Genome-Wide Analysis of Polycistronic MicroRNAs in Cultivated and Wild Rice

Patricia Baldrich[1], Yue-le Caroline Hsing[2], and Blanca San Segundo[1,*]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Edifici CRAG, Campus UAB, Bellaterra (Cerdanyola del Vallés), Barcelona, Spain

[2]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

*Corresponding author: E-mail: blanca.sansegundo@cragenomica.es.

Accepted: March 7, 2016

Abstract

MicroRNAs (miRNAs) are small noncoding RNAs that direct posttranscriptional gene silencing in eukaryotes. They are frequently clustered in the genomes of animals and can be independently transcribed or simultaneously transcribed into single polycistronic transcripts. Only a few miRNA clusters have been described in plants, and most of them are generated from independent transcriptional units. Here, we used a combination of bioinformatic tools and experimental analyses to discover new polycistronic miRNAs in rice. A genome-wide analysis of clustering patterns of *MIRNA* loci in the rice genome was carried out using a criterion of 3 kb as the maximal distance between two miRNAs. This analysis revealed 28 loci with the ability to form the typical hairpin structure of miRNA precursors in which 2 or more mature miRNAs mapped along the same structure. RT-PCR provided evidence for the polycistronic nature of seven miRNA precursors containing homologous or nonhomologous miRNA species. Polycistronic miRNAs and candidate polycistronic miRNAs are located across different rice chromosomes, except chromosome 12, and resided in both duplicated and nonduplicated chromosomal regions. Finally, most polycistronic and candidate polycistronic miRNAs showed a pattern of conservation in the genome of rice species with an AA genome. The diversity in the organization of *MIR* genes that are transcribed as polycistrons suggests a versatile mechanism for the control of gene expression in different biological processes and supports additional levels of complexity in miRNA functioning in plants.

Key words: genome duplication, *Oryza sativa*, polycistronic miRNAs, target gene, wild rice.

Introduction

Rice is one of the world's most important food crops and a main food source for more than half of the world's population. The genus *Oryza* comprises 2 cultivated and 22 wild species representing different genome types, both diploid and tetraploid (AA, BB, CC, EE, FF, GG, BBCC, CCDD, KKLL, and HHJJ) genomes. The AA genome includes two cultivated species, *Oryza sativa* (Asian rice) and *Oryza glaberrima* (African rice), and their closest relatives and ancestors. The *Oryza rufipogon* (perennial) and *Oryza nivara* (annual) species have been proposed to be the direct ancestors of *O. sativa* (Sweeney and McCouch 2007; Dogara and Jumare 2014). *Oryza glaberrima* was independently domesticated from the wild progenitor *Oryza barthii* after the domestication of Asian rice (Wang et al. 2014). Furthermore, the cultivated rice varieties belonging to the *O. sativa* group include *japonica* and *indica* subspecies. *Oryza sativa japonica* was first domesticated around the middle region of the Pearl River in Southern China and that *O. sativa indica* rice was subsequently developed from crosses between *japonica* rice and local wild rice as the initial cultivars spread into southern Asia (Huang et al. 2012). In addition to its long history of natural selection and domestication, rice represents the model plant for research in monocotyledoneous species. The genome sequence is available for cultivated rice varieties (*indica* and *japonica*) and wild rice species.

MicroRNAs (miRNAs) are small noncoding RNAs that have emerged as important players in posttranscriptional gene silencing (Jones-Rhoades et al. 2006). They have critical roles during plant development and adaptation to environmental stress (Aukerman and Sakai 2003; Palatnik et al. 2003; Mallory et al. 2004; Chiou et al. 2006; Navarro et al. 2006; Sunkar et al. 2007). Genes encoding miRNAs (*MIR* genes) are transcribed as long primary transcripts (pri-miRNAs) with unique stem-loop structures that are sequentially processed by the

RNase III DICER-like (DCL) enzyme, typically DCL1, which results in a miRNA-5p/miRNA-3p duplex (also named miRNA/miRNA*) (Kurihara and Watanabe 2004; Arikit et al. 2013). The duplex is then exported to the cytoplasm, where the active miRNA sequence is loaded into the RNA-induced silencing complex and directs cleavage or translational repression of the target mRNA (Llave et al. 2002; Brodersen et al. 2008).

The prevalent model to explain the origin of plant miRNAs is that they arise from inverted duplication of their target genes, which generates a proto-*MIR* gene (Allen et al. 2004; Fahlgren et al. 2007; Axtell and Bowman 2008). Accumulation of mutations would shape the proto-*MIR* into a young *MIR* gene and, eventually, an ancient *MIR* gene. In addition, the spontaneous evolution from random sequences with hairpin-like structures or derivation from transposable elements has been proposed to explain the origin of plant miRNAs (Felippes et al. 2008; Piriyapongsa and Jordan 2008). Evidence for frequent birth and death of Arabidopsis *MIRNA* genes exists (Fahlgren et al. 2007; Nozawa et al. 2012) which proposed that the number and diversity of miRNAs has changed in a lineage-specific manner after the divergence of dicotyledonous and monocotyledonous species.

Processes driving expansion and evolution of miRNAs in plants include whole-genome duplication (polyploidization), duplication of subchromosomal regions (segmental duplications), and local duplications that involve one or two genes (tandem duplications). Segmental and tandem duplication events occurred during the evolution of the rice genome (Guyot and Keller 2004; Thiel et al. 2009), and members of distinct miRNA families have been found to locate in duplicated genomic sequences in rice (Jiang et al. 2006). Duplication events may also lead to diversification in the expression of miRNAs (Baker et al. 2005; Guo et al. 2005; Campo et al. 2013).

Occurrence of miRNA clusters is common in animals, but only a few miRNA clusters have been found in plants (i.e., miR166, miR169, and miR395) (Boualem et al. 2008; Zhao et al. 2009; Calviño and Messing 2013; Patanun et al. 2013; Barik et al. 2014). Evidence in animals also supports that clustered miRNAs can be independently transcribed or simultaneously transcribed into single polycistronic transcripts (Altuvia et al. 2005). The precursor structures of these polycistronic miRNAs might contain copies of the same miRNA family member (e.g., homologous miRNAs) or unrelated miRNAs (e.g., nonhomologous miRNA clusters). Plant miRNA precursors are generally much longer and heterogeneous than animal miRNA precursors. It is then plausible that long miRNA precursors in plants might contain more than one miRNA that can be transcribed as a single transcriptional unit. However, a systematic analysis of polycistronic miRNAs in plants is still lacking.

In this work, we present evidence for the existence of polycistronic miRNA precursors encoding homologous or nonhomologous miRNAs in rice. We discuss the importance of the genomic organization of miRNAs as polycistrons for regulation of gene expression. The chromosomal distribution of polycistronic miRNAs was investigated in the context of the duplication history of the rice genome. The conservation of polycistronic miRNAs in *Oryza* species, both cultivated and wild species, is presented.

Materials and Methods

RNA Isolation and RT-PCR Analysis

Rice (*O. sativa* cv. Nipponbare) plants were grown at 28 ± 2°C with a 16-h/8-h light/dark cycle. Leaves from 3-week-old plants were harvested and used for total RNA extraction with TriReagent (Ambiont), followed by DNase treatment (Turbo Free DNase, Ambiont). Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA). Nested RT-PCR (reverse transcription- polymerase chain reaction) was used for obtaining polycistronic DNA sequences with specific primers for each polycistron (supplementary table S3, Supplementary Material online). Sequencing DNA fragments confirmed the specific amplification of the entire precursor sequences.

Expression Analysis of miRNAs Encoded in Policistronic miRNA Precursors

MiRNA expression data were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/) and described by Baldrich et al., (2015). Data sets correspond to transcript profiles of leaves and roots of 3-week-old rice plants (accession numbers GSM1626119, GSM1626120, GSM1626121, GSM1626131, GSM1626132, and GSM1626133). We considered miRNA species with ≥50 reads in the Illumina sequencing data. The degradome sequencing data sets were also retrieved from GEO (accession numbers GSM1626143 and GSM1626144).

Genome-Wide Identification of Polycistronic miRNAs

Mature miRNAs were downloaded from the miRBase registry (http://www.mirbase.org, version 21). To identify candidate polycistronic miRNAs, we used a stringent criterion of 3 kb as the interval between mature miRNAs. Distances between miRNAs were calculated by subtracting the start coordinate of the downstream miRNA and the end coordinate of the upstream miRNA. Secondary structure prediction involved use of RNAFold with default parameters (Vienna 2.1.0; http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). Following the current annotation, the miRNAs originating from opposite arms of the same miRNA precursor are denoted with a -5p or -3p suffix. Nucleotide sequences were retrieved by using miRBase annotation coordinates. MEGA 6.0 (http://www.megasoftware.net/; Tamura et al. 2013) was used for sequence alignment of miR1861 precursors.

To investigate whether the polycistronic miRNAs arose or evolved from segmental duplication events, we examined their chromosomal location. For this, the genomic map of duplicated blocks in the rice genome was taken from Thiel et al. (2009).

Conservation Analysis of Polycistronic miRNAs in Rice Species

Genome sequences for *O. sativa* ssp *japonica* (cv Nipponbare) (AA genome), *O. sativa* ssp *indica* (AA), *O. rufipogon* (AA), *O. barthii* (AA), *Oryza glumaepatula* (AA), *O. nivara* (AA), *O. glaberrima* (AA), *Oryza punctata* (BB) and *Oryza brachyantha* (FF) were retrieved from Ensembl Plants release 25 (http://plants.ensembl.org/index.html). Orthologous polycistronic miRNA precursors were identified by using a similarity criterion with BLAST release 27 (Camacho et al. 2009) with default conditions.

Results

In this study, we performed a genome-wide analysis of clustering patterns for *MIRNA* loci representing potential polycistronic miRNA precursors in the rice genome. To prevent overestimation of the number of clustered miRNAs, we used a stringent criterion of a maximal distance of 3 kb between same-strand consecutive miRNAs. This analysis scheme was previously used to identify clustered miRNAs in the human genome (Altuvia et al. 2005). Then, we analyzed the genomic region encompassing each of these miRNA clusters for its ability to form hairpin-forming precursor structures. In this way, 28 loci that contain two or more annotated miRNAs mapping into a common precursor structure were identified (table 1; their RNA secondary structures and nucleotide sequences are shown in supplementary figs. S1 and S2, Supplementary Material online). We named these precursor structures as candidate polycistronic miRNAs. Among them, there were precursor structures with homologous miRNAs (e.g., precursors encoding two or more members of the same family), and precursors encoding miRNAs belonging to different miRNA families (table 1).

A region containing 15 miR2118 sequences spanning approximately 20 kbs was identified at chromosome 4 (supplementary fig. S3, Supplementary Material online). For prediction of secondary structures, we considered regions encompassing miR2118 mature miRNAs within this super cluster located at a maximum distance of the 3 kb. In this way, five precursors containing two or three mature miR2118 sequences in the same structure were identified (table 1 and supplementary figs. S1–S3, Supplementary Material online). A clustering pattern of miR2118 sequences was previously reported in *Brachypodium dystachium*, *Panax ginseng*, and *Panicum virgatum* (Xie et al. 2010; Wu et al. 2012; Jeong and Green 2013). Similarly, miR395 appears to be organized in compact clusters in plants (Guddeti et al. 2005; Merchan et al. 2009).

To note, we identified four candidate polycistronic miRNAs encoding nonhomologous miRNAs. They were miR1423-1868, miR1876-1862d, miR5147-437, and miR6255-6253 (table 1 and supplementary figs. S1 and S2, Supplementary Material online). Two precursors each comprising two sequences that were annotated in miRBase as two different members of the miR1440 family or the miR5512 family (e.g., "a" and "b" members in each family) were also found in this study (table 1 and supplementary figs. S4 and S5, Supplementary Material online). However, folding analysis revealed that the two miRNA sequences annotated in miRBase represented the miRNA-5p/miRNA-3p species for the corresponding miRNA mapping opposite to each other with the characteristic 2-nt 3′-overhangs (a signature of DICER cleavage for miRNA/miRNA* duplexes) (Kozomara and Griffiths-Jones 2014). Therefore, the miRNA annotation for these particular miRNAs is likely incorrect, and these sequences can no longer be considered different family members.

By using the same criterion of 3 kb as the maximal distance between 2 miRNAs, 14 of the 28 candidate polycistronic miRNAs identified in the rice genome were also found in the genome of other plant species, both monocotyledonous and dicotyledonous species (table 1). For this analysis, plant species with the highest number of miRNAs registered in miRBase were considered. Thus, 14 candidate polycistronic miRNAs, including the 4 precursors containing nonhomologous sequences, were found only in the rice genome (table 1). Whether these candidate polycistronic miRNA loci are present in plant species not included in this analysis or are not clustered by the criteria used in this study remains to be determined.

To investigate whether candidate polycistronic miRNAs are transcribed as polycistronic units, we carried out nested RT-PCR followed by sequencing of PCR products. Total RNA was isolated from leaves of 3-week-old rice plants. We confirmed transcripts for seven miRNA precursors, each comprising the expected mature miRNAs, thus supporting that these miRNA precursors are transcribed as single transcriptional units (fig. 1). The experimentally validated polycistronic miRNAs included three homologous (miR166k-166h, miR1428e-1428d, and miR1861b-1861c) and four nonhomologous (miR1423-1868, miR1876-1862d, miR5147-437, and miR6255-6253) polycistronic precursors. Their corresponding secondary structures are presented in figure 2. A search in small RNA databases, Expressed Sequence Tag (EST), and transcriptomic data sets did not give indication on the polycistronic transcription of any other candidate polycistronic miRNAs. Further studies are then needed to provide evidence on the polycistronic nature of the precursor structures that have not been experimentally validated in this work. As only RNA samples from rice leaves were assayed in this work, a spatiotemporal or conditional expression pattern of these candidate rice tissues might well

Table 1

MiRNA Precursor Structures Containing Two or More miRNAs in Rice

	miRNA 1	miRNA 2	miRNA 3	miRNA 4	miRNA 5	miRNA 6	miRNA 7	Chromosome	Start	End	Length (nt)	Conservation
Homologous miRNAs	miR156c	miR156b						1	4,665,975	4,666,516	541	Sb–Gm
	miR166k	**miR166h**						2	3.2×10^7	3.2×10^7	289	Sb–Gm–Zm–At–Mt
	miR169j	miR169k						9	2×10^7	2×10^7	3427	Sb–Gm–Zm–At–Mt
	miR169m	miR169l	miR169q					8	2.7×10^7	2.7×10^7	3698	Sb–Gm–Zm–At–Mt
	miR395a	miR395b	miR395c	miR395d	miR395e	miR395f	miR395g	4	3.2×10^7	3.2×10^7	1002	Sb–Gm–Zm–At–Mt
	miR395h	miR395i	miR395j	miR395k	miR395y	miR395l		8	3,299,144	3,300,090	946	Sb–Gm–Zm–At–Mt
	miR395t	miR395u	miR395v	miR395w				9	6,606,291	6,607,495	1204	Sb–Gm–Zm–At–Mt
	miR399c	miR399h						5	2.6×10^7	2.6×10^7	2096	Sb–Zm–At–Mt
	miR399e	miR399a						1	3×10^7	3×10^7	1172	Sb–Zm–At–Mt
	miR1428e	**miR1428d**						3	2.3×10^7	2.3×10^7	521	—
	miR1861b	**miR1861c**						2	2.1×10^7	2.1×10^7	318	—
	miR1861d	miR1861e						4	6,536,337	6,536,647	310	—
	miR1861f	miR1861g						5	1.9×10^7	1.9×10^7	319	—
	miR1861h	miR1861p						6	2.7×10^7	2.7×10^7	311	—
	miR1861k	miR1861j						8	1.5×10^7	1.5×10^7	307	—
	miR1861l	miR1861m						9	1.4×10^7	1.4×10^7	318	—
	miR2118c	miR2118d	miR2118e					4	2.2×10^7	2.2×10^7	637	—
	miR2118f	miR2118g						4	2.2×10^7	2.2×10^7	382	Zm–Sb–Mt–Gm
	miR2118h	miR2118i						4	2.2×10^7	2.2×10^7	421	Zm–Sb–Mt–Gm
	miR2118k	miR2118l						4	2.2×10^7	2.2×10^7	413	Zm–Sb–Mt–Gm
	miR2118n	miR2118o						4	2.2×10^7	2.2×10^7	401	Zm–Sb–Mt–Gm
	miR2118p	miR2118r	miR2118q					11	7,807,433	7,810,885	3452	Zm–Sb–Mt–Gm
	miR5143b	miR5143a						1	8,415,808	8,417,315	1507	—
	miR5534a	miR5534b						11	6,404,295	6,404,519	224	—
Nonhomologous miRNAs	**miR1423**	**miR1868**						4	2×10^7	2×10^7	499	—
	miR1876	**miR1862d**						10	4,833,365	4,833,760	395	—
	miR5147	**miR437**						2	1.7×10^7	1.7×10^7	689	—
	miR6255	**miR6253**						7	2.5×10^7	2.5×10^7	306	—
Misannotated miRNAs	miR1440-5p/miR1440-3p[a]						9	5,980,506	5,980,694	188	—	
	miR5512-5p/miR5512-3p[b]						4	1.8×10^7	1.8×10^7	134	—	

Note.—The criterion of a 3-kb distance between consecutive miRNAs was used to identify candidate polycistronic miRNAs. This analysis was based on miRNAs annotated in the miRBase registry (release 21). For miR1861, a precursor containing miR1861h-1861p was identified, with miR1861p not registered in the current version of miRBase (Baldrich et al. 2015). The precursor structures and nucleotide sequences are given in supplementary figures S1 and S2, Supplementary Material online, respectively. In bold are polycistronic miRNAs experimentally validated in this study by RT-PCR. The chromosomal coordinates of the precursor comprising two or more miRNAs (start and end) are indicated. Conservation among different plant species, both monocotyledonous (Zm, *Zea mays*; Sb, *Sorghum bicolor*) and dicotyledonous (At, *Arabidopsis thaliana*; Mt, *Medicago truncatula*; Gm, *Glycine max*), is shown.

[a]Annotated in miRBase as miR1440b and miR1440a.

[b]Annotated in miRBase as miR5512a and miR5512b.

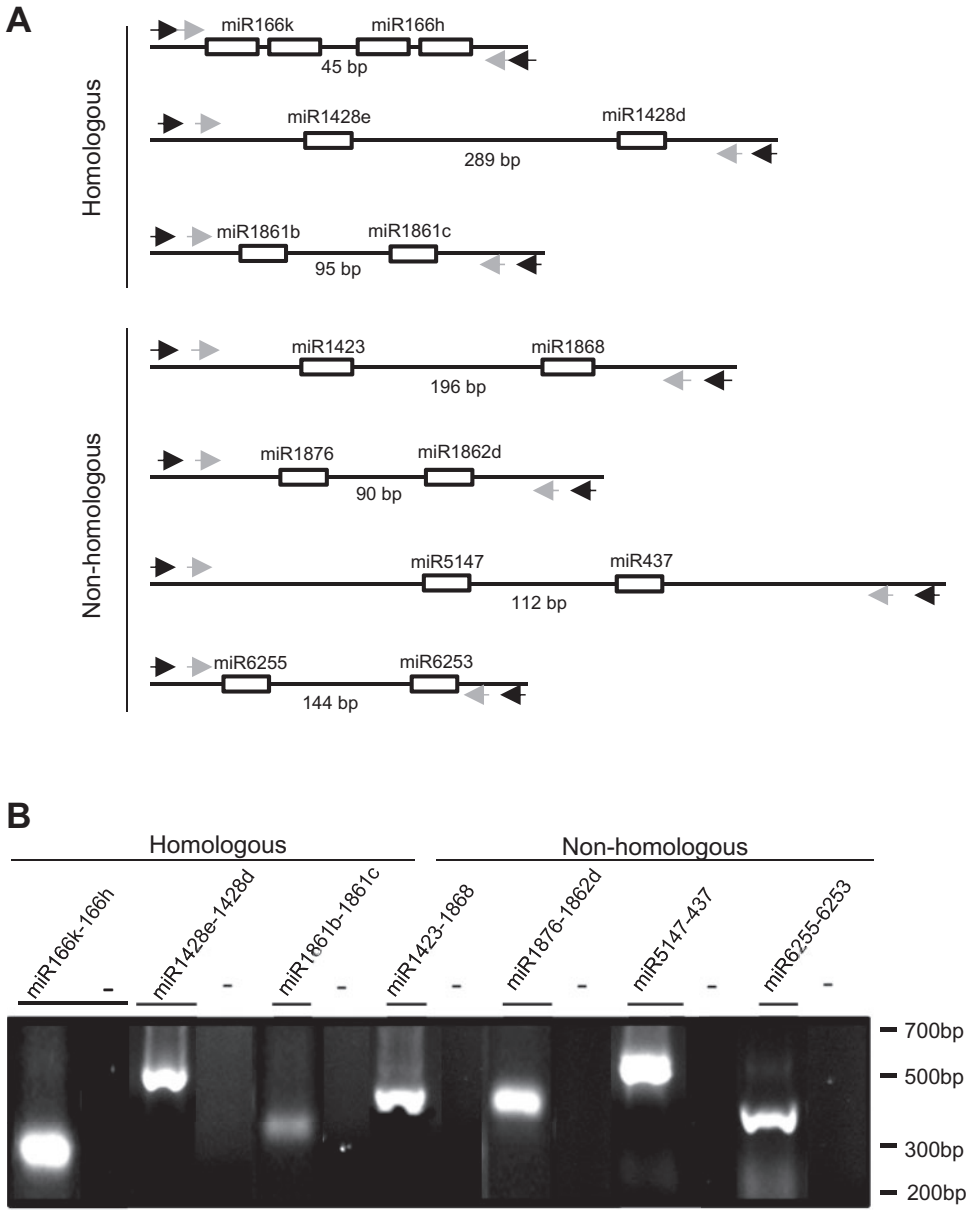

Fig. 1.—Experimental validation of polycistronic miRNAs in rice. (*A*) Schematic representation of miRNA precursors showing the positions of mature miRNAs (boxes) and distances between them. Arrows indicate the position of primers used for nested RT-PCR in *B*. (*B*) Nested RT-PCR detection of polycistronic miRNA transcripts. Total RNAs were extracted from leaves of 3-week-old rice seedlings, subjected to DNase treatment and used for cDNA synthesis. (−) negative controls (samples with no reverse transcriptase added).

explain our failure to detect transcripts for other candidate polycistronic miRNAs. Also, the complex secondary structure of some of these precursors or their low-level expression might hamper their detection in rice tissues. Finally, due to the high similarity in nucleotide sequence of miR1861 precursors (supplementary fig. S6, Supplementary Material online) the presence of less abundant miR1861 polycistronic transcripts other than those encoding miR1861b-1861c in the same PCR reaction cannot be excluded.

On the other hand, coexpression of clustered miRNAs has been widely used as evidence that they derive from a common primary transcript in eukaryotes (Bartel, 2004). In order to explore the potential for coexpression of miRNAs contained in polycistronic precursors validated in this work, we interrogated publicly available miRNA profiling data sets for leaf and root tissues of 3-week-old rice plants (miRNA species with $\geq$50 reads in Illumina sequencing data sets were condirered). This analysis demonstrated that mature miRNAs in the miR166k-166h, miR1423-1868, and miR1876-1862d polycistrons are indeed coexpressed in leaf and root tissues (fig. 3). In the case of miR5147-437, the mature miRNA species located in the 5′ region of this precursor (e.g., miR5147) was barely

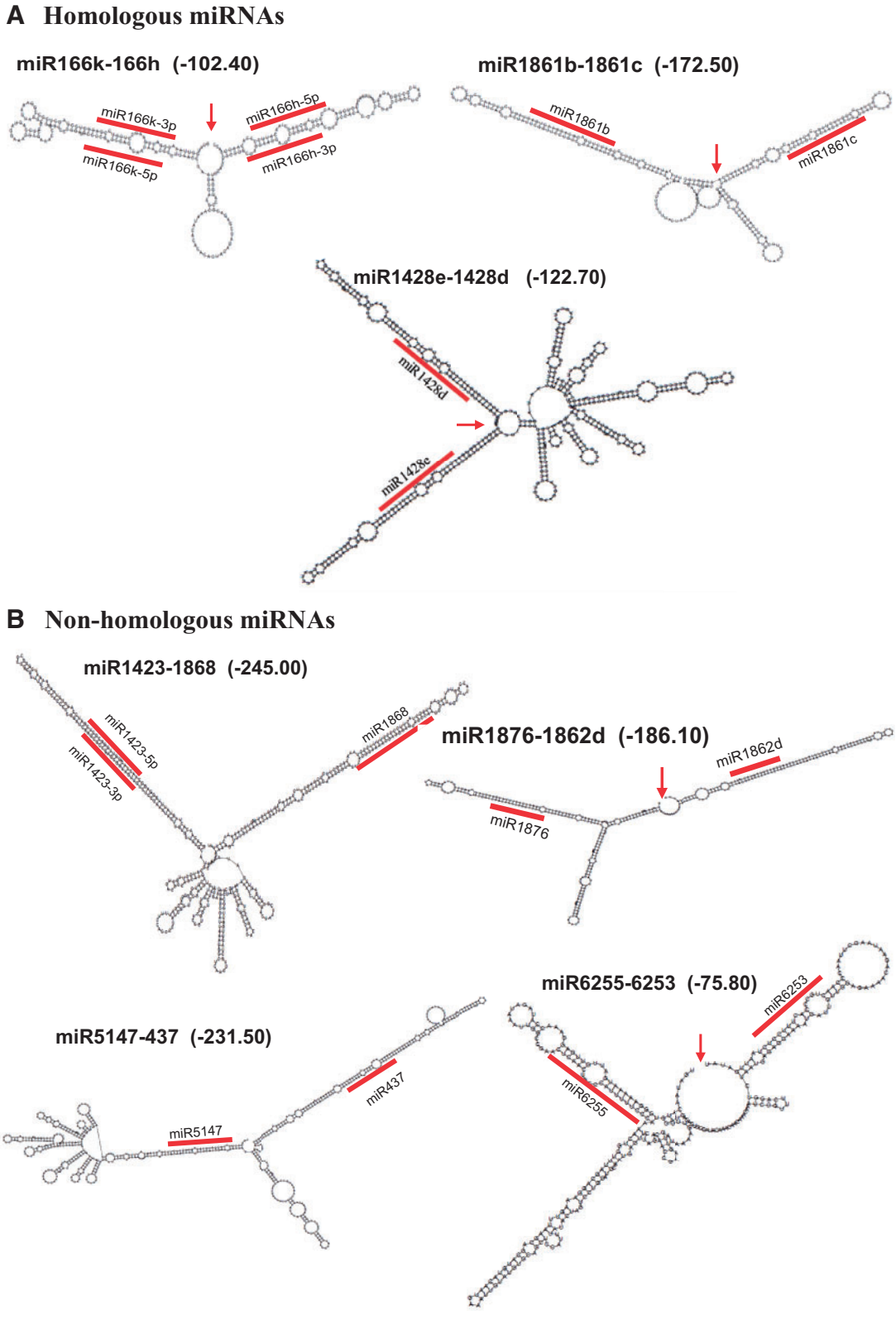

Fig. 2.—Secondary structures of polycistronic miRNAs experimentally validated in this work. (*A*) Homologous polycistronic miRNAs. (*B*) Nonhomologous polycistronic miRNAs. Nucleotide sequences are shown in supplementary figure S2*A* and *B*, Supplementary Material online. Mature miRNA sequences are indicated with a red line. Minimal free energy for each structure is indicated in parentheses.

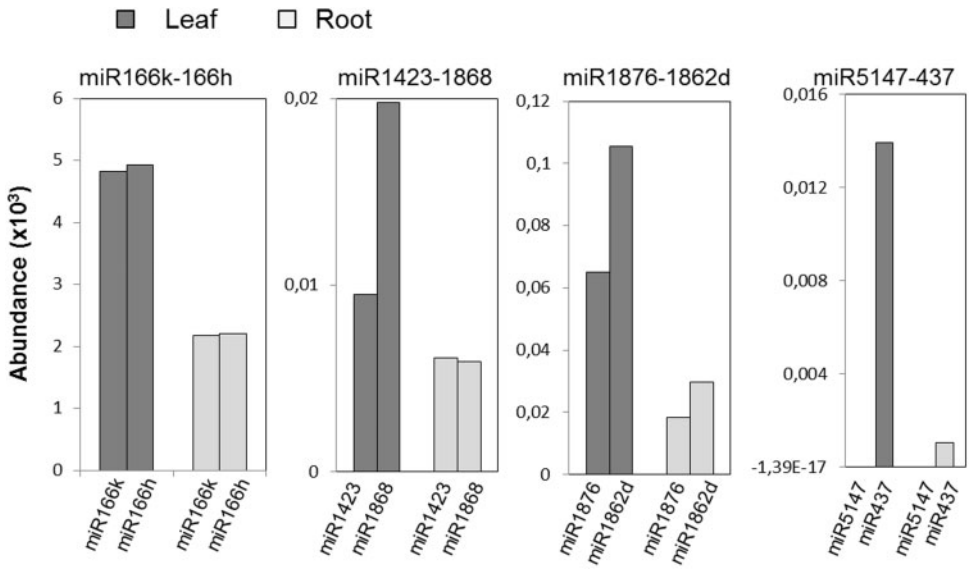

Fig. 3.—Coexpression of miRNAs produced by polycistronic precursors in rice. Expression in leaf and root tissues. Histograms indicate the abundance of mature miRNAs encoded in each polcistronic precursor in leaves (dark gray) or roots (light gray), as revealed by Solexa/Illumina sequencing data sets deposited in GEO (accession numbers GSM1626119, GSM1626120, GSM1626121, GSM1626131, GSM1626132, and GSM1626133).

detected. This observation might reflect a different turnover rate of mature miRNAs derived from this polycistron. Posttranscriptional regulatory mechanisms might also explain a differential accumulation of miRNA species produced by a common precursor transcript. Unfortunately, very little is known about the posttranscriptional regulation of polycistronic miRNAs. Additionally, the structure of primary transcripts (pri- and/or pre-miRNAs) for polycistronic precursors might establish a differential processing (and hence a differential accumulation) of individual miRNAs. As an additional complexity, the individual miRNAs of a particular polycistronic miRNA might also be independently transcribed from other loci (as well as being transcribed as polycistrons).

A search in the literature and publicly available information on degradome analysis of rice tissues allowed us to examine the target genes for mature miRNAs encoded by the experimentally validated polycistronic miRNAs (supplementary table S1, Supplementary Material online). The target genes for miRNAs encoded by homologous polycistronic miRNA precursors are known to regulate the expression of genes that have diverse functions during development and stress responses. For instance, miR1867 family members are known to target glyoxalase transcripts, glyoxalases playing an important role in oxidative stress tolerance by recycling reduced glutathione (Kaur et al. 2014). As for the target genes for miRNAs encoded by nonhomologous polycistronic miRNAs, they are known to be involved in the control of a range of biochemical processes, such as calcium signaling, protein synthesis, and degradation and transcriptional regulation (supplementary table S1, Supplementary Material online). In some cases, however, the target gene for miRNAs encoded by nonhomologous polycistronic miRNAs has not been identified, whereas in other cases, no function has been assigned for the target gene.

Next, we examined the chromosomal distribution of rice polycistronic miRNAs. Also, by taking advantage of the known duplication history of the rice genome (Guyot and Keller 2004; Yu et al. 2005), we investigated the relationship between the location of polycistronic miRNAs and the designated genome-wide intra- and interchromosomally duplicated regions. Except for chromosome 12, the various polycistronic miRNAs (and candidate polycistronic miRNAs) distributed among the different rice chromosomes. They are located in both duplicated (large and small segmental duplications) and nonduplicated regions of the reference genome for *O. sativa* (*japonica* cv Nipponbare) (fig. 4). For instance, miR169-containing precursor structures are located in the duplicated regions of chromosomes 8 and 9. Chromosomes 8 and 9 have been reported to have extensive collinearity of marker loci arrangement, which indicates a possible common origin (Wang et al. 2000). Two miR1861 precursors are located within chromosomal regions with large genomic duplication events (chromosomes 8 and 9), whereas the other miR1861 precursors are located in regions with no segmental duplication history (chromosomes 2, 4, 5, and 6) (fig. 4). Thus, polycistronic miRNAs identified in this study are located in both duplicated and nonduplicated regions of the reference genome for *O. sativa* cv. Nipponbare.

To further explore the evolutionary history of the rice polycistronic miRNAs (and candidate polycistronic miRNAs), we investigated their presence/absence in the genome of *Oryza* species from different geographical regions representing the

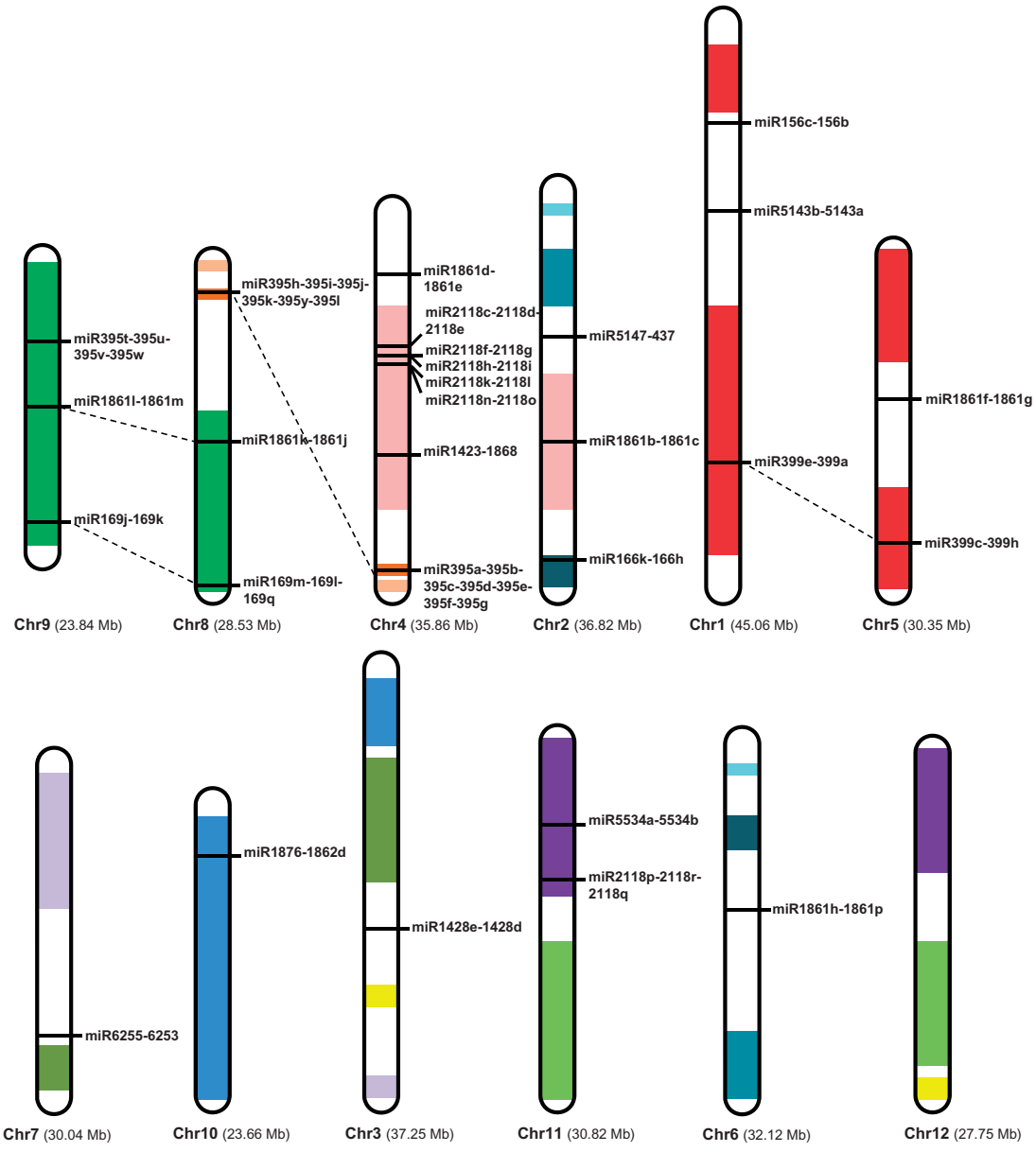

Fig. 4.—Chromosomal location of polycistronic miRNAs and candidate polycistronic miRNAs in rice. The relative locations of polycistronic miRNAs identified by 3-kb distance between mature miRNAs is shown. Segmentally duplicated regions in the 12 rice chromosomes are indicated by the same color. Connecting lines indicate a correlation between duplicated regions and the presence of the indicated miRNA precursors.

AA genome, both cultivated and wild species. The *Oryza* species examined were the following: *O. sativa* spp. *japonica* (cv. Nipponbare), *O. sativa* spp. *indica* (cv. 9311), and their wild relatives *O. rufipogon* and *O. nivara*; *O. glumaepatula* (cultivated in Central and South America) and *O. glaberrima* (African rice) and its wild relative *O. barthii*. Additionally, *Oryza* species representing the BB (*O. punctata*) and FF (*O. brachyantha*) genome types were included in this study. Most polycistronic miRNAs, either experimentally validated or candidate polycistronic miRNAs, showed a pattern of conservation in their length, sequence, and chromosomal location in the rice species with an AA genome (e.g., *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, *O. rufipogon*, and to a lesser extent *O. nivara*) (table 2; chromosomal coordinates and lengths are shown in supplementary table S2, Supplementary Material online). Some of them, although present, had a different chromosomal location as compared with *O. sativa* ssp. *japonica*. Not all polycistronic miRNA precursors were identified in the genome of *O. glaberrima*, the African rice species with an AA genome, even though all were present in its wild ancestor *O. barthii*. Finally, *O. punctata* (BB genome) had 16 of the 23 polycistronic miRNA precursors

Table 2
Polycistronic miRNAs in *Oryza* Species

Polycistronic miRNA ***Oryza sativa* ssp. *japonica* (AA)**	*Oryza sativa ssp. indica (AA)*	*Oryza rufipogon (AA)*	*Oryza nivara (AA)*	*Oryza glumaepatula (AA)*	*Oryza glaberrima (AA)*	*Oryza barthii (AA)*	*Oryza punctata (BB)*	*Oryza brachyantha (FF)*
miR156c-156b (Chr01)	98%	98%	98%	98%	—	98%	86%	—
miR166k-166h (Chr02)	94%	98%	98% (Chr01)	98%	98%	98%	88%	83%
miR169j-169k (Chr09)	98%	99%	98%	99%	97%	98%	—	—
miR169m-169l-169q (Chr08)	99%	100%	98%	99%	97%	96%	—	—
miR395a-395b-395c-395d-395e-395f-395g (Chr04)	99%	99%	99%	99%	90%	99%	82%	78%
miR395h-395i-395j-395k-395y-395j (Chr08)	99%	100%	98%	99%	97%	99%	93%	81%
miR395t-395u-395v-395w (Chr09)	100%	99%	98% (Chr01)	98%	—	98%	—	—
miR399c-399h (Chr05)	99%	100%	99% (Chr08)	97%	—	99%	86%	—
miR399e-399a (Chr01)	99%	99%	—	99%	—	99%	85%	—
miR1423-1868 (Chr04)	—	100%	98%	98%	98%	97%	—	—
miR1428e-1428d (Chr03)	98%	99%	98% (Chr11)	99%	99%	98%	—	—
miR1861b-1861c (Chr02)	99%	100%	99%	98%	99%	94% (Chr05)	93%	—
miR1861d-1861e (Chr04)	98%	97%	97% (Chr01)	99%	99%	98%	95%	—
miR1861f-1861g (Chr05)	99%	99%	94% (Chr02)	99%	99%	96%	90%	—
miR1861h-1861p (Chr06)	99%	99%	99%	99%	98%	98%	92%	—
miR1861k-1861j (Chr08)	97%	100%	97%	98%	98%	98%	91%	—
miR1861l-1861m (Chr09)	98% (Chr04)	99%	92%	98%	99%	96% (Chr08)	95%	—
miR1876-1862d (Chr10)	99%	99%	99%	99%	99%	99%	93%	—
miR2118c-2118d-2118e (Chr04)	100%	100%	99%	99%	98%	98%	94%	89%
miR2118f-2118g (Chr04)	100%	100%	99%	—	96%	96%	96%	92%
miR2118h-2118i (Chr04)	100%	99%	99%	99%	98%	98%	93%	—
miR2118k-2118l (Chr04)	100%	99%	99%	99%	99%	99%	92%	—
miR2118n-2118o (Chr04)	99%	100%	99%	99%	97%	97%	93%	—
miR2118p-2118r-2118q (Chr11)	99%	99%	99%	99%	99%	99%	92%	—
miR5143b-5143a (Chr01)	99%	99%	99%	99%	99%	99%	93%	—
miR5147-437 (Chr02)	99%	94%	99% (Chr11)	97%	—	97%	—	—
miR5534a-5534b (Chr11)	93%	90%	93%	90%	—	78%	—	—
miR6255-6253 (Chr07)	99%	100%	99% (Chr06)	99%	97%	97%	99%	82%

NOTE.—*Oryza sativa* ssp. *japonica* corresponds to cv. Nipponbare, and *Oryza sativa* ssp. *indica* corresponds to cv. 9311. Includes polycistronic miRNAs experimentally validated in this work (in bold) as well as candidate polycistronic miRNAs. For each precursor, the percentage of sequence homology relative to the sequence in *O. sativa* ssp. japonica (cv Nipponbare) is shown. In parenthesis, precursors with a different chromosomal location relative to *O. sativa* cv. Nipponbare. —, not present.

found in *O. sativa* cv. Nipponbare, whereas only 4 polycistronic miRNA precursors were identified in the *O. brachyantha* genome (FF genome). The low sequence similarity observed between *O. brachyantha* (FF genome) and rice species with an AA genome might indicate primitive forms of these *MIR* genes in the FF genome. Similar to what was proposed for *Arabidopsis MIR* genes (Fahlgren et al. 2007), polycistronic miRNA might have spawned and been lost frequently during evolution of rice. However, we should be cautious before concluding that the observed differences among *Oryza* species are biologically relevant because for wild species only one accession has been sequenced. Furthermore, information on the miRNA transcriptome is publicly available only for four of the nine *Oryza* species. The generation of genome sequences for more *Oryza* accessions (cultivated and wild relatives of *Oryza* species) and their miRNA transcriptomes is a requisite to get reliable information on the evolutionary history of polycistronic *MIR* genes.

Discussion

Evidence is presented on the occurrence of polycistronic miRNAs in the rice genome. By using a criterion of 3-kb distance between 2 consecutive mature miRNAs, we identified 28 loci encompassing 2 or more miRNAs mapping into a hairpin-forming precursor structure. Seven of these miRNA precursor structures were experimentally validated as polycistronic miRNAs encoding homologous (miR166k-166h,

miR1428e-1428d, and miR1861b-1861c) or nonhomologous (miR1423-1868, miR1876-1862d, miR5147-437, and miR6255-6253) miRNAs. Analysis of miRNA transcriptomes allowed us to demonstrate coexpression of mature miRNAs encoded in miR166k-166h, miR1423-1868, and miR1876-1862 precursor transcripts further supporting that these precursors represent bona fide polycistronic miRNAs.

Most plant miRNAs, in contrast with animal miRNAs, are encoded as single units in a miRNA precursor. Several miRNA clusters containing copies of the same miRNA which are independently transcribed have been identified in plants (i.e., miR156, miR166, miR395, or miR159a) (Guddeti et al. 2005; Wang et al. 2007; Boualem et al. 2008; Lacombe et al. 2008; Merchan et al. 2009; Barik et al. 2014), and the few polycistronic miRNAs so far described in plants generally contain homologous miRNAs. Only a few examples of nonhomologous polycistronic miRNAs have been reported in plants (mainly in *A. thaliana*) (Guddeti et al. 2005; Wang et al. 2007; Boualem et al. 2008; Lacombe et al. 2008; Merchan et al. 2009; Barik et al. 2014). Also in Arabidopsis, two functionally related miRNAs, miR846 and miR842, were found to be generated by the same transcriptional unit but from alternative splicing isoforms of the primary transcript (Jia and Rock 2013). The polycistronic nature of miR166k-166h, miR1428e-1428d, and miR1861b-1861c in rice has been reported (Zhu et al. 2008; Barik et al. 2014).

Because cotranscription is often used to imply a functional relationship of miRNAs in polycistrons, there is the possibility that polycistronic miRNAs have evolved to fine tune gene expression in pathways or regulatory networks controlling related processes. Of particular interest are polycistronic miRNAs containing nonhomologous miRNAs potentially controlling target genes involved in different physiological processes. Because polycistronic transcription eliminates the need for individual promoters and regulatory factors, polycistronic miRNAs provide opportunities to coordinate the spatial and/or temporal gene expression patterns. A better knowledge of target genes for rice miRNAs is, however, needed to infer the functional significance of the organization of these miRNAs as polycistrons in rice.

When analyzing the distribution of polycistronic miRNAs in the rice genome, they were found to locate in both duplicated and nonduplicated chromosomal regions. Thus, genomic segmental duplication events cannot be considered the only factor contributing to the expansion and diversification of polycistronic *MIR* genes in rice. Similar to the evolution of protein gene families, miRNAs might evolve by a process of gene duplication, including segmental duplication and tandem duplication, followed by dispersal and diversification (Maher et al. 2006; Nozawa et al. 2012). An increase in gene copy number of polycistronic miRNAs might also cause a dosage effect in the accumulation of individual miRNAs produced by polycistrons.

Conclusion

The results here reported demonstrate the existence of polycistronic transcripts encoding homologous and nonhomologus miRNAs in rice. The simultaneous production of several miRNAs from a common transcript with the ability to regulate the expression of genes involved in diverse functions during plant development and adaptation to stress conditions expands the gene regulatory capacity of this class of riboregulators. Investigating potential functional interactions between individual miRNAs encoded by a common precursor operating in various biological contexts is then of great interest. The information here presented provides a foundation for further investigations on the biological significance of polycistronic miRNAs in plants while being potentially useful in future studies on the evolution and functionality of polycistronic *MIR* genes in rice, a species with agronomical importance and a model in cereal research.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S6 are available at *Genome Biology and Evolution online* (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and the European Regional Development funds (FEDER) (BIO2012-32838, BIO2015-67212-R to B.S.S.) and the CSIC/NSC (Spanish National Research Council/National Science Council of Taiwan)-Cooperative Research Project-Formosa Program (2009TW0041 to B.S.S. and Y.I.C.H]. P.B. received a Ph.D. grant from the Ministerio de Ciencia e Innovación, Formación de Personal Investigador (BES-2010-032879). We thank J.M. Casacuberta for critically reading this manuscript and Ms Laura Smales (BioMedEditing, Toronto, Canada) for English editing.

Literature Cited

Allen E, et al. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. Nat Genet. 36:1282–1290.

Altuvia Y, et al. 2005. Clustering and conservation patterns of human microRNAs. Nucleic Acids Res. 33:2697–2706.

Arikit S, Zhai J, Meyers BC. 2013. Biogenesis and function of rice small RNAs from non-coding RNA precursors. Curr Opin Plant Biol. 16:170–179.

Aukerman MJ, Sakai H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell 15:2730–2741.

Axtell MJ, Bowman JL. 2008. Evolution of plant microRNAs and their targets. Trends Plant Sci. 13:343–349.

Baker CC, Sieber P, Wellmer F, Meyerowitz EM. 2005. The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. Curr Biol. 15:303–315.

Baldrich P, et al. 2015. MicroRNA-mediated regulation of gene expression in the response of rice plants to fungal elicitors. RNA Biol. 12:847–863.
Barik S, et al. 2014. Phylogenetic analysis reveals conservation and diversification of micro RNA166 genes among diverse plant species. Genomics 103:114–121.
Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297.
Boualem A, et al. 2008. MicroRNA166 controls root and nodule development in *Medicago truncatula*. Plant J. 54:876–887.
Brodersen P, et al. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. Science 320(80):1185–1190.
Calviño M, Messing J. 2013. Discovery of microRNA169 gene copies in genomes of flowering plants through positional information. Genome Biol Evol. 5:402–417.
Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
Campo S, et al. 2013. Identification of a novel microRNA (miRNA) from rice that targets an alternatively spliced transcript of the *Nramp6* (Natural resistance-associated macrophage protein 6) gene involved in pathogen resistance. New Phytol. 199:212–227.
Chiou TJ, et al. 2006. Regulation of phosphate homeostasis by microRNA in *Arabidopsis*. Plant Cell 18:412–421.
Dogara AM, Jumare AI. 2014. Origin, distribution and heading date in cultivated rice. Int J Plant Biol Res. 2:2–6.
Fahlgren N, et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2:1–14.
Felippes FF, et al. 2008. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. RNA 14:2455–2459.
Guddeti S, et al. 2005. Molecular evolution of the rice miR395 gene family. Cell Res. 15:631–638.
Guo HS, Xie Q, Fei JF, Chua NH. 2005. MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for arabidopsis lateral root development. Plant Cell 17:1376–1386.
Guyot R, Keller B. 2004. Ancestral genome duplication in rice. Genome 47:610–614.
Huang X, et al. 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501.
Jeong DH, Green PJ. 2013. The role of rice microRNAs in abiotic stress responses. J Plant Biol. 56:187–197.
Jiang D, et al. 2006. Duplication and expression analysis of multicopy miRNA gene family members in *Arabidopsis* and rice. Cell Res. 16:507–518.
Jia F, Rock CD. 2013. MIR846 and MIR842 comprise a cistronic MIRNA pair that is regulated by abscisic acid by alternative splicing in roots of Arabidopsis. Plant Mol Biol. 81:447–460.
Jones-Rhoades MW, Bartel DP, Bartel B. 2006. MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol. 57:19–53.
Kaur C, et al. 2014. Glyoxalases and stress tolerance in plants. Biochem Soc Trans. 42:485–490.
Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42:68–73.
Kurihara Y, Watanabe Y. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A. 101:12753–12758.
Lacombe S, et al. 2008. Identification of precursor transcripts for 6 novel miRNAs expands the diversity on the genomic organisation and expression of miRNA genes in rice. BMC Plant Biol. 8:123.
Llave C, Xie Z, Kasschau KD, Carrington JC. 2002. Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science 297:2053–2056.
Maher C, Stein L, Ware D. 2006. Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. 16:510–519.
Mallory AC, et al. 2004. MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. EMBO J. 23:3356–3364.
Merchan F, Boualem A, Crespi M, Frugier F. 2009. Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. Genome Biol. 10:R136.
Navarro L, et al. 2006. A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312(80):436–439.
Nozawa M, Miura S, Nei M. 2012. Origins and evolution of microRNA genes in plant species. Genome Biol Evol. 4:230–239.
Palatnik JF, et al. 2003. Control of leaf morphogenesis by microRNAs. Nature 425:257–263.
Patanun O, Lertpanyasampatha M, Sojikul P, Viboonjun U, Narangajavana J. 2013. Computational identification of microRNAs and their targets in cassava (*Manihot esculenta* Crantz.). Mol Biotechnol. 53:257–269.
Piriyapongsa J, Jordan IK. 2008. Dual coding of siRNAs and miRNAs by plant transposable elements. RNA 14:814–821.
Sunkar R, Chinnusamy V, Zhu J, Zhu JK. 2007. Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends Plant Sci. 12:301–309.
Sweeney M, McCouch S. 2007. The complex history of the domestication of rice. Ann Bot. 100:951–957.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30:2725–2729.
Thiel T, et al. 2009. Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. BMC Evol Biol. 9:209.
Wang SP, Liu KD, Qi-Fa Z. 2000. Segmental duplications are common in rice genome. Acta Bot Sin. 42:1150–1155.
Wang M, et al. 2014. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. Nat Genet. 9:982–988.
Wang S, et al. 2007. Molecular evolution and selection of a gene encoding two tandem microRNAs in rice. FEBS Lett. 581:4789–4793.
Wu B, Wang M, Ma Y, Yuan L, Lu S. 2012. High-throughput sequencing and characterization of the small RNA transcriptome reveal features of novel and conserved microRNAs in *Panax ginseng*. PLoS One 7:e44385.
Xie F, Frazier TP, Zhang B. 2010. Identification and characterization of microRNAs and their targets in the bioenergy plant switchgrass (*Panicum virgatum*). Planta 232:417–434.
Yu J, et al. 2005. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 3:e038.
Zhao B, et al. 2009. Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. BMC Mol Biol. 10:29.
Zhu QH, et al. 2008. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res. 18:1456–1465.

Associate editor: Laura Rose