Architectural of the PPR gene family in the moss Physcomitrella patens

Mamoru Sugita,* Mizuho Ichinose, Mizuki Ide, and Chieko Sugita
Center for Gene Research; Nagoya University; Chikusa-ku; Nagoya, Japan

Keywords: moss, Physcomitrella patens, PPR protein, RNA editing, RNA splicing, RNA cleavage, DYW domain, targeted gene disruption, homologous recombination

Pentatricopeptide repeat (PPR) proteins are widespread in eukaryotes and in particular, include several hundred members in land plants. The majority of PPR proteins are localized in mitochondria and plastids, where they play a crucial role in various aspects of RNA metabolism at the post-transcriptional level in gene expression. However, many of their functions remain to be characterized. In contrast to vascular plants, the moss Physcomitrella patens has only 105 PPR genes. This number may represent a minimum set of PPR proteins required for post-transcriptional regulation in plant organelles. Here, we review the overall structure of the P. patens PPR gene family and the current status of the functional characterization of moss PPR proteins.

Introduction

Pentatricopeptide repeat (PPR) proteins are nucleus-encoded and constitute an extraordinarily large family in land plants, composed of more than 450 members in vascular plants.1,2 Surprisingly, the lycophyte Selaginella moellendorffii has over 800 PPR genes.3 The majority of plant PPR proteins are localized in mitochondria or plastids, and play important roles in a wide range of physiological and developmental functions such as cytoplasmic male fertility, photosynthesis, respiration, and embryogenesis.4 Most PPR proteins that have been investigated are required for various post-transcriptional steps associated with RNA in plant organelles (for recent review, see ref. 5).

The PPR proteins are structurally divided into four classes, P, PLS, E/E+, and DYW, based on their PPR motif and characteristic C-terminal domain structures.1 In Arabidopsis and rice, P-class PPR proteins represent half of all PPR proteins and the remaining half are the E/E+ and DYW-class proteins (Table 1). Extensive functional analyses of PPR proteins have been performed using flowering plants: Arabidopsis, rice, and maize. However, the function of most PPR proteins is unknown, and their characterization remains one of the major challenges in plant science. In contrast to studies performed in flowering plants, knowledge regarding the PPR proteins required for organelle biogenesis in early land plants is limited. However, studies on the moss P. patens organelles have made rapid progress using recently established technologies that generated a wealth of information on the genomes of the nucleus and organelles.6,7 Here, we describe the current data of the overall structure of the P. patens PPR protein family and their function in plastids and mitochondria, and attempt to highlight the differences and similarities of mosses and angiosperms.

The Physcomitrella PPR protein family. Moss PPR genes were first described in 2004 by Hattori et al., who identified over 30 PPR genes in P. patens.8 Subsequently, the whole genome sequence of this moss was disclosed9 and a total of 103 PPR genes were annotated in the genome.2 The P. patens PPR genes are named PpPPR_#, and are numbered sequentially (Table 2). The genome database was updated (The Physcomitrella patens resource COSMOS, http://www.cosmoss.org/) and two additional PPR proteins were identified and designated PpPPR_104 and PpPPR_105. The Physcomitrella PPR gene family is rather small compared with PPR gene families in vascular plants, and contains only 10 DYW-class PPR proteins and no E/E+-class proteins (Table 1).

Subcellular localization of the Physcomitrella PPR proteins. In silico and in vivo analyses have shown that most PPR proteins are localized in either mitochondria or chloroplasts.1 Similarly, we checked the subcellular localization of 105 PpPPR proteins using in silico analysis and in vivo analyses using transient assay or transgenic moss plants expressing PpPPR-green fluorescent protein (GFP) fusion proteins. The subcellular localization of 29 PpPPR proteins was determined experimentally and that of 68 proteins was derived from prediction (Table 2). At least 95 out of 105 PPR proteins are presumably localized in chloroplasts or mitochondria, or both. The number of plastid-targeted PPR proteins is nearly the same as mitochondrial PPR proteins. PpPPR_63 is localized in the nucleus and its paralogs (PpPPR_67 and 104) are located in both chloroplasts and mitochondria. PpPPR_86 is predictably targeted to the endoplasmic reticulum (ER). Subcellular localization of eight PpPPR proteins was not predicted. This could be why PPR ORF models lack the correct initiation codon and, thus, are missing a potential targeting peptide.

Physcomitrella PPR proteins diverge from Arabidopsis and rice proteins. The number of PPR genes in Arabidopsis and rice are strikingly similar (Table 1). More than 80% of Arabidopsis and rice PPR proteins are orthologous pairs.2 In contrast, the

*Correspondence to: Mamoru Sugita; Email: sugita@gene.nagoya-u.ac.jp
Submitted: 02/21/2013; Revised: 04/10/2013; Accepted: 04/22/2013
http://dx.doi.org/10.4161/rna.24772
one-third of the PpPPR genes were tagged by an antibiotic-resistant gene cassette and characterization of their mutants are in progress. This reverse-genetics approach has revealed the function of several PpPPR genes as described below.

**P-class PPR proteins in Physcomitrella.** The P-class PPR proteins are characterized by 35 canonical amino-acid PPR (P) motif. P-class PPR proteins are usually involved in RNA cleavage, RNA splicing, RNA stability, or translation. More than half (55%) of the Arabidopsis PPR proteins are grouped into the P-class. Some contain an additional conserved motif or domain, such as an RNA recognition motif (RRM), small MutS-related (Smr) domain, cystathione β synthase (CBS) domain, or NYN metallonuclease domain. In Physcomitrella, most (85%) of the PpPPR proteins are P-class proteins, and 40% of the P-class PPR proteins show high amino acid identities with Arabidopsis PPR protein sequences, including EMBRYO-DEFECTIVE (EMB) genes, AtPPR4, AtPPR5, MRL1, AtCBS1, GUN1, pTAC2, and PRORP1 (Table 2).

In Physcomitrella, the first functional analysis was achieved for P-class PpPPR_38. PpPPR_38 is involved in splicing and cleavage of the clpP pre-mRNA and binds specifically to the intergenic spacer of chloroplast clpP-5'-rps12 dicistronic mRNA. Although the gene organization of clpP-5'-rps12 is conserved in Physcomitrella and Arabidopsis, PpPPR_38 orthologs are not identified in Arabidopsis. This suggests that an Arabidopsis protein involved in clpP maturation is highly diverged from PpPPR_38. Several P-class PPR proteins are known to be splicing factors for plastid or mitochondrial pre-mRNA in Arabidopsis and maize. Among these, at least AtPPR4 and AtPPR5 homologs are found in Physcomitrella.

The nucleus-localized PpPPR_63 possesses a NYN-metallonuclease domain and is likely orthologous to Arabidopsis PRORP2 that possesses RNase P activity. Disruption of PpPPR_63 gene resulted in abnormal formation of the branched filaments of protonemata (Komura and Sugita, unpublished), suggesting the involvement of PpPPR_63 in the growth and development of protonemal filaments. PpPPR_59 contains an Smr domain and is plastid-localized. Disruption of the PpPPR_59 gene did not result in a different phenotype than wild-type moss plants (Ide and Sugita, unpublished). Arabidopsis GUN1,

---

*Physcomitrella* PPR proteins are somewhat diverged from the Arabidopsis and rice PPR proteins. Intron-containing PPR genes represent three-fourths in *Physcomitrella* but only one-fourth in Arabidopsis and rice. *Physcomitrella* PPR genes are generally intron-rich and alternative splicing variants are often found in PpPPR genes, including PpPPR_38 and PpPPR_43. The gene structure and encoded amino acid sequence of many PPR proteins are well conserved in *Physcomitrella* and Arabidopsis plants (Fig. 1). This conservation suggests that such homologous PPR proteins have the same or similar function in moss and flowering plants. Presumably, intron-rich PPR genes may represent “ancient” PPR genes that pre-dated the occurrence of retrotransposition-mediated expansion of the PPR gene family in land plants.

Although a large mutant collection of targeted gene knockout lines was produced for *P. patens*, no PPR gene-targeted lines are identified. To characterize the function of each PpPPR protein, we constructed gene-targeted knockout or knockdown mutant lines via homologous recombination. To date, in our laboratory, one-third of the PpPPR genes were tagged by an antibiotic-resistant gene cassette and characterization of their mutants are in progress. This reverse-genetics approach has revealed the function of several PpPPR genes as described below.

**Figure 1.** Examples of intron conservation in the homologous PPR genes. In each figure, the first panel shows the gene structure and the second panel shows the motif structure of the predicted PPR proteins.
pTAC2, and SUPPRESSOR OF VARIEGATION7 (SVR7) are a P-class PPR protein with an Smr domain. GUN1 is known to be involved in plastid-to-nucleus retrograde signaling\(^{28}\) and pTAC2 is a component of the transcriptionally active plastid chromosome and might be involved in plastid gene expression.\(^{29}\) SVR7 could be required for FtsH-mediated chloroplast biogenesis.\(^{28}\) PPR motifs likely act as sequence-specific RNA-binding proteins, and non-PPR domains may take part in some RNA processing steps. At least 11 paralogous pairs are found in *Physcomitrella* P-class proteins, including PpPPR\(_3\) and 76, PpPPR\(_{75}\) and 85, and PpPPR\(_{19}\) and 51 (Fig. 1). Their paralogous pairs may have redundant function.

**PLS-class PPR proteins in *Physcomitrella***. This class of PPR proteins is characterized by canonical PPR (P), PPR long (L), and PPR short (S) motifs. Six PLS-class PPR proteins are present in *Physcomitrella* (Table 1) but their functions have not been identified. Three proteins are predicted to be localized in the mitochondrion and two are predicted to be localized in the plastid. Disruption of the *PpPPR*\(_{31}\) gene encoding a mitochondrial protein resulted in severe protonemal growth retardation (Tasaki and Sugita, unpublished). *Physcomitrella* PLS-class proteins are structurally unrelated to the Arabidopsis proteins.

**DYW-class PPR proteins are involved in RNA editing and RNA splicing in *Physcomitrella***. *P. patens* has 10 DYW-class PPR proteins. The DYW domains are 95 amino acids and are named after its patens erolobosean protist *Naegleria gruberi*.\(^{29}\) The protist DYW-class protein is hypothesized to have horizontal gene transfer from plants in very early land plant evolution.\(^{29}\) *Funaria hygrometrica*, a closely related species of *P. patens*, has nine DYW-class PPR proteins homologous to the *P. patens* proteins but lacks the PpPPR\(_{56}\) ortholog.\(^{10}\) In contrast, marchantiid liverworts do not possess DYW-class proteins.\(^{31}\)

In seed plants, more than 400 C-to-U RNA editing sites have been identified in the mitochondria. To date, more than 30 E/ E+- and DYW-class PPR proteins have been identified as editing site-specific factors in flowering plants.\(^{5}\) In contrast, RNA editing occurs at only 11 sites in *P. patens* mitochondrial mRNAs.\(^{32,33}\) To date, eight out of 10 *Physcomitrella* DYW-class proteins have been identified as RNA editing factors. PpPPR\(_{56}\) is involved in editing at the nad3 and nad4 sites,\(^{34}\) PpPPR\(_{77}\) at the cox2 and cox3 sites\(^{34}\) and PpPPR\(_{91}\) at the nad5-2 site.\(^{34}\) PpPPR\(_{78}\) and PpPPR\(_{79}\) are required for editing at the rps14 and cox1 sites\(^{30,35}\) and the nad5-1 site, respectively. PpPPR\(_{71}\) is a sequence-specific recognition factor for editing at the ccmF-2 site of ccmFc mRNA.\(^{33,36}\) In fact, this was demonstrated using electrophoresis mobility shift assays for detection of RNA binding of PpPPR\(_{71}\) protein to the target RNA. PpPPR\(_{65}\) targets the ccmF-1 editing site (Ichinose and Sugita, unpublished; Rüdinger and Knoop, unpublished) and PpPPR\(_{98}\) is responsible for atp9 editing (Ichinose and Sugita, unpublished). Thus, eight DYW-class PPR proteins function in editing all 11 sites in *P. patens* mitochondrial transcripts. Among these 11 editing sites, editing at the ccmF-1, ccmF-2, and nad5-1 sites also occurs in Arabidopsis mitochondria. Interestingly, the moss *F. hygrometrica* lacks both the PpPPR\(_{56}\) ortholog and its target nad3 and nad4 editing sites.\(^{30}\) This suggests that PPR genes and their cognate editing sites are mutually constrained in evolution.\(^{30,37}\)

E/E+-class PPR proteins are required for RNA editing in plastids and mitochondria in flowering plants.\(^{38,39}\) However, no E/ E+-class PPR proteins exist in *P. patens* mitochondria. In addition, DYW\(^{140}\) and multiple organellar editing factor (MORF)\(^{31}\) proteins have recently been identified as editing factors in Arabidopsis but are not present in *Physcomitrella*. This suggests that DYW-class PPR proteins are a sole key player required for RNA editing in

---

**Table 1. Number of PPR genes in Arabidopsis, rice, and moss**

| Plant species       | Total | P    | PLS | E/E+ | DYW |
|---------------------|-------|------|-----|------|-----|
| *Arabidopsis thaliana* | 450   | 250  | 7   | 106  | 87  |
| *Oryza sativa*      | 477   | 235  | 14  | 138  | 90  |
| *Physcomirella patens* | 105   | 89   | 6   | 0    | 10  |

P, PLS, E/E+ and DYW classes are defined by Lurin et al. (2004).\(^{1}\)
| Name       | Class | Auxiliary domain/motif | Location* | Arabidopsis PPR proteinb homologus to PpPPR_# |
|------------|-------|------------------------|-----------|---------------------------------------------|
| PpPPR_1    | P     | M                      | At5g50280 (EMB1006)16 |
| PpPPR_2    | P     | C                      |            |
| PpPPR_3    | P     | RRM                    | At5g04810 (AtPPR4)19 |
| PpPPR_4    | P     | C                      |            |
| PpPPR_5    | P     | M                      |            |
| PpPPR_6    | P     | C                      |            |
| PpPPR_7    | P     | LAGLIDADG              |            |
| PpPPR_8    | P     | M                      |            |
| PpPPR_9    | PLS   | M                      |            |
| PpPPR_10   | P     | C or M                 | At4g21190 (EMB1417)17 |
| PpPPR_11   | P     | M                      |            |
| PpPPR_12   | P     | C                      |            |
| PpPPR_13   | P     | C                      |            |
| PpPPR_14   | P     | C                      | At3g46610  |
| PpPPR_15   | P     | C*                     |            |
| PpPPR_16   | P     | M                      |            |
| PpPPR_17   | P     | C                      | At4g39620 (AtPPR5)10 |
| PpPPR_18   | P     | M                      |            |
| PpPPR_19   | P     | C*                     | At3g34830 (MRL1)11 |
| PpPPR_20   | P     | C*                     | At1g01970  |
| PpPPR_21   | P     | C*                     | At5g02860  |
| PpPPR_22   | P     | LAGLIDADG              | M          |
| PpPPR_23   | P     | C                      | At3g59040  |
| PpPPR_24   | P     | M                      |            |
| PpPPR_25   | PLS   | M                      |            |
| PpPPR_26   | P     | -                      |            |
| PpPPR_27   | P     | -                      | At3g53170  |
| PpPPR_28   | P     | C                      |            |
| PpPPR_29   | P     | M                      |            |
| PpPPR_30   | P     | Smr                    | M          | At1g18900 |
| PpPPR_31   | PLS   | M*                     |            |
| PpPPR_32   | P     | C                      |            |
| PpPPR_33   | P     | M                      |            |
| PpPPR_34   | PLS   | C                      |            |
| PpPPR_35   | P     | M                      | At3g53170  |
| PpPPR_36   | P     | M                      |            |
| PpPPR_37   | P     | -                      |            |
| PpPPR_38   | P     | C*                     |            |

*Location indicates chloroplast (C), mitochondria (M), both C and M (C/M), either C or M, endoplasmic reticulum (ER), nucleus (Nuc) or unknown (-). Asterisks indicate location experimentally determined. Arabidopsis PPR proteins listed show more than 35% amino acid identity with PpPPR proteins, excluding PLS and DYW-class proteins.

### Table 2. List of Physcomitrella PPR proteins (continued)

| Name       | Class | Auxiliary domain/motif | Location* | Arabidopsis PPR proteinb homologus to PpPPR_# |
|------------|-------|------------------------|-----------|---------------------------------------------|
| PpPPR_1    | P     | M                      | At5g50280 (EMB1006)16 |
| PpPPR_2    | P     | C                      |            |
| PpPPR_3    | P     | RRM                    | At5g04810 (AtPPR4)19 |
| PpPPR_4    | P     | C                      |            |
| PpPPR_5    | P     | M                      |            |
| PpPPR_6    | P     | C                      |            |
| PpPPR_7    | P     | LAGLIDADG              |            |
| PpPPR_8    | P     | M                      |            |
| PpPPR_9    | PLS   | M                      |            |
| PpPPR_10   | P     | C or M                 | At4g21190 (EMB1417)17 |
| PpPPR_11   | P     | M                      |            |
| PpPPR_12   | P     | C                      |            |
| PpPPR_13   | P     | C                      |            |
| PpPPR_14   | P     | C                      | At3g46610  |
| PpPPR_15   | P     | C*                     |            |
| PpPPR_16   | P     | M                      |            |
| PpPPR_17   | P     | C                      | At4g39620 (AtPPR5)10 |
| PpPPR_18   | P     | M                      |            |
| PpPPR_19   | P     | C*                     | At3g34830 (MRL1)11 |
| PpPPR_20   | P     | C*                     | At1g01970  |
| PpPPR_21   | P     | C*                     | At5g02860  |
| PpPPR_22   | P     | LAGLIDADG              | M          |
| PpPPR_23   | P     | C                      | At3g59040  |
| PpPPR_24   | P     | M                      |            |
| PpPPR_25   | PLS   | M                      |            |
| PpPPR_26   | P     | -                      |            |
| PpPPR_27   | P     | -                      | At3g53170  |
| PpPPR_28   | P     | C                      |            |
| PpPPR_29   | P     | M                      |            |
| PpPPR_30   | P     | Smr                    | M          | At1g18900 |
| PpPPR_31   | PLS   | M*                     |            |
| PpPPR_32   | P     | C                      |            |
| PpPPR_33   | P     | M                      |            |
| PpPPR_34   | PLS   | C                      |            |
| PpPPR_35   | P     | M                      | At3g53170  |
| PpPPR_36   | P     | M                      |            |
| PpPPR_37   | P     | -                      |            |
| PpPPR_38   | P     | C*                     |            |

*Location indicates chloroplast (C), mitochondria (M), both C and M (C/M), either C or M, endoplasmic reticulum (ER), nucleus (Nuc) or unknown (-). Asterisks indicate location experimentally determined. Arabidopsis PPR proteins listed show more than 35% amino acid identity with PpPPR proteins, excluding PLS and DYW-class proteins.
Physcomitrella. However, we cannot exclude the possibility that non-DYW class PPR or non-PPR proteins (e.g., RRM type RNA-binding proteins) are necessary for recognition of the RNA editing site or the efficiency of RNA editing events together with DYW-class proteins in Physcomitrella.

In Physcomitrella plastids, RNA editing occurs at only one site in the translated region of rps14 mRNA. Plastid-localized PpPPR_45 is predicted to be a plastid rps14 RNA editing factor.

Table 2. List of Physcomitrella PPR proteins (continued)

| Name    | Class | Auxiliary domain/ motif | Location | Arabidopsis PPR protein homologous to PpPPR_# |
|---------|-------|-------------------------|----------|-----------------------------------------------|
| PpPPR_76 | P     | RRM M                  | At5g04810 (AtPPR4) |
| PpPPR_77 | DYW   | M*                     |          |
| PpPPR_78 | DYW   | M*                     |          |
| PpPPR_79 | DYW   | M*                     |          |
| PpPPR_80 | P     | C                      | At4g39620 (AtPPRS) |
| PpPPR_81 | P     | Smr M                  |          |
| PpPPR_82 | P     | C*                     |          |
| PpPPR_83 | P     | -                      | At2g41720 (EMB2651) |
| PpPPR_84 | P     | -                      |          |
| PpPPR_85 | P     | Smr C                  | At2g31400 (GUN1) |
| PpPPR_86 | P     | ER                     |          |
| PpPPR_87 | P     | M                      |          |
| PpPPR_88 | P     | M                      |          |
| PpPPR_89 | P     | M                      |          |
| PpPPR_90 | P     | C                      | At5g42310 |
| PpPPR_91 | DYW   | M*                     |          |
| PpPPR_92 | P     | C                      | At4g308252 |
| PpPPR_93 | P     | C                      |          |
| PpPPR_94 | P     | C                      | At4g308252 |
| PpPPR_95 | P     | C                      |          |
| PpPPR_96 | P     | Smr C*                 |          |
| PpPPR_97 | P     | M                      |          |
| PpPPR_98 | DYW   | M*                     |          |
| PpPPR_99 | P     | C                      |          |
| PpPPR_100 | P    | C                      | At2g30100 |
| PpPPR_101 | P    | C                      |          |
| PpPPR_102 | P    | C*                     |          |
| PpPPR_103 | P    | -                      |          |
| PpPPR_104 | P    | NYN C/M*               | At2g32230 (RPORP1) |
| PpPPR_105 | PLS  | C*                     |          |

*Location indicates chloroplast (C), mitochondria (M), both C and M (C/M), either C or M, endoplasmic reticulum (ER), nucleus (Nuc) or unknown (-). Asterisks indicate location experimentally determined. Arabidopsis PPR proteins listed show more than 35% amino acid identity with PpPPR proteins, excluding PLS and DYW-class proteins.
In contrast, PpPPR_43 protein is not involved in RNA editing, but is required for group II intron splicing of the mitochondrialdicistronic pre-mRNA. The DYW domain of PpPPR_43 is distinct from the other nine DYW domains of Physcomitrella PPR proteins.

**Function of the DYW domain.** DYW-class proteins are involved in RNA editing, RNA splicing, and RNA cleavage. This suggests that the DYW domain itself may have certain catalytic activity for target RNA species. There is a correlation between the presence of nuclear DYW genes and the occurrence of organelle RNA editing among land plants. Therefore, a hypothesis was provided in which the DYW domains are responsible for RNA editing in plant organelles and catalyze RNA editing. In fact, the DYW domain contains a conserved region, which includes invariant residues that match the active site of cytidine deaminases (C/HxE……CxxxC) from various organisms. However, cytidine deaminase activity was not detected by an in vitro assay using the recombinant DYW domain of Arabidopsis protein (At2g02980). Alternatively, recombinant DYW domains are found to possess endoribonuclease activity. Arabidopsis CRR2, a DYW-class PPR protein, is required for intergenic RNA cleavage of plastid rps7 and ndhB dicistronic pre-mRNA. The DYW domain of CRR2 has been shown to be indispensable for cleavage of the target RNA in vivo. The DYW domain contains the cytochrome f family heme-binding site signature (CxxCH), which overlaps with the active site of cytidine deaminase. Mutation of this signature to GxxGH resulted in a significant reduction of RNA cleavage activity. This indicates that the CxxCH motif is required for endoribonuclease activity of the DYW domain.

**Physcomitrella** DYW domains are well conserved (60–80% amino acid identities among DYW-class proteins, excluding PpPPR_43) and contain HSE….CxDCH residues. This suggests that **Physcomitrella** DYW domains may have potential endoribonuclease activity and/or cytidine deaminase activity. This possibility was tested and at least three DYW domains of PpPPR_56, 71, and 77 showed endoribonuclease-like activity (Fig. 2). Interestingly, its activity tightly depends on the substrate RNA used for the assay. For instance, the DYW of PpPPR_56 (r-56) digested ccmFe RNA but not nad3 RNA, whereas that of PpPPR_77 (r-77) rapidly cleaved the nad3 RNA but not ccmFe RNA. In contrast, r-71 efficiently degraded both RNAs. This implies that some DYW domains have potential RNA degrada- tion activity. In contrast, no cytidine deaminase activity was detected. As described above, Physcomitrella DYW-class proteins are involved in RNA editing but also may function in certain RNA processing events in organelles. This possibility will be further investigated.

**Perspectives**

Plastid genomes of land plants are relatively uniform in size and their gene content and organization are well conserved. However, mitochondrial genome structures largely differ between Physcomitrella and flowering plants. The extraordinarily large number of E/E+ and DYW-class PPR proteins in vascular plants can be correlated with large number of RNA editing sites in mitochondria. The number of P-class PPR proteins in Physcomitrella is less than half of those of flowering plants. This may reflect certain differences of regulatory processes in organelar gene expression between early land plants and flowering plants. Identification of all target RNA molecules recognized by Physcomitrella PPR proteins and characterization of their functions will provide clues for understanding the basal molecular mechanism of post-transcriptional regulation that evolved in land plant organelles.

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We thank Prof. Volker Knoop and Dr. Merike Rüdinger for communicating their unpublished results. This work was supported by JSPS KAKENHI Grant Number 23657003, 25291059 and by a Research Grant from DAIKO FOUNDATION (Nagoya).

**References**

1. Lurin C, André C, Aubourg S, Bellaux M, Bitton F, Bruyère C, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16:2089-103; PMID:15269332; http://dx.doi.org/10.1105/tpc.104.022236.

2. O’Toole N, Hattori M, Andrés C, Iida K, Lurin C, Schmitz-Linneweber C, et al. Large-scale analysis of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 2008; 25:1120-8; PMID:18343892; http://dx.doi.org/10.1093/molbev/mss057.

3. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribkova M, dePamphilis CW, et al. The Selaginella genome identifies genetic changes associated with the transition to vascular plants. Genome 2011; 54:960-3; PMID:21551031; http://dx.doi.org/10.1139/g10-034.

4. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 2008; 13:663-70; PMID:19004664; http://dx.doi.org/10.1016/j.tplants.2008.10.001.

5. Gutmann B, Gobert A, Giegé P. Mitochondrial genome evolution and the emergence of PPR proteins. Adv Bot Res 2012; 63:253-313; http://dx.doi.org/10.1016/B978-0-12-394279-1.00010-7.

6. Cene D, Bezaudin M, Harries P, Quatrano R, Moses S. As a model system for the study of metabolism and development. Annu Rev Plant Biol 2006; 57:497-520; PMID:16669772; http://dx.doi.org/10.1146/annurev.arplant.57.032905.105358.

7. Rensing SA, Lang D, Zimmer AD, Terry A, Salanova A, Shapiro H, et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 2008; 319:64-9; PMID:18079367; http://dx.doi.org/10.1126/science.1158666.

8. Hattori M, Hasebe M, Sugita M. Identification and characterization of cDNAs encoding pentatricopeptide repeat proteins in the moss Physcomitrella patens. Gene 2004; 343:305-11; PMID:15388585; http://dx.doi.org/10.1016/j.gene.2004.09.015.

9. Hattori M, Miyake H, Sugita M, A Pentatricopeptide repeat protein is required for RNA processing of cpDNA Pre-mRNA in moss chloroplasts. J Biol Chem 2007; 282:10773-82; PMID:17283080; http://dx.doi.org/10.1074/jbc.M608034200.

10. Ichinose M, Tasaki E, Sugita C, Sugita M. A PPR-DYW protein is required for splicing of a group II intron of cbc1 pre-mRNA in Physcomitrella patens. Plant J 2012; 70:271-8; PMID:22117821; http://dx.doi.org/10.1111/j.1365-313X.2011.04869.x.

11. Schwen G, Egner T, Fritawsky D, Granado J, Guirton MC, Hartmann N, et al. Large-scale analysis of 733 Physcomitrella plants transformed with different gene disruption libraries: production parameters and mutant phenotypes. Plant Biol (Stuttg) 2005; 7:228-37; PMID:15912442; http://dx.doi.org/10.1055/s-2005-837692.

12. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science 1994; 265:615-21; PMID:8036511; http://dx.doi.org/10.1126/science.265.5186.615.

13. Moreira D, Philippe H. Smr: a bacterial and eukaryotic homologue of the C-terminal region of the MutS family. Trends Biochem Sci 1999; 24:298-300; PMID:10411722; http://dx.doi.org/10.1016/S0968-0004(99)01419-X.

14. Bateman A. The structure of a domain common to archaeabacteria and the homocystinuria disease protein. Trends Biochem Sci 1997; 22:12-3; PMID:9020585; http://dx.doi.org/10.1016/S0968-0004(96)0046-7.
15. Anantharaman V, Aravind L. The NYN domains: novel predicted RNAs with a PIN domain-like fold. RNA Biol 2006; 3:18-27; PMID:17114934; http://dx.doi.org/10.4161/rna.3.1.2548.

16. Bryant N, Lloyd J, Sweeney C, Myouga F, Minke D. Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in Arabidopsis. Plant Physiol 2011; 155:1678-89; PMID:21139883; http://dx.doi.org/10.1104/pp.111.186120.

17. Majeron W, Frizzo G, Asakura Y, Qu X, Huang M, Ponnula L, et al. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. Plant Physiol 2012; 158:156-89; PMID:22065420; http://dx.doi.org/10.1104/pp.111.188474.

18. Christian JO, Braginets R, Schulze WX, Walther D. Characterization and prediction of protein phosphorylation hotspots in Arabidopsis thaliana. Front Plant Sci 2012; 3:207; PMID:22973286; http://dx.doi.org/10.3389/fpls.2012.00207.

19. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Carrier RE, Williams-Voelker PM, Krogger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rpl23 pre-mRNA. Plant Cell 2006; 18:1365-70; PMID:17041147; http://dx.doi.org/10.1105/tpc.106.046101.

20. Beick S, Schmitz-Linneweber C, Williams-Carrier R, Jensen B, Barkan A. The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. Mol Cell Biol 2008; 28:5357-47; PMID:18591259; http://dx.doi.org/10.1128/MCB.00563-08.

21. Johnson X, Westrkillof K, Finazzi G, Albers MR, Nott A, Mockler TC, Hong F, Wostrikoff K, Finazzi G, Kuras R, Johnson X, Wostrikoff K, Finazzi G, Kuras R, Kushwaha HR, Singh AK, Sopory SK, Singla-Pareek L, et al. Nucleoid-enriched proteomes in plastid biogenesis. Plant Physiol 2010; 154:1588-107; PMID:20888816; http://dx.doi.org/10.1104/pp.110.164111.

22. Kushwaha HR, Singh AK, Sopory SK, Singla-Pareek L. Reveals their developmental and selective pressures. BMC Genet 2005; 10:200; PMID:16030475; http://dx.doi.org/10.1111/j.1471-419X.2005.00222.x.

23. Koussevitzky S, Nott A, Mockler TC, Hong F, Koussevitzky S, Nott A, Mockler TC, Hong F, Monroy-Hernandez A, Hanson MR, et al. A chloroplast editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Natl Acad Sci USA 2012; 109:5104-9; PMID:22411807; http://dx.doi.org/10.1073/pnas.1202492109.

24. Sun T, Germain A, Gilette LS, Hammani K, Barkan A, Hanson MR, et al. An RNA recognition motif-containing protein is required for plastid RNA editing in Arabidopsis and maize. Proc Natl Acad Sci USA 2013; 110:E1169-78; PMID:23487777; http://dx.doi.org/10.1073/pnas.1202612110.

25. Miyata Y, Sugita C, Kobayashi Y, Hagiwara S, Sugita M. Chloroplast ribosomal 51S protein transcript is edited to create a translation initiation codon in the moss Physcomitrella patens. Biochim Biophys Acta 2002; 1576:346-9; PMID:12084583; http://dx.doi.org/10.1016/S0006-8995(01)00942-8.

26. Hattori M, Sugita M. A moss pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. Plant Physiol 2010; 154:1588-601; PMID:20935174; http://dx.doi.org/10.1104/pp.110.164111.

27. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Carrier RE, Williams-Voelker PM, Krogger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rpl23 pre-mRNA. Plant Cell 2006; 18:1365-70; PMID:17041147; http://dx.doi.org/10.1105/tpc.106.046101.

28. Liu X, Yu F, Roderer M. An Arabidopsis pentatricopeptide repeat protein, SUPPRESSOR OF VARIEGATION7, is required for FtsH-mediated chloroplast biogenesis. Plant Physiol 2010; 154:1588-601; PMID:20935174; http://dx.doi.org/10.1104/pp.110.164111.

29. Knopf V, Rüdinger M. DYT-type PPR proteins in a heterologous system: plant RNA editing factors involved in an ancient horizontal gene transfer? FEBS Lett 2010; 584:4287-91; PMID:20888816; http://dx.doi.org/10.1101/tpc.1097090; http://dx.doi.org/10.1101/tpc.1097090.

30. Rüdinger M, Szövényi P, Rensing SA, Knoop V. Assigning DYT-type PPR proteins to RNA editing sites in the funaria moss Physcomitrella patens and Funaria hygrometrica. Plant Cell 2011; 67:570-80; PMID:21466601; http://dx.doi.org/10.1111/j.1365-313X.2011.04600.x.

31. Rüdinger M, Polsakiewicz M, Knopf V. Organelar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jumengianthus but not in marchantid liverworts. Mol Ecol Evol 2008; 25:1450-14; PMID:18400790; http://dx.doi.org/10.1111/j.1365-313X.2008.04175.x.

32. Rüdinger M, Funk HT, Rensing SA, Maier UG, Knopf V. RNA editing: only eleven sites are present in the Physcomitrella patens mitochondrial transcriptome and a universal nomenclature proposal. Mol Genet Genomics 2009; 281:473-81; PMID:19106717; http://dx.doi.org/10.1007/s00438-009-0424-z.

33. Tasaki E, Hattori M, Sugita M. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmF transcript. Plant J 2010; 62:560-70; PMID:20163555; http://dx.doi.org/10.1111/j.1365-313X.2010.04417.x.

34. Ohtani S, Ichinose M, Tasaki E, Aoki Y, Komura Y, Sugita M. Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. Plant Cell Physiol 2010; 51:942-9; PMID:20283750; http://dx.doi.org/10.1093/pcp/pcp142.

35. Tasaki E, Hattori M, Sugita M. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmF transcript. Plant J 2010; 62:560-70; PMID:20163555; http://dx.doi.org/10.1111/j.1365-313X.2010.04417.x.

36. Tasaki E, Hattori M, Sugita M. The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmF transcript. Plant J 2010; 62:560-70; PMID:20163555; http://dx.doi.org/10.1111/j.1365-313X.2010.04417.x.

37. Hayes ML, Giang K, Mulligan RM. Molecular evolution of pentatricopeptide repeat genes reveals translocation in species lacking an editing target and structural domains under distinct selective pressures. BMC Evol Biol 2012; 12:66; PMID:22583635; http://dx.doi.org/10.1186/1471-2148-12-66.

38. Takemura K, Zehren M, Verbrink D, Kugelmugel H, Martel B, Brounnicke A. Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. Proc Natl Acad Sci USA 2012; 109:5104-9; PMID:22411807; http://dx.doi.org/10.1073/pnas.1202492109.

39. Fujii S, Small I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 2011; 191:37-47; PMID:21557747; http://dx.doi.org/10.1111/j.1469-8137.2011.03746.x.