An evolutionarily conserved P-subfamily pentatricopeptide repeat protein is required to splice the plastid *ndhA* transcript in the moss *Physcomitrella patens* and *Arabidopsis thaliana*

Ayaka Ito[1], Chieko Sugita[1], Mizuho Ichinose[1,2], Yoshinobu Kato[3], Hiroshi Yamamoto[3], Toshiharu Shikanai[3], and Mamoru Sugita[1,*]

[1]Center for Gene Research, Nagoya University, Nagoya, 464-8602 Japan

[2]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, 464-8602 Japan

[3]Department of Botany, Graduate School of Science, Kyoto University, Kyoto, 606-0076, Japan

(The first two authors equally contributed to this work)

*Corresponding author:

E-mail. sugita@gene.nagoya-u.ac.jp

Running title: A plastid *ndhA* intron-specific PPR splicing factor

Keywords: P-subfamily PPR protein, splicing, *ndhA*, plastid, *Physcomitrella patens*, *Arabidopsis thaliana*

Abstract

Pentatricopeptide repeat (PPR) proteins are known to play important roles in posttranscriptional regulation in plant organelles. However, the function of the majority of PPR proteins remains unknown. To examine their functions, *Physcomitrella patens PpPPR_66* knockout (KO) mutants were generated and characterized. KO mosses exhibited a wild type-like growth phenotype but showed aberrant chlorophyll fluorescence due to defects in chloroplast NADH dehydrogenase-like (NDH) activity. Immuno-blot analysis suggested that disruption of *PpPPR_66* led to a complete loss of the chloroplast NDH complex. To examine whether the loss of PpPPR_66 affects the expression of plastid *ndh* genes, the transcript levels of 11 plastid *ndh* genes were analyzed by reverse-transcription PCR. This analysis indicated that splicing of the *ndhA* transcript was specifically impaired while the mRNA accumulation levels as well as the processing patterns of other plastid *ndh* genes were not affected in the KO mutants. Complemented *PpPPR_66* KO lines transformed with the *PpPPR_66* full-length cDNA rescued splicing of the *ndhA* transcript. *Arabidopsis thaliana* T-DNA tagged lines of a *PPR_66* homolog (*At2g35130*) showed deficient splicing of the *ndhA* transcript. This indicates that the two proteins are functionally conserved between bryophytes and vascular plants. An in vitro RNA binding assay demonstrated that the recombinant PpPPR_66 bound preferentially to the region encompassing from a part of exon 1 to a 5' part of the *ndhA* group II intron. Taken together, these results indicate that PpPPR_66 acts as a specific factor to splice *ndhA* pre-mRNA.

Introduction

The nuclear-encoded pentatricopeptide repeat (PPR) proteins are synthesized in the cytoplasm, but are then posttranslationally imported into plastids or mitochondria or both, where they function in various RNA processing steps (Small and Peeters, 2000, Lurin *et al*. 2004, Colcombet *et al*. 2013). The PPR gene family exists ubiquitously in eukaryotes, but has expanded considerably to 450 PPR genes in *Arabidopsis thaliana* to over 1,000 in the spikemoss *Selaginella moellendorffii* (Cheng *et al*. 2016). The PPR proteins are divided into P and PLS subfamilies. The P subfamily contains only canonical PPR (P) motifs while the PLS subfamily consists of repeated blocks of P and PPR-like (L, S) motifs. The PLS subfamily is unique to the plant kingdom and has been further divided into PLS, E/E+, and DYW classes (Lurin *et al*. 2004). Among them, P subfamily PPR proteins make up more than half of seed plants. The loss-of-function of PPR proteins often leads to defects in organellar function, such as photosynthesis and respiration (Schmitz-Linneweber and Small 2008).

The P-subfamily PPR proteins are reportedly involved in intergenic RNA processing and translation (Barkan *et al*. 1994, Fisk *et al*. 1999, Meierhoff *et al*. 2003), *trans*-splicing (Schmitz-Linneweber *et al*. 2006) and *cis*-splicing (Falcon de Longevialle *et al*. 2008, Khrouchtchova *et al*. 2012), as well as the stabilization of plastid mRNAs (Pfalz *et al*. 2009, Johnson *et al*. 2010) and tRNA (Beick *et al*. 2008). In addition, P-subfamily PPR proteins with a small MutS-related (Smr) domain were shown to be involved in processing of chloroplast 23*S*-4.5*S* pre-rRNA (Zoschke *et al*. 2016, Wu *et al*. 2016). On the other hand, the PLS-subfamily proteins mostly function as an RNA editing site-recognition factor for both mitochondrial and plastid transcripts (Fujii and Small 2011, Takenaka *et al*. 2013, Ichinose and Sugita 2017). Besides, some PLS members are reportedly required for RNA splicing of specific transcripts (Chateigner-Boutin *et al*. 2011, Ichinose *et al*. 2012, Zhang *et al*. 2015). Thus, both subfamilies bind to RNAs in a gene-specific manner and contribute to various types of RNA processing steps (Barkan and Small 2014). However, the function of most P-subfamily PPR proteins is unknown.

The bryophyte moss *Physcomitrella patens* possesses ~105 PPR proteins, over 80% of which are members of the P subfamily member (Sugita *et al*. 2013). To date, the following four P-subfamily PPR proteins were investigated to reveal their function in *P. patens*. PpPPR_38 was involved in the maturation of *clpP* pre-mRNA (Hattori *et al*. 2007, Hattori and Sugita 2009). PpPPR_67 and 104 were required for plastid tRNA maturation (Sugita *et al*. 2014). PpPPR_4 was recently shown to play a role in plastid $tRNA^{Ile}$ splicing (Goto *et al*. 2016). Several moss P-subfamily proteins are conserved at the

amino acid sequence level to functionally characterized Arabidopsis and maize PPR proteins (Sugita *et al*. 2013). PpPPR_67 and 104 are functionally related to Arabidopsis PRORP1 to 3 (Gobert *et al*. 2010). However, the function of putative orthologous PPR proteins has yet to be elucidated. To clarify the function of P-subfamily proteins, we constructed a series of *PPR* gene knockout (KO) mutants from *P. patens*. Here we describe the functional characterization of P-subfamily PpPPR_66 protein, which is essential for splicing of plastid *ndhA* transcript. The recombinant PpPPR_66 bound preferentially to the 5' part of domain I of the *ndhA* group II intron. PpPPR_66 may function as a specific factor for RNA splicing of *ndhA* pre-mRNA.

Results

PpPPR_66 is targeted to chloroplasts

The *PpPPR_66* gene (Pp3c16_5890/Pp1s15_385) encodes a polypeptide of 578 amino acids (aa) that consists of an N-terminal transit peptide and a PPR tract composed of 11 PPR motifs (Figs. 1a and S1). The TargetP program (Emanuelsson *et al*. 2000) predicted PpPPR_66 to be localized in plastids. To investigate its subcellular location, a fusion protein, which is composed of its N-terminal 121 aa and green fluorescent protein (GFP), was transiently expressed in the moss protonemal cells. GFP fluorescence was observed in the chloroplasts but not in mitochondria or the cytoplasm (Fig. S2). The PpPPR_66 homologs, which we here refer to as PPR66L, are found in a wide range of land plants, including the liverwort *Marchantia polymorpha*, the spikemoss *Selaginella moellendorffii*, *Arabidopsis thaliana*, and *Zea mays* (maize) (Fig. S1). The *PpPPR_66* gene is interrupted by eight introns (Fig. S1). Intron positions of *PPR66L* genes are conserved with those of the *PpPPR_66* gene. This indicates that PpPPR_66 and PPR66Ls are orthologous. In addition, PpPPR_72 (Pp3c6_26210/Pp1s53_63) is likely a paralog of PpPPR_66 (Fig. S1).

Chloroplast NDH activity is defective in *PpPPR_66* KO mutants

For loss-of-function analysis of *PpPPR_66*, we generated *PpPPR_66* KO lines by replacing its coding region with a cassette carrying the *gfp* and drug resistant (*hpt*) genes via homologous recombination (Fig. 1b). We confirmed, by genomic-PCR analysis, that recombination occurred in each of the designed targeted loci (Fig. S3). In the KO mutant lines (Δ*66*-2, Δ*66*-3), we also confirmed, by RT-PCR analysis, that *PpPPR_66* transcript was not detected (Fig. 1c). The *PpPPR_66* KO mosses displayed a wild type-like growth phenotype (Fig. 1b). Then, we investigated the photosynthetic status by kinetics

multispectral fluorescence imaging. Chlorophyll fluorescence parameters, such as F_v/F_m, (the integrity of photosystem II (PSII)) and ΦPSII (the effective quantum yield of PSII), were almost the same in WT and KO mosses (Table S1). This suggests that photosynthesis was not impaired in the KO mutants. To investigate chloroplast NDH activity, we monitored chlorophyll fluorescence in the mosses with a pulse amplitude modulation (PAM) chlorophyll fluorometer. This measurement showed that the transient increase of chlorophyll fluorescence after turning off actinic light appeared in the WT but not in the KO mutants (Fig. 2). It is well known that this change in fluorescence represents NDH activity in chloroplasts (Shikanai 2016). Thus, this result indicates that chloroplast NDH activity was lost in the *PpPPR_66* KO mutants.

Immuno-blot analysis did not detect NdhM and PnsB1 subunits of the NDH complex in the KO mutants (Fig. 3). In contrast, cytochrome *f* of the cytochrome b_6f complex (Cyt*f*), the β subunit of H^+-ATP synthase (AtpB), and the PSII reaction center D1 protein (PsbA), accumulated at similar levels in WT and KO mosses. Accumulation of NdhM and PnsB1 depends of the core NDH subunit, NdhB, in the liverwort *Marchantia polymorpha* (Ueda *et al*. 2012). This result suggests that the entire NDH complex was likely lost or drastically decreased in the KO mutants.

_PpPPR_66_ KO mutants completely lost splicing of plastid _ndhA_ transcript

Loss of PPR proteins frequently leads to impaired RNA processing, including splicing of plastid transcripts (Barkan and Small 2014). Since a loss of NDH activity and the NDH complex was observed in the *PpPPR_66* KO mosses, RNA maturation of plastid *ndh* genes, such as RNA stability and/or RNA splicing, might be affected in the KO mosses. To verify this, RT-PCR analysis was performed to investigate mRNA levels of 11 *ndh* genes, *ndhA* to *ndhK*, which are located at four different positions in the plastid genome (Fig. 4). This analysis showed that spliced *ndhA* transcript (895-bp amplicon by RT-PCR) did not accumulate while unspliced *ndhA* transcript (1,585-bp amplicon) accumulated considerably in the KO mosses (Fig. 4). In contrast, the other *ndh* transcripts accumulated at similar levels in both WT and KO mosses. To verify this result, we performed RNA gel blot hybridization of an *ndhA*-containing gene cluster. Probing with the *rps15*, *ndhH*, *ndhA*, *ndhI-G-E* and *psaC* sequences, a 6.3-kb transcript was detected in the WT but not in the KO mosses. Instead of the 6.3-kb transcript, a longer transcript (7 kb) was detected in the KO mosses (Fig. 5). The 7-kb transcript might be a primary transcript encompassing an entire gene cluster from *rps15* to *ndhD*. The 6.3-kb transcript could be produced from the 7-kb primary transcript after splicing of the *ndhA* intron. To verify this result, an *ndhA* intron-specific probe (Int) and a 3' exon-specific probe (Ex) were used for northern blot analysis.

Probing with the probe Ex, the 6.3-kb transcript was detected in the WT but not in the KO mosses and the 7-kb band was detected in the KO mosses but not in WT. When the probe Int was used, the 6.3-kb transcript was not detected in both the WT and KO mosses. On the other hand, the 7-kb band was detected weakly in the WT but strongly in the KO mosses. This result confirmed that the 6.3-kb and 7-kb bands were spliced and unspliced *ndhA* transcripts, respectively. A 0.7-kb band was detected in WT but not in KO mutants, suggesting an exised intron. Approximately 3-kb transcripts were shown to be unspliced *ndhA* precursors because they were detected by both Int and Ex probes. However, shorter *ndhA* transcripts, which could be produced from the 3-kb unspliced *ndhA* precursors, were not detected, in *P. patens*.

The *psaC* probe detected a strong band of 0.4 kb in both WT and KO mutants, which might be produced from a polycistronic transcript or could be transcribed by using a dedicated promoter. Although we do not know whether the 0.4-kb transcript was transcriptionally or posttranscriptionally produced, the strong signal that was detected suggests that *psaC* mRNA is extremely stable, unlike *ndh* transcripts, in *P. patens*.

As a result of defects to *ndhA* splicing in the KO mosses, we examined the possibility that PpPPR_66 is involved in splicing of other intron-containing transcripts. There are 12 protein-coding genes and six tRNA genes, which contain intron(s) in *P. patens* plastids (Sugiura *et al*. 2003). To assess whether splicing of mRNAs and pre-tRNAs was affected in the KO mutants, RT-PCR analysis was carried out using exon-specific primers. Spliced transcripts of intron(s)-containing genes except for *ndhA* in the KO mutants accumulated to similar levels as those in the WT (Figs. S4, S5). These results strongly suggest that PpPPR_66 is specifically required in splicing of the *ndhA* intron. However, we cannot exclude the possibility that plastid genes other than the genes examined were affected in the KO mutants.

To confirm that PpPPR_66 is essential for *ndhA* splicing, we generated moss transformants that expressed *PpPPR_66* full-sized cDNA in the KO mutant Δ*66*-3. Two independent complemented moss plants restored splicing of *ndhA* (Fig. S6). This complementation experiment confirmed that the *ndhA* splicing defect in the KO mutants was caused by a loss-of-function of the *PpPPR_66* gene.

A. thaliana PPR66L is involved in _ndhA_ splicing

PpPPR_66 showed 44% aa identity and 81% similarity to Arabidopsis PPR66L (At2g35130). The *AtPPR66L* gene is interrupted by seven introns and their positions are identical to the intron positions of the *PpPPR_66* gene (Fig. S1). This strongly suggests that Arabidopsis PPR66L may be a functional

ortholog of PpPPR_66. To investigate this possibility, we analyzed the Arabidopsis *PPR66L* (*At2g35130*) KO null mutant lines, SALK_043507 and SALK_065137 (Fig. S7**)**. We measured chloroplast NDH activity in vivo by a PAM chlorophyll fluorometer and found that both KO mutants exhibited no chloroplast NDH activity (Fig. 6a). However, photosynthetic parameters were not affected in the KO mutants (Fig. S8). Furthermore, both KO mutants showed no visible phenotype under our growth conditions as previously reported in other *ndh* mutants (Hashimoto *et al*. 2003, Yamamoto *et al*. 2011). Then, we investigated the splicing status of *ndhA* transcript by RT-PCR. This analysis showed that *ndhA* transcript was not spliced in the KO mutants (Fig. 6b). To verify the splicing defect suggested by RT-PCR analysis, we carried out RNA gel blot hybridization. In Arabidopsis, the *ndhA*-containing gene cluster was also transcribed as a polycistronic precursor, but was then heavily processed (Fig. 6c). An exon-specific antisense RNA probe (Exon probe) detected several discrete *ndhA* transcripts, of which four RNA bands (marked by open circles in Fig. 6c) were detected in the WT but not in the KO mutants. The 1.2-kb RNA detected in the WT is presumably a mature *ndhA* mRNA and other RNA bands are likely spliced *ndhA* transcript precursors. Instead of these RNA bands, four transcripts (marked by closed circles) detected in the KO mutants were each 1-kb longer than the WT-specific transcripts. These shifted RNAs are expected to be unspliced *ndhA* pre-mRNAs because the Arabidopsis *ndhA* intron is 1,080 nucleotides (nt) in length. This was confirmed by RNA gel blot hybridization using the intron-specific probe (Intron probe, Fig. 6c). A 1-kb strong hybridization signal (marked by an arrowhead) detected in WT probably represents an excised intron. These results indicate that the Arabidopsis PPR66L is also involved in *ndhA* splicing, similar to PpPPR_66.

To test whether Arabidopsis *PPR66L* rescues the splicing defect of *ndhA* pre-mRNA in the moss *PpPPR_66* KO mutant, we performed a complementation experiment. However, Arabidopsis *PPR66L* did not restore *ndhA* splicing in the moss KO mutant (Fig. S6).

PpPPR_66 binds to the 5' half of domain I of the *ndhA* intron

Since PpPPR_66 is required for *ndhA* splicing, PpPPR_66 is expected to bind to some region within the *ndhA* intron as occurs in other PPR splicing factors (Falcon de Longevialle *et al*. 2008, Khrouchtchova *et al*. 2012, Goto *et al*. 2016). To investigate whether PpPPR_66 binds to the 690-nt group II intron of *ndhA*, we carried out an RNA electrophoresis mobility shift assay (REMSA) using three overlapping RNA probes, RNA1 to RNA3 (Fig. 7a). RNA1 (280 nt) covered a 3' part (56 nt) of exon 1 and a 5' part (224 nt) of the intron. RNA2 (250 nt) covered the middle part of the intron, and RNA3 (292 nt) encompassed a 3' part (238 nt) of the intron and a 5' part (54 nt) of exon 2. RNA1 and

RNA2 can form a part of domain I of the group II intron. RNA3 consists of domains II to VI. For REMSA, we prepared 70-kDa recombinant PpPPR_66 (rPPR66) fused to thioredoxin (Trx) at its N-terminus. As shown in Fig. 7b, clear shifted bands were detected by the RNA1 probe. This suggests that PpPPR_66 binds preferentially to the 5' half of domain I of the *ndhA* group II intron with high affinity (using only 12.5 nM of rPPR66). To further investigate which part of the 5' half of domain I was involved, we performed REMSA using three overlapping RNA probes, RNA1a (115 nt), 1b (90 nt) and 1c (100 nt). RNA1a bound preferentially to rPPR66 while RNA1b was weakly but RNA1c was rarely bound. RNA1a covered a 3' part (56 nt) of exon 1 and a 5' part (59 nt) of the intron, which can form a long stem-loop structure. This result suggests that PpPPR_66 may preferentially bind to some site in the 115-nt region extending from a part of exon 1 to the intron.

To specify possible binding sites, we predicted a target sequence, 5'-Y-Y-Y-A-G-U-Y-x-U-U-G-3', recognized by the 11 PPR motifs of PpPPR_66 according to the RNA recognition code of PPR (Barkan *et al.* 2012, Yagi *et al*. 2013). We scanned this predicted target sequence through the *P. patens* plastid genome and found many matching sites on the plastid genome but no matching sequences within the *ndhA* intron. This suggests that PpPPR_66 might not bind to a co-linear RNA sequence but instead interact with complex RNA structures in the group II intron.

Discussion

In this study, we identified PpPPR_66 as a plastid *ndhA*-specific splicing factor. A loss-of-function of PpPPR_66 resulted in impaired splicing of the *ndhA* intron and subsequently led to the loss of chloroplast NDH activity and accumulation of the NDH complex. With the exception of the lack of NDH activity, KO mutants did not display any phenotype in photosynthetic electron transport, which is consistent with other PPR mutants with specifically impaired NDH activity (Hashimoto *et al*. 2003, Kotera *et al*. 2005, Shikanai 2016). In the Arabidopsis *crr2* mutants defective in CHLORORESPIRATORY REDUCTION 2 (CRR2) PPR protein, NDH activity was lost and accumulation of the NDH complex was impaired but photosynthetic electron transport was unaffected (Hashimoto *et al*. 2003). CRR2 functions in the intergenic processing of plastid RNA between *rps7* and *ndhB*, which may be essential for *ndhB* translation (Hashimoto *et al*. 2003). Unspliced *ndhA* mRNA could not be translated in the *PpPPR_66* KO mutants, and rested in the absence of NDH activity, as observed in the *crr2* mutant. Likewise, PPR66L proteins from other plant species might also be involved in *ndhA* intron splicing and NDH activity as shown by analysis of the Arabidopsis *PPR66L* KO

mutants. However, we cannot exclude the possibility that PpPPR_66 and PPR66L participate in certain functions other than *ndhA* splicing.

PpPPR_66-like homologs are widely distributed in land plants but not in green algae, *Chlamydomonas*, *Volvox*, and *Chlorella*. Streptophyta (charophytes and land plants) have plastid *ndhA* genes with a group II intron while green algae have an intron-less *ndhA* gene in the plastid genome. Thus, there is likely coevolution of PPR66L and the plastid *ndhA* intron in the land plant lineage. Unlike most land plants, *P. patens* has PpPPR_66 and its paralog, PpPPR_72. Since their RNA recognition codes are almost identical (Fig. S1), PpPPR_72 might be functionally redundant with PpPPR_66. As presented in this study, however, *ndhA* splicing was impaired in the *PpPPR_66* KO mutants, which retained an intact *PpPPR_72* gene. It is possible that PpPPR_72 may be localized in other intracellular compartments rather than in chloroplasts. If this is true, PpPPR_66 and 72 may have different functions in these respective compartments.

Several P-subfamily PPR proteins involved in splicing of plastid group II introns have been identified. Maize PPR4, which harbors an RNA recognition motif and 16 PPR motifs, was shown to facilitate *rps12 trans*-splicing through direct interaction with intron RNA (Schmitz-Linneweber *et al*. 2006). Maize PPR5 is involved in splicing or stability of pre-tRNA^{Gly} (Beick *et al*. 2008). Maize 9, line 258ORGANELLE TRANSCRIPT PROCESSING 51 (OTP51) with 10 PPR motifs and two C-terminal LAGLIDADG motifs is required for *cis*-splicing of the *ycf3*-2 intron (Falcon de Longevialle *et al*. 2008). Maize and *Arabidopsis* THYLAKOID ASSEMBLY 8 (THA8), which possesses four PPR motifs, are involved in splicing of both *ycf3*-2 and tRNA^{Ala} introns (Khrouchtchova *et al*. 2012). A gene disruption mutant *osppr4* by insertion of *Tos17* from rice OsPPR4, a homolog of maize PPR4, and its knockdown mutants produced by RNAi led to a strong defect in the *cis*-splicing of *atpF*, *ndhA*, *rpl2*, and *rps12* introns (Asano *et al*. 2013). In addition, *trans*-splicing of *rps12* was also defected in *osppr4*. *P. patens* PpPPR_4, which is not related to maize and rice PPR4, was recently identified as an RNA binding protein required for splicing of pre-tRNA^{Ile} (Goto *et al*. 2016). In addition to P-subfamily proteins, PLS-subfamily PPR proteins also are known as splicing factors. For instance, a mutation of the Arabidopsis PLS-class PPR gene *PIGMENT-DEFICIENT MUTANT 1* (*PDM1*), which is also known as *SEEDLING LETHAL 1* (*SEL1*) (Pyo *et al*. 2013), resulted in splicing deficient of *ndhA*, *trnK*, and *rps12*-2 introns (Zhang *et al*. 2015). Splicing efficiency of the *ndhA* transcript in the *pdm1-1* mutant was reduced to 10% relative to that in WT plants (Zhang *et al*. 2015). SEL1 protein has previously been shown to be an RNA editing factor for *accD* sites (Pyo *et al*. 2013). Among these PPR proteins involved in splicing, PPR5, THA8, OTP51, and PpPPR_4 were demonstrated to bind their target introns *in vitro*.

Maize OTP51 and THA8 bound the first 197 nt of the *ycf3*-2 intron with high and low affinity, respectively (Khrouchtchova *et al*. 2012). PPR5 binds to a 40-nt single-strand region within domain I of the $tRNA^{Gly}$ group II intron (Williams-Carrier *et al*. 2008). Thus, these three PPR proteins bind some region in domain I of the respective group II intron. In contrast, *P. patens* PPR_4 binds to domain III of its target $tRNA^{Ile}$ group II intron (Goto *et al*. 2016). In the present study, PpPPR_66 was shown to bind preferentially to the 115-nt region encompassing from a part of exon 1 to the *ndhA* group II intron.

In seed plants, splicing of the *ndhA* transcript is known to require several nuclear-encoded factors, including CHLOROPLAST RNA SPLICING 2 (CRS2) (Jenkins et al. 1997), CRS2-ASSOCIATED FACTOR 1 and 2 (CAF1 and CAF2) (Ostheimer *et al*. 2003), and CHLOROPLAST RNA SPLICING AND RIBOSOME MATURATION (CRM) FAMILY MEMBER 2 (CFM2) (Asakura and Barkan 2007). These factors do not possess PPR motifs. A mutation in the *crs2* gene partially or completely blocks the splicing of nine respective plastid group II introns, including the *ndhA* intron (Jenkins et al. 1997). In this case, the defects of mRNA splicing in *crs2* mutants is likely to be a consequence of the plastid ribosome deficiency that in turn results from a failure to splice the *rps16* and *rpl16* mRNAs (Jenkins et al. 1997). Like *crs2* mutants, *ndhA* splicing is likely sensitive in stressed plants such as *dpm1*/*sel1* mutants and *ppr4* mutants that display an albino or pale green phenotype. *ndhA* splicing defects in these mutants may be a secondary effect due to ribosome deficiency, loss of RNA editing and/or photosynthetic activity. In contrast, PpPPR_66 binds to domain I of the *ndhA* intron as shown in this study and *ndhA* splicing was almost completely blocked in the *PpPPR_66* KO mutants. This supports that PpPPR_66 is an *ndhA* intron-specific splicing factor, even if it interacts with other splicing factors.

As presented here, Arabidopsis PPR66L is also involved in *ndhA* splicing. However, Arabidopsis *PPR66L* cDNA did not rescue *ndhA* splicing deficient in the *PpPPR_66* KO mutant (Fig. S6). It is intriguing that crucial 11 RNA recognition codes differ by three between PpPPR_66 and AtPPR66L (Fig. S1). PpPPR_66 has a PPR2 (N/G)-PPR3 (N/A)-PPR4 (T/N) tract while AtPPR66L has a PPR2 (A/S)-PPR3 (N/N)-PPR4 (N/N) repeat. These differences may be related to the result in which AtPPR66L was unable to complement the *PpPPR_66* KO phenotype. Although PpPPR_66 and Arabidopsis PPR66L are required for *ndhA* splicing, their mode of action for splicing *ndhA* transcript might differ slightly between *P. patens* and Arabidopsis. This possibility remains to be further assessed.

Experimental Procedures

312

313 **Plant growth condition**

314 *P. patens* was grown in a controlled-environmental culture room at 25°C under continuous light (~30
315 μmol photons m^{-2} s^{-1}) either on the minimal medium (BCD) or on the minimal medium supplemented
316 with 5 mM diammonium (+)-tartrate (BCDAT) (Nishiyama *et al.* 2000). *A. thaliana* wild type (Columbia)
317 and T-DNA tagged lines were grown in soil at 23°C for 3 to 4 weeks as described (Yamamoto *et al*.
318 2011). The T-DNA-tagged lines SALK_043507 and SALK_065137 were provided by Arabidopsis
319 Biological Resource Center (ABRC).

320

321 **Subcellular localization**

322 DNA-free RNA was isolated and reverse-transcribed to synthesize cDNA (Ichinose *et al*. 2013). A
323 cDNA coding for the N-terminal 121 aa of PpPPR_66 was amplified using specific primers (Table S2),
324 and cloned into pKSPGFP9 (Tasaki *et al.* 2010). The obtained p66N-GFP was introduced by particle
325 bombardment using the IKDA GIE-III biolistic gun (Tanaka Co., Ltd., Hokkaido) into the transgenic
326 Mt-RFP OX moss expressing the mitochondria-localized RFP protein and fluorescence emitted from
327 expressed fusion proteins was observed as described (Ichinose *et al*. 2013).

328

329 **Plasmid construction and moss transformation**

330 The 1,002 bp region upstream from *PpPPR_66* was PCR-amplified from genomic DNA with 66KO-1
331 and 66KO-2 primers (Table S2) and the 1,157 bp region downstream of *PpPPR_66* was amplified with
332 66KO-3 and 66KO-4 primers. The amplified 1,002 bp DNA was cloned into pNGH4 (a derivative of
333 pKI-GFP, Ichinose *et al*. 2013). The resulting plasmid was cut with *Sma*I and ligated with the amplified
334 1,157 bp downstream DNA. The resultant plasmid, p66KO, carried the 1,002 bp upstream region, the
335 *gfp* and *hpt* gene expression cassette derived from pNGH4, and 1,157 bp downstream of *PpPPR_66*
336 in this order. p66KO was cut with *Nae*I and introduced into the protonemata using the biolistic gun, and
337 hygromycin-resistant moss colonies were selected. Gene disruption in transformants was verified by
338 genomic PCR with appropriate primers (Table S2, Fig. S3). To verify null KO mutants, RT-PCR was
339 performed using cDNA, SapphireAmp Fast Master Mix (TaKaRa Bio Inc.), and the 66P3 and 66P4
340 primers (Table S2).

341

342 **Generation of complemented mosses**

343 Full-length *PpPPR_66* cDNA was PCR-amplified with the primers 66P3 and 66P4 and *AtPPR66L*

cDNA with primers AtPPR66h-F and AtPPR66h-stopR (Table S2), and inserted into p9WmycZ3 (Goto *et al*. 2016). The resulting plasmid was cut with *Not*I and introduced into the KO mutant Δ*66*-3 using the biolistic gun and zeocin-resistant mosses were selected. The *PpPPR_66* cDNA was sequenced in its entirety and deposited in the DDBJ DNA database under accession number LC335802.

Analysis of chlorophyll fluorescence

Chlorophyll fluorescence from mosses was analyzed with a FluorCam 800MF (Photon System Instruments) and Mini-PAM (WALZ). Three-week-old protonemata were adapted in the dark for 10 min before measurement of F_v/F_m and ΦPSII. Minimum fluorescence (F_o) was determined by a weak red light and maximum fluorescence of the dark-adapted state (F_m) was measured during a subsequent saturating pulse (SP, 1400 μmol photons m^{-2} s^{-1} for 0.8 s). The protonemata were then illuminated with actinic red light (AL, 96 μmol photons m^{-2} s^{-1}) for 1 min. The activity of chloroplast NDH in vivo was analyzed by monitoring the transient increase in chlorophyll fluorescence after turning AL (apparent F_o) off.

Immuno-blot analysis

Total membrane proteins were extracted from the moss protonemata. Membrane protein extracts corresponding to 2 μg chlorophyll (100%) were separated on 0.1% SDS-14% polyacrylamide gels and blotted to nylon membranes. Immunodetection was carried out according to ECL Prime (GE Healthcare UK Ltd.) protocols. Antibodies against PsbA (AS05084A, Agrisera), anti-β-subunit of H^+-ATP synthase (provided by T. Hisabori, Tokyo Institute of Technology), and anti-liverwort NdhM (Ueda *et al*. 2012) were used. Antibodies against Arabidopsis PnsB1 and rice cytochrome *f* were kindly provided by T. Endo (Kyoto University) and A. Makino (Tohoku University), respectively.

RT-PCR and northern blot analysis

Total cellular RNA was isolated from four-day-old *P. patens* protonemata and two-week-old *A. thaliana* plants. Preparation of cDNA was performed using the ReverTra Ace qPCR RT Kit (TOYOBO). RT-PCR was performed using the primers listed in Tables S2 and S3, as described previously (Goto *et al*. 2016). For northern blot analysis, RNA (10 μg) was loaded onto a 1% agarose gel containing formaldehyde and transferred to a nylon membrane. The blotted RNAs were hybridized with gene-specific DNA probes amplified using appropriate primers (Goto *et al*. 2016, Table S2). Antisense RNA probes labeled with digoxigenin-UTP (Roche) were prepared as follows. For the exon-specific RNA probe, a

spliced *ndhA* was amplified from Arabidopsis cDNA with the primers, ndhA-At172F and ndhA-At2062R, and was cloned into the *Sma*I site of pBluescript SK(+). The resultant plasmid was linearized by *Bam*HI digestion and used as a DNA template. For the intron-specific RNA probe, an *ndhA* intron region containing T7 promoter was amplified from the Arabidopsis genome DNA with the primers, T7-AtndhAi-1071R and AtndhAi298F. Respective DNA templates were transcribed with T7 RNA polymerase (TaKaRa) and antisense RNA probes were used for northern blot hybridization.

Recombinant protein

cDNA encoding the PpPPR_66 without its N-terminal 52 aa was amplified using specific primers (Table S2) and was cloned into pBAD/Thio-TOPO (Invitrogen). The recombinant protein, rPPR66, was expressed at 16°C for 16 h in *Escherichia coli* XL1-blue, and recovered using Ni-NTA agarose (Qiagen).

REMSA

DNA templates for *in vitro* transcription were prepared by PCR using appropriate primers (Table S2) and were transcribed with T7 RNA polymerase (TaKaRa) and ribonucleotides including [^{32}P] UTP. For REMSA, the recombinant protein was incubated for 10 min at 25°C in the reaction mixture (Goto et al. 2016) and then heat-denatured ^{32}P-labeled *in vitro* transcribed RNA probes (or chemically synthesized oligo RNA probes) were added, and incubated for 15 min. The reaction mixture was applied to 4% polyacrylamide gel electrophoresis and ^{32}P-labeled RNAs in the gel were detected using a STORM 820 Phosphorimager (GE Healthcare).

Acknowledgements

We are grateful to Prof. Toru Hisabori, Dr. Tsuyoshi Endo, and Prof. Amane Makino for providing the antibodies, and Mr. Yoshitaka Hayakawa for construction of plasmids to disrupt *PpPPR_66* gene. We also thank ABRC for providing seeds of Arabidopsis T-DNA-tagged lines. This work was supported by JSPS KAKENHI Grant Numbers 15K14917, 17K08195 (to MS), 16H06555 (to TS). The authors declare no conflict of interest.

Short supporting information legends

Figure S1. Multiple sequence alignment of PpPPR_66 and its homologs.

Figure S2. Chloroplast localization of PpPPR_66 protein.

Figure S3. Generation of *PpPPR_66* KO mutants.

Figure S4. RT-PCR analysis of plastid intron-containing mRNAs.

Figure S5. RT-PCR analysis of plastid intron-containing tRNAs.

Figure S6. RT-PCR analysis of complemented *PpPPR_66* KO mutants.

Figure S7. Isolation of Arabidopsis *PPR66L* KO mutants.

Figure S8. Light intensity dependence of chlorophyll fluorescence parameters in Arabidopsis wild type (WT) and *AtPPR66L* KO mutants.

Table S1. Chlorophyll fluorescence parameters in wild type and mutant mosses.

Table S2. Primers used for plasmid construction and DNA/RNA analyses.

Table S3. Primers used for analysis of RNA splicing in *P. patens* plastids.

References

Asakura, Y. and Barkan, A. (2007) A CRM domain protein functions dually in group I and group II intron splicing in land plant chloroplasts. *Plant Cell* **19**, 3864-3875.

Asano, T., Miyao, A., Hirochika, H., Kikuchi, S. and Kadowaki, K. (2013) A pentatricopeptide repeat gene of rice is required for splicing of chloroplast transcripts and RNA editing of *ndhA*. *Plant Biotech*. **30**, 57-64.

Barkan, A., Walker, M., Nolasco, M. and Johnson, D. (1994) A nuclear mutation in maize blocks the processing and translation of several chloroplast mRNAs and provides evidence for the differential translation of alternative mRNA forms. *EMBO J.* **13**, 3170-3181.

Barkan, A. and Small, I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* **65**, 415-442.

Barkan, A., Rojas, M., Fujii, S., Yap, A., Chong, Y.S., Bond, C.S. and Small, I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.* **8**, e1002910.

Beick, S., Schmitz-Linneweber, C., Williams-Carrier, R., Jensen, B. and Barkan, A. (2008) The pentatricopeptide repeat protein PPR5 stabilizes a specific tRNA precursor in maize chloroplasts. *Mol. Cell. Biol*. **28**, 5337-5347.

Chateigner-Boutin, A.L., des Francs-Small, C.C., Delannoy, E., Kahlau, S., Tanz, S.K., de Longevialle, A.F., Fujii, S. and Small, I. (2011) OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript *rpoC1*. *Plant J*. **65**, 532–542.

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M.F., Bai, F., Castleden, I., Song, Y., Song, B, Huang, J., Liu, X., Xu, X., Lim. B.L., Bond, C.S., Yiu, S.M., and Small I. (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J*. **85**, 532-547.

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

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol. Biol.* **300**, 1005-1016.

Falcon de Longevialle, A., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H. and Small, I. (2008) The pentatricopeptide repeat gene *OTP51* with two LAGLIDADG motifs is required for the *cis*-splicing of plastid *ycf3* intron 2 in *Arabidopsis thaliana. Plant J.* **56**, 157-168.

Fisk, D.G., Walker, M.B. and Barkan, A. (1999) Molecular cloning of the maize gene *crp1* reveals similarity between regulators of mitochondrial and chloroplast gene expression. *EMBO J.* **18**, 2621-2630.

Fujii, S. and Small, I. (2011) The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol*. **191**, 37-47.

Gobert, A., Gutmann, B., Taschner, A., Gößringer, M., Holzmann, J., Hartmann, R.K., Rossmanith, W. and Giegé, P. (2010) A single Arabidopsis organellar protein has RNase P activity. *Nature Struct. Mol. Biol*. **17**, 740-744.

Goto, S., Kawaguchi, Y., Sugita, C., Ichinose, M. and Suigta, M. (2016) P-class pentatricopeptide repeat protein PTSF1 is required for splicing of the plastid pre-tRNA^{Ile} in *Physcomitrella patens*. *Plant J.* **86**, 493-503.

Hashimoto, M., Endo, T., Peltier, G., Tasaka, M. and Shikanai, T. (2003) A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast *ndhB* in *Arabidopsis*. *Plant J*. **36**, 541-549.

Hattori, M., Miyake, H. and Sugita, M. (2007) A pentatricopeptide repeat protein is required for RNA processing of *clpP* pre-mRNA in moss chloroplasts. *J. Biol. Chem.* **282**, 10773-10782.

Hattori, M. and Sugita, M. (2009) A moss pentatricopeptide repeat protein binds to the 3'-end of plastid *clpP* pre-mRNA and assists with the mRNA maturation. *FEBS J.* **276**, 5860-5869.

Ichinose, M. and Sugita, M. (2017) RNA editing and its molecular mechanism in plant organelles. *Genes (Basel)* **8**, 5.

Ichinose, M., Sugita, C., Yagi, Y., Nakamura, T. and Sugita, M. (2013) Two DYW subclass PPR proteins are involved in RNA editing of *ccmFc* and *atp9* transcripts in the moss *Physcomitrella patens*: First complete set of PPR editing factors in plant mitochondria. *Plant Cell Physiol.* **54**, 1907-1916.

Ichinose, M., Tasaki, E., Sugita, C. and Sugita, M. (2012) A PPR-DYW protein is required for splicing of a group II intron of *cox1* pre-mRNA in *Physcomitrella patens*. *Plant J.* **70**, 271-278.

Jenkins, B., Kulhanek, D. and Barkan, A. (1997). Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* **9**, 283–296.

Johnson, X., Wostrikoff, K., Finazzi, G., Kuras, R., Schwarz, C., Bujaldon, S., Nickelsen, J., Stern, D.B., Wollman, F.A. and Vallon, O. (2010) MRL1, a conserved pentatricopeptide repeat protein, is required for stabilization of *rbcL* mRNA in *Chlamydomonas* and *Arabidopsis*. *Plant Cell* **22**, 234-248.

Khrouchtchova, A., Monde, R.A. and Barkan, A. (2012) A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. *RNA* **18**, 1197-1209.

Kotera, E., Tasaka, M. and Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433, 326–330.

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, ML., Mireau, H., Peeters, N., Renou, J-P., Szurek, B., Taconnat, L. and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089-2103.

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N. and Schuster, G. (2003) HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. *Plant Cell* **15**, 1480-1495.

Michel, F., Umesono, K. and Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns – a review. *Gene* **82**, 5-30.

Nishiyama, T., Hiwatashi, Y., Sakakibara, K., Kato, M., and Hasebe, M. (2000) Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. *DNA Res.* **7,** 9-17.

Ostheimer, G.J., Williams−Carrier, R., Belcher, S., Osborne, E., Gierke, J. and Barkan, A. (2003) Group II intron splicing factors derived by diversification of an ancient RNA−binding domain. *EMBO J*. **22**, 3919-3929.

Pfalz, J., Bayraktar, O.A., Prikryl, J. and Barkan, A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5' and 3' mRNA termini in chloroplasts. *EMBO J.* **28**, 2042-2052.

Pyo, Y.J., Kwon, K.C., Kim, A. and Cho, M.H. (2013) Seedling lethal1, a pentatricopeptide repeat protein lacking an E/E+ or DYW domain in Arabidopsis, is involved in plastid expression and early chloroplast development. *Plant Physiol*. **163**, 1844-1858.

Schmitz-Linneweber, C., Williams-Carrier, R.E, Williams-Voelker, P.M., Kroeger, T.S., Vichas, A. and Barkan, A. (2006) A pentatricopeptide repeat protein facilitates the *trans*-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell* **18**, 2650-2663.

Schmitz-Linneweber, C. and Small, I. (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci.* **13**, 663-670.

Schmitz-Linneweber, C., Lampe, M-K., Sultan, L.D. and Ostersetzer-Biran, O. (2015) Organellar maturases: A window into the evolution of the spliceosome. *Biochim. Biophys. Acta* **1847**, 798-808.

Shikanai, T. (2016) Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase. *Biochim. Biophys. Acta* **1857**, 1015-1022.

Small, I.D. and Peeters, N. (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**, 46-47.

Sugita, C., Komura, Y., Tanaka, K., Kometani, K., Satoh, H. and Sugita, M. (2014) Molecular characterization of three PRORP proteins in the moss *Physcomitrella patens*: nuclear PRORP protein is not essential for moss viability. *PLoS ONE* **9**, e108962.

Sugita, M., Ichinose, M., Ide, M. and Sugita, C. (2013) Architecture of the PPR gene family in the moss *Physcomitrella patens*. *RNA Biol.* **10**, 1439-1445.

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C. and Sugita, M. (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucleic Acids Res.* **31**, 5324-5331.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B., Brennicke, A. (2013) RNA editing in plants and its evolution. *Annu. Rev. Genet*. **47**, 335-352.

Tasaki, E., Hattori, M. and Sugita, M. (2010) The moss pentatricopeptide repea protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J .* **62**, 560-570.

534 **Ueda, M., Kuniyoshi, T., Yamamoto, H., Sugimoto, K., Ishizaki, K., Kohchi, T., Nishimura, Y. and**
535 **Shikanai, T.** (2012) Composition and physiological function of the chloroplast NADH
536 dehydrogenase-like complex in *Marchantia polymorpha*. *Plant J*. **72**, 683-693.
537 **Williams-Carrier, R., Kroeger, T. and Barkan, A.** (2008) Sequence-specific binding of a chloroplast
538 pentatricopeptide repeat protein to its native group II intron ligand. *RNA* **14**, 1930-1941.
539 **Wu, W., Liu, S., Ruwe, H., Zhang, D., Melonek, J., Zhu, Y., Hu, X., Gusewski, S., Yin, P., Small, I.,**
540 **Howell, K.A. and Huang, J.** (2016) SOT1, a pentatricopeptide repeat protein with a small
541 MutS-related domain, is required for correct processing of plastid 23S-4.5S rRNA precursors in
542 *Arabidopsis thaliana*. *Plant J.* **85**, 607-621.
543 Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA
544 recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants.
545 *PLoS ONE* **8**, e57286.
546 Yamamoto, H., Peng, L., Fukao, Y. and Shikanai, T. (2011) An Src homology 3 domain-like fold protein
547 forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in
548 Arabidopsis. *Plant Cell* **23**, 1480–1493.
549 **Zhang, H.D., Cui, Y.L., Huang, C., Yin, Q.Q., Qin, X.M., Xu, T., He, X.F., Zhang, Y., Li, Z.R. and**
550 **Yang, Z.N.** (2015) PPR protein PDM1/SEL1 is involved in RNA editing and splicing of plastid
551 genes in *Arabidopsis thaliana*. *Photosynth. Res*. **126**, 311-321.
552 **Zoschke, R., Watkins, K.P., Miranda, R.G. and Barkan, A.** (2016) The PPR-SMR protein PPR53
553 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J*. **85**, 594-606.
554

555 **Figure legends**

556

557 **Figure 1.** Phenotype of *PpPPR_66* gene KO mutants.

558 (a) Predicted PpPPR_66 consists of a putative transit peptide (TP) and 11 PPR motifs. (b) Moss
559 colonies of wild type (WT) and KO mutants (Δ*66*-2, Δ*66*-3). The mosses were grown for 3 weeks on
560 BCD or BCDAT medium without hygromycin B. Scale bars = 1 cm. (c) RT-PCR for detection of
561 *PpPPR_66* transcript in WT and KO mosses. *PpActin1* transcript was also amplified as a control.

562

563 **Figure 2.** Analysis of chloroplast NDH activity in vivo.

564 Wild-type (WT) and KO mutants (Δ*66*-2, Δ*66*-3) of moss colonies were exposed to actinic light (AL) (50
565 μmol photons m^{-2} s^{-1}) for 5 min. AL was turned off and the subsequent transient rise in chlorophyll
566 fluorescence (boxed region) ascribed to chloroplast NDH activity was monitored using PAM chlorophyll
567 fluorometry. F_m, maximum chlorophyll fluorescence; F_o, minimal chlorophyll fluorescence; ML,
568 measuring light; SP, saturating light pulse of white light. Prior to AL illumination, SP was applied to
569 monitor the F_m level. The bottom curve indicates a typical trace of chlorophyll fluorescence change in
570 the WT. Insets are magnified traces from the boxed area.

571

572 **Figure 3.** Western blot analysis of chloroplast proteins from *P. patens*.

573 Total proteins (the indicated dilution of the WT sample) were subjected to immunoblot analysis with
574 antibodies for PnsB1 and NdhM subunits of the NDH complex, cytochrome *f* (Cyt*f*), H^+-ATP synthase
575 β-subunit (AtpB), and PSII D1 protein (PsbA). The blotted membrane was stained with Ponceau S and
576 the light-harvesting chlorophyll binding protein (LHCII) is indicted (bottom).

577

578 **Figure 4.** RT-PCR analysis of *ndh* transcripts.

579 Eleven *ndh* genes are located in different positions in the *P. patens* plastid genome. The amplified
580 cDNA regions are shown as horizontal bars with double arrowheads including length (bp). *ndhA* and
581 *ndhB* contain a group II intron. Spliced and unspliced forms were amplified by RT-PCR. Primer
582 sequences are listed in Table S3. Lanes 1, 2, and 3 indicate PCR products amplified for 26, 30, and 34
583 cycles, respectively.

584

585 **Figure 5.** Northern blot analysis of the *ndhA*-containing gene cluster.

586 Total RNA (10 μg) from wild type (WT) and KO mutant mosses (Δ*66*-2 and Δ*66*-3) was analyzed by

RNA gel blot hybridization using DNA probes 1 to 5, Int, and Ex. The gels stained with ethidium bromide are also shown. RNA size markers (0.2 to 8.0 kb) are indicated on the right. Stars and dots indicate a 7-kb and 6.3-kb transcript, respectively. Arrowheads indicate an excised intron.

Figure 6. Physiological and molecular characterization of the Arabidopsis *PPR66L* KO mutants.
(a) Analysis of chloroplast NDH activity in vivo. Chlorophyll fluorescence from an Arabidopsis leaf was monitored as described in Figure 2. The bottom curve indicates a typical trace of chlorophyll fluorescence change in the Arabidopsis wild type (WT). Insets are magnified traces from the boxed area. The fluorescence levels were normalized by the F_m levels. F_m, F_o, ML, SP, and AL are as described in Fig. 2.
(b) Spliced and unspliced *ndhA* transcripts were amplified by RT-PCR. Primer sequences are listed in Table S2. *AtPPR66L* (*At2g35130*) and *rbcL* transcripts were also amplified by RT-PCR as controls.
(c) Total RNA (10 μg) from the Arabidopsis WT and KO mutants was analyzed by northern blot hybridization using antisense exon or intron RNA probes. RNA size markers (1.0 to 8.0 kb) are indicated on the right. Open and closed dots indicate the spliced and unspliced *ndhA* transcripts, respectively. The arrowhead indicates an excised intron.

Figure 7. In vitro binding of recombinant PpPPR_66 to the *ndhA* intron.
(a) Schematic secondary structure of the *P. patens ndhA* group II intron. Terminal thick lines indicate exons 1 and 2 of *ndhA*. DI to DVI in the 690-bp intron indicate domains I to VI (Michel *et al*. 1989). Exon-binding sequence (EBS) 1 and intron-binding sequence (IBS) 1, EBS2 and IBS2, α and α', γ and γ' refer to three-dimensional pairings (Michel *et al*. 1989). A bulging adenosine (A) in DVI is enclosed by a circle. Numbering of 100 to 600 indicates nucleotide positions from the 5' end of the 690-bp intron. Regions of probes RNA1 to 3 and RNA1a to 1c used for REMSA are indicated by arrows and dashed lines, respectively. (b) REMSA using recombinant proteins (rTrx, rPPR66) and ^{32}P-labeled RNA probes (100 pM each). The amount of recombinant proteins is shown above each lane.

Figure 1

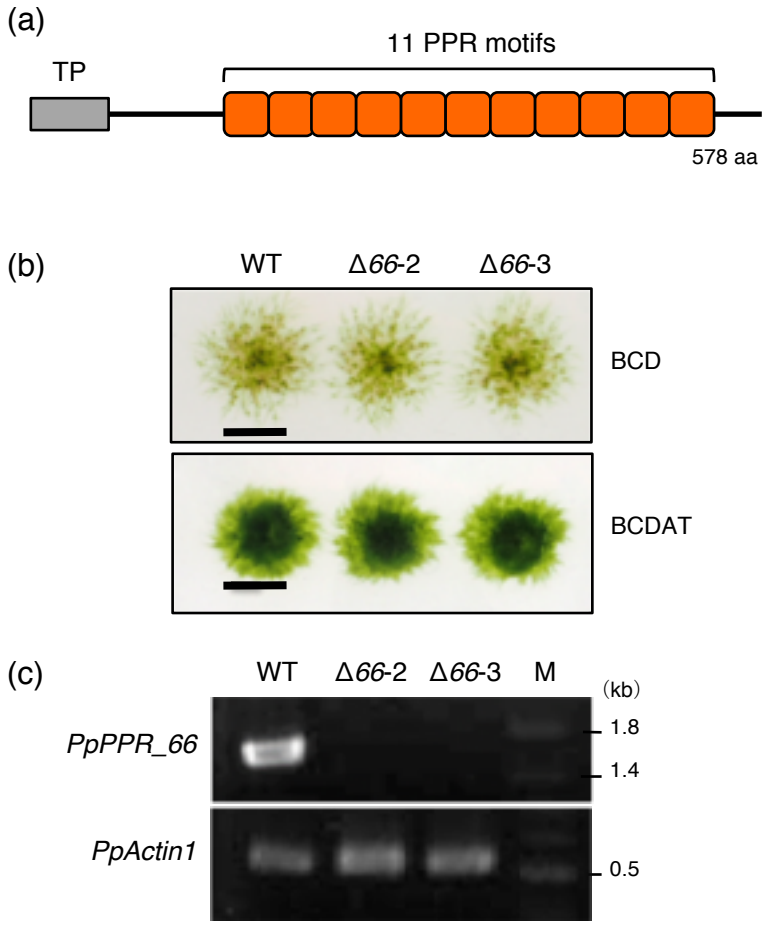

(a)
11 PPR motifs
TP
578 aa
(b)
WT
Δ66-2
Δ66-3
BCD
BCDAT
(c)
WT
Δ66-2
Δ66-3
M
(kb)
PpPPR_66
PpActin1
1.8
1.4
0.5

Figure 2

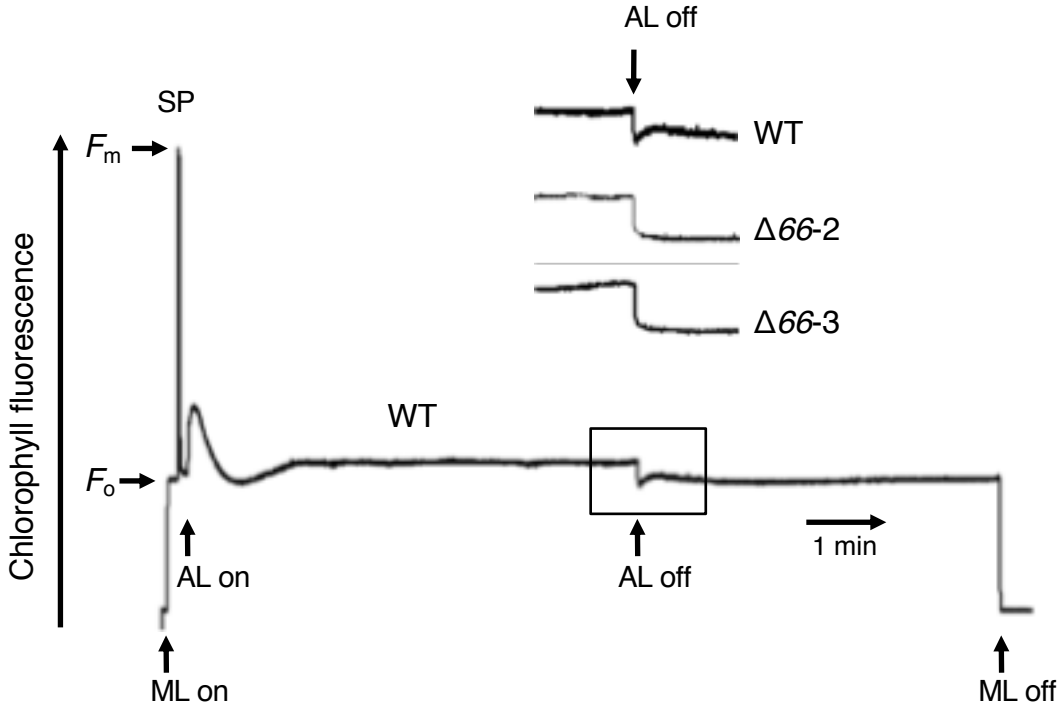

AL off
SP
WT
Δ66-2
Δ66-3
Fm
WT
Chlorophyll fluorescence
Fo
AL on
AL off
1 min
ML on
ML off

Figure 3

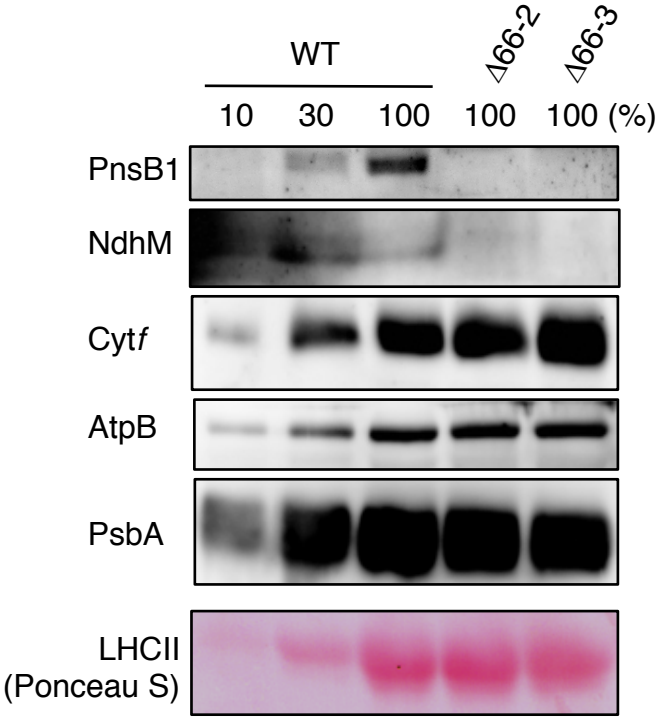

WT
Δ66-2
Δ66-3
10
30
100
100
100 (%)
PnsB1
NdhM
Cytf
AtpB
PsbA
LHCII
(Ponceau S)

Figure 4

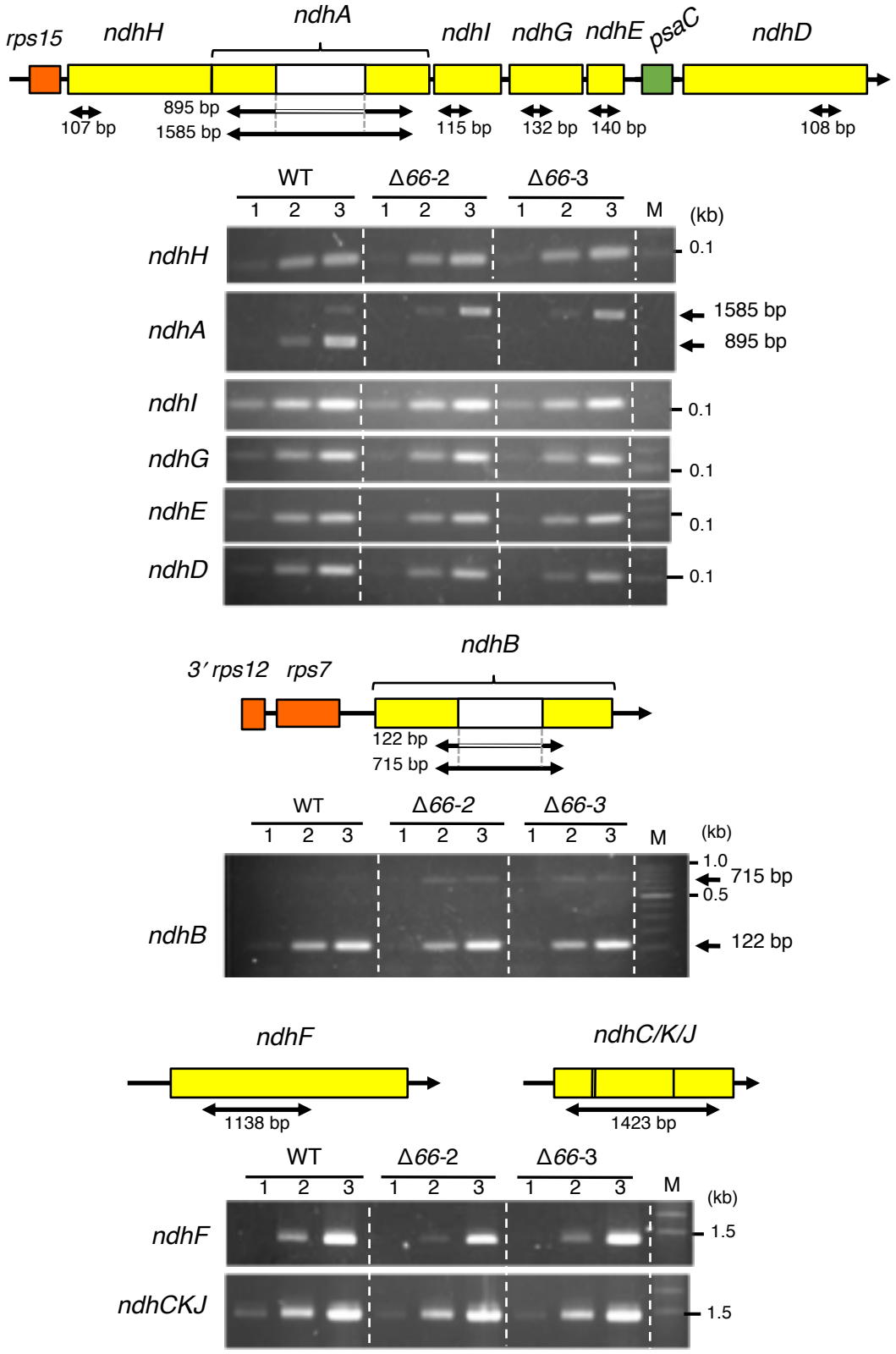

rps15
ndhH
ndhA
ndhI
ndhG
ndhE
psaC
ndhD
107 bp
895 bp
1585 bp
115 bp
132 bp
140 bp
108 bp
WT
Δ66-2
Δ66-3
1 2 3 1 2 3 1 2 3 M
(kb)
ndhH
0.1
ndhA
1585 bp
895 bp
ndhI
0.1
ndhG
0.1
ndhE
0.1
ndhD
0.1
3′ rps12
rps7
ndhB
122 bp
715 bp
WT
Δ66-2
Δ66-3
1 2 3 1 2 3 1 2 3 M
(kb)
ndhB
1.0
715 bp
0.5
122 bp
ndhF
ndhC/K/J
1138 bp
1423 bp
WT
Δ66-2
Δ66-3
1 2 3 1 2 3 1 2 3 M
(kb)
ndhF
1.5
ndhCKJ
1.5

Figure 5

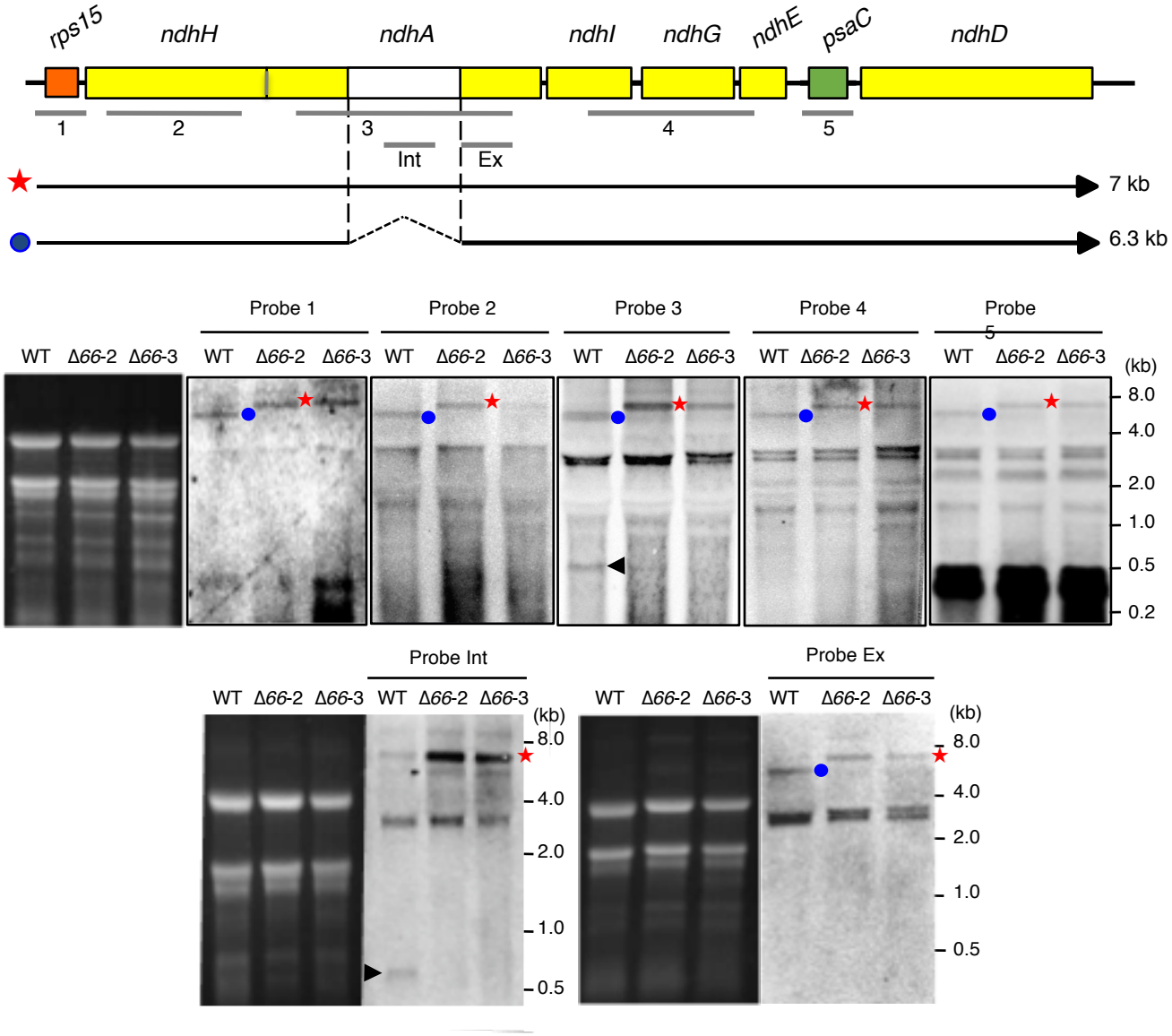

rps15
ndhH
ndhA
ndhI
ndhG
ndhE
psaC
ndhD
1
2
3
4
5
Int
Ex
7 kb
6.3 kb
Probe 1
Probe 2
Probe 3
Probe 4
Probe 5
WT
Δ66-2
Δ66-3
(kb)
8.0
4.0
2.0
1.0
0.5
0.2
Probe Int
Probe Ex

Figure 6

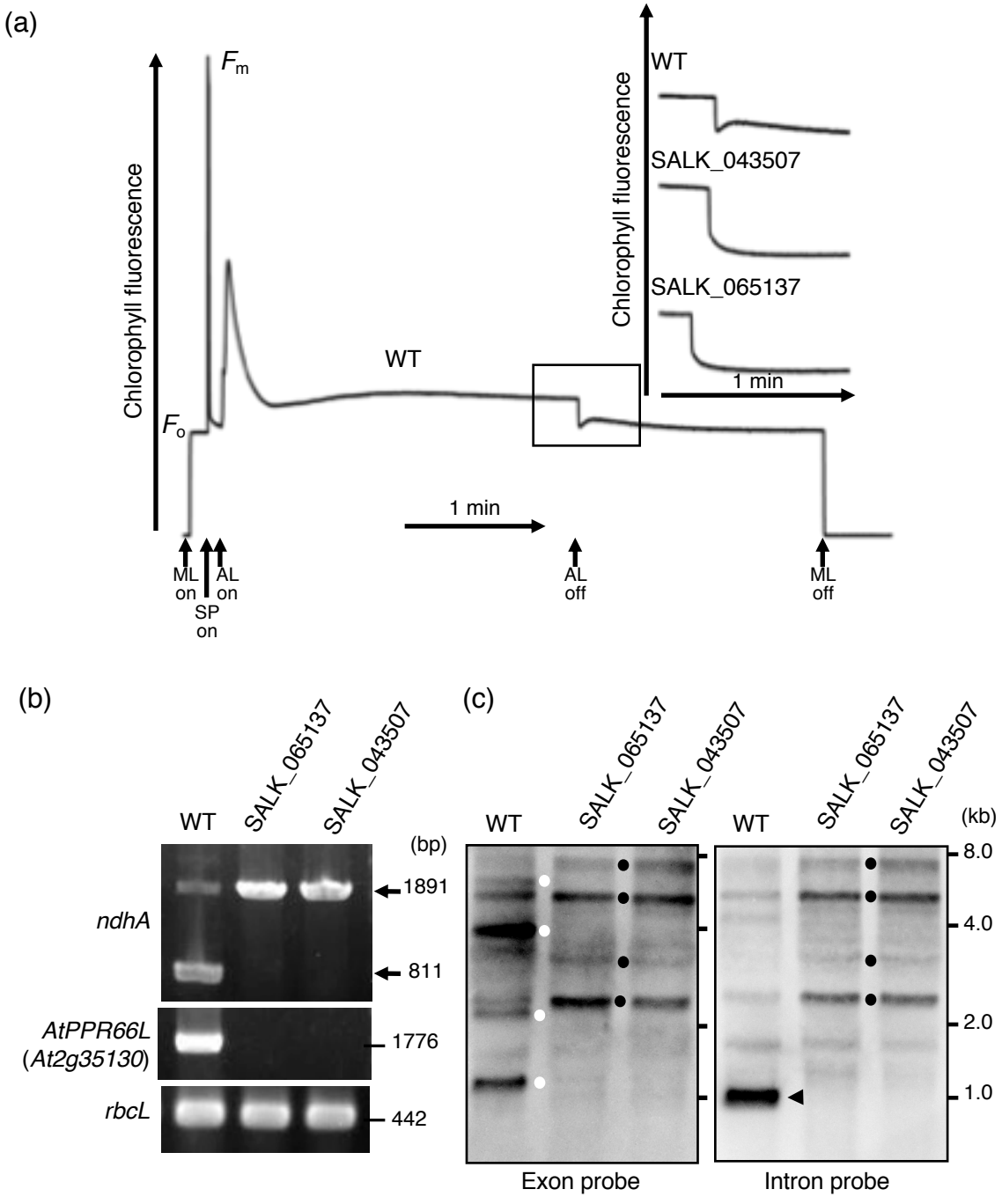
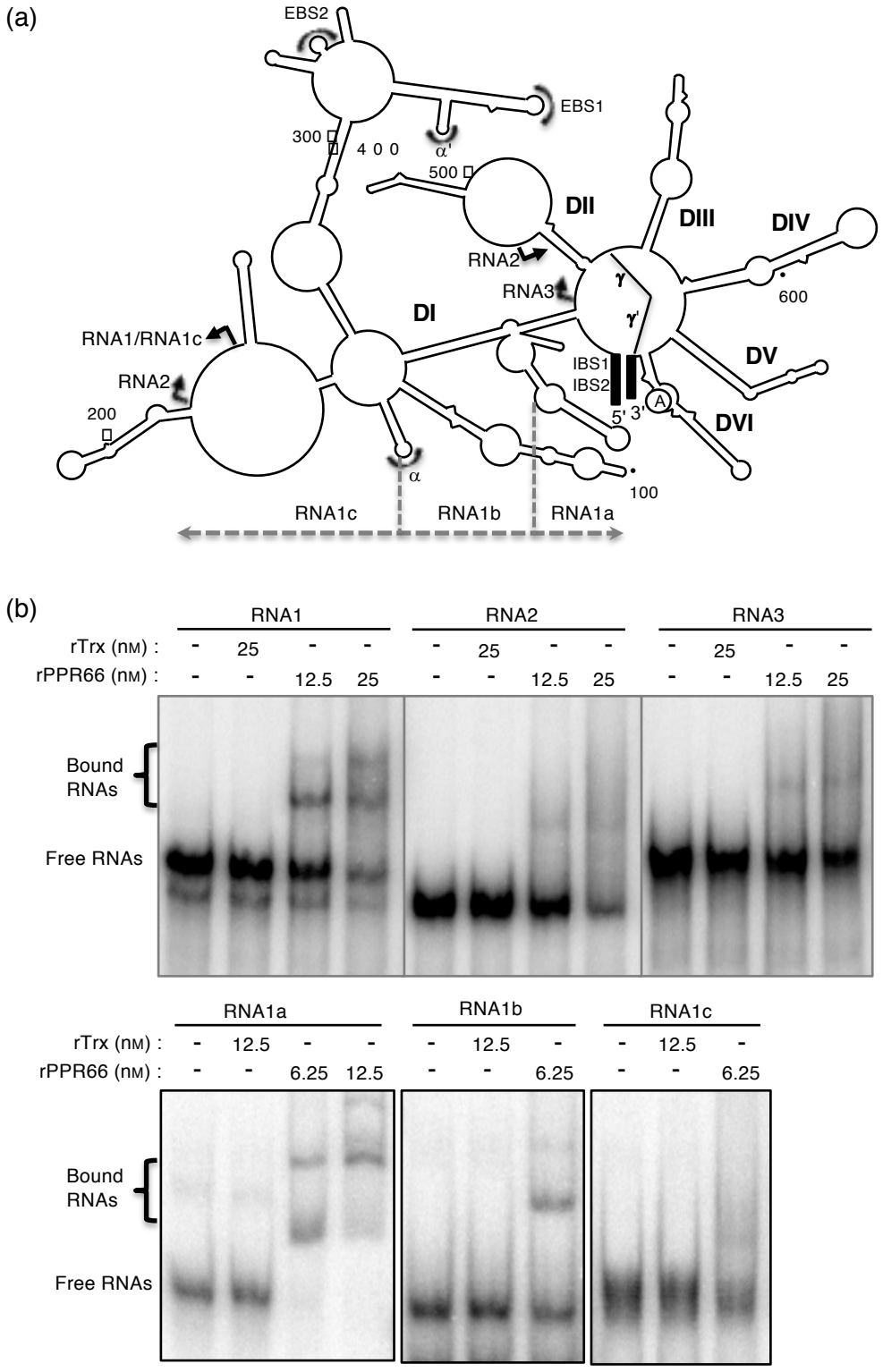

(a)
Chlorophyll fluorescence
F_m
F_o
WT
1 min
ML on
AL on
SP on
AL off
ML off
WT
SALK_043507
SALK_065137
1 min
(b)
WT
SALK_065137
SALK_043507
(bp)
ndhA
1891
811
AtPPR66L
(At2g35130)
1776
rbcL
442
(c)
WT
SALK_065137
SALK_043507
(kb)
8.0
4.0
2.0
1.0
Exon probe
Intron probe

Figure 7


(a)
EBS2
EBS1
300
400
α'
500
DII
DIII
DIV
RNA2
RNA3
γ
γ'
600
DI
RNA1/RNA1c
RNA2
IBS1
IBS2
DV
200
5' 3'
A
DVI
α
100
RNA1c
RNA1b
RNA1a
(b)
RNA1
RNA2
RNA3
rTrx (nM) : - 25 - - - 25 - - - 25 - -
rPPR66 (nM) : - - 12.5 25 - - 12.5 25 - - 12.5 25
Bound RNAs
Free RNAs
RNA1a
RNA1b
RNA1c
rTrx (nM) : - 12.5 - - - 12.5 - - 12.5 -
rPPR66 (nM) : - - 6.25 12.5 - - 6.25 - - 6.25
Bound RNAs
Free RNAs

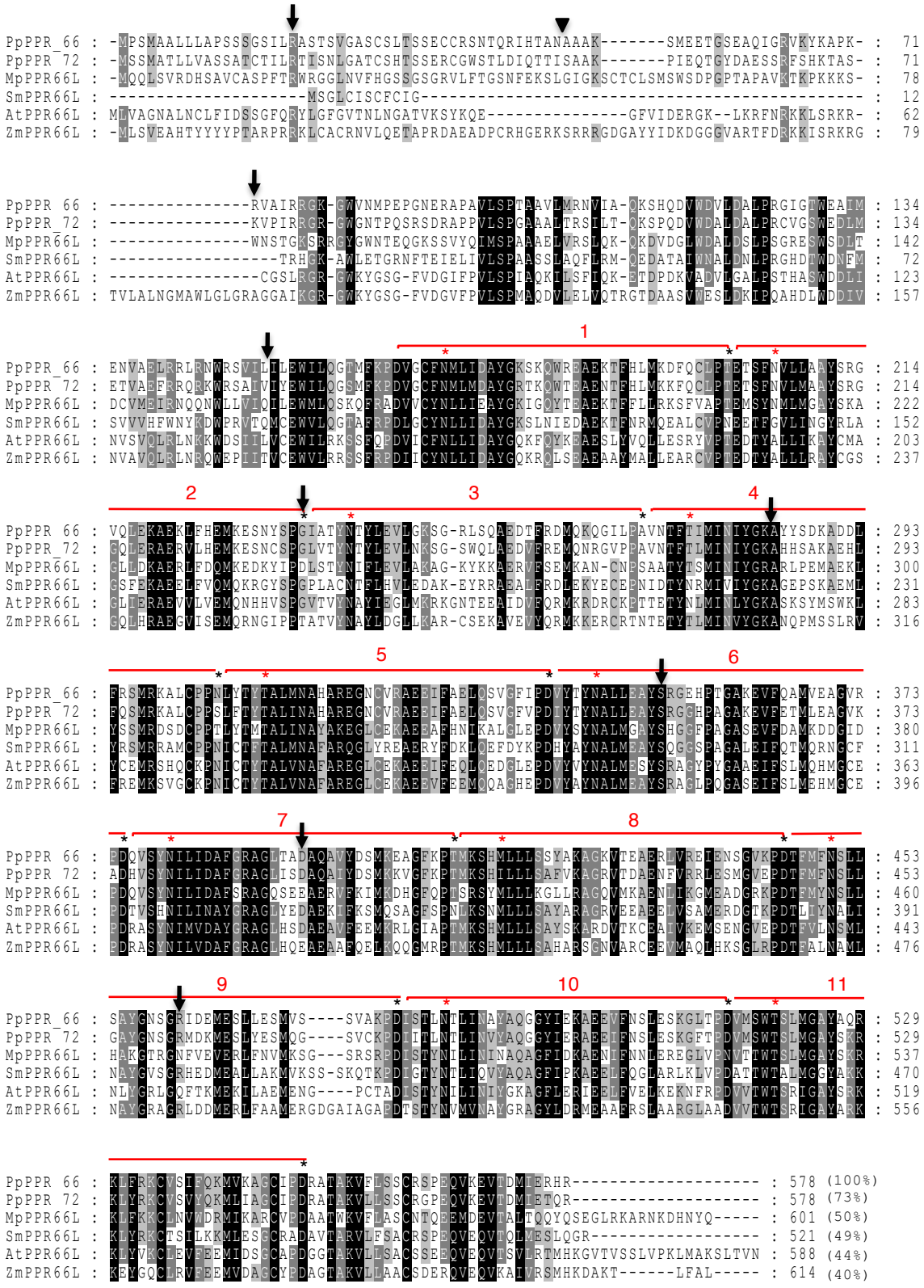


Figure S1. Multiple sequence alignment of PpPPR_66 and its homologs. Amino acid (aa) sequences were aligned with ClustalW (http://clustalw.ddbj.nig.ac.jp/index.php?lang=en). Identical and conserved aa residues are shaded in black and grey, respectively. PpPPR_66 (Pp3c16_5890); PpPPR_72 (Pp3c6_26210); MpPPR66L, *Marchantia polymorpha* PPR_66-like (Mapoly0002s0014.1); SmPPR66L, *Selaginella moellendorffii* (109632); AtPPR66L, *Arabidopsis thaliana* (At2g35130); ZmPPR66L, *Zea mays* (GRMZM2G007372_T01). PPR motifs are marked in red brackets 1 to 11. The position 5 and 35 aa residues in each PPR motif are indicated by red and black asterisks, respectively. An arrowhead indicates the predicted cleavage site of the transit peptide. The intron positions of *PpPPR_66* are indicated by downwards arrows.

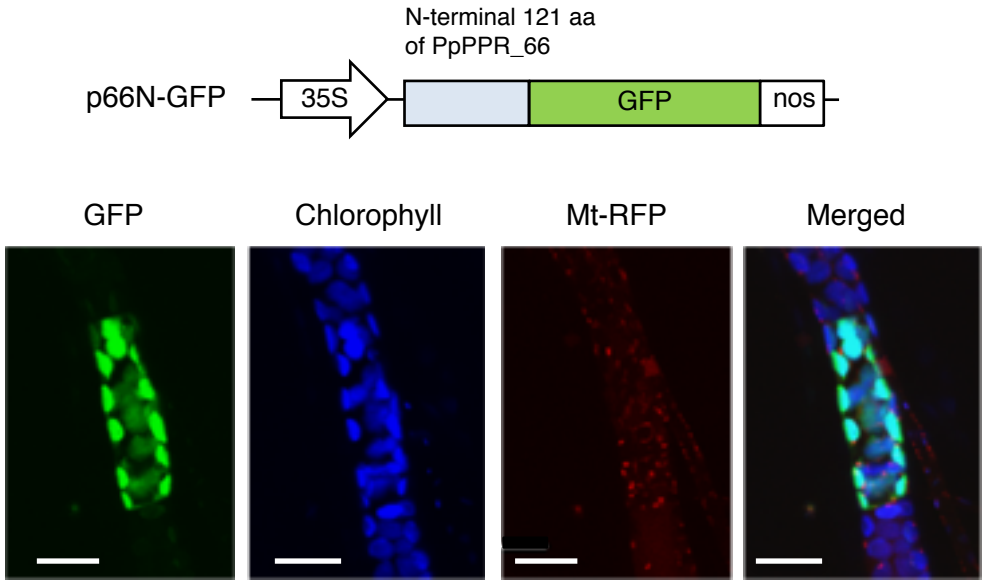


Figure S2. Chloroplast localization of PpPPR_66 protein.
Chimeric protein was transiently expressed in the Mt-RFP OX moss. Fluorescence of 66N-GFP (GFP), RFP (Mt-RFP) and chlorophyll autofluorescence were detected by confocal fluorescent microscopy. An overlay of fluorescence images (Merged) is shown. Scale bars = 20 μm.

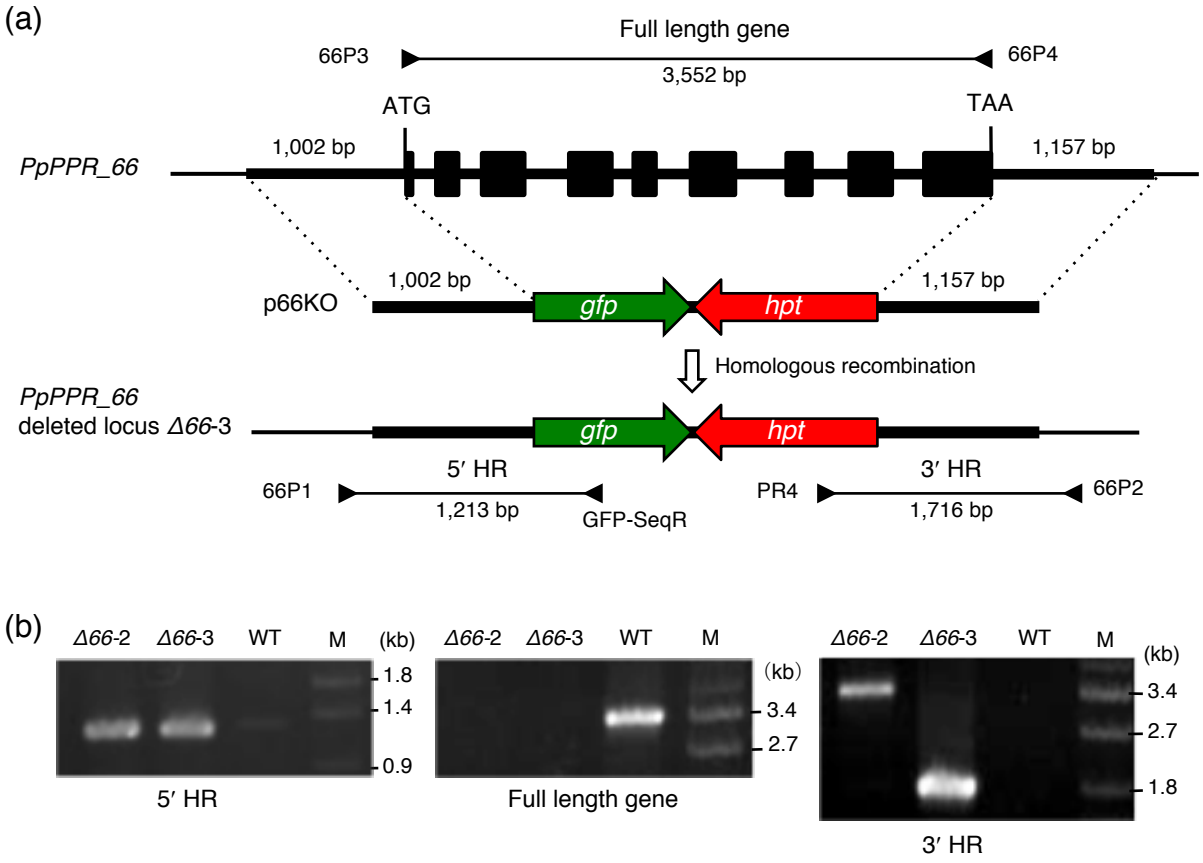


Figure S3. Generation of *PpPPR_66* KO mutants.
(a) Structures of wild-type (WT) and the altered genomic locus after replacement of the *gfp-hpt* gene cassette by homologous recombination (HR) are illustrated. Primers and the expected fragment sizes for PCR analysis are also shown. Primer sequences are listed in Table S2. The DNA regions for HR are represented in thick horizontal lines.
(b) Genomic PCR analysis of WT and KO mutants. The predicted 1,213- (5' HR) and 1,716-bp (3' HR) fragments were amplified from the KO lines and the 3,552-bp fragment (full length gene) was amplified from the WT. DNA size marker is the λ DNA *Sty*I-digest (lanes M).

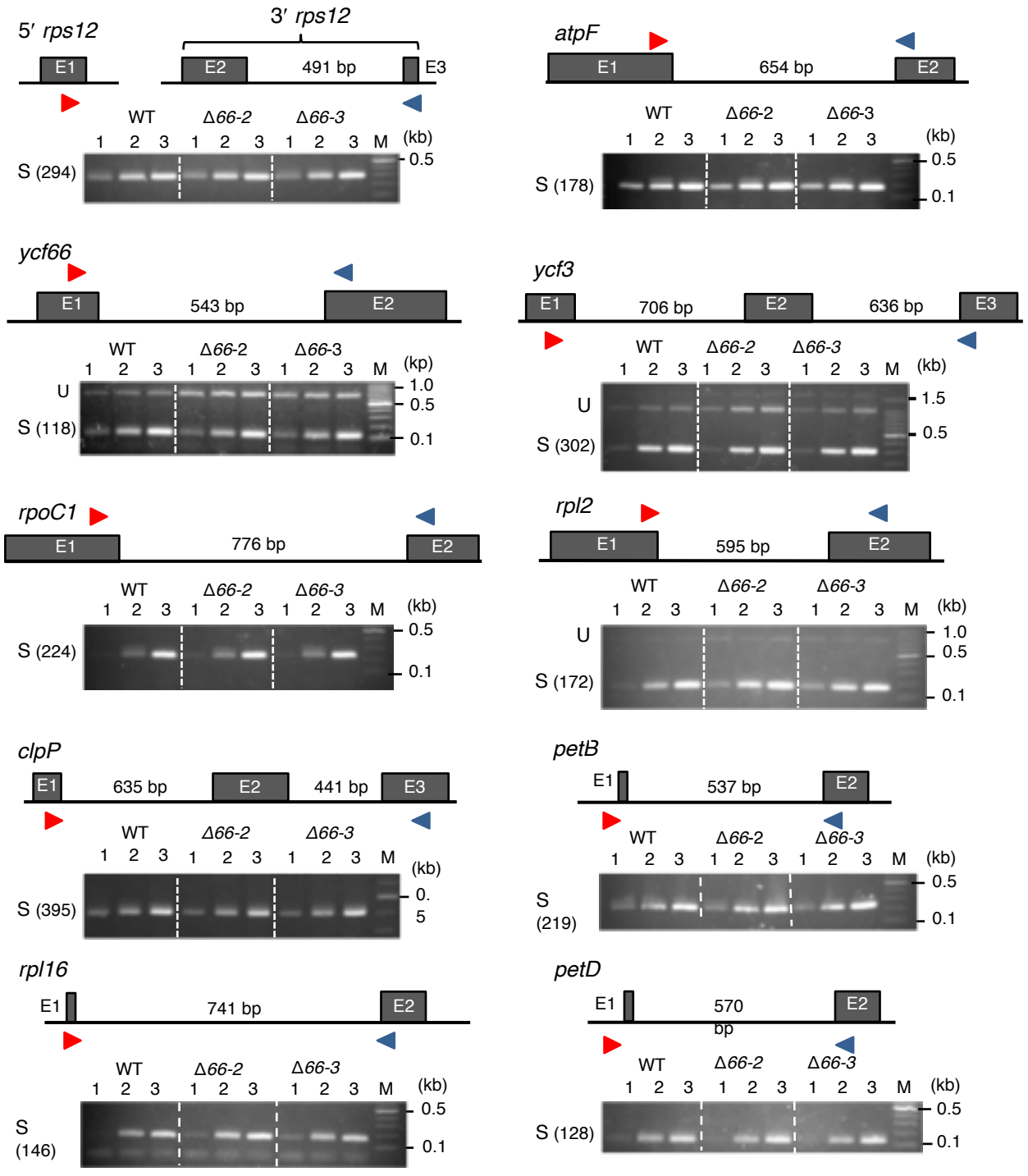


Figure S4. RT-PCR analysis of plastid intron-containing mRNAs.
Schematic gene structures are shown as exons (E1, E2, and E3) and intron(s) including length (bp). Forward and reverse primers used for RT-PCR are shown in red and blue arrowheads, respectively. Primer sequences are listed in Table S3. The amplified fragments derived from spliced and unspliced transcripts are indicated as S and U, respectively. Spliced transcript sizes (bp) are indicated in parentheses. Lanes 1, 2, and 3 indicate the number of PCR reaction cycles, 26, 30, and 34, respectively. Lane M indicates DNA size marker.

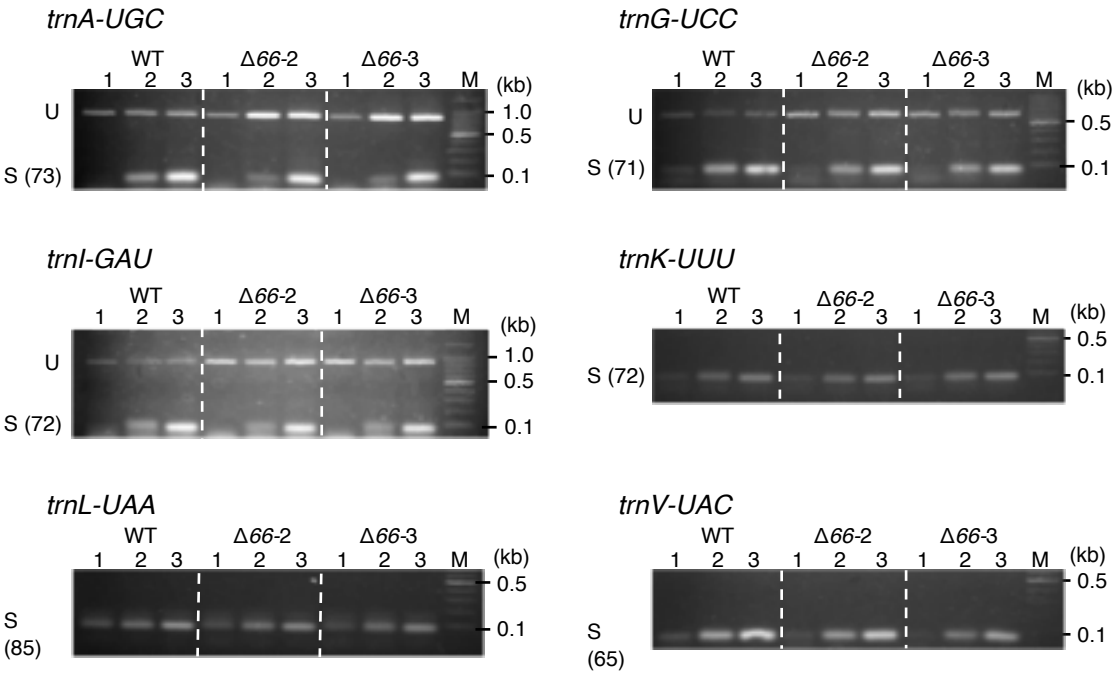


Figure S5. RT-PCR analysis of plastid intron-containing tRNAs.
Forward and reverse primers used for RT-PCR are listed in Table S3. The amplified fragments are indicated as spliced (S) and/or unspliced (U) tRNAs. The size (bp) of spliced tRNAs is indicated in parentheses. Lanes 1, 2, and 3 indicate the number of PCR reaction cycles, 26, 30, and 34, respectively. Lane M indicates DNA size marker.

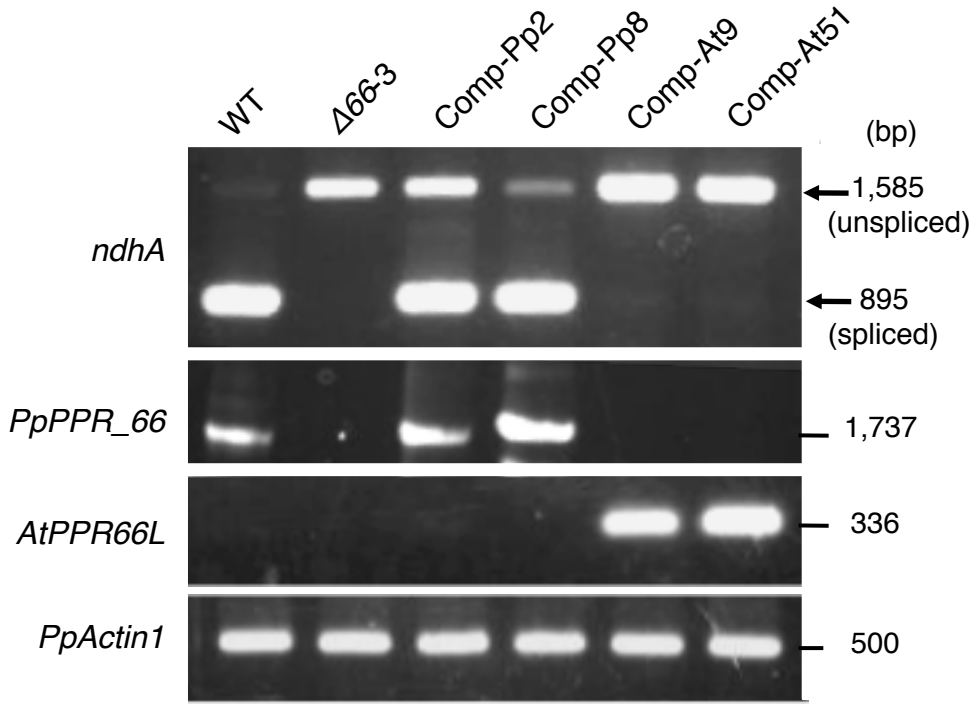


Figure S6. RT-PCR analysis of complemented *PpPPR_66* KO mutants. RT-PCR on cDNA from wild type (WT), KO mutant (Δ*66*-3), and Δ*66*-3 complemented with the full-length cDNA of *PpPPR_66* (Comp-Pp2, Comp-Pp8) or *AtPPR66L* (Comp-At9, Comp-At51). Primer sets used for RT-PCR are listed in Table S2. Arrows indicate unspliced (1,585 bp) and spliced (895 bp) *ndhA* products.

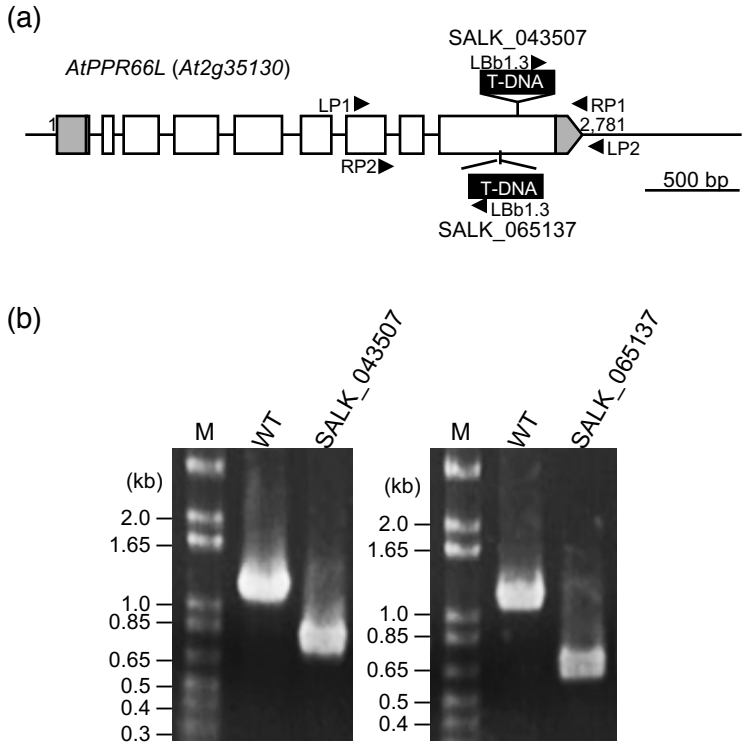
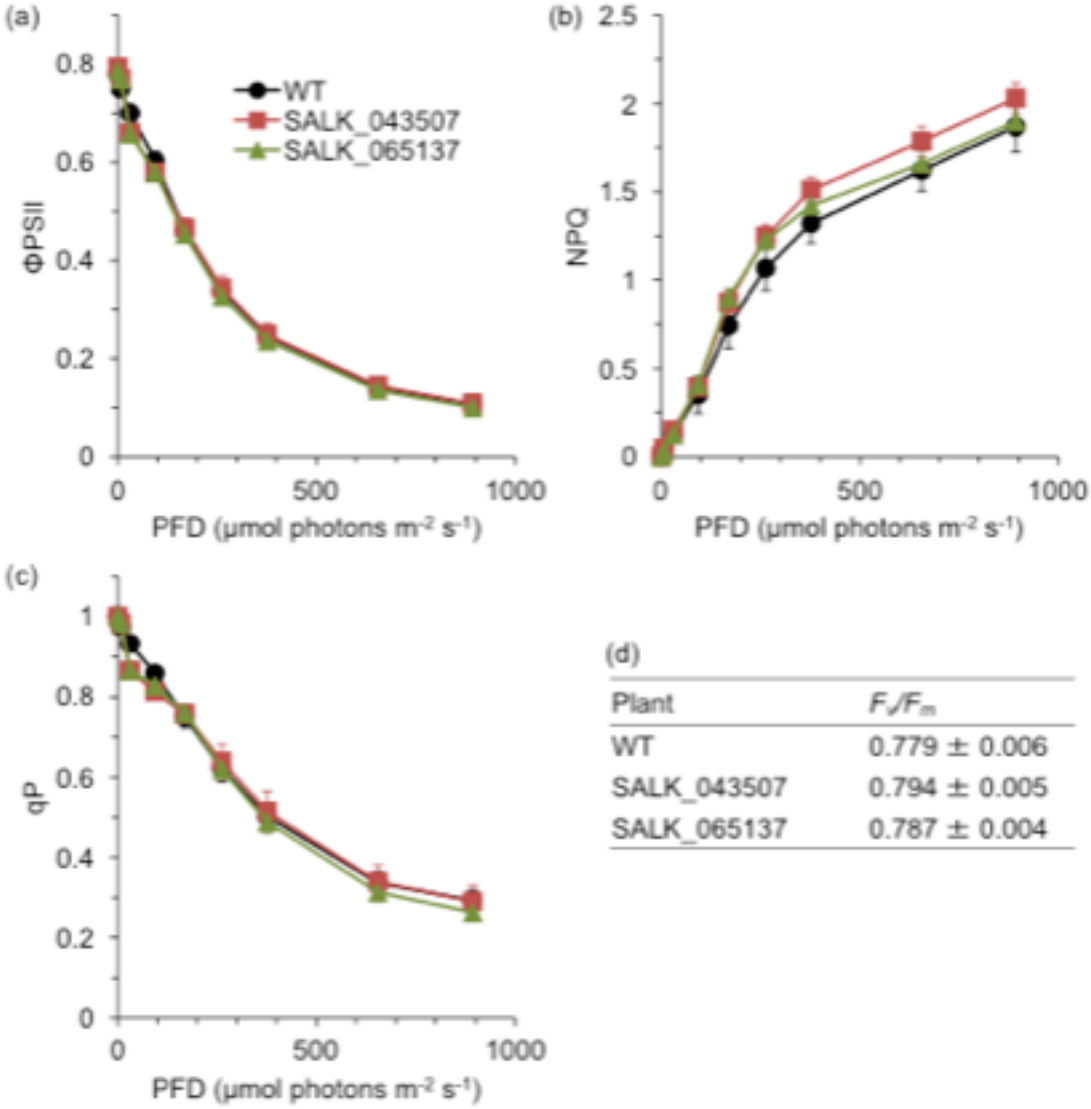


Figure S7. Isolation of Arabidopsis *PPR66L* KO mutants.
(a) Schematic gene structure of Arabidopsis *PPR66L* (*At2g35130*). Positions of T-DNA insertions are indicated. Open and gray boxes indicate translated and untranslated regions, respectively. Positions of primers used for PCR in (b) are indicated by arrowheads.
(b) Genotyping of Arabidopsis *PPR66L* gene KO mutants. PCR was performed on genomic DNA to detect homozygosity of the T-DNA-tagged lines SALK_043507 and SALK_065137. T-DNA-specific primer LBb1.3 and gene-specific primer sets (LP1 and RP1 for SALK_043507 and LP2 and RP2 for SALK_065137) were used for PCR. Primer sequences are listed in Table S2. The amplified fragments were separated by agarose gel electrophoresis. Lane M indicates DNA size marker.



(d)

Plant	F_v/F_m
WT	0.779 ± 0.006
SALK_043507	0.794 ± 0.005
SALK_065137	0.787 ± 0.004

Figure S8. Light intensity dependence of Chl fluorescence parameters in Arabidopsis wild type (WT) and *AtPPR66L* KO mutants.
Measurement of ΦPSII (a), NPQ (b), qP (c) and F_v/F_m (d) in WT, SALK-043507 and SALK_065137 leaves. Values are means ± SD (n = 6 or 7). The horizontal axis indicates photon flux density (PFD).