PPR proteins – orchestrators of organelle RNA metabolism

Aleix Gorchs Rovira and Alison G Smith*

Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK

Correspondence

*Corresponding author,

e-mail: as25@cam.ac.uk

Pentatricopeptide repeat (PPR) proteins are important RNA regulators in chloroplasts and mitochondria, aiding in RNA editing, maturation, stabilisation or intron splicing, and in transcription and translation of organellar genes. In this review, we summarise all PPR proteins documented so far in plants and the green alga *Chlamydomonas.* By further analysis of the known target RNAs from *Arabidopsis thaliana* PPR proteins, we find that all organellar-encoded complexes are regulated by these proteins, although to differing extents. In particular, the orthologous complexes of NADH dehydrogenase (Complex I) in the mitochondria and NADH dehydrogenase-like (NDH) complex in the chloroplast were the most regulated, with respectively 60 and 28% of all characterised *A. thaliana* PPR proteins targeting their genes.

Abbreviations – CMS**,** Cytoplasm Male Sterility; PMT, Post-transcriptional modification; T/P/OPR, tetra-, penta-, octo-tricopeptide repeat.

The need to regulate organelle genetic expression

Photosynthesis and respiration, which arose initially in prokaryotic organisms, requires the assembly of many different components to form functional protein complexes (Vermaas 2001) and networks for effective electron transfer (Anraku 1988). After the endosymbiotic events in which a free-living cyanobacterium and an alpha-proteobacterium gave rise respectively to the chloroplasts and mitochondria found in eukaryotic cells, there was migration of essential genes from the "new organelle" to the nucleus. In the majority of present-day photosynthetic eukaryotes we find only around 5% of the

several thousand chloroplast proteins encoded in the organelle (Martin et al. 2002), and even fewer of the mitochondrial proteins. Nonetheless, the genes that remain encode many of the key proteins of the photosynthetic and respiratory processes. Having three different compartments synthesising organellar proteins poses a major challenge to regulate precisely and efficiently the expression of the various components in the proper dosage. A key question therefore is how the nucleus orchestrates the production and assembly of these complex structures, in particular to achieve a fine-tuned balance of the different components of the photosynthetic and respiratory machinery.

The organelle environment has retained many bacterial properties from the endosymbiotic events, and there are many similarities between the genetic machinery of chloroplasts and mitochondria and in bacteria. Nonetheless there is little regulation of transcriptional processes in organelles, and mRNA levels do not correlate with protein abundance (del Campo 2009). Instead, eukaryotes have focussed their efforts on developing highly complex and effective post-transcriptional factors that influence photosynthetic and mitochondrial component dosages and correct assembly of the complexes by RNA stabilisation, modification, editing, and translation (Rochaix 2002). One group of proteins intimately involved in this system are the so-called pentatricopeptide repeat (PPR) proteins, which are RNA binding and are involved in the maturation of organelle RNA in eukaryotic organisms(Barkan and Small 2014). These proteins are closely related to tetraricopeptide repeat (TPR) proteins involved in protein-protein interactions and share functionality with the octotricopeptide repeat (OPR) proteins (Rahire et al. 2012. Bohne et al. 2016).

T/P/OPR proteins are members of the alpha-solenoid family of proteins. They contain between 2 and 30 degenerate 34, 35 or 38 amino acid helical repeats respectively, which stack together to form extended surfaces that recognise DNA, RNA or other proteins (Kobe and Kajava 2000) (Fig. S1). This repeat is often degenerate, although it can be easily recognised by its repetition and some consensus amino acids that are highly conserved (Small and Peeters 2000). Each repeat contains two antiparallel alpha helixes that, in the case of PPRs, are thought specifically to recognise an RNA base (Delannoy et al. 2007). In higher plants such as *A. thaliana*, there are hundreds of PPR proteins encoded in the nucleus, the vast majority of them targeted to mitochondria or chloroplasts (Colcombet et al. 2013). The high number of PPRs in land plants been linked to the extent of post-transcriptional editing of mRNA from C to U of the different organellar mRNA (Barkan and Small 2014). This is supported by the fact that there are no

reports of post-transcriptional editing of chloroplast or mitochondrial mRNAs in the green alga *Chlamydomonas reinhardtii,* which has only 14 PPR proteins (Tourasse et al. 2013). However, *A. thaliana* encodes just one OPR protein, RAP, which is involved in the processing of the chloroplast 16S ribosomal RNA (Kleinknecht et al. 2014), whereas *C. reinhardtii* has more than forty OPR proteins (Eberhard et al. 2011). This might therefore indicate different evolutionary routes for the two photosynthetic organisms in the control of their organelle genes: expansion of either PPRs or OPRs.

Investigation of the function of PPR proteins, either individually or in systematic studies such as by Cheng et al. (2016), have revealed that their effect is not limited to RNA stability or RNA editing, but includes processing of polycistronic mRNAs (Meierhoff et al. 2003), translation of various mRNA with which they associate (Yamazaki et al. 2004), and intron splicing of organellar group II introns (Khrouchtchova et al. 2012). There have been several detailed reviews of PPR proteins and their functioning over the last few years (Schmitz-Linneweber and Barkan 2007, Stern et al. 2010, Shikanai and Fujii 2013, Tourasse et al. 2013, Barkan and Small 2014, Hammani et al. 2014, Manna 2015). Here, we gather what is known of PPR proteins specifically in photosynthetic organisms, documenting where known their organelle targeting, their mechanism of function, their target genes and what effect knocking out or down that specific PPR has on the organelle and/or the host organism. We then analyse further the PPR proteins in *A. thaliana* to provide a first detailed examination of their role in a single organism.

PPR proteins are found throughout the plant kingdom

PPR proteins play crucial roles in plant function and development, with more evidence gained every year. Since their discovery in 2000 (Small and Peeters 2000), the realisation of the scale of their representation in plant genomes has been recognised, with ~450 in *A. thaliana* and close to 600 in maize (Lurin et al. 2004). The majority have been identified simply from their predicted amino acid sequences, but a total of 188 PPR proteins from photosynthetic organisms have been further characterised to some extent. Supplementary Dataset 1 lists all these documented PPR proteins, with a summary in Table 1. More than half of the studies describe a PPR protein from *A. thaliana,* reporting a total of 41 chloroplast-targeted PPR proteins and 63 mitochondrially localised. There are also reports on agriculturally relevant species such as *Zea mays* (maize) or *Oryza sativa* (rice), which combined represent 25% of the total

(Table 1). For those from "other plant species" the majority did not characterise the PPR function directly but instead linked it to an observed phenotype, most commonly cytoplasmic male sterility (CMS) (Gillman et al. 2007, Uyttewaal et al. 2008, Liu et al. 2016). This is where ordinarily monoecious plants are converted to females because of pollen abortion, a desirable agronomic trait for ease of production of hybrid lines (reviewed in Budar et al. 2003). In order to consider the molecular details of the function and target genes of PPRs, and to gain relevant insight into the overall role of PPRs more generally, we decided to focus solely on the *A. thaliana* data due to the high numbers of PPRs characterised, 108 in total, representing almost a quarter of all PPRs in this plant.

Editing is the most prevalent function of *A. thaliana* PPR proteins

Based on analyses of knockout lines for each of the characterised *A. thaliana* PPR proteins, it is possible to determine their targets and which aspect of post-transcriptional modification (PTM) they affect (Table 2). The different PTMs that PPR proteins are important for are mRNA stabilisation, transcript processing, intron splicing, and RNA editing, as well as effects on the translation of certain mRNAs (for a comprehensive review of PTMs refer to Stern et al. 2010). Inspection of the various PTMs of mitochondrial and chloroplast genes by the different 108 PPR proteins, presented in Table 2, reveals some interesting observations. Firstly, all organelle encoded processes are affected by at least one PPR protein. We also found that the frequency of PPRs involved in mRNA stabilisation and translation is quite low in both organelles, whereas that for mRNA processing and intron splicing is more common. By far and away the most frequent prevalent effect is that of RNA editing, accounting for 40 of 76 (53%) of the PTMs identified in the chloroplast, and 65 of 95 (68%) in the mitochondria.

This high frequency of RNA editing is found in land plants and it is often correlated with the high abundance of PPR proteins. The reason for so many editing PPRs has been proposed to be as a means for the cell to 'debug' the genetic material in the organelles, which is transmitted by asexual reproduction, so there is no opportunity for gene conversion during recombination (Maier et al. 2008). This repair would be more necessary for the land plants due to the higher rates of UV radiation. The debugging theory states that after the mutation has occurred, a PPR protein would evolve to fix it. However, mutants incapable of editing the chloroplast-targeted CLB19 (Ramos-Vega et al. 2015) or the mitochondrial

REME2 (Bentolila et al. 2013) resulted in lethality, implying the reverse order of events. Moreover, on four other occasions mutations of these PPR proteins led to impaired gametogenesis but not lethality (Lu et al. 2011, Sun et al. 2018) making it impossible to select for these organellar mutations in the first place. Instead, the theory of neutral evolution theory is more likely to be operating here. This theory (reviewed in Barkan and Small 2014) proposes that biology tends to evolve complexity by having processes in the cell that are redundant or do not have any effect in the functioning of the cell. This could explain why in a third of the cases analysed in this review, 59 of 171 (34%), there is no physiological effect when editing of a particular gene is impaired (Robbins et al. 2009, Brehme et al. 2015) (Table 2 and Supplementary Dataset 1).

However, from the data presented in Table 2, we can see that there is a higher proportion of "no phenotype" in the mitochondria than in the chloroplast, which might be due to the greater number of editing by PPR proteins that happens in this organelle. Interestingly, if we link the post-transcriptional processes to the "no phenotype", we can see that editing accounts for 15 of 16 cases in the chloroplast and 34 of 41 in the mitochondria. This suggests that editing is favoured by neutral evolution probably because it provides more opportunity for evolution via the degenerate code and synonymous and neutral amino acid mutations. However, "no phenotype" is not restricted to editing, it is also seen for PPRs involved in RNA stability (Stoll et al. 2014), processing (Fujii et al. 2016) and intron splicing (Yap et al. 2015) although to a significantly lesser extent. Nonetheless, it is possible that it was not possible to quantify the effect of the knock-out or that these proteins work in a highly complex network where they have a dose effect in the function of the target gene. For example, *A. thaliana nad4* is edited by eight different PPRs (Fig. 1) and, in half of the cases, the lack of editing has no significant phenotype. This could provide a safety net for possible random mutations in the PPR transcripts or in the target mRNA. Despite the high correlation of editing and "no phenotype" seen in our analysis, it is worth pointing out that editing can play a crucial role in plants, especially in ferns where it has been reported that editing for the creation of the start codon is conserved and poses a selective advantage over having a translatable sequence encoded in the genome (Li et al. 2018).

NADH dehydrogenase: the main target?

Figure 2 shows a schematic of the complexes found in the chloroplast thylakoid and mitochondrial inner membranes as well as soluble proteins, together with a compilation of the numbers of PPR proteins known to act on each one. Each complex has components whose transcripts are influenced by at least 2 PPR proteins. Surprisingly, the vast majority of the PPR proteins, a total of 78 of 108 analysed so far, are linked to two similar complexes in the organelles, the mitochondrial NADH dehydrogenase (Complex I) and the NADH dehydrogenase-like (NDH) complex in the chloroplast. These complexes share a common L-shaped form (Shikanai 2016) and both catalyse proton translocation across the membrane during electron transfer to a quinone, but their roles are quite different. The chloroplast NDH is a ferredoxin-dependent plastoquinone reductase involved in PSI cyclic electron transfer and is important for optimising the induction of photosynthesis in water stress. Nevertheless, the complex is not necessary for growth in optimal conditions (Burrows et al. 1998) and its genes are the first to be lost in nonphotosynthetic orchids (Schelkunov et al. 2015). In contrast, Complex I is responsible for oxidation of NADH produced during respiration and it is the main point of entry of electrons into the mitochondrial electron transport chain (mETC) providing up to 40% of the protons for the formation of ATP (Fromm et al. 2016). Unlike its chloroplast counterpart, lack of this complex affects carbon metabolism and photosynthesis (Fromm et al. 2016), and entails profound changes leading to curled leaves and delayed reproductive and vegetative development phenotype (de Longevialle et al. 2007). Surprisingly, recent studies have found that hemiparasitic mistletoe has lost the availability to produce a functional Complex I (da Fonseca-Pereira et al. 2018) pointing to more similarities to the NDH complex.

In the chloroplast, 24 of 85 (28%) of the total characterised chloroplast PTM events are involved with the NDH complex, for its mitochondrial counterpart, NADH dehydrogenase, this number is increased to 60 of 99 (60%) (Fig. 2). The number of PPRs affecting the NDH/NADH complexes is substantially greater than for the other complexes. Similarly, in *Oryza sativa* and *Zea mays*, PPRs affecting the NDH and NADH dehydrogenase are 7 of 27 (26%) and 15 of 24 (62%) respectively. These high numbers could be attributed to the theory of neutral evolution: these complexes are not vital for the plant, in contrast to the ATPases or the two photosystems, allowing minor deficiencies from mutations and other post-transcriptional processes to be better tolerated.

Alignment of NDH and NADH dehydrogenase PPR proteins with each other does not reveal any

similarity, neither between PPRs targeted to the same organelle, nor linked to orthologous subunits (data not shown). The absence of similar patterns is illustrated by consideration of how two gene orthologues from NADH dehydrogenase (*nad4)* and NDH (*ndhD*) are modified, processed and protected by their respective PPRs. Both transcripts are heavily edited, but the editing sites are not localised in the same area (Fig. 1). Moreover, mitochondrial *nad4* also has PPR-aided intron splicing, 5' processing and 3' stabilisation, whereas *ndhD* does not (Fig. 1). Thus, it is likely that the PPR and the post-transcriptional modification mechanisms evolved separately for these two genes.

Nevertheless, it is worth noting that in the *A. thaliana* chloroplast and mitochondrion all complexes that have at least one element encoded in the organelle genomes are influenced by PPR proteins. The extent of involvement in the post-transcriptional processes varies depending on the complex involved, with ATP synthase being the least regulated in both organelles, perhaps because of the vital function that it carries out in the cell. We found similar numbers of PPRs involved in the chloroplast cytochrome b_6f complex and the mitochondrial counterpart the cytochrome bc_1 complex. In the mitochondria, more than half of the PPR proteins associated with the cytochrome bc_1 (Complex III) and cytochrome oxidase (Complex IV) interact with genes involved in biogenesis (Tang et al. 2010) and maturation (Chateigner-Boutin et al. 2013), rather than the structural subunits, demonstrating PPR proteins have acquired a more complex role in the production of the functional complexes. In the chloroplast both photosystems, PSI and PSII, have the same number of PPRs interacting with them (Fig. 2A). However, although the targets were identified from unbiassed mutant libraries, it should be pointed out that many more PPR proteins remain to be characterised, so this relative proportion may change in the future.

It is not just photosynthetic and respiratory complexes that interact with PPRs in the organelles. Factors involved with both ribosomal proteins and RNA are prevalent in both organelles. In the chloroplast PPRs influence other processes such as protein degradation by proteases (Ramos-Vega et al. 2015) or carbon metabolism with the regulation of *rbcL* (Luro et al. 2013) or *accD* involved in fatty acid metabolism (Du et al. 2017). Involvement of the PPR proteins in mRNA translation is also more prevalent in the chloroplast (Table 2). Lastly, we found that four chloroplast PPRs: DG1 (Chi et al. 2008), PDM2 (Du et al. 2017), PDM3 (Zhang et al. 2017) and VAC1 (Tseng et al. 2010), influence overall gene expression (nuclear encoded polymerase (NEP)- and PEP-related genes) but no such PPRs were found to have a similar effect in the mitochondria. This could be explained partially by the fact that mitochondria use

only nucleus-encoded RNA polymerases (review in Börner et al. 2015), whereas chloroplasts have in addition a plastid-encoded polymerase (PEP) (Kremnev and Strand 2014). This could mean that there has been a possible co-evolution of PPRs and PEPs in the photosynthetic organisms.

Conclusions

PPR proteins are multifaceted factors that are important for essential processes in plants, including photosynthesis, respiration, organelle development and gametogenesis. In our meta-analysis of characterised *A. thaliana* PPR proteins we found that there is a significant preference in the involvement of these proteins for the chloroplast NDH and the mitochondrial NADH dehydrogenase compared to other organellar complexes probably due to their non-vital role in the cell. Furthermore, editing seems to be very widespread in both organelles even though there is a preference bias in the mitochondria of *A. thaliana*. In some cases, knock-out of the PPR protein has no identifiable effect on the photosynthetic organism, providing strong evidence for the theory of neutral evolution in this protein family. However, there are only few studies on cumulative effects of these PPRs. It could be that the "no phenotype" PPR proteins are part of a much larger network that gradually improves gene translation by editing or removing certain introns or processing the 5' or 3' end of transcripts. Moreover, even in *A. thaliana,* only a proportion of the entire PPR complement has been characterised. Therefore, there is now a need to push for a more comprehensive knowledge of PPR proteins: how they are organised in their own interaction network, and how they interact with other proteins to achieve their final functionality. Finally, one question remains unanswered in all these studies: if these PPRs are involved in PTMs in the organelles, which transcription factors or external cues regulate their expression in turn? Clearly, there is still much to learn about these enigmatic nucleus-encoded factors, and their orchestration of organelle function.

Acknowledgements – AGR is supported by a studentship from the Cambridge BBSRC Doctoral Training Partnership.

Figure legends

Fig. 1. Representation of the effect of the PPR proteins on two orthologous genes in *A. thaliana*, the mitochondrial *nad4* and chloroplast *ndhD*. These two genes are regulated by various PPR proteins the effects of which are depicted by the key. If the diamonds are coloured, it means that without the PPR protein, there is a phenotype that can be seen in the mutant strain such aberrant chlorophyll levels or respiratory impairment. Translation start is indicated by +1.

Fig. 2. Representation of the number of PPR proteins involved with different complexes in the chloroplast (A) and mitochondria (B). The different complexes are coloured according to the proportion of the total of PPR proteins involved with their plastid-encoded genes. The values used for colouring can be found in the tables showing the total number of PPR proteins, which do not add up to 108 because one PPR can affect multiple complexes. Carbon metabolism in A involves the genes *rbcL* and *accD*.

Supporting information

Additional supporting information may be found in the online version of this article:

Fig. S1. Schematic representation of the repeats found in TPRs, PPRs and OPRs.

Supplementary Dataset 1. Compilation of the information found in the papers reporting a PPR in several photosynthetic organisms.

References

Anraku Y (1988) Bacterial Electron Transport Chains. Annu Rev Biochem 57:101–132

Barkan A, Small I (2014) Pentatricopeptide Repeat Proteins in Plants. Annu Rev Plant Biol 65:415–442

Bentolila S, Babina AM, Germain A, Hanson MR (2013) Quantitative trait locus mapping identifies REME2, a PPR-DYW protein required for editing of specific C targets in Arabidopsis mitochondria. RNA Biol 10:1520–5

Bohne A-V, Schwenkert S, Grimm B, Nickelsen J (2016) Roles of Tetratricopeptide Repeat Proteins in

Biogenesis of the Photosynthetic Apparatus. Int Rev Cell Mol Biol 324:187–227

Börner T, Aleynikova AY, Zubo YO, Kusnetsov V V. (2015) Chloroplast RNA polymerases: Role in chloroplast biogenesis. Biochim Biophys Acta - Bioenerg 1847:761–769

Brehme N, Bayer-Császár E, Glass F, Takenaka M (2015) The DYW Subgroup PPR Protein MEF35 Targets RNA Editing Sites in the Mitochondrial rpl16, nad4 and cob mRNAs in Arabidopsis thaliana. PLoS One 10:e0140680

Budar F, Touzet P, De Paepe R (2003) The Nucleo-Mitochondrial Conflict in Cytoplasmic Male Sterilities Revisited. Genetica 117:3–16

Burrows PA, Sazanov LA, Svab Z, Maliga P, Nixon PJ (1998) Identification of a functional respiratory complex in chloroplasts through analysis of tobacco mutants containing disrupted plastid ndh genes. EMBO J 17:868–76

Chateigner-Boutin A-L, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I (2013) The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J 74:935–945

Cheng S, Gutmann B, Zhong X, Ye Y, Fisher MF, Bai F, Castleden I, Song Y, Song B, Huang J, Liu X, Xu X, Lim BL, Bond CS, Yiu S-M, 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

Chi W, Ma J, Zhang D, Guo J, Chen F, Lu C, Zhang L (2008) The pentratricopeptide repeat protein DELAYED GREENING1 is involved in the regulation of early chloroplast development and chloroplast gene expression in Arabidopsis. Plant Physiol 147:573–84

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

da Fonseca-Pereira P, Silva WB, Araújo WL, Nunes-Nesi A (2018) How Does European Mistletoe Survive Without Complex I? Trends Plant Sci.

de Longevialle AF, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small ID (2007) The Pentatricopeptide Repeat Gene *OTP43* Is Required for *trans* -Splicing of the Mitochondrial *nad1* Intron 1 in *Arabidopsis thaliana*. Plant Cell 19:3256–3265

del Campo EM (2009) Post-transcriptional control of chloroplast gene expression. Gene Regul Syst Bio 3:31–47

Delannoy E, Stanley WA, Bond CS, Small ID (2007) Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem Soc Trans 35:

Du L, Zhang J, Qu S, Zhao Y, Su B, Lv X, Li R, Wan Y, Xiao J (2017) The Pentratricopeptide Repeat Protein Pigment-Defective Mutant2 is Involved in the Regulation of Chloroplast Development and Chloroplast Gene Expression in Arabidopsis. Plant Cell Physiol 58:747–759

Eberhard S, Loiselay C, Drapier D, Bujaldon S, Girard-Bascou J, Kuras R, Choquet Y, Wollman F-A (2011) Dual functions of the nucleus-encoded factor TDA1 in trapping and translation activation of atpA transcripts in Chlamydomonas reinhardtii chloroplasts. Plant J 67:1055–1066

Fromm S, Senkler J, Eubel H, Peterhänsel C, Braun H-P (2016) Life without complex I: proteome analyses of an Arabidopsis mutant lacking the mitochondrial NADH dehydrogenase complex. J Exp Bot 67:3079–3093

Fujii S, Suzuki T, Giegé P, Higashiyama T, Koizuka N, Shikanai T (2016) The Restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial *orf291* RNA in Arabidopsis. Plant J 86:504–513

Gillman JD, Bentolila S, Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. Plant J 49:217–227

Hammani K, Bonnard G, Bouchoucha A, Gobert A, Pinker F, Salinas T, Giegé P (2014) Helical repeats modular proteins are major players for organelle gene expression. Biochimie 100:141–150

Khrouchtchova A, Monde R-A, Barkan A (2012) A short PPR protein required for the splicing of specific group II introns in angiosperm chloroplasts. RNA 18:1197–209

Kleinknecht L, Wang F, Stübe R, Philippar K, Nickelsen J, Bohne A-V (2014) RAP, the sole octotricopeptide repeat protein in Arabidopsis, is required for chloroplast 16S rRNA maturation. Plant Cell 26:777–87

Kobe B, Kajava A V (2000) When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. Trends Biochem Sci 25:509–15

Kremnev D, Strand A (2014) Plastid encoded RNA polymerase activity and expression of photosynthesis genes required for embryo and seed development in Arabidopsis. Front Plant Sci 5:385

Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z, Simenc M, Small I, Wafula E, Angarita S, Barker MS, Bräutigam A, dePamphilis C, Gould S, Hosmani PS, Huang Y-M, Huettel B, Kato Y, Liu X, Maere S, McDowell R, Mueller LA, Nierop KGJ, Rensing SA, Robison T, Rothfels CJ, Sigel EM, Song Y, Timilsena PR, Van de Peer Y, Wang H, Wilhelmsson PKI, Wolf PG, Xu X, Der JP, Schluepmann H, Wong GK-S, Pryer KM (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. Nat Plants 4:460–472

Liu Z, Yang Z, Wang X, Li K, An H, Liu J, Yang G, Fu T, Yi B, Hong D (2016) A Mitochondria-Targeted PPR Protein Restores pol Cytoplasmic Male Sterility by Reducing orf224 Transcript Levels in Oilseed Rape. Mol Plant 9:1082–1084

Lu Y, Li C, Wang H, Chen H, Berg H, Xia Y (2011) AtPPR2, an Arabidopsis pentatricopeptide repeat protein, binds to plastid 23S rRNA and plays an important role in the first mitotic division during gametogenesis and in cell proliferation during embryogenesis. Plant J 67:13–25

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 M-L, Mireau H, Peeters N, Renou J-P, Szurek B, Taconnat L, Small I (2004) Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis. PLANT CELL ONLINE 16:2089–2103

Luro S, Germain A, Sharwood RE, Stern DB (2013) RNase J participates in a pentatricopeptide repeat protein-mediated 5′ end maturation of chloroplast mRNAs. Nucleic Acids Res 41:9141–9151

Maier UG, Bozarth A, Funk HT, Zauner S, Rensing SA, Schmitz-Linneweber C, Börner T, Tillich M (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? BMC Biol 6:36

Manna S (2015) An overview of pentatricopeptide repeat proteins and their applications. Biochimie 113:93–99

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (2002) Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc Natl Acad Sci U S A 99:12246–51

Meierhoff K, Felder S, Nakamura T, Bechtold N, 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 15:1480–95

Rahire M, Laroche F, Cerutti L, Rochaix J-D (2012) Identification of an OPR protein involved in the translation initiation of the PsaB subunit of photosystem I. Plant J 72:652–661

Ramos-Vega M, Guevara-García A, Llamas E, Sánchez-León N, Olmedo-Monfil V, Vielle-Calzada JP, León P (2015) Functional analysis of the *Arabidopsis thaliana CHLOROPLAST BIOGENESIS 19* pentatricopeptide repeat editing protein. New Phytol 208:430–441

Robbins JC, Heller WP, Hanson MR (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

Rochaix J-D (2002) Chlamydomonas, a model system for studying the assembly and dynamics of photosynthetic complexes. FEBS Lett 529:34–38

Schelkunov MI, Shtratnikova VY, Nuraliev MS, Selosse M-A, Penin AA, Logacheva MD (2015) Exploring the Limits for Reduction of Plastid Genomes: A Case Study of the Mycoheterotrophic Orchids Epipogium aphyllum and Epipogium roseum. Genome Biol Evol 7:1179–1191

Schmitz-Linneweber C, Barkan A (2007) RNA splicing and RNA editing in chloroplasts. Springer, Berlin, Heidelberg, pp 213–248

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

Shikanai T, Fujii S (2013) Function of PPR proteins in plastid gene expression. RNA Biol 10:1446–56

Small ID, Peeters N (2000) The PPR motif – a TPR-related motif prevalent in plant organellar proteins

Stern DB, Goldschmidt-Clermont M, Hanson MR (2010) Chloroplast RNA Metabolism. Annu Rev Plant Biol 61:125–155

Stoll B, Zendler D, Binder S (2014) RNA Processing Factor 7 and Polynucleotide Phosphorylase Are Necessary for Processing and Stability of nad2 mRNA in Arabidopsis Mitochondria. RNA Biol 11:968–976

Sun Y, Huang J, Zhong S, Gu H, He S, Qu L-J (2018) Novel DYW-type pentatricopeptide repeat (PPR) protein BLX controls mitochondrial RNA editing and splicing essential for early seed development of Arabidopsis. J Genet Genomics 45:155–168

Tang J, Kobayashi K, Suzuki M, Matsumoto S, Muranaka T (2010) The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. Plant J 61:456–466

Tourasse NJ, Choquet Y, Vallon O (2013) PPR proteins of green algae. RNA Biol 10:1526–1542

Tseng C-C, Sung T-Y, Li Y-C, Hsu S-J, Lin C-L, 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

Uyttewaal M, Arnal N, Quadrado M, Martin-Canadell A, Vrielynck N, Hiard S, Gherbi H, Bendahmane A, Budar F, Mireau H (2008) Characterization of Raphanus sativus Pentatricopeptide Repeat Proteins Encoded by the Fertility Restorer Locus for Ogura Cytoplasmic Male Sterility. PLANT CELL ONLINE 20:3331–3345

Vermaas WF (2001) Photosynthesis and Respiration in Cyanobacteria. In: Encyclopedia of Life Sciences. John Wiley & Sons, Ltd, Chichester, UK

Yamazaki H, Tasaka M, Shikanai T (2004) PPR motifs of the nucleus-encoded factor, PGR3, function

in the selective and distinct steps of chloroplast gene expression in *Arabidopsis*. Plant J 38:152–163

Yap A, Kindgren P, Colas des Francs-Small C, Kazama T, Tanz SK, Toriyama K, Small I (2015) AEF1/MPR25 is implicated in RNA editing of plastid *atpF* and mitochondrial *nad5* , and also promotes *atpF* splicing in Arabidopsis and rice. Plant J 81:661–669

Zhang J, Xiao J, Li Y, Su B, Xu H, Shan X, Song C, Xie J, Li R (2017) PDM3, a pentatricopeptide repeat-containing protein, affects chloroplast development. J Exp Bot 68:5615–5627

Table 1. Experimentally characterised PPR proteins in different plants and *Chlamydomonas reinhardtii*, and their subcellular location. The species included in "Other species" are *Raphanus sativus*, *Sorghum bicolor, Hordeum vulgare, Cucumis sativus, Brassica napus, Gossypium hirsutum, Brachypodium distachyon* and *Solanum tuberosum*. The majority of which are associated with cytoplasm male sterility (CMS). These represent a fraction of the total number of likely PPR proteins encoded by each organism. Some PPR proteins are dually localised (Supplementary Dataset 1).

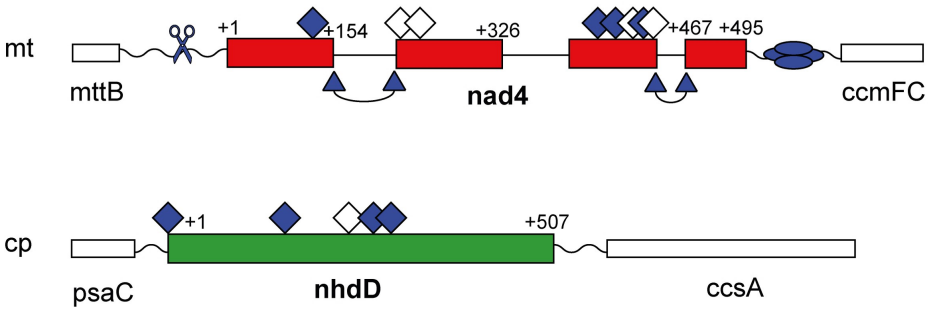
Characterised PPR proteins in photosynthetic organisms

Organism	Total	Chloroplast	Mitochondria	Nucleus
Arabidopsis thaliana	108	41	63	4
Oryza sativa	25	15	10	1
Zea mays	24	9	15	0
Physcomitrella patens	15	5	10	0
Chlamydomonas reinhardtii	3	3	0	0
Other plant species	18	2	15	1

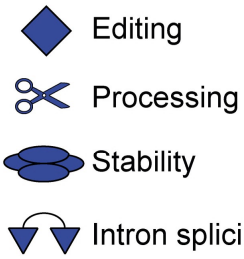
Table 2. Knockout of A thaliana PPR proteins enables targets to be established. All organelle encoded processes are affected by at least one PPR protein. In some cases, a single PPR protein targets several genes, and/or more than one post-transcriptional modification (PTM). This accounts for the fact that there are more events than total number of characterised proteins (see Supplementary Dataset). Functions of the PPRs have been categorised in mRNA stabilisation, translation, processing, splicing and editing (for detailed explanation of the functions please refer to Stern et al 2010). The "no phenotype editing" column highlights how many events do not result in a measurable phenotype due to lack of editing; the "no phenotype other PTMs" are all other knockouts with no phenotype.

	mRNA stabilisation	Processing	Splicing	Translation	Editing	No phenotype editing	No phenotype other PTMs
Chloroplast							
ATP synthase	0	0	0	3	1	1	0
Carbon metabolism	0	1	0	0	6	1	1
Cytochrome b6f	2	5	2	2	0	0	0
NDH	0	1	3	0	20	7	0
PEP RNA polymerase	0	1	1	0	5	2	0
Protease	0	0	0	0	1	0	0
PSI	0	1	2	0	0	0	0
PSII	0	0	0	0	2	1	0
Ribosomes	2	2	8	0	5	3	0
Total for each PTM	**4**	**11**	**16**	**5**	**40**	**15**	**1**
Mitochondria							
ATP synthase	0	0	1	0	2	1	1
Cytochrome c oxidase	0	4	0	0	11	6	3
Cytochrome c reductase	0	0	0	0	6	5	3
NADH dehydrogenase	4	3	15	1	36	18	2
Ribosomes	0	2	0	0	10	4	0
Total for each PTM	**4**	**9**	**16**	**1**	**65**	**34**	**9**

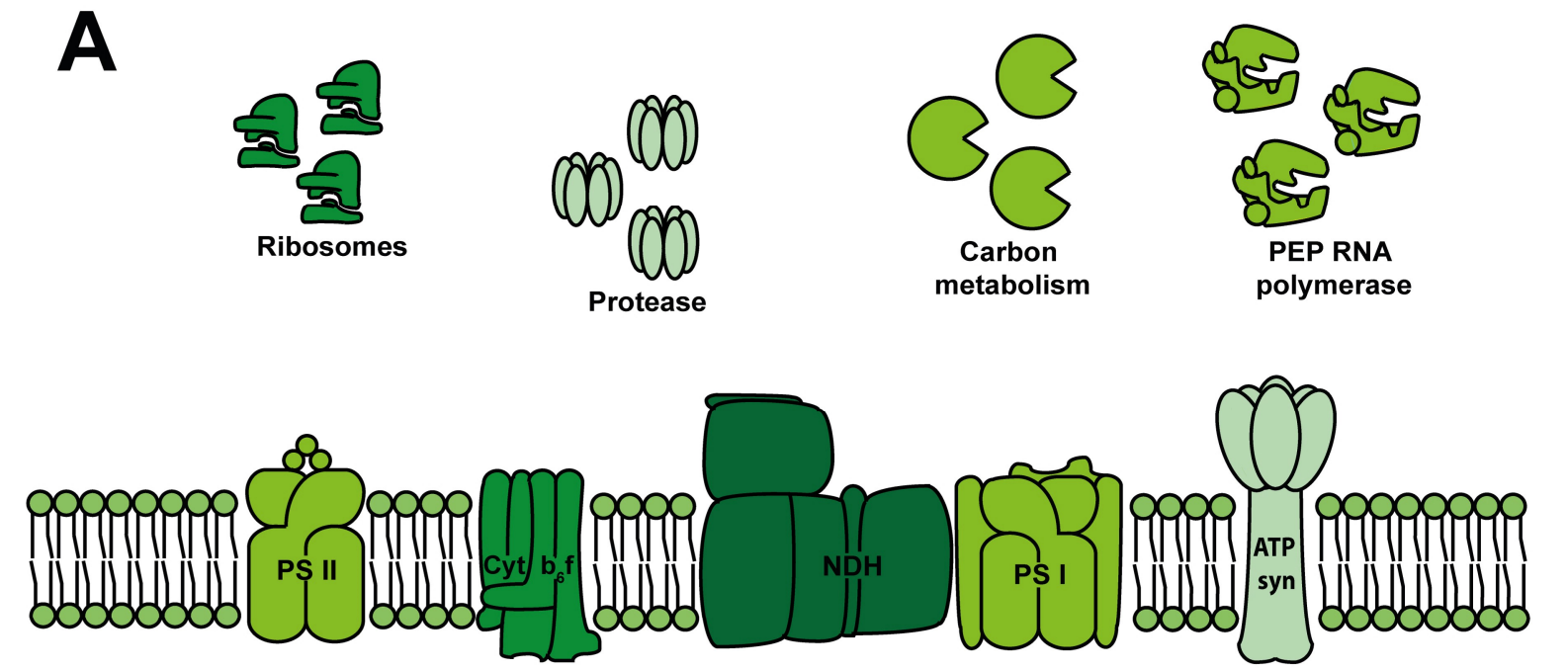
Gene orthologs in *A. thaliana*



Legend



Colour = change phenotype
No colour = no phenotype



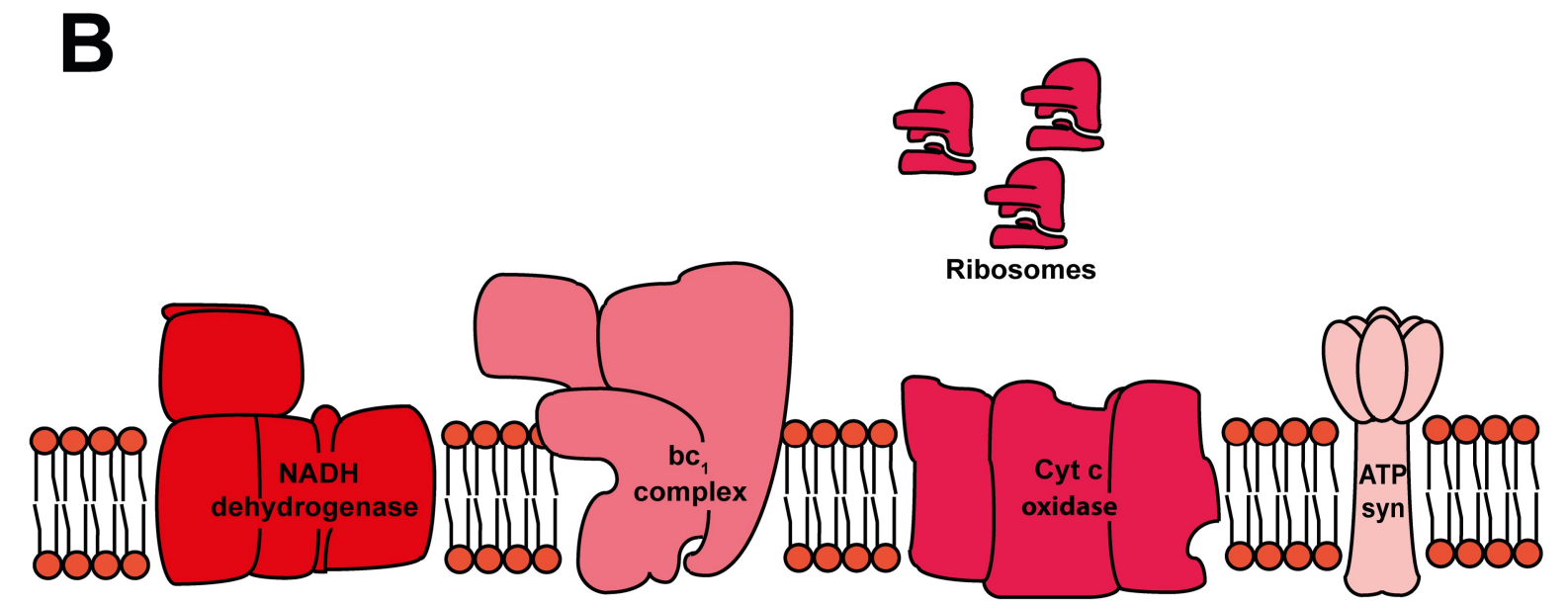
High influence by PPRs

Medium influence by PPRs

Medium/Low influence by PPRs

Low influence by PPRs

	Number of PPRs
NDH	24
Ribosomes	17
Cyt b_6f	14
PEP RNA Polymerase	7
Carbon metabolism	7
PS I	5
PS II	5
ATP synthase	4
Protease	2



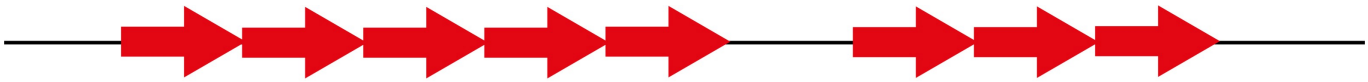
High influence by PPRs

Medium influence by PPRs

Medium/Low influence by PPRs

Low influence by PPRs

	Number of PPRs
NADH dehydrogenase	60
Cyt c oxidase	15
Ribosomes	12
bc_1 complex	9
ATP synthase	3



2-30 helical repeats composed of 34 amino acids - TPR

35 amino acids - PPR

38 amino acids - OPR