Article

MISF2 Encodes an Essential Mitochondrial Splicing Cofactor Required for *nad2* mRNA Processing and Embryo Development in *Arabidopsis thaliana*

Tan-Trung Nguyen [1,†]**, Corinne Best** [2,†]**, Sofia Shevtsov** [2]**, Michal Zmudjak** [2]**, Martine Quadrado** [1]**, Ron Mizrahi** [2]**, Hagit Zer** [2]**, Hakim Mireau** [1,*] **and Oren Ostersetzer-Biran** [2,*]

1   Institut Jean-Pierre Bourgin INRAE, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France
2   Department of Plant and Environmental Sciences, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Givat-Ram, Jerusalem 91904, Israel
†   These authors have equally contributed to the MS.
*   Correspondence: Hakim Mireau, E-mail, hakim.mireau@versailles.inra.fr, and Oren Ostersetzer-Biran. E-mail, oren.ostersetzer@mail.huji.ac.il.

Abstract: Mitochondria play key roles in cellular energy metabolism in eukaryotes. Mitochondria of most organisms contain their own genome and specific transcription and translation machineries. The expression of angiosperm mtDNA involves extensive RNA-processing steps, such as RNA trimming, editing, and the splicing of numerous group II-type introns. Pentatricopeptide repeat (PPR) proteins are key players of plant organelle gene expression and RNA metabolism. In the present analysis, we reveal the function of the <u>MI</u>TOCHONDRIAL <u>S</u>PLICING <u>F</u>ACTOR 2 gene (*MISF2*, *AT3G22670*) and show that it encodes a mitochondria-localized PPR protein that is crucial for early embryo-development in Arabidopsis. Molecular characterization of embryo-rescued *misf2* plantlets indicates that the splicing of *nad2* intron 1 and thus respiratory complex I biogenesis are strongly compromised. Moreover, the molecular function seems conserved between MISF2 protein in Arabidopsis and its orthologous gene (*EMP10*) in maize, suggesting that the ancestor of MISF2/EMP10 was recruited to function in *nad2* processing before the monocot-dicot divergence, ~200 million years ago. These data provide new insights into the function of nuclear-encoded factors in mitochondrial gene expression and respiratory chain biogenesis during plant embryo development.

Keywords: group II; intron; splicing; PPR; respiration; complex I; mitochondria; embryogenesis; arabidopsis; angiosperms

1. Introduction

Mitochondria are key sites of cellular energy metabolism (i.e., ATP production), as well as for the biosynthesis of various essential metabolites. Most modern mitochondria contain vestigial genomes (mtDNA, mitogenome) derived from that of their ancestral bacterial progenitor, and which vary quite widely in size between organisms. In plants, angiosperm mtDNAs are remarkably large and complex in structure [1], encoding rRNAs, tRNAs, ribosomal proteins, as well as various subunits of respiratory complexes (CI to CIV), the ATP synthase enzyme (CV), cofactors of the cytochrome c biogenesis (CCM) machinery, and at least one component of the twin-arginine protein translocation system [2].

In *Arabidopsis*, the oxidative phosphorylation (OXHPOS) machinery is composed of >100 different subunits, most of which are encoded by nuclear loci and about 20 are expressed from the mitogenome. The biogenesis of the respiratory chain machinery thus involves various mechanisms for regulating the expression of subunits that are encoded

by two physically separate genetic compartments [3-5]. However, the identity of factors and pathways involved in these regulations remain largely elusive.

The expression of mitochondrial genes in plants involves extensive RNA-processing steps. These include RNA trimming, RNA editing, and the removal of many intron sequences that reside in various genes [6-9]. These RNA processing steps are essential for mt-RNAs to synthesize the protein they encode. While the excision of canonical group II introns rely on proteins encoded by the introns themselves (*i.e.*, IEPs, or maturases) [10, 11], the splicing of the plant organellar group II introns involves a repertoire of nuclear-encoded factors that assist splicing reactions and which may serve as key control points in plant mitochondrial gene expression [1, 12-14]. These belong to diverse families of RNA binding factors. A few are related to maturases [14, 15], whereas others were identified as *e.g.* RNA helicases [16-18], PORR-related proteins [19], and relevantly to our study also pentatricopeptide repeat (PPR) proteins [20].

The PPR family constitutes a large protein family in land plants, with approximately 450 members identified in Arabidopsis and 490 genes in maize [21-23]. PPR proteins are recognized by a degenerate 35 amino-acid motif folding into two antiparallel helices connected by a short loop/turn [24, 25]. PPRs have been shown to play multifarious functions in organellar RNA metabolism, such as RNA stability and protection [8, 26], C-to-U RNA editing [9], mRNA translation (Haili et al 2016, Waltz et al, 2019, Nguyen et al, 2021) and group II intron splicing [6, 8, 27-29]. Members of the PPR family are also linked to fertility restoration, where they regulate the expression of mitochondrial CMS-associated ORFs [30, 31].

Genetic and biochemical studies indicated that PPRs are sequence-specific RNA-binding trans-factors, and that RNA recognition is mostly mediated by amino acids found at positions 5 and 35 in each PPR repeat. Association with each of the four RNA bases involves specific amino acid combinations that are the basis of a PPR protein-RNA recognition code [32, 33]. These data were further supported by the analysis of PPR protein-RNA crystal structures [24, 25, 34-37]. PPRs are classified into two main groups: P and PLS type proteins, which in addition to canonical 35-amino acid PPR motifs (P) include long (L) or short (S) repeat variants [22, 38]. While PLS-type proteins are almost exclusively associated with RNA editing [9, 39], P-type PPR factors facilitate a wide range of organellar RNA expression steps going from stabilization to translation [27, 40]. In this work, we analyzed the function of a mitochondrial P-type PPR factor that we named *MITOCHONDRIAL INTRON SPLICING FACTOR* 2 (*MISF2*). The *Arabidopsis* MISF2 protein is homologous to the PPR protein EMP10 in *Zea mays* [41]. We show that MISF2 is required for the splicing of *nad2* intron 1, and that the respiratory complex I biogenesis is strongly destabilized in *misf2* mutants. The molecular function of *MISF2* is conserved in both Arabidopsis and maize   [41], suggesting that the common ancestor *MISF2* was recruited to function in *nad2* splicing prior to the divergence of monocot and dicot plants, *i.e.* about ~200 million years ago [42].

2. Results

2.1. The topology of MISF2 protein

To better understand processes associated with mitochondrial RNA (mt-RNA) expression in plants, we assembled a collection of Arabidopsis T-DNA mutants affected in genes encoding mitochondria-targeted P-type PPR proteins and identified that heterozygous plants carrying insertions in the At3g22670 gene could not set homozygous mutants in their progeny. Domain search analysis using the PPRfinder [43], PPRCODE [33], SMART [44] and CDD [45] algorithms indicated that the deduced product of *AT3G22670* gene (Figure 1, Supplementary Figure S1) encodes a 562 amino acid PPR protein with a predicted topology of NH_2-165-P-3-P-P-P-P-P-P-P-P-P-42-COOH (where 'P' designates P-type PPR motifs and amino acids not assigned to any defined domain are specified by numbers) (Figure 1, Figure S1a).

Subcellular localization prediction algorithms, available at the ExPASy portal (https://www.expasy.org), UniProt [46] and the 'SUBcellular location database for Arabidopsis proteins' (SUBA4, http://suba.live) [47], indicate the presence of a predicted 24-amino acid mitochondrial targeting sequence in the N-terminal region of MISF2 (Figure S1a). In silico 3D structure prediction, using the AlphaFold server [48], suggested that MISF2 harbors a typical PPR helical fold (Figure S1a), with an inner basic core representing the RNA binding surface, as previously indicated from the analysis of the plant PPR10 protein [25].
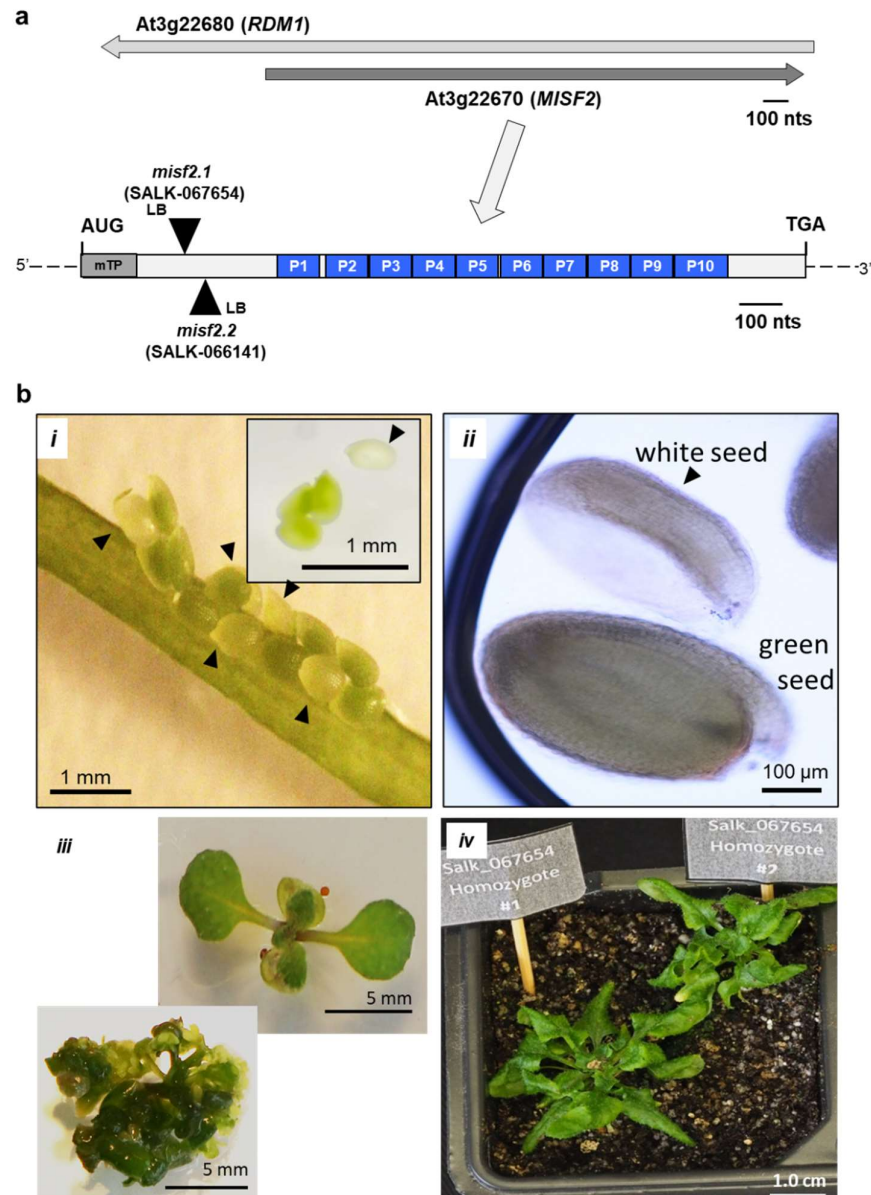

Figure 1. MISF2 (At3g22670) gene topology and *misf2* mutant phenotypes. (**a**) Scheme of the At3g22670 locus and gene structure. The position of the two T-DNA insertion sites in the coding region of *MISF2* (*i.e.*, SALK-line 067654, *misf2.1* and SALK-line 066141, *misf2.2*) are located 324 and 350 nucleotides downstream to the ATG start codon and which correspond to the N-terminal region, upstream to the PPR motifs of the MISF2 protein. (**b**) Morphologies of *misf2* mutants. Green (wild type or heterozygous) and white (homozygous) seeds were collected from surface-sterilized immature siliques of heterozygous *misf2* plants (*i*) and sown on MS agar media supplemented with

vitamins plants. Arrows point toward white seeds. Panel B-***ii*** shows differential interference contrast microscopy images (*i.e.*, Nomarski) of the embryos found in green or white seeds. Following germination, few rescued *misf2* plantlets (***iii***) were able to survive on soil, although failed to flower and set viable seeds (***iv***).

2.2. MISF2 encodes a lowly-expressed P-type PPR protein that is localized in mitochondria

Expression analysis of *MISF2* were made using publicly available microarray and high-throughput sequencing databases. The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org) and 'Genevestigator analysis toolbox' [49] databases indicated differential expression of *MISF2* gene throughout development, with *MISF2* expression being dominant in embryonic organs, young developing leaves, apical root tissues, flowers and the shoot apex (Figure S2). To further investigate the intracellular location of MISF2, a fragment comprising the first 203 amino acids of MISF2 was fused in-frame to GFP (MISF2-GFP), expressed in Arabidopsis plants and the subcellular localization of the resulting fluorescence examined by confocal microscopy (Figure 2). In agreement with the *in-silico* data, the MISF2-GFP signal was detected as round-shaped particles that co-localized with those of the MitoTracker® marker, a mitochondrion-specific fluorescent probe (Figure 2). These results are consistent with the predicted mitochondrial targeting of MISF2.
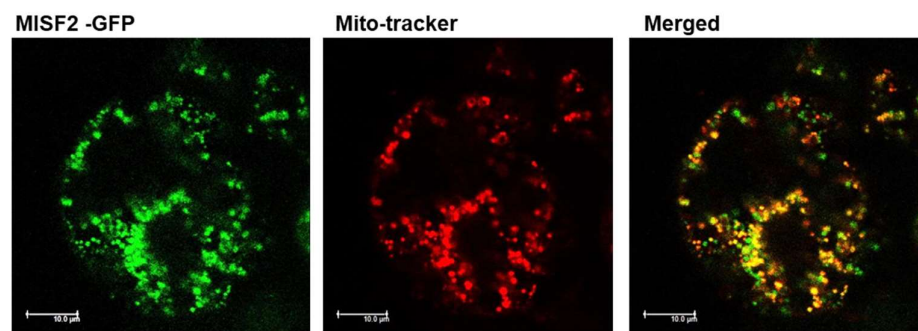

Figure 2. MISF2 is localized to the mitochondria. Arabidopsis plant cells were transformed with a construct expressing the GFP fused in frame to the N-terminal region (*i.e.*, 203 amino acids) of the Arabidopsis MISF2 protein sequence. The fluorescence corresponding to the GFP (green, left), the MitoTracker® marker (red, center) and the merged signals (right) are shown.

2.3. MISF2 is required for early embryo development in Arabidopsis thaliana

Two T-DNA insertion lines were identified within the *MISF2* gene. The SALK_067654 (*misf2.1*) and SALK_066141 (*misf2.2*) contain T-DNA insertions located 324 and 350 nucleotides downstream of *MISF2* translational start, respectively (Figure 1a, and Figure S3a). No homozygous plants could be recovered from the progeny of heterozygote *misf2* lines, suggesting that the At3g22670 gene is an essential gene that is critical during embryogenesis. Heterozygous *misf2.1* and *misf2.2* mutant lines did not show any obvious phenotypes under normal growth conditions (see Material and Methods) suggesting that homozygous mutants could be embryonic lethal. To test this assumption, we compared the developmental phenotypes of embryos contained in immature seeds (10 days after pollination) of heterozygous *misf2* and wild type plants. Siliques of heterozygous *misf2* plants contained about one-quarter of yellow to white seeds (Figure 1b-*i*), which later degenerated into shrunken and brown mature seeds. Microscopy analyses further indicated that green seeds in siliques of heterozygous *misf2* plants contained fully developed embryos, while white seeds had embryos arrested at the late torpedo/walking stick stages (Figures 1b-*i* and 1b-*ii*).

2.4. MISF2 is essential for nad2 pre-mRNAs processing in Arabidopsis mitochondria

Although *misf2* is not found among the 32 Arabidopsis embryo-defective *ppr* mutants of the 'SeedGene' database [50], our genetic and microscopic analyses indicate that *MISF2* is essential for proper embryo development (Figure 1b). Embryo rescue approaches enable the establishment of certain Arabidopsis mutants showing germination-defective phenotypes [51]. Among these are a few mutants affected in mitochondria biogenesis and function, such as the *cod1* [52], *ndufv1* [53] *cal1/cal2* [54, 55] or *nmat3* mutants [56]. Therefore, white seeds contained in young siliques of heterozygous m*isf2* plants (*i.e.,* 10~12 days post-anthesis, DPA), were sown on MS-agar plates supplemented with 1% sucrose and various vitamins (see Materials and Methods) and then transferred to controlled growth chamber. Indeed, under these conditions, 30% of the white seeds have germinated after 3 months of culture and were then transferred to liquid culture using the same medium (see Materials and Methods). PCR genotyping indicated that while green seeds derived from *misf2.1* or *misf2.2* heterozygote plants were either wild type or heterozygous for the mutations, plantlets obtained from white seeds were all homozygous for either of the two *misf2* mutant alleles.

The optimal conditions for rescue of *misf2.1* seedlings were similar to those reported for the *nmat3* mutant [56], whereas rescue of *misf2.2* was achieved following the rescue method previously described for the Arabidopsis *cod1* mutant [52]. Phenotypical variations between rescued *misf2* plantlets were visible, as certain seedlings developed into slow-growing normal-looking plants with twisted leaves (Figure 1b-*iii*), while others produced miniature bushy-like structures (Figure 1b-*iii*), as previously reported for several other *emb* mutants [57], notably affected in mitochondria biogenesis [7, 52, 56]. Few homozygous *misf2* plantlets (*e.g.*, Figure 1b-*iv*) could be further transferred and cultivated on soil, but none could produce viable seeds. For RNA and protein analyses, we used 3-week-old mutant plantlets [57]. *MISF2* gene is partially overlapping with *RDM1* (At3g22680) (Figure 1a), a nuclear DNA methyltransferase, which its functions in Arabidopsis are not essetial during embryogenesis [58]. To further establish the specific roles of MISF2 in mitochondria biogenesis, we generated a 'complemented' homozygous *misf2.2* line expressing the native *AtMISF2* gene (*misf2.2/35S::MISF2*).

The steady-state levels of mitochondrial mRNAs in *misf2.1* and *misf2.2* mutants were analyzed by RT-qPCR in comparison to wild type (Col-0) plants. This analysis revealed a strong reduction (*i.e.*, about 70 to 1,200 folds) in the accumulation of mature *nad2* transcripts spliced from its first intron in *misf2.1* and *misf2.2* mutants, respectively (*i.e.*, *nad2* exons a-b, Figure 3a). The steady-state levels of most other mitochondrial transcripts, including *nad2* transcripts spliced from its other introns were found to over-accumulate from 2 to 5 folds in both *misf2* mutant lines (Figure 3a). The RNA profiles in plantlets derived from immature wild type embryos and grown under the same conditions as *misf2* mutants did not show any significant reductions in the accumulation of mitochondrial transcripts, including *nad2* (Figure 3b). Similarly, the accumulation of *nad2* transcripts in functionally complemented *misf2.2* plants were globally equivalent to those in wild type plants (Figure 3b and Figure S3b). Based on these data, we concluded that the maturation defects observed for *nad2* transcripts in *misf2.1* and *misf2.2* plants relate to the functions of MISF2 and not to physiological differences between the embryo-rescued plantlets and 3-week-old Arabidopsis seedlings germinated on MS-media plates.
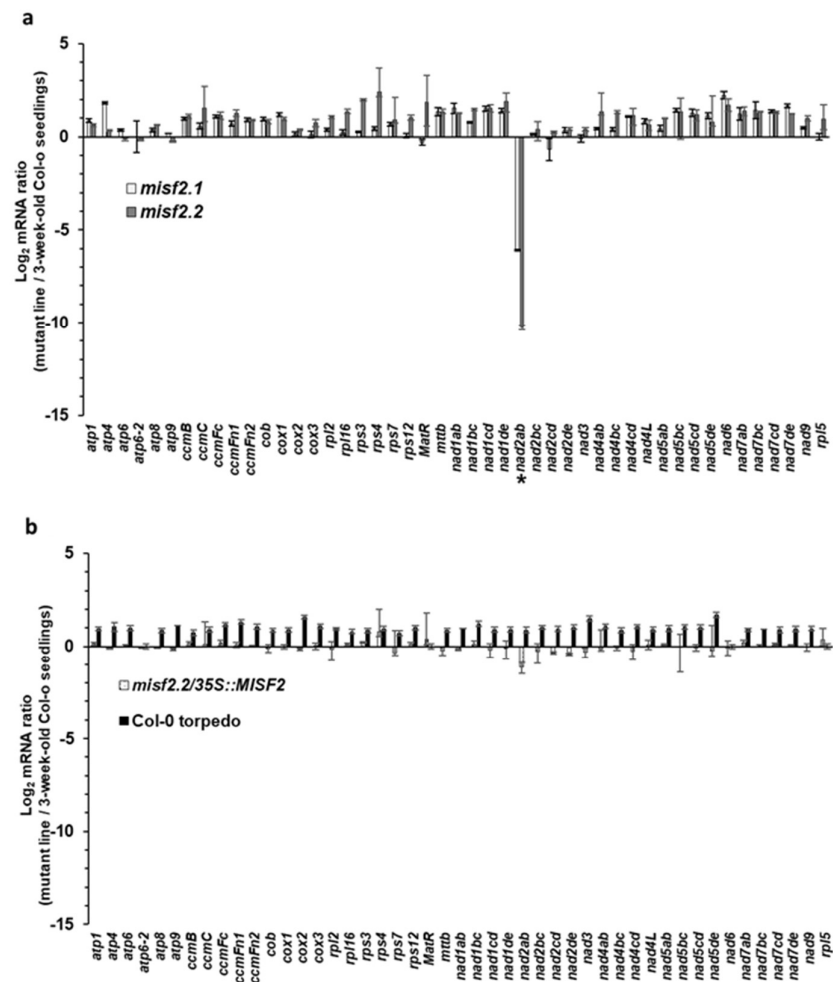
Figure 3. Relative accumulation of mitochondrial mRNAs in *misf2* mutants. Transcriptome analyses of mt-RNAs levels in *Arabidopsis* wild type (Col-0) and *misf2* mutant plants by RT-qPCR. RNA extracted from 3-week-old wild type seedlings (Col-0), 4-months-old rescued *misf2* mutants, plantlets derived from immature Col-0 seeds (*i.e.*, at the torpedo stage) and functionally complemented *misf2* mutants were reverse-transcribed and the relative steady-state levels of cDNAs corresponding to mitochondrial mRNAs evaluated by qPCR. Log$_2$ ratios of mt mRNA abundances in *misf2.1* and *misf2.2* mutant lines (**a**), plantlets derived from immature Col-0 seeds and complementation line (**b**) to those of 3-week-old MS-agar grown wild type plants. The values are means of three biological replicates (error bars indicate one standard deviation).

2.5. MISF2 is required for efficient splicing of nad2 intron 1

We reasoned that the reduced steady-state levels observed for the upstream region of mature *nad2* transcripts (*i.e.*, spliced exons 'a' and 'b') in *misf2* mutants likely relate to defects in the excision of the first intron in *nad2*. We thus determined the splicing efficiencies of *nad2* intron 1 and that of the other 22 mitochondrial introns in wild type plants and germinated embryos, as well as in *misf2* mutant and functionally complemented *misf2.2* plants by RT-qPCR. The obtained data revealed a strong reduction in the splicing efficiency of *nad2* intron 1, with splicing reductions reaching about 360 and 11,000 folds in *misf2.1* and *misf2.2* plants, respectively, compared with the wild type (Figure 4a).
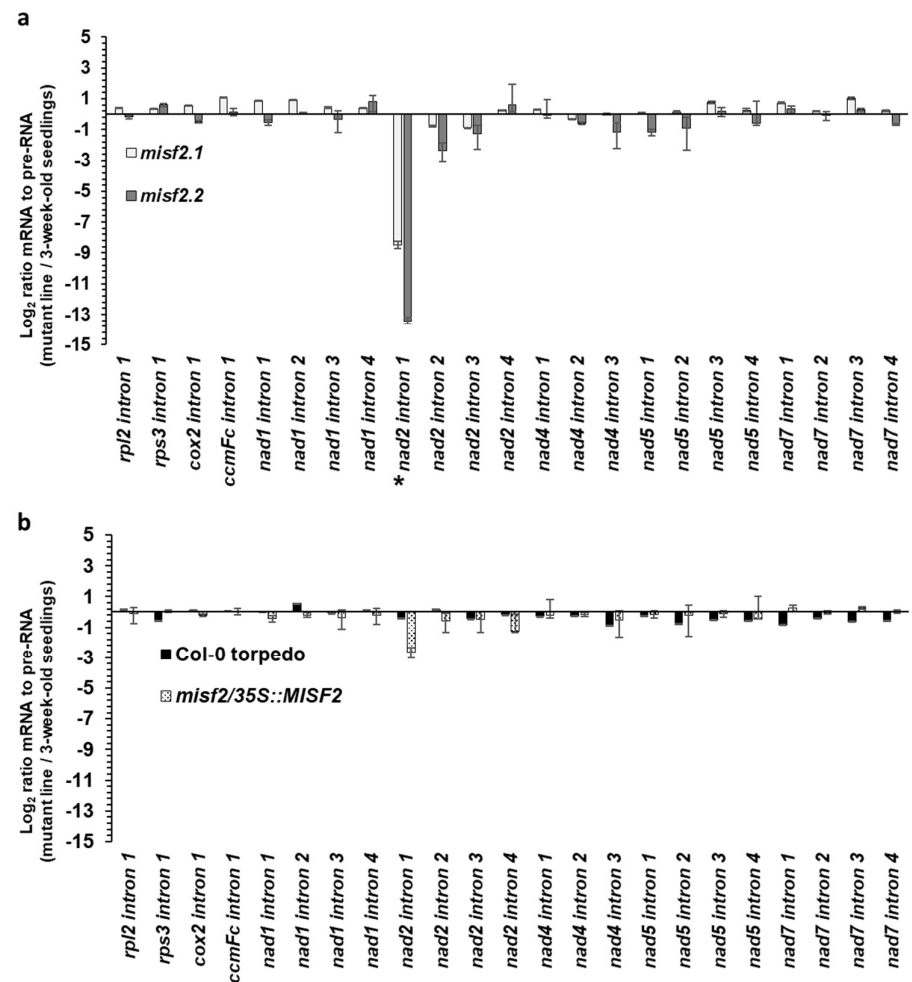
Figure 4. Splicing efficiencies in *misf2* mutants. The relative accumulation of mRNA and pre-RNA transcripts in wild type and *misf2* plants, corresponding to the 23 group II intron sequences in Arabidopsis, was evaluated by RT-qPCR. The histogram shows the splicing efficiencies as indicated by the $\log_2$ ratios of pre-RNA to mature mRNA transcript abundance in *misf2-1* and *misf2-2* mutant lines compared with those in wild type plants (**a**), as well as germinated wild type seeds collected at the torpedo stage (Col-0-torpedo) and complemented *misf2-2* line compared with those of wild type plants (**b**). The values are means of three (*misf2.1*) and five (*misf2.2*) biological replicates (error bars indicate one standard deviation). .

In contrast to *nad2* intron 1, the splicing efficiency of other mitochondrial transcripts was not significantly affected in *misf2* mutants, although small reductions (*i.e.*, 2.5 to 6.7 folds) in the splicing efficiencies of *nad2* introns 2 and 3 were seen in *misf2.2* plants. The reduction in *nad2* intron 1 splicing observed in *misf2* plants was largely corrected in complemented *misf2.2* plants expressing the native *AtMISF2* gene (*misf2.2/35S::MISF2*), strongly supporting the role for MISF2 in the processing of *nad2* intron 1 pre-mRNA (Figure 4b).

2.6. The MISF2 protein associates with nad2 intron 1 in vivo

PPR proteins are known to be sequence-specific RNA-binding factors [24, 27, 32, 59-61]. A combinatorial code for RNA-recognition by PPR proteins was proposed, based on combinations of amino acids found at positions 5 and 35 of each PPR repeat [32, 61, 62]. Using this code for the prediction of binding sites for the 10 PPR repeats of MISF2 (Figure 1, Figure S1) suggested the following sequence: 5′-N-(G>>A>U)-(U>C>G)-(U>C>G)-

(A>G)-(G>>A)-(G>>A>U)-(C>U)-G-N-3'. BLAST search along the updated *Arabidopsis* mtDNA (BK010421) revealed a matching 10-nucleotide sequence within domain I of *nad2* intron 1 (Figure 5a).
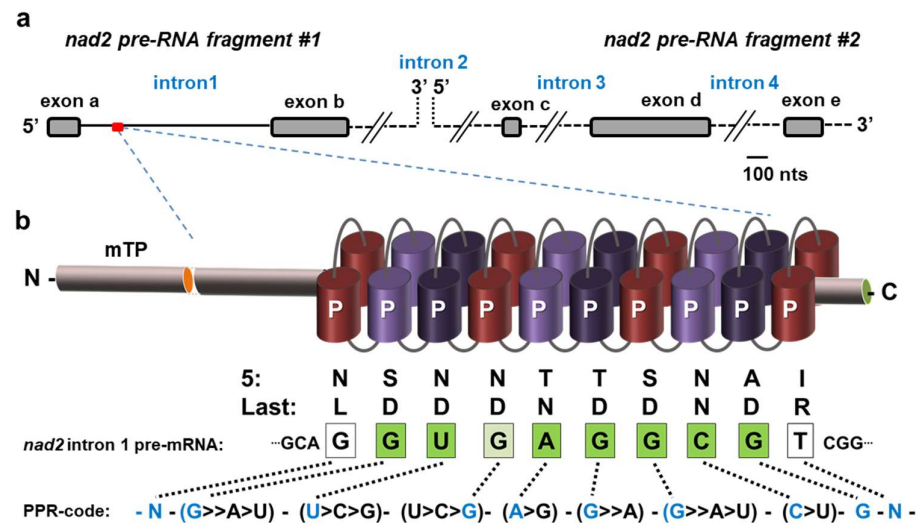

Figure 5. The predicted MISF2 binding site within domain I of *nad2* intron 1. (a) The expression of *nad2* in Arabidopsis mitochondria involves the transcription of two precursor RNA transcripts, which are divided by the second intron. The maturation of and mRNA requires the splicing of four group II introns found in both *cis* (*nad2* introns 1, 3 and 4) and *trans* (intron 2) configurations. The first pre-RNA fragment consists of two exons separated by intron 1, while the second fragment harbors three exons separated by introns 3 and 4. (b) MISF2 is a P-type PPR protein, which harbors a mitochondrial localization signal (mTP) and 10 PPR motifs. The fifth and the last amino acids of each PPR repeat (Figure S1) are indicated below each PPR repeats. The best corresponding RNA binding site (*i.e.*, 5'- GGUGAGGCGU -3') is indicated within the first intron of *nad2* pre-RNA fragment #1, with bases marked in green for perfect matches to the proposed binding site, in pale green for partial matches, and white for unassigned nucleotides.

No other sequences of 10 bases long corresponding to the predicted MISF2 binding site could be identified elsewhere in the plant mitogenome. A model for the asscoation of MISF2 with its predicted RNA bindng site within *nad2* intron 1 is illustrated in (Figure 5b). The in-silico data, therefore, correlated with the 'genetically defined' RNA target of MISF2, *nad2* intron 1 (Figures 4 and 5).

To further examine the in vivo targets of MISF2, an HA-tagged version of MISF2 was expressed in Arabidopsis. After confirming the expression of the tagged protein in vivo (Figure 6a), the MISF2-3HA protein was immunoprecipitated from total extracts (Figure 6b) and co-purified RNAs were analyzed by RT-qPCR (Figure 6c).
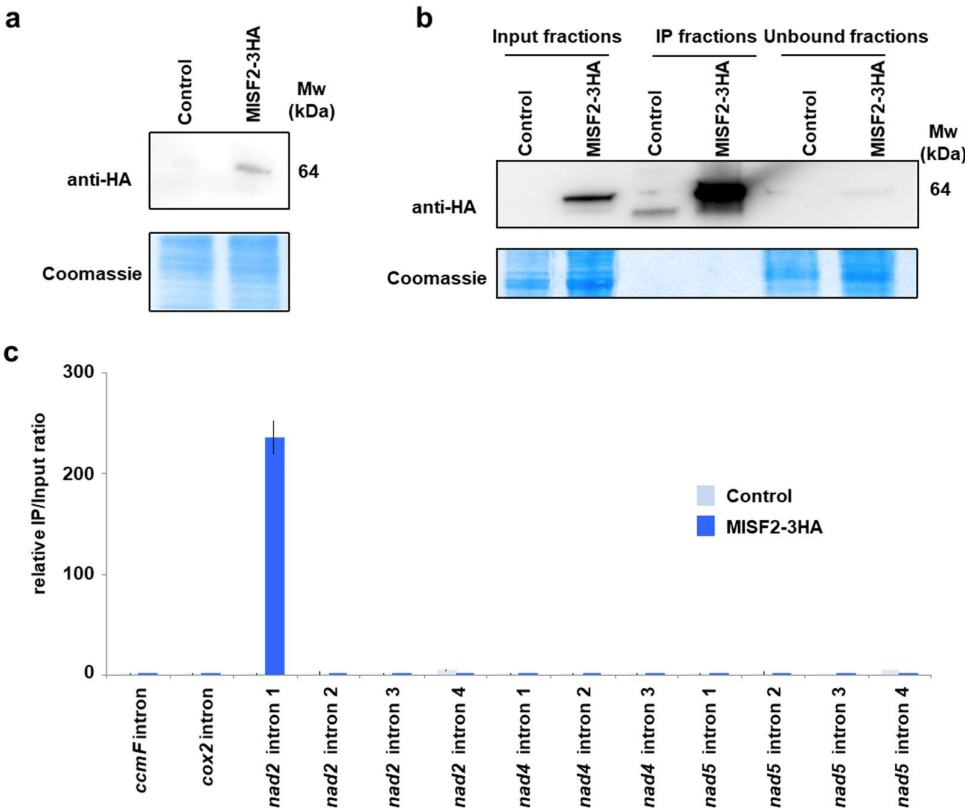
Figure 6. The MISF2 protein associates with *nad2* intron 1 in vivo. (**a**) Immunodetection of the MISF2-3HA fusion protein in protein extracts prepared form untransformed (control) and transformed (MISF2-3HA) Arabidopsis cell cultures (**b**) Immunoprecipitation assays were conducted with the anti-HA antibody and the shown immunoblot analysis attests for the strong enrichment of the MISF2-3HA fusion in the immunoprecipitated (IP) fraction derived from the Arabidopsis transgenic cell line expressing the fusion. The weak MISF2-3HA signal in the unbound fraction demonstrates the efficiency of the immunoprecipitation.   Parts of the blots stained with Coomassie blue are shown to display equal loading between samples. (**c**) Co-immunoprecipitated RNAs were analyzed by qRT-PCR using primer pairs specific to the indicated mitochondrial introns and relative enrichment ratios (immunoprecipitation fraction/input fraction) are shown.

Primers amplifying *nad2* intron 1 was used in this analysis, along with other primers pairs targeting introns whose splicing was found to be slightly reduced in *misf2* plants and additional controls. The obtained results revealed a very strong co-enrichment of *nad2* intron 1 specifically in the MISF2-3HA immunoprecipitate. None of the other tested introns (i.e., the single introns within *ccmF* or *cox2* mRNAs, *nad2* introns 2, 3 and 4, *nad4* introns 1 to 3, or *nad5* introns 1 to 4) were co-enriched with MISF2-3HA, strongly supporting that *nad2* intron1 is the *in vivo* RNA target of this PPR protein and thereby confirming that MISF2 specifically associates with its genetically defined intron RNA.

2.7. Analysis of the biogenesis of the respiratory chain in misf2 mutants

The respiratory system of plant cells is made of 5 major protein complexes, termed as complex I (CI, about 1,000 kDa in size), CII (160 kD), dimeric complex III (III$_2$, 500 kDa), CIV (200 and 220 kDa forms), and the ATP synthase (CV, 660 kDa) [63]. Plant mitochondria also harbor various enzymes that belong to the 'alternative electron transport' pathway, involving alternative NADH dehydrogenases and the alternative cytochrome oxydase [64]. Genetic and biochemical studies showed that Nad2 is essential for complex I (CI) biogenesis and function [65-72]. The reduction in *nad2* splicing (Figures 3 and 4) suggests that the CI Nad2 subunit likely accumulates to very low levels in *misf2* plants. Indeed, BN-PAGE analysis of Arabidopsis respiratory complexes indicated that CI is below

detectable levels in *misf*2 mutant plants (Figure 7). Immunoblots made with antibodies against the carbonic anhydrase CA2 [73] further indicated the accumulation of several complex I assembly intermediates of about 610, 230 and 85 kDa in both *misf*2 mutants. While CI was considerably reduced in both *misf*2 mutants, BN-PAGE analyses indicated that other respiratory complexes, including CIII, CV, and particularly CIV were rather upregulated in *misf*2 plants (Figure 7).
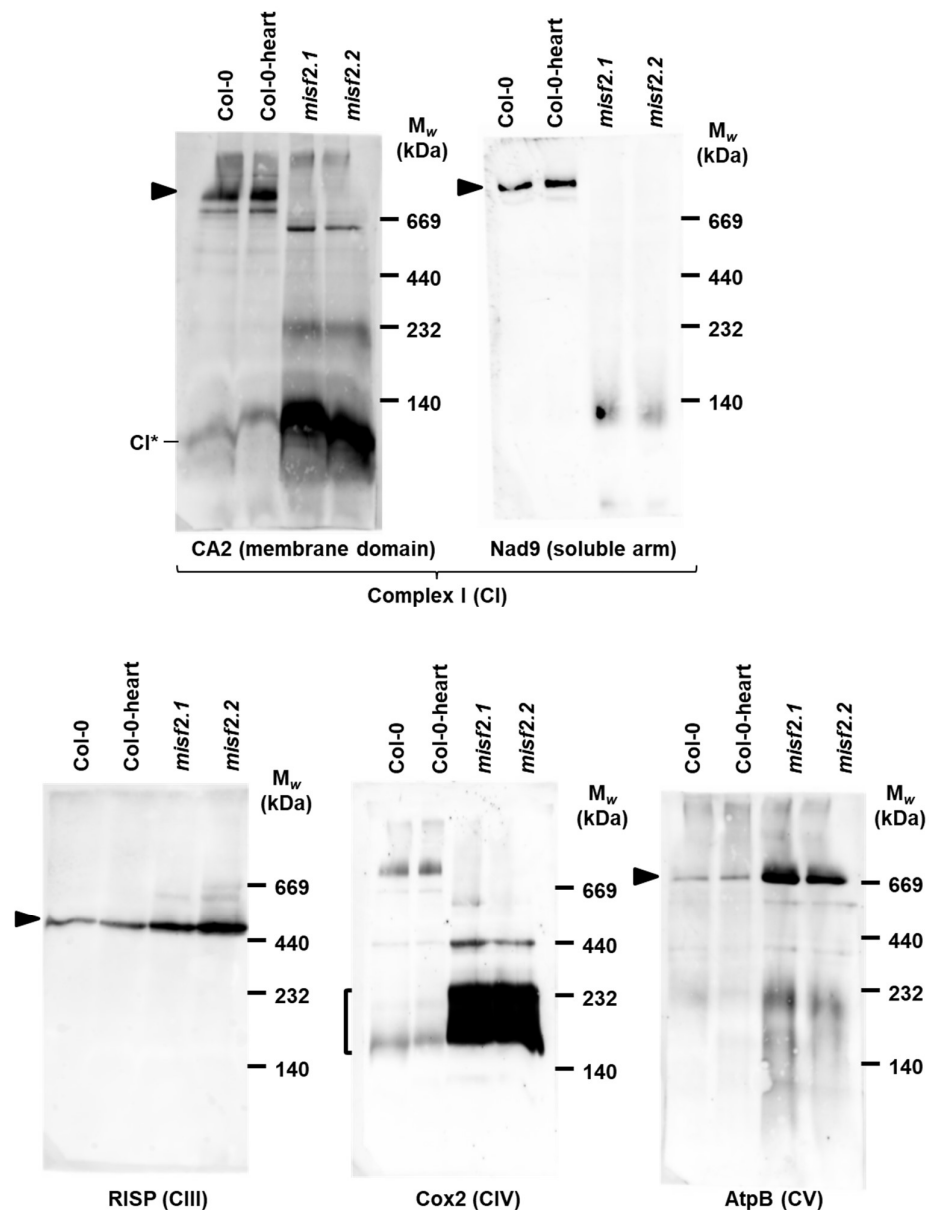

Figure 7. Holo-complex I is below detectable levels in *misf*2 mutants. Blue native (BN)-PAGE analysis of crude organellar fractions was performed as described by [74]. Aliquots, equivalent to 40 mg of crude organellar membrane extracts, obtained from wild type and *misf*2 plants, were solubilized with digitonin and resolved by BN-PAGE. For immunodetection, the proteins were transferred onto PVDF membranes and probed with the antibodies indicated below each blot (Table S2), as indicated. Arrows point toward native complexes I (~1,000 kDa), CIII dimer (III$_2$, ~500 kDa), CIV (about 200 and 220 kDa forms), and CV (~660 kDa) [63]. CI* indicates the ~85 kDa sub-CI assembly intermediate [67].

We further analyzed the relative accumulation of various mitochondrial proteins in Col-0, *misf*2 mutants and the functionally-complemented *misf2.2/35S::MISF2* line by im-munoblotting analysis using various antibodies raised against different plant mitochon-drial proteins. The data indicated that the CI-subunit CA2 and Nad9 accumulate in similar quantities in *misf*2 and wild type plants. The levels of various other mitochondrial pro-teins, including the Rieske iron-sulfur protein (RISP) of CIII, the Cox2 subunit of CIV, the AtpB subunit of CV, and the mitochondrial outer-membrane voltage-dependent anion channel (VDAC, or PORIN) proteins were upregulated in *misf*2 mutants, as compared with the wild type plants (Figure 8a).
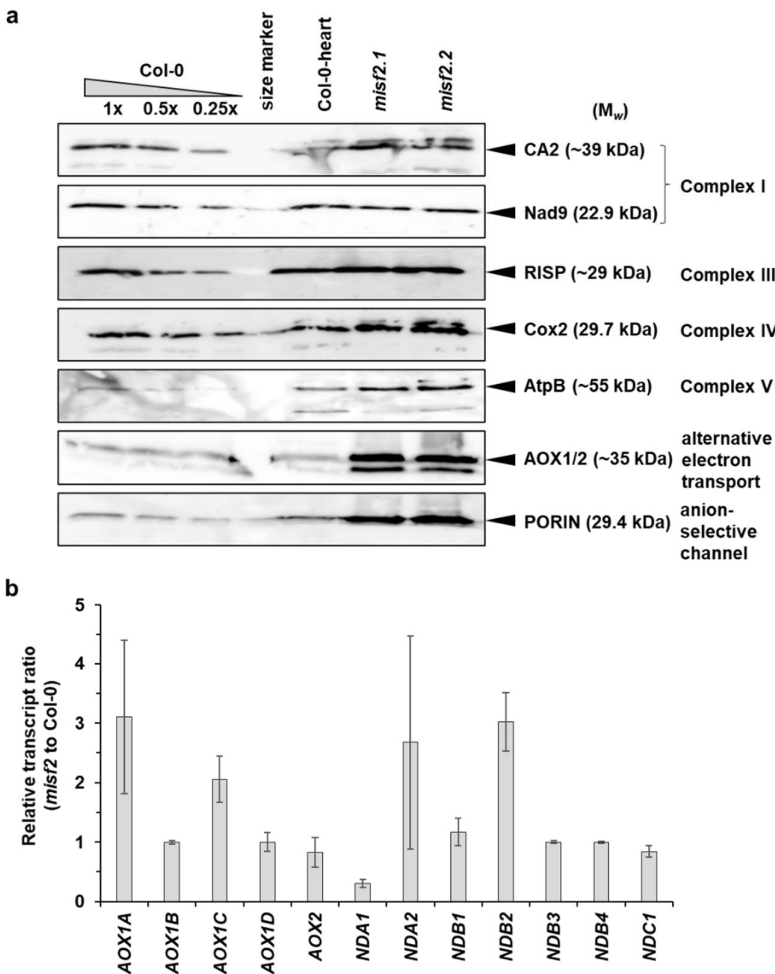

Figure 8. Relative accumulation of different mitochondrial proteins and *AOX* or *ND* transcripts in wild type and *misf*2 plants. (**a**) Immunoblots made with crude organellar fractions (equivalent to ~40 mg FW) extracted from 3-week-old MS-grown wild type plants, *in vitro* germinated wild type embyos at the heart to torpedo stages (Col-0 heart) and homozygous *misf*2 plantlets. The blots were probed with antibodies raised against the indicated mitochondrial proteins. (**b**) Analysis of the steady-state levels of various alternative oxidase (AOX) and rotenone-insensitive NAD(P)H dehy-drogenase (ND) mRNAs by RT-qPCR. The histogram displays relative mRNAs levels in *misf*2 plant-lets to 3-week-old MS-grown wild type plants.

In contrast, the accumulation of all tested mitochondrial proteins was equivalent be-tween the complemented line (*misf2.2/35S::MISF2*) and wild type plants (Figure S4).

Arabidopsis mutants affected in CI biogenesis undergo oxidative stress and often show subsequently a strong induction of the alternative respiratory pathways [53, 55, 72, 75, 76]. Accordingly, the relative accumulation of transcripts corresponding to various

alternative oxidase (AOX) and rotenone-insensitive NAD(P)H dehydrogenase (NDs) mRNAs in *misf2* was generally higher than in wild type plants (Figure 8b). Similarly, immunoblot assays indicated that the steady-state levels of AOX1 or AOX2 proteins was higher in *misf2* compared to the wild type (Figure 8a and S4).

3. Discussion

3.1. The MISF2 gene encodes a mitochondrion localized PPR protein, which is playing essential roles during early embryo-development in Arabidopsis plants

Mitochondria play key roles in energy metabolism and are thus vital organelle for plant life. During the evolution, the mitochondrial genomes of land plants have undergone increased plasticity, showing substantial variations in genome size and structures, and gene expression patterns between species (reviewed by *e.g.*, [1]). Angiosperm mtDNAs are the largest and least gene-dense genomes among eukaryotes [2]. mRNA production and expression in land plant mitochondria involve extensive processing steps, which include endonucleolytic cleavages, 5' and 3' mRNA trimming, extensive sequence editing and, relevantly to our study, the removal of intron (mostly group II-type) sequences that interrupt the coding regions of many mitochondrial genes (reviewed by e.g., [6]). These essential activities may serve as key control points of plant mitochondrial gene expression and are facilitated by numerous RNA binding cofactors [6, 8].

In this study, we assigned a role to an Arabidopsis PPR protein, namely the MISF2 encoded by the *AT3G22670* gene-locus in mt-RNA metabolism (*i.e.*, *nad2* maturation) and respiratoty complex I assembly. MISF2, like its maize orthologue, EMP10 [41], is a P-type PPR protein comprising 10 PPR motifs that is localized in mitochondria. As for *EMP10* in maize, downregulation of *MISF2* expression results in premature arrest of Arabidopsis embryo development at the late torpedo stage (Figure 1B). The loci of *MISF* and *RDM1* genes in Arabidopsis are partially overlapping. It was therefore impotrant to confirm that the developmental defect phenotypes and alterd mt-RNA metabolism we see in *misf2* mutants result directly from the downregulation of *MISF2* expression. To this aim, we analyzed the growth phenotypes and organellar RNA and protein profiles associated with a functionaly complemented *misf2* line, and further showed that MISF2 is directly required for *nad2* RNA maturation and resiratory complex I biogenesis and respiratory functions.

3.2. MISF2 is required for the splicing of nad2 intron 1

Most mitochondrial introns in angiosperms are classified as group II-type [12]. Model introns belonging to this class are large catalytic RNAs that are characterized by a conserved secondary structure consisting of six double-helical domains (named domains I to VI), radiating from a central hub [77, 78]. The excision of group II introns *in vivo* in bacteria and in the organelles of eukaryotic cells requires the action of various RNA binding protein cofactors. In canonical group II introns, these at least include maturase proteins that are most often encoded by the introns themselves [79]. In plant mitochondria, many additional proteinaceous splicing factors are required, which either derive from an ancient group of maturases [14], or from various other RNA binding cofactors that were recruited along evolution to facilitate mitochondrial intron splicing   [13, 80].

The PPR protein family is the largest RNA binding protein family known in plants, with about 400 to 600 members targeted to mitochondria and plastids [81]. PPR proteins bind their RNA substrates in a sequence specific manner, and were shown to play pivotal roles in various aspects of posttranscriptional RNA processing, including the excision of group II introns in land plant organelles [6, 9, 27, 40]. Here, we analyzed the molecular functions of the Arabidopsis MISF2 protein by characterizing loss-of-function mutants. As no homozygous mutant individuals could be identified among mature seeds of self-fertilized heterozygous *misf2* progenies, we used embryo rescue approaches [52, 56] to generate homozygous mutant plant material, which allowed us to analyze the role of MISF2 in mitochondrial RNA metabolism.

Analysis of mitochondrial RNA profiles in wild type and *misf2* plants showed a large reduction in the accumulation of mature *nad2* mRNA in both mutant lines (Figure 3). The RT-qPCR analyses further revealed a strong reduction in the splicing efficiency of *nad2* intron 1 in *misf2* plants (Figure 4). The most probable RNA-binding site for MISF2 protein (*i.e.*, GGUGAGGCGU) resides within the domain I of *nad2* intron 1 (Figure 5), which also corresponds to the genetic and biochemical RNA target of MISF2 (Figures 3, 4 and 6). In model group II introns, maturases were shown to bind with great affinity and specificity to their cognate intron-RNAs, in particular to regions of domain I and around the maturase coding sequences within domain IV [82]. It will therefore be interesting to investigate whether sequence changes within plant *nad2* intron 1 was accompanied by the recruitment of the PPR MISF2 factor to facilitate its splicing, possibly to stabilize (or nucleate) *nad2* intron 1 folding into a catalytically active structure.

Our data provide strong evidence that MISF2 is specifically required for *nad2* intron 1 splicing and that this RNA processing step is essential for early embryogenesis in Arabidopsis.

3.3. Embryo development and complex I biogenesis and activity

The electron transport chain is made of four major multi-subunit protein complexes, denoted as CI to CIV. Plants also possess several enzymes corresponding to non-phosphorylating bypasses of the electron transport chain (ETC), namely the alternative oxidase (AOX) and several rotenone-insensitive NAD(P)H dehydrogenases (NDs) [64, 83-86]. The biogenesis of respiratory CI in angiosperms involves the incorporation of ~50 different subunits encoded by both mitochondrial (*i.e.*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4l*, *nad5*, *nad6*, *nad7* and *nad9*) and nuclear gene loci [87]. These are incorporated into two main different CI domains, consisting in a membrane domain and the matrix (or peripheral) arm [67, 76, 88, 89].

Nad2 is a pivotal subunit of CI, that is suggested to be incorporated very early during the assembly of the membrane arm [67, 88, 90, 91]. The early steps of CI biogenesis involve the production of an ~85 kDa assembly intermediate of the membrane arm, which contains various gamma-type carbonic anhydrase subunits. Subsequently, Nad2 and few other subunits are incorporated to form a ~200 kDa membrane-bound CI assembly intermediate [67]. It is therefore anticipated that a strong reduction in Nad2 would interfere with the assembly of the CI membrane arm, and hence with the biogenesis of the ~1.0 MDa holo-CI. Consequently, BN-PAGE analysis of wild type and mutant plants revealed a major reduction in CI abundance in both *misf2* mutant lines (Figure 7). Immunoblot analysis with anti-CA2 antibodies further revealed the existence of various CI intermediates in *misf2* mutants, among which a major particle of about 85 kDa, which was also observed in the *abo5* mutant that is impaired in *nad2* expression [69]. The CI particles of higher mass (*i.e.*, 230 kDa and 610 kDa) detected in the mutants may correspond to Nad2-deprived assembly intermediates that are less stable than the ~85 kDa particles [67].

It has been demonstrated that the severity of CI production defects correlates with the gravity of the phenotypes displayed by corresponding plant mutants [53, 72, 76]. Severe CI mutants are impaired in the storage of essential nutrients, but not in the mobilization of the stored reserves [53]. Accordingly, mutants affected in β-oxidation, a metabolic process by which fatty acids are broken down by various tissues to produce energy, contain embryos that are typically arrested at earlier developmental stages compared with CI mutants [92]. Embryo maturation is often incomplete in various CI mutants, leading to the production of seeds with reduced reserves and germination capacity. One can anticipate that altered respiration interferes with numerous essential metabolic activities, resulting in altered embryo development.

In our study, we noticed that a severe defect in the production of the Nad2 subunit of CI results in impaired embryogenesis and a loss of germination capacity of mutant Arabidopsis seeds. However, most characterized plant CI mutants are generally able to germinate under standard culture conditions (see *e.g.*, [17, 71, 72, 93-95]. The inability of *misf2*

mutants to germinate under normal conditions is expected to result from an early arrest of the development of mutant embryos, placing *misf*2 mutants among the most severe CI mutants reported so far. We currently do not know the role that the embryo-rescue medium plays in improving the seed germination of *misf*2 mutants. It may be due to the presence of certain important chemicals in the rescue medium, or simply to a weakening of the seed coat by the high sugar concentration of the medium. Once germination has been induced, we could observe that *misf*2 mutants often show growth phenotypes like other Arabidopsis CI mutants (Figure 1b-*iv*). It was previously suggested that once photosynthesis is established, growth is to a lesser extent dependent on the application of external vitamins and/or sugars [53]. Subsequently, rescued *misf*2 mutants can slowly proceed with their vegetative growth phase, but remain unable to complete their life cycle, flower and produce viable seeds.

4. Materials and Methods

4.1. Plant material and growth conditions

Arabidopsis thaliana of the Columbia (Col-0) accession was used in all experiments. The wild type (Col-0 line) and the SALK-067654 (*misf2.1*) and SALK-066141 (*misf2.2*) mutants were obtained from the Arabidopsis Biological Resource Center (ABRC). Prior to germination, seeds obtained from wild type and mutant lines were surface-sterilized with Cl_2 gas, generated by the addition of 1 ml HCl per 50 mL of bleach (sodium hypochlorite 4.7%), for 4 hours at room-temperature (RT). The seeds were then sown on MS-agar plates containing 1% (w/v) sucrose or rescued by a method described in detail below. For synchronized germinations, the seeds were kept in the dark for 5 days at 4°C and then grown under long day condition (LD, 16:8-hour) in a growth chamber (Percival Scientific, Perry, IA, USA) at 22°C and under light intensity of 300 μE m^{-2} s^{-1}. PCR was used for genotyping the plants using specific oligonucleotides listed in Table S1. Sequencing of specific PCR products was used to check the T-DNA insertion site in both mutant lines.

4.2. GFP localization assay

The DNA region encoding the first 203 amino acids of MISF2 was PCR amplified with specific oligonucleotides (*i.e.*, *misf*2-B1 and *misf*2-B2; Table S1C). The 609 nts PCR DNA fragment was cloned into pDONR207 vector using the Gateway BP clonase enzyme mix and verified by Sanger sequencing. The entry clone was then transferred into pGWB5 vector by Gateway LR reaction to create a GFP translational fusion between the MISF2 N-terminus sequence and the GFP coding sequence. The vector was transformed into *Agrobacterium tumefaciens* (strain C58C51) and used to transform Arabidopsis plant cells. Transgenic cells were selected on hygromycin, and GFP fluorescence was visualized by confocal microscopy Leica TCS SP8. To visualize mitochondria *in vivo*, plant cells were incubated with 1 μM MitoTracker® Red (Thermofisher) for 10 min at room temperature prior to observation under confocal microscopy.

4.3. Embryo-rescue and establishment of homozygous misf2 mutants

Siliques from wild type and heterozygous *misf*2 plants were surfaced sterilized with 6% bleach solution for 10 min at RT. The seeds were then soaked in a 70% ethanol solution for 10 min at RT, washed briefly with sterile DDW, and opened in a sterile hood. Green and white seeds obtained from siliques of heterozygous *misf*2 plants 10 days after self-fertilization were sown on MS-agar plates supplemented with 1% (w/v) sucrose and 10 mg myoinositol, 100 µg thiamine, 100 µg pyridoxine and 100 µg nicotinic acid. For DNA and RNA analysis we used Arabidopsis wild type and *misf*2 plantlets at stage R6 (*i.e.*, 6 to 8 leaves) [57]. To obtain larger quantities of plant material, plantlets at stage R6 were grown on MS-agar plates and then transferred to MS-based liquid medium supplemented with 1% (w/v) sucrose and 10 mg myoinositol 100 µg Thiamine, 100 µg Pyridoxine, 100

µg nicotinic acid and incubated at 22℃ and under a light intensity of 300 µE m^{-2} s^{-1} with moderate (50~100 RPM) shaking.

4.4. Microscopic analyses of Arabidopsis wild type and mutant plants

Analysis of whole plant morphology, roots, leaves, siliques and seeds of wild type and mutant lines were examined under Stereoscopic (dissecting) microscope or light microscope at the Bio-Imaging unit of the Institute of Life Sciences (The Hebrew University of Jerusalem). Seeds were incubated with Hoyer solution for 30 minutes and the cleared samples were analyzed by differential interference contrast (Nomarski) microscopy.

4.5. RNA extraction and analysis

RNA extraction and analysis was performed essentially as previously described [17, 19, 96-98]. Total RNA was prepared from 200 mg seedlings grown on MS-agar plates supplemented with 1% sucrose, using the RNAzol RT reagent (Sigma-Aldrich). The RNA was then treated with RNase-free DNase I prior to its use in the assays. RT-qPCR was performed with specific oligonucleotides designed to exon-exon (mRNAs) regions corresponding to mitochondrial genes and intron-exon regions (pre-mRNAs) within each of the 23 group II introns in *Arabidopsis thaliana* (Tables S1). cDNA was synthesized by reverse transcription with the Superscript III reverse transcriptase (Invitrogen), using 1 - 2 µg of total RNA and 250 ng of a mixture of random hexanucleotides (Promega) and incubated for 50 min at 50ºC. Reactions were stopped by 15 min incubation at 70°C and the RT samples served directly for real-time PCR on a LightCycler 480 (Roche) using 2.5 µL of LightCycler 480 SYBR Green I Master mix and 2.5 µM of primers in a final volume of 5 µL. Reactions were performed in triplicate in the following conditions: pre-heating at 95°C for 10 min, followed by 40 cycles of 10 sec at 95°C, 10 sec at 58°C and 10 sec at 72°C. The nucleus-encoded 18S rRNA (At3g41768) and the mitochondrial 26S ribosomal rRNA subunit (ArthMr001) were used as reference genes.

4.6. Crude mitochondria preparations

Crude mitochondria extracts prepared essentially as described previously [74]. To this end, 200 mg of plantlets grown in liquid culture were harvested and homogenized in 2 ml of 75 mM MOPS-KOH, pH 7.6, 0.6 M sucrose, 4 mM EDTA, 0.2% polyvinylpyrrolidone-40, 8 mM L-cysteine, 0.2% bovine serum albumin and protease inhibitor cocktail 'complete Mini' from Roche Diagnostics GmbH (Mannheim, Germany). The lysate was filtrated through one layer of Miracloth and centrifuged at 1,300g for 4 min at 4ºC (to remove cell debris). The supernatant was then centrifuged at 22,000g for 10 min at 4ºC. The resulting pellet containing thylakoid and mitochondrial membranes was washed twice with 1ml of wash buffer 37.5 mM MOPS-KOH, 0.3 M sucrose and 2mM EDTA, pH 7.6 prior to use.

4.7. Blue native PAGE analysis of respiratory complexes

Blue native (BN)-PAGE of crude organellar membranous fractions was performed according to the method described in Pineau, Layoune, Danon and De Paepe [74]. An aliquot equivalent to 40 mg of crude Arabidopsis mitochondria extracts was solubilized with 5% (w/v) digitonin in BN-solubilization buffer (30 mM HEPES, pH 7.4, 150 mM potassium acetate, 10% [v/v] glycerol) and then incubated on ice for 30 min. The samples were centrifuged 8 min at 20,000 xg to pellet non-solubilized material and 0.2% [v/v] of Serva Blue G was added to the supernatant. The samples were then loaded onto a native 4 to 16% linear gradient gel.   For 'non-denaturing-PAGE' immunoblotting, the gel was transferred to a PVDF membrane (Bio-Rad) in Cathode buffer (50 mM Tricine and 15 mM Bis-Tris-HCl, pH 7.0) for 16 h at 4°C at constant current of 40 mA. The blots where then incubated with antibodies against mitochondrial proteins (Table S2) and hybridization

signals were identified by chemiluminescence assay after incubation with an appropriate horseradish peroxidase (HRP)-conjugated secondary antibody.

4.8. RNA co-immunoprecipitation assays

Immunoprecipitation of MISF2-3HA were performed using the µMACS HA-Tagged Protein Isolation Kit (Miltenyi Biotec) following a procedure previously described in Wang et al, 2020 [80].

5. Conclusions

Angiosperms encode numerous PPR proteins that are predominantly localized in plastids and mitochondria, and which carry essential roles in organellar RNA metabolism. These include the EMP10 protein, which regulates the maturation of *nad2* in maize mitochondria [41]. Analysis of protein and RNA profiles of mutants affected in the Arabidopsis orthologous gene, designated *MITOCHONDRIAL SPLICING FACTOR* 2 (*MISF2*, encoded by At3g22670 gene), indicates that MISF2 also functions specifically in the excision of the first intron of *nad2*. Plant mutants affected in *MISF2* accumulate high levels of *nad2* pre-RNA *in vivo* and show a notable reduction in *nad2* intron 1 splicing. The splicing defect found in *misf2* (or *emp10*) is associated with altered CI biogenesis and arrested embryonic development. Together, these data show that the molecular functions are conserved between the Arabidopsis MISF2 protein and its related EMP10 homolog in maize [41], which suggests that the common PPR ancestor of MISF2 and EMP10 has been recruited to act in *nad2* intron 1 splicing before the divergence of monocot and dicot plant species [42]. Our results provide important insights into the roles of nuclear-encoded PPR factors in mitochondria gene expression and the biogenesis of the respiratory system during early plant life.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: The topology and structure of MISF2 protein.; Figure S2: MISF2 gene expression pattern during Arabidopsis development.; Figure S3: Nucleotide sequence of the *MISF2* gene and precise location of T-DNA insertion site.; Figure S4: Steady-state level analysis of various mitochondrial proteins in wild type, *misf2.2* and functionally complemented *misf2.2* plants.; Table S1: List of oligonucleotides used in this study.; Table S2: List of antibodies used in this study.

Author Contributions: methodology, T-T.N. C.B., S.S., M.Z., M.Q., R.M., H.Z., H.M. and O.O.B.; formal analysis, T-T.N. C.B., S.S., M.Z., M.Q., R.M., H.Z., H.M. and O.O.B.; data curation, T-T.N., H.M., R.M. and O.O.B.; writing—original draft preparation, H.M. and O.O.B.; writing—review and editing, H.M., T-T.N., R.M. and O.O.B.; supervision, H.M. and O.O.B.; funding acquisition, H.M. and O.O.B. All authors have read and agreed to the published version of the manuscript.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Best, C.; Mizrahi, R.; Ostersetzer-Biran, O., Why so complex? the intricacy of genome structure and gene expression, associated with angiosperm mitochondria, may relate to the regulation of embryo quiescence or dormancy—intrinsic blocks to early plant life. *Plants* **2020,** 9, (5), 598.

2.   Gualberto, J. M.; Newton, K. J., Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu Rev Plant Biol* **2017,** 68, 225-252.

3.   Woodson, J. D.; Chory, J., Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* **2008,** 9, (5), 383-395.

4.   Kleine, T.; Leister, D., Retrograde signaling: Organelles go networking. *BBA* **2016,** 1857, (8), 1313-25.

5.   Fuchs, P.; Rugen, N.; Carrie, C.; Elsasser, M.; Finkemeier, I.; Giese, J.; Hildebrandt, T. M.; Kuhn, K.; Maurino, V. G.; Ruberti, C.; Schallenberg-Rudinger, M.; Steinbeck, J.; Braun, H. P.; Eubel, H.; Meyer, E. H.; Muller-Schussele, S. J.; Schwarzlander, M., Single organelle function and organization as estimated from Arabidopsis mitochondrial proteomics. *Plant J* **2020,** 101, (2), 420-441.

6.   Zmudjak, M.; Ostersetzer-Biran, O., *RNA metabolism and transcript regulation*. Chichester John Wiley & Sons, Ltd: 2017; Vol. 50, p 143-184.

7.   Colas des Francs-Small, C.; Small, I., Surrogate mutants for studying mitochondrially encoded functions. *Biochimie* **2014,** 100, (0), 234-242.

8.   Hammani, K.; Giege, P., RNA metabolism in plant mitochondria. *Trends Plant Sci* **2014,** 19, (6), 380-9.

9.   Small, I. D.; Schallenberg-Rudinger, M.; Takenaka, M.; Mireau, H.; Ostersetzer-Biran, O., Plant organellar RNA editing: what 30 years of research has revealed. *Plant J* **2020,** 101, (5), 1040-1056.

10. Michel, F.; Lang, B. F., Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature* **1985,** 316, (6029), 641-3.

11. Sharp, P. A., On the origin of RNA splicing and introns. *Cell* **1985,** 42, (2), 397-400.

12. Bonen, L., Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion* **2008,** 8, (1), 26-34.

13. Brown, G. G.; Colas des Francs-Small, C.; Ostersetzer-Biran, O., Group II intron splicing factors in plant mitochondria. *Front Plant Sci* **2014,** 5, (35), 1-13.

14. Schmitz-Linneweber, C.; Lampe, M.-K.; Sultan, L. D.; Ostersetzer-Biran, O., Organellar maturases: A window into the evolution of the spliceosome. *BBA - Bioenergetics* **2015,** 1847, ((9)), 798-808.

15. Mohr, G.; Lambowitz, A. M., Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res* **2003,** 31, (2), 647-652.

16. Köhler, D.; Schmidt-Gattung, S.; Binder, S., The DEAD-box protein PMH2 is required for efficient group II intron splicing in mitochondria of *Arabidopsis thaliana*. *Plant Mol Biol* **2010,** 72, (4), 459-467.

17. Zmudjak, M.; Shevtsov, S.; Sultan, L. D.; Keren, I.; Ostersetzer-Biran, O., Analysis of the roles of the Arabidopsis nMAT2 and PMH2 proteins provided with new insights into the regulation of group II intron splicing in land-plant mitochondria. *Int J Mol Sci* **2017,** 18, (11), 1-25.

18. He, J.; Duan, Y.; Hua, D.; Fan, G.; Wang, L.; Liu, Y.; Chen, Z.; Han, L.; Qu, L.-J.; Gong, Z., DEXH box RNA helicase–mediated mitochondrial reactive oxygen species production in Arabidopsis mediates crosstalk between abscisic acid and auxin signaling. *Plant Cell* **2012,** 24, (5), 1815-1833.

19. Colas des Francs-Small, C.; Kroeger, T.; Zmudjak, M.; Ostersetzer-Biran, O.; Rahimi, N.; Small, I.; Barkan, A., A PORR domain protein required for rpl2 and ccmFc intron splicing and for the biogenesis of c-type cytochromes in Arabidopsis mitochondria. *Plant J* **2012,** 69, (6), 996-1005.

20. Peeters, N.; Small, I., Dual targeting to mitochondria and chloroplasts. *BBA* **2001,** 1541, 54-63.

21. Chen, L.; Li, Y.-x.; Li, C.; Shi, Y.; Song, Y.; Zhang, D.; Li, Y.; Wang, T., Genome-wide analysis of the pentatricopeptide repeat gene family in different maize genomes and its important role in kernel development. *BMC Plant Biol* **2018,** 18, (1), 366.

22. Lurin, C.; Andres, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyere, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; Lecharny, A.; Le Ret, M.; Martin-Magniette, M.-L.; Mireau, H.; Peeters, N.; Renou, J.-P.; Szurek, B.; Taconnat, L.; Small, I., Ge-nome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **2004,** 16, (8), 2089-2103.

23. Schmitz-Linneweber, C.; Small, I., Pentatricopeptide repeat proteins: a socket set for organelle gene expression. *Trends Plant Sci* **2008,** 13, (12), 663-670.

24. Coquille, S.; Filipovska, A.; Chia, T.; Rajappa, L.; Lingford, J. P.; Razif, M. F. M.; Thore, S.; Rackham, O., An artificial PPR scaffold for programmable RNA recognition. *Nat Comm* **2014,** 5, 5729.

25. Gully, B. S.; Cowieson, N.; Stanley, W. A.; Shearston, K.; Small, I. D.; Barkan, A.; Bond, C. S., The solution structure of the pentatricopeptide repeat protein PPR10 upon binding atpH RNA. *Nucleic Acids Res* **2015,** 43, (3), 1918-26.

26. Binder, S.; Hölzle, A.; Jonietz, C., RNA processing and RNA stability in plant mitochondria. In *Plant mitochondria*, Springer: 2011; pp 107-130.

27. Barkan, A.; Small, I., Pentatricopeptide Repeat Proteins in Plants. *Annu Rev Plant Biol* **2014,** 65, (1), 415-442.

28. Shikanai, T.; Fujii, S., Function of PPR proteins in plastid gene expression. *RNA Biol* **2013,** 10, (9), 1446-1456.

29. Geddy, R.; Brown, G. G., Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant ge-nomes and may be subject to diversifying selection. *BMC Genomics* **2007,** 8, 130-130.

30. Dahan, J.; Mireau, H., The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. *RNA biol* **2013,** 10, (9), 1469-76.

31. Zhao, N.; Wang, Y.; Hua, J., Genomewide identification of PPR gene family and prediction analysis on restorer gene in Gossy-pium. *J Genet* **2018,** 97, (5), 1083-1095.

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

33. Shen, C.; Zhang, D.; Yan, J.; Zhang, Q.; Hong, S.; Yang, Y.; Yao, Y.; Yin, P.; Zou, T., Delineation of pentatricopeptide repeat codes for target RNA prediction. *Nucleic Acids Res* **2019,** 47, (7), 3728-3738.

34. Yin, P.; Li, Q.; Yan, C.; Liu, Y.; Liu, J.; Yu, F.; Wang, Z.; Long, J.; He, J.; Wang, H.-W.; Wang, J.; Zhu, J.-K.; Shi, Y.; Yan, N., Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* **2013,** 504, 168.

35. Ke, J.; Chen, R. Z.; Ban, T.; Zhou, X. E.; Gu, X.; Tan, M. H.; Chen, C.; Kang, Y.; Brunzelle, J. S.; Zhu, J. K.; Melcher, K.; Xu, H. E., Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat Struc Mol Biol* **2013,** 20, (12), 1377-82.

36. Gully, B. S.; Shah, K. R.; Lee, M.; Shearston, K.; Smith, N. M.; Sadowska, A.; Blythe, A. J.; Bernath-Levin, K.; Stanley, W. A.; Small, I. D.; Bond, C. S., The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystal* **2015,** 71, (Pt 2), 196-208.

37. Shen, C.; Zhang, D.; Guan, Z.; Liu, Y.; Yang, Z.; Yang, Y.; Wang, X.; Wang, Q.; Zhang, Q.; Fan, S.; Zou, T.; Yin, P., Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeat proteins. *Nat Comm* **2016,** 7, 11285.

38. Brehme, N.; Zehrmann, A.; Verbitskiy, D.; Hartel, B.; Takenaka, M., Mitochondrial RNA editing PPR proteins can tolerate protein tags at E as well as at DYW domain termini. *Front Plant Sci* **2014,** 5 (127), 1-4.

39. Takenaka, M.; Jörg, A.; Burger, M.; Haag, S., RNA editing mutants as surrogates for mitochondrial SNP mutants. *Plant Physiol Biochem* **2019,** 135, 310-321.

40. Rovira, A. G.; Smith, A. G., PPR proteins - orchestrators of organelle RNA metabolism. *Physiol Plant* **2019,** 166, (1), 451-459.

41. Cai, M.; Li, S.; Sun, F.; Sun, Q.; Zhao, H.; Ren, X.; Zhao, Y.; Tan, B. C.; Zhang, Z.; Qiu, F., Emp10 encodes a mitochondrial PPR protein that affects the cis-splicing of nad2 intron 1 and seed development in maize. *Plant J* **2017,** 91, (1), 132-144.

42. Wolfe, K. H.; Gouy, M.; Yang, Y. W.; Sharp, P. M.; Li, W. H., Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *P Natl Acad Sci USA* **1989,** 86, (16), 6201-6205.

43. Gutmann, B.; Royan, S.; Schallenberg-Rüdinger, M.; Lenz, H.; Castleden, I. R.; McDowell, R.; Vacher, M. A.; Tonti-Filippini, J.; Bond, C. S.; Knoop, V.; Small, I. D., The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Mol Plant* **2020,** 13, (2), 215-230.

44. Letunic, I.; Doerks, T.; Bork, P., SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res* **2012,** 40, D302-D305.

45. Marchler-Bauer, A.; Anderson, J. B.; DeWeese-Scott, C.; Fedorova, N. D.; Geer, L. Y.; He, S.; Hurwitz, D. I.; Jackson, J. D.; Jacobs, A. R.; Lanczycki, C. J.; Liebert, C. A.; Liu, C.; Madej, T.; Marchler, G. H.; Mazumder, R.; Nikolskaya, A. N.; Panchenko, A. R.; Rao, B. S.; Shoemaker, B. A.; Simonyan, V.; Song, J. S.; Thiessen, P. A.; Vasudevan, S.; Wang, Y.; Yamashita, R. A.; Yin, J. J.; Bryant, S. H., CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res* **2003,** 31, (1), 383-387.

46. Wu, C. H.; Apweiler, R.; Bairoch, A.; Natale, D. A.; Barker, W. C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; Martin, M. J.; Mazumder, R.; O'Donovan, C.; Redaschi, N.; Suzek, B., The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res* **2006,** 34, (Database issue), D187-91.

47. Hooper, C. M.; Castleden, I. R.; Tanz, S. K.; Aryamanesh, N.; Millar, A. H., SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res* **2017,** 45, (D1), D1064-D1074.

48. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D., Highly accurate protein structure prediction with AlphaFold. *Nature* **2021,** 596, (7873), 583-589.

49. Hruz, T.; Laule, O.; Szabo, G.; Wessendorp, F.; Bleuler, S.; Oertle, L.; Widmayer, P.; Gruissem, W.; Zimmermann, P., Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinform* **2008,** 2008, 420747-420751.

50. Meinke, D. W., Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *New Phytol* **2020,** 226, (2), 306-325.

51. Franzmann, L.; Patton, D. A.; Meinke, D. W., In vitro morphogenesis of arrested embryos from lethal mutants of Arabidopsis thaliana. *Theor Appl Genet* **1989,** 77, (5), 609-616.

52. Dahan, J.; Tcherkez, G.; Macherel, D.; Benamar, A.; Belcram, K.; Quadrado, M.; Arnal, N.; Mireau, H., Disruption of the CYTOCHROME C OXIDASE DEFICIENT1 gene leads to cytochrome c oxidase depletion and reorchestrated respiratory metabolism in Arabidopsis. *Plant Physiol* **2014,** 166, (4), 1788-802.

53. Kuhn, K.; Obata, T.; Feher, K.; Bock, R.; Fernie, A. R.; Meyer, E. H., Complete mitochondrial complex I deficiency induces an up-regulation of respiratory fluxes that is abolished by traces of functional complex I. *Plant Physiol* **2015,** 168, (4), 1537-49.

54. Cordoba, J. P.; Marchetti, F.; Soto, D.; Martin, M. V.; Pagnussat, G. C.; Zabaleta, E., The CA domain of the respiratory complex I is required for normal embryogenesis in Arabidopsis thaliana. *J Exp Bot* **2016,** 67, (5), 1589-603.

55. Fromm, S.; Going, J.; Lorenz, C.; Peterhansel, C.; Braun, H. P., Depletion of the "gamma-type carbonic anhydrase-like" subunits of complex I affects central mitochondrial metabolism in Arabidopsis thaliana. *BBA* **2016,** 1857, (1), 60-71.

56. Shevtsov-Tal, S.; Best, C.; Matan, R.; Chandran, S. A.; Brown, G. G.; Ostersetzer-Biran, O., nMAT3 is an essential maturase splicing factor required for holo-complex I biogenesis and embryo development in Arabidopsis thaliana plants. *Plant J* **2021,** 106, (4), 1128-1147.

57. Boyes, D. C.; Zayed, A. M.; Ascenzi, R.; McCaskill, A. J.; Hoffman, N. E.; Davis, K. R.; Görlach, J., Growth Stage–Based Pheno-typic Analysis of Arabidopsis. A Model for High Throughput Functional Genomics in Plants. *Plant Cell* **2001**, 13, (7), 1499-1510.

58. Sasaki, T.; Lorković, Z. J.; Liang, S.-C.; Matzke, A. J. M.; Matzke, M., The Ability to form homodimers is essential for RDM1 to function in RNA-directed dna methylation. *PLoS ONE* **2014**, 9, (2), e88190.

59. 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.

60. Yagi, Y.; Nakamura, T.; Small, I., The potential for manipulating RNA with pentatricopeptide repeat proteins. *Plant J* **2014**, 78, (5), 772-782.

61. Takenaka, M.; Zehrmann, A.; Brennicke, A.; Graichen, K., Improved Computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS ONE* **2013**, 8, (6), e65343.

62. Cheng, S.; Gutmann, B.; Zhong, X.; Ye, Y.; Fisher, M. F.; Bai, F.; Castleden, I.; Song, Y.; Song, B.; Huang, J.; Liu, X.; Xu, X.; Lim, B. L.; Bond, C. S.; Yiu, S. M.; Small, I., Redefining the structural motifs that determine RNA binding and RNA editing by pen-tatricopeptide repeat proteins in land plants. *Plant J* **2016**, 85, (4), 532-47.

63. Klodmann, J.; Senkler, M.; Rode, C.; Braun, H.-P., Defining the "protein complex proteome" of plant mitochondria. *Plant Physiol* **2011**, 157, (2), 587-98.

64. Millar, A. H.; Whelan, J.; Soole, K. L.; Day, D. A., Organization and regulation of mitochondrial respiration in plants. *Ann Rev Plant Biol* **2011**, 62, (1), 79-104.

65. Zhao, P.; Wang, F.; Li, N.; Shi, D.-Q.; Yang, W.-C., Pentatricopeptide repeat protein MID1 modulates *nad2* intron 1 splicing and Arabidopsis development. *Scientific Reports* **2020**, 10, (1), 2008.

66. Marchetti, F.; Cainzos, M.; Shevtsov, S.; Cordoba, J. P.; Sultan, L. D.; Brennicke, A.; Takenaka, M.; Pagnussat, G.; Ostersetzer-Biran, O.; Zabaleta, E., Mitochondrial Pentatricopeptide repeat protein, EMB2794, plays a pivotal role in nadh dehydrogenase subunit *nad2* mRNA maturation in *Arabidopsis thaliana*. *Plant Cell Physiol* **2020**, 61, (6), 1080-1094.

67. Ligas, J.; Pineau, E.; Bock, R.; Huynen, M. A.; Meyer, E. H., The assembly pathway of complex I in *Arabidopsis thaliana*. *Plant J* **2019**, 97, (3), 447-459.

68. Xiu, Z.; Sun, F.; Shen, Y.; Zhang, X.; Jiang, R.; Bonnard, G.; Zhang, J.; Tan, B. C., EMPTY PERICARP16 is required for mitochon-drial *nad2* intron 4 cis-splicing, complex I assembly and seed development in maize. *Plant J* **2016**.

69. Liu, Y.; He, J.; Chen, Z.; Ren, X.; Hong, X.; Gong, Z., ABA overly-sensitive 5 (ABO5), encoding a pentatricopeptide repeat protein required for cis-splicing of mitochondrial *nad2* intron 3, is involved in the abscisic acid response in Arabidopsis. *Plant J* **2010**, 63, (5), 749-65.

70. Hsu, Y. W.; Wang, H. J.; Hsieh, M. H.; Hsieh, H. L.; Jauh, G. Y., Arabidopsis mTERF15 is required for mitochondrial *nad2* intron 3 splicing and functional complex I activity. *PLoS ONE* **2014**, 9, (11), e112360.

71. Nakagawa, N.; Sakurai, N., A mutation in At-nMat1a, which encodes a nuclear gene having high similarity to group II Intron maturase, causes impaired splicing of mitochondrial *nad4* transcript and altered carbon metabolism in *Arabidopsis thaliana*. *Plant Cell Physiol* **2006**, 47, (6), 772-783.

72. Keren, I.; Tal, L.; Colas des Francs-Small, C.; Araújo, W. L.; Shevtsov, S.; Shaya, F.; Fernie, A. R.; Small, I.; Ostersetzer-Biran, O., nMAT1, a nuclear-encoded maturase involved in the *trans*-splicing of *nad1* intron 1, is essential for mitochondrial complex I assembly and function. *Plant J* **2012**, 71, (3), 413-426.

73. Perales, M.; Parisi, G.; Fornasari, M. S.; Colaneri, A.; Villarreal, F.; Gonzalez-Schain, N.; Echave, J.; Gomez-Casati, D.; Braun, H. P.; Araya, A.; Zabaleta, E., Gamma carbonic anhydrase like complex interact with plant mitochondrial complex I. *Plant Mol Biol* **2004**, 56, (6), 947-57.

74. Pineau, B.; Layoune, O.; Danon, A.; De Paepe, R., L-galactono-1,4-lactone dehydrogenase is required for the accumulation of plant respiratory complex I. *J Biol Chem* **2008**, 283, (47), 32500-32505.

75. Karpova, O. V.; Kuzmin, E. V.; Elthon, T. E.; Newton, K. J., DifferentialeExpression of alternative oxidase genes in maize mito-chondrial mutants. *Plant Cell* **2002**, 14, (12), 3271-3284.

76. Ostersetzer-Biran, O., Respiratory complex I and embryo development. *J Exp Bot* **2016**, 67, (5), 1205-7.

77. Michel, F.; Umesono, K.; Ozeki, H., Comparative and functional anatomy of group II catalytic introns — a review. *Gene* **1989**, 82, (1), 5-30.

78. Michel, F.; Ferat, J. L., Structure and activities of group-II introns. *Annual review of biochemistry* **1995**, 64, 435-461.

79. Lazowska, J.; Jacq, C.; Slonimski, P. P., Sequence of introns and flanking exons in wild-type and *box3* mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell* **1980**, 22, (2 Pt 2), 333-48.

80. Wang, C.; Fourdin, R.; Quadrado, M.; Dargel-Graffin, C.; Tolleter, D.; Macherel, D.; Mireau, H., Rerouting of ribosomal proteins into splicing in plant organelles. *P Natl Acad Sci USA* **2020**, 117, (47), 29979-29987.

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

82. Matsuura, M.; Noah, J. W.; Lambowitz, A. M., Mechanism of maturase-promoted group II intron splicing. *EMBO J* **2001**, 20, (24), 7259-7270.

83. Schertl, P.; Braun, H. P., Respiratory electron transfer pathways in plant mitochondria. *Front Plant Sci* **2014**, 5, 163.

84. Senkler, J.; Senkler, M.; Eubel, H.; Hildebrandt, T.; Lengwenus, C.; Schertl, P.; Schwarzländer, M.; Wagner, S.; Wittig, I.; Braun, H.-P., The mitochondrial complexome of *Arabidopsis thaliana*. *Plant J* **2017**, 89, (6), 1079-1092.

85. Subrahmanian, N.; Remacle, C.; Hamel, P. P., Plant mitochondrial complex I composition and assembly: A review. *BBA - Bio-energetics* **2016**, 1857, (7), 1001-14.

86. Jacoby, R. P.; Li, L.; Huang, S.; Pong Lee, C.; Millar, A. H.; Taylor, N. L., Mitochondrial composition, function and stress response in plants. *J Integr Plant Biol* **2012**, 54, (11), 887-906.

87. Lee, C. P.; Taylor, N. L.; Millar, A. H., Recent advances in the composition and heterogeneity of the Arabidopsis mitochondrial proteome. *Front Plant Sci* **2013,** 4.

88. Braun, H.-P.; Binder, S.; Brennicke, A.; Eubel, H.; Fernie, A. R.; Finkemeier, I.; Klodmann, J.; König, A.-C.; Kühn, K.; Meyer, E.; Obata, T.; Schwarzländer, M.; Takenaka, M.; Zehrmann, A., The life of plant mitochondrial complex I. *Mitochondrion* **2014**, 19, Part B, (0), 295-313.

89. Klodmann, J.; Sunderhaus, S.; Nimtz, M.; Jänsch, L.; Braun, H.-P., Internal architecture of mitochondrial complex I from *Arabidopsis thaliana*. *Plant cell* **2010,** 22, (3), 797-810.

90. Soufari, H.; Parrot, C.; Kuhn, L.; Waltz, F.; Hashem, Y., Specific features and assembly of the plant mitochondrial complex I revealed by cryo-EM. *Nat Comm* **2020**, 11, (1), 5195.

91. Maldonado, M.; Padavannil, A.; Zhou, L.; Guo, F.; Letts, J. A., Atomic structure of a mitochondrial complex I intermediate from vascular plants. *eLife* **2020,** 9, e56664.

92. Pinfield-Wells, H.; Rylott, E. L.; Gilday, A. D.; Graham, S.; Job, K.; Larson, T. R.; Graham, I. A., Sucrose rescues seedling establishment but not germination of Arabidopsis mutants disrupted in peroxisomal fatty acid catabolism. *Plant J* **2005,** 43, (6), 861-72.

93. Koprivova, A.; des Francs-Small, C. C.; Calder, G.; Mugford, S. T.; Tanz, S.; Lee, B. R.; Zechmann, B.; Small, I.; Kopriva, S., Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial *nad7* transcripts. *J Biol Chem* **2010,** 285, (42), 32192-9.

94. Colas des Francs-Small, C.; Falcon de Longevialle, A.; Li, Y.; Lowe, E.; Tanz, S. K.; Smith, C.; Bevan, M. W.; Small, I., The Pentatricopeptide Repeat Proteins TANG2 and ORGANELLE TRANSCRIPT PROCESSING439 are involved in the splicing of the multipartite *nad5* transcript encoding a subunit of mitochondrial complex I. *Plant Physiol* **2014,** 165, (4), 1409-1416.

95. Weissenberger, S.; Soll, J.; Carrie, C., The PPR protein SLOW GROWTH 4 is involved in editing of *nad4* and affects the splicing of *nad2* intron 1. *Plant Mol Biol* **2017,** 93, (4-5), 355-368.

96. Zmudjak, M.; Colas des Francs-Small, C.; Keren, I.; Shaya, F.; Belausov, E.; Small, I.; Ostersetzer-Biran, O., mCSF1, a nucleus-encoded CRM protein required for the processing of many mitochondrial introns, is involved in the biogenesis of respiratory complexes I and IV in Arabidopsis. *New Phytol* **2013,** 199, (2), 379-394.

97. Cohen, S.; Zmudjak, M.; Colas des Francs-Small, C.; Malik, S.; Shaya, F.; Keren, I.; Belausov, E.; Many, Y.; Brown, G. G.; Small, I.; Ostersetzer-Biran, O., nMAT4, a maturase factor required for *nad1* pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in Arabidopsis mitochondria. *Plant J* **2014,** 78, (2), 253-268.

98. Sultan, L. D.; Mileshina, D.; Grewe, F.; Rolle, K.; Abudraham, S.; Głodowicz, P.; Khan Niazi, A.; keren, I.; Shevtsov, S.; Klipcan, L.; Barciszewski, J.; Mower, J. P.; Dietrich, A.; Ostersetzer, O., The reverse-transcriptase/RNA-maturase protein MatR is required for the splicing of various group II introns in Brassicaceae mitochondria. *Plant cell* **2016,** 28, (11), 2805-2829.