Gene 431 (2009) 61–66

Contents lists available at ScienceDirect

Gene

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

Genomic analysis of rice microRNA promoters and clusters

Xiao Cui [1], Si Min Xu [1], Dao Shuai Mu, Zhi Min Yang *

Department of Biochemistry and Molecular Biology, College of Life Science, Nanjing Agricultural University, Nanjing, 210095, China

ARTICLE INFO

Article history:
Received 8 August 2008
Received in revised form 16 October 2008
Accepted 13 November 2008
Available online 24 November 2008

Received by A. Rynditch

Keywords:
miRNA gene
Rice
Promoter
miRNA cluster

ABSTRACT

MicroRNAs (miRNAs) are endogenous single-stranded non-coding small RNAs with a length of about 21 nt, that negatively regulate development and stresses. Rice miRNAs are representative of the monocot miRNAs, and many of them are non-conserved in *Arabidopsis* and the other plant species. Previous studies have shown that a majority of plant miRNAs are expressed from independent transcription units, whereas some others are transcribed with their host genes. Despite of the fact that a growing number of rice miRNAs are discovered, little is known about the transcriptional regulation of miRNA genes. In this study, we performed genomic analysis of rice miRNA transcripts surrounding the regions of promoter/transcription start site (TSS) and TATA-box, and organization of miRNA clusters. We detected 249 promoters for 212 rice pre-miRNA sequences via bioinformatics approach and found that the conserved rice miRNA genes have a greater proportion of promoters than the non-conserved miRNA genes. We further globally analyzed the genomic organization of pri-miRNAs and found that 52 rice miRNA genes appear in 18 clusters. Alignment of the miRNA sequences in these clusters shows a number of miRNA paralogs within the cluster. The data obtained may aid our understanding of the specific sequences upstream of pre-miRNAs and the functional implications of miRNA clusters in rice plants.

1. Introduction

MicroRNAs (miRNAs) are endogenous single-stranded non-coding RNAs with a length of about 21 nt. They are known to negatively regulate post-transcription of genes through interactions with 3′ untranslated regions or coding regions of their targets (Bartel, 2004). In plant, miRNAs regulate the expression of a large number of genes which control development (Chen, 2004; Laufs et al., 2004) and stress responses (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Liu et al., 2008). Overexpression or knockout of miRNA genes disturbs metabolisms and consequently results in the abnormal phenotypes (Palatnik et al., 2003; Chen, 2004; Guo et al., 2005). Expression of miRNAs in plants involves several steps. First, a miRNA gene is transcribed in the nucleus as a primary transcript (pri-miRNA), which is usually a long sequence of more than several hundred nucleotides. This step is controlled by Pol II enzymes (Bartel, 2004; Kurihara and Watanabe, 2004). Then, the pri-miRNA is cleaved to an intermediate named miRNA precursor (or pre-miRNA) with stem–loop structures. While the pre-miRNAs in animals are transported by Exportin 5 from the nucleus into the cytoplasm (Yi et al., 2003; Lund et al., 2004), the plant miRNA precursors are processed by Dicer-like enzyme 1 (DCL1) in the nucleus (Tang et al., 2003; Kurihara and Watanabe, 2004; Allen et al., 2005) and transported into the cytoplasm by HASTY (a plant orthologue of exportin 5) (Park et al., 2005). The single-stranded miRNAs associate with ARGONAUTE (AGO) proteins in a complex termed RNA-induced silencing complex (RISC), where it guides the cleavage or translational repression of its target by base-pairing with the target (Bartel, 2004).

Several lines of evidence have shown that a majority of miRNAs are encoded in their own genes located in the intergenic regions (Lee and Ambros, 2001; Qiu et al., 2007; Xie et al., 2007), suggesting that they exist as independent transcription units. Another class of miRNA genes can be found in the intronic regions, which may be transcribed as a part of the annotated genes (Lee et al., 2004). Unlike animal miRNAs, miRNAs in plant are primarily encoded in intergenic regions (Jones-Rhoades and Bartel, 2004), suggesting that plant pri-miRNA transcription may differ from that of animals. Pri-miRNAs are typically transcribed by RNA polymerase II and have promoter elements that are similar to those of protein-coding genes (Smale, 2001). It is known that the class II promoters have two parts: the core promoter and upstream element. The core promoter has at least two elements: a TATA box beginning at approximate position −30 and an initiator centered on the transcription start site (TSS). A recent investigation identified more than 52 promoters in *Arabidopsis*, and most of them were found to contain TATA-boxes in their core promoter regions (Xie et al., 2005). However, there are exceptions to this rule. For example, identification of human miRNA gene mir-23a-27a-24-2 reveals that its promoter has no common elements like TATA-boxes or initiator elements required for initiating transcription (Lee et al., 2004). Such

Abbreviations: miRNAs, MicroRNAs; TSS, transcription start site; DCL1, Dicer-like1; AGO, ARGONAUTE; RISC, RNA induce silencing complex; GFF file, Gene-Finding Format or General Feature Format.

* Corresponding author.
E-mail address: zmyang@njau.edu.cn (Z.M. Yang).

[1] These authors contributed equally to this work.

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.11.016

TATA-less promoters are often found in housekeeping genes (Smale, 2001). Understanding the features of upstream sequences including promoters, transcription start sites, or diversity of the specific elements is necessary to understand the location and extent of pri-miRNAs, expression patterns of miRNAs, and miRNA-mediated regulatory pathways and networks.

Understanding of the mechanism for miRNA gene expression is fundamentally important. One of the main goals for miRNA research is to elucidate how pri-miRNA genes are transcribed and how complex gene regulatory networks evolve (Chen and Rajewsky, 2007). Recently, analysis by cloning and computational approaches resulted in identification of many miRNA promoters in *Arabidopsis* (Xie et al., 2005; Shahmuradov et al., 2005; Megraw et al., 2006; Wang et al., 2006). Xie and co-workers (2005) identified 63 TSSs in *Arabidopsis* via 5′-RACE amplification, which provide opportunity for computational analysis of miRNA promoter regions in plants. To elucidate the transcriptional regulation of miRNA genes, Zhou et al. (2007) identified the promoters of intergenic microRNA genes in *Caenorhabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana* and *Oryza sativa*, and found that most known miRNA genes in the four species have the same type of promoters as protein-coding genes. However, relative to *Arabidopsis*, little is known about promoters of miRNAs in rice. Recently, a study performed via high-throughput sequencing identified millions of small RNAs from rice (Lu et al., 2008). The alignment of full-length cDNA sequences may allow to identify some promoters in rice. But, most of cDNA sequences clones do not extend to the transcriptional start site (TSS) (Ohler et al., 2001; Guddeti et al., 2005). Bioinformatic prediction of miRNA promoter sequences is proved to be efficient to elucidate regulatory mechanisms for miRNA expression. It generates a number of candidates to save the initial testing of thousands of potential genomic fragments including sequences of promoter regions (Shahmuradov et al., 2005; Wang et al., 2006; Saini et al., 2007; Xie et al., 2007; Yin et al., 2008; Zhou et al., 2008).

Rice (*O. sativa*) is one of the most importantly economical crops because it provides the major portion of calories in human diet in Asia and the other parts of the world. In addition to its agricultural importance, it is a model system for monocot species with complete genome sequences, thus making rice a favorite for functional genomic research. To date, there are at least 1160 higher plant miRNAs registered in the miRNA database (http://microrna.sanger.ac.uk), of which a total of 269 loci have been found in *O. sativa* and 184 in *A. thaliana* (miRBase, Release 11.0, Griffiths-Jones et al., 2006). Although more miRNAs have been discovered in rice than in *Arabidopsis*, there is a growing recognition of the significant numbers of rice miRNAs that are non-conserved in *Arabidopsis* and the other plant species (Wang et al., 2004; Sunkar et al., 2005; Griffiths-Jones et al., 2006; Liu et al., 2008). It is shown that there have been 90 rice miRNAs representing 40 families that have no homologous counterparts in *Arabidopsis*. These results imply that non-conserved miRNAs in rice may play species-specific roles in development and stressful responses. From an evolutionary perspective, these non-conserved miRNAs may have expanded after the divergence of the monocot and dicot plant lineages. However, this is only the beginning to be studied. Also, it is unclear about the upstream regulatory sequence of miRNAs in rice genome. Therefore, the aim of this study is to identify: (1) specific sequences or motifs adjacent to independent and co-transcribed pre-miRNAs, which are associated with their expression of conserved and non-conserved miRNAs, and (2) the pattern of miRNA clusters associated with the upstream specific sequences of pre-miRNA in rice plants. We developed multiple strategies based on the rice genome data to describe the features of pri-miRNA transcripts and map them to the surrounding regions: transcription start site, TATA box binding motif, promoter region of cluster miRNA families. The outcome of the study may aid our understanding of the features outside of the actual pre-miRNAs and provide the insight into the structure of primary transcripts and the transcription of clustered miRNAs as polycistronic transcripts.

2. Materials and methods

2.1. Upstream sequences of rice miRNAs

We obtained rice (*O. sativa*) pre-miRNA sequences from the miRBase database (version 10.1, http://microrna.sanger.ac.uk/seguence/index.html) (Griffiths-Jones et al., 2006). The rice genome sequences were downloaded from TIGR Oryza Pseudomolecules (Release 4.0) (http://rice.plantbiology.msu.edu) and TIGR Oryza Genome Browser (Ouyang et al., 2007). The conservation of miRNAs was deduced from miRNA distribution across plant kingdom. Only the specific miRNAs were defined as non-conserved miRNAs. We then divided the miRNAs into two groups based on their genomic background. The first group was composed of 162 miRNAs, that reside between protein-coding genes and were defined as “intergenic miRNAs”; the second group consists of 50 miRNAs, that overlap protein-coding genes and were defined as “intronic miRNAs” (Ouyang et al., 2007). The definition of protein-coding genes is in accordance with TIGR automated annotations of rice genome, thus including verified protein genes as well as hypothetical protein genes.

Sequences in the intergenic regions upstream of pre-miRNAs were organized according to the method described previously (Zhou et al., 2007). Briefly, if a pre-microRNA and its upstream gene were in the

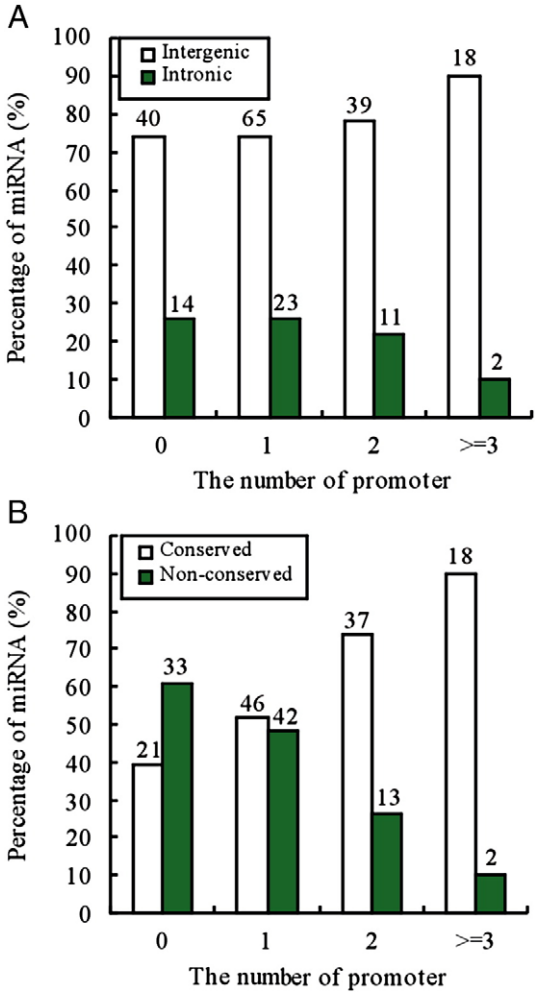

Fig. 1. Distribution of rice miRNA genes with the same amount of promoters. (A) The percentage of rice pri-miRNAs located within the intergenic or intronic regions. (B) The percentage of pri-miRNAs in relation to the conserved or non-conserved miRNAs. The numbers on the bars indicate the amounts of promoters contained by pri-miRNAs derived from intergenic/intronic regions or by conserved/non-conserved pri-miRNAs.