Gene 531 (2013) 15–22

Contents lists available at ScienceDirect

Gene

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

Identification of abiotic stress miRNA transcription factor binding motifs (TFBMs) in rice

S.J.S. Rama Devi [a], M.S. Madhav [a,*], G. Ram Kumar [a], A.K. Goel [b], B. Umakanth [a], B. Jahnavi [a], B.C. Viraktamath [a]

[a] Crop Improvement section, Directorate of Rice Research, Rajendranagar, Hyderabad 500030, India
[b] Department of Statistics, Acharya N. G. Ranga Agricultural University, Rajendranagar, Hyderabad 500030, India

ARTICLE INFO

Article history:
Accepted 19 August 2013
Available online 28 August 2013

Keywords:
miRNA gene
Promoter
Transcription factor
Cis-regulatory elements
Overrepresented motifs
Positional preference

ABSTRACT

Plant growth and yield are affected by many abiotic stresses like salinity, drought, cold and heavy metal; these stresses trigger up and down-regulate several genes through various transcription factors (TFs). Transcription factor binding motifs (TFBMs), located in the upstream region of the genes, associate with TFs to regulate the gene expression. Many factors, including the activation of miRNAs, which are encoded by genes having independent transcription units, regulate the gene expression. TFBMs in the regulatory region of miRNA sequences influence the miRNA expression, which in turn influences the expression of other genes in the cell. However, the current level of information available on TFBMs of miRNA involved in abiotic stress related defense pathway(s) is limited and in-depth studies in this direction may lead to a better understanding of their role in expression and regulation of defense responses in plants. In this study, various aspects related to genomic positions of pre-miRNA, prediction of TSS and TATA box positions and identification of known, unique motifs at regulatory regions of all the reported miRNAs of rice associated with different abiotic stresses are discussed. Sixteen motifs were identified in this study, of which nine are known *cis*-regulatory elements associated with various stresses, two strong motifs, (CGCCGCCG, CGGCGGCG) and five unique motifs which might play a vital role in the regulation of abiotic stresses related miRNA genes. Common motifs shared by miRNAs that are involved in more than one abiotic stresses were also identified. The motifs identified in this study will be a resource for further functional validation.

1. Introduction

Rice is one of the important food crops in several countries, a major cereal crop and also a good model for understanding complex biological functions. Plant growth and development is regulated at many stages spatiotemporally (Bian et al., 2012; Lan et al., 2012; Willmann and Poethig, 2007). Recent studies demonstrated the involvement of microRNAs (miRNAs) at different stages of plant growth (Meng et al., 2009; Reinhart et al., 2002; Wang et al., 2004; Xue et al., 2009; Zhang et al., 2006). These miRNAs have attained significant attention during the recent past, since they play an active role in gene expression and are known to be involved in various environment stress responses in plants that includes salinity, cold, heavy metal, hydrogen peroxide, ABA, drought and different biotic stresses (Ding et al., 2011; Huang et al., 2009; Jian et al., 2010; Li et al., 2011; Liu et al., 2009; Lu and Huang, 2008; Lv et al., 2010; Sanan-Mishra et al., 2009; Zhao et al., 2007, 2009; Zhou et al., 2010). MiRNAs are located at diverse genomic positions, they may be intergenic (between two genes) or genic (within genes) or in clusters (Cui et al., 2009; Zhang et al., 2009). Though many miRNAs are identified, their transcriptional regulation is poorly understood. But recent studies indicate that miRNAs are known to be encoded by their own genes and transcribed by RNA polymerase II to form long primary transcripts (pri-miRNAs) (Dugas and Bartel, 2004; Kurihara and Watanabe, 2004; Lee and Ambros, 2001; Xie et al., 2007) suggesting their existence as independent transcription units. However, the promoters of miRNAs are not well characterized (Ozsolak et al., 2008) and miRNA genes are believed to be regulated by various transcription factors (TFs) (Hackenberg et al., 2012; Sun et al., 2012). Transcription factor binding motifs (TFBMs), that are located in the upstream region of the genes, regulate the gene expression in association with TFs. The architecture of TFBMs may be conserved to a particular class of genes or genes co expressed to a particular stress, and that phenomenon may also operate even in the regulation of miRNA genes. Towards understanding their transcriptional regulation, the miRNA promoter elements were discovered in four model organisms viz., *Caenorabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana* and *Oryza sativa*. Similar types of promoters were identified in most of the known miRNA genes as well as Protein Coding Genes (PCGs) (Zhou et al., 2007). Further, putative core promoters of most known miRNA genes were also identified and many significant motifs were discovered, some of them

Abbreviations: miRNA, microRNA; TF, transcription factor; TFBMs, transcription factor binding motifs; ASR miRNA, abiotic stress miRNA; TSS, transcription start site; PCGs, protein coding genes.

* Corresponding author. Tel.: +91 40 24015036x208; fax: +91 40 24015308.
E-mail addresses: sheshu24@gmail.com, sheshu_24@yahoo.com (M.S. Madhav).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.08.060

were species specific and some were conserved across different species (Zhou et al., 2007). Megraw et al. (2006) compared the representation of the TFBMs in promoters of *Arabidopsis* miRNA, protein coding genes and random genomic sequences and reported five TFBM's that were overrepresented in promoters of miRNAs than promoters of PCG's. However, limited information is available on TFs involved in miRNA regulation and their binding motifs. Hence a comprehensive study in this direction may lead to a better understanding of their expression and regulation.

The objectives of this study is to identify and determine the positional conservation of overrepresented, unique as well as known *cis* motifs present in ASR (abiotic stress) miRNAs and to validate such overrepresented motifs with other miRNAs including maize and *Arabidopsis* ASR miRNAs. A systematic approach of in silico identification of the TFBMs in the promoter region of abiotic stress miRNA (ASR miRNA) genes of rice is demonstrated in this study. Various bioinformatic tools available in the public domain were exploited to identify the reliable candidate TFBMs involved in ASR miRNA.

2. Materials and methods

2.1. Datasets

Six datasets were used to identify the unique motifs associated with rice ASR miRNA. These datasets were (1) abiotic stress specific rice miRNAs selected from literature (i.e. salinity, drought, cold and heavy metal stresses) (Supplementary Table 1); (2) rice miRNAs (301 no's) which are not involved in abiotic stresses (Supplementary Table 2), were retrieved from the miRBase database (http://www.mirbase.org; Release 16, 2010); (3) random protein coding genes (PCGs) covering all the chromosomes of rice (10 per chromosome); (4) PCGs involved in various abiotic stresses (Supplementary Table 3); (5) *Arabidopsis* abiotic stress miRNA (Supplementary Table 4) and (6) maize abiotic stress miRNA (Supplementary Table 5). The datasets (4), (5), and (6) were also selected based on literature survey.

2.2. Retrieval of promoter sequences

The genomic locations and upstream sequences (2 kb) of rice miRNAs were retrieved from the RAP-DB (http://rapdb.dna.affrc.go.jp/). The miRNAs were assumed as independent transcription units to have uniformity. 2 kb upstream sequences were retrieved at the beginning of the pre-miRNA for the prediction of TSS (Transcription Start Site) for all the types of miRNA (genic, intergenic and cluster miRNA). Two kb upstream sequences were also retrieved for *Arabidopsis* and maize pre-miRNA from *Arabidopsis thaliana* (TAIRv10) and maizeGDB databases (http://gbrowse.arabidopsis.org/cgi-bin/gbrowse/arabidopsis/ and http://gbrowse.maizegdb.org/gb2/gbrowse/maize_v2/). The TSS and TATA-box predictions were made using TSSP web tool (http://linux1.softberry.com/berry.phtml? topic = tssp& group = programs&subgroup = promoter). Putative promoter sequences from −1,000 to −1 from the TSS were retrieved for all classes of miRNA and used for motif search and identification of strong motifs. For random and abiotic stress PCGs, 1 kb upstream sequences were directly downloaded from RAP-DB and used for the identification of TFBMs.

2.3. Scanning for transcription factor binding motifs (TFBMs)

The ASR miRNA of rice and maize were scanned for putative TFBM's using two modules (1) PLACE (Plant *cis*-acting regulatory DNA elements) signal scan search, to identify the known *cis*-regulatory elements (http://www.dna.affrc.go.jp/PLACE/signalscan.html) and (2) MELINA — II a web based tool, to run and compare the motif prediction algorithms at once (http://melina2.hgc.jp/public/index.html) (Okumura et al., 2007). The four motif prediction algorithms used in this study were: (1) Consensus, (2) Gibbs Sampler, (3) MD SCAN [these three programs were used with default parameters] and (4) MEME [with a cut-off E-value of 1 with anr (any number of repetitions) mode]. The motifs identified by at least two programs were considered as strong motifs. All the other four datasets were also scanned for the identification of TFBMs using MELINA-II. The positional conservation of motifs was analyzed manually using the data output of MELINA II by dividing 1 kb upstream region into 5 windows with equal lengths of 200 bp. If a motif had a higher frequency in a particular window than others, it was considered to be conserved in that particular window.

2.4. Identification of unique motifs

All the motifs identified through MELINA-II among ASR miRNA were checked with the TFBM database (PLACE) to know their function. The motifs which were not detected in PLACE database were considered as unique (novel) motifs.

2.5. Validation of strong motifs

In order to predict precisely that the motifs identified were specific to the stress miRNAs, the overrepresented motifs were validated with the motifs identified from the upstream sequence (1 kb from TSS) from five sets of controls (1) miRNA of rice which are not involved in the stresses, (2) 120 random protein coding genes (PCGs) covering all the chromosomes, (3) protein coding genes (PCGs) specifically involved in different abiotic stress, (4) *Arabidopsis* stress miRNAs and (5) maize stress miRNAs.

2.6. Statistical analysis

The motifs obtained were analyzed using Z-test to determine the significance of presence or absence of the unique motifs obtained through MELINA II. The analysis was carried out by comparing the ratio of occurrence of motif in the target sequence to the ratio of occurrence of motif in the five control sequences using the formula,

$$Z = |X/n - P| / \sqrt{PQ/n}$$

where P is proportion of the presence of motif in target sequences, Q is 1-P, which indicates the proportion of motif in non target sequences, n is number of sequences tested, X is number of sequences showing the presence of a particular motif. The calculated Z values were compared with the standard Z tabulated value, 1.96 ($P < 0.05$) & 2.576 ($P < 0.01$) to find the significance of the motifs identified.

3. Results

3.1. Characterization and identification of genomic locations for rice ASR miRNAs

One hundred and thirty six miRNAs that were involved in salinity, cold, heavy metal stress, hydrogen peroxide treatment, ABA treatment and drought stress (Ding et al., 2011; Huang et al., 2009; Jian et al., 2010; Li et al., 2011; Liu et al., 2009; Lv et al., 2010; Zhao et al., 2007, 2009; Zhou et al., 2010) were selected. [The summary of the miRNAs used in the present study is given in Supplementary Table1]. The miRNAs were classified based on their involvement in one or more than one stresses and they are represented in Table 1. The upstream and downstream regions of ASR miRNA were checked to get the information regarding the presence of nearest genes and their positions (Supplementary Table 1). The results indicate that 80 miRNAs (58.82%) were intergenic, 33 (24.26%) were in clusters and 23 (16.91%) were genic in their location (Supplementary Fig. 1). The cluster miRNAs were either intergenic (e.g. mir2124d and mir812i) or genic (e.g. mir172b and mir806a). Some instances, one member of miRNA cluster may be present in genic while others may be intergenic