This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author’s institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier’s archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright
Molecular evolution and selection of a gene encoding two tandem microRNAs in rice

Sheng Wang\textsuperscript{a}, Qian-Hao Zhu\textsuperscript{b}, Xingyi Guo\textsuperscript{a}, Yijie Gui\textsuperscript{a}, Jiandong Bao\textsuperscript{a},
Chris Helliwell\textsuperscript{b}, Longjiang Fan\textsuperscript{a,*}

\textsuperscript{a} Institute of Crop Science, Zhejiang University, Hangzhou 310029, China
\textsuperscript{b} CSIRO Plant Industry, Canberra, ACT 2601, Australia

Received 25 June 2007; revised 31 August 2007; accepted 4 September 2007
Available online 12 September 2007

Abstract It has been shown that overexpression of MIR156bc resulted in a bushy phenotype in maize and rice. Our results indicated that the MIR156bc locus was highly conserved among cereals, but not in dicots and that genome duplication events played an important role in the evolution of the miR156 family. Genetic diversity investigation at the locus indicated that only ~9% of nucleotide diversity observed in wild rice (O. rufipogon) was maintained in the cultivated rice and the neutral model was rejected (\(P < 0.05\)) based on Tajima’s D and Fu and Li’s D* and F* tests. To our knowledge, this is the first example of miRNA gene to be targeted by both natural and domestication selection in plants.

© 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Keywords: MicroRNA; MIR156bc; Selection; Genome duplication; Oryza sativa

1. Introduction

Small RNAs transcriptionally or posttranscriptionally regulate gene expression in eukaryotes. In plants, these small regulatory and noncoding RNAs are classified into microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs are ~21 base-pair (bp) long. They derive from hairpin structured miRNA precursors (pre-miRNAs) that are processed by an RNAse III-like enzyme called DCL1 (DICER-LIKE 1) from primary miRNAs (pri-miRNAs) transcribed by RNA polymerase II [1]. miRNAs posttranscriptionally down-regulate gene expression by cleavage or translational repression of target mRNAs. In plants, miRNAs were first found in 2002 [2–5]. To date, 4584 miRNAs have been identified in animals and plants (miRBase, http://microrna.sanger.ac.uk/, Release 9.2). In rice, 242 miRNAs of 62 families have been identified by cloning or computational prediction [6–10].

The miR156 family was one of the first characterized miRNA families in plants [5,11], and has 12 members in rice [5,7]. It is highly conserved in the plant kingdom and has been identified in 45 different plant species [12]. miR156 has been demonstrated to target SPL genes, which are plant specific transcription factors containing an SBP box [6,13–15]. Recently, evolution of miRNA families with multiple members (such as miR156, miR169 and miR395) has been investigated in Arabidopsis. It was found that duplication events played an important role in diversification and evolution of these miRNA families [16]. Duplication was also one of the main mechanisms involved in the evolution of several rice miRNA families, such as the miR159 and miR395 families [17,18]. Of 12 miR156 family members located on six chromosomes, miR156b and miR156c (miR156b/c hereafter) are tandem miRNAs on chromosome 1. A full-length cDNA (AK110797) encodes both miRNAs. Overexpression of miR156b resulted in reduced plant height and increased number of tillers in rice [13]. In maize the Corngrass1 (Cgl) gene also encodes tandem MIR156 genes (MIR156b and MIR156c). The dominant Cgl mutant shows dwarfing, multiple-tillers and a bushy phenotype due to overexpression of miR156b/c [19].

Of the six domestication genes identified to date, a notable feature is that five of them encode transcription factors that regulate or target other genes by directly binding to their DNA [20]. Transcription factors, in turn, are major targets of miRNAs. For example, tga1, a transcription factor of the SPL family known to have had a role in the domestication of maize from teosinte, is one of the targets of Cgl [19].

In this study, we show that tandem and whole genome duplication are the driving force for expansion of the miR156 family in rice, and that genomic organization around the MIR156bc locus is highly conserved in cereals. Sequencing the MIR156bc locus in 30 cultivated rice cultivars and 15 wild rice accessions revealed that this locus experienced strong natural and potential domestication selection.

2. Materials and methods

2.1. Plant materials

Forty-five diverse Oryza accessions were selected from a wide range of geographical locations to represent a broad range of the genetic diversity within cultivated rice (Oryza sativa) and its wild ancestor, O. rufipogon. O. sativa includes 30 domesticated lines (15 indica and 15 japonica cultivars). Detailed information of the 45 rice accessions is listed in supplemental Table S1.

2.2. PCR and DNA sequencing

To investigate sequence variation in the MIR156bc locus among the selected rice lines, a ~900bp genomic fragment that covers miR156b and miR156c was amplified from each cultivar/accession using the following pair of primers. Forward primer: 5’ TGGCTAGCTAATCTGAGA 3’; reverse primer: 5’ TCAGAAATACCTCACAGAGAGTGATACG 3’. Primers were designed based on the genomic sequence

0014-5793/$32.00 © 2007 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.
of japonica cultivar Nipponbare using Primer3 [21]. The primers were compared to the rice genome sequence (NCBI GenBank) to ensure their specificity.

Genomic DNA was extracted from fresh rice leaves using a cetyltrimethylammonium bromide (CTAB) protocol [22]. PCR reactions were carried out on a thermocycler (Eppendorf) under the following conditions: 95 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 53 °C for 30 s and extension at 72 °C for 90 s, with a final extension at 72 °C for 10 min. PCR products were visualized on 0.8% agarose gel. A product of expected size was amplified from each sample. The amplified products were purified using a glassmilk PCR purification Kit (BioDev-Tech, China). For O. sativa cultivars, purified PCR products were directly sequenced on both strands using the forward or reverse primer. For O. rufipogon, in which either homozygous or heterozygous individuals might exist, PCR fragments were cloned into pGEM T-easy vectors (PMD19-T, Takara) and sequenced using the forward or reverse primer. At least three independent clones were sequenced. All sequences were deposited into GenBank, with GenBank accession numbers EU004233-EU004275.

2.3. Genome data

The TIGR rice pseudomolecules (release 5) and sorghum (Sorghum bicolor) genome segments were downloaded from http://www.tigr.org/tdb/e2k1/osa1/ and http://www.phytozome.net/sorghum, respectively.

2.4. Sequence analysis

MIR156 precursor sequences were aligned using MAFFT (version 5.8) [23] for the construction of phylogenetic trees. Neighbor-joining (NJ) phylogenies based on the Kimura 2-parameter distance matrix were generated by MEGA version 3.1 [24]. Bootstrap confidence values were obtained by 1000 replicates. Over 60% supporting notes were shown. Watterson’s estimator of θ [25], the average pairwise nucleotide diversity [26] and Tajima’s D and Fu and Li’s D* and F* tests of neutrality [27,28] were estimated using DNASP version 4.10.2 [29].

For rice syntenic analysis, the duplicated blocks among different chromosomes available in TIGR’s Rice Genome Annotation (http://www.tigr.org/tdb/e2k1/osa1/) were used as guides, and the corresponding genomic sequences containing miR156 family members were downloaded from TIGR. Repeats were filtered using Repeatmasker (tigr_plant_repeat.lib, mips_REdat_4.3_rptmsk.lib, repbase) before the duplicated genomic sequence pairs were aligned using BLASTN. The hits with an E value less than 1e-10 were used to identify syntenic regions containing miR156 members (Fig. 1 A and Figure S1). Only MIR156k on chromosome 9 did not have a counterpart at its syntenic region on chromosome 8 (Figure S1). Phylogenetic analysis indicated that MIR156a-k locate at the same genomic position. The transcript of MIR156bc containing miR156b and miR156c is shown, with the arrow indicating their transcription direction. Detailed alignments of duplicated segmental pairs containing miR156 members see Fig. 2 and Fig. S1; (B) Phylogenetic tree of miR156 family. MIR156 precursor sequences were used to generate the NJ tree.

Fig. 1. Genomic localization and phylogenetic tree of the miR156 family in rice. (A) Distribution of miR156 family members and duplication events contributed to evolution of miR156 family in rice genome. a–k refer to miR156a-k (miR156h and miR156j locate at the same genomic position). The transcript of MIR156bc containing miR156b and miR156c is shown, with the arrow indicating their transcription direction. Detailed alignments of duplicated segmental pairs containing miR156 members see Fig. 2 and Fig. S1; (B) Phylogenetic tree of miR156 family. MIR156 precursor sequences were used to generate the NJ tree.

3. Results and discussion

3.1. Evolution of MIR156bc through duplication events

Twelve MIR156 genes located on six chromosomes have been identified in rice. MIR156b and MIR156c are located close to each other on chromosome 1 and are within a single transcript AK110797 (Fig. 1A). MIR156j and MIR156h, which have the same mature miRNAs but different length precursors, are located at the same genomic region on chromosome 6 (Fig. 1A). Based on analysis of rice genomic sequences, a whole genome duplication event that occurred about 70 million years ago (mya) has been well documented [30–33], which predates the divergence of cereals (~50 mya; [34]) but postdates the divergence of monocots and dicots (~200 mya; [35]). Seventeen large duplicated blocks have been maintained [36]. We found that 11 of 12 MIR156 loci (except MIR156a) are located in the duplicated regions (Fig. 1A). The collinear regions containing MIR156 genes were further investigated in detail; a perfect syntenic relationship was found for four MIR156 pairs (MIR156bc and MIR156l, MIR156i and MIR156c, MIR156d and MIR156hi, and MIR156f and MIR156g) (Fig. 2A and Figure S1).
monocots and dicots. Interestingly, MIR156l is the only MIR156 member in the syntenic region of MIR156b/c on chromosome 5 (Fig. 2A). To see whether there is another MIR156 member near MIR156l, a 1 Mb genomic syntenic region (TIGR chr5_4054114_5935558, see Fig. 2A) containing MIR156l was used for prediction of putative miRNAs using RNA folding software (http://www.tbi.univie.ac.at/~ivo/RNA/). No new putative MIRNA gene was found (data not shown), suggesting that the putative MIRNA corresponding to MIR156b or MIR156c on chromosome 5 might have been lost after the whole genome duplication event. Similarly, the paralog of MIR156k on chromosome 8 that arose from whole genome duplication was found to be lost (Fig. 1A). The miR156 family of Arabidopsis also experienced a large expansion via segmental duplication events and loss of family members [16]. In addition, tandem or segmental duplication events have been shown to have had a role in evolution of other miRNA families in rice, such as miR160, miR162, miR167, miR169, miR171 and miR395 [17,18]. Therefore, as for protein-coding genes, duplication and loss of duplicates may represent one of the main evolutionary routes for birth and death of MIRNAs in plants.

Our analysis showed that over half of the duplicated MIR156 genes have survived after genome or segmental duplication in rice. The retained proportion of duplicated MIR156 genes is significantly higher than that of protein-coding genes, for which less than 25% of the duplicates are thought to have survived following the genome duplication in rice [30]. This estimation is much lower (3.2–9.5%) if the initial numbers of genes in a duplicated region are considered [37]. The rice SPL family, targets of miR156, has also undergone dramatic expansion and has at least 19 members [13], and derived from the genome duplication event (data not shown). Identification and characterization of the functions of the SPL genes should shed light on the co-evolution relationship between the miR156 and the SPL gene family.

3.2. Highly conserved MIR156b/c among cereals

Large-scale genome sequencing and well documented genetic maps have made the investigation of synteny among cereals possible. A 3.3 Mb genomic alignment containing the MIR156b/c locus in rice, maize and sorghum is shown in Fig. 2B. A large sorghum genomic segment (Super_67, ~3.3 Mb) sequenced by the Sorghum Genome Project, DOE-JGI Community Sequencing Program (CSP, http://www.phytozome.net/sorghum), which contains the MIR156b/c locus, showed high synteny to the corresponding genomic region of
Coalescent simulations. Detailed accessions for two groups see Table S1. *P < 0.05 and **P < 0.02.

Simulations of Tajima's


in transgenic plants overexpressing miR156b (rice;[13]) or cereals. The function of MIR156b/c served at least in rice, maize and sorghum and perhaps in all cereals. However the spacer size between the two miRNA sequences of the two mature miRNAs are identical in three cereals. The conserved function of MIR156b/c locus was highly conserved at least in rice, maize and sorghum and perhaps in all cereals. The function of MIR156b/c also appears to be conserved in rice and maize as the phenotypic changes observed in transgenic plants overexpressing miR156b (rice; [13]) or miR156b/c (maize; [19]) are similar.

MIR156b/c are encoded by a single transcript (full-length cDNA AK110797), suggesting they are co-expressed. However, the mature miR156b and miR156c sequences are identical so it is not possible to determine if both miRNAs are expressed. Co-transcription of MIRNAs is not unique to MIR156b/c. For example, 24 MIR395 genes are organized into four compact clusters, each transcribed as a single transcript in rice [17]. Co-transcription of similar or identical MIRNA genes might have a dosage effect in plants. In contrast, animal MIRNA genes in a cluster usually are not homologs although they are evolutionarily related, such as those in the mir-17 gene cluster. These non-homologous MIRNA genes could regulate multiple functionally related genes simultaneously [38].

3.3. Molecular diversity and selection of MIR156b/c

The conserved function of MIR156b/c in maize and rice led us to investigate whether selection forces have acted on it during the origin and domestication of rice [13,19]. To answer this question, genomic sequences of the MIR156b/c locus were amplified from 30 rice cultivars and 15 wild rice accessions. Several methods were used to examine whether the nucleotide polymorphism observed in the MIR156b/c locus fits the neutral model. For loci selected only by domestication and genetic improvement, a significant selection signal should be detected in cultivated rice but not in wild rice. In contrast, for the loci that are important for both cultivated and wild rice, a selection signal should be detected in both populations [39]. Our results indicated that the MIR156b/c locus experienced strong natural selection in O. rufipogon, and natural and/or domestication selection in the cultivated rice because the neutral model was rejected (P < 0.05) based on significant parameters of Tajima’s D test, and Fu and Li’s D* and F* test (Table 1). Coalescent Simulations of Tajima’s D test also supported this result. In view of the importance of the MIR156b/c locus in determination of plant appearance in rice and maize and the fact that O. rufipogon has a similar appearance to that of cultivated rice [40], it is no surprise to see strong natural selection on the MIR156b/c locus in O. rufipogon. To know which region of the MIRNA gene is the selection target, Tajima’s D test was carried out across the MIR156b/c locus using sequences amplified from 15 accessions of O. rufipogon. The test indicated that a 100–150 bp region (including mature miR156) upstream of the mature miR156b and miR156c in each of their precursors experienced stronger selection. This result implies the importance of mature miRNA sequences and their 5’ flanking regions in miRNA biogenesis and function. Nucleotide diversity of the MIR156b/c locus in the cultivated rice was 8.9% of that in O. rufipogon according to an estimator of average pairwise nucleotide diversity (π). In other words, MIR156b/c lost >90% of its nucleotide diversity during domestication (Table 1). Another estimator (Watterson’s θ) also showed the same trend as π. Across the MIR156b/c locus, the nucleotide diversity was relatively lower in MIR156b and the MIR156c precursors. The mature sequences of miR156b and miR156c were extremely conserved among the cultivated and wild rice with no nucleotide mutation or insertions/deletions in the 45 samples from the two groups.

Genetic diversity in neutral (unselected) genes is expected to be reduced only by bottleneck or demographic effects, therefore retaining more diversity than selected genes [39]. A recent investigation on domestication bottlenecks based on 10 neutral loci in rice indicated that domesticated rice maintains about 52.1% (π) of the variability found in its progenitor O. rufipogon [41]. In the MIR156b/c locus, however, only 8.9% diversity was maintained in the cultivated rice. Therefore the bottleneck effect is not enough to explain this dramatic diversity loss. We believe that the MIR156b/c locus was a target of domestication selection, although bottleneck effects could have also narrowed the genetic diversity of this locus in the cultivated rice during domestication process. This result suggests that domestication selection could target not only transcription factors as shown before [20] but their upstream regulators, for example, MIRNAs.

Table 1

Nucleotide polymorphisms and selection test

<table>
<thead>
<tr>
<th>Group</th>
<th>Sample size (n)</th>
<th>π (×1000)</th>
<th>θ (×1000)</th>
<th>D</th>
<th>P</th>
<th>D'</th>
<th>F*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cultivated rice</td>
<td>30</td>
<td>0.43 (0.16)</td>
<td>1.61 (0.84)</td>
<td>−2.01*</td>
<td>0.000</td>
<td>−3.24**</td>
<td>−3.34**</td>
</tr>
<tr>
<td>Wild species</td>
<td>15</td>
<td>4.81 (0.87)</td>
<td>9.66 (3.92)</td>
<td>−2.09*</td>
<td>0.004</td>
<td>−2.85**</td>
<td>−3.04**</td>
</tr>
</tbody>
</table>

Note: π, the average pairwise nucleotide diversity [20]; θ, Watterson’s estimator [19]; D, Tajima’s D [21]; D’ and F*, Fu and Li’s D* and F* [22]; P, Coalescent simulations. Detailed accessions for two groups see Table S1. *P < 0.05 and **P < 0.02.

This work was supported by the National High Technology Research and Development Program of China (2006AA10A102) and 973 Program (2006CB101700). Dr. Fan thanks Pao Zhao-long and Pao Yu-kong Scholarship for Chinese Students Studying Abroad for travel grants.

Appendix A. Supplementary data

References


