

## Molecular evolution of the rice miR395 gene family

Sreelatha GUDDETI<sup>1,¶</sup>, De Chun ZHANG<sup>2,¶</sup>, Ai Li LI<sup>3,¶</sup>, Chuck H. LESEBERG<sup>1</sup>, Hui KANG<sup>1</sup>, Xiao Guang LI<sup>3</sup>, Wen Xue ZHAI<sup>2</sup>, Mitrick A. JOHNS<sup>1</sup>, Long MAO<sup>1,3,\*</sup>

<sup>1</sup>Department of Biological Sciences, Northern Illinois University, DeKalb, IL 60115, USA

<sup>2</sup>Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup>Institute of Crop Science and National Key Facility for Crop Gene Resource and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China

### ABSTRACT

MicroRNAs (miRNAs) are 20-22 nucleotide non-coding RNAs that play important roles in plant and animal development. They are usually processed from larger precursors that can form stem-loop structures. Among 20 miRNA families that are conserved between Arabidopsis and rice, the rice miR395 gene family was unique because it was organized into compact clusters that could be transcribed as one single transcript. We show here that in fact this family had four clusters of total 24 genes. Three of these clusters were segmental duplications. They contained miR395 genes of both 120 bp and 66 bp long. However, only the latter was repeatedly duplicated. The fourth cluster contained miR395 genes of two different sizes that could be the consequences of intergenic recombination of genes from the first three clusters. On each cluster, both 1-duplication and 2-duplication histories were observed based on the sequence similarity between miR395 genes, some of which were nearly identical suggesting a recent origin. This was supported by a miR395 locus survey among several species of the genus *Oryza*, where two clusters were only found in species with an AA genome, the genome of the cultivated rice. A comparative study of the genomic organization of *Medicago truncatula* miR395 gene family showed significant expansion of intergenic spaces indicating that the originally clustered genes were drifting away from each other. The diverse genomic organizations of a conserved microRNA gene family in different plant genomes indicated that this important negative gene regulation system has undergone dramatic tune-ups in plant genomes.

**Keywords:** gene duplication, genome evolution, *Oryza sativa*, microRNA.

### INTRODUCTION

MicroRNAs (miRNAs) are 20-22 nucleotide (nt) noncoding RNAs that were firstly discovered in *Caenorhabditis elegans* [1, 2]. miRNAs compose a new class of genes that play important roles in silencing target genes and thus in negatively regulating gene expression in plant and animal development [3-5]. In Arabidopsis plants, miR159 and miR172 are involved in the control of leaf morphology and flower development respectively [6, 7]. miRNA genes are characterized by their stem-loop secondary structures. The 20-22 nt miRNAs, also called mature miRNAs, are usually located at the stem region of

larger precursors, or miRNA genes. In plants, the sizes of miRNA precursors range from 60 to a few hundred nt [5, 8]. Mature miRNAs are processed from their precursors by Dicer-like enzymes [9, 10]. These short, single-stranded RNAs are incorporated into the RNA-induced silencing complex (RISC, [11]) and used as a guide to direct post-transcriptional regulation either by cleaving the target mRNA when identical to the pairing site or by repressing protein translation when the miRNA has mismatches with the target sequences [4]. Due to their relative small sizes, many known miRNAs genes in animal genomes form clusters that can be transcribed as a single polycistronic transcript [3, 10, 12, 13]. Co-transcription of functionally different miRNAs provides the opportunity to target several categories of genes simultaneously. The miRNA clusters are conserved in animal genomes as shown by the human miR17 cluster. The human miR17 gene family has three clusters of ~1 kb located on chromosome 7 (three members), chro-

<sup>¶</sup>These authors contributed equally to this work.

\*Correspondence: Long MAO

Tel: +1-815-7537947; Fax: +1-815-7530461;

E-mail: lmao@niu.edu

mosome 13 (six members), and the X chromosome (five members) respectively [14]. Similar clusters are largely conserved in the genomes of mouse, rat, dog, and frog.

To date, hundreds of miRNAs have been identified in various organisms as shown by the collection in the microRNA Registry [15]. In plants, miRNA genes have been readily discovered based on their cross species sequence conservation and stem-loop structures in completely sequenced genomes of Arabidopsis and rice [8, 16, 17]. Recently, rice specific or monocot specific microRNAs have been identified which suggests the differentiated evolution of microRNAs in individual plant genome [18, 19]. So far, the miRNA Registry collected about 28 rice miRNA families from 134 precursors. Thus, one mature miRNA on average have nearly five precursors. When searching the japonica rice Nipponbare DNA sequences using miRNA gene sequences collected in the RNA Registry, we found that most of the rice miRNA genes were scattered around the genome. Except for the miR395 family, only a limited number of miRNA families had members located in close neighborhood and none of them contained three members in a region of 1 kb (Tab. 1). Two clusters of miR395 genes have been reported in the genome of indica rice subspecies [5]. We found, using more complete japonica rice sequences generated by the International Rice Sequencing Project (IRGSP), that a total of 24 miR395 genes in the rice genome were in fact organized into four clusters. We then further studied in detail miR395 gene sequences and their genomic organization

and demonstrated the molecular evolutionary history of this gene family in the rice genome. For comparison, a study of *Medicago truncatula* miR395 gene family was also carried out. The evolution of structural organization and potential functional differentiation of miR395 genes were discussed.

## MATERIALS AND METHODS

### Sequence resource

Rice pseudo-chromosome sequences were downloaded from the TIGR web site (<http://www.tigr.org>). Arabidopsis and rice miRNA sequences both mature and the precursors, were downloaded from the microRNA Registry (<http://www.sanger.ac.uk/Software/Rfam/mirna>). The rice coordinates were obtained by a BlastN [20] search of rice pseudo-chromosome sequences using the miRNA dataset. Medicago BAC sequences were retrieved from GenBank at National Center for Biotechnology Information (NCBI). The Medicago miR395 clusters were on BACs mth2-10i14 (acc# AC146854) for locus a and mth2-28b24 (acc# AC149601) for locus b. The fifth locus of rice miR395 was located on chromosome 11 and was found to be derived from BAC Ba0016J08 and identical to cluster d. The BAC contains three contigs by the time of this analysis and has large pieces of identical sequences with BAC B1151D08 of chromosome 9. Therefore, Ba0016J08 was considered to be mistakenly incorporated into the pseudo-chromosome and excluded for further investigation.

### Multiple sequence alignment, phylogenetic analysis, and secondary structure prediction

Multiple sequence alignment (MSA) was performed using ClustalX [21] with a gap opening penalty value of 5 and an extension penalty value of 2. The low gap penalty values are to maximize the alignment at the loop regions of miRNAs where the similarity was usually low. Most of MSAs use miRNA sequences that were delineated at one end by the mature miRNA sequence and the other by the sequence that form stem with the mature miRNA in the secondary structure. BOXSHADE ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)) was used to display multiple sequence alignments. Dendrograms or phylograms of Neighbor-Joining trees were generated using PAUP\* v4.0b10 [22]. Secondary structures of miRNAs were predicted using *Mfold* [23, 24] web service (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi>).

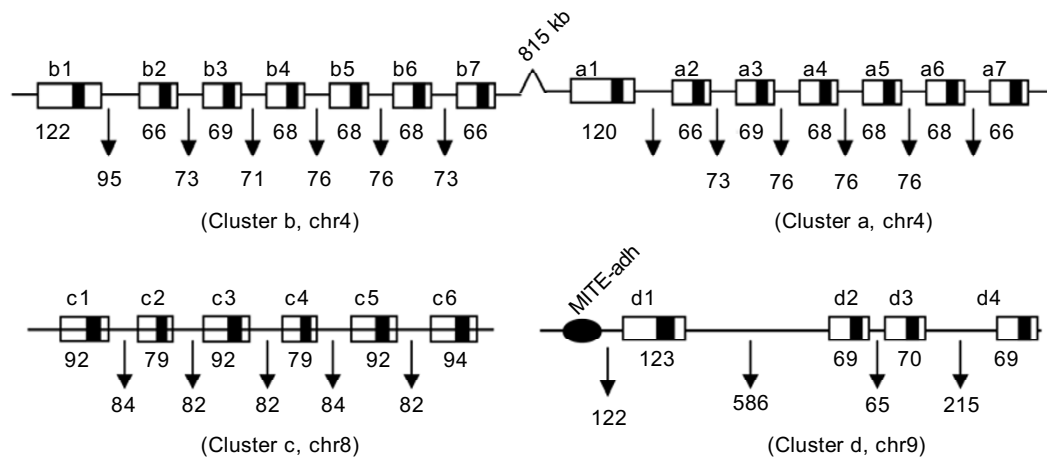
### Rice materials and PCR reactions

Rice leaf materials were obtained from *O. sativa* cv. Nipponbare, *O. nivara* (AA), *O. punctata* (BB), *O. punctata* (tetraploid, BBCC), *O. officinalis* (CC), *O. latifolia* (CCDD), *O. australiensis* (EE), *O. brachyantha* (FF), *O. meyeriana* (GG) and were kindly provided by Dr. Zhu Kuan CHENG of Institute of Genetics and Developmental Biology, Chinese Academy of Sciences. Genomic DNAs were extracted according to the description of [25]. Due to the strong secondary structure of the miR395 loci, nested PCR was carried out when needed to achieve clearer and stronger bands. Primer sequences were selected in the region unique to each cluster such that cluster-specific bands could be amplified. Primers used were: cluster a, 207A-F1: 5' CAT ATT TTT CAA CTG TGG TAG GTG3' and 207A-R1: 5' TTC CCT TCA ATT CAA ATT GTG G 3'; Cluster b, 207B-F1: 5' CTA CTG CAG TAG TAA GCA CTA C3' and 207B-R2: 5' CGT

**Tab. 1** Additional clustered miRNA genes in the rice genome<sup>a</sup>

Gene	Rice pseudo-chromosome	Begin	Distance (bp) to the previous gene
156c	chr1	4662795	
156b	chr1	4663175	380
399k	chr5	25257015	
399c	chr5	25262220	5105
169i	chr8	26638889	
169h	chr8	26643089	4200
169l	chr8	26653228	10139
169q	chr8	26653335	107
169j	chr9	17900693	
169k	chr9	17903965	3272

<sup>a</sup>Only those with two genes located within at least 10 kb are listed. See Fig. 1 for the genomic organization of the miR395 gene family. Gene names follow Jones-Rhoades and Bartel, 2004.



**Fig. 1** Genomic organization of the rice miR395 gene family. Black areas in the boxes represent miR395 location on each gene. Numbers below each box represent the gene size delineated by stem sequences in the secondary structure of miR395 genes. Arrows point to the intergenic sequences.

AGT TTG AAA TAT CAC ATG TTA AAT GCC3'; (for nested PCR) 207B-F2: 5' TGT CCA CTG GAG TTC TCC TCA ATC CAC TTC3' and 207B-R2; Cluster c: 207C-F1: 5' AGG GAG CTG AGC TGA GAA TTA G3' and 207C-R2: 5' GTA GAA TAC AGA TAT ACT ACT GC3'; Cluster d: 207D-F1: 5' GAG AGT AGC CAC CGG CTC TGT G3' and 207D-R2: 5' GAG AGT TCA TTC CCC AAC ACT TCC3'; (for nested PCR) 207D-F2: 5' GTC GAT CTA GGA GTC ATA ATT TGG3' and 207D-R1: 5' ATG ACG TCA AGT GCT TGA AGG G 3'. Standard PCR reaction was performed with 5 min of 94°C followed by 35 cycles 94°C, 1 min, 55°C, 1 min, and 72°C, 2 min. Annealing temperatures were optimized until clear bands were produced.

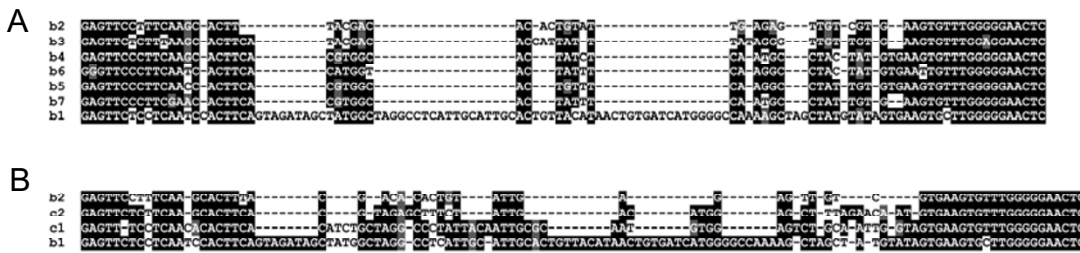
## RESULTS

### Genomic organization of the rice miR395 gene family

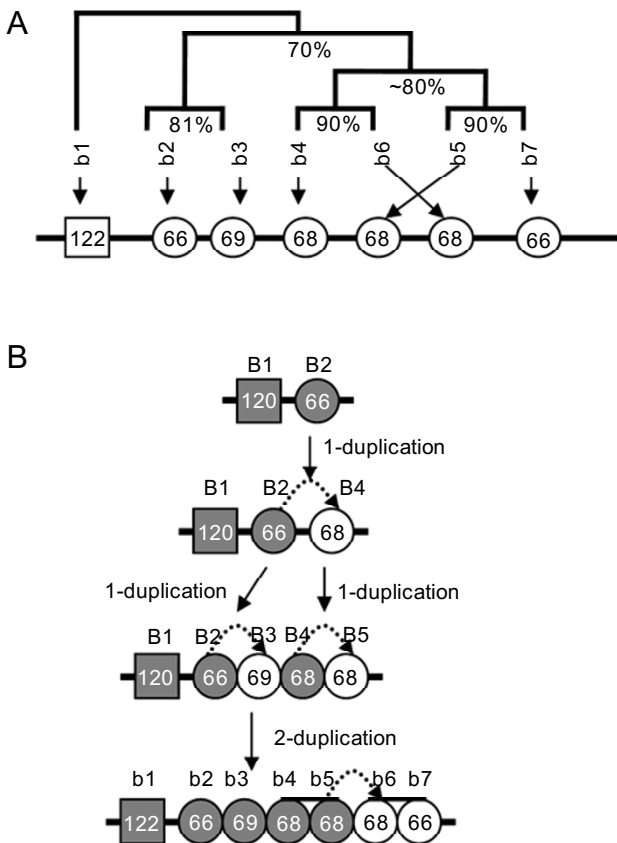
When using known miR395 gene sequences as queries to search the rice genome, we found a total of 24 near perfect matches. These 24 matches were then found to be located on four different BACs. Stem-loop structures were predicted using *mfold* from all the matching regions. To get a better picture of their genetic location, we searched the rice pseudo-chromosomes developed by TIGR (version 2) and found that the four clusters were about 1 kb in size and located on three chromosomes. As shown in Fig. 1, clusters a and b were located on chromosome 4. These two clusters were separated by around 815 kb. Cluster c was on chromosome 8 whereas cluster d on chromosome 9. We further looked at the number of miR395 genes on each cluster and found that clusters a and b had an exact miR395 gene arrangement – one ~120 bp and six ~66 bp. Cluster d contained only four miR395 genes, however, with similar gene sizes to those on clusters a and b: one ~120 bp and three ~66 bp. Cluster c had six miR395 genes with two novel gene sizes – two ~79 bp and four ~92 bp.

### The relationship of miR395 genes on cluster b

Formation of gene families is naturally considered as a result of gene duplications. Tandem gene duplication can involve one gene (so-called 1-duplication), or two genes (so-called 2-duplication), or the whole gene clusters (i.e. segmental duplication, [26]). As mentioned before, cluster b contained miR395 genes of two sizes, one ~120 bp and six ~66 bp. Multiple sequence alignment showed that these miR395 genes had nearly identical sequence at the mature miRNA region and were highly conserved at the region pairing with the mature miRNA (Fig. 2). Gene a1 was the only one among the seven genes on this cluster that was ~120 bp long. The rest was much smaller. In the multiple sequence alignment of genes from this cluster (Fig. 2A), small patches of relatively conserved regions were still observable between the 122 bp gene b1 and the rest 66 bp genes, suggesting that despite of the significant size difference between these two kinds of miR395 genes they may be derived from the same ancestral copy. It is interesting, though, that only the smaller 66 bp copy was significantly amplified to six genes, whereas the larger 122 bp gene remained un-amplified. To study the relationship among these genes, we performed a pair wise sequence similarity comparison. As shown in Fig. 3A, the overall similarity among six ~66 bp genes was higher than 70%, suggesting that they were products of rather recent tandem duplication events. The similarity between b4, b5, b6, and b7 were even higher, with the similarity between b4/b6 and b5/b7 higher than 90%. This was apparently due to a 2-duplication of b4 and b5. According to the above sequence similarity, the evolution history of cluster b was depicted in Fig. 3B. Because b1 was about 122 bp long,



**Fig. 2** Multiple sequence alignment of miR395 genes from cluster b (A) and comparison of the representative genes from clusters b and c (B).



**Fig. 3** A possible evolution history of cluster b. (A) Relationship of cluster b miR395 genes based on their sequence similarity as shown by percentage numbers and (B) A possible evolution path for cluster b. Putative ancestral genes were labeled with capital letters.

nearly double of the size of the rest genes, it is not clear when and how the ancestral 66 bp gene was derived. Therefore, we assume that b1 and at least one 66 bp gene (presumably b2) pre-existed in the ancestor of cultivated rice. Sequence similarity among the genes provides a possible scenario, i.e. b4 could be the result of the first 1-duplication on this cluster. b2 and b4 were respectively duplicated generating b3 and b5. A 2-duplication event in-

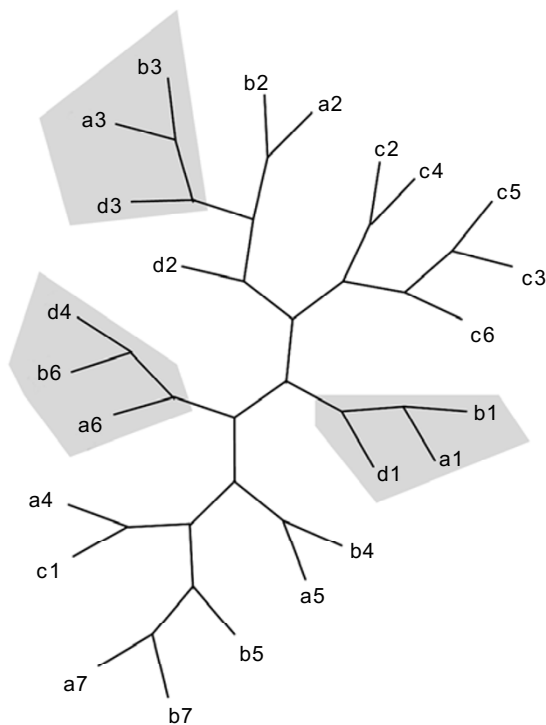
volving b4 and b5 then produced b6 and b7. In summary, cluster b began with two miR395 genes of 120 bp and 66 bp respectively and experienced three rounds of 1-duplication and one round of 2-duplication.

**Clusters a and d were segmental duplicates of cluster b**

Comparison of the genomic sequences corresponding to the four clusters revealed significant similarity between clusters a, b, and d, indicating that these three clusters were derived from segmental duplication. Cluster d was different to clusters a and b by having four, instead of seven, miR395 genes. The similarity of these three clusters extended beyond the first miR395 gene into putative promoter regions, except that cluster d had a ~150 bp insertion in its promoter. A search of the TIGR repeat database showed that the insert represented a miniature inverted transposable element (MITE), MITE-adh [27]. The insertion of this MITE-like element explains the facilitated degeneration of cluster d causing three miR395 genes to degenerate to an extent beyond recognition. Nevertheless, miR395 genes remained on cluster d still had good similarity with genes on the clusters a and b as shown in Fig. 4. A Neighbor-Joining tree was developed for miR395 genes to show the close relationship among members from these three clusters, where corresponding members on clusters a, b and d such as a1, b1, and c1 and a3, b3, and d3 could be clearly grouped together.

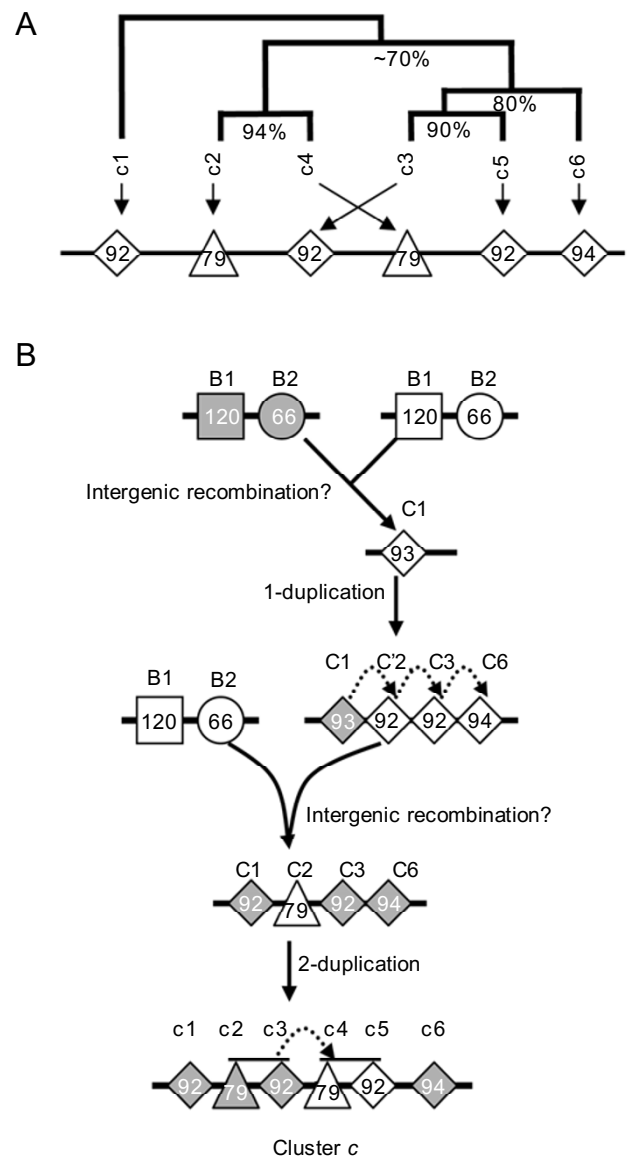
**Cluster c came into being with a different evolutionary path**

Cluster c is interesting because it had low sequence similarity with the other three clusters. Cluster c had miR395 genes of two different sizes. Also, the putative 5' promoter region of cluster c lacked similarity to clusters a, b, and c as well. These characteristics suggested that cluster c was not a duplicate from cluster b. In other words, it could arise with a different evolutionary path. As shown in Fig. 5, miR395 genes on cluster c were in two novel sizes: ~79 bp (c2 and c4) and 92 bp (c1, c3, c5, and c6). Although different in sizes, c1 showed small patches



**Fig. 4 A** Neighbor-Joining tree showing the close relationship between corresponding miR395 genes on clusters a, b, and d (shaded). MiR395 genes from cluster c form a separate clade, except c1.

of similarity with members from cluster b as shown in Fig. 2B, suggesting they were by some means related members. For the appearance of the novel gene sizes of cluster c, a simple arithmetic calculation provided interesting clues: i.e. the length of c1 (92 bp) equals approximately half of the sum of that of b1 (122 bp) and b2 (66 bp), i.e.  $(122 + 66)/2 = 94$ . Similarly, the length of c2 (79 bp) equals approximately half of the length of c1 (92 bp) and b2 (66 bp) added together, i.e.  $(92 + 66)/2 = 79$ . If this arithmetic result is not a sheer coincidence, cluster c members could be assumed arising by intergenic recombination events. Combining the evidence from the sequence similarity among genes on this cluster (Fig. 5A), we propose a plausible scenario of the evolution history for cluster c, as depicted in Fig. 5B. The first founding member of cluster c, the ~92 bp c1, was derived from an intergenic recombination occurred between 120 bp and 66 bp copies of cluster b. C1 then underwent three 1-duplications forming a four-member cluster. A second intergenic recombination then occurred between ancestral gene C2 of 92-bp and a 66-bp copy of cluster b resulting in the first 79 bp gene on cluster c, the current version of gene c2. A final step of a 2-duplication involving c2 and c3 completed the generation procedure of cluster c. The extremely high similarity between cluster c members, 94% identical



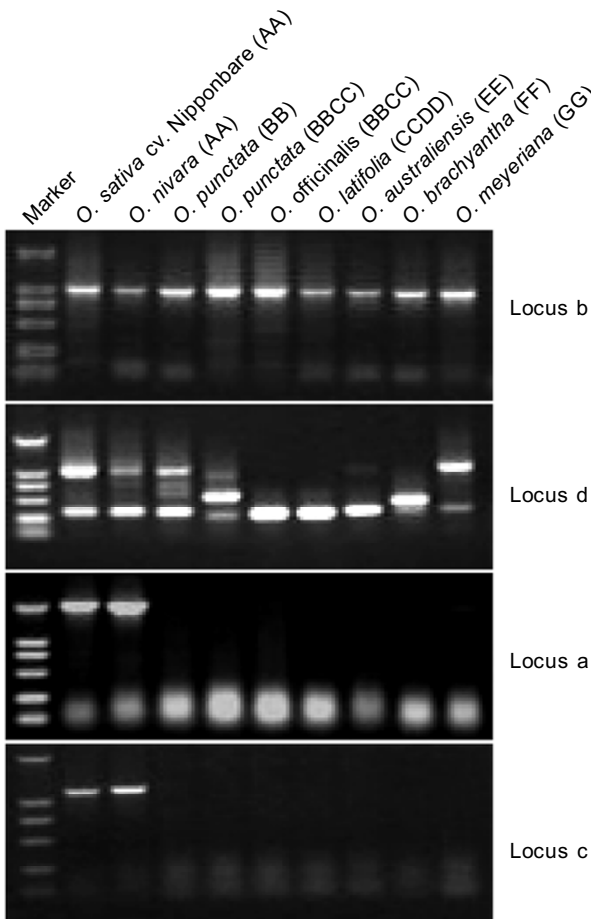
**Fig. 5 A** possible evolution history of cluster c. **(A)** Relationship of cluster b miR395 genes based on their sequence similarity as shown by percentage numbers and **(B)** A possible evolution path for cluster c. Putative ancestral genes were labeled with capital letters.

between c2 and c4 and 92% identical between c3 and c5 (Fig. 5A), indicates that the 2-duplication of c2-c3 and c4-c5 had a very recent origin.

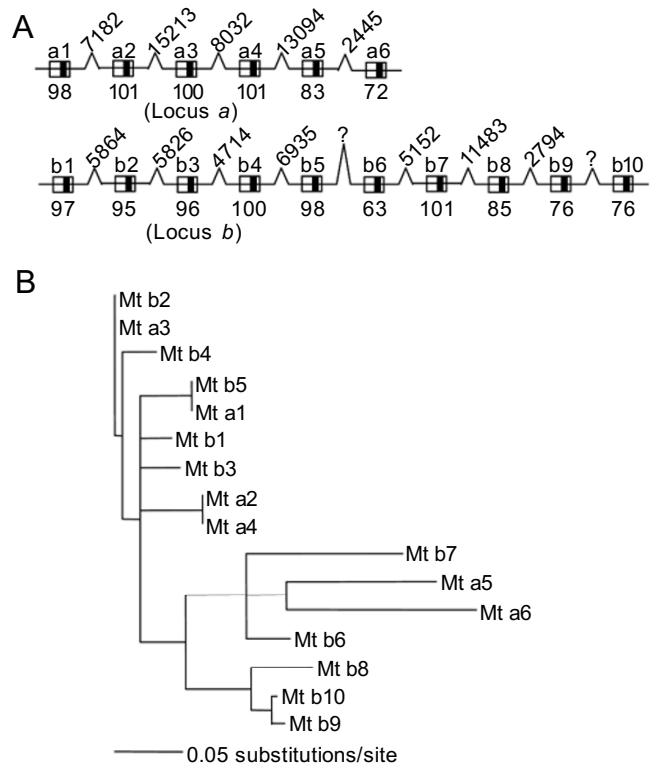
### Organization miR395 gene families in the genus *Oryza*

In order to gain an insight about the timing of the genetic events forming miR395 gene family, we surveyed the genetic loci corresponding to the four miR395 clusters in the genus *Oryza*. In addition to the sequenced *O. sativa* cv. Nipponbare which has a genome designated as AA ( $2n=24$ ), other rice species used include diploid spe-

cies *O. nivara* (AA), *O. punctata* (BB), *O. officinalis* (CC), *O. australiensis* (EE), *O. brachyantha* (FF), *O. meyeriana* (GG), and tetraploid species *O. punctata* (tetraploid, BBCC), *O. latifolia* (CCDD). As shown in Fig. 6, cluster b specific primers were able to amplify a DNA fragment common to all rice species investigated, suggesting that this cluster was established early among the species of the genus *Oryza*. The next cluster that was present in all the species was cluster d. Clusters a and c however were only present in rice species of an AA genome, which is shared by the cultivated rice *O. sativa* Nipponbare and its close relative *O. nivara*. This is consistent with the observation that sequences of cluster d was more divergent than that of clusters b and a, implying that the segmentation duplication leading to cluster d was earlier than the cluster b/a duplication event. Cluster d was divergent among rice species, indicating that this locus had experienced a fast divergence, probably due to the insertion of a copy of transposon-like MITE-adh at its 5' region.



**Fig. 6** A locus survey of four miR395 gene clusters among species of the genus *Oryza*. Molecular weight markers (in bp) from the top to the bottom are 2000, 1000, 750, 500, 250, 100.



**Fig. 7** The *Medicago truncatula* miR395 gene family. (A) Genomic organization. Numbers indicate DNA sequence length in bp and the question marks indicate gaps of unknown distance and (B) A Neighbor-Joining tree displaying the relationships between *Medicago* miR395 genes (see text for detail).

**Genomic organization of the miR395 gene in *M. truncatula***

In *Arabidopsis*, six miR395 genes organized into two clusters, each spanning about 4 kb [5]. The miR395 gene family as shown in this study has significantly expanded in the rice genome, although they were organized in even more compact manner. It will be interesting to know the genomic organization of miR395 gene families in other plant genomes.

A sequence search of the high throughput genomic sequence section of GenBank found a total of 16 miR395 genes in *M. truncatula*, indicating that the miR395 gene family was significantly amplified when compared with the *Arabidopsis* genome. These genes were clustered on two unfinished bacterial artificial chromosome (BAC) clones (mth2-10i14 and mth2-28b24). However, in contrast to their compact organization in the rice genome, *M. truncatula* miR395 genes seemed to widely spread out on two loci (named as a (on mth2-10i14) and b (on mth2-28b24)). As shown in Fig. 7A, locus a contained six members spanning more than 45 kb, whereas locus b consisted of 10 recognizable members that were located in

three sequence contigs of at least 50 kb long (the BAC was still under sequencing). The average distance between neighboring genes was more than 6 kb, with the longest distance of more than 15 kb.

Despite the large distance between the miR395 genes, sequences of some members of *Medicago* miR395 gene family were highly similar, both inside each cluster and between the clusters. For example, among the six members of locus a, genes a1, a2, a3, and a4 were all more than 90% identical, with a2 and a4 100% identical! Between the two loci, sequences of a1 and b5, a3 and b2 were identical or nearly identical to each other (data not shown). This suggests that these two loci could originally be two segmental duplicates before the genes drifted away from each other. Like rice miR395 family, a variety of groups of closely related genes on the dendrogram of *Medicago* miR395 gene family members indicated multiple rounds of duplication events (Fig. 7B).

## DISCUSSION

miRNA-mediated gene regulation is an ancient gene regulation mechanism conserved among all lineages of land plants [28] and even predates the divergence of plant and animal kingdoms. The plant genomes have experienced significant rearrangement as demonstrated by the fact that members of most miRNA gene families were scattered around the plant genomes. Since duplication events have been widely considered as the initial force for gene family formation [26, 29], it is almost certain that these scattered plant miRNA genes were originally located close to each other like rice miR395 genes. Protein genes in tandem arrangement are common in plant genomes [30, 31] and so are segmental duplication [32]. Although sequences related to miRNA function are extremely short (only 21-23 nt), the miRNA families were generated using usual duplication mechanisms common to much larger protein genes. The rice miR395 gene family was amplified through reiterated duplication events involving one gene, two genes, or the entire cluster.

The expansion of miRNA gene families indicated a contiguous effort in the genome as a consequence of positive selection. From Arabidopsis to rice and *Medicago*, the miR395 gene family was significantly increased. The influx of the genes was achieved by various duplication events. Novel genes could be generated by intergenic recombination between extant genes and then subject to similar amplification via gene duplication. Although the expansion of the miR395 gene family seems to be positively selected, a downsizing mechanism was also observed on cluster d, where insertion of MITE-adh at the promoter region seemed to facilitate sequence degradation. Three out of the seven miR395 genes were lost on this cluster

indicating a good correlation between the two events. Further, the insertion of such a transposable like element may cause the whole cluster not being able to be actively transcribed. Therefore, functionally, more miR395 genes were removed from this family.

A cascade of sequence similarity between members of the miR395 family, some of which were even identical, indicates duplication events that occurred at different time during evolution. Many of them seem to be rather recent. Although not absolutely confirmed, at least two clusters of miR395 genes seem to present only in rice species with an AA genome that is shared by the cultivated rice. This may indicate that the influx for miR395 genes was increased in the process of domestication. In other words, the inclusion of additional miR395 genes may contribute to advantageous traits for human selection.

The rice miR395 gene clusters, like human miR17 clusters, were compact in a range of about 1 kb, making it possible to be transcribed as single polycistronic transcripts. In fact, one EST with a GenBank accession number of CA764701 was identical to part of cluster b, suggesting that the seven members of miR395 genes can be transcribed as a single polycistronic unit [5]. The genomic organization of human miR17 clusters were largely conserved among various vertebrate species [14]. In contrast, the compact structure of miR395 gene family was not conserved in *M. truncatula*. Although the sequences where miR395 genes reside are still in the processing of being sequenced, it is clear that miR395 genes were scattered in a range of more than a hundred kb.

miR395 targets ATP sulfurylases that are involved in sulfate assimilation [33]. Sulfur is one of the most essential macronutrients required by plants. The importance to maintain normal and efficient sulfur assimilation is evident. There are three ATP sulfurylase proteins in Arabidopsis [5]. A preliminary GenBank search showed that at least two similar genes in the rice genome that have been annotated with similar function (data not shown). Since Arabidopsis and rice seem to have a similar number of target genes for miR395, the increase of gene copy number may cause a dosage effect should all genes be expressed at the same time in each genome. On the other hand, the diversity in the genomic organizations of miR395 gene families in different plant genomes may provide different spatial and temporal expression patterns of miR395 genes. The spreading of *Medicago* miR395 genes, for example, provided the opportunities for each miR395 gene to accommodate a different promoter. The diverse distribution of miRNA genes explains the tissue specific or even cell type specific expression patterns [34]. It would be interesting to investigate the consequences of the dosage effect and differential expression patterns of miR395 genes

in these different plant genomes.

## ACKNOWLEDGEMENTS

We thank Dr. Tom SIMS for comments on the manuscript. Dr. Zhu Kuan CHENG kindly provided the wild rice materials. This work is supported in part by a grant from Northern Illinois University Foundation to Long MAO, and National Institutes of Health (NIH) grant to Mitrick JOHNS and Long MAO (No. 44-G1A62164), and a grant from the National Natural Science Foundation of China for oversea young scholars to Long MAO (No. 30228022).

Received, July 22, 2005

Revised, July 26, 2005

Accepted, Aug 10, 2005

## REFERENCES

- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 1993; **75**:843-54.
- Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 2001; **294**:858-62.
- Lai EC, Tomancak P, Williams RW, Rubin GM. Computational identification of *Drosophila* microRNA genes. *Genome Biol* 2003; **4**:R42.
- Bartel B, Bartel DP. MicroRNAs: at the root of plant development? *Plant Physiol* 2003; **132**:709-17.
- Jones-Rhoades MW, Bartel DP. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 2004; **14**:787-99.
- Palatnik JF, Allen E, Wu X, *et al.* Control of leaf morphogenesis by microRNAs. *Nature* 2003; **425**:257-63.
- Chen X. A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* 2004; **303**:2022-5.
- Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. *Genes Dev* 2002; **16**:1616-26.
- Grishok A, Pasquinelli AE, Conte D, *et al.* Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 2001; **106**:23-34.
- Lee Y, Jeon K, Lee JT, Kim S, Kim VN. MicroRNA maturation: stepwise processing and subcellular localization. *Embo J* 2002; **21**:4663-70.
- Hammond SM, Bernstein E, Beach D, Hannon GJ. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 2000; **404**:293-6.
- Mourelatos Z, Dostie J, Paushkin S, *et al.* miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev* 2002; **16**:720-8.
- Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. New microRNAs from mouse and human. *RNA* 2003; **9**:175-9.
- Tanzer A, Stadler PF. Molecular evolution of a microRNA cluster. *J Mol Biol* 2004; **339**:327-35.
- Griffiths-Jones S. The microRNA Registry. *Nucleic Acids Res* 2004; **32 Database issue**:D109-11.
- Llave C, Xie Z, Kasschau KD, Carrington JC. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 2002; **297**:2053-6.
- Park W, Li J, Song R, Messing J, Chen X. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 2002; **12**:1484-95.
- Sunkar R, Zhu JK. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* 2004; **16**:2001-19.
- Sunkar R, Girke T, Jain PK, Zhu J-K. Cloning and Characterization of MicroRNAs from Rice. *Plant Cell* 2005; **17**:1397-411.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-10.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994; **22**:4673-80.
- Swofford DL. 4.0 ed. Sinauer Associates, Sunderland, Massachusetts. 2002.
- Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 1999; **288**:911-40.
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; **31**:3406-15.
- McCouch SR, Kochert G, Yu ZH, *et al.* Molecular mapping of rice chromosomes. *Theor Appl Genet* 1988; **76**:815-29.
- Elemento O, Gascuel O, Lefranc M-P. Reconstructing the duplication history of tandemly repeated genes. *Mol Biol Evol* 2002; **19**:278-88.
- Hass BL, Pires JC, Porter R, Phillips RL, Jackson SA. Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). *Theor Appl Genet* 2003; **107**:773-82.
- Floyd SK, Bowman JL. Gene regulation: ancient microRNA target sequences in plants. *Nature* 2004; **428**:485-6.
- Zhang P, Min W, Li WH. Different age distribution patterns of human, nematode, and *Arabidopsis* duplication genes. *Genes Dev* 2004; **342**:263-8.
- Mayer K, Murphy G, Tarchini R, *et al.* Conservation of microstructure between a sequenced region of the genome of rice and multiple segments of the genome of *Arabidopsis thaliana*. *Genome Res* 2001; **11**:1167-74.
- Bancroft I. Duplicate and diverge: the evolution of plant genome microstructure. *Trends Genet* 2001; **17**:89-93.
- AGI. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000; **408**:796-815.
- Rotte C, Leustek T. Differential subcellular localization and expression of ATP sulfurylase and 5'-adenylylsulfate reductase during ontogenesis of *Arabidopsis* leaves indicates that cytosolic and plastid forms of ATP sulfurylase may have specialized functions. *Plant Physiol* 2000; **124**:715-24.
- Babak T, Zhang W, Morris Q, Blencowe BJ, Hughs TR. Probing microRNAs with microarrays: Tissue specificity and functional inference. *RNA* 2004; **10**:1813-9.