

Note

Identification of new pentatricopeptide repeat proteins, MREF1 and 2, involved in mitochondrial RNA editing, using computational target RNA prediction

Takehito Kobayashi, Yusuke Yagi, Taizo Tamura, Takahiro Nakamura*

Faculty of Agriculture, Kyushu University, Fukuoka 819-0395, Japan
*E-mail: tnaka@agr.kyushu-u.ac.jp Tel & Fax: +81-92-802-4721

Received November 5, 2019; accepted January 16, 2020 (Edited by Y. Chiba)

Abstract C-to-U RNA editing has been widely observed in mitochondrial and plastid RNAs in plants. The editing sites are known to be recognized by pentatricopeptide repeat (PPR) proteins, which belong to one of the largest protein families in vascular plants. PPR proteins are sequence-specific RNA-binding proteins that participate in various steps of organelle RNA metabolism, such as cleavage, stabilization, splicing, translation, and editing. Elucidating the underlying mechanisms of sequence-specific RNA recognition by PPR proteins expanded our understanding of the role of PPR proteins in plant organellar RNA editing and enabled the computational prediction of target RNA-editing sites for PPR proteins of interest. Combining computational prediction and experimental verification, we identified three new PPR proteins involved in mitochondrial RNA editing: At1g56570, known as PGN for RNA editing of *nad6_leader_-73* and *cox2_742*, At4g04370 for RNA editing of *nad5_242*, and At2g41080 for *atp1_1292*. Therefore, At4g04370 and At2g41080 were designated as mitochondrial RNA-editing factor 1 (MREF1) and MREF2, respectively. This study supports the use of computational prediction in establishing connections between PPR proteins and specific RNA-editing sites, which are important for maintaining various physiological processes, such as plant development, embryogenesis, and biotic- and abiotic-stress responses.

Key words: computational prediction, mitochondria, PPR protein, RNA editing.

Vascular plants convert hundreds of cytidines (Cs) to uridines (Us) at the RNA level in mitochondria and plastids (Ichinose and Sugita 2016). These C-to-U RNA editing sites are known to be recognized by pentatricopeptide repeat (PPR) proteins (Sun et al. 2016). PPR proteins are characterized by the presence of tandem arrays of a degenerate 35-amino-acid motif, termed the PPR motif (Small and Peeters 2000). PPR proteins form one of the largest protein families in land plants (Cheng et al. 2016). PPR proteins are nuclear-encoded, but can localize to mitochondria or chloroplasts and bind organelle transcripts in a sequence-specific manner (Barkan and Small 2014; Nakamura et al. 2012). The PPR protein family is categorized into P and PLS subclasses, based on the PPR motif composition (Cheng et al. 2016; Lurin et al. 2004; O'Toole et al. 2008). P-class PPR proteins contain canonical 35-amino acid PPR motifs and are involved in a wide range of organelle RNA metabolism processes, such as cleavage, stabilization, splicing, and translation. Meanwhile, PLS-class PPR proteins are composed of longer or shorter PPR motifs, in addition to the 35-amino acid PPR motifs, and largely

function in plant organellar RNA editing.

The principle of specific RNA recognition by PPR proteins, designated as the PPR code, has recently been elucidated by computational, biochemical, and structural analyses (Barkan et al. 2012; Takenaka et al. 2013; Yagi et al. 2013a; Yin et al. 2013). When PPR proteins bind RNA, the N- to C-terminus of the PPR protein is aligned with the 5' to 3' end of the target RNA, with a one-motif to one-nucleotide correspondence. The RNA base specificity is determined by a combination of a few amino acid residues, specifically located within the PPR motif. The PLS-class PPR proteins involved in RNA editing recognize the 5' upstream sequence of an editable C, while the last C-terminal PPR motif is located at the nucleotide position -4 (four nucleotides upstream), with respect to the editing site. This rule of PPR–RNA alignment in RNA editing has enabled the computational prediction of target RNA-editing sites for PPR proteins of interest (Takenaka et al. 2013; Yagi et al. 2013a).

Arabidopsis thaliana contains approximately 200 PLS-class PPR protein genes, most of which are thought to be involved in organellar RNA editing (Ichinose and

Sugita 2016; Sun et al. 2016). To date, approximately 60 *A. thaliana* PLS-class PPR proteins have been experimentally identified as RNA-editing PPR proteins, along with their corresponding specific RNA-editing sites. A single RNA-editing PPR protein has been shown to be potentially responsible for one or multiple (up to eight) specific RNA-editing events that are essential for maintaining various physiological processes, such as plant development and responses to environmental stimuli (Chateigner-Boutin et al. 2008; Sechet et al. 2015; Zhu et al. 2014). These studies suggest the importance of PLS-class PPR proteins and organellar RNA editing in understanding plant physiology.

In the present study, we identified three new PLS-class PPR proteins involved in mitochondrial RNA editing. Two of these proteins (At2g41080 and At4g04370) were selected due to their significant prediction scores by computational analysis. The remaining PPR protein, At1g56570, has been previously shown to be involved in plant defense and abiotic stress responses, but its target RNA has not yet been identified (Laluk et al. 2011).

The three PLS-class PPR genes identified in this study exhibit domain architecture characteristic of RNA-editing PPR proteins. The two PPR genes, At1g56570 and At4g04370, consist of 13 and 18 PPR motifs, respectively, along with E and E+ domains (Figure 1A). The remaining PPR gene, At2g41080, consists of 12 PPR motifs, with E, E+, and DYW domains (Figure 1A). The E and E+ domains have been shown to be essential for RNA-editing activity (Ichinose and Sugita 2016; Sun et al. 2016; Takenaka 2014). The DYW domain is known as a candidate for the catalytic cytidine deaminase function of the editing reaction (Ichinose and Sugita 2016; Sun et al. 2016; Takenaka 2014).

The PLS-class domain architecture of these proteins suggests that they are involved in RNA editing. To identify the target RNA-editing sites, we performed computational target RNA prediction, which successfully identified target RNA-editing sites for several uncharacterized PLS-class PPR proteins (Brehme et al. 2015; Takenaka et al. 2013; Yagi et al. 2013a, 2013b). The computational prediction was performed as previously described (Yagi et al. 2013a). To obtain the predicted RNA recognition sequences for the candidate PPR proteins, the amino acid residues determining the nucleotide base specificity were extracted from the PPR motifs. These specificity-determining residues were then converted to nucleotide base preferences, based on the PPR code (Figure 1A). In parallel, a dataset of RNA sequences was prepared by extracting the 5' upstream sequences before the editable C, for 499 RNA-editing sites that consist of 465 and 34 sites in *Arabidopsis* mitochondria and chloroplasts, respectively. The predicted RNA recognition sequence for the PPR protein and the RNA sequence dataset were used as queries to

obtain a prediction score, represented as a *p*-value, using the FIMO program in the MEME suite software toolkit (<http://meme-suite.org>; Grant et al. 2011). The detailed method of the latest computational PPR–RNA prediction can be found in Kobayashi et al. 2019.

The two PPR genes, At4g04370 and At2g41080, showed highly significant *p*-values ($p < 10^{-5}$) for the mitochondrial RNA-editing sites of *nad5_242* and *atp1_1292*, respectively (Figure 1B). In contrast, the PPR gene, At1g56570, involved in plant defense and abiotic-stress responses, showed significant *p*-values ($p < 0.01$) for several RNA-editing sites (Figure 1B).

The prediction results were verified in gene-deficient plants for the respective PPR genes. Three T-DNA insertion lines, SALK_141937 for At1g56570, SALK_025427C for At4g04370, and SALK_087125C for At2g41080 were obtained from ABRC stock (<https://www.arabidopsis.org>). Homozygous T-DNA insertion lines were isolated by performing genomic PCR analyses with the appropriate primers (Figure 2A; Supplementary Table S1). Additionally, reverse transcription PCR (RT-PCR) analyses confirmed the absence of intact RNAs for the PPR proteins in the respective T-DNA insertion lines (Figure 2A). Thus, the T-DNA insertion lines were used as PPR gene-deficient lines and were designated as *at1g56570*, *at4g04370*, and *at2g41080*.

PPR-deficient lines and wild-type plants (WT, Columbia) were grown on half-strength Murashige and Skoog (MS) medium (Merck KGaA, Darmstadt, Germany) containing 1% sucrose, under continuous white light in a growth chamber at 22°C, after incubation at 4°C. Two-week-old seedlings of the gene-deficient lines displayed no visible aberrant phenotypes under these experimental conditions. Total cellular RNA was extracted from the two-week-old seedlings using a Direct-zol™ RNA MiniPrep kit (Zymo Research, Irvine, CA, USA), following the manufacturer's protocol. RNA was reverse transcribed into cDNA using Superscript III reverse transcriptase (Thermo Fisher Scientific, Waltham, MA, USA) and a random 7-mer primer. PCR amplicons for organelle transcripts were generated using previously published gene-specific primer pairs (Bentolila et al. 2013; Supplementary Table S1). RNA editing was analyzed using Sanger sequencing of the candidate RNA-editing sites within the top 25 ranking sites. The top 25 ranking sites include 86% (95 out of 111) of the experimentally characterized PPR–RNA matches for RNA editing in *Arabidopsis* to date (Kobayashi et al. 2019).

Editing deficiencies were not observed in most of the examined RNA-editing sites. However, RNA-editing defects were indeed observed at *nad6_leader_-73* and *cox2_742* in *at1g56570* (SALK_141937). These RNA-editing sites were ranked 1st and 24th, respectively, in the prediction (Figure 2B). The RNA editing of *nad5_242*,

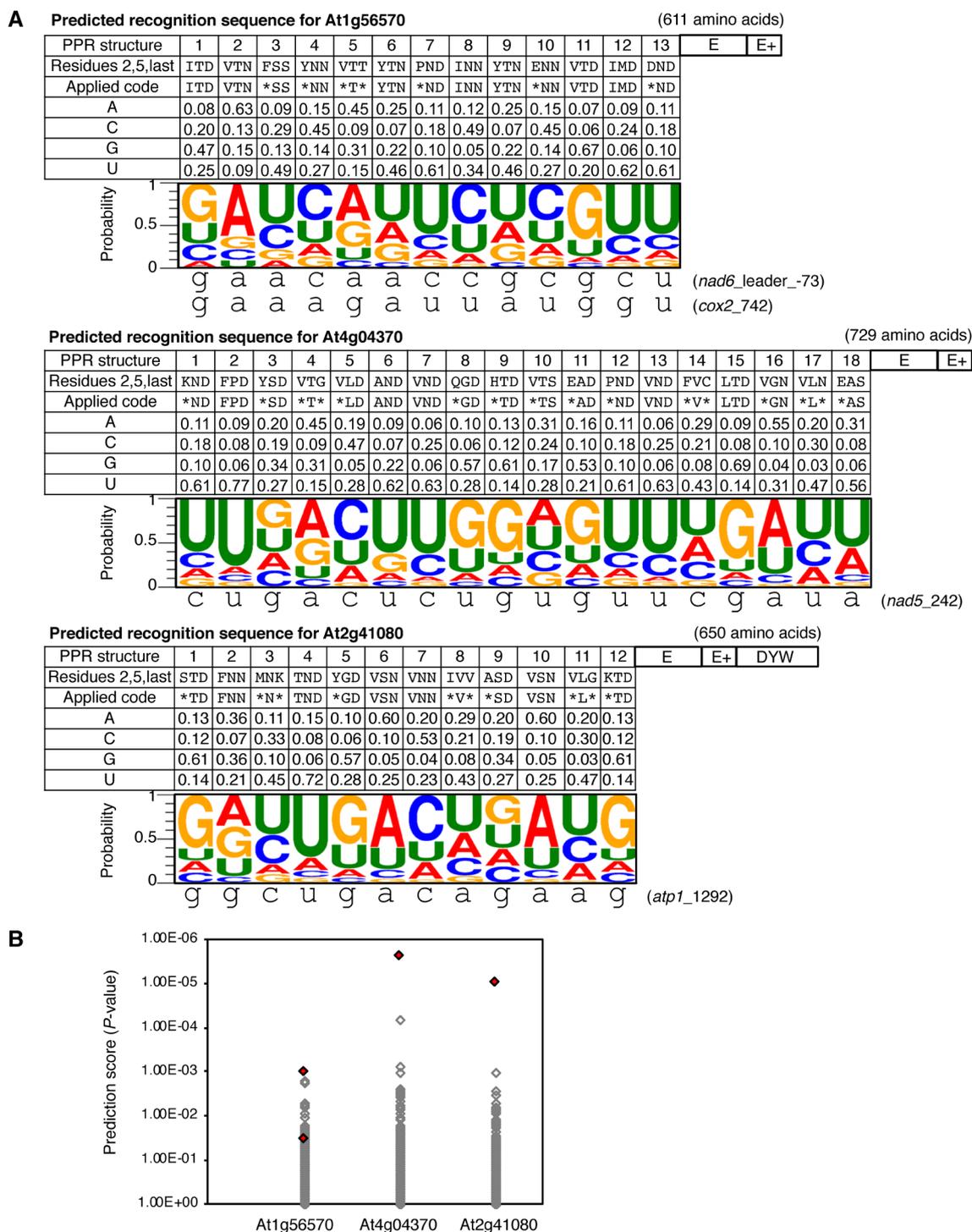


Figure 1. Computational target RNA-editing site predictions for three PPR proteins. (A) The predicted recognition RNA sequences for At1g56570, At4g04370, and At2g41080. The PPR protein structure is schematically represented by following the PPR motif definition of Lurin et al. (2004) and Cheng et al. (2016). The three amino acid residues determining the nucleotide base specificity were extracted from each PPR motif. The specificity-determining residues were converted to the interacting base preference by applying the PPR code dataset of Table S4 in Yagi et al. (2013a). The asterisk indicates any amino acid. The RNA sequence before the editing site in which the impairment/abolishment was observed in the gene-deficient line is also shown. (B) The result of computational target RNA-editing site prediction. The prediction was performed using the predicted recognition sequence against 499 organelle RNA-editing sites in *A. thaliana*. The grey diamonds represent the *p*-values of the prediction score for the respective RNA-editing sites. The red diamond indicates the editing site where the impairment/abolishment was observed in the gene-deficient line.

which ranked 1st in the prediction, was completely abolished in *at4g04370* (SALK_025427C, Figure 2B). The extent of *atp1_1292* RNA editing, which ranked 1st

in the prediction, was significantly different between *at2g41080* (SALK_087125C) and WT (Figure 2B) plants. This partial defect may be due to the expression

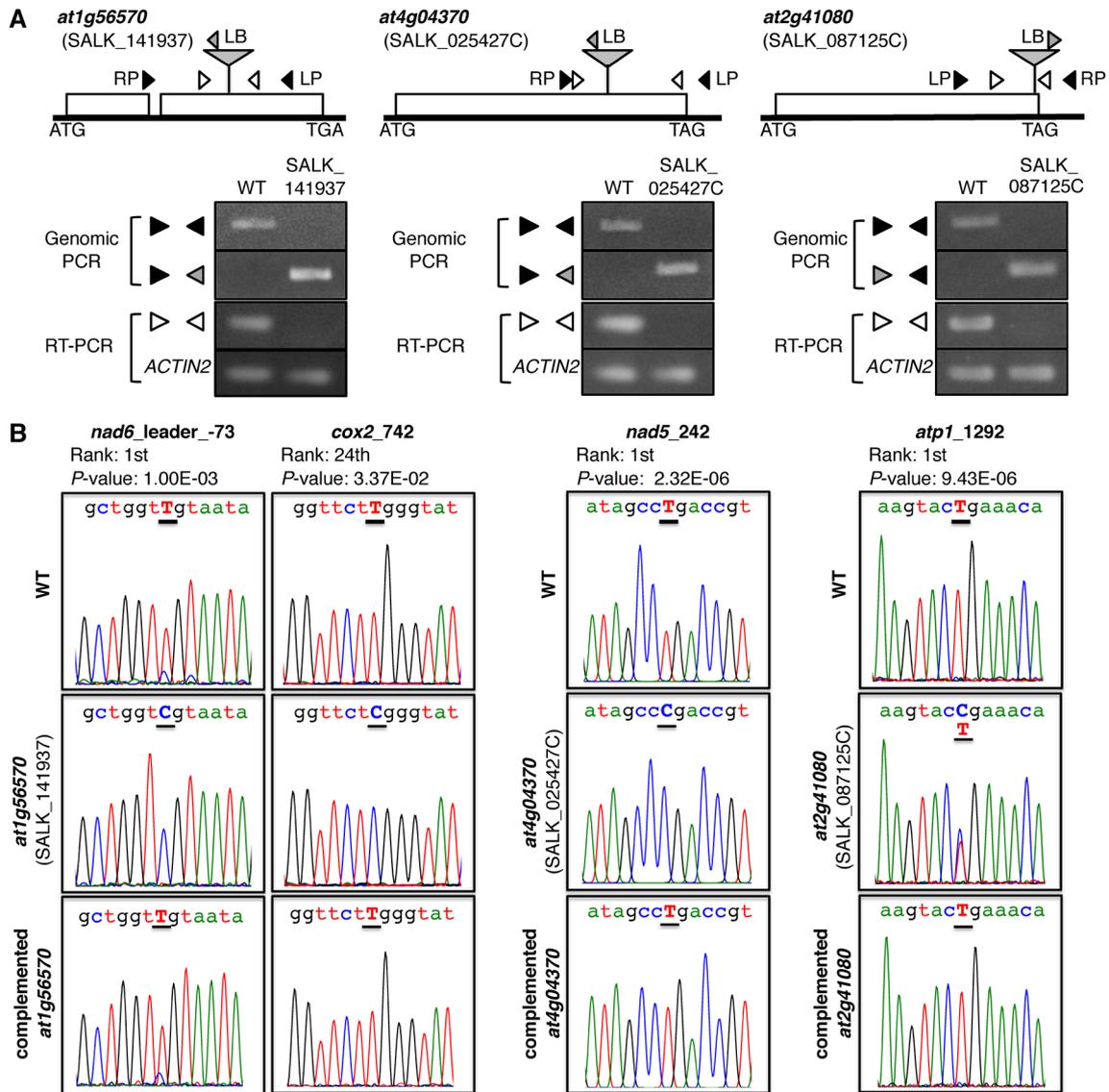


Figure 2. Identification of PPR proteins involved in mitochondrial RNA editing. (A) Genotyping of the gene-deficient lines. The location of coding regions, T-DNA insertions, and primers is schematically represented by white boxes, gray inverted triangles, and arrowheads, respectively. Genomic PCR analyses were performed using the gene-specific primer pair, LP and RP (black arrowheads) and the primer pair, RP and LB (black and gray arrowheads). The primer sequences were obtained from a T-DNA primer design tool (<http://signal.salk.edu/tdnaprimers.2.html>), and listed in Supplementary Table S1. The location of the T-DNA insertion was determined by sequencing the PCR amplicon, using the primer pair, RP and LB. RT-PCR analyses were performed using gene-specific primers indicated by white arrowheads. *ACTIN 2* (*At3g18780*) was used as an internal control. (B) RNA-editing deficiency in PPR-deficient lines. Sequence chromatograms are shown, where abnormalities of RNA editing (indicated by underlines) were detected between wild-type (WT) and the gene-deficient lines: *nad6_leader_-73* and *cox2_742* for *at1g56570*; *nad5_242* for *at4g04370*; *atp1_1292* for *at2g41080*. RNA-editing activity in the PPR-deficient lines was restored to wild-type levels by complementation with the corresponding PPR protein gene.

of a truncated *At2g41080* gene, because the T-DNA insertion was located at the vicinity of the stop codon in the T-DNA insertion line used in this study (Figure 2A). Another possibility is that another PPR protein partially edits the same site.

We next verified the relationships between three PPR proteins and their corresponding RNA-editing activities using complementation experiments. PPR gene-deficient lines were transformed using the floral dip method (Clough and Bent 1998), with a construct (based on the pZP vector; Hajdukiewicz et al. 1994) that expresses the

corresponding wild-type PPR protein under the control of the *A. thaliana* ubiquitin-10 promoter. Transformed T1 seedlings were selected on half-strength MS agar plates supplemented with 25 mg/l hygromycin (Fujifilm Wako Pure Chemical Industries, Ltd., Osaka, Japan). The presence of the introduced T-DNA fragment was confirmed by genomic PCR analysis (Supplementary Table S1). When we examined RNA-editing in five independent transformants for each PPR expression construct, the RNA-editing ability at their respective sites was restored to WT levels (Figure 2B).

Taken together, these results showed that the three PLS-class PPR proteins were involved in mitochondrial RNA editing at the predicted sites. Among them, the newly characterized At4g04370 and At2g41080 were designated as mitochondrial RNA editing factor 1 (MREF1) and MREF2, respectively, due to their involvement in mitochondrial RNA editing. However, further analyses are required to characterize the physiological phenotypes of these gene-deficient lines. In addition, the possibility that these PPR proteins are involved in other RNA-editing site(s) cannot be excluded.

It has been demonstrated that At1g56570, designated as PGN (pentatricopeptide repeat protein for germination on NaCl), is a mitochondria-localized, PLS-class PPR protein that plays an important role in plant defense against necrotrophic pathogens and the response to abiotic stress, such as abscisic acid, glucose, and salinity (Laluk et al. 2011). However, the molecular function of PGN remains unknown. The present study provided additional information that At1g56570 (PGN) participates in the RNA editing of the *cox2* coding region and 5'UTR of *nad6* mRNA (position at -73 from the translational start codon). It is possible that the RNA-editing defect, in which the amino acid sequence of COX2-248 was changed from tryptophan to arginine, may be involved in the physiological phenotypes of *pgn* and in COX2 protein function, although the loss of RNA editing at the 5' UTR of *nad6* mRNA or the multiple RNA editing defects may participate(s) in the *pgn* phenotype.

Here, we have identified three PLS-class PPR proteins involved in mitochondrial RNA editing, as indicated by computational target RNA predictions. Although further research including analysis of all RNA-editing sites will be required to elucidate their functions, these results confirmed the effectiveness of computational methods for establishing connections between PLS-class PPR proteins and specific RNA-editing sites. Further, the results of this study were utilized for the improvement and verification of PPR codes, and the comprehensive prediction of all *Arabidopsis* PLS-PPR proteins (Kobayashi et al. 2019).

Loss of the mitochondrial RNA-editing PPR protein gene is known to affect a wide variety of physiological processes, including abiotic- and biotic-stress responses (Barkan and Small 2014). This may be due to a change in mitochondrial respiratory activity, although the precise mechanisms remain largely unknown. Further studies would facilitate our understanding of not only the mechanisms of plant organellar RNA editing by PPR proteins, but also the role of mitochondria in plant physiology.

Acknowledgements

T.K. was supported by Research Fellowships from the Japan Society for the Promotion of Science for Young Scientists. We thank Atsuko Tsutsumi, Konatsu Sasamoto, and Masuyo Kawabata for her excellent technical assistance.

References

- Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet* 8: e1002910
- Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol* 65: 415–442
- Bentolila S, Oh J, Hanson MR, Bukowski R (2013) Comprehensive high-resolution analysis of the role of an *Arabidopsis* gene family in RNA editing. *PLoS Genet* 9: e1003584
- Brehme N, Bayer-Császár E, Glass F, Takenaka M (2015) The DYW subgroup PPR protein MEF35 targets RNA editing sites in the mitochondrial *rpl16*, *nad4* and *cob* mRNAs in *Arabidopsis thaliana*. *PLoS One* 10: e0140680
- Chateigner-Boutin AL, Ramos-Vega M, Guevara-Garcia A, Andres C, de la Luz Gutierrez-Nava M, Cantero A, Delannoy E, Jimenez LF, Lurin C, Small I, et al. (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* 56: 590–602
- Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, Castleden I, Song Y, Song B, Huang J, et al. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J* 85: 532–547
- Clough SJ, Bent AF (1998) Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16: 735–743
- Grant CE, Bailey TL, Noble WS (2011) FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27: 1017–1018
- Hajdukiewicz P, Svab Z, Maliga P (1994) The small, versatile *pPZP* family of *Agrobacterium* binary vectors for plant transformation. *Plant Mol Biol* 25: 989–994
- Ichinose M, Sugita M (2016) RNA editing and its molecular mechanism in plant organelles. *Genes (Basel)* 8: 5
- Kobayashi T, Yagi Y, Nakamura T (2019) Comprehensive prediction of target RNA editing sites for PLS-Class PPR proteins in *Arabidopsis thaliana*. *Plant Cell Physiol* 60: 862–874
- Laluk K, Abuqamar S, Mengiste T (2011) The *Arabidopsis* mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiol* 156: 2053–2068
- Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al. (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* 16: 2089–2103
- Nakamura T, Yagi Y, Kobayashi K (2012) Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. *Plant Cell Physiol* 53: 1171–1179
- O'Toole N, Hattori M, Andrés C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I (2008) On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* 25: 1120–1128
- Sechet J, Roux C, Plessis A, Effroy D, Frey A, Perreau F, Biniek C,

- Krieger-Liszkay A, Macherel D, North HM, et al. (2015) The ABA-deficiency suppressor locus *HAS2* encodes the PPR protein LOI1/MEF11 involved in mitochondrial RNA editing. *Mol Plant* 8: 644–656
- Small ID, Peeters N (2000) The PPR motif: A TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci* 25: 45–47
- Sun T, Bentolila S, Hanson M (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci* 21: 962–973
- Takenaka M (2014) How complex are the editosomes in plant organelles? *Mol Plant* 7: 582–585
- Takenaka M, Zehrmann A, Brennicke A, Graichen K (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* 8: e65343
- Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T (2013a) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One* 8: e57286
- Yagi Y, Tachikawa M, Noguchi H, Satoh S, Obokata J, Nakamura T (2013b) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol* 10: 1419–1425
- Yin P, Li Q, Yan C, Liu Y, Liu J, Yu F, Wang Z, Long J, He J, Wang HW, et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504: 168–171
- Zhu Q, Dugardeyn J, Zhang C, Mühlenbock P, Eastmond PJ, Valcke R, De Coninck B, Öden S, Karampelias M, Cammue BPA, et al. (2014) The *Arabidopsis thaliana* RNA editing factor SLO2, which affects the mitochondrial electron transport chain, participates in multiple stress and hormone responses. *Mol Plant* 7: 290–310