An Rtf2 domain-containing protein influences pre-mRNA splicing and is essential for embryonic development in *Arabidopsis thaliana*

Taku Sasaki\(^1,2\), Tatsuo Kanno\(^2\), Shih-Chieh Liang, Pao-Yang Chen, Wen-Wei Liao, Wen-Dar Lin, Antonius J.M. Matzke and Marjori Matzke

Institute of Plant and Microbial Biology
Academia Sinica
128, Section 2, Academia Road
Nangang District
Taipei 115, Taiwan

\(^1\)Present address: Department of Integrated Genetics, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan

\(^2\)These authors contributed equally to this study
Running Title: AtRTF2 in RNA splicing and development

5 key words or phrases: alternative splicing, C2HC2 zinc finger, intron retention, Rtf2 domain, ubiquitin ligase

Corresponding author:

Marjori Matzke
Institute of Plant and Microbial Biology
Academia Sinica
128, Section 2, Academia Road, Nangang District
Taipei 115, Taiwan
Tel: +886-2787-1135
Email: marjorimatzke@gate.sinica.edu.tw

Co-corresponding author:

Antonius Matzke
Institute of Plant and Microbial Biology
Academia Sinica
128, Section 2, Academia Road, Nangang District
Taipei 115, Taiwan
Tel: +886-2787-1136
Email: antoniusmatzke@gate.sinica.edu.tw
Abstract

Alternative splicing is prevalent in plants, but little is known about its regulation in the context of developmental and signaling pathways. We describe here a new factor that influences pre-mRNA splicing and is essential for embryonic development in *Arabidopsis thaliana*. This factor was retrieved in a genetic screen that identified mutants impaired in expression of an alternatively spliced *GFP* reporter gene. In addition to the known spliceosomal component PRP8, the screen recovered *Arabidopsis* RTF2 (AtRTF2), a previously uncharacterized, evolutionarily conserved protein containing a Replication termination factor2 (Rtf2) domain. A homozygous null mutation in *AtRTF2* is embryo-lethal, indicating that AtRTF2 is an essential protein. Quantitative RT-PCR demonstrated that impaired expression of *GFP* in *atrtf2* and *prp8* mutants is due to inefficient splicing of the *GFP* pre-mRNA. A genome-wide analysis using RNA-seq indicated that 13-16% of total introns are retained to a significant degree in *atrtf2* mutants. Considering these results and previous suggestions that Rtf2 represents an ubiquitin-related domain, we discuss the possible role of AtRTF2 in ubiquitin-based regulation of pre-mRNA splicing.
Introduction

It is increasingly recognized that co-transcriptional and post-transcriptional gene regulation is comparable to transcriptional regulation in intricacy and importance. Pre-mRNA splicing is a co-transcriptional process and a major determinant of transcript abundance and complexity (Reddy et al. 2013). Constitutive splicing refers to the use of only one set of splice sites to generate a single mature mRNA. By contrast, alternative splicing occurs when variable splice sites are selected, leading to the generation of more than one processed RNA product from a single pre-mRNA. An individual gene can thus potentially encode multiple proteins, leading to a substantial increase in proteomic diversity (Chen and Manley, 2009; Syed et al. 2012; Reddy et al. 2013).

Recent work has established that alternative splicing is common in plants, affecting around 60% of intron-containing genes (Marquez et al. 2012). Alternative splicing has important roles in plant growth, development, abiotic stress tolerance, circadian rhythms and pathogen defense (Staiger and Brown, 2013). The most common outcome of alternative splicing in plants is intron retention (Marquez et al. 2012; Lan et al. 2013), which occurs when an intron fails to be spliced out of the pre-mRNA. Retained introns frequently contain premature termination codons (PTCs) that can channel the transcript into the nonsense-mediated decay (NMD) pathway. Intron retention provides a means for ‘transcriptome tuning’ (Braunschweig et al. 2014) and contributes to the post-transcriptional regulation of gene expression by reducing levels of inappropriately expressed transcripts (Kalyna et al. 2012; Ge and Porse, 2013).

Alternative splicing is subject to elaborate regulation that relies on general and specific trans-acting factors as well as cis-acting sequence elements, epigenetic modifications of DNA and chromatin, and post-translational modifications of splicing proteins (Chen and Manley, 2009; Reddy et al. 2013). However, the mechanistic roles of diverse splicing regulators and the means by which internal and external signals are conveyed to the splicing machinery are
not yet fully understood (Heyd and Lynch, 2011; Reddy et al. 2013). Moreover, given the prevalence of alternative splicing, it is likely that additional proteins contributing to this process remain to be discovered (Chen and Manley, 2009).

In this paper, we report the finding of a new factor that influences pre-mRNA splicing and is required for embryonic development in *Arabidopsis thaliana* (*Arabidopsis*). This factor was identified during a genetic suppressor screen originally intended to detect mutations that suppress the effects of the *defective in meristem silencing4-1* (*dms4-1*) mutation on RNA-directed DNA methylation (RdDM) and plant development (Kanno et al. 2010; Sasaki et al. 2012). During the course of this screen, however, it became apparent that the *GFP* reporter gene under investigation is subject to alternative splicing. Hence, in addition to authentic suppressor mutations that suppress the effects of the *dms4-1* mutation on RdDM and/or development, our screen is capable of identifying mutations in genes encoding proteins required for productive splicing of the *GFP* pre-mRNA. We describe here the identification of two splicing factors, one known and one novel, from this genetic screen.

**Materials and Methods**

**Plant materials and generation of transgenic plants**

The *sdr1-1/atrf2-1* and *sdr4-1/prp8-7* mutants were screened from an ethyl methanesulfonate (EMS) mutagenized population of the *dms4-1* mutant (Sasaki et al. 2012). Mapping of the *sdr1-1/atrf2-1* and *sdr4-1/prp8-7* mutations was carried out on F2 mapping populations using co-dominant markers as described previously (Kanno et al. 2008). Whole genome sequencing and data analysis to locate the causal mutations in these mutants were performed according to a prior protocol (Eun et al. 2011). Two T-DNA insertion alleles, *atrf2-2* (SALK_040515) and *atrf2-3* (SALK_081659), were obtained from the Arabidopsis Biological Resource Center (Alonzo et al. 2003).
For complementation tests of the defect in GFP expression in the \textit{atrf2-1} mutant, constructs encoding C-terminal HA-tagged AtRTF2 and HA-tagged AtRTF2 \textit{N} - which lacks amino acids 6-73 of an approximately 80 amino acid, plant specific N-terminal extension in the AtRTF2 protein - were cloned under the control of the native \textit{AtRTF2} promoter into a modified binary vector based on pCB302 (Xiang \textit{et al.} 1999) and introduced into the homozygous \textit{atrf2-1} mutant (hypomorphic allele) in a \textit{T} locus background (WT \textit{DMS4}, no \textit{S} locus) using the floral dip method (Clough and Bent, 1998). Transformed seedlings (T1) were selected on solid Murashige and Skoog (MS) medium containing 20 mg/l phosphinothricine (PPT). Complementation of the defect in GFP expression by the HA-tagged constructs in the \textit{atrf2-1} mutant was assessed by Western blotting using a GFP antibody in T2 or T3 progeny of selected lines.

Complementation of developmental defects conditioned by the null \textit{atrf2-2} mutation was tested in two ways. In one approach, homozygous \textit{atrf2-1} plants containing either the AtRTF2-HA or AtRTF2 \textit{N}-HA construct were crossed with plants heterozygous for the \textit{atrf2-2} null allele. The resulting F1 progeny were allowed to self-fertilize, producing F2 progeny. Normal-looking F2 seedlings were genotyped for the T-DNA insertion in the \textit{atrf2-2} mutant. In a second approach, AtRTF2-GFP and AtRTF2 \textit{N}-GFP constructs were introduced into the heterozygous \textit{atrf2-2} mutant using the floral dip method. T1 plants (selected by their resistance to PPT) were allowed to self-fertilize to produce T2 progeny. Normal-looking T2 progeny were genotyped for the T-DNA insertion in the \textit{atrf2-2} mutant. Primers used for genotyping are shown in Table S1. In both cases, successful complementation by a particular transgene construct was determined by recovery of normal-looking progeny that were homozygous for the \textit{atrf2-2} allele.

\textbf{Fluorescence microscopy}
Fluorescence images of GFP expression in roots of PPT-resistant seedlings expressing the AtRTF2-GFP and AtRTF2 ΔN-GFP fusion genes under the control of the native AtRTF2 promoter were made using a Leica TCS LSI-III Confocal Microscope System.

**DNA methylation analysis**

DNA methylation analysis of the target enhancer by bisulfite sequencing was carried out as described previously (Kanno et al. 2008; Daxinger et al. 2009; Sasaki et al. 2012, 2014). Genomic DNA was isolated using a Plant Genomic DNA Purification Kit (GeneMark, Taiwan). For bisulfite sequencing, 1 μg of genomic DNA was digested with HindIII. Purified DNA was converted using an EpiTect Bisulfite Kit (Qiagen). PCR products using primers listed in Table S1 were cloned with pGEM T-Easy Vector System (Promega) and transformed into competent E. coli cell. Sixteen to twenty independent clones were sequenced. Exon 15 of the PHAVOLUTA gene was used as a control for complete bisulfite conversion (primers in Table S1) (Daxinger et al. 2009).

**Whole-genome bisulfite sequencing**

Approximately 3 μg of genomic DNA was sonicated to approximately 250 bp before it was ligated to Illumina adaptors, then size-selected, denatured and treated with sodium bisulfite (BS) to reveal their methylation status. The BS-seq libraries were sequenced using the Illumina HiSeq 2000 platform to generate up to 100 cycles in paired ends. The reads were aligned to the reference genome (TAIR10) using BS Seeker 2 (Guo et al. 2013). To profile genome-wide DNA methylation, the methylation level for each covered cytosine in the genome is estimated as #C/(#C + #T), where #C represents the number of methylated reads and #T corresponds to the number of unmethylated reads. The methylation level per cytosine serves as an estimate of the percentage of cells containing methylation at this cytosine. The raw reads and the processed data set for the new methylomes (dms4-1, sdr1-1, and dms4-1...
sdrl-1) are publicly available from NCBI GEO under accession GSE63238. The raw reads and the processed dataset for the wild-type methylome (Sasaki et al. 2014) can be downloaded from NCBI GEO under accession GSE47453.

**Western blotting**

Protein extraction, SDS PAGE and Western blotting to detect GFP and tubulin proteins were carried out according to published procedures (Eun et al. 2011; Sasaki et al. 2012).

**RT-PCR and Quantitative RT-PCR**

Total RNAs were isolated from approximately three-week-old seedlings using a Plant Total RNA Miniprep Purification Kit (GeneMark, Taiwan) and treated with RQ1 DNase (Promega) according to the manufacturer's instructions. cDNA was synthesized using Transcriptor First Strand cDNA Synthesis Kit (Roche) following the manufacturer’s protocol using an oligonucleotide d(T) primer and 1 g of total RNA. One microliter of cDNA was used as a template for RT-PCR. The PCR conditions for detecting GFP transcripts were as follows: 94 °C for 2 min followed by 30 cycles of 94 °C for 10 s, 58°C for 20 s, and 72 °C for 2 min, and finally 72 °C for 7 min.

Quantitative RT-PCR was performed using a 7500 Real time PCR System (Applied Biosystems) with the program recommended by the manufacturer using 1 μl of cDNA as a PCR template and SYBR Green PCR Master Mix (Applied Biosystems). To validate intron retention events identified from the RNAsseq data, several endogenous genes were selected on the basis of their p-values (Table S2) for quantitative RT-PCR analysis. Stably expressed At5G60390 was used for normalization (Wang et al. 2014). Three biological replicates were carried out for each sample. Error bars indicate standard error. Primer sets for RT-PCR and quantitative RT-PCR are shown in Table S1.
**5’ RACE and cloning of GFP transcripts**

To determine the transcription start site for each GFP transcript, cDNA synthesis, 5’ RACE, and cloning the RACE product into the vector were carried out using SMARTer RACE 5’/3’ Kit (Clontech) according to the manufacturer’s instructions using 1ug of total RNA as starting material. The PCR conditions for the first amplification were 35 cycles of 94°C for 30 s, 68°C for 30 s, and 72°C for 3 min, and for the nested PCR were 20 cycles of 94°C for 30 s, 68°C for 30 s, and 72°C for 2 min. At least three clones for each cDNA (long, middle, short) were sequenced. Gene specific primers are listed in Table S1.

**Library preparation and RNA sequencing**

Total RNA was extracted from approximately 14 day-old seedlings of the T line, sdr1-1/atrtf2-1 and sdr4-1/prp-8-7 mutants (T background) and from newly germinated seedlings of the heterozygous (normal phenotype) and homozygous (arrested development) atrtf2-2 mutant using a Plant Total RNA Miniprep Purification Kit (TR02; GeneMark, Taiwan). The protocol with RNA lysis Solution B was used.

Libraries for RNA-seq were prepared following Illumina TruSeq stranded mRNA sample preparation kit (RS-122-2103) according to the manufacturer’s protocol. Briefly, 4 ug of total RNA were used for library construction. PolyA RNA was captured by oligodT beads and fragmented after elution from the beads. The first-strand cDNA was synthesized by reverse transcriptase (SuperScrip III, 18080-093, Invitrogen) using dNTP and random primers. The second-strand cDNA was generated using a dUTP mix. The double-stranded cDNA was subjected to the addition of a single ‘A’ base to the 3’ end followed by ligation of the barcoded Truseq adapters. Finally, the products were purified and enriched with 12 cycles of PCR to create the final double-stranded cDNA library. A final size selection was performed by 2% low-range agarose (161-3107, Bio-Rad) gel electrophoresis to yield a library of inserts.
250-350 bases in length. The library was extracted from the agarose gel using MinElute PCR purification kit (28004, QIAGENE). Final libraries were analyzed using Agilent High Sensitivity DNA analysis chip (5067-4626, Agilent) to estimate the quantity and check size distribution, and were then quantified by qPCR using the KAPA Library Quantification Kit (KK4824, KAPA). The prepared library was pooled for paired-end sequencing using IlluminaHiSeq 2500 at YOURGENE BIOSCIENCE Co. (New Taipei City, Taiwan) with 126 bp paired-end sequence reads. Approximately one hundred million reads per sample were sequenced. All RNAseq FASTQ files are available from the Sequence Read Archive (SRA) at NCBI under the following accession numbers: T line (SRR1652313); atrf2-1 (SRR1652314); prp8-7 (SRR1652315); atrf2-2 heterozygous (SRR1652316); atrf2-2 homozygous (SRR1652317).

**Detection of intron retention events**

Reads from RNA sequencing were mapped to the TAIR10 genome using BLAT (Kent, 2002) with the default setting. Under a threshold of 95% mapping identity, around 92% of reads per sample were accepted. RackJ (http://rackj.sourceforge.net/) was then used to compute average depths of all exons and all introns.

Given a control sample and a mutant sample, the preference of an intron retention event was measured using chi-squared test for goodness-of-fit by comparing the depths of the intron in the two samples taking the depths of neighboring exons in the two samples as the background. Here, depth was defined as total covering read bases divided by intron (or exon) length. In so doing, an intron has been retained in a higher chance than in the other sample would be inferred and an intron retention event was identified as significantly increased if its P-value based on the chi-squared value was less than or equal to 0.05 and the ratio intron depth/exon_depth in mutant was larger than that in control. See Table S2 for the comparison tables. Selected examples of intron retention were chosen for validation by quantitative RT-PCR according to the procedure described above. Primers used are shown in Table S1.
Protein immunoprecipitation (IP)

Total proteins were extracted from two-week-old seedlings of T, AtRTF2-GFP and AtRTF N-GFP transgenic plants. The T line serves as a negative control to eliminate non-specific interactions with GFP. Five grams of seedlings were ground in liquid nitrogen. After grinding, 10 ml of IP buffer [50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5mM MgCl$_2$, 10% glycerol, 0.1% NP-40, 2mM DTT, 1mM PMSF, 0.7 $\mu$g/ml pepstatin A, 10 $\mu$g/ml MG132, and protease inhibitor (Roche)] were added to extract proteins. The solution was centrifuged twice at 9,000 g for 10 min at 4°C to remove insoluble cell debris. The supernatant comprised the protein extract that was used for immunoprecipitation.

One hundred microliters of GFP-trap magnetic beads (ChromoTek) were washed three times with 1 ml of IP buffer and then added to the 10 ml protein extract. The whole solution was incubated at 4°C overnight with 20 rpm agitation. Following this step, the beads were collected using a magnetic stand and washed three times with 1 ml of IP buffer. The bound proteins were eluted by boiling in 30 µl of 4% SDS for 10 min.

Mass-spectrometry analysis

A linear ion trap mass spectrometer (Q Exactive MS, Dionex nanoUHPLC, Thermo Scientific) combined with an on-line nano-scale high performance liquid chromatography system (nanoACQUITY UPLC, Waters Corp., MA U.S.A) was used for protein identification and analysis. The liquid chromatography (LC) system consists of an autosampler and a binary pump, a C18 trap (Symmetry C18, 180 µm x 2.0 cm, 5 µm, Waters Corp.) and a C18 nanoACQUITY UPLC column (BEH130 C18, 75 µm x 10 cm, 1.7 µm, Waters Corp.). The linear gradient applied to separate tryptic peptides was from 5% to 40% of acetonitril in 0.1% (v/v) formic acid. LC was run for 90 min at nanoflow (300 nl/min) rate. The C18 reverse phase column was coupled to a nano-electrospray ionization (nESI) source.
Prior to LC/MS/MS analysis, the affinity purified proteins were digested with trypsin according to the filter-aided sample preparation method (Wiśniewski et al. 2009). Four microliters of the resulting peptide solution was loaded onto the column. The data acquisition was performed with a full MS scan followed by four MS/MS scans. The MS scan range was over the mass-to-charge (m/z) from 400 to 1800 using the data-dependent data acquisition mode with dynamic exclusion enabled. The MS/MS data were converted to peak list files with extract_msn script (Thermo Fisher Scientific) and then analyzed by a Mascot search program (Matrix Science Inc., Boston MA) against the *Arabidopsis* database. Proteins with more than two matching peptides were retained, and the proteins with peptide score over 10 and protein score over 20 were considered as reliable hits. Only the proteins with unique peptides identified were included in further analysis. The results shown in Supplemental Table S3 are derived from three independent biological replicates.

**Results**

**DMS4 suppressor screen**

We previously used a two-component transgene silencing system - Target plus Silencer (T+S) – to identify factors required for RdDM and transcriptional gene silencing (TGS) of a GFP reporter gene under the control of an upstream enhancer active in meristem regions (Figure 1). Forward genetic screens using this system retrieved thirteen defective in meristem silencing (dms) mutants that are deficient in Pol V-associated components of the RdDM pathway (Eun et al. 2012; Matzke and Mosher, 2014). One mutant, *dms4*, which is disabled in a putative IWR1 (interacts with RNA polymerase II) transcription factor, is unique in displaying defects not only in RdDM/TGS but also plant development (Kanno et al. 2010).

To dissect the roles of DMS4 in RdDM and development, we initiated a genetic suppressor screen using a mutant harboring the *dms4-1* allele (Sasaki et al. 2012). One type of suppressor mutation anticipated in this screen restores RdDM/TGS but not normal
development (sdr, suppressor of dms4, RdDM) (Figure S1). In addition to bona fide sdr mutants displaying the expected phenotypes (restoration of GFP silencing and enhancer methylation but not normal development), we identified two unusual mutants, sdr1-1 and sdr4-1. Although these two mutants were GFP-negative and resembled phenotypically the dms4-1 mutant (Figure 2A, B), the apparent silencing of the GFP reporter gene occurred without increases in DNA methylation at the target enhancer (Figure 2C). Moreover, subsequent genetic crosses revealed that the sdr1-1 and sdr4-1 mutations were able to impair GFP expression in a DMS4 wild-type background lacking the S locus (Figure 3A) and without an increase in DNA methylation (Figure 3B, sdr1-1 shown only). Thus, the sdr1-1 and sdr4-1 mutations diminish expression of the GFP reporter gene at the T locus in a manner that is independent of DNA methylation, the dms4-1 mutation, and the S locus that triggers RdDM.

**SDR1 is an Rtf2 domain-containing protein**

Using classical mapping with co-dominant markers, we mapped the sdr1-1 mutation to the bottom arm of chromosome 5. Subsequent Illumina whole genome sequencing revealed a G to A nucleotide change (chr5: 23,485,838) in the gene At5g58020. This gene encodes a 354 amino acid protein that contains an Rtf2 (Replication termination factor 2) domain (Pfam PF04641), which is defined by a C2HC2 motif related to the C3HC4 RING-finger motif (Inagawa et al. 2009) (Figure 4A). Rtf2 was discovered in fission yeast, where it is needed to stabilize a paused DNA replication fork to establish imprinting at the mating type locus (Inagawa et al. 2009). Although Rtf2 proteins are found in eukaryotes ranging from fission yeast to humans (Figure 4B), the Rtf2 orthologs in plants have an N-terminal extension of approximately 80 amino acids (Figure 4C). In contrast to the high amino acid sequence similarity in the Rtf2 domain among different plant species, the N-terminal extension is only weakly conserved at the amino acid sequence level (Figure 4C). The sdr1-1 mutation results
in substitution of glycine for glutamic acid at position 85 (G85E), which is located at the beginning of the conserved Rtf2 domain (Figure 4A, C). We will refer to the SDR1 protein hereafter as AtRTF2 and the sdr1-1 mutation as atrf2-1.

To complement the atrf2-1 mutation and test a functional requirement for the N-terminal extension, we generated constructs encoding HA-tagged versions of full-length AtRTF2 (AtRTF2-HA) and a truncated form lacking most of the N-terminal region (AtRTF2 ΔN-HA) under the control of the endogenous AtRTF2 promoter (Figure 5A). These constructs were introduced into the atrf2-1 mutant containing the T locus. As assessed by GFP protein accumulation using Western blotting, AtRTF2-HA but not AtRTF2 ΔN-HA complemented the atrf2-1 mutation and restored wild-type levels of GFP expression (Figure 5B). These results demonstrate that full-length AtRTF2 including the N-terminal segment is required for correct GFP expression.

Two transfer-DNA (T-DNA) insertion alleles, atrf2-2 and atrf2-3, were obtained from a public stock center. Both strains harbor T-DNA insertions within the Rtf2 motif (Figure 4A). When homozygous, the putative null atrf2-2 allele conditions defects in embryogenesis and lethality shortly after germination (Figure 6A, B), confirming that AtRTF2 is an essential gene (Savage et al. 2013). By contrast, the other two alleles, atrf2-1 and atrf2-3, are hypomorphic and display only mild developmental phenotypes. The developmental defect in atrf2-2 was complemented by transgene constructs encoding full-length AtRTF2 but not the truncated version lacking most of the N-terminal extension (Table 1). These results demonstrate that full-length AtRTF2 is essential for both normal development and for proper expression of the GFP reporter gene.

Although AtRTF2 has been considered a plastid-targeted protein (Savage et al. 2013) we found that an AtRTF2-GFP fusion protein under the control of the native promoter accumulates predominantly in the nucleus (Figure 6C). Expression of AtRTF2 is ubiquitous
and particularly strong in developing embryos during seed maturation (*Arabidopsis* eFP Browser, Winter *et al.* 2007).

**SDR4 is the core spliceosomal protein PRP8**

Using classical mapping with co-dominant markers, we mapped the *sdr4-1* mutation to a genetic interval on the bottom arm of chromosome 1. Subsequent Illumina whole genome sequencing region revealed a G to A mutation (chr1: 30,125,295) in the gene At1g80070. This gene encodes the core splicing factor PRP8 (pre-mRNA processing 8), which is one of the largest and most highly conserved proteins of the spliceosome (Grainger and Beggs, 2005). We will refer hereafter to SDR4 as PRP8 and the *sdr4-1* mutation as *prp8-7* (**Figure 7**). The mutation we recovered results in the substitution of a highly conserved glycine residue at position 1820 to glutamic acid (G1820E), which is in the RNase H domain of the PRP8 protein (**Figure 7**). Similarly to *AtRTF2*, PRP8 is expressed ubiquitously and shows particularly strong expression during seed maturation (*Arabidopsis* eFP Browser, Winter *et al.* 2007). Homozygous null *prp8* mutations result in an abnormal suspensor and embryo-lethality (Schwartz *et al.* 1994). Plants homozygous for the *prp8-7* mutation grow and reproduce normally, although they are somewhat late flowering, indicating that the *prp8-7* allele is hypomorphic.

**AtRTF2 and PRP8 are required for splicing of GFP pre-mRNA**

*AtRTF2* was identified in the same screen as PRP8, a known splicing factor, and *GFP* expression is impaired in a DNA methylation-independent manner in both *atrtf2-1* and *prp8-7* mutants. Moreover, both genes have similar expression patterns and null mutations are embryo-lethal. These observations suggest that *AtRTF2* and PRP8 act in the same pathway, prompting us to test for defects in *GFP* pre-mRNA splicing in the *atrtf2-1* and *prp8-7* mutants.
We had originally predicted from the structure of the T construct that GFP transcription would initiate in the minimal promoter upstream of the GFP coding sequence (Figure 1). In practice, however, this minimal promoter does not appear to be used and we detected instead three major GFP transcripts that are likely to result from alternative splicing of a pre-mRNA initiating upstream in the distal enhancer region (Kanno et al. 2008) (Figure 8A; Figure S3). The ‘short’ transcript, which results from productive splicing of a cryptic intron with non-canonical donor and acceptor sites (AT-AC), can be translated into GFP protein using the first methionine codon after the transcription start site (Figure S4). Mutational analysis confirmed that the short transcript is indeed the major GFP mRNA (Figure S5). By contrast, the ‘long’ unspliced transcript and the ‘middle’ transcript, which results from unproductive splicing of a conventional GT-AG intron (Figure 8A; Figure S3), contain a number of PTCs after the initiating methionine and cannot be translated into a functional GFP protein (Figure S4).

Consistent with a role for AtRTF2 and PRP8 in splicing the GFP pre-mRNA, the ratio of the short translatable and long untranslatable transcripts varied in wild-type and mutant plants. In GFP-positive plants (T line and atrtf2-1 mutant complemented with the wild-type AtRTF2-HA sequence; Figure 8B, lanes 1 and 4), the short translatable transcript was prominent. By contrast, in GFP-negative plants (atrtf2-1 and prp8-7 mutants, atrtf2 mutant complemented with the truncated AtRTF2ΔN-HA construct; Figure 8B, lanes 2, 3 and 5, respectively), the long untranslatable transcript was the predominant form. AtRTF2 and PRP8 are thus required to splice the AT-AC intron to generate a translatable GFP mRNA. Quantitative RT-PCR confirmed the different ratios of short and long transcripts in wild-type plants and the atrtf2-1 mutant (Figure 8C).

*Genome-wide requirement for AtRTF2 in splicing*
To determine whether \textit{atrtf2} mutations affect splicing of introns genome-wide, we performed RNA-seq on total RNA extracted from seedlings of the \textit{T} line and homozygous \textit{atrtf2-1} and \textit{prp8-7} mutants, as well as newly germinated embryos that were either heterozygous (normal) or homozygous (arrested; \textbf{Figure 6A}) for the \textit{atrtf2-2} null mutation. Because the main splicing defect observed for the \textit{GFP} reporter gene in the \textit{atrtf2-1} and \textit{prp8-7} mutants was intron retention (IR), which is also the most common outcome of alternative splicing in plants (Marquez \textit{et al.} 2012), we focused our analysis of the RNA-seq data on IR events. In \textit{atrtf2-1}, \textit{atrtf2-2} and \textit{prp8-7} mutants, 13.6\%, 15.7\% and 6.7\% of total introns, respectively, displayed a significant degree of increased intron retention (\textbf{Fig. 9A}). Both major U2 and minor U12 introns were affected to a comparable degree. A 62.1\% overlap in IR events was observed between \textit{atrtf2-1} and \textit{atrtf2-2} using the total number of IRs in \textit{atrtf2-2} as a denominator, and around 38.6\% overlap in IRs was observed between \textit{atrtf2} mutants and \textit{prp8-7} using the total number of IRs in \textit{prp8-7} as a denominator (\textbf{Figure 9A}; \textbf{Table S2}).

Following the definition of IR fold change in a previous publication (Lan \textit{et al.} 2013), the distributions of fold changes of significantly increased IR events follow the power-law distribution, which shows linear distributions in the log-log scale (\textbf{Table S2, sheet 5}), where the maximum fold changes were 3291.5, 10127.3 and 2718.1 for \textit{atrtf2-1}, \textit{atrtf2-2} and \textit{prp8-7}, respectively. The median fold changes of the three samples were all above five, which means that more than half of significantly increased IR events showed fold changes greater than five in each mutant. Additionally, more than 98\% of significantly increased IR events represent introns fully covered by reads of the sample, suggesting that almost all increased IR events retain full introns. Our method identifies significantly increased and reduced IR events equally. However, in the mutants, increased IRs predominate among significant IR events (16467 vs 919 for \textit{atrtf2-1}; 19031 vs 874 for \textit{atrtf2-2}; and 8124 vs 1875 for \textit{prp8-7}), indicating that the mutations in \textit{AtRTF2} and \textit{PRP8} are associated with reduced splicing efficiency. \textbf{Table S2, sheet 6} gives a summary table of significantly increased IR events.
Several cases of IR events affecting endogenous genes in the *atrtf2-1* and *prp8-7* mutants were validated using qRT-PCR (Figure 9B).

**AtRTF2-interacting proteins**

To initiate studies on proteins that may associate with AtRTF2 in a complex, we carried out affinity purification of the AtRTF2-GFP fusion protein followed by mass spectrometry. This analysis revealed a number of predicted and known splicing factors, including PRP8, as well as other proteins not directly related to splicing (Table S3).

**Discussion**

Our study has identified a new factor, AtRTF2, which influences pre-mRNA splicing and is essential for embryo development in *Arabidopsis*. A splicing-related role was initially suggested by the identification of the hypomorphic *atrtf2-1* mutation in the same genetic screen as the hypomorphic *prp8-7* mutation, which impairs the activity of the core spliceosomal constituent PRP8. Further work showed that the *atrtf2-1* and *prp8-7* mutations have identical effects on the splicing pattern of *GFP* pre-mRNA: productive splicing of a cryptic intron with non-canonical AT-AC termini is less efficient in the two mutants, leading to lower levels of a translatable *GFP* mRNA and increased accumulation of an unspliced, untranslatable *GFP* transcript. In addition, mutations in *atrtf2* disrupt splicing genome-wide, leading to a significant degree of increased retention for approximately 13-16% of total introns. Mass spectrometry-based profiling suggests that AtRTF2 potentially associates with several predicted and known splicing factors, including PRP8, although these remain to be substantiated using additional approaches. Collectively, the findings implicate AtRTF2, which was functionally uncharacterized prior to our study, in pre-mRNA splicing.

In contrast to PRP8, which acts directly in the splicing reaction by providing a scaffold for spliceosome assembly as well as amino acids for catalysis (Chen and Moore, 2014),
AtRTF2 may have a more transient regulatory role during the spliceosome cycle. One possibility is that AtRTF2 contributes to ubiquitin-based modulation of spliceosomal proteins. The Rtf2 domain consists of a C2HC2 zinc finger that is related to the C3HC4 RING-finger motif but folds in a way to create only one functional Zn$^{+2}$ ion binding site. The founding member of the Rtf2 protein family was discovered in fission yeast, where it is important for stabilizing a paused DNA replication fork during imprinting at the mating type locus, possibly by facilitating sumoylation of PCNA (Inagawa et al. 2009; Komander and Rape, 2012).

With regard to pre-mRNA splicing, the Rtf2 domain has been described as an ubiquitin-related domain in a structural bioinformatics analysis of splicing factors (Korneta et al. 2012). Another example noted in the study of Korneta and coworkers is NOSIP (nitric oxide synthase interacting protein) (CG6179) (Korneta et al. 2012), which contains an Rtf2 domain and is a component of the Drosophila melanogaster spliceosome (Herold et al. 2009). The Rtf2 domain has been annotated in association with RING E3 ubiquitin ligases (Choy et al. 2013) and suggested to act as an ubiquitin ligase in the context of splicing (Korneta et al. 2012). A divergent cyclophilin that may be involved in splicing also contains an Rtf2 domain (Page and Winter, 1998). In addition to other reversible post-translational modifications, such as acetylation, methylation and phosphorylation, ubiquitination is increasingly recognized for its role in regulating the spliceosomal cycle (Bellare et al. 2005; Song et al. 2010; Mishra et al. 2011; Korneta et al. 2012; Oka et al. 2014; Ammon et al. 2014; Chen and Moore, 2014). Notably, PRP8, which can bind ubiquitin through its conserved JAMM (JAB1/MPN/Mov34 metalloenzyme) domain, was detected as an ubiquitin conjugate in affinity-purified particles in budding yeast, suggesting a means to reversibly modulate the activity of this protein (Bellare et al. 2008). A recent advanced proteomics analysis identified PRP8 as a target of ubiquitination in Arabidopsis (Kim et al., 2013). Further work is needed to determine whether AtRTF2 has ubiquitin ligase activity, and if so, whether PRP8 and other splicing factors are among its substrates.
The embryonic lethality of null atrtf2-2 mutations is consistent with disruptions in splicing of key transcripts important for early stages of development. Further work is required to understand in more detail the nature of the splicing defects and their impact on development. Even the hypomorphic atrtf2-1 mutation decreases the efficiency of intron removal in more mature plants, demonstrating that AtRTF2 is required continuously during plant growth to maintain optimal splicing activity.

In our system, the AtRTF2-dependent, productive splicing event excises a cryptic intron with non-canonical AT-AC sites in the GFP pre-mRNA. Although AT-AC termini are a feature of U12 introns removed by the minor spliceosome, the AT-AC intron in the GFP pre-mRNA lacks additional highly conserved U12 intron sequences at the 5’ splice site and branch point (Burge et al. 1998; Lin et al. 2010; Turunen et al. 2013). Therefore, this intron is likely to be processed primarily by the major U2 spliceosome (Wu and Krainer, 1997). AtRTF2 is not specialized for a particular class of intron because our genome-wide analysis of intron retention found that atrtf2 mutations affect splicing of both U2 and U12 introns to a similar extent. Although the PTC-containing long and middle GFP transcripts are potentially targets of NMD, they nevertheless accumulate to detectable levels. This observation is consistent with previous work indicating that many intron-containing transcripts are retained in the nucleus and hence not degraded by the NMD machinery, which is located in the cytoplasm (Kalyna et al. 2012).

PRP8 was identified in a previous screen for factors affecting splicing of the COOLAIR antisense transcript involved in epigenetic regulation of the FLOWERING LOCUS C (FLC) gene in Arabidopsis (Marquardt et al. 2014). Like the prp8-7 mutation we recovered (G1802E), the mutation identified by Marquardt and coworkers, prp8-6 (G1891E), is also present in the RNaseH domain of PRP8 (Marquardt et al. 2014). The physiological significance of finding two distinct hypomorphic mutations in the RNase H domain of PRP8
in independent screens is not yet clear, but these mutations should prove useful for further dissecting PRP8 function in the plant spliceosome.

Although our GFP reporter gene system illuminates a role for AtRTF2 in pre-mRNA splicing, the function of this protein may not be limited to this process. Additional roles are suggested by the identification of proteins not known to be involved in splicing in the affinity purification-mass spectrometry analysis. Moreover, the Rtf2 protein in fission yeast is involved in an activity unrelated to pre-mRNA splicing (Inagawa et al. 2009). The plant-specific N-terminal extension is essential for AtRTF2 function for reasons that are not yet known. This extension may interact with certain factors or be modified in a way that is important for the regulation of AtRTF2 activity or stability. Given the evolutionary conservation of the RTF2 protein and the presence of the Rtf2 domain in several splicing proteins (Korneta et al. 2012), it will be interesting to determine the degree to which AtRTF2 orthologs and other Rtf2 domain-containing proteins are involved in the regulation of pre-mRNA splicing in different organisms.

The identification of AtRTF2 in our screen demonstrates the usefulness of the alternatively spliced GFP reporter gene for uncovering novel proteins involved in pre-mRNA splicing. To retrieve additional components acting in the AtRTF2 pathway, we recently initiated a new forward genetic screen based on the T line and recovered a number of putative mutants in which the wild-type GFP gene is silenced or only weakly expressed. The identification of the causal mutations conditioning weak GFP expression in these mutants has the potential to unveil more new splicing factors and provide mechanistic insights into the regulation of splicing efficiency and alternative splicing in plants.

Acknowledgements

We are grateful to Academia Sinica and the Taiwan Ministry of Science and Technology (NSC Project number MOST 103-2311-B-001-004-MY3) for financial support. TS was
supported by the Japan Society for the Promotion of Science Postdoctoral Fellowships for Research Abroad. We thank David Meinke for helpful discussions, the Proteomics Core Lab of the Institute of Plant and Microbial Biology (IPMB) at Academia Sinica (AS), for mass spectrometry analysis, the DNA Microarray Core Laboratory (IPMB, AS) for library preparation for RNA sequencing, and the sequencing services provided by the National Center for Genome Medicine of the National Core Facility Program for Biotechnology, Ministry of Science and Technology, Taiwan.

References


Grainger, R.J. and J.D. Beggs, 2005 Prp8 protein: at the heart of the spliceosome. RNA 1: 533-557.


Kent, W.J., 2002 BLAT--the BLAST-like alignment tool. Genome Res. 12:656-64.


**Figure Legends**

**Figure 1. T+S transgene silencing system to study RdDM**

The two-component transgene silencing system consists of a Target locus, \( T \), which contains a *GFP* reporter gene downstream of a minimal promoter (narrow gray bar) and an upstream virus-derived enhancer (containing a short tandem repeat) that drives *GFP* expression in shoot and root meristem regions. The Silencer locus, \( S \), contains an inverted DNA repeat of distal enhancer sequences (black arrowheads corresponding to thick black bar in \( T \)) that is transcribed by RNA polymerase II (Pol II) from a constitutive viral promoter (35S). The resulting hairpin RNA is processed by DICER-LIKE3 (DCL3) to produce 24-nt small interfering RNAs that induce Pol V complex-mediated *de novo* DNA methylation (blue ‘m’) of the target enhancer region, leading to transcriptional silencing of *GFP* expression (Kanno *et al.* 2008, 2010; Sasaki *et al.* 2014). Figure not drawn to scale.

**Figure 2. Phenotypes of sdr1 and sdr4 mutants**

A. *GFP* expression is silenced in ‘wild-type’ (WT) T+S seedlings whereas silencing is released in the *dms4-1* mutant, which is GFP-positive. *GFP* silencing appears to be re-established in the *dms4-1 sdr1-1* and *dms4-1 sdr4-1* double mutants, which are GFP-negative (all mutations in T+S background).

B. The *dms4-1 sdr1-1* double mutant displays delayed growth and other phenotypic features of the single *dms4-1* mutant compared to the age-matched WT control. The *dms4-1 sdr4-1* double mutant also appears similar to the *dms4-1* single mutant (not shown).

C. Bisulfite sequencing of the target enhancer demonstrates heavy cytosine methylation in all sequence contexts (black, CG; blue CHG; red, CHH; \( H \) is A, T or C) in WT T+S plants, which are GFP-negative (part A). Release of *GFP* silencing in the *dms4-1* mutant is associated with substantial loss of methylation. The double mutants, *dms4-1 sdr1-1* and *dms4-1 sdr4-1*...
l sdr4-1, appear GFP-negative (part A) but this is not accompanied by restoration of the WT DNA methylation level in the target enhancer.

**Figure 3. Impaired GFP expression in sdr1 and sdr4 mutants in the T line**

A. The wild-type T line expresses GFP in the shoot and root meristem regions. GFP expression is impaired in sdr1-1 and sdr4-1 seedlings (T background only, WT DMS4, S locus absent).

B. Bisulfite sequencing of the target enhancer demonstrated that the impairment of GFP expression in the sdr1-1 single mutant is not accompanied by any significant DNA methylation at the target enhancer (right). The unmethylated enhancer in SDR1 wild-type plants containing the T locus is shown as a control (left). The sdr1-1 mutation also has no effect on DNA methylation genome-wide as indicated by a methylome analysis (Figure S2).

**Figure 4. SDR1 is an evolutionarily conserved Rtf2 domain-containing protein**

A. SDR1 (At5g58020), which is 354 amino acids in length, contains an Rtf2 domain (amino acids 84-338) and hence renamed here AtRTF2. AtRTF2 is a single copy gene in Arabidopsis. The position of the G85E amino acid substitution resulting from the sdr1-1/atrtf2-1 point mutation identified in our screen and the sites of two T-DNA insertion alleles (atrtf2-2 and atrtf2-3) are indicated.

B. Phylogenetic tree of Rtf2 orthologs in different organisms. With the exception of budding yeast, Rtf2 orthologs are present in other eukaryotes examined. An unrooted phylogenetic tree was generated by the neighbor-joining method using ClustalW and visualized with TreeView.

C. Amino acid sequence alignments of AtRTF2 orthologs in plants shows high similarity in the Rtf2 domain but only partial conservation in the plant-specific N-terminal extension, which is approximately 80 amino acids in length. The red arrowhead indicates the location of the sdr1-1/atrtf2-1 G85E mutation at the beginning of the Rtf2 motif. The multiple sequence
alignment by ClustalW was performed using GenomeNet (http://www.genome.jp/tools/clustalw/) and consensus amino acid residues were highlighted using BoxShade (http://www.ch.embnet.org/software/BOX_form.html).

**Figure 5. Full length AtRTF2 is required for proper GFP expression**

A. Schematic drawing of C-terminal HA-tagged constructs encoding full-length AtRTF2 (ATRTF2-HA) or a truncated version lacking amino acids 7-63 of the N-terminal extension (ATRTF2 N^{7-63}-HA). These constructs were introduced into the homozygous atrtf2-1 mutant.

B. Western blots probed with a GFP antibody revealed little accumulation in the T+S line, in which GFP expression is silenced (Figure 2A) but wild-type accumulation in the T line. GFP accumulation is low in the atrtf2-1 mutant but returns to the wild-type level when the mutant is complemented with the full length AtRTF2-HA construct (2 examples shown). Levels of GFP protein remain low when using the AtRTF2 N-HA construct in the complementation test (2 examples shown). RT-PCR (bottom) confirmed that the HA-tagged transgenes are transcribed. The blot was probed with an antibody to tubulin as a control for protein stability and loading levels. The stained membrane is also shown as a loading control. Abbreviation: gDNA, genomic DNA (shown for plants without and with the AtRTF2-HA transgene).

**Figure 6. AtRTF2 encodes an essential nuclear protein**

A. Self-fertilized plants heterozygous for the null atrtf2-2 mutation (Figure 4A) produce approximately 25% defective seedlings (arrowheads, top, and enlargement, bottom), which are homozygous for the atrtf2-2 mutation. These seedlings die shortly after germination. The developmental defect can be complemented by full-length AtRTF2 transgenes but not by truncated AtRTF2 N versions lacking most of the plant-specific N-terminal extension (Table 1).
B. A seed pigment defect is visible in siliques of heterozygous atrf2-2 mutant plants (Savage et al. 2013). In our experiment, around 10-14 days after flowering, approximately 25% of the seeds (196/831 counted) appeared white (red arrowheads) whereas 100% of the seeds in an age-matched wild-type control (882/882 counted) were green.

C. AtRTF2-GFP (left) and AtRTF2 ΔN-GFP (right) fusion proteins (constructs below) are located predominately in nuclei (shown here in root tip cells).

Figure 7. SDR4 is PRP8

Intron-exon structure of the PRP8 gene (At1g80070) (top) and domain structure of PRP8 (bottom), a core spliceosomal protein of 2359 amino acids. The sdr4-1/prp8-7 mutation (G to A at position 30,125,295 on chromosome 1) creates a G1820E amino acid substitution in the RNase H-like domain. A second point mutation in this region, prp8-6 (G1891E), was reported recently (Marquardt et al. 2014). PRP8 domains were identified in Pfam (http://pfam.xfam.org/).

Figure 8. Alternative splicing of GFP pre-mRNA

A. As shown by cDNA cloning and sequencing, the T locus encodes three polyadenylated GFP transcripts that result from alternative splicing: a ‘long’ unspliced transcript; a ‘middle’ transcript that results from splicing a canonical GT-AG intron; and a ‘short’ transcript that results from splicing a cryptic intron with non-canonical AT-AC splice sites. Although AT-AC termini are a feature of U12 introns removed by the minor spliceosome, this intron does not contain the highly conserved U12 intron 5’ splice site or branch point sequence and therefore is not a U12 intron (Figure S3). The long and middle transcripts are not translatable into GFP protein (GFP-) owing to the presence of numerous PTCs after the initiating methionine (Figure S4). The short transcript can be translated into GFP protein (GFP+) (Figure S4) and mutational analysis indicates that it is indeed the major transcript encoding
GFP protein (Figure S5). A 5’-RACE experiment demonstrated that transcription initiates in the distal enhancer region around 45 bp downstream of the short tandem repeat in the target enhancer (arrowheads) (Figure S3). ‘TATA’ indicates an apparently unused minimal promoter directly upstream of the GFP coding region (Figure 1). Arrows indicate the positions of the primers used for the amplification of the three GFP transcripts (part B), and the individual ‘long’ and ‘short’ transcripts (part C). We did not analyze in detail the ‘middle’ transcript because it is the least abundant and accumulates inconsistently.

B. Semi-quantitative RT-PCR showing accumulation of long and short GFP transcripts in the indicated genotypes. Actin is shown as a constitutively expressed control. –RT, no reverse transcriptase.

C. Quantitative RT-PCR showing accumulation of long and short GFP transcripts in the indicated genotypes. Stably expressed At5g60390 was used for normalization (Wang et al. 2014).

Figure 9. Intron retention in atrf2 and prp8 mutants

A. Venn diagrams indicate significantly increased IR events in homozygous atrf2-1 (hypomorphic allele), atrf2-2 (null allele), and prp8-7 (hyomorphic allele) mutants. Total introns (U2 and U12) are shown at the left; U12 introns only are shown to the right. The full list is provided in Table S2.

B. Validation of IR events by quantitative RT-PCR. Genes were selected from Table S2 based on p-values. Types of IR events observed are exemplified by At3g63140, At2g47940 and At4g33150 (IR in atrf2-1 only) and At1g09340, At5g16050 and At3g25690 (IR in prp8-7 only). At3g13920 is shown as a control gene that shows no statistically significant IR changes in the wild-type T line and the mutants. The Y- axis indicates the relative IR level normalized to stably expressed At5g60390 (Wang et al. 2014). The primers were designed so that one was inside the target intron and the other was in an adjacent exon. The numbers in
parenthesis after each gene ID indicate the target intron number for validation as counted from the genomic 5’ end. The error bars indicate standard error of the mean (SEM) of three independent biological replicates. Letters above each bar indicate statistical significance tested by Tukey’s HSD test (p < 0.05). The same letter (‘a’) indicates no statistically significant difference between the two samples. A different letter (‘b’) indicates a statistical difference between the two samples.

Table 1. Full length AtRTF2 complements developmental defect in homozygous atrf2-2 mutants

Supplementary files

Figure S1. dms4-1 suppressor screen

Figure S2. Genome wide analysis of DNA methylation

Figure S3. Sequence of GFP reporter gene plus upstream enhancer and GFP RNAs

Figure S4. PTCs in ‘long’ and ‘middle‘ GFP transcripts

Figure S5. Mutational analysis verifying ‘short’ transcript is the major GFP mRNA

Table S1-Primers

Table S2 – Excel file of IR retention analysis in mutants

Table S3 – Excel file of candidate AtRTF2-interacting proteins identified by affinity purification-mass spectrometry
Table 1. Full length AtRTF2 complements developmental defect in homozygous atrf2-2 mutants

**Top:** In one complementation test, homozygous atrf2-1 plants containing either the AtRTF2-HA or AtRTF2 ΔN-HA construct were crossed with plants heterozygous for the atrf2-2 null allele. The resulting F1 progeny were allowed to self-fertilize, producing F2 progeny. Normal-looking F2 seedlings were genotyped for the T-DNA insertion in atrf2-2. Approximately 25% of the normal-looking F2 seedlings recovered from the AtRTF2-HA lines were homozygous for the atrf2-2 mutation, indicating successful complementation of the developmental defect by the full length AtRTF2-HA construct. By contrast, no homozygous atrf2-2 plants were recovered from the AtRTF2 ΔN-HA lines, indicating unsuccessful complementation with the truncated construct.

**Bottom:** In a second complementation test, AtRTF2-GFP and AtRTF2 ΔN-GFP constructs were introduced into the heterozygous atrf2-2 mutant using the floral dip method. T1 plants (selected by their resistant to PPT) were allowed to self-fertilize to produce T2 progeny. Normal-looking T2 progeny were genotyped for the T-DNA insertion in atrf2-2. Consistent
with the result described above, normal-looking progeny that were homozygous for the *atrft2-2* mutation were only obtained with the construct encoding the full-length AtRTF2-GFP fusion protein.

Results of Chi-squared tests carried out to determine if the segregation ratio differs from the expected 1:2:1 ratio are shown.
Figure 1 Sasaki et al.
Figure 2 Sasaki et al.

A

WT (T+S)  
dms4-1  
dms4-1 sdr1-1  
dms4-1 sdr4-1

B

WT  
dms4-1  
dms4-1 sdr1-1

C

DNA methylation (%)  
WT (T+S)  
dms4-1 (T+S)  
dms4-1 sdr1-1 (T+S)  
dms4-1 sdr4-1 (T+S)
Figure 3 Sasaki et al.

A

B

DNA methylation (%)
Figure 4 Sasaki et al.

A

\[ \text{sdr1-1/atrtf2-1 (G85E)} \]

\[ \text{T-DNA atrtf2-2} \]

\[ \text{T-DNA atrtf2-3} \]

(B)

\[ \text{Anopheles} \]

\[ \text{Caenorhabditis} \]

\[ \text{Arabidopsis} \]

\[ \text{Oryza} \]

\[ \text{Magnaporthe} \]

\[ \text{Neurospora} \]

\[ \text{Schizosaccharomyces} \]

\[ \text{Danio} \]

\[ \text{Gallus} \]

\[ \text{Mus} \]

\[ \text{Rattus} \]

\[ \text{Homo} \]

\[ \text{Pan} \]

\[ \text{Macaca} \]

\[ \text{Bos} \]

\[ \text{Canis} \]

\[ \text{Drosophila} \]

\[ 0.1 \]

C

<table>
<thead>
<tr>
<th>A. thaliana</th>
<th>1</th>
<th>--------------</th>
<th>MHIR</th>
<th>QI</th>
<th>FO</th>
<th>V</th>
<th>KSP</th>
<th>D</th>
<th>CQKV-VA</th>
<th>L</th>
<th>QL</th>
<th>DPAQL</th>
<th>LL</th>
<th>S-</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. sitchensis</td>
<td>1</td>
<td>--------------</td>
<td>MNPK</td>
<td>S</td>
<td>D</td>
<td>LQ</td>
<td>I</td>
<td>PV</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
<td>S-</td>
<td></td>
</tr>
<tr>
<td>G. max</td>
<td>1</td>
<td>--------------</td>
<td>MRHK</td>
<td>Q</td>
<td>SLE</td>
<td>LQ</td>
<td>P</td>
<td>LFKSY</td>
<td>L</td>
<td>Q</td>
<td>D-----</td>
<td>S-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. truncatula</td>
<td>1</td>
<td>--------------</td>
<td>MRHK</td>
<td>K</td>
<td>SLE</td>
<td>LQ</td>
<td>P</td>
<td>LFKSY</td>
<td>L</td>
<td>Q</td>
<td>D-----</td>
<td>S-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. trichocarpa</td>
<td>1</td>
<td>--------------</td>
<td>MRHK</td>
<td>S</td>
<td>LQ</td>
<td>I</td>
<td>PQ</td>
<td>L</td>
<td>SP</td>
<td>L</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
</tr>
<tr>
<td>R. communis</td>
<td>1</td>
<td>--------------</td>
<td>MHTR</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>K</td>
<td>T</td>
<td>SP</td>
<td>L</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
</tr>
<tr>
<td>V. vinifera</td>
<td>1</td>
<td>--------------</td>
<td>MHTR</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>K</td>
<td>T</td>
<td>SP</td>
<td>L</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
</tr>
<tr>
<td>O. sativa</td>
<td>1</td>
<td>--------------</td>
<td>MKRAT</td>
<td>T</td>
<td>V</td>
<td>L</td>
<td>Q</td>
<td>PT</td>
<td>L</td>
<td>DK</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>H. vulgare</td>
<td>1</td>
<td>--------------</td>
<td>MAPA</td>
<td>G</td>
<td>T</td>
<td>L</td>
<td>Q</td>
<td>PT</td>
<td>L</td>
<td>DK</td>
<td>L</td>
<td>S</td>
<td>F</td>
<td>P</td>
</tr>
<tr>
<td>S. moellendorfii</td>
<td>1</td>
<td>--------------</td>
<td>MVRD</td>
<td>R</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>L</td>
<td>P</td>
<td>L</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
</tr>
<tr>
<td>P. patens</td>
<td>1</td>
<td>--------------</td>
<td>MVRD</td>
<td>R</td>
<td>Q</td>
<td>S</td>
<td>D</td>
<td>L</td>
<td>P</td>
<td>L</td>
<td>TGD</td>
<td>E</td>
<td>F</td>
<td>N-----</td>
</tr>
</tbody>
</table>

| A. thaliana       | 44 | ISF | S | AC | S | T | L | G | L | L | NG | ST | R | I | QV | SK | L | PS | V | M | T | L | F | R | RG | GGGAT | GAESR | DCY | LNM |
| P. sitchensis    | 48 | KEMG | Q | FY | F | CG | A | L | D | N | CR | L | MD | MD | V | GH | N | S | L | I | Q | L | IP | RV | C | GGGAT | GAESR | DYL | K |
| G. max           | 40 | HR | SF | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| M. truncatula    | 40 | HR | SF | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| P. trichocarpa   | 44 | QNP | SF | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| R. communis      | 43 | QNL | S | F | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| V. vinifera      | 44 | QTLD | S | F | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| O. sativa        | 45 | LES | S | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| H. vulgare       | 44 | SS | S | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| S. moellendorfii | 43 | QNL | S | F | FT | FN | GK | P | D | K | T | P | L | SQ | FP | PLST | L | V | L | Q | V | R | RL | GGGAT | GAESR | DCY | LNM |
| P. patens        | 36 | CVW | H | V | T | F | GR | P | S | D | A | GR | L | P | S | A | S | L | P | SG | SL | Q | L | RL | GAESR | DCY | LNM |
Figure 5 Sasaki et al.

A

AtRTF2-HA
AtRTF2ΔN-HA

B

T+S  T  atrtf2-1  AtRTF2-HA  Atrtf2ΔN-HA

(kDa)  55  40  35  55  40  35

Tubulin
GFP

φDNA  gDNA (+AtRTF2-HA)
Figure 6 Sasaki et al.

A

B

<table>
<thead>
<tr>
<th></th>
<th># of aborted embryos</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>0/882</td>
</tr>
<tr>
<td>sdr1-2 +/-</td>
<td>196/831</td>
</tr>
</tbody>
</table>

C

WT

sdr1-2 +/-
Figure 7 Sasaki et al.

(sdr4-1/prp8-7)

U5-snRNA binding site
U6-snRNA interacting site
PRP8 domain IV

PRO8NT - PROCN - RRM - JAB - PROCT

RNase H like
RNase H like

2359aa

(chr 1: 30,125,295)
Figure 8 Sasaki et al.

A

![Diagram showing relative mRNA levels and transcription factor binding sites.](image)

B

![Gel images showing GFP and Actin expression levels with RT and -RT controls.](image)

C

![Bar graph showing relative RNA levels with long and short transcripts.](image)
Figure 9 Sasaki et al.

A

Total introns – 120,998

U12 introns - 2069

prp8-7

atrtf2-2 (embryo-lethal)

prp8-7

atrtf2-1 (embryo-lethal)

prp8-7

atrtf2-2

B

AT1G09340 (1)

ATG16050 (2)

AT4G33150 (5)

AT3G13920 (2)

AT3G25690 (4)

AT3G63140 (5)

AT2G47940 (16)

T

atrtf2-1

prp8-7