Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) 2021 22(5):366-382
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2000519

Genome-wide analysis of microRNA156 and its targets, the genes encoding SQUAMOSA promoter-binding protein-like (SPL) transcription factors, in the grass family Poaceae

Erkui YUE[1], Hua TAO[2], Jianhong XU[1]✉

[1]*Institute of Crop Science, Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Hangzhou 310058, China*
[2]*Henan Agricultural Radio and Television School, Zhengzhou 450008, China*

Abstract: MicroRNAs (miRNAs) are endogenous small non-coding RNAs that play an important role in post-transcriptional gene regulation in plants and animals by targeting messenger RNAs (mRNAs) for cleavage or repressing translation of specific mRNAs. The first miRNA identified in plants, miRNA156 (miR156), targets the SQUAMOSA promoter-binding protein-like (SPL) transcription factors, which play critical roles in plant phase transition, flower and plant architecture, and fruit development. We identified multiple copies of *MIR156* and *SPL* in the rice, *Brachypodium*, sorghum, maize, and foxtail millet genomes. Sequence and chromosomal synteny analysis showed that both *MIR156*s and *SPL*s are conserved across species in the grass family. Analysis of expression data of the *SPL*s in eleven juvenile and adult rice tissues revealed that four non-miR156-targeted genes were highly expressed and three miR156-targeted genes were only slightly expressed in all tissues/developmental stages. The remaining *SPL*s were highly expressed in the juvenile stage, but their expression was lower in the adult stage. It has been proposed that under strong selective pressure, non-miR156-targeted mRNA may be able to re-structure to form a miRNA-responsive element. In our analysis, some non-miR156-targeted *SPL*s (*SPL5*/*8*/*10*) had gene structure and gene expression patterns similar to those of miR156-targeted genes, suggesting that they could diversify into miR156-targeted genes. DNA methylation profiles of *SPL*s and *MIR156*s in different rice tissues showed diverse methylation patterns, and hypomethylation of non-CG sites was observed in rice endosperm. Our findings suggested that *MIR156*s and *SPL*s had different origination and evolutionary mechanisms: the *SPL*s appear to have resulted from vertical evolution, whereas *MIR156*s appear to have resulted from strong evolutionary selection on mature sequences.

Key words: MicroRNA156 (miR156); SQUAMOSA promoter-binding protein-like (*SPL*) gene; DNA methylation; Gene expression; Grass genome

1 Introduction

MicroRNAs (miRNAs) are short non-coding RNAs that range in length from 20 to 24 nucleotides (nt) and are derived from single-stranded precursors with stable hairpin (HP) structures (Bartel, 2004; Voinnet, 2009). In plants, miRNAs are highly complementary to their targets, and their binding can cause cleavage of the target (Reinhart et al., 2002). miRNAs have been identified in nearly all eukaryotes and act in concert with transcription factors to achieve a dynamic equilibrium of gene expression and to allow adaptation to environmental stressors (Zhang et al., 2006; Martin et al., 2010; Sollome et al., 2016).

miRNA156 (miR156) was originally identified in *Arabidopsis thaliana* because of its striking similarity to the corresponding animal miRNA (Reinhart et al., 2002). miR156 is relatively conserved in higher plants (Willmann and Poethig, 2007) and regulates the expression of more than half of the SQUAMOSA promoter-binding protein-like (*SPL*) genes in rice (*Oryza sativa*) and Arabidopsis (Xie et al., 2006; Guo et al., 2008; Miao et al., 2019). SPLs are plant-specific transcription factors that are characterized by a highly conserved SQUAMOSA promoter-binding protein (SBP) domain which binds to a specific *cis*-element glycidyltrimethylammonium chloride (GTAC)-binding

✉ Jianhong XU, jhxu@zju.edu.cn
Jianhong XU, https://orcid.org/0000-0002-5620-0410

Received Sept. 1, 2020; Revision accepted Nov. 2, 2020;
Crosschecked Mar. 5, 2021

domain (Birkenbihl et al., 2005; Kropat et al., 2005). They have essential roles in plant phase transition (Fornara and Coupland, 2009; Huijser and Schmid, 2011), flower and fruit development (Manning et al., 2006; Lal et al., 2011; Yang et al., 2019), plant architecture (Jiao et al., 2010; Miura et al., 2010), trichome distribution (Yu et al., 2010), anthocyanin biosynthesis (Gou et al., 2011), gibberellin signaling (Zhang et al., 2007), shade avoidance (Xie et al., 2017), antiviral and disease defense (Wang et al., 2018; Liu et al., 2019; Yao et al., 2019), and the copper response (Kropat et al., 2005; Gielen et al., 2016; Yan et al., 2017; Ramamurthy et al., 2018). The 17 and 19 *SPL*s identified in Arabidopsis and rice, respectively, fall into several subgroups according to the amino acid sequences of their SBP domains and the exon-intron structure of their encoding genes (Guo et al., 2005; Xie et al., 2006).

Although *SPL* genes have been systematically studied in Arabidopsis and rice, there have been few studies in other grass species. The grass family Poaceae is a large and ubiquitous family of monocots, which includes species that are considered to be the most important crops for human nutrition and economies. With the advent of large-scale sequencing technologies, whole-genome sequences are now available for rice (International Rice Genome Sequencing Project and Sasaki, 2005), *Brachypodium distachyon* (The International Brachypodium Initiative, 2010), foxtail millet (*Setaria italica*) (Bennetzen et al., 2012; Zhang et al., 2012), sorghum (*Sorghum bicolor*) (Paterson et al., 2009), and maize (*Zea mays*) (Schnable et al., 2009), which we used to analyze the copy number, genomic location, and synteny of *MIR156*s and *SPL*s in the grass family.

One possible alternative regulatory mechanism may be epigenomic modifications, including DNA methylation and histone modifications (Feng et al., 2010; Xu et al., 2020). In plants, DNA methylation tends to occur at cytosine bases in the context of CG, CHG, and CHH (H refers to A, T, or C) (Law and Jacobsen, 2010; Teixeira and Colot, 2010). In this study, we analyzed available methylation data for *SPL*s and *MIR156*s in four rice tissues. Our systematic analysis of *MIR156*s and *SPL*s in grasses provides insight into the conservation and evolution of these genes in plants and the complementary roles of miRNAs and DNA methylation in regulating gene expression.

2 Materials and methods

2.1 Sequence retrieval

The following databases were used to retrieve sequences for Arabidopsis (TAIR 9; http://www.arabidopsis.org), rice (http://rice.plantbiology.msu.edu (release 7) and http://www.phytozome.net/rice), Brachypodium (http://www.brachypodium.org and http://www.phytozome.net/brachypodium (JGI v1.2)), foxtail millet (http://foxtailmillet.genomics.org.cn/page/species/index.jsp and http://www.phytozome.net/foxtailmillet (JGI v2.1)), sorghum (http://www.phytozome.net/sorghum (JGI v1.0)), and maize (http://www.maizegdb.org, http://www.maizesequence.org (AGP v2), and http://www.phytozome.net/maize).

2.2 Identification of miR156 members in grass genomes

Precursor and mature miR156 sequences from grasses were retrieved from the miRBase database (http://www.mirbase.org (release 22.1)) (Kozomara and Griffiths-Jones, 2011). The reverse-complement of RNA sequences was determined by a custom Perl script when necessary. To identify all putative miR156s in grass genomes, BlastN with an *E*-value cutoff of 1×10^{-3} was conducted. The upstream 20 nt and downstream 200 nt of the putative miR156 genes were also extracted for secondary structure analysis. The HP structure was computationally predicted with the Mfold package (http://www.bioinfo.rpi.edu/applications/mfold) with manual modifications (Zuker, 2003). Potential miR156s that originated from intergenic and intronic loci were retained for further analysis. For the prediction of new miR156s, filtering criteria were similar to those of Meyers et al. (2008), as follows: (1) mature sequences of miR156 reside in the stem arm of the stem-loop structure; (2) no more than five unpaired nucleotides between the miRNA and miRNA*; (3) no more than three consecutive unpaired nucleotides; (4) folding free energy should be lower than −25 kcal/mol; and (5) homologs of miR156s should be contained in related monocot plants.

2.3 Identification of *SPL*s in grass genomes

Nineteen *O. sativa SPL* (*OsSPL*) sequences were used as references to search for related orthologous *SPL* copies in *Brachypodium*, foxtail millet, sorghum, and maize (Xie et al., 2006). BlastX with an *E*-value cutoff of 1×10^{-3} was performed against respective

grass genomes to locate the *SPL*s, and was combined with an HMMER (http://hmmer.janelia.org) search (Pfam: PF03110) to find all possible SBPs. Search results were manually modified to reduce the number of false positives.

2.4 miR156 target prediction

The potential targets of miR156 were predicted by the psRNATarget (http://plantgrn.noble.org/psRNATarget (Dai and Zhao, 2011), and TAPIR (http://bioinformatics.psb.ugent.be/webtools/tapir (Bonnet et al., 2010)) programs with default parameters. Unique mature miR156 sequences were chosen to predict their targets in rice, *Brachypodium*, sorghum, maize, and foxtail millet. To gain high-confidence predictions, no more than three mismatches were allowed between miR156 and its targets. A G: U wobble was regarded as a 0.5 mismatch, and no mismatches were allowed at positions 10 and 11 (from the 5' end of the sequence), which are the canonical cleavage sites in plants (Franco-Zorrilla et al., 2007).

2.5 Phylogenetic analysis of *SPL*s and *MIR156*s

Because neither the gene sequences nor the protein sequences of SPLs could be aligned exactly, only the amino acid sequences of the conserved SBP domains were used for phylogenetic analysis. SBP sequences (119) were extracted from Arabidopsis, rice, *Brachypodium*, foxtail millet, sorghum, and maize, and were aligned with the MUSCLE algorithm with manual modifications (Edgar, 2004). Evolutionary histories were inferred using the neighbor-joining method. A bootstrap test with 1000 replicates was used to build the tree, and the evolutionary distances were computed using the Poisson correction method and were expressed as the number of amino acid substitutions per site in MEGA X (https://www.megasoftware.net) (Kumar et al., 2018).

To construct the phylogenetic tree of *MIR156*s, stem-loop sequences were aligned by Clustal W with manual modifications (Larkin et al., 2007). MEGA X (Kumar et al., 2018) was then used to construct a maximum-likelihood phylogenetic tree based on the Kimura-2-parameter model and a bootstrap test with 1000 replicates (Kimura, 1980).

2.6 Expression profile of *SPL*s in rice

To obtain the hierarchical clustering of *OsSPL* expression profiles among diverse tissues, expression data for *OsSPL*s were obtained from the Gene Expression Omnibus (GEO) (Barrett et al., 2011) database under accession numbers GSE19602 and GSE19824, and RNAseq data were obtained from the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/index.shtml). The data were transformed into a matrix to satisfy the condition of MultiExperiment Viewer (http://mev.tm4.org) (Saeed et al., 2006) and $\log_2$-transformed for better visualization. Hierarchical clustering of the genes was assessed using a Euclidean distance metric in MultiExperiment Viewer.

2.7 DNA methylation patterns of *SPL*s and *MIR156*s in rice

To examine methylation of *MIR156*s and *OsSPL*s in rice, we retrieved sequences covering a region 2 kb upstream from the start codon, the gene body, and the 3'-untranslated region (3'-UTR) of *OsSPL*s, and about 900 bp upstream and downstream of *O. sativa-MIR156* (*osa-MIR156*) sequences (2 kb in total including the stem-loop sequence) from genome databases. Bisulfite sequencing data were downloaded from GEO under accession number GSE22591. DNA methylation patterns of *OsSPL*s and *osa-MIR156*s were explored according to Zemach et al. (2010) with minor modifications. Briefly, (1) extracted sequences were converted to FASTQ format; (2) low-quality reads were removed; (3) adaptors were trimmed, and sequences of <18 nt were discarded; (4) custom Perl scripts were used to convert all Cs to Ts and Gs to As for subsequent alignment with *OsSPL*s using Bowtie (Langmead et al., 2009), allowing up to two mismatches; (5) duplicate simple reads were filtered to single reads; and (6) methylcytosines in the context of CG, CHG, and CHH were counted to calculate the methylation rate of *OsSPL*s and *osa-MIR156*s in the four tissues.

3 Results

3.1 Identification and analysis of miR156 in the grass family

The miRBase database includes 12 miR156 sequences from rice, 10 from *Brachypodium*, 9 from sorghum, and 12 from maize (Table 1). These miR156 precursor sequences were used to identify ten in foxtail millet (Table 1 and Fig. S1). The chromosomal distribution of miR156 orthologs is shown in Table S1.

Table 1 Orthologous *MIR156* copies identified in rice, *Brachypodium*, foxtail millet, sorghum, and maize

Rice	*Brachypodium*	Foxtail millet	Sorghum	Maize 1	Maize 2
osa-MIR156a		***sit-MIR156a***			
osa-MIR156b	*bdi-MIR156e*	***sit-MIR156b***	*sbi-MIR156b*	*zma-MIR156b*	
osa-MIR156c	*bdi-MIR156g*	***sit-MIR156c***	*sbi-MIR156c*	*zma-MIR156c*	
osa-MIR156d	*bdi-MIR156c*	***sit-MIR156d***	*sbi-MIR156a*	*zma-MIR156d*	
osa-MIR156e	*bdi-MIR156d*	***sit-MIR156e***	*sbi-MIR156h*	*zma-MIR156h*	*zma-MIR156e*
osa-MIR156f	*bdi-MIR156b*	***sit-MIR156f***	*sbi-MIR156i*	*zma-MIR156a*	
osa-MIR156g	*bdi-MIR156i*	***sit-MIR156g***	*sbi-MIR156f*	*zma-MIR156g*	
osa-MIR156h_j			*sbi-MIR156e*	*zma-MIR156k*	
osa-MIR156i	*bdi-MIR156h*	***sit-MIR156i***	*sbi-MIR156g*	*zma-MIR156l*	*zma-MIR156i*
osa-MIR156k	*bdi-MIR156j*	***sit-MIR156h***	*sbi-MIR156d*	*zma-MIR156j*	
osa-MIR156l					
		sit-MIR156j			
				zma-MIR156f	

The orthologous *MIR156* gene copies are listed within a single row. The names of previously identified *MIR156*s are those used in miRBase, and the newly identified *MIR156*s were named according to the syntenic analysis with bold fonts. As maize is allotetraploid, the two homeologous chromosomes were named Maize 1 and Maize 2.

Only one of the two identical tandem copies of *osa-MIR156f* was used for further analysis. In addition, one of the *Brachypodium* miR156 sequences, *bdi-MIR156a* (MI0011566), appeared to be misannotated in miRBase since it does not contain the canonical precursor stem-loop structure and appears to be a fragment of *Bradi4g33770* (an *SPL*). The folding free energy of all the newly identified miR156 members was lower than −50 kcal/mol, and they were all located in intergenic regions, both typical features of miRNAs in plants (Nozawa et al., 2012).

Alignment of the *MIR156* orthologous regions (130−1500 kb) from rice, *Brachypodium*, foxtail millet, sorghum, and maize revealed that all but *osa-MIR156l*, *sit-MIR156j*, and *zma-MIR156f* are located within syntenic regions (Figs. 1 and S2). We found that the rice locus *osa-MIR156a* has a corresponding copy only in foxtail millet, not in *Brachypodium*, sorghum, or maize; and *osa-MIR156h-j* has orthologous copies in sorghum and maize, but not in *Brachypodium* or foxtail millet (Figs. 1 and S2).

Alignment of all 52 miR156 precursor sequences from the five grass species showed that their sequences were not similar outside of the stem-loop sequence. There was little conservation of precursor sequences, but the mature miR156 sequences were extremely well conserved. Therefore, phylogenetic analysis was performed using only the stem-loop sequences. All miR156 genes could be divided into five groups and ten subgroups, with nearly all the orthologous *MIR156* copies clustered in the same subgroup (Fig. 2). Three species-specific genes, *osa-MIR156l*, *sit-MIR156j*, and *zma-MIR156f* clustered with the *osa-MIR156a*, *sit-MIR156f*, and *zma-MIR156g* subgroups, respectively, suggesting that the former set was duplicated from the latter (Fig. 2). Furthermore, two homologous *MIR156* copies were present at the *osa-MIR156i* (*zma-MIR156i* and *zma-MIR156l*) and *osa-MIR156e* (*zma-MIR156e* and *zma-MIR156h*) loci in maize, supporting previous findings that maize is allotetraploid (Gaut et al., 2000).

3.2 Predicted targets of miR156

Putative targets of miR156 were predicted by psRNATarget (Dai and Zhao, 2011) and TAPIR (Bonnet et al., 2010), both of which predicted that most targets would be *SPL*s (Fig. 3a). A total of 12 *SPL*s were predicted in rice, 9 in *Brachypodium*, 10 in sorghum, 19 in maize, and 10 in foxtail millet (Figs. 3b and S3). Further analysis showed that the miRNA responsive elements (MREs) in most of these *SPL*s were located in the last exon of the gene. In contrast, the MREs of *OsSPL4*, *OsSPL13*, *BdSPL11*, *SiSPL12*, *ZmSPL12a*, and *ZmSPL12b* were located in the 3'-UTR (Table S2), apparently because of the insertion of an upstream terminator UAA codon.

3.3 Identification and analysis of *SPL*s in grass genomes

Previous studies revealed 19 *SPL*s in the rice genome (Xie et al., 2006), which we used as reference

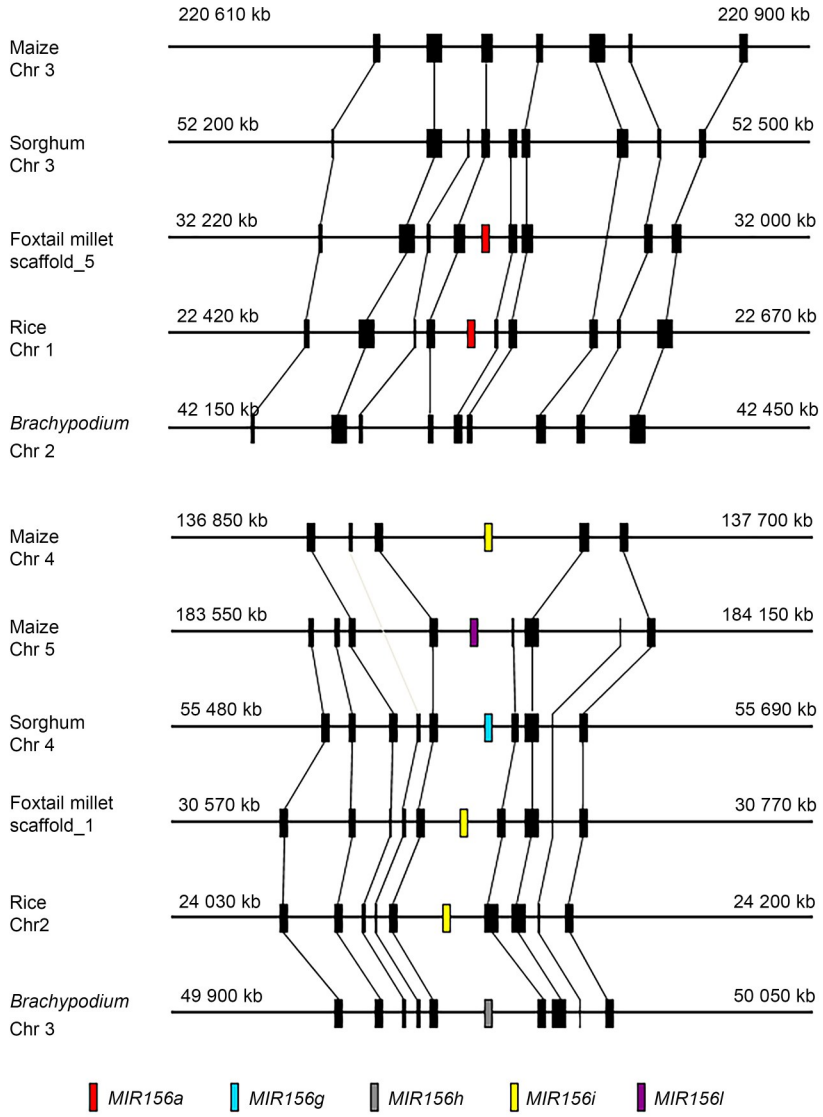

Fig. 1 Sequence alignment of *MIR156* collinear regions in maize, sorghum, foxtail millet, rice, and *Brachypodium*. *osa-MIR156*s were used as the reference sequences to search for orthologous genes in other grass species. The *MIR156*s are shown in colored boxes according to their annotation names in miRBase, whereas other conserved syntenic genes are in black. The numbers represent the relative positions of regions on the chromosomes (Chrs).

sequences to identify 17 *SPL*s in *Brachypodium*, 18 in sorghum, 31 in maize, and 18 in foxtail millet (Table 2). The *BdSPL17* identified in this study is annotated in the *Brachypodium* database as two genes, *Bradi4g18890* and *Bradi4g18900*. However, further analysis showed it to be one continuous gene that is orthologous to *OsSPL19*. Alignment of orthologous regions of all the *SPL*s revealed that only *OsSPL11* and *ZmSPL19* have no orthologous copies in other species and that all other *SPL*s are present in the syntenic regions (Fig. S4). Furthermore, we did not find an ortholog of *OsSPL12* in *Brachypodium*, and a tandem duplication event appears to have occurred at the *ZmSPL15* (*ZmSPL15a* and *ZmSPL15b*) locus in maize (Table 2 and Fig. S4).

The SBP domain of all the SPLs was encoded by the first and second exons despite considerable variation in the gene lengths (from 648 bp to >10 kb; Table S2). Alignment of the nucleotide and protein sequences showed large variation overall, but there was greater conservation within the 80-amino-acid SBP domain. In addition, 14 *SPL*s were identified in moss (*Physcomitrella patens*) and 24 in green algae (*Chlamydomonas reinhardtii*). Therefore, a phylogenetic tree was constructed from the SBP domains of 102 *SPL* sequences

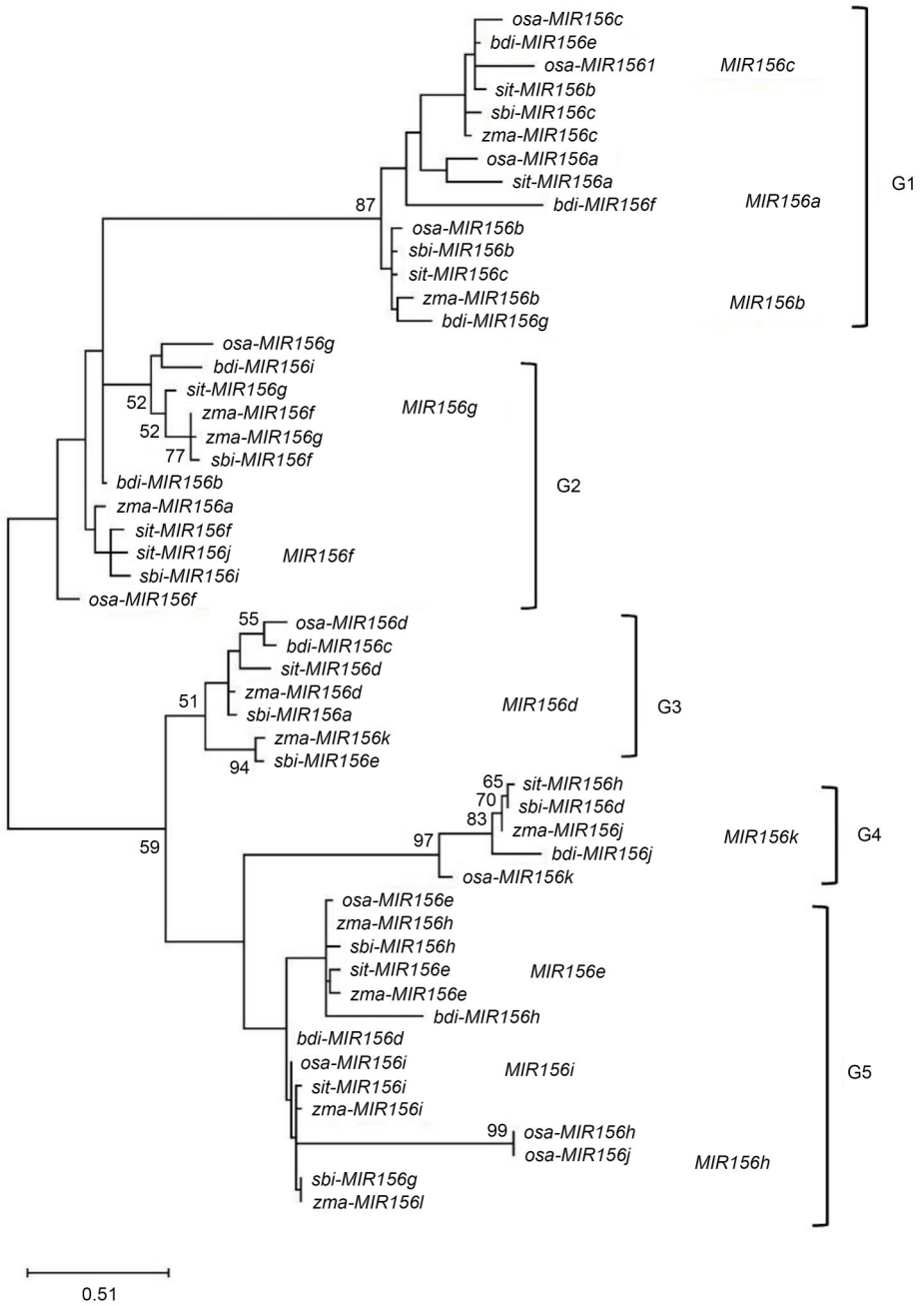

Fig. 2 Phylogenetic analysis of *MIR156*s from members of the grass family. The phylogenetic tree was constructed with 52 *MIR156* stem-loop sequences from the grass family. G1, *MIR156a*/*b*/*c*; G2, *MIR156f*/*g*; G3, *MIR156d*; G4, *MIR156k*; G5, *MIR156e*/*h*/*i*. The tree shown had the highest log likelihood (−891.7215) and bootstrap values of >50%.

from grasses, 17 from Arabidopsis, 14 from moss, and 1 from green algae (as the out-group; Fig. 4a). We found that the *SPL*s fell into eight groups (G1–G8). All of the non-miR156-targeted *SPL*s clustered in groups G2, G3, G4, and G8, and all of the miR156-targeted *SPL*s clustered in groups G1, G5, G6, and G7 (Fig. 4a). G8 contained *SPL*s from all species, including green algae, suggesting that this group may represent the ancestral *SPL*. Furthermore, gene structure analysis showed that members within each group had a similar exon-intron structure and that all non-miR156-targeted *SPL*s had more than 10 exons (except for *SPL*s from angiosperms in G3), whereas miR156-targeted *SPL*s contained only 2–4 exons (Fig. 4b).

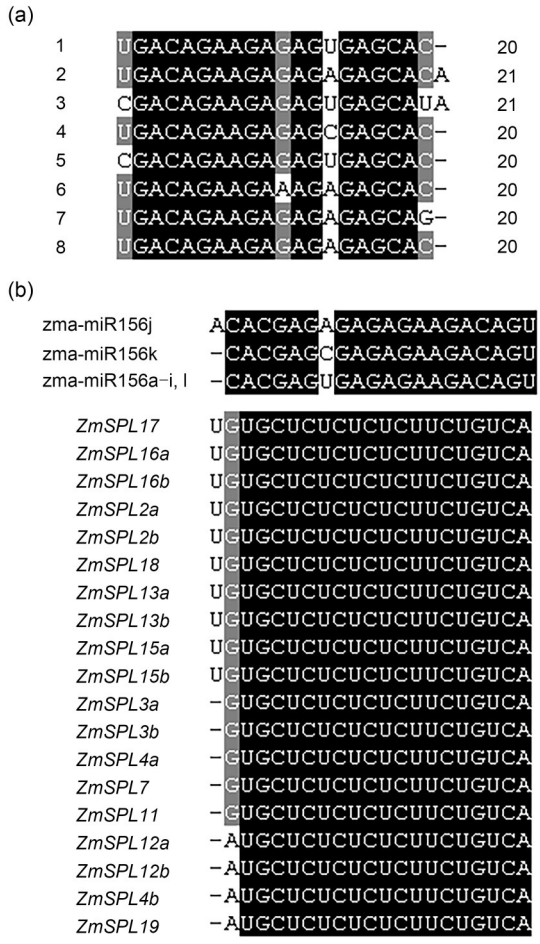

Fig. 3 Alignments of miR156 mature sequences and putative targets in maize. (a) Comparison of miR156 mature sequences from grass species and Arabidopsis. Most mature sequences start with a U residue, which is one of the defining characteristics of miRNAs in plants (Wu et al., 2009). 1: bdi-miR156b–i, osa-miR156a–j, sbi-miR156a–c, sbi-miR156f–i, zea-miR156a–i, zea-miR156l, ath-miR156a–f; 2: bdi-miR156j, osa-miR156k, sbi-miR156d, zea-miR156j; 3: osa-miR156l; 4: sbi-miR156e, zea-miR156k; 5: ath-miR156g; 6: ath-miR156h; 7: ath-miR156i; 8: ath-miR156j. (b) Predicted putative targets of miR156 in maize. Black and gray shading indicate 100% and 80% conserved nucleotides, respectively.

Two species-specific genes, *OsSPL11* and *ZmSPL19*, clustered in the same clade with *OsSPL4* and *ZmSPL4*, respectively (Fig. 4a), suggesting that they were duplicated from *OsSPL4* and *ZmSPL4*, respectively, after their divergence from other species. We also observed that a common homologous region was not present in seven loci (*OsSPL7*, *OsSPL10*, *OsSPL11*, *OsSPL12*, *OsSPL16*, *OsSPL18*, and *OsSPL19*), and one orthologous gene was deleted at the *OsSPL9* locus in the maize genome (Figs. 4a and S4).

The SBP domains of the *SPL*s were rich in cysteine, arginine, and histidine residues, which suggests the presence of two Zn^{2+}-binding sites (Birkenbihl et al., 2005; Kropat et al., 2005), and a putative nuclear localization signal (NLS) that partly overlaps with the second Zn^{2+}-binding site was located in the C-terminus of the SBP domain (Fig. 5). Intriguingly, the conserved SBP domains of some SPLs, such as BdSPL7, ZmSPL2a, and ZmSPL15b, were separated by additional exon or sequence insertions, which presumably come together in the three-dimensional (3D) conformation of the protein to form one active DNA-binding site.

3.4 Expression profiles of *SPL*s in different rice tissues and at different developmental stages

Because rice miR156 expression is higher in the juvenile phase than in adult tissues (Sunkar et al., 2008; Paterson et al., 2009), we hypothesized that miR156-targeted *SPL*s would show a converse trend. Expression data for *SPL*s in 11 rice tissues/developmental stages were retrieved from the GEO database. The *OsSPL* expression patterns were classified into four types (Fig. 6 and Table S3). The non-miR156-targeted genes *OsSPL1*, *OsSPL6*, *OsSPL9*, and *OsSPL15* showed relatively stable and high expression throughout all plant developmental phases (Fig. 6). In particular, *OsSPL15* was expressed at a very high level in all 11 tissues/developmental stages, suggesting that it probably acts as a general housekeeping gene in plant growth and development. In contrast, the miR156-targeted genes *OsSPL2*, *OsSPL3*, *OsSPL12*, and *OsSPL13* were more highly expressed in early stages (from shoots to post-emergence inflorescence) than in later stages (anther to 25-DAP (day after pollination) endosperm; Fig. 6). Eight genes, including five miR156-targeted genes (*OsSPL4*, *OsSPL11*, *OsSPL14*, *OsSPL16*, and *OsSPL18*) and three non-miR156-targeted genes (*OsSPL5*, *OsSPL8*, and *OsSPL10*) were more highly expressed in pistils and in pre- and post-emergence inflorescence stages than in other tissues/developmental stages. The remaining miR156-targeted genes (*OsSPL7*, *OsSPL17*, and *OsSPL19*) were expressed at low levels in all 11 tissues/developmental stages.

Expression levels of almost all the *SPL*s peaked at the pre-emergence inflorescence stage, which strongly correlates with flowering in rice. They gradually decreased through the post-emergence inflorescence

Table 2 *SPL*s identified in orthologous regions of rice, *Brachypodium*, foxtail millet, sorghum, and maize

No.	Rice		*Brachypodium*		Foxtail millet	
	Standard *SPL* gene name	Gene name from the respective genome database	Standard *SPL* gene name	Gene name from the respective genome database	Standard *SPL* gene name	Gene name from the respective genome database
1	*OsSPL1*	*LOC_Os01g18850*	*BdSPL1*	*Bradi2g11240*	*SiSPL1*	*Si000259m.g*
2	*OsSPL2*	*LOC_Os01g69830*	*BdSPL2*	*Bradi2g59110*	*SiSPL2*	*Si001804m.g*
3	*OsSPL3*	*LOC_Os02g04680*	*BdSPL3*	*Bradi3g03510*	*SiSPL3*	*Si017520m.g*
4	*OsSPL4*	*LOC_Os02g07780*	*BdSPL4*	*Bradi3g05510*	*SiSPL4*	*Si017749m.g*
5	*OsSPL5*	*LOC_Os02g08070*	*BdSPL5*	*Bradi3g05720*	*SiSPL5*	*Si017338m.g*
6	*OsSPL6*	*LOC_Os03g61760*	*BdSPL6*	*Bradi1g02760*	*SiSPL6*	*Si034095m.g*
7	*OsSPL7*	*LOC_Os04g46580*	*BdSPL7*	*Bradi5g17720*	*SiSPL7*	*Si010262m.g*
8	*OsSPL8*	*LOC_Os04g56170*	*BdSPL8*	*Bradi5g24670*	*SiSPL8*	*Si022315m.g*
9	*OsSPL9*	*LOC_Os05g33810*	*BdSPL9*	*Bradi2g25580*	*SiSPL9*	*Si021754m.g*
10	*OsSPL10*	*LOC_Os06g44860*	*BdSPL10*	*Bradi1g31390*	*SiSPL10*	*Si006559m.g*
11	*OsSPL11*	*LOC_Os06g45310*				
12	*OsSPL12*	*LOC_Os06g49010*			*SiSPL11*	*Si006472m.g*
13	*OsSPL13*	*LOC_Os07g32170*	*BdSPL11*	*Bradi1g26720*	*SiSPL12*	*Si031154m.g*
14	*OsSPL14*	*LOC_Os08g39890*	*BdSPL12*	*Bradi3g40030*	*SiSPL13*	*Si013870m.g*
15	*OsSPL15*	*LOC_Os08g40260*	*BdSPL13*	*Bradi3g40240*	*SiSPL14*	*Si013151m.g*
16	*OsSPL16*	*LOC_Os08g41940*	*BdSPL14*	*Bradi3g41250*	*SiSPL15*	*Si013747m.g*
17	*OsSPL17*	*LOC_Os09g31438*	*BdSPL15*	*Bradi4g33770*	*SiSPL16*	*Si030195m.g*
18	*OsSPL18*	*LOC_Os09g32944*	*BdSPL16*	*Bradi4g34667*	*SiSPL17*	*Si030892m.g*
19	*OsSPL19*	*LOC_Os11g30370*	*BdSPL17*	*Bradi4g18900*	*SiSPL18*	*Si026656m.g*
20						

No.	Sorghum		Maize 1		Maize 2	
	Standard *SPL* gene name	Gene name from the respective genome database	Standard *SPL* gene name	Gene name from the respective genome database	Standard *SPL* gene name	Gene name from the respective genome database
1	*SbSPL1*	*Sb03g011920*	*ZmSPL1a*	*GRMZM2G133646*	*ZmSPL1b*	*GRMZM2G081127*
2	*SbSPL2*	*Sb03g044160*	*ZmSPL2a*	*GRMZM5G878561*	*ZmSPL2b*	*GRMZM2G371033*
3	*SbSPL3*	*Sb04g003175*	*ZmSPL3a*	*GRMZM2G065451*	*ZmSPL3b*	*GRMZM2G097275*
4	*SbSPL4*	*Sb04g004940*	*ZmSPL4a*	*GRMZM2G163813*	*ZmSPL4b*	*GRMZM2G126827*
5	*SbSPL5*	*Sb04g005180*	*ZmSPL5a*	*GRMZM2G168229*	*ZmSPL5b*	*GRMZM2G111136*
6	*SbSPL6*	*Sb01g002530*	*ZmSPL6a*	*GRMZM2G156756*	*ZmSPL6b*	*GRMZM2G138421*
7	*SbSPL7*	*Sb06g024630*	*ZmSPL7*	*GRMZM2G148467*		
8	*SbSPL8*	*Sb06g031290*	*ZmSPL8a*	*GRMZM2G036297*	*ZmSPL8b*	*GRMZM2G058588*
9	*SbSPL9*	*Sb09g020110*	*ZmSPL9*	*GRMZM2G109354*		
10	*SbSPL10*	*Sb10g026200*	*ZmSPL10*	*GRMZM2G101499*		
11						
12	*SbSPL11*	*Sb10g029190*	*ZmSPL11*	*GRMZM2G414805*		
13	*SbSPL12*	*Sb02g034180*	*ZmSPL12a*	*GRMZM2G113779*	*ZmSPL12b*	*GRMZM2G067624*
14	*SbSPL13*	*Sb07g027740*	*ZmSPL13a*	*GRMZM2G460544*	*ZmSPL13b*	*GRMZM2G160917*
15	*SbSPL14*	*Sb07g027420*	*ZmSPL14a*	*GRMZM2G098557*	*ZmSPL14b*	*GRMZM2G169270*
16	*SbSPL15*	*Sb07g026220*	*ZmSPL15a*	*GRMZM2G101511*	*ZmSPL15b*	*AC233751.1_FGT002*
17	*SbSPL16*	*Sb02g028420*	*ZmSPL16a*	*GRMZM2G126018*	*ZmSPL16b*	*GRMZM2G307588*
18	*SbSPL17*	*Sb02g029300*	*ZmSPL17*	*GRMZM2G061734*		
19	*SbSPL18*	*Sb05g017510*	*ZmSPL18*	*GRMZM2G106798*		
20			*ZmSPL19*	*GRMZM2G156621*		

Rice SQUAMOSA promoter-binding protein-like (*SPL*) genes (Xie et al., 2006) were used as reference genes to search for orthologs in other grass genomes. The two homeologous chromosomes were named Maize 1 and Maize 2.

stage and then dropped dramatically (Fig. 6). As expected, miR156-targeted *OsSPL*s were more highly expressed in the adult reproductive phase than in the vegetative phase, and *OsSPL*s that contain MREs had similar expression patterns (Fig. 6).

3.5 Methylation patterns of *SPL*s and *MIR156*s in rice tissues

To investigate whether DNA methylation correlates with the expression of *MIR156*s and *SPL*s, we

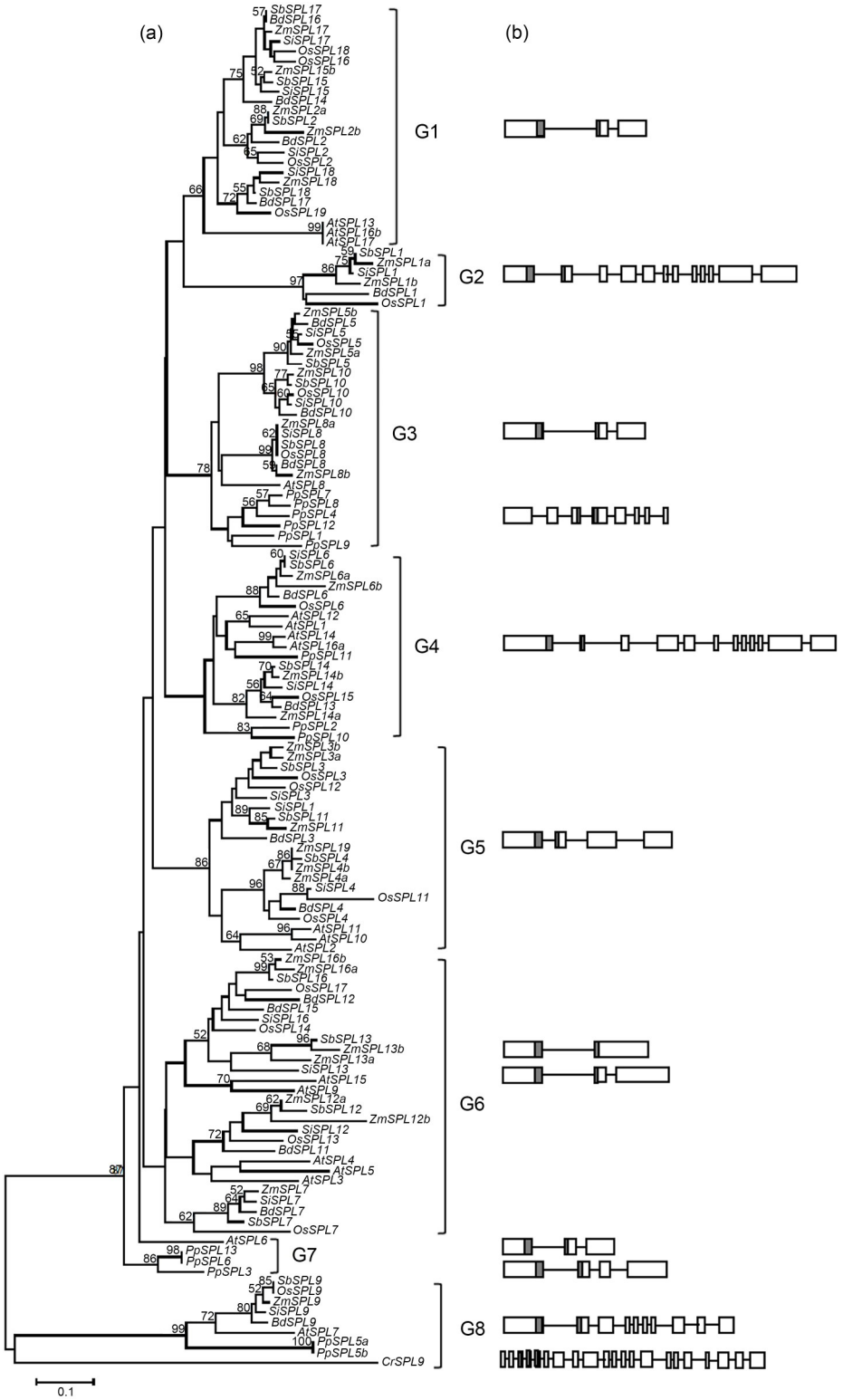

Fig. 4 Phylogenetic analysis of SQUAMOSA promoter-binding protein-like (*SPL*) genes from grass, Arabidopsis, and moss. (a) A phylogenetic tree constructed from 102 *SPL* sequences from grass, 17 from Arabidopsis, 14 from moss, and 1 from green algae (as an out-group). The genes fall into eight lineages (G1–G8), some of which are targeted by miRNA156 (G1, G5−G7) and some of which are not (G2–G4, G8). (b) Gene structure of the *SPL*s. Gray filled boxes, SBP domains; White boxes, other exon regions; Lines, introns.

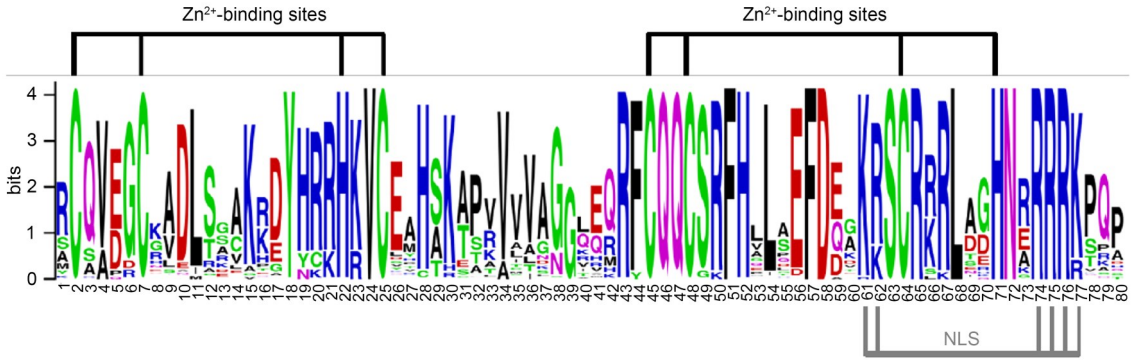

Fig. 5 WebLogo representation of the SQUAMOSA promoter-binding protein (SBP) domain sequences within SQUAMOSA promoter-binding protein-like (SPL) proteins of several grass species. The overall height of each stack within the 80-amino-acid SBP domain sequence reflects the conservation at the designated position, whereas the height of each single-letter amino acid symbol within each stack indicates the frequency of that amino acid. Two Zn^{2+}-binding sites, Cys-Cys-His-Cys and Cys-Cys-Cys-His, are marked by the black lines above the sequence, and the nuclear localization signal (NLS) is marked by the gray lines below the sequence.

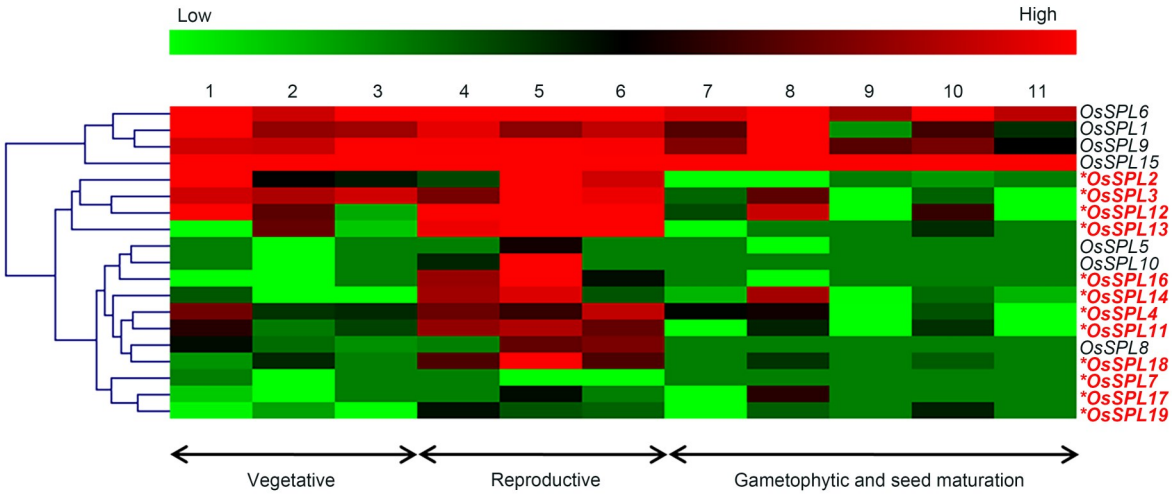

Fig. 6 Heatmap representation of SQUAMOSA promoter-binding protein-like (*SPL*) expression in different rice tissues and at developmental stages. The rice life cycle is divided into three phases: vegetative (1–3), reproductive (4–6), and gametophytic phase and seed maturation (9–11). Expression data for *OsSPL*s in 11 tissues/developmental stages were obtained from the Gene Expression Omnibus (GEO) database. Relative gene expression levels from low (green) to high (red) are shown. The miR156-targeted genes are indicated by asterisks with red bold font, and non-targeted genes with normal font. 1. shoots; 2. leaves-20 d; 3. seedling four-leaf stage; 4. pistil; 5. pre-emergence inflorescence; 6. post-emergence inflorescence; 7. anther; 8. seed-5 DAP; 9. seed-10 DAP; 10. embryo-25 DAP; 11. endosperm-25 DAP. DAP: day after pollination.

examined the DNA methylation status of the promoter, gene body, and 3'-UTRs of all *MIR156*s and *SPL*s in four rice tissues (root, shoot, embryo, and endosperm) using data from GEO (Tables S4 and S5). The promoters (2 kb upstream from the start codon) of *SPL*s exhibited variable patterns of methylation. Almost no (<0.5%) methylation was seen in the promoters of some genes (*OsSPL3*, *OsSPL7*, *OsSPL12*, and *OsSPL18*), slight methylation (1%–10%) in others (*OsSPL1*, *OsSPL2*, *OsSPL6*, *OsSPL8*, *OsSPL10*, *OsSPL13*, *OsSPL15*, *OsSPL16*, and *OsSPL19*), and a high degree of methylation (>15%) in the remaining genes (*OsSPL4*, *OsSPL5*, *OsSPL9*, *OsSPL11*, *OsSPL14*, and *OsSPL17*). More than 50% of the methylation in the highly methylated promoters occurred within distinct methylation blocks (Fig. 7a, Tables S4 and S6). Methylation levels of all *SPL* promoters were lower in endosperm than in other tissues except for *OsSPL2* and the genes with little or no promoter methylation. Furthermore, overall promoter methylation at CG sites was 2-fold higher than

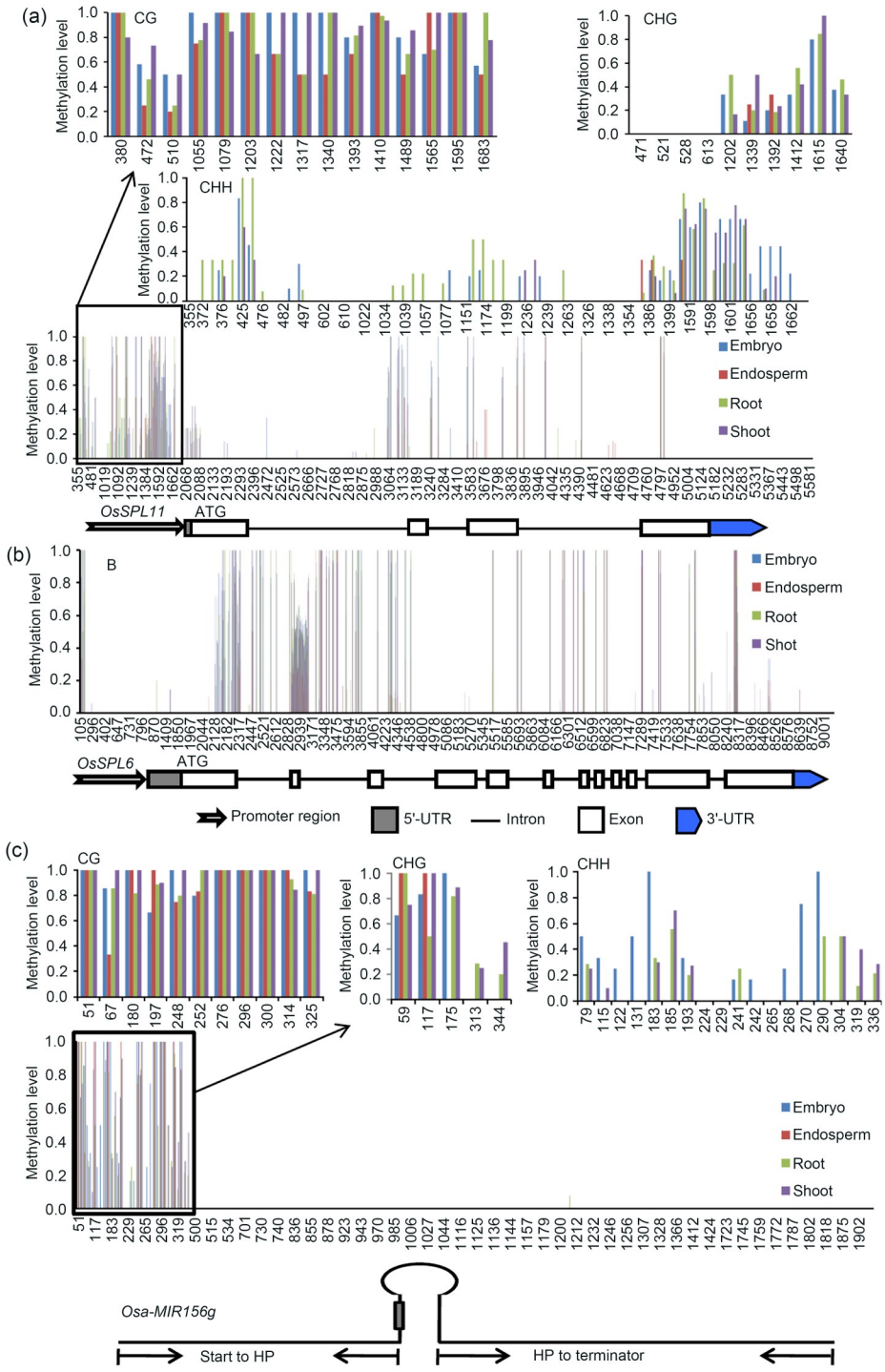

Fig. 7 Methylation analysis of *OsSPL*s and *osa-MIR156*s in four rice tissues (embryo, endosperm, root, and shoot). (a, b) *OsSPL11* (a) is an example of an SQUAMOSA promoter-binding protein-like (*SPL*) that is highly methylated within its putative promoter region and sparsely methylated within its gene body and 3'-untranslated region (3'-UTR), whereas *OsSPL6* (b) is an example of biased methylation within the gene body and hypomethylation in the promoter and 3'-UTR; (c) *osa-MIR156g* exemplifies the methylation bias upstream of the hairpin (HP) structure and the absence of methylcytosine in the stem-loop sequences that are observed in *osa-miR156* genes. The boxed regions of the methylation graphs indicate which regions are presented at higher resolution in the accompanying graphs. Note that the full-length sequence shows all three methylation sites. The gene structures are shown at the bottom of each subpanel. Both the *SPL*s and *MIR156*s are less frequently methylated in endosperm than in other tissues.

at CHG sites and 3-fold higher than at CHH sites. However, CG-site methylation was 5-fold higher than at CHG sites and 12-fold higher than at CHH sites in the endosperm (Table S6), which suggested a more pronounced hypomethylation in CHG and CHH contexts in the endosperm.

In terms of gene body methylation, the *SPL*s fell into two groups: those that were highly methylated (>5%; *OsSPL1*, *OsSPL3*, *OsSPL4*, *OsSPL6*, *OsSPL9*, *OsSPL12*, *OsSPL14*, *OsSPL15*, *OsSPL17*, and *OsSPL19*) and those with little or no methylation (<5%; Fig. 7b, Tables S4 and S7). About 40% of the gene body methylation in the highly methylated group was at CG sites, whereas CHG and CHH methylation made up <1%. We also found that no cytosines were methylated in MREs.

In contrast to the methylation patterns in promoters and gene bodies, there was almost no methylcytosine in the 3'-UTR of the *SPL*s, except for three methylated CG sites in *OsSPL4*, five in *OsSPL14*, and one in *OsSPL15* (Table S4).

In the case of *MIR156*, the methylation frequency in the putative promoter regions (1 kb upstream of the stem-loop structure) of *osa-MIR156a*/*f*/*g*/*h*/*j*/*l* was >10% (Fig. 7c and Table S5), and methylation at CG sites was 2-fold higher than that at CHG sites and 3-fold higher than that at CHH sites (Tables S5 and S8), similar to the *SPL* promoters. In contrast, there were almost no methylcytosines in the gene bodies (stem-loop sequences) except in *osa-MIR156d*, in which gene body methylation reached 75% in the endosperm (Table S5). However, because all mature miR156 sequences are nearly identical, we were unable to assess the effect of DNA methylation on their expression.

4 Discussion

4.1 Origin and evolution of *SPL*s

In plants, miRNAs tend to facilitate the cleavage of their target mRNA, rather than inhibit translation, as is more common in animals (Llave et al., 2002; Palatnik et al., 2003). *SPL*s, which encode a conserved SBP domain, are the primary targets of miR156 (Figs. 3b and S3). SBP-box genes have been identified in moss and green algae but not in other algae, which indicates that they are plant-specific and likely originated before the divergence of green plants (Cardon et al., 1999; Riese et al., 2007; Guo et al., 2008; Chang et al., 2016; Lei and Liu, 2016; Morea et al., 2016; Peng et al., 2016; Wang et al., 2018; Song et al., 2020). Our phylogenetic analysis showed that most *CrSPL*s clustered together as an out-group (data not shown), but some clustered with the non-miR156-targeted *SPL*s (Fig. 4a). We therefore hypothesize that the original *SPL* was a non-miR156-targeted gene that then diversified into non-miR156-targeted and miR156-targeted *SPL*s that contained MREs after the divergence of green algae from the progenitor of land plants, but before land plants diverged. Similar phylogenetic histories are thought to have occurred in other plant-specific transcription factors such as DNA-binding with one finger (DOF) and APETALA2 (AP2) (Shigyo et al., 2006; Moreno-Risueno et al., 2007; Agarwal et al., 2016; Jin et al., 2018).

Our analysis suggests that only four *SPL*s were present before the divergence of land plants, three of which were non-miR156-targeted and one of which was miR156-targeted. The *SPL*s then expanded to at least 7 copies before the divergence of monocots and dicots, and to 18 copies in the grass family (Fig. 4a). Moreover, we found that the miR156-targeted and non-miR156-targeted *SPL*s had distinguishable gene structures (compare G1 and G5–G7 with G2–G4 and G8, respectively; Fig. 4). However, the gene structure of one group of the non-miR156-targeted genes (G3) showed similarities to that of the miR156-targeted genes (Fig. 4b). Further analysis showed that *SPL8* in this group had a putative MRE in its 3'-UTR (Fig. S5), suggesting that it could be converted into an miR156-targted gene by alternative splicing. We also found that three genes (*OsSPL5*, *OsSPL8*, and *OsSPL10*) in this group had the same expression pattern as miR156-targeted *SPL*s (Fig. 6). Our data suggested that non-miR156-targeted genes could diversify into miR156-targeted genes by simultaneous evolution toward fewer exons and the formation of MREs (Fig. 8).

4.2 Origin and evolution of *MIR156* in grass genomes

One proposed mechanism for the evolutionary origins of miRNAs is that they were initially generated from the inverted duplication of a target gene. That no miR156 sequence was identified in green algae suggests that *MIR156* arose from its target *SPL* after the divergence of green algae, but before land plants emerged (Tang, 2010). This raises the intriguing possibility

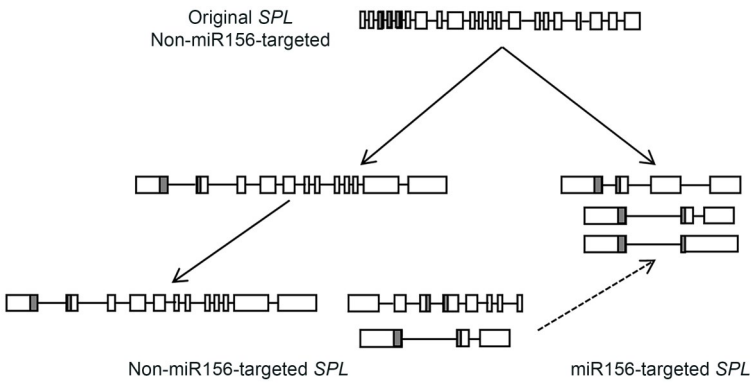

Fig. 8 Model for the evolutionary diversification of SQUAMOSA promoter-binding protein-like (*SPL*) gene structure. In this model, the original *SPL* was not targeted by miR156 and was made up of multiple exons (>10 exons). This gene then diversified into non-miR156-targeted (>10 exons) and miR156-targeted (2–4 exons) genes containing microRNA responsive elements (MREs). Under strong environmental pressure, however, alternate splicing may have allowed non-miR156-targeted SPLs to reduce their number of exons, which could potentially have enabled them to form MREs, thus transforming them into miR156-targeted genes. The gray boxes indicate the SQUAMOSA promoter-binding protein (SBP) domains, and all exons and introns are shown in boxes and lines.

that the co-evolution of *MIR156*s and its target SPLs may have been an important regulatory mechanism during evolution. Mature miR156 sequences are highly conserved in grasses and Arabidopsis (Fig. 4a), but their gene sequences are extremely diverse. This indicates that there may have been strong selective pressure to maintain the secondary structure of miR156 species, further suggesting that miR156 has served a critical role in the development of plants throughout their evolutionary history (Axtell et al., 2007). Moreover, our results demonstrated that tandem, segmental, and whole-genome duplications appear to have played important roles in the expansion of the *MIR156* gene family in grass genomes.

Almost all orthologous *MIR156* genes clustered together in our phylogenetic analysis, and the original *MIR156* had expanded to at least ten genes before the divergence of the grasses. However, the *MIR156*s from Arabidopsis clustered together as an out-group (data not shown), suggesting that no orthologous *MIR156* genes are present in monocots and dicots and that miRNAs in this branch may have resulted from a different evolutionary pathway.

Interestingly, we observed that miR529 and miR156 were very similar, differing only at nucleotides 10 and 14 (Fig. S6). Furthermore, some *SPL*s (such as *OsSPL14* and *OsSPL17*) are also targeted by miR529 (Jeong et al., 2011; Yue et al., 2017). In contrast to the decline in miR156 after seedling development, miR529 expression is sustained throughout the growth and development of plants, and *OsSPL14* is predominantly cleaved by miR529 in panicles (Jeong et al., 2011). miR156 is highly expressed in various plants and tissues (Sunkar et al., 2008; Gou et al., 2011; Jiao et al., 2011; Schreiber et al., 2011), and in some cases accounts for nearly half of the total miRNA expression (Barrett et al., 2011), suggesting that it may be a housekeeping miRNA. Furthermore, *MIR156* and *MIR529* are both present in the common ancestor of embryophytes, whereas *MIR529* is not present in the common ancestor of Eudicots-Asterids (Cuperus et al., 2011; Axtell and Meyers, 2018), which suggests that these two genes originated from the same ancestor but had different evolutionary fates.

4.3 Correlation of DNA methylation with gene expression

A large number of *OsSPL*s were methylated within their gene bodies, primarily at CG sites. DNA methylation was markedly less prevalent in the 5' and 3' regions of the transcription unit, where it could affect transcription factor binding (Zilberman et al., 2007). In addition, there was a positive correlation between gene body methylation and promoter hypomethylation and transcription (Figs. 6 and 7, Tables S4–S6) (Zilberman et al., 2007). *OsSPL*9 was highly methylated in certain regions upstream of the start site, but it had a very high expression level, suggesting that the core promoter could lie outside of these methylation regions. Another six *OsSPL*s showed higher promoter

methylation but no difference in their levels of expression compared with genes that were less frequently methylated in their promoters (Fig. 6 and Table S4). We hypothesized that the methylation within these upstream regions may not be in functional motifs and/or promoter regions. Furthermore, we observed methylation patterns that were indistinguishable between some miR156-targeted *OsSPL*s and the rest of the non-miR156-targeted *OsSPL*s (*OsSPL5*, *OsSPL8*, and *OsSPL10*). This confirmed our assumption that DNA methylation functions as a control mechanism at a pre-transcriptional level, whereas miRNAs are involved at a post-transcriptional level. These mechanisms may thus provide complementary gene regulation.

Non-CG methylation was reduced in all *OsSPL*s in the endosperm compared with the other tissues (shoot, root, and embryo), especially within the promoter, whereas CG hypomethylation was localized (Fig. 7, Tables S4 and S6). This is consistent with a previous study (Zemach et al., 2010) and suggests that hypomethylation in rice endosperm is a mechanism for activating genes.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 31901515).

Author contributions

Jianhong XU designed the study, wrote and edited the manuscript. Erkui YUE performed the experiments, analyzed the data, and wrote the manuscript. Hua TAO performed the experiments and analyzed the data. All authors have read and approved the final manuscript and, therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethic guidelines

Erkui YUE, Hua TAO, and Jianhong XU declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Agarwal G, Garg V, Kudapa H, et al., 2016. Genome-wide dissection of AP2/ERF and HSP90 gene families in five legumes and expression profiles in chickpea and pigeonpea. *Plant Biotechnol J*, 14(7):1563-1577. https://doi.org/10.1111/pbi.12520

Axtell MJ, Meyers BC, 2018. Revisiting criteria for plant microRNA annotation in the era of big data. *Plant Cell*, 30(2):272-284. https://doi.org/10.1105/tpc.17.00851

Axtell MJ, Snyder JA, Bartel DP, 2007. Common functions for diverse small RNAs of land plants. *Plant Cell*, 19(6): 1750-1769. https://doi.org/10.1105/tpc.107.051706

Barrett T, Troup DB, Wilhite SE, et al., 2011. NCBI GEO: archive for functional genomics data sets—10 years on. *Nucleic Acids Res*, 39(S1):D1005-D1010. https://doi.org/10.1093/nar/gkq1184

Bartel DP, 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281-297. https://doi.org/10.1016/S0092-8674(04)00045-5

Bennetzen JL, Schmutz J, Wang H, et al., 2012. Reference genome sequence of the model plant setaria. *Nat Biotechnol*, 30(6):555-561. https://doi.org/10.1038/nbt.2196

Birkenbihl RP, Jach G, Saedler H, et al., 2005. Functional dissection of the plant-specific SBP-domain: overlap of the DNA-binding and nuclear localization domains. *J Mol Biol*, 352(3):585-596. https://doi.org/10.1016/j.jmb.2005.07.013

Bonnet E, He Y, Billiau K, et al., 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, 26(12):1566-1568. https://doi.org/10.1093/bioinformatics/btq233

Cardon G, Höhmann S, Klein J, et al., 1999. Molecular characterisation of the *Arabidopsis* SBP-box genes. *Gene*, 237(1):91-104. https://doi.org/10.1016/S0378-1119(99)00308-X

Chang JZ, Yan FX, Qiao LY, et al., 2016. Genome-wide identification and expression analysis of SBP-box gene family in *Sorghum bicolor* L. *Hereditas (Beijing)*, 38(6): 569-580 (in Chinese). https://doi.org/10.16288/j.yczz.16-008

Cuperus JT, Fahlgren N, Carrington JC, 2011. Evolution and functional diversification of *MIRNA* genes. *Plant Cell*, 23(2):431-442. https://doi.org/10.1105/tpc.110.082784

Dai XB, Zhao PX, 2011. PsRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res*, 39(S2):W155-W159. https://doi.org/10.1093/nar/gkr319

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792-1797. https://doi.org/10.1093/nar/gkh340

Feng SH, Jacobsen SE, Reik W, 2010. Epigenetic reprogramming in plant and animal development. *Science*, 330(6004): 622-627. https://doi.org/10.1126/science.1190614

Fornara F, Coupland G, 2009. Plant phase transitions make a SPLash. *Cell*, 138(4):625-627. https://doi.org/10.1016/j.cell.2009.08.011

Franco-Zorrilla JM, Valli A, Todesco M, et al., 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat Genet*, 39(8):1033-1037. https://doi.org/10.1038/ng2079

Gaut BS, le Thierry D'Ennequin M, Peek AS, et al., 2000. Maize as a model for the evolution of plant nuclear genomes. *Proc Natl Acad Sci USA*, 97(13):7008-7015. https://doi.org/10.1073/pnas.97.13.7008

Gielen H, Remans T, Vangronsveld J, et al., 2016. Toxicity responses of Cu and Cd: the involvement of miRNAs and the transcription factor SPL7. *BMC Plant Biol*, 16:145. https://doi.org/10.1186/s12870-016-0830-4

Gou JY, Felippes FF, Liu CJ, et al., 2011. Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. *Plant Cell*, 23(4):1512-1522. https://doi.org/10.1105/tpc.111.084525

Guo AY, He K, Liu D, et al., 2005. DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics*, 21(10): 2568-2569. https://doi.org/10.1093/bioinformatics/bti334

Guo AY, Zhu QH, Gu XC, et al., 2008. Genome-wide identification and evolutionary analysis of the plant specific SBP-box transcription factor family. *Gene*, 418(1-2):1-8. https://doi.org/10.1016/j.gene.2008.03.016

Huijser P, Schmid M, 2011. The control of developmental phase transitions in plants. *Development*, 138(19):4117-4129. https://doi.org/10.1242/dev.063511

International Rice Genome Sequencing Project, Sasaki T, 2005. The map-based sequence of the rice genome. *Nature*, 436(7052):793-800. https://doi.org/10.1038/nature03895

Jeong DH, Park S, Zhai JX, et al., 2011. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. *Plant Cell*, 23(12):4185-4207. https://doi.org/10.1105/tpc.111.089045

Jiao YP, Song WB, Zhang M, et al., 2011. Identification of novel maize miRNAs by measuring the precision of precursor processing. *BMC Plant Biol*, 11:141. https://doi.org/10.1186/1471-2229-11-141

Jiao YQ, Wang YH, Xue DW, et al., 2010. Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. *Nat Genet*, 42(6):541-544. https://doi.org/10.1038/ng.591

Jin JH, Wang M, Zhang HX, et al., 2018. Genome-wide identification of the AP2/ERF transcription factor family in pepper (*Capsicum annuum* L.). *Genome*, 61(9):663-674. https://doi.org/10.1139/gen-2018-0036

Kimura M, 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*, 16(2):111-120.

Kozomara A, Griffiths-Jones S, 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*, 39(S1):D152-D157. https://doi.org/10.1093/nar/gkq1027

Kropat J, Tottey S, Birkenbihl RP, et al., 2005. A regulator of nutritional copper signaling in *Chlamydomonas* is an SBP domain protein that recognizes the GTAC core of copper response element. *Proc Natl Acad Sci USA*, 102(51): 18730-18735. https://doi.org/10.1073/pnas.0507693102

Kumar S, Stecher G, Li M, et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, 35(6):1547-1549. https://doi.org/10.1093/molbev/msy096

Lal S, Pacis LB, Smith HMS, 2011. Regulation of the *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE genes/microRNA156* module by the homeodomain proteins PENNYWISE and POUND-FOOLISH in *Arabidopsis*. *Mol Plant*, 4(6):1123-1132. https://doi.org/10.1093/mp/ssr041

Langmead B, Trapnell C, Pop M, et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3):R25. https://doi.org/10.1186/gb-2009-10-3-r25

Larkin MA, Blackshields G, Brown NP, et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21): 2947-2948. https://doi.org/10.1093/bioinformatics/btm404

Law JA, Jacobsen SE, 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*, 11(3):204-220. https://doi.org/10.1038/nrg2719

Lei KJ, Liu H, 2016. Research advances in plant regulatory hub miR156 and targeted SPL family. *Chem Life*, 36(1): 13-20 (in Chinese). https://doi.org/10.13488/j.smhx.20160103

Liu MM, Shi ZY, Zhang XH, et al., 2019. Inducible overexpression of *Ideal Plant Architecture1* improves both yield and disease resistance in rice. *Nat Plants*, 5(4):389-400. https://doi.org/10.1038/s41477-019-0383-2

Llave C, Xie ZX, Kasschau KD, et al., 2002. Cleavage of *Scarecrow-like* mRNA targets directed by a class of *Arabidopsis* miRNA. *Science*, 297(5589):2053-2056. https://doi.org/10.1126/science.1076311

Manning K, Tör M, Poole M, et al., 2006. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet*, 38(8):948-952. https://doi.org/10.1038/ng1841

Martin RC, Liu PP, Goloviznina NA, et al., 2010. microRNA, seeds, and darwin?: diverse function of miRNA in seed biology and plant responses to stress. *J Exp Bot*, 61(9): 2229-2234. https://doi.org/10.1093/jxb/erq063

Meyers BC, Axtell MJ, Bartel B, et al., 2008. Criteria for annotation of plant microRNAs. *Plant Cell*, 20(12):3186-3190. https://doi.org/10.1105/tpc.108.064311

Miao CB, Wang Z, Zhang L, et al., 2019. The grain yield modulator miR156 regulates seed dormancy through the gibberellin pathway in rice. *Nat Commun*, 10:3822. https://doi.org/10.1038/s41467-019-11830-5

Miura K, Ikeda M, Matsubara A, et al., 2010. *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat Genet*, 42(6):545-549. https://doi.org/10.1038/ng.592

Morea EGO, da Silva EM, Silva GFFE, et al., 2016. Functional and evolutionary analyses of the miR156 and miR529

families in land plants. *BMC Plant Biol*, 16:40. https://doi.org/10.1186/s12870-016-0716-5

Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, et al., 2007. The family of DOF transcription factors: from green unicellular algae to vascular plants. *Mol Genet Genomics*, 277(4):379-390. https://doi.org/10.1007/s00438-006-0186-9

Nozawa M, Miura S, Nei M, 2012. Origins and evolution of microRNA genes in plant species. *Genome Biol Evol*, 4(3): 230-239. https://doi.org/10.1093/gbe/evs002

Palatnik JF, Allen E, Wu XL, et al., 2003. Control of leaf morphogenesis by microRNAs. *Nature*, 425(6955):257-263. https://doi.org/10.1038/nature01958

Paterson AH, Bowers JE, Bruggmann R, et al., 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 457(7229):551-556. https://doi.org/10.1038/nature07723

Peng H, He XJ, Gao J, et al., 2016. Genome-wide identification and function analysis of SBP gene family in maize. *Acta Agron Sin*, 42(2):201-211 (in Chinese). https://doi.org/10.3724/SP.J.1006.2016.00201

Ramamurthy RK, Xiang QY, Hsieh EJ, et al., 2018. New aspects of iron-copper crosstalk uncovered by transcriptomic characterization of Col-0 and the copper uptake mutant *spl7* in *Arabidopsis thaliana*. *Metallomics*, 10(12): 1824-1840. https://doi.org/10.1039/c8mt00287h

Reinhart BJ, Weinstein EG, Rhoades MW, et al., 2002. MicroRNAs in plants. *Genes Dev*, 16(13):1616-1626. https://doi.org/10.1101/gad.1004402

Riese M, Höhmann S, Saedler H, et al., 2007. Comparative analysis of the SBP-box gene families in *P. patens* and seed plants. *Gene*, 401(1-2):28-37. https://doi.org/10.1016/j.gene.2007.06.018

Saeed AI, Bhagabati NK, Braisted JC, et al., 2006. TM4 microarray software suite. *Methods Enzymol*, 411:134-193. https://doi.org/10.1016/S0076-6879(06)11009-5

Schnable PS, Ware D, Fulton RS, et al., 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326(5956):1112-1115. https://doi.org/10.1126/science.1178534

Schreiber AW, Shi BJ, Huang CY, et al., 2011. Discovery of barley miRNAs through deep sequencing of short reads. *BMC Genomics*, 12:129. https://doi.org/10.1186/1471-2164-12-129

Shigyo M, Hasebe M, Ito M, 2006. Molecular evolution of the AP2 subfamily. *Gene*, 366(2):256-265. https://doi.org/10.1016/j.gene.2005.08.009

Sollome J, Martin E, Sethupathy P, et al., 2016. Environmental contaminants and microRNA regulation: transcription factors as regulators of toxicant-altered microRNA expression. *Toxicol Appl Pharmacol*, 312:61-66. https://doi.org/10.1016/j.taap.2016.06.009

Song J, Cao XN, Wang HL, et al., 2020. Genome wide identification and expression analysis of the SBP gene family in foxtail millet. *J Nucl Agric Sci*, 34(7):1409-1420 (in Chinese). https://doi.org/10.11869/j.issn.100-8551.2020.07.1409

Sunkar R, Zhou XF, Zheng Y, et al., 2008. Identification of novel and candidate miRNAs in rice by high throughput sequencing. *BMC Plant Biol*, 8:25. https://doi.org/10.1186/1471-2229-8-25

Tang GL, 2010. Plant microRNAs: an insight into their gene structures and evolution. *Semin Cell Dev Biol*, 21(8): 782-789. https://doi.org/10.1016/j.semcdb.2010.07.009

Teixeira FK, Colot V, 2010. Repeat elements and the Arabidopsis DNA methylation landscape. *Heredity (Edinb)*, 105(1):14-23. https://doi.org/10.1038/hdy.2010.52

The International Brachypodium Initiative, 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 463(7282):763-768. https://doi.org/10.1038/nature08747

Voinnet O, 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell*, 136(4):669-687. https://doi.org/10.1016/j.cell.2009.01.046

Wang J, Zhou L, Shi H, et al., 2018. A single transcription factor promotes both yield and immunity in rice. *Science*, 361(6406):1026-1028. https://doi.org/10.1126/science.aat7675

Willmann MR, Poethig RS, 2007. Conservation and evolution of miRNA regulatory programs in plant development. *Curr Opin Plant Biol*, 10(5):503-511. https://doi.org/10.1016/j.pbi.2007.07.004

Wu L, Zhang QQ, Zhou HY, et al., 2009. Rice microRNA effector complexes and targets. *Plant Cell*, 21(11):3421-3435. https://doi.org/10.1105/tpc.109.070938

Xie KB, Wu CQ, Xiong LZ, 2006. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiol*, 142(1):280-293. https://doi.org/10.1104/pp.106.084475

Xie YR, Liu Y, Wang H, et al., 2017. Phytochrome-interacting factors directly suppress *MIR156* expression to enhance shade-avoidance syndrome in *Arabidopsis*. *Nat Commun*, 8:348. https://doi.org/10.1038/s41467-017-00404-y

Xu L, Yuan K, Yuan M, et al., 2020. Regulation of rice tillering by RNA-directed DNA methylation at miniature inverted-repeat transposable elements. *Mol Plant*, 13(6):851-863. https://doi.org/10.1016/j.molp.2020.02.009

Yan JP, Chia JC, Sheng HJ, et al., 2017. Arabidopsis pollen fertility requires the transcription factors CITF1 and SPL7 that regulate copper delivery to anthers and jasmonic acid synthesis. *Plant Cell*, 29(12):3012-3029. https://doi.org/10.1105/tpc.17.00363

Yang RX, Li PC, Mei HL, et al., 2019. Fine-tuning of miR528 accumulation modulates flowering time in rice. *Mol Plant*, 12(8):1103-1113. https://doi.org/10.1016/j.molp.2019.04.009

Yao SZ, Yang ZR, Yang RX, et al., 2019. Transcriptional regulation of miR528 by OsSPL9 orchestrates antiviral response in rice. *Mol Plant*, 12(8):1114-1122. https://doi.org/10.1016/j.molp.2019.04.010

Yu N, Cai WJ, Wang SC, et al., 2010. Temporal control of trichome distribution by microRNA156-targeted *SPL* genes in *Arabidopsis thaliana*. *Plant Cell*, 22(7):2322-2335. https://doi.org/10.1105/tpc.109.072579

Yue EK, Li C, Li Y, et al., 2017. MiR529a modulates panicle architecture through regulating SQUAMOSA PROMOTER BINDING-LIKE genes in rice (*Oryza sativa*). *Plant Mol Biol*, 94(4-5):469-480. https://doi.org/10.1007/s11103-017-0618-4

Zemach A, Kim MY, Silva P, et al., 2010. Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci USA*, 107(43):18729-18734. https://doi.org/10.1073/pnas.1009695107

Zhang BH, Pan XP, Cobb GP, et al., 2006. Plant microRNA: a small regulatory molecule with big impact. *Dev Biol*, 289(1):3-16. https://doi.org/10.1016/j.ydbio.2005.10.036

Zhang GY, Liu X, Quan ZW, et al., 2012. Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol*, 30(6):549-554. https://doi.org/10.1038/nbt.2195

Zhang Y, Schwarz S, Saedler H, et al., 2007. SPL8, a local regulator in a subset of gibberellin-mediated developmental processes in Arabidopsis. *Plant Mol Biol*, 63(3):429-439. https://doi.org/10.1007/s11103-006-9099-6

Zilberman D, Gehring M, Tran RK, et al., 2007. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet*, 39(1):61-69. https://doi.org/10.1038/ng1929

Zuker M, 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 31(13):3406-3415. https://doi.org/10.1093/nar/gkg595

Supplementary information

Figs. S1–S6; Tables S1–S8