majority of plus strands was observed for predicted miRNAs. Mature miRNA sequences showed variations up to 20–21 nucleotides, where the 20 nt length was occupied by few miRNA family. As compared to predicted mature miRNAs length, precursor miRNAs showed large variation in their length (Table 1). Large variations in the length of pre-miRNAs were reported in earlier studies also (Barozai et al., 2012; Wang et al., 2012; Patanun et al., 2013).

3.1.1.2. GC content. Pairing of three hydrogen bonds between G and C contributes to the formation and stabilization of the secondary structure of stem–loop hairpins. With this logic, the stability of secondary structure of RNA should contain high GC content in the sequence. In this study, predicted miRNA family 160 alone showed 40% of all GC contents in their pre-miRNA sequence. In this study, predicted conserved miRNA family miR160 and miR164 showed majority of GC content in their pre-miRNA sequences. The range of GC varied from 30.71 to 59.75 (Table 1). Most conserved identified miRNA families 160 and 164 were GC rich as reported earlier in *Helianthus* and *Nicotiana tabacum* (Frazier et al., 2010; Barozai et al., 2012), supporting the notion that conserved miRNAs may tend to have high GC content.

Compared to GC content AU content was high and varied from 40.00 to 69.28 (Table 1). Uracil was dominant in the first position of mature miRNA sequence (Table 1), suggesting its important role in miRNA mediated regulation in plants (Zhang et al., 2008; Unver et al., 2010; Dhandapani et al., 2011; Luo et al., 2013). In the case of pre-miRNA, predominance of Adenine and Uracil was observed, which is in accordance to the earlier report in *Gossypium arboretum* L. and *Brassica rapa* L. (Dhandapani et al., 2011; Wang et al., 2012).

3.1.1.3. MFE and MFEI. MFE is another criterion for measuring the stability of a RNA or secondary structure. It is reported that precursor-microRNAs have lower folding energies than other non-coding RNAs (Bonnet et al., 2004b). The MFE of the 9 predicted pre-miRNAs varied from 09.20 to 81.34 (−kcal/mol). Due to the variation in the length of precursor miRNAs, it is not enough to characterize miRNA on the basis of MFE. The MFEI resolution for length of variation was also calculated in order to distinguish miRNA from RNAs (Zhang et al., 2006b). The range of MFEI of predicted pre-miRNAs in *O. basilicum* varied from 0.34 to 0.81 (−kcal/mol).