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

Umme Qulsum¹, Toshifumi Tsukahara¹,²,³,*

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 (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-
The PPR subfamily members are mainly involved in RNA splicing, stabilization, cleavage, and the activation or repression of translation. PPR proteins primarily form complexes with other family members or sometimes with associated group members to form editosome complexes. In addition to the PLS subfamily, the P subfamily of PPR proteins is involved in RNA editing in plants.

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. The PPR motif can be reprogrammed to target any RNA molecule. 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. Tissue-specific alternative splicing affects protein phosphorylation, which ultimately alters protein stability, enzymatic activity, subcellular localization in mammalian tissues. 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: GTTGTCACTCTCTCGCCCCCCAG, reverse primer: TGCAACTAGCGTTGGAAAACA.

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.
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 genes

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, and 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, 27, 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 increased 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: (CGFLKCYSDDYITR; 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 (VNFVNPVVLKLIENLKYKADLVHTIQFQ; 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.
The major role of these residues exhibit 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 alternative 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).
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

---

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.
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

This work was supported by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (26670167, 17H02204 and 18K19288).

References


(Received August 1, 2018; Revised December 9, 2018; Accepted December 11, 2018)