Original Article

Tissue-specific alternative splicing of pentatricopeptide repeat (PPR) family genes in *Arabidopsis thaliana*

Umme Qulsum[1], Toshifumi Tsukahara[1,2,3,*]

[1] *School of Materials Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi City, Ishikawa, Japan;*
[2] *Area of Bioscience and Biotechnology, School of Materials Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi City, Ishikawa, Japan;*
[3] *Division of Transdisciplinary Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi City, Ishikawa, Japan.*

Summary

Alternative splicing is a post- and co-transcriptional regulatory mechanism of gene expression. Pentatricopeptide repeat (PPR) family proteins were recently found to be involved in RNA editing in plants. The aim of this study was to investigate the tissue-specific expression and alternative splicing of PPR family genes and their effects on protein structure and functionality. Of the 27 *PPR* genes in *Arabidopsis thaliana*, we selected six *PPR* genes of the P subfamily that are likely alternatively spliced, which were confirmed by sequencing. Four of these genes show intron retention, and the two remaining genes have 3' alternative-splicing sites. Alternative-splicing events occurred in the coding regions of three genes and in the 3' UTRs of the three remaining genes. We also identified five previously unannotated alternatively spliced isoforms of these *PPR* genes, which were confirmed by PCR and sequencing. Among these, three contain 3' alternative-splicing sites, one contains a 5' alternative-splicing site, and the remaining gene contains a 3'-5' alternative-splicing site. The new isoforms of two genes affect protein structure, and three other alternative-splicing sites are located in 3' UTRs. These findings suggest that tissue-specific expression of different alternatively spliced transcripts occurs in Arabidopsis, even at different developmental stages.

Keywords: RNA editing, alternative splicing, PPR, Arabidopsis

1. Introduction

Alternative splicing is a bimolecular process in which multiple mRNAs are generated from the same gene through the selection of different splicing sites. This key process occurs in living organisms during development and is regulated in response to environmental factors (*1-3*). Surprisingly, up to 60% of multiexon-containing genes undergo alternative splicing (*4*). Core circadian clock genes are also controlled by alternative splicing in Arabidopsis under different environmental conditions

(*4,5*). The functional consequences and effects of alternative splicing on plant phenotypes are an important focus of study (*4*).

Arabidopsis contains approximately 450 pentatricopeptide repeat (PPR) proteins (*6*). PPR proteins were first identified almost two decades ago (*7*). This important protein family is involved in a wide variety of cellular processes in plants, such as RNA editing (*8-10*), RNA stabilization (*11,12*), post-transcriptional RNA maturation (*13,14*), seed development (*15*), endonuclease activity (*16*), and various phenotypic effects (*17*). The PPR protein family comprises the P and PLS subfamilies. P proteins contain a 35 amino-acid classical tandem repeat known as the P motif. PLS subfamily proteins consist of a P motif, an S (short) 31 amino-acid motif, and an L (long) 35-36 amino-acid motif. The PLS subfamily can be divided into the PLS, DYW, and E subclasses. DYW and E proteins contain an additional, conserved C-terminal motif with deaminase properties that plays a distinct role in RNA editing in plants (*18-

21). The DYW domain has C-to-U converting activity and shares similarity with the domain found in cytidine deaminase (*22*). The P subfamily members are mainly involved in RNA splicing, stabilization, cleavage, and the activation or repression of translation (*23*). PPR proteins primarily form complexes with other family members or sometimes with associated group members to form editosome complexes (*9*). In addition to the PLS subfamily, the P subfamily of PPR proteins is involved in RNA editing in plants (*21,24*).

RNA editing efficiency varies in different tissues and developmental stages, such as non-green tissues in seedlings and mature plants, although most RNA editing sites are found in green tissues (*25*). The PPR motif can be reprogrammed to target any RNA molecule (*26*). Recent advancements in the use of synthetic PPR molecules have created interest in programmable targeting to RNA substrates. Most PPR proteins localize to the mitochondria, chloroplasts, or both organelles, whereas very few localize to the cytoplasm and nucleus (*27*). Tissue-specific alternative splicing affects protein phosphorylation, which ultimately alters protein stability, enzymatic activity, subcellular localization in mammalian tissues (*28*). However, the role of extensive tissue-specific alternative splicing in plants is less well understood.

To date, most studies investigating alternative splicing have been performed using whole plants. However, we were interested in investigating this process in specific tissues and at particular developmental stages. The aim of the current study was to investigate tissue-specific alternative splicing of selected *PPR* genes in Arabidopsis seedling, leaf, stem, stipe, and root tissue. We also investigated the expression patterns of alternatively spliced transcripts in various Arabidopsis tissues on day 4, 8, 12, 16, 21, 27, and 32 of plant development. Finally, we performed bioinformatics analysis to investigate how alternative splicing affects protein diversity in this plant, shedding light on this important process during plant development.

2. Materials and Methods

2.1. *Plant growth conditions and sample collection*

Arabidopsis thaliana ecotype Colombia (Col-0) seeds were sown in paper pots containing a 1:2:1 mixture of horticultural perlite, peat moss, and vermiculite, covered with plastic wrap, and incubated for 3 to 4 days in the dark. The pots were transferred to a U-ING Green Farm hydroponic grow box (Osaka, Japan) in a growth room at 22°C, relative humidity 45%, and a 16 h light/8 h dark cycle. After germination, the plants were watered every morning and evening and fertilized twice weekly. Seedlings (whole plants on days 4, 8, and 12) and 16-, 21-, 27-, and 32-day-old leaf, stipe, stem, and root tissue were collected for analysis.

2.2. *RNA extraction and cDNA synthesis*

RNA was extracted from the samples using a Qiagen Plant Mini kit (Hilden, Germany, catalog no. 74904) according to the manufacturer's instructions. The RNA was treated with DNase (RQ1 RNase free DNase; Promega, Madison, WI, USA) to digest contaminating genomic DNA. After DNase treatment, the samples were purified by phenol-chloroform and ethanol precipitation. The final purified RNA was quantified using a NanoDrop (Thermo Fisher Scientific, Waltham, MA, USA). The cDNA was synthesized using reverse transcriptase (Superscript III, Invitrogen, Carlsbad, CA, USA) with oligo dT primers, which was confirmed using primers for the Arabidopsis housekeeping gene GAPDH: forward primer: GTTGTCATCTCTGCCCCAAG, reverse primer: TGCAACTAGCGTTGGAAACA.

2.3. *Selection of PPR candidate genes from the Arabidopsis genome and mRNA database*

Candidate of alternatively spliced genes were selected from various databases as follows: The accession numbers of 425 *PPR* genes were obtained from *https://www.ncbi.nlm.nih.gov/*. The accession numbers were used as queries against the www.plantgdb.org/AtGDB/ database to obtain a genomic map of each gene. Based on these genomic maps, likely alternatively spliced genes (whole genome model) were selected using At-TAIR10. After identifying 27 genes that are alternatively spliced and expressed in different tissues from the Arabidopsis information resource (TAIR; *https://www.arabidopsis.org/*), the full-length genomic DNA, mRNA, and cDNA of each gene was identified as well. Finally, for each gene, the whole genome sequence, intron-exon sequences, CDS, transcript sequence, and deduced protein sequence were downloaded from *http://atgenie.org and crosschecked*.

2.4. *Primer design*

Primers were designed using Primer3 primer design software (bioinfo.ut.ee/primer3-0.4.0/primer3/), and BLAST analysis was performed using NCBI/Primer BLAST. When the primers failed to produce the desired product, another set of primers was designed and tested. The primers (in TE buffer at a concentration of 50 pmol/μL under salt-free conditions) were purchased from Eurofins Genomics (Tokyo, Japan). Each primer set was diluted to 10 pmol/μL with TE buffer (working concentration).

2.5. *PCR and polyacrylamide gel image analysis*

PCR was performed using 30-35 cycles at a denaturation temperature of 94°C and an elongation temperature

Table 1. Candidates studied genes and accession number of PPR family proteins

Tentative Gene Name	Accession number	AT°C	Cycle	Sequence (5' → 3')	
				Forward	Reverse
PPR1	AT5G24060	55	30	GAACACAGAGGACGGAGGAG	TTGACTTCCCAGTTGGCTTC
PPR2	AT5G27300	58	30	GTTTTGATCCCACCCATCAC	TGAACGTGTTGGGACTGAAG
PPR3	AT2G19280	58	30	TGATCCCGAACAGTCTCAAGT	CTCTCATCTTCCTCCGGCTA
PPR4	AT4G38150	59	35	ATTTCCGTCTTTGGCTTGG	CAGTGCCTTTGGAGGATGAT
PPR5	AT1G30610	58	30	GGATATGGAAAGCCGTGGTA	GCCAGTGTAAGTCACCACGA
PPR6	AT1G05670	58	30	AAGAACTGATGCTGGTGCAA	TGGAGAGATGATGTGGCTCTT

All genes belong to the P-subfamily of PPR family proteins which were confirmed by PCR and sequencing. The table indicates annealing temperature, thermal cycle and primer sequences.

of 72°C (Table 1); the annealing temperature varied depending on the primer. The PCR products were subjected to electrophoresis on a ~6% polyacrylamide gel and stained with SYBR Green Dye (Lonza, Rockland, ME, USA). An equal amount of PCR product was subjected to polyacrylamide gel electrophoresis in a gel containing ethidium bromide. Each gel image was photographed using different exposure times to obtain high-quality images for analysis. The gel images were analyzed using LAS3000 software (Fujifilm, Tokyo, Japan). The experiments were conducted thrice (*n* = 3).

2.6. *Transcript sequencing*

The PCR products were sequenced on an Applied Biosystems 3130xl Genetic Analyzer (Foster City, CA, USA). The desired bands from PAGE were excised, transferred to a disposable pellet pestle/tissue grinder (Kimble®, Capitol Scientific, Inc., Austin, TX, USA, catalog no. 749520-0090), and incubated in a -80°C freezer for 1 h. The frozen gel piece was ground well with a pestle. Approximately 10 µL of 0.1× TE was added to the gel powder, followed by additional grinding. After discarding the pestle, the tube was vortexed for 10 min and centrifuged at full speed at 4°C for 20 min in a tabletop centrifuge. The supernatant was transferred to another tube, and 3 µL of sample was used for sequencing with a BigDye Terminator V3.1 Sequencing Standard kit (Applied Biosystems, Austin, TX, USA, catalog no. 4336935). However, in some samples where the PCR product was unable to sequence, in that case the PCR product was amplified by TA cloning using the pGEMT-Easy vector system (Promega, catalog no. A1360). The sequencing results using the reverse primers were reverse complemented using the online software, *http://www.bioinformatics.org/sms2/reference.html* (*29*). All sequencing results were aligned with the Arabidopsis genome *via* BLAST searches (*30*). The experiments were conducted thrice (*n* = 3).

2.7. *Analyzing the tissue-specific expression patterns of the isoforms of different PPR genes*

The PCR products were subjected to 6% polyacrylamide

gel electrophoreses in 1× TBE buffer at 200 volts for 20 min. The gel was stained with SYBR Green Dye (Lonza) for 20 min in 1× TBE buffer in a constant rotating shaker. The gels were imaged using an LAS3000 Imaging System (Fujifilm). Image J software (NIH, MD, USA) was used for densitometry analysis of the bands for comparative expression analysis of the different alternatively spliced transcripts, which were confirmed by sequencing.

2.8. *Determining the effect of alternative splicing on protein diversity*

I-TASSER (Iterative Threading ASSEmbly Refinement) bioinformatics tools (*31-33*) were used to examine the effects of alternatively spliced isoforms on protein levels, as well as ligand-binding sites and the distances between residues. Each full-length amino-acid sequence was submitted to the I-TASSER server. The amino-acid sequence of each isoform was obtained from *http://atgenie.org*. However, the sequences of newly identified isoforms were transcribed using the ExPASy Translate tool. The amino-acid sequences were then submitted to the server with the desired output parameters, and data were obtained 1-2 weeks later. Various parameters were comparatively analyzed between the alternatively spliced isoforms.

3. Results

We identified 27 alternatively spliced genes of PPR family proteins from an Arabidopsis database (*https://www.ncbi.nlm.nih.gov*) that were differentially expressed in different tissues. Of these genes, we have selected six genes due to their functional importance for further analysis by semi-quantitative PCR and sequencing. We have represented the alternatively spliced isoforms of a gene in alphabetic order a, b, c, d.

3.1. *Tissue-specific expression patterns of the isoforms of different PPR geness*

In *PPR1* harboring a 3' alternative-splicing site, a 39 nucleotide (nt) sequence was added to exon 3, creating

PPR1c. This 3' alternative-splicing site was identified in 4, 8 and 12-day-old seedlings, in leaves on days 16, 21, 27, and 32, in stipes on days 16, 21, 27, and 32, in stems on days 16, 21, 27, and 32, and in roots on days 16, 21, 27, and 32 (Figure 1A-C). We detected another new, unannotated splice isoform containing a 3' alternatively spliced site resulting in an additional 15 nt sequence in exon 3, creating *PPR1b*, which was expressed in leaves on day 21 (Figure 1A-C). *PPR1c* expression gradually increased in stipes but gradually decreased in stems and roots; however, this isoform was highly expressed in all of these tissues on day 27 (Figure 1B).

In *PPR2* harboring a 3' alternatively spliced site, a 26 nt sequence was added to exon 2, creating *PPR2b*, a previously unannotated splicing product. The only 3' alternative-splicing site was activated in leaves on day 32 and in stems on days 16, 21, 27, and 32, but not in leaves on day 16 (Figure 2A-C). Both isoforms were expressed in leaf 21 and 27 days but *PPR2b* expression found higher in 21 days whereas *PPR2a* expression higher in 27 days. In leaves, *PPR2b* was expressed at gradually decreasing levels on days 21, 27, and 32 whereas *PPR2a* highly expressed on day 16 (Figure 2B). In stems, *PPR2a* was not expressed, but *PPR2b* was highly expressed. The expression of *PPR2b* in stems gradually increased on days 16, 21, and 27 but decreased on day 32 (Figure 2B).

In *PPR3*, a 73 bp intron sequence was retained within exons 3 and 4, creating *PPR3b*. This isoform was not expressed in seedlings on days 4 and 12, leaves on day 27, stipes on days 16, 21, and 32, stems on days 21 and 32, or roots on days 16, 21, and 32 (Figure 3A). In leaves, *PPR3a* expression gradually increased on days 16, 21, 27, and 32, with the expression level on day 32 almost twice that on day 16. In stipes, the expression of this splice isoform gradually decreased on days 16, 21, 27, and 32 (Figure 3B); this expression pattern is opposite to that detected in leaves.

In *PPR4*, a 325 bp intron sequence was retained between exons 1 and 2, producing *PPR4d* (Figure 4A). This splice isoform was expressed in almost all tissues except 21 days old stem and root (Figure 4A). We detected two new, unannotated splice isoforms, including one with an additional 5' 80 nt sequence in exon 1, creating *PPR4b*, and another with a 201 bp 5'-3' sequence within exons 1 and 2, creating *PPR4c*, as confirmed by Sanger sequencing (Figure 4A-C). *PPR4d* expression was higher in almost all tissues but *PPR4c* highly expressed in stipe 16 days and stem 32 days (Figure 4B).

In *PPR5*, an 84 bp intron sequence was retained between exons 7 and 8, creating *PPR5b* (Figure 5A). This splice isoform was not activated in seedlings on day 4, leaves on days 16 and 27, stipes on days 21 and 32, stems on day 21, or roots on days 21, 27, and 32 (Figure 5A). In *PPR6*, a 106 bp intron sequence was retained between exons 2 and 3, creating *PPR6c*. This isoform was not activated in seedlings on days 4, 8, and 12, leaves on days 16 and 32, stipes on days 16, 21, and 27, stems on days 16, 21, 27, and 32, or roots on days 16, 21, 27, and 32 (Figure 6A).

Finally, we detected another unannotated alternative-splicing event: a 3' alternative-splicing site resulting in a 9 nt deletion in exon 3, which was detected in stems on day 16, as confirmed by Sanger sequencing (Figure 6A-C). *PPR6b*, containing a 3' alternative-splicing site, is expressed in all tissues. On the other hand, a splice isoform, *PPR6c* with intron retention between exons 2 and 3 was expressed only in leaves on days 21 and 27 and in stipes on day 32 (Figure 6B).

3.2. *Determining the effect of alternative splicing on protein structure*

We detected alternative-splicing events in the coding regions of PPR1, PPR2, and PPR5. Next, we investigated whether alternative splicing affects protein structure. The secondary structure of PPR1 differs among the three isoforms (Table 2). Reference isoform PPR1c, a 3' alternative-splicing site leads to the in-frame addition of 39 nt encoding 13 amino acids: (CGFLKCYSDYITR; amino acids 48th-60th). Due to a 13 amino-acid addition to PPR1a, the number of alpha helices, beta sheets, and random coils is altered in the resulting isoform, PPR1c; PPR1c lacks a beta sheet in its upstream region but contains an additional alpha helix at its C-terminus (Figure 7D). Moreover, the distance among the first 100 residues in the upstream regions of these isoforms is nearly identical (Figure 7D). The addition or deletion of 13 amino acids may alter the ligand-binding ability of the isoforms (Figure 7F-G). PPR1b harbors a new 3' alternative-splicing site leading to the addition of a 15 nt sequence with an in-frame addition of five amino acids: DYITR (48th-52th) (Table 2). Interestingly, however, despite the addition of five amino acids in PPR1a, the number of alpha helices, beta sheets, and random coils remains unchanged (Table 2).

In PPR2, the reference isoform is PPR2a, and another new isoform, PPR2b (Table 2). In PPR2, a 26 nt addition resulted in a frameshift. In this case, the number of alpha helices and beta sheets changed, but the number of coils remained the same (Table 2). In PPR5, due to the addition of 28 amino acids (VNFVNPVVLKLIENLIYKADLVHTIQFQ; amino acids 719th–746th), number of alpha helices, beta sheets and random coil changed (Table 2).

4. Discussion

Much is known about alternative splicing in humans and animals, but little is known about this process in plants. In plants, alternative splicing is a highly diversified process, which greatly affects transcript diversity; in Arabidopsis, even a single nucleotide exon has been reported (*34*). In this study, we focused on
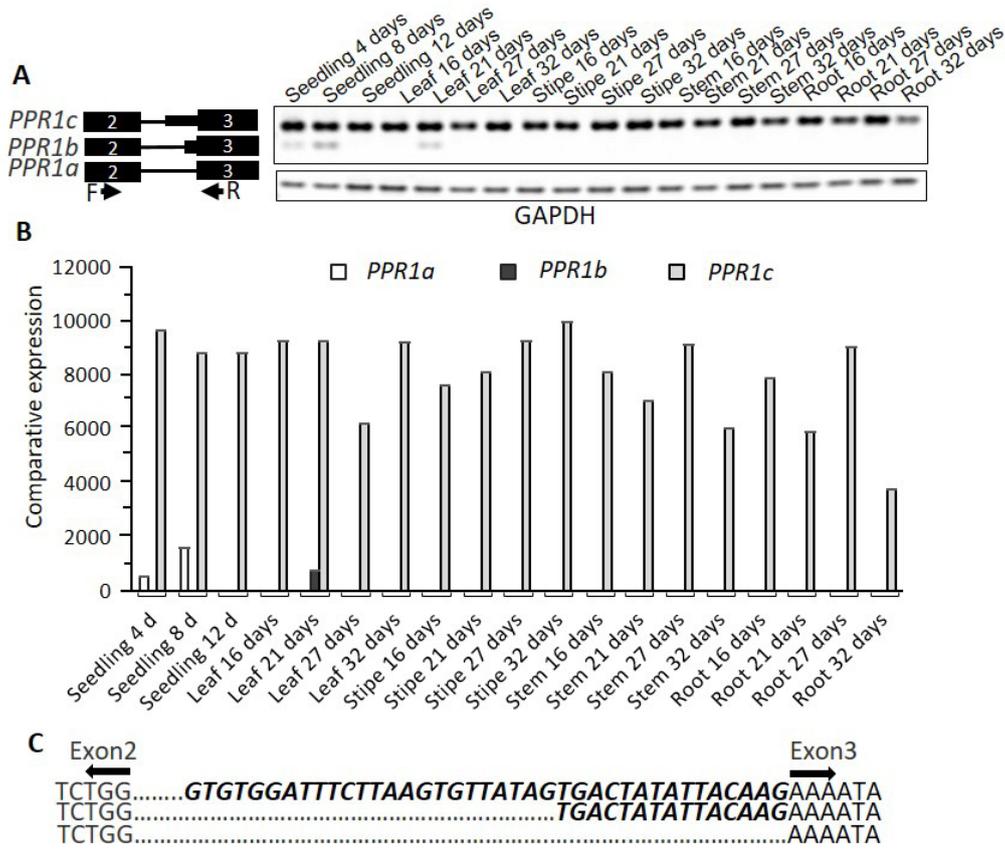
Figure 1. PCR and sequence analysis of PPR1 (accession no. AT5G24060). (A) PCR analysis of seedling, leaf, stipe, stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of three alternatively spliced isoforms in different tissues. **(C)** Genomic sequence of PPR1 from exon 2 to exon 3. Arrow indicates exon 2 and 3 boundary, and dots indicate intron sequences. Sequences in bold and italics indicate a 3′ alternative-splicing site.
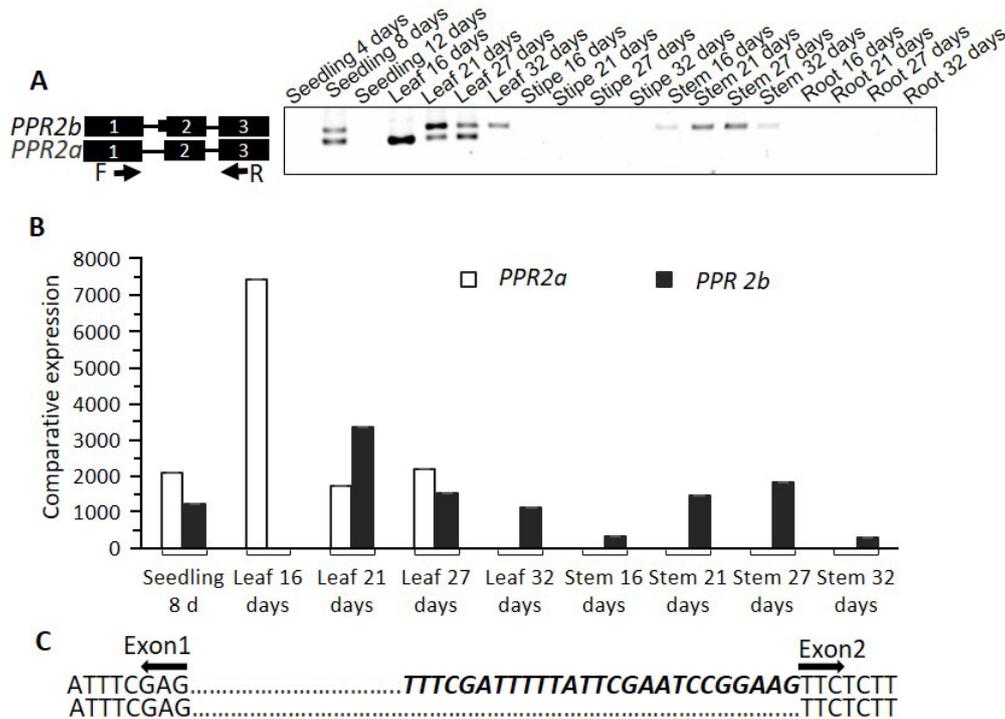

Figure 2. PCR and sequence analysis of PPR2 (accession no. AT5G27300). (A) PCR analysis of seedling, leaf, stipe, stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of two splice isoforms in different tissues. **(C)** Genomic sequence of PPR2 from exon 1 to exon 2. Arrow indicates the exon 1 and 2 boundary, and dots indicate intron sequences. Sequences in bold and italics indicate a 3′ alternative-splicing site.
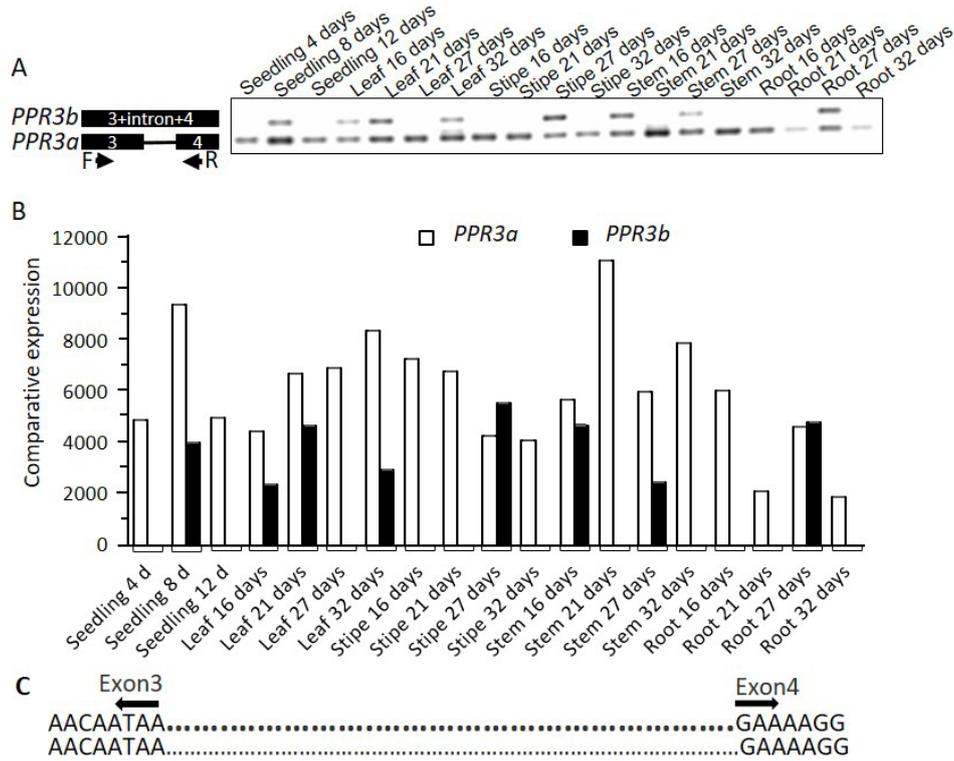
Figure 3. PCR and sequence analysis of PPR3 (accession no. AT2G19280). (A) PCR analysis of seedling, leaf, stipe, stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of the two splice isoforms in different tissues. **(C)** Genomic sequence of PPR3 from exon 3 to exon 4. Arrow indicates the exon 3 and 4 boundary, and dots indicate intron sequences. Bold dot indicate intron retained.
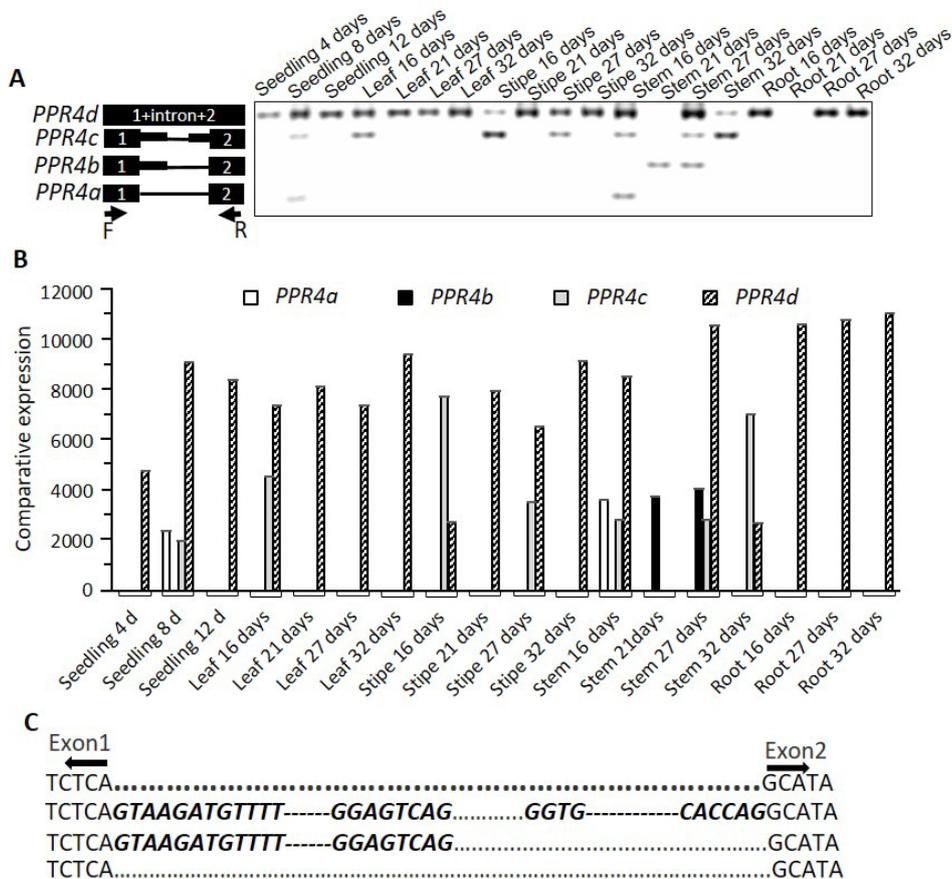

Figure 4. PCR and sequence analysis of PPR4 (accession no. AT4G38150). (A) PCR analysis of seedling, leaf, stipe, stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of the splice isoforms in different tissues. **(C)** Genomic sequence of PPR4 from exon 1 to exon 2. Arrow indicates the exon 1 and 2 boundary, and dots indicate intron sequences. Sequences in bold and italics and dashes indicate a 3′-5′ alternative-splicing site. Bold dot indicate intron retained.
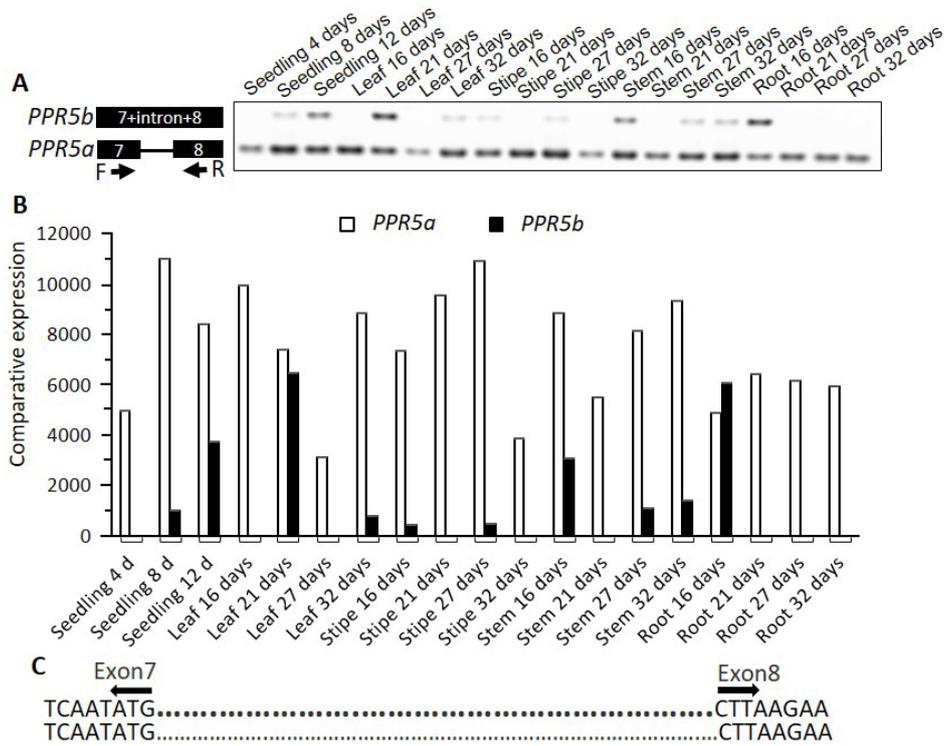
Figure 5. PCR and sequence analysis of PPR5 (accession no. AT1G30610). (A) Polyacrylamide gel electrophoresis of PCR product amplified from seedling, leaf, stipe, stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of the two splice isoforms in different tissues. **(C)** Genomic sequence of PPR5 from exon 7 to exon 8. Arrow indicates the exon 7 and 8 boundary, and dots indicate intron sequences. Bold dot indicate intron retained.
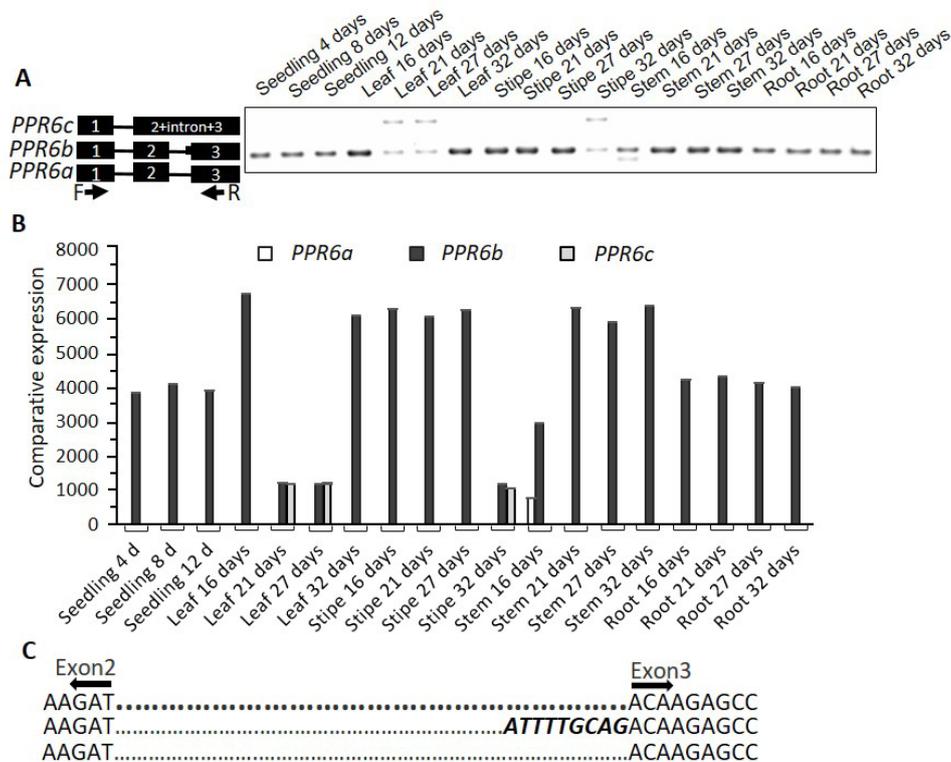

Figure 6. PCR and sequence analysis of PPR6 (accession no. AT1G05670). (A) PCR analysis of seedling, leaf, stipe stem, and root tissue. Fragment size estimated from the gel in base pairs (bp). **(B)** Comparative expression analysis of splice isoforms in different tissues. **(C)** Genomic sequence of PPR6 from exon 2 to exon 3. Arrows indicate the exon 2 and 3 boundaries, and dots indicate intron sequences. Sequences in bold and italics indicate a 3′ alternative-splicing site. Bold dot indicate intron retained.

Table 2. Summary of alternative-splicing events affecting PPR protein structure

Tentative Gene Name	Alternative-splicing event	Effect on		No. of alpha helices	No. of beta sheets	No. of random coils
		Nucleotides	Amino acids			
PPR1a	Reference isoform shorter	–	–	18	14	31
PPR1b	New 3' alternative-splicing site	15 nucleotide addition (TGACTATATTACAAG)	5 amino-acid addition (D Y I T R)	18	14	31
PPR1c	Reference isoform 3' alternative-splicing site	39 nucleotide addition (GTGTGGATTTCTTAAGTGT TATAGTGACTATATTACAAG)	13 amino-acid addition (C G F L K C Y S D Y I T R)	20	13	33
PPR2a	Reference isoform shorter	–	–	32	1	33
PPR2b	New 3' alternative-splicing site	26 nucleotide addition (TTTCGATTTTTATTCGAATC CGGAAG)	Frameshift	27	1	28
PPR5a	Reference isoform shorter	–	–	48	5	54
PPR5b	Reference isoform intron retention	84 nucleotide addition (GTAAATTTTGTAAATCCTG TAGTCCTTAAGCTTATCGA GAATTTGATTTACAAAGCT GATCTTGTTCATACCATCCA ATTTCAG)	28 amino-acid addition (V N F V N P V V L K L I E N L I Y K A D L V H T I Q F Q)	55	6	58

"–" indicate not applicable.

alternative splicing of *PPR* genes. According to the *http://www.uniprot.org* database, *PPR1* binds with iron as a cofactor, but its biological function is unknown. *PPR2*, *PPR3*, *PPR4*, *PPR5*, and *PPR6* exhibit endonuclease activity and bind to RNA, and function in RNA modification. *PPR5* likely functions in plant embryo development (*17,35*). This protein is expressed in the female gametophyte and is responsible for the initiation of gametogenesis (*36*). PPR4 is predicted to be localized to the mitochondria (*6*).

The tissue-specific expression patterns of *PPR* genes differ due to alternative splicing. In the current study, we found that retained introns were highly expressed in all tissues than 3′ and 5′ alternative-splicing sites. In different cell of Arabidopsis root tissue, intron retention is more common than other types of alternative splicing (*37*). However, it is interesting that in almost all type of tissues even in different aged condition, we found intron retention is the common phenomenon. In the current study, we detected higher tissue-specific expression of *PPR1* and *PPR2* than of the other *PPR* genes. RNA editing in *ndhB-149*, *ndhB-1255*, and *ndhD-2* does not occur in roots or in lincomycin-treated seedlings (*25*). In addition, *matK-640* is fully edited, and *accD-794*, *atpF-92*, *psbE-214*, *psbF-77*, *psbZ-50*, and *rps14-50* are completely or highly edited, in both green and non-green tissues (*25*). Based on these findings and the current results, it is likely that alternative splicing of *PPR* genes affects tissue-specific RNA editing.

We also investigated the effects of alternative splicing on protein structure and functionality. PPR1c contains a 13 amino-acid sequence, CGFLKCYSDYITR (48th-60th). Compared with the shorter isoforms, PPR1c contains a higher number of alpha helices and random coils but fewer beta sheets (Table 2). This amino-acid sequence is homologous to that of THCA synthase in *Cannabis sativa*. The major role of these residues is substrate binding rather than direct catalysis (*38*). Another homolog of this peptide is found in a juvenile hormone esterase-related protein in *Operophtera brumata*. This protein plays a major role in controlling growth and development (*39*). In PPR1b, the addition of the amino-acid sequence DYITR (48th-52th) creates a long random coil. This peptide is homologous to non-ribosomal peptide synthases (NRPSs), which help produce natural products with antimicrobial and anticancer properties (*40*). The additional 26 nt sequence in PPR2b is homologous to that in *Apis cerana* endothelin-converting enzyme 1 (ECE-1), as revealed by BLAST searches (*https://www.ncbi.nlm.nih.gov*). ECE-1 and ECE-2 can both cleave amyloid-β in mouse brain (*41*).

In this study, we found that alternative splicing in *PPR* genes adversely affects their structure and functionality. PPR proteins are highly diversified due to alternative splicing. I-TASSER analysis showed that the structures of these proteins, especially their folding ability and sometimes their ligand-binding ability, were completely altered due to alternative splicing. This process can also affect the number of alpha helices, beta sheets, and coils in a protein. Indeed, alternatively spliced exons were previously shown to affect the tertiary structures of proteins and have a great impact on protein folding ability and (ultimately) functionality (*42*). Protein-protein interaction pathways can also be altered due to tissue-specific alternative splicing (*43*).

Therefore, it is obvious that due to alterative splicing our studied PPR proteins might be greatly affected with altered even opposite function as these proteins bind to its target RNA with an algorithmic manner (*8,44*).
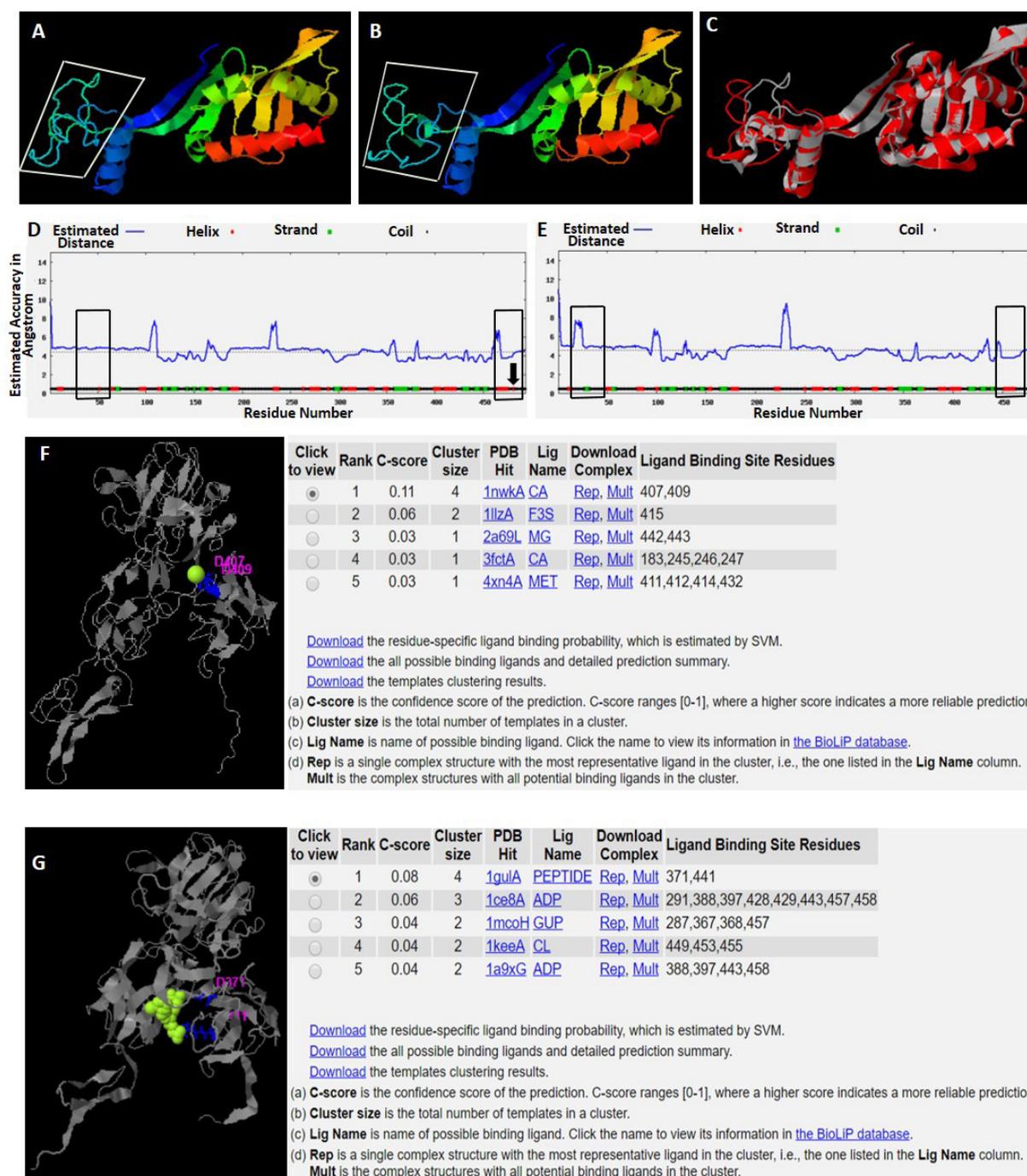
Figure 7. Effects of the addition and deletion of 13 amino acids on the structures and changes in ligand-binding site residues in PPR1, as revealed using the I-TASSER bioinformatics tool. (A) Addition of 13 amino acids (48th–60th), **(B)** Deletion of 13 amino acids, **(C)** Superposition of the two isoforms. **(D)** The beta sheet is absent in the N-terminus and an alpha helix is present in the C-terminus due to the addition of 13 amino acids, arrow indicates additional alpha helix in C-terminus. **(E)** A beta sheet is present in the N-terminus and the alpha helix is absent in the C-terminus due to a deletion of 13 amino acids. **(F-G)** Ligand-binding site is completely altered in the two isoforms.

Therefore, alternative splicing may not only alter binding location but also binding affinity. In the current study, we also determined that the secondary structures and ligand-binding activities of different isoforms from the same gene can be altered due to alternative splicing. Indeed, in the nuclear transcripts of AT1G29930.1 and AT1G52400.1 from Arabidopsis, C-to-U and U-to-C RNA editing occurs in the translation borders (*45*). These deamination and amination reactions occur in

highly adjacent sites, suggesting that the deamination reaction serves as the donor of amino groups for the amination reaction (*45*), although the amination frequency is higher. Another factor might also function as an amino-group donor. The interesting thing is that the same *PPR* gene may produce the enzyme with amination and deamination activity that may be generated due to alternative splicing.

In conclusion, the findings of this study indicate that

tissue-specific alternative splicing highly diversified in plants. Alternative-splicing events in PPR transcripts have strong effects on protein diversity. More investigations of the localizations of alternatively spliced transcripts and proteins, as well as site-specific and tissue-specific RNA editing, are needed to further understand their effects on growth and development. PPRs are modular proteins that are highly reprogrammable. Therefore, it would be interesting to investigate how these types of natural editing events in transcripts affect substrate recognition and plant physiology. Additional experiments are needed to understand the precise effects of alternative-splicing and editing events on the highly programmable PPR protein family.

Acknowledgement

References

1. Graveley BR. Alternative splicing: Increasing diversity in the proteomic world. Trends Genet. 2001; 17:100-107.
2. Lareau LF, Green RE, Bhatnagar RS, Brenner SE. The evolving roles of alternative splicing. Curr Opin Struct Biol. 2004; 14:273-282.
3. Kelemen O, Convertini P, Zhang Z, Wen Y, Shen M, Falaleeva M, Stamm S. Function of alternative splicing. Gene. 2013; 514:1-30.
4. Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW. Alternative splicing in plants--coming of age. Trends Plant Sci. 2012; 17:616-623.
5. James AB, Syed NH, Bordage S, Marshall J, Nimmo GA, Jenkins GI, Herzyk P, Brown JW, Nimmo HG. Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. Plant Cell. 2012; 24:961-981.
6. Lurin C, Andres C, Aubourg S, *et al.* Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell. 2004; 16:2089-2103.
7. Small ID, Peeters N. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci. 2000; 25:45-47.
8. Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T. A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. J Biol Chem. 2006; 281:37661-37667.
9. Andres-Colas N, Zhu Q, Takenaka M, De Rybel B. Multiple PPR protein interactions are involved in the RNA editing system in Arabidopsis mitochondria and plastids. Proc Natl Acad Sci U S A . 2017; 114:8883-8888.
10. Bayer-Csaszar E, Haag S, Jorg A, Glass F, Hartel B, Obata T, Meyer EH, Brennicke A, Takenaka M. The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. Biochim Biophys Acta. 2017; 1860:813-828.
11. Choquet Y. 5' and 3' ends of chloroplast transcripts can

both be stabilised by protein 'caps': A new model for polycistronic RNA maturation. EMBO J. 2009; 28:1989-1990.
12. Prikryl J, Rojas M, Schuster G, Barkan A. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. Proc Natl Acad Sci U S A. 2011; 108:415-420.
13. Delannoy E, Stanley WA, Bond CS, Small ID. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem Soc Trans. 2007; 35:1643-1647.
14. Williams-Carrier R, Kroeger T, Barkan A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA. 2008; 14:1930-1941.
15. Gutiérrez-Marcos JF, Dal Prà M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, Dickinson HG, Consonni G. empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. Plant Cell. 2007; 19:196-210.
16. Zhou W, Lu Q, Li Q, Wang L, Ding S, Zhang A, Wen X, Zhang L, Lu C. PPR-SMR protein SOT1 has RNA endonuclease activity. Proc Natl Acad Sci U S A. 2017; 114:E1554-e1563.
17. Cushing DA, Forsthoefel NR, Gestaut DR, Vernon DM. Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. Planta. 2005; 221:424-436.
18. Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. Plant Cell. 2009; 21:146-156.
19. Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J. 2013; 74:935-945.
20. Schallenberg-Rudinger M, Lenz H, Polsakiewicz M, Gott JM, Knoop V. A survey of PPR proteins identifies DYW domains like those of land plant RNA editing factors in diverse eukaryotes. RNA Biol. 2013; 10:1549-1556.
21. Leu KC, Hsieh MH, Wang HJ, Hsieh HL, Jauh GY. Distinct role of Arabidopsis mitochondrial P-type pentatricopeptide repeat protein-modulating editing protein, PPME, in nad1 RNA editing. RNA Biol. 2016; 13:593-604.
22. Shikanai T. RNA editing in plants: Machinery and flexibility of site recognition. Biochim Biophys Acta. 2015; 1847:779-785.
23. Barkan A, Small I. Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol. 2014; 65:415-442.
24. Doniwa Y, Ueda M, Ueta M, Wada A, Kadowaki K, Tsutsumi N. The involvement of a PPR protein of the P subfamily in partial RNA editing of an Arabidopsis mitochondrial transcript. Gene. 2010; 454:39-46.
25. Tseng C-C, Lee C-J, Chung Y-T, Sung T-Y, Hsieh M-H. Differential regulation of Arabidopsis plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol. 2013; 82:375-392.
26. Coquille S, Filipovska A, Chia T, Rajappa L, Lingford JP, Razif MFM, Thore S, Rackham O. An artificial

PPR scaffold for programmable RNA recognition. Nat Commun. 2014; 5:5729.

27. Colcombet J, Lopez-Obando M, Heurtevin L, Bernard C, Martin K, Berthome R, Lurin C. Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. RNA Biol. 2013; 10:1557-1575.

28. Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science. 2012; 338:1593-1599.

29. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 2000; 28:1102-1104.

30. Kent WJ. BLAT – the BLAST-like alignment tool. Genome Res. 2002; 12:656-664.

31. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. 2008; 9:40.

32. Roy A, Kucukural A, Zhang Y. I-TASSER: A unified platform for automated protein structure and function prediction. Nat Protoc. 2010; 5:725-738.

33. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. Nat Methods. 2015; 12:7-8.

34. Guo L, Liu C-M. A single-nucleotide exon found in Arabidopsis. Sci Rep. 2015; 5:18087.

35. Meinke D, Sweeney C, Muralla R. Integrating the genetic and physical maps of Arabidopsis thaliana: Identification of mapped alleles of cloned essential (EMB) genes. PLoS One. 2009; 4:e7386.

36. Greco M, Chiappetta A, Bruno L, Bitonti MB. In *Posidonia oceanica* cadmium induces changes in DNA methylation and chromatin patterning. J Exp Bot. 2012; 63:695-709.

37. Li S, Yamada M, Han X, Ohler U, Benfey PN. High-resolution expression map of the Arabidopsis root reveals alternative splicing and lincRNA regulation. Dev Cell. 2016; 39:508-522.

38. Shoyama Y, Tamada T, Kurihara K, Takeuchi A, Taura F, Arai S, Blaber M, Shoyama Y, Morimoto S, Kuroki R. Structure and function of 1-tetrahydrocannabinolic acid (THCA) synthase, the enzyme controlling the psychoactivity of Cannabis sativa. J Mol Biol. 2012; 423:96-105.

39. Kontogiannatos D, Swevers L, Maenaka K, Park EY, Iatrou K, Kourti A. Functional characterization of a juvenile hormone esterase related gene in the moth Sesamia nonagrioides through RNA interference. PLoS One. 2013; 8:e73834.

40. Agrawal S, Acharya D, Adholeya A, Barrow CJ, Deshmukh SK. Nonribosomal peptides from marine microbes and their antimicrobial and anticancer potential. Front Pharmacol. 2017; 8:828.

41. Palmer JC, Tayler HM, Love S. Endothelin-converting enzyme-1 activity, endothelin-1 production, and free radical-dependent vasoconstriction in Alzheimer's disease. J Alzheimers Dis. 2013; 36:577-587.

42. Birzele F, Csaba G, Zimmer R. Alternative splicing and protein structure evolution. Nucleic Acids Res. 2008; 36:550-558.

43. Ellis JD, Barrios-Rodiles M, Colak R, Irimia M, Kim T, Calarco JA, Wang X, Pan Q, O'Hanlon D, Kim PM, Wrana JL, Blencowe BJ. Tissue-specific alternative splicing remodels protein-protein interaction networks. Mol Cell. 2012; 46:884-892.

44. Nakamura T, Yagi Y, Kobayashi K. Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. Plant Cell Physiol. 2012; 53:1171-1179.

45. Meng Y, Chen D, Jin Y, Mao C, Wu P, Chen M. RNA editing of nuclear transcripts in Arabidopsis thaliana. BMC Genomics. 2010; 11:S12-S12.