AEF1/MPR25 is implicated in RNA editing of plastid *atpF* and mitochondrial *nad5* and also promotes *atpF* splicing in *Arabidopsis* and rice

Aaron Yap¹, Peter Kindgren¹, Catherine Colas des Francs-Small¹, Tomohiko Kazama³, Sandra K. Tanz¹, Kinya Toriyama³, Ian Small^{1,2}

¹Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley 6009 WA, Australia

²Centre of Excellence in Computational Systems Biology, The University of Western Australia, 35 Stirling Highway, Crawley 6009 WA, Australia

³Laboratory of Environmental Biotechnology, Graduate School of Agricultural Science, Tohoku University, Sendai 981-8555, Japan

Corresponding author:

Ian Small ian.small@uwa.edu.au

Australian Research Council Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway Crawley 6009 WA, Australia

Phone: +61 86488 4499

Fax: + 61 86488 4401

Aaron Yap aaron.yap@uwa.edu.au; Peter Kindgren peter.kindgren@uwa.edu.au; Catherine Colas des Francs-Small catherine.colasdesfrancs-small@uwa.edu.au; Tomohiko Kazama tomo-kazama@bios.tohoku.ac.jp; Sandra Tanz sandra.tanz@uwa.edu.au; Kinya Toriyama torikin@bios.tohoku.ac.jp

Running title: Chloroplast RNA editing factor AEF1

Key words: RNA editing, splicing, dual-targeting, chloroplast, mitochondria, *Arabidopsis thaliana*, ATP synthase

Word count: 4607

Summary: 185

Significance statement: 52

Introduction: 746

Results: 929

Discussion: 1019

Materials & Methods: 786

Acknowledgements: 36

Short legends for Supporting Information: 87

References: 1824

Figure Legends: 819

Summary

RNA editing is an essential mechanism that modifies target cytidines to uridine in both mitochondrial and plastid mRNA. Target sites are recognised by pentatricopeptide repeat (PPR) proteins. Using bioinformatics predictions based on the code describing sequence recognition by PPR proteins, we have identified an Arabidopsis editing factor required for the editing of *atpF* in plastids. A loss of function mutation in *ATPF EDITING FACTOR 1* (*AEF1*, *AT3G22150*) results in severe variegation, presumably due to decreased plastid ATP synthase levels. Loss of editing at the *atpF* site is coupled with a large decrease in splicing of the *atpF* transcript, even though the editing site is within an exon and 53 nucleotides distant from the splice site. The rice orthologue of AEF1, MPR25, has been reported to be required for editing of a site in mitochondrial *nad5* transcripts and we confirm that editing of the same site is affected in the Arabidopsis *aef1* mutant. We also show that splicing of chloroplast *atpF* transcripts is affected in the rice *mpr25* mutant. AEF1 is thus highly unusual for an RNA editing specificity factor in having functions in both organelles.

Significance statement

The RNA editing factor AEF1/MPR25 is a sequence-specific RNA binding protein with an unusual multi-functional role in editing and splicing of the chloroplast *atpF* transcript. It is unique (so far) amongst editing specificity factors in also acting in mitochondria, where it recognizes a similar sequence to induce editing of the *nad5* transcript.

Introduction

RNA editing in plants occurs in plastids and mitochondria where specific cytidine nucleotides in mRNAs are deaminated to uridines (Takenaka et al., 2013a). There are 34 major editing sites in *Arabidopsis* plastids (Chateigner-Boutin and Small, 2007) plus at least 9 minor sites (Ruwe et al., 2013), and over 600 editing sites in *Arabidopsis* mitochondria (Bentolila et al., 2013). Editing of the RNA sequence usually alters codons in such a way as to correct the implied amino acid sequence towards one that is evolutionarily conserved (Jiang et al., 2012). Null mutants exhibiting defects in plastid RNA editing at specific sites often have deleterious phenotypes, such as *clb19* (defective in the editing of *clpP* and *rpoA* (Chateigner-Boutin et al., 2008)) and *ysl* (which does not edit *rpoB* (Zhou et al., 2009)), showing the importance of the editing process for normal growth and development.

So far, 18 plastid editing specificity factors have been reported, recognising 25 of the 34 major editing sites (Chateigner-Boutin et al., 2008; Hammani et al., 2009; Okuda et al., 2010; Yagi et al., 2013a; Zhou et al., 2009; Okuda et al., 2009; Kotera et al., 2005; Robbins et al., 2009; Cai et al., 2009; Okuda et al., 2007; Hayes et al., 2013). All of them are pentatricopeptide repeat (PPR) proteins, consisting of tandem repeats of ~35 residues (Small and Peeters, 2000), each of which folds into 2 anti-parallel alpha helices and together forming an alpha solenoid superhelix (Fujii et al., 2011; Yin et al., 2013). PPR proteins bind RNA (Barkan and Small, 2014) and can be divided into 2 subclasses: the P subclass consisting only of P motifs and the PLS subclass consisting of P, L, and S motifs (Lurin et al., 2004). PLS-PPR proteins are found predominantly in land plants where they can be extremely numerous (Fujii and Small, 2011); they include all known editing specificity factors (Takenaka et al., 2013a).

PLS-PPR proteins can be further divided into 3 subgroups based on the extra C-terminal domains that are usually present. All PLS-PPR RNA editing factors contain an E domain at or near the C-terminus, and many contain (in addition) a C-terminal DYW domain named for the characteristic Asp-Tyr-Trp C-terminal triplet (Lurin et al., 2004). Although all editing specificity factors are PLS-PPR proteins, the converse is not true. A few examples of PLS-PPR proteins with other roles have been described, such as CRR2 (implicated in intercistronic processing (Hashimoto et al., 2003; Okuda et al., 2009)), OTP70 (Chateigner-Boutin et al., 2011) and PpPPR_43 (Ichinose et al., 2012) (both implicated in intron splicing). These alternative roles still require site-specific RNA recognition, and recent breakthroughs have given insights into how the remarkable specificity of PPR proteins is achieved.

Modelling based on patterns of sequence variation suggested a 1 motif : 1 nucleotide correspondence in PPR:RNA complexes (Fujii et al., 2011) and further work showed strong correlations between the amino acid residues at positions 6 and 1' of each repeat and the nature of the aligned nucleotide (Barkan et al., 2012; Yagi et al., 2013b). This led to a proposed 'code' by which PPR proteins might recognise their target sequences (Barkan et al., 2012; Takenaka et al., 2013b; Yagi et al., 2013b). A crystal structure of the P-class PPR protein PPR10 has confirmed most of these hypotheses (Yin et al., 2013). From the proposed alignments, PLS-class editing factors are expected to align 5' of the cytidine to be edited with the terminal S motif aligned with the nucleotide at -4 with respect to the editing site (Barkan et al., 2012; Takenaka et al., 2013b; Yagi et al., 2013b). Using this information, new editing factors have been predicted and verified experimentally (Takenaka et al., 2013b; Yagi et al., 2013b), raising the possibility of being able to predict all of the remaining uncharacterised editing factors in plants. Progressing towards this goal, we report the discovery and characterization of an Arabidopsis editing factor required for the editing of a site in plastid *atpF* transcripts (at genome position 12707). This is the only site in *atpF* that is edited in Arabidopsis. Unexpectedly, the factor is an apparent orthologue of the rice editing factor MPR25, reported to be required for editing

a site in mitochondrial *nad5* transcripts (Toda et al., 2012). We show that the Arabidopsis and rice editing factors have conserved functions in both *atpF* and *nad5* transcript processing in chloroplasts and mitochondria.

Results

Prediction of an atpF recognition factor

Alignments of putative editing factors (i.e. all 193 Arabidopsis PLS-PPR proteins containing E domains) against the *atpF* editing site were scored using a table of log likelihood ratios (Table S1) derived from observed frequencies of association between amino acid combinations at position 6 and 1' and the four RNA nucleotides (Barkan et al., 2012). PLS-PPR proteins were aligned such that the terminal S motif aligned with the nucleotide at -4 with respect to the edited C at genome position 12707. The score for each protein was calculated by adding the score for each aligned motif:nucleotide. The top score (3.32) was seen for the alignment to At3g22150 (Figure 1). The next highest score was ~30% lower (2.35 for At2g34400). The distribution of possible scores for each protein was obtained by aligning each PLS-PPR sequence to all other possible sequences from the Arabidopsis chloroplast genome and calculating each respective alignment score. The score distributions were found to be approximately Gaussian; the score for At3g22150 was 2.82 standard deviations above the mean score obtained on the rest of the genome.

Characterization of the aef1-1 mutant

We isolated an insertion mutant line, SALK_041683, hereafter called *atpF editing factor1-1* (*aef1-1*). *aef1-1* displayed variegation, retarded growth and pale leaves (Figure 2). The insertion was mapped to the coding sequence of the gene At3g22150 (Figure 2b) at nucleotide 252 from the start codon, within the first PPR motif. The *aef1-1* mutant was complemented with the *AEF1* gene

including its promoter. The complemented lines recovered the wild type phenotype (Figure 2), lacking the strong variegated phenotype as well as all the molecular defects discussed below.

AEF1 is targeted to plastids

The full-length *AEF1* coding sequence was fused before a GFP open reading frame and transiently expressed in Arabidopsis suspension culture cells. The fusion protein accumulated specifically in plastids as shown by the co-location with a Rubisco small subunit control construct (Figure 3A-C). No targeting to mitochondria was observed under these conditions as gauged by a lack of co-location with a *cox4*-RFP marker (Figure 3D-F).

AEF1 is required for editing and efficient splicing of atpF transcripts

Total RNA was isolated from *aef1-1* mutants and converted to cDNA and analysed for loss of editing at the *atpF* site. Using poisoned primer extension (PPE) we found that *aef1-1* mutants are completely defective in the editing of *atpF* (Figure 4). Editing was restored to wild type levels in the complemented *aef1-1* plants expressing AEF1. High Resolution Melting (HRM) analysis of the other major editing sites in plastids revealed that *aef1-1* is specifically defective in the editing of the *atpF* site; all other sites are edited normally (Figure S1).

When amplifying the sequence surrounding the *atpF* editing site from *aef1-1* plants, we discovered that the mutant does not splice the *atpF* transcript efficiently. To investigate the extent of the splicing defect, we quantified spliced and unspliced transcripts by qPCR (Figure 5 and Figure S2), comparing *aef1-1* to *crs1-1*, a known *atpF* splicing mutant (Jenkins et al., 1997; Till et al., 2001). The *aef1-1* mutant shows a strong *atpF* splicing defect, but not as strong as the defect in *crs1-1*. The *atpF* editing defect in *aef1-1* is not a secondary effect of the splicing defect as *atpF* is still edited in *crs1-1*, albeit at a reduced level compared to wild type (Figure S3).

In addition to the defect in processing of *atpF* in *aef1-1*, we also found a minor apparent defect in the splicing of *ndhB* that is not seen in *crs1-1* (Fig. 5). However, although *aef1-1* contains

significantly decreased levels of spliced *ndhB* transcripts (Figure S2), there is no accumulation of the unspliced variant, suggesting that the defect is in stabilisation of the spliced transcript rather than the splicing process itself.

AEF1 is required for normal accumulation of AtpF

Antibodies against AtpF, ATPG and AtpA were used for western blot analysis of plastid ATP synthase levels in *aef1-1*. Neither *aef1-1* nor *crs1-1* accumulate AtpF to a detectable level (Figure 6). Accumulation of the plastid-encoded subunit AtpA and nucleus-encoded subunit ATPG is less affected in both mutants, confirming that the primary defect is likely to be in AtpF synthesis.

AEF1 is an orthologue of MPR25

AEF1 has previously been suggested to be an orthologue of the rice protein MPR25 (Toda et al., 2012). In rice, MPR25 is an editing factor required for the proper editing of the mitochondrial site *nad5-1580*, which is also edited in *Arabidopsis*. PPE analysis of the *nad5* site showed that the editing at this site is reduced by about 50% in *aef1-1* compared to wild type (Figure 7). Normal levels of editing are restored in the complemented plants. Two other editing sites within *nad5* are not affected in the mutant suggesting that this editing defect is specific (Figure S4). The *nad5* sites in *Arabidopsis* and rice are excellent matches for AEF1 and MPR25 (Figure 7).

Although rice does not require editing of *atpF* (there is a T at the site corresponding to the edited C in the rice *atpF* gene), the *atpF* transcript contains a sequence that aligns to MPR25 (Figure 7). As the rice *mpr25* mutant has a physiological phenotype suggesting a defect in photosynthesis, we checked *atpF* splicing in the mutant and found it was significantly retarded, with an over-accumulation of unspliced products and little accumulation of spliced transcripts (Figure 8).

Discussion

Correct production of AtpF is required for the assembly of the ATP synthase complex and thus production of ATP within the plastid. AtpF is an essential part of the peripheral stator linking the F1 and F0 complexes together with ATPG, the product of the nuclear gene *ATPG/PDE334* (At4g32260) (Kong et al., 2013). Translation of functional AtpF requires editing and splicing of the plastid *atpF* transcript. Editing of the site at position 92 alters codon 31 from CCA (proline) to CUA (leucine); leucine is evolutionarily conserved at this position. *AtpF* also contains a single group II intron that requires a specific splicing factor (CRS1) for processing (Jenkins et al., 1997; Till et al., 2001). Splicing of *atpF* (at least in maize chloroplasts) also requires several more protein factors including WHY1 (Prikryl et al., 2008), RNC1 (Watkins et al., 2007), WTF1 (Kroeger et al., 2009), and probably the plastid-encoded maturase MatK (Zoschke et al., 2010).

The *atpF* intron is widely conserved across angiosperms, but the editing site is less so. Monocots such as rice and maize have a genome-encoded T at the editing site and thus do not require RNA editing of the *atpF* transcript. However, phylogenetic analysis within the Malphigiales has shown that there is a strong association between C-to-T substitutions at position 92 and the loss of the *atpF* intron (Daniell et al., 2008). Reverse transcription of edited and spliced transcripts could be one explanation for these correlated losses of both *atpF* editing and splicing (Daniell et al., 2008), but the mechanistic link involving AEF1 suggests that an additional factor may be reduced selection pressure for retention of AEF1 once either the intron or the editing site has been lost. We cannot tell from our data whether the lack of editing or the defect in splicing is primarily responsible for the striking phenotypes shown by the *aef1-1* plants. The mutant phenotype of *crs1* mutants is considerably more severe, so we conclude that the low levels of spliced but unedited transcripts can produce some AtpF, indicating that the protein with proline 92 is at least partially functional.

An interesting question is whether it is the alteration of sequence induced by RNA editing that facilitates *atpF* splicing or the binding of AEF1. Several examples of intron splicing dependent on RNA editing events have been reported in which it seems likely that it is the sequence alteration that

is crucial (Borner et al., 1995; Castandet et al., 2010; Farre et al., 2012). In these cases, the sequence alterations are at key sites within the introns, whereas in the case of AEF1, the editing site is in the first exon, over 50 nucleotides from the splice site. Thus we suspect that it is the binding of AEF1 upstream of the editing site that influences splicing through effects on RNA folding or accessibility rather than the editing event itself. This interpretation is strongly supported by the data from rice, where *atpF* splicing is also severely affected in the absence of MPR25 even though there is no difference in sequence of the *atpF* transcript. Many other non-editing PPR proteins are known to influence splicing, although the mechanisms by which they act are unclear (Falcon de Longevialle et al., 2010).

AEF1 is clearly an orthologue of MPR25 in rice (O'Toole et al., 2008), originally reported as a mitochondrion-specific PPR protein required for the editing of *nad5-1580* (Toda et al., 2012). We found that *aef1-1* has a defect in editing of the orthologous editing site in *Arabidopsis* mitochondria, which is restored in the complemented line. The predicted binding sites for AEF1 and MPR25 in *atpF* and *nad5* are similar, consistent with this editing factor having a dual function and being present in both organelles. Arguing against dual-targeting of AEF1 and MPR25, we saw no AEF1-GFP signal in mitochondria and Toda et al. did not see MPR25_FLAG in plastids (Toda et al., 2012). We cannot therefore entirely rule out some strange two-way indirect connection between *atpF* processing in plastids and *nad5* processing in mitochondria, but we favour the interpretation that the protein localisation experiments are giving incomplete results. GFP fusions do not always reliably report both locations of dual-targeted proteins (Duchêne et al., 2005), so our results with AEF1-GFP are not unprecedented. These targeting questions could be tested by attempting to complement *aef1* with AEF1-GFP fusions containing various targeting sequences of known specificity. Finally, an explanation is needed for why the *aef1-1* mutant only shows a partial *nad5* editing defect. We suggest that the T-DNA insertion in *aef1-1* affects plastid function of AEF1 more severely than it does the putative mitochondrial function; the insertion site is near the start of the coding sequence

and would be expected to remove the plastid targeting sequence but not the RNA binding and editing functions, and possibly not mitochondrial targeting. Our attempts to express the truncated protein as a GFP fusion were unsuccessful.

The retention of editing factors for long evolutionary periods despite the apparent ease with which the need for them could be eliminated (by reversion of the edited C to a T in the genome) has puzzled researchers for some time. Multi-functionality might provide an explanation for unusually conserved editing factors. AEF1/MPR25 provides an interesting example in which the same factor provides editing or splicing functions to both organelles and has presumably done so since before the monocot/dicot split.

Until recently, the discovery of new RNA editing factors has relied on extensive forward (Takenaka et al., 2010; Takenaka and Brennicke, 2012) or reverse genetic screens (Hammani et al., 2009). The ‘PPR code’ (Barkan et al., 2012) allows the prediction of which factors bind which sites, and the accuracy of these predictions is illustrated by several recent reports (Takenaka et al., 2013b; Yagi et al., 2013a; Yagi et al., 2013b), including this work. This will accelerate the discovery of the remaining editing factors and supports attempts to design bespoke RNA binding factors based on the same understanding of RNA recognition by PPR proteins (Yagi et al., 2014).

Materials and methods

Plant materials and growth conditions

Arabidopsis thaliana Col-0 was used as a control. Mutant *crs1-1* seeds (At5g16180, SALK_026861) were selected from heterozygous populations as previously reported (Asakura and Barkan, 2006). The SALK_041683 line (*aef1-1*, At3g22150) was selected for homozygosity using LP 5’- TCACTCTGACAAAATCAAACGC-3’, RP 5’- TCATCATTATCCCAAATTGCC-3’ and Lba1. All plants were ethanol-sterilized and sown on ½ Gamborg (Duchefa Biochemie) plus 1% sucrose

plates. Plants were then stratified at 4 °C for 2 days and grown under long (16 hr light cycle) and short (8 hr light cycle) conditions at 40 µE, 22 °C with 60% humidity. Aerial tissue was harvested at 14 days or equivalent biological age in the case of mutants exhibiting growth retardation. For complementation, AEF1 including the endogenous promoter was cloned from genomic DNA (Col-0) using F 5'-

GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATTAATTATGCAATATTTA-3' and R 5'- GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTCATCCAATTCCAAACTT-3'. The amplified fragment was cloned into pDONR207 using BP clonase II (Invitrogen). Following sequencing, the AEF1 fragment was sub-cloned into the binary pGWB1 vector (Nakagawa et al., 2007). Complemented lines were harvested 14 days post hygromycin selection (Harrison et al., 2006).

The rice mutant *mpr25* contains a *Tos17* retrotransposon insertion in LOC_Os04g51350 (Toda et al., 2012). The *mpr25* mutant was complemented with the full-length *MPR25* ORF driven by the maize ubiquitin promoter (Toda et al., 2012). *Oryza sativa* L. cv Nipponbare was used as a control for rice experiments. Rice plants were grown under short day conditions (8 hr light cycle) at 100 µmol/sec/m². The third leaf was harvested for RNA extraction.

Molecular material

Total RNA was extracted from Arabidopsis tissues using the Qiagen RNeasy Plant kit following the manufacturer's protocol, quantified using a Nanodrop ND-1000 spectrophotometer and visualised on a 2% agarose gel to ensure integrity. The Ambion Turbo DNase kit was used to remove DNA contamination according to the manufacturer's specifications followed by ethanol precipitation. Random-primed cDNA was generated using Invitrogen SuperScriptTM III Reverse Transcriptase as per the manufacturer's recommendations. Total protein was extracted from 50 mg tissue using standard acetone precipitation.

Molecular phenotype analysis

Poisoned primer extension was done as initially described by Peeters and Hanson (2002), with modifications as published by Chateigner-Boutin and Small (2007). High-resolution melting was carried out according to the method detailed in Chateigner-Boutin and Small (2007). Initial analysis of *nad5* editing used Sanger sequencing by Macrogen Inc. Quantitative PCR of chloroplast splicing products was carried out using the Roche LightCycler® 480 Instrument as previously described (Falcon de Longevialle et al., 2008). For western analysis, AtpF antibody was obtained from Agrisera (AS10 1604), TUB α was obtained from Sigma (Sigma T5168), ATPG and AtpA antibodies were gifts from Professor Alice Barkan's laboratory (McCormac and Barkan, 1999). For northern analysis, three μ g of RNA was fractionated in denaturing formaldehyde gels, transferred to a nylon membrane and hybridized with an *atpF*-specific probe, which was PCR-labelled with digoxigenin (Roche) using primers 5'- TGGGAGTTTCGGGCTTAATA-3' and 5'- CATTCCATGGCCCCGAGAAT -3'.

Subcellular localisation of AEF1 by GFP tagging

The full-length coding sequence of *AEF1* was amplified using the Expand High Fidelity PCR system (Roche Diagnostics, Germany) and the primers 5'- GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATTAATTATGCAATATTTA-3' and 5'- GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACTCATCCAATTCCAAACTT-3'. The PCR products were cloned into the Gateway® vector pDONR207 (Invitrogen, USA) and sequenced. The entry vector and a Gateway® cloning cassette (Carrie et al., 2009) were recombined to clone *AEF1* in frame with the coding region of the green fluorescent protein (GFP) under control of the viral 35S promoter. For co-localization studies, the small subunit of Arabidopsis ribulose biphosphate carboxylase fused to the red fluorescent protein (RFP) was used as a plastid control and the mitochondrial targeting sequence of yeast Cox4 fused to mCherry in pBIN20 (Nelson et al.,

2007) was used as a mitochondrial control. The fusion constructs were biolistically transformed into cultured *Arabidopsis* cells. Five µg of each of the GFP and RFP/mCherry plasmids were co-precipitated onto 1 µm gold particles and transformed using the biolistic PDS-1000/He system (Bio-Rad). Particles were bombarded onto 2 mL of *Arabidopsis* cultured cells resting on filter paper on osmoticum plates (2.17 g/L Murashige & Skoog Modified Basal Salt Mixture, 30 g/L sucrose, 0.5 mg/L naphthalene acetic acid, 0.05 mg/L kinetin, 36.44 g/L mannitol). After bombardment, the *Arabidopsis* cells were placed in the dark at 22 °C. Fluorescence images were obtained 24 h after transformation using an Olympus BX61 epifluorescence microscope with excitation wavelengths of 460/480 nm (GFP) and 535/555 nm (RFP and mCherry), and emission wavelengths of 495–540 nm (GFP) and 570–625 nm (RFP and mCherry). Subsequent images were captured using Cell® imaging software.

Acknowledgements

We are grateful for the financial support for this work by the Australian Research Council (DP120102870 and CE140100008 to IS; DE120100307 to SKT). PK was supported through a post-doctoral fellowship from the Wenner-Gren foundation.

Supporting Information

Table S1. Log likelihood ratios derived from observed frequencies of association between amino acid combinations at position 6 and 1' and the four RNA nucleotides.

Figure S1. High Resolution Melting (HRM) screen of known editing sites in *aefl-1* and *crs1-1*.

Figure S2. Accumulation of spliced and unspliced transcripts in *aefl-1* and *crs1-1* mutants.

Figure S3. Quantification of *atpF* editing in *aefl-1* in comparison to *crs1-1*.

Figure S4. Quantification of editing in *aei1-1* and Col-0 at multiple sites in *nad5*.

References

- Asakura, Y. and Barkan, A.** (2006) Arabidopsis orthologs of maize chloroplast splicing factors promote splicing of orthologous and species-specific group II introns. *Plant Physiol* **142**, 1656–1663.
- 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.
- Bentolila, S., Oh, J., Hanson, M.R. and Bukowski, R.** (2013) Comprehensive high-resolution analysis of the role of an *Arabidopsis* gene family in RNA editing. *PLoS Genet* **9**, e1003584.
- Borner, G.V., Morl, M., Wissinger, B., Brennicke, A. and Schmelzer, C.** (1995) RNA editing of a group II intron in *Oenothera* as a prerequisite for splicing. *Mol Gen Genet* **246**, 739–744.
- Cai, W., Ji, D., Peng, L., Guo, J., Ma, J., Zou, M., Lu, C. and Zhang, L.** (2009) LPA66 is required for editing *psbF* chloroplast transcripts in *Arabidopsis*. *Plant Physiol* **150**, 1260–1271.
- Carrie, C., Kuhn, K., Murcha, M.W., Duncan, O., Small, I.D., O'Toole, N. and Whelan, J.** (2009) Approaches to defining dual-targeted proteins in *Arabidopsis*. *Plant J* **57**, 1128–1139.
- Castandet, B., Choury, D., Begu, D., Jordana, X. and Araya, A.** (2010) Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res* **38**, 7112–7121.
- Chateigner-Boutin, A.L., Colas des Francs-Small, C., Delannoy, E., Kahlau, S., Tanz, S.K., Falcon de Longevialle, A., 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.

- Chateigner-Boutin, A.L., Ramos-Vega, M., Guevara-Garcia, A. et al.** (2008) CLB19, a pentatricopeptide repeat protein required for editing of *rpoA* and *clpP* chloroplast transcripts. *Plant J* **56**, 590–602.
- Chateigner-Boutin, A.L. and Small, I.** (2007) A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. *Nucleic Acids Res* **35**, e114.
- Daniell, H., Wurdack, K.J., Kanagaraj, A., Lee, S.B., Saski, C. and Jansen, R.K.** (2008) The complete nucleotide sequence of the cassava (*Manihot esculenta*) chloroplast genome and the evolution of *atpF* in Malpighiales: RNA editing and multiple losses of a group II intron. *Theor Appl Genet* **116**, 723–737.
- Duchêne, A.M., Giritch, A., Hoffmann, B., Cognat, V., Lancelin, D., Peeters, N.M., Zaepfel, M., Maréchal-Drouard, L. and Small, I.D.** (2005) Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **102**, 16484–16489.
- 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.
- Falcon de Longevialle, A., Small, I.D. and Lurin, C.** (2010) Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol Plant* **3**, 691–705.
- Farre, J.C., Akinin, C., Araya, A. and Castandet, B.** (2012) RNA editing in mitochondrial *trans*-introns is required for splicing. *PLoS One* **7**, e52644.
- Fujii, S., Bond, C.S. and Small, I.D.** (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc Natl Acad Sci U S A* **108**, 1723–1728.

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

Hammani, K., Okuda, K., Tanz, S.K., Chateigner-Boutin, A.L., Shikanai, T. and Small, I. (2009) A study of new *Arabidopsis* chloroplast RNA editing mutants reveals general features of editing factors and their target sites. *Plant Cell* **21**, 3686–3699.

Harrison, S.J., Mott, E.K., Parsley, K., Aspinall, S., Gray, J.C. and Cottage, A. (2006) A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation. *Plant Methods* **2**, 19.

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.

Hayes, M.L., Giang, K., Berhane, B. and Mulligan, R.M. (2013) Identification of two pentatricopeptide repeat genes required for RNA editing and zinc binding by C-terminal cytidine deaminase-like domains. *J Biol Chem* **288**, 36519–36529.

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

Jenkins, B.D., Kulhanek, D.J. 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.

Jiang, Y., Fan, S.L., Song, M.Z., Yu, J.N. and Yu, S.X. (2012) Identification of RNA editing sites in cotton (*Gossypium hirsutum*) chloroplasts and editing events that affect secondary and three-dimensional protein structures. *Genet Mol Res* **11**, 987–1001.

Kong, M., Wang, F., Yang, Z. and Mi, H. (2013) ATPG is required for the accumulation and function of chloroplast ATP synthase in *Arabidopsis*. *Chin Sci Bull* **58**, 3224–3232.

- Kotera, E., Tasaka, M. and Shikanai, T.** (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326–330.
- Kroeger, T.S., Watkins, K.P., Friso, G., van Wijk, K.J. and Barkan, A.** (2009) A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc Natl Acad Sci U S A* **106**, 4537–4542.
- Lurin, C., Andres, C., Aubourg, S. et al.** (2004) Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **16**, 2089–2103.
- McCormac, D.J. and Barkan, A.** (1999) A nuclear gene in maize required for the translation of the chloroplast *atpB/E* mRNA. *Plant Cell* **11**, 1709–1716.
- Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T. and Kimura, T.** (2007) Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34–41.
- Nelson, B.K., Cai, X. and Nebenfuhr, A.** (2007) A multicolored set of *in vivo* organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J* **51**, 1126–1136.
- Okuda, K., Chateigner-Boutin, A.L., Nakamura, T., Delannoy, E., Sugita, M., Myouga, F., Motohashi, R., Shinozaki, K., Small, I. and Shikanai, T.** (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. *Plant Cell* **21**, 146–156.
- Okuda, K., Hammani, K., Tanz, S.K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I. and Shikanai, T.** (2010) The pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid *ndhB* and *ndhG* transcripts. *Plant J* **61**, 339–349.

- Okuda, K., Myouga, F., Motohashi, R., Shinozaki, K. and Shikanai, T.** (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. *Proc Natl Acad Sci U S A* **104**, 8178–8183.
- O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M. and Small, I.** (2008) On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol* **25**, 1120–1128.
- Peeters, N.M. and Hanson, M.R.** (2002) Transcript abundance supercedes editing efficiency as a factor in developmental variation of chloroplast gene expression. *RNA* **8**, 497–511.
- Prikryl, J., Watkins, K.P., Friso, G., van Wijk, K.J. and Barkan, A.** (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. *Nucleic Acids Res* **36**, 5152–5165.
- Robbins, J.C., Heller, W.P. and Hanson, M.R.** (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast *accD* transcript. *RNA* **15**, 1142–1153.
- Ruwe, H., Castandet, B., Schmitz-Linneweber, C. and Stern, D.B.** (2013) *Arabidopsis* chloroplast quantitative editotype. *FEBS Lett* **587**, 1429–1433.
- 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.
- Takenaka, M. and Brennicke, A.** (2012) Using multiplex single-base extension typing to screen for mutants defective in RNA editing. *Nat Protoc* **7**, 1931–1945.
- Takenaka, M., Verbitskiy, D., Zehrmann, A. and Brennicke, A.** (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. *J Biol Chem* **285**, 27122–27129.
- Takenaka, M., Zehrmann, A., Verbitskiy, D., Hartel, B. and Brennicke, A.** (2013a) RNA editing in plants and its evolution. *Annu Rev Genet* **47**, 335–352.

Takenaka, M., Zehrmann, A., Brennicke, A. and Graichen, K. (2013b) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One* **8**, e65343.

Till, B., Schmitz-Linneweber, C., Williams-Carrier, R. and Barkan, A. (2001) CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* **7**, 1227–1238.

Toda, T., Fujii, S., Noguchi, K., Kazama, T. and Toriyama, K. (2012) Rice MPR25 encodes a pentatricopeptide repeat protein and is essential for RNA editing of *nad5* transcripts in mitochondria. *Plant J* **72**, 450–460.

Watkins, K.P., Kroeger, T.S., Cooke, A.M., Williams-Carrier, R.E., Friso, G., Belcher, S.E., van Wijk, K.J. and Barkan, A. (2007) A ribonuclease III domain protein functions in group II intron splicing in maize chloroplasts. *Plant Cell* **19**, 2606–2623.

Yagi, Y., Nakamura, T. and Small, I. (2014) The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J* **78**, 772–782.

Yagi, Y., Tachikawa, M., Noguchi, H., Satoh, S., Obokata, J. and Nakamura, T. (2013a) Pentatricopeptide repeat proteins involved in plant organellar RNA editing. *RNA Biol* **10**,

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

Yin, P., Li, Q., Yan, C. et al. (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **504**, 168–171.

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A.L., Delannoy, E., Hammani, K., Small, I. and Huang, J. (2009) The Arabidopsis gene *YSI* encoding a DYW protein is required for editing of *rpoB* transcripts and the rapid development of chloroplasts during early growth. *Plant J* **58**, 82–96.

Zoschke, R., Nakamura, M., Liere, K., Sugiura, M., Borner, T. and Schmitz-Linneweber, C. (2010) An organellar maturase associates with multiple group II introns. *Proc Natl Acad Sci U S A* **107**, 3245–3250.

Figure Legends

Figure 1. Predicted alignment of AEF1 to the target *atpF* RNA. The 6/1' amino acid combinations for each motif are presented aligned to the *atpF* RNA sequence upstream of the editing site at nucleotide 12707. Open boxes indicate nucleotides that match the expectation based on correlations seen for other editing factors, shaded boxes indicate mismatches (Barkan et al., 2012).

Figure 2. Characterization of the *aef1-1* mutant. (a) The visible phenotype of *aef1-1* mutants. After one week growth, *aef1-1* mutants are pale and visibly retarded in growth in comparison to wild type Col-0 plants. After 25 days growth on soil, *aef1-1* mutants display extensively variegated true leaves and continue to grow slowly. These phenotypes are completely restored in complemented plants expressing an introduced *AEF1* gene (compl.). (b) Schematic representation of AEF1 (At3g22150). The protein comprises 6 PLS triplets terminating with an E domain. The T-DNA insertion in SALK_041683 (*aef1-1*) is within the sequence encoding the first P motif.

Figure 3. Subcellular localisation of AEF1. An AEF1-GFP fusion was expressed in Arabidopsis cells by biolistic transformation using a Rubisco small subunit (SSU) RFP fusion as a marker for plastid targeting (A-C) and ScCOX4-mCherry as a marker for mitochondrial targeting (D-F). The overlays suggest that AEF1-GFP is targeted to plastids (C) but not to mitochondria (F).

Figure 4. Lack of editing of *atpF* in *aef1-1*. Poisoned primer extension gel showing the extent of editing of *atpF* transcripts in *aef1-1* mutants compared to that in complemented plants (AEF1) and

wild type (Col-0). The top band (E) represents edited transcripts as indicated by the lane with the edited template control (Ed), the lowest band (P) is unextended primer, and the middle darker band (U) represents unedited transcripts as indicated by the lane with the unedited template control (Un). The percentage editing (% E) is calculated as the proportion of fluorescence of the edited band over the total fluorescence of both the edited and unedited bands. Editing of *atpF* is completely abolished in the *aef1-1* mutant and is restored to wild type levels in the complemented plants.

Figure 5. Relative splicing efficiency of chloroplast introns in *aef1-1* and *crs1-1*. Relative splicing efficiency was calculated as $\log_2$ of the ratio of spliced to unspliced transcripts in the mutants over the ratio of spliced to unspliced transcripts in wild type. White bars represent the relative splicing efficiency in *aef1-1* and grey bars represent the relative splicing efficiency in *crs1-1*. Error bars indicate standard deviation in biological replicates (n=3). The *aef1-1* editing mutant has a serious defect in *atpF* splicing (although not to the extent seen in *crs1-1*) and a minor splicing defect in *ndhB*.

Figure 6. Levels of plastid ATP synthase subunits in *aef1-1* and *crs1-1*. Western blots of proteins from *aef1-1*, complemented *aef1-1* (*AEF1*), Col-0 wild type (WT) and *crs1-1* leaves were probed with antibodies against AtpF, ATPG and AtpA (and TUB α as a loading control). Three replicates are shown for each experiment. Both *aef1-1* and *crs1-1* lack AtpF, but AtpF is restored to wild type levels in the *AEF1* complemented plants.

Figure 7. Editing of mitochondrial *nad5* (21975) in *aef1-1*. (a) Alignment of rice MPR25 to its editing site in *nad5* and to rice chloroplast *atpF*, and alignment of AEF1 to the orthologous Arabidopsis mitochondrial *nad5* editing site at nucleotide 21975 (sequence accession NC_001284).

(b) Poisoned primer extension gel showing the extent of editing of *nad5* transcripts in *aef1-1* mutants compared to that in complemented plants (*AEF1*) and wild type (*Col-0*). The upper band (E) represents edited transcripts as indicated by the lane with the edited template control (Ed), the lowest band (P) is unextended primer, and the middle darker band (U) represents unedited transcripts as indicated by the lane with the unedited template control (Un). This editing site shows decreased editing in *aef1-1* mutants and is restored to wild type levels in the complemented plants.

Figure 8. Splicing of chloroplast *atpF* in rice *mpr25*. a) Northern blot of rice total RNA from wild type (WT), *mpr25* and *mpr25* complemented with the full-length *MPR25* ORF driven by the maize ubiquitin promoter (compl.). RNA size markers are shown in the left panel and the stained rRNA below to show loading. The blot was hybridised with a probe covering the whole *atpF* gene. Bands are numbered to show which signals correspond to which transcripts. b) Schematic depiction of the possible transcripts covering the *atpF* locus. Spliced-out introns are indicated by dotted lines. Numbering of the transcripts matches the numbering of the bands in a). Splicing of *atpF* is severely affected in *mpr25*, as shown for example by the ratio between transcript 7 (unspliced) and transcript 8 (spliced) compared with the signals seen in the wild type.

Figure 1

AEF1

motif		P	L	S	P	L	S	P	L	S	P	L	S	P	L	S	P	L	S				
pos. 6		N	S	N	N	V	S	N	L	N	N	T	S	N	A	S	T	V	C				
pos. 1'		D	S	N	S	D	N	D	P	D	D	F	D	N	N	N	D	S					
<i>atpF</i> (12707)		a	c	c	g	a	u	a	u	u	u	a	g	c	a	a	c	a	a	a	u	c	C

Figure 2

(a)

(b)

Figure 3

Figure 4

Figure 5

Figure 6

Figure 7

(a)

MPR25

motif	S	P	L	S	P	L	S	P	L	S	P	L	S	P	L	S	P	L	S				
pos. 6	A	N	S	N	N	V	S	N	L	N	N	T	S	N	A	T	T	L	C				
pos. 1'	P	D	T	N	T	D	N	D	P	D	D	G	D	T	N	S	D	T					
<i>Os-atpF</i>	a	c	c	g	a	u	a	u	u	u	u	a	g	c	a	a	c	a	a	a	u	c	U
<i>Os-nad5</i>	a	u	a	c	c	u	a	u	u	c	u	g	u	u	u	a	g	u	a	c	u	u	C

AEF1

motif	P	L	S	P	L	S	P	L	S	P	L	S	P	L	S	P	L	S					
pos. 6	N	S	N	N	V	S	N	L	N	N	T	S	N	A	S	T	V	C					
pos. 1'	D	S	N	S	D	N	D	P	D	D	F	D	N	N	N	D	S						
<i>nad5</i> (21975)	a	u	a	c	c	u	a	u	u	c	u	g	u	u	u	a	g	u	a	c	u	u	C

(b)

(a)

(b)