Computational Biology and Chemistry 59 (2015) 131–141

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research article

Computational exploration of microRNAs from expressed sequence tags of *Humulus lupulus*, target predictions and expression analysis

Ajay Kumar Mishra, Ganesh Selvaraj Duraisamy, Anna Týcová, Jaroslav Matoušek*

Biology Centre ASCR, v.v.i, Department of Molecular Genetics, Institute of Plant Molecular Biology, Branišovská 31, České Budějovice 37005, Czech Republic

ARTICLE INFO

Article history:
Received 4 May 2015
Received in revised form 2 September 2015
Accepted 7 September 2015
Available online 9 September 2015

Keywords:
microRNA
Humulus lupulus
Comparative genomics
Posttranscriptional gene regulation
Expressed sequence tags

ABSTRACT

Among computationally predicted and experimentally validated plant miRNAs, several are conserved across species boundaries in the plant kingdom. In this study, a combined experimental-*in silico* computational based approach was adopted for the identification and characterization of miRNAs in *Humulus lupulus* (hop), which is widely cultivated for use by the brewing industry and apart from, used as a medicinal herb. A total of 22 miRNAs belonging to 17 miRNA families were identified in hop following comparative computational approach and EST-based homology search according to a series of filtering criteria. Selected miRNAs were validated by end-point PCR and quantitative reverse transcription-polymerase chain reaction (qRT-PCR), confirmed the existence of conserved miRNAs in hop. Based on the characteristic that miRNAs exhibit perfect or nearly perfect complementarity with their targeted mRNA sequences, a total of 47 potential miRNA targets were identified in hop. Strikingly, the majority of predicted targets were belong to transcriptional factors which could regulate hop growth and development, including leaf, root and even cone development. Moreover, the identified miRNAs may also be involved in other cellular and metabolic processes, such as stress response, signal transduction, and other physiological processes. The *cis*-regulatory elements relevant to biotic and abiotic stress, plant hormone response, flavonoid biosynthesis were identified in the promoter regions of those miRNA genes. Overall, findings from this study will accelerate the way for further researches of miRNAs, their functions in hop and shows a path for the prediction and analysis of miRNAs to those species whose genomes are not available.

1. Introduction

Hop (*Humulus lupulus* L., Cannabaceae) is a dioecious, rhizomatous, twining perennial flowering plant, native to Europe, western Asia and North America. Lupulin glands of hop cones (female inflorescences) are composed of biosynthetic cells that secret a specific but complex metabolites consisting mainly of terpenophenolics (hop bitter acids and prenylflavonoids) and terpenoids (essential oil components), which serve essential raw ingredient in beer, contributing the distinctive bitterness, flavour and aroma, as well as preservative activity (Matoušek et al., 2006). In addition, hop has been traditionally used for variety of health benefits that lately has been confirmed through research. Among these health benefits from hops include relaxation and sleep inducer, anti-inflammatory effect, estrogenic effect, antioxidant activity and as an anti-tumor properties (Van Cleemput et al., 2009). Until recently limited progress has been made in basic gene discovery and gene regulation in hop, and none of the miRNAs have been reported. Understanding gene function and their regulatory mechanisms could provide new insight into the mechanisms of hop growth and development, the response to environmental abiotic/biotic stresses and particularly on cones initiation and development.

MicroRNAs (miRNAs) are a subset of endogenous approximate 22 nucleotide (nt) small non-coding regulatory RNA molecules that regulate gene expression post-transcriptionally by mediating mRNA degradation or translational repression in a sequence specific manner (Filipowicz et al., 2008). In plants, most miRNA genes are intergenic and transcribed individually from their own region, but a few genes are organized into polycistronic transcription units and co-transcribed from a single promoter at the end of a miRNA gene cluster (Voinnet, 2009). Like other RNAs and protein-coding genes, miRNA genes are transcribed by RNA polymerase II (Pol II) in which the general transcriptional co-activator and mediator are used to recruit Pol II to miRNA gene promoters for transcription initiation (Kim et al., 2011). Increasing evidence suggests that miRNAs play pivotal roles in multiple

* Corresponding author. Fax: +420 385310356.
E-mail address: jmat@umbr.cas.cz (J. Matoušek).

http://dx.doi.org/10.1016/j.compbiolchem.2015.09.005
1476-9271/

biological processes in plants, especially controlling tissue (leaf, root, stem, and flower) differentiation and development, self-negative feedback regulation of metabolism, phase transition from vegetative growth to reproductive growth. signal transduction and response to environmental conditions such as biotic/abiotic stresses (Lakhotia et al., 2014), including nutrient stresses, for example, phosphorus starvation (Gu et al., 2010), nitrogen starvation (Zhao et al., 2011), and micronutrient deficiency or toxicity (Valdes-Lopez et al., 2010). Considering the importance of miRNAs several approaches have been established for identifying miRNAs in various plant species, such as, direct cloning, small RNA (sRNA) high-throughput sequencing by next generation sequencer (NGS) and computational prediction. Since some miRNAs are expressed at a low level and the expression of many miRNAs has spatio-temporal specificity, it is difficult to clone low-abundance miRNAs through direct cloning approach (Llave et al., 2002). High-throughput sequencing of sRNA has great promise to generate an accurate and comprehensive picture of the sRNA transcriptome in different plants, tissues, and at different developmental stages. Meanwhile, such a tool provides the direct way of discovering miRNAs as well as to quantify them, some of which facilitates the investigation of sRNA populations in economically important species which lack complete genome information. It is a high throughput and simple method, but costly for miRNA discovery (Martin et al., 2013).

A comparative genomics study across phylogenetically distant taxa has shown that many miRNAs are highly evolutionarily conserved in the plant and animal kingdom (Axtell et al., 2006). This feature of extensive evolutionary conservation of these miRNAs among themselves renders a powerful approach to their identification using computational based approaches. On the basis of this strategy, researchers developed an expressed sequence tag (EST) and a genome survey sequence (GSS) approach to identify miRNAs (Zhang et al., 2005) from plants or even from various animals. EST analysis has some substantial advantages over the other approaches such as: (1) conserved miRNAs can be identified without whole genome sequences, (2) direct evidence for miRNA expression could be obtained without genomic sequence surveys, and (3) miRNA identification can be conducted without highly specialized software. Based on this approach, thousands of miRNAs have been systematically identified in more than 42 plant species, including many important plant and agricultural crop species, such as maize (Zhang et al., 2006a), wheat (Han et al., 2009), soybean (Chen et al., 2009), tobacco (Frazier et al., 2010), potato (Xie et al., 2011) and Asiatic cotton (Wang et al., 2012).

In this study, we employed a comparative genome-based homology search to identify miRNAs in hop using their current expressed sequence tags (ESTs) available in the NCBI GenBank database. The new miRNAs identified in this study would enable investigation of the complexity of miRNA-mediated genes network in understanding their roles in regulating growth, development, metabolism, and other physiological processes such as cone development and transcriptional factors (TFs) involved in secondary metabolic pathway including a prenylflavonid, xanthohumol which has various useful biological activity (Matoušek et al., 2006).

2. Materials and methods

2.1. Sequences and software

The known plant miRNA sequences from *Arabidopsis*, *Brassica*, *Glycine*, *Saccharum*, *Sorghum*, *Vitis*, *Solanum*, *Oryza*, *Triticum*, *Chlamydomonas*, and other plant species were downloaded from the miRNA database miRBase (http://www.mirbase.org/) (miRBase version 21.0, accessed June 2014) and plant microRNA database (http://bioinformatics.cau.edu.cn/PMRD/). After removing redundant sequences, the unique sequences were used as reference miRNA set for identification of hop miRNAs using Blastn searches. Hop expressed sequence tags (ESTs), cDNAs, and mRNAs were downloaded from the GenBank nucleotide databases from the National Centre for Biotechnology Information (NCBI). Currently, a total of 25,692 hop ESTs are available in the NCBI EST database (dbEST release 130101, January 1, 2013; http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). Comparative software mpiBLAST-1.6.0 (Darling et al., 2003) consist of core algorithm similar to BLAST-2.2.22 of NCBI GenBank (Altschul et al., 1997) was downloaded and set up locally. To predict the secondary structure of pre-miRNA, Zuker folding algorithm software MFOLD 3.2 (http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form) was used online to analyze the secondary structure and calculating the minimum free energy (Zuker, 2003).

2.2. Identification of potential miRNAs in hop

The mature sequences of all currently available plant miRNAs after removing redundant sequence were subjected to a Blastn search against all of the currently available hop EST sequences using mpiBLAST-1.6.0 algorithm in Linux based 32 core cluster system. Adjusted blast parameter settings were as follows: (1) the default word-match size between query and database sequences was set at seven; (2) the expected values were set at 1000 to increase the hit chance for more potential sequences and (3) the sequence number of the BLASTn search and the sequence alignments were set to 1000. If only a partial reference mature miRNA sequence was aligned to a EST sequence, the non-aligned parts were manually inspected and compared to determine the number of matching nucleotides. These sequences were considered as potential miRNA candidates only if they fullfil the following criteria with slight modification from a previous study (Zhang et al., 2007): (1) at least 18–22 nt length was assumed between the predicted mature miRNAs, and (2) allowed to have 0–4 nt mismatches in sequence with all previously known plant mature miRNAs. These whole EST sequences were used for BLAST*x* analysis using NCBI BLAST+2.2.29 against NCBI nr (non-redundant protein) database for removing the protein-coding sequences and retaining only the non-protein coding sequences (Altschul et al., 1997). In total, 47 EST sequences were found to be non-protein coding sequences.

The secondary structure of candidate pre-miRNA sequences of these potential miRNA homologs was predicted using the Zuker folding algorithm with MFOLD-3.2 (Zuker, 2003). All parameters were set to default values. All outputs obtained from mfold were recorded into an excel file, which included the EST ID numbers, respective miRNA homologs, total length of the sequences, the number of each nucleotide (A, G, C and U), the number of arms per structure, location of the matching regions, percentage (%) of (A + U/T) and (G + C) content and minimal folding free energy (MFE, DG in kcal/mol). Then, the adjusted minimal folding free energy (AMFE) [AMFE = (MFE/length of a potential pre-miRNA) × 100] and the minimal folding free energy index (MFEI) [MFEI = AMFE/(G + C)%, where (G + C)% represents% GC content over pre-miRNA sequence] were calculated according to a previous report (Zhang et al., 2006b). An EST was considered a miRNA candidate when it fit all of the following criteria: (1) the predicted mature miRNAs had no more than three nucleotide substitutions compared with a known mature miRNAs; (2) the EST sequence could fold into an appropriate stem-loop hairpin secondary structure; (3) the mature miRNA was localized in one arm of the stem-loop structure; (4) there was no loop or break in the miRNA or miRNA* sequences; (5) there were no more than 6 mismatches between the predicted mature miRNA sequence and its opposite miRNA* sequence in the secondary structure; and (6) the predicted secondary structure had