Mol Genet Genomics (2015) 290:593–602
DOI 10.1007/s00438-014-0943-0

ORGINAL PAPER

Genome-wide identification and evolutionary analysis of positively selected miRNA genes in domesticated rice

Qingpo Liu · Hong Wang · Haichao Hu · Hengmu Zhang

Received: 18 May 2014 / Accepted: 17 October 2014 / Published online: 2 November 2014

Abstract The next-generation sequencing of tens to hundreds of plant genotypes made the uncovering of miRNA genes evolution available at the genome-wide level. Using the combinations of population genetics and evolutionary biology approaches, we have identified 21 miRNA loci having significant negative Tajima's *D* and Fu and Li's *D** and *F** values, of which 14 miRNAs (ps-miRNAs) showing clear signatures of positive selection in domesticated rice. The average sequence diversity (π) of the 21 miRNAs in cultivated rice is only 13.8 % of that in their wild progenitors. Interestingly, protein-coding genes immediately flanking these ps-miRNAs are apparently under weaker selective constraints. Totally, the 21 miRNAs are predicted to target 68 mRNA genes, of which 12 targets are estimated to have endured positive selection during rice evolution. In addition, the expression pattern and potential biological functions of ps-miRNAs targets are further investigated by searching published micro-array data and different mutant databases, respectively. We conclude that miRNAs, like protein-coding genes, should be crucial for driving rice evolution. These analyses may deepen our understanding on the miRNA genes evolution and functions during rice domestication.

Keywords Rice · miRNA · Positive selection · Evolution · Domestication

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-014-0943-0) contains supplementary material, which is available to authorized users.

Q. Liu (✉) · H. Wang · H. Hu
Department of Agronomy, College of Agriculture and Food Science, Zhejiang A & F University, Lin'an, Hangzhou 311300, People's Republic of China
e-mail: liuqp@zafu.edu.cn

Q. Liu · H. Wang · H. Hu
The Key Laboratory for Quality Improvement of Agricultural Products of Zhejiang Province, College of Agriculture and Food Science, Zhejiang A & F University, Lin'an, Hangzhou 311300, People's Republic of China

H. Zhang (✉)
Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, People's Republic of China
e-mail: zhhengmu@tsinghua.org.cn

Introduction

MiRNAs (microRNAs) are a class of endogenous small non-coding RNAs that are ubiquitously present in living organisms except for fungi (Chen 2005; Zhuo et al. 2013). These small RNAs are key regulator of gene expression by mediating target cleavage or translational repression, and play important roles in regulation of a diverse of biological processes consequently (Gielen et al. 2012; Pashkovskiy and Ryazansky 2013). In a species, the miRNA repertoire always consists of a set of deeply conserved and many newly evolved miRNAs (Rajagopalan et al. 2006). Conserved miRNAs are often under stronger selective constraints (Cuperus et al. 2011; Liu et al. 2013), and have experienced a concerted evolution with their binding target sites (Ehrenreich and Purugganan 2008; Guo et al. 2008), whereas the young miRNAs are always lineage- or species-specific, more divergent, and tend to lack targets (Cuperus et al. 2011).

Growing evidence shows that single nucleotide polymorphisms (SNPs), like in protein-coding genes, widely exist in miRNAs (Ehrenreich and Purugganan 2008; Wang

et al. 2010; Gong et al. 2012; Liu et al. 2013). MiRNA-related SNPs, especially those in stems and mature miRNAs, may change the stability of secondary structure, and further affect the miRNA biogenesis and function by modifying the miRNA-target interactions (Gong et al. 2012; Liu et al. 2013). For example, a naturally occurred C/T transition within the mature miR164a, which reduced the mature miRNAs generation, significantly affects the leaf shape and shoot architecture in *Arabidopsis thaliana* (Todesco et al. 2012). To date, a large-scale investigation of miRNA-related SNPs has been performed in *Arabidopsis* (Ehrenreich and Purugganan 2008) and rice (Liu et al. 2013), respectively, where 364 out of 591 rice miRNA precursors are reported to possess one or more SNPs (Liu et al. 2013). Particularly, in some miRNA loci, a stronger positive selection signature has been detected (Wang et al. 2007; de Meaux et al. 2008; Wang et al. 2010), a result inferring the complicated evolutionary mechanisms of miRNAs in plants.

Recently, the whole genomic variations of several plant species, including rice (Huang et al. 2012; Xu et al. 2012), foxtail millet (Jia et al. 2013), maize (Hufford et al. 2012), cucumber (Qi et al. 2013), watermelon (Guo et al. 2013), and soybean (Li et al. 2014) have been extensively investigated by re-sequencing tens to hundreds of cultivated and wild plant genotypes, and a plenty of regions with low sequence diversity (LDRs) in cultivated genotypes that are probably associated with domestication, have been identified subsequently (Huang et al. 2012; Xu et al. 2012; Guo et al. 2013; Jia et al. 2013; Qi et al. 2013). The amounts of high-throughput SNP data provide us the unprecedented opportunity for further understanding the miRNA-related SNPs, their evolution, and possible relationships with plant domestication (Henry 2012).

The modern rice (*Oryza sativa*) is commonly considered to be originated from its Asian wild ancestor (*Oryza rufipogon*) thousands of years ago (Kovach et al. 2007). During the periods of rice domestication, tens of protein-coding genes associated with important agronomic traits have experienced strong artificial selection and/or nature selection (Asano et al. 2011; Sang and Ge 2013; Zhu et al. 2013). Interestingly, by analyzing the previously identified domestication-related selective sweeps (He et al. 2011; Huang et al. 2012; Xu et al. 2012; Yang et al. 2012; Yonemaru et al. 2012), as well as comparing with the whole miRNA repertoire of the deeply sequenced Chinese wild rice genome (*O. rufipogon*; Wang et al. 2012), some miRNAs are considered to be putative domestication-related candidates (Wang et al. 2012; Liu et al. 2013), which may have played important roles for driving rice evolution. Given the functional importance of miRNAs, it is intriguing to further uncover their evolutionary patterns and mechanisms in domesticated rice. In this study, based on the 5.04 millions SNPs generated by Huang et al. (2012) and 592 miRNA precursors (release 20), 21 miRNAs evolving non-neutrally have been characterized in domesticated rice, where several miRNAs and their specific target genes may become common targets for human selection during rice domestication.

Materials and methods

Sequence data

A total of 592 rice miRNA precursors (pre-miRNAs) and 5.04 millions high-quality SNPs were downloaded from the miRBase database (release 20.0; Kozomara and Griffiths-Jones 2014) and the RiceHap3 database (http://202.127.18.221/RiceHap3/index.php), respectively. The SNP data that were derived from 462 cultivated rice and 446 *O. rufipogon* genotypes was based on the IRGSP v4 rice genomic sequences (Huang et al. 2012).

Tests of neutrality

The 592 pre-miRNA sequences were BLASTed against the IRGSP v4 genomic sequences with *E* value cutoff 10^{-10}. These miRNAs that were exactly mapped to the reference genome were recorded. After that, the chromosomal locations of pre-miRNAs and SNPs were compared to map SNPs onto pre-miRNAs. Then, the SNPs were used as input for the following analyses. MAFFT v7 (Katoh and Standley 2013) was utilized for multiple sequence alignments. The software DnaSP v5.10 (Librado and Rozas 2009) was employed to calculate the sequence diversity parameters π and θ to assess the population mutation rate per locus, and to perform the Tajima's *D* (Tajima 1989) and Fu and Li's *D** and *F** tests (Fu and Li 1993) to estimate the neutrality of DNA polymorphisms within a given population. The two neutrality tests were all calculated based on the segregating sites. In addition, the Fay and Wu's *H* test was performed using DnaSP v5.10 to measure an excess of high-frequency mutations in the cultivated and wild rice populations, respectively (Fay and Wu 2000). In this analysis, the predicted miRNAs of the Africa rice *Oryza glaberrima* were used as outgroup sequences.

Identification of orthologous gene pairs

The amino acid and protein-coding sequences (CDSs) of *Brachypodium distachyon* were downloaded from the Phytozome database (version 8.0; http://www.phytozome.net/), and the orthologous gene pairs of rice and *B. distachyon* were downloaded from the PhylomeDB (Huerta-Cepas et al. 2011) and MetaPhOrs (Pryszcz et al. 2011) databases,

respectively. The protein-coding genes immediately up- and down-stream of each of the identified miRNAs being under selection in rice were extracted based on their chromosomal coordinates, and then were used as query to search against the *B. distachyon* sequences to identify their orthologous counterparts. The amino acid sequences of orthologs were aligned using MAFFT v.7 (Katoh and Standley 2013), and then the codon alignments of CDS sequences were generated based on the resulting amino acid alignments using a custom PERL program. The yn00 program implemented in the PAML v4.4 package (Yang 2007), with the Yang and Nielsen method (Yang and Nielsen 2000), was adopted to calculate the pair-wise synonymous (d_S) and non-synonymous (d_N) distance between the orthologous genes of rice and *B. distachyon*.

Target gene prediction, expression, and functional analysis

A plant miRNA-target prediction server, psRNATarget (Dai and Zhao 2011), was employed to predict potential targets for the identified putative positively selected miRNAs. To reduce false positives, the parameter "maximum expectation" value was set to 2.0 (Dai and Zhao 2011). The micro-array dataset GSE7951 was downloaded from the GEO (Gene Expression Omnibus) database in NCBI (National Center for Biotechnology Information) to investigate the expression pattern of target genes in nine rice tissues (Li et al. 2007). The software dChip 2008 (Li 2008) was used to perform the cluster analysis and display the expression patterns of tested target genes. In addition, the PMRD database (Cui et al. 2012) was searched to examine whether some target genes were involved in plant male reproduction.

To further understand the biological functions of targets, the OryGenesDB database (Droc et al. 2009) was searched using the gene locus ID to obtain the information on the inserted mutants of target genes. Then, the rice mutant databases including OTL (http://oryzatagline.cirad.fr/), NIAS (https://tos.nias.affrc.go.jp/), and RMD (http://rmd.ncpgr.cn/), were further searched to examine whether these mutants have obviously observed phenotypes recorded (Wei et al. 2013).

Results

Identification of positively selected miRNAs (ps-miRNAs) in cultivated rice genotypes

It is found that 541 out of the 592 pre-miRNAs are exactly mapped to the IRGSP v4 rice genome. The pre-miRNAs are on average 4,784.9 and 4,639.5 bp away from their immediately up- and down-stream protein-coding genes, respectively, if the ones located at intron sequences are excluded. Considering that pre-miRNAs, like protein-coding genes, possess promoter sequences and other regulatory elements, 1,500 nt flanking sequences on each side of the pre-miRNAs together with the precursors were retrieved and analyzed further. By comparing the 5.04 millions of SNPs (Huang et al. 2012) and 541 pre-miRNAs, the SNPs were projected onto each of the concatenated sequences.

To test whether the miRNA genes evolve neutrally, two neutrality tests including Tajima's *D* (Tajima 1989) and Fu and Li's *D** and *F** (Fu and Li 1993) were performed for cultivated and wild rice, respectively. In total, twenty-one miRNA loci were found to have significant negative Tajima's *D* and Fu and Li's *D** and *F** values in cultivated but not in the wild rice genotypes ($p < 0.05$), indicative of their non-neutral evolutionary pattern (Table 1). At these miRNA loci, the sequence diversity (π) was estimated at ~0.00029 in cultivated rice, which is 6.26-fold lower than that in common wild rice (~0.0021). In other words, these loci reduced, on average, approximately 86.2 % of sequence diversity in the cultivated rice population compared to their wild progenitors, where *MIR167i* is a typical example that has lost 98.1 % of its sequence diversity during evolution. Moreover, the sequence diversity of these miRNA loci was significantly lower than that of 111 randomly chosen STS (sequence-tagged sites) fragments ($\pi = 0.00229$; Caicedo et al. 2007), and that of 631 and 614 randomly selected protein-coding gene fragments (≤1,500 bp, with or without intronic regions; $\pi = 0.00048$) and intergenic sequences (≤5,000 bp; $\pi = 0.00054$), respectively. The results suggest that these miRNA loci may have potentially undergone selection during rice evolution.

In addition to the Tajima's *D* and Fu and Li's *D** and *F** tests, another independent test, Fay and Wu's *H* test (Fay and Wu 2000) was performed to examine whether the 21 miRNA loci had experienced positive selection in their evolutionary history. Under expectation, all these miRNA loci had significant negative Fay and Wu's *H* values, except for 7 miRNAs having no homologs in *O. glaberrima* (Table 1), suggesting that most of these miRNA loci should have experienced positive selection in cultivated rice. Notably, 9 miRNAs were also detected to have the signatures of positive selection in the wild rice population (Table 1). Accordingly, *MIR821c*, *MIR160f*, *MIR395c*, *MIR395d*, *MIR1436* might be possibly under direct selection in cultivated rice, based on the role proposed by Yamasaki et al. (2005). In addition, by searching against the domestication-related low sequence diversity regions (LDRs) identified previously (He et al. 2011; Huang et al. 2012; Xu et al. 2012), five miRNA loci including *MIR821b*, *MIR167i*, *MIR160f*, *MIR815a*, and *MIR1436* are considered as ones having experienced selective sweeps during rice domestication.

Table 1 Summary of nucleotide polymorphisms and neutrality tests at 21 miRNA loci

MIRNA	*O. rufipogon*					*O. sativa*							
	S	π	θ	*D*	*H*	*S*	π	θ	*D*	CS*p*	*D**	*F**	*H*
MIR5533	87	0.0017	0.0040	−1.662	−11.043***	50	0.0003	0.0023	−2.441***	0.000	−3.089*	−3.367**	−10.724***
MIR1320	45	0.0011	0.0022	−1.373	−13.466***	15	0.0000	0.0007	−2.204**	0.000	−6.615**	−5.915**	−9.848***
MIR821b	76	0.0032	0.0035	−0.219	NA	37	0.0005	0.0017	−1.958*	0.001	−2.949*	−3.036**	NA
MIR821c	76	0.0023	0.0035	−0.947	−1.538	33	0.0003	0.0015	−2.063*	0.000	−2.576*	−2.849*	−8.565***
MIR167i	50	0.0016	0.0023	−0.904	−6.874*	14	0.0000	0.0007	−2.198**	0.000	−4.869**	−4.639**	−9.823*
MIR160f	59	0.0014	0.0029	−1.479	0.138	15	0.0001	0.0007	−2.185**	0.000	−3.226*	−3.414**	−7.808***
MIR815a	56	0.0017	0.0027	−1.042	−4.561*	15	0.0000	0.0007	−2.207**	0.000	−2.549*	−2.926*	−9.789***
MIR5489	74	0.0021	0.0034	−1.110	NA	33	0.0004	0.0015	−1.876*	0.000	−2.576*	−2.749*	NA
MIR395c	79	0.0023	0.0038	−1.168	1.802	55	0.0007	0.0027	−2.110**	0.000	−2.670*	−2.903*	−3.277*
MIR395d	78	0.0023	0.0038	−1.131	−0.122	55	0.0007	0.0027	−2.110**	0.000	−2.670*	−2.903*	−6.991***
MIR156b	50	0.0020	0.0024	−0.400	NA	19	0.0002	0.0009	−1.873*	0.001	−3.055*	−3.123**	NA
MIR156c	49	0.0021	0.0023	−0.284	NA	20	0.0002	0.0010	−1.912*	0.000	−2.879*	−3.013**	NA
MIR156i	55	0.0020	0.0027	−0.700	−7.144*	36	0.0004	0.0017	−2.011*	0.000	−3.061*	−3.143**	−17.065***
MIR1436	60	0.0018	0.0028	−1.092	−3.557	25	0.0001	0.0012	−2.288**	0.000	−3.152*	−3.382**	−7.431***
MIR1872	58	0.0025	0.0028	−0.370	−8.427*	30	0.0003	0.0015	−2.161**	0.000	−2.485*	−2.843*	−16.409***
MIR1846c	80	0.0032	0.0039	−0.494	−13.277**	35	0.0004	0.0017	−2.092*	0.000	−2.360*	−2.716*	−29.268***
MIR5154	66	0.0023	0.0032	−0.769	−15.985***	19	0.0001	0.0009	−2.253**	0.000	−4.238**	−4.163**	−13.691***
MIR1882c	79	0.0019	0.0038	−1.450	NA	33	0.0003	0.0016	−2.182**	0.000	−2.576*	−2.912*	NA
MIR1881	87	0.0034	0.0039	−0.353	−32.971***	41	0.0006	0.0018	−1.883*	0.001	−2.540*	−2.714*	−37.101***
MIR812f	71	0.0019	0.0034	−1.207	NA	23	0.0003	0.0011	−1.880*	0.000	−2.943*	−3.033**	NA
MIR1883b	63	0.0015	0.0030	−1.436	NA	32	0.0002	0.0015	−2.254**	0.000	−3.122*	−3.325**	NA

S number of segregating sites

π average number of nucleotide differences per site between two sequences (Nei 1987) calculated on the total number of polymorphic sites

θ Watterson's estimator (Watterson 1975)

D Tajima's *D* (Tajima 1989)

H Fay and Wu' *H* test (Fay and Wu 2000)

*D** and *F** Fu and Li's *D** and *F** (Fu and Li 1993)

CS*p* Coalescent simulations for Tajima's *D*

NA not available

$^{*}P < 0.05$, $^{**}P < 0.02$, and $^{***}P < 0.001$

Selective constraint of ps-miRNAs flanking protein-coding genes

To further examine the evolutionary mechanisms of ps-miRNAs, the protein-coding genes immediately flanking ps-miRNAs were retrieved, and the synonymous (d_S) and non-synonymous (d_N) substitution patterns between rice and *B. distachyon* orthologs were compared subsequently. In the analysis, the protein-coding genes immediately flanking 21 randomly selected non-ps-miRNAs were retrieved as control. The comparison of d_N and d_S rate values of orthologous gene pairs shows that ps-miRNA flanking genes are apparently under weaker selective constraints, as reflected from their significant higher average ω value (0.188 ± 0.025 vs. 0.144 ± 0.017, one-tail *t* test $p = 0.035$) and non-synonymous substitution rate value (d_N; 0.293 ± 0.061 vs. 0.132 ± 0.015, one-tail *t* test $p = 0.009$) when compared with non-ps-miRNA flanking genes. However, no significant difference in the synonymous rate (d_S) values was observed between ps- and non-ps-miRNAs flanking orthologous pairs (1.695 ± 0.258 vs. 1.171 ± 0.246, one-tail *t* test $p = 0.143$).

Target prediction and selection analysis

To better understand the biological function of ps-miRNAs, the target genes of ps-miRNAs were predicted using psRNATarget (Dai and Zhao 2011) to search against the MSU rice transcripts database. A total of 68 targets with "maximum expectation" value ≤2.0 were predicted for the 21 ps-miRNAs, which were subsequently subjected to the DnaSP v5.10 (Librado and Rozas 2009) to test whether

they evolved neutrally. It was found that 12 out of 68 target genes had significant negative Tajima's *D* and/or Fu and Li's *D** and *F** values (Table 2), suggesting that some targets of ps-miRNAs might have also experienced positive selection during rice evolution.

Functional annotation of putative positively selected target genes

The expression profiling of target genes that were under selection was investigated by searching against the GSE7951 micro-array data. In the first glance, most of targets were found to be expressed in more than two tissues, with the exception of *LOC_Os04g59430.1* that showed a tissue-specific expression pattern (Fig. 1). However, five genes including *LOC_Os08g39890.1*, *LOC_Os12g16350.1*, *LOC_Os02g04680.1*, *LOC_Os12g16350.13*, and *LOC_Os02g07960.2* were substantially expressed in embryo, shoot, ovary, and 5d-seed, respectively. In addition, we examined their expression pattern during different stages of anther and pollen development, and found that at least 6 targets were highly abundant during the whole pollen developmental stages (Supplementary Fig. S1).

The OryGenesDB database was searched to obtain more information about the biological function of putative positively selected target genes. Inserted mutants of all 12 targets were found in different mutant databases (Supplementary Table S1), of which two target genes (*LOC_Os02g04680.1* and *LOC_Os03g50620.1*) with observed phenotypes were recorded in the mutant databases. The mutant of *OsSPL3* (*LOC_Os02g04680.1*) that belongs to the SBP-box gene family displayed abnormal phenotypes, including dwarf, low fertility, and lesion mimic, etc., while yellow leaf is the only record for the ATP-binding protein-coding gene *LOC_Os03g50620.1* mutant so far (Supplementary Table S1).

Discussion

Recently, amounts of domestication-related chromosomal regions with tens to hundreds Kb in length have been identified, which comprises of protein-coding genes and non-coding sequences including some miRNAs (Huang et al. 2012; Xu et al. 2012; Guo et al. 2013; Jia et al. 2013; Liu et al. 2013; Qi et al. 2013). Using the combinations of deep sequencing, evolutionary biology and population genetics approaches, hundreds of protein-coding genes are supposed to be rice domestication related, although their biological functions need to be validated experimentally (He et al. 2011; Huang et al. 2012; Xu et al. 2012; Yang et al. 2012; Yonemaru et al. 2012). It seems that the numbers of domestication-related genes are far more than what we previously expected. Thus, given their crucial regulatory functions, it is intriguing to uncover whether miRNAs were under natural and/or artificial selection during rice domestication. In human, a number of miRNA-containing regions were found to have undergone events of positive selection (Quach et al. 2009). In this study, significant negative Tajima's *D* values for 190 miRNA loci were detected (Supplementary Table S2), of which some miRNAs, including *MIR156b*/*c* (Wang et al. 2007), *MIR395a*/*b*, *MIR399d* (Wang et al. 2010), *MIR5513*, *MIR818e*, *MIR5158*, *MIR1847*, *MIR1865*, *MIR160f*, and *MIR5143* (Liu et al. 2013), had been identified to be either under positive selection or putative domestication-related candidates in previous studies. However, to reduce false positives, only 21 out of the 190 miRNA loci were retained in this study. In addition, as discussed by Wang et al. (2012), expression of miRNAs could be a target of domestication. We further compared the 21 miRNAs with those identified by Wang et al. (2012) and Zhu et al. (2008), and found that there were 8 (*MIR1872*, *MIR395c*, *MIR1846c*, *MIR1881*, *MIR815a*, *MIR160f*, *MIR1436*, *MIR812f*) and 11 (*MIR156b*, *MIR156c*, *MIR156i*, *MIR160f*, *MIR167i*, *MIR1846c*, *MIR1881*, *MIR1883b*, *MIR395d*, *MIR812f*, and *MIR815a*) miRNAs being differentially expressed in cultivated and wild rice, respectively. These analyses strongly support the speculation of selection in miRNAs and their importance for driving rice evolution (Wang et al. 2010, 2012).

It is generally thought that cultivated rice has experienced a severe genetic bottleneck, and their sequence diversity reduced dramatically during domestication compared to their wild progenitors. Zhu et al. (2007) found that in cultivated rice, four neutral genes (*Adh1*, *Waxy*, *ks1* and *RGRC2*) have lost <2.6-folds of their sequence diversity (π) compared with that in wild rice. In this study, we found that with the exception of *MIR395c*/*d*, other miRNAs have sequence diversity (π) reduction >2.6-folds (from 3.55 to 51.7 folds) in domesticated rice relative to the wild rice population. Compared with the average π (0.0024) of the whole genome of *O. sativa* (Huang et al. 2010), a ~7.3-fold reduction of sequence diversity was observed in the 21 ps-miRNAs. In particular, the average nucleotide diversity of ps-miRNAs was more than 17.6-fold lower than that of 111 randomly selected rice STS sequences (Caicedo et al. 2007), with 6 ps-miRNAs (*MIR167i*, *MIR815a*, *MIR1320*, *MIR160f*, *MIR5154*, and *MIR1436*) having 16.6–75.3-folds reduction. For comparison, moreover, we completely randomly chose 631 and 614 protein-coding gene fragments and intergenic regions from 462 *O. sativa* genomes. Even though sequence regions being under potential positive

Table 2 Summary of nucleotide polymorphisms and neutrality tests at 12 target gene loci

MIRNA	Target gene	*O. rufipogon*				*O. sativa*							Target annotation
		S	π	θ	*D*	*S*	π	θ	*D*	CS*p*	*D**	*F**	
miR156b/c/i	*Os02g07780.1*	79	0.0019	0.0029	−1.038	33	0.0004	0.0012	−1.754*	0.002	−1.303	−1.820	OsSPL4—SBP-box gene family member, expressed
	Os08g39890.1	71	0.0020	0.0025	−0.627	26	0.0002	0.0009	−2.019*	0.000	−0.550	−1.424	OsSPL14—SBP-box gene family member, expressed
	Os02g04680.1	75	0.0011	0.0023	−1.536	22	0.0002	0.0007	−1.940*	0.000	−0.931	−1.637	OsSPL3—SBP-box gene family member, expressed
miR156c	*Os07g40390.1*	43	0.0016	0.0032	−1.370	13	0.0002	0.0010	−1.763*	0.000	−0.044	−0.856	Expressed protein
miR160f	*Os04g59430.1*	65	0.0022	0.0027	−0.491	26	0.0003	0.0011	−1.818*	0.000	−2.513*	−2.691*	Auxin response factor, putative, expressed
miR821b/c	*Os12g16350.1*	298	0.0021	0.0033	−1.123	154	0.0004	0.0017	−2.232**	0.000	−2.402*	−2.745*	Enoyl-CoA hydratase/isomerase family protein, putative, expressed
miR812f	*Os02g07960.4*	153	0.0012	0.0023	−1.399	50	0.0001	0.0007	−2.448***	0.000	−1.146	−2.114	STRUBBELIG-RECEPTOR FAMILY 3 precursor, putative, expressed
	Os03g50620.1	122	0.0026	0.0031	−0.444	62	0.0003	0.0016	−2.315**	0.000	−1.327	−2.167	ATP-binding protein, putative, expressed
	Os02g07960.3	150	0.0012	0.0023	−1.397	48	0.0001	0.0007	−2.436***	0.000	−0.944	−1.974	STRUBBELIG-RECEPTOR FAMILY 3 precursor, putative, expressed
	Os02g07960.2	151	0.0012	0.0023	−1.396	49	0.0001	0.0007	−2.444***	0.000	−0.882	−1.940	STRUBBELIG-RECEPTOR FAMILY 3 precursor, putative, expressed
miR1436	*Os03g63370.1*	73	0.0017	0.0023	−0.760	38	0.0003	0.0012	−2.038*	0.000	−3.229*	−3.264**	CRAL/TRIO domain-containing protein, expressed
	Os12g16350.13	157	0.0020	0.0031	−1.038	78	0.0005	0.0016	−2.014*	0.000	−3.174*	−3.127**	Enoyl-CoA hydratase/isomerase family protein, putative, expressed

S number of segregating sites

π average number of nucleotide differences per site between two sequences (Nei 1987) calculated on the total number of polymorphic sites

θ Watterson's estimator (Watterson 1975)

D Tajima's *D* (Tajima 1989)

*D** and *F** Fu and Li's *D** and *F** (Fu and Li 1993)

CS*p* Coalescent simulations for Tajima's *D*

*$P < 0.05$, **$P < 0.02$, and ***$P < 0.001$

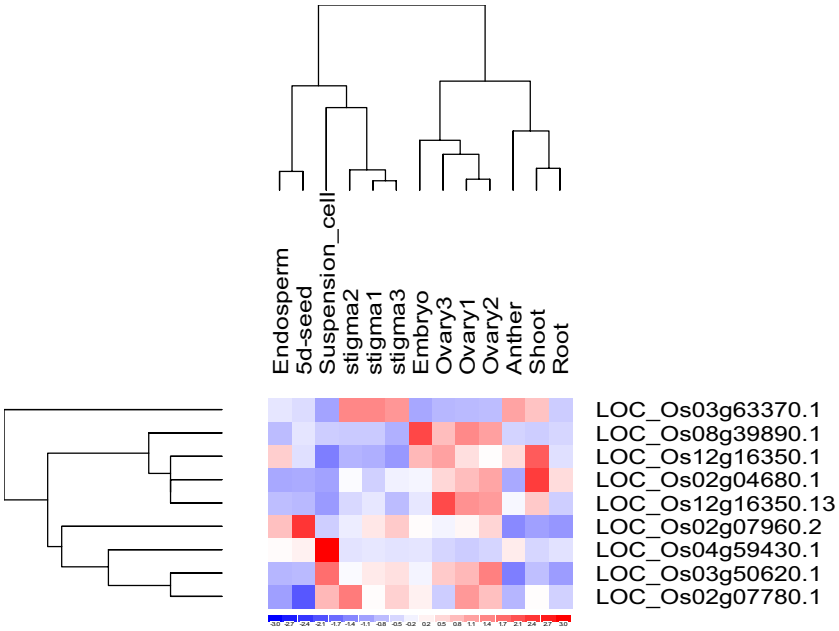

Fig. 1 Expression profiles of target genes of several miRNA loci that are under potential positive selection in nine rice tissues. The data were from the published micro-array dataset GSE7951 (Li et al. 2007)

or negative selection are probably included because of the complete randomness of sequences collection, the mean nucleotide diversity of the two samples is remarkably higher than that in 21 ps-miRNAs. These comparisons suggest that the low sequence diversity observed in most of the 21 ps-miRNAs cannot be solely explained by a population bottleneck in that this phenomenon would cause a reduction in sequence diversity throughout the genome (Asano et al. 2011).

Modern Asian cultivated rice (*O. sativa*) was domesticated from common wild rice, *O. rufipogon* (Kovach et al. 2007), but the geographical range of modern rice is far beyond its wild ancestor (Khush 1997). For better adaptation to diverse environments, like in soybean (Li et al. 2014), selection must have imposed on a set of protein-coding genes and regulatory elements in cultivated rice and its wild relatives. In consistent with this speculation, herein, 9 out of 21 ps-miRNAs were also detected to have undergone potential positive selection in common wild rice (Table 1). Interestingly, He et al. (2011) identified four LDRs with length longer than 200 Kb in *O. rufipogon*. Because LDRs are a possible signature of selective sweep (He et al. 2011), it seems that stronger selection should have occurred in common wild rice too, although both the selection strength and wideness are not as common as that in cultivated rice (He et al. 2011). On the other hand, Allaby et al. (2008) argued that before wild crops were domesticated to the modern cultivated crops, they were domesticated to predomestication crops first. It was estimated that it took about 3,000 years from the time of initial cultivation (~12,000 years ago) until the appearance of modern domesticated rice (Zhao 1998; Lu et al. 2002). It can be inferred that during predomestication of the progenitor rice, some sequence regions including non-coding sequences should have endured stronger selection, and some of these functional regions were subsequently retained in the rice genome. Thus, these detected miRNAs being positively selected in both wild and cultivated rice should have played crucial roles during rice evolution.

One of the mechanisms for the origination and expansion of miRNAs in plants is by duplication of preexisting miRNA genes with subsequent mutations (Nozawa et al. 2012). Whole genome duplication (WGD) should be involved in this process (Zhang et al. 2009; Sun et al. 2012). As reported, the rice genome has undergone two rounds of WGDs during evolution (Guyot and Keller 2004). We searched against the segmentally duplicated chromosomal regions identified previously in rice, and found that 4 out of the 21 ps-miRNAs (*MIR156i*, *MIR167i*, *MIR821c*, *MIR1846c*) fell into the chromosomal regions that were supposed to have undergone large-scale segmental duplications (Supplementary Fig. S2). Our observation is consistent with Wang et al. (2007), where *MIR156i* and *MIR156e* are paralogs that were produced by segmental duplication which occurred between rice chromosomes 2 and 4 (Wang et al. 2007). Thus, whole genome duplication had a role in the evolution of miRNAs with signatures of positive selection in rice.

During the past decades, a few of protein-coding genes associated with rice domestication have been identified and characterized in rice (Izawa et al. 2009; Jiao et al. 2010; Asano et al. 2011; Sang and Ge 2013; Zhu et al. 2013). One interesting result in this study is that the target genes of several ps-miRNAs are also

potentially positively selected during rice evolution. As known, plant architecture and yield are important traits for rice domestication and improvement. Jiao et al. (2010) reported that a point mutation in the SBP-box-containing transcription factor *OsSPL14* that perturbs the normal miR156-*OsSPL14* interaction, leads to improved plant architecture in rice. Coincidently, evidence for stronger positive selection was found to potentially act on *MIR156i* and its target gene *OsSPL14* in this study. Another typical example is the *MIR397a* gene that targets laccase genes in rice and poplars. The signature of stronger positive selection was detected in the *MIR397a*-containing sequence region. However, we cannot arbitrarily exclude the possibility that this signature is caused by selection on its down-stream protein-coding gene *LOC_Os06g46970*, because of their closeness in physical distance. Although *MIR397a* was not considered as ps-miRNAs in this study, its biological functions are reported to be very important for plants growth and development. It was established that in poplars, ptr-miR397a acts as a negative regulator for lignin biosynthesis (Lu et al. 2013), while over-expression of miR397a/b in rice has enlarged grain size, and promoted panicle branching, and thereby leading to an increase of overall grain yield (Zhang et al. 2013). In addition, 6 out of the 21 ps-miRNAs (*MIR1436*, *MIR156b/c/i*, *MIR160f*, and *MIR167i*) are identified to be related to flowering, an important domestication trait, in common wild rice (Chen et al. 2013). It appears that miRNA genes including *MIR397a* may serve as one of the driving evolutionary forces for rice domestication (Wang et al. 2012). However, Wang et al. (2010) argued that positive selection signatures detected on some miRNA loci might have arisen by selective sweeps rather than direct artificial selection. Whether artificial and/or natural selection imposed directly on miRNA genes, needs to be further validated and confirmed by phenotypic experiments in the near future.

In conclusion, herein, 21 out of 541 miRNA loci have been characterized to be under strong selection in cultivated rice, whose nucleotide diversity is only 13.8 % of that in their wild progenitors. The selection signatures detected in ps-miRNAs would result from either selective sweep or direct selection. In particular, 12 out of 68 target genes of ps-miRNAs are also found to be non-neutrally evolved, which always exhibit tissue-specific expression pattern and may play crucial roles in the processes of plants reproduction. The biological functions of some ps-miRNAs and their specific targets are being investigated accordingly.

Acknowledgments This work was financially supported by grants from the National Natural Science Foundation of China (31000170, 31471431), Zhejiang Provincial Natural Science Foundation of China (LY14B070015, Y3090152), and the Training Program for Outstanding Young Talents of Zhejiang A&F University.

References

Allaby RG, Fuller DQ, Brown TA (2008) The genetic expectations of a protracted model for the origins of domesticated crops. Proc Natl Acad Sci USA 105:13982–13986

Asano K, Yamasaki M, Takuno S, Miura K, Katagiri S, Ito T, Doi K, Wu J, Ebana K, Matsumoto T, Innan H, Kitano H, Ashikari M, Matsuoka M (2011) Artificial selection for a green revolution gene during japonica rice domestication. Proc Natl Acad Sci USA 108:11034–11039

Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, Bustamante CD, Purugganan MD (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet 3:e163

Chen XM (2005) MicroRNA biogenesis and functions in plants. FEBS Lett 579:5923–5931

Chen Z, Li F, Yang S, Dong Y, Yuan Q, Wang F, Li W, Jiang Y, Jia S, Pei X (2013) Identification and functional analysis of flowering related microRNAs in common wild rice (*Oryza rufipogon* Griff.). PLoS ONE 8:e82844

Cui X, Wang Q, Yin W, Xu H, Wilson ZA, Wei C, Pan S, Zhang D (2012) PMRD: a curated database for genes and mutants involved in plant male reproduction. BMC Plant Biol 12:215

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 Res 39:W155–W159

de Meaux J, Hu JY, Tartler U, Goebel U (2008) Structurally different alleles of the ath-MIR824 microRNA precursor are maintained at high frequency in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 105:8994–8999

Droc G, Périn C, Fromentin S, Larmande P (2009) OryGenesDB 2008 update: database interoperability for functional genomics of rice. Nucleic Acids Res 37:D992–D995

Ehrenreich IM, Purugganan MD (2008) Sequence variation of microRNAs and their binding sites in *Arabidopsis*. Plant Physiol 146:1974–1982

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gielen H, Remans T, Vangronsveld J, Cuypers A (2012) MicroRNAs in metal stress: specific roles or secondary responses? Int J Mol Sci 13:15826–15847

Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, Sun J, Guo AY (2012) Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat 33:254–263

Guo X, Gui Y, Wang Y, Zhu QH, Helliwell C, Fan L (2008) Selection and mutation on microRNA target sequences during rice evolution. BMC Genom 9:454

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, Min J, Guo X, Murat F, Ham BK, Zhang Z, Gao S, Huang M, Xu Y, Zhong S, Bombarely A, Mueller LA, Zhao H, He H, Zhang Y, Zhang Z, Huang S, Tan T, Pang E, Lin K, Hu Q, Kuang H, Ni P, Wang B, Liu J, Kou Q, Hou W, Zou X, Jiang J, Gong G, Klee K, Schoof H, Huang Y, Hu X, Dong S, Liang D, Wang J, Wu K, Xia Y, Zhao X, Zheng Z, Xing M, Liang X, Huang B, Lv T, Wang J, Yin Y, Yi H, Li R, Wu M, Levi A,

Zhang X, Giovannoni JJ, Wang J, Li Y, Fei Z, Xu Y (2013) The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45:51–58

Guyot R, Keller B (2004) Ancestral genome duplication in rice. Genome 47:610–614

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu CI, Shi S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7:e1002100

Henry RJ (2012) Next-generation sequencing for understanding and accelerating crop domestication. Brief Funct Genomics 11:51–56

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967

Huang X, Kurata N, Wei X, Wang Z, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W, Guo Y, Lu Y, Zhou C, Fan D, Weng Q, Zhu C, Huang T, Zhang L, Wang Y, Feng L, Furuumi H, Kubo T, Miyabayashi T, Yuan X, Xu Q, Dong G, Zhan Q, Li C, Fujiyama A, Toyoda A, Lu T, Feng Q, Qian Q, Li J, Han B (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501

Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T (2011) PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. Nucleic Acids Res 39:D556–D560

Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM, Lai J, Morrell PL, Shannon LM, Song C, Springer NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, Ross-Ibarra J (2012) Comparative population genomics of maize domestication and improvement. Nat Genet 44:808–811

Izawa T, Konishi S, Shomura A, Yano M (2009) DNA changes tell us about rice domestication. Curr Opin Plant Biol 12:185–192

Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H, Zhu C, Lu Y, Zhou C, Fan D, Weng Q, Guo Y, Huang T, Zhang L, Lu T, Feng Q, Hao H, Liu H, Lu P, Zhang N, Li Y, Guo E, Wang S, Wang S, Liu J, Zhang W, Chen G, Zhang B, Li W, Wang Y, Li H, Zhao B, Li J, Diao X, Han B (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). Nat Genet 45:957–961

Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. Nat Genet 42:541–544

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:25–34

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73

Li C (2008) Automating dChip: toward reproducible sharing of microarray data analysis. BMC Bioinformatics 9:231

Li M, Xu W, Yang W, Kong Z, Xue Y (2007) Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice. Plant Physiol 144:1797–1812

Li Y, Zhou G, Ma J, Jiang W, Jin L, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S, Zuo Q, Shi X, Li Y, Zhang W, Hu Y, Kong G, Hong H, Tan B, Song J, Liu Z, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Zhang L, Guan R, Wang K, Li W, Chen S, Chang R, Jiang Z, Jackson SA, Li R, Qiu L (2014) De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. Nat Biotechnol. doi:10.1038/nbt.2979

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Liu Q, Wang H, Zhu L, Hu H, Sun Y (2013) Genome-wide identification and analysis of miRNA-related single nucleotide polymorphisms (SNPs) in rice. Rice 6:10

Lu HY, Liu ZX, Wu NQ, Berné S, Saito Y, Liu B, Wang L (2002) Rice domestication and climatic change: phytolith evidence from East China. Boreas 31:378–385

Lu S, Li Q, Wei H, Chang MJ, Tunlaya-Anukit S, Kim H, Liu J, Song J, Sun YH, Yuan L, Yeh TF, Peszlen I, Ralph J, Sederoff RR, Chiang VL (2013) Ptr-miR397a is a negative regulator of laccase genes affecting lignin content in *Populus trichocarpa*. Proc Natl Acad Sci USA 110:10848–10853

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nozawa M, Miura S, Nei M (2012) Origins and evolution of microRNA genes in plant species. Genome Biol Evol 4:230–239

Pashkovskiy PP, Ryazansky SS (2013) Biogenesis, evolution, and functions of plant microRNAs. Biochemistry 78:627–637

Pryszcz LP, Huerta-Cepas J, Gabaldón T (2011) MetaPhOrs: orthology and paralogy predictions from multiple phylogenetics evidence using a consistency-based confidence score. Nucleic Acids Res 39:e32

Qi J, Liu X, Shen D, Miao H, Xie B, Li X, Zeng P, Wang S, Shang Y, Gu X, Du Y, Li Y, Lin T, Yuan J, Yang X, Chen J, Chen H, Xiong X, Huang K, Fei Z, Mao L, Tian L, Städler T, Renner SS, Kamoun S, Lucas WJ, Zhang Z, Huang S (2013) A genome variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet 45:1510–1515

Quach H, Barreiro LB, Laval G, Zidane N, Patin E, Kidd KK, Kidd JR, Bouchier C, Veuille M, Antoniewski C, Quintana-Murci L (2009) Signatures of purifying and local positive selection in human miRNAs. Am J Hum Genet 84:316–327

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionary fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev 20:3407–3425

Sang T, Ge S (2013) Understanding rice domestication and implications for cultivar improvement. Curr Opin Plant Biol 16:139–146

Sun J, Zhou M, Mao Z, Li C (2012) Characterization and evolution of microRNA genes derived from repetitive elements and duplication events in plants. PLoS ONE 7:e34092

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Todesco M, Balasubramanian S, Cao J, Ott F, Sureshkumar S, Schneeberger K, Meyer RC, Altmann T, Weigel D (2012) Natural variation in biogenesis efficiency of individual *Arabidopsis thaliana* microRNAs. Curr Biol 22:166–170

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 Lett 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 Evol Biol 10:119

Wang Y, Bai X, Yan C, Gui Y, Wei X, Zhu QH, Guo L, Fan L (2012) Genomic dissection of small RNAs in wild rice (*Oryza rufipogon*): lessons for rice domestication. New Phytol 196:914–925

Watterson GA (1975) On the number of segregating sites in genetical models without recombinations. Theor Popul Biol 7:256–276

Wei FJ, Droc G, Guiderdoni E, Hsing YC (2013) International consortium of rice mutagenesis: resources and beyond. Rice 6:39

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30:105–111

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell 17:2859–2872

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yang C, Kawahara Y, Mizuno H, Wu J, Matsumoto T, Itoh T (2012) Independent domestication of Asian rice followed by gene flow from japonica to indica. Mol Biol Evol 29:1471–1479

Yonemaru J, Yamamoto T, Ebana K, Yamamoto E, Nagasaki H, Shibaya T, Yano M (2012) Genome-wide haplotype changes produced by artificial selection during modern rice breeding in Japan. PLoS ONE 7:e32982

Zhang L, Chia JM, Kumari S, Stein JC, Liu Z, Narechania A, Maher CA, Guill K, McMullen MD, Ware D (2009) A genome-wide characterization of microRNA genes in maize. PLoS Genet 5:e1000716

Zhang YC, Yu Y, Wang CY, Li ZY, Liu Q, Xu J, Liao JY, Wang XJ, Qu LH, Chen F, Xin P, Yan C, Chu J, Li HQ, Chen YQ (2013) Overexpression of microRNA OsmiR397 improves rice yield by increasing grain size and promoting panicle branching. Nat Biotechnol 31:848–852

Zhao ZJ (1998) The middle Yangtze region in China is one place where rice was domesticated: phytolith evidence from the Diaotonghuan cave, northern Jiangxi. Antiquity 72:885–897

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. Mol Biol Evol 24:875–888

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 Res 18:1456–1465

Zhu Z, Tan L, Fu Y, Liu F, Cai H, Xie D, Wu F, Wu J, Matsumoto T, Sun C (2013) Genetic control of inflorescence architecture during rice domestication. Nat Commun 4:2200

Zhuo Y, Gao G, Shi J, Zhou X, Wang X (2013) miRNAs: biogenesis, origin and evolution, functions on virus-host interaction. Cell Physiol Biochem 32:499–510