RESEARCH PAPER

RNA Biology 10:9, 1557–1575; September 2013; © 2013 Landes Bioscience

Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles

Jean Colcombet[§], Mauricio Lopez-Obando[§], Laure Heurtevin[†,§], Clément Bernard, Karine Martin, Richard Berthomé[‡], and Claire Lurin*

Unité de Recherche en Génomique Végétale (URGV); UMR INRA/UEVE - ERL CNRS 91057; CP 5708; 91057 EVRY CEDEX, France

Current affiliation: [†]Génétique et Amélioration des Fruits et Légumes (GAFL); INRA - UR1052; 84143 MONTFAVET CEDEX, France; [‡]Laboratoire des Interactions Plantes Micro-organismes (LIPM); UMR INRA 441/CNRS 2594; CS 52627; 31326 CASTANET TOLOSAN CEDEX, France

[§]These authors contributed equally to this work.

Keywords: PPR proteins, organelle genome expression, *Arabidopsis thaliana*, targeting peptide, subcellular localization

Abbreviations: PPR, PentatricoPeptide Repeat; RFP, red fluorescent protein; ORF, open reading frame; TP, targeting peptide

Four hundred and fifty-eight genes coding for PentatricoPeptide Repeat (PPR) proteins are annotated in the *Arabidopsis thaliana* genome. Over the past 10 years, numerous reports have shown that many of these proteins function in organelles to target specific transcripts and are involved in post-transcriptional regulation. Therefore, they are thought to be important players in the coordination between nuclear and organelle genome expression. Only four of these proteins have been described to be addressed outside organelles, indicating that some PPRs could function in post-transcriptional regulations of nuclear genes.

In this work, we updated and improved our current knowledge on the localization of PPR proteins of Arabidopsis within the plant cell. We particularly investigated the subcellular localization of 166 PPR proteins whose targeting predictions were ambiguous, using a combination of high-throughput cloning and microscopy. Through systematic localization experiments and data integration, we confirmed that PPR proteins are largely targeted to organelles and showed that dual targeting to both the mitochondria and plastid occurs more frequently than expected. These results allow us to speculate that dual-targeted PPR proteins could be important for the fine coordination of gene expressions in both organelles.

Introduction

Plant nuclear genomes code for more than 99% of the 25 000–30 000 proteins required to build plant cells and tissues.[1] These proteins are addressed to various cell compartments to ensure specific cellular processes. Two other small genomes, formed by primary endo-symbiosis events, which led to the organelle formation, are found in mitochondria and plastids.[2] Throughout evolution, organelles have lost much of their original genomes by the transfer of genetic material to the nucleus. However, they have retained small genomes encoding key proteins and RNAs necessary for their biology. In Arabidopsis, 57 mitochondrial genes and 128 chloroplast genes have been annotated on the corresponding genomes (TAIRv10). The proteins encoded by these genes, acting together with nuclear imported proteins, play an important role in mitochondria and plastid functions.[3,4] Many of the proteins encoded by genes transferred from organelles to the nucleus are important for organelle gene expression or metabolism and need to be targeted back to their original compartment. In addition, many other nuclearly encoded proteins have acquired functions in different steps of organelle biology. Overall, more than 3000 proteins encoded by the nuclear genome are predicted to be targeted to the organelles,[5] creating a requirement for a coordinated regulation of nuclear and organellar gene expression and a precise control of protein addressing and import into the organelles. Several import systems exist in mitochondria and plastids where translocation is mediated mainly by co-translational and post-translational machineries. The main machineries are well known.[6-8] They are named Translocase of the Outer/Inner Mitochondria membrane complexes (TOM/TIM) in mitochondria and Translocase of the Outer/Inner Chloroplast membrane complexes (TOC/TIC) in plastids. TOM/TIM and TOC/TIC account for the targeting of most organellar proteins. These two Translocase complexes share both similar structural conformations and import mechanisms with the recognition of a Targeting Peptide (TP) and the involvement of chaperones, receptor, and

*Correspondence to: Claire Lurin; Email: lurin@evry.inra.fr
Submitted: 03/29/2013; Revised: 08/11/2013; Accepted: 08/12/2013
http://dx.doi.org/10.4161/rna.26128

pore type proteins.[8] Despite these similarities, the mechanisms of translocation are specific to each Translocase. For example, the translocation into plastids requires GTP hydrolysis whereas it is not the case in mitochondria.[8]

Organelle physiological processes are under the control of proteins expressed from distinct genomes suggesting a tight and complex coordination in gene expression and, therefore, intracellular signaling pathways between cell compartments. Whereas nuclear genes are largely regulated at the transcriptional level, organelle genes are often constitutively expressed but tightly regulated at post-transcriptional levels.[9] Imported nuclear proteins are necessary for a wide range of organellar transcriptional and post-transcriptional processes, including RNA transcription, RNA processing, RNA editing, RNA splicing, and translation. Among these nuclear factors, the large family of PentatricoPeptide Repeat (PPR) proteins are emerging more and more as central actors of the inter-compartmental coordination of gene expression.[10] As expected for proteins involved in complex genome regulations, they define one of the largest families encoded by the nuclear genome with 458 members in Arabidopsis, 477 in rice, and up to 800 in *Selaginella moellendorffii*.[11-14]

A typical PPR protein is constructed from a stretch (2–30) of 35-amino acid motifs (known as PPR motifs) often merged in N terminus with a targeting peptide thought to allow an organelle subcellular localization. Several studies confirmed that the targeting peptide is functional, suggesting that PPR proteins are massively targeted to mitochondria or plastids.[10,11] Based on the PPR motif sequences and their relative serial organization, we proposed a classification of PPR proteins in two main subfamilies.[11] In Arabidopsis, the largest one, named the P-type subfamily, contains 255 PPR proteins harboring tandem repeats of a simple canonical PPR motif (the P-type motif). The second one is known as the PLS-type subfamily and contains the remaining 203 PPR proteins.[11] Their module-based structures but also biochemical and genetic data indicate that PPR proteins are able to interact in a sequence-specific way with organelle RNAs to assure various post-transcriptional functions.[10,15] Recently, through computational and molecular biology approaches, a RNA recognition code was proposed for PPR proteins where two adjacent PPR motifs are able to recognize one specific nucleotide.[16,17] The specificity of the base recognition is accomplished by the combination of three amino acids, two located in the first PPR motif (third and sixth positions) and the third at the first position in the subsequent PPR motif.[16,17]

PPR proteins have largely been associated with transcriptional, post-transcriptional, and translational regulation of organellar expression.[10] A growing number of PPR proteins have been shown to be required for editing. For example, CHLORORESPIRATORY REDUCTION 4 (CRR4) is necessary for editing of the chloroplast *ndhD* transcript[18] and MITOCHONDRIAL RNA EDITING FACTOR1 (MEF1) is required for editing of three mitochondrial transcripts.[19] Arabidopsis PPR proteins are also involved in splicing of organelle transcripts: ORGANELLAR TRANSCRIPT PROCESSING43 (OPT43) and OTP51 are necessary for the correct trans-splicing of *nad1* and cis-splicing of *ycf3* transcripts, respectively.[20,21] Finally, PPR proteins are involved in translation processes. For example, CHLOROPLAST RNA PROCESSING 1 (CRP1) has been proposed to be a chloroplast translation regulator[22] and PPR336 is associated with mitochondrial polysomes.[23] As expected with essential players in gene expression involved in respiration and photosynthesis, a large proportion of mutants in PPR genes are embryo or gametophyte lethal.[11,24,25]

Despite the growing PPR literature indicating that PPR proteins function mainly in organelles, some members could also have targets in the nucleus or the cytoplasm. In Arabidopsis, four PPR proteins were shown to be localized out of organelles. Two of them, PROTEINACEOUS RNase P 2 (PRORP2) and PRORP3, are localized exclusively in the nucleus where they are needed to achieve RNase P activity.[26] The two others have a more complex subcellular localization with a dual targeting to both mitochondria and nucleus. The GLUTAMINE-RICH PROTEIN23 (GRP23) interacts in nucleus with RNA polymerase II but its nuclear and mitochondrial functions are not yet understood.[27] Similarly, Hammani and co-workers showed that PPR PROTEIN LOCALIZED TO THE NUCLEUS AND MITOCHONDRIA1 (PNM1) is involved in protein translation in mitochondria whereas it physically interacts with two proteins in the nucleus, NUCLEOSOME ASSEMBLY PROTEIN1 and the transcription factor TCP8.[28] In animals, one example of a PPR protein localized out of the organelles has also been reported but its localization is still a matter of debate. This PPR protein, named BICOID STABILIZATION FACTOR (BSF) in *Drosophila,* as well as Leucine-Rich Repeat PentatricoPeptide Repeat Cassette (LRPPRC) in humans, was localized in the cytoplasm and nucleus of early *Drosophila* embryo cells[29] with roles in transcription and RNA transport. Other authors showed the protein to be localized in mitochondria where it would be involved in mRNAs maturation, poly-adenylation, and translation.[30]

Only a handful of PPR proteins were shown to function out of organelles. Many post-transcriptional processes are being shared by both the organelles and nucleus; therefore, this number may be underestimated. In order to identify new Arabidopsis PPR proteins addressed out of the organelles but also to improve our general knowledge on PPR targeting, we systematically investigated the subcellular localization of a third of the PPR family whose addressing prediction was ambiguous. We took advantage of a high-throughput cloning strategy combined with a transient expression system to elucidate whether the N terminus targeting peptides of candidate PPR proteins were functional to address the protein into organelles. We report in this work that, despite erroneous predictions of subcellular localization, most PPR proteins are addressed to one of the organelles and showed that a fraction of them, probably underestimated, are addressed to both mitochondria and plastids.

Results

Localization study of PPR proteins with ambiguous predictions of localization. We previously published a manually curated list of Arabidopsis PPR gene models.[12] When this work was initiated, the most accurate algorithms to predict subcellular localization of plant proteins were TargetP v1.01[31] and Predotar v1.03.[32] Therefore, we used them to identify Arabidopsis genes coding for PPR proteins with ambiguous localization predictions (**Table 1**). TargetP was recently improved with the TargetP v1.1 version of the software. Among the 458 PPR genes, Predotar predicts that 244 and 92 PPR proteins are addressed, respectively, to mitochondria and plastids, whereas 122 PPRs would not have any organelle localization (**Table 2**). TargetP v1.1 gives similar results with 232 and 123 PPRs localized to mitochondria and plastids, respectively, and 103 PPRs without organelle localizations (**Table 1**). Taken together, 166 PPR proteins were predicted not to be addressed to either of the two main plant organelles by at least one of the two software (Predotar v1.03 and TargetP v1.01). Among them, 53 PPR proteins were not predicted to be addressed in the organelles by both algorithms. We chose to experimentally investigate the subcellular localization of those 166 PPR proteins as they were good candidates to have atypical functions out of the organelles.

Almost all the proteins addressed to organelles contain a targeting peptide in their N terminus extremity, which is cleaved during the transfer through the organelle membranes.[8] A mitochondrial Targeting Peptide (mTP) is typically 40–50 amino acid long,[33-36] whereas a chloroplast Targeting Peptide (cTP) is usually up to 60 amino acid long.[37] To assess the targeting peptide functionality, we systematically merged in frame the first 300 bp, coding for the first 100 amino acids of each candidate PPR ORFs to the Red Fluorescent Protein (RFP) coding sequence using the Gateway technology. The aim of this approach was to experimentally detect any mTP or cTP present in the first 100 amino acids but not recognized by the prediction software. Vector cloning based on Gateway recombination technology was successful for 162 genes (97%). After agro-infiltration of *Nicotiana benthamania* leaves with these constructs and subsequent protoplasts preparation, we were able to detect RFP signals for 131 constructs (79%) using either epifluorescent or confocal microscope. All localization experiments were repeated at least three times and observed independently by two of the authors. **Table 1** summarizes all predicted and experimental data obtained during this study. Presented in **Figure 1A** are examples of typical subcellular localizations observed using 300 bp constructs. In **Figure 1**, RFP fluorescence was visualized using a confocal microscope and compared with the distribution of the mitochondrion-specific probe MitoTracker Green and the chlorophyll autofluorescence. In the overlay panels, combined fluorescence from RFP (in red), MitoTracker (in green), and chlorophyll autofluorecence (in blue) appears in yellow when RFP signal co-localizes with MitoTracker staining indicating a localization of the fusion protein in mitochondria whereas it appears in violet when RFP signal is localized in plastids. It was detected that 68 and 31 300 bp-PPR constructs gave an exclusive mitochondrial and plastid localization, respectively, as exemplified by AT3G15130 and AT3G46610 in **Figure 1A**. Interestingly, 24 constructs exhibited a signal in both organelles (see for example, AT2G36240 and AT5G47460 in **Fig. 1A**) and nine constructs gave localizations out of organelles, appearing as typical nuclear and cytosolic signals (AT1G06150 and AT1G06580 in **Fig. 1A**). These localization results in the nucleus and the cytosol of the protoplasts suggest that the first 100 amino acids of these proteins do not code for a functional peptide targeting to organelles and that the RFP fusion proteins are localized where the translation occurs (in the cytosol) and in the nucleus by passive diffusion of small proteins through nuclear pores.

As addressing signals could be outside the first 100 amino acids and because using the first 100 amino acids may induce addressing artifacts, we decided to investigate in more detail the subcellular localization of the 33 PPR proteins that did not show a simple single organellar localization. Out of the 24 PPR proteins localized in both organelles and the nine proteins appearing outside of the organelles, we successfully cloned the whole ORFs and created RFP fusions for 19. Subcellular localizations of these fusions were monitored using Agro-infiltrated *N. benthamiana*-derived protoplasts observed under epifluorescent and confocal microscope (examples in **Fig. 1B**). Results are summarized in **Table 1**. We confirmed the dual subcellular localization for six out of the 11 ORFs successfully expressed and encoding full-length proteins thought to be addressed in both organelles (see AT5G47460 in **Fig. 1B** for example). As for AT2G36240 (**Fig. 1B**), we showed a single localization in mitochondria for the other five. Among the nine PPR proteins thought to be out of the organelles on the base of the first 100 amino acids, seven ORFs were successfully cloned but no cytosolic localization was confirmed: the whole proteins fused to RFP were systematically addressed to one or both organelles, as exemplified by AT1G06150 and AT1G06580 in **Figure 1B**. Surprisingly, six of them were localized in both organelles (AT1G06150 in **Fig. 1B**). Altogether, 12 PPRs proteins were verified as being localized in both mitochondria and plastids using the full-length protein.

Integrative overview of the subcellular localizations of PPR protein family. In order to provide a general overview of the localization of the whole PPR protein family, we aggregated our results concerning one-third of the family, with all available data from published studies and accessible databases (**Table 2**).

We first re-examined the localization prediction of the 458 PPR proteins encoded in the Arabidopsis genome using six available bio-informatics prediction tools: Predotar,[32] TargetP,[31] iPSORT,[38] Loctree,[39] Multiloc,[40] and AtSubP.[41] Despite using distinct algorithms, those tools largely provide similar results and **Table 2** gives a single localization prediction aggregating the six software results following a rule emphasized in the caption of **Table 2**. Overall, 65% and 17% of the PPR proteins are predicted to function in mitochondria and in plastids, respectively. For 18%, the results are unclear either because a majority of the software was unable to provide an organellar prediction or because they provide overmuch diverging organellar predictions.

We also added the growing data coming from the proteomics identification of organelle proteins in Arabidopsis mitochondria (SUBA3[5]) and chloroplast (SUBA3,[5] AT_Chloro,[42] PPDB[43]), also including localization data obtained from maize chloroplast (PPDB[43]) and rice mitochondria,[44] according to the recent concept of orthoproteomics.[45] As published,[12] a very good level of orthology observed between the members of PPR families in *A. thaliana* and *Oriza sativa* suggests that both their function and their

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data

PPR	Gene model		Prediction		Fluorescent signal		Conclusion
	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	
At1g01970	AT1G01970.1		M	none	M	C	C
At1g02420	AT1G02420.1		ER	M	M/C	c.u.	m/c
At1g04840	AT1G04840.1		none	M	C	n.a.	C
At1g05670		AtPPR_1g05670	M	none	N/Ct	M/C	M/C
At1g06150		AtPPR_1g06150	C	none	N/Ct	M/C	M/C
At1g06580	AT1G06580.1		M	ER	N/Ct + M	M/C	M/C
At1g08610	AT1G08610.1		none	C	no signal	n.a.	-
At1g09190	AT1G09190.1		M	none	M	n.a.	M
At1g09410	AT1G09410.1		M	M	no signal	n.a.	pM
At1g09900	AT1G09900.1		none	C	M	n.a.	M
At1g10270	AT1G10270.1		M	M	M	n.a.	M
At1g10330	AT1G10330.1		ER	none	no signal	n.a.	-
At1g11290	AT1G11290.1		C	none	C	n.a.	C
At1g14470	AT1G14470.1		ER	none	C	n.a.	C
At1g15480		AtPPR_1g15480	M	ER	no signal	n.a.	-
At1g18485	AT1G18485.1		C	ER	N/Ct	c.u.	-
At1g19290	AT1G19290.1		M	ER	M	n.a.	M
At1g19720	AT1G19720.1		none	none	M	n.a.	M
At1g20230	AT1G20230.1		none	none	M	n.a.	M
At1g22830	AT1G22830.1		none	M	M	n.a.	M
At1g25360	AT1G25360.1		M	none	M/C	M/C	M/C
At1g31430	AT1G31430.1		ER	none	M	n.a.	M
At1g31790	AT1G31790.1		C	none	C	n.a.	C
At1g31840		AtPPR_1g31840	ER	none	M	n.a.	M
At1g31920	AT1G31920.1		none	none	no signal	n.a.	-
At1g33350	AT1G33350.1		none	C	M	n.a.	M
At1g50270	AT1G50270.1		none	none	M	n.a.	M
At1g53330	AT1G53330.1		M	none	c.u.	n.a.	-
At1g56570	AT1G56570.1		C	none	M	n.a.	M
At1g59720	AT1G59720.1		C	none	C	n.a.	C
At1g60770	AT1G60770.1		none	M	M	n.a.	M
At1g62260	AT1G62260.1		ER	M	M	n.a.	M
At1g62590	AT1G62590.1		M	none	M/C	M	M
At1g63330	AT1G63330.1		none	none	M	n.a.	M
At1g63400	AT1G63400.1		M	none	M/C	M	M
At1g64100		AtPPR_1g64100	C	none	M/C	c.u.	m/c
At1g68930	AT1G68930.1		M	none	M	n.a.	M
At1g69290		AtPPR_1g69290	ER	ER	M	n.a.	M
At1g71490	AT1G71490.1		none	C	M	n.a.	M
At1g73710	AT1G73710.1		M	none	C	n.a.	C
At1g74400	AT1G74400.1		none	M	M/C	M	M
At1g74580	AT1G74580.1		none	none	no signal	n.a.	-
At1g74630	AT1G74630.1		ER	C	no signal	n.a.	-
At1g76280		AtPPR_1g76280	M	ER	M	n.a.	M
At2g01360		AtPPR_2g01360	ER	ER	no signal	n.a.	-
At2g01740	AT2G01740.1		none	M	M	n.a.	M
At2g02750	AT2G02750.1		M	M	no signal	n.a.	pM
At2g04860	AT2G04860.1		none	M	M	n.a.	M
At2g06000	AT2G06000.1		none	M	C	n.a.	C
At2g13600	AT2G13600.1		none	none	M	n.a.	M
At2g15820	AT2G15820.1		C	none	C	n.a.	C
At2g15980	AT2G15980.1		none	none	no signal	n.a.	-
At2g16880	AT2G16880.1		none	none	no signal	n.a.	-
At2g20540	AT2G20540.1		none	none	M	n.a.	M
At2g21090	AT2G21090.1		M	none	M	n.a.	M
At2g22070	AT2G22070.1		none	C	M	n.a.	M
At2g26790	AT2G26790.1		M	ER	M	n.a.	M
At2g27610	AT2G27610.1		M	none	M	n.a.	M
At2g28050	AT2G28050.1		none	M	M	n.a.	M
At2g32630	AT2G32630.1		M	none	M	n.a.	M
At2g33680	AT2G33680.1		none	none	c.u.	n.a.	-
At2g33760	AT2G33760.1		none	C	C	n.a.	C
At2g34400	AT2G34400.1		none	M	M/C	c.u.	m/c
At2g35130		AtPPR_2g35130	ER	none	C	n.a.	C
At2g36240	AT2G36240.1		none	none	M/C	M	M
At2g36730	AT2G36730.1		none	none	no signal	n.a.	-
At2g37230	AT2G37230.1		M	none	M	n.a.	M
At2g37310	AT2G37310.1		none	none	C	n.a.	C
At2g39620		AtPPR_2g39620	M	ER	C	n.a.	C
At2g40720	AT2G40720.1		none	none	M	n.a.	M
At2g41080		AtPPR_2g41080	C	none	M/C	no signal	m/c

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data (continued)

	Gene model		Prediction		Fluorescent signal		
PPR	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	Conclusion
At2g41720	AT2G41720.1		none	C	C	n.a.	C
At2g44880	AT2G44880.1		ER	none	N/Ct	M/C	M/C
At2g45350	AT2G45350.1		none	C	c.u.	n.a.	-
At3g01580	AT3G01580.1		M	none	M	n.a.	M
At3g02010	AT3G02010.1		M	none	M	n.a.	M
At3g05240	AT3G05240.1		none	M	no signal	n.a.	-
At3g06920	AT3G06920.1		none	none	M	n.a.	M
At3g08820	AT3G08820.1		none	C	M/C	M	M
At3g09060	AT3G09060.1		none	M	M	n.a.	M
At3g09650	AT3G09650.1		C	none	C	n.a.	C
At3g12770	AT3G12770.1		none	none	M	n.a.	M
At3g14330	AT3G14330.1		none	M	M	n.a.	M
At3g15130	AT3G15130.1		none	M	M	n.a.	M
At3g15930	AT3G15930.1		none	M	M	n.a.	M
At3g16610	AT3G16610.1		M	none	M	n.a.	M
At3g18840	AT3G18840.2		C	none	N/Ct	c.u.	-
At3g20730		AtPPR_3g20730	ER	ER	M	n.a.	M
At3g21470	AT3G21470.1		none	none	M/C	c.u.	m/c
At3g23020	AT3G23020.1		none	C	c.u.	n.a.	-
At3g23330	AT3G23330.1		none	M	M/C	c.u.	m/c
At3g25970	AT3G25970.1		ER	none	M	n.a.	M
At3g26540	AT3G26540.1		none	C	M	n.a.	M
At3g28640	AT3G28640.1		M	none	no signal	n.a.	-
At3g28660	AT3G28660.1		M	none	no signal	n.a.	-
At3g29290		AtPPR_3g29290	M	none	no signal	n.a.	-
At3g42630	AT3G42630.1		none	M	C	n.a.	C
At3g46610	AT3G46610.1		none	ER	C	n.a.	C
At3g46790	AT3G46790.1		C	none	C	n.a.	C
At3g47530	AT3G47530.1		none	none	M/C	M/C	M/C
At3g47840	AT3G47840.1		M	none	C	n.a.	C
At3g48810	AT3G48810.1		ER	none	M	n.a.	M
At3g49240	AT3G49240.1		M	none	M/C	M/C	M/C
At3g49710	AT3G49710.1		none	none	N/Ct	M/C	M/C
At3g49740	AT3G49740.1		C	none	M	n.a.	M
At3g50420	AT3G50420.1		none	none	M/C	M/C	M/C
At3g53170		AtPPR_3g53170	none	none	C	n.a.	C
At3g56550	AT3G56550.1		none	C	no signal	n.a.	-
At3g57430	AT3G57430.1		C	ER	C	n.a.	C
At3g58590	AT3G58590.1		none	none	M	n.a.	M
At3g62890		AtPPR_3g62890	C	none	N/Ct	M/C	M/C
At4g01570	AT4G01570.1		C	none	C	n.a.	C
At4g02750	AT4G02750.1		M	none	M	n.a.	M
At4g04370	AT4G04370.1		none	C	M/C	c.u.	m/c
At4g08210	AT4G08210.1		none	none	M	n.a.	M
At4g11690	AT4G11690.1		ER	ER	M	n.a.	M
At4g13650	AT4G13650.1		none	M	no signal	n.a.	pM
At4g14820	AT4G14820.1		none	C	M	n.a.	M
At4g14850	AT4G14850.1		C	none	M/C	c.u.	m/c
At4g15720	AT4G15720.1		none	none	C	n.a.	C
At4g16470	AT4G16470.1		ER	M	M	n.a.	M
At4g18840		AtPPR_4g18840	none	C	N/Ct	C	C
At4g20090	AT4G20090.1		C	ER	M	n.a.	M
At4g20740	AT4G20740.1		C	none	C	n.a.	C
At4g21065	AT4G21065.1		ER	none	no signal	n.a.	-
At4g21880	AT4G21880.1		ER	M	M	n.a.	M
At4g22760	AT4G22760.1		ER	M	M	n.a.	M
At4g28010	AT4G28010.1		M	none	M	n.a.	M
At4g30700	AT4G30700.1		none	M	M/C	M/C	M/C
At4g33170	AT4G33170.1		C	none	M/C	c.u.	m/c
At4g37170	AT4G37170.1		M	none	M	n.a.	M
At4g38010	AT4G38010.1		none	none	no signal	n.a.	-
At5g03800	AT5G03800.1		C	none	M/C	c.u.	m/c
At5g04810	AT5G04810.1		C	none	C	n.a.	C
At5g06540	AT5G06540.1		ER	M	C	n.a.	C
At5g08310		AtPPR_5g08310	none	M	no signal	n.a.	-
At5g08490	AT5G08490.1		M	none	no signal	n.a.	-
At5g08510	AT5G08510.1		none	none	M	n.a.	M
At5g10690	AT5G10690.1		C	ER	C	n.a.	C
At5g14080		AtPPR_5g14080	M	none	M/C	c.u.	m/c
At5g15300	AT5G15300.1		none	none	no signal	n.a.	-
At5g16860	AT5G16860.1		none	M	no signal	n.a.	-

Table 1. Subcellular localization study of 166 PPR proteins with ambiguous prediction data (continued)

PPR	Gene model		Prediction		Fluorescent signal		Conclusion
	TAIR v10	O'Toole	Target P	Predotar	Targeting Peptide	FL Protein	
At5g18475	AT5G18475.1		C	none	M	n.a.	M
At5g18950	AT5G18950.1		M	none	M	n.a.	M
At5g21222	AT5G21222.1		none	none	M/C	c.u.	m/c
At5g25630		AtPPR_5g25630	none	none	no signal	n.a.	-
At5g27270	AT5G27270.1		C	none	C	n.a.	C
At5g37570	AT5G37570.1		none	M	no signal	n.a.	-
At5g38730	AT5G38730.1		none	C	no signal	n.a.	-
At5g39680	AT5G39680.1		none	none	M	n.a.	M
At5g40405	AT5G40405.1		none	none	M	n.a.	M
At5g43790	AT5G43790.1		none	none	M/C	c.u.	m/c
At5g46100	AT5G46100.1		none	M	no signal	n.a.	-
At5g46680	AT5G46680.1		none	M	M	n.a.	M
At5g47460	AT5G47460.1		M	none	M/C	M/C	M/C
At5g48910	AT5G48910.1		C	none	C	n.a.	C
At5g50990		AtPPR_5g50990	none	none	no signal	n.a.	-
At5g52630	AT5G52630.1		none	none	M	n.a.	M
At5g55840		AtPPR_5g55840	none	none	M	n.a.	M
At5g56310	AT5G56310.1		M	none	M	n.a.	M
At5g59600	AT5G59600.1		none	none	C	n.a.	C
At5g59900	AT5G59900.1		M	none	M	n.a.	M
At5g65570	AT5G65570.1		M	none	no signal	n.a.	-
At5g65820	AT5G65820.1		M	none	M	n.a.	M
At5g66520	AT5G66520.1		none	none	C	n.a.	C
At5g67570		AtPPR_5g67570	none	none	no signal	n.a.	-

Manually curated Arabidopsis PPR gene models were used.[12] Most of them are identical to TAIR v10 gene models but 22 models are different and are indicated with their AtPPR codes. Predictions of localization using Predotar v1.03 and Target P v1.1 software are listed. Experimental fluorescent signals observed in protoplasts expressing Targeting Peptide or Full-Length (FL) protein fused to RFP are shown. Two independent observations by two of the authors were done on at least three independent agro-infiltrations. For each PPR, a tentative conclusion is proposed with the following rules: (1) if available, the observation of FL-protein fusion is considered as the true localization, (2) if a mitochondrial or a chloroplastic localization was observed for the targeting peptide and no observation was recorded for the full-length protein, the result of TP is indicated as conclusion, (3) if a dual localization was observed and no observation was obtained with the full-length protein, the result of TP is indicated as probable in lowercase, (4) if no experimental observation was obtained, the predicted localization is indicated with a preceding "p"., . M, mitochondria; C, chloroplasts; N/Ct, nucleus and cytoplasm; M/C, dual localization in mitochondria and chloroplasts; pM, predicted in mitochondria (conclusion column); pC, predicted in chloroplasts (conclusion column); m, probably in mitochondria (conclusion column); c, probably in chloroplasts (conclusion column); m/c, probably in mitochondria and chloroplasts (conclusion column); -, no conclusion; c.u., cloning unsuccessful; n.a., not attempted.

subcellular localization are largely conserved between species even between monocotyledons and dicotyledons. Overall, 83 (about 18%) of the Arabidopsis PPR proteins, or PPR orthologs in other species, were identified either in the plastidial or the mitochondrial proteomes, providing a very useful set of PPR protein localization data (**Table 2**). Three and five PPR proteins were identified during proteomics characterization of Arabidopsis nuclear and cytosolic proteins, respectively.[46-49] Surprisingly, 28 PPR proteins were characterized in plasma membrane or vacuole extracts.[50-56] Without functional characterization of any of these membrane PPR proteins, these observations cannot be solved. They may be due to intrinsic technical limitations of proteomics approaches; in contrast, their number may indicate unsuspected localizations and functions. However, proteome-based localizations validate many of the prediction results of bio-informatics software as 48 (71%) of them matched the available predictions (**Table 3**).

A growing number of PPR proteins were subjected to in planta functional characterization either in dedicated studies (see references in **Table 2**) or in systematic studies including the work reported here and three previous ones[11,57] unpublished data in SUBA3 (**Table 2**). Authors usually characterized localizations by microscopy using fusions between PPR proteins or, if suspected, putative targeting peptides and a fluorescent reporter. Including the work reported here, 208 PPR localizations were experimentally determined using fluorescent fusion proteins, largely correlating with both bio-informatics and proteomics approaches (**Table 2 and 3**). Among the 159 PPRs proteins for which both experimental localization data based on protein fusion and predictions using bio-informatics tools are available, 135 (85%) have a similar localization. In addition, among the 36 PPR proteins being both identified in sub-proteomes and subjected to experimental localization studies using fluorescent protein fusion, 30 (83%) were compatible. The last set of data comes from the identification of the molecular functions of PPR proteins using reverse genetics, providing very important data about their localization (**Table 2**). As largely reported in the literature, PPR proteins are involved in regulating gene expression by acting through direct interaction with specific RNAs. A literature survey indicates that molecular roles were assigned to 68 PPR proteins (**Table 2**), occurring in plastids (31), in mitochondria (34), in both mitochondria and plastids (1), or in the nucleus (2). These reverse genetics studies are very strong statements of PPR localization, which could be considered as true localization. When compared with this very high quality data set, our data as well as all data of fluorescent protein localization appeared as very highly correlated

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins

AGI	Gene Annotation[1]	Domains 2	Localization				EMB[7]	Molecular Function (localization)[8]	References
			Predictions 3	Proteomics 4	Experimental[5]	Conclusion 6			
At1g01970	PPR containing protein	P	M		C[a]	c			[a]this report
At1g02060	PPR containing protein	P	M			pM			
At1g02150	PPR containing protein	P	M	C (At[a,b,c], Zm[c])		C			[a]AT_Chloro, [b]Kong et al 2011, [c]PPDB
At1g02370	PPR containing protein	P	M	M (At[a])		M			[a]Klodmann et al 2011
At1g02420	PPR containing protein	P	M		m/c[a]	m/c			[a]this report
At1g03100	PPR containing protein	P	M	Ct (At[a])		pM			[a]Hummel et al 2012
At1g03510	PPR containing protein	PLS-E	C			pC			
At1g03540	PPR containing protein	PLS-E	M			pM			
At1g03560	PPR containing protein	P	M			pM			
At1g04840	PPR containing protein	PLS-E-DYW	m	PM (At[a])	C[b]	-			[a]Mitra et al 2009, [b]this report
At1g05600	EMB3101	P	M			pM	confirmed[a]		[a]SeedGenes
At1g05670	PPR containing protein	P	M		M/C[a]	m/c			[a]this report
At1g05750	CLB19/PDE247	PLS-E	c		C[a,b]	C		Editing *rpoA* and *clpP* (C[a])	[a]Chateigner-Boutin et al 2008, [b]in house SUBA3
At1g06140	MEF3	PLS-E	M			M		Editing *atp4* (M[a])	[a]Verbistkiy et al 2012
At1g06150	EMB1444	PLS-E	C		M/C[a]	m/c	potential[b]		[a]this report, [b]Cushing et al 2005
At1g06270	PPR containing protein	P	M		ER/C[a]	-			[a]Narsai et al 2011
At1g06580	PPR containing protein	P	M		M/C[a]	m/c			[a]this report
At1g06710	MTSF1	P	m		M[a]	M		Processing and stability of *nad4* (M[a])	[a]Haïli et al 2013
At1g07590	PPR containing protein	P	M			pM			
At1g07740	PPR containing protein	P	M			pM			
At1g08070	OTP82	PLS-E-DYW	C		C[a]	C		Editing *ndhG, ndhB* (C[b,c])	[a]in house SUBA3, [b]Hammani et al 2009, [c]Okuda et al 2010
At1g08610	PPR containing protein	P	C			pC			
At1g09190	PPR containing protein	PLS-E	-		M[a]	m			[a]this report
At1g09220	PPR containing protein	PLS-E	M			pM			
At1g09410	PPR containing protein	PLS-E-DYW	M			c			
At1g09680	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At1g09820	PPR containing protein	P	M			pM			
At1g09900	PPR containing protein	P	c	C (Zm[a])	M[b]	m/c			[a]PPDB, [b]this report
At1g10270	GRP23	P	M		N[a], M[b,c]	M/N	confirmed[d]		[a]Ding et al 2006, [b]Narsai et al 2011, [c]this report, [d]SeedGenes
At1g10330	PPR containing protein	PLS-E	m			pM			
At1g10910	EMB3103	P	M	C (Zm[a])		c	confirmed[b]		[a]PPDB, [b]SeedGenes
At1g11290	CRR22	PLS-E-DYW	c		C[a,b]	C		Editing *ndhB, ndhD, rpoB* (C[c])	[a]this report, [b]in house SUBA3, [c]Okuda et al 2009
At1g11630	PPR containing protein	P	M	M (At[a])		M			[a]Heazlewood et al 2004
At1g11710	PPR containing protein	P	M			pM			
At1g11900	PPR containing protein	P	M			pM			
At1g12300	PPR containing protein	P	M			pM			
At1g12620	PPR containing protein	P	M			pM			
At1g12700	RPF1	P	M		M[a]	M		Processing *nad4* transcript (M[a])	[a]Holze et al 2011
At1g12775	EMB1586	P	M			pM	confirmed[a]		[a]SeedGenes
At1g13040	PPR containing protein	P	M	V (At[a])		pM			[a]Jaquinod et al 2007
At1g13410	PPR containing protein	PLS-E	m			pM			
At1g13630.1	PPR containing protein	P	-			-			
At1g13800	FAC19	P	M			pM	confirmed[a]		[a]Yu et al J 2011
At1g14470	PPR containing protein	PLS	m		C[a]	c			[a]this report
At1g15480	PPR containing protein	P	m	M (At[a]) PM (At[b])		M			[a]Klodmann et al 2011, [b]Mitra et al 2009
At1g15510	AtECB2 /VAC1	PLS-E-DYW	M		C[a,b,c]	C		Editing *accD* and *ndhF* (C[b,c])	[a]in house SUBA3, [b]Yu et al 2009, [c]Tseng et al 2010
At1g16480	pseudogene	PLS-E-DYW	M			pM			
At1g16830	PPR containing protein	P	M			pM			
At1g17630	PPR containing protein	PLS-E	M			pM			
At1g18485	PPR containing protein	PLS-E-DYW	C			pC			
At1g18900	PPR containing protein	P-D	M			pM			
At1g19290	PPR containing protein	P	m		M[ab]	M			[a]this report, [b]in house SUBA3
At1g19520	NFD5	P	M	PM (At[a])		pM	potential[b]		[a]Zhang et al 2011, [b]Portereiko et al 2006
At1g19720	PPR containing protein	PLS-E-DYW	-	C (At[a,b], Zm[b])	M[c]	m/c			[a]Kong et al 2011, [b]PPDB, [c]this report
At1g20230	PPR containing protein	PLS-E-DYW	-		M[a]	m			[a]this report
At1g20300	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At1g22830	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At1g22960	PPR containing protein	P	M			pM			
At1g25360	PPR containing protein	PLS-E-DYW	M		M/C[a]	m/c			[a]this report
At1g26460	PPR containing protein	P	M	M (At[a,b], Os[c]) PM (At[d])		M			[a]Heazlewood et 2004, [b]Klodmann et al 2011, [c]Huang et al 2009, [d]Zhang et al 2011
At1g26500	PPR containing protein	P	M			pM			
At1g26900	PPR containing protein	PLS-E	M			pM			
At1g28020	PPR containing protein	P	M			pM			
At1g28690	PPR containing protein	PLS-E	M			pM			
At1g29710	PPR containing protein	PLS-E-DYW	M			pM			
At1g30290	pseudogene	P	M			pM			
At1g30610	EMB2279	P	C	C (Zm[a])		C	confirmed[b]		[a]PPDB, [b]SeedGenes
At1g31430	PPR containing protein	PLS-E	c		M[a]	m			[a]this report
At1g31790	PPR containing protein	PLS	c		C[a]	C			[a]this report
At1g31840	PPR containing protein	P	-		M[a]	m			[a]this report
At1g31920	PPR containing protein	PLS-E-DYW	-	C (Zm[a])		c			[a]PPDB
At1g32415	PPR containing protein	PLS-E	M			pM			
At1g33350	PPR containing protein	PLS-E	m		M[a]	M			[a]this report
At1g34160	OGR1	PLS-E-DYW	M		M[a]	M		Editing *nad4, nad2, ccmC, cox2, cox3* (M[a])	[a]Kim et al 2009
At1g43010	PPR containing protein	P	M			pM			
At1g43980	PPR containing protein	PLS-E	M			pM			
At1g47580	DYW1	PLS-E-DYW	c		C[a,b]	C		Editing *ndhD* (C[a])	[a]Boussardon et al 2012, [b]in house SUBA3
At1g50270	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At1g51965	ABO5	P	M	C (At[a])	M[b]	M		Splicing *nad2* intron3 (M[b])	[a]AT_Chloro,[b]Liu et al 2010
At1g52620	PPR containing protein	P	M	C (At[a])		c			[a]PPDB
At1g52640	PPR containing protein	P	M			pM			
At1g53330	CB_1265	P	M			pM	confirmed[a]		[a]Kocàbek et al 2006
At1g53600	PPR containing protein	PLS-E	M			pM			
At1g55630	PPR containing protein	P	M			pM			
At1g55890	PPR containing protein	P	M	M (At[a,b]) PM (At[c])		M			[a]Heazlewood et 2004, [b]Klodmann et al 2011, [c]Mitra et al 2009
At1g56570	PGN	PLS-E	-		M[a,b]	M			[a]Laluk et al 2011, [b]this report

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation[1]	Domains [2]	Localization Predictions [3]	Localization Proteomics [4]	Localization Experimental[5]	Localization Conclusion [6]	EMB[7]	Molecular Function (localization)[8]	References
At1g56690	PPR containing protein	PLS-E-DYW	M			pM			
At1g59720	CRR28	PLS-E-DYW	c		M[a],C[b,c]	C		Editing *ndhB, ndhD* (C[d])	[a]Lurin et al 2004, [b]in house SUBA3, [c]this report, [d]Okuda et al 2009
At1g60770	PPR containing protein	P	m	M (At[a,b], Os[c]) PM (At[d])	M[e]	M			[a]Heazlewood et 2004, [b]Klodmann et al 2011, [c]Huang et al 2009, [d]Mitra et al 2009, [e]this report
At1g61870	PPR336	P	M	M (At[a,b,c])	M[d]	M			[a]Heazlewood et al 2004, [b]Uyttewaal 2007, [c]Klodmann et al 2011, [d]Lurin et al 2004
At1g62260	MEF9	PLS-E	-		M[a]	M		Editing *nad7* (M[b])	[a]this report, [b]Takenaka et al 2010
At1g62350	THA8-LIKE3	P-D	M			pM			
At1g62590	PPR containing protein	P	m		M[a]	M			[a]this report
At1g62670	RPF2	P	M		M[a]	M		Processing *nad9* & *cox3* (M[a])	[a]Jonietz et al 2010
At1g62680	PPR containing protein	P	M			pM			
At1g62720	AtNG1	P	-		M[a]	M			[a]Yang et al 2011
At1g62910	PPR containing protein	P	M	PM (At[a])		pM			[a]Mitra et al 2009
At1g62930	RPF3	P	C		M[a]	M		Processing *ccmC* (M[a])	[a]Jonietz et al 2011
At1g63070	PPR containing protein	P	M			pM			
At1g63080	PPR containing protein	P	M			pM			
At1g63130	PPR containing protein	P	M			pM			
At1g63150	PPR containing protein	P	M			pM			
At1g63320	PPR containing protein	P	m			pM			
At1g63330	PPR containing protein	P	-		M[a]	m			[a]this report
At1g63400	PPR containing protein	P	-		M[a]	m			[a]this report
At1g64100	PPR containing protein	P	-	M (At[a])	m/c[b]	M/c			[a]Klodmann et al 2011, [b]this report
At1g64310	OTP71	PLS-E	m		M[a], M/C[b]	M/c		Editing *ccmFN2* (M[a])	[a]Chateigner-Boutin et al 2013, [b]in house SUBA3
At1g64430	PPR containing protein	P	C			pC			
At1g64580	PPR containing protein	P	M			pM			
At1g66345	PPR containing protein	P	M			pM			
At1g68930	PPR containing protein	PLS-E-DYW	m		M[a]	M			[a]this report
At1g68980	PPR containing protein	P	M			pM			
At1g69290	PPR containing protein	P	M	PM (At[a])	M[b]	M			[a]Li et al 2012, [b]this report
At1g69350	PPR containing protein	PLS-E	M		M[a]	M			[a]in house SUBA3
At1g71060	PPR containing protein	P	M			pM			
At1g71210	PPR containing protein	P	M			pM			
At1g71420	PPR containing protein	PLS-E-DYW	M	Ct (At[a])		c			[a]Hummel et al 2012
At1g71460	PPR containing protein	PLS	C	C (At[a], Zm[a]) Ct(At[b])		C			[a]PPDB, [b]Hummel et al 2012
At1g71490	PPR containing protein	PLS-E	-		M[a]	m			[a]this report
At1g73400	PPR containing protein	P	M			pM			
At1g73710	PPR containing protein	P	-		C[a]	c			[a]this report
At1g74400	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At1g74580	PPR containing protein	P	-			-			
At1g74600	OTP87 = OsPPR1	PLS-E	C		M/C[a]	M/c		Editing *nad7, atp1* (M[a])	[a]Hammani et al 2011
At1g74630	PPR containing protein	PLS-E-DYW	-			-			
At1g74750	PPR containing protein	P-D	C			pC			
At1g74850	PTAC2	P-D	C	C (At[a,b,c,d,e,f], Zm[f]) N (At[g])		C	confirmed[c]		[a]Kleffmann et al 2004, [b]AT_Chloro, [c]Pfalz et al 2006, [d]Kong et al 2011, [e]Ingelsson et al 2012, [f]PPDB, [g]Sakamoto et al 2013
At1g74900	OTP43	P	M		M[a]	M		Splicing *nad1* intron1 (M[b])	[a]in house SUBA3, [b]de longevialle et al 2007
At1g76280.1	PPR containing protein	P	M		M[a]	M			[a]this report
At1g77010	PPR containing protein	PLS-E	M			pM			
At1g77170	PPR containing protein	PLS-E	C			pC			
At1g77340	PPR containing protein	P	m			pM			
At1g77360	APPR6	P	M		M[a]	M		Processing and translation stabilisation of *rps3* (M[a])	[a]Manavski et al 2012
At1g77405	PPR containing protein	P	C			pC			
At1g79080	PPR containing protein	P	C			pC			
At1g79490	EMB2217	P-D	M		M[a]	M	potential[b]		[a]Narsai et al 2011, [b]SeedGenes
At1g79540	PPR containing protein	P	M			pM			
At1g80150	PPR containing protein	P	M			pM			
At1g80270.1	PPR_596	P	M	M(At[a], Os[b]) C (At[c]) PM (At[d])	M[e]	M/c			[a]Klodmann et al 2011, [b]Huang et al 2009, [c]Froehlich et al 2003, [d]Zhang et al 2011, [e]Narsai et al 2011
At1g80550	PPR containing protein	P	M			pM			
At1g80880	PPR containing protein	P	M			pM			
At2g01360	PPR containing protein	P	-			-			
At2g01390	PPR containing protein	P	M		M[a]	M	confirmed[b]		[a]Lurin et al 2004, [b]SeedGenes
At2g01510	PPR containing protein	PLS-E-DYW	M			pM			
At2g01740	PPR containing protein	P	M		M[a]	M			[a]this report
At2g01860	EMB975	P	-		C[a]	c	confirmed[b]		[a]Lurin et al 2004, [b]SeedGenes
At2g02150	PPR containing protein	P	C			pC			
At2g02750	PPR containing protein	PLS-E	M			pM			
At2g02980	OTP85	PLS-E-DYW	M		C[a]	C		Editing *ndhD* (C[b])	[a]in house SUBA3, [b]Hammani et al 2009
At2g03380	PPR containing protein	PLS-E	M			pM			
At2g03880	REME1	PLS-E-DYW	M	PM (At[a])	M[b], C[c]	M		Editing *nad2, mttB* (M[b])	[a]Alexandersson et al 2004, [b]Bentolila et al 2010, [c]Lurin et al 2004
At2g04860	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At2g06000.1	PPR containing protein	P	M		C[a]	c			[a]this report
At2g13420	PPR containing protein	P	M			pM			
At2g13600	SLO2	PLS-E	-		M[a,b]	M		Editing *mttB, nad1, nad4L, nad7, nad1* (M[a])	[a]Zhu et al 2012, [b]this report
At2g15630	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At2g15690	PPR containing protein	PLS-E-DYW	M	M (Os[a]) C(Zm[b])	M/C[c]	M/C			[a]Huang et al 2009, [b]PPDB, [c]in house SUBA3
At2g15820	OTP51	P	M	C (Zm[a])	C[b]	C		Splicing *ycf3* intron2 (C[c])	[a]PPDB, [b]this report, [c]de Longevialle et al 2008
At2g15980	PPR containing protein	P	-			-			
At2g16650	PRORP2	P-D	-		N[a]	N		Processing tRNA and maturation of RNA (N[b])	[a]Gobert et al 2010, [b]Gutmann et al 2012
At2g16880	PPR containing protein	P	m			pM			
At2g17033.1	PPR containing protein	P-D	M	C (Zm[a])		c			[a]PPDB
At2g17140	PPR containing protein	P	m			pM			
At2g17210	PPR containing protein	PLS-E	-			-			
At2g17525	PPR containing protein	P	M			pM			
At2g17670.1	PPR containing protein	P	M			pM			
At2g18520	PPR containing protein	PLS	M	M (At[a])		M			[a]Klodmann et al 2011
At2g18940	ZmPPR10	P	C	C (Zm[a])		C		Translation stabilisation of *atpI-AtpH* & *psaJ-rpl33* (C[a])	[a]Pfalz et al 2009

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation[1]	Domains[2]	Localization				EMB[7]	Molecular Function (localization)[8]	References
			Predictions[3]	Proteomics[4]	Experimental[5]	Conclusion[6]			
At2g19280	PPR containing protein	P	M	C (At[a])	M[a]	M/c			[a]PPDB, [b]Lurin et al 2004
At2g20540	MEF21	PLS-E	-		M[a,b]	M		Editing *cox3* (M[c])	[a]Lurin et al 2004, [b]this report, [c]Takenaka et al 2010
At2g20710.1	PPR containing protein	P	M			pM			
At2g21090	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At2g22070	PPR containing protein	PLS-E-DYW	C		M[a]	m			[a]this report
At2g22410	SLO1	PLS-E	M		M[a]	M		Editing *nad4, nad9* (M[a])	[a]Sung et al 2010
At2g25580	MEF8	PLS-E-DYW	M			M		Editing *nad5, nad6* (M[a,b])	[a]Takenaka et al 2010, [b]Vervitskiy et al 2012
At2g26790	PPR containing protein	P	M		M[a]	M			[a]this report
At2g27610	PPR containing protein	PLS-E-DYW	-	Ct (At[a])	M[b]	M			[a]Hummel et al 2012, [b]this report
At2g27800	PPR containing protein	P	M			pM			
At2g28050	PPR containing protein	P	M	C (At[a])	M[b]	M/c			[a]AT_Chloro, [b]this report
At2g29760	OTP81	PLS-E-DYW	C		C[a,b]	C		Editing *rps12* (C[c])	[a]Lurin et al 2004, [b]in house SUBA3, [c]Hammani et al 2009
At2g30100	PPR containing protein	P	C			pC			
At2g30780	PPR containing protein	P	M			pM			
At2g31400	GUN1	P-D	C		C[a]	C			[a]Koussevitzky et al 2007
At2g32230	PRORP1	P-D	M	C (Zm[a])	M/C[b]	M/C	confirmed[b]	Processing tRNA elements (M/C[b])	[a]PPDB, [b]Gobert et al 2010
At2g32630	PPR containing protein	P	M		M[a]	M			[a]this report
At2g33680	PPR containing protein	PLS-E	-			-			
At2g33760	PPR containing protein	PLS-E-DYW	c		C[a,b]	C			[a]this report, [b]in house SUBA3
At2g34370	PPR containing protein	PLS-E-DYW	M		M[a]	M			[a]Lurin et al 2004
At2g34400	PPR containing protein	PLS-E	M		m/c[a]	m/c			[a]this report
At2g35030	PPR containing protein	PLS-E	M		M[a]	M			[a]in house SUBA3
At2g35130	PPR containing protein	P	-	C (Zm[a])	C[b]	c			[a]PPDB, [b]this report
At2g36240	PPR containing protein	P	m		M[a]	M			[a]this report
At2g36730	PPR containing protein	PLS-E	m			pM			
At2g36980	PPR containing protein	PLS-E	M			pM			
At2g37230	PPR containing protein	P	M	C (At[a,b]) M(Os[c]) PM (At[d])	M[e]	M/c			[a]AT_Chloro, [b]PPDB, [c]Huang et al 2009, [d]Zhang et al 2011, [e]this report
At2g37310	PPR containing protein	PLS-E	M		C[a]	c			[a]this report
At2g37320	PPR containing protein	PLS-E	M			pM			
At2g38420	PPR containing protein	P	M			pM			
At2g39230	LOJ	P	M			pM			
At2g39620	PPR containing protein	PLS-E	M		C[a]	c			[a]this report
At2g40240	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At2g40720	PPR containing protein	PLS-E	m		M[a]	M			[a]this report
At2g41080	PPR containing protein	PLS-E-DYW	-		m/c[a]	m/c			[a]this report
At2g41720.1	EMB2654	P	-		C[a]	c	potential[b]		[a]this report, [b]SeedGenes
At2g42920	PPR containing protein	PLS-E	C			pC			
At2g44880	AHG11	PLS-E	-	PM (At[a])	M[b], M/C[c]	M/c		Editing *nad4* (M[b])	[a]Mitra et al 2009, [b]Murayama et al 2012, [c]this report
At2g45350	CRR4	PLS-E	-	V (At[a])		C		Editing *ndhD* (C[b])	[a]Szponarski et al 2004, [b]Kotera et al 2004
At2g46050	PPR containing protein	PLS-E	M			pM			
At2g48000	PPR containing protein	P	M			pM			
At3g01580	PPR containing protein	PLS-E	M	C (At[a])	M[b]	M/c			[a]Kong et al 2011, [b]this report
At3g02010	PPR containing protein	PLS-E-DYW	M		M[a,b]	M			[a]Lurin et al 2004, [b]this report
At3g02330	PPR containing protein	PLS-E	m		M/C[a]	m/c			[a]in house SUBA3
At3g02490	PPR containing protein	P	M			pM			
At3g02650	PPR containing protein	P	M			pM			
At3g03580	PPR containing protein	PLS-E-DYW	-			-			
At3g04130.1	PPR containing protein	P	M			pM			
At3g04750	PPR containing protein	PLS-E	M			pM			
At3g04760	PPR containing protein	P	C	C (At[a,b])		C			[a]PPDB, [b]Kleffman et al 2004
At3g05240	MEF19	PLS-E	m			M		Editing *ccb206* (M[a])	[a]Takenaka et al 2010
At3g05340	PPR containing protein	PLS-E	M			pM			
At3g06430	EMB2750 /AtPPR2	P	C	C (Zm[a])	C[b,c]	C		Translation stabilisation (C[c,d])	[a]PPDB, [b]in house SUBA3, [c]Williams & Barkan 2003, [d]Lu et al 2011
At3g06920	PPR containing protein	P	-		M[a]	m			[a]this report
At3g07290	PPR containing protein	P	M			pM			
At3g08820	PPR containing protein	PLS-E-DYW	m		M[a,b]	M			[a]this report, [b]in house SUBA3
At3g09040	PPR containing protein	PLS-E	M			pM			
At3g09060	PPR containing protein	P	M		M[a]	M			[a]this report
At3g09650	HCF152/CRM3	P	C	C (Zm[a])	C[b,c]	C		Processing *petB* (C[b])	[a]PPDB, [b]Meierhoff et al 2003, [c]this report
At3g11460	MEF10	PLS-E-DYW	M		M[a]	M		Editing *nad2* (M[b])	[a]Lurin et al 2004, [b]Hartel et al 2013
At3g12770	MEF22	PLS-E-DYW	-	Ct (At[a])	M[b]	M		Editing *nad3* (M[c])	[a]Hummel et al 2012, [b]this report, [c]Takenaka et al 2010
At3g13150	PPR containing protein	P	-			-			
At3g13160	PPR containing protein	P	M	M (At[a,b,c])	M[d]	M			[a]Heazlewood et al 2004, [b]Klodmann et al 2011, [c]Taylor et al 2011, [d]in house SUBA3
At3g13770	PPR containing protein	PLS-E-DYW	M		M[a]	M			[a]Lurin et al 2004
At3g13880	OTP72	PLS-E	M		M[a]	M		Editing *rpl16* (M[b])	[a]Lurin et al 2004, [b]Chateignier-Boutin et al 2013
At3g14330	CREF3	PLS-E-DYW	M		M[a]	m/C		Editing *psbE* (C[b])	[a]this report, [b]Yagi et al 2013
At3g14580	PPR containing protein	P	M			pM			
At3g14730	PPR containing protein	PLS-E	-			-			
At3g15130	PPR containing protein	PLS-E-DYW	M		M[a], M/C[b]	M/c			[a]this report, [b]in house SUBA3
At3g15200	PPR containing protein	P	M			pM			
At3g15590	PPR containing protein	P	M	M (At[a,b,c])		M			[a]Brugiere et al 2004, [b]Klodmann et al 2011, [c]Taylor et al 2011
At3g15930	PPR containing protein	PLS-E	C		M[a]	m			[a]this report
At3g16010	PPR containing protein	P	M			pM			
At3g16610	PPR containing protein	PLS-E	m		M[a]	M			[a]this report
At3g16710	PPR containing protein	P	M			pM			
At3g16890	PPR40	P	M		M[a,b]	M			[a]Zsigmond et al 2008, [b]in house SUBA3
At3g18020	PPR containing protein	P	M			pM			
At3g18110	EMB1270	P-D	C	C (Zm[a])		C	confirmed[b]		[a]PPDB, [b]SeedGenes
At3g18840	PPR containing protein	PLS-E	-			-			
At3g18970	MEF20	PLS-E	M	C (At[a])		M/c		Editing *rps4* (M[b])	[a]AT_Chloro, [b]Takenaka et al 2010
At3g20730	PPR containing protein	PLS-E	m		M[a]	M			[a]this report
At3g21470	PPR containing protein	PLS-E	-		m/c[a]	m/c			[a]this report
At3g22150	MPR25	PLS-E	C	C (At[a])	M[b], C[c]	M/c		Editing *nad5* (M[b])	[a]AT_Chloro, [b]Toda et al 2012, [c]in house SUBA3
At3g22470	PPR containing protein	P	M			pM			

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

			Localization						
AGI	Gene Annotation[1]	Domains 2	Predictions 3	Proteomics 4	Experimental[5]	Conclusion 6	EMB[7]	Molecular Function (localization)[8]	References
At3g22670	PPR containing protein	P	M			pM			
At3g22690	YS1	PLS-E-DYW	C		C[a]	C		Editing *rpoB* (C[a])	[a]Zhou et al 2008
At3g23020	PPR containing protein	P	m	C (At[a,b])		c			[a]PPDB, [b]Kleffman et al 2004
At3g23330	PPR containing protein	PLS-E-DYW	-		m/c[a]	m/c			[a]this report
At3g24000	PPR containing protein	PLS-E-DYW	M		M[a]	M			[a]Lurin et al 2004
At3g25060	PPR containing protein	PLS-E	M			pM			
At3g25210	PPR containing protein	P	-			-			
At3g25970	PPR containing protein	PLS-E	-		M[a]	m			[a]this report
At3g26540	PPR containing protein	PLS	-		M[a]	m			[a]this report
At3g26630	PPR containing protein	PLS	C			pC			
At3g26782	MEF14	PLS-E-DYW	-	M (At[a])		M		Editing *matR* (M[b])	[a]Heazlewood et al 2004, [b]Verbitskiy et al 2011
At3g27750	EMB3123/THA8	P-D	C	C (At[a])	C[b]	C		Splicing *ycf3* intron 2 & *trnA* (C[b])	[a]PPDB, [b]Khrouchtchova et al 2012
At3g28640	PPR containing protein	PLS-E	-			c			
At3g28660	PPR containing protein	PLS-E	C			pC			
At3g29230	PPR containing protein	PLS-E	C	C (At[a])	C[b]	C			[a]PPDB, [b]in house SUBA3
At3g29290	EMB2076	P	M			pM	potential[a]		[a]SeedGenes
At3g42630	PPR containing protein	P	M	C (At[a,b], Zm[a])	C[c]	C			[a]PPDB, [b]Kleffman et al 2004, [c]this report
At3g46610	PPR containing protein	P	-	C (Zm[a])	C[b]	C			[a]PPDB, [b]this report
At3g46790	CRR2	PLS-E-DYW	C		C[a,b]	C		Processing *ndhB* (C[c])	[a]this report, [b]in house SUBA3, [c]Hashimoto et al 2003
At3g46870	THA8-LIKE2	P	M	C (At[a,b]) PM (At[c])		c			[a]PPDB, [b]Kleffman et al 2004, [c]Mitra et al 2009
At3g47530	PPR containing protein	PLS-E-DYW	m		M/C[a]	m/c			[a]this report
At3g47840	PPR containing protein	PLS-E	M		C[a]	c			[a]this report
At3g48250	BIR6	P	C		M[a]	M		Splicing *nad7* intron1 (M[a])	[a]Koprivova et al 2010
At3g48810	PPR containing protein	P	m		M[a,b]	M			[a]this report, [b]in house SUBA3
At3g49140	PPR containing protein	PLS-E-DYW	M	C (At[a,b])		c			[a]AT_Chloro, [b]PPDB
At3g49170	EMB2261	PLS-E-DYW	C		C[a]	C	confirmed[b]		[a]in house SUBA3, [b]SeedGenes
At3g49240	EMB1796	P	M	M (At[a,b]) PM (Atd) C (At[c], Zm[c])	M/C[e]	M/C	confirmed[f]		[a]Ito et al 2006, [b]Klodmann 2011, [c]PPDB, [d]Zhang et al 2011, [e]this report, [f]SeedGenes
At3g49710	PPR containing protein	PLS-E-DYW	-		M/C[a]	m/c			[a]this report
At3g49730	Zmempp4 orthologous	P	-		M (Zm[a])	M			[a]Gutierrez-marcos et al 2007
At3g49740	PPR containing protein	PLS-E	m		M[a,b]	M			[a]this report, [b]in house SUBA3
At3g50420	PPR containing protein	PLS-E	m		M/C[a]	m/c			[a]this report
At3g51320	PPR containing protein	PLS-E	M			pM			
At3g53170	PPR containing protein	P	-	N (At[a]) C (Zm[b])	C	n/C			[a]Pendle et al 2005, [b]PPDB
At3g53360	PPR containing protein	PLS-E	M			pM			
At3g53700	MEE40	P	-	C (At[a], Zm[a])		C	confirmed[b]		[a]PPDB, [b]Pagnussat et al 2005
At3g54980	PPR containing protein	P	M			pM			
At3g56030	PPR containing protein	P	M			pM			
At3g56550	PPR containing protein	PLS-E-DYW	m			pM			
At3g57430	OTP84	PLS-E-DYW	C	PM (At[a])	C[b,c]	C		Editing *psbZ, ndhB, ndhF* (C[d])	[a]Li et al 2012, [b]this report, [c]in house SUBA3, [d]Hammani et al 2009
At3g58590	PPR containing protein	P	m		M[a]	M			[a]this report
At3g59040.1	PPR containing protein	P	C	C (Zm[a])		C			[a]PPDB
At3g60050	PPR containing protein	P	M			pM			
At3g60960	PPR containing protein	P	M	M (At[a])		M			[a]Heazlewood et al 2004
At3g60980	PPR containing protein	P	M			pM			
At3g61170	PPR containing protein	PLS-E-DYW	M			pM			
At3g61360	PPR containing protein	P	M			pM			
At3g61520	PPR containing protein	P	M			pM			
At3g62470	PPR containing protein	P	M			pM			
At3g62540	PPR containing protein	P	M			pM			
At3g62890	PPR containing protein	PLS-E-DYW	C		M/C[a], C[b]	m/C			[a]this report, [b]in house SUBA3
At3g63370	OTP86	PLS-E-DYW	C		C[a]	C		Editing *rps14* (C[b])	[a]in house SUBA3, [b]Hammani et al 2009
At4g01030	PPR containing protein	PLS-E-DYW	M	C (Zm[a])		c			[a]PPDB
At4g01400.1	PPR containing protein	P-D	M	PM (At[a])		pM			[a]Mitra et al 2009
At4g01570	PPR containing protein	P	M		C[a]	c			[a]this report
At4g01990	PPR containing protein	PLS-E	M			pM			
At4g02750	PPR containing protein	PLS-E-DYW	M		M[a,b]	M			[a]Lurin et al 2004, [b]this report
At4g02820	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At4g04370	PPR containing.protein.	PLS-E	m		M/C[a,b]	M/C			[a]this report, [b]in house SUBA3
At4g04790	PPR containing protein	P	M			pM			
At4g08210	PPR containing protein	PLS-E	m		M[a]	M			[a]this report
At4g11690	PPR containing protein	P	m		M[a]	M			[a]this report
At4g13650	PPR containing protein	PLS-E-DYW	M			pM			
At4g14050	PPR containing protein	PLS-E-DYW	M			pM			
At4g14170	PPR containing protein	PLS-E	M			pM			
At4g14190	PPR containing protein	P	M			pM			
At4g14820	PPR containing protein	PLS-E-DYW	C		M[a]	m			[a]this report
At4g14850	LOI1/MEF11	PLS-E-DYW	M		M[a,] M/C[b]	M/c		Editing *cox3, nad4, ccb203* (M[c])	[a]Tang et al 2010, [b]this report, [c]Verbitskiy et al 2010
At4g15720	PPR containing protein	PLS-E-DYW	-		C[a]	c			[a]this report
At4g16390	SVR7 /RNA binding P67	P-D	C	C (At[a,b], Zm[b])	C[c,d]	C			[a]AT_Chloro, [b]PPDB, [c]Lurin et al 2004, [d]Liu et al 2010
At4g16470	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At4g16835	PPR containing protein	PLS-E-DYW	m			pM			
At4g17616	PPR containing protein	P	M			pM			
At4g17910	PPR containing protein	P	M			pM			
At4g18520	PDM1	P	C	C (Zm[a])		C		Processing *rpoA* transcript (C[b])	[a]PPDB, [b]Hao et al 2010
At4g18750	DOT4	PLS-E-DYW	C	C (Zm[a])	C[b]	C			[a]PPDB, [b]in house SUBA3
At4g18840	PPR containing protein	PLS-E	-		C[a,b]	c			[a]this report, [b]in house SUBA3
At4g18975.1	PPR containing protein	P	C			pC			
At4g19191	PPR containing protein	PLS-E	M			pM			
At4g19220	PPR containing protein	PLS-E	M			pM			
At4g19440	PPR containing protein	P	-			-			
At4g19900	Glycosyl transferase-related	P-D	m			pM			
At4g20090	EMB1025	P	c		M[a,b]	M	confirmed[c]		[a]Lurin et al 2004, [b]this report, [c]SeedGenes
At4g20740	EMB3131	P	c	PM (At[a])	C[b]	C	confirmed[c]		[a]Li et al 2012, [b]this report, [c]SeedGenes
At4g20770	PPR containing protein	PLS-E	M			pM			
At4g21065.1	PPR containing protein	PLS-E-DYW	-			-			
At4g21170	PPR containing protein	P	M		M/C[a]	M/C			[a]Narsai et al 2011
At4g21190	EMB1417	P	M	C (Zm[a])		c	confirmed[b]		[a]PPDB, [b]SeedGenes

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation[1]	Domains 2	Localization: Predictions 3	Localization: Proteomics 4	Localization: Experimental[5]	Localization: Conclusion 6	EMB[7]	Molecular Function (localization)[8]	References
At4g21300	PPR containing protein	PLS-E	M	C (Zm[a])	C[b]	C			[a]PPDB, [b]in house SUBA3
At4g21705	PPR containing protein	P	M			pM			
At4g21880	PPR containing protein	P	M		M[a]	M			[a]this report
At4g21900	PRORP3	P-D	-		N[a]	N		Processing tRNA and maturation of RNA (N[b])	[a]Gobert et al 2010, [b]Gutmann et al 2012
At4g22760	PPR containing protein	PLS-E	-		M[a]	m			[a]this report
At4g25270	OTP70	PLS-E	C		C[a,b]	C		Splicing *rpoC1* intron (C[a])	[a]Chateigner-Boutin et al 2011, [b]in house SUBA3
At4g26680	PPR containing protein	P	M			pM			
At4g26800	PPR containing protein	P	M			pM			
At4g28010	RFP5	P	M		M[a]	M		processing *nad6, atp9*, *26S rRNA* (M[b])	[a]this report, [b]Hauler et al 2013
At4g30700	MEF29/ ZmPPR2263	PLS-E-DYW	M		M/C[a]	M/c		Editing *nad5, cob* (M[a])	[a]Sosso et al 2012
At4g30825	PPR containing protein	P	-	C (Zm[a])		c			[a]PPDB
At4g31070	PPR containing protein	PLS-E	M			pM			
At4g31850	PGR3	P	-		C[a]	C		Translation stabilisation *petL* and *ndhA* (C[b,c])	[a]Lurin et al 2004, [b]Yamazaki et al 2004, [c]Cai et al 2011
At4g32430	PPR containing protein	PLS-E	M			pM			
At4g32450	MEF8S	PLS-E-DYW	M		M[a]	M		Editing *nad5, nad6* (M[b])	[a]in house SUBA3, [b]Vervitskiy et al 2012
At4g33170	PPR containing protein	PLS-E-DYW	M		M/C[a,b]	M/C			[a]this report, [b]in house SUBA3
At4g33990	EMB2758	PLS-E-DYW	M			pM	potential[a]		[a]SeedGenes
At4g34830	MRL1	P	C	C (At[a,b]) PM (At[c])		C		Processing stabilisation *rbcL* (C[d])	[a]PPDB, [b]AT_Chloro, [c]Li et al 2012, [d]Johnson et al 2010
At4g35130	PPR containing protein	PLS-E-DYW	C		C[a]	C			[a]in house SUBA3
At4g35850	PPR containing protein	P	M	M (At[a,b,c,d], Os[e])		M			[a]Millar et al 2001, [b]Heazlewood et al 2004, [c]Klodmann et al 2011, [d]Taylor et al 2011, [e]Huang et al 2009
At4g36680	PPR containing protein	P	M	M (At[a,b])		M			[a]Heazlewood et 2004, [b]Klodmann et al 2011
At4g37170	PPR containing protein	PLS-E-DYW	-		M[a]	m			[a]this report
At4g37380	PPR containing protein	PLS-E-DYW	C			pC			
At4g38010	PPR containing protein	PLS-E	M			pM			
At4g38150.1	PPR containing protein	P	-	M (At[a])		m			[a]Taylor et al 2011
At4g39530	PPR containing protein	PLS-E	M			pM			
At4g39620	EMB2453 /ZmPPR5	P	C	C (Zm[a])		C	confirmed[b]	splicing *trnG* (C[c])	[a]PPDB, [b]SeedGenes, [c]Beick et al 2008
At4g39952	PPR containing protein	PLS-E	M			pM			
At5g01110	PPR containing protein	P	C			pC			
At5g02830	PPR containing protein	P	C	C (Zm[a])		C			[a]PPDB
At5g02860	PPR containing protein	P	M	C (Zm[a])		c	potential[b]		[a]PPDB, [b]Myouga et al 2010
At5g03560.2	PPR containing protein	P	M			pM			
At5g03800	EMB175	PLS-E-DYW	C	C (Zm[a]) PM (At[b])	M/C[c,d]	m/C	confirmed[e]		[a]PPDB, [b]Keinath et al 2010, [c]in house SUBA3, [d]this report, [e]SeedGenes
At5g04780	PPR containing protein	PLS-E-DYW	-	C (At[a])		c			[a]Kleffman et al 2004
At5g04810	ZmPPR4	P-D	C	C (At[a,c])	C[b]	C		Splicing *rps12* intron1 (C[c])	[a]PPDB, [b]this report, [c]Schmitz-linneweber et al 2006
At5g06540	PPR containing protein	PLS-E-DYW	M		C[a], M/C[b]	m/c			[a]this report, [b]in house SUBA3
At5g08310	PPR containing protein	PLS-E	M			pM			[a]this report
At5g08490	SLG1	PLS-E	M		M[a]	M		Editing *nad3* (M[a])	[a]Yuan & Liu 2012
At5g08510	PPR containing protein	PLS-E	-		M[a]	m			[a]this report
At5g09450	PPR containing protein	P	M	M (At[a])		M			[a]Klodmann et al 2011
At5g09950	MEF7	PLS-E-DYW	-		M[a]	M		Editing *nad2, nad4L, cob, ccb206* (M[b])	[a]Lurin et al 2004, [b]Zehrmann et al 2012
At5g10690	PPR containing protein	P-D	-	Ct (At[a])	C[b]	C			[a]Ito et al 2011, [b]this report
At5g11310	PPR containing protein	P	M			pM			
At5g12100	PPR containing protein	P	M	M (At[a])		M			[a]Tan et al 2009
At5g13230	PPR containing protein	PLS-E-DYW	M		M[a]	M			[a]Lurin et al 2004
At5g13270	RARE1	PLS-E-DYW	-		C[a]	C		Editing *accD* (C[b])	[a]Lurin et al 2004, [b]Robbins et al 2009
At5g13770	PPR containing protein	P	C	C (Zm[a])		C			[a]PPDB
At5g14080	PPR containing protein	P	C	C (At[a])	m/c[b]	m/C			[a]PPDB, [b]this report
At5g14770	PPR containing protein	P	M	PM (At[a])	M[b]	M			[a]Li et al 2012, [b]Lurin et al 2004
At5g14820	PPR containing protein	P	M			pM			
At5g15010	PPR containing protein	P	C			pC			
At5g15280	PPR containing protein	P	M			pM			
At5g15300	PPR containing protein	PLS-E	-	PM (At[a])		-			[a]Mitra et al 2009
At5g15340	PPR containing protein	PLS-E-DYW	M		M[a]	M			[a]Lurin et al 2004
At5g15980	PPR containing protein	P	M	M (At[a]) PM (At[b])		M			[a]Klodmann et al 2011, [b]Zhang et al 2011
At5g16420	PPR containing protein	P	M			pM			
At5g16640	PPR containing protein	P	M			pM			
At5g16860	PPR containing protein	PLS-E-DYW	-			-			
At5g18390	PPR containing protein	P	M			pM			
At5g18475	PPR containing protein	P	M		M[a]	M			[a]this report
At5g18950	PPR containing protein	P	M		M[a]	M			[a]this report
At5g19020	MEF18	PLS-E	C			M		Editing *nad4* (M[a])	[a]Takenaka et al 2010
At5g21222	AtC401	P-D	-	C (At[a])	m/c[b]	m/C			[a]PPDB, [b]this report
At5g24830	PPR containing protein	P	-			-			
At5g25630	PPR containing protein	P	-	C (Zm[a])		c			[a]PPDB
At5g27110	PPR containing protein	PLS-E	M		M/C[a]	m/c			[a]in house SUBA3
At5g27270	EMB976	P	C	C (Zm[a])	C[b]	C	potential[c]		[a]PPDB, [b]this report, [c]SeedGenes
At5g27460	PPR containing protein	P	M			pM			
At5g28460	PPR containing protein	P	M			pM			
At5g36300	pseudogene	P	-			-			
At5g37570	PPR containing protein	PLS-E	M			pM			
At5g38730	PPR containing protein	P	M			pM			
At5g39350	PPR containing protein	PLS-E	M		M[a]	M			[a]in house SUBA3
At5g39680	EMB2744	PLS-E-DYW	-		M[a,b]	m	potential[c]		[a]Lurin et al 2004, [b]in house SUBA3, [c]SeedGenes
At5g39710	EMB2745	P	M		M[a]	M	potential[b]		[a]Narsai et al 2011, [b]SeedGenes
At5g39980	EMB3140	P	-	C (Zm[a])		c	confirmed[b]		[a]PPDB, [b]SeedGenes
At5g40400	PPR containing protein	P	-			-			
At5g40405	PPR containing protein	PLS-E-DYW	-		M[a]	m			[a]this report
At5g40410	PPR containing protein	PLS-E-DYW	m	C (At[a,b])		c			[a]AT_Chloro, [b]Kong et al 2011
At5g41170	PPR containing protein	P	M			pM			
At5g42310	Ortholog of Z. Mays CRP1	P	M	C (At[a])		C		Translation stabilisation *petA* and *psaC* (C[b,c])	[a]PPDB, [b]Fisk et al 1999, [c]Schmitz-linneweber et al 2005
At5g42450	PPR containing protein	PLS-E	-			-			
At5g43790	PPR containing protein	PLS-E	-		m/c[a]	m/c			[a]this report
At5g43820	PPR containing protein	P	M			pM			

Table 2. Prediction and experimental localization data of *Arabidopsis thaliana* PPR proteins (continued)

AGI	Gene Annotation[1]	Domains [2]	Localization: Predictions [3]	Localization: Proteomics [4]	Localization: Experimental[5]	Localization: Conclusion [6]	EMB[7]	Molecular Function (localization)[8]	References
At5g44230	PPR containing protein	PLS-E-DYW	C			pC			
At5g46100	PPR containing protein	P	M			pM			
At5g46460	PPR containing protein	PLS-E-DYW	M			pM			
At5g46580	PPR containing protein	P-D	C	C (At[a,b], Zm[b])		C			[a]AT_Chloro, [b]PPDB
At5g46680	PPR containing protein	P	M	PM (At[a])	M[b]	M			[a]Li et al 2012, [b]this report
At5g47360	PPR containing protein	P	M			pM			
At5g47460	PPR containing.protein.	PLS-E	M		M/C[a]	m/c			[a]this report
At5g48730	PPR containing protein	P	C	C (Zm[a])		C			[a]PPDB
At5g48910	LPA66	PLS-E-DYW	C		C[a,c], M[b]	m/C		Editing *psbF* (C[c])	[a]this report, [b]in house SUBA3, [c]Cai et al 2009
At5g50280	EMB1006	P	C	C (Zm[a]) PM (At[b])		C	potential[c]		[a]PPDB, [b]Mitra et al 2009, [c]SeedGenes
At5g50390	EMB3141	PLS-E-DYW	C		C[a]	C	potential[b]		[a]in house SUBA3, [b]SeedGenes
At5g50990	PPR containing protein	PLS-E-DYW	-			-			
At5g52630	MEF1	PLS-E-DYW	c		C[a], M[b]	M		Editing *rps4, nad7, nad2* (M[c])	[a]Lurin et al 2004, [b]this report, [c]Zehrmann et al 2009
At5g52850	PPR containing protein	PLS-E-DYW	-			-			
At5g55740	CRR21	PLS-E	C		M[a],C[b]	C		Editing *ndhD* (C[c])	[a]Lurin et al 2004, [b]in house SUBA3, [c]Okuda et al 2007
At5g55840	PPR containing protein	P	-		M[a]	m			[a]this report
At5g56310	PPR containing protein	PLS-E	M		M[a]	M			[a]this report
At5g57250	PPR containing protein	P	M			pM			
At5g59200	OTP80	PLS-E	c		C[a]	C		Editing *rpl23* (C[b])	[a]in house SUBA3, [b]Hammani et al 2009
At5g59600	PPR containing protein	PLS-E	-		C[a]	c			[a]this report
At5g59900	PPR containing protein	P	M		M[a]	M			[a]this report
At5g60960	PNM1	P	M	M (Os[a])	M/N[b], M[c]	M/N	confirmed[b]		[a]Huang et al 2009, [b]Hamanni et al 2011, [c]Narsai et al 2011
At5g61370	PPR containing protein	P	M		M[a]	M			[a]Narsai et al 2011
At5g61400	PPR containing protein	P	M			pM			
At5g61800	PPR containing protein	PLS-E	M	PM (At[a])		pM			[a]Li et al 2012
At5g61990	PPR containing protein	P	M			pM			
At5g62370	PPR containing protein	P	M			pM			
At5g64320	PPR containing protein	P	M			pM			
At5g65560	PPR containing protein	P	M			pM			
At5g65570	PPR containing protein	PLS-E-DYW	m	PM (At[a])		pM			[a]Mitra et al 2009
At5g65820	Zmempp4 ortholog 2	P	M		M (Zm[a], At[b])	M			[a]Gutierrez-marcos et al 2007, [b]this report
At5g66500	PPR containing protein	PLS-E	M			pM			
At5g66520	CREF7	PLS-E-DYW	-		C[a]	C		Editing *ndhB* (C[b])	[a]this report, [b]Yagi et al 2013
At5g66631	PPR containing protein	P	C			pC			
At5g67570	EMB1408/DG1/ ZmPPR8852	P	-	C (Zm[a])	C[b]	C			[a]PPDB, [b]Chi et al 2008

(1) Functional annotations were obtained from TAIR web site using the *Arabidopsis* Genome Initiative (AGI) genome release ver10. ABO5, ABA OVERLAY-SENSITIVE; AtECB, EARLY CHLOROPLAST BIOGENESIS; BIR, BSO-INSENSITIVE-ROOTS; CLB, CHLOROPLAST BIOGENESI; CREF, CHLOROPLAST RNA EDITING FACTOR; CRR, CHLORORESPIRATORY REDUCTION; DG, DELAYED GREENING; DOT, DEFECTIVELY ORGANIZED TRIBUTARIES; EMB, EMBRYO DEFECTIVE; FAC, EMBRYONIC FACTOR; GRP, GLUTAMINE-RICH PROTEIN; GUN, GENOME UNCUPLED; HCF, HIGH CHLOROPHYLL FLUORESCENCE; LOI, LOVASTATINE INSENSITIVE; LOJ, LATERAL ORGAN JUNCTION; LPA, LOW PSII ACCUMULATION; MEF, MITOCONDRIAL RNA EDITING FACTOR; MPR25, MITOCHONDRIAL PPR 25; MTSF, MITOCHONDRIAL STABILITY FACTOR; NFD, NUCLEAR FUSION DEFECTIVE; OGR1, OPAQUE AND GROWTH RETARDATION; OTP, ORGANELLE TRANSCRIPT PROCESSING, PDE: PIGMENT DEFECTIVE; PDM, PIGMENT DEFICIENT MUTANT; PGN, PENTATRICOPEPTIDE GERMINATION ON NaCl; PGR, PROTON GRADIENT REGULATION; PNM, PROTEIN LOCALIZED TO THE NUCLEUS AND MITOCHONDRIA; PPR, PENTATRICOPEPTIDE REPEAT; PRORP, PROTEINACEUS RNASE P; PTAC, PLASTID TRANSCRIPTIONALLY ACTIVE; REME, REQUIERED FOR EFFICENCY OF MITOCHONDRIAL EDITING; RPF, RNA PROCESSING FACTOR; SLG, SLOW GROWTH; SVR, SUPRESSOR OF VARIEGATION; VAC, VANILLA CREAM; YS, YELLOW SEEDLING; Zmempp, Z. mays EMPTY PERICARP, ZmPPR, Zea mays PPR. (2) PPR domains were recovered from FLAGdb++ v5 (http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml) and from manually curated published evidences. Domain identifiers are according to Lurin and co-workers:[11] "P" for PPR P-type domains, "P-D" for PPR P-type with additional atypical domain, "PLS" for PPR PLS-type domains, "PLS-E" for PPR PLS-type with an E- or EE+- type additional domain, and "PLS-E-DYW" for PPR PLS-type containing EE+ and DYW additional domains. (3) Localization predictions were aggregated from the independent predictions provided by the following software: Predotar v1.03, TargetP server v1.1, iPSORT, Multi Loc, LocTree, and AtsubP server with the complete *Arabidopsis* proteome using default settings. The rules to propose a conclusive prediction were as follows: if four or more software give the same prediction, this prediction is proposed and noted in uppercase; if three software give the same prediction and the three others do not predict any localization, the prediction is proposed and noted in uppercase; if two software give the same prediction and the four others do not predict any localization, the prediction is proposed and noted in lowercase; if three software give the same prediction and another predict a different localization, the main prediction is proposed and noted in lowercase; in the other cases, no prediction is proposed (-). (4) Proteomic localizations were gathered from published studies and from organelle proteomic databases as indicated in corresponding references in the last column of the table. Additional information in brackets states in which specie(s) the proteomic investigation was (were) performed: "At" stands for *Arabidopsis thaliana*, "Zm" for *Zea mays*, and "Os" for *Oriza sativa*. (5) Experimental localizations of fluorescent proteins were collected from targeted published studies and systematic approaches,[11,57] this report, unpublished data from SUBA3 either using targeting peptides or full-length proteins. (6) Conclusion column gives a probable subcellular localization by integrating prediction, proteomic, genetics, and fluorescent proteins data. The decision rule is as follows: reverse genetics is prevalent followed by fluorescent proteins, proteomic data, and prediction. The conclusion is indicated in uppercase if reverse genetics data is available, if two experimental results are identical, or if the experimental data fit with the prediction. If not, the conclusion is indicated in lowercase. If only predictions are available, the predicted localization is indicated with a preceding "p". (7) Data of PPR Embryo defective mutants (EMB) was obtained from SeedGenes database (http://www.seedgenes.org/index.html) and manually curated mutants from published studies. (8) Molecular function based on reverse genetics approaches were obtained from literature, the localization of the molecular function is indicated in brackets. Localization data is indicated as followed. M, mitochondria; C, chloroplasts; N, nucleus; V, vacuole; Ct, cytosol; PM, plasma membrane. N/Ct, nucleus and cytoplasm; M/C, mitochondria and chloroplasts; lower case, "probably"; "pX", predicted in compartment X (conclusion column).

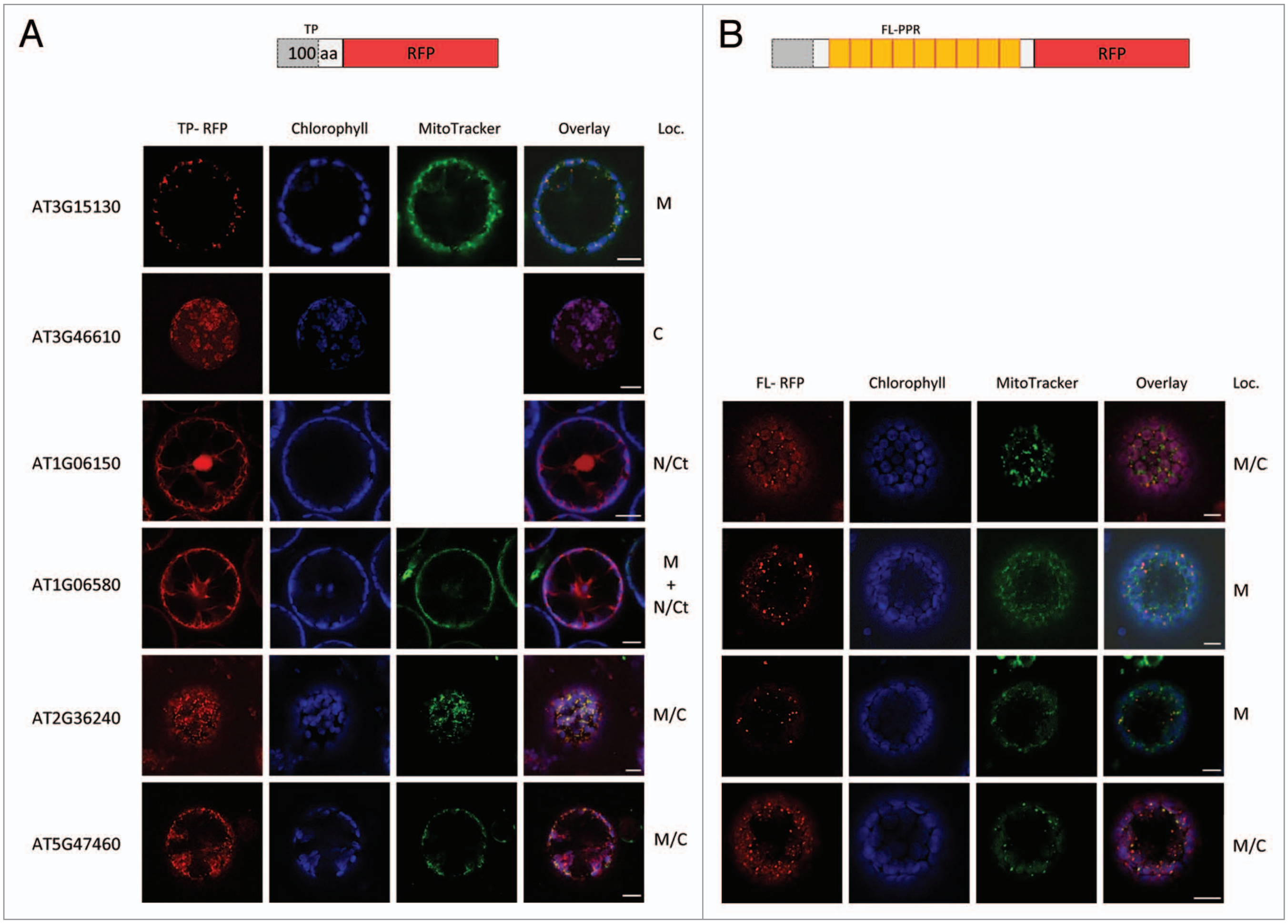

Figure 1. Examples of typical sub-cellular localizations observed using confocal microscopy. Confocal images of protoplasts obtained from *Nicotiana benthamiana* leaves infiltrated with constructs containing (**A**) the first 300 bp, coding for the first 100 amino acids, of six PPR ORFs, fused to the RFP coding sequence (TP-RFP) or (**B**) the full-length sequence of four PPR ORFs fused to the RFP coding sequence (FL-RFP). The RFP fluorophore (in red), the MitoTracker Green staining (in green), and the chlorophyll autofluorescence (in blue) were simultaneously visualized. Overlay panels show combined fluorescence from RFP, MitoTracker, and chlorophyll autofluorecence. Loc, deduced subcellular localization; M, mitochondria; C, chloroplasts; N/Ct, nucleus and cytosol; M/C, mitochondria and chloroplasts. Bars: 10 μm.

with 17 out of 18 (94%) and 53 out of 57 (93%) compatible localization, respectively (**Table 3**).

As concluded in **Table 2** and taking into account all the above depicted approaches, we assigned all Arabidopsis PPR proteins a probable localization depending on the strength of the available data. The localization based on reverse genetics, when available, prevailed over any other approaches. Because we showed that the experimental localizations of fusion proteins were highly correlated with the localization of the molecular function when identified (**Table 3**), this data prevailed over the proteomics and bio-informatics ones. Additionally, PPR protein identification in organellar proteomes, though showing some discrepancies with functional data suggesting some errors of localization linked to this technique, was as far as we know more trustable that bio-informatics predictions. Finally, when no experimental data was available, we proposed a predicted localization in mitochondria or chloroplast (pM or pC). **Figure 2** gives a graphical view of these results. The number of PPR proteins with a suspected or proved subcellular localization in at least one of the two organelles increased significantly with our study. For example, the experimental mitochondrial and chloroplast localization data increased by 50% (from 134 to 212) and with the addition 19 PPRs with experimental dual targeting to mitochondria and chloroplast to the previously 10 known. Overall, 275 PPR proteins (60%) are expected to function in mitochondria, with 44% of them being validated in experimental studies. Additionally, 109 PPR proteins (24%) are expected to function in plastids, 82% being demonstrated experimentally. Forty-five PPR proteins (10%) are suspected to have a dual addressing to both plastids and mitochondria. Finally, five PPR proteins have been shown to have atypical localization: PROPR2 and PROPR3 were shown to be addressed to the nucleus,[58] PNM1 and GRP23 to both nucleus and mitochondria,[27,28,57] and AT3G53170 was observed in both nuclear and chloroplastic extracts during proteomics studies.[48,59] Only 24 PPR proteins (5%) do not have any clear localization based on experimental or bio-informatics reported investigations.

Table 3. Correlations between localization data sets

Data sets (number of PPR proteins with data in this set)	Fusion proteins: This study (126)	Fusion proteins: All data (208)	Proteomics (84)	Reverse genetics (68)
Predictions (377)	70/87 (80%)	135/159 (85%)	48/68 (71%)	44/55 (80%)
Reverse genetics (68)	17/18 (94%)	53/57 (93%)	12/15 (80%)	
Proteomics (84)	15/19 (79%)	30/36 (83%)		

In each cell of the table, the number and the percent of compatible localizations among the intersection of data available in both data sets are indicated. Two results are considered as compatible when their localizations are coherent: for example, experimental localization in both organelles and prediction or proteomics indicating only one of the two organelles.

Discussion

RFP fusions with PPR-targeting peptides allowed us to study the subcellular localization of many members of the PPR family. Our aim in this study was to clarify the subcellular localization of 166 members of the large PPR family selected to have ambiguous localization predictions when we started the approach. In order to determine this, we used a strategy of high-throughput gateway cloning of the first 300 bp of PPR ORFs (corresponding to the N-terminal 100 amino acids of proteins) combined to a systematic microscopy investigation of the localization of transiently expressed RFP-tagged proteins. When it was determined that the first 100 amino acids displayed an interesting localization pattern, we performed in a second step a similar study using the whole ORF. Overall, with this work, we provided experimental information on the localization of 131 PPR proteins.

We have shown that 129 PPR proteins have functional targeting peptides able to address the RFP protein in one or both organelles. Seventeen have been previously published in dedicated studies and were shown to localize in agreement with our systematic results (**Table 2 and 3**).[19,21,60-71] Additionally, 15 PPR proteins (HCF152 and OTP51 included) were identified in the same compartment using untargeted proteomic approaches (**Table 2 and 3**).[44,59,72-76] These independent localization results largely validate our systematic strategy.

The strategy we used to study the localization of proteins can be performed at large scale to provide rapid functional information for organellar proteins. Nonetheless, some limitations have to be kept in mind when considering the results: first of all, the use of *Nicotiana benthamiana* is convenient as leaves are very comfortable to work with, but the evolution of addressing signals might be slightly different in distinct dicotyledonous species, explaining some discrepancies in the results. Second of all, the agro-infiltration to transform plant cells and generation of protoplasts to visualize expression are two steps known to generate stresses which, in some cases, may affect the conclusions. At least, the use of the very strong 2X35S promoter to trigger chimerical protein expression may overwhelm the translation and import machineries, leading to erroneous localization. However, the low number of discrepancy cases between our results and published information gained using a very large set of techniques largely validate our strategy and strengthen our results (**Table 3**).

Most discrepancies between our work and previous experimental localizations concern dual-localized proteins. Four of our dual-localized candidates (EMB175, AT5G14080, AT1G64100, AtC401) were previously shown in a single organelle using proteomic approaches[74] PPDB. Similarly, MEF11, and AHG11 were functionally characterized in mitochondria editing,[77-79] and AT3G62890-GFP fusion was previously observed in plastids in house SUBA3, whereas our results suggested a dual localization in both organelles for these three proteins. In contrast, three PPR proteins (AT2G37230, AT3G15130, AT5G06540) are suspected to have a dual localization because of proteomics results PPDB,[42,44] or expression of fusion proteins (unpublished result from SUBA3), and were observed only in one of the two organelles in our study. Finally, five proteins previously observed in plastid extracts (AT1G09900, AT1G19720, AT2G28050, AT3G01580) or shown to be involved in plastid editing (AT3G14330) were observed in mitochondria in our study. Without any functional characterization, these differences cannot be definitively solved. Erroneous dual localization based on RFP-fusion localization could be explained by artifacts triggered by overexpression, whereas erroneous dual localization based on proteomics experiments could be due to sample contaminations. On the other hand, erroneous single localization might be common because of limitation in protein detection in one of the compartments during proteomics or microscopy experiments. The functional characterization of a protein in one of the two organelles does not refute the localization in the other one. Due to these experimental detection limitations, as well as the fact that we believe that dual-localized PPR proteins are mostly underestimated (see below), we have tentatively concluded that these 14 PPR proteins are localized in both organelles.

During this work, we did not observe the nuclear localization of GRP23 published by Ding and co-workers;[27] however, we did observe a mitochondrial localization of the TP fused to RFP, as described previously by Narsai and co-workers.[57] The GRP23 Nuclear Localization Signal, located at position 99–108, was not included in the 100 amino acid fragment used in our experiments.[27] Taken together, these results suggest that GRP23, as PNM1, may localize in both mitochondria and nucleus.

Addressing of PPR proteins to both organelles is underestimated. We identified 19 new PPR proteins that could have a role in both organelles. Integration of proteomic data and previous fluorescent subcellular localization studies suggest that overall at least 45 PPR proteins could be dual targeted. Recently, about 100 nuclear-encoded proteins were shown to be targeted to both mitochondria and plastids.[80] They are proposed to code for important cellular housekeeping activities. In addition, a study showed that in many cases, the dual targeting of proteins is conserved in three distant *Viridiplantae* species,[81] allowing to assume that some PPR proteins could have the same dual localization in several species and probably with related functions.

Among the PPR family, five proteins were published to be dually addressed into mitochondria and plastids.[57,58,69,82] The two orthologs, PPR2263 of maize and MITOCHONDRIAL EDITING FACTOR29 of Arabidopsis (included in our study), were shown to localize mainly in mitochondria, in which they edit *nad5* and *cob* transcripts, but also in plastids, in which their function remain to be elucidated.[69] Four other PPRs (PRORP1, OTP87, AT1G06270, AT4G21170) were not assayed in our investigation because their predicted localizations were not ambiguous according to our criteria. AT1G06270 and AT4G21170 are uncharacterized P-type PPR proteins shown as dually localized by Narsai and co-workers.[57] PROTEINACEOUS RNASE P 1 (PRORP1) was the first PPR protein shown to be dually addressed.[58] PRORP1 is an atypical PPR protein composed of 5.5 consecutive PPR repeats linked to a carboxyl-terminal (C-terminal) metallonuclease domain by a structural zinc-binding domain.[83] This protein is responsible for the nucleolitic maturation of tRNAs, an activity required in both organelles. By the use of targeting peptides fused to GFP protein, three proteins (OTP87, AT1G06270, AT4G21170) were also found in both organelles.[57,82] OTP87 is an essential PPR protein required for RNA editing of mitochondrial *nad7* and *atp1* transcripts in *A. thaliana*. However, the depletion by an antisense strategy of OSPPR1, the ortholog of OTP87 in *O. sativa*, was described to affect the chloroplast biogenesis.[84] The predictions of localization corresponding to these five dual-localized proteins are either mitochondrial or plastidial (**Table 2**). Similarly, among 45 PPR proteins suspected to be localized in both organelles, eight are predicted in chloroplasts, 28 in mitochondria, and only nine do not have any predicted subcellular localization (**Table 2**). This suggests that many dual-targeted PPR proteins might be still unidentified. In particular, we suspect that many might be included in the 172 PPR proteins having a clear localization prediction in one of the two organelles. Moreover, although different mechanisms of dual targeting exist in the plant cell,[85] the current information does not help to hypothesize by which mechanism PPR proteins could be dual targeted, preventing the predictions of these dual localizations.

Dual targeting to mitochondria and chloroplast is an emerging class of localization in the plant cell and the PPR family seems to have an important contribution. Taking into account the functions of PPR proteins in RNA editing, RNA processing, and translation, this type of localization in the PPR family is not surprising and could be seen as a way to control or coordinate organelle RNA metabolism.[86,87] However, this hypothesis requires testing because, until now, only one PPR protein has been shown to function in both organelles.[58] The analysis of domains in a PPR protein could help to infer its putative function. PPR proteins with dual localization seem to be present in all types of functional categories. However, among 45 dual-localized PPR proteins, 31 belong to the PPR-PLS subclass showing a probable overrepresentation of this subclass in the dual-targeted

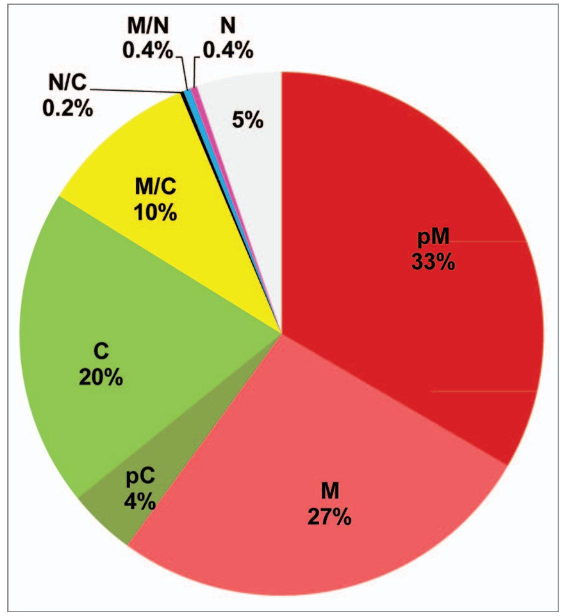

Figure 2. Distribution of the localization of *Arabidopsis thaliana* PentatricoPeptide Repeat (PPR) proteins. Classes of localization and percentage of each class in the PPR family are shown. pM, predicted mitochondria localization in dark red; M, mitochondria localization in light red; pC, predicted plastid localization in dark green; C, plastid localization in light green; M/C, mitochondria and plastid localization in yellow; N/C, nuclear and chloroplastic localization in black; M/N, mitochondria and nuclear localization in blue; N, nuclear localization in pink, unclear localization in light gray.

PPR proteins. Nevertheless, it is important to note that the localization of many PPR-P proteins (115) were not characterized yet, probably biasing this observation.

PPR proteins localized out of organelles seem to represent atypical examples in the family. Using the first 300 bp, we also identified nine PPR proteins potentially addressed out of the organelles, i.e. giving a nuclear and cytosolic localization. None were confirmed using the whole ORFs (**Table 1, Fig. 1**). This suggests that the number of PPR proteins being out of organelles is smaller than we thought when this work was initiated. In total, less than 1–2% of PPR proteins could function in the cytoplasm and/or the nucleus (**Fig. 2**). This value may be still overestimated as the model gene loci are sometimes miss-predicted, in particular, concerning the initiation codon. This may also suggest that the correct targeting sometime needs a peptide longer than the 100 amino acids we used for our work. Huang and co-workers showed that the length of mitochondrial presequence varied greatly from 19–109 amino acids.[36] For GRP23, the beginning of the NLS signal has been located at the amino acid 99. Using the first 100 amino acids, we observed RFP signal into mitochondria (as previously described by Narsai and coworkers[57]) whereas the full-length protein localizes in the nucleus.[27] This findings confirm that systematic localization using the whole proteins could give more accurate information on PPR localizations.

The case of PNM1 is even more complicated. The PNM1 nuclear localization is controlled by a NLS sequence in the C terminus of the protein[82] but the whole protein is addressed to mitochondria. The nuclear localization was only obtained with a truncated form of the protein without the predicted targeting peptide fused with the reporter fluorescent protein. This nuclear localization was confirmed using a specific antibody. The meaning of such a complex addressing system is still a matter of debate but suggests that a few very interesting PPR could be involved in signaling between organelles and nucleus.[86]

Materials and Methods

Bioinformatic predictions and data collection. Subcellular localization prediction of the PPR proteins were performed using TargetP server (http://www.cbs.dtu.dk/services/TargetP/) (version 1.01 was used when we initiated this work to select the 166 PPRs and version 1.1 was used when we built **Tables 1 and 2**), Predotar v1.03 (http://urgi.versailles.inra.fr/predotar/predotar.html), iPSORT (http://ipsort.hgc.jp/), Loctree (https://www.rostlab.org/owiki/index.php/Loctree), Multiloc (http://abi.inf.uni-tuebingen.de/Services/MultiLoc/), and AtSubP (http://bioinfo3.noble.org/AtSubP/?dowhat=About) software using default setting. Proteomic data was recovered from published proteomic references and subcellular proteome databases: PPDB (Plant Proteome Database http://ppdb.tc.cornell.edu/),[43] SUBA3 (Subcellular location database for Arabidopsis proteins http://suba.plantenergy.uwa.edu.au/),[5] and AT_CHLORO (http://www.grenoble.prabi.fr/at_chloro/).[42]

Subcellular localization of proteins. The first 100 codons or the whole PPR ORFs were PCR amplified from *Arabidopsis thaliana* (ecotype Columbia-0) genomic DNA or cDNA using iProof DNA polymerase (Bio-Rad), specific primers (listed in **Table S1**) and a two-step amplification protocol as described previously.[11] PCR products were recombined into pDONR207 (Invitrogen) using Gateway® BP Clonase® II Enzyme mix (Invitrogen) as described.[11] For microscopic investigation, LR recombination reactions were performed using Gateway® LR Clonase® Enzyme Mix (Invitrogen) in order to transfer PPR sequences from Entry vectors to the pGREENII-derived destination vector p0229-RFP2[11] allowing C-terminal translational fusion with the RFP protein under the control of the 2X35S promoter. The proper ORF fusion was confirmed by sequencing using P35STL (5'-CGAATCTCAA GCAATCAAGC-3') and RFP2rev (5'-TGAACTCGGT GATGACGTTC-3') primers.

Binary vectors were introduced into thermo-competent *Agrobacterium tumefaciens* strain C58C1 harboring the helper plasmid pSOUP.[88] A single resistant colony was then used to inoculate 5 mL of Luria Bertani medium supplemented with 5 mg L^{-1} Tetracycline, 50 mg L^{-1} Kanamycine, and 2.5 mg L^{-1} Rifampicine. This overnight pre-culture was then diluted 10 times and further grown overnight in similar conditions. After centrifugation, Agrobacterium cells were re-suspended in agro-infiltration buffer (10 mM MES/KOH pH 5.6, 10 mM $MgCl_2$, 150 μM 3',5'-Dimethoxy-4'-hydroxyacetophenone -Sigma-Aldrich-) with a final OD_{600} between 0.2–0.3, and incubated at room temperature for 2 h. Agrobacterium suspensions were infiltrated using 1 mL syringes without needle in leaves of *Nicotana benthamiana*.

Protoplasts were prepared from leaf material (harvested 48–96 h after infiltration), cut into thin strips, and incubated in enzyme solution containing 4.3 g.L^{-1} Murashige and Skoog Basal Salt Mixture (ICN Biomedicale), 0.5 g.L^{-1} MES, 20 g.L^{-1} sucrose, 80 g.L^{-1} mannitol, KOH to pH 5.6, 0.4 g.L^{-1} Pectinase from *Rhizopus sp.* (Sigma-Aldrich), 1 g.L^{-1} Driselase® *Basidiomycetes sp.* (Sigma-Aldrich) and 2 g.L^{-1} Cellulase Onozuka RS from *Trichoderma viride* (SERVA Electrophoresis GmbH) at 28 °C for 2–4 h.[89] Protoplasts were observed using an Eclipse TE2000S inverted microscope (Nikon) and RFP signal monitored using a custom filter block (exciter HQ546/12, emitter HQ605/75, beam-splitter Q560lp; Chroma Technology). For each construction, at least three independent agro-infiltrations were realized and each of them was observed independently by two of the authors. To confirm mitochondrial localizations, protoplasts were stained with 1 μM MitoTracker Green (Invitrogen) for 15–30 min. For confocal microscopy, proteins were visualized using a spectral Leica SP2 AOBS confocal microscope (Leica Microsystems) equipped with argon and HeNe lasers. Fluorescent signals were detected with a sequential configuration using a 488 nm laser line (MitoTracker Green: excitation/emission 488/510–530 nm) and a 543 nm laser line (RFP: excitation/emission 543/570–600 nm and chlorophyll autofluorescence: excitation/emission 543/600–700 nm). The images were coded red (RFP), green (MitoTracker Green), and blue (chlorophyll autofluorescence), giving yellow co-localization in mitochondria when green and red signals overlap in merged images and violet co-localization in plastid when blue and red signals overlap. Microscopic observations were performed using a Leica HCPL APO 633/1.20 Water Corr/0.17 Lbd.BL objective. Each image shown represents the projection of optical sections taken as a Z series.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Etienne Delannoy (Unité de Recherche en Génomique Végétale) for valuable discussions and comments on the article. We thank Olivier Grandjean (Institut Jean Pierre Bourgin), and Lionel Gissot (Institut Jean Pierre Bourgin) for their help with confocal microscopy. We thank Andeol Falcon de Longevialle (Genopole Plant Process Innovation) for providing us with beautiful healthy *Nicotiana benthamiana* plants. We thank Jennifer Yansouni for English editing. This research was supported a PhD fellowship to ML-O from the Université d'Evry Val-d'Essonne and the French Ministry of Research and Higher Education.

Supplemental Materials

Supplemental materials may be found here:
www.landesbioscience.com/journals/rnabiology/article/26128

References

1. Initiative AG; Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature 2000; 408:796-815; PMID:11130711; http://dx.doi.org/10.1038/35048692
2. Gray MW. Evolution of organellar genomes. Curr Opin Genet Dev 1999; 9:678-87; PMID:10607615; http://dx.doi.org/10.1016/S0959-437X(99)00030-1
3. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nat Genet 1997; 15:57-61; PMID:8988169; http://dx.doi.org/10.1038/ng0197-57
4. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res 1999; 6:283-90; PMID:10574454; http://dx.doi.org/10.1093/dnares/6.5.283
5. Tanz SK, Castleden I, Hooper CM, Vacher M, Small I, Millar HA. SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in Arabidopsis. Nucleic Acids Res 2013; 41(Database issue):D1185-91; PMID:23180787; http://dx.doi.org/10.1093/nar/gks1151
6. Duncan O, Murcha MW, Whelan J. Unique components of the plant mitochondrial protein import apparatus. Biochim Biophys Acta 2013; 1833:304-13; PMID:22406071; http://dx.doi.org/10.1016/j.bbamcr.2012.02.015
7. Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science 2013; 339:571-4; PMID:23372012; http://dx.doi.org/10.1126/science.1229262
8. Schleiff E, Becker T. Common ground for protein translocation: access control for mitochondria and chloroplasts. Nat Rev Mol Cell Biol 2011; 12:48-59; PMID:21139638; http://dx.doi.org/10.1038/nrm3027
9. Binder S, Brennicke A. Gene expression in plant mitochondria: transcriptional and post-transcriptional control. Philos Trans R Soc Lond B Biol Sci 2003; 358:181-8, discussion 188-9; PMID:12594926; http://dx.doi.org/10.1098/rstb.2002.1179
10. Schmitz-Linneweber C, Small I. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 2008; 13:663-70; PMID:19004664; http://dx.doi.org/10.1016/j.tplants.2008.10.001
11. Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, et al. Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 2004; 16:2089-103; PMID:15269332; http://dx.doi.org/10.1105/tpc.104.022236
12. O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 2008; 25:1120-8; PMID:18343892; http://dx.doi.org/10.1093/molbev/msn057
13. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, et al. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 2011; 332:960-3; PMID:21551031; http://dx.doi.org/10.1126/science.1203810
14. Fujii S, Small I. The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 2011; 191:37-47; PMID:21557747; http://dx.doi.org/10.1111/j.1469-8137.2011.03746.x
15. de Longevialle AF, Small ID, Lurin C. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. Mol Plant 2010; 3:691-705; PMID:20603383; http://dx.doi.org/10.1093/mp/ssq025
16. Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. PLoS Genet 2012; 8:e1002910; PMID:22916040; http://dx.doi.org/10.1371/journal.pgen.1002910
17. Yagi Y, Hayashi S, Kobayashi K, Hirayama T, Nakamura T. Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. PLoS One 2013; 8:e57286; PMID:23472078; http://dx.doi.org/10.1371/journal.pone.0057286
18. Kotera E, Tasaka M, Shikanai T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. Nature 2005; 433:326-30; PMID:15662426; http://dx.doi.org/10.1038/nature03229
19. Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. Plant Cell 2009; 21:558-67; PMID:19252080; http://dx.doi.org/10.1105/tpc.108.064535
20. de Longevialle AF, Meyer EH, Andrés C, Taylor NL, Lurin C, Millar AH, Small ID. The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. Plant Cell 2007; 19:3256-65; PMID:17965268; http://dx.doi.org/10.1105/tpc.107.054841
21. de Longevialle AF, Hendrickson L, Taylor NL, Delannoy E, Lurin C, Badger M, Millar AH, Small I. 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 2008; 56:157-68; PMID:18557832; http://dx.doi.org/10.1111/j.1365-313X.2008.03581.x
22. Schmitz-Linneweber C, Williams-Carrier R, Barkan A. RNA 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 2005; 17:2791-804; PMID:16141451; http://dx.doi.org/10.1105/tpc.105.034454
23. Uyttewaal M, Mireau H, Rurek M, Hammani K, Arnal N, Quadrado M, Giegé P. PPR336 is associated with polysomes in plant mitochondria. J Mol Biol 2008; 375:626-36; PMID:18054044; http://dx.doi.org/10.1016/j.jmb.2007.11.011
24. Lloyd J, Meinke D. A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. Plant Physiol 2012; 158:1115-29; PMID:22247268; http://dx.doi.org/10.1104/pp.111.192393
25. Muralla R, Lloyd J, Meinke D. Molecular foundations of reproductive lethality in Arabidopsis thaliana. PLoS One 2011; 6:e28398; PMID:22164284; http://dx.doi.org/10.1371/journal.pone.0028398
26. Gutmann B, Gobert A, Giegé P. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes Dev 2012; 26:1022-7; PMID:22549728; http://dx.doi.org/10.1101/gad.189514.112
27. Ding YH, Liu NY, Tang ZS, Liu J, Yang WC. Arabidopsis GLUTAMINE-RICH PROTEIN23 is essential for early embryogenesis and encodes a novel nuclear PPR motif protein that interacts with RNA polymerase II subunit III. Plant Cell 2006; 18:815-30; PMID:16489121; http://dx.doi.org/10.1105/tpc.105.039495
28. Hammani K, Gobert A, Hleibieh K, Choulier L, Small I, Giegé P. An Arabidopsis dual-localized pentatricopeptide repeat protein interacts with nuclear proteins involved in gene expression regulation. Plant Cell 2011; 23:730-40; PMID:21297037; http://dx.doi.org/10.1105/tpc.110.081638
29. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS Biol 2007; 5:e117; PMID:17456005; http://dx.doi.org/10.1371/journal.pbio.0050117
30. Bratic A, Wredenberg A, Grönke S, Stewart JB, Mourier A, Ruzzenente B, Kukat C, Wibom R, Habermann B, Partridge L, et al. The bicoid stability factor controls polyadenylation and expression of specific mitochondrial mRNAs in Drosophila melanogaster. PLoS Genet 2011; 7:e1002324; PMID:22022283; http://dx.doi.org/10.1371/journal.pgen.1002324
31. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007; 2:953-71; PMID:17446895; http://dx.doi.org/10.1038/nprot.2007.131
32. Small I, Peeters N, Legeai F, Lurin C. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 2004; 4:1581-90; PMID:15174128; http://dx.doi.org/10.1002/pmic.200300776
33. Gavel Y, Nilsson L, von Heijne G. Mitochondrial targeting sequences. Why 'non-amphiphilic' peptides may still be amphiphilic. FEBS Lett 1988; 235:173-7; PMID:3402595; http://dx.doi.org/10.1016/0014-5793(88)81257-2
34. Roise D. Recognition and binding of mitochondrial presequences during the import of proteins into mitochondria. J Bioenerg Biomembr 1997; 29:19-27; PMID:9067798; http://dx.doi.org/10.1023/A:1022403604273
35. Waltner M, Hammen PK, Weiner H. Influence of the mature portion of a precursor protein on the mitochondrial signal sequence. J Biol Chem 1996; 271:21226-30; PMID:8702895; http://dx.doi.org/10.1074/jbc.271.35.21226
36. Huang S, Taylor NL, Whelan J, Millar AH. Refining the definition of plant mitochondrial presequences through analysis of sorting signals, N-terminal modifications, and cleavage motifs. Plant Physiol 2009; 150:1272-85; PMID:19474214; http://dx.doi.org/10.1104/pp.109.137885
37. Bionda T, Tillmann B, Simm S, Beilstein K, Ruprecht M, Schleiff E. Chloroplast import signals: the length requirement for translocation in vitro and in vivo. J Mol Biol 2010; 402:510-23; PMID:20688079; http://dx.doi.org/10.1016/j.jmb.2010.07.052
38. Bannai H, Tamada Y, Maruyama O, Nakai K, Miyano S. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 2002; 18:298-305; PMID:11847077; http://dx.doi.org/10.1093/bioinformatics/18.2.298
39. Nair R, Rost B. Mimicking cellular sorting improves prediction of subcellular localization. J Mol Biol 2005; 348:85-100; PMID:15808855; http://dx.doi.org/10.1016/j.jmb.2005.02.025
40. Höglund A, Dönnes P, Blum T, Adolph HW, Kohlbacher O. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. Bioinformatics 2006; 22:1158-65; PMID:16428265; http://dx.doi.org/10.1093/bioinformatics/btl002
41. Kaundal R, Saini R, Zhao PX. Combining machine learning and homology-based approaches to accurately predict subcellular localization in Arabidopsis. Plant Physiol 2010; 154:36-54; PMID:20647376; http://dx.doi.org/10.1104/pp.110.156851
42. Ferro M, Brugière S, Salvi D, Seigneurin-Berny D, Court M, Moyet L, Ramus C, Miras S, Mellal M, Le Gall S, et al. AT_CHLORO, a comprehensive chloroplast proteome database with subplastidial localization and curated information on envelope proteins. Mol Cell Proteomics 2010; 9:1063-84; PMID:20061580; http://dx.doi.org/10.1074/mcp.M900325-MCP200

43. Sun Q, Zybailov B, Majeran W, Friso G, Olinares PD, van Wijk KJ. PPDB, the Plant Proteomics Database at Cornell. Nucleic Acids Res 2009; 37(Database issue):D969-74; PMID:18832363; http://dx.doi.org/10.1093/nar/gkn654
44. Huang S, Taylor NL, Narsai R, Eubel H, Whelan J, Millar AH. Experimental analysis of the rice mitochondrial proteome, its biogenesis, and heterogeneity. Plant Physiol 2009; 149:719-34; PMID:19010998; http://dx.doi.org/10.1104/pp.108.131300
45. Huang M, Friso G, Nishimura K, Qu X, Olinares PD, Majeran W, Sun Q, van Wijk KJ. Construction of plastid reference proteomes for maize and Arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. J Proteome Res 2013; 12:491-504; PMID:23198870; http://dx.doi.org/10.1021/pr300952g
46. Hummel M, Cordewener JH, de Groot JC, Smeekens S, America AH, Hanson J. Dynamic protein composition of Arabidopsis thaliana cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics. Proteomics 2012; 12:1024-38; PMID:22522809; http://dx.doi.org/10.1002/pmic.201100413
47. Ito J, Batth TS, Petzold CJ, Redding-Johanson AM, Mukhopadhyay A, Verboom R, Meyer EH, Millar AH, Heazlewood JL. Analysis of the Arabidopsis cytosolic proteome highlights subcellular partitioning of central plant metabolism. J Proteome Res 2011; 10:1571-82; PMID:21166475; http://dx.doi.org/10.1021/pr1009433
48. Pendle AF, Clark GP, Boon R, Lewandowska D, Lam YW, Andersen J, Mann M, Lamond AI, Brown JW, Shaw PJ. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol Biol Cell 2005; 16:260-9; PMID:15496452; http://dx.doi.org/10.1091/mbc.E04-09-0791
49. Sakamoto Y, Takagi S. LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana. Plant Cell Physiol 2013; 54:622-33; PMID:23396599; http://dx.doi.org/10.1093/pcp/pct031
50. Alexandersson E, Saalbach G, Larsson C, Kjellbom P. Arabidopsis plasma membrane proteomics identifies components of transport, signal transduction and membrane trafficking. Plant Cell Physiol 2004; 45:1543-56; PMID:15574830; http://dx.doi.org/10.1093/pcp/pch209
51. Jaquinod M, Villiers F, Kieffer-Jaquinod S, Hugouvieux V, Bruley C, Garin J, Bourguignon J. A proteomics dissection of Arabidopsis thaliana vacuoles isolated from cell culture. Mol Cell Proteomics 2007; 6:394-412; PMID:17151019; http://dx.doi.org/10.1074/mcp.M600250-MCP200
52. Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Grossniklaus U, Schulze WX, Robatzek S, Panstruga R. PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. J Biol Chem 2010; 285:39140-9; PMID:20843791; http://dx.doi.org/10.1074/jbc.M110.160531
53. Li B, Takahashi D, Kawamura Y, Uemura M. Comparison of plasma membrane proteomic changes of Arabidopsis suspension-cultured cells (T87 Line) after cold and ABA treatment in association with freezing tolerance development. Plant Cell Physiol 2012; 53:543-54; PMID:22318864; http://dx.doi.org/10.1093/pcp/pcs010
54. Mitra SK, Walters BT, Clouse SD, Goshe MB. An efficient organic solvent based extraction method for the proteomic analysis of Arabidopsis plasma membranes. J Proteome Res 2009; 8:2752-67; PMID:19334764; http://dx.doi.org/10.1021/pr801044y
55. Szponarski W, Sommerer N, Boyer JC, Rossignol M, Gibrat R. Large-scale characterization of integral proteins from Arabidopsis vacuolar membrane by two-dimensional liquid chromatography. Proteomics 2004; 4:397-406; PMID:14760709; http://dx.doi.org/10.1002/pmic.200300607
56. Zhang ZJ, Peck SC. Simplified enrichment of plasma membrane proteins for proteomic analyses in Arabidopsis thaliana. Proteomics 2011; 11:1780-8; PMID:21433285; http://dx.doi.org/10.1002/pmic.201000648
57. Narsai R, Law SR, Carrie C, Xu L, Whelan J. In-depth temporal transcriptome profiling reveals a crucial developmental switch with roles for RNA processing and organelle metabolism that are essential for germination in Arabidopsis. Plant Physiol 2011; 157:1342-62; PMID:21908688; http://dx.doi.org/10.1104/pp.111.183129
58. Gobert A, Gutmann B, Taschner A, Gössringer M, Holzmann J, Hartmann RK, Rossmanith W, Giegé P. A single Arabidopsis organellar protein has RNase P activity. Nat Struct Mol Biol 2010; 17:740-4; PMID:20473316; http://dx.doi.org/10.1038/nsmb.1812
59. Majeran W, Friso G, Asakura Y, Qu X, Huang M, Ponnala L, Watkins KP, Barkan A, van Wijk KJ. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. Plant Physiol 2012; 158:156-89; PMID:22065420; http://dx.doi.org/10.1104/pp.111.188474
60. Hashimoto M, Endo T, Peltier G, Tasaka M, Shikanai T. A nucleus-encoded factor, CRR2, is essential for the expression of chloroplast ndhB in Arabidopsis. Plant J 2003; 36:541-9; PMID:14617084; http://dx.doi.org/10.1046/j.1365-313X.2003.01900.x
61. Meierhoff K, Felder S, Nakamura T, Bechtold N, Schuster G. HCF152, an Arabidopsis RNA binding pentatricopeptide repeat protein involved in the processing of chloroplast *psbB-psbT-psbH-petB-petD* RNAs. Plant Cell 2003; 15:1480-95; PMID:12782738; http://dx.doi.org/10.1105/tpc.010397
62. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast rps12 pre-mRNA. Plant Cell 2006; 18:2650-63; PMID:17041147; http://dx.doi.org/10.1105/tpc.106.046110
63. Gutiérrez-Marcos JF, Dal Prà M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, et al. empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. Plant Cell 2007; 19:196-210; PMID:17259266; http://dx.doi.org/10.1105/tpc.105.039594
64. Cai W, Ji D, Peng L, Guo J, Ma J, Zou M, Lu C, Zhang L. LPA66 is required for editing psbF chloroplast transcripts in Arabidopsis. Plant Physiol 2009; 150:1260-71; PMID:19448041; http://dx.doi.org/10.1104/pp.109.136812
65. Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T. Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in Arabidopsis chloroplasts. Plant Cell 2009; 21:146-56; PMID:19182104; http://dx.doi.org/10.1105/tpc.108.064667
66. Laluk K, Abuqamar S, Mengiste T. The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. Plant Physiol 2011; 156:2053-68; PMID:21653783; http://dx.doi.org/10.1104/pp.111.177501
67. Takenaka M. MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the nad7 transcript in mitochondria of Arabidopsis. Plant Physiol 2010; 152:939-47; PMID:20018598; http://dx.doi.org/10.1104/pp.109.151175
68. Hammani K, Okuda K, Tanz SK, Chateigner-Boutin AL, Shikanai T, Small I. A study of new Arabidopsis chloroplast RNA editing mutants reveals general features of editing factors and their target sites. Plant Cell 2009; 21:3686-99; PMID:19934379; http://dx.doi.org/10.1105/tpc.109.071472
69. Sosso D, Mbelo S, Vernoud V, Gendrot G, Dedieu A, Chambrier P, Dauzat M, Heurtevin L, Guyon V, Takenaka M, et al. PPR2263, a DYW-Subgroup Pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. Plant Cell 2012; 24:676-91; PMID:22319053; http://dx.doi.org/10.1105/tpc.111.091074
70. Zhu Q, Dugardeyn J, Zhang C, Takenaka M, Kühn K, Craddock C, Smalle J, Karampelias M, Denecke J, Peters J, et al. SLO2, a mitochondrial pentatricopeptide repeat protein affecting several RNA editing sites, is required for energy metabolism. Plant J 2012; 71:836-49; PMID:22540321; http://dx.doi.org/10.1111/j.1365-313X.2012.05036.x
71. Chateigner-Boutin AL, Colas des Francs-Small C, Fujii S, Okuda K, Tanz SK, Small I. The E domains of pentatricopeptide repeat proteins from different organelles are not functionally equivalent for RNA editing. Plant J 2013; 74:935-45; PMID:23521509; http://dx.doi.org/10.1111/tpj.12180
72. Heazlewood JL, Tonti-Filippini JS, Gout AM, Day DA, Whelan J, Millar AH. Experimental analysis of the Arabidopsis mitochondrial proteome highlights signaling and regulatory components, provides assessment of targeting prediction programs, and indicates plant-specific mitochondrial proteins. Plant Cell 2004; 16:241-56; PMID:14671022; http://dx.doi.org/10.1105/tpc.016055
73. Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjölander K, Gruissem W, Baginsky S. The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol 2004; 14:354-62; PMID:15028209; http://dx.doi.org/10.1016/j.cub.2004.02.039
74. Klodmann J, Senkler M, Rode C, Braun HP. Defining the protein complex proteome of plant mitochondria. Plant Physiol 2011; 157:587-98; PMID:21841088; http://dx.doi.org/10.1104/pp.111.182352
75. Peltier JB, Ytterberg AJ, Sun Q, van Wijk KJ. New functions of the thylakoid membrane proteome of Arabidopsis thaliana revealed by a simple, fast, and versatile fractionation strategy. J Biol Chem 2004; 279:49367-83; PMID:15322131; http://dx.doi.org/10.1074/jbc.M406763200
76. Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS One 2008; 3:e1994; PMID:18431481; http://dx.doi.org/10.1371/journal.pone.0001994
77. Murayama M, Hayashi S, Nishimura N, Ishide M, Kobayashi K, Yagi Y, Asami T, Nakamura T, Shinozaki K, Hirayama T. Isolation of Arabidopsis ahg11, a weak ABA hypersensitive mutant defective in nad4 RNA editing. J Exp Bot 2012; 63:5301-10; PMID:22821940; http://dx.doi.org/10.1093/jxb/ers188
78. Tang J, Kobayashi K, Suzuki M, Matsumoto S, Muranaka T. The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. Plant J 2010; 61:456-66; PMID:19929879; http://dx.doi.org/10.1111/j.1365-313X.2009.04082.x
79. Verbitskiy D, Zehrmann A, van der Merwe JA, Brennicke A, Takenaka M. The PPR protein encoded by the LOVASTATIN INSENSITIVE 1 gene is involved in RNA editing at three sites in mitochondria of Arabidopsis thaliana. Plant J 2010; 61:446-55; PMID:19919573; http://dx.doi.org/10.1111/j.1365-313X.2009.04076.x
80. Carrie C, Small I. A reevaluation of dual-targeting of proteins to mitochondria and chloroplasts. Biochim Biophys Acta 2013; 1833:253-9; PMID:22683762; http://dx.doi.org/10.1016/j.bbamcr.2012.05.029
81. Xu L, Carrie C, Law SR, Murcha MW, Whelan J. Acquisition, conservation, and loss of dual-targeted proteins in land plants. Plant Physiol 2013; 161:644-62; PMID:23257241; http://dx.doi.org/10.1104/pp.112.210997

82. Hammani K, des Francs-Small CC, Takenaka M, Tanz SK, Okuda K, Shikanai T, Brennicke A, Small I. The pentatricopeptide repeat protein OTP87 is essential for RNA editing of nad7 and atp1 transcripts in Arabidopsis mitochondria. J Biol Chem 2011; 286:21361-71; PMID:21504904; http://dx.doi.org/10.1074/jbc.M111.230516
83. Howard MJ, Lim WH, Fierke CA, Koutmos M. Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5' processing. Proc Natl Acad Sci U S A 2012; 109:16149-54; PMID:22991464; http://dx.doi.org/10.1073/pnas.1209062109
84. Gothandam KM, Kim ES, Cho H, Chung YY. OsPPR1, a pentatricopeptide repeat protein of rice is essential for the chloroplast biogenesis. Plant Mol Biol 2005; 58:421-33; PMID:16021404; http://dx.doi.org/10.1007/s11103-005-5702-5
85. Yogev O, Pines O. Dual targeting of mitochondrial proteins: mechanism, regulation and function. Biochim Biophys Acta 2011; 1808:1012-20; PMID:20637721; http://dx.doi.org/10.1016/j.bbamem.2010.07.004
86. Duchêne AM, Giegé P. Dual localized mitochondrial and nuclear proteins as gene expression regulators in plants? Front Plant Sci 2012; 3:221; PMID:23056004; http://dx.doi.org/10.3389/fpls.2012.00221
87. Krause K, Krupinska K. Nuclear regulators with a second home in organelles. Trends Plant Sci 2009; 14:194-9; PMID:19285907; http://dx.doi.org/10.1016/j.tplants.2009.01.005
88. Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. Plant Mol Biol 2000; 42:819-32; PMID:10890530; http://dx.doi.org/10.1023/A:1006496308160
89. Chupeau Y, Bourgin JP, Missonier C, Dorion N, Morel G. Preparation and culture of protoplasts from various Nicotiana species. Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences 1974; D 278:1565-8