Gene 537 (2014) 333–342

Contents lists available at ScienceDirect

Gene

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

Short Communication

Computational identification and characterization of conserved miRNAs and their target genes in garlic (*Allium sativum* L.) expressed sequence tags

Debashis Panda, Budheswar Dehury, Jagajjit Sahu, Madhumita Barooah, Priyabrata Sen, Mahendra K. Modi *

Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, Assam, India

ARTICLE INFO

Article history:
Accepted 4 January 2014
Available online 13 January 2014

Keywords:
miRNA
Garlic
Comparative genomics
Expressed sequence tag (EST)
psRNAtarget server

ABSTRACT

The endogenous small non-coding functional microRNAs (miRNAs) are short in size, range from ~21 to 24 nucleotides in length, play a pivotal role in gene expression in plants and animals by silencing genes either by destructing or blocking of translation of homologous mRNA. Although various high-throughput, time consuming and expensive techniques like forward genetics and direct cloning are employed to detect miRNAs in plants but comparative genomics complemented with novel bioinformatic tools pave the way for efficient and cost-effective identification of miRNAs through homologous sequence search with previously known miRNAs. In this study, an attempt was made to identify and characterize conserved miRNAs in garlic expressed sequence tags (ESTs) through computational means. For identification of novel miRNAs in garlic, a total 3227 known mature miRNAs of plant kingdom Viridiplantae were searched for homology against 21,637 EST sequences resulting in identification of 6 potential miRNA candidates belonging to 6 different miRNA families. The psRNATarget server predicted 33 potential target genes and their probable functions for the six identified miRNA families in garlic. Most of the garlic miRNA target genes seem to encode transcription factors as well as genes involved in stress response, metabolism, plant growth and development. The results from the present study will shed more light on the understanding of molecular mechanisms of miRNA in garlic which may aid in the development of novel and precise techniques to understand some post-transcriptional gene silencing mechanism in response to stress tolerance.

1. Introduction

Among the various types of RNAs the microRNAs (miRNA) are small, endogenous, non-coding and regulatory RNAs. These small miRNA (~21–24 nucleotides long) molecules generally originate from long self-complementary precursor sequences (Bartel, 2004). The long primary precursor miRNAs are first transcribed by RNA polymerase II and then processed into hairpin stem loop structures to form the precursor miRNAs (pre-miRNAs). Further the loop regions of pre-miRNAs are cut into small, double-stranded RNAs using a dicer-like-enzyme (DCL1) in plants (Kurihara and Watanabe, 2004). Finally these mature miRNA molecules are incorporated into an RNA-induced silencing (RISC) complex that negatively controls the target gene expression either by inhibition of translation elongation process or by destruction of the messenger RNA (mRNA) on the basis of the degree of complementarity of miRNA within the target mRNA (Aukerman and Sakai, 2003; Bartel, 2004; Tang et al., 2003; Voinnet, 2009). Generally the miRNA target sites are present at the 3′ un-translated regions (UTRs) of the mRNAs (Carrington and Ambros, 2003). When the degree of complementarity of miRNA within its target mRNA is weak, gene expression is suppressed by blocking the translation of the mRNA (Carrington and Ambros, 2003; Novina and Sharp, 2004). However, when the mRNA target has a single and perfect or nearly perfect miRNA complementary site, miRNA triggers mRNA degradation (Kidner and Martienssen, 2005).

MiRNAs play a pivotal role in regulation of gene expression at the post-transcriptional levels and execute multipurpose functions in plants. It includes the development of plant (Yang et al., 2007), hormone signaling, response to biotic and abiotic stresses (Frazier et al., 2011; Jagadeeswaran et al., 2009; Phillips et al., 2007; Zhang et al., 2006a), signal transduction, protein degradation (Zhang et al., 2008; Zhou et al., 2010), transgene suppression (Allen et al., 2005), disease development (Johnson et al., 2005) and defense against invading viruses (Bennasser et al., 2004).

During recent years the identification and characterization of miRNA (Zhang et al., 2006a) and their target genes from plants has been extensively studied (Jones-Rhoades and Bartel, 2004; Unver et al., 2009).

Abbreviations: EST, expressed sequence tag; GSS, genome survey sequence; miRNA, microRNA; psRNATarget, plant small RNA target analysis server; MFE, minimal folding free energy; MFEI, minimal folding free energy index; qRT-PCR, quantitative real-time polymerase chain reaction.

* Corresponding author. Tel.: +91 376 2340095 (O); fax: +91 376 2340001, 2340101.
E-mail addresses: mkmodi@gmail.com, mkmodi@aau.ac.in (M.K. Modi).

0378-1119/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2014.01.010

Several high-throughput techniques for experimental determination of miRNAs *viz*., direct cloning and deep sequencing has been used predominantly along with homology based comparative genome analysis from expressed sequence tag (EST) and genome survey sequences (GSS). A large number of miRNAs are evolutionarily conserved in the plant kingdom, which ranges from mosses and ferns to higher flowering plants (Zhang et al., 2006b). This attribute has been used as a practical indicator for the identification and prediction of miRNAs by homology searches in other species. Identification miRNAs using EST analysis has some advantages over other methods (Zhang et al., 2008). It has been suggested that most of the miRNAs predicted by EST analysis can be recovered by high throughput deep sequencing (Kwak et al., 2009).

Besides producing a plethora of biologically active primary metabolites, medicinal aromatic plants are also able to accumulate secondary metabolites naturally. Among the secondary metabolites, essential oils are volatile substances with characteristic odor, and considered most potential source of natural bio-molecules responsible particularly for treatment of infectious diseases (Amagase et al., 2001; Avato et al., 2000 and Harris et al., 2001). *Allium sativum* L. a monocot plant commonly known as garlic belongs to the economically important family Liliaceae. Garlic belongs to the large and diverse genus onion (*Allium*), which includes more than 1250 species. Garlic has been used throughout recorded history for culinary and medicinal use and health benefits (Banerjee et al., 2003). Interest in garlic has dramatically increased in recent years due to its nutritional and pharmaceutical value. It has tremendous impact on controlling high blood pressure and cholesterol (Yeh and Yeh, 1994), shows anti-inflammatory (Baek et al., 2001), antioxidant, and antiviral activity (Galana and Marquez, 2009) and used immensely for cancer treatment (Kaschula et al., 2010; Unnikrishnan and Kuttan, 1990). The rekindling of interest in garlic is due to recent revelations of its beneficial effects in the treatment of various human and animal diseases and also due to the availability of more than 2000 publications on different aspects of garlic which validates the claims made in traditional systems of medicine.

On the basis of sequence conservation between plant species, computational approaches yield reasonable results in the identification of miRNAs and their structure prediction (Floyd and Bowman, 2004; Zhang et al., 2006b). For a species whose genome sequences are not yet completely mapped, sequence databases available for GSSs, ESTs and bacterial artificial clones (BAC) in the public domain can provide adequate resources to identify the conserved miRNAs (Dhandapani et al., 2011; Xie et al., 2007, 2010; Zhang et al., 2006a).

Recently, genome-wide comparative analysis was used to identify the conserved miRNAs in various plant species including cotton (Zhang et al., 2007), soybean (Zhang et al., 2008), wheat (Jin et al., 2008), potato (Xie et al., 2011), apple (Gleave et al., 2008; Yu et al., 2011), Solanaceae (Kim et al., 2011), switchgrass (Xie et al., 2010), citrus (Song et al., 2010), Chinese cabbage (Wang et al., 2011) and *Brassica rapa* L. (Dhandapani et al., 2011). According to the latest release of miRBase (Release 20: June 2013), a total of 7385 miRNAs from 72 different plant species belonging to kingdom Viridiplantae have been identified and deposited.

Although a plethora of publications is available on the pharmacological properties of garlic and their beneficial health effects, there are no reports on comparative genomic studies on garlic. To date, no miRNAs within the Liliaceae family have been identified. Considering the medicinal value and economic importance of garlic, we used computational comparative genomics approach using EST sequences to identify potentially conserved miRNAs and their putative target genes. In this study, we performed a computational analysis and a homolog search of 3227 unique known miRNAs from Viridiplantae, against all EST sequences (21,637) of garlic available in dbEST of NCBI incorporating BLAST search. From 23 pre-miRNA predicted through Blast search, 6 potential miRNAs were identified based on their folding characteristics and minimal folding free energy (MFE). All these pre-miRNAs exhibit stable MFE and form stem loop structures. The potential target genes of new miRNAs of garlic were also identified and their putative functions were investigated to understand the roles of miRNAs in stress response, physiology, growth and development of garlic.

2. Materials and methods

2.1. Sequence database and reference miRNAs

A mature miRNA set comprising of 5939 known miRNAs belonging to Viridiplantae (miRBase Release 19) was downloaded from the publicly available miRBase (http://www.mirbase.org/cgi-bin/browse.pl) database (Griffiths-Jones et al., 2008). Out of the 5939 known miRNAs, only 3227 unique (non-redundant) miRNAs were selected as reference miRNA for conserved miRNA search in garlic. As comparative-genome based miRNA prediction relies on EST or GSS, a total of 21,637 ESTs of garlic were downloaded from dbEST (http://www.ncbi.nlm.nih.gov/nucest) of the NCBI (Release 130101, 01 January 2013) and used for miRNA prediction.

2.2. Computational resources

The alignment tool BLAST version 2.2.27 (Altschul et al., 1990) which was used for conserved miRNA prediction of garlic ESTs was downloaded from the NCBI website (http://ftp.ncbi.nih.gov/blast/executables/blast+). Assembly of ESTs was done using CAP3 (Huang and Madan, 1999) and it was downloaded from (http://seq.cs.iastate.edu/cap3.html), whereas the secondary structures of pre-miRNAs were performed through online version of MFOLD available at (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) (Zuker, 2003). The target genes for the putative miRNAs of garlic were identified using plant small RNA (psRNA) target server (http://plantgrn.noble.org/psRNATarget/) (Dai and Zhao, 2011).

2.3. Identification of the conserved miRNAs via BLASTn search

Redundancies in the ESTs of garlic were removed by assembling the 21,637 ESTs into contigs and singletons respectively via locally installed software CAP3, with default parameters. These assembled sequences (contigs and singletons) were further used for homolog search against the known unique mature plant miRNA sequences. The pairwise alignment of assembled transcripts against known miRNAs was achieved using BLASTn algorithm with a threshold of E value at 10. Low complexity was chosen as the sequence filter, the word-match size between the query and database was kept at 7. The following rules were considered for selecting the candidate miRNA from the homolog search:

1. The candidate miRNA should contain at least 18 nt length with no gap in between.
2. The number of mismatches between the known miRNAs and assembled EST sequences (contigs and singletons) was taken as less than three. The contigs and singleton sequences which closely matched (n/n, n-1/n and n-2/n nucleotide matches, whereas n represents known miRNA length) with only the known miRNAs selected for further study.

After selecting the candidate miRNAs by following the stringent criteria mentioned above, the protein coding sequences within the candidate miRNAs were discarded by BLASTx search against non-redundant (NR) protein database. The remaining precursor sequences of potential miRNA homologs were assessed for secondary structure using the Zuker folding algorithm in MFOLD software. The following parameters were used for secondary structure prediction in MFOLD: linear RNA sequence; folding temperature fixed at 37 ºC; ionic conditions of 1 M NaCl without divalent ions; percent sub-optimality number of 5; maximum interior/bulge loop size of 30; energy dot plot was turned on and the other parameters were set as default. After prediction of the secondary structure of the precursor sequence of potential miRNA