Published online 8 March 2017

Nucleic Acids Research, 2017, Vol. 45, No. 10 ***6119–6134***
doi: 10.1093/nar/gkx162

The pentatricopeptide repeat protein MTSF2 stabilizes a *nad1* precursor transcript and defines the 3′ end of its 5′-half intron

Chuande Wang[1,2,†], Fabien Aubé[1,†], Noelya Planchard[1,2,†], Martine Quadrado[1], Céline Dargel-Graffin[1], Fabien Nogué[1] and Hakim Mireau[1,*]

[1]Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, RD10, 78026 Versailles Cedex, France and [2]Paris-Sud University, Université Paris-Saclay, 91405 Orsay Cedex, France

Received January 11, 2017; Revised February 22, 2017; Editorial Decision February 24, 2017; Accepted February 28, 2017

ABSTRACT

RNA expression in plant mitochondria implies a large number of post-transcriptional events in which transcript processing and stabilization are essential. In this study, we analyzed the function of the Arabidopsis mitochondrial stability factor 2 gene (*MTSF2*) and show that the encoded pentatricopeptide repeat protein is essential for the accumulation of stable *nad1* mRNA. The production of mature *nad1* requires the assembly of three independent RNA precursors via two trans-splicing reactions. Genetic analyses revealed that the lack of *nad1* in *mtsf2* mutants results from the specific destabilization of the *nad1* exons 2–3 precursor transcript. We further demonstrated that MTSF2 binds to its 3′ extremity with high affinity, suggesting a protective action by blocking exoribonuclease progression. By defining the 3′ end of *nad1* exons 2–3 precursor, MTSF2 concomitantly determines the 3′ extremity of the first half of the trans-intron found at the end of the transcript. Therefore, binding of the MTSF2 protein to *nad1* exons 2–3 precursor evolved both to stabilize the transcript and to define a 3′ extremity compatible with the trans-splicing reaction needed to reconstitute mature *nad1*. We thus reveal that the range of transcripts stabilized by association with protective protein on their 3′ end concerns also mitochondrial precursor transcripts.

INTRODUCTION

Mitochondria are essential organelles that are the site of cellular respiration, and a variety of other important biochemical processes such as heme biosynthesis, part of the urea cycle and Fe–S cluster formation. They descent from ancient bacteria and contain their own genome that is separate from the nuclear chromatin (1,2). Modern mitochondria encode around 50 genes in most eukaryotes. Despite this conserved and limited coding capacity, the structure of mitochondrial genomes is highly variable among organisms, going from very small and compact genomes in animals to huge and likely multi-circular structures in higher plants (3). The size increase of plant mitochondrial genomes results mostly from the presence of extended intergenic regions that are rich in repeated sequences (4,5). The high recombinogenic activity of plant mitochondrial DNA results in genomic organizations involving complex arrays of inter-convertible linear and circular DNA molecules. Subsequently, mitochondrial gene order is poorly conserved among plant species and examples of gene fragmentations are frequent (5–10). In response, complex RNA expression processes involving a considerable number of post-transcriptional events have evolved to maintain proper expression of mitochondrial genes and generate mature mRNAs. These steps include over 400 C-to-U RNA editing events, cis and trans-intron splicing, 5′ and 3′ RNA processing as well as mRNA stabilization. The underlying mechanisms are still poorly understood at the molecular level and the vast majority of the involved protein factors have not been identified. However, in the last years, the pentatricopeptide repeat (PPR) protein family has shown to figure prominently in these processes (11,12). Nuclear genomes of higher plant encode over 450 PPR genes, and the vast majority encode proteins that are predicted to target mitochondria or chloroplasts (11,13). PPR proteins are almost exclusively composed of tandem repeats of a highly degenerate 35-amino-acid motifs (14). Length variants of PPR repeats have been detected however, with repeats that are longer (L) or shorter (S) than the initially-recognized 35-amino-acid (P) motif (11,15). PLS PPR proteins, which contain ordered series of P, L and S motifs, have been almost exclusively associated with RNA editing in both mitochon-

*To whom correspondence should be addressed. Tel: +33 130 833 070; Fax: +33 130 833 319; Email: hakim.mireau@inra.fr
†These authors contributed equally to the paper as first authors.
Present address: Fabien Aubé, Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure 69007 Lyon, France.

dria and chloroplasts (16,17). In contrast, PPR proteins exclusively composed of P-type repeats were shown to participate in various aspects of organellar RNA expression ranging from gene transcription to mRNA translation (for review see (12,18)). Recent crystal structures of PPR proteins showed that PPR repeats fold into a hairpin formed by two antiparallel alpha helices whose repetition assemble into a right-handed superhelical spiral (19–23). Indeed, PPR domains organize highly specific RNA binding interfaces and PPR–RNA co-crystallization studies revealed that the solenoid like structure formed by the succession of PPR repeats wraps around single-stranded RNA (24,25). Evolutionary, biochemical and structural data support that combinations involving amino acid 5 and 35 in each repeat are critical for RNA recognition and form contacts with the RNA base to be bound (15,26–30).

The production of translation-competent mRNAs implies the correct processing of 5′ and 3′ extremities as well as the stabilization of mature transcripts. The definition of mRNA ends appears to be particularly essential in plant mitochondria where genes are transcribed from multiple promoters and transcription termination is not tightly controlled (31–33). A global analysis of mitochondrial mRNA ends in *Arabidopsis thaliana* revealed that most protein-coding transcripts exhibit multiple 5′ extremities whereas their 3′ end is generally very well defined and unique (34). Moreover, the relative abundance of individual 5′ ends identified for some transcripts can differ among Arabidopsis natural accessions (35,36). Further analyses have revealed that most plant mitochondrial mRNA extremities are generated post-transcriptionally. Several P-type PPR proteins—most of them belonging to the restorer of fertility-like subfamily—have been found to be involved in 5′ end formation of mitochondrial transcripts in Arabidopsis (37–43). However, the functional significance of 5′ formation for most mitochondrial mRNAs remains unclear as transcripts with incorrect 5′ ends (accumulating in *ppr* mutants) remain stable and correctly translated in most cases, and have no impact on plant development. Most of these 5′-processing PPR proteins are predicted to bind near the cleavage site they condition and the detection of upstream cleavage products support the recruitment of an endoribonuclease to trigger the RNA clipping. The associated RNase activities have long remained mysterious but recently, two mitochondria-targeted nucleases of the Nedd4-BP1, YacP Nuclease (NYN) family were shown to be involved in transcript 5′ end formation in plant mitochondria (44). Much less is known on the formation of mRNA 3′ ends in plant mitochondria. A single factor involved in the 3′ end processing of the *nad4* mitochondrial transcript has been identified so far (45). This factor, called mitochondrial stability factor 1 (MTSF1), encodes a PPR protein whose binding site coincides with the 3′ end of *nad4* mRNAs. The strong destabilization of *nad4* transcript in *mtsf1* mutants revealed a tight connection between 3′ end formation and mRNA stabilization in plant mitochondria. The associated mechanism supposes an early association of MTSF1 with the 3′ region of precursor *nad4* mRNA and that 3′-to-5′ exoribonucleases like PNP1 or RNR1 degrade unnecessary 3′ sequence extension until their progression is stopped by the presence of MTSF1. This protein-mediated protection process was initially described to occur in plastids and was shown to apply for both 5′ and 3′ end formation in this organelle, whereas it seems limited to 3′ processing in mitochondria (46–48).

Still, much remains to be understood regarding RNA processing and stabilization in plant mitochondria. In particular, the type and number of mitochondrial transcripts that are processed and stabilized by helical repeat proteins like PPR sitting on their 3′ end awaits clarification. In this study, we analyzed the function of the mitochondria stability factor 2 (MTSF2) protein in Arabidopsis and demonstrated that this mitochondria-targeted PPR protein is required for the 3′ processing and stabilization of the *nad1* exons 2–3 precursor transcript. This pre-mRNA is one of the three *nad1* precursor transcripts that are unified by two trans-splicing reactions to create the mature *nad1* mitochondrial mRNA. Our analysis reveals that the protein-based mechanism permitting the stabilization and the 3′ end processing is not limited to mature mitochondrial mRNA but concerns also precursor transcripts that are substrates of trans-splicing reactions.

MATERIALS AND METHODS

Plant material

The N654650 (*mtsf2-1*) Arabidopsis mutant line was obtained from the European Arabidopsis Stock Centre (http://arabidopsis.info/). *mtsf2-1* mutant plants were genotyped with the MTSF2-3 and the MTSF2-5 primers combined with the LB-SALK2 primer.

The *mtsf2-2* mutant was generated using the CRISPR-CAS9 technology. The guide sequence 5′-GTAGGTAGAAAGCTGATTGA-3′ specific to *MTSF2* and inducing a cleavage 684 nucleotides downstream of the ATG codon was chosen using the CRISPOR program on the Tefor web site (http://crispor.tefor.net/). A synthetic gRNA (sgRNA) expression cassette flanked by *attB* recombination sites and comprising the promoter of the Arabidopsis U6-26 snRNA (49), the 5′-G-N(19)-3′ guide sequence targeting the *MTSF2* gene, the tracrRNA scaffold (50) was designed. The sgRNA cassette was chemically synthesized (Intregrated DNA Technologies), subcloned into pDONR207 by Gateway™ BP reaction (Invitrogen) and subsequently combined with a CAS9 expression cassette by Gateway™ LR reaction (Invitrogen) into pDe-Cas9 binary vector (51). Col-0 plants were transformed and transgenic plants were selected on Basta (6 mg/l).

RNA extraction and analysis

Depending on the amount needed, total RNAs were isolated from flower buds using either TRIzol reagent (Life Technologies) or the RNeasy Plant Mini kit (Qiagen) following the manufacturer's recommendations and treated with Room Temperature Stable (RTS) DNase only when RNA were prepared for quantitative (Q)-RT PCR (Mobio). RNA gel blot analyses were performed as indicated in (52). Gene-specific probes were amplified by polymerase chain reaction (PCR) using primers indicated in Supplementary Table S1 and radiolabeled with the Prime A Gene labeling

kit (Promega) following the recommendations of the manufacturer. Quantitative RT-PCR was performed as previously explained in (52), using the same set of primers.

Primers

The DNA oligonucleotides used in this work are listed in Supplementary Table S1.

Circular RT-PCR

Five microgram of total RNA were circularized with 40 U of T4 RNA ligase (New England Biolabs), following the manufacturer's instructions. Circularized RNA were phenol extracted and precipitated in one volume of isopropanol, for 2h at −20°C. Circularized RNA were resuspended in 14.5 μl of water and first strand complementary DNA (cDNA) synthesis was done for 3 h at 40°C using 400 U of M-MLV reverse transcriptase (Fermentas), 8 mM of random hexamers (Eurofins), 1X M-MLV buffer, 0.5 mM dNTPs and 40 U of Riboblock RNase inhibitor (Fermentas). The obtained cDNA were diluted four times, and 5 μl of the obtained cDNA solution were used for PCR amplification with divergent primers. The primers used for mapping *nad1* exon 1, exons 2–3 and exons 4–5 precursor transcripts are listed in Supplementary Table S1. Amplified PCR products were gel purified, cloned into pCR2.1®-TOPO® TA vector (ThermoFisher Scientific) and inserts of independent recombinant plasmids were sequenced after *Escherichia coli* transformation.

Blue native gel and respiratory complex in-gel activity assay

Crude organelle extracts enriched in mitochondria were prepared as previously described (53) using young Arabidopsis seedlings grown *in vitro* for 10 days. Succinctly, 400 mg of Arabidopsis plantlets were ground in 4 ml of extraction buffer (75 mM MOPS-KOH pH 7.6, 0.6 M Sucrose, 4 mM ethylenediaminetetraacetic acid (EDTA), 0.2% [w/v] polyvinylpyrrolidone 40, 8 mM Cys and 0.2% [w/v] bovine serum albumin (BSA)). The lysate was filtered through one layer of Miracloth and centrifuged for 4 min at 1300 *g*, and the resulting supernatant was centrifuged again for 20 min at 25 000 *g*. Organelle containing pellets were resuspended in 400 μl of 10 mM MOPS-KOH pH 7.2, and 0.3 M Sucrose, completed with three volumes of water (1.2 ml) and then centrifuged to 25 000 *g* for 20 min. The pellets were finally resuspended in 200 μl of resuspension buffer (50 mM BisTris/HCl pH 7, 0.5 M 6-aminohexanoic acid, 1 mM EDTA and 0.3 M Sucrose) before subsequent analysis. Organelle suspensions were then complemented with 1% of n-dodecyl β-D-maltoside (DDM), incubated on ice for 30 min and centrifuged for 10 min at 100 000 *g*. The resulting supernatant was supplemented with a third volume of 5% (w/v) Coomassie Blue G-250 prepared in 20 mM BisTris/HCl pH 7, and 0.5 M 6-aminohexanoic acid. Subsequently, one-third of the sample volume was completed with 4× loading buffer (200 mM BisTris, 64 mM HCl, 200 mM NaCl, 0.004% [w/v] Ponceau S and 40% [v/v] glycerol). Twenty microliter of protein samples per lane were loaded on a precast 3–12% (w/v) polyacrylamide NativePAGE™ Bis/Tris gel (Invitrogen). Gels were run at 80 V overnight at 4°C and then for an additional 2 h at 200 V the next day following manufacturer's instructions.

Following electrophoresis, BN-PAGE gels were either transferred to polyvinylidene difluoride (PVDF) membranes (see below), or stained to reveal respiratory complexes. Complex I activity was revealed by incubating the gel in the presence of 0.2% nitroblue tetrazolium, 0.2 mM of NADH and 0.1 M Tris–HCl (pH 7.4). For complex IV, the gel was incubated in 50 mM KH_2PO_4 (pH 7.4), 1 mg/ml reduced cytochrome c and 0.1% 3,3′-diaminobenzidine. When colorations were sufficient, gels were transferred to a fixing solution containing 30% (v/v) methanol and 10% (v/v) acetic acid to stop the reactions.

Immunoblot analysis

Denaturing protein gels were electro-transferred to PVDF membrane under semidry conditions in transfer buffer (25 mM Tris, 192 mM Gly, 20% [v/v] methanol and 0.025% [w/v] sodium dodecyl sulphate (SDS)) at 15 mA for 30 min. BN-PAGE gels were transferred under liquid conditions in 50 mM Bis/Tris and 50 mM Tricine at 20 V overnight at 4°C. Membranes were incubated with antibodies specific to RISP (diluted 1:5000; 54), COX2 (diluted 1:2000; Agrisera), NAD9 (1:100 000; 55), alternative oxidase (AOX) (diluted 1:50; a gift of David Day, University of Western Australia) or ATPβ (diluted 1:5000; Agrisera). Detection was carried out using peroxidase-conjugated secondary antibodies and enhanced chemiluminescence reagents (Western Lightning Plus ECL, Perkin Elmer). For SDS gels, the apparent molecular masses of the proteins were estimated with Pageruler prestained protein ladder (Fermentas).

Production of recombinant MTSF2 and gel shift assays

The *MTSF2* coding sequence deprived of the upstream region encoding the mitochondrial presequence was PCR amplified using the GWMTSF2-5 and GWMTSF2-3 primers (Supplementary Table S1). The obtained PCR product was subcloned into pDONR207 by Gateway™ BP reaction (Invitrogen) and subsequently transferred into pMAL-TEV (homemade) by Gateway™ LR reaction (Invitrogen). The resulting MBP-MTSF2 protein fusion was expressed in Rosetta™ (DE3) *E. coli* strain for 3h at 37°C. Bacterial pellets were lysed in 50 mM HEPES-KOH (pH7.8), 150 mM NaCl, 1% glycerol, 0.01% CHAPS and 5 mM β-mercaptoethanol using the One Shot cell disruption system (Constant Systems). After centrifugation at 20 000 *g* for 15 min, the MBP-MTSF2 fusion was purified from the supernatant on amylose resin (Brand) following manufacturer's recommendations. Protein purity was verified on SDS-polyacrylamide gel electrophoresis (PAGE) before proceeding to gel shift experiment.

DNA fragments spanning the 3′ end of *nad1* exons 2–3 pre-mRNA were PCR amplified using primers indicated in Supplementary Table S1. Radiolabeled RNA probes were generated by *in vitro* transcription with T7 RNA polymerase (Promega) in the presence of 30 μCi of [α-32P]UTP according to manufacturer's instructions. Probes were gel purified on 5% polyacrylamide gel, precipitated and quantified on a NanoDrop apparatus (Thermo Scientific). Gel

shift reactions were performed in 50 mM HEPES-KOH (pH 7.8), 100 mM NaCl, 0.1 mg/ml BSA, 4 mM DTT, 10% glycerol, 2 mg/ml heparin, 10 U Riboblock (Fermentas), containing 100 pM radiolabeled RNA and protein concentrations indicated in figure legends. Reactions were carried out for 30 min at 25°C, and run on 5% polyacrylamide gel at 150 volts in 1 × THE (66 mM HEPES, 34 mM Tris (pH 7.7), 0.1 mM EDTA). The gels were then dried and exposed to a phosphorimager screen (FLA-9500 Fujifilm).

Prediction of MTSF2 RNA binding site

A putative RNA binding motif for MTSF2 was generated by using the combinations of amino acids at positions 5 and 35 of each PPR repeat (which were predicted with TPRpred (https://toolkit.tuebingen.mpg.de/tprpred)) and weighting A/U/G/C nucleotides for each combination according to Barkan *et al.*, (27). To predict the best potential binding sites for MTSF2, we used the FIMO program (http://meme-suite.org/tools/fimo), which searches motifs within sequence databases and ranks them by *P*-values (56). The RNA binding motif obtained for MTSF2 was then used to search the entire Col-0 mitochondrial genome (JF729201.1) and its fully edited version (homemade) with FIMO and get a list of potential binding sites ranked by *P*-values.

RESULTS

Arabidopsis *mtsf2* display a globally retarded growth phenotype

In a global effort to better understand gene expression in higher plant mitochondria, we assembled a large collection of Arabidopsis mutants bearing transfer DNA (T-DNA) insertion in nuclear genes encoding mitochondrially targeted P-type PPR proteins. Its characterization revealed that the *mtsf2* mutant line for which homozygous mutants displayed a global retarded growth phenotype compared to wild-type plants (Figure 1A). The PPR gene affected in this line corresponded to the AT1G52620 gene and encoded a 92-kDa protein comprising 20 canonical P-type PPR repeats according to TPRpred predictions (Figure 1B; https://toolkit.tuebingen.mpg.de/tprpred). Reverse transcription (RT)-PCR analysis indicated that no detectable full-length mRNA derived from the AT1G52620 gene accumulated in the mutant, strongly suggesting that it corresponded to a null mutant (Supplementary Figure S1). To confirm that the observed phenotype was effectively associated with inactivation of the AT1G52620 gene and since no other T-DNA insertion within the coding sequence could be identified in public mutant collections, we decided to create a second mutant line using the CRISPR-Cas9 technology. A guide RNA mediating DNA cleavage at position 684 downstream of the ATG codon of AT1G52620 and with very unlikely off-target effects was chosen using the CRISPOR on-line tool (http://crispor.tefor.net/). The guide RNA coding sequence was cloned downstream of the Arabidopsis U6 promoter and inserted into the pDe-Cas9 binary vector (51). Nine independent Arabidopsis transformants were obtained and 200 plants derived from five independent progenies were phenotyped. Out of these 1000 plants, 164 plants displayed developmental alterations that were strictly identical to the

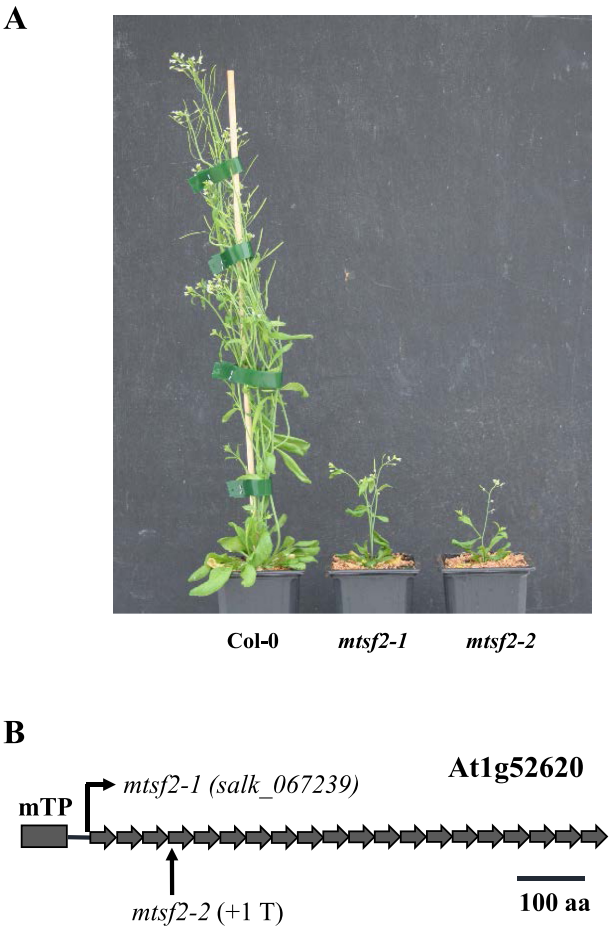

Figure 1. The Arabidopsis *mtsf2* mutants are strongly delayed in their development. (**A**) Eight-week-old plants showing the global vegetative phenotype of *mtsf2* mutants compared to the wild-type (Col-0). *mtsf2* mutant plants exhibit strongly reduced size and bear twisted and dark green rosette leaves. They are fertile but their siliques are shorter compared to Col-0. (**B**) Schematic representation of the MTSF2 protein depicting the 20 predicted PPR repeats and the positions of *mtsf2-1* (T-DNA insertion) and *mtsf2-2* (+1T at position 685) mutations. mTP: mitochondrial transit peptide.

mtsf2-1 mutant line. A DNA fragment of AT1G52620 covering the guide RNA cleavage site was amplified by PCR from 20 of these plants and sequenced. Sequence analysis revealed that most mutant plants contained a single T insertion right after the induced DNA cleavage site. The associated frameshift created an in-frame TGA stop codon immediately downstream of the T insertion. This second mutant allele was named *mtsf2-2* and was chosen for further phenotypical and molecular characterization (Figure 1A and B).

mtsf2 mutant plants exhibited large developmental differences compared with Col-0 reference plants. In particular, both mutant lines grew much slower than the wild-type and reached about half the size of Col-0 plants when cultured on soil in the greenhouse for 3 months (Figure 1A). Additionally, *mtsf2* plants showed deformed and dark green rosette leaves and were late flowering but remained fertile (Figure 1A).

The Arabidopsis subcellular database SUBA (http://suba3.plantenergy.uwa.edu.au/) strongly supported the mi-

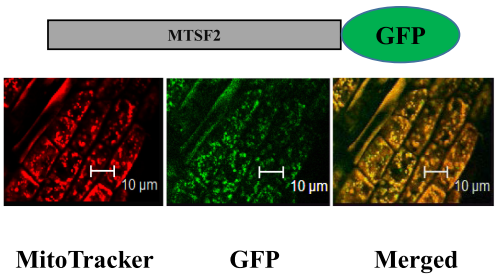

Figure 2. The *MTSF2* gene encodes a mitochondria-targeted PPR protein. Confocal microscope images analyzing the subcellular distribution of an MTSF2:GFP translational fusion in root cells of young Arabidopsis transgenic plants. Prior to observation, roots were soaked in a solution of Mitotracker Red to label mitochondria. The left panel shows the red fluorescence of mitochondria revealed by the Mitotracker Red. The center panel presents the green fluorescence of the GFP. Merged signals are shown in the right panel.

tochondrial localization of the MTSF2 protein. To verify the cellular distribution of MTSF2, a GFP translational fusion comprising the complete coding sequence of MTSF2 was generated and expressed in transgenic Arabidopsis plants. Root cells of transformed plants were analyzed by confocal microscopy and the GFP fluorescence distribution appeared to concentrate in small cell bodies throughout the cytosol. The use of MitoTracker Red as control on the same samples indicated that these signals corresponded effectively to mitochondria (Figure 2). This analysis confirmed that MTSF2 was effectively targeted to Arabidopsis mitochondria.

mtsf2 mutants do not accumulate detectable respiratory complex I

The confirmation of the mitochondrial localization of MTSF2 prompted us to consider that the developmental alterations observed in *mtsf2* mutants could result from a respiratory dysfunction in these plants. To identify the cause of this potential respiratory deficiency, we examined the steady-state levels of the different respiratory chain complexes on blue-native gels. Mitochondria-enriched pellets were prepared from both wild-type and mutant plants and solubilized in 1% DDM. Following migration, the abundance of respiratory complexes was either visualized by in-gel activity staining or by western blot analysis with appropriate antibodies. This allowed us to observe that most respiratory complexes accumulated to normal levels in both *mtsf2* mutants compared to the wild-type (Figure 3A and B). The only exception concerned complex I, which could not be detected in *mtsf2* extracts at all (Figure 3A and B). To further analyze the respiratory activity in *mtsf2* plants, we assessed the induction of the alternative respiratory pathways in the mutants by measuring the expression levels of alternative NADH dehydrogenases (NDA and NDB) and AOX genes. Quantitative RT-PCR analysis revealed the overexpression of several components of alternative respiratory pathways notably the *AOX1A*, *AOX1D*, *NDA1* and *NDB4* genes (Supplementary Figure S2A). The two *AOX* transcripts accumulated respectively four and eight times more compared to the wild-type. The NADH dehydrogenase genes over-accumulated respectively 2 and 64 times more than in wild-type plants. The overexpression of AOX was further visualized by probing total mitochondrial protein extracts with an antibody to AOX (Supplementary Figure S2B), confirming the strong engagement of the alternative respiratory pathways in *mtsf2* plants.

Altogether, these observations allowed us to conclude that *mtsf2* mutants represent complex I deficient plants. The developmental abnormalities of *mtsf2* plants (notably their slow growth, curly leaves and late flowering phenotype) are in complete agreement with this kind of respiratory alteration since they are typical of complex I mutants as previously observed either by us (6,45,52) or by other groups (see for instance (57–59)).

The stability of the *nad1* exons 2–3 precursor is compromised in *mtsf2* mutants

The role of PPR proteins in the processing of organellar RNA led us to hypothesize that the processing of one or several mitochondrial mRNAs encoding complex I subunit was defective in *mtsf2* mutants. In a first approach, the steady state levels of all mitochondria-encoded protein-coding mRNAs were measured by quantitative RT-PCR in *mtsf2* and wild-type plants. This approach revealed a major difference for the *nad1* transcript whose mature form was found to accumulate 500 times less in *mtsf2* mutants compared to the wild-type (Figure 4). This reduction was observed across the entire *nad1* transcript (for all exons), which strongly suggested a global destabilization of *nad1* mRNA in *mtsf2* plants. A slight decrease of *nad2* mature mRNA abundance was also detected, but this only concerned the upstream mRNA region encompassing the first two *nad2* exons. All other mitochondria encoded mRNAs accumulated to almost normal levels compared to the wild-type. In Arabidopsis, the mature form of *nad1* results from the fusion of three distinct pre-mRNA molecules named *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 that are reunified by two trans-splicing reactions (Figure 5A). To better understand the cause of *nad1* destabilization in *mtsf2* plants, the steady levels of these three precursors were measured by quantitative RT-PCR. Interestingly, whereas the steady state levels of *nad1* exon 1 and *nad1* exons 4–5 pre-mRNAs were increased, *nad1* exons 2–3 transcript abundance showed a significant reduction in *mtsf2* plants compared to the wild-type (Figure 5B). These observations were corroborated by RNA gel blot analysis, which confirmed the absence of detectable mature *nad1* mRNA and *nad1* exons 2–3 precursor forms in the two mutant lines (Figure 6A). A probe specific to the first portion of *nad1* intron 3 detected two *nad1* exons 2–3 transcripts in Col-0, which were both missing in the *mtsf2* mutants. The slight over-accumulation of *nad1* exon 1 and *nad1* exons 4–5 transcripts detected by quantitative RT-PCR was also visible in this second approach. The reduction of mature *nad2* mRNA abundance was noticeable in *mtsf2* plants, as well as the concomitant over-accumulation of unspliced precursor forms (Figure 6B). Quantitative RT-PCR further indicated that splicing of *nad2* intron 1 and 2 were significantly diminished in *msf2* mutants (Supplementary Figure S3).

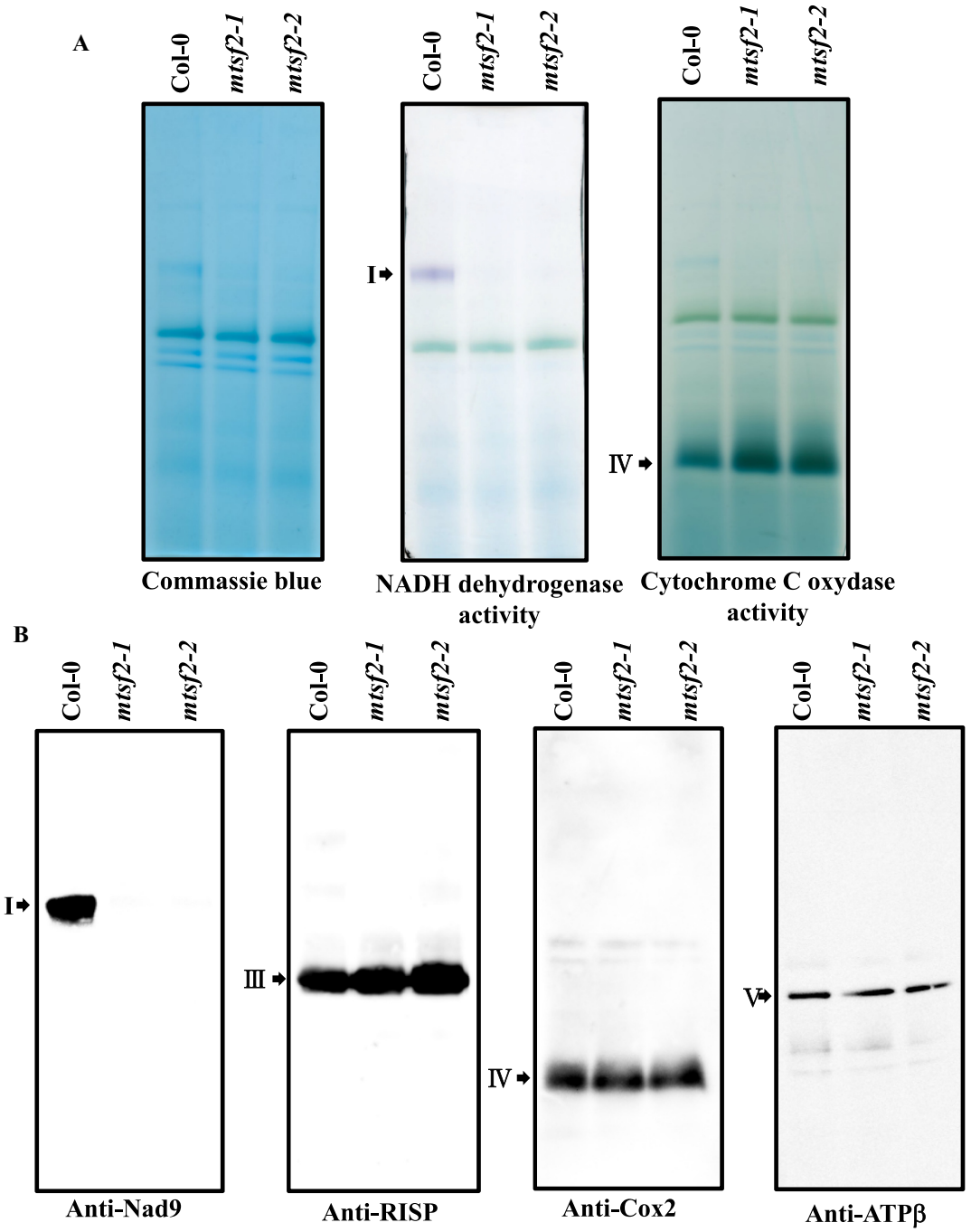

Figure 3. Arabidopsis *mtsf2* lack detectable respiratory complex I. (**A**) Crude mitochondrial extracts prepared from wild-type and *mtsf2* plants, separated on BN-PAGE gels and stained with Coomassie blue. In-gel staining revealing the NADH dehydrogenase activity of complex I and cytochrome c oxidase activity of complex IV are also presented. (**B**) Following migration, BN-PAGE gels were transferred to membranes and blots were hybridized with antibodies to mitochondrial NAD9, RISP, COX2 and ATPβ. The different respiratory complexes are designated by their roman numeral.

Overall, these results strongly suggested a global instability of *nad1* mRNA in *mtsf2* mutants that likely resulted from a specific destabilization of *nad1* exons 2–3 precursor transcript. The moderate increase (4–16 times) of *nad1* exon 1 and *nad1* exons 4–5 pre-mRNA abundance were likely consecutive to this absence since the two trans-splicing reactions involving *nad1* introns 1 and 3 cannot occur in the absence of *nad1* exons 2–3 precursor. Additionally, a partial effect of *mtsf2* inactivation on *nad2* intron 1 and 2 splicing was also detected from these analyses.

The MTSF2 binding site lies in the 3′ region of *nad1* exons 2–3 transcript

To better understand the role of MTSF2 in the stabilization of *nad1* exons 2–3 pre-mRNA, we took advantage of recent progress on the recognition of RNA by PPR proteins to pre-

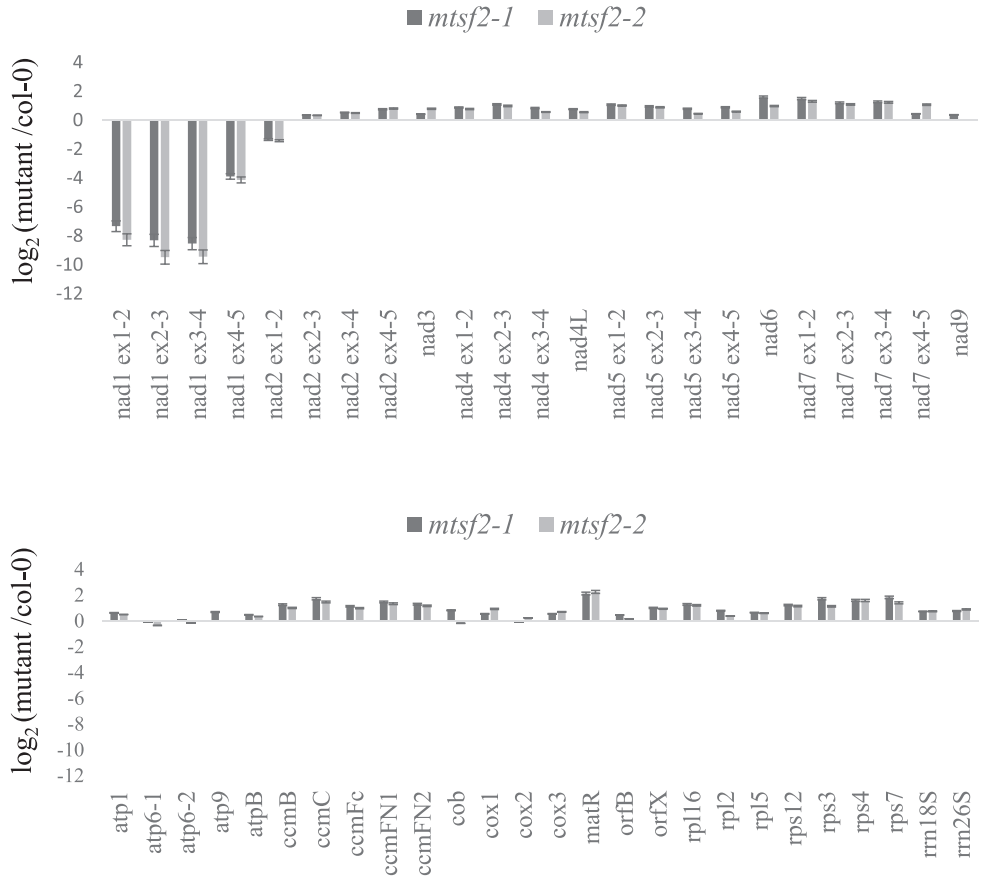

Figure 4. Mature *nad1* mRNA strongly under-accumulate in *mtsf2* mutants. Quantitative RT-PCR measuring the steady state levels of mature mitochondrial mRNAs in Col-0 and *mtsf2* plants. The histograms show log2 ratios of mutants to wild-type. A single PCR was considered for mRNAs carrying no introns, whereas the accumulation of individual exons was analyzed for intron-containing transcripts. Three biological replicates and three technical replicates were used per genotype; standard errors are indicated.

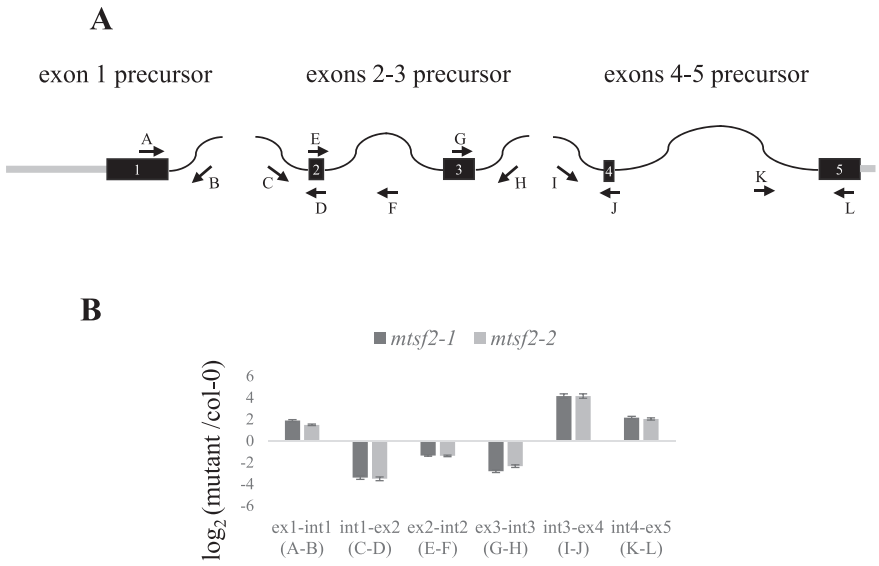

Figure 5. The under-accumulation of mature *nad1* mRNA results from the specific destabilization of *nad1* exons 2–3 precursor transcript in *mtsf2* plants. (**A**) Schematic representation of *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 precursor transcripts and positions of primers used for quantitative RT-PCR analysis. (**B**) Quantitative RT-PCR measuring the steady state levels of *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 precursor RNA in Col-0 and *mtsf2* plants. The histograms show log2 ratios of mutants to wild-type. The PCR reactions were performed with the indicated primer pairs. Three biological replicates and three technical replicates were used per genotype; standard errors are indicated.

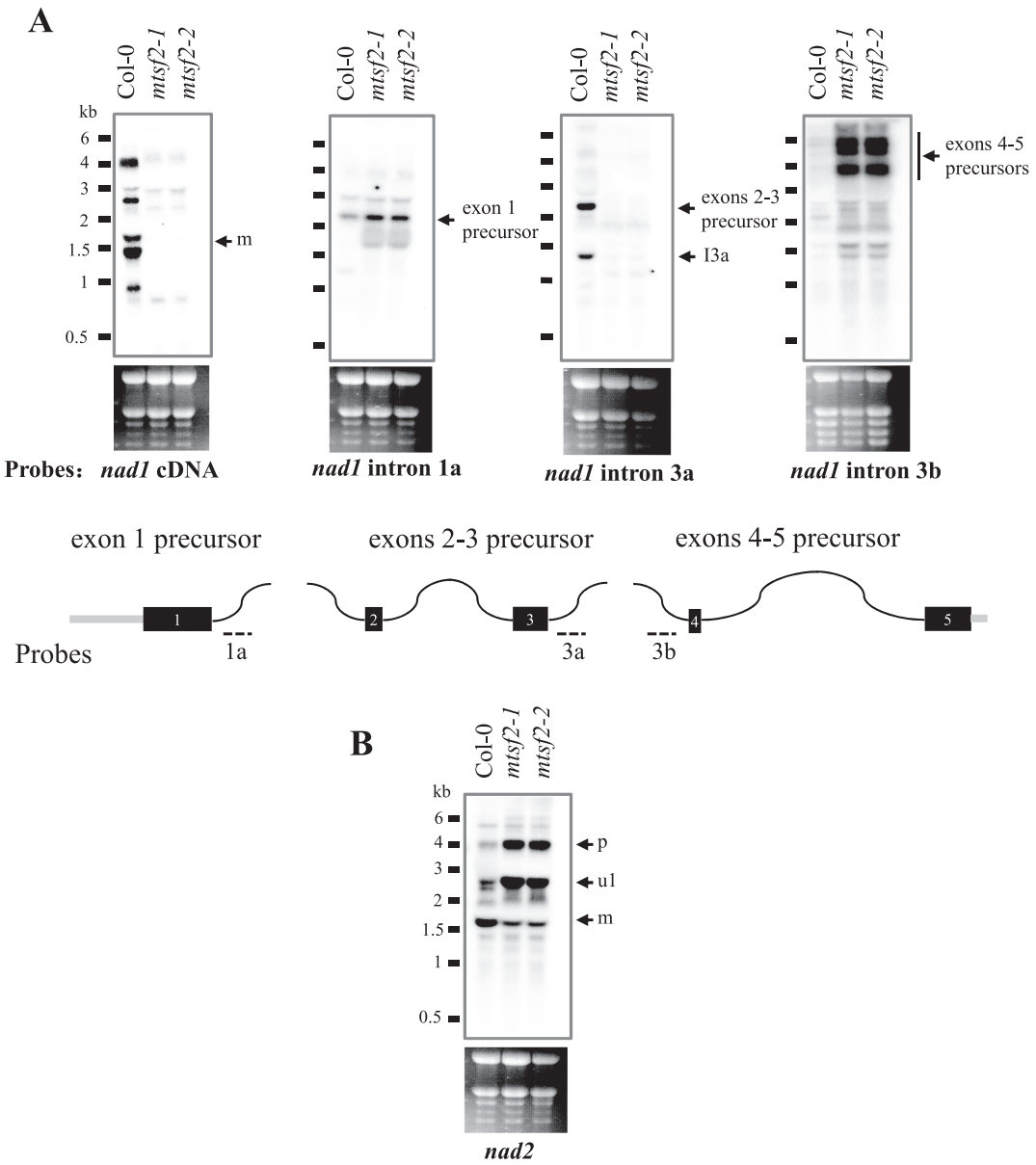

Figure 6. MTSF2 is essential to produce stable *nad1* exons 2–3 precursor RNA and mature *nad2* transcript. (**A**) RNA gel blot analyses detecting the mature form of *nad1* and the different *nad1* precursor transcripts (*nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5) in wild-type (Col-0) and *mtsf2* plants. The blots were performed on total RNA and hybridized with the indicated probes. A schematic representation showing the three *nad1* precursor transcripts is shown in the center with UTRs materialized as gray boxes, exons as black boxes and introns as curved lines. The probes used to detect the three different pre-mRNAs are also indicated. (**B**) RNA gel blot detecting the different forms of *nad2* mRNA accumulating in Col-0 and *mtsf2* mutants. A full-length *nad2* cDNA probe was used in this experiment. Ethidium bromide staining of ribosomal RNAs is shown below the blots and serves as a loading control. m: mature mRNAs; I3a: 5′-half of *nad1* trans-intron 3 accumulating as a linear molecule; u1: unspliced *nad2* transcript containing intron 1; p: *nad2* unspliced precursor transcript.

dict which mitochondrial sequence in Arabidopsis could be the most likely target of MTSF2. It has been recently established that each PPR motif aligns to a single nucleotide of the RNA target in modular patterns, and that the combinatorial action of amino acids 5 and 35 of each repeat are critical for base selection (26,27,29). This has been the basis of an RNA recognition code for PPR proteins that established relationships between the identity of these two amino acids and each RNA base (25–30). We applied this code to the 20 PPR repeats of MTSF2 and obtained a degenerate RNA recognition sequence integrating all potential binding sites for MTSF2 (Figure 7A). The obtained motif was then used to scan the non-edited and fully edited Arabidopsis mitochondrial genome and the sequence showing the highest probability for a possible association with MTSF2 was found to locate 1376 nt downstream of the third exon of *nad1* (Figure 7B). Interestingly, this sequence (GAUACAUCGGUGUAAAAGGU) is found in the region encoding the first half of *nad1* intron 3 and in contrast to all other predicted binding sites was the only one that could be con-

A

B

Rank	P-value	Genome position	Location	Strand	Sequence
1	5.74E-07	118933-118952	nad1-intron3	-	GATACATCGGTGTAAAAGGT
2	5.31E-06	289277-289296	intergenic	-	GGTTCACCGGTGAAAACTGA
3	1.23E-05	186582-186601	intergenic	+	TCCCCATTGCTGTAACCCGG
4	1.57E-05	63099-63118	intergenic	-	GAAACAGTGGTGCAAACTCT
5	1.57E-05	276959-276978	intergenic	+	CATATACCGATGTAAACAAA

Figure 7. Prediction of the MTSF2 binding site according to the PPR code. (**A**) The amino acid at positions 5 and 35 of each MTSF2 PPR repeat were extracted and are listed from N to C-terminus. The obtained combinations were then used to calculate the probabilities of nucleotide recognition by each individual PPR repeat according to the PPR code (27,28,30). The sequence logo depicting these probabilities was obtained with http://weblogo.berkeley.edu/. (**B**) The obtained recognition motif was then used to scan both strands of the complete non-edited and fully edited Arabidopsis mitochondrial genome from Col-0. The five most-probable MTSF2 binding sequences are shown. The *P*-values were determined with the FIMO program.

tained in the *nad1* exons 2–3 transcript (Figure 7B). To see whether MTSF2 showed significant affinity for this potential RNA target, gel shift assays were performed using a series of *in vitro*-transcribed RNA probes located around this predicted binding site. A slightly truncated version of MTSF2 lacking its mitochondrial targeting sequence but containing all PPR repeats present in the wild-type protein was expressed in fusion to the maltose binding protein in *E. coli* and purified to homogeneity (Figure 8A). Among the different RNA probes tested, MTSF2 showed specific binding affinity to the one containing the predicted binding site (Figure 8B and C). In contrast, MTSF2 displayed no binding to probes derived from the same region of *nad1* intron 3 but lacking the GAUACAUCGGUGUAAAAGGU segment. We then looked for evidence supporting an association of MTSF2 with this RNA segment *in vivo*. The binding of RNA-stabilizing factors like PPR proteins have been shown to lead to the accumulation of short RNA sequences in plant organelles corresponding to their binding sites (45,47,60). On the basis of RNA-seq data, many such RNA footprints were recently mapped to the Arabidopsis mitochondrial genome to identify potential binding sites of RNA stability factors (61). We searched this database to see if any RNA footprints mapped to the 3′ region of *nad1* exons 2–3 pre-mRNA. A 35-nucleotide RNA fragment whose 3′ extremity coincided with the MTSF2 binding site could be identified, strongly supporting association of MTSF2 with this RNA region *in vivo* (Supplementary Figure S4).

Overall, these observations revealed that the MTSF2 PPR protein associates *in vivo* with a 20-base RNA segment located 1376 bases downstream of *nad1* exon 3, a region that also comprises the sequences encoding the first portion of *nad1* trans-intron 3.

The 3′ end of *nad1* exons 2–3 precursor coincides with the MTSF2 binding site

To further elucidate the origin of *nad1* mRNA destabilization in *mtsf2* plants, we also mapped the 5′ and 3′ extremities of the *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 precursor transcripts. We chose circular RT-PCR because it seemed to be well suited for mapping the extremities of the low abundant *nad1* precursor transcripts. The amplification profiles obtained for *nad1* exon 1 and *nad1* exons 4–5 pre-mRNAs were identical in Col-0 and *mtsf2* plants except that some of the bands were much more visible in the mutant samples on agarose gels (Figure 9A). This difference of amplification efficiency likely results from the higher accumulation of *nad1* exon 1 and *nad1* exons 4–5 pre-mRNAs in *mtsf2* plants (Figures 5 and 6), making some fragments easier to amplify from the mutant extract. Each of these amplification products were sequenced and revealed no difference between the two genotypes regarding the different 5′ and 3′ termini of *nad1* exon 1 and *nad1* exons 4–5 pre-mRNAs (Figure 9A). In contrast, an amplification product corresponding to *nad1* exons 2–3 could only be produced from

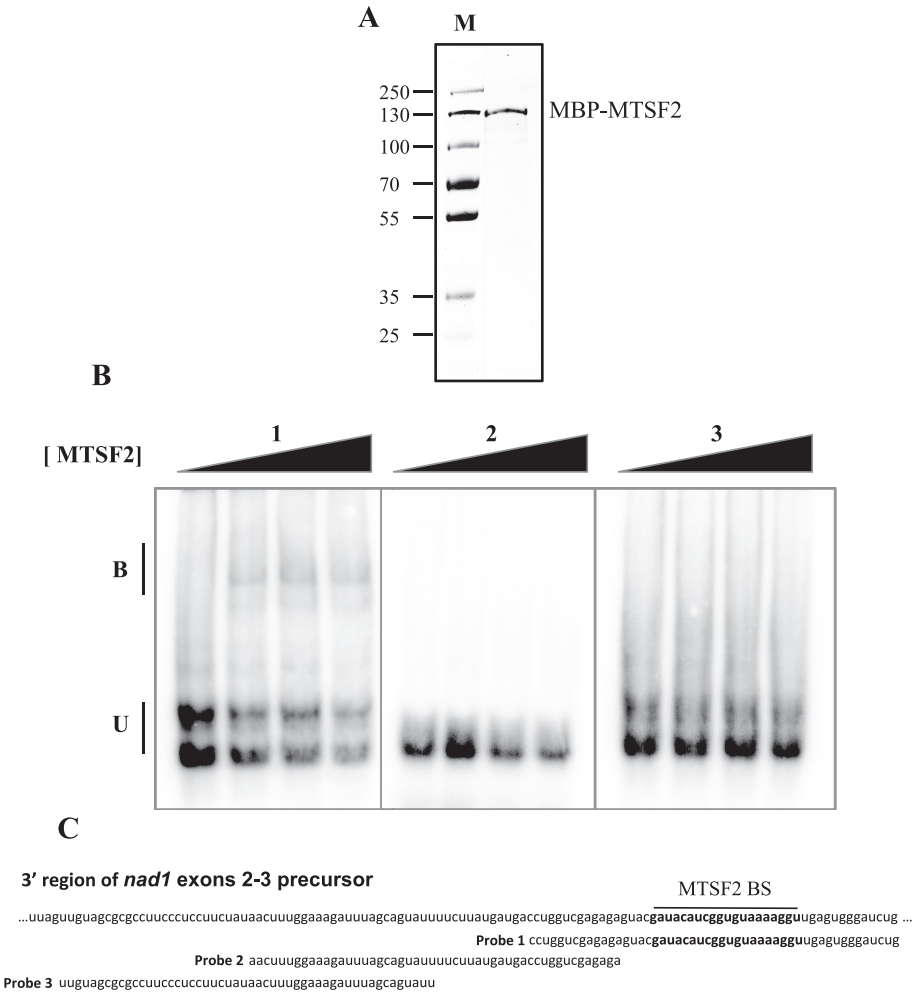

Figure 8. Gel mobility shift assays showing the binding of MTSF2 to a short RNA segment of the *nad1* exons 2–3 precursor 3′ region and corresponding to its best-predicted binding site according to the PPR code. (**A**) Sodium dodecyl sulphate-polyacrylamide gel electrophoresis protein gel stained with Coomassie blue showing the purity of the MBP-MTSF2 fusion protein overexpressed and purified from *Escherichia coli*. Ten micrograms of the purified protein were loaded on the gel. M: protein size marker (kDa). (**B**) Gel mobility shift assays performed with MBP-MRSF2 and the RNA probes depicted in panel C. For each probe, the reactions contained (from left to right) 0, 50 100 and 150 nM of pure MBP-MTSF2 protein, 100 pM of gel-purified RNA probe and 2 mg/ml heparin. U: unbound probe; B: bound probe. (**C**) Sequences and positions of the different probes used in the gel shift reactions. The MTSF2 binding site is indicated (MTSF2 BS).

Col-0 plants (Figure 9A). The natural low abundance of *nad1* exons 2–3 precursor, along with its high instability in *mtsf2* plants, likely explains this lack of amplification from the mutant extract. Cloning and sequencing of the fragment generated from wild-type plants indicated that the vast majority of *nad1* exons 2–3 transcripts bears a 5′ end starting 716 bases upstream of exon 2 and terminated 1395 or 1396 nt downstream of exon 3 (Figure 9B). Remarkably, these two 3′ ends mapped at the end, or one base downstream, of the identified MTSF2 RNA binding site. These results therefore revealed that the MTSF2 binding site coincided with the 3′ extremities of *nad1* exons 2–3 pre-mRNA and that the binding of MTSF2 to this region is required for the stabilization of *nad1* exons 2–3 precursor in Arabidopsis mitochondria.

DISCUSSION

MTSF2 stabilizes the mitochondrial *nad1* exons 2–3 precursor transcript in Arabidopsis

The molecular processes leading to transcript stabilization in plant mitochondria have been poorly explored up to now. However, a mechanism involving the association of RNA binding proteins to transcript extremities has been described for a few transcripts in chloroplasts and mitochondria (12,45,62–64). The underlying mechanisms propose that these proteins bind to mRNA 5′ or 3′ UTRs and stabilize them by blocking the progression of 5′-to-3′ or 3′-to-5′ exonucleases. This mechanism implies that the binding site of these stability factors coincides with the 5′ or 3′ extremity of their target transcript. They are thus responsible for both mRNA stabilization and transcript end processing. Among the identified factors acting this way, a single one was previously found to be essential for mRNA

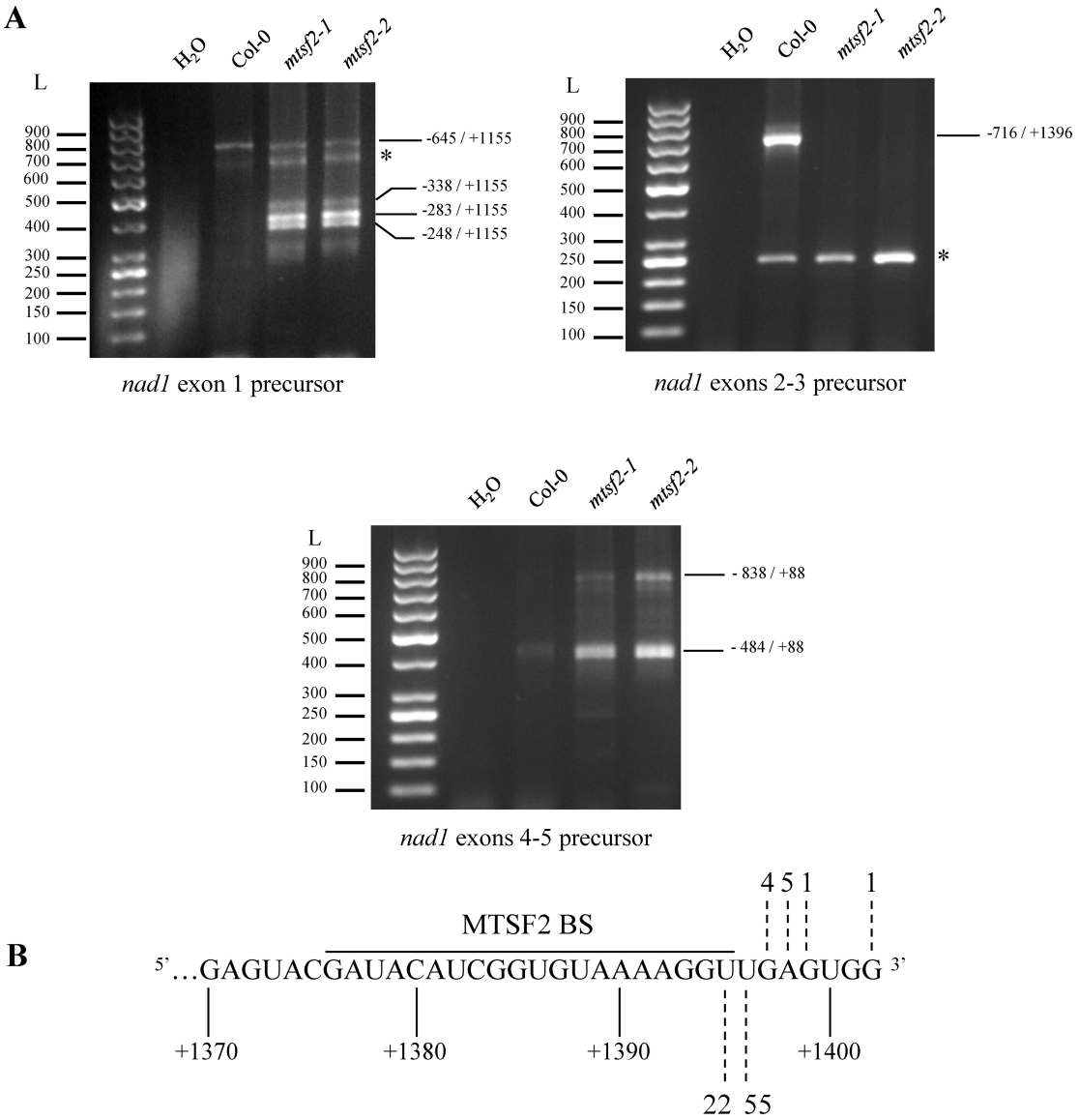

Figure 9. The *nad1* exons 2–3 transcript 3′ extremity coincides with the MTSF2 RNA binding site. (**A**) Agarose gel showing circular-RT PCR amplification products obtained for the *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 precursor transcripts in wild-type (Col-0) and *mtsf2* mutants. All visible bands were cloned and sequenced in the different genotypes. The coordinates of 5′ and 3′ extremities of each fragment are indicated with respect to the first and last exon contained in each precursor transcript. *: artifact amplification products not related to *nad1*. L: DNA size ladder, H_2O: negative control. (**B**) Partial RNA sequence showing the two major 3′ ends of *nad1* exons 2–3 pre-mRNA. The indicated coordinates are numbered with respect to the last exon present in the *nad1* exons 2–3 precursor (formally *nad1* exon 3). The dotted lines indicate the positions of the *nad1* exons 2–3 transcript 3′ ends found in the wild-type (Col-0). Numbers next to dotted lines represent how often a particular end was found after sub-cloning and sequencing the major band shown for *nad1* exons 2–3 in panel A. The sequence fragment corresponding to the MTSF2 binding site is also indicated (MTSF2 BS).

stabilization in plant mitochondria. It corresponded to the MTSF1 PPR protein that is required for the stability of *nad4* and whose binding site corresponded to the last nucleotides of mature *nad4* mRNA (45). Therefore, the number and more importantly the type of transcripts that are stabilized by such protein-mediated process in plant mitochondria remain largely unclear. The RNA segments corresponding to the binding sites of several of these stability factors were shown to survive degradation and to accumulate in plant organelles in the form of short RNA footprints (47,60). The mapping of a large number of such footprints in Arabidopsis mitochondria identified 14 protein-coding mature mRNAs that are potentially stabilized by association with protective proteins at their 3′ end (61). The present study reveals for the first time that the range of transcripts stabilized by this mechanism is not limited to mature mRNAs but that precursor transcripts whose 3′ extremity terminates in the first half of a trans-splicing intron are also stabilized by the association with a protective protein. We effectively revealed that the mitochondria-targeted MTSF2 PPR protein is required for both the stabilization and concomitantly the 3′ end processing of the *nad1* exons 2–3 precursor transcript. *nad1* is one of the three mRNAs of mitochondria whose mature form in most seed plants results

from the fusion of three independently-transcribed precursor transcripts called *nad1* exon 1, *nad1* exons 2–3 and *nad1* exons 4–5 via two trans-splicing reactions (Figure 5). We observed that the lack of *nad1* mature mRNA in *mtsf2* mutants correlated with the specific destabilization of the central precursor *nad1* exons 2–3 (Figures 5 and 6). Moreover, the *nad1* exons 4–5 transcript, which covers the 3′ region of *nad1*, slightly over-accumulated in *mtsf2* plants and therefore could not explain the general destabilization of mature *nad1* mRNA in the mutants. Additionally, we observed that *nad1* exon 1 and *nad1* exons 4–5 precursors bore 5′ and 3′ extremities identical to the wild-type in the absence of MTSF2, implying that these precursor RNAs were both normally stabilized and processed in *mtsf2* mutants (Figure 9). Further, by predicting candidate binding sites for MTSF2, we demonstrated that this PPR protein associates to the 3′ extremity of *nad1* exons 2–3 pre-mRNA with high affinity. A low-abundance short RNA matching the MTSF2 binding site could be found in a dedicated database listing potential binding sites of organellar RNA stabilizing factors (61), providing very strong support for an effective association of MTSF2 to the 3′ extremity of *nad1* exons 2–3 RNA *in vivo*. Altogether, our results strongly support the model implying an association of MTSF2 with the 3′ end *nad1* exons 2–3 precursor RNA and acting as a steric block against exoribonucleolytic activity. Interestingly, this mechanism appears to be conserved in angiosperms as suggested by the almost perfect sequence conservation of the MTSF2-binding site among monocot and dicot plant species (Figure 10). The high level of sequence conservation of MTSF2 in seed plants and more importantly the amino acids involved in PPR–RNA interaction further support conserved function of MTSF2 in angiosperms (Supplementary Figure S6). Our study thus reveals that 3′ extremity of 5′-half introns of mitochondrial precursors requires active stabilization to survive degradation and that the association with their cognate 3′-half is insufficient to protect them from the degrading activity of 3′-to-5′ mitochondrial exonucleases. A sufficient stability of trans-precursor transcripts is therefore necessary to leave enough time for these molecules to find the precursor RNA containing their intron partner and re-assemble functional trans-introns. Cis-elements like stem loop structures present on the 3′ extremities of trans-precursor RNA may also trigger a similar protective activity in some instances, but this needs to be proven.

Further, our results revealed that MTSF2 is not only essential for *nad1* exons 2–3 precursor RNA stability but also for the stable accumulation of a 1.3 kb RNA species whose size corresponds exactly to the 5′-half of *nad1* intron 3 (Figure 6). This observation supports that released 5′-half intron molecules do accumulate stably in mitochondria after splicing as recently proposed (65) and proves that PPR proteins are responsible for such stable accumulations. This also indicates that the 5′-half of *nad1* intron 3 is not fused with its 3′-half intron partner upon splicing and therefore that the *nad1* intron 3 splicing mechanism occurs mostly through the hydrolytic pathway and not via the transesterification pathway which would have released an open lariat combining both intron halves, unlike previously proposed (66).

Upon binding to *nad1* exons 2–3 precursor, MTSF2 simultaneously defines the 3′ end of the first half of *nad1* intron 3

The high recombinogenic activity of plant mitochondrial genomes has resulted in various reported examples of gene fragmentation. The generated gene fragments are expressed from separate DNA regions and the resulting transcript or protein pieces must find their appropriate partners to reconstitute the entire mRNA or protein (6,7,9). The most frequent case of gene partitioning is associated with DNA breakage occurring in mitochondrial introns. Transcript reconstruction is permitted by trans-splicing reactions in which portions of introns re-assemble in a correct structure permitting the splicing reaction to occur and subsequently the ligation of exons. With the exception of the group I intron found in the *cytochrome oxidase 1* gene, all other mitochondrial introns in plants belong to the class II (6). Although they share a common evolutionary origin, group II introns have low sequence conservation but organize themselves into a phylogenetically conserved secondary structure comprising 6 helical domains that are distributed around a central hub (67). Throughout evolution, the sequence and structural changes of group II introns have been accompanied with the appearance of a large number of protein factors facilitating the splicing reaction (68–70). While the catalysis of splicing is very likely mediated by the introns themselves, there are several lines of evidence indicating that proteinaceous factors are required to help intron folding and to stabilize the structure in their catalytically active form (68,70–72). It is still unclear how the appropriate intron fragments functionally re-associate and the relative contributions of protein chaperones and base pairing in this process remain largely unknown. Factors like maturases, RNA helicases or CRM (chloroplast RNA splicing and ribosome maturation) proteins are required for the splicing of a large panel of mitochondrial introns in plants (68–70). Other classes of site-specific RNA binding proteins, among which PPR proteins are the most frequently represented, facilitate the splicing of one or two introns in most cases (18,59,66,68,73). There are five bi-partite trans-splicing introns in Arabidopsis mitochondria and they all contain disruptions that have occurred within their domain IV (74). Our analysis unambiguously revealed an essential role of MTSF2 in the stability of *nad1* exons 2–3 precursor RNA, but this protective activity is concomitantly associated with the definition of the 3′ end of the first half of *nad1* intron 3. A sequence comparison of DNA regions covering the 5′ half of *nad1* intron 3 in monocot and dicot species revealed that the 3′ extremity of *nad1* exons 2–3 transcript marks a clear breakpoint in DNA homology. Effectively, the DNA region located between *nad1* exon 3 and the MTSF2 binding site shows much higher sequence conservation than the one situated immediately downstream of *nad1* exons 2–3 pre-mRNA 3′ end (Supplementary Figure S5). Thus, not only the binding site of MTSF2 appears to be conserved in angiosperms but also its position relatively to the end of *nad1* exon 3 shows very little variation among seed plants. This observation suggests that the splitting of *nad1* intron 3 occurred likely once prior to the separation of monocot and dicot plants and that MTSF2 has been likely selected to allow this fragmentation event by stabilizing the newly-

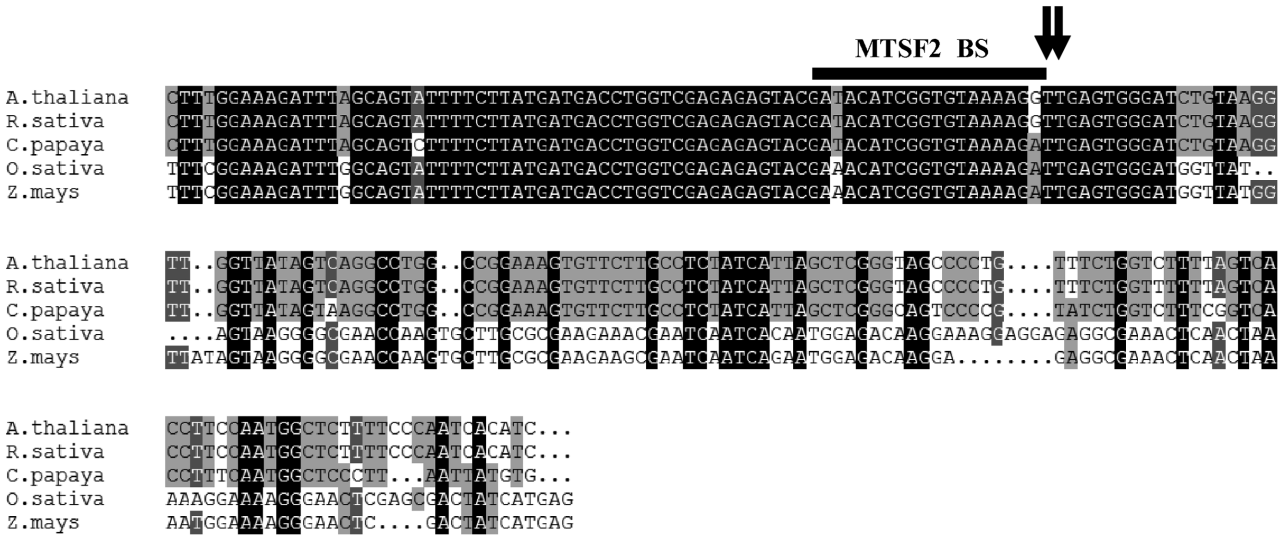

Figure 10. The MTSF2 binding site is conserved in angiosperms. Multiple sequence alignment of *nad1* exons 2–3 transcript 3′ end from a representative selection of dicot (*Arabidopsis thaliana*, *Rafanus sativa* and *Carina papaya*) and monocot (*Oryza sativa*, *Zea mays*) species. The sequence corresponding to the MTSF2 binding site is shown above the sequences and the downward arrows indicates the two 3′ ends of *nad1* exons 2–3 RNA mapped in *A. thaliana*.

created *nad1* exons 2–3 transcript and thereby permitting the trans-splicing of *nad1* intron 3. The evolutionary history of *nad1* intron 3 in land plants supports this hypothesis and suggests that the splitting of this intron occurred effectively early in the evolution of seed plants (75).

MSTF2 may play an active role in *nad1* intron 3 trans-splicing

The position of the MTSF2 binding site delimits the end of a conserved region comprising intron domains I–IV that are known to be essential for the splicing reaction to occur. By defining an extremity to *nad1* exons 2–3 precursor that is compatible with the progression of splicing, one can consider that MTSF2 plays an indirect role in splicing. However, MTSF2 may carry a more direct role in *nad1* intron 3 splicing. As previously proposed for the PPR4 protein in maize (76), MTSF2 may help the two halves of *nad1* intron 3 to recognize each other, either alone or through interactions with other potentially involved splicing factors. Alternatively and because its binding site lies within domain IV, MTSF2 may facilitate the splicing reaction of *nad1* intron 3 by stabilizing the structure of domains V or VI, which are part of the intron catalytic core and contain the bulged adenosine attacking the 5′ splice site junction respectively (77). However, the strong destabilization of *nad1* exons 2–3 transcript in *mtsf2* mutants makes the validation of a possible role of MTSF2 in *nad1* intron 3 intron splicing very difficult. The identification of MTSF2 protein partners or separation-of-function mutations in *MTSF2* could help to distinguish between the different roles that this protein may play in *nad1* exons 2–3 pre-mRNA processing. Nonetheless, a role of MTSF2 in intron splicing is suggested by the reduction of *nad2* intron 1 and intron 2 splicing in *mtsf2* mutants (Figure 6 and Supplementary Figure S3). Though, decreased splicing efficiency for these specific introns is frequently observed in independent Arabidopsis complex I mutants (45,52,59,78) and may just result from secondary effects to the loss of respiratory complex I.

Other trans-precursor mitochondrial transcripts are also likely stabilized by association of RNA binding proteins at their 3′ extremity

As indicated earlier, RNA binding proteins involved in transcript stabilization lead to the accumulation of short RNA footprints in mitochondria and plastids (47,60). These short sequences correspond to the binding sites of RNA stability trans-factors that are permanently protected against degradation during RNA turnover. In plant mitochondria, these short RNA footprints show typical characteristics comprising well defined 3′ ends and 5′ extremities spreading over several bases (61). A quite complete list of such footprints was recently established in Arabidopsis and mapped to the mitochondrial chromosome. We used this on-line available resource (https://www.molgen.hu-berlin.de/projects-jbrowse-athaliana.php) to identify RNA footprints in the five Arabidopsis 5′-half intron sequences to see whether any of the corresponding trans-precursor RNAs could be stabilized by association of a protective RNA binding protein at their 3′ end. Three of such RNA footprints were found to map to DNA regions located downstream of *nad1* exon1, *nad5* exons 1–2 and *nad5* exons 3 precursor RNAs (Supplementary Figure S4). The footprint identified downstream of *nad5* exon 3 precursor was mapped 57 bases downstream of exon 3, which is likely incompatible with the production of a functional 5′-half intron. However, the short RNAs that mapped downstream of *nad1* exon 1 and *nad5* exons 1–2 regions show accumulation levels that are similar to the short RNA mapping to the MTSF2 binding site (Supplementary Figure S4). They are 17 and 21 bp long and are located 1286 and 674 bases downstream of *nad1* exon1 and *nad5* exon 2 respectively. These distances would leave sufficient sequence space downstream of each

exon to organize intron domains I–IV, which spread over 500 bases for most group II introns (6). These observations strongly suggest that the stabilization of mitochondrial trans-precursor RNAs is not limited to *nad1* exons 2–3 pre-mRNA and MTSF2 and that *nad1* exon 1 and *nad5* exons 1–2 precursor transcripts are likely stabilized by the association of PPRs or related proteins to their 3′ extremity. Future investigations are required to validate or invalidate these assumptions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Institut National de la Recherche Agronomique [INRA UMR 1318]; China Scholarship Council (to C. W.); French Ministère de l'Enseignement et de la Recherche; Fondation pour la Recherche Médicale [#FDT20160736470 to N.P.]. Funding for open access charge: Institut Jean-Pierre Bourgin [INRA UMR1318].
Conflict of interest statement. None declared.

REFERENCES

1. Gray,M.W. (1999) Mitochondrial Evolution. *Science*, **283**, 1476–1481.
2. Lang,B.F., Gray,M.W. and Burger,G. (1999) Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.*, **33**, 351–397.
3. Burger,G., Gray,M.W. and Lang,B.F. (2003) Mitochondrial genomes: anything goes. *Trends Genet.*, **19**, 709–716.
4. Kubo,T. and Newton,K.J. (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion*, **8**, 5–14.
5. Gualberto,J.M., Mileshina,D., Wallet,C., Niazi,A.K., Weber-Lotfi,F. and Dietrich,A. (2014) The plant mitochondrial genome: dynamics and maintenance. *Biochimie*, **100**, 107–120.
6. Bonen,L. (2008) Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*, **8**, 26–34.
7. Rayapuram,N., Hagenmuller,J., Grienenberger,J.M., Bonnard,G. and Giegé,P. (2008) The three mitochondrial encoded CcmF proteins form a complex that interacts with CCMH and c-type apocytochromes in *Arabidopsis*. *J. Biol. Chem.*, **283**, 25200–25208.
8. Choi,B., Acero,M.M. and Bonen,L. (2012) Mapping of wheat mitochondrial mRNA termini and comparison with breakpoints in DNA homology among plants. *Plant Mol. Biol.*, **80**, 539–552.
9. Grewe,F., Edger,P.P., Keren,I., Sultan,L., Pires,J.C., Ostersetzer-Biran,O. and Mower,J.P. (2014) Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion*, **19**, 135–143.
10. Kim,B., Kim,K., Yang,T.-J. and Kim,S. (2016) Completion of the mitochondrial genome sequence of onion (*Allium cepa L.*) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the $ccmF_N$ gene split. *Curr. Genet.*, **62**, 873–885.
11. Lurin,C., Andrés,C., Aubourg,S., Bellaoui,M., Bitton,F., Bruyère,C., Caboche,M., Debast,C., Gualberto,J., Hoffmann,B. *et al.* (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089–2103.
12. Barkan,A. and Small,I. (2014) Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.*, **65**, 415–442.
13. O'Toole,N., Hattori,M., Andrés,C., Iida,K., Lurin,C., Schmitz-Linneweber,C., Sugita,M. and Small,I. (2008) On the expansion of the pentatricopeptide repeat gene family in plants. *Mol. Biol. Evol.*, **25**, 1120–1128.
14. Small,I.D. and Peeters,N. (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.*, **25**, 46–47.
15. Cheng,S., Gutmann,B., Zhong,X., Ye,Y., Fisher,M.F., Bai,F., Castleden,I., Song,Y., Song,B., Huang,J. *et al.* (2016) Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.*, **85**, 532–547.
16. Takenaka,M., Zehrmann,A., Verbitskiy,D., Härtel,B. and Brennicke,A. (2013) RNA editing in plants and its evolution. *Annu. Rev. Genet.*, **47**, 335–352.
17. Sun,T., Bentolila,S. and Hanson,M.R. (2016) The unexpected diversity of plant organelle RNA editosomes. *Trends Plant Sci.*, **21**, 962–973.
18. Hammani,K. and Giegé,P. (2014) RNA metabolism in plant mitochondria. *Trends Plant Sci.*, **19**, 380–389.
19. Ringel,R., Sologub,M., Morozov,Y.I., Litonin,D., Cramer,P. and Temiakov,D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*, **478**, 269–273.
20. Howard,M.J., Lim,W.H., Fierke,C.A. and Koutmos,M. (2012) Mitochondrial ribonuclease P structure provides insight into the evolution of catalytic strategies for precursor-tRNA 5′ processing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16149–16154.
21. Ke,J., Chen,R.-Z., Ban,T., Zhou,X.E., Gu,X., Tan,M.H.E., Chen,C., Kang,Y., Brunzelle,J.S., Zhu,J.-K. *et al.* (2013) Structural basis for RNA recognition by a dimeric PPR-protein complex. *Nat. Struct. Mol. Biol.*, **20**, 1377–1382.
22. Coquille,S., Filipovska,A., Chia,T., Rajappa,L., Lingford,J.P., Razif,M.F.M., Thore,S.E.P. and Rackham,O. (2014) An artificial PPR scaffold for programmable RNA recognition. *Nat. Commun.*, **5**, 1–9.
23. 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. *et al.* (2015) The design and structural characterization of a synthetic pentatricopeptide repeat protein. *Acta Crystallogr. D Biol. Crystallogr.*, **71**, 196–208.
24. Yin,P., Li,Q., Yan,C., Liu,Y., Liu,J., Yu,F., Wang,Z., Long,J., He,J., Wang,H.-W. *et al.* (2013) Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature*, **5**, 168–171.
25. Shen,C., Zhang,D., Guan,Z., Liu,Y., Yang,Z., Yang,Y., Wang,X., Wang,Q., Zhang,Q., Fan,S. *et al.* (2016) Structural basis for specific single-stranded RNA recognition by designer pentatricopeptide repeatproteins. *Nat. Commun.*, **7**, 1–8.
26. Fujii,S., Bond,C.S. and Small,I.D. (2011) Selection patterns on restorer-like genes reveal a conflict between nuclear and mitochondrial genomes throughout angiosperm evolution. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1723–1728.
27. Barkan,A., Rojas,M., Fujii,S., Yap,A., Chong,Y.S., Bond,C.S. and Small,I. (2012) A combinatorial amino acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet.*, **8**, e1002910.
28. Takenaka,M., Zehrmann,A., Brennicke,A. and Graichen,K. (2013) Improved computational target site prediction for pentatricopeptide repeat RNA editing factors. *PLoS One*, **8**, e65343.
29. Kindgren,P., Yap,A., Bond,C.S. and Small,I. (2015) Predictable alteration of sequence recognition by RNA editing factors from arabidopsis. *Plant Cell*, **27**, 403–416.
30. Yagi,Y., Hayashi,S., Kobayashi,K., Hirayama,T. and Nakamura,T. (2013) Elucidation of the RNA recognition code for pentatricopeptide repeat proteins involved in organelle RNA editing in plants. *PLoS One*, **8**, e57286.
31. Lupold,D.S., Caoile,A.G. and Stern,D.B. (1999) The maize mitochondrial *cox2* gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units. *J. Biol. Chem.*, **274**, 3897–3903.
32. Holec,S., Lange,H., Kühn,K., Alioua,M., Börner,T. and Gagliardi,D. (2006) Relaxed transcription in *Arabidopsis* mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol. Cell. Biol.*, **26**, 2869–2876.
33. Holec,S., Lange,H., Canaday,J. and Gagliardi,D. (2008) Coping with cryptic and defective transcripts in plant mitochondria. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1779**, 566–573.
34. Forner,J., Weber,B., Thuss,S., Wildum,S. and Binder,S. (2007) Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5′ and 3′ end formation. *Nucleic Acids Res.*, **35**, 3676–3692.

35. Forner,J., Hölzle,A., Jonietz,C., Thuss,S., Schwarzländer,M., Weber,B., Meyer,R.C. and Binder,S. (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis accessions. *Plant Physiol.*, **148**, 1106–1116.
36. Stoll,B., Stoll,K., Steinhilber,J., Jonietz,C. and Binder,S. (2013) Mitochondrial transcript length polymorphisms are a widespread phenomenon in *Arabidopsis thaliana*. *Plant Mol. Biol.*, **81**, 221–233.
37. Jonietz,C., Forner,J., Hölzle,A., Thuss,S. and Binder,S. (2010) RNA PROCESSING FACTOR2 is required for 5′ end processing of *nad9* and *cox3* mRNAs in mitochondria of *Arabidopsis thaliana*. *Plant Cell*, **22**, 443–453.
38. Hölzle,A., Jonietz,C., Törjek,O., Altmann,T., Binder,S. and Forner,J. (2011) A RESTORER OF FERTILITY-like PPR gene is required for 5′-end processing of the nad4 mRNA in mitochondria of *Arabidopsis thaliana*. *Plant J.*, **65**, 737–744.
39. Jonietz,C., Forner,J., Hildebrandt,T. and Binder,S. (2011) RNA PROCESSING FACTOR3 is crucial for the accumulation of mature *ccmC* transcripts in mitochondria of Arabidopsis accession Columbia. *Plant Physiol.*, **157**, 1430–1439.
40. Hauler,A., Jonietz,C., Stoll,B., Stoll,K., Braun,H.-P. and Binder,S. (2013) RNA PROCESSING FACTOR 5 is required for efficient 5′ cleavage at a processing site conserved in RNAs of three different mitochondrial genes in *Arabidopsis thaliana*. *Plant J.*, **74**, 593–604.
41. Stoll,B., Zendler,D. and Binder,S. (2014) RNA Processing Factor 7 and Polynucleotide Phosphorylase Are Necessary for Processing and Stability of *nad2* mRNA in Arabidopsis Mitochondria. *RNA Biol.*, **11**, 968–976.
42. Arnal,N., Quadrado,M., Simon,M. and Mireau,H. (2014) A restorer-of-fertility like pentatricopeptide repeat gene directs ribonucleolytic processing within the coding sequence of *rps3-rpl16* and *orf240a* mitochondrial transcripts in *Arabidopsis thaliana*. *Plant J.*, **78**, 134–145.
43. Stoll,K., Jonietz,C. and Binder,S. (2014) In *Arabidopsis thaliana* two co-adapted cyto-nuclear systems correlate with distinct *ccmC* transcript sizes. *Plant J.*, **81**, 247–257.
44. Stoll,B. and Binder,S. (2016) Two NYN domain containing putative nucleases are involved in transcript maturation in Arabidopsis mitochondria. *Plant J.*, **85**, 278–288.
45. Haili,N., Arnal,N., Quadrado,M., Amiar,S., Tcherkez,G., Dahan,J., Briozzo,P., Colas des Francs-Small,C., Vrielynck,N. and Mireau,H. (2013) The pentatricopeptide repeat MTSF1 protein stabilizes the *nad4* mRNA in Arabidopsis mitochondria. *Nucleic Acids Res.*, **41**, 6650–6663.
46. Barkan,A. (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.*, **155**, 1520–1532.
47. Zhelyazkova,P., Hammani,K., Rojas,M., Voelker,R., Vargas-Suarez,M., Borner,T. and Barkan,A. (2011) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Res.*, **40**, 3092–3105.
48. Hammani,K.K., Cook,W.B.W. and Barkan,A.A. (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 5651–5656.
49. Li,X., Jiang,D.H., Yong,K. and Zhang,B.D. (2007) Varied transcriptional efficiencies of multiple *Arabidopsis* U6 small nuclear RNA genes. *J. Integr. Plant Biol.*, **49**, 222–229.
50. Mali,P., Esvelt,K.M. and Church,G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat. Meth.*, **10**, 957–963.
51. Fauser,F., Schiml,S. and Puchta,H. (2014) Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.*, **79**, 348–359.
52. Haïli,N., Planchard,N., Arnal,N., Quadrado,M., Vrielynck,N., Dahan,J., des Francs-Small,C.C. and Mireau,H. (2016) The MTL1 pentatricopeptide repeat protein is required for both translation and splicing of the mitochondrial *NADH DEHYDROGENASE SUBUNIT7* mRNA in *Arabidopsis*. *Plant Physiol.*, **170**, 354–366.
53. Dahan,J., Tcherkez,G., Macherel,D., Benamar,A., Belcram,K., Quadrado,M., Arnal,N. and Mireau,H. (2014) Disruption of the *CYTOCHROME C OXIDASE DEFICIENT1* gene leads to cytochrome c oxidase depletion and reorchestrated respiratory metabolism in *Arabidopsis*. *Plant Physiol.*, **166**, 1788–1802.
54. Carrie,C., Giraud,E., Duncan,O., Xu,L., Wang,Y., Huang,S., Clifton,R., Murcha,M., Filipovska,A., Rackham,O. *et al.* (2010) Conserved and Novel Functions for Arabidopsis thaliana MIA40 in Assembly of Proteins in Mitochondria and Peroxisomes. *J. Biol. Chem.*, **285**, 36138–36148.
55. Lamattina,L., Gonzalez,D., Gualberto,J. and Grienenberger,J.M. (1993) Higher plant mitochondria encode an homologueof the nuclear-encoded 30-kDa subunit of bovine mitochondrial complex I . *Eur. J. Biochem.*, **217**, 831–838.
56. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
57. Sung,T.-Y., Tseng,C.-C. and Hsieh,M.-H. (2010) The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. *Plant J.*, **63**, 499–511.
58. Zhu,Q., Dugardeyn,J., Zhang,C., Takenaka,M., Kühn,K., Craddock,C., Smalle,J., Karampelias,M., Denecke,J., Peters,J. *et al.* (2012) SLO2, a mitochondrial PPR protein affecting several RNA editing sites, is required for energy metabolism. *Plant J.*, **71**, 836–849.
59. Colas des Francs-Small,C., Falcon de Longevialle,A., Li,Y., Lowe,E., Tanz,S.K., Smith,C., Bevan,M.W. and Small,I. (2014) The pentatricopeptide repeat proteins TANG2 and ORGANELLE TRANSCRIPT PROCESSING 439 are involved in the splicing of the multipartite *nad5* transcript encoding a subunit of mitochondrial complex I. *Plant Physiol.*, **165**, 1409–1416.
60. Ruwe,H. and Schmitz-Linneweber,C. (2011) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? *Nucleic Acids Res.*, **40**, 3106–3116.
61. Ruwe,H., Wang,G., Gusewski,S. and Schmitz-Linneweber,C. (2016) Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Res.*, **44**, 7406–7417.
62. Hattori,M. and Sugita,M. (2009) A moss pentatricopeptide repeat protein binds to the 3′ end of plastid *clpP* pre-mRNA and assists with mRNA maturation. *FEBS J.*, **276**, 5860–5869.
63. Pfalz,J., Bayraktar,O.A., Prikryl,J. and Barkan,A. (2009) Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. *EMBO J.*, **28**, 2042–5211.
64. Hammani,K., Takenaka,M., Miranda,R. and Barkan,A. (2016) A PPR protein in the PLS subfamily stabilizes the 5′-end of processed *rpl16* mRNAs in maize chloroplasts. *Nucleic Acids Res.*, **44**, 4278–4288
65. Massel,K., Silke,J.R. and Bonen,L. (2016) Multiple splicing pathways of group II trans-splicing introns in wheat mitochondria. *Mitochondrion*, **28**, 23–32.
66. Gualberto,J.M., Le Ret,M., Beator,B. and Kühn,K. (2015) The RAD52-like protein ODB1 is required for the efficient excision of two mitochondrial introns spliced via first-step hydrolysis. *Nucleic Acids Res.*, **43**, 6500–6510.
67. Pyle,A.M., Fedorova,O. and Waldsich,C. (2007) Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.*, **32**, 138–145.
68. Brown,G.G., Colas des Francs-Small,C. and Ostersetzer-Biran,O. (2014) Group II intron splicing factors in plant mitochondria. *Front. Plant Sci*, **5**, 35.
69. Schmitz-Linneweber,C., Lampe,M.-K., Sultan,L.D. and Ostersetzer-Biran,O. (2015) Organellar maturases: A window into the evolution of the spliceosome. *Biochim. Biophys. Acta Bioenerg.*, **1847**, 798–808.
70. Sultan,L.D., Mileshina,D., Grewe,F., Rolle,K., Abudraham,S., Głodowicz,P., Niazi,A.K., Keren,I., Shevtsov,S., Klipcan,L. *et al.* (2016) The reverse transcriptase/RNA maturase protein MatR is required for the splicing of various group II introns in brassicaceae mitochondria. *Plant Cell*, **28**, 2805–2829.
71. Noah,J.W. and Lambowitz,A.M. (2003) Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*, **42**, 12466–12480.
72. Qu,G., Kaushal,P.S., Wang,J., Shigematsu,H., Piazza,C.L., Agrawal,R.K., Belfort,M. and Wang,H.-W. (2016) Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.*, **23**, 549–557.
73. Hsu,Y.-W., Wang,H.-J., Hsieh,M.-H., Hsieh,H.-L. and Jauh,G.-Y. (2014) Arabidopsis mTERF15 is required for mitochondrial *nad2*

intron 3 splicing and functional complex I activity. *PLoS One*, **9**, e112360.
74. Glanz,S. and Kück,U. (2009) Trans-splicing of organelle introns—a detour to continuous RNAs. *Bioessays*, **31**, 921–934.
75. Malek,O. and Knoop,V. (1998) Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA*, **4**, 1599–1609.
76. Schmitz-Linneweber,C., Williams-Carrier,R.E., Williams-Voelker,P.M., Kroeger,T.S., Vichas,A. and Barkan,A. (2006) A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell*, **18**, 2650–2663.
77. Bonen,L. and Vogel,J. (2001) The ins and outs of group II introns. *Trends Genet.*, **17**, 322–331.
78. Koprivova,A., des Francs-Small,C.C., Calder,G., Mugford,S.T., Tanz,S., Lee,B.-R., Zechmann,B., Small,I. and Kopriva,S. (2010) Identification of a pentatricopeptide repeat protein implicated in splicing of intron 1 of mitochondrial *nad7* transcripts. *J. Biol. Chem.*, **285**, 32192–32199.