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.

© 2008 Elsevier B.V. All rights reserved.

## 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

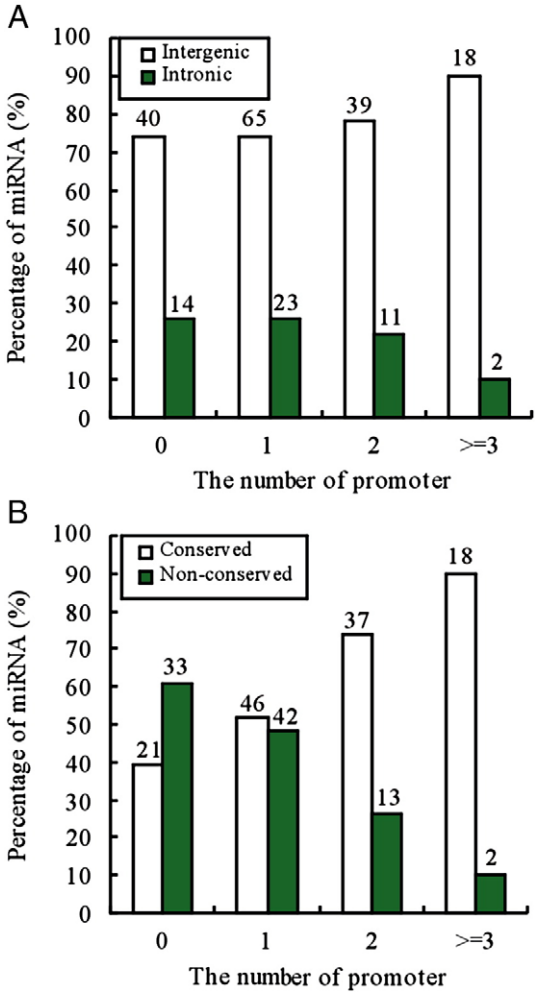


**Fig. 1.** Distribution of rice miRNA genes with the same amount of promoters. (A) The percentage of rice pri-miRNAs located within the intergenic or intronic regions. (B) The percentage of pri-miRNAs in relation to the conserved or non-conserved miRNAs. The numbers on the bars indicate the amounts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

same direction and the distance between them was longer than 2400 bp, the 2000 bp sequence upstream of the pre-miRNA was retrieved; otherwise, the sequence between 400 bp downstream of the upstream gene and the precursor was used. Similarly, if a pre-miRNA and its upstream gene were in the opposite direction and the distance between them was longer than 4000 bp, the 2000 bp sequence upstream of the precursor was obtained; otherwise, the sequence from the precursor and the middle point between the upstream gene and precursor was retrieved. All these sequences were extracted by Apollo (a genome annotation tool) on the base of GFF file ('Gene-Finding Format' or 'General Feature Format') of rice miRNAs from genome sequences (Lewis et al., 2002). Three hundred random sequences with length of 2000 bp were automatically generated by computer as a control.

### 2.2. Prediction of specific sequences upstream of rice miRNA genes

Sequences of transcription start site (TSS) and TATA-box were predicted using TSSP (http://mendel.cs.rhul.ac.uk/mendel.php?topic=gen). It is one of the best promoter prediction methods for plant miRNAs and the predictions were obtained at the default thresholds of TSSP. The program has been trained and tested on independent sets of well-known promoters (Shahmuradov et al., 2005). We also adopted our "miRNAassist" for supplemental predictions (Xie et al., 2007).

### 2.3. Clustering of rice miRNA genes

For analysis of miRNA clustering, both upstream and downstream sequences with pairwise distance less than 10 kb were considered as clustered miRNAs. When the clustered miRNAs were organized, the 5′-end sequences of the first miRNAs in the upstream regions were fetched following the same rule mentioned above. Sequences within less than 150 bp and between any two clustered miRNAs were removed. Some miRNAs that overlap transposons were excluded from our data set.

## 3. Results and discussion

### 3.1. Analysis of rice miRNA genes with relation to the promoters

We first searched for the putative promoters for all of the 243 rice pre-miRNA sequences, which are publicly available (Griffiths-Jones et al., 2006). We analyzed the upstream sequences up to 2000 bp of

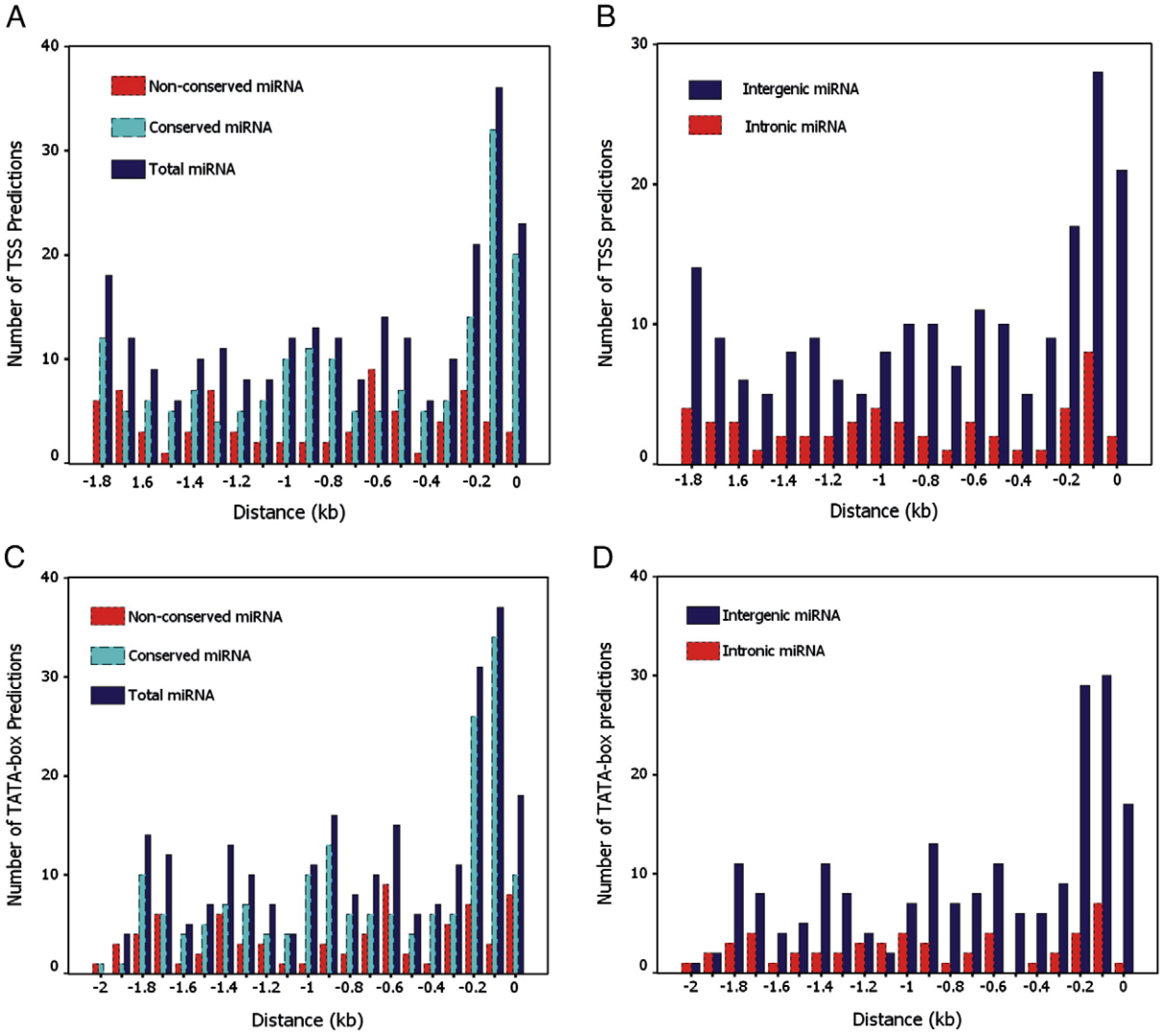


**Fig. 2.** Genomic distribution of TSS and TATA-box from the 5′ end of the pre-miRNAs sequences. (A) TSS predictions in the conserved and non-conserved miRNAs; (B) TSS predictions from pri-miRNAs in intronic and intergenic regions; (C) TATA-box predictions in the conserved and non-conserved miRNAs; and (D) TATA-box predictions in intronic and intergenic regions. For the random intergenic sequences, see Supplemental Fig. 2.

the pre-miRNA sequences. A total of 212 (87.2%) upstream sequences were obtained. These candidates were used for further analysis. The other 31 (12.8%) pre-miRNA sequences were unsuitable for promoter prediction due to the following reasons: (1) some of these miRNA genes overlap each other; (2) some miRNAs (*e.g.* Osa-miR395) are located in the upstream region with very short distance; and (3) some miRNAs cannot be located in the chromosome and hence further detection was impossible. Our analysis reveals that the 212 pri-miRNA sequences correspond to the 249 predicted promoters. Of the 212 sequences, 54 (25.5%) were found to contain no promoters (Fig. 1). Within the other 158 (74.5%) promoter-containing sequences, 88 (41.5%) were predicted to have only one promoter, 50 (23.6%) contain two promoters, and the rest 20 (9.7%) contain three or more promoters. miRNA genes in plants and animals are found in diverse genomic locations. In animals, they are either encoded in protein-non-coding regions, or hosted within the introns (Altuvia et al., 2005). In plants, the majority of miRNA genes are found to occur in the intergenic regions. To understand the genomic distribution of rice pri-miRNAs in more detail, we divided these pri-miRNA sequences into two groups: intergenic/intronic and conserved/non-conserved pri-miRNAs. The pri-miRNAs from intergenic regions usually contain more promoters than the intronic pri-miRNAs (Figs. 1A and 2B). But the proportions of the two groups of pri-miRNAs remain relatively unchanged depending on the number of promoters they contain. The pri-miRNAs also can be classified into conserved and non-conserved groups. In this case, the proportion of the two pri-miRNAs considerably varies (Fig. 1B). The percentage of conserved pri-miRNAs tends to increase with the promoters they contain, whereas the proportion of non-conserved pri-miRNA shows a decrease. These data suggest that rice pri-miRNAs that contain more than one promoter appear to be conserved.

### 3.2. Analysis of specific sequences upstream of pre-miRNA

We subsequently analyzed the distribution of the putative promoter positions within the corresponding pri-miRNA sequences. We were able to obtain a total of 249 TSS predictions within 2000 bp upstream of pre-miRNAs for the 212 miRNA genes. A profound peak is observed near the start positions of the transcripts. The vast majority of predicted TSSs are found to lie within 400 bp upstream of pre-miRNAs (Fig. 2, Fig. S1). For the TSSs contained by the conserved miRNA genes, the distribution can be separated into three distinct regions (Fig. 2A). The first peak is found close to the pre-miRNA within the 400 bp upstream region, containing 41% of total (249) TSS predictions. The second broad region from −680 bp to −1300 bp contains 19% of TSSs. The third region ranges from −1700 bp to −2000 bp, containing 11.3%. The remaining has 28.7% of the total predictions. Also, it is found that TSSs are located in the different regions of the upstream non-conserved miRNA genes. The four small peaks (within 0 bp to −400 bp, −400 bp to −750 bp, −1200 bp to −1500 bp, and −1500 bp to −2000 bp) together contain 65 (26.1%) of the total TSS predictions.

It is generally accepted that miRNA genes in organisms can be classified into two groups, which are transcribed from either intergenic or intronic regions. The intergenic miRNA genes are located far away from any annotated genes, implying the independent transcription from their own transcription regulatory elements. The intronic miRNA genes lie within introns of protein-coding genes and in the same orientation as the miRNAs, suggesting that the protein-coding genes may serve as host genes for co-expressed miRNAs (Lee et al., 2004). We analyzed the distribution of TSSs derived from the two categories of miRNA genes. It is shown that there are two dominant peaks with TSSs from intergenic pri-miRNA sequences (Fig. 2B). This pattern of TSS distribution is very similar to that from conserved pre-miRNAs (Fig. 2A). For intronic pri-miRNAs, there is a single TSS-containing peak. Although the pattern of TSS distribution differs between the intergenic and intronic pri-miRNAs, TSSs in the two pri-miRNAs lie predominantly in the region from 0 to −400 bp. Overall, 30.5% and 8% of total TSSs in these regions belong to the intergenic and intronic miRNA genes, respectively.

The TATA-box is a well characterized promoter element. In this study, we have predicted 238 TATA-boxes for 212 upstream pri-miRNAs. Of the 212 rice miRNA genes, 202 (95.3%) have a TATA-box. This finding suggests that the majority of rice miRNA genes have the same promoters as the protein-coding genes and transcribed by RNA polymerase II. However, there are 10 pri-miRNA sequences including MIR171a, MIR164a, MIR426, MIR438, MIR439h, MIR439f, MIR445a, MIR812b, MIR819h and MIR820c, which were predicted to contain no TATA-boxes. The non-TATA box-containing miRNA genes may fall into the class termed TATA-less promoters, which can be found in *Arabidopsis* (Xie et al., 2005). Such TATA-less promoters tend to be found in two classes of genes: (1) the first class comprises the housekeeping genes that are constitutively active in virtually all cells because they control common biochemical pathways, needed to sustain cellular life; (2) the second class of genes with TATA-less promoters are developmentally regulated genes such as the homeotic genes that control development of the fruit fly or genes that are active during the development of the immune system in mammals (Smale, 2001; Weaver, 2001). Analysis of TATA-box distribution in the regions upstream of the rice pre-miRNAs shows that the locations of

**Table 1**
Features of predicted rice miRNA clusters on the genome

| No. | Cluster | Cluster length | Distance between miRNAs | Locations | Number of promoter |
|---|---|---|---|---|---|
| 1 | MIR156b-c | 541 | bc:218 | Intergenic | Single |
| 2 | MIR159c-d | 7896 | cd:7512 | Intergenic | Multiple |
| 3 | MIR166h-k | 289 | hk:45 | Intergenic | Single |
| 4 | MIR167a-h | 3929 | ah:3670 | Intergenic<br>h:intronic | Multiple |
| 5 | MIR169h-i | 2775 | hi:1190 | Intergenic | Multiple |
| 6 | MIR169j-k | 3427 | jk:3148 | Intergenic | Multiple |
| 7 | MIR169n-o | 3900 | no:3552 | Intergenic | Single |
| 8 | MIR171f-443 | 6451 | 171f-443:6178 | Intergenic | Multiple |
| 9 | MIR172b-806a | 4513 | 172b-806a:4025 | Intergenic | Single |
| 10 | MIR172c-820b | 9244 | 172c-820b:8944 | Intergenic | Multiple |
| 11 | MIR395a-b-c-d-e-f-g | 1002 | ab:75; bc:53; cd:51; de:56; ef:57; fg:56 | Intergenic | Single |
| 12 | MIR395h-i-j-k-l | 946 | hi:64; ij:62; jk:62; kl:247 | Intergenic | Single |
| 13 | MIR395m-n-o-p-q-r-s | 1035 | mn:131; no:73; op:76; pq:76; qr:76; rs:65 | Intergenic | Single |
| 14 | MIR395t-u-v-w | 1204 | wv:216; vu:58; ut:517 | Intergenic | None |
| 15 | MIR396a-c | 7629 | ac:7336 | Intergenic | Multiple |
| 16 | MIR399a-e | 1172 | ae:907 | Intergenic | Multiple |
| 17 | MIR399c-h-k | 5420 | ch:1889; hk:3219 | Intergenic<br>k:intronic | Multiple |
| 18 | MIR399d-439h | 8955 | 399d-439h:8572 | Intergenic | Multiple |

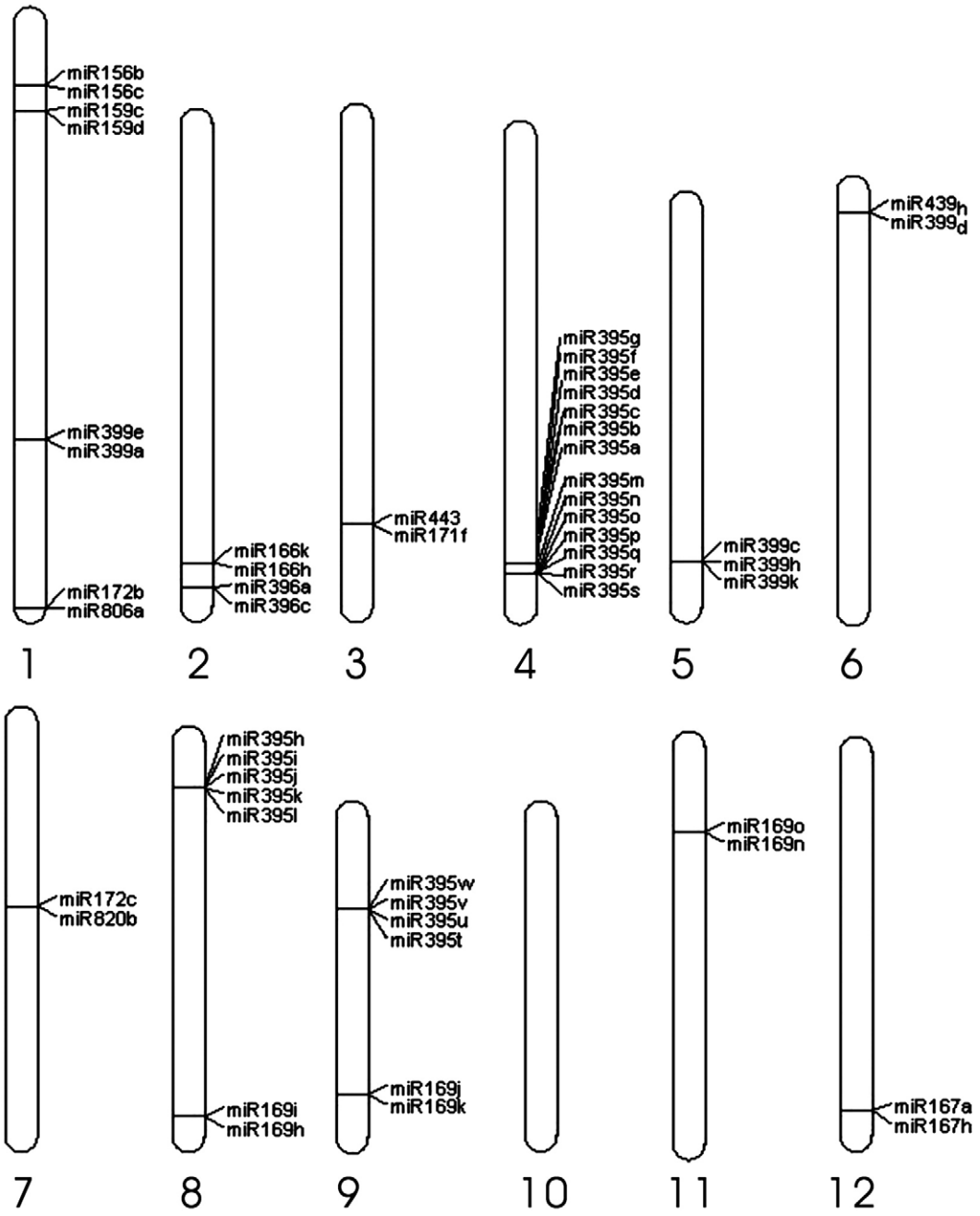


**Fig. 3.** Chromosomal location of the predicted rice miRNA clusters. The relative locations of individual miRNAs are shown across the 11 rice chromosomes except for chromosome 10, where no rice miRNA clusters are detected.

TATA-box are very similar to those of TSSs (Figs. 2C, D). The majority of TATA-boxes occur within 0 to −400 bp of pri-miRNAs. As a control, we predicted TSSs and TATA-boxes in randomly selected sequences of length 2000 bp. The distribution of predictions in random intergenic sequences is apparently different from that of the surrounding miRNAs (Fig. S2), suggesting that the distribution of TSSs or TATA-boxes at the indicated regions is specific.

### 3.3. Clustering of rice miRNA genes

In human genome, many of known miRNAs are shown to form clusters. However, very little is known about miRNA clustering in plants. Some clusters reflect the processing of miRNAs from a single polycistronic transcript, implying that more than one pre-miRNA may be processed from the same primary transcript (Altuvia et al., 2005). Also, this process implies that the co-transcription of functionally different miRNAs in a cluster may simultaneously target several categories of genes (Guddeti et al., 2005; Wang et al., 2007). In contrast, some other miRNA clusters demonstrate the independent and transcriptional regulation and relative long-distance physical location (Tanzer and Stadler, 2004). We examined the potential clustering of miRNAs on the rice genome and analyzed the genomic organization of all registered miRNA genes within 10 kb. Because rice miRNA genes lie in the protein-coding regions or in the intergenic regions, we treated the pri-miRNAs as the two groups (miRNA clusters from intergenic regions and introns). Our analysis reveals that a total of 18 clusters were predicted for 52 rice miRNA genes (Table 1). Within these miRNA genes, 50 are located in intergenic regions and only 2 are found in introns. The 18 rice miRNA clusters contain 13 pairs, 1 triplet, one group of four, one group of five, and two groups of seven, and are

differentially located onto the 11 chromosomes (Fig. 3). The rice miRNA clusters are diverse in structure. The length of miRNA clusters varies from 289 to 9244 bp, with an average of 3907. About 67% of the clusters are within the length of 946–6451 bp. The varied size of these miRNA clusters suggests that they may be differentially transcribed and may offer unique functions for regulating expression of miRNAs. The diversity of the rice miRNA clusters also can be found in the aspect of the amount of promoters they contain. For instance, the clusters of MIR156b-c, MIR166k-h, MIR169n-o, MIR172b-806a, MIR395a-b-c-d-e-f-g, MIR395h-i-j-k-l, and MIR395m-n-o-p-q-r-s contain only one promoter, suggesting such clustered miRNAs may be transcribed under the control of a single promoter; the other miRNA clusters (MIR159c-d, MIR167a-h, MIR169h-i, MIR169j-k, MIR171f-443, MIR172c-820d, MIR396a-c, MIR399a-e, MIR399c-h-k and MIR399d-439h) have multiple promoters. However, no promoter was predicted for MIR395t-u-v-w. We found that most of the clustered rice miRNAs are derived from the same family (Table 1). Only MIR172c-820d, MIR172b-806a, MIR399d-439h, and MIR171f-443 are the clusters that have two different miRNA members. This feature is demonstrated in the rice miR395 cluster (Guddeti et al., 2005). In human, most miRNA clusters contain miRNAs from the different families and may consist of several different groups of non-homologous miRNAs (Yu et al., 2006). Additionally, the human miRNA clusters are frequently observed to have super sized clusters. The human hsa-miR-127 (also known as hsa-miR134), that resides on an imprinted region of human chr. 14q32, contains more than 50 members, and has been found as the largest miRNA cluster to date (Volinia et al., 2006). However, such a large miRNA cluster has not been identified yet in plant kingdom.

In conclusion, we have performed genomic analysis of upstream specific regulatory sequences (TSS, TATA-box) of rice pre-miRNAs and detected 249 promoters for 212 rice pre-miRNA sequences. It is found that the conserved rice miRNA genes have a greater proportion of promoters than the non-conserved miRNA genes. Further, we globally analyzed the genomic organization of registered miRNA genes and found 18 clusters containing 52 miRNA genes in the rice genome. These findings may help our understanding of the specific sequence/motif upstream of pre-miRNAs and functional implications of miRNA clusters in rice plants. Also, they may provide a foundation for further investigation of the functional role of miRNA promoters and the regulation of miRNA expression in rice.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (30671249) and the Natural Science Foundation of Jiang Su, China (BK2006137), as well as the Student Research Training program from NAU.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2008.11.016.

## References

Allen, E., Xie, Z.X., Gustafson, A.M., Carrington, J.C., 2005. mRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121, 207–221.
Altuvia, Y., et al., 2005. Clustering and conservation patterns of human microRNAs. Nucleic Acids Res. 33, 2697–2706.
Bartel, D.P., 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
Chen, X., 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. Science 303, 2022–2025.
Chen, K., Rajewsky, N., 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. 8, 93–103.
Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. Database Issue 34, D140–144.
Guddeti, S., et al., 2005. Molecular evolution of the rice miR395 gene family. Cell Res. 15, 631–638.
Guo, H.S., Xie, Q., Fei, J.F., Chua, N.H., 2005. MicroRNA164 directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell 17, 1376–1386.
Jones-Rhoades, M.W., Bartel, D.P., 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14, 787–799.
Kurihara, Y., Watanabe, Y., 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. Proc. Natl. Acad. Sci. 101, 12753–12758.
Laufs, P., Peaucelle, A., Morin, H., Traas, J., 2004. MicroRNA regulation of the CUC genes is required for boundary size control in *Arabidopsis* meristems. Development 131, 4311–4322.
Lee, R., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.
Lee, Y., et al., 2004. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 23, 4051–4060.
Lewis, S.E., et al., 2002. Apollo: a sequence annotation editor. Genome Biol. 3, 1–14.
Liu, H.H., Tain, X., Li, Y.J., Wu, C.A., Zheng, C.C., 2008. Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. RNA 14, 1–8.
Lu, C., et al., 2008. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc. Natl. Acad. Sci. 105, 4951–4956.
Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., Kutay, U., 2004. Nuclear export of microRNA precursors. Science 303, 95–98.
Megraw, M., Bave, V., Rusinov, V., Jensen, S.T., Kalantidis, K., Hatzigeorgiou, A.G., 2006. MicroRNA promoter element discovery in *Arabidopsis*. RNA 12, 1612–1619.
Ohler, U., Niemann, H., Lia, G.C., Rubin, G.M., 2001. Joint modeling of DNA sequence and physical properties to improve eukaryotic promoter recognition. Bioimfomatics 17, S199–206.
Qiu, C.X., et al., 2007. Computational identification of microRNAs and their targets in *Gossypium hirsutum* expressed sequence tags. Gene 395, 49–61.
Ouyang, S., et al., 2007. The TIGR rice genome annotation resource: improvements and new features. Nucleic Acids Res. 35, D883–D887.
Palatnik, J.F., et al., 2003. Control of leaf morphogenesis by microRNAs. Nature 425, 257–263.
Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., Poethig, R.S., 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. Proc. Natl. Acad. Sci. 102, 3691–3696.
Saini, H.K., Griffithos-Jones, S., Enright, A.J., 2007. Genomic analysis of human microRNA transcripts. Proc. Natl. Acad. Sci. 104, 17719–17724.
Shahmuradov, I.A., Solovye, V.V., Gammerman, A.J., 2005. Plant promoter prediction with confidence estimation. Nucleic Acid Res. 33, 1069–1076.
Smale, S.T., 2001. Core promoters: active contributors to combinatorial gene regulation. Gene Dev. 15, 2503–2508.
Sunkar, R., Zhu, J.K., 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell 16, 2001–2019.
Sunkar, R., Girke, T., Jain, P.K., Zhu, J.K., 2005. Cloning and characterization of microRNAs from rice. Plant Cell 17, 1397–1411.
Tang, G., Reinhart, B.J., Bartel, D.P., Zamore, P.D., 2003. A biochemical framework for RNA silencing in plants. Genes Dev. 17, 49–63.
Tanzer, A., Stadler, P.F., 2004. Molecular evolution of a microRNA cluster. J. Mol. Biol. 39, 327–335.
Volinia, S., et al., 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl. Acad. Sci. 103, 2257–2261.
Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J., Qu, L.H., 2004. Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acid Res. 32, 1688–1695.
Wang, S., et al., 2007. Molecular evolution and selection of a gene encoding two tandem microRNAs in rice. FEBS Lett. 581, 4789–4793.
Wang, Y., Hindemitt, T., Mayer, K.F.X., 2006. Significant sequence similarities in promoters and precursors of *Arabidopsis thaliana* non-conserved microRNAs. Bioinformatics 22, 2585–2589.
Weaver, R.F., 2001. Molecular Biology. McGraw-Hill Companies, Inc., pp. 279–280.
Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., Carrington, J.C., 2005. Expression of *Arabidopsis* miRNA genes. Plant Physiol. 138, 2145–2154.
Xie, F.L., Huang, S.Q., Guo, K., Zhu, Y.Y., Nie, L., Yang, Z.M., 2007. Computational identification of novel microRNAs and targets in *Brassica napus*. FEBS Lett. 581, 1464–1473.
Yi, R., Qin, Y., Macara, I.G., Cullen, B.R., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 17, 3011–3016.
Yin, Z., Li, C., Han, X., Shen, F., 2008. Identification of conserved microRNAs and their target genes in tomato (*Lycopersicon esculentum*). Gene 414, 60–66.
Yu, J., et al., 2006. Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. Biochem. Biophy. Res. Comm. 349, 59–68.
Zhou, X., Ruan, J., Wang, G., Zhang, W., 2007. Characterization and identification of miRNA core promoters in four model species. Plos Computational Biology 3, 412–423.
Zhou, Z.S., Wang, S.J., Yang, Z.M., 2008. Bioinformatic identification and expression analysis of new microRNAs from *Medicago truncatula*. Biochem. Biophy. Res. Comm. 374, 538–542.