

Genomic dissection of small RNAs in wild rice (*Oryza rufipogon*): lessons for rice domestication

Yu Wang¹, Xuefei Bai¹, Chenghai Yan¹, Yiejie Gui¹, Xinghua Wei², Qian-Hao Zhu³, Longbiao Guo² and Longjiang Fan¹

¹Department of Agronomy & Key Laboratory of Crop Germplasm Resource of Zhejiang Province, Zhejiang University, Hangzhou, 310058, China; ²State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, China; ³CSIRO Plant Industry, Canberra, ACT, 2601, Australia

Authors for correspondence:

Longjiang Fan

Tel: +86 571 88982730

Email: fanlj@zju.edu.cn

Longbiao Guo

Tel: +86 571 63370537

Email: longbiaoguo@yahoo.com.cn

Received: 31 May 2012

Accepted: 28 July 2012

New Phytologist (2012) **196**: 914–925

doi: 10.1111/j.1469-8137.2012.04304.x

Key words: degradome, genome, *MIRNA*, *Oryza rufipogon*, rice domestication.

Summary

- The lack of a *MIRNA* set and genome sequence of wild rice (*Oryza rufipogon*) has prevented us from determining the role of *MIRNA* genes in rice domestication.
- In this study, a genome, three small RNA populations and a degradome of *O. rufipogon* were sequenced by Illumina platform and the expression levels of microRNAs (miRNAs) were investigated by miRNA chips.
- A *de novo* *O. rufipogon* genome was assembled using *c.* 55× coverage of raw sequencing data and a total of 387 *MIRNA*s were identified in the *O. rufipogon* genome based on *c.* 5.2 million unique small RNA reads from three different tissues of *O. rufipogon*. Of these, *O. rufipogon* *MIRNA*s, 259 were not found in the cultivated rice, suggesting a loss of these *MIRNA*s in the cultivated rice. We also found that 48 *MIRNA*s were novel in the cultivated rice, suggesting that they were potential targets of domestication selection. Some miRNAs showed significant expression differences between wild and cultivated rice, suggesting that expression of miRNA could also be a target of domestication, as demonstrated for the miR164 family.
- Our results illustrated that *MIRNA* genes, like protein-coding genes, might have been significantly shaped during rice domestication and could be one of the driving forces that contributed to rice domestication.

Introduction

Noncoding small RNAs, including microRNAs (miRNAs) and small-interfering RNAs (siRNAs), are critical regulators in plants. They regulate the expression of their target genes by mRNA cleavage or repressing mRNA translation at the post-transcriptional level (Bartel, 2004). miRNAs are processed from precursors with a hairpin structure by Dicer-like enzymes (Voinnet, 2009). Plant miRNAs play critical roles in a variety of development processes (Llave *et al.*, 2002; Aukerman & Sakai, 2003; Chen, 2004). Thousands of miRNAs have been identified in a wide range of plant species, from unicellular green alga *Chlamydomonas reinhardtii* to higher plants such as rice (Kozomara & Griffiths-Jones, 2011). However, new miRNAs are still being identified even for the well characterized plant species. For example, a recent study identified 76 new rice miRNAs by massive sequencing of small RNA populations from different tissues (Jeong *et al.*, 2011).

In view of the comprehensive regulatory function of miRNAs in rice, whether *MIRNA* genes were under artificial selection during domestication is an intriguing question. Investigation of population genetics has illustrated a strong positive selection at the miR156b/c locus in rice (Wang *et al.*, 2007). Recent studies have

shown that miR156 controls the ideal plant architecture, an important agronomic trait, by regulating the expression levels of *OsSPL14*, and was a target of artificial selection (Jiao *et al.*, 2010; Miura *et al.*, 2010). Meanwhile, a large-scale investigation of selection signals at *MIRNA* genes indicated that several *MIRNA* genes might undergo artificial selection during rice domestication (Wang *et al.*, 2010). Based on ANOVA of miRNAs and their binding sites in target genes and paralogs of target genes, a dynamic gain and loss of miRNA binding sites caused by nucleotide substitution or insertion/deletion (indel) during rice evolution was detected (Guo *et al.*, 2008).

The wild rice population holds great genetic diversity and is an important genetic resource for rice breeding. Comparative analysis of the genomes from cultivated and wild rice could provide crucial insights into rice domestication. Although the genomes of two cultivated rice species (*Oryza sativa* ssp. *indica* and *japonica*) have been fully sequenced, and five accessions of *Oryza rufipogon*, a wild ancestor of *O. sativa*, have recently been sequenced to a mean coverage of 10.3× (Xu *et al.*, 2012), the whole-genome sequence of *O. rufipogon* has not yet been generated. Meanwhile, no miRNA has been identified in *O. rufipogon*. A more complete *O. rufipogon* genome will offer a good opportunity to identify all *O. rufipogon*

miRNAs and to study the evolutionary dynamics of miRNAs during rice domestication at a genome-wide scale.

In this work, we sequenced an accession of *O. rufipogon* to *c.* 55× coverage, identified miRNAs using small RNAs generated from three different tissues of *O. rufipogon* and identified miRNA targets in *O. rufipogon* by degradome sequencing. Using a comparative genomics approach, we found that rice *MIRNA* genes have experienced a complex evolutionary process during domestication.

Materials and Methods

Plant material

A Chinese accession of the *O. rufipogon* Griff. wild rice, collected from Dongxiang, Jiangxi Province, and provided by China National Rice Research Institute was used in genome, small RNA and degradome sequencing. Six cultivars (*Oryza sativa* L.) and six accessions of *O. rufipogon* (Supporting Information, Table S1), randomly selected from our collections kindly provided by the International Rice Research Institute (IRRI) and the China Rice Research Institute, were used in miRNA microarray experiments, as previously described (Wang *et al.*, 2010).

Genome sequencing and annotation

Genomic DNA of *O. rufipogon* was extracted from young seedlings using the CTAB protocol (Chong, 2001). Fragmented DNA was fractionated by electrophoresis and DNA fragments of the desired length were excised and gel-purified. Purified DNA was then used to generate two paired-end sequencing libraries with an insert size of *c.* 500 bp and 2 kb, respectively. DNA sequencing was performed on the Illumina HiSeq2000 platform.

Sequencing reads were aligned onto the rice reference genome sequence (*O. sativa*, *japonica*; <http://rice.plantbiology.msu.edu/>, Release 7 of the MSU Rice Genome Annotation Project) using SOAP2 (<http://soap.genomics.org.cn/>; Li *et al.*, 2009b) and the bwa package (<http://bio-bwa.sourceforge.net/index.shtml>; Li & Durbin, 2009). The sequencing depth and coverage relative to the rice reference genome were calculated based on the alignment. Genomic variation detection and visualization were performed using samtools-0.1.18 (<http://samtools.sourceforge.net/>; Li *et al.*, 2009a,b) and inGAP-sv (<http://ingap.sourceforge.net/>; Qi *et al.*, 2010). The raw sequence data have been deposited into the NCBI Short Read Archive under the accession number SRA055709.

The *de novo* assembly of the *O. rufipogon* genome was performed using SOAPdenovo (Li *et al.*, 2010a). Raw reads were preprocessed to remove adaptors, to filter out reads of low quality ($\geq 50\%$ of its nucleotides with quality value ≤ 5). Error corrections were performed with the 'Correction' program to reduce the low frequency K-mer for better assembly (Li *et al.*, 2010a,b). A best assembly result (i.e. its contigs with the longest N50) was obtained by using the 49-mer parameter and its contigs were used for further scaffold assembly based on the paired-end information of short reads. The gaps between the assembled scaffolds were then closed by GapCloser v1.1. Repeat sequences were annotated by searching against the Repbase version 16.06

(<http://www.girinst.org/>) and Oryza Repeat Database version 3.3 (http://rice.plantbiology.msu.edu/annotation_oryza.shtml) using RepeatMasker (<http://www.repeatmasker.org/>) and masked. Repeat-masked contigs were subjected to gene prediction using Augustus (<http://augustus.gobics.de/>). Known noncoding RNAs were also annotated by searching against the Rfam database (version 10.0, <http://rfam.sanger.ac.uk/>).

Sequencing of small RNA populations and degradome

Total RNA was extracted from leaves and roots of four-leaf stage seedlings (*c.* 20 d), and dehusked 5–10 d postanthesis (DPA) developing grains were collected from *O. rufipogon*. Small RNA sequencing was performed using the Illumina GA II. Two corresponding small RNA datasets (GSE11014 and GSE19602; Zhu *et al.*, 2008; He *et al.*, 2010) from Nipponbare (*O. sativa*, *japonica*) were downloaded from GEO databases.

A degradome sequencing library was constructed using mRNA isolated from leaves of four-leaf stage seedlings used in small RNA sequencing according to the protocols described previously (Addo-Quaye *et al.*, 2008; German *et al.*, 2008). The raw data of small RNA and degradome sequencing have been submitted to the NCBI GEO datasets with the accession numbers of GSE39309.

Analyses of small RNA and degradome data

Raw small RNA reads and degradome segments were trimmed to remove adaptors, and filtered to remove low-quality reads using custom perl scripts. The clean reads with a size of 20–24 nt were then mapped to the *de novo* assembled *O. rufipogon* contig sequences using BLAST ($W = 7$, $E = 1000$). Reads mapped to coding regions and known noncoding RNAs (rRNA, tRNA, snRNA, snoRNA) were excluded for further analysis. The known rice miRNAs were downloaded from miRBase (version 18.0, <http://www.mirbase.org>).

A bioinformatics pipeline (Fig. 1) designed following the criteria previously described (Meyers *et al.*, 2008) was used to predict new miRNAs in *O. rufipogon*. Each step was executed with custom perl scripts. The core program used was the modified version of MIREAP (<https://sourceforge.net/projects/mireap/>) developed by Jeong *et al.* (2011). However, three more criteria were adopted to guarantee the quality of the predicted *MIRNAs*: the abundance of the small RNA reads; $> 90\%$ of the reads from each pre-miRNA mapped to the same strand; and the two most abundant reads accounting for $> 70\%$ of the total reads mapped to each pre-miRNA. The RNA secondary structure of the candidate pre-miRNAs was predicted using the Vienna RNA package (<http://www.tbi.univie.ac.at/~ivo/RNA/>; Hofacker, 2003) and manually checked.

Targets of known rice miRNAs and newly identified wild rice miRNAs were predicted by searching against the coding regions of *O. rufipogon* annotated in this study using psRNATarget (<http://plantgrn.noble.org/psRNATarget/>; Dai & Zhao, 2011). Evidence for miRNA-mediated cleavage was obtained based on analysis of the degradome sequencing data using the CleaveLand pipeline (Addo-Quaye *et al.*, 2009).

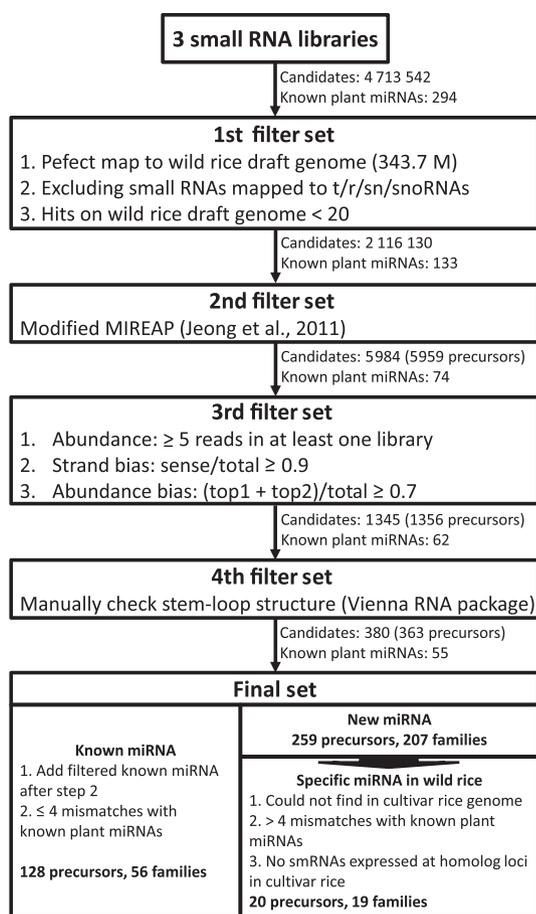


Fig. 1 The bioinformatics pipeline for identification of microRNAs (miRNAs) from the small RNA populations of *Oryza rufipogon*. MicroRNAs were identified by a series of filtering steps shown in the diagram and described in the text. The numbers of candidate small RNAs and known rice miRNAs identified at each step are indicated.

Analysis of miRNA expression

Twelve accessions, including six *O. sativa* (three *indica* and three *japonica*) and six *O. rufipogon* lines, were cultivated under the same conditions (Table S1). Fresh leaves from the four-leaf stage seedlings were used in total RNA extraction. miRNA chips were designed by the LC Sciences Company, China, based on the rice miRNAs deposited in miRBase version 13.0. Each chip contains 253 miRNA probes, representing 414 rice miRNAs. For each chip, small RNA samples from a cultivar and an *O. rufipogon* accession were hybridized separately.

Microarray assay was performed by the LC Sciences Company. Five micrograms of total RNA was size-fractionated using a YM-100 Microcon centrifugal filter (Millipore), and the low-molecular-weight RNAs (< 300 nt) isolated were 3'-extended with a poly(A) tail using poly(A) polymerase. An oligonucleotide tag was then ligated to the poly(A) tail for later fluorescent dye staining; two different tags were used for the two RNA samples in dual-sample experiments. Chip hybridization was performed overnight using a microcirculation pump (Atactic Technologies, China). The hybridization melting temperatures of each probe were balanced by chemical modifications of the detection probes. Hybridization was performed

in 100 μ l 6 \times SSPE buffer (0.90 M NaCl, 60 mM Na₂HPO₄, 6 mM EDTA, pH 6.8) containing 25% formamide at 34°C. Hybridization images were collected using a laser scanner (GenePix 4000B, Molecular Devices, China) and digitized using Array-Pro image analysis software (Media Cybernetics, China). Data were analyzed by first subtracting the background and then normalizing the signals using a LOWESS filter (locally weighted regression). For two-color experiments, the ratio of the two sets of detected signals (log₂-transformed and balanced) and *P*-values of the *t*-test were calculated; a *P*-value < 0.01 was considered as differentially expressed. The raw chip data have been deposited into the NCBI GEO datasets under the accession number GSE39309.

Results

De novo assembly of the *O. rufipogon* genome

To perform genome-wide comparison of *MIRNA* genes between cultivated and wild rice and to investigate the effects of domestication on evolution of *MIRNA*s in rice, we sequenced the genome of *O. rufipogon* (collected from Dongxiang, China) using the next-generation sequencing technology. Two paired-end (PE) libraries with an insert size of *c.* 500 bp and 2 kb were constructed and sequenced. In total, 22.4 Gb of sequence data were generated, which corresponded to a *c.* 55 \times coverage of the reference rice genome (*O. sativa*, ssp. *japonica*, TIGR V6.1) (Table S2). Sequenced short reads were *de novo* assembled using SOAPdenovo (<http://soap.genomics.org.cn/>). We obtained 343.7 Mb of genome scaffold sequences with an N50 of 27 880 and 1101 bp for the scaffolds and contigs, respectively. To annotate the genome, first, repeat sequences (25.2%) were identified by RepeatMasker and

Table 1 Summary of high-throughput sequencing of small RNAs and degradome from *Oryza rufipogon*

Category	Small RNA sequencing			Degradome sequencing*
	Leaf	Root	Developing seed	Leaf
Total				
Raw reads	7 572 483	6 626 103	6 427 458	12 048 037
Unique reads	1 570 177	1 274 554	2 350 532	470 548
Mapped to contig[†]				
Reads	4 531 410	3 576 573	3 609 178	2 282 736
Specific in each library	1 060 734	602 238	2 475 430	NA
Conserved in all libraries	2 750 528	2 647 377	773 948	NA
Unique	1 018 843	773 319	1 191 354	81 657
Specific in each libraries	775 330	546 705	1 024 137	NA
Conserved in all libraries	63 660	63 660	63 660	NA
Singleton	772 715	640 686	888 682	7155

NA, not available.

*Including 18–33 nt reads.

[†]Contigs from the wild rice genome generated in this study; for small RNAs, only 20–24 nt reads were counted.

Table 2 Summary of novel microRNAs (miRNAs) unique to the wild rice (*Oryza rufipogon*)

miRNA	Mature sequence	Size (nt)	No. of reads	Predicted target*	Score	Homolog in <i>O. sativa</i>	Score	Annotation	Evidence†
oru-miRX5e	GAATGTCGAAAATGATAGAACGTC	24	14	NA					NY
oru-miRX5f	GAATGTCGAAAATGATAGAACGTC	24	14	NA					NY
oru-miRX30	AACGGTGTAAATCCGATAGTAGATC	24	29	C7180253_g269	3	Os09g37200.1	NA	Transferase family protein, putative, expressed	NY
oru-miRX65	CGGATTTGAGCTTGCCGACGGC	22	32	scaffold13410_g38	0.5	Os10g40770.1	0	Expressed protein	Y
oru-miRX69	ACGGTATGATCGGATTGTAGATTG	24	7	scaffold3144_g418	3	Os03g17350.1	3	White-brown complex homolog protein, putative, expressed	Y
oru-miRX83	TTATGTTTTTGGAGAGGGGGC	21	5	scaffold3653_g404	3	Os04g50920.1	3	WRKY37, expressed	Y
oru-miRX96	TTACGAGAAATGACTAGTAGC	21	7	scaffold7063_g237	3	Os08g17390.1	NA	Expressed protein	Y
oru-miRX105	AAAGTATTTTGTATCGGATGGG	21	22	scaffold8192_g41	3	NA		NA	Y
oru-miRX109	TTGGAACATAGGAATTTTACA	21	16	scaffold461_g560	3	Os03g51020.1	NA	Expressed protein	Y
oru-miRX113	TACGGGACGAATCTATTAAGG	21	17	scaffold591_g301	2.5	Os05g25890.1	NA	Expressed protein	Y
oru-miRX121	ATATCCGTAGATGTTAATGAATC	24	10	scaffold1019_g92	2.5	Os02g57090.1	2.5	Anthranilate phosphoribosyltransferase, putative, expressed	Y
oru-miRX139	GTCTCTATACATTTGTAGATGTGGC	24	35	NA					Y
oru-miRX141	ATAAAGTCTTGTACTGTGAACCGG	24	54	NA					NY
oru-miRX143	TTAGGACAAGGATGTGGCATC	21	165	NA					NY
oru-miRX158	TTATCTAACTGTTTTTTGTGTAG	23	7	NA					Y
oru-miRX160	AATATAGAGGATTTTGGTAGATG	24	7	NA					Y
oru-miRX179	TATCCGGCATTATTTAGGGC	21	6	NA					Y
oru-miRX185	GACGGTTACACCTGACATGTGGC	24	6	NA					Y
oru-miRX188	GAAGATACGTAACAAGATGAGC	24	7	NA					NY
oru-miRX192	GACAGGTGTAGACACAGTAGAC	24	5	NA					NY
oru-miRX283	TTTTATAGGATTTTGTGGGA	21	40	NA					Y

*NA, no target predicted or no homolog found in *Oryza sativa*; only targets with the best scores are shown.

†Y, not found in the dataset, including 40 rice cultivars sequenced by Xu *et al.* (2012); NY, not determined yet.

masked. The remaining sequences were used in gene prediction using Augustus (<http://augustus.gobics.de/>). In total, 29 660 genes or protein coding sequences were predicted. These genes were further annotated by BLASTX/BLASTP search against the nucleotide (NCBI) and protein (SWISSPROT) databases. Known noncoding RNAs (rRNAs, tRNAs, snRNAs and snoRNAs) were also annotated based on the Rfam database by BLAST.

Identification of miRNAs and their targets in *O. rufipogon*

To identify miRNAs in *O. rufipogon*, we generated three small RNA sequencing libraries using RNAs isolated from leaves and roots of four-leaf stage seedlings and developing seeds. In total, 7 572 483, 6 626 103 and 6 427 458 clean reads with a size between 20 and 24 nt were generated from each library. Consistently, 24 and 21 nt small RNAs were the most abundant in each library (Table 1). These reads were mapped to the draft genome sequences of *O. rufipogon* by BLAST ($E = 1000$, $W = 7$); 59.8, 54.0 and 56.2% of the reads could be exactly mapped. For miRNA identification, reads that mapped to the known noncoding RNAs were eliminated for further analysis.

MIRNA genes were predicted using the pipeline illustrated in Fig. 1. To guarantee the quality of the predicted MIRNAs, in addition to the criteria described previously (Meyers *et al.*, 2008), three additional criteria (the third filter in Fig. 1) were used. Using this pipeline, we identified 387 MIRNAs, including 128 known MIRNAs from 56 families, and 259 novel MIRNAs from 207 families. Of the 387 MIRNAs, 28 were present in all three tissues (leaf, root and developing grain) and 51 in at least two tissues. The precursor sequences, stem-loop structures and read numbers in three tissues of these MIRNAs are shown in Table S3. Of the 259 novel MIRNAs, at least 20 (Table 2) could be *O. rufipogon*-specific for the following reasons: they are absent in the cultivated rice genomes (from 40 cultivars; Xu *et al.*, 2012); each has over four mismatches in the mature sequences with any known plant miRNAs (miRBase Version 18); and no small RNA reads were detected in their homologous pre-miRNA loci in the cultivated rice. Three examples of such *O. rufipogon*-specific miRNAs are shown in Fig. 2. Based on homolog search and sequence comparison, big deletion (> 20 nt; for 16 MIRNAs) or nucleotide mutation events (> 4 nt; for four MIRNAs) were found to be the reason for their loss or dysfunction in the cultivated rice. For the remaining 239 novel *O. rufipogon* MIRNAs, small (< 4 nt) sequence deletions or nucleotide mutations seem to be the reason for their dysfunction in the cultivated rice. To know whether the mutations found in *O. rufipogon*-specific MIRNAs were caused by misassembly of sequences, the precursor sequences of 15 *O. rufipogon*-specific MIRNAs were amplified and sequenced, and in 14 cases, the assembly could be confirmed, suggesting that these MIRNAs have indeed been lost in the cultivated rice.

All 387 miRNAs newly identified in *O. rufipogon* were subjected to target prediction and 169 (including 105 novel miRNAs) were predicted to target annotated genes in *O. rufipogon* (Table S4). Of the 20 *O. rufipogon*-specific miRNAs, six have targets identified in the annotated gene set based on the *O. rufipogon* genome, and most of them were not targeting transcription factors (Table 2), a

characteristic of nonconserved miRNAs. Degradome sequencing matched harmoniously with small RNA sequencing technology for high-throughput prediction of miRNA target genes. To confirm the predicted targets, we constructed a degradome sequencing library using total RNA from seedlings of four-leaf stage *O. rufipogon* plants. In total, 12 048 037 raw clean degradome fragments were generated. Based on the approaches described previously (Addo-Quaye *et al.*, 2008, 2009), targets were predicted for 88 miRNAs (including 11 novel miRNAs and one *O. rufipogon*-specific miRNA) (Fig. 3; Table S5). As expected, novel targets were found for novel miRNAs identified in *O. rufipogon*. For instance, oru-miRX35 was found to target tryptophan synthase beta chain 1 (scaffold2742_g70) (Fig. 3). In addition, conserved miRNAs were found to regulate conserved targets in *O. rufipogon* (Table S5); however, novel targets were also found for conserved miRNAs—for example, genes encoding glucan endo-1,3-beta-glucosidase precursor (scaffold485_g607, homolog of Os02g04670) and Bowman–Birk type bran trypsin inhibitor precursor (scaffold4757_g143, homolog of Os01g03380) were found to be targeted by miR156 and miR408, respectively. These results suggest a dynamic nature of miRNA and target interaction during rice domestication.

Genomic variations of MIRNAs between wild and cultivated rice

To further investigate the dynamics of MIRNAs during rice domestication, we mapped *O. rufipogon* genomic and small RNA reads onto the MIRNAs previously identified in the cultivated rice (*O. sativa*). Of the 543 rice MIRNAs deposited in miRBase (version 18), 495 were covered by both genomic (to their whole precursors) and small RNA reads (to their mature sequences) from *O. rufipogon*, and 48 were not covered by any *O. rufipogon*

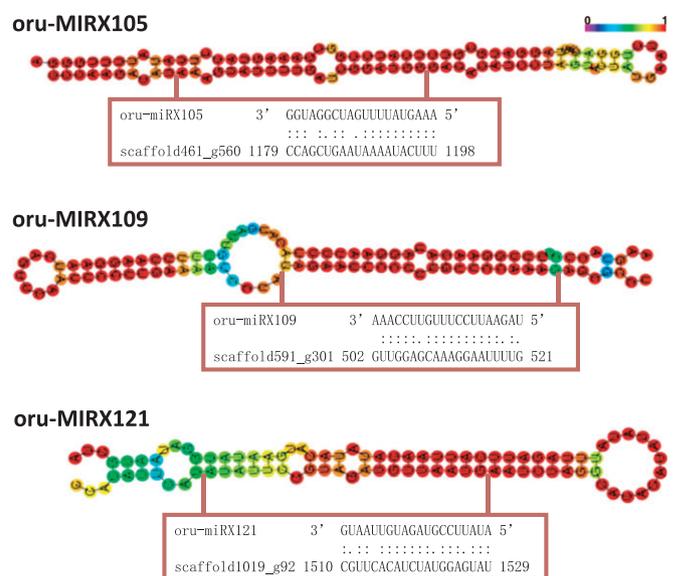


Fig. 2 Stem-loop structures of three selected novel microRNAs (miRNAs) in *Oryza rufipogon*. Base pairings between miRNA and their predicted targets are shown with two dots representing a Watson–Crick pair and one dot representing a G–U (guanine and uracil) pair.

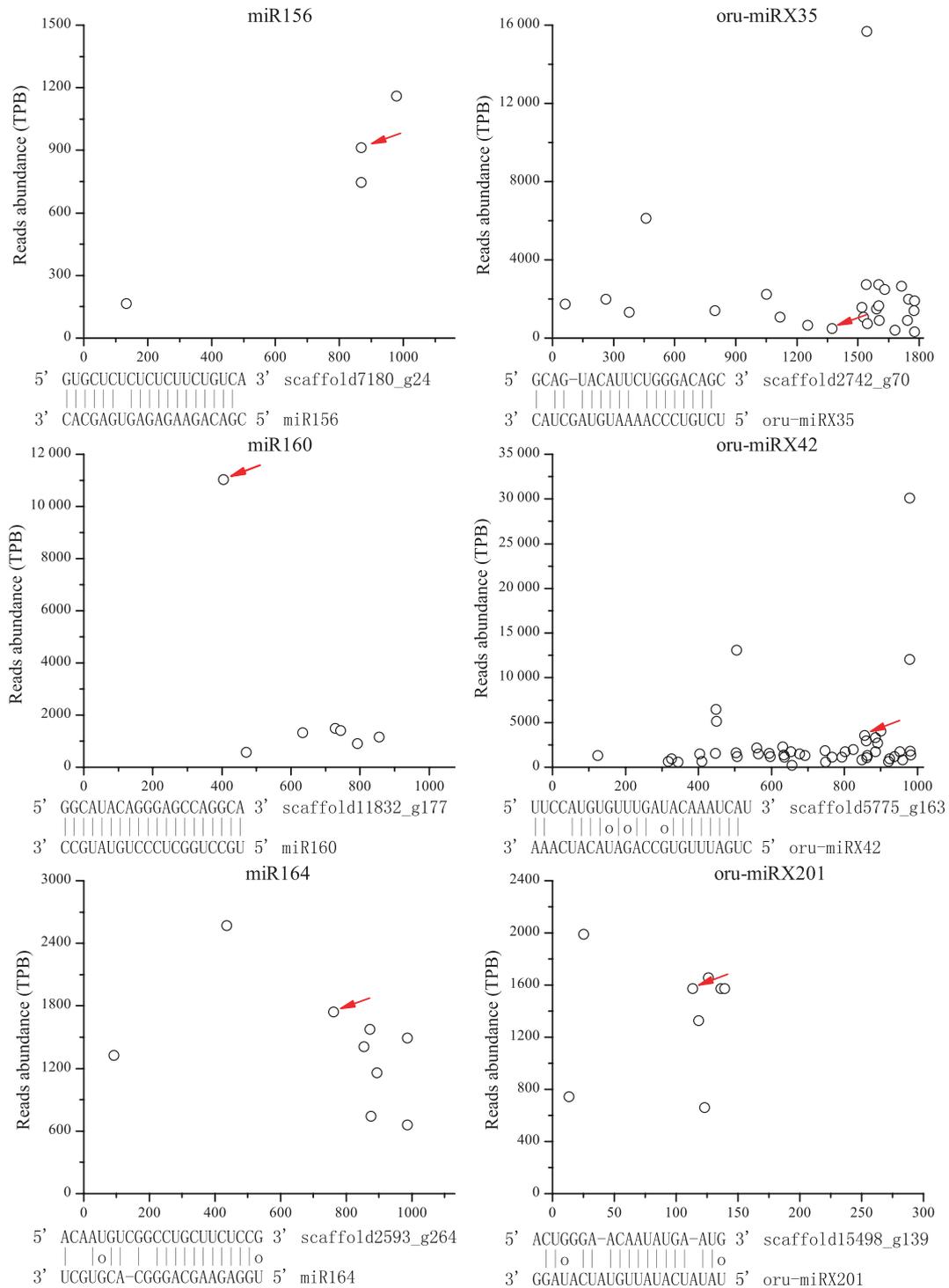


Fig. 3 Plot signals of the candidate targets of microRNAs (miRNAs) by degradome sequencing. Targets of three conserved and novel miRNAs are shown in the left and right panels, respectively. The x-axis displays the nucleotide position of the target genes while the y-axis indicates the abundance of reads converted into transcripts per billion (TPB). Each circle represents a degradome fragment mapped to the target gene and the circle indicated by the red arrow represents the expected miRNA cleavage product. Base pairing between miRNAs and their predicted targets are shown below each panel, with one vertical representing a Watson–Crick pair and one circle representing a G–U (guanine and uracil) pair.

sequence (Table 3). Of the 495 common *MIRNAs* found in both the cultivated and wild rice, 272 were unchanged during rice domestication while 223 had small indels or mutations in their precursors or the mature sequences (Table S6), but these changes did not seem to affect the processing of these *MIRNAs* in the

cultivated rice. Of the 48 *MIRNAs* found only in *O. sativa*, 25 had no corresponding pri-miRNAs found in *O. rufipogon*. Two such *MIRNAs* (*MIR1435* and *MIR413*) were shown in Fig. 4(a). After mapping the *O. rufipogon* PE reads onto the *japonica* genome sequence (Nipponbare), it is clear that these two *MIRNAs* are

Table 3 Classification of 543 known *Oryza sativa* microRNAs (miRNAs)

Category	Number*
Common in <i>O. sativa</i> and <i>Oryza rufipogon</i>	495
No mutation between <i>O. sativa</i> and <i>O. rufipogon</i>	272 (128)
SNP/indels in miRNA	50 (1)
SNP/indels in pre-miRNA (except miRNA)	173 (36)
Novel in <i>O. sativa</i> †	48
Unable to form hairpin structure in <i>O. rufipogon</i>	23 (0)
No corresponding <i>MIRNA</i> in <i>O. rufipogon</i>	25 (1)
Total	543 (166)

*Numbers in parentheses are the number of miRNAs conserved in plants.

†No small RNA reads were found in our *Oryza rufipogon* small RNA dataset.

located at a region with a big deletion in *O. rufipogon*. For the remaining 23 *MIRNAs*, their ancestral sequences were found in *O. rufipogon* but they were unable to form the stem-loop structures required for miRNA production because of big indels (≥ 20 nt) or a mutation (> 4 nt) in the mature sequences in *O. rufipogon*

(Table 3; Table S6). Two such *MIRNAs* are shown in Fig. 4(b). Apparently, parts of *MIR1442* and *MIR439h*, including the whole mature sequences, were missing in the wild rice genome. Genomic variations in precursors of 38 previously identified conserved miRNA families in wild rice are shown in Fig. 5. These results demonstrate that most *MIRNAs* have been adopted or fixed in the cultivated rice population.

Previous investigations of miRNA::target interaction suggest a coevolution relationship between miRNAs and their targets (Guo *et al.*, 2008; Cuperus *et al.*, 2011). Our results support this notion because no target was found in *O. rufipogon* for about half of the newly evolved *MIRNAs* in *O. sativa*, and when a target was predicted in *O. rufipogon*, it usually had a higher score (less likely to be a real target) than the target predicted in *O. sativa* (Table 4).

Expression profiles of miRNAs in wild and cultivated rice

To investigate the expression changes of miRNA in the cultivated and wild rice, six miRNA chips were used to compare the



Fig. 4 Examples of genomic variations between wild rice *Oryza rufipogon* microRNAs (miRNAs) and their counterparts in cultivated rice (*Oryza sativa*). (a) Two miRNA loci in cultivated rice (*osa-miR1435* and *osa-miR413*) were not found in *O. rufipogon*. Mapping results of the wild rice paired-end reads to the Nipponbare genome are shown. (b) Two examples of disabled miRNAs in the wild rice as a result of sequence deletions. Two big deletions covered the whole mature sequences of *osa-miR1442* and *osa-miR439h*. For each miRNA, mapping results of the wild rice reads on pre-miRNA sequences are shown.

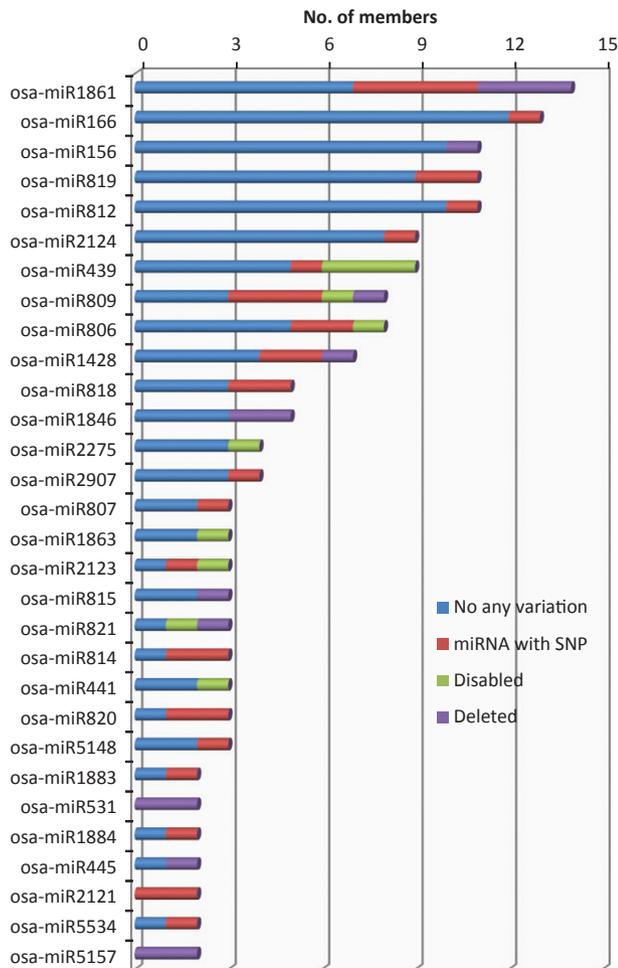


Fig. 5 Genomic variations in precursors of 38 previously identified conserved microRNA (miRNA) families in wild rice (*Oryza rufipogon*). Also see Table 4 for details.

expression levels of miRNAs in six cultivars (three each from the *indica* and *japonica* group) and six accessions of *O. rufipogon*. Each chip was separately hybridized with low-molecular-weight RNAs (< 300 nt) extracted from four-leaf seedlings of a cultivar and a wild accession (Table S1). Expression divergence between the cultivated and wild rice was calculated with a false discovery rate (FDR) of < 3%. MicroRNAs with significant expression changes between cultivated and wild rice were clustered by hierarchical clustering (Fig. 6). In total, 87 miRNA families showed altered expression levels in the cultivated and wild rice (seven selected miRNA families are shown in Table S7). Interestingly, more miRNA families (71) tended to be up-regulated in the cultivars. Of these 71 families, 20 were conserved miRNAs. For example the miR164 family was significantly up-regulated in all six cultivars tested (Table S7). Meanwhile, the expression levels of several miRNAs showed divergence between the *indica* and *japonica* group. For example, a significant down-regulation of miR156, miR162, miR167 and miR397 was only observed in the *indica* but not in the *japonica* group, and miR396 was significantly up-regulated in the *indica* group but down-regulated in the *japonica* group. These differences might be a result of the independent domestication routes of the two subspecies of *O. sativa*.

Although microarray is a powerful tool for investigation of the expression profiles of miRNAs, it could only detect relatively highly expressed miRNAs and miRNAs with their sequences represented on the chip. Because the design of our miRNA chip was based on rice miRNAs in version 13 of miRBase and more new miRNAs have been identified since then, to have a comprehensive view of the changes of all miRNAs in wild and cultivated rice, we compared their read numbers (reads per million (RPM)) in *O. rufipogon* (this study) with those generated previously from the equivalent tissues in *O. sativa* (*japonica*; Zhu *et al.*, 2008; He *et al.*, 2010). In general, the results from high-throughput sequencing were consistent with our miRNA microarray results (Table S8). For example, the up-regulation of the miR164 family in the cultivars analyzed was supported by the small RNA sequencing results (Fig. S1). In addition, the small RNA sequencing approach was able to distinguish the expression levels of individual members of this family, which is hard to be achieved by the miRNA chip.

To test whether the miRNAs with significant expression differences between wild and cultivated rice were potential targets of domestication selection, a neutrality test for positive selection was carried out for the miR164 family in the wild and cultivated populations. To this end, we sequenced the precursors of all six miR164 members in 54 cultivars (*O. sativa*, *indica* and *japonica*) and 15 accessions of *O. rufipogon* (Wang *et al.*, 2010). A significant reduction of nucleotide diversity was observed in all six members in the two cultivar groups, varying from 1.32- to 7.31-fold (as measured by θ_w) relative to the wild group (Table S9), suggesting a domestication bottleneck effect on these *MIRNA* loci. A similar effect has been observed in protein-coding genes (Zhu *et al.*, 2007). Besides using Tajima's *D* (Tajima, 1989) as a neutrality test, we also employed Fay and Wu's *H* test, which detects the positive selection by measuring the relatively high frequency of the departure of derived alleles from neutrality (Fay & Wu, 2000), rather than tracing an excess of low-frequency polymorphisms after selective sweep by the *D* test. Significant positive selection was detected in *MIR164c*, *d* and *e* in the cultivated population but not in the wild rice population, suggesting that these three members were putative targets of artificial selection during rice domestication. Furthermore, using the published genomic sequence information of 20 domesticated and five wild rice lines (Xu *et al.*, 2012), we performed Tajima's *D* test using Zhu *et al.* (2007)'s method for the 188 miRNAs with expression differences detected between the wild and cultivated lines by the miRNA array experiment (Table S8), and found significant positive selection at regions (< 10 kb) containing 29 of these *MIRNAs* (data not shown). This result suggests that these *MIRNAs* might be targets of domestication selection.

Taken together, these results provide evidence of expression divergence of miRNAs in cultivated and wild rice and demonstrate that in some cases the expression divergence of miRNAs could be a result of domestication selection.

Discussion

In this study, a *de novo* assembly of the *O. rufipogon* (the ancestor of the cultivated rice, *O. sativa*) genome was generated, and a genome-

Table 4 The predicted microRNA (miRNA) targets in *Oryza rufipogon* and *Oryza sativa*

miRNA*	<i>O. sativa</i>		<i>O. rufipogon</i>		Homolog	Annotation
	Target	Score	Target	Score		
osa-miR1435	Os03g42280.1	1.5	scaffold9489_g182	3	Y	B3 DNA binding domain containing protein, expressed
osa-miR2097-3p	Os01g14130.1	2.5	scaffold13939_g54	2	Y	Alpha/beta hydrolase fold, putative, expressed
osa-miR2097-5p	Os08g43920.1	3	scaffold15723_g69	2	ND	Carrier, putative, expressed
osa-miR2103	Os03g50830.1	0	scaffold9433_g80	2	Y	Conserved hypothetical protein
osa-miR2120	Os11g35750.1	0	Not Found	–	ND	Conserved hypothetical protein
osa-miR2928	Os01g22954.1	2	Not Found	–	ND	Serine carboxypeptidase, putative, expressed
osa-miR413	Os08g17400.1	2.5	Not Found	–	ND	WRKY DNA-binding domain containing protein, expressed
osa-miR416	Os08g31780.1	2.5	Not Found	–	ND	MLA1, putative, expressed
osa-miR5079	Os02g27130.1	1.5	Not Found	–	ND	Expressed protein
osa-miR5526	Os02g21009.2	2.5	Not Found	–	ND	Sodium/calcium exchanger protein, putative, expressed
osa-miR1427	Os02g16030.1	2.5	scaffold13728_g98	3	ND	Hairpin-induced protein 1 domain containing protein, expressed
osa-miR1851	Os01g61520.1	0.5	scaffold28_g88	2.5	Y	Hypothetical protein
osa-miR414	Os12g01320.1	0	scaffold2732_g45	0	Y	Retrotransposon protein, putative, Ty3-gypsy subclass
osa-miR5155	Os04g40660.1	2.5	scaffold6007_g136	2.5	Y	MA3 domain containing protein, expressed
osa-miR5162	Os10g41920.1	1.5	scaffold762_g56	3	ND	Expressed protein
osa-miR5484	Os02g50140.1	3	scaffold12274_g127	3	Y	Caleosin related protein, putative, expressed
osa-miR5532	Os07g38664.1	3	scaffold16393_g42	3	ND	Expressed protein
osa-miR1318	Os03g59790.1	0.5	Not Found	–	ND	EF hand family protein, putative
osa-miR1442	Os02g28720.1	2.5	Not Found	–	ND	Spotted leaf 11, putative, expressed
osa-miR2872	Os07g28050.1	0.5	Not Found	–	ND	Retrotransposon protein, putative, unclassified
osa-miR5074	Os10g06480.1	0	Not Found	–	ND	Retrotransposon protein, putative, Ty3-gypsy subclass
osa-miR5160	Os10g36130.1	1	Not Found	–	ND	Hypothetical protein

*miRNAs identified only in the cultivated rice (*O. sativa*) and their best predicted targets are listed. Y, yes; ND, not determined.

wide investigation of miRNAs was carried out in *O. rufipogon* using the next-generation sequencing technology and miRNA microarray. To our knowledge, this is the deepest sequenced (> 50×) genome of *O. rufipogon* so far and the first large-scale attempt at miRNA identification in wild rice. The deeper sequencing promises a better *de novo* assembly of the wild rice genome. Li *et al.* (2010a,b) suggested 30× as a minimal sequence depth required for achieving a proper assembly of the human genome. Meanwhile, the deeper sequencing also guarantees a proper sequence depth (e.g. > 4×) for single nucleotide polymorphism (SNP) between the wild and cultivated rice. For example, our sequencing coverage (*c.* 55× of *O. rufipogon* provided at least 4.5× in all 548 known rice miRNA loci. However, there were *c.* 40 miRNA loci with < 4× read coverage for each of the five wild lines that were sequenced to 8.5–11.5× of genomic coverage (Xu *et al.*, 2012) (Fig. S2). Comparative analyses of *MIRNAs*, miRNA::target interactions and miRNA expression in the cultivated and wild rice revealed a complex and dynamic evolutionary process of miRNAs during rice domestication and provided lessons for rice domestication, that is, *MIRNA* genes, like protein-coding genes, were significantly shaped by artificial selection.

Gain and loss of *MIRNAs* as a driven force for rice domestication

In rice, exhaustive miRNA investigations have been carried out in many tissues and developmental stages in the past 10 yr (www.mirbase.org). In this study, we identified 754 *MIRNAs* in *O. rufipogon*, 387 via small RNA population sequencing (i.e.

de novo approach) and 495 via a comparative genomic approach. Of the 754 *O. rufipogon MIRNAs*, 128 were identified by both approaches. Many *O. rufipogon MIRNAs* failed to be identified by the *de novo* approach although small RNAs were detected, most likely because of the high stringency of our *MIRNA* prediction program and/or low expression levels of these *MIRNAs* in the tissues we analyzed in the wild rice. Of the 387 *O. rufipogon MIRNAs* identified using the *de novo* approach, 259 were not found in the cultivated rice, suggesting that a number of *MIRNAs* have been lost during rice domestication. We also found that 48 *MIRNAs* were only present in *O. sativa* but not in *O. rufipogon*. In about half of these *MIRNAs*, mutations disabling the stem-loop structure in *O. rufipogon* were observed (Fig. 5); for the remainder, no corresponding genomic sequences were found in *O. rufipogon*, suggesting that these were most likely newly evolved *MIRNAs* through accumulation of mutations and/or genome rearrangement in the cultivated rice.

Functions of *MIRNAs* are achieved through regulating the expression levels of their target genes. Some targets of the 259 miRNAs identified only in *O. rufipogon* are transcription factors, such as the WRYK and MYB family genes, which could be involved in important developmental processes; however, the vast majority of their targets seem to be related to specific biological pathways (Table S4). Loss of these miRNAs means loss of regulation of their targets in the cultivated rice. Although it is not clear which particular developmental events and biological processes have been affected by loss of these miRNA::target interactions in cultivated rice and how the cultivated rice was shaped by these changes, the findings

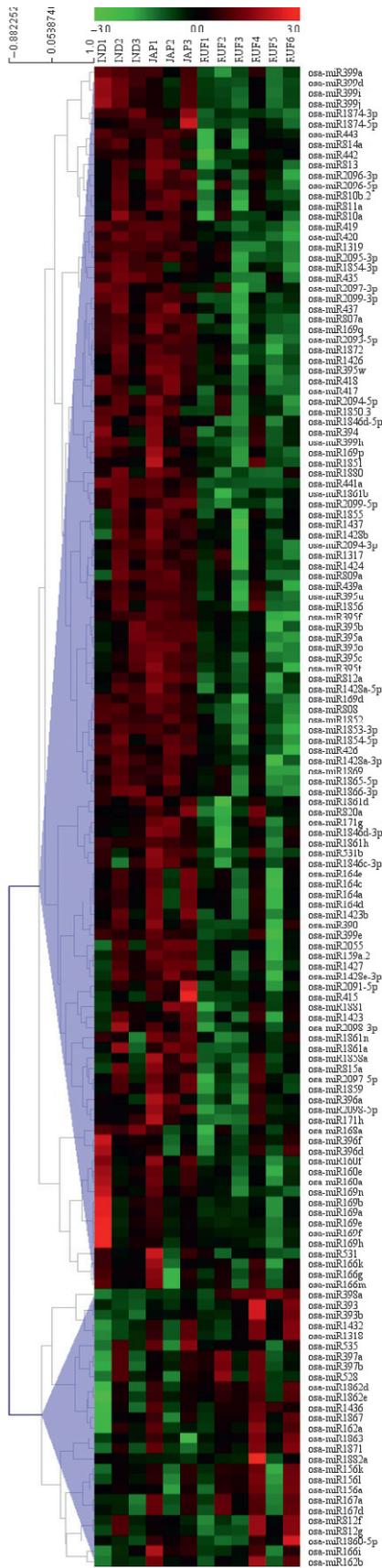


Fig. 6 Clustering of microRNAs (miRNAs) based on miRNA microarray chip results. Six cultivars and six wild rice (*Oryza rufipogon*) accessions are labeled at the top of the figure, while miRNAs are listed on the right.

demonstrate that evolution of *MIRNAs* share a similar trend as protein-coding genes, for which many have been lost in the cultivated populations during domestication (Lam *et al.*, 2010; Xu *et al.*, 2012). On the other hand, new miRNA::target interactions might be established for the fixed and newly evolved *MIRNAs* in cultivated rice. If the interaction plays an important role during rice domestication, we expected it to be stable in various sources of the cultivated rice. After checking the *MIRNA* sequences in the 40 domesticated rice accessions that were generated recently (Xu *et al.*, 2012), we found no mutations in 15 of the 48 newly evolved *MIRNAs*, suggesting that these *MIRNAs* might have experienced strong artificial selection and were important for rice domestication, as shown before for other *MIRNAs* (Wang *et al.*, 2007, 2010; Jiao *et al.*, 2010).

A similar phenomenon of birth and death of *MIRNAs* has been observed in *Arabidopsis* (Fahlgrén *et al.*, 2007, 2010; Ehrenreich & Purugganan, 2008; Zhang *et al.*, 2011). A frequent birth and death of *MIRNA* genes and dynamic miRNA::target interactions were observed through comparing miRNAs in *A. thaliana* with those in its close relative *A. lyrata*. These results suggest that both natural and domestication selection could play a significant role in shaping the small RNA populations in plants.

Expression of miRNAs was also shaped during domestication

We found that the sequences of about half of the 543 *O. sativa* *MIRNAs* remain unchanged during rice domestication, and that for the remaining *MIRNAs* common to *O. sativa* and *O. rufipogon*, most mutations (SNP or small indels) between them were heterogenous (Table S6), implying that the miRNA::target interactions for these *MIRNAs* were stable in the cultivated rice and its ancestor. However, both the miRNA microarray and high-throughput sequencing showed significant expression changes for some miRNA families in the cultivated rice and its ancestor. This could be caused by different transcriptional strengths of *MIRNAs* as a result of mutations in the promoter region that affect the recruitment efficiency of the transcription machinery or by mutations in the pri- or pre-miRNAs that affect miRNA biogenesis. These mutations could be the targets of artificial selection and be fixed in the cultivated populations. For the first possibility, no such case has been reported in *MIRNAs*, but it has been documented for protein-coding genes in rice. For example, a SNP in the 5' regulatory region of the rice-shattering gene *qSH1* caused loss of seed shattering owing to the absence of abscission layer formation in the cultivated rice (Konishi *et al.*, 2006). For the second possibility, it has been demonstrated that mutations in the pre-miRNAs, particularly in the loop-distal region of the hairpin, would affect the secondary structure and processing of pre-miRNAs, and consequently the biogenesis of miRNAs, including the cleavage accuracy and rate of DCL1 (Mateos *et al.*, 2010; Song *et al.*, 2010; Werner *et al.*, 2010). In addition, our population genetics investigation of the miR164 family has suggested a significant positive selection on at least three of its members in the domesticated rice population (Wang *et al.*, 2010 and this study). It has been shown that miR164 targets *NAC* domain transcription factors, including *Os12g41680*, *Os06g23650* and *Os04g38720*

(Wu *et al.*, 2009; Li *et al.*, 2010a,b). Overexpression of miR164b caused semi-dwarf and lower fertility in rice (Zhu *et al.*, 2012). Therefore, miR164 might be important for establishing the proper plant architecture in the cultivated rice.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31071393) and the PhD programs of the Foundation of Ministry of Education of China (20100101110096).

References

- Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. 2008. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. *Current Biology* 18: 758–762.
- Addo-Quaye C, Miller W, Axtell MJ. 2009. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* 25: 130–131.
- 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.
- Bartel DP. 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.
- Chong L. 2001. Molecular cloning—a laboratory manual, 3rd edn. *Science* 292: 547–610.
- Cuperus JT, Fahlgren N, Carrington JC. 2011. Evolution and functional diversification of *MIRNA* genes. *Plant Cell* 23: 431–442.
- Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* 39: W155–W159.
- Ehrenreich IM, Purugganan MD. 2008. Sequence variation of microRNAs and their binding sites in Arabidopsis. *Plant Physiology* 146: 1974–1982.
- Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangel JL *et al.* 2007. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of *MIRNA* genes. *PLoS ONE* 2: e219.
- Fahlgren N, Jogdeo S, Kasschau KD, Sullivan CM, Chapman EJ, Laubinger S, Smith LM, Dasenko M, Givan SA, Weigel D *et al.* 2010. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Plant Cell* 22: 1074–1089.
- Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
- German MA, Pillay M, Jeong DH, Hetawal A, Luo SJ, Janardhanan P, Kannan V, Rymarquis LA, Nobuta K, German R *et al.* 2008. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nature Biotechnology* 26: 941–946.
- Guo X, Gui Y, Wang Y, Zhu QH, Helliwell C, Fan L. 2008. Selection and mutation on microRNA target sequences during rice evolution. *BMC Genomics* 9: 454.
- He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F *et al.* 2010. Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22: 17–33.
- Hofacker IL. 2003. Vienna RNA secondary structure server. *Nucleic Acids Research* 31: 3429–3431.
- Jeong DH, Park S, Zhai J, Gurazada SG, De Paoli E, Meyers BC, Green PJ. 2011. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell* 23: 4185–4207.
- Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X *et al.* 2010. Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nature Genetics* 42: 541–544.
- Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312: 1392–1396.
- Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 39: D152–D157.
- Lam HM, Xu X, Liu X, Chen W, Yang G, Wong FL, Li MW, He W, Qin N, Wang B *et al.* 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics* 42: 1053–1059.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009a. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009b. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25: 1966–1967.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K *et al.* 2010a. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research* 20: 265–272.
- Li YF, Zheng Y, Addo-Quaye C, Zhang L, Saini A, Jagadeeswaran G, Axtell MJ, Zhang W, Sunkar R. 2010b. Transcriptome-wide identification of microRNA targets in rice. *Plant Journal* 62: 742–759.
- Llave C, Kasschau KD, Rector MA, Carrington JC. 2002. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14: 1605–1619.
- Mateos JL, Bologna NG, Chorostecki U, Palatnik JF. 2010. Identification of microRNA processing determinants by random mutagenesis of *Arabidopsis MIR172a* precursor. *Current Biology* 20: 49–54.
- Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, Cao X, Carrington JC, Chen X, Green PJ *et al.* 2008. Criteria for annotation of plant microRNAs. *Plant Cell* 20: 3186–3190.
- Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M. 2010. *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nature Genetics* 42: 545–549.
- Qi J, Zhao F, Buboltz A, Schuster SC. 2010. inGAP: an integrated next-generation genome analysis pipeline. *Bioinformatics* 26: 127–129.
- Song L, Axtell MJ, Fedoroff NV. 2010. RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Current Biology* 20: 37–41.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* 136: 669–687.
- Wang S, Zhu QH, Guo X, Gui Y, Bao J, Helliwell C, Fan L. 2007. Molecular evolution and selection of a gene encoding two tandem microRNAs in rice. *FEBS Letters* 581: 4789–4793.
- Wang Y, Shen D, Bo S, Chen H, Zheng J, Zhu QH, Cai D, Helliwell C, Fan L. 2010. Sequence variation and selection of small RNAs in domesticated rice. *BMC Evolutionary Biology* 10: 119.
- Werner S, Wollmann H, Schneeberger K, Weigel D. 2010. Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Current Biology* 20: 42–48.
- Wu L, Zhang Q, Zhou H, Ni F, Wu X, Qi Y. 2009. Rice MicroRNA effector complexes and targets. *Plant Cell* 21: 3421–3435.
- Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L *et al.* 2012. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnology* 30: 105–111.
- Zhang Y, Jiang W, Gao L. 2011. Evolution of microRNA genes in *Oryza sativa* and *Arabidopsis thaliana*: an update of the inverted duplication model. *PLoS ONE* 6: e28073.
- Zhu Q, Zheng X, Luo J, Gaut BS, Ge S. 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Molecular Biology and Evolution* 24: 875–888.
- Zhu Q-H, Curaba J, Lima JC, Helliwell C. 2012. Functions of miRNAs in rice. In: Sunkar R, ed. *MicroRNAs in plant development and stress responses*. Berlin, Heidelberg, Germany: Springer-Verlag, 149–176.
- Zhu QH, Spriggs A, Matthew L, Fan L, Kennedy G, Gubler F, Helliwell C. 2008. A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Research* 18: 1456–1465.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Expression divergence of the miR164 family in cultivated and wild rice.

Fig. S2 Number of miRNA loci with $< 4\times$ of coverage by Illumina reads from wild rice as found in this study and Xu *et al.* (2012).

Table S1 Accession number and geographic origin of the cultivated and wild rice used for microarrays in this study

Table S2 Global statistics of the *O. rufipogon* genome

Table S3 List of miRNAs identified in the wild rice in this study

Table S4 miRNA targets predicted by bioinformatic approaches

Table S5 miRNA targets predicted by degradome sequencing

Table S6 Genomic variations of 534 known rice miRNAs in the wild rice

Table S7 Selected microRNAs loci with significant signals in more than four chips and the signal value in each chip

Table S8 Expression of known miRNA families in developing grain and seedlings in *O. sativa* (Nipponbare) and *O. rufipogon* small RNA populations through Illumina sequencing

Table S9 Sequence diversity and neutrality test of miR164 in rice

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Trust, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews.
- Regular papers, Letters, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <25 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@ornl.gov)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**