The P-class pentatricopeptide repeat protein PpPPR_21 is needed for accumulation of the *psbI-ycf12* dicistronic mRNA in *Physcomitrella* chloroplasts

Tetsuo Ebihara[1], Takuya Matsuda[1], Chieko Sugita[1], Mizuho Ichinose[1,2], 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, and*

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

*For correspondence

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

Running title: *A PPR stabilization factor for the* psbI-ycf12 *mRNA*

Running authors: *Tetsuo Ebihara* et al.

Numbers of Figures: 8

Supporting information: Figures S1 to S6, Tables S1

SUMMARY

Chloroplast gene expression is controlled by numerous nuclear-encoded RNA-binding proteins. Among them, pentatricopeptide repeat (PPR) proteins are known to be key players of posttranscriptional regulation in chloroplasts. However, the functions of many PPR proteins remain unknown. In this study, we characterized the function of a chloroplast-localized P-class PPR protein PpPPR_21 in *Physcomitrella patens*. Knockout (KO) mutants of *PpPPR_21* exhibited reduced protonemata growth and lower photosynthetic activity. Immunoblot analysis and blue-native gel analysis showed a remarkable reduction of the photosystem II (PSII) reaction center protein and poor formation of the PSII super-complexes in the KO mutants. To assess whether PpPPR_21 is involved in chloroplast gene expression, chloroplast genome-wide microarray analysis and Northern blot hybridization were performed. These analyses indicated that the *psbI-ycf12* transcript encoding the low molecular weight subunits of PSII did not accumulate in the KO mutants while other *psb* transcripts accumulated at similar levels in wild type and KO mutants. A complemented *PpPPR_21* KO moss transformed with the cognate full-length *PpPPR_21* cDNA rescued the level of accumulation of *psbI-ycf12* transcript. RNA binding experiments showed that the recombinant PpPPR_21 bound efficiently to the 5′-untranslated and translated regions of *psbI* mRNA. The present study suggests that PpPPR_21 may be essential for the accumulation of a stable *psbI-ycf12* mRNA.

Keywords: P-class PPR protein, mRNA stability, *psbI*, *ycf12*, chloroplast, *Physcomitrella patens*, *Arabidopsis thaliana*

Significance Statement: Pentatricopeptide repeat (PPR) proteins play important roles in various RNA processing steps in chloroplasts. Here we identified a P-class PPR protein which is essential for the accumulation of stable *psbI-ycf12* mRNA in chloroplasts of the moss *Physcomitrella patens*. This PPR protein binds specifically to the 5′-untranslated and translated regions of *psbI* RNA.

INTRODUCTION

Chloroplast gene expression depends on a number of post-transcriptional steps including site-specific RNA cleavage, RNA splicing and RNA editing, which are essential. Stabilization of RNA is also important for efficient translation. The nucleus-encoded pentatricopeptide repeat (PPR) proteins that are imported into chloroplasts or mitochondria are involved in post-transcriptional and translational regulation at various steps of RNA processing (Small and Peeters, 2000, Lurin *et al*. 2004, Schmitz-Linneweber and Small 2008, Gutmann *et al*. 2012, Barkan and Small 2014, Manavski *et al*. 2018). In land plants, PPR proteins constitute an extremely large family that is encoded by 450 PPR genes in *Arabidopsis* and over 1,000 in lycophytes (Cheng *et al*. 2016). Among the 450 Arabidopsis PPR proteins, one-third are localized in chloroplasts (Colcombet *et al*. 2013). The PPR proteins harbor tandem arrays of a degenerate 31–36-amino acid motif and are roughly grouped into two classes. The P class consists of canonical 35-amino acid PPR (P) motifs while the PLS class is organized by repeated units of P and PPR-like (L, S) motifs. The PLS class mostly functions as an RNA editing site-recognition factor in chloroplasts and mitochondria (Fujii and Small 2011, Takenaka *et al*. 2013b, Ichinose and Sugita 2017).

The loss-of-function PPR proteins frequently result in defects in chloroplast biogenesis (Barkan and Small 2014, Belcher *et al*. 2015). The P-class PPR proteins are involved in translational control and intergenic RNA cleavage (Barkan *et al*. 1994, Fisk *et al*. 1999, Meierhoff *et al*. 2003) and splicing (Schmitz-Linneweber *et al*. 2006, Falcon de Longevialle *et al*. 2008, Khrouchtchova *et al*. 2012, Goto *et al*. 2016, Ito *et al*. 2018), as well as the stabilization of chloroplast mRNAs (Pfalz *et al*. 2009, Johnson *et al*. 2010, Zhelyazkova *et al*. 2012, Zoschke *et al*. 2016, Rajas *et al*. 2018). In all these cases, PPR proteins bind in a gene-specific manner to target RNAs. Bioinformatic analyses of PPR proteins and target RNA sequences have proposed an RNA recognition code for PPR proteins involving a combination of amino acid residues at two or three positions in a PPR motif that determines the base preference (Barkan *et al*. 2012, Yagi *et al*. 2013, Takenaka *et al*. 2013a). However, it is not easy to identify target RNAs recognized by a PPR protein simply on the basis of its PPR code. The functions of many P-class PPR proteins are unknown and the RNA recognition mode of PPR motifs to the target RNA sequence is not fully understood.

The moss *Physcomitrella patens* genome encodes at least 89 P-class members, half of which are targeted to chloroplasts (Sugita *et al*. 2013). To date, our reverse-genetics approaches succeeded in

clarifying the contribution of five canonical P-class PPR proteins to splicing or processing of a specific RNA species in the *P. patens* chloroplasts (Hattori *et al*. 2007, Sugita *et al*. 2014, Goto *et al*. 2016, Ito *et al*. 2018). Here, we characterized P-class PpPPR_21 KO mutants and show that PpPPR_21 is essential for the accumulation of chloroplast *psbI-ycf12* dicistronic transcript. In addition, we demonstrated that PpPPR_21 bound specifically to the 5′ untranslated and translated regions of *psbI* transcript. Thus, PpPPR_21 may function as a stabilization factor specifically for the *psbI-ycf12* dicistronic transcript.

RESULTS

PpPPR_21 is localized in chloroplasts

The *PpPPR_21* gene (Pp3c22_3230V3.3/Pp1s266_45) was deduced to encode a protein of 860 amino acids (aa) with 19 PPR motifs (Figures 1a and S1). To examine its intracellular localization, a chimeric protein of the N-terminal PpPPR_21 (102 aa) fused to green fluorescent protein (GFP) was transiently expressed in moss cells. Green fluorescence was detected in the chloroplasts but not in mitochondria (Fig. 1b). The PpPPR_21 homologs, which are referred to as PPR21L, are widely distributed in bryophytes, ferns, and seed plants *A. thaliana* and *Zea mays* but not in the green alga *Chlamydomonas reinhardtii* (Fig. S1).

Photosynthesis was impaired in *PpPPR_21* knockout mosses

To examine the function of PpPPR_21, we generated *PpPPR_21* knockout (KO) mutants by replacing its coding region with the *gfp-hpt* gene cassette (Fig. S2a). Using PCR analysis, we verified that expected homologous recombination occurred in the targeted loci (Fig. S2b) and that *PpPPR_21* transcript was not amplified by RT-PCR in the KO mutants, Δ*21*-10 and Δ*21*-182 (Fig. S2c), indicating that they are null mutants. Both KO mutants grew more slowly than the wild type (WT) under phototrophic conditions (Fig. 2a). To investigate the photosynthetic status of these KO mutants, we analyzed them by kinetics multispectral fluorescence imaging. A chlorophyll fluorescence parameter, F_v/F_m, representing the integrity of photosystem II (PSII) was reduced to between 0.59 and 0.61 in KO mosses relative to the WT level of 0.75 (Fig. 2b). The effective quantum yield of PSII (ΦPSII) and non-photochemical quenching (NPQ) also decreased in the KO mutants at the range of light intensity analyzed. Photochemical quenching (qP) decreased slightly in the KO mutants. These results imply

that PSII activity was primarily affected, resulting in a reduction of other electron transport parameters. However, the plant growth and chlorophyll fluorescence parameters were rescued in a complemented moss, Comp-17, that was transformed with *PpPPR_21* full-length cDNA in the KO mutant Δ*21*-182 as a background line. The complementation experiment confirmed that the defect in growth and impaired photosynthesis in the mutant were caused by a loss-of-function of *PpPPR_21*.

We then performed immunoblot analysis to investigate the steady-state levels of the thylakoid membrane proteins (Fig. 3). The level of the PSII reaction center D1 protein (PsbA) was reduced to less than 10% of the WT level in the KO mutants. The level of the oxygen-evolving complex (OEC) extrinsic protein PsbO and cytochrome *f* (Cyt*f*) of the cytochrome b_6f complex were slightly reduced in the KO mosses. On the other hand, the PsaA protein of photosystem I (PSI) and the β subunit of H^+-ATP synthase (AtpB) in KO mosses accumulated to similar levels as WT. To analyze the formation of protein complexes in thylakoid membranes, thylakoid membrane proteins were separated by blue-native polyacrylamide gel electrophoresis (BN-PAGE). Consistent with the results of immuno-blotting (Fig. 3), the levels of PSII super-complexes (PSII SC) and PSII dimers were considerably reduced in the KO mutants, whereas the level of PSII monomer in the KO mutants was comparable to the WT level (Fig. 4).

_PpPPR_21_ KO mutants lost chloroplast _psbI-ycf12_ dicistronic mRNA

The loss of a PPR protein often leads to aberrant RNA processing or translation in chloroplasts. Since a remarkable reduction of PSII integrity and the PSII complex level was observed in the *PpPPR_21* KO mutants, RNA maturation steps, such as RNA stability and RNA splicing, could be affected in the PSII subunit gene(s) of the KO mutants. To assess this possibility, we performed a chloroplast transcriptomic analysis using a tiling array covering the chloroplast genome DNA of both strands (Sugiura *et al*. 2003). This analysis showed that transcript levels from the chloroplast genome positions 60,000 to 61,000 were markedly decreased in the KO mosses (Fig. 5a). To examine which gene expression was affected in the KO mutants, we performed Northern hybridization of a *psbK-psbI-ycf12-trnG* gene cluster, which is positioned at positions 60,000 to 61,000. The *psbK* transcripts (0.8 kb and 0.5 kb), *trnG*-UCC precursor (0.7 kb) and mature (0.1 kb) transcripts were observed at similar levels in WT and KO mosses (indicated as RNA1, 2, 5 and 6, Fig. 5b). In contrast, probing with *psbI* (probe B) and *ycf12* (probe C), a 1.1-kb transcript was observed in the WT but not in

the KO mosses (referred to as RNA3 in Fig. 5b). In addition, two transcripts longer than 1.1 kb were detected clearly in KO mutants while more less in WT and Comp-17. These may be primary and process transcripts from the *psbK-psbI-ycf12* gene cluster. The 1.1-kb RNA band might be a dicistronic *psbI-ycf12* transcript. The *ycf12* probe detected not only the 1.1-kb RNA but also an approximately 0.5-kb transcript (RNA4) in the WT. The 0.5-kb RNA accumulated in both WT and KO mutants and might be transcribed from its own dedicated promoter of the *ycf12* gene. The Comp-17 moss restored accumulation of the 1.1-kb *psbI-ycf12* transcript. This reconfirmed that the lack of accumulation of *psbI-ycf12* transcript in the KO mutants was due to a loss-of-function of the *PpPPR_21* gene. Thus, although the *psbI-ycf12* transcript was lost in the KO mutants, the monocistronic *ycf12* transcript accumulated normally, even in the KO mutants. This suggests that *ycf12* expression is not affected by the loss of PpPPR_21.

We then determined the 5′ and 3′ ends of the *psbI-ycf12* transcript (RNA3 in Fig. 5b) by sequencing DNA fragments amplified by circular RT (cRT)-PCR. In the WT and Comp-17 mosses, approximately 1.0-kb DNA fragments were amplified using two primers facing outward (black arrowheads) from RNA sample treated with T4 RNA ligase, but not from T4 RNA ligase-untreated RNA sample (Fig. 6a). In contrast, such DNA fragments were not amplified by cRT-PCR in the KO mutants. Accumulation of the 3′ *rps12-rps7* transcript amplified using gene-specific primers was not affected in the *PpPPR_21* KO mosses as expected. This result also supports no accumulation of the 1.1-kb *psbI* transcript in the KO mosses. DNA sequencing of the amplified fragments by cRT-PCR revealed that the 5′ ends of *psbI* transcript were positioned at 45 nucleotides (nt) upstream from the *psbI* translated region and that the 3′ ends were mostly 65 nt downstream from the TAA stop codon of the *ycf12* translated region (Fig. 6c, d). From this result, the length of *psbI-ycf12* transcript was calculated as 1060 nt, matching well with the 1.1 kb transcript (RNA3) detected by RNA gel blot hybridization (Fig. 5). Similarly, the 5′ and 3′ ends of RNA4 detected by RNA blot hybridization analysis were determined by cRT-PCR using a different set of primers (white arrowheads in Fig. 6b) and DNA sequencing of the amplified DNA fragments. The 5′ ends of RNA4 were at 295 nt upstream of the *ycf12* translated region and its 3′ ends were at the same positions as those of the 1.1-kb *psbI-ycf12* transcript. The length of RNA4 was calculated as 461 nt (Fig. 6c).

As a result of the lack of accumulation of *psbI* transcript in the KO mutants, we investigated the possibility that PpPPR_21 is also involved in the accumulation of other *psb* transcripts. To assess

whether expression of other *psb* genes was affected in the KO mutants, northern blot hybridization was performed using *psb* gene-specific probes. The analyzed *psb* gene transcripts in the KO mutants accumulated at levels similar to WT (Fig. S3). This result strongly suggests that PpPPR_21 is specifically involved in the accumulation of *psbI-ycf12* transcript.

PpPPR_21 binds to the 5′ untranslated and translated regions of the *psbI* transcript

Since PpPPR_21 is required for the accumulation of the 1.1 kb *psbI* transcript, PpPPR_21 may bind to some region in the *psbI* transcript as was observed in other chloroplast PPR stabilization factors (Pfalz et al. 2009, Johnson et al. 2010, Zhelyazkova et al. 2012, Zoschke et al. 2016).

As shown in Fig. S4a, we predicted a target sequence, 5′-CGAUUCUAUUUCUxUUUUx-3′, of PpPPR_21 according to the amino acid code for nucleotide recognition by PPR motifs (Yagi *et al*. 2013). Predicted nucleotide binding intensities were scanned through the *P. patens* chloroplast genome, resulting in the identification of four matching sites on the *psbI* gene (Fig. S4b). Two sites (ranked at 3 and 52 out of 60 top ranked sites) were overlapped by six nucleotides and were within the translated region of *psbI*. Another two (ranked at 28 and 56) were also overlapped by 14 nucleotides and positioned in the 5′-untranslated region (UTR) of *psbI*.

To confirm whether PpPPR_21 binds directly to the putative binding sites, we performed an RNA electrophoresis mobility shift assay (REMSA) using two 29-nt synthetic RNAs, PppsbI-RNA1 and PppsbI-RNA2, which correspond to the predicted binding sequence in the translated region and the 5′-UTR of *psbI* mRNA, respectively (Figs. 6d and 7a). As a negative control, two non-related sequences, AtpsbI-RNA3 and PpndhA-RNA4, were synthesized and used as probes (Fig. 7a). AtpsbI-RNA3 corresponds to the proximal region of the 5′-UTR of Arabidopsis processed *psbI* transcript because the 5′ end of Arabidopsis *psbI* transcript was mapped at 55-nt upstream from the ATG start codon of *psbI* by means of cRT-PCR and DNA sequencing (Fig. S5). PpndhA-RNA4 is a sequence located in the *P. patens* chloroplast *ndhA* intron. For REMSA, we expressed the recombinant thioredoxin (Trx)-PpPPR_21 fusion protein (rPp21) in *Escherichia coli* and recovered it by Ni-NTA agarose. Clear shifted bands were detected by PppsbI-RNA1 and PppsbI-RNA2 (Fig. 7b). However, rPp21 did not bind to AtpsbI-RNA3 and PpndhA-RNA4. This result suggests that PpPPR_21 binds specifically to the translated region and the 5′-UTR of *psbI* mRNA.

DISCUSSION

In this study, we showed that PpPPR_21 is required for the accumulation of chloroplast *psbI-ycf12* transcript encoding low molecular weight subunits of PSII. The *PpPPR_21* KO mutants lacking *psbI* transcripts grew slowly and exhibited a significant reduction of PSII core protein D1 (PsbA) level (less than 10% of the WT level), and a concomitantly poor level of PSII supercomplexes. Similar phenotypic features were reported in the green alga *C. reinhardtii* and tobacco *psbI* gene KO mutants (Künstner *et al*. 1995, Schwenkert *et al*. 2006). The *C. reinhardtii psbI* KO mutant was shown to grow photoautotrophically but showed a 10 to 20% decrease in the level of oxygen evolution relative to the WT level (Künstner *et al*. 1995). Tobacco *psbI* KO mutants were photoautotrophically viable under greenhouse conditions, but the levels of PsbA and PsbO were reduced to 50% compared with WT levels and the PSII complexes were poorly formed, suggesting that PsbI is essential for the stability of dimeric PSII and PSII supercomplexes (Schwenkert *et al.* 2006). Unlike the tobacco *psbI* KO mutants, the *PpPPR_21* KO mutants lacking *psbI* transcripts exhibited a significant and slight reduction of PsbA and PsbO, respectively. In cyanobacteria, PsbI was shown to be located on the periphery of the PSII dimer, and is required for the assembly process of PSII dimers (Dobáková *et al*. 2007, Kawakami *et al*. 2011). Likewise, the PSII supercomplex assembly is likely severely affected in the *PpPPR_21* KO mosses.

Ycf12 (Psb30) is also known to be essential for the optimal functionality of the PSII complex in high light intensity in *C. reinhardtii* (Inoue-Kashino *et al*. 2011). Although the *psbI* gene is present in the chloroplast genome of almost all oxygenic-photosynthetic organisms, the *ycf12* gene is distributed in oxygenic-photosynthetic organisms except for angiosperms, such as Arabidopsis. As presented in this study, *ycf12* (*psb30*) is not only co-transcribed with *psbI* but is also transcribed by its own promoter to produce an *ycf12* monocistronic form in the WT and KO mutant mosses. This suggests that expression of *ycf12* might not be affected in the *PpPPR_21* KO mosses. Cotranscription of *psbK* and *psbI* genes was reported in barley (Sexton *et al*. 1990), mustard (Neuhaus and Link 1990) and Arabidopsis (Tseng *et al*. 2010). In these plants, although the *psbK-psbI* intergenic region is site-specifically cleaved, both primary and processed transcripts accumulated at detectable levels following Northern blot analysis, indicating they are relatively stable. In contrast, *psbK-psbI-ycf12* tricistronic transcripts could be faintly detected but the processed 0.5 kb *psbK* transcript and 1.1 kb *psbI-ycf12* transcript substantially

accumulated in the *P. patens* chloroplasts. This suggests that *psbK-psbI* intergenic cleavage may proceed more rapidly in *P. patens* than in seed plants.

Our in vitro RNA binding experiment showed that PpPPR_21 bound specifically to both the 5′-UTR and the translated region of the *psbI* transcript. The 5′-UTR of the processed *psbI-ycf12* RNA can be folded into a possible secondary structure where a ribosome-binding site (RBS) is masked (Fig. 8a). Once PpPPR_21 bound to its target sequence and refolded the secondary structure to open the RBS, it is presumed that binding of PpPPR_21 to this site stabilized *psbI-ycf12* RNA by blocking 5′ → 3′ exonucleolytic degradation and mediated efficient translation (Fig. 8b). Likewise, it is well known that maize PPR10 prevents the formation of an RNA structure that masks the *atpH* RBS from an RNA duplex (Pfalz *et al*. 2009, Prikryl *et al*. 2011). Other P-class PPR proteins are known to stabilize RNAs and define the positions of processed RNA termini by blocking exoribonucleases (Schmitz-Linneweber *et al*. 2005, Fujii *et al.* 2013, Rojas *et al*. 2018). PpPPR_21 also binds the translated region of *psbI* mRNA in vitro and this binding may contribute to either stabilization or translation, or both, of mRNA. Most PPR stabilizers bind either the untranslated or intercistronic regions of transcripts while no PPR protein has been identified that binds the translated region in vivo and in vitro (Barkan and Small 2014, Manavski *et al*. 2018). To clarify whether PpPPR_21 binds the translated region of *psbI* mRNA in vivo and contributes to the control of translational efficiency, further analyses need to be performed.

Homologous sequences of *PpPPR_21* are found in land plants ranging from bryophytes to seed plants (Fig. S1). The *PpPPR_21* gene is interrupted by four introns while *PPR21L* genes, except for the moss *Sphagnum fallax* gene, are intron-less. PpPPR_21 showed 45% aa identity and 85% similarity to Arabidopsis PPR21L (At5g02860), which was predicted to be localized in chloroplasts (Colcombet *et al.* 2013). A predicted target sequence of Arabidopsis PPR21L is 5′-CGAUUCUAUAUCUxCUUCx-3′, which is similar to the sequence of 5′-GAAUCUAUUCUCUUUUUU-3′ sequence in the 5′-UTR of the Arabidopsis *psbI* transcript (Fig. S5c). This suggests that AtPPR21L is a functional ortholog of PpPPR_21. To investigate this possibility, we analyzed the T-DNA tagged lines of Arabidopsis *PPR21L* (*At5g02860*), GABI_290B09, SALK_087900 and SALK_089346 (Fig. S6a). GABI_290B09 and SALK_087900 have T-DNA inserted into the translated region. However, no homozygous mutants of these two T-DNA tagged lines were obtained

by analyzing the genotype of the next generations of heterozygous plants. A problem was encountered when attempting to isolate a homozygous mutant in the two independent T-DNA insertion lines, so the loss of function of *AtPPR21L* may lead to embryonic lethality. A *Ds*/*Spm*-tagged mutant of At*PPR21L* (*At5g02860*) lacking homozygotes was previously reported as mutant stock 54-4198-1 (Myouga *et al*. 2010). In contrast, a homozygous mutant of the SALK_089346 line, in which T-DNA was inserted at 24-bp upstream of the *AtPPR21L* translated region, was obtained (Fig. S6b). This homozygous mutant did not show a reduction of *AtPPR21L* transcript level (Fig. S6c) or a visible phenotype but showed a normal level of chlorophyll fluorescence F_v/F_m (the integrity of PSII), under our growth conditions (Fig. S6d, e). From these observations, we cannot conclude that *AtPPR21L* is a functional ortholog of PpPPR_21. To fortify this conclusion, we will further generate and characterize *AtPPR21L* mutants, for example by generating knockdown mutants.

In angiosperms and green algae, the RNA binding of PPR proteins serving as cap for ribonucleases protection leave small RNA footprints that accumulate in vivo (Ruwe *et al*. 2012, Cavaiuolo *et al*. 2017). In Arabidopsis, a small RNA footprint (5′-CCAUACUAAAUCUGGAUCAUUUC-3′) was identified in the chloroplasts and mapped at positions 261 to 240-bp upstream of the *psbI* coding region and 118 to 143-bp downstream of *psbK* coding region (Ruwe *et al.* 2012). This small RNA sequence was not matched with the predicted binding site of AtPPR21L. The other small RNAs were not identified in the *psbK-psbI-trnS* gene cluster in Arabidopsis. Identification of small RNAs resulting from protective action of PpPPR_21 needs to be addressed for elucidating the precise function of PpPPR_21 in the *P. patens* chloroplasts.

EXPERIMENTAL PROCEDURES

Plant growth conditions

P. patens was phototrophically grown at 25°C as described in Ito *et al*. (2018) and *A. thaliana* was grown in soil at 23°C as described in Yamamoto *et al*. (2011). The T-DNA-tagged lines (GABI_290B09, SALK_087900 and SALK_089346) were provided by the Arabidopsis Biological Resource Center (ABRC; https://abrc.osu.edu/).

Intracellular localization

Total cellular RNA was reverse-transcribed to synthesize cDNA. A cDNA encoding the N-terminal 102

aa sequence of PpPPR_21 was amplified using specific primers (Table S1), and cloned into pKSPGFP9 (Tasaki *et al.* 2010). The obtained plasmid p21N-GFP was introduced by particle bombardment into the transgenic Mt-RFP OX moss and fluorescence was detected as described by Ichinose *et al*. (2013).

Generation of *PpPPR_21* KO moss and complemented moss

The DNA fragments of the 1087-bp region upstream and the 1028-bp region downstream from *PpPPR_21* were amplified by PCR from genomic DNA with gene-specific primers (Table S1). Respective DNA fragments were cloned into pNGH4 (Ito *et al*. 2018). The resultant plasmid p21KO was linearized with *Nae*I and used to transform the moss protonemata, and hygromycin-resistant mosses were selected. Gene disruption in transformants was confirmed by genomic PCR and null KO mutants were verified by RT-PCR with appropriate primers (Table S1, Fig. S2).

In order to complement the KO mosses, *PpPPR_21* cDNA was amplified by 21P3 and 21P4 primers (Table S1), and cloned into p9WmycZ3 (Goto *et al*. 2016). The obtained plasmid was digested with *Not*I and introduced into the KO mutant Δ*21*-182 and zeocin-resistant mosses were selected. The nucleotide sequence of *PpPPR_21* cDNA was deposited in the DDBJ DNA database under accession number LC380412.

Analysis of chlorophyll fluorescence

Chlorophyll fluorescence from moss colonies and Arabidopsis leaves was measured as described previously (Ito *et al*. 2018).

Immunoblot analysis

Total cellular proteins were extracted from the moss protonemata and separated on 0.1% SDS–14% polyacrylamide gels. Blotting of proteins to nylon membranes and immunodetection were carried out as described in Ito *et al*. (2018). The anti-PsbA (AS05084A, Agrisera, http://www.agrisera.com/), anti-PsaA (AS06172, Agrisera), anti-PsbO (provided by F. Sato), anti-cytochrome *f* (provided by A. Makino) and anti-β-subunit of chloroplast H^+-ATP synthase (provided by T. Hisabori) were used.

Blue native gel electrophoresis

Thylakoid membrane preparation, BN-PAGE and 2D-PAGE were performed as described earlier (Shimizu *et al.* 2008). Thylakoids (10 μg of chlorophyll) in 25 mM Bis Tris-HCl (pH 7.0) and 20% glycerol were solubilized with 1% *n*-dodecyl-β-D-maltoside on ice for 10 min in the dark. After centrifugation, the supernatant was supplemented with one-tenth BN sample buffer (5% Serva blue G, 100 mM BisTris-HCl (pH 7.0), 0.5 M 6-amino-n-caproic acid, and 30% (w/v) glycerol) and electrophoresis was performed in the first (5 to 13.5% BN-PAGE) and second (14% SDS-PAGE) dimension. Immunological analysis of the PsbA protein was performed for the second dimension.

Microarray analysis

Agilent-084494 Physcomitrella Custom Microarray was a tiling array covering the chloroplast and mitochondrial genome DNA of both strands, and whose probes were 60 nt in length with a 20 nt overlap. Total cellular RNA was extracted from 4 day-old WT or KO mosses using Isogen II (Nippongene, http://www.nippongene.com/) and treated with RNase-free DNase I. About 150 ng of total cellular RNA was reverse transcribed using random primers binding to the T7 promoter sequence. Complementary RNAs (cRNAs) were transcribed and labeled from these cDNAs by a Low Input Quick Amp WT Labeling kit (Agilent, https://www.agilent.com/). Six-hundreds ng of Cyanine-3 (Cy-3) labeled cRNA was used for hybridization, according to the manufacturers' protocol using the Gene Expression Hybridization Kit and the Gene Expression Wash Pack (Agilent). The microarrays were scanned with an Agilent SureScan and processed using Agilent Feature Extraction software. The expression values were quantile-normalized. The microarray data are available on the Gene Expression Onmibus (GEO accession ID: GSE121554).

RNA gel blot hybridization

RNA (10 or 15 μg) was loaded onto a 1% agarose gel and transferred to a nylon membrane. The blotted RNAs were hybridized with gene-specific DNA probes (Table S1) as described previously (Goto *et al*. 2016).

cRT–PCR

Two μg of total RNA treated with DNase I was self-ligated using T4 RNA ligase (TaKaRa, http://www.takara-bio.co.jp). RNA 5′ pyrophosphohydrolase (RppH, New England Biolabs,

https://international.neb.com/) was used to remove pyrophosphates from the 5′ end of triphosphate RNA (Hetzel et al. 2016). The circularized RNA was reverse-transcribed using random primers and ReverTra Ace (TOYOBO, http://lifescience.toyobo.co.jp). For the *psbI-ycf12* transcript, cDNA was amplified with the gene-specific primers (Table S1) and PrimeSTAR GXL DNA polymerase (TaKaRa). The cRT–PCR products were separated by 1% agarose gel, recovered from the gel, and cloned into the *Sma*I site of pUC18 and sequenced.

Prediction of PpPPR_21 RNA binding site

Prediction of the binding sites for PpPPR_21 was performed according to Yagi *et al*. (2013). The nucleotide-specifying residues (NSRs; positions 2, 5, 35) were extracted from each PPR motif, which were defined by PPR database (http://ppr.plantenergy.uwa.edu.au; Cheng *et al*. 2016). The NSRs were converted into a probability matrix that indicated the decoding nucleotide frequency according to the PPR code using table S4 in Yagi *et al.* (2013). This probability matrix was used to search PpPPR_21 RNA binding sites against both strands of the entire *P. patens* chloroplast genome (AP005672.1) by the FIMO program in the MEME suite (http://meme.nbcr.net/meme/fimo-intro.html). The predicted binding sites were ranked by *P*-values calculated by FIMO.

Recombinant protein and REMSA

To express mature PpPPR_21 as a fusion protein with Trx at its N-terminus in *Escherichia coli*, complementary DNA coding for PpPPR_21(excluding its N-terminal 102 aa) was amplified using specific primers (Table S1) and was cloned into pBAD/Thio-TOPO (Invitrogen, http://www.introgen.com/). The recombinant protein, rPp21, was expressed at 16°C for 16 h in *E. coli* BL21 in the presence of 0.2% arabinose, and expressed protein was isolated using Ni-NTA agarose (Qiagen, http://www.qiagen.com/). ^{32}P-labeling of RNA probes and REMSA were carried out as described in Goto *et al*. (2016).

ACKNOWLEDEMENTS

We are grateful to Fumihiko Sato, Toru Hisabori and Amane Makino for providing the antibodies. We also thank ABRC for providing seeds of Arabidopsis T-DNA-tagged lines. This work was supported by JSPS KAKENHI grant nos 17K08195 (to MS), 16H06555 (to TS) and 18K14435 (to MI).

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

SUPPORTING INFORMATION

Figure S1. Multiple sequence alignment and phylogenetic tree of PpPPR_21 and its homologs.

Figure S2. Generation of *PpPPR_21* knockout (KO) mutants.

Figure S3. RNA gel blot hybridization of *Physcomitella patens psb* genes.

Figure S4. Prediction of the PpPPR_21-binding site according to its PPR code.

Figure S5. Determination of 5' and 3' ends of Arabidopsis *psbK-psbI-trnS* transcript.

Figure S6. Isolation and characterization of Arabidopsis *PPR21L* T-DNA tagged mutants.

Table S1. Primers used for plasmid construction and DNA/RNA analyses.

REFERENCES

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.

Belcher, S., Williams-Carrier, R., Stiffler, N. and Barkan, A. (2015) Large-scale genetic analysis of chloroplast biogenesis in maize. *Biochim. Biophys. Acta* **1847**, 1004-1016.

Cavaiuolo, M., Kuras, R., Wollman, F.A., Choquet, Y. and Vallon, O. (2017) Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism. *Nucleic Acids Res*. **45**, 10783-10799.

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.

Dobáková, M., Tichy, M. and Komenda, J. (2007) Role of the PsbI protein in photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol*, **145**, 1681-1691.

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., Sato, N. and Shikanai, T. (2013) Mutagenesis of individual pentatricopeptide repeat motifs affects RNA binding activity and reveals functional partitioning of Arabidopsis PROTON GRADIENT REGULATION3. *Plant Cell* **25**, 3079-3088.

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

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.

Gutmann, B., Gobert, A. and Giege, P. (2012) Mitochondrial genome evolution and the emergence of PPR proteins. *Adv. Bot. Res.* 63, 253-313.

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.

Hetzel, J., Duttke, S. H., Benner, C. and Chory. J. (2016) Nascent RNA sequencing reveals distinct features in plant transcription. *Pro. Natl. Acad. Sci. USA* **113**, 12316–12321.

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

Inoue-Kashino, N., Kashino, Y. and Takahashi, Y. (2011) Psb30 is a photosystem II reaction center subunit and is required for optimal growth in high light in *Chlamydomonas reinhardtii*. *J. Photochem. Photobiol. B: Biol.* **104**, 220-228.

Ito, A., Sugita, C., Ichinose, M., Kato, Y., Yamamoto, H., Shikanai, T. and Sguita, M. (2018) An evolutionarily conserved P-subfamily pentatricopeptide repeat protein is required to splice the plastid *ndhA* transcript in the moss *Physcomitrella patens* and *Arabidopsis thaliana*. *Plant J*. **94**, 638-648.

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.

Kawakami, K., Umena, Y., Iwai, M., Kawabata, Y., Ikeuchi, M., Kamiya, N. and Shen, J.R. (2011) Roles of PsbI and PsbM in photosystem II dimer formation and stability studied by deletion mutagenesis and X-ray crystallography. *Biochim. Biophys. Acta* **1807**, 319-325.

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.

Künstner, P., Guardiola, A. Takahashi, Y. and Rochaix, J.D. (1995) A mutant strain of *Chlamydomonas reinhardtii* lacking the chloroplast photosystem II *psbI* gene grows photoautotrophically. *J. Biol. Chem*. **270**, 9651-9654.

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.

Manavski, N., Schmid, L. and Meurer, J. (2018) RNA-stabilization factors in chloroplasts of vascular plants. *Essays Biochem*. doi: 10.1042/EBC20170061.

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.

Myouga, F., Akiyama, K., Motohashi, R., Kuromori, T., Ito, T., Iizumi, H., Ryusui, R., Sakurai, T. and Shinozaki, K. (2010) The chloroplast function database: a large-scale collection of Arabidopsis *Ds*/*Spm*- or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. *Plant J*. **61**, 529-542.

Neuhaus, H. and Link, G. (1990) The chloroplast *psbK* operon from mustard (*Sinapis alba* L.): multiple transcripts during seedling development and evidence for divergent overlapping transcription. *Curr. Genet*. **18**, 377-383.

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.

Prikryl, J., Rojas, M., Schuster, G. and Barkan, A. (2011) Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc. Natl. Acad. Sci. USA* **108**, 415–420.

Ruwe, H. and Schmitz-Linneweber, C. (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res*. **40**, 3106-3116.

Rojas, M., Ruwe, H., Miranda, R.G., Zoschke, R., Hase, N., Schmitz-Linneweber, C. and Barkan, A. (2018) Unexpected functional versatility of the pentatricopeptide repeat proteins PGR3, PPR5 and PPR10. *Nucleic Acids Res*. doi: 10.1093/nar/gky737.

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., Williams-Carrier, R. and Barkan, A. (2005) Immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5’ region of mRNAs whose translation it activates. *Plant Cell* **17**, 2791-2804.

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.

Schwenkert, S., Umate, P., Dal Bosco, C., Volz, S., Mlçochová, L., Zoryan, M., Eichacker, L.A., Ohad, I., Herrmann, R.G. and Meuer, J. (2006) PsbI affects the stability, function, and phosphorylation patterns of photosystem II assemblies in tobacco. *J. Biol. Chem*. **281**,

34227-34238.

Sexton, T.B., Jones, J.T. and Mullet, J.E. (1990) Sequence and transcriptional analysis of the barley ctDNA region upstream of *psbD-psbC* encoding *trnK* (UUU), *rps16*, *trnQ* (UUG), *psbK*, *psbI*, and *trnS* (GCU). *Curr. Genet*. **17**, 445-454.

Shimizu, H., Peng, L., Myouga, F., Motohashi, R., Shinozaki, K. and Shikanai, T. (2008) CRR23/NdhL is a subunit of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis. *Plant Cell Physiol*. **49**, 835–842.

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., Brennicke, A. and Graichen, K. (2013a) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* **8**, e65343.

Takenaka, M., Zehrmann, A., Verbitskiy, D., Härtel, B. and Brennicke, A. (2013b) 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.

Tseng, C.C., Sung, T.Y., Li, Y.C., Hsu, S.J., Lin, C.L. and Hsieh, M.H. (2010) Editing of *accD* and *ndhF* chloroplast transcripts is partially affected in the *Arabidopsis vanilla cream1* mutant. *Plant Mol. Biol*. **73**, 309-323.

Yagi, Y., Hayashi, S., Kobayashi, K., Hirayama, T. and Nakamura, T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS ONE* **8**, e57286.

Yamamoto, H., Peng, L., Fukao, Y. and Shikanai, T. (2011) An Src homology 3 domain-like fold

protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in Arabidopsis. *Plant Cell* **23**, 1480–1493.

Z**helyazkova, P., Hammani, K., Rojas, M., Voelker, R., Vargas-Suárez, M., Börner, T., Barkan, A.** (2012) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.*, **40**, 3092–3105.

Zoschke, R., Watkins, K.P., Miranda, R.G. and Barkan, A. (2016) The PPR-SMR protein PPR53 enhances the stability and translation of specific chloroplast RNAs in maize. *Plant J*. **85**, 594-606.

Figure legends

Figure 1. Chloroplast localization of PpPPR_21 protein.

(a) Diagram of PpPPR_21 composed of a putative transit peptide (TP) and 19 pentatricopeptide repeat motifs. Fusion protein 21N-GFP is shown below.

(b) Chimeric protein was transiently expressed in the mitochondria-localized red fluorescence protein (Mt-RFP) overexpressing moss. Fluorescence of 21N-GFP (GFP), RFP (Mt-RFP) and chlorophyll fluorescence were detected by confocal fluorescent microscopy. An overlay of fluorescence images (Merged) is shown. Scale bars = 20 μm.

Figure 2. Phenotype of *PpPPR_21* gene knockout mutants and a complemented moss.

(a) Wild type (WT), knockout (KO) mutants (Δ*21*-10, Δ*21*-182) and complemented moss (Comp-17) were grown on BCDAT plates for 2 weeks without hygromycin B. Scale bars = 10 mm.

(b) Measurement of the maximum PSII yield in dark-adapted state (F_v/F_m) and light intensity dependence of effective quantum yield of PSII (ΦPSII), non-photochemical quenching (NPQ) and photochemical quenching (qP) in WT, KO mutants, and a complemented moss. Values are means ± SD (n = 6). The horizontal axis indicates photon flux density (PFD).

Figure 3. Immunoblot analysis of chloroplast proteins from *Physcomitrella patens*.

Total cellular proteins [the indicated dilution of the wild-type (WT) sample] were subjected to immunoblot analysis with antibodies for PSII core D1 protein (PsbA), PSII oxygen-evolving complex (OEC) extrinsic protein (PsbO), cytochrome *f* of the cytocrome b_6f complex (Cyt*f*), PSI core (PsaA) and β-subunit of ATP synthase (AtpB). The gel was stained with Coomassie brilliant blue G-250 (CBB) and the large subunit of RuBisCO (RbcL) and the light-harvesting chlorophyll binding protein (LHCII) are indicated (bottom).

Figure 4. Detection of PSII complexes by blue-native/SDS polyacrylamide gel electrophoresis.

(a) Blue native (BN)-polyacrylamide gel electrophoresis (PAGE) analysis of thylakoid membrane protein complexes. Thylakoids (10 μg of chlorophyll) were solubilized with 1% *n*-dodecyl-β-D-maltoside, and electrophoresis was performed.

(b) Two-dimensional separation [BN-gel follows SDS-PAGE] of thylakoid proteins complexes. Complexes in BN-gel were subsequently separated by SDS-PAGE. Immunological analysis of the PsbA protein was performed in the second dimension.

Figure 5. RNA gel blot hybridization of the *psbK-psbI-ycf12-trnG* cluster in *Physcomitrella patens*.
(a) Ratio of RNA abundance in the knockout (KO) mutants relative to the wild type (WT) measured by a replicate microarray. Microarray data from chloroplast genome positions 57,500 to 62,500 is shown.
(b) Total RNA (10 µg) from WT, KO mutant mosses (Δ*21*-10 and Δ*21*-182) and complemented moss (Comp-17) was analyzed by RNA gel blot hybridization using DNA probes A (358 bp), B (315 bp), C (386 bp) and D (642 bp). The detected RNA bands are indicated as numbers 1 to 6 and the gels stained with ethidium bromide are shown below. RNA size markers (0.2 to 4.0 kb) are indicated on the right.

Figure 6. Determination of the 5′ and 3′ end positions of *psbI-ycf12* transcript in *Physcomitrella patens*.
(a) Total RNA was treated with (+) or without (-) T4 RNA ligase to form circular RNAs, then subjected to RT-PCR using two different primer sets as indicated by black or white arrowheads in panel (c). Circular RT (cRT)-PCR products of *psbI-ycf12* and 3′ *rps12-rps7* transcripts were detected on 2% agarose gels. RNA ligase treatment did not affect amplification of RT-PCR product as shown in *PpActin1*.
(b) Total RNA was treated with (+) or without (-) RNA 5′ pyrophosphohydrolase (RppH) and T4 RNA ligase and the resultant circular RNAs were subjected to RT-PCR. cRT-PCR products were detected in wild type (WT) and Δ*21*-10 moss.
(c) cRT-PCR products were cloned and sequenced. The 5′ ends of *psbI-ycf12* transcripts were mapped at 45 nucleotides (nt) upstream from the *psbI* translated region and their 3′ ends were mostly mapped at 65-nt downstream from the *ycf12* translated region. When a different primer set (white arrowheads) was used, cRT-PCR products corresponding to 461-nt long transcripts and their 5′ and 3′ ends were mapped at 295-nt upstream and 65-nt downstream of the *ycf12* translated region, respectively.
(d) Nucleotide sequence of *psbI-ycf12* transcript mapped at 5′ end mapped by cRT-PCR and sequencing. Ribosome-binding site (RBS) is boxed and the translated region of *psbI* is bold-faced. RNA probe sequences used for REMSA (Fig. 7) are also underlined.

Figure 7. *In vitro* binding of the recombinant PpPPR_21 to RNA probes containing the predicted RNA binding site.

(a) Coomassie brilliant blue-stained recombinant PpPPR_21 protein (rPp21, 2 µg) separated on an 8% polyacrylamide-SDS gel.

(b) The nucleotide sequences of RNA probes used for REMSA are shown together with their location within respective RNA in parenthesis. In PppsbI-RNA1 and -RNA2, the nucleotides matching the predicted binding site of PpPPR_21 are shown in red.

(c) REMSA was performed with recombinant proteins (rTrx or rPp21) and ^{32}P-labeled RNA probe. All RNA probes were used at 50 pM with protein concentrations ranging from 0 (-) to 5 nM as shown above each lane.

Figure 8. Model of PpPPR_21 function.

(a) Predicted secondary structure of the 5'-UTR of the *P. patens psbI*. Putative binding site of PpPPR_21 and ribosome-binding site (RBS) are indicated.

(b) Once PpPPR_21 binds to the 5′-UTR of the *psbI-ycf12* RNA, it may prevent the formation of an RNA structure that masks the *psbI* RBS and also may stabilize *psbI-ycf12* mRNA by blocking 5′ → 3′ exonucleolytic degradation. PpPPR_21 also binds to the translated region of *psbI* and mediates either stabilization or translation, or both, of *psbI* RNA.

Supplementary figures

Figure S1. Multiple sequence alignment and phylogenetic tree of PpPPR_21 and its homologs.

(a) 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_21 (Pp3c22_3230V3.3); SfPPR21L, *Sphagnum fallax* PPR_21-like (Sphfalx0062s0115); MpPPR21L, *Marchantia polymorpha* (Mapoly0063s0054.1); SmPPR21L, *Selaginella moellendorffii* (61162); AtPPR21L, *Arabidopsis thaliana* (At5g02860); OsPPR21L, *Oryza sativa* (LOC_Os07g40120); ZmPPR21L, *Zea mays* (GRMZM2G092739_T01), PPR motifs are marked in red brackets 1 to 19. The position 5 and 35 aa residues in each PPR motif are indicated by red and black asterisks, respectively. The arrowhead indicates the predicted cleavage site

628 of the transit peptide.
629 (b) The phylogenetic tree including PPR21L and two most closely related PPR paralogues within each
630 taxon was constructed using MEGA X program by neighbor-joining method (Saitou and Nei 1987).
631 Bootstrap values from 1000 replicates are indicated at each branch as percentages.
632 Pp3c11_7720V3.2(PpPPR_45) was used as an outgroup.
633
634 **Figure S2.** Generation of *PpPPR_21* knockout (KO) mutants.
635 (a) Structures of wild-type (WT) and the altered genomic locus after replacement of the *gfp-hpt* gene
636 cassette by homologous recombination (HR) are illustrated. Primers and the expected fragment sizes
637 for PCR analysis are also shown. Primer sequences are listed in Table S1. The DNA regions for HR
638 are represented as thick horizontal lines.
639 (b) Genomic PCR analysis of WT and KO mutants. The predicted 1308- (5′ HR) and 1605-bp (3′ HR)
640 fragments were amplified from the KO lines while the 3629-bp fragment (full length gene) was
641 amplified from WT. DNA size marker is the λ DNA *Sty*I-digest (lanes M).
642 (c) RT-PCR for detection of *PpPPR_21* transcript in WT, KO and complemented (Comp-17) mosses.
643 *PpActin1* transcript was also amplified as a control.
644
645 **Figure S3.** RNA gel blot hybridization of *Physcomitrella patens psb* genes.
646 Total RNA (15 µg or 10 µg for *psbA*) from *P. patens* wild type (WT) and knockout mutant (Δ*21*-10,
647 Δ*21*-182) mosses was analyzed by northern blot hybridization using gene-specific probes (Table S1).
648 RNA size markers (0.2 to 8 kb) are indicated on the right. Lanes 1, 2 and 3 indicate WT, Δ*21*-10 and
649 Δ*21*-182, respectively.
650
651 **Figure S4.** Prediction of the PpPPR_21-binding site according to its PPR code.
652 (a) The amino acids at positions 2, 5 and 35 of each PpPPR_21 PPR motif were extracted and are
653 listed from the N- to the C-terminus. The obtained combinations were then used to calculate the
654 probabilities of nucleotide recognition by each individual PPR motif according to the PPR code (Yagi *et*
655 *al*. 2013).
656 (b) This predicted target sequence was scanned through the *P. patens* chloroplast genome and four
657 matching sites were found on the *psbI–ycf12* transcript. The numbers indicate the chloroplast genome

position (position column).

Figure S5. Determination of 5′ and 3′ ends of Arabidopsis *psbK-psbI-trnS* transcript.

(a) Total RNA was treated with (+) or without (-) RNA 5′ pyrophosphohydrolase (RppH) or T4 RNA ligase to form circular RNAs, then subjected to RT-PCR using a set of primer sets (black arrowheads) as indicated in (b). Circular RT (cRT)-PCR products of *psbI-trnC* transcripts were detected on a 2% agarose gels. RppH and T4 RNA ligase treatment did not affect amplification of RT-PCR product as shown in *AtpsbI*. The asterisk indicates non-specific PCR products.

(b) cRT-PCR products were cloned and sequenced. The 5′ ends of *psbI-trnC* transcript were mapped at 55 nucleotides (nt) upstream from the *psbI* translated region and the 3′ ends were mapped at 751 and 858 nt downstream from A of the translation start codon of the *psbI* translated region. The 5′ end of the *psbK-psbI-trnC* transcript was mapped at 164-nt upstream from the *psbK* translated region.

(c) Nucleotide sequence of the 5′-UTR and a part of of the *psbI* translated region. Ribosome-binding site (RBS) is boxed and the *psbI* translated region is bold-faced. The AtpsbI-RNA3 sequence used for REMSA (Fig. 7) is also underlined.

Figure S6. Isolation and characterization of Arabidopsis *PPR21L* T-DNA tagged mutants.

(a) Schematic gene structure of Arabidopsis *PPR21L* (*At5g02860*). This structure is based on representative gene model reported in The Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org/). Positions of T-DNA insertion relative to translational start codon (+1) in GABI_290B09, SALK_087900 and SALK_089346 are indicated. Open box indicates a translated region. Positions of primers used for PCR in (b) are indicated by arrowheads.

(b) Genotyping of Arabidopsis *PPR21L* gene KO mutants. PCR was performed on genomic DNA to detect homozygosity of the T-DNA-tagged line SALK_089346. T-DNA-specific primer LBb1.3 and gene-specific primers (LP and RP) were used for PCR. Primer set used for PCR are indicated in the bottom of the figure. Primer sequences are listed in Table S1. The amplified fragments were separated by agarose gel electrophoresis. Lane M indicates the DNA size marker.

(c) RT-PCR for detection of *AtPPR21L* and *rbcL* transcripts in wild type (WT) and SALK_089346. Lane numbers indicate reaction cycles of PCR.

(d) Growth phenotype of the WT and *Arabidopsis PPR21L* T-DNA insertion line. Plants were grown in a

688 growth chamber at 50 µmol photons m^{-2} s^{-1} under long-day conditions (16-hr light/8-hr dark) for 3
689 weeks.
690 (e) Measurement of F_v/F_m in WT and SALK_089346 leaves. Values are means ± SD (n = 12).

Figure 1

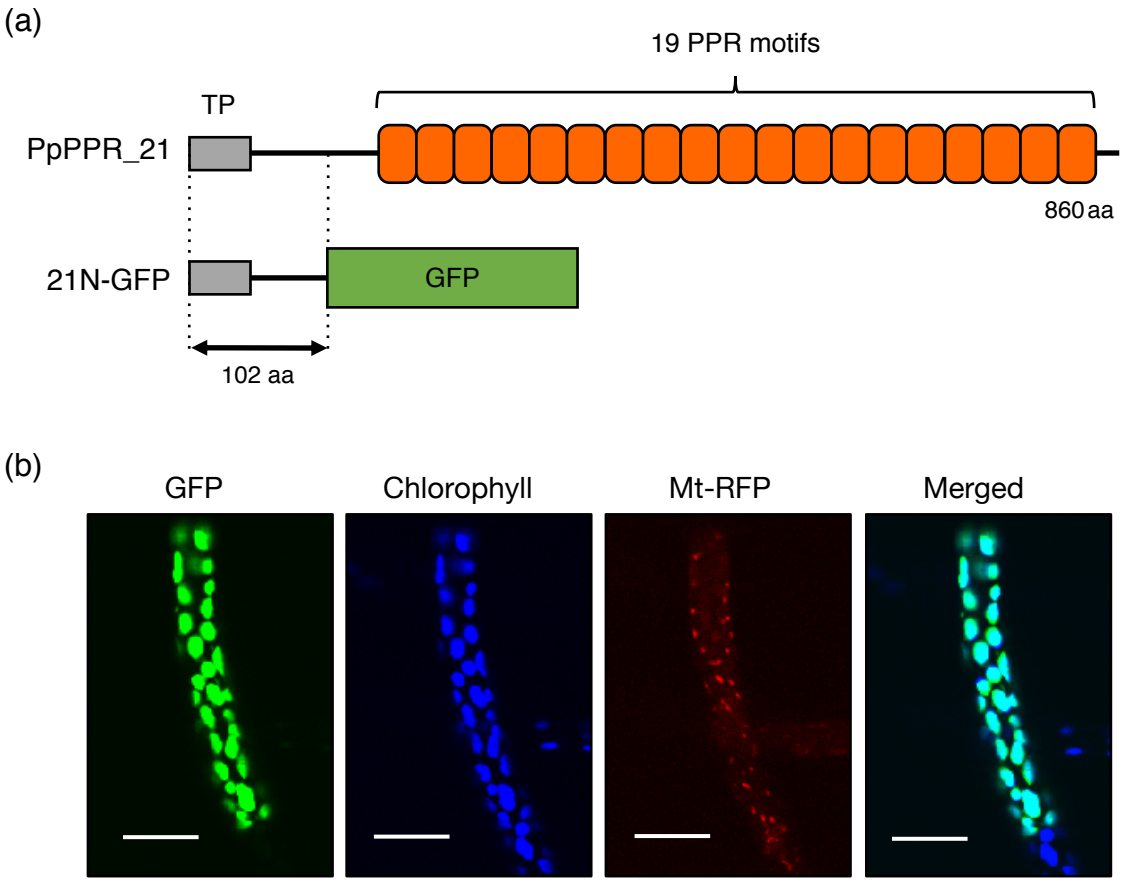
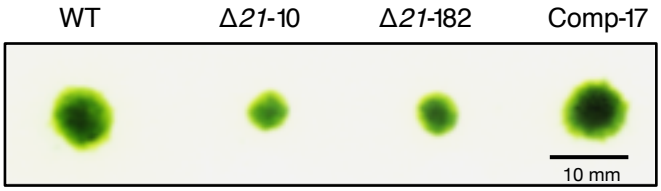

(a)
19 PPR motifs
TP
PpPPR_21
860 aa
21N-GFP
GFP
102 aa
(b)
GFP
Chlorophyll
Mt-RFP
Merged

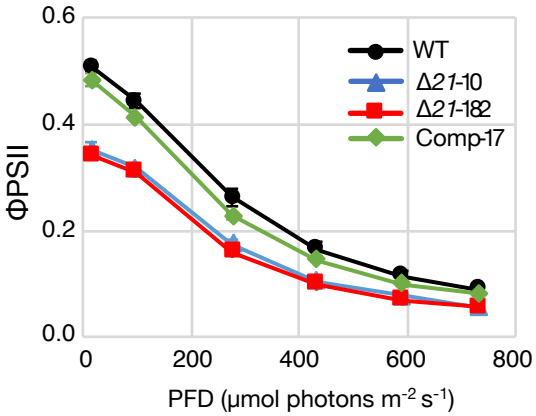
Figure 2

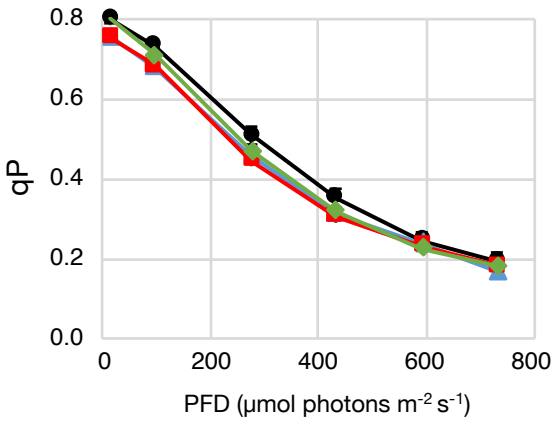
(a)



(b)

Plant	F_v/F_m
WT	0.753 ± 0.005
Δ*21*-10	0.596 ± 0.012
Δ*21*-182	0.613 ± 0.005
Comp-17	0.757 ± 0.005





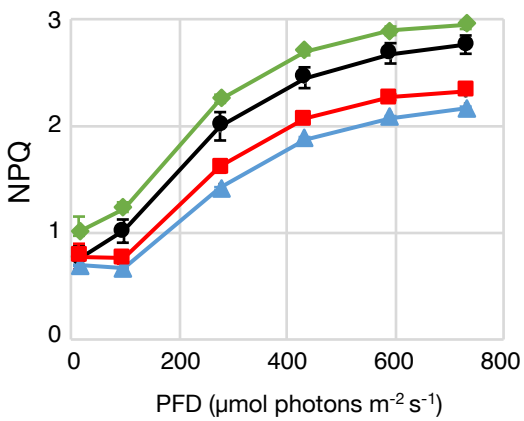


Figure 3

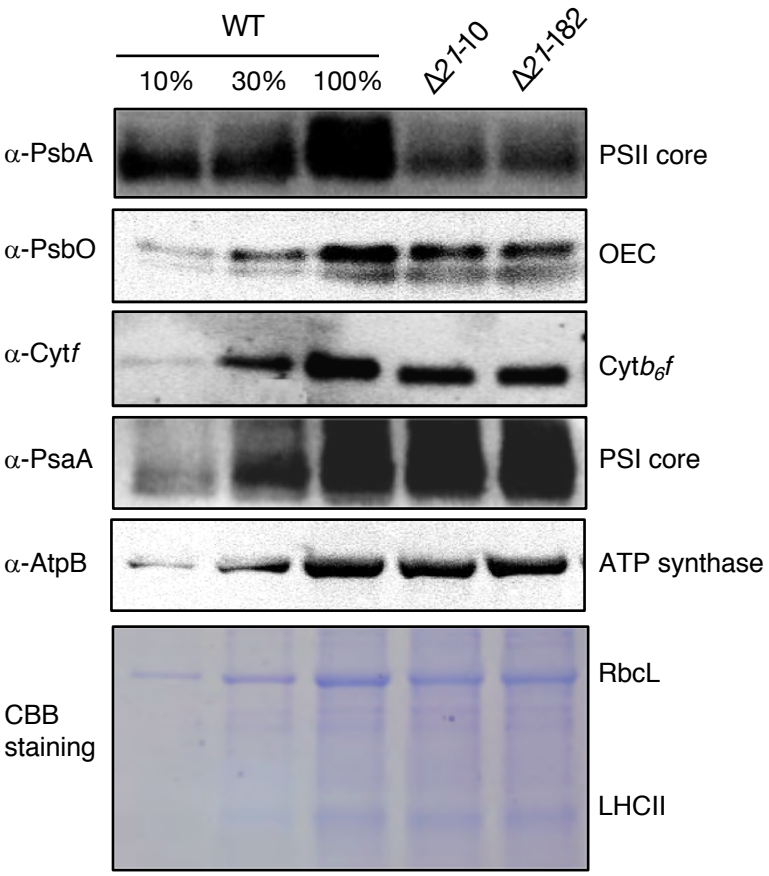

WT
10%
30%
100%
Δ21-10
Δ21-182
α-PsbA
PSII core
α-PsbO
OEC
α-Cytf
Cytb6f
α-PsaA
PSI core
α-AtpB
ATP synthase
CBB staining
RbcL
LHCII

Figure 4

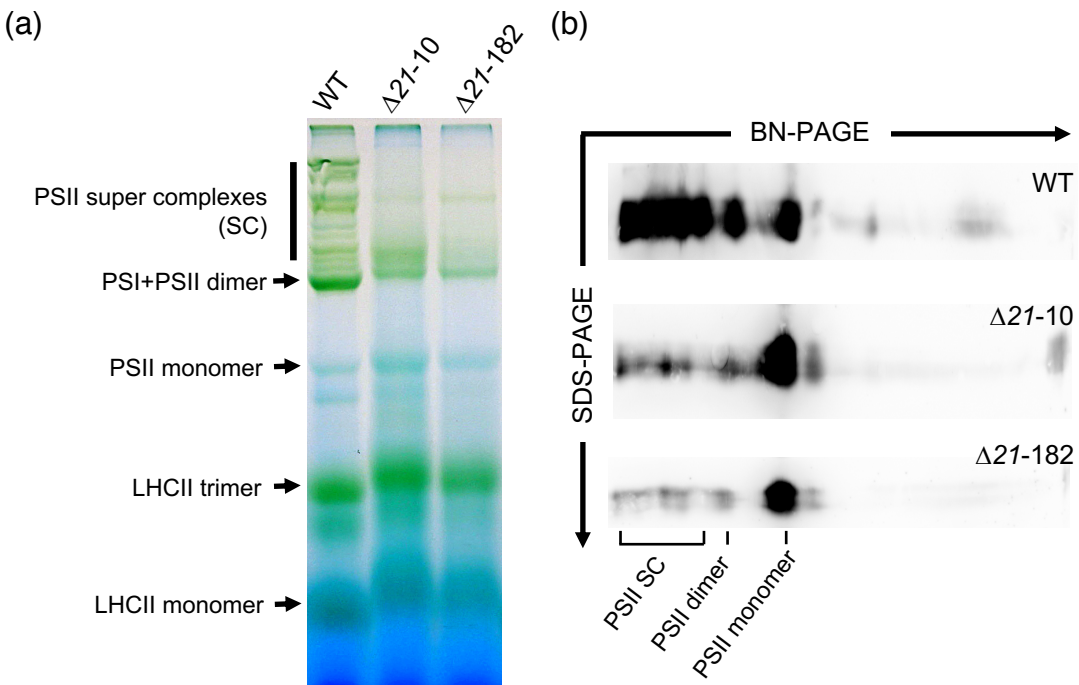

(a)
WT
Δ21-10
Δ21-182
PSII super complexes (SC)
PSI+PSII dimer
PSII monomer
LHCII trimer
LHCII monomer
(b)
BN-PAGE
SDS-PAGE
WT
Δ21-10
Δ21-182
PSII SC
PSII dimer
PSII monomer

Figure 5

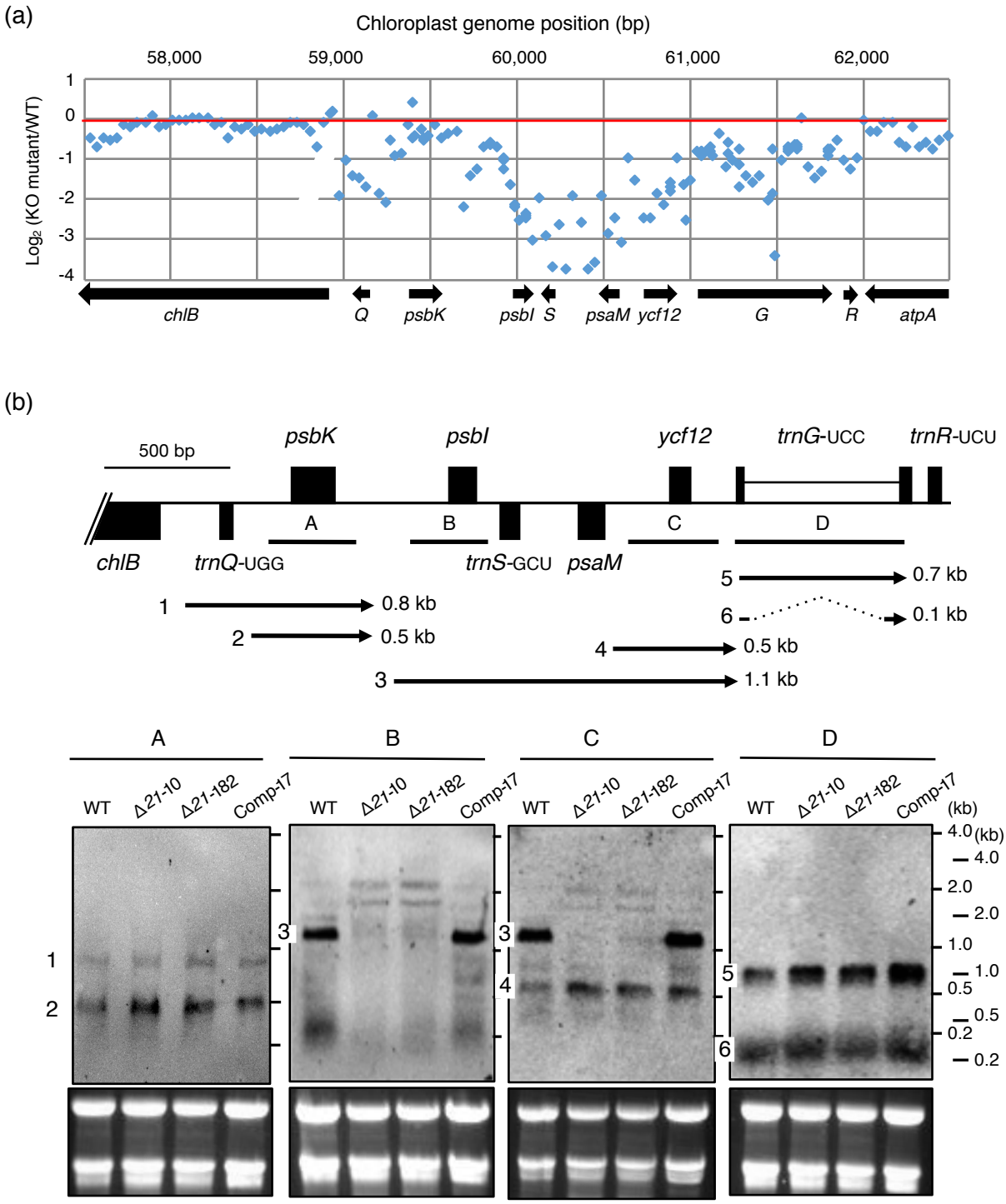

(a)
Chloroplast genome position (bp)
58,000
59,000
60,000
61,000
62,000
Log2 (KO mutant/WT)
1
0
-1
-2
-3
-4
chlB
Q
psbK
psbI
S
psaM
ycf12
G
R
atpA
(b)
500 bp
psbK
psbI
ycf12
trnG-UCC
trnR-UCU
A
B
C
D
chlB
trnQ-UGG
trnS-GCU
psaM
1
0.8 kb
2
0.5 kb
3
1.1 kb
4
0.5 kb
5
0.7 kb
6
0.1 kb
A
B
C
D
WT
Δ21-10
Δ21-182
Comp-17
(kb)
4.0
2.0
1.0
0.5
0.2
1
2
3
4
5
6

Figure 6

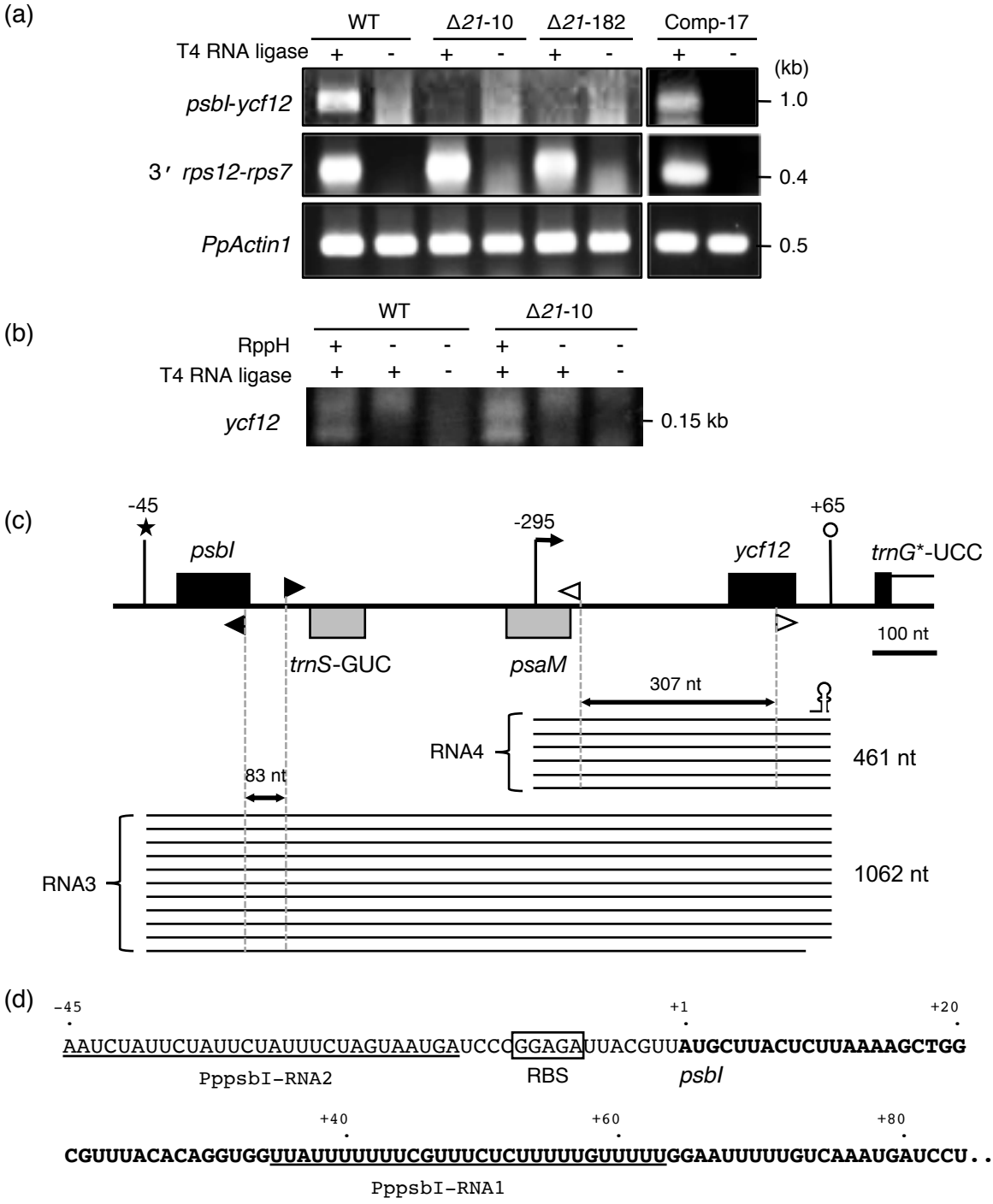
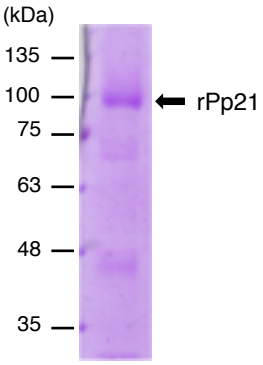

(a)
WT
Δ21-10
Δ21-182
Comp-17
T4 RNA ligase
+ - + - + - + -
(kb)
psbI-ycf12
1.0
3′ rps12-rps7
0.4
PpActin1
0.5
(b)
WT
Δ21-10
RppH
T4 RNA ligase
ycf12
0.15 kb
(c)
-45
-295
+65
psbI
ycf12
trnG*-UCC
100 nt
trnS-GUC
psaM
307 nt
RNA4
461 nt
83 nt
RNA3
1062 nt
(d)
-45
+1
+20
AAUCUAUUCUAUUCUAUUUCUAGUAAUGAUCCCGGAGAUUACGUUAUGCUUACUCUUAAAAGCTGG
PppsbI-RNA2
RBS
psbI
+40
+60
+80
CGUUUACACAGGUGGUUAUUUUUUUCGUUUCUCUUUUUGUUUUUGGAAUUUUUGUCAAAUGAUCCU..
PppsbI-RNA1

Figure 7

(a)



(b)

PppsbI-RNA1 5′–UUAUUUUUUUCGUUUCUCUUUUUGUUUUU–3′

PppsbI-RNA2 5′–AAUCUAUUCUAUUCUAUUUCUAGUAAUGA–3′

AtpsbI-RNA3 5′–AGAGAAUCUAUUCUCUUUUUUUUUAGUAA–3′

PpndhA-RNA4 5′–UUAUUUAUCCGUGGAAUGUUCUUAU–3′

(c)

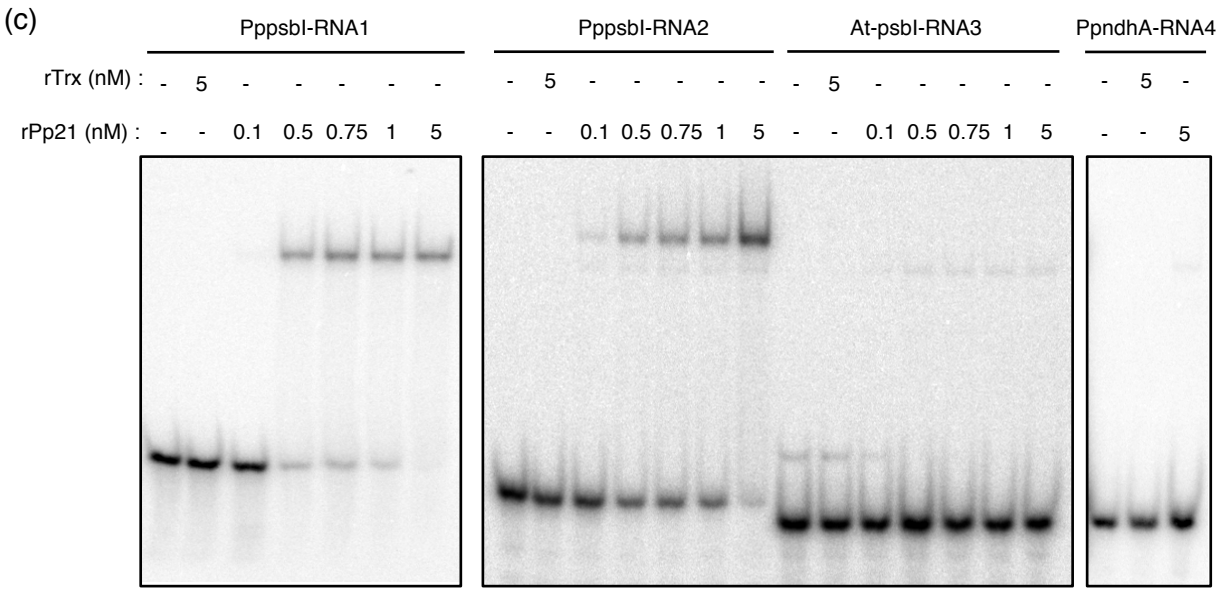
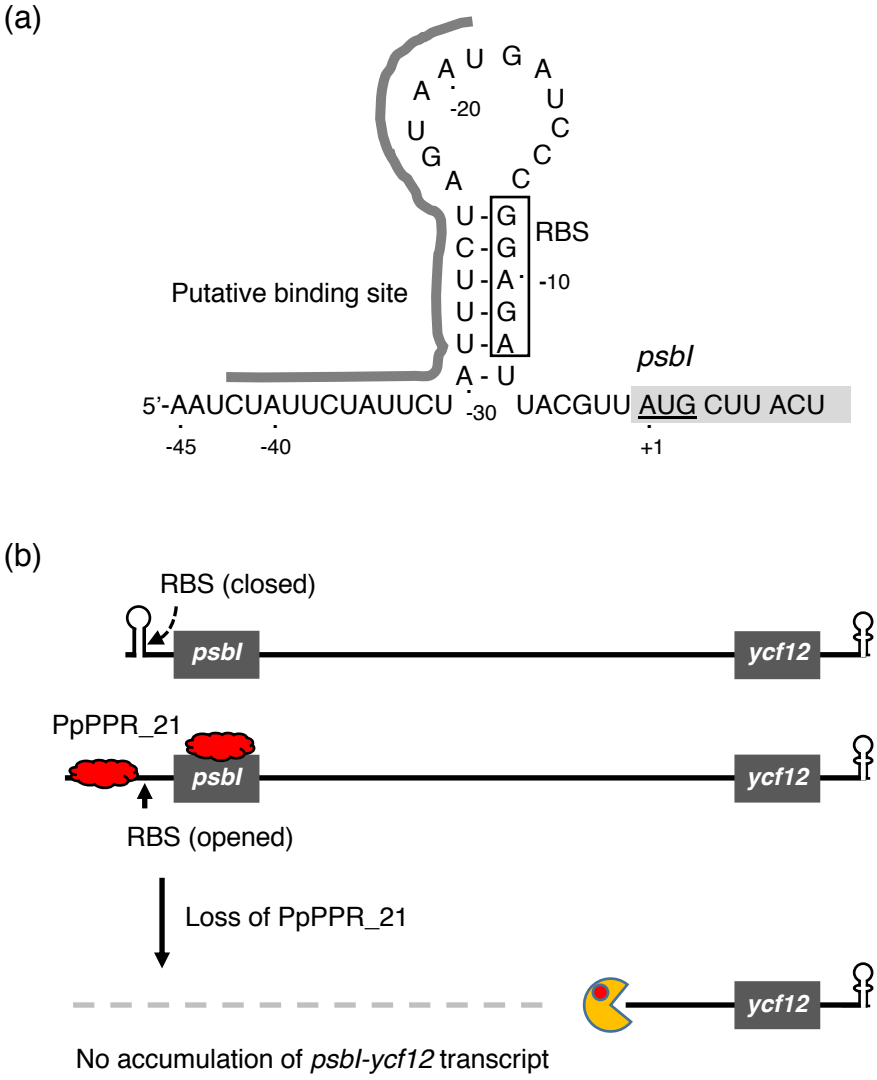


Figure 8


(a)
A U G A
A
-20
U
C
U
G
C
A
C
U - G
C - G
U - A
U - G
U - A
A - U
RBS
-10
Putative binding site
psbI
5'-AAUCUAUUCUAUUCU UACGUU AUG CUU ACU
-45
-40
-30
+1
(b)
RBS (closed)
psbI
ycf12
PpPPR_21
psbI
ycf12
RBS (opened)
Loss of PpPPR_21
ycf12
No accumulation of psbI-ycf12 transcript

(a)

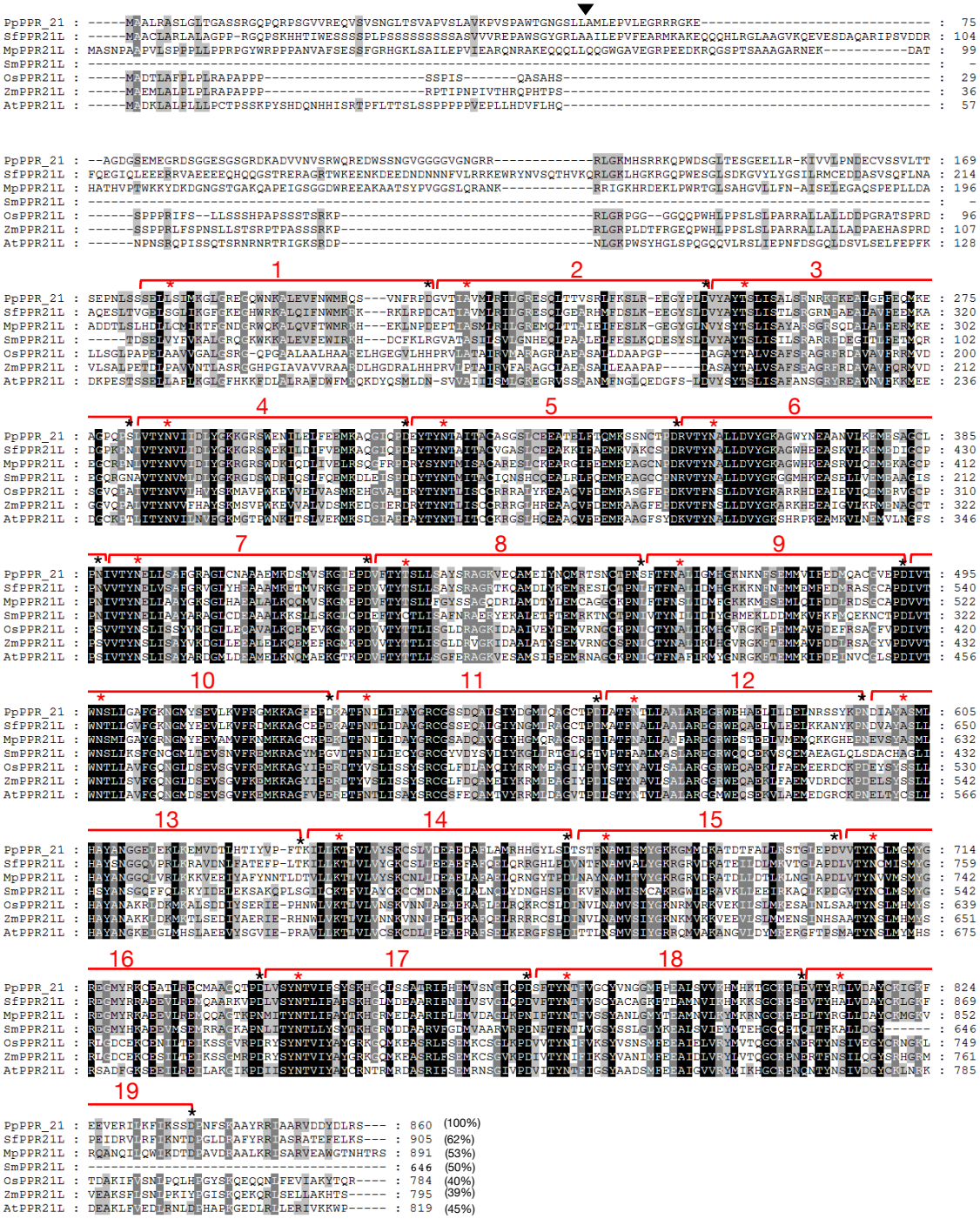
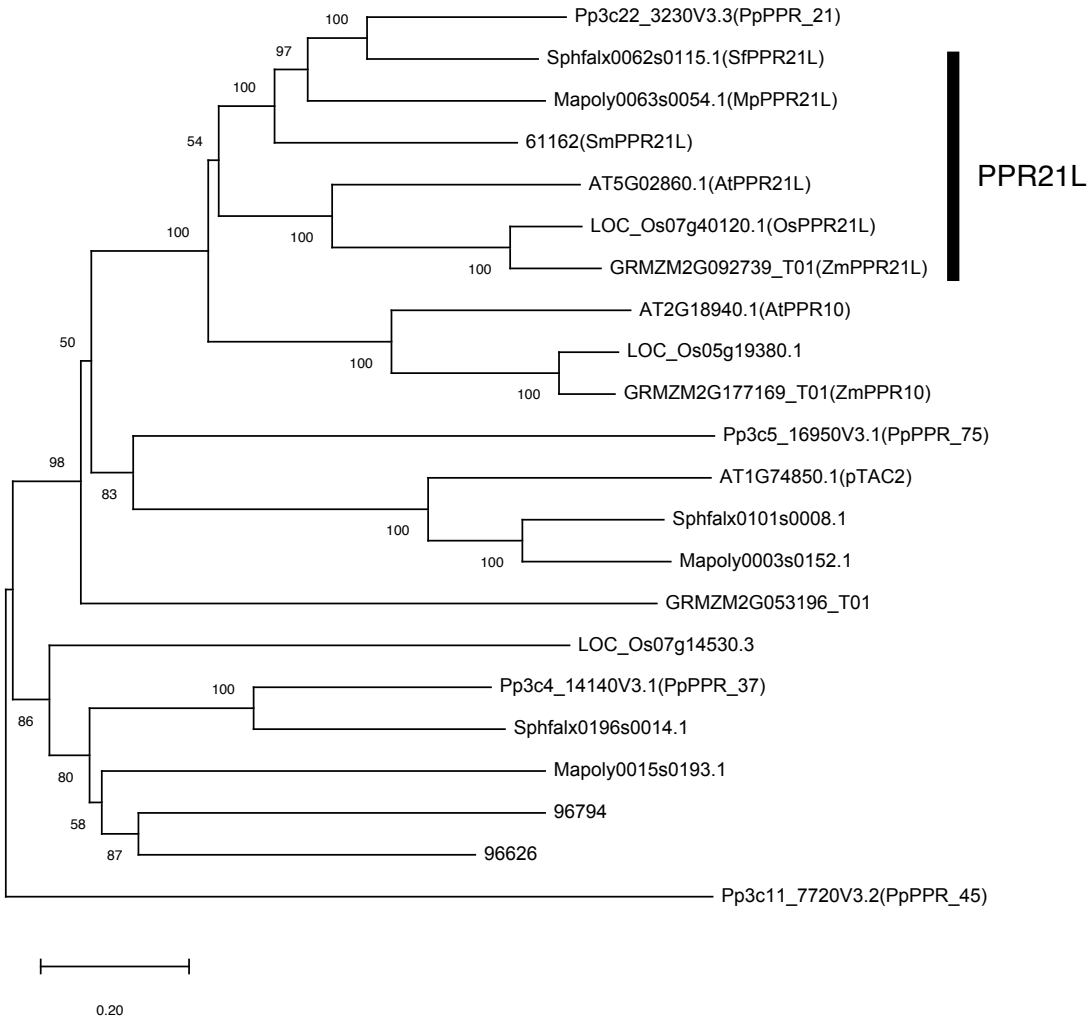
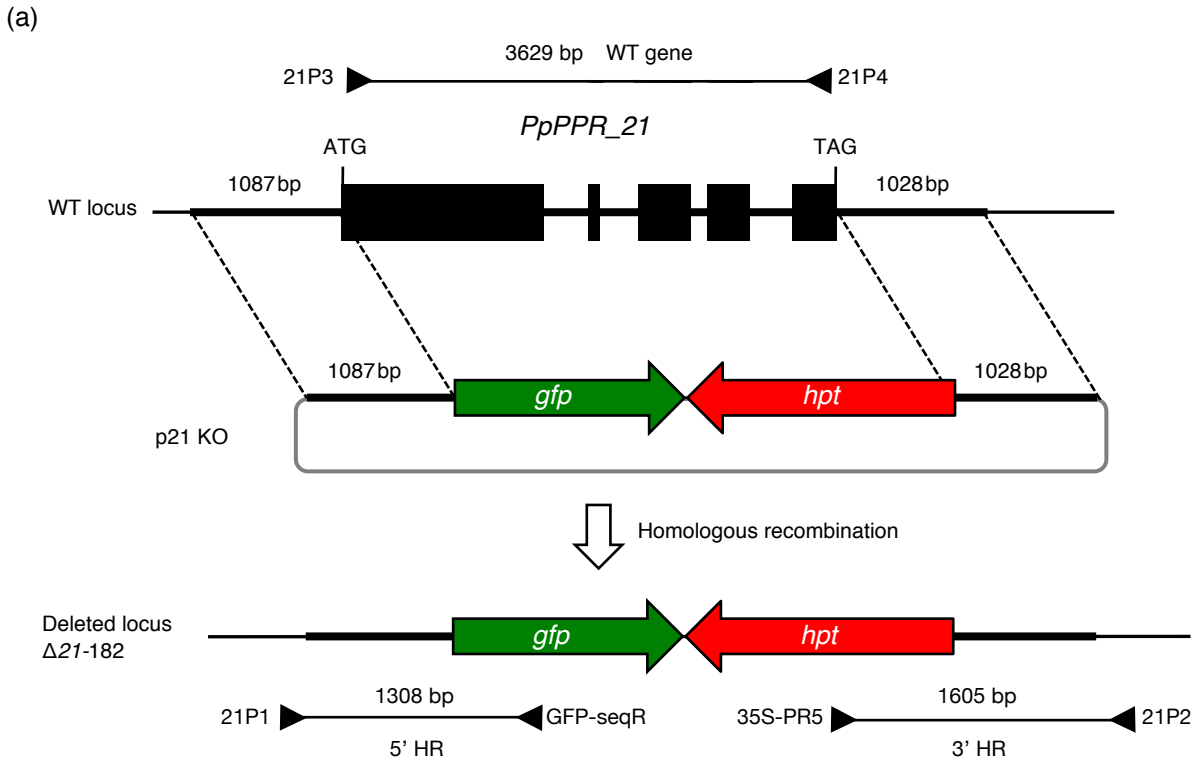


Figure S1. Multiple sequence alignment and phylogenetic tree of PpPPR_21 and its homologs.
(a) Amino acid (aa) sequences were aligned with ClustalW (http://clustalw.ddbj.nig.ac.jp/index.php?lang=en). Identical and conserved aa residues are shaded in black and grey, respectively. PpPPR_21 (Pp3c22_3230V3.3); SfPPR21L, *Sphagnum fallax* PPR_21-like (Sphfalx0062s0115); MpPPR21L, *Marchantia polymorpha* (Mapoly0063s0054.1); SmPPR21L, *Selaginella moellendorffii* (61162); AtPPR21L, *Arabidopsis thaliana* (At5g02860); OsPPR21L, *Oryza sativa* (LOC_Os07g40120); ZmPPR21L, *Zea mays* (GRMZM2G092739_T01), PPR motifs are marked in red brackets 1 to 19. The position 5 and 35 aa residues in each PPR motif are indicated by red and black asterisks, respectively. The arrowhead indicates the predicted cleavage site of the transit peptide.
(b) The phylogenetic tree including PPR21L and two most closely related PPR paralogues within each taxon was constructed using MEGA X program by neighbor-joining method (Saitou and Nei 1987). Bootstrap values from 1000 replicates are indicated at each branch as percentages. Pp3c11_7720V3.2(PpPPR_45) was used as an outgroup.

(b)


Pp3c22_3230V3.3(PpPPR_21)
Sphfalx0062s0115.1(SfPPR21L)
Mapoly0063s0054.1(MpPPR21L)
61162(SmPPR21L)
AT5G02860.1(AtPPR21L)
LOC_Os07g40120.1(OsPPR21L)
GRMZM2G092739_T01(ZmPPR21L)
PPR21L
AT2G18940.1(AtPPR10)
LOC_Os05g19380.1
GRMZM2G177169_T01(ZmPPR10)
Pp3c5_16950V3.1(PpPPR_75)
AT1G74850.1(pTAC2)
Sphfalx0101s0008.1
Mapoly0003s0152.1
GRMZM2G053196_T01
LOC_Os07g14530.3
Pp3c4_14140V3.1(PpPPR_37)
Sphfalx0196s0014.1
Mapoly0015s0193.1
96794
96626
Pp3c11_7720V3.2(PpPPR_45)
100
97
100
54
100
100
100
50
100
100
98
83
100
100
100
86
80
58
87
0.20



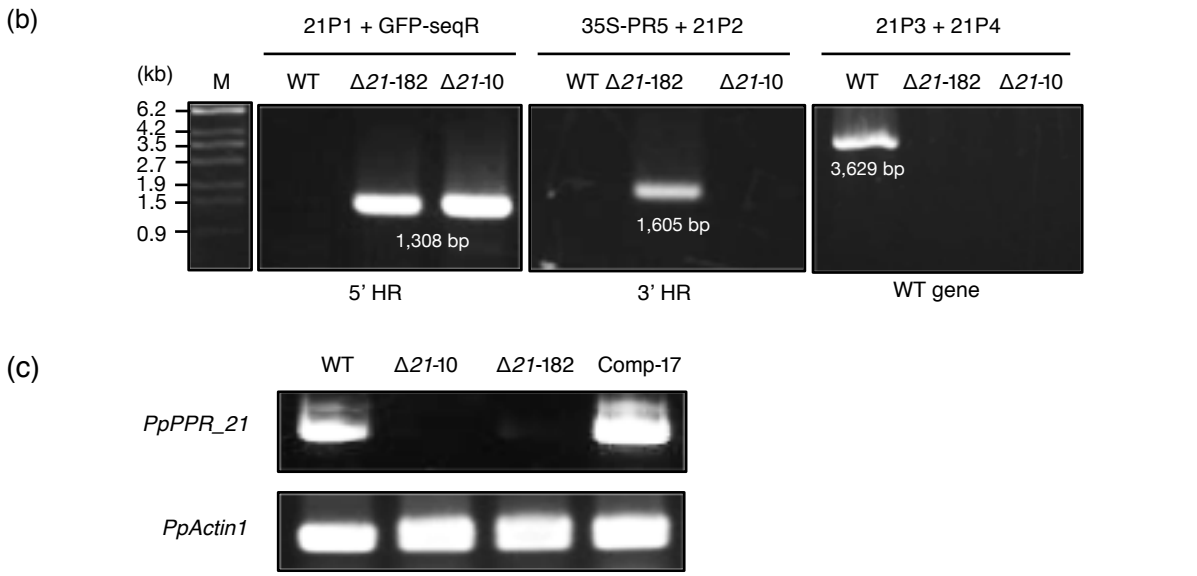


Figure S2. Generation of *PpPPR_21* knockout (KO) mutants.
(a) Structures of wild-type (WT) and the altered genomic locus after replacement of the *gfp-hpt* gene cassette by homologous recombination (HR) are illustrated. Primers and the expected fragment sizes for PCR analysis are also shown. Primer sequences are listed in Table S1. The DNA regions for HR are represented as thick horizontal lines.
(b) Genomic PCR analysis of WT and KO mutants. The predicted 1308- (5′ HR) and 1605-bp (3′ HR) fragments were amplified from the KO lines while the 3629-bp fragment (full length gene) was amplified from WT. DNA size marker is the λ DNA *Sty*I-digest (lanes M).
(c) RT-PCR for detection of *PpPPR_21* transcript in WT, KO and complemented (Comp-17) mosses. *PpActin1* transcript was also amplified as a control.

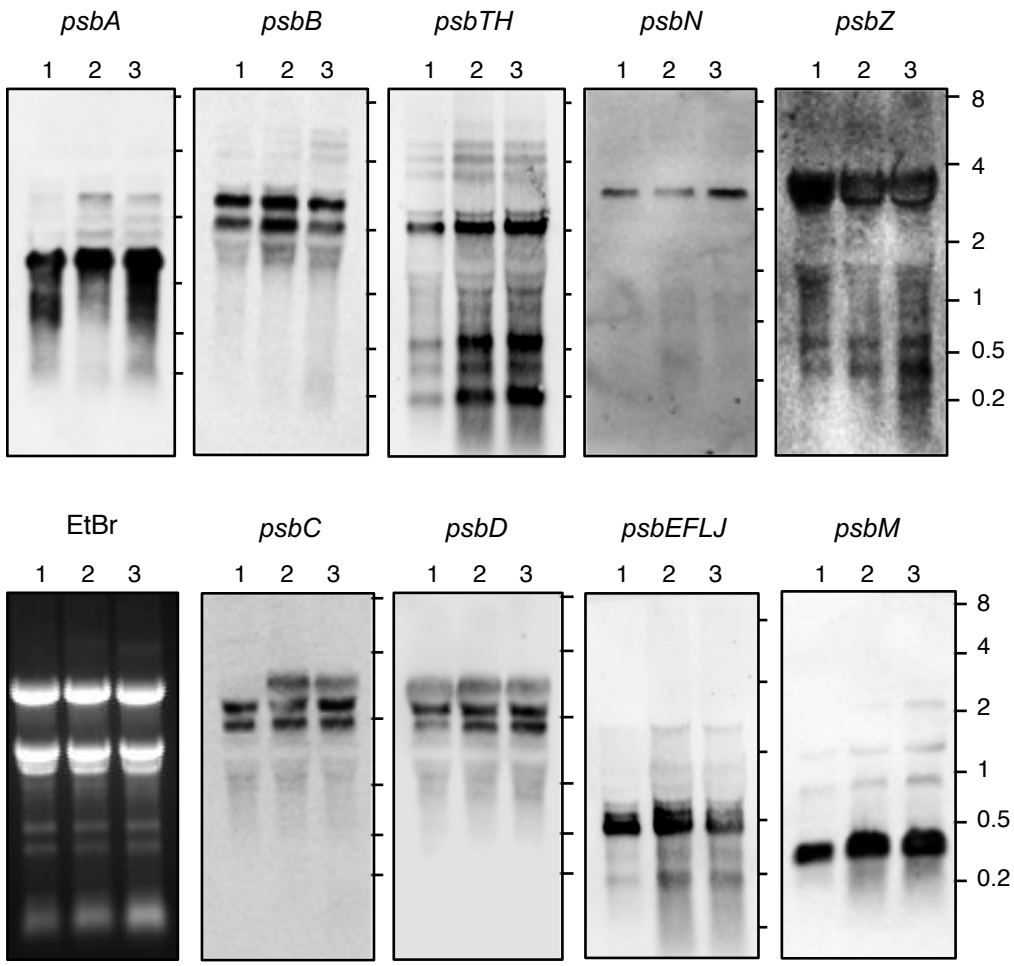


Figure S3. RNA gel blot hybridization of *Physcomitrella patens psb* genes.
Total RNA (15 µg or 10 µg for *psbA*) from *P. patens* wild type (WT) and knockout mutant (Δ*21*-10, Δ*21*-182) mosses was analyzed by northern blot hybridization using gene-specific probes (Table S1). RNA size markers (0.2 to 8 kb) are indicated on the right. Lanes 1, 2 and 3 indicate WT, Δ*21*-10 and Δ*21*-182, respectively.

(a)

PPR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Position 2	S	V	Y	V	Y	V	V	F	F	V	A	A	I	I	S	V	V	F	V
Position 5	L	A	T	N	N	N	N	T	N	N	N	N	A	K	N	N	N	N	R
Position 35 (Last)	D	D	S	D	D	N	D	N	D	D	D	N	T	D	D	D	D	D	D
code	*LD	*AD	*TS	VND	YND	VNN	VND	FTN	FND	VND	AND	*NN	*A*	***	SND	VND	VND	FND	***
A	0.19	0.16	0.31	0.06	0.08	0.2	0.06	0.75	0.23	0.06	0.09	0.15	0.26	0.25	0.09	0.06	0.06	0.23	0.25
C	0.47	0.1	0.24	0.25	0.3	0.53	0.25	0.07	0.19	0.25	0.07	0.45	0.19	0.25	0.18	0.25	0.25	0.19	0.25
G	0.05	0.53	0.17	0.06	0.11	0.04	0.06	0.06	0.1	0.06	0.22	0.14	0.19	0.25	0.27	0.06	0.06	0.1	0.25
U	0.28	0.21	0.28	0.63	0.52	0.23	0.63	0.12	0.49	0.63	0.62	0.27	0.36	0.25	0.45	0.63	0.63	0.49	0.25
Prediction	C	G	A	U	U	C	U	A	U	U	U	C	U	–	U	U	U	U	–

(b)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19			
	C	G	A	U	U	C	U	A	U	U	U	C	U	x	U	U	U	U	x			
position																				position	rank	p-value
59952	C	T	A	T	T	C	T	A	T	T	C	T	A	T	T	T	C	T	A	59970	56	8.77E-05
59957	C	T	A	T	T	C	T	A	T	T	T	C	T	A	G	T	A	A	T	59975	28	4.44E-05
60025	T	G	G	T	T	A	T	T	T	T	T	T	T	C	G	T	T	T	C	60043	52	8.01E-05
60038	C	G	T	T	T	C	T	C	T	T	T	T	T	G	T	T	T	T	T	60056	3	1.98E-06

Figure S4. Prediction of the PpPPR_21-binding site according to its PPR code.
(a) The amino acids at positions 2, 5 and 35 of each PpPPR_21 PPR motif were extracted and are listed from the N- to the C-terminus. The obtained combinations were then used to calculate the probabilities of nucleotide recognition by each individual PPR motif according to the PPR code (Yagi *et al*. 2013).
(b) This predicted target sequence was scanned through the *P. patens* chloroplast genome and four matching sites were found on the *psbI–ycf12* transcript. The numbers indicate the chloroplast genome position (position column).

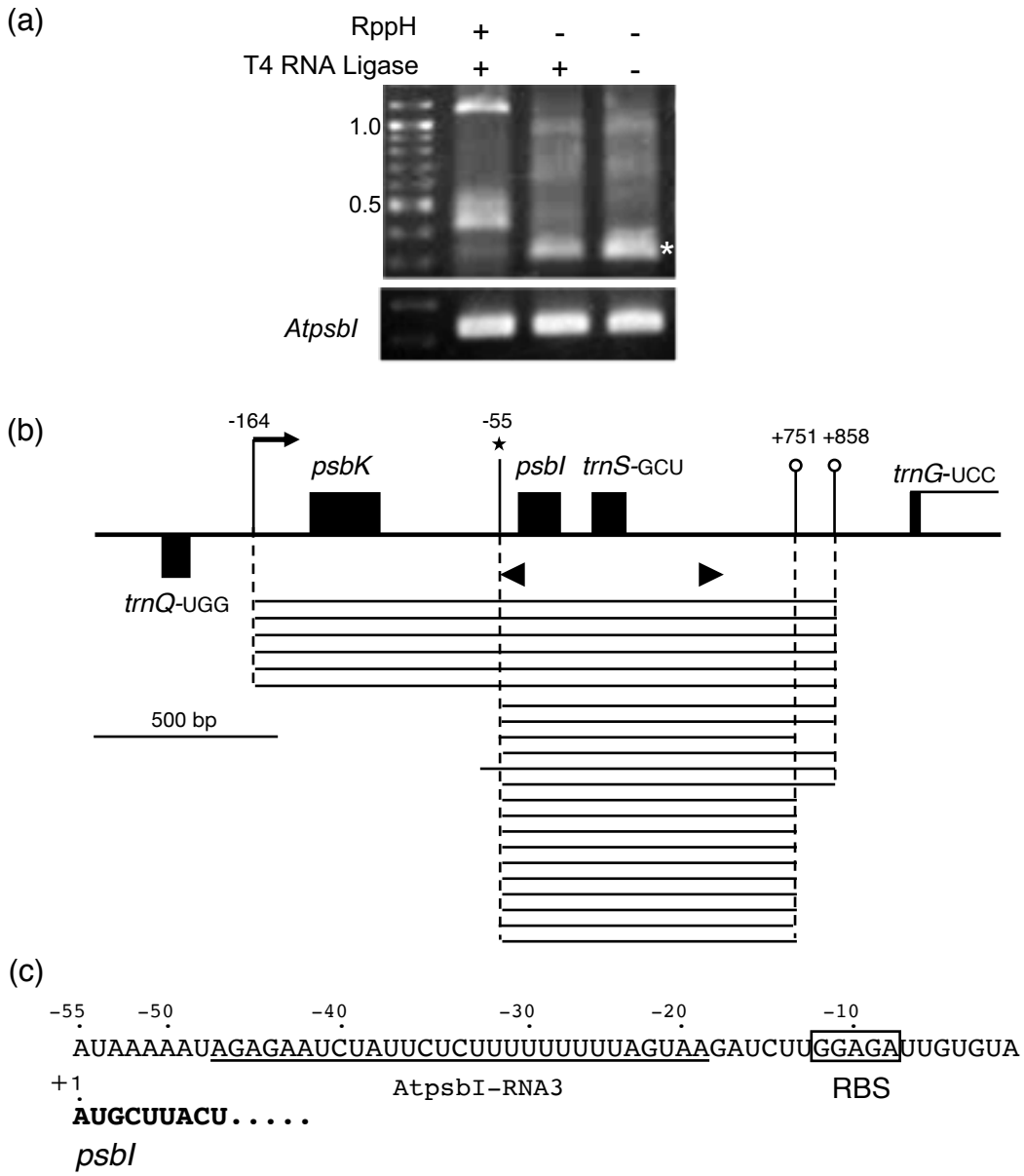


Figure S5. Determination of 5′ and 3′ ends of Arabidopsis *psbK-psbI-trnS* transcript.
(a) Total RNA was treated with (+) or without (-) RNA 5′ pyrophosphohydrolase (RppH) or T4 RNA ligase to form circular RNAs, then subjected to RT-PCR using a set of primer sets (black arrowheads) as indicated in (b). Circular RT (cRT)-PCR products of *psbI-trnC* transcripts were detected on a 2% agarose gels. RppH and T4 RNA ligase treatment did not affect amplification of RT-PCR product as shown in *AtpsbI*. The asterisk indicates non-specific PCR products.
(b) cRT-PCR products were cloned and sequenced. The 5′ ends of *psbI-trnC* transcript were mapped at 55 nucleotides (nt) upstream from the *psbI* translated region and the 3′ ends were mapped at 751 and 858 nt downstream from A of the translation start codon of the *psbI* translated region. The 5′ end of the *psbK-psbI-trnC* transcript was mapped at 164-nt upstream from the *psbK* translated region.
(c) Nucleotide sequence of the 5′-UTR and a part of of the *psbI* translated region. Ribosome-binding site (RBS) is boxed and the *psbI* translated region is bold-faced. The AtpsbI-RNA3 sequence used for REMSA (Fig. 7) is also underlined.

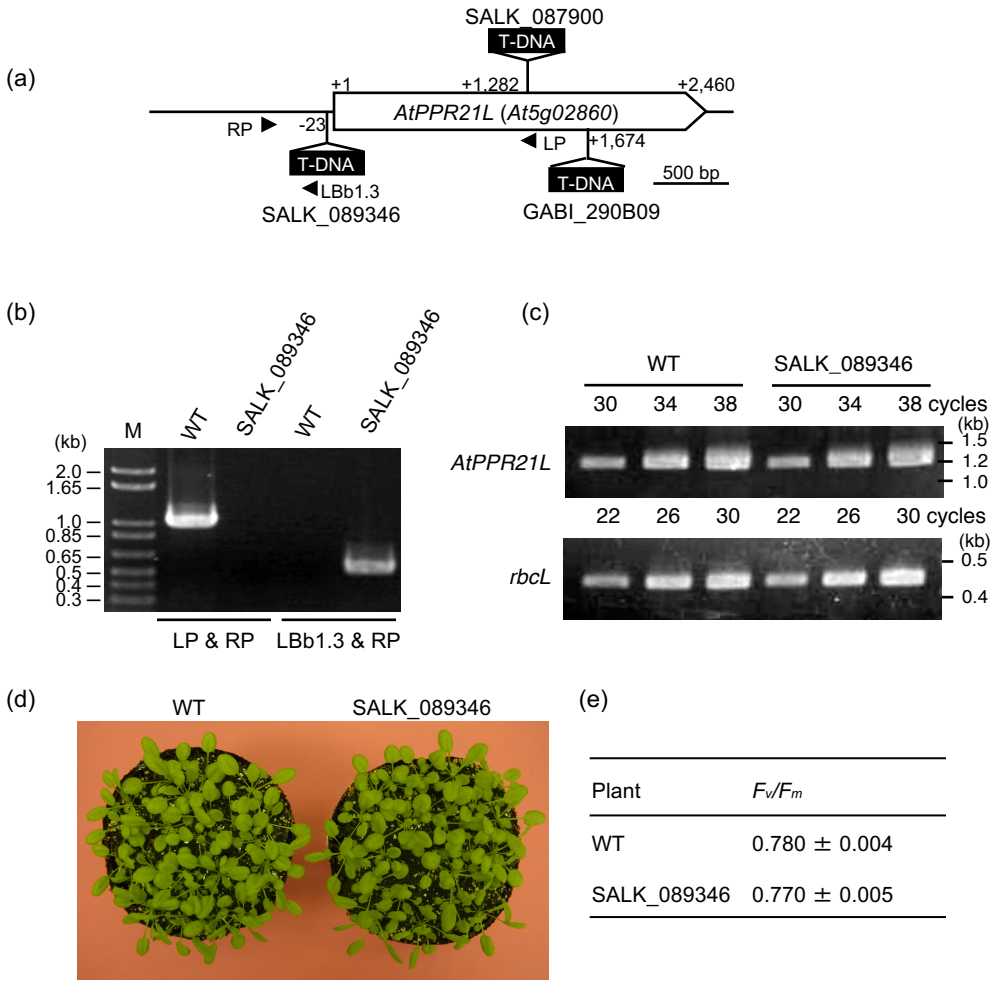


Figure S6. Isolation and characterization of Arabidopsis *PPR21L* T-DNA tagged mutants.
(a) Schematic gene structure of Arabidopsis *PPR21L* (*At5g02860*). This structure is based on representative gene model reported in The Arabidopsis Information Resource (TAIR, https://www.arabidopsis.org/). Positions of T-DNA insertion relative to translational start codon (+1) in GABI_290B09, SALK_087900 and SALK_089346 are indicated. Open box indicates a translated region. Positions of primers used for PCR in (b) are indicated by arrowheads.
(b) Genotyping of Arabidopsis *PPR21L* gene KO mutants. PCR was performed on genomic DNA to detect homozygosity of the T-DNA-tagged line SALK_089346. T-DNA-specific primer LBb1.3 and gene-specific primers (LP and RP) were used for PCR. Primer set used for PCR are indicated in the bottom of the figure. Primer sequences are listed in Table S1. The amplified fragments were separated by agarose gel electrophoresis. Lane M indicates the DNA size marker.
(c) RT-PCR for detection of *AtPPR21L* and *rbcL* transcripts in wild type (WT) and SALK_089346. Lane numbers indicate reaction cycles of PCR.
(d) Growth phenotype of the WT and *Arabidopsis PPR21L* T-DNA insertion line. Plants were grown in a growth chamber at 50 µmol photons m^{-2} s^{-1} under long-day conditions (16-hr light/8-hr dark) for 3 weeks.
(e) Measurement of F_v/F_m in WT and SALK_089346 leaves. Values are means ± SD (n = 12).