A Case of Gene Fragmentation in Plant Mitochondria Fixed by the Selection of a Compensatory Restorer of Fertility-Like PPR Gene

Tan-Trung Nguyen,[†,1] Noelya Planchard,[†,1,2] Jennifer Dahan,[†,‡,1] Nadège Arnal,[§,1] Sandrine Balzergue,[3] Abdelilah Benamar,[3] Pierre Bertin,[4] Véronique Brunaud,[5] Céline Dargel-Graffin,[1] David Macherel,[3] Marie-Laure Martin-Magniette,[5] Martine Quadrado,[1] Olivier Namy,[4] and Hakim Mireau *,[1]

[1]Institut Jean-Pierre Bourgin (IJPB), INRAE, AgroParisTech, Université Paris-Saclay, Versailles, France

[2]Paris-Sud University, Université Paris-Saclay, Orsay, France

[3]Unité Mixte de Recherche 1345, Institut de Recherche en Horticulture et Semences, Université d'Angers, Angers, France

[4]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Saclay, Gif-sur-Yvette, France

[5]Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, University of Evry, Orsay, France

[†]These authors contributed equally to this work as first authors.

[‡]Present address: Department of Entomology, Plant Pathology and Nematology, University of Idaho, Moscow, ID, USA

[§]Present address: Centre National de Ressources Génomiques Végétales, INRAE, Auzeville-Tolosane, France

***Corresponding author:** E-mail: hakim.mireau@inrae.fr.

Associate editor: Victoria Sork

Abstract

The high mutational load of mitochondrial genomes combined with their uniparental inheritance and high polyploidy favors the maintenance of deleterious mutations within populations. How cells compose and adapt to the accumulation of disadvantageous mitochondrial alleles remains unclear. Most harmful changes are likely corrected by purifying selection, however, the intimate collaboration between mitochondria- and nuclear-encoded gene products offers theoretical potential for compensatory adaptive changes. In plants, cytoplasmic male sterilities are known examples of nucleo-mitochondrial coadaptation situations in which nuclear-encoded restorer of fertility (*Rf*) genes evolve to counteract the effect of mitochondria-encoded cytoplasmic male sterility (CMS) genes and restore fertility. Most cloned *Rfs* belong to a small monophyletic group, comprising 26 pentatricopeptide repeat genes in Arabidopsis, called *Rf*-like (RFL). In this analysis, we explored the functional diversity of *RFL* genes in Arabidopsis and found that the *RFL8* gene is not related to CMS suppression but essential for plant embryo development. In vitro-rescued *rfl8* plantlets are deficient in the production of the mitochondrial heme–lyase complex. A complete ensemble of molecular and genetic analyses allowed us to demonstrate that the *RFL8* gene has been selected to permit the translation of the mitochondrial $ccmF_{N2}$ gene encoding a heme–lyase complex subunit which derives from the split of the $ccmF_N$ gene, specifically in *Brassicaceae* plants. This study represents thus a clear case of nuclear compensation to a lineage-specific mitochondrial genomic rearrangement in plants and demonstrates that *RFL* genes can be selected in response to other mitochondrial deviancies than CMS suppression.

Key words: **mitochondria, mitochondrial translation, Rf-like PPR proteins, c-type cytochrome maturation, plant respiratory mutant.**

Article

Introduction

The functioning of eukaryotic cells relies on the tight collaboration between the nucleus and one or more genome-containing cytoplasmic organelles. Among these, mitochondria (mt), which produce most of the energy of eukaryotic cells through respiration, derive from an ancient endosymbiotic event that involved an α-proteobacteria (Roger et al. 2017). Although mt of all eukaryotes have a unique origin, modern mt display a huge diversity in genome organization, size, and gene expression processes (Burger et al. 2003; Neupert 2016). For instance, mitochondrial genomes of photosynthetic organisms span over hundreds of kilobases (Sloan, Alverson, et al. 2012; Gualberto and Newton 2017), whereas those in animals are extremely minimalist (14–20 kb) and comprise virtually no intergenic regions (Boore 1999). Nevertheless, the coding capacity of mitochondrial genomes is similar and very limited across eukaryotes with often less than 50 genes (Sloan et al. 2018). Subsequently, mt functioning relies on the import of an extensive number of nuclear-encoded proteins from the cytoplasm which tightly

collaborate with mt-encoded gene products to orchestrate respiratory metabolism and mitochondrial gene expression. Large differences concerning the evolutionary pattern of mitochondrial DNA are found among eukaryotes. The size expansion of seed plant mitochondrial genomes results mostly from the acquisition of large intergenic sequences of unknown origin (Mower et al. 2012; Yurina and Odintsova 2016) or acquired by horizontal gene transfer (Bergthorsson et al. 2003). Compared with animals, synonymous substitution rates in plant mitochondria are generally extremely low (Wolfe et al. 1987; Drouin et al. 2008; Richardson et al. 2013). Indeed, flowering plant mitochondrial genomes evolve mostly by genome rearrangements resulting in highly variable organizations, even between closely related individuals (Palmer and Herbon 1988; Allen et al. 2007; Darracq et al. 2011; Sloan, Alverson, et al. 2012; Cole et al. 2018). The structural variability of plant mitochondrial DNAs is due to the presence of numerous recombinationally active DNA repeats. Large repeats (>500 bp) produce circular and linear mitochondrial DNA subgenomic forms by homologous recombination, whereas intermediate (50–500 bp) and small (<50 bp) repeats promote ectopic homologous or microhomology-based recombinations, respectively (Arrieta-Montiel et al. 2009; Gualberto and Newton 2017). Although highly regulated, elevated homologous recombination activity and break-induced DNA replication thus render plant mitogenomes more prone to sequence rearrangements than in most other eukaryotes. Stoichiometric shifting and selective amplification of subgenomic products are also major driving forces leading to the emergence of alternative mitotypes in plants (Gualberto and Newton 2017; Havird et al. 2019). Most sequence rearrangements in plant mitochondrial genomes occur outside of coding regions and result in mitochondrial gene shuffling with no phenotypical consequence, although few changes affecting gene function have been described (Kubo and Newton 2008). Certain rearrangements can also produce new genes like cytoplasmic male sterility (CMS) ORFs which are often made of pieces of essential mitochondrial genes and impede normal pollen production (Chen and Liu 2014). The high mutational load of mitochondrial genomes combined with their uniparental mode of inheritance are likely key conditions leading to rapid evolution of mitochondrial DNA (Sloan et al. 2017). The polyploid nature of mt is also thought to favor the propagation of feebly deleterious mitochondrial mutations, which can be maintained more or less silently at low copy numbers in cells leading to heteroplasmic cells. How cells compose and subsequently adapt to the accumulation of deleterious mitochondrial alleles is still largely unclear. Most harmful mitochondrial genome changes are likely rapidly corrected by purifying selection involving template-based DNA repair or gene conversion that are known to be prominent in plant mt (Castellana et al. 2011; Davila et al. 2011; Sloan et al. 2018). However, selection appears less efficient to eliminate weakly deleterious mutations from mitochondrial genome, which frequently cohabit with wild-type alleles in cells (Kubo and Newton 2008; Payne et al. 2013). It has been proposed that the intimate collaboration between mitochondria- and nuclear-encoded gene products offers also great potential for compensatory adaptive changes in nuclear genes encoding mt-targeted proteins in response to mitochondrial sequence mutations (Sloan et al. 2017). This mitonuclear compensatory coevolution model finds supports in growing number of correlative studies but most of them remain to be rigorously tested (Osada and Akashi 2012; Sloan et al. 2018). In flowering plants, nucleo-cytoplasmic male sterilities are well known examples of nucleo-mitochondrial coadaptation situations where nuclear-encoded restorer of fertility (*Rf*) genes evolve to counteract the effect or the expression of mitochondrial CMS genes and thus restore male fertility (Chen and Liu 2014). Most cloned *Rf* belong to a small monophyletic group of pentatricopeptide repeat (PPR) genes—called *Rf*-like (RFL)—exhibiting extensive diversifying selection (Geddy and Brown 2007; Fujii et al. 2011; Dahan and Mireau 2013). In *Arabidopsis thaliana*, this PPR subgroup comprises 26 members and none seems involved in CMS suppression (Arnal et al. 2014). In this analysis, we explored the functional diversity of *RFL* genes in Arabidopsis and surprisingly found that the *RFL8* gene is essential for mitochondrial activity. Our results show that this gene was selected to permit the expression of the $ccmF_{N2}$ mitochondrial gene which corresponds to the downstream split product of the $ccmF_N$ gene in the *Brassicaceae* plant family (Rayapuram et al. 2008). We show that the role of RFL8 is to permit the translation of $ccmF_{N2}$ by setting the translational initiation site. This study represents thus a potential case of nuclear compensation to a recent deleterious mitochondrial genomic rearrangement in plants and shows that *RFL* genes can be selected for other mitochondrial deviancies than just the suppression of CMS gene expression.

Results

The Arabidopsis *RFL8* PPR Gene Is Essential for Embryo Development

To get insight into the functional diversity of *RFL* genes, Arabidopsis T-DNA mutants affected in each gene of this PPR subfamily were genetically and phenotypically characterized. Most of the mutants showed no phenotypic alterations, except for the one affecting the *RFL8* gene and for which no homozygous mutant individuals could be identified in the progeny of hemizygous *rfl8* plants. The T-DNA in this line was inserted 1138 bases downstream of *RFL8* start codon, just upstream of the 11th PPR repeat coding sequence (fig. 1*A*). Young siliques of *rfl8* hemizygous plants displayed one-quarter of white to yellow seeds harboring embryos arrested at the early-bent cotyledon stage (fig. 1*B* and *C*). A similar embryo-lethality was previously observed in a second mutant allele of *RFL8*, corroborating our observations (Yang et al. 2011). When rescued by tissue culture (Dahan et al. 2014), the embryos contained in immature discolored seeds produced small and bushy plantlets that were found homozygous for the *SALK_0015489* T-DNA insertion. They grew extremely slowly as compared with the wild type and developed short leaves with limited stems and a few flowers but no roots after 3–4 months of culture (fig. 1*D* and *E*). They could, however,

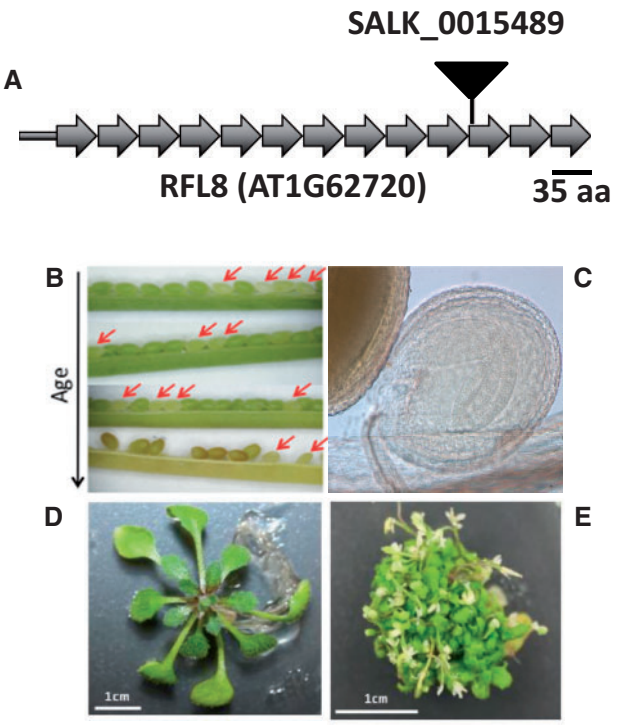

Fig. 1. The Arabidopsis *rfl8* mutant is embryonic lethal but can be rescued by in vitro culture of immature embryos. (*A*) Schematic representation of the RFL8 PPR protein and its 13 P-type PPR repeats (shown as dark gray arrows). The position of the SALK_0015489 T-DNA insertion with respect to the PPR motif organization of RFL8 is indicated by an inverted black triangle. (*B*) Arabidopsis seeds in siliques of *rfl8* heterozygous plants showing discolored seeds (red arrows) containing homozygous *rfl8* mutant embryos. (*C*) Nomarski photograph of an *rfl8* mutant embryo arrested at the bent cotyledon stage. (*D*) Wild-type Col-0 Arabidopsis plant grown for 5 weeks on embryo-rescue culture medium after germination. (*E*) *rfl8* mutant plantlet grown for 22 weeks after germination on embryo-rescue culture medium after germination.

be vegetatively propagated to produce sufficient mutant material for molecular characterization. Homozygous mutant plants with a phenotype indistinguishable from the wild type could be generated upon expression of an ectopic copy of *RFL8* fused in frame with the triple hemagglutinin epitope tag (3HA) coding region, supporting that the observed embryonic lethality resulted from the inactivation of *RFL8* (supplementary fig. 1*A*, Supplementary Material online). Subcellular distribution analysis confirmed the mitochondrial localization of the RFL8 protein (supplementary fig. 1*B*, Supplementary Material online), corroborating targeting predictions for this protein (Dahan and Mireau 2013).

rfl8 Plants Do Not Produce Mitochondrial c-Type Cytochromes

The mitochondrial localization of RFL8 suggested that the growth defects of *rfl8* plants may result from an improper functioning of the respiratory chain. To identify the origin of such deficiency, the steady-state levels of respiratory chain complexes were first visualized on blue-native gels. Because *rfl8* plants developed as miniature plants, crude membrane extracts were used in this approach. In-gel activity and western blot results revealed a significant decrease in complex I accumulation and barely detectable levels of complexes III and IV in *rfl8* plants (fig. 2*A* and supplementary fig. 2, Supplementary Material online). Further analysis of respiratory chain subunits revealed near-normal accumulation of all tested complex I subunits (Nad3, Nad6, and Nad9), whereas complex III (RISP and Cob) and complex IV (Cox2) proteins were undetectable in *rfl8* plants (fig. 2*B*). The lack of detectable complexes III and IV in *rfl8* plants suggested strong alterations of respiration, which was comparatively measured in freshly harvested wild-type and *rfl8* seedlings using a Clark's electrode in the dark. Although the oxygen consumption rates of wild-type and *rfl8* explants were virtually identical in the absence of respiratory inhibitors, their impact on oxygen uptake differed drastically. In contrast to the wild type, complex III and complex IV inhibitors impacted very moderately oxygen consumption in *rfl8* plants, suggesting that the cytochrome respiratory pathway contributes very moderately to dark oxygen uptake in these plants (supplementary fig. 3*A*, Supplementary Material online). These observations correlated with the highly reduced levels of complex III and complex IV in *rfl8* plants, as observed in BN–PAGE analysis (fig. 2*A* and supplementary fig. 2, Supplementary Material online). The contribution of the cyanide-insensitive alternative pathway in *rfl8* respiration was then investigated by adding the inhibitor of alternative oxidase (AOX), n-propylgallate (nPG), to the reaction medium. The addition of nPG caused a very significant drop in oxygen consumption in *rfl8* (supplementary fig. 3*A*, Supplementary Material online), confirming that respiratory activity in these plants is for the most part supported by the alternative respiratory pathway. The strong overaccumulation of AOX and alternative NADH dehydrogenase (*NDA* and *NDB*) transcripts and AOX protein in *rfl8* plants strongly supported these observations (fig. 2*B* and supplementary fig. 3*B*, Supplementary Material online). Complex III and complex IV being both involved in cytochrome c (CYT_C) oxydo-reduction, we next wondered if the concomitant destabilization of these two complexes in *rfl8* plants could result from alterations in the production of c-type cytochromes. Western-blot analyses confirmed these assumptions as no trace of CYT_C and CYT_{C1} could be detected in *rfl8* mitochondrial preparations (fig. 2*B*). C-type cytochromes are hemoproteins that are mitochondrially imported from the cytosol and then matured in the intermembrane space (IMS) to become active. This maturation process involves at least two functional modules consisting in a heme-handling complex responsible for heme transport across the inner mitochondrial membrane and delivery, and a heme–ligation complex in charge of maintaining apocytochromes in a reduced state and linking heme to apocytochromes (Verissimo and Daldal 2014). These two functional modules can be separated as two multiprotein complexes on BN–PAGE gels with apparent molecular weights close to 500 kDa each in Arabidopsis (Meyer et al. 2005; Rayapuram et al. 2007). Whereas the heme-handling complex could be detected in *rfl8* plant extracts with a size and abundance equivalent to the wild type, no detectable traces of heme–lyase complex appeared to accumulate in the *rfl8* mutant (fig. 2*C*). Interestingly, the heme lyase subunit $CcmF_{N2}$

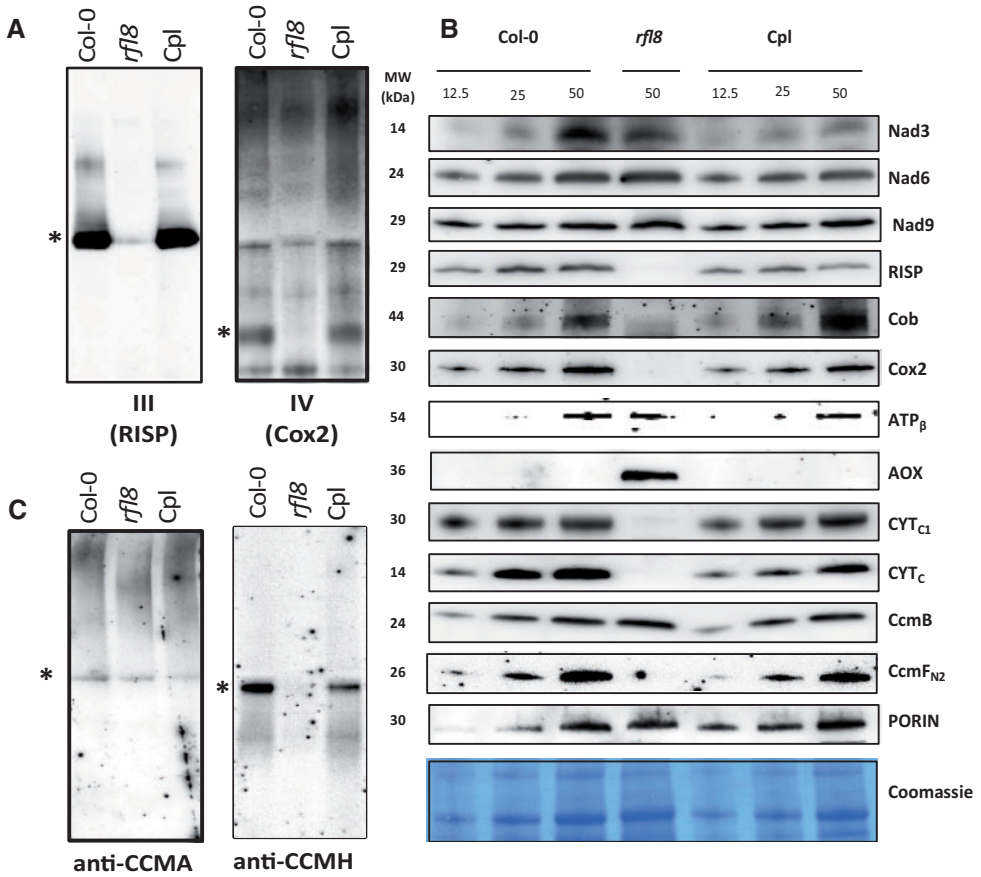

Fig. 2. *rfl8* plants contain dramatically reduced levels of complexes III, IV, and cytochrome c/c1 heme lyase. (*A*) Immunoblots of BN–PAGE gels probed with antibodies against the Rieske iron–sulfur (RISP, a subunit of complex III) and Cox2 (a subunit of complex IV) proteins. About 100 µg of crude mitochondrial extracts prepared from wild-type (Col-0), *rfl8*, and complemented *rfl8* plants (Cpl) were used in the analysis. * shows bands corresponding to the indicated respiratory complexes. (*B*) Steady-state level analysis of various mitochondrial proteins in wild-type, *rfl8*, and functionally complemented *rfl8* plants. About 12.5, 25, or 50 µg of proteins from crude mitochondrial preparations were loaded in each lane and probed with antibodies specific to the indicated mitochondrial proteins (right). PORIN was used as loading control to verify equal loading across samples. Molecular weight (MW) of detected proteins is indicated (left). (*C*) Same analysis as shown in (*A*) except that antibodies to CCMA and CCMH were used to detect the heme-handling and heme–lyase complexes, respectively.

protein could not be also detected in *rfl8* plants by immunoblot analysis (fig. 2*B*).

RFL8 Is Required for $ccmF_{N2}$ Translation and Possibly *nad2* Intron 1 Splicing

The essential roles of PPR proteins in organellar mRNA expression (Barkan and Small 2014) led us to search for mitochondrial mRNA processing defects in *rfl8* plants. The steady-state levels of all mitochondria-encoded mRNAs and pre-mRNAs were evaluated by RT–qPCR (supplementary fig. 4, Supplementary Material online). This approach revealed a global overaccumulation of virtually all mitochondrial mRNAs in *rfl8* that was also readily visible in RNA gel blot experiments (supplementary fig. 5, Supplementary Material online). This was the case for all mitochondria-encoded complexes III, IV, and cytochrome c maturation (CCM) mRNAs whose mature forms accumulated at the same size as in the wild type. RT–qPCR results allowed us to calculate the splicing efficiency for all mitochondrial introns and the most significant detected decrease concerned the first intron of *nad2* with a 10-fold reduction in splicing efficiency (supplementary fig. 6, Supplementary Material online) that was confirmed by RNA-gel blot analysis (supplementary fig. 7, Supplementary Material online). Since none of these observations could explain the loss of c-type cytochrome production in *rfl8*, we next used the Ribo-seq approach to measure the translation efficiency of all mitochondria-encoded mRNAs in both wild-type and mutant plants. Translational efficiencies (TE) were determined as the relative number of ribosome footprints along mitochondrial mRNAs normalized by their respective abundance in each genotype, as detailed in Planchard et al. (2018). Comparative TE results showed that among transcripts corresponding to complexes that were barely detectable in *rfl8* (e.g., cIII, cIV, and the CCM heme–lyase complex) only $ccmF_{N2}$ showed a significant reduction in ribosome loading (fig. 3*A*). Most other implicated transcripts like *cob*, *cox2*, or *cox3* appeared indeed to be more translated in *rfl8* compared with wild-type plants. Several other mRNAs like *matR*, *nad7*, *rps3*, or *rpl16* were also less translated in *rfl8* but the decrease in TE was slightly weaker compared with $ccmF_{N2}$ (fig. 3*A*). Moreover, the ribosome footprint distributions along these undertranslated mRNAs revealed no major

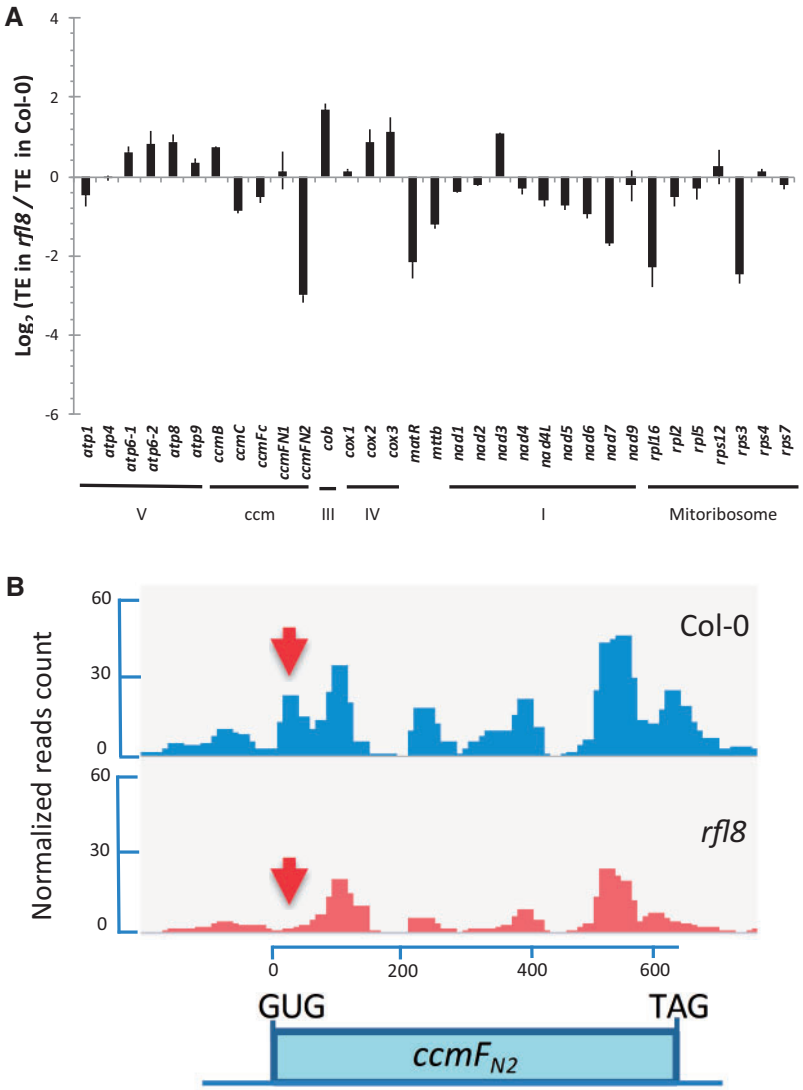

Fig. 3. Translation of the mitochondrial $ccmF_{N2}$ transcript is reduced in *rfl8* plants. (*A*) Ribo-Seq analysis of mitochondrial mRNAs in *rfl8* plants compared with the wild type. The bars depict $\log_2$ ratios of ribosome footprint abundance for each mitochondrial mRNA normalized to both mRNA length and abundance in *rfl8* plants relative to the wild type (Col-0). Values between genotypes were also normalized to the numbers of Ribo-Seq reads mapping to mitochondrial ORFs. The reported values are means of three independent biological replicates (error bars indicate SD). (*B*) Screenshots from the Integrated Genome Viewer software showing the distributions of ribosome footprints along the $ccmF_{N2}$ transcript in both wild-type (Col-0) and the *rfl8* mutant. The distributions were normalized to the number of reads mapping to the mitochondrial genome. The red arrow points to a ribosome footprint peak near the proposed translation start site of $ccmF_{N2}$ that is visible in the wild type (Col-0) but undetectable in the *rfl8* mutant.

difference with the wild type, except for $ccmF_{N2}$ for which a ribosome peak in the vicinity of its proposed GUG start codon (Rayapuram et al. 2008) is visible in the control but absent in the mutant (fig. 3*B* and supplementary fig. 8, Supplementary Material online). Altogether, these results indicated that the translation of $ccmF_{N2}$ is reduced in *rfl8* plants and that this perturbation likely explained the loss of the cytochrome heme–lyase complex in the mutant.

RFL8 Binds Downstream of the 5′ Region of $ccmF_{N2}$ Transcript

To further clarify the role of RFL8 in mitochondrial mRNA expression, its in vivo RNA targets were identified using two complementary approaches. $ccmF_{N2}$ being a first likely candidate, we sought to specify the RFL8 binding site within this transcript by taking advantage of the dual observation showing that RFL8 appears functionally conserved in *Brassicaceae* (fig. 4*A* and *B*; supplementary fig. 9, Supplementary Material online) but that the $ccmF_{N2}$ gene harbors unrelated 5′ noncoding regions in *A. thaliana* and most other *Brassicaceae* plants such as *Brassica rapa* (fig. 4*C* and supplementary fig. 10, Supplementary Material online). We therefore tested the ability of the *B. rapa RFL8* to complement the Arabidopsis *rfl8* mutant and used it as a genetic means to delineate the RFL8 binding zone within the $ccmF_{N2}$ transcript. The *B. rapa RFL8* gene was thus transformed in Arabidopsis *rfl8* heterozygous

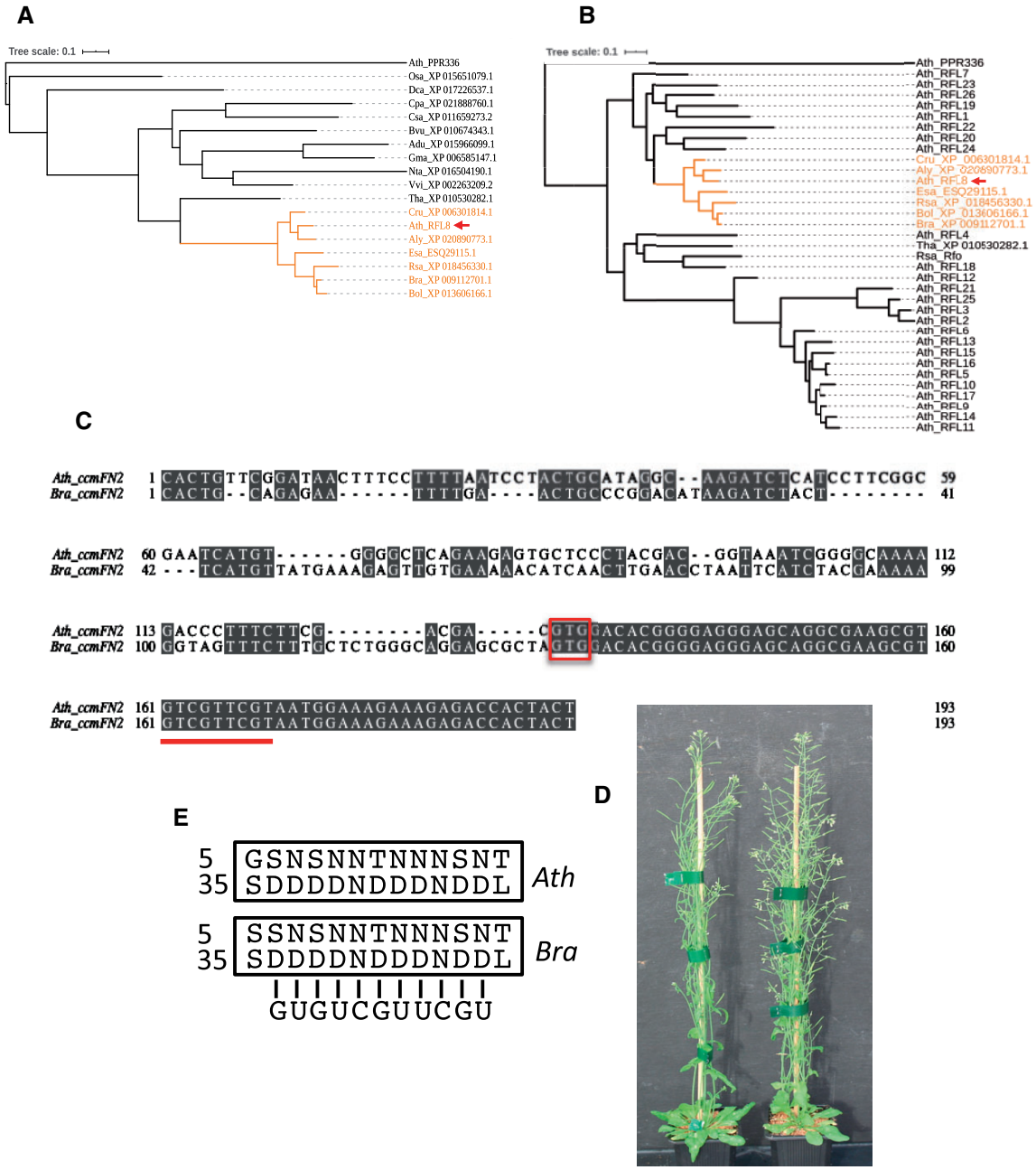

Fig. 4. The *Brassica rapa RFL8* homolog can functionally complement the Arabidopsis *rfl8* mutant. (*A*) Along with $ccmF_{N2}$, *Brassicaceae* plants all encode a very close homolog to the RFL8 protein. Phylogenetic relationship between the closest RFL8 homologs identified in a representative panel of *Brassicaceae* (orange) and non-*Brassicaceae* (black) angiosperm plants (see details below). The Arabidopsis PPR protein PPR336 was chosen as outgroup. (*B*) *Brassicaceae* closest RFL8 homologs are more closely related to RFL8 than to other Arabidopsis Rf-like (RFL) proteins. Phylogenetic relationship between *Brassicaceae* RFL8 homologs and all other *Arabidopsis thaliana* RFL proteins. The red arrow points to the Arabidopsis RFL8 protein and the RFL8 clade is shown in orange. Sequence alignments were done with T-coffee and the tree constructed with iTOL. (*C*) Sequence alignment of a part of $ccmF_{N2}$ 5′ region from *Arabidopsis thaliana* and *Brassica rapa*. The proposed $ccmF_{N2}$ translational start codon (GTG) is boxed in red and the predicted RFL8 binding site is underlined in red. A multiple sequence alignment of 1 kb of DNA sequence upstream of the GTG codon from a representative panel of *Brassicaceae* plants is shown in supplementary figure S10, Supplementary Material online. (*D*) Comparative growth of a wild-type (Col-0) plant and a homozygous Arabidopsis *rfl8* mutant expressing the *B. rapa RFL8* homolog, 8 weeks of culture after sowing. (*E*) Prediction of the Arabidopsis and *Brassica rapa* RFL8 RNA binding sites. The amino acids at positions 5 and 35 of each RFL8 PPR repeat are shown from N- to C-terminus. The obtained amino acid combinations were used to calculate the probabilities of nucleotide recognition by each individual PPR repeat according to the PPR code (Barkan et al. 2012) and the most likely target sequence identified in the mitochondrial genome of Arabidopsis and *B. rapa* is indicated. This sequence is found in the 5′ region of $ccmF_{N2}$ (see panel *C*). *Ath*, *Arabidopsis thaliana*; *Osa*, *Oryza sativa*; *Dca*, *Daucus carota*; *Csa*, *Cucumis sativus*; *Bvu*, *Beta vulgaris*; *Vvi*, *Vitis vinifera*; *Nta*, *Nicotiana tabacum*; *Gma*, *Glycine max*; *Adu*, *Arachis duranensis*; *Tha*, *Tarenaya hassleriana*; *Cru*, *Capsela rubella*; *Aly*, *Arabidopsis lyrata*; *Esa*, *Eutrema salsugineum*; *Rsa*, *Raphanus sativus*; *Bol*, *Brassica oleraceae*; *Bra*, *Brassica rapa*.

plants and homozygous *rfl8* plants with a recovered wild-type phenotype could be identified in the progenies of most primary transformants (fig. 4*D*). This ability of the *B. rapa RFL8* gene to induce the translation of Arabidopsis $ccmF_{N2}$ suggested that the binding site of RFL8 resided in a $ccmF_{N2}$ region that is identical in both Arabidopsis and *B. rapa* (supplementary fig. 11, Supplementary Material online), namely the sequence located downstream of the theoretical start codon (GTG) of $ccmF_{N2}$ (Rayapuram et al. 2008). Interestingly, an 11 base fragment situated 25 bases downstream of this GTG codon is predicted by the PPR code (Barkan et al. 2012) to be the most likely binding site of RFL8 (fig. 4*E* and supplementary fig. 12, Supplementary Material online). The search for RFL8 in vivo RNA targets was next determined by RNA-immunoprecipitation and sequencing assay (RIP-seq). Immunoprecipitations were conducted with the HA antibody from mitochondrial extracts extracted from wild-type and *RFL8-3HA* complemented plants (supplementary fig. 13, Supplementary Material online) and coenriched RNAs species were deep sequenced after reverse transcription. After mapping, enrichment ratios were calculated for all mitochondrial genes and a statistically significant coenrichment was only observed for the $ccmF_{N2}$ transcript (fig. 5*A*). Enrichment ratios were also calculated for all mitochondrial introns, but no coassociation with RFL8 was detected for any of these regions (supplementary fig. 14, Supplementary Material online) which did not support a direct role of RFL8 in intron splicing. To determine in detail which parts of $ccmF_{N2}$ transcript associate with RFL8 in vivo, RIP-Seq assays were redone by including an RNAse I treatment prior to the immunoprecipitation step, thereby limiting coenriched RNA species to the region physically covered by RFL8 and protected from RNAse I digestion. Subsequent mapping results revealed a single protected zone within the 5′ region of $ccmF_{N2}$ mRNA and highly enriched in RFL8 immunoprecipitation (fig. 5*B*). This stretch of 40 bases starts 18 bases downstream of the GTG codon mentioned above and coincides with the predicted binding site of RFL8 (fig. 5*B* and supplementary fig. 15, Supplementary Material online). Binding assays further demonstrated the capacity of RFL8 to specifically associate with this 11-base predicted binding site in vitro (fig. 5*C*), confirming that it corresponded to the binding site of RFL8 within $ccmF_{N2}$.

Discussion

Rf-Like Genes as Compensatory Responses to Various Mitochondrial Genetic Deviancies

Genetic incompatibilities and disruption of gene flows between populations are major drivers of speciation (Presgraves 2010; Fishman and Sweigart 2018). They often result from changes in epistatic nuclear loci appearing in independent populations, which could produce unfavorable genetic associations when reassociated by hybridization. The interacting loci may, however, locate in different cellular genetic compartments and, in fact, mitonuclear interactions have been identified as an important source in the establishment of reproductive barriers (Sloan et al. 2017; Havird et al. 2019; Postel and Touzet 2020). Mitochondria are prone to the accumulation of mutations and their typical uniparental mode of transmission makes the separation of favorable and unfavorable alleles impossible to allow natural selection to operate separately on them. Even if purifying selection seems strong to remove mitochondrial mutations, deleterious mutations do accumulate in mitochondrial genomes (Stewart et al. 2008; Popadin et al. 2013). Compensatory coevolution proposes that nuclear-encoded mitochondrial functions may either appear or get modified to compensate or nullify the effect of deleterious mitochondrial genetic changes, allowing their accumulation without negatively impacting mitochondrial activity (Sloan et al. 2018; Hill 2020). In plants, illegitimate recombination produces mitochondrial sequence reshuffling leading most mitochondrial intergenic sequences to be unalignable even between closely related individuals (Sloan, Müller, et al. 2012). In contrast, the impact on essential mitochondrial coding sequences is much less frequent but few cases of disruption of gene continuity, most often within introns, have been described (Knoop 2004). The $ccmF_{N2}$ gene that derives from the fragmentation of *ccmF* is a clear example of recent gene modification in the evolution of angiosperms (Rayapuram et al. 2008). A first fission event of *ccmF* traces back to the early evolution of land plants and split the gene into $ccmF_N$ and $ccmF_C$ which encode N- and C-terminal portions of the CcmF protein, respectively (Knoop 2004). The $ccmF_C$ gene was further split into two genes in *Marchantia* (Oda et al. 1992) whereas $ccmF_N$ was independently divided in $ccmF_{N1}$ and $ccmF_{N2}$ in *Brassicaceae* (Handa et al. 1996; Unseld et al. 1997), Allium (Kim et al. 2016), and *Fabaceae* (Choi et al. 2020). Interestingly, the split of $ccmF_{N1}$ and $ccmF_{N2}$ in *Brassicaceae* has resulted in separate and functionally independent loci whereas the two genes are still adjacent in *Fabaceae* and may represent an early stage of $ccmF_N$ fission (Choi et al. 2020). In the present analysis, we reveal that the Arabidopsis *RFL8* PPR protein is essential for the translation of the $ccmF_{N2}$ mitochondrial mRNA in Arabidopsis. The selection of *RFL8* has been obviously determinant to render the $ccmF_{N2}$ gene fragment functional and permit the production of the $CcmF_{N2}$ protein which upon association with $CcmF_{N1}$ is able to reconstitute the complete $CcmF_N$ portion of CcmF (Rayapuram et al. 2008). It is difficult to know if the selection of $ccmF_{N2}$ has resulted from any kind of selective advantage or if $ccmF_{N2}$ along with $ccmF_{N1}$ replaced $ccmF_N$ by simple genetic drift. Whatever the reasons that led to the maintenance of $ccmF_{N2}$ in mitochondria, *RFL8* remains a nuclear response aiming at promoting its expression and thus counteracting the deleterious effects associated with the truncation of $ccmF_N$. $ccmF_{N2}$ and *RFL8* are thus epistatic loci located in separate genetic compartments whose products interact in mitochondria and have become indispensible for c-type cytochrome biogenesis and respiration in *Brassicaceae* plants. $ccmF_{N2}$ and *RFL8* are thus an interesting example of nucleo-mitochondrial coadaptation that may have played role in the reproductive isolation of *Brassicaceae*. Interestingly, RFL8 belongs to the fast-evolving RFL gene family (Fujii et al. 2011). This subfamily forms a separate monophyletic group

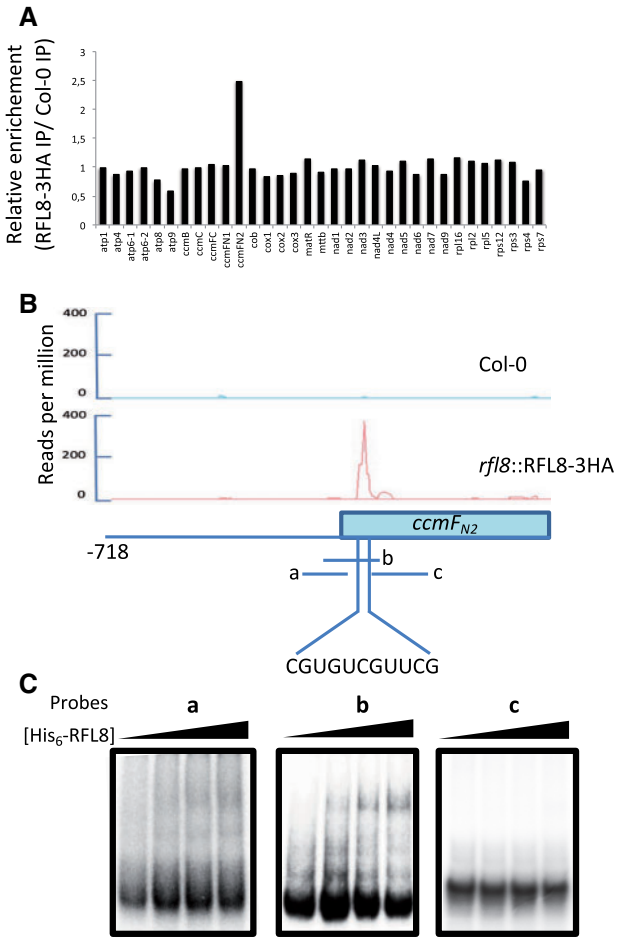

Fig. 5. The RFL8 protein associate with the *ccmF*$_{N2}$ mRNA in vivo. (*A*) RIP-Seq assay on mitochondrial extracts prepared from complemented *rfl8*::RFL8-3HA and wild-type plants. Coimmunoprecipitated RNA was used for cDNA synthesis and then analyzed by Illumina deep sequencing. Obtained reads were mapped to the Arabidopsis mitochondrial genome. The reported values are ratios of read counts per mitochondrial ORF between *rfl8*::RFL8-3HA and wild-type plants (RFL8-3HA/Col-0 ratios). They are means of two independent biological replicates performed on each genotype. Differential enrichment performed with edgeR identified *ccmF*$_{N2}$ with a false discovery rate control at 5%. (*B*) Similar RIP-Seq experiment as performed in (*A*) except that initial mitochondrial extracts were predigested with RNase-I prior to immunoprecipitation. Sequencing reads were mapped to the mitochondrial genome and screenshots from the Integrated Genome Viewer software showing read distributions along the *ccmF*$_{N2}$ coding sequence and 5′-UTR in wild type (Col-0) and complemented *rfl8*::RFL8HA plants are depicted. The diagram below the plots materializes the proposed position of the *ccmF*$_{N2}$ ORF (Rayapuram et al. 2008) and its 5′-UTR as well as that of the three RNA probes (a, b, and c) used in gel mobility shift assay used in panel (*C*). The location and sequence of the RFL8 binding site predicted by the PPR code are also indicated. (*C*) Gel mobility shift assays confirming the ability of RFL8 to associate with the RNA sequence predicted the PPR code. 0, 100, 200, or 400 nM of recombinant RFL8 were assayed in the shown lanes combined with gel-purified RNA probes (a, b, and c) indicated in panel (*B*). U, unbound probe; B, bound probe.

in the PPR family from which many *Rf* genes have evolved in various crop species (Dahan and Mireau 2013). Unlike most PPR genes for which interspecies orthologous relationships can be easily found (O'Toole et al. 2008), *RFL* genes are poorly conserved even at the species level, underlining diversifying constraints for these genes (Fujii et al. 2011). Most *RFL* genes were found associated with nonessential mitochondrial mRNA 5′ processing events (Jonietz et al. 2010; Hölzle et al. 2011; Jonietz et al. 2011; Arnal et al. 2014; Fujii et al. 2016) and the analysis of *RFL8* reveals that certain RFL genes fulfill essential functions. The election of *RFL8* as an essential factor for *ccmF*$_{N2}$ expression may have been fortuitous but the rapid sequence divergence of *RFL* genes has likely been advantageous to rapidly create a specific translational activator to render *ccmF*$_{N2}$ functional. *RFL8* belongs to a clade of eight RFLs for which potential gene orthologs in other *Brassicaceae* outside of *Arabidopsis thaliana* can be found (Fujii et al. 2011; Arnal et al. 2014). It will be interesting to investigate the function of other *RFL* genes from this clade to see whether they also fulfill essential function for mitochondria biogenesis in *Brassicaceae* plants.

C-Type Cytochrome Biogenesis Is Essential for Plant Embryo Development

C-type cytochromes are widespread heme-containing proteins involved in a variety of critical redox reactions (Welchen and Gonzalez 2016). CYT-c/c1 biogenesis is an intricate and complex process ensuring the covalent attachment of a protoporphyrin IX-Fe prosthetic group (heme b) to two conserved cysteines *via* thioester bonds in apocytochromes, conferring them stability (Travaglini-Allocatelli 2013; Babbitt et al. 2015). In plant mitochondria, CYT-c/c1 maturation (CCM) occurs in the intermembrane space *via* the most complex of the known CCM machineries, called CCM-System I, that is also found in proteobacteria, archea, and red algae (Verissimo and Daldal 2014). It is based on three highly integrated functional domains which are responsible for 1) transporting heme b to the IMS and loading it onto the heme chaperone CcmE, 2) maintaining the apocytochromes in a reduced state, and 3) the covalent ligation of apocytochromes with heme b. Although the presence of a single large maturase complex engaging all CCM modules has been postulated, BN–PAGE analysis on Arabidopsis extracts could separate two large membrane-bound CCM complexes of yet unclear compositions. The CcmF$_{N1}$, CcmF$_{N2}$, CcmF$_{C}$, and CCMH were shown to colocalize in a 500-kDa complex corresponding to the heme–lyase complex (Meyer et al. 2005; Rayapuram et al. 2008) and CCMA was found in a separate complex of about the same size likely corresponding to the heme-handling complex (Rayapuram et al. 2007). The analysis of the Arabidopsis *rfl8* mutant revealed that the heme–lyase complex is lacking in these plants, resulting in undetectable levels of CYTc and CYTc1 and mutant embryos to stop their development at the early cotyledon stage unless they are saved by tissue culture (fig. 1C). This embryonic-lethal phenotype is identical to that of a second mutant allele previously identified in *RFL8* (Yang et al. 2011) and that of an Arabidopsis mutant affected in the *CCMH* gene encoding another subunit of the heme–lyase complex (Meyer et al. 2005). They underline the essential character of c-type

cytochromes for mitochondrial activity and show that strong shortage in holocytochrome c and c1 content leads to premature arrest of plant embryo development. Analysis of small rescued *rfl8* plantlets revealed that such deficiency induces strong reductions in respiratory complex III and undetectable levels of complex IV (fig. 2*A* and supplementary fig. 2, Supplementary Material online) that are concordant with the structural role of CYTc1 in complex III (Ndi et al. 2018) and the electron shuttling activity of CYTc between complex III and complex IV (Welchen and Gonzalez 2016). The survival of *rfl8* miniature plants in vitro is certainly permitted by the strong induction of the AOX that can recover electrons from the ubiquinone pool to reduce O_2 into water instead of leaving them diffuse freely to make ROS in the absence of complex IV (fig. 2*B* and supplementary fig. 3, Supplementary Material online). Interestingly, we could produce similar plantlets with an Arabidopsis mutant impaired in the production of the complex-IV subunit Cox2 that is also strongly deprived in both complex III and complex IV (Dahan et al. 2014), further confirming the ability of the alternative respiratory pathway to maintain complex III- and IV-deficient miniature plants alive, provided that they are cultivated in vitro on a sucrose-rich medium. Such resulting physiological conditions appear to not only affect plant development but also mitochondrial gene expression as virtually all protein-encoding mRNAs strongly overaccumulate in *rfl8* plants compared with the wild type (supplementary figs. 4 and 5, Supplementary Material online). Mitochondrial transcript overaccumulation is often observed in plant mitochondrial mutants (Kwasniak et al. 2013; Hsieh et al. 2015; Haïli et al. 2016; Wang et al. 2018), but the increase does not reach the extent found in *rfl8* plants. The mitochondrial translatome in *rfl8* was also found profoundly perturbed, with many genes (including $ccmF_{N2}$) being less-efficiently translated as compared with the wild type (fig. 3). The cause of these translational changes is currently unclear, but it could result from a triple conjuncture that may involve overabundant mitochondrial transcripts, a limited number of available mitoribosomes, and a differential recruitment of mitoribosomes on mitochondrial mRNAs for translation initiation. In such disturbed molecular context, the lack of the heme–lyase complex and $CcmF_{N2}$ in *rfl8* plants and the identification of the RFL8 in vivo binding site within $ccmF_{N2}$ transcript strongly support that the decrease in $ccmF_{N2}$ translation is the direct and primary consequence of *RFL8* loss. The negative impact on the translation of other mitochondrial transcripts as well as *nad2* intron 1 splicing are most likely indirect effects of the *rfl8* mutation as RFL8 does not associate with any of these RNA species in vivo.

Translation Initiation of $ccmF_{N2}$ Occurs at a Non-AUG Codon and in a Region Corresponding to Coding Sequence in the Nonsplit $ccmF_N$ Gene

The translational start of $ccmF_{N2}$ has always been questionable as no in-frame AUG codon can be identified in the upstream region of this ORF (Rayapuram et al. 2008). Multiple sequence alignments revealed that $ccmF_{N2}$ ORF extends up to an in-frame GTG codon that has been proposed to maybe serve as a potential translation start codon, although no experimental data supported this hypothesis (fig. 4*C* and supplementary fig. S11, Supplementary Material online). In this analysis, we show that to activate $ccmF_{N2}$ translation, RFL8 binds to an 11-nucleotide sequence located 25 bases downstream of this codon and no in-frame AUG is found within 100 bases downstream of this site. However, this RNA region is subjected to translation according to our Ribo-seq analysis (fig. 3), thereby suggesting that $ccmF_{N2}$ translation initiates at a non-AUG codon that needs to be identified. A multiple sequence alignment involving $CcmF_{N2}$ and $CcmF_N$ from various plant species shows that the recognized W-rich heme-interacting motif in these proteins begins 67 amino acids downstream of the GTG codon (supplementary fig. 16, Supplementary Material online), leaving a stretch of about 200 nucleotides of RNA sequence for translation initiation to produce a potentially active $CcmF_{N2}$ protein. Interestingly, this region is a coding sequence in $ccmF_N$ and our findings suggest that part of it has obviously become the 5′-UTR of $ccmF_{N2}$. How this RNA region of $ccmF_{N2}$ and the binding of RFL8 to it have allowed the creation of a translation start site for $ccmF_{N2}$ is currently unclear. This segment of $ccmF_{N2}$ mRNA may contain cis-elements permitting the recruitment of the mitochondrial translation initiation machinery. In particular, the two triple AAA motifs that are found a few bases downstream of the RFL8 binding sites (supplementary fig. 15, Supplementary Material online) may serve as mitoribosome recruitment sequences as recently proposed (Waltz et al. 2020). However, their perfect conservation in $ccmF_N$ coding sequences makes this hypothesis rather unlikely as it would imply that translation initiation occurs also internally within $ccmF_N$ coding sequence. It seems more likely that the selection of the RFL8 PPR protein as a trans-factor being able to bind within this RNA region has been the determining parameter to permit the creation of a translation start site for $ccmF_{N2}$ ORF. PPR proteins are effectively known to play essential roles in organellar translation, notably in plastids (Barkan and Small 2014). In plant mitochondria, besides RFL8, a single other PPR protein (MTL1) was shown to play a direct role in mitochondrial translation initiation (Haïli et al. 2016). The predicted MTL1 binding site is located 35 bases upstream of the AUG codon of the mRNA whose translation it facilitates, suggesting that the translation start site in $ccmF_{N2}$ should occur within a similar distance downstream of the RFL8 binding site. Indeed, the distribution of ribosome footprints revealed two ribosome peaks in this region of $ccmF_{N2}$, the first of which being RFL8-dependant (fig. 3*B* and supplementary fig. 15, Supplementary Material online). This first peak overlaps with the RFL8 binding site and may correspond to the footprint of an RFL8-containing complex, whereas the second one, located just upstream of the W-rich motif coding region, may represent the translation initiation site in $ccmF_{N2}$. The way by which translation initiation occurs in plant mitochondria and the role that translational PPR plays in this process remain unclear. In particular, the emblematic role of the plastid PPR10 whose binding would prevent the formation of a stem-loop structure in the 5′ leader of

its target mRNA *atpH* to permit the recruitment of ribosomes through the liberation of a Shine and Dalgarno (S/D) motif (Prikryl et al. 2011) cannot apply to plant mitochondria as no such ribosome binding sequence is found upstream of plant mitochondrial mRNAs and no RNA secondary structure can be predicted in the 5′ region of $ccmF_{N2}$. The recently determined high-resolution structure of the plant mitoribosome shows the incorporation of many plant-specific subunits, including ten ribosomal PPR (rPPR) proteins (Waltz et al. 2019). The roles that rPPRs may play in ribosome functioning are still elusive. However, one of them, rPPR10, is found near the small ribosomal subunit mRNA exit channel and seems well placed to play a role in mRNA recruitment, maybe through an association with A-rich motifs (Waltz et al. 2020). PPR proteins like RFL8 or MTL1 may collaborate with rPPR10 in translation initiation or act independently by attracting mitoribosomes near mRNA translation start sites as mRNA-specific translation activators act in yeast mitochondria (Desai et al. 2017). Plant PPR proteins facilitating mitochondrial translation could also act as simple molecular barriers whose binding could prevent the passage of mitoribosomes initiating translation upstream of their binding site and thereby defining translatable regions in mitochondrial mRNAs which could correspond to PPR-free mRNA regions. Further biochemical analysis of PPR proteins such as RFL8 or MTL1 should be instrumental to elucidate how translation initiation is molecularly orchestrated in plant mitochondria.

Materials and Methods

Plant Material and Growth

Arabidopsis (*Arabidopsis thaliana*) Col-0 plants were obtained from the INRAE Arabidopsis stock center in Versailles (http://dbsgap.versailles.inra.fr/portail/). The Arabidopsis *rfl8* (N515489) mutant was obtained from the SALK mutant collection (Alonso et al. 2003). Plants were grown in a greenhouse in long-day conditions for 10–12 weeks before use. For complementation tests, the Arabidopsis *RFL8* promoter and coding sequence with no stop codon were coamplified by PCR using the GWRFL8-F and GWRFL8-R primers. The obtained amplification production was cloned by BP Gateway (Invitrogen) reaction into pDONR207 (Invitrogen) and subsequently transferred by LR Gateway reaction into pGWB13 (Nakagawa et al. 2007), creating an C-terminal fusion with the 3HA protein tag. The *B. rapa RFL8* gene construct was obtained by overlapping PCR amplification. To this end, the *B. rapa RFL8* coding sequence was amplified by PCR with the RFL8Rapa-GW3 and RFL8Rapa-GW2 primers and the Arabidopsis *RFL8* promoter region with the RFL8Rapa-GW5 and RFL8Rapa-GW4 primers. The obtained amplification products overlapped on 40 bases and carried half of an *att*B1 site upstream of the *RFL8* promoter region and half of an *att*B2 site downstream of the *B. rapa RFL8* coding sequence. About 10 ng of each PCR product were mixed and subjected to a third PCR amplification with the GW3 and GW5 primers to fuse them in a single DNA fragment bearing complete *att*B sites. After gel purification, the generated amplification product was cloned by BP Gateway cloning into pDONR207 (Invitrogen) and then by LR Gateway reaction into pGWB1 (Nakagawa et al. 2007). Once transferred into *Agrobacterium tumefaciens*, both constructs were used to transform Arabidopsis *rfl8* heterozygous plants by floral dip (Clough and Bent 1998).

RNA Analysis

Quantitative RT–PCR analysis measuring the steady-state levels of mitochondria-encoded pre-mRNAs, mRNAs, and intron splicing efficiencies were done as previously described in Haïli et al. (2016). Ribo-Seq analysis was developed as previously described in Planchard et al. (2018). For RIP-Seq analysis, mitochondrial pellets representing ≈ 4 mg of total mitochondrial proteins were prewashed two times in 500 µl of 12.5 mM HEPES-KOH pH 7.5 and 0.3 M sucrose and then lysed in 200 µl of 30 mM HEPES-KOH pH 7.5, 100 mM KCl, 1 mM EDTA, 5 mM $MgCl_2$, 0.5% (w/v) CHAPS, 1 mM AEBSF, 1 mM DTT, Complete Mini Protease inhibitor cocktail (Roche), and 1 U/µl Riboblock (Fermentas) for 30 min at 4 °C with occasional vortexing. The mitochondrial lysate was cleared by centrifuging at 13,000 × g for 15 min at 4 °C. Approximately 2 mg of recovered mitochondrial lysate were used for each coimmunoprecipitation repeat. For RIP-seq with RNase I treatment, no Ribolock was added in the extraction buffer and the mitochondrial extracts were pretreated with 1 U of RNase I (Ambion) per µl of extract at 25 °C for 10 min and then placed on ice. Then, 150 µl of Dynabeads protein G (Invitrogen) per CoIP were equilibrated three times in a same volume of CoIP buffer (50 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 5 mM $MgCl_2$, 0.5% [v/v] Nonidet P-40, 1 mM AEBSF, 1 mM DTT, 1 U/µl Riboblock [Fermentas] and Complete Mini Protease inhibitor cocktail [Roche]) at 4 °C and then resuspended in 150 µl of CoIP buffer. The mitochondrial lysates were precleared with 50 µl of equilibrated Dynabeads protein G by rotating for 5 min at 4 °C and lysates were collected by pelleting the beads with a magnet. About 10% of the precleared lysates were kept as input fraction. Then, 10 µg of ChIP grade anti-HA antibody (Abcam) were added to the lysates and incubated for 3 h at 4 °C with gentle rotation. The protein extracts were then transferred onto 100 µl of equilibrated protein G beads and further incubated for 60 min at 4 °C with gentle rotation. The protein G beads were collected using a magnet and washed three times with 500 µl of CoIP buffer. About 10% of the washed beads and 10% of the supernatant were kept as CoIP and unbound fractions, respectively, for immunoblot analysis. Next, 500 µl of RNA elution buffer (20 mM Tris–HCl pH 7.5, 10 mM EDTA, 200 mM NaCl, and 0.2% [w/v] SDS) supplemented with 0.2 µg/µl of Proteinase K was added to the protein G beads and incubated for 30 min at 37 °C. RNAs were then purified by TRI-reagent extraction and ethanol precipitated with 40 µg of glycogen. Purified RNAs were resuspended in 14 µl of water and quantified using Qubit microRNA assay (Thermo Fisher Scientific). About 10 ng of RNAs were used to prepare the sequencing libraries. For samples not treated with RNase I, purified RNAs were first sheared by incubation at 95 °C for 2 min with NEBNext Magnesium RNA fragmentation kit (New England Biolabs) according to manufacturer's

instructions. The RNAs were then ethanol precipitated along with 40 μg of glycogen and phosphorylated with 1 µl of T4 polynucleotide kinase (New England Biolabs) in 20 µl for 60 min at 37 °C. A final purification of RNAs was performed with the RNeasy Plant Mini kit (Qiagen) by following the procedure described for RNA clean-up. The sequencing libraries were prepared with the TruSeq Small RNA Library preparation kit (Illumina) with some modifications. In brief, all adapters for ligation and the STP stop solution were diluted four times before use and the obtained libraries were PCR amplified with 25 cycles. Libraries were gel purified to enrich for inserts between 15 and 100 nucleotides and then sequenced on a NextSeq 500 sequencer (Illumina) with a read length of 75 nucleotides (single end). Sequencing data were processed as described in Planchard et al. (2018) except that reads of all lengths were aligned to the mitochondrial genome. Enrichment values were calculated as the normalized abundance of reads per gene (RPKM) compared with a nontransgenic control sample (Col-0). Mapped read distributions were visualized with the IGV software. Differential enrichment analysis of RIP-Seq data was done using the differential analysis procedure detailed in Rigaill et al. (2016). Briefly, genes with less than 1 read after a count per million (CPM) normalization in at least one half of the samples were discarded. Library size was normalized using the trimmed mean of *M* value (TMM) method and count distribution was modeled with a negative binomial generalized linear model with a single genotype and replicate effects. Dispersion was estimated by the edgeR method (version 3.28.0; McCarthy et al. 2012) in the statistical software "R" (version 3.6.1). Enrichment differences compared two genotypes using likelihood ratio test and *P* values were adjusted by the Benjamini–Hochberg procedure to control the false discovery rate. A gene was declared differentially enriched if its adjusted *P* value <0.05.

Protein Extraction and Analysis

Mitochondria were purified as previously described in Haïli et al. (2013). Crude mitochondrial extracts were prepared from 800 mg of in vitro-grown seedlings that were ground in a mortar with sand and 8 ml of extraction buffer (75 mM MOPS-KOH pH 7.6, 0.6 M sucrose, 4 mM EDTA, 0.2% [w/v] polyvinylpyrrolidone-40, 0.2% [w/v] BSA, and 8 mM cysteine). Cell debris was removed by filtering through two layers of Miracloth (Calbiochem) and the homogenate was centrifuged at 1,300 × g for 5 min at 4 °C. The supernatant was centrifuged at 25,000 × g for 20 min at 4 °C and the obtained pellet resuspended in 1 ml of buffer I (10 mM MOPS-KOH pH 7.2, 0.3 M sucrose) and then aliquoted to 500 µl in Eppendorf tube. Crude mitochondria were recovered by centrifugation at 13,000 × g for 15 min at 4 °C and lysed in 400 µl of buffer II (10 mM BisTris–HCl pH 7.5, 0.3 M sucrose 0.5 M 6-aminohexanoic acid, 1 mM EDTA, and 1% DDM) for blue native gel analysis or in 200 µl of 0.5× RIPA buffer (25 mM Tris–HCl pH 7.5, 75 mM NaCl, 0.5 mM EDTA, 0.5% [v/v] Triton X-100, 0.05% [v/v] SDS, 0.25% sodium deoxycholate, 1 mM DTT, AEBSF 1 mM, Complete Mini Protease inhibitor cocktail [Roche]) for SDS–PAGE gel analysis. For blue native gel, 100 μg of crude mitochondrial proteins were separated on 4–16% (w/v) polyacrylamide Native PAGE gels (Invitrogen) and then either transferred to PVDF membranes or subjected to in-gel activity staining as previously described in Dahan et al. (2014). For immunoblots, proteins were separated on 4–20% (w/v) polyacrylamide SDS–PAGE gels (Biorad) and transferred to PVDF membranes prior to incubation with antibodies.

Gel Mobility Shift Assay

The *RFL8* coding sequence lacking the region encoding the mitochondrial presequence was amplified by PCR with the GWRFL8-8 and RFL8-Cpl-B2 primer (supplementary table 1, Supplementary Material online). The obtained PCR product was cloned into pDONR207 by Gateway BP reaction (Invitrogen) and subsequently subcloned into pDEST17 by Gateway LR reaction (Invitrogen). The resulting 6xHis-RFL8 protein was expressed in Rosetta (DE3) *Escherichia coli* cells for 3 h at 20 °C using 0.2 mM of IPTG. Bacterial pellets were lysed in 50 mM HEPES-KOH (pH7.5) and 150 mM NaCl with the One Shot cell disruption system (Constant Systems). The 6xHis-RFL8 protein revealed to be highly insoluble and was solubilized from *Escherichia coli* inclusion bodies in 50 mM HEPES-KOH (pH7.5), 150 mM NaCl, 2% N-lauryl sarcosine, and 10 mM β-mercaptoethanol by rotation overnight at 4 °C. The obtained protein solution was then enriched to high purity using a Ni-Sepharose column (GE Healthcare). Fractions containing highly pure 6xHis-RFL8 protein were identified by SDS–PAGE gel electrophoresis. The purified protein solution was then dialyzed overnight against the same buffer in which the N-lauryl sarcosine concentration was changed to 0.01%. Before performing gel shift experiment, the concentration of protein was determined by Bradford protein assay (Bio-Rad). Radiolabeled RNA probes were generated by in vitro transcription with T7 RNA polymerase (Promega) in the presence of 30 μCi of [α-32P] rUTP according to manufacturer's instructions. Probes were purified on 5% denaturing urea polyacrylamide gel, eluted in 0.5 M ammonium acetate, 0.1% sodium dodecyl sulfate, and 1 mM EDTA, and then purified with TRI-reagent before precipitation with isopropanol. The RNA probes were quantified on a NanoDrop apparatus. Before gel mobility shift assay, 100 pM of purified probes were heated to 95 °C for 1 min and snap-cooled on ice. The RNA-binding reactions were performed in 50 mM HEPES-KOH (pH 7.5), 100 mM NaCl, 0.1 mg/ml BSA, 10% glycerol, 0.5 mg/ml heparin, 4 mM DTT, 10 U Riboblock (Fermentas), 100 pM radiolabeled RNA, and protein concentrations indicated in figure legends. The reactions were carried out 30 min at room temperature, and then samples were loaded on a 5% polyacrylamide (29:1 acrylamide/bis-acrylamide) gel in 1xTHE (66 mM HEPES, 34 mM Tris [pH 7.7], 0.1 mM EDTA). The gels were run at 140 V in 30 min, dried, and then exposed to a phosphorimager screen (FLA-9500 Fujifilm).

Respiration Measurements

Oxygen consumptions were measured with a liquid phase Oxytherm oxygen electrode system (Hansatech) as previously described in Dahan et al. (2014).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

This work was supported by ANR MITRA (ANR-16-CE11-0024-01) to H.M. The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work has benefited from the support of IJPB's Plant Observatory technological platforms.

Author Contributions

T.T.N., N.P., and J.D. performed the majority of the experiments. M.Q., C.G., and A.B. provided technical assistance. H.M. designed the study. H.M. and T.T.N. analyzed the data with contribution of all authors. P.B and O.N analyzed the RIP-Seq data and performed bioinformatics analysis. H.M. and T.T.N. wrote the paper.

Data Availability

The data underlying this article (Ribo-Seq and RIP-Seq data) have been submitted to NCBI Gene Expression Omnibus (GEO) under accession number GSE166459 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166459). All steps of the experiment, from growth conditions to bioinformatic analyses, were managed in CATdb database (Gagnot et al. 2008; http://tools.ips2.u-psud.fr/CATdb/) with ProjectID Blanc09_RIPseq_RFL8.

References

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, et al. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* 177(2):1173–1192.

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301(5633):653–657.

Arnal N, Quadrado M, Simon M, 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(1):134–145.

Arrieta-Montiel MP, Shedge V, Davila J, Christensen AC, Mackenzie SA. 2009. Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics* 183(4):1261–1268.

Babbitt SE, Sutherland MC, Francisco BS, Mendez DL, Kranz RG. 2015. Mitochondrial cytochromec biogenesis: no longer an enigma. *Trends Biochem Sci*. 40(8):446–455.

Barkan A, Rojas M, Fujii S, Yap A, Chong YS, Bond CS, Small I. 2012. A combinatorial amino Acid code for RNA recognition by pentatricopeptide repeat proteins. *PLoS Genet*. 8(8):e1002910.

Barkan A, Small I. 2014. Pentatricopeptide repeat proteins in plants. *Annu Rev Plant Biol*. 65:415–442.

Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424(6945):197–201.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res*. 27(8):1767–1780.

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet*. 19(12):709–716.

Castellana S, Vicario S, Saccone C. 2011. Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein–coding genes. *Genome Biol Evol*. 3:1067–1079.

Chen L, Liu Y-G. 2014. Male sterility and fertility restoration in crops. *Annu Rev Plant Biol*. 65:579–606.

Choi I-S, Ruhlman TA, Jansen RK. 2020. Comparative mitogenome analysis of the genus *Trifolium* reveals independent gene fission of ccmFn and intracellular gene transfers in Fabaceae. *Int J Mol Sci*. 21(6):1959.

Clough SJ, Bent AF. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J*. 16(6):735–743.

Cole LW, Guo W, Mower JP, Palmer JD. 2018. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Mol Biol Evol*. 29:380.

Dahan J, Mireau H. 2013. The Rf and Rf-like PPR in higher plants, a fast-evolving subclass of PPR genes. *RNA Biol*. 10(9):1469–1476.

Dahan J, Tcherkez G, Macherel D, Benamar A, Belcram K, Quadrado M, Arnal N, 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(4):1788–1802.

Darracq A, Varré JS, Marechal-Drouard L, Courseaux A, Castric V, Saumitou-Laprade P, Oztas S, Lenoble P, Vacherie B, Barbe V, et al. 2011. Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. *Genome Biol Evol*. 3:723–736.

Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu Y-Z, Weigel D, Mackenzie SA. 2011. Double-strand break repair processes driveevolution of the mitochondrial genome in *Arabidopsis*. *BMC Biol*. 9:64.

Desai N, Brown A, Amunts A, Ramakrishnan V. 2017. The structure of the yeast mitochondrial ribosome. *Science* 355(6324):528–531.

Drouin G, Daoud H, Xia J. 2008. Molecular phylogenetics and evolution. *Mol Phylogenet Evol*. 49(3):827–141.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 30(1):207–210.

Fishman L, Sweigart AL. 2018. When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu Rev Plant Biol*. 69:707–731.

Fujii S, Bond CS, Small ID. 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(4):1723–1728.

Fujii S, Suzuki T, Giegé P, Higashiyama T, Koizuka N, Shikanai T. 2016. The restorer-of-fertility-like 2 pentatricopeptide repeat protein and RNase P are required for the processing of mitochondrial orf291RNA in *Arabidopsis*. *Plant J*. 86(6):504–513.

Gagnot S, Tamby JP, Martin-Magniette ML, Bitton F, Taconnat L, Balzergue S, Aubourg S, Renou JP, Lecharny A, Brunaud V. 2008. CATdb: a public access to Arabidopsis transcriptome data from the URGV-CATMA platform. *Nucleic Acids Res*. 36(Database issue):D986–D990.

Geddy R, Brown GG. 2007. Genes encoding pentatricopeptide repeat (PPR) proteins are not conserved in location in plant genomes and may be subject to diversifying selection. *BMC Genomics* 8(1):130.

Gualberto JM, Newton KJ. 2017. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol*. 68:225–252.

Haïli N, Arnal N, Quadrado M, Amiar S, Tcherkez G, Dahan J, Briozzo P, Colas des Francs-Small C, Vrielynck N, Mireau H. 2013. The pentatricopeptide repeat MTSF1 protein stabilizes the nad4 mRNA in Arabidopsis mitochondria. *Nucleic Acids Res*. 41(13):6650–6663.

Haïli N, Planchard N, Arnal N, Quadrado M, Vrielynck N, Dahan J, des Francs-Small CC, 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(1):354–366.

Handa H, Bonnard G, Grienenberger JM. 1996. The rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 is divided into two independently transcribed reading frames. *Mol Gen Genet*. 252(3):292–302.

Havird JC, Forsythe ES, Williams AM, Werren JH, Dowling DK, Sloan DB. 2019. Selfish mitonuclear conflict. *Curr Biol*. 29(11):R496–R511.

Hill GE. 2020. Mitonuclear compensatory coevolution. *Trends Genet*. 36(6):403–414.

Hölzle A, Jonietz C, Törjek O, Altmann T, Binder S, 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(5):737–744.

Hsieh W-Y, Liao J-C, Chang C-Y, Harrison T, Boucher C, Hsieh M-H. 2015. The slow growth3 pentatricopeptide repeat protein is required for the splicing of mitochondrial NADH dehydrogenase subunit7intron 2 in *Arabidopsis*. *Plant Physiol*. 168(2):490–501.

Jonietz C, Forner J, Hildebrandt T, Binder S. 2011. RNA processing factor3 is crucial for the accumulation of mature ccmC transcripts in mitochondria of *Arabidopsis* accession Columbia. *Plant Physiol*. 157(3):1430–1439.

Jonietz C, Forner J, Hölzle A, Thuss S, 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(2):443–453.

Kim B, Kim K, Yang T-J, 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. *Curr Genet*. 62(4):873–885.

Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr Genet*. 46(3):123–144.

Kubo T, Newton KJ. 2008. Angiosperm mitochondrial genomes and mutations. *Mitochondrion* 8(1):5–14.

Kwasniak M, Majewski P, Skibior R, Adamowicz A, Czarna M, Sliwinska E, Janska H. 2013. Silencing of the nuclear RPS10 gene encoding mitochondrial ribosomal protein alters translation in Arabidopsis mitochondria. *Plant Cell* 25(5):1855–1867.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 40(10):4288–4297.

Meyer EH, Giegé P, Gelhaye E, Rayapuram N, Ahuja U, Thöny-Meyer L, Grienenberger JM, Bonnard G. 2005. AtCCMH, an essential component of the c-type cytochrome maturation pathway in Arabidopsis mitochondria, interacts with apocytochrome c. *Proc Natl Acad Sci U S A*. 102(44):16113–16118.

Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel J, Greilhuber J, Dolezel J, Leitch IJ, editors. Plant genome diversity. Vol. 1. Vienna: Springer. p. 123–144.

Nakagawa T, Kurose T, Hino T, Tanaka K, Kawamukai M, Niwa Y, Toyooka K, Matsuoka K, Jinbo T, Kimura T. 2007. Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J Biosch Bioeng*. 104(1):34–41.

Ndi M, Marin-Buera L, Salvatori R, Singh AP, Ott M. 2018. Biogenesis of the bc. *J Mol Biol*. 430(21):3892–3905.

Neupert W. 2016. Mitochondrial gene expression: a playground of evolutionary tinkering. *Annu Rev Biochem*. 85:65–76.

O'Toole N, Hattori M, Andrés C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. 2008. On the expansion of the pentatricopeptide repeat gene family in plants. *Mol Biol Evol*. 25(6):1120–1128.

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T. 1992. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol*. 223(1):1–7.

Osada N, Akashi H. 2012. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol*. 29(1):337–346.

Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol*. 28(1–2):87–97.

Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF. 2013. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet*. 22(2):384–390.

Planchard N, Bertin P, Quadrado M, Dargel-Graffin C, Hatin I, Namy O, Mireau H. 2018. The translational landscape of Arabidopsis mitochondria. *Nucleic Acids Res*. 283:1476.

Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE. 2013. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol Biol Evol*. 30(2):347–355.

Postel Z, Touzet P. 2020. Cytonuclear genetic incompatibilities in plant speciation. *Plants* 9(4):487.

Presgraves DC. 2010. The molecular evolutionary basis of species formation. *Nat Rev Genet*. 11(3):175–180.

Prikryl J, Rojas M, Schuster G, Barkan A. 2011. Mechanism of RNA stabilization and translational activation by a pentatricopeptide repeat protein. *Proc Natl Acad Sci U S A*. 108(1):415–420.

Rayapuram N, Hagenmuller J, Grienenberger JM, Bonnard G, 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(37):25200–25208.

Rayapuram N, Hagenmuller J, Grienenberger JM, Giegé P, Bonnard G. 2007. AtCCMA interacts with AtCcmB to form a novel mitochondrial ABC transporter involved in cytochrome c maturation in *Arabidopsis*. *J Biol Chem*. 282(29):21015–21023.

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol*. 11:29.

Rigaill G, Balzergue S, Brunaud V, Blondet E, Rau A, Rogier O, Caius J, Maugis-Rabusseau C, Soubigou-Taconnat L, Aubourg S. 2016. Synthetic data sets for the identification of key ingredients for RNA-seq differential analysis. *Brief Bioinform*. 19:65–76.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. 2017. The origin and diversification of mitochondria. *Curr Biol*. 27(21):R1177–R1192.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLoS Biol*. 10(1):e1001241.

Sloan DB, Havird JC, Sharbrough J. 2017. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol*. 26(8):2212–2236.

Sloan DB, Müller K, McCauley DE, Taylor DR, Štorchová H. 2012. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol*. 196(4):1228–1239.

Sloan DB, Warren JM, Williams AM, Wu Z, Abdel-Ghany SE, Chicco AJ, Havird JC. 2018. Cytonuclear integration and co-evolution. *Nat Rev Genet*. 19(10):635–648.

Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, Larsson N-G. 2008. Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS Biol*. 6(1):e10.

Travaglini-Allocatelli C. 2013. Protein machineries involved in the attachment of heme to cytochrome c: protein structures and molecular mechanisms. *Scientifica* 2013:505714–505717.

Unseld M, Marienfeld JR, Brandt P, Brennicke A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet*. 15(1):57–61.

Verissimo AF, Daldal F. 2014. Cytochrome c biogenesis system I: an intricate process catalyzed by a maturase supercomplex? *Biochim Biophys Acta*. 1837(7):989–998.

Waltz F, Nguyen T-T, Arrivé M, Bochler A, Chicher J, Hammann P, Kuhn L, Quadrado M, Mireau H, Hashem Y, et al. 2019. Small is big in Arabidopsis mitochondrial ribosome. *Nat Plants*. 5(1):106–117.

Waltz F, Soufari H, Bochler A, Giegé P, Hashem Y. 2020. Cryo-EM structure of the RNA-rich plant mitochondrial ribosome. *Nat Plants*. 6(4):377–383.

Wang C, Aubé F, Quadrado M, Dargel-Graffin C, Mireau H. 2018. Three new pentatricopeptide repeat proteins facilitate the splicing of mitochondrial transcripts and complex I biogenesis in *Arabidopsis*. *J Exp Bot*. 69(21):5131–5140.

Welchen E, Gonzalez DH. 2016. Cytochrome c, a hub linking energy, redox, stress and signaling pathways in mitochondria and other cell compartments. *Physiol Plant*. 157(3):310–321.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84(24):9054–9058.

Yang L, Peng X, Sun M-X. 2011. AtNG1 encodes a protein that is required for seed germination. *Plant Sci*. 181(4):457–464.

Yurina NP, Odintsova MS. 2016. Mitochondrial genome structure of photosynthetic eukaryotes. *Biochemistry* 81(2):101–113.